

## Computational $^{59}\text{Co}$ NMR Spectroscopy: Beyond Static Molecules<sup>†</sup>

Sonja Grigoleit and Michael Bühl\*

Max-Planck Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1,  
D-45470 Mülheim an der Ruhr, Germany

Received October 12, 2004

**Abstract:** GIAO–B3LYP computations of  $^{59}\text{Co}$  NMR chemical shifts are reported for  $\text{CoH}(\text{CO})_4$ ,  $\text{Co}(\text{CO})_4^-$ ,  $\text{CoCp}(\text{C}_2\text{H}_4)_2$ ,  $\text{Co}(\text{CN})_6^{3-}$ ,  $\text{Co}(\text{NH}_3)_3(\text{CN})_3$ ,  $\text{Co}(\text{NH}_3)_6^{3+}$ ,  $\text{Co}(\text{NH}_3)_4(\text{CO}_3)^+$ ,  $\text{Co}(\text{acac})_3$ , and  $\text{Co}(\text{H}_2\text{O})_6^{3+}$ , employing both static calculations for equilibrium geometries as well as methods which include zero-point and classical thermal effects. The zero-point effects were computed by applying a perturbational approach, and the classical thermal effects were evaluated using Car-Parrinello molecular dynamics simulations. Both methods lead to a downfield shift of  $\delta(^{59}\text{Co})$  with respect to the equilibrium values, which can be attributed to a large extent to cobalt–ligand bond elongation. In some cases the zero-point and classical thermal corrections improve the agreement between computed and experimental values, but especially for complexes where the experimental NMR data were obtained in aqueous solution, the error increases somewhat. Mean absolute deviations between averaged and experimental  $\delta(^{59}\text{Co})$  values are on the order of 500–760 ppm over a chemical shift range of almost 20 000 ppm. The computed structures and properties of three  $\text{Co}_2(\text{CO})_8$  tautomers reproduce the experimental data very well. Two transition states for interconversion of these tautomers were located: low barriers are obtained, consistent with the observed fluxionality on the NMR time scale. Two model cobaloximes were taken as test cases to study the change of  $\delta(^{59}\text{Co})$  upon deuteration three bonds away from the metal. The sizable downfield shift of  $\delta(^{59}\text{Co})$  observed on going from H to D is attributed to a changed vibrational wave function, which causes a noticeable cobalt–ligand bond elongation.

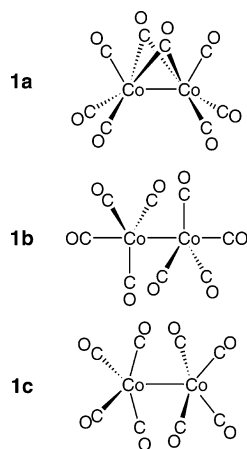
### Introduction

The  $^{59}\text{Co}$  isotope had been prominent since the dawn of NMR spectroscopy, because it was one of the first nuclei for which the phenomenon of the chemical shift was observed.<sup>1</sup> To this day, the large chemical shift of this nucleus, on the order of 20 000 ppm,<sup>2–4</sup> together with rather favorable NMR properties, make  $^{59}\text{Co}$  NMR spectroscopy a highly sensitive probe for the electronic structure, geometrical parameters, and reactivities of cobalt complexes.<sup>5</sup> In the surge of the blossoming life sciences, interest in  $^{59}\text{Co}$  NMR of Co-containing biomolecules or model complexes thereof has

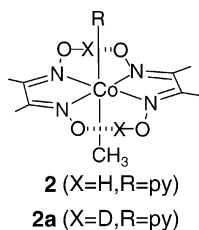
recently been renewed.<sup>6–9</sup> Since the mid-1990s, the tools of density functional theory (DFT) have successfully been employed to calculate transition-metal chemical shifts,<sup>10–13</sup> and the  $^{59}\text{Co}$  nucleus has been an early target for such computations.<sup>14–17</sup> In these studies, which confirmed the suitability of the B3LYP hybrid functional for chemical shift calculations of transition metals,<sup>18</sup> static computations were performed for optimized or experimental geometries. Current developments are directed to go beyond such a static picture in order to account for the dynamic nature of matter, thereby striving for an increase in the accuracy of theoretical NMR parameters. Methods that have been successfully applied to chemical shift calculations of lighter nuclei comprise solutions of the nuclear Schrödinger equations,<sup>19</sup> Quantum Monte Carlo Calculations,<sup>20</sup> perturbational zero-point corrections,<sup>21–24</sup> and molecular dynamics (MD) simulations.<sup>25,26</sup> The latter

\* Corresponding author fax: (+49)208-306-2996; e-mail: buehl@mpi-muelheim.mpg.de.

<sup>†</sup> Dedicated to Prof. R. Ahlrichs on the occasion of his 65th birthday.



**Figure 1.** Tautomeric binuclear cobalt carbonyls.



**Figure 2.** Cobaloximes of this study.

approach has also been used to evaluate solvent effects on transition-metal chemical shifts.<sup>27–29</sup>

In two recent studies<sup>30,31</sup> we have assessed the performance of the latter two methods, zero-point corrections and MD simulation, for a large number of Ti, V, Mn, and Fe complexes covering a total chemical shift range of ca. 4000 ppm. According to these systematic investigations, inclusion of zero-point and classical thermal effects does not lead to generally improved results but can afford more accurate  $\delta$ -(M) values in some cases. We now extend this systematic investigation to  $\delta(^{59}\text{Co})$  in order to test if the same conclusions are valid for this nucleus with its very large chemical shift range. For this purpose we have chosen a representative set of inorganic and organometallic cobalt complexes, which span almost 20 000 ppm in  $\delta(^{59}\text{Co})$ , thereby increasing the range of chemical shifts covered so far by almost an order of magnitude. In addition to this systematic performance test, we report specific structural applications of  $^{59}\text{Co}$  chemical shift calculations. The first of these applications concerns the dinuclear complex  $\text{Co}_2(\text{CO})_8$ , a textbook example of a fluxional carbonyl complex, which exists in the form of at least two rapidly interconverting isomers in solution (see Figure 1). NMR spectra are thus observed as a dynamic average of the equilibrium mixture. If the isomers of this mixture would differ considerably in their  $\delta(^{59}\text{Co})$  values, sufficiently accurate computations of the latter could afford information on the relative isomer population.

The second application is related to an intriguing isotope effect in cobaloximes (see Figure 2). Asaro et al. have investigated  $^{59}\text{Co}$  chemical shifts in cobaloximes [ $\text{MeCo}(\text{Hdmg})_2\text{py}$ ] (Hdmg = dimethylglyoximate, py = pyridine)<sup>7</sup> and found that the substitution of the two bridging hydrogen atoms by deuterium leads to a remarkably large change in

$\delta(^{59}\text{Co})$  ( $\Delta\delta \approx 50$  ppm),<sup>32</sup> even though this substitution takes place three bonds away from the metal! This finding was rationalized in terms of a secondary geometrical effect, in which the fundamental vibrational wave function is modified, such that the mean  $r(\text{O}\cdots\text{D})$  distance becomes shorter than  $r(\text{O}\cdots\text{H})$ , which, in turn, affects the  $r(\text{O}\cdots\text{O})$  distance and, ultimately the  $r(\text{Co}\cdots\text{N})$  bond length. For a deeper analysis of experimental results, we computed the vibrational averaged geometries and NMR chemical shifts of cobaloximes using the above-mentioned perturbational approach. Analysis of respective anharmonic contributions to the zero-point corrected geometries affords a deeper understanding of the geometrical effects on the  $^{59}\text{Co}$  chemical shifts of cobaloximes. This case study of remote isotope effects is a stringent test for the perturbational zero-point corrections, a test which they are shown to pass.

## Computational Methods

Geometries have been fully optimized employing the gradient-corrected exchange–correlation functionals of Becke<sup>33</sup> and Perdew,<sup>34,35</sup> denoted BP86, together with a fine integration grid (75 radial shells with 302 angular points per shell). For the optimization we employed basis AE1, that is, Wachters’ all-electron basis augmented with one additional diffuse d and two p functions with the contraction scheme (14s11p6d)/[8s7p4d] for cobalt<sup>36,37</sup> and 6-31G\* basis for all other elements. For the cobaloxime complexes we employed additionally basis AE1(\*), that is, the same basis as AE1 and an additional p-polarization function on the two bridging hydrogen atoms. As shown before<sup>31</sup> the 6-31G\* basis on the ligands is sufficient to compute molecular properties which depend on the curvature of the potential energy surface (PES). All structures were characterized as minima on the PES by the absence of imaginary harmonic vibrational frequencies or as transition states by the presence of exactly one imaginary frequency.

Magnetic shielding tensors have been computed with the gauge-including atomic orbitals method (GIAO) as implemented<sup>38</sup> in the Gaussian98 program,<sup>39</sup> employing the B3LYP hybrid functional<sup>40,41</sup> and Basis II’, that is, the same augmented Wachters basis set for cobalt and the IGLO–II basis<sup>42</sup> for the ligands except H: a (9s5p)/[5s4p] basis augmented with one set of d-polarization functions for C,N,O, and a (3s)/[2s] basis for H.

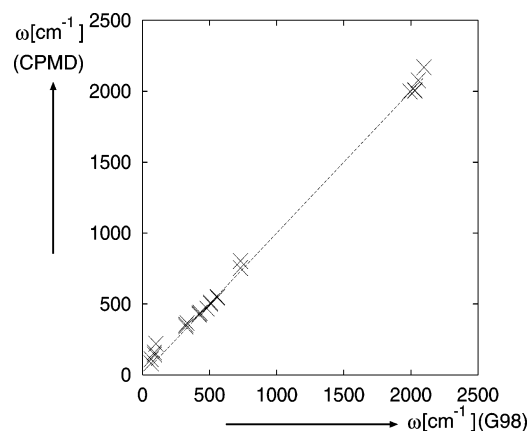
Molecular dynamics simulations were performed using the density-functional based Car-Parrinello scheme<sup>43</sup> as implemented in the CPMD program.<sup>44</sup> The BP86 functional was used, together with norm-conserving Troulier-Martins pseudo-potentials in the Kleinman-Bylander form.<sup>45,46</sup>

Periodic boundary conditions were imposed by using supercells with box sizes between 10 and 14.5 Å so that the minimum distance between atoms in neighboring virtual boxes is larger than 6.3 Å. Kohn–Sham orbitals were expanded in plane waves up to a kinetic energy cutoff of 80 Ry. In the dynamic simulations a fictitious electronic mass of 600 au and a time step of 0.121 fs were used. From the microcanonical runs with an average temperature of 300 K snapshots were collected for the NMR calculations: after an equilibration time of 0.5 ps, 41 snapshots were taken every

**Table 1:** Equilibrium ( $r_e$ ), Effective ( $r_{\text{eff}}$ ), and Averaged ( $r_{\text{av}}$ ) Geometrical Parameters for Co Complexes [in Å] BP86 Level<sup>a</sup>

|    |   |                    | $r_e(\text{AE1})$ | $r_{\text{eff}}$ | $r_e(\text{CP-opt})$ | $r_{\text{av}}(\text{CPMD})$ | exp                               |
|----|---|--------------------|-------------------|------------------|----------------------|------------------------------|-----------------------------------|
| 3  | CoH(CO) <sub>4</sub>  | Co–H               | 1.495             | 1.503            | 1.496                | 1.500                        | 1.556(18) <sup>77</sup>           |
|    |   | Co–C <sub>eq</sub> | 1.799             | 1.805            | 1.794                | 1.802                        | 1.818(3) <sup>77</sup>            |
|    |   | Co–C <sub>ax</sub> | 1.804             | 1.809            | 1.805                | 1.817                        | 1.764(10) <sup>77</sup>           |
|    |   | C–O <sub>eq</sub>  | 1.163             | 1.154            | 1.154                | 1.156                        | 1.141 <sup>77</sup>               |
|    |   | C–O <sub>ax</sub>  | 1.161             | 1.152            | 1.152                | 1.153                        | 1.141 <sup>77</sup>               |
| 4  | Co(CO) <sub>4</sub> <sup>−</sup>                                  | Co–C               | 1.777             | 1.782            | 1.771                | 1.777                        | 1.735–1.769 <sup>78</sup>         |
|    |   | C–O                | 1.183             | 1.180            | 1.175                | 1.176                        | 1.132–1.172 <sup>78</sup>         |
| 5  | CoCp(C <sub>2</sub> H <sub>4</sub> ) <sub>2</sub>                 | Co–C(Cp)           | 2.107             | 2.116            | 2.109                | 2.126                        |                                   |
|    |   | Co–C(Et)           | 2.028             | 2.041            | 2.021                | 2.042                        |                                   |
| 6  | Co(CN) <sub>6</sub> <sup>3−</sup>                                 | Co–C               | 1.929             | 1.937            | 1.918                | 1.935                        | 1.890 <sup>79</sup>               |
|    |   | C–N                | 1.188             | 1.174            | 1.175                | 1.176                        | 1.160 <sup>79</sup>               |
| 7  | Co(NH <sub>3</sub> ) <sub>3</sub> (CN) <sub>3</sub><br>(fac)      | Co–N               | 2.070             | 2.087            | 2.077                | 2.113                        |                                   |
|    |   | Co–C               | 1.851             | 1.856            | 1.844                | 1.851                        |                                   |
|    |   | C–N                | 1.184             | 1.181            | 1.172                | 1.173                        |                                   |
|    |   | N–H                | 1.029             | 1.018            | 1.027                | 1.030                        |                                   |
| 8  | Co(NH <sub>3</sub> ) <sub>6</sub> <sup>3+</sup>                   | Co–N               | 2.032             | 2.047            | 2.019                | 2.045                        | 1.972 <sup>79</sup>               |
|    |   | N–H                | 1.036             | 1.021            | 1.032                | 1.035                        | 0.980 <sup>79</sup>               |
| 9  | Co(NH <sub>3</sub> ) <sub>4</sub> (CO <sub>3</sub> ) <sup>+</sup> | Co–N <sub>1</sub>  | 1.983             | 1.993            | 1.979                | 1.998                        | 1.931(18)/1.953(14) <sup>80</sup> |
|    |   | Co–N <sub>2</sub>  | 2.042             | 2.057            | 2.033                | 2.051                        | 2.031(22) <sup>80</sup>           |
|    |   | Co–O               | 1.892             | 1.900            | 1.920                | 1.921                        | 1.905(11) <sup>80</sup>           |
|    |   | O–C                | 1.365             | 1.368            | 1.361                | 1.364                        | 1.336(18) <sup>80</sup>           |
|    |   | N–H                | 1.032             | 1.017            | 1.028                | 1.032                        |                                   |
| 10 | Co(acac) <sub>3</sub>   | Co–O               | 1.909             | 1.914            | 1.923                | 1.932                        | 1.885/1.923 <sup>79</sup>         |
| 11 | Co(H <sub>2</sub> O) <sub>6</sub> <sup>3+</sup>                   | Co–O               | 1.957             | 1.966            | 1.954                | 1.975                        | 1.873(5) <sup>81</sup>            |
|    |   | O–H                | 0.995             | 0.984            | 0.990                | 0.993                        |                                   |

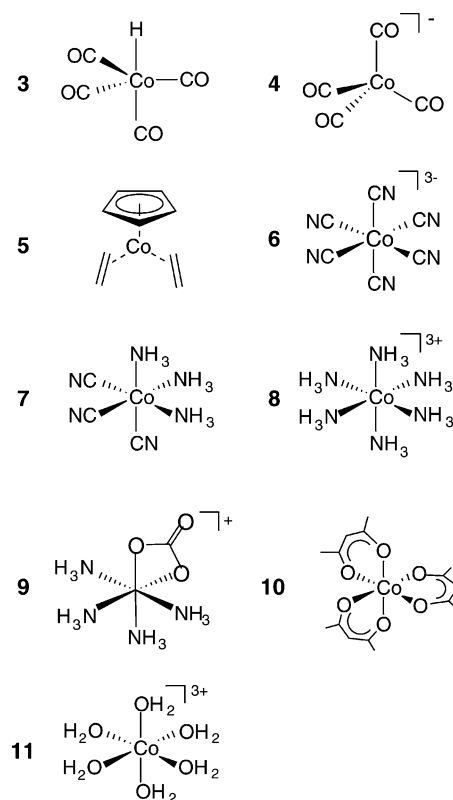
<sup>a</sup> For simplicity the averaged bond distance are given where appropriate.



**Figure 3.** Harmonic frequencies of CoH(CO)<sub>4</sub> computed with G98 and CPMD to test the performance of the generated cobalt pseudopotential.

24 fs (total time ca. 1 ps). Equilibrium geometries for the compounds were obtained by optimizing the forces on all atoms with the CPMD program using the setup detailed above (denoted CP–Opt); additionally the averaged structural parameters were computed from the microcanonical run (denoted CPMD).

For cobalt no suitable pseudopotential was available; thus, a semicore potential was generated following the method described previously.<sup>27,47</sup> We used cutoff radii of 1.5 au for all three s-, p-, and d-channels. To test the performance of the generated pseudopotential we compared the computed optimized geometries of the test set in Figure 4 to those optimized using the above nonperiodic all-electron method.



**Figure 4.** Co complexes of this study.

The bond lengths for both methods agree well within 1–2 pm (compare  $r_e(\text{CP-opt})$  and  $r_e(\text{AE1})$  data in Table 1). In addition, the vibrational frequencies of CoH(CO)<sub>4</sub> and

$\text{Co}(\text{CN})_6^{3-}$  were computed. Judging from the slope of the regression line ( $a = 0.99$ ) and the correlation coefficient ( $r = 1.00$ ), the agreement of the harmonic frequencies computed with G98 and CPMD is excellent (Figure 3). Thus, the use of the generated pseudopotential (which is available from the authors upon request) seems to be well justified.

Vibrational corrections were computed using the perturbational approach of Ruud et al.<sup>21–23</sup> In this method, the molecule is first shifted from its equilibrium geometry  $r_e$  to an effective geometry  $r_{\text{eff}}$  via the harmonic frequencies  $\omega_e$  and the cubic force field  $V^{(3)}$ :

$$r_{\text{eff},j} = r_{e,j} - \frac{1}{4\omega_{e,j}^2} \sum_m V_{e,jmm}^{(3)} \omega_{e,m} \quad (1)$$

This effective geometry corresponds to the vibrationally averaged structure of a system at 0 K.<sup>48</sup> Due to the anharmonicity of the PES the effective bond lengths are typically slightly longer than the equilibrium ones. Second, the magnetic shielding tensor is expanded in a Taylor series around this effective geometry. Thus, the expansion term containing the perturbed vibrational wave function to first order vanishes, and for the computation of the magnetic shielding to second order only the zeroth order vibrational wave function is needed. The equation for the calculation of the vibrationally averaged magnetic shielding is thus simplified to

$$\sigma_0 = \sigma_{\text{eff}} + \frac{1}{4} \sum_i \frac{\sigma_{\text{eff},ii}^{(2)}}{\omega_{\text{eff},i}} \quad (2)$$

where  $\sigma_{\text{eff}}^{(2)}$  is the second derivative of the magnetic shielding, evaluated numerically, and  $\sigma_{\text{eff}}$  and  $\omega_{\text{eff}}$  are magnetic shielding constant and the harmonic frequencies, both computed at the effective geometry.

In essence the vibrationally averaged magnetic shielding includes the leading contributions from both the anharmonicity of the PES (through the use of the effective geometry as an expansion point) and from the curvature of the magnetic shielding hypersurface.

For the computation of  $r_{\text{eff}}$  and  $\sigma_0$  the corresponding parts of the Dalton program package<sup>49</sup> had been adapted so that energies, energy derivatives, and properties produced with Gaussian98 can be processed.<sup>50</sup>  $V^{(3)}$  is obtained numerically at the BP86/AE1 level using the gradient technique, and a stepsize of 0.25 au for the finite displacements, as recommended by the test calculations in an earlier study.<sup>50</sup> For the computation of  $\sigma_{\text{eff}}^{(2)}$  we used a stepsize of 0.1 au as recommended by the test calculations in the same study.<sup>50</sup>

To compute the effective geometries of  $\text{Co}(\text{NH}_3)_6^{3+}$  with sufficient numerical precision, it was necessary to employ an ultrafine integration grid (99 radial shells with 590 angular points per shell). In this case the use of a smaller integration grid produces error-prone low-frequency modes of  $-\text{NH}_3$  rotations which lead to abnormal, long N–H bond lengths ( $>1.3$  Å). No such artifacts were encountered in the remaining molecules of the present study.

The magnetic shielding constants of the reference for relative  $\delta$  values (aqueous  $\text{K}_3[\text{Co}(\text{CN})_6]$  in the experiment) have been evaluated by correlation of the computed shieldings  $\sigma(\text{calc})$  of the complexes versus the experimental chemical shifts  $\delta(\text{exp})$  and by taking the respective axis intercepts of the linear regression lines as the estimated reference magnetic shielding for that particular level. Following this procedure we obtain  $\sigma_e(\text{BP86/AE1}) = -5837$  ppm,  $\sigma_{\text{eff}} = -6060$  ppm,  $\sigma_0 = -6124$  ppm,  $\sigma_e(\text{CP-opt}) = -5753$  ppm, and  $\sigma_{\text{av}}(\text{CPMD}) = -6203$  ppm. Similar procedures had been adopted before in cases when a direct computation for the experimental standard is difficult or impossible.<sup>18,51,52</sup> As has been noted earlier,<sup>14</sup> these data are in good qualitative accord with experimental estimates for this absolute shielding value, around  $-5400$  ppm.<sup>53</sup>

## Results

**Mononuclear Cobalt Complexes.** The test set of mononuclear cobalt complexes is displayed in Figure 4. This set comprises electron-rich organometallic and high-valent inorganic Co(III) complexes and covers a chemical shift range of nearly 20 000 ppm.

The structural parameters of the equilibrium geometries, the zero-point corrected effective geometries, and the thermal averaged geometries of this test set are listed in Table 1. Where available, the corresponding experimental values are given. The computed bond lengths are in good accord with the experimental ones. The mean deviations from experiment are around 3.6 and 3.4 pm for the BP86/AE1 and CP-opt equilibrium geometries, respectively, where the computed bond lengths tend to be overestimated. This degree of agreement is typical for the type of density functional employed.<sup>54</sup> Both quantum-mechanical zero-point correction and classical thermal averaging lead to elongation of the cobalt–ligand bonds, thereby worsening the accord with experiment in comparison to the equilibrium values. The average Co–L bond elongation due to the zero-point effect is around 0.9 pm and due to thermal averaging 1.5 pm, leading to mean deviations from experiment of 3.8 and 4.0 pm, respectively (compare  $r_{\text{eff}}$  and  $r_{\text{av}}$  values with  $r_{\text{exp}}$  in Table 1). Similar results of bond elongation due to zero-point correction and thermal averaging were observed previously<sup>31</sup> for Ti, V, Mn, and Fe complexes.

An unusual effect is observed for the zero-point corrected structures. While all the cobalt–ligand bonds are elongated, the bonds between the ligand atoms coordinated directly to the metal center and an atom further away from the cobalt center are slightly contracted, on average by  $-0.9$  pm (compare  $r_e(\text{AE1})$  and  $r_{\text{eff}}$  in Table 1). The classical thermal effect, on the other side, leads to the expected bond elongation of intraligand bonds by an average amount of  $\Delta r = +0.2$  pm (compare  $r_e(\text{CP-opt})$  and  $r_{\text{av}}$  in Table 1). It is not clear at this point whether the extent of this unusual intraligand bond contraction is real or whether it is an artifact of the perturbational approach itself.

During the CPMD simulations of the cobalt complexes ligand rotations were observed. While the cyclopentadienyl ligand within the  $\text{CoCp}(\text{C}_2\text{H}_4)_2$  complex rotates rather slowly ( $72^\circ$  in 1.5 ps), the  $-\text{NH}_3$ ,  $-\text{CH}_3$ , or  $-\text{OH}_2$  groups within

**Table 2:** Equilibrium, Effective, Zero-Point Corrected, and Averaged Chemical Shifts in ppm of the Cobalt Complexes<sup>d</sup>

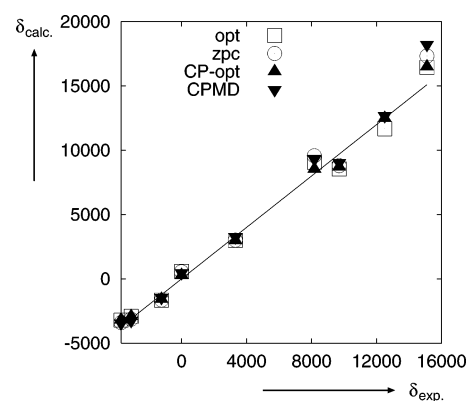
|                             |   | $\delta_e(\text{AE1})$ | $\delta_{\text{eff}}$ | $\delta_0$ | $\delta_e(\text{CP-opt})$ | $\delta_{\text{av}}(\text{CPMD})$ | exp                 |
|-----------------------------|---|------------------------|-----------------------|------------|---------------------------|-----------------------------------|---------------------|
| 3                           | $\text{CoH}(\text{CO})_4$                 | -3207                  | -3354                 | -3402      | -3187                     | -3442                             | -3720 <sup>2</sup>  |
| 4                           | $\text{Co}(\text{CO})_4^-$                | -2917                  | -3081                 | -3134      | -2924                     | -3244                             | -3100 <sup>82</sup> |
| 5                           | $\text{CoCp}(\text{C}_2\text{H}_4)_2$     | -1653                  | -1599                 | -1642      | -1586                     | -1460                             | -1235 <sup>83</sup> |
| 6                           | $\text{Co}(\text{CN})_6^{3-}$             | 576                    | 586                   | 546        | 272                       | 477                               | 0                   |
| 7                           | $\text{Co}(\text{NH}_3)_3(\text{CN})_3$   | 2994                   | 3055                  | 3019       | 2980                      | 3260                              | 3313 <sup>2</sup>   |
| 8                           | $\text{Co}(\text{NH}_3)_6^{3+}$           | 9066                   | 9596                  | 9572       | 8537                      | 9306                              | 8176 <sup>2</sup>   |
| 9                           | $\text{Co}(\text{NH}_3)_4(\text{CO}_3)^+$ | 8549                   | 8834                  | 8806       | 8703                      | 9032                              | 9700 <sup>84</sup>  |
| 10                          | $\text{Co}(\text{acac})_3$                | 11670                  | 11770                 | -          | 12486                     | 12686                             | 12500 <sup>14</sup> |
| 11                          | $\text{Co}(\text{H}_2\text{O})_6^{3+}$    | 16445                  | 17344                 | 17330      | 16493                     | 18224                             | 15100 <sup>85</sup> |
| slope <sup>a</sup>          |   | 1.00                   | 1.04 <sup>b</sup>     | 1.08       | 1.02                      | 1.08                              |                     |
| axis intercept <sup>a</sup> |   | 94                     | 95                    | 89         | 44                        | 78                                |                     |
| mean abs. dev. <sup>c</sup> |   | 692                    | 761                   | 765        | 492                       | 698                               |                     |

<sup>a</sup> From a linear regression analysis with respect to the experimental value (using the estimated reference). <sup>b</sup> Without  $\delta$  value of  $\text{Co}(\text{acac})_3$ : 1.08. <sup>c</sup> Mean absolute deviation from experiment (using the estimated reference). <sup>d</sup> GIAO-B3LYP/II level, in parentheses: source of equilibrium, effective, or snapshot geometries.

$\text{Co}(\text{NH}_3)_3(\text{CN})_3$ ,  $\text{Co}(\text{NH}_3)_6^{3+}$ ,  $\text{Co}(\text{NH}_3)_4(\text{CO}_3)^+$ ,  $\text{Co}(\text{acac})_3$ , and  $\text{Co}(\text{H}_2\text{O})_6^{3+}$  rotate much faster, so that one or two complete turns occur during the simulation time of 1.5 ps.  $\text{Co}(\text{H}_2\text{O})_6^{3+}$  and  $\text{Co}(\text{NH}_3)_6^{3+}$  are interesting in that respect because the minimum geometries display, due to small displacements of the ligands, symmetries which are lower than the highest possible ones. For the hexaquo complex, the preferred symmetry is  $S_6$  instead of  $T_h$ , while for the hexammin complex it is  $C_3$  instead of  $S_6$ . The respective energy differences between these forms are small, 6.5 and 1.6 kcal/mol for water and ammonia, respectively (BP86/AE1 level). Thus, interconversions via ligand rotations can occur readily in MD simulations at room temperature, and no specific symmetries can be assigned to the dynamically averaged structures.<sup>55</sup>

The chemical shifts of the cobalt centers computed at the equilibrium and the effective geometries as well as the zero-point corrected and thermally averaged chemical shifts are listed in Table 2. Since the experimental standard, aqueous  $\text{K}_3[\text{Co}(\text{CN})_6]^2$ , is difficult to model computationally, we have evaluated the reference magnetic shielding from a correlation of theoretical  $\sigma$  versus experimental  $\delta$  values (see computational details). The  $\delta$  values for pristine  $\text{Co}(\text{CN})_6^{3-}$  are included in Table 2. The noticeable deviation from zero of these values indicates that, as expected, gaseous **6** is a poor model for the experimental standard and that solvent effects on this highly charged anion may be sizable.<sup>56</sup> For the isoelectronic  $\text{Fe}(\text{CN})_6^{4-}$ , remarkably large solvent effects on the metal shielding constant, exceeding 1000 ppm, have been reported using suitable MD-based methods.<sup>28</sup> Similar simulations are in progress to model the aqueous solutions of Co complexes; the results of these demanding calculations will be reported in due course.

At the equilibrium geometries—computed both at BP86/AE1 and BP86/CP-opt levels, the chemical shifts tend to be shifted upfield with regard to the experimental values. The mean absolute error, 692 and 492 ppm, respectively, seems quite large at first sight. These errors, however, amount to but a few percent of the total  $^{59}\text{Co}$  chemical shift range (ca. 19 000 ppm), a relative accuracy typical for DFT computed transition-metal chemical shifts so far.<sup>12,13</sup>



**Figure 5.** Plot of computed—GIAO—B3LYP for BP86/AE1 optimized (opt), zero-point corrected geometries (zpc), CP optimized (CP-opt), and thermally averaged geometries (CPMD)—versus experimental chemical shifts; solid line: ideal slope.

The slope of the  $\delta_e(\text{calc})$  versus  $\delta(\text{exp})$  regression lines are very close to the ideal value of 1 (see Table 2, Figure 5). Thus, our particular combination of DFT methods (B3LYP shifts for BP86 geometries) appears to perform very well in such static computations, without showing any signs of degradation at the deshielded end of the  $\delta(^{59}\text{Co})$  range.

Both zero-point correction as well as thermal averaging lead to a decrease of the shielding constants. Typically this downfield shift is larger in the substrates than that in the reference shieldings, resulting in an increase of the  $\delta$  values. The two exceptions are  $\text{CoH}(\text{CO})_4$  and  $\text{Co}(\text{CO})_4^-$ , where the downfield shifts are smaller than that of the reference, so that the overall effect is a decrease of  $\delta$ . The mean absolute error increases somewhat upon zero-point correction and thermal averaging, namely from 692 to 765 and from 492 to 698 for  $\delta_0$  and  $\delta_{\text{av}}$ , respectively. The larger slope of the corresponding regression lines (1.08 for the zero-point corrected and the thermal averaged chemical shifts, as compared to the near-ideal values for the corresponding  $\delta_e$  data, see Table 2) also indicates that the zero-point corrected and thermal averaged chemical shift agree somewhat less perfectly with the experimental data than the chemical shifts at the equilibrium geometry.

**Table 3:** Shielding/Bond-Length Derivative  $\partial\sigma_{\text{Co}}/\partial r_{\text{Co-L}}$  in ppm/pm of Cobalt Compounds<sup>b</sup>

|    |   |  | $\partial\sigma_{\text{Co}}/\partial r_{\text{Co-L}}$ | "per bond" <sup>a</sup> |
|----|---|--|---|-------------------------|
| 3  | CoH(CO) <sub>4</sub>  | Co-(CO) <sub>4</sub>                             | -118.0  | -29.5                   |
|    |   | Co-H   | -23.4   | -23.1                   |
| 4  | Co(CO) <sub>4</sub> <sup>-</sup>                                  | Co-(CO) <sub>4</sub>                             | -132.3  | -33.1                   |
| 5  | CoCp(C <sub>2</sub> H <sub>4</sub> ) <sub>2</sub>                 | Co-Cp  | -103.3  | -103.3                  |
|    |   | Co-(C <sub>2</sub> H <sub>4</sub> ) <sub>2</sub> | -113.2  | -56.6                   |
| 6  | Co(CN) <sub>6</sub> <sup>3-</sup>                                 | Co-(CN) <sub>6</sub>                             | -311.9  | -52.0                   |
| 7  | Co(NH <sub>3</sub> ) <sub>3</sub> (CN) <sub>3</sub>               | Co-(NH <sub>3</sub> ) <sub>3</sub>               | -108.1  | -36.0                   |
|    |   | Co-(CN) <sub>3</sub>                             | -245.4  | -81.8                   |
| 8  | Co(NH <sub>3</sub> ) <sub>6</sub> <sup>3+</sup>                   | Co-(NH <sub>3</sub> ) <sub>6</sub>               | -471.1  | -78.5                   |
| 9  | Co(NH <sub>3</sub> ) <sub>4</sub> (CO <sub>3</sub> ) <sup>+</sup> | Co-(NH <sub>3</sub> ) <sub>4</sub>               | -334.8  | -83.7                   |
|    |   | Co-(CO <sub>3</sub> )                            | -109.6  | -54.8                   |
| 10 | Co(acac) <sub>3</sub>   | Co-(acac) <sub>3</sub>                           | -630.8  | -105.1                  |
| 11 | Co(H <sub>2</sub> O) <sub>6</sub> <sup>3+</sup>                   | Co-(H <sub>2</sub> O) <sub>6</sub>               | -692.6  | -115.4                  |

<sup>a</sup>  $\pi$ -ligands counted as single ligand. <sup>b</sup> Explanation see text.

The largest offset due to zero-point correction and thermal averaging is obtained for Co(H<sub>2</sub>O)<sub>6</sub><sup>3+</sup> and Co(NH<sub>3</sub>)<sub>6</sub><sup>3+</sup>, for which the experimental NMR shifts are obtained in aqueous solution. Together with Co(CN)<sub>6</sub><sup>3-</sup>, these are the complexes with the largest overestimation of the cobalt–ligand bond lengths with respect to experiment (Table 1). It is well possible that also for these highly charged cations, interactions with a protic solvent (or polar surrounding) can affect geometries and NMR chemical shifts, as had been noted for highly charged anions.<sup>28</sup> The relatively good performance of the corresponding gas-phase equilibrium structures in the chemical shift calculations could then benefit to some extent from fortuitous error cancellation. Further investigations along this line are in progress.

As in our previous report,<sup>31</sup> we investigated the influence of the metal–ligand bond elongation on the NMR chemical shifts in a more quantitative manner. To this end, we have computed the metal shielding/bond-length derivatives employing the following procedure: the magnetic shielding constants  $\sigma$  have been computed at five different geometries, in which the bond length to each of symmetry-equivalent ligand atoms have been distorted from their equilibrium values by  $-1$  pm to  $+3$  pm while leaving all other parameters unchanged. The shielding/bond-length derivatives  $\partial\sigma_{\text{Co}}/\partial r_{\text{Co-L}}$  have been determined by linear regression (see Table 3). Qualitatively we obtained similar results as for the Ti, V, Mn, and Fe complexes studied previously. The largest

$\partial\sigma_{\text{Co}}/\partial r_{\text{Co-L}}$  values are attributed to the 6- $\pi$ -ligand Cp and to the oxygen ligands,  $-\text{OH}_2$  and (acac). For Co(CN)<sub>6</sub><sup>3-</sup>, an experimental estimate is available for this quantity, based on isotope effects on the force field and the <sup>59</sup>Co chemical shift.<sup>57</sup> As has been noted before,<sup>14</sup> this experimental value,<sup>57</sup>  $\partial\sigma_{\text{Co}}/\partial r_{\text{Co-L}} = -75$  ppm pm<sup>-1</sup> is somewhat underestimated at the B3LYP level, cf. the value of  $-52$  ppm pm<sup>-1</sup> given in Table 3. As is the case for other transition-metal complexes,<sup>31</sup> the individual  $\partial\sigma_{\text{Co}}/\partial r_{\text{Co-L}}$  increments for a given ligand L are not transferable between complexes but increase in absolute value with the deshielding of the metal. For instance, the  $\partial\sigma_{\text{Co}}/\partial r_{\text{Co-N}}$  values of the ammonia ligands in Co(NH<sub>3</sub>)<sub>3</sub>(CN)<sub>3</sub> and Co(NH<sub>3</sub>)<sub>6</sub><sup>3+</sup> are ca.  $-36$  and  $-79$  ppm, respectively (see per bond values in Table 3), paralleling the increase in  $\delta(^{59}\text{Co})$  in this series, from  $\delta = 3313$  to  $8176$  ppm (Table 2). The shielding/bond-length derivatives are negative for all cobalt complexes in this study. However, this is not generally the case, for example when spin–orbit effects come to play they could be positive.<sup>58,59</sup>

To investigate to what extent the change in chemical shift due to zero-point correction and thermal averaging can be traced back to bond elongations, we have multiplied the elongation of each metal–ligand bond with the corresponding shielding/bond-length derivative and summed up all the products for each complex. The estimated differences  $(\sigma_e - \sigma_{\text{av}})_{\text{estd}}$  and  $(\sigma_e - \sigma_{\text{eff}})_{\text{estd}}$  are compared to the actual calculated difference in the magnetic shielding  $(\sigma_e - \sigma_{\text{av}})_{\text{calc}}$  and  $(\sigma_e - \sigma_{\text{eff}})_{\text{calc}}$  in Table 4, and are plotted in Figure 6.

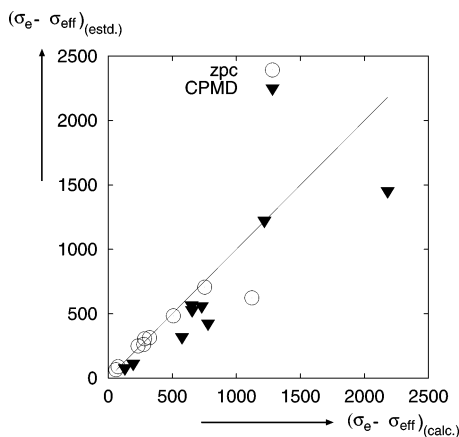
For the zero-point corrected chemical shifts, these estimated differences agree well with the calculated values, except for the Co(H<sub>2</sub>O)<sub>6</sub><sup>3+</sup> complex. Judging from the slope of the regression line, the bond elongation accounts for 91% of the downfield shift due to zero-point correction (neglecting the Co(H<sub>2</sub>O)<sub>6</sub><sup>3+</sup> values). It is unclear why the estimated difference  $(\sigma_e - \sigma_{\text{eff}})_{\text{estd}}$  of the Co(H<sub>2</sub>O)<sub>6</sub><sup>3+</sup> complex only covers approximately half of the value of the actual calculated difference  $(\sigma_e - \sigma_{\text{eff}})_{\text{calc}}$ . It is possible that in this case, the influence of bond angles on the chemical shift is larger than assumed.

For the thermally averaged chemical shifts, the estimated and calculated differences agree to a somewhat lesser extent. In this case the bond elongation accounts, on average, for 72% of the downfield shift. We obtained similar results for the complexes of Ti, V, Mn, and Fe in our previous report,<sup>31</sup>

**Table 4:** Difference between Thermally Averaged and Equilibrium Magnetic Shieldings  $\sigma_e - \sigma_{\text{av}}$  and between Magnetic Shieldings at Effective and Equilibrium Geometry<sup>a</sup>

|    |   | $(\sigma_e - \sigma_{\text{eff}})_{\text{estd}}$ | $(\sigma_e - \sigma_{\text{eff}})_{\text{calc}}$ | $(\sigma_e - \sigma_{\text{av}})_{\text{estd}}$ | $(\sigma_e - \sigma_{\text{av}})_{\text{calc}}$ |
|----|---|--|--|---|---|
| 3  | CoH(CO) <sub>4</sub>  | 89   | 76   | 115   | 195   |
| 4  | Co(CO) <sub>4</sub> <sup>-</sup>                                  | 66   | 59   | 79  | 130   |
| 5  | CoCp(C <sub>2</sub> H <sub>4</sub> ) <sub>2</sub>                 | 262  | 277  | 321   | 576   |
| 6  | Co(CN) <sub>6</sub> <sup>3-</sup>                                 | 250  | 233  | 530   | 655   |
| 7  | Co(NH <sub>3</sub> ) <sub>3</sub> (CN) <sub>3</sub>               | 306  | 284  | 561   | 730   |
| 8  | Co(NH <sub>3</sub> ) <sub>6</sub> <sup>3+</sup>                   | 707  | 753  | 1225  | 1219  |
| 9  | Co(NH <sub>3</sub> ) <sub>4</sub> (CO <sub>3</sub> ) <sup>+</sup> | 484  | 508  | 427   | 779   |
| 10 | Co(acac) <sub>3</sub>   | 315  | 323  | 568   | 650   |
| 11 | Co(H <sub>2</sub> O) <sub>6</sub> <sup>3+</sup>                   | 623  | 1122   | 1454  | 2181  |

<sup>a</sup> estd: estimated using the shielding/bond-length derivatives from Table 3 multiplied with the corresponding differences in bond lengths from Table 1 and calc: actual calculated differences from Table 2.



**Figure 6.** Plot of estimated vs calculated differences between effective and equilibrium magnetic shieldings  $\sigma_e - \sigma_{\text{eff}}$  (zpc) and between thermally averaged and equilibrium magnetic shieldings  $\sigma_e - \sigma_{\text{av}}$  (CPMD); data from Table 4; line: ideal slope.

**Table 5:** Computed (GIAO–B3LYP for BP86/AE1 Optimized Geometries) and Experimental Elements of the  $^{59}\text{Co}$  Chemical Shielding Tensor ( $\sigma_{11}$ ,  $\sigma_{22}$ ,  $\sigma_{33}$ )<sup>c</sup>

|           | $\sigma_{ii}(\text{opt})$ | $\sigma_{ii}(\text{eff})$ | $\sigma_{ii}(\text{CP-opt})$ | $\sigma_{ii}(\text{CPMD})$ | $\sigma_{ii}(\text{exp})$ |
|-----------|---------------------------|---------------------------|------------------------------|----------------------------|---------------------------|
| <b>9</b>  | -875                      | -969                      | -871                         | -1210                      | -667 <sup>84</sup>        |
|           | -220                      | -294                      | 12                           | -174                       | -167                      |
|           | 1095                      | 1262                      | 859                          | 1383                       | 833                       |
| <b>10</b> | -449                      | -487                      | -538                         | -962                       | -686 <sup>a</sup>         |
|           | 220                       | 231                       | 231                          | 20                         | 212                       |
|           | 230                       | 255                       | 306                          | 944                        | 473                       |
| <b>12</b> | -672                      | -536                      |                              |                            | -696 <sup>b</sup>         |
|           | 211                       | 176                       |                              |                            | 232                       |
|           | 463                       | 362                       |                              |                            | 464                       |

<sup>a</sup>  $\delta_{\text{iso}} = 12505(1)$  ppm,  $\Omega = 1159(46)$  ppm,  $\kappa = -0.55(5)$ .<sup>4</sup> <sup>b</sup> Exptl data for **2**,  $\delta_{\text{iso}} = 3640(100)$  ppm,  $\Omega = 1160(50)$  ppm,  $\kappa = -0.6(2)$ .<sup>4</sup> <sup>c</sup> Traceless representation.

where an average slope of 63% for the corresponding linear regression line has been obtained. Thus, we can also state for the cobalt complexes that it is indeed the metal–ligand bond distances that are most important for the chemical shift of the metal nucleus and its sensitivity toward temperature effects.<sup>60</sup> The remaining discrepancies between estimated and calculated effects on  $\sigma(\text{Co})$  are probably due to nonadditivity of the increments and the neglect of L–Co–L bendings.

**Table 6:** Largest Component  $q_{zz}$  of the B3LYP/II' Computed Electric Field Gradient at the  $^{59}\text{Co}$  Nucleus in a.u., Square of  $q_{zz}$ , the Thermally Averaged  $q_{zz}$ , the Experimental  $q_{zz}$  Computed from the Nuclear Quadrupole Coupling Constant  $e q_{zz} Q/h$  in MHz and the Experimental Line Widths in Hz

|           | at $r_e(\text{BP86/AE1})$ |            | CPMD                       |                            | exp                          |             |  |
|-----------|---------------------------|------------|----------------------------|----------------------------|------------------------------|-------------|--|
|           | $ q_{zz} $                | $q_{zz}^2$ | $ \langle q_{zz} \rangle $ | $\langle  q_{zz}  \rangle$ | $e q_{zz} Q/h$               | $q_{zz}$    | $\Delta\nu_{1/2}$                              |
| <b>3</b>  | 1.365                     | 1.863      | 0.931                      | 1.370                      | 101–163 <sup>a</sup>         | 1.023–1.652 | 3265 <sup>82</sup>                             |
| <b>5</b>  | 1.833                     | 3.360      | 1.925                      | 1.925                      |                              |             | 6800 <sup>83</sup>                             |
| <b>6</b>  | 0.000                     | 0.000      | 0.041                      | 0.371                      | 6.2 <sup>2</sup>             | 0.063       | $6 \pm 1.5$ ; <sup>2</sup> 202 <sup>86</sup>   |
| <b>7</b>  | 0.062                     | 0.004      | 0.094                      | 0.310                      |                              |             | $750 \pm 50$ ; <sup>2</sup> 1007 <sup>86</sup> |
| <b>8</b>  | 0.011                     | 0.000      | 0.079                      | 0.367                      | 0.97(1)–3.37(1) <sup>2</sup> | 0.010–0.034 | $183 \pm 10$ ; <sup>2</sup> 202 <sup>86</sup>  |
| <b>9</b>  | 0.427                     | 0.182      | 0.042                      | 0.503                      | 18.82(1) <sup>2</sup>        | 0.191       | $5476 \pm 400$ <sup>2</sup>                    |
| <b>10</b> | 0.022                     | 0.000      | 0.003                      | 0.235                      | 5.53(7) <sup>2</sup>         | 0.056       |  |
| <b>12</b> | 0.492                     | 0.242      |                            |                            | 29.6(4) <sup>2</sup>         | 0.300       | 3100 <sup>87</sup>                             |

<sup>a</sup> For  $\text{XCo}(\text{CO})_4$  with  $\text{X} = \text{SiPh}_3$  and  $\text{X} = \text{SnCl}_3$ .<sup>2</sup>

**The Magnetic Shielding Tensor.** In a number of cases, information on the full  $^{59}\text{Co}$  magnetic shielding tensor is known from solid-state  $^{59}\text{Co}$  NMR spectroscopy.<sup>4</sup> To investigate the influence of zero-point vibration and thermal averaging in more detail we have taken a look at individual shielding-tensor elements of three example compounds, namely of **9**, **10**, and **12**, a model for the cobaloxime **2** lacking the four peripheral methyl groups at the glyoximate ligands (see section on cobaloximes below). The other complexes of the present study were not included as either there were no shielding tensors reported in the literature or there were multiple sites in the NMR spectrum. The computed and experimental elements of the traceless shielding tensor (i.e. the values relative to the isotropic average) are listed in Table 5. The slopes of the  $\sigma_{ii}(\text{calc})$  versus  $\sigma_{ii}(\text{exp})$  regression lines are 1.03 and 0.97 at the BP86/AE1 and CP-optimized geometries, respectively, which is very close to the ideal slope of 1. The tensor elements at the effective geometry do not differ significantly from those at the static optimized geometry. The slope of the regression line is also quite similar with a value of 1.06. Also, there is no general trend in the change of the shielding tensor elements from the optimized to the effective geometry. For **9** the range of the values covered by the tensor elements is increased, while for **12** the range is reduced. The thermal averaged tensor elements agree significantly less with the experimental data (slope = 1.63). In general, the thermal averaging leads to a broadened span of the shielding tensor elements, and the deviation from the experimental values increases in the smallest and largest tensor elements. It is possible, that due to the thermal movement in the gas phase “extreme” geometries are obtained which lead to far higher (or lower) values in the tensor elements in comparison to those at the static optimized geometries. The experimental data, on the other hand, are taken from crystals or powder samples, where the thermal movement does not allow such “extreme” geometries.

Finally we have investigated the tensor of the electric field gradient (EFG). First, the largest component of the EFG tensor  $q_{zz}$  can be compared directly to experimental values from the nuclear quadrupole coupling constant,  $e^2 q_{zz} Q/h$  ( $Q$ : nuclear quadrupole moment). The computed  $q_{zz}$  values at the optimized geometry agree both in sequence as in order of magnitude with the experimental data (see Table 6).<sup>61</sup>

**Table 7:** Equilibrium ( $r_e$ ) Geometrical Parameters in Å and Relative Energy (BP86/AE1) of **1a–1e**<sup>a</sup>

|                                 | $r_e(\text{Co–Co})$ | $r_e(\text{Co–C}_a)$ | $r_e(\text{Co–C}_b)$ | $r_e(\text{Co–C}_c)$ | $r_e(\text{Co–C}_d)$ | $E$ [kJ/mol] |
|---------------------------------|---------------------|----------------------|----------------------|----------------------|----------------------|--------------|
| <b>1a</b>                       | 2.550               | 1.809                | 1.820                | 1.958                |                      | 0            |
| exp <sup>88</sup> ( <b>1a</b> ) | 2.528(1)            | 1.815(1)             | 1.832(4)             | 1.939(4)             |                      |              |
| <b>1b</b>                       | 2.697               | 1.812                | 1.775                |                      |                      | 27.0         |
| <b>1c</b>                       | 2.633               | 1.797                | 1.803                |                      |                      | 16.7         |
| <b>1d (1b → 1c)</b>             | 2.694               | 1.798                | 1.804                | 1.776                | 1.832                | 36.1         |
| <b>1e (1c → 1a)</b>             | 2.613               | 1.808                | 1.801                | 1.807                | 1.802                | 18.6         |

<sup>a</sup> See Figure 7 for numbering of atoms.

When analyzing the thermally averaged  $q_{zz}$  values, we should distinguish between highly symmetric molecules, which have tensor elements of approximately zero at the optimized geometry and unsymmetric molecules with high  $q_{zz}$  values. In the first case, we notice a strong oscillation of  $q_{zz}$  around zero, whereas the average value  $|\langle q_{zz} \rangle|$  is very small ( $<0.08$ ). The unsymmetric molecules, on the other hand, oscillate around the respective  $q_{zz}$  value at the optimized geometry (e.g. for **3** the value oscillates between  $|1.1|$  and  $|1.8|$ ), so that the average absolute value is only slightly larger than the  $q_{zz}$  value at the static geometry (1.370 instead of 1.365 at the optimized geometry).

We also attempted to correlate the intense line broadening which is caused by the interaction of the quadrupole moment of the <sup>59</sup>Co nucleus ( $I=7/2$ ) with the EFG at the nucleus. However, we did not obtain a meaningful correlation between  $q_{zz}$  (either computed or taken from nuclear quadrupole resonance spectroscopy) and the experimental line width,  $\Delta\nu_{1/2}$ , which for a quadrupole relaxation is expected to be<sup>62</sup>

$$\Delta\nu_{1/2} \propto q_{zz}^2(1 + \eta^2/3)\tau_c \quad (3)$$

( $\eta$ : asymmetry parameter of the EFG tensor,  $\tau_c$ : molecular correlation time). The main reason is probably that the line width also depends on other factors such as temperature, solvent or molecular size, and the EFG need not always be decisive. Similar findings have been reported for other nuclei such as <sup>99</sup>Ru<sup>51</sup> and <sup>49</sup>Ti.<sup>30</sup>

## Dicobalt Octacarbonyl

Dicobalt octacarbonyl is not included in the test set discussed so far, because its structure in solution is fluxional. For instance only one signal is found in the <sup>13</sup>C NMR spectrum.<sup>63</sup> Doubly bridged **1a** is found in the crystal,<sup>64</sup> but other forms such as **1b** and **1c** have been proposed in matrix-isolation studies.<sup>65</sup> Earlier DFT studies have confirmed that **1a–1c** are quite similar in energy, but the relative stabilities can vary with the particular functional employed.<sup>66,67</sup> If the chemical shift of the individual isomers would be sufficiently different from each other, comparison of computed and experimental  $\delta$  values could afford new evidence concerning the equilibrium mixture in the NMR experiment. Similar structural applications of chemical shift calculations have been reported before.<sup>12,68</sup> In addition to the minima **1a–1c**, we have also located transition structures connecting them, which—to our knowledge—have not been reported yet.

The optimized geometrical parameters of stable cobalt carbonyls as well as those of the transition states are listed in Table 7. The computed values of **1a** agree very well with

**Table 8:** Chemical Shifts<sup>89</sup> of **1a–1e**<sup>a</sup>

|                     | Co                  | C <sub>a</sub> | C <sub>b</sub> | C <sub>c</sub> | C <sub>d</sub> | C <sub>average</sub> |
|---------------------|---------------------|----------------|----------------|----------------|----------------|----------------------|
| <b>1a</b>           | −2265               | 186.8          | 197.0          | 233.5          |                | 203.6                |
| <b>1b</b>           | −2399               | 193.7          | 217.0          |                |                | 199.5                |
| <b>1c</b>           | −2354               | 208.7          | 198.8          |                |                | 203.8                |
| <b>1d (1b → 1c)</b> | −2326               | 203.2          | 200.3          | 213.4          | 187.1          | 201.1                |
| <b>1e (1c → 1a)</b> | −2282               | 194.7          | 201.3          | 204.7          | 211.5          | 203.1                |
| exp                 | −2200 <sup>71</sup> |                |                |                |                | 203.2 <sup>63</sup>  |

<sup>a</sup> See Figure 7 for numbering of atoms; GIAO–B3LYP/II level.

the experimental values of the crystal structure; the computed and experimental bond length agree better than 1%. Kenny et al. reported about similar good values using the BP86 density functional and a double- $\zeta$  plus polarization basis.<sup>66</sup> Energetically the  $C_{2v}$  symmetric **1a** is the most stable structure—**1b** and **1c** lie higher in energy by 27.0 and 16.7 kJ/mol, respectively, very close to the corresponding values of 26.4 and 15.5 kJ/mol obtained by Kenny et al.<sup>66</sup>

We have located 2 different transition states (**1d** and **1e**), one between **1b** and **1c** and one between **1a** and **1c**. Despite considerable effort, we could not locate a transition state connecting **1a** and **1b**. Even in the absence of the latter, interconversion between the two isomers **1a** and **1b** is possible via isomer **1c**. Both transition states were identified by a single imaginary vibrational frequency each, which describes the movement of transformation between the two minima they connect. Additionally, each transition state was checked by following the intrinsic reaction coordinate (IRC)<sup>69,70</sup> from the transition state to the two reactants. The two transition states are  $C_2$ -symmetric, i.e. one  $C_2$  axis is preserved during the transitions  $D_{3d} \rightarrow D_{2d}$  and  $C_{2v} \rightarrow D_{2d}$ . The Co–Co bond length of the transition states is of a similar value as in the unbridged structures **1b** and **1c**. Energetically these transition states do not lie much higher than the reactants. The largest energy difference (between **1a** and **1d**) amounts to 36 kJ/mol (see Table 7). Thus, the interconversion of all three isomers should readily take place at a reasonable rate at room temperature. These results are fully consistent with the observed fluxionality on the NMR time scale.

Our computed NMR chemical shifts show that the  $\delta(^{59}\text{Co})$  values of **1a–1c** lie close together (see Table 8), between −2265 ppm and −2399 ppm, and agree well with the experimental value of −2200 ppm.<sup>71</sup> The differences between the chemical shifts of **1a–1c** are much smaller than the mean absolute errors in the computed  $\delta(^{59}\text{Co})$  values (Table 2). Thus, no structural assignments can be made on the basis of the  $\delta(^{59}\text{Co})$  values. Likewise, the computed <sup>13</sup>C chemical shifts do not appear to be of diagnostic value for



the structure in solution. While in the static structures, the chemical shifts of terminal C–O groups (between 186 and 217 ppm) could be discriminated from those of a bridging carbonyl group (234 ppm in **1a**), the  $\delta(^{13}\text{C})$  value averaged over all C atoms of a given isomer (which should be compared to the single, dynamic average observed in solution) shows only fairly little variation between **1a** and **1c**.

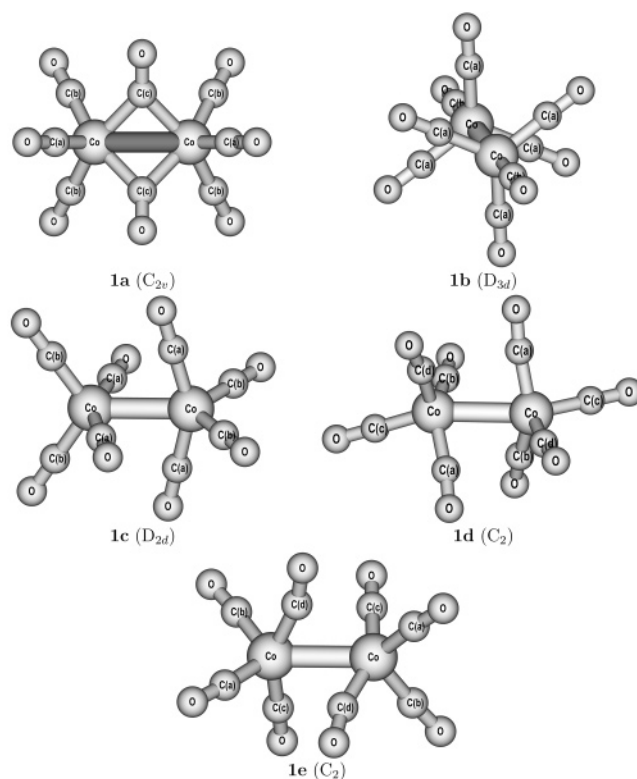
As the average Co–C bond length in structure **1a** is 5 pm longer than in **1b** and **1c**, we would expect a shift to higher fields for **1b** and **1c**. As this is not the case, it is very probable that the low field shift is due to the increase in the Co–Co bond length from **1a** to **1b** and **1c**. Using the computed shielding/bond-length derivative for the Co–Co bond of **1b** ( $-11.6 \text{ ppm pm}^{-1}$ ), we obtain an estimated value  $\Delta\delta(^{59}\text{Co}) = 171 \text{ ppm}$  for **1b** and  $\Delta\delta(^{59}\text{Co}) = 96 \text{ ppm}$  for **1c**, which resembles the calculated values of 134 and 89 ppm, respectively.

The computed vibrational harmonic frequencies of **1a** to **1c** agree very well with the data of Kenny et al.,<sup>66</sup> obtained at a similar level, and with experiment, where available. We provide these data as Supporting Information and refrain from a deeper discussion.

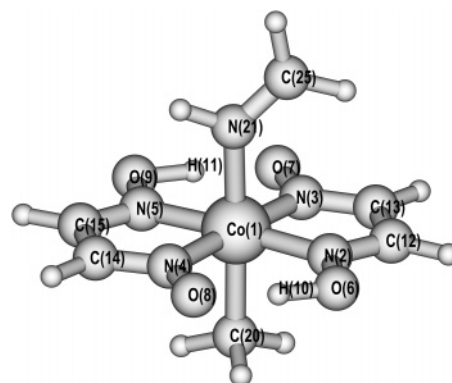
## Cobaloximes

Cobaloximes are important model complexes for cobalamines and vitamin B<sub>12</sub>.<sup>72</sup> Consequently, a sizable amount of spectroscopical data, including  $\delta(^{59}\text{Co})$  values are available.<sup>2,6,73</sup> Here we pay special attention to the effect of deuteration of a remote H-bond on the  $\delta(^{59}\text{Co})$  chemical shift, as observed in the dimethylglyoximato complex ( $[\text{MeCo}(\text{Hdmg})_2\text{py}]^7$  (Hdmg = dimethylglyoximate, py = pyridine, **2**). We have considered two smaller model systems, **12** and **13**, with glyoximato ligands (i.e. lacking the four methyl groups at the periphery) and pyridine (**12**) or imine  $\text{H}_2\text{C}=\text{NH}$  (**13**) as a model for the latter. The deuterated compounds in which H<sub>10</sub> and H<sub>11</sub> (see Figure 8) are substituted by deuterium are labeled as **12a** and **13a**, respectively. Changing an isotope within a molecule has no effect on the equilibrium geometry but alters the vibrational wave function of the molecule. Thus, the effect of deuteration has to be investigated by including the anharmonic vibrations and computing the vibrationally averaged geometry. Secondary isotope effects (i.e. by changing an isotope next to the resonating nucleus) have been reproduced with a variety of approaches.<sup>24</sup> An isotope effect over three bonds, however, as in **2**, is a much more challenging task. We have computed the zero-point corrected effective geometries using the same perturbational approach as discussed above and computed the NMR chemical shift  $\delta_{\text{eff}}$  at these effective geometries. We have not computed the fully zero-point corrected value,  $\delta_0$ , as the differences between  $\delta_0$  and  $\delta_{\text{eff}}$  tend to be very small (see Table 2) and do not warrant the extra effort that would be necessary to evaluate  $\delta_0$  for a system as large as **12**. In Table 9 the BP86/AE1 and BP86/AE1(\*) optimized geometries as well as the effective geometries of **12**, **12a**, **13**, and **13a** are listed.

On going from the equilibrium to the effective geometry of **13**, the bridging proton is shifted toward a more symmetric



**Figure 7.** BP86/AE1 optimized geometries of  $\text{Co}_2(\text{CO})_8$  isomers (**1a–1c**) and transition states (**1d** and **1e**).



**Figure 8.** BP86/AE1(\*) optimized geometry of **2a**.

position between the two oxygen atoms, as evidenced by the change in the O–H/H $\cdots$ O distances from 1.09/1.43 Å ( $r_e$ ) to 1.18/1.27 Å ( $r_{\text{eff}}$ , Table 9). Upon deuteration, the hydrogen atom is shifted back from this more mid-positioned site between the two oxygen atoms to a more asymmetric position close to one of the oxygen atoms but not as asymmetric as in the equilibrium structure (cf. the O–H/H $\cdots$ O distances of 1.15/1.32 Å for **13a** in Table 9). This finding is consistent with a potential energy surface which forms a double well along the O $\cdots$ H–O path with a low central barrier and a pronounced anharmonicity.<sup>74</sup> We have located the transition state for migration of the hydrogen between the two oxygen atoms in **13**, (O $\cdots$ H–O)  $\leftrightarrow$  (O $\cdots$ H $\cdots$ O)  $\leftrightarrow$  (O–H $\cdots$ O). The transition state is nearly  $C_s$ -symmetric except for the methyl group, which is rotated out of the symmetry plane, and shows a more mid-positioned

**Table 9:** Equilibrium ( $r_e$ ) and Effective ( $r_{\text{eff}}$ ) Geometrical Parameters in Å of **12** and **13**<sup>c</sup>

|         | atoms | $r_e$ ( <b>12</b> ) <sup>a</sup> | $r_{\text{eff}}$ ( <b>12</b> ) <sup>a</sup> | $r_{\text{eff}}(\text{D})$ ( <b>12a</b> ) <sup>a</sup> | $r_e$ ( <b>13</b> ) <sup>b</sup> | $r_{\text{eff}}$ ( <b>13</b> ) <sup>b</sup> | $r_{\text{eff}}(\text{D})$ ( <b>13a</b> ) <sup>b</sup> |
|---------|-------|----------------------------------|---|--|----------------------------------|---|--|
| Co–N    | 1–2   | 1.898                            | 1.903                                       | 1.903  | 1.892                            | 1.899                                       | 1.900  |
|         | 1–3   | 1.922                            | 1.918                                       | 1.921  | 1.912                            | 1.905                                       | 1.911  |
|         | 1–4   | 1.922                            | 1.919                                       | 1.922  | 1.920                            | 1.913                                       | 1.918  |
|         | 1–5   | 1.897                            | 1.902                                       | 1.902  | 1.901                            | 1.907                                       | 1.907  |
| Co–Imin | 1–21  | 2.069                            | 2.083                                       | 2.084  | 1.980                            | 1.995                                       | 1.996  |
| Co–Me   | 1–20  | 2.012                            | 2.026                                       | 2.025  | 2.009                            | 2.021                                       | 2.020  |
| O – H,D | 6–10  | 1.066                            | 1.122                                       | 1.104  | 1.085                            | 1.180                                       | 1.149  |
|         | 8–10  | 1.497                            | 1.393                                       | 1.428  | 1.430                            | 1.278                                       | 1.331  |
|         | 7–11  | 1.484                            | 1.381                                       | 1.418  | 1.420                            | 1.262                                       | 1.317  |
|         | 9–11  | 1.068                            | 1.127                                       | 1.108  | 1.089                            | 1.188                                       | 1.156  |
| O ... O | 6–8   | 2.551                            | 2.506                                       | 2.522  | 2.506                            | 2.452                                       | 2.473  |
|         | 7–9   | 2.540                            | 2.499                                       | 2.515  | 2.499                            | 2.444                                       | 2.465  |

<sup>a</sup> AE1 basis. <sup>b</sup> AE1(\*) basis. <sup>c</sup> BP86 level; see Figure 8 for numbering of atoms.

**Table 10:** Equilibrium ( $\delta_e$ ) and Effective ( $\delta_{\text{eff}}$ ) <sup>59</sup>Co Chemical Shifts in ppm (GAIO–B3LYP/II Level) of **12** and **13**

|                                   | basis <sup>a</sup> | $\delta_e$ | $\delta_{\text{eff}}$ |
|-----------------------------------|--------------------|------------|-----------------------|
| <b>12</b>                         | AE1                | 3362       | 3211                  |
| <b>12a</b>                        | AE1                | 3362       | 3276                  |
| <b>13</b>                         | AE1                | 3120       | 2977                  |
| <b>13</b>                         | AE1(*)             | 2943       | 2761                  |
| <b>13a</b>                        | AE1                | 3120       | 3043                  |
| <b>13a</b>                        | AE1(*)             | 2943       | 2847                  |
| <b>2</b> <i>exp</i> <sup>b</sup>  |                    |            | ≈3640                 |
| <b>2a</b> <i>exp</i> <sup>b</sup> |                    |            | ≈3700                 |

<sup>a</sup> Used in energy evaluations together with BP86 functional.

<sup>b</sup> Estimated from the spectrum plotted in ref 7, cf. footnote 32.

site of the migrating H atoms between the O-termini ( $r_e = 1.27/1.19$  Å). This transition state is only 2.4 kJ/mol higher in energy than the ground state, so that the hydrogen can readily exchange between the two oxygen atoms at room temperature. Due to the dependence of the vibrational energy on the nuclear mass ( $E \propto 1/\sqrt{m}$ ), the vibrational levels of the deuterated compound lie “deeper” inside the anharmonic potential well, so that in the vibrationally averaged geometry the deuterium is located closer to one of the two oxygen atoms. Similar findings have been reported for hydrogen bond model complexes (e.g., ClH:NH<sub>3</sub><sup>75,76</sup>).

The computed <sup>59</sup>Co chemical shifts of **12**, **12a**, **13**, and **13a** are summarized in Table 10. Employing Basis AE1 we compute an isotope effect  $\Delta\delta = \delta_{\text{eff}}(\text{Co,D}) - \delta_{\text{eff}}(\text{Co,H})$  of  $\Delta\delta = 65$  ppm for **12** and  $\Delta\delta = 66$  ppm for **13**. Employing Basis AE1(\*),  $\Delta\delta = 86$  ppm is obtained for compound **13**, in very good agreement with the experimental value of  $\Delta\delta \approx 50$  ppm.<sup>7</sup> Thus, both sign and magnitude of  $\Delta\delta$  are correctly reproduced by the perturbational approach, which thus offers a faithful qualitative description of this effect.

Upon going from H to D, the main change in the coordination geometry about Co is found for the Co–N distances to the glyoximate ligands, which increase on average by 0.003 Å (Table 9). If we suppose that the corresponding shift in  $\delta(^{59}\text{Co})$  is based only on the increase of the Co–N bond length, then we can arrive at an estimated shielding/bond-length derivative of  $\approx 72$  ppm pm<sup>-1</sup> per bond (using the  $\Delta\delta$  value of **13** with AE1(\*) basis). This value

lies within the typical range of  $\partial\sigma_{\text{Co}}/\partial r_{\text{Co-N}}$  derivatives (36–84 ppm/pm, see Table 3). Thus, the shift of  $\delta(^{59}\text{Co})$  due to the substitution of hydrogen by deuterium 3 atoms away from the metal center can be explained by an increase in the cobalt–ligand bond length due to a modified vibrational wave function, as anticipated during analysis of the experimental data.<sup>7</sup> According to our results, however, the extent of this bond elongation, well below 1 pm, is much less than estimated before.<sup>7</sup>

As a further test of the performance of the perturbational approach, we investigated the primary isotope effect,  $^p\Delta = \delta(\text{H}) - \delta(\text{D})$ . Our computed values of  $^p\Delta = +0.57$  ppm for **12**, Basis AE1 and  $^p\Delta = +0.49$  ppm for **13**, Basis AE(\*) agree very well with the experimental value of +0.38 ppm.<sup>7</sup> Thus, the task to reproduce such a small effect both in sign as well as in magnitude has been successfully accomplished.

## Conclusions

We have applied different DFT-based methods to investigate the  $\delta(^{59}\text{Co})$  chemical shifts of a set of cobalt complexes that span nearly the entire chemical shift range of this nucleus. For this purpose we have performed static NMR chemical shift calculations at optimized geometries as well as computations which include zero-point or classical thermal effects. Overall, zero-point effects and thermal averaging result in a deshielding of <sup>59</sup>Co nuclei and also in an increase of most  $\delta(^{59}\text{Co})$  values. To a large extent these downfield shifts can be rationalized in terms of the bond elongation of the cobalt–ligand bonds. However, it is only in some cases that zero-point or thermally averaged chemical shifts are improved over the respective static equilibrium values. In general, the mean absolute error of the computed chemical shifts increases slightly, by 70 to 200 ppm, upon inclusion of zero-point or thermal effects, where the largest errors are observed mainly for those complexes that were measured in aqueous solution. In these cases, solvation effects are expected to be substantial, effects which will be addressed in future work.

In addition, we have investigated three tautomeric forms of the binuclear cobalt octacarbonyl and have located, for the first time, salient transition states connecting these minima. These transition states are computed ca. 2 to 36

kJ/mol above any of the tautomers, consistent with the fluxionality of this molecule on the NMR time scale. The  $\delta(^{59}\text{Co})$  chemical shifts differ only slightly between the tautomeric forms and fall in a short-range less than 200 ppm, thereby preventing conclusions concerning the composition of the equilibrium mixture.

As further cobalt compounds we have investigated model cobaloximes, paying special attention to the effect of deuteration 3 bonds away from the metal center. Due to the changes in the vibrational wave function, the structure of the ( $\text{O}\cdots\text{H}-\text{O}$ )-system is altered, such that the O–D bond is 3 pm shorter than the former O–H bond. This in turn leads to a slightly larger Co–N bond separation, which causes a considerable downfield shift of the cobalt center. Both sign and magnitude of this isotope effect,  $\Delta\delta(^{59}\text{Co})$ , agrees well with the experimental values.

Quantitative predictions of  $^{59}\text{Co}$  chemical shifts, and those of transition metals in general, remain a challenge for approximate density functional theory, and the applied methods to include zero-point and classical thermal effects are no panacea and need further improvement, such as the inclusion of solvation effects. Nevertheless, the possibility to go beyond static molecules in  $^{59}\text{Co}$  chemical shift calculations can open new possibilities for applications of such computations, thus affording an ever more refined theoretical complement to experimental  $^{59}\text{Co}$  NMR spectroscopy.

**Supporting Information Available:** Harmonic vibrational frequencies at BP86/AE1 level of **1a–1c** (Table S1) and plot of computed - GIAO-B3LYP for BP86/AE1 optimized geometries - versus experimental elements of the chemical shielding tensor (Figure S1). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- Proctor, W. G.; Yu, F. C. *Phys. Rev.* **1951**, *81*, 20.
- Pregosin, P. S. In *Transition Metal Nuclear Magnetic Resonance*; Pregosin, P. S., Ed.; Elsevier: Amsterdam, 1991; p 144.
- Yamasaki, A. *J. Coord. Chem.* **1991**, *24*, 211.
- Chan, J. C. C.; Au-Yeung, S. C. F. *Ann. Rep. NMR Spectrosc.* **2000**, *41*, 1.
- von Philipsborn, W. *Chem. Soc. Rev.* **1999**, *28*, 95.
- Medek, A.; Frydman, V.; Frydman, L. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14237.
- Asaro, F.; Liguori, L.; Pellizer, G. *Angew. Chem., Int. Ed.* **2000**, *39*, 1932.
- Medek, A.; Frydman, L. *J. Am. Chem. Soc.* **2000**, *122*, 684.
- Pellizer, G.; Asaro, F.; Pergolese, B. *Magn. Reson. Chem.* **2004**, *42*, 756.
- Kaupp, M.; Malkin, V. G.; Malkina, O. L. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Kollman, P. A., Clark, T., Schaefer, H. F., Gasteiger, J., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; Vol. 3, p 1857.
- Schreckenbach, G.; Ziegler, T. *Theor. Chem. Acc.* **1998**, *99*, 71.
- Bühl, M. In *Calculation of NMR and EPR Parameters. Theory and Applications*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH: Weinheim, 2004; p 421.
- Bühl, M.; Kaupp, M.; Malkin, M.; Malkina, V. G. *J. Comput. Chem.* **1999**, *20*, 91.
- Godbout, N.; Oldfield, E. *J. Am. Chem. Soc.* **1997**, *119*, 8065.
- Chan, J. C. C.; Au-Yeung, S. C. F.; Wilson, P. J.; Webb, G. A. *J. Mol. Struct. (THEOCHEM)* **1996**, *365*, 125.
- Chan, J. C. C.; Au-Yeung, S. C. F. *J. Mol. Struct. (THEOCHEM)* **1997**, *393*, 93.
- Chan, J. C. C.; Au-Yeung, S. C. F. *J. Phys. Chem.* **1997**, *101*, 3637.
- Bühl, M. *Chem. Phys. Lett.* **1997**, *267*, 251.
- Sundholm, D.; Gauss, J.; Schäfer, A. *J. Chem. Phys.* **1996**, *105*, 11051.
- Böhm, M. C.; Schulte, J.; Ramirez, R. *Int. J. Quantum Chem.* **2002**, *86*, 28.
- Ruud, K.; Åstrand, P.-O.; Taylor, P. R. *J. Chem. Phys.* **2000**, *112*, 2668.
- Ruud, K.; Åstrand, P.-O.; Taylor, P. R. *J. Am. Chem. Soc.* **2001**, *123*, 4826.
- Ruden, T.; Lutnæs, O. B.; Helgaker, T. *J. Chem. Phys.* **2003**, *118*, 9572.
- Ruden, T. A.; Ruud, K. In *Calculation of NMR and EPR Parameters: Theory and Applications*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH: Weinheim, 2004; p 153.
- Malkin, V. G.; Malkina, O. L.; Steinebrunner, G.; Huber, H. *Chem. Eur. J.* **1996**, *2*, 452.
- Searles, R. D. I.; Huber, H. In *Calculation of NMR and EPR Parameters. Theory and Applications*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH: Weinheim, 2004; p 175.
- Bühl, M.; Parrinello, M. *Chem. Eur. J.* **2001**, *7*, 4487.
- Bühl, M.; Mauschick, F. T. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5508.
- Bühl, M.; Schurhammer, R.; Imhof, P. *J. Am. Chem. Soc.* **2004**, *126*, 3310.
- Bühl, M.; Mauschick, F. T. *Magn. Reson. Chem.* **2004**, *42*, 737.
- Grigoleit, S.; Bühl, M. *Chem. Eur. J.* **2004**, *10*, 5541.
- The value for single H/D substitution is given in ref 7 as  $\Delta\delta = 25$  ppm; in the plotted spectrum,  $\Delta\delta \approx 55$  ppm can be estimated for double substitution.
- Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.
- Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406.
- Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033.
- Hay, P. J. *J. Chem. Phys.* **1977**, *66*, 4377.
- Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497.
- Gaussian98 (revision A.7). Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratman, R.

- E.; Burant, J. C.; Dapprich, S.; Milliam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, C.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Chen, B. G. J. W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian, Inc., Pittsburgh, PA, 1998.
- (40) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (41) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (42) Kutzelnigg, W.; Fleischer, U.; Schindler, M. *NMR: Basic Principles and Progress*; 1990; Vol. 23, p 165.
- (43) Car, R.; Parrinello, P. *Phys. Rev. Lett.* **1985**, *55*, 2471.
- (44) CPMD Version 3.3a. Hutter, J.; Alavi, A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckermann, M.; Parrinello, M. Max-Planck-Institut für Festkörperforschung (Stuttgart) and IBM Research Laboratory (Zürich), 1995–1999.
- (45) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993.
- (46) Kleinman, L.; Bylander, D. M. *Phys. Rev. Lett.* **1982**, *48*, 1425.
- (47) Bühl, M. *J. Phys. Chem. A* **2002**, *106*, 10505.
- (48) Åstrand, P.-O.; Ruud, K.; Sundholm, D. *Theor. Chem. Acc.* **2000**, *103*, 365.
- (49) Dalton, a molecular electronic structure program, release 1.2. Helgaker, T.; Jensen, H. J. A.; Jørgenson, P.; Olsen, J.; Ruud, K.; Ågren, H.; Auer, A. A.; Bak, K. L.; Bakken, V.; Christiansen, O.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Hättig, C.; Hald, K.; Halkier, A.; Heiberg, H.; Hettema, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Pedersen, T. B.; Ruden, T. A.; Sanchez, A.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sylvester-Hvid, K. O.; Taylor, P. R.; Vahtras, O. 2001.
- (50) Bühl, M.; Imhof, P.; Repisky, M. *Chem. Phys. Chem.* **2004**, *5*, 410.
- (51) Bühl, M.; Gaemers, S.; Elsevier: C. *J. Chem. Eur. J.* **2000**, *6*, 3272.
- (52) Bühl, M. *Theor. Chem. Acc.* **2002**, *107*, 336.
- (53) Bramley, R.; Brorson, M.; Sargeson, A. M.; Schäffer, C. E. *J. Am. Chem. Soc.* **1985**, *107*, 2780.
- (54) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: 2000.
- (55) Quantum chemical zero-point corrections essentially preserve the symmetries of the equilibrium geometries.
- (56) Isolated **6** is not stable in the gas phase, where it would ionize spontaneously. In the calculations with atom-centered Gaussian-type basis sets, it is the finite size of the latter that prevents electron detachment. In the plane-wave calculations, the same is achieved by the specific treatment of the electrostatic potentials under the cluster boundary conditions, where compensating background charges are involved. No artifacts pointing to such a spontaneous electron detachment are encountered in the CPMD calculations.
- (57) Jameson, C. J.; Rehder, D.; Hoch, M. *J. Am. Chem. Soc.* **1987**, *109*, 2589.
- (58) Minaev, B.; Vaara, J.; Ruud, K.; Ågren, H. *Phys. Lett.* **1998**, *295*, 455.
- (59) Crompton, B.; Carrington, T.; Salahub, D. R.; Malkina, O. L.; Malkin, V. G. *J. Chem. Phys.* **1999**, *110*, 7153.
- (60) A referee suggested to apply some empirical scaling to the metal–ligand bond distances before computing the chemical shifts. While such a procedure may be feasible for highly symmetrical complexes, we have not attempted to do this, because it is not clear how to implement it in a general, rigorous way without impeding the predictive power.
- (61) Accurate computations of EFGs require large basis sets and sophisticated treatments of electron correlation and relativity, see e.g. Schwerdtfeger, P.; Pernpointner, M.; Nazarewicz, W. In *Calculation of NMR and EPR Parameters. Theory and Applications*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH: Weinheim, 2004; p 279.
- (62) *The Principles of Nuclear Magnetism*; Abragam, A., Ed.; Oxford University Press: Oxford, 1961.
- (63) *C-13-NMR Spectroskopie*; Kalinowski, H.-O., Berger, S., Braun, S., Eds.; Thieme: Stuttgart, 1984.
- (64) Sumner, G. G.; Klug, H. P.; Alexander, L. E. *Acta Crystallogr.* **1964**, *17*, 732.
- (65) Sweany, R. L.; Brown, T. L. *Inorg. Chem.* **1977**, *16*, 415.
- (66) Kenny, J. P.; King, R. B.; Schaefer, H. F., III *Inorg. Chem.* **2001**, *40*, 900.
- (67) Folga, E.; Ziegler, T. *J. Am. Chem. Soc.* **1993**, *115*, 5169.
- (68) Bühl, M. In *Encyclopedia of Computational Chemistry*; Schleyer, v. R. P., Schreiner, P. R., Alhinger, N. L., Clark, T., Gasteiger, J., Kollman, P., III, H. F. S., Eds.; Wiley: 1996; p 1835.
- (69) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154.
- (70) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523.
- (71) Mooberry, E. S.; Pupp, M.; Slater, J. L.; Sheline, R. K. *J. Chem. Phys.* **1971**, *55*, 3655.
- (72) Stryer, L. *Biochemistry*, 4th ed.; Freeman: New York, 1995.
- (73) Tavagnacco, C.; Balducci, G.; Costa, G.; Taschler, K.; von Philipsborn, W. *Helv. Chim. Acta* **1990**, *73*, 1469.
- (74) Altman, L. J.; Laungani, D.; Gunnarsson, G.; Wennerström, H.; Forsén, S. *J. Am. Chem. Soc.* **1978**, *100*, 8264.
- (75) Jordan, M. J. T.; Del Bene, J. E. *J. Am. Chem. Soc.* **2000**, *122*, 2101.
- (76) Del Bene, J. E.; Jordan, M. J. T. *J. Phys. Chem. A* **2002**, *1–6*, 5385.
- (77) McNeill, E. A.; Scholer, F. R. *J. Am. Chem. Soc.* **1977**, *99*, 6243.
- (78) Klüfers, P. *Z. Kristallogr.* **1984**, *167*, 253.
- (79) Xie, X.; Au-Yeung, S. C. F.; Liu, H. *J. Mol. Struct. (THEOCHEM)* **1995**, *331*, 181.
- (80) Barclay, G. A.; Hoskins, B. F. *J. Chem. Soc.* **1962**, 586.
- (81) Beatti, J. K.; Best, S. P. *Coord. Chem. Rev.* **1997**, *166*, 391.
- (82) Lucken, E. A. C.; Noack, K.; Williams, D. F. *J. Chem. Soc. (A)* **1967**, p 148.
- (83) Benn, R.; Cibura, K.; Hoffmann, P.; Jonas, K.; Rufinska, A. *Organometallics* **1985**, *4*, 2214.

- (84) Spiess, H. W.; Haas, H.; Hartmann, H. *J. Chem. Phys.* **1969**, *50*, 3057.
- (85) Juranić, N. *Inorg. Chem.* **1983**, *22*, 521.
- (86) Juranić, N.; Čelap, M. B.; Vučelić, D.; Malinar, M. M.; Radivojša, P. N. *Spectrochim. Acta* **1979**, *35A*, 997.
- (87) Asaro, F.; Liguori, L.; Pellizer, G. *Phys. Chem. Chem. Phys.* **1999**, *1*, 4981.
- (88) Leung, P. C.; Coppers, C. *Acta Crystallogr.* **1983**, *B39*, 535.
- (89) The chemical shift was obtained by using the mean computed <sup>13</sup>C shielding of Fe(CO)<sub>5</sub> as a reference and the experimental chemical shift of the latter ( $\delta = 211.0$ )[63] for conversion.

CT0499200

## Understanding the Cosolvation Effect of Dendrimers

Gilberto Teobaldi, Manuel Melle-Franco, and Francesco Zerbetto\*

*Dipartimento di Chimica “G. Ciamician”, Università di Bologna,  
V. F. Selmi 2, 40126 Bologna, Italy*

Received September 23, 2004

**Abstract:** Molecules that are virtually insoluble in certain solvents may be uploaded to “hostile” phases by dendrimers. Prime examples of this phenomenon are Eosin Y, EY, and Rose Bengal, RB, that are not soluble in  $\text{CH}_2\text{Cl}_2$  where they can, however, be solvated through the interaction with a fourth generation dendrimer of polypropylene amine, POPAM-4D. The two dyes share the same carbon framework and differ for the pattern of halogenation, and yet their cosolvation varies over a factor of 4: six Eosin Y and  $\sim 25$  Rose Bengals are solvated by the macromolecule. Leveraging on a previous report where molecular dynamics simulations of 12 EY@POPAM-4D in  $\text{CH}_2\text{Cl}_2$  showed a reduction to the experimental limit of 6, we now perform similar calculations with an excess, i.e., 40, of RB@POPAM-4D. The simulations quantitatively reproduce the cosolvation effect. They also provide a microscopic understanding of its origin and of motions-interactions of the macromolecule and both of its guests.

### Introduction

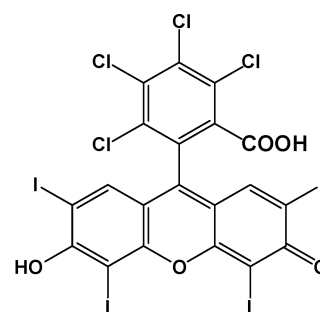
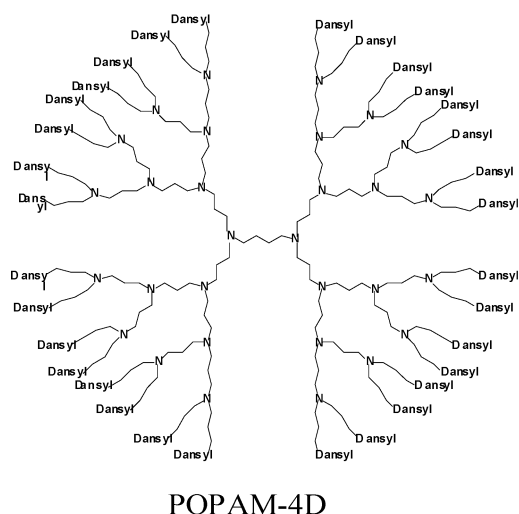
One of the several, potentially important, applications of dendrimers<sup>1–6</sup> is the encapsulation of molecules. An important practical use of the entrapment could be transporting and delivering drugs to specific targets inside the body.<sup>7</sup> To entrap a guest, the macromolecule must interact better with the guest than the surrounding environment. In practice, the carrier must be more “attractive” to the guest than the outside world. Another application of the same phenomenon is the increase of solubility, or even the solvation of molecules in solvents where it is usually not possible. Peculiarly, small variations of the chemical structure of the guests lead to substantial variations of their average number “inside” dendrimers.<sup>8</sup> The case of a 4-th generation polypropylene amine dendrimer, in short POPAM-4D, and two rather similar dyes is illustrative. The dyes are water-soluble and  $\text{CH}_2\text{Cl}_2$  insoluble; POPAM-4D is  $\text{CH}_2\text{Cl}_2$  soluble but is water insoluble. When an aqueous solution of the dyes is shaken with an organic solution of the dendrimer, up to 6 Eosin Y molecules or  $\sim 25$  molecules of Rose Bengal are transferred to  $\text{CH}_2\text{Cl}_2$ .<sup>8</sup> All the molecular structures are given in Scheme 1. Eosin Y and Rose Bengal share the same carbon framework and differ for the substitution of 4 hydrogens with

4 chlorines, and 4 bromines with 4 iodines, with Rose Bengal heavier by  $\sim 50\%$ . And yet, POPAM-4D uploads four times as many heavier molecules.<sup>8</sup> *The relatively small structural change varies by half an order of magnitude the cosolvation effect of the dendrimer!*

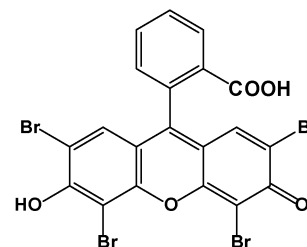
Recently,<sup>9</sup> we studied the dynamics of this dendrimer with a varying number of eosins Y. The intent was to verify if computational methods could reproduce and provide insight into the 1:6 POPAM-4D:Eosin Y ratio. The initial “computer” encapsulation of twelve dyes evolved in less than a nanosecond to a stable arrangement with only six guests, while the other six fled away in the solvent.<sup>9</sup> In practice, the host cannot harbor more than six Eosin Y’s and since this dye is not  $\text{CH}_2\text{Cl}_2$ -soluble, this is the maximum number that can be uploaded to the organic phase per dendrimer molecule. The work focused mainly on dynamical aspects of the guest–host system and recorded a multitude of entrances and exits in the macromolecule. The origin of the hustle and bustle of molecules around the dendrimer is the “unfriendly” solvent environment. Low solubility arises – although not only – from weak intermolecular interactions with the solvent, which imply high mobility in it. If, after departure from POPAM-4D, a dye comes again in contact with the macromolecule, favorable stabilizing interaction forces “suck” it in. The molecule may then displace another

\* Corresponding author fax: +39 051 2099473; e-mail: francesco.zerbetto@unibo.it.

Scheme 1



Rose Bengal



Eosin Y

guest, or, alternatively, if the closest guest molecule is solidly embedded in the macromolecule, it may be expelled again.

The ability of a simulation to reproduce the 4-fold increase of concentration of dye molecules caused by relatively small structural changes would broaden our understanding of the guest–host dynamics of the dendrimer (and strongly vouch for the accuracy of the computational model). Through the comparison of the results for the two types of guests, one could also understand the origin of the very different cosolvation effect obtained for the two structurally similar dyes. It is believed that the elucidation of the structural and dynamical properties of dendrimers and their guest–host systems can be partly achieved using computer modeling.<sup>10</sup>

## Results and Discussion

The flexible branches that characterize the majority of dendrimers can rearrange to form a myriad of conformations. Their interconversion is rapid in a liquid, and this exchange can be partly transmitted to the solid, making the structural characterization very difficult (see, however, ref 11). Here we attempt to assist the process of developing a quantitative understanding of the geometrical and dynamical properties of their guest–host systems and of the cosolvation effect in particular. Before presenting the results of the simulations, it is worth summarizing the experiments. Aqueous solutions of water-soluble dyes that are not soluble in  $\text{CH}_2\text{Cl}_2$  are shaken together with a  $\text{CH}_2\text{Cl}_2$ +dendrimer solution. The dyes move from acidic water,  $\text{pH}=5$ , whose color fades, the  $\text{CH}_2\text{Cl}_2$  solution, which becomes colored. The concentration of the dendrimer in  $\text{CH}_2\text{Cl}_2$  is less than  $2.5 \times 10^{-6}$  M. The concentration of the dyes in water is  $7.5 \times 10^{-5}$  M. Aggregation and dishomogeneity due to concentration should therefore be unimportant on long time scales, although they may occur locally either at or around the dendrimer. The

measurements indicate that each molecule of dendrimer can interact, on average, with up to six eosin Y and  $\sim 25$  Rose Bengal molecules.

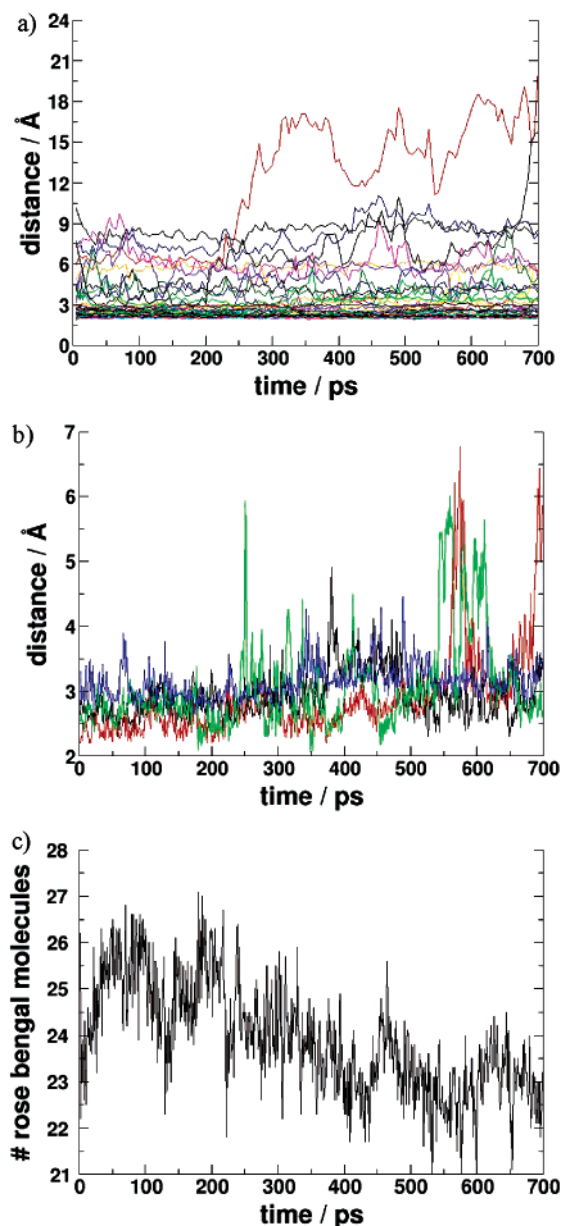
In analogy with what done for the case of Eosin Y, we performed new molecular dynamics, MD, simulations of the dendrimer in  $\text{CH}_2\text{Cl}_2$  solution with an excess number, i.e., 40, of molecules of Rose Bengal. The present comparison of the two sets of simulations for Eosin Y and Rose Bengal implies the assumption that the equilibria in water play a similar role for the two dyes.

The previous simulations with Eosin Y's<sup>9</sup> showed that the excess of twelve dyes is rapidly reduced to the experimental limit of six. Rose Bengal behaves similarly. The 40 guests swarm around the dendrimer, Figure 1a, establishing and breaking contact several times. A clearer view of the behavior of selected few molecules is presented in Figure 1b.

The average number of guests in contact with the hyperbranched molecule is shown in Figure 1c. To assist the eye, the numbers represent averages over 5 ps of dynamics. Despite the highly dynamical evolution characterized by a series of departures and arrivals, the macromolecule is able to host between 21 and 27 Rose Bengals during the whole simulation.

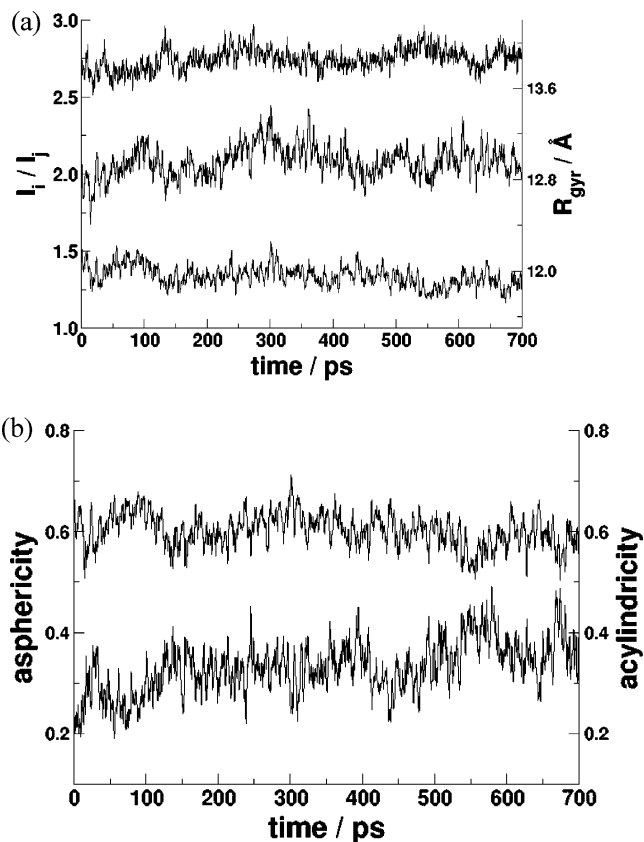
The modeling reproduces the experimental observation of a substantial difference in the interaction with the macromolecule upon relatively small structural changes of the host. The calculations may also provide complementary information about three main aspects:

1. the shape of the macromolecule, and its variation, during the dynamics in the absence and in the presence of the guests,
2. the solvent molecules dynamics inside and around the macromolecule, both in the absence and in the presence of the guests,
3. the dynamics of the guest molecules.



**Figure 1.** Simulation of an excess of Rose Bengal molecules in a  $\text{CH}_2\text{Cl}_2$  box with POPAM-4D: (a) instantaneous shortest distance between an atom of each of the 40 Rose Bengals and one of the dendrimer; (b) zoom on four of the guests, which show multiple entrances and exits from contact with the dendrimer; (c) average number over 5 ps of the Rose Bengal molecules in contact with the dendrimer.

**Shape Variation of POPAM-4D.** In solvents, dendrimers tend to have collapsed, compact conformations<sup>12</sup> that are characterized by high internal mobility. The presence of guests may affect the conformational equilibria. Figure 2a shows the instantaneous ratios of the largest moment of inertia divided by either one of the other two, together with the gyration radius. A spherical, static dendrimer would have both ratios constant and equal to one; a cylindrical dendrimer would have only one of the ratios equal to one, see computational section for details. Figure 2b displays the deviations from perfect spherical and cylindrical shapes as a function of time.

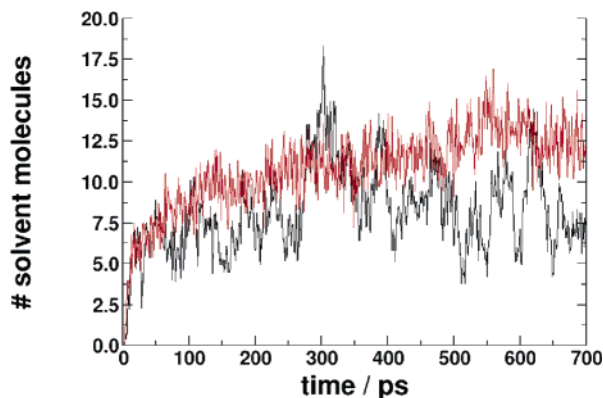


**Figure 2.** Dynamics of POPAM-4D in  $\text{CH}_2\text{Cl}_2$  solution in the presence of an excess of Rose Bengal molecules: (a) instantaneous ratios of moments of inertia  $I_1/I_2$  and  $I_1/I_3$ , with  $I_1 > I_2 > I_3$ , the top curve is the gyration radius; (b) top, deviation from spherical shape, bottom, deviation from cylindrical shape.

During the dynamics  $I_1/I_2$  ranges between 1.25 and 1.5, while  $I_1/I_3$  ranges between 1.7 and 2.5: the macromolecule is hardly spherical. The average values of  $r_{\text{gyr}}$  give additional information. For POPAM-4D, EY@POPAM-4D, and RB@POPAM-4D, they were 17.17, 15.93 and 13.86 Å with largest instantaneous deviations from the average values of 2.23, 1.27, and 0.40 Å. Therefore the guests tend to make the macromolecule more compact. Because of the lower  $r_{\text{gyr}}$ , one of the results of the presence of Rose Bengal is the effective decrease of the size and number of the pockets that can encapsulate the guests inside the macromolecule. It can be surmised that the increased solubility of RB must have an origin different from the direct interaction between guest and host. So far as one can think of a dendrimer as a box, the presence of guests effectively close the dendrimer box (It may even be thought that the molecules pull the door with them after entering!).

The nonspherical shape of the macromolecule appears in the top line of Figure 2b. In the same order as above (POPAM-4D, EY@POPAM-4D, RB@POPAM-4D), the average values are 0.62, 0.50, 0.60 with largest instantaneous deviations of 0.19, 0.19, 0.11. The macromolecules are therefore hardly spherical. A better description of the dendrimer is as a cylinder. Indeed, the values of departure from the perfect cylindrical shape are small and are 0.36, 0.27, 0.33. The relative largest instantaneous deviations are





**Figure 3.** Number of solvent molecules, averaged over 5 ps, around POPAM-4D in  $\text{CH}_2\text{Cl}_2$  solution in the presence of an excess of Rose Bengal molecules: the red line is the number of solvent molecules with at least one atom less than 2.8 Å away from an atom of the dendrimer, and the black line is the number of solvent molecules inside  $r_{\text{gyr}}$ .

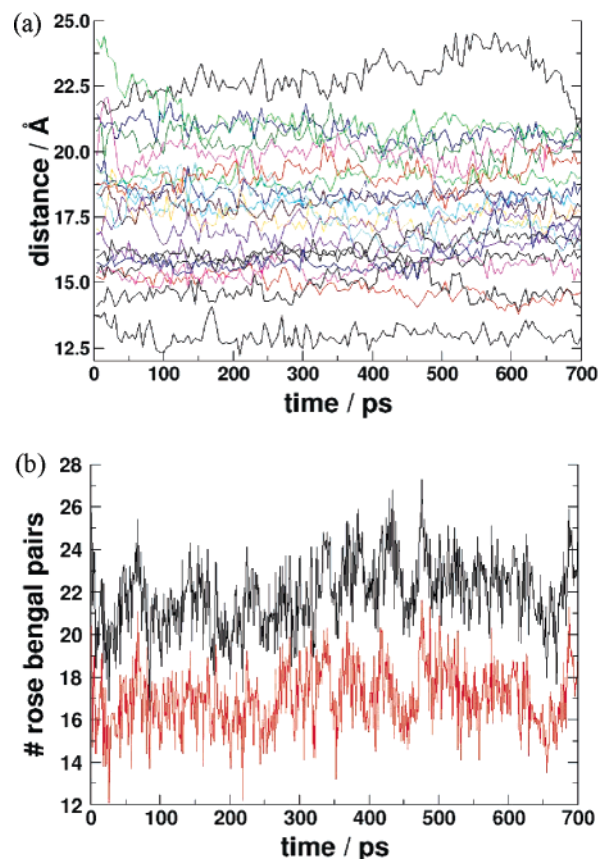
0.29, 0.22, 0.16. Overall, it appears that the general shape of empty POPAM-4D and RB@POPAM-4D departs from that of EY@POPAM-4D.

The conclusion of this subsection is that the presence of the guests is a way to manipulate the structure of the dendrimer, which becomes more cylindrical and denser.

**Solvent Dynamics inside POPAM-4D.** In solution, the free dendrimer has  $\sim 30$   $\text{CH}_2\text{Cl}_2$  molecules in contact with it at any one time. They can be either on the surface or inside. The formation of the guest–host system with 6 Eosin Y decreases them to about 24. Figure 3 shows that in the case of Rose Bengal, they undergo a further drop to  $\sim 12$ . Because of the way the simulation is set up,  $\text{CH}_2\text{Cl}_2$  molecules are not initially present inside POPAM-4D. However, the solvent molecules penetrate very rapidly inside the dendrimer and, in 200 ps, reach an average number of 12. The process corresponds to a rather fast equilibration: during this time the dynamics also reaches a near constant value of the potential energy. This percolation process does not correspond to a real physical process. It is, however, useful to understand the rapidity of the solvent movements inside the dendrimer. In the case of Eosin Y, the same equilibration, with more molecules was reached in  $\sim 60$  ps. The different time frames agree with the more compact structure of POPAM-4D in the presence of Rose Bengal, which, in turn, makes movements inside it slower.

The solvent molecules inside  $r_{\text{gyr}}$  tend to be fewer than those in contact with the hyper-branched system. The trend is opposite to that recorded when the guests are Eosin Y's.<sup>9</sup> Although  $r_{\text{gyr}}$  does not perfectly match the surface of the dendrimer (the nonspherical shape makes the radius include regions where there may not be dendrimer atoms and exclude zones where they are still present), the inversion with respect to the case of Eosin Y signals that Rose Bengal displaces  $\text{CH}_2\text{Cl}_2$  in the neighborhood of the surface (see below).

**Dynamics of the Guest Molecules.** While the number of “permanent” guests at POPAM-4D is roughly constant, their position is hardly stationary. Figures 4a and 4b give more details. Figure 4a shows the evolution of the distances



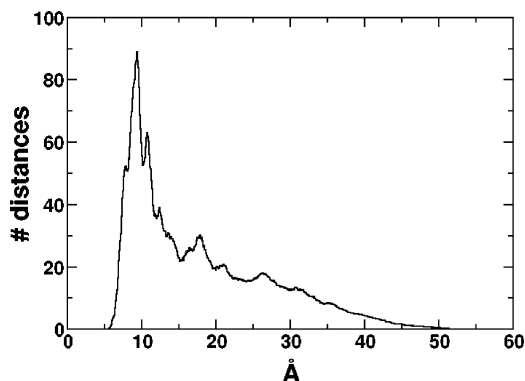
**Figure 4.** (a) Instantaneous Rose Bengals-POPAM-4D center-to-center of mass distance for the dyes in contact with the dendrimer; (b) Rose Bengal physical dimerization: in black the total number of pairs, in red the number of dimers in contact with the dendrimer.

between the centers of mass of the guest Rose Bengals and that of the dendrimer. At odds with the case of the Eosins where two molecules resided in the inner core of the dendrimer (5–10 Å from the center of mass),<sup>9</sup> RB's distribute rather evenly. This agrees with the more tightly packed structure of the macromolecule, which does not allow the dyes to percolate inside it although they do show a substantial mobility.

Figure 4b shows that a number of guest pairs (or even trimers) are instantaneously formed. Some of the pairs permanently leave the macromolecule. The physical dimerization of the guests indicates that in  $\text{CH}_2\text{Cl}_2$ , Rose Bengal prefers to interact with itself rather than with the solvent.

**Origin of the Cosolvation.** The simulations satisfactorily describe the different nature of the guest–host systems of POPAM-4D with the two dyes and reproduce their different amount of solvated molecules. A single bottom line able to explain the larger solvation of RB's is, however, still required.

To obtain it, we calculated the average energies, over time, of guest–host, guest–solvent, and guest–guest interactions for each one of the dyes of 12EY@POPAM-4D and 40RB@POPAM-4D. A table with the results is given in the Supporting Information. Both  $\langle V_{\text{guest–host}} \rangle$  and  $\langle V_{\text{guest–solv}} \rangle$  resulted quite similar, while  $\langle V_{\text{guest–guest}} \rangle$  is substantially larger for Rose Bengal. Indeed, the average of the time-averages



**Figure 5.** Radial distribution of the distance of the centers of mass of the Rose Bengals. The peak of the dimers  $\sim 10$  Å is a doublet due to the presence of two types of aggregates.

of the interactions for the two dyes are

$$\text{for EY, } \langle V_{\text{guest-host}} \rangle = -13.9; \langle V_{\text{guest-solv}} \rangle = -35.0; \langle V_{\text{guest-guest}} \rangle = -4.37 \text{ kcal mol}^{-1}$$

$$\text{for RB, } \langle V_{\text{guest-host}} \rangle = -12.3; \langle V_{\text{guest-solv}} \rangle = -32.6; \langle V_{\text{guest-guest}} \rangle = -25.3 \text{ kcal mol}^{-1}$$

It appears that the most remarkable difference between the two dyes arises from  $\langle V_{\text{guest-guest}} \rangle$ , with a factor of 5–6 in favor of Rose Bengal. Even discounting the larger number of Rose Bengals in the simulations, it can be concluded that the origin of the larger cosolvation effect for this dye arises from its tendency to form pairs in  $\text{CH}_2\text{Cl}_2$ . In practice, POPAM-4D solvates some RB molecules, which subsequently assist the solvation of other molecules giving a true collective cosolvation effect. Interestingly, the radial distribution of the relative distances of the centers of mass of the Rose Bengals, Figure 5, shows that the dimers are of two types. The more abundant type is due to the formation of a hydrogen bond between the carboxylic hydrogen and a carbonyl located between the two Iodine atoms, see Scheme 1. The less abundant dimer is due to  $\pi$ -stack formation, which is hindered by the nonplanar conformation of the molecule.

To support further the results, we also estimated the free energy of cavitation in  $\text{CH}_2\text{Cl}_2$ <sup>17f</sup> which resulted in 22.8 kcal mol<sup>-1</sup> for the monomer and 37.8 kcal mol<sup>-1</sup> for the dimer. The sum of electrostatic and van der Waals terms gives  $-69.9$  kcal mol<sup>-1</sup> for the monomer and  $-74.6$  kcal mol<sup>-1</sup> for the dimer. Solubilizing a RB dimer involves a stabilization of  $(37.8 - 2 \times 74.6) - 2 \times (22.8 - 69.9) = -17.2$  kcal mol<sup>-1</sup>.

Finally, we calculated the rates of entrance-exits or better contact formation and breaking,  $k_{e-e}$ , of the guest molecules using a time-honored approach.<sup>13</sup> As a rule, the kinetic constants are related to thermodynamics only when the forward and backward processes are considered at the same time. However, some general information can still be acquired from their analysis. Only molecules that spent between 30 and 70% of their time at POPAM-4D were considered. The results must be considered semiquantitative. For EY,  $k_{ee} = 1.5 \times 10^{11} \text{ s}^{-1}$ , while for RB,  $k_{ee} = 4.4 \times 10^{11} \text{ s}^{-1}$ . These high rates are compatible with molecular

motions in solvents and imply that the behavior of the macromolecule with respect to its guest is analogous to that of a real solvent. The larger rate for RB originates from the presence of dye molecules external to the dendrimer, which facilitate the extraction of the guests (see below).

## Conclusion

MD calculations find that POPAM-4D is a flexible system where a large number of solvent molecules move nearly freely. Introduction of extra dye molecules in a thought-experiment performed using Molecular Dynamics shows that they tend to leave the macromolecule, first tentatively with several returns and then permanently, until they reach the experimental limit of six for Eosin Y and  $>20$  for Rose Bengal. The presence of the guest molecules compacts the dendrimer structure and dislodges the solvent molecules inside it. The guest molecules can still move inside and around POPAM-4D substantially, but Rose Bengal tends to form pairs. They swarm in and out of the dendrimer very rapidly, with similar rate constants, of the order of what is expected for solvent–solute interactions. It is the tendency of Rose Bengal molecules to undergo physical dimerization, which is due to their higher interaction energy, that ultimately makes POPAM-4D extract more than 20 of them from the aqueous solution, compared to the lower number reported for Eosin Y.

## Computational Details

All the calculation were performed with the MM3 force field that has been found to be accurate for organic systems<sup>14</sup> and is parametrized explicitly to describe the  $\pi$ - $\pi$  stacking interactions that govern a sizable part of the interactions in this system. The parametrization of the S–C and S–N torsional potential energy curves of the sulfonamide of the dansyl units that terminate the branches of POPAM-4D was presented before.<sup>9</sup> The molecular dynamics calculations were run at constant volume and at a temperature equal to 300 K, using periodic boundary conditions (PBC), with the Tinker program.<sup>15</sup> The temperature was kept constant using the algorithm proposed by Berendsen et al.<sup>16</sup> Periodic boundary conditions were used throughout with a cubic box with a linear dimension of 80.58 Å. The approach based on MM3 implemented in the Tinker program has been rather successful in a variety of applications carried out by our group.<sup>17</sup> As in the previous work, solvent molecules of  $\text{CH}_2\text{Cl}_2$  were treated explicitly as dipoles.<sup>9</sup> Optimized geometries of POPAM-4D and 40 Rose Bengal molecules were inserted in an equilibrated box of solvent molecules. All solvent molecules with center of mass less than 4.2 Å away from any atom (hydrogens included) of the dendrimer were removed. The 4.2 Å value was determined by trial and error to replace an equal number of atoms, which makes for the most stable simulation box. This leads to a system of 3880  $\text{CH}_2\text{Cl}_2$  molecules plus the guest–host system. Equilibration was performed in the NVT ensemble at 300 K until the potential energy converged. During the simulation a plateau for the total energy is reached (see Figure 1 in the Supporting Information). This condition is not sufficient per se to prove that the guest–host arrangement is stable, but it shows that

the guest dynamics do not involve strong nonequilibrium conditions. We also verified that the rotational autocorrelation function of the molecules of Rose Bengal that form and break contacts with the dendrimer gives a lifetime shorter than the simulation time. The correlation time for these motions is shorter than 200 ps (see Figure 2 in the Supporting Information). This is less than one-third of the time of the simulations and is the time needed by the guests, which wander the most, to explore the different interactions both with the macromolecule and the solvent. A simulation three times as long should acquire sufficient statistics.

To analyze quantitatively the results, one can use two distances. The first is the gyration radius,  $r_{\text{gyr}}$ ,

$$r_{\text{gyr}} = \left[ \frac{\sum_i^{N_d} m_i (r_i - r_{\text{CM},d})^2}{M_d} \right]^{1/2} \quad (1)$$

where  $m_i$  and  $r_i$  run over all the masses and the positions of the atoms of the dendrimer,  $r_{\text{CM},d}$ , is the location of the center of mass of the dendrimer, and  $M_d$  is the dendrimer mass; the second and more relevant distance is the shortest distance between an eosin atom and a dendrimer atom. Two other useful quantities are the departure from perfect spherical shape,  $b$ , and from perfect cylindrical shape,  $c$ ,

$$b = I_1^2 - \frac{1}{2}(I_2^2 + I_3^2) \quad (2)$$

$$c = I_2^2 - I_3^2 \quad (3)$$

where the three components of moments of inertia,  $I$ , are  $I_1 \geq I_2 \geq I_3$ ; in the ideal shape, both  $b$  and  $c$  are zero.

Finally, the evaluation of the rates of multiple entrance-exits was calculated using the time correlation model of Allen and Schofield.<sup>13</sup>

**Acknowledgment.** We would like to thank Professor Vincenzo Balzani and Drs. Paola Ceroni and Veronica Vicinelli for constant encouragement and useful discussions on the topic of this work. Financial support from MURST and Università di Bologna (Funds for selected topics) is greatly appreciated.

**Supporting Information Available:** A file with 1) a figure with the energy in time during the simulation; 2) the autocorrelation function of the rotational motions of the RB that move the most during the simulation; and 3) a table with quantitative details of the energies of guest–host, guest–solvent, and guest–guest interactions of 12EY@POPAM-4D and 4ORB@POPAM-4D and two animations with EY@POPAM-4D and RB@POPAM-4D where the solvent molecules have been removed. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Newkome, G. R.; Charles, M. N.; Vogtle, F. *Dendrimers and Dendrons*, Wiley-VCH: Weinheim, 2001.

- (2) Tomalia, D. A.; Hedstrand, D. M.; Ferritto, M. S. *Macromolecules* **1991**, *24*, 1435–1438.
- (3) (a) Frechet, J. M.; *Science* **1994**, *263*, 1710–1715. (b) Frechet, J. M.; *Science* **1995**, *269*, 1080–1083.
- (4) Zimmerman, S. C.; Zeng, F.; Richert, D. E. C.; Kolotuchin, S. V.; *Science* **1996**, *271*, 1095–1098.
- (5) (a) Bell, T. W. *Science* **1996**, *271*, 1077–1078. (b) Li, H.; Kang, D.-J.; Blamire, M. G.; Huck, W. T. S. *Nano Lett.* **2002**, *2*, 347–349.
- (6) (a) Auletta, T.; Dordi, B.; Mulder, A.; Sartori, A.; Onclin, S.; Bruinink, C. M.; Peter, M.; Nijhuis, C. A.; Beijleveld, H.; Schoenherr, H.; Vansco, G. J.; Casnati, A.; Ungaro, R.; Ravoo, B. J.; Huskens, J.; Reinhoudt, D. N.; *Angew. Chem., Int. Ed.*, **2004**, *43*, 369–373. (b) van Hest, J. C. M.; Delnoye, D. A. P.; Baars, M. W. P. L.; van Genderen, M. H. P.; Meijer, E. W. *Science* **1995**, *268*, 1592–1595. (c) Johan, F. G. A.; Jansen, J.; Meijer, E. W.; de Brabander-van den Berg, E. M. M. *J. Am. Chem. Soc.* **1995**, *117*, 4417–4418. (d) Boas, U.; Sontjens, S. H. M.; Jensen, K. J.; Christensen, J. B.; Meijer, E. W. *ChemBioChem*, **2002**, *3*, 433–439. (e) Elemans, J. A. A. W.; Boerakker, M. J.; Holder, S. J.; Rowan, A. E.; Cho, W.-D.; Percec, V.; Nolte, R. J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5093–5098. (f) Chen, S.; Yu, Q.; Li, L.; Boozer, C. L.; Homola, J.; Yee, S. S.; Jiang, S. *J. Am. Chem. Soc.* **2002**, *124*, 3395–3401. (g) Boas, U.; Karlsson, A. J.; de Waal, B. F. M.; Meijer, E. W. *J. Org. Chem.* **2001**, *66*, 2136–2145. (h) Kleij, A. W.; Van de Coevering, R.; Gebbink, R. J. M. K.; Noordman, A.-M.; Spek, A.-L.; Van Koten, G. *Chem.-Eur. J.* **2001**, *7*, 181–192. (i) Kleinman, M. H.; Flory, J. H.; Tomalia, D. A.; Turro, N. J. *J. Phys. Chem. B* **2000**, *104*, 11472–11479. (j) Cardona, C. M.; Alvarez, J.; Kaifer, A. E.; McCarley, T. D.; Pandey, S.; Baker, G. A.; Bonzagni, N. J.; Bright, F. V. *J. Am. Chem. Soc.* **2000**, *122*, 6139–6144. (k) Baars, M. W. P. L.; Kleppinger, R.; Koch, M. H. J.; Yeu, S.-L.; Meijer, E. W. *Angew. Chem., Int. Ed.* **2000**, *39*, 1285–1288. (l) Stephan, H.; Spies, H.; Johannsen, B.; Klein, L.; Vogtle, F.; *Chem. Comm.* **1999**, 1875–1876. (m) Zimmerman, S. C.; Wang, Y.; Bharathi, P.; Moore, J. S. *J. Am. Chem. Soc.* **1998**, *120*, 2172–2173. (n) Castro, R.; Cuadrado, I.; Alonso, B.; Casado, C. M.; Moran, M.; Kaifer, A. E. *J. Am. Chem. Soc.* **1997**, *119*, 5760–5761. (o) Stelvekens, S.; van Hest, J. C. M.; Jansen, J. F. G. A.; van Boxel, D. A. F. J.; de Brabander-van den Berg, E. M. M.; Meijer, E. W. *J. Am. Chem. Soc.* **1996**, *118*, 7398–7399.
- (7) (a) DeLong, R.; Stephenson, K.; Loftus, T.; Fisher, M.; Alahari, S.; Nolting, A.; Juliano, R. L. *J. Pharm. Sci.* **1997**, *86*, 762–764. (b) Neelov, I. M.; Adolf, D. D.; McLeish, T. C. B. *Computer simulation of dendrimers and hyperbranched polymers and its application in drug delivery*, Programme and Proceedings of 6<sup>th</sup> International Symposium on Polymer Therapeutics, 7–9 January, 2004 Cardiff, UK, p 41. (c) Balabaev, N.; Neelov, I.; Bessonov, V.; Mazo, M. *Computer simulation of dendrimers in explicit water and in complexes with drug molecules*, Euroconference on Multiscale Phenomena in material Structure Formation, Bled, Slovenia, May 10–16, 2004, p 13.
- (8) (a) Balzani, V.; Ceroni, P.; Gestermann, S.; Gorka, M.; Kauffmann, C.; Maestri, M.; Vogtle, F. *ChemPhysChem* **2000**, *1*, 224–227. (b) Balzani, V.; Ceroni, P.; Gestermann, S.; Gorka, M.; Kauffmann, C.; Vogtle, F. *Tetrahedron* **2002**, *58*, 629–637.
- (9) Teobaldi, G.; Zerbetto, F. *J. Am. Chem. Soc.* **2003**, *125*, 7388–7393.

- (10) (a) Miklis, P.; Cagin, T.; Goddard, W. A., III. *J. Am. Chem. Soc.* **1997**, *119*, 7458–7462. (b) Karatasos, K.; Adolf, D. B.; Davis, G. R. *J. Chem. Phys.* **2001**, *115*, 5310–5318. (c) Maiti, P. K.; Cagin, T.; Wang, G.; Goddard, W. A., III. *Macromolecules* **2004**, *37*, 6236–6254. (d) Terao, T.; Nakayama, T. *Macromolecules* **2004**, *37*, 4686–4694. (e) Neelov, I. M.; Adolf, D. B. *J. Phys. Chem. B* **2004**, *108*, 7627–7636. (f) Lee, I.; Athey, B. D.; Wetzal, A. W.; Meixner, W.; Baker, J. R., Jr. *Macromolecules* **2002**, *35*, 4510–4520.
- (11) (a) Kohn, F.; Hofkens, J.; Wiesler, U.-M.; Cotlet, M.; van der Auweraer, M.; Mullen, K.; de Schryver, F. C. *Chem.-Eur. J.* **2001**, *7*, 4126–4133. (b) Hofkens, J.; Verheijen, W.; Shukla, R.; Dehaen, W.; De Schryver, F. C. *Macromolecules* **1998**, *31*, 4493–4497. (c) Masuo, S.; Yoshikawa, H.; Asahi, T.; Masuhara, H.; Sato, T.; Jiang, D.-L.; Aida, T. *J. Phys. Chem. B* **2001**, *105*, 2885–2889. (d) Stiriba, S.-E.; Kautz, H.; Frey, H. *J. Am. Chem. Soc.* **2002**, *124*, 9698–9699.
- (12) (a) Laufersweiler, M. J.; Rohde, J. M.; Chaumette, J.-L.; Sarazin, D.; Parquette, J. R. *J. Org. Chem.* **2001**, *66*, 6440–6452. (b) Tande, B. M.; Wagner, N. J.; Mackay, M. E.; Hawker, C. J.; Jeong, M. *Macromolecules* **2001**, *34*, 8580–8585. (c) Trollss, M.; Atthof, B.; Wuersch, A.; Hedrick, J. L.; Pople, J. A.; Gast, A. P. *Macromolecules*, **2000**, *33*, 6423–6438.
- (13) Allen, M. P.; Schofield, P. *Mol. Phys.* **1980**, *39*, 207–215.
- (14) (a) Allinger, N. L.; Yuh, Y.-H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566. (b) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566–8575. (c) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576–8582.
- (15) (a) Ponder, J. W.; Richards, F.; *J. Comput. Chem.* **1987**, *8*, 1016–1024. (b) Kundrot, C.; Ponder, J. W.; Richards, F. *J. Comput. Chem.* **1991**, *12*, 402–409. (c) Dudek, M. J.; Ponder, J. W. *J. Comput. Chem.* **1995**, *16*, 791–816.
- (16) Berendsen, H. J. C.; Postma, J. P. M.; van Gusteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (17) (a) Bermudez, V.; Capron, N.; Gase, T.; Gatti, F. G.; Kajzar, F.; Leigh, D. A.; Zerbetto, F.; Zhang, S. W. *Nature* **2000**, *406*, 608–611. (b) Fustin, C.-A.; Leigh, D. A.; Rudolf, P.; Timpel, D.; Zerbetto, F. *ChemPhysChem* **2000**, *1*, 97–100. (c) Cavallini, M.; Lazzaroni, R.; Zamboni, R.; Biscarini, F.; Timpel, D.; Zerbetto, F.; Clarkson, G. J.; Leigh, D. A. *J. Phys. Chem. B* **2001**, *105*, 10826–10830. (d) Biscarini, F.; Cavallini, M.; Leigh, D. A.; Léon, S.; Teat, S. J.; Wong, J. K. Y.; Zerbetto, F. *J. Am. Chem. Soc.* **2002**, *124*, 225–233. (e) Baxter, R. J.; Rudolf, P.; Teobaldi, G.; Zerbetto, F. *ChemPhysChem* **2004**, *5*, 245–248. (f) Höfinger, S.; Zerbetto, F. *J. Phys. Chem. A* **2003**, *107*, 11253–11257.

CT0499332

## Direct Dynamics Studies on the Hydrogen Abstraction Reactions of an F Atom with CH<sub>3</sub>X (X = F, Cl, and Br)

Li Wang, Jing-yao Liu, Ze-sheng Li,\* and Chia-chung Sun

*Institute of Theoretical Chemistry, State Key Laboratory of Theoretical and Computational Chemistry, Jilin University, Changchun 130023, P. R. China*

Received September 26, 2004

**Abstract:** The hydrogen abstraction reactions of F + CH<sub>3</sub>F (R1), F + CH<sub>3</sub>Cl (R2), and F + CH<sub>3</sub>Br (R3) are investigated by the dual-level direct dynamics method. Optimized geometries and frequencies of all the stationary points and extra points along the minimum-energy path (MEP) are obtained at the MP2/6-311G(d, p) level of theory, and then the energy profiles are refined at the CCSD(T)/6-311++G(3df, 2pd) level of theory. The basis set superposition error (BSSE) on the energy changes is corrected by means of the counterpoise method. Using the variational transition state theory (VTST) with the inclusion of the small-curvature tunneling correction, the rate constants are calculated over a wide temperature range of 189–2000 K. It is found that the activation energies for the title reactions are on the order of R1 > R2 > R3 and the rate constants exhibit just the opposite order of  $k_3 > k_2 > k_1$ . Both the activation energies and the rate constants show the clear-cut linear correlations with the hardness  $\eta$  of the halomethane molecules. Good agreement between the calculated and experimental rate constants is obtained at the measured temperatures. Furthermore, we hope that the theoretical studies for these compounds can give further information concerning the effects of halogen substitution on the rate constants of this class of hydrogen abstraction reactions.

### Introduction

The adverse effect of halogen-substituted hydrocarbons has attracted international attention.<sup>1–3</sup> The important atmospheric species, halomethanes, should be responsible for the depletion of the ozone layer in the stratosphere and the greenhouse effect. Thus, to estimate the atmospheric lifetimes of such species, accurate data for the rate constants as well as their temperature dependencies are needed. The reactions including fluorine atoms attract lots of attentions because fluorine atoms can react with most trace atmospheric compounds with the rapid rate constants. In early literature, the considerable experimental investigations<sup>4–10</sup> have been devoted to the kinetics of the hydrogen abstraction reactions of F + CH<sub>3</sub>X (X=F, Cl, and Br). For the reaction F + CH<sub>3</sub>F → HF + CH<sub>2</sub>F (R1), there is larger discrepancy among the measured rate constants. The experimented value of  $(6.59 \pm 1.51) \times 10^{-12} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ <sup>10</sup> is much lower than

that of  $8.8 \times 10^{-11} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ .<sup>4</sup> Very recently, Persky determined the rate constant of  $(2.8 \pm 0.2) \times 10^{-11} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$  at 298 K,<sup>9</sup> which agrees well with other measured values. However, almost all the studies are carried out at room temperature except that Persky<sup>9</sup> reported the temperature dependence of rate constants from 189 to 298 K and the Arrhenius expression is given. With respect to the reaction F + CH<sub>3</sub>Cl → HF + CH<sub>2</sub>Cl (R2), there are three experimental values available at 295 or 298 K,<sup>7,11–12</sup> which show well mutual consistency. Also some experiments are conducted on the reaction of F + CH<sub>3</sub>Br → HF + CH<sub>2</sub>Br (R3), in which the results determined by Iyer and Rowland<sup>13</sup> are somewhat higher than other values.<sup>7,14–15</sup> Contrary to the considerable experiments, no theoretical information is available on the reactions of F atoms with halomethanes.

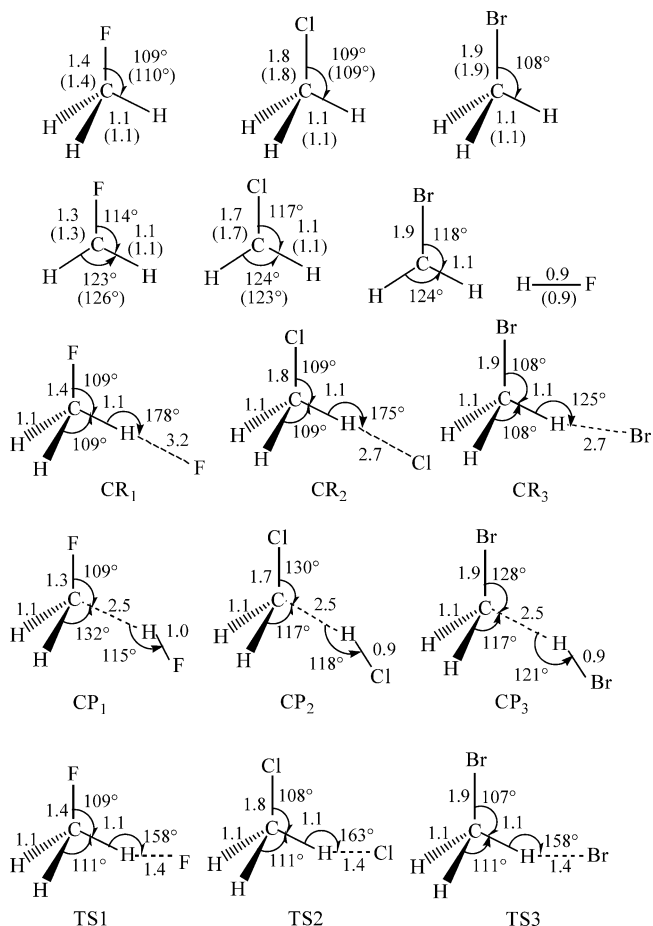
Here, a dual-level (X/Y)<sup>16–18</sup> direct dynamics method is applied to study the kinetic nature of the reactions over a wide temperature range. In this approach, the required electronic structure information for the stationary points and a series of extra points along the minimum energy path

\* Corresponding author fax: +86–431–8498026; e-mail: lly121@mail.jlu.edu.cn, liujy121@163.com.

(MEP) are obtained directly from ab initio calculations. Subsequently, by means of the Polyrate 8.4.1 program<sup>19</sup> the rate constants are carried out using the variational transition state theory (VTST)<sup>20,21</sup> proposed by Truhlar and co-workers for each reaction. The comparison between theory and experiment is discussed. Due to lack of the temperature dependence kinetic data of these reactions, the present theoretical results are expected to be useful and reasonable to estimate the dynamical properties of these reactions over a wide temperature range where no experimental value is available.

## Calculation Methods

All of the electronic structure calculations are carried out with the Gaussian 98 program.<sup>22</sup> The geometries, energies, and frequencies of all the stationary points, including the reactants, transition states (TSs), and products involved in the three reactions, are computed by using restricted or unrestricted second-order Møller–Plesset perturbation theory with the 6-311G(d, p) basis set (MP2/6-311G(d, p)). To obtain more reliable reaction enthalpies and barrier heights, high level single-point calculations for the stationary points are performed at the CCSD(T) level (coupled-cluster approach with single and double substitutions including a perturbative estimate of connected triples substitutions) with the flexible 6-311++G(3df, 2pd) basis set using the MP2 optimized geometries. To test the consistency of the calculated geometries and energies, we also employ higher electronic correlation level QCISD (quadratic configuration interaction with single and double substitutions) for the  $\text{CH}_3\text{F} + \text{F}$  reaction. The minimum-energy path (MEP) is calculated by the intrinsic reaction coordinate (IRC) theory at the MP2 level to confirm that the TS really connects with minima along the reaction path. The first and second energy derivatives at geometries along the MEP are obtained to calculate the curvature of the reaction path and to calculate the generalized vibrational frequencies along the reaction path. The potential profile is further refined at the CCSD(T)/6-311++G(3df, 2pd)//MP2/6-311G(d, p) level. Furthermore, the effect of the basis set superposition error on the energies is considered by means of the counterpoise method proposed by Boys and Bernardi.<sup>23</sup> The initial information on the potential energy surface is used to evaluate the rate constants by means of the Polyrate 8.4.1 program. The rate constants are calculated by using the variational transition state theory (VTST)<sup>20,21</sup> proposed by Truhlar and co-workers. The specific level of VTST that we used is canonical variational transition-state theory (CVT)<sup>24</sup> with the small-curvature tunneling (SCT)<sup>25</sup> method. The  $^2\text{P}_{1/2}$  and  $^2\text{P}_{3/2}$  electronic states of the fluorine atom, with a splitting of  $\Delta E = 404 \text{ cm}^{-1}$  ( $1.15 \text{ kcal mol}^{-1}$ ) due to the spin–orbit coupling, are used in the calculation of the electronic partition functions. It should be noted that the spin–orbit effect would lower the energy of the lower state (i.e.,  $^2\text{P}_{3/2}$  electronic state) of the F atom. Furthermore, as pointed out by Truhlar et al.,<sup>26</sup> the spin–orbit coupling is essentially fully quenched at the transition state and the nonrelativistic treatment will give a good approximation to the correct energy. With the same consideration, in this article the effect of the spin–orbit coupling is considered in the reactant partition function



**Figure 1.** Optimized geometries of  $\text{CH}_3\text{F}$ ,  $\text{CH}_3\text{Cl}$ ,  $\text{CH}_3\text{Br}$ ,  $\text{HF}$ ,  $\text{CH}_2\text{F}$ ,  $\text{CH}_2\text{Cl}$ ,  $\text{CH}_2\text{Br}$ , complexes, and three transition states at the MP2/6-311G(d, p) level. The values in the parentheses are the experimental values.<sup>27–29</sup> Bond lengths are in angstroms and angles are in degrees.

and is neglected in the saddle point, that is, the effective barrier heights are increased. As a result, the rate constants can be expected to slightly decrease when the spin–orbit coupling is considered only in the reactant partition function. The curvature components are calculated by using a quadratic fit to obtain the derivative of the gradient with respect to the reaction coordinate.

## Results and Discussion

**1. Stationary Points.** The geometric parameters of all the stationary points including the reactants, complexes, products, and transition states optimized at the MP2/6-311G(d, p) level are shown in Figure 1 as well as the available experimental values<sup>27–29</sup> for comparison. As can be seen from Figure 1, the largest deviation between theoretical bond lengths and experimental values is  $0.1 \text{ \AA}$ , and the largest deviation of the angle is about  $3^\circ$ . It is clear that the theoretical values are in reasonable accord with the experimental ones. At the MP2 level, complexes  $\text{CR}_1$ ,  $\text{CR}_2$ , and  $\text{CR}_3$  are located at the entrance channel of reactions R1, R2, and R3, respectively, in which the H–F bond lengths are  $3.2$ ,  $2.7$ , and  $2.7 \text{ \AA}$ , respectively. Also, there exist three complexes  $\text{CP}_1$ ,  $\text{CP}_2$ , and  $\text{CP}_3$  located at the product sides of three reactions. With respect to the three transition states, all of the breaking C–H

**Table 1.** Calculated and Experimental Frequencies (in  $\text{cm}^{-1}$ ) of the Reactants, Complexes, Products and Transition States at the MP2/6-311G(d, p) Level

|                    | MP2/6-311G(d, p)   | exptl.  |
|--------------------|--|---|
| CH <sub>3</sub> F  | 1106,1224,1224,1519,1519,1537,3088,3186,3186             | 1049,1182,1182,1464,1464,1467,2965,3006,3006 <sup>a</sup> |
| CH <sub>3</sub> Cl | 784,1064,1064,1441,1496,1496,3119,3227,3227              | 732,1017,1017,1355,1455,1455,2968,3054,3054 <sup>b</sup>  |
| CH <sub>3</sub> Br | 648,990,990,1381,1496,1496,3121,3237,3237                | 611,955,955,1306,1443,1443,2935,3056,3056 <sup>c</sup>    |
| HF                 | 4252   | 4138 <sup>d</sup>   |
| CH <sub>2</sub> F  | 713,1206,1212,1520,3193,3351                             |   |
| CH <sub>2</sub> Cl | 161,879,1046,1474,3239,3394                              |   |
| CH <sub>2</sub> Br | 98,732,967,1438,3232,3390                                |   |
| CR <sub>1</sub>    | 10,18,20,1107,1224,1224,1519,1519,1536,3088,3186,3187    |   |
| CP <sub>1</sub>    | 60,91,116,288,315,779,1209,1212,1516,3193,3350,4195      |   |
| CR <sub>2</sub>    | 19,67,87,783,1065,1069,1442,1498,1500,3122,3226,3237     |   |
| CP <sub>2</sub>    | 36,75,112,282,285,561,880,1051,1471,3222,3371,4200       |   |
| CR <sub>3</sub>    | 27,28,78,647,991,994,1379,1495,1498,3125,3238,3246       |   |
| CP <sub>3</sub>    | 30,74,108,279,289,559,735,972,1437,3214,3364,4196        |   |
| TS1                | 866i,103,144,853,1146,1207,1307,1388,1523,1864,3133,3224 |   |
| TS2                | 954i,77,149,780,835,1050,1198,1349,1466,1746,3162,3252   |   |
| TS3                | 1028i,81,164,603,724,981,1171,1346,1444,1784,3167,3260   |   |

<sup>a</sup> From ref 27. <sup>b</sup> From ref 28. <sup>c</sup> From ref 31. <sup>d</sup> From ref 29.

**Table 2.** Enthalpies (in  $\text{kcal mol}^{-1}$ ) and Barrier Heights (in  $\text{kcal mol}^{-1}$ ) at MP2/6-311G(d, p) and CCSD(T)/6-311++G(3df, 2pd)//MP2/6-311G(d, p) Levels and Available Experimental Values

| levels              | CH <sub>3</sub> F + F → HF + CH <sub>2</sub> F |   | CH <sub>3</sub> Cl + F → HF + CH <sub>2</sub> Cl |                            | CH <sub>3</sub> Br + F → HF + CH <sub>2</sub> Br |                            |
|---------------------|--|---|--|----------------------------|--|----------------------------|
|                     | $\Delta H_{298}^0$                             | $\Delta E(0\text{ K})$                          | $\Delta H_{298}^0$                               | $\Delta E(0\text{ K})$     | $\Delta H_{298}^0$                               | $\Delta E(0\text{ K})$     |
| MP2                 | -32.58   | 3.48  | -32.71   | 3.86                       | -31.75   | 3.90                       |
| CCSD(T)             | -34.80   | -4.02 (-2.27) <sup>b</sup> [-0.81] <sup>c</sup> | -36.73   | -5.12 (-3.18) <sup>b</sup> | -35.58   | -5.16 (-3.30) <sup>b</sup> |
| exptl. <sup>a</sup> | -36.14 ± 3.3                                   |   | -35.54 ± 2.3                                     |                            | -35.64 ± 2.3                                     |                            |

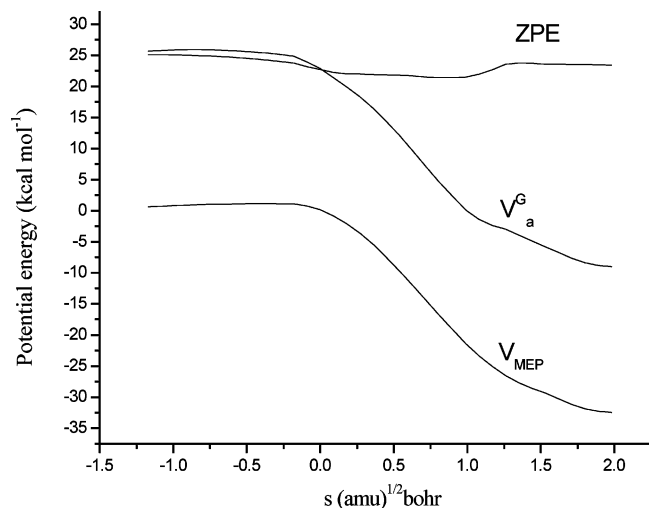
<sup>a</sup> From ref 32. <sup>b</sup> The values in the parentheses including BSSE corrections. <sup>c</sup> The values in the square bracket obtained at the CCSD(T)/6-311++G(3df, 2pd)//QCISD/6-311G(d, p) level with BSSE correction.

bonds are elongated by 4.6% compared to the C–H equilibrium bond length in isolated CH<sub>3</sub>F, CH<sub>3</sub>Cl, and CH<sub>3</sub>Br, respectively; and the forming H–F bond is elongated by 55, 53 and 53% with respect to the H–F equilibrium bond length in isolated HF, respectively. The elongation of the forming bond is greater than that of the breaking bond, indicating that the three TSs are reactant-like, i.e., all these reactions proceed via “early” transition states. This rather early character in these transition states is in accordance with the low barrier heights and the exothermicity of these reactions, in keeping with Hammond’s postulate.<sup>30</sup>

The harmonic vibrational frequencies are calculated at the same level of theory to characterize the nature of each critical point and to make zero-point energy (ZPE) corrections. Table 1 lists the harmonic vibrational frequencies of all the stationary points along with the available experimental data.<sup>27–29,31</sup> Our calculated frequencies are in good agreement with the experimental values, with the largest deviation within 9%. The number of imaginary frequencies (0 or 1) indicates whether a minimum or a transition state has been located. All of the complexes have only real frequencies. The transition state is confirmed by normal-mode analysis to have only one imaginary frequency, which takes the values of 866i, 954i, and 1028i  $\text{cm}^{-1}$ , respectively.

The reaction enthalpies ( $\Delta H_{298}^0$ ) and classical barrier heights ( $\Delta E(0\text{K})$ ) calculated at the MP2/6-311G(d, p) and CCSD(T)/6-311++G(3df, 2pd)//MP2 levels with ZPE and BSSE corrections are listed in Table 2. The calculated enthalpies of -34.80, -36.73, and -35.58  $\text{kcal mol}^{-1}$  for

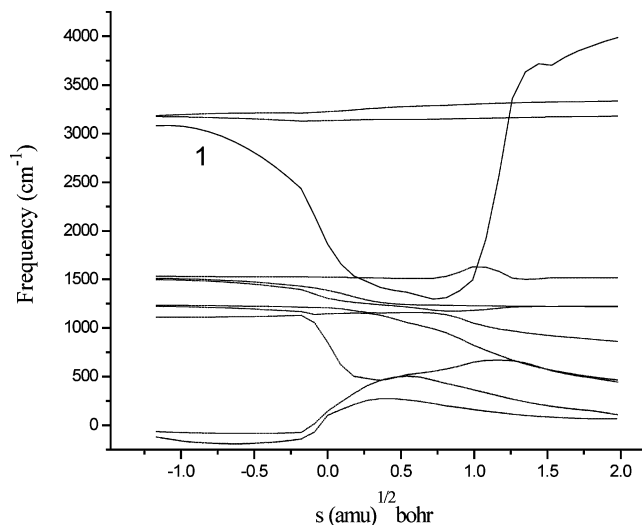
the series of reactions at the higher level are in good agreement with the experimental values of -36.14 ± 3.3, -35.54 ± 2.3, and -35.64 ± 2.3  $\text{kcal mol}^{-1}$ , respectively, which is derived from the experimental standard heats of formation (CH<sub>3</sub>F, -56 ± 1  $\text{kcal mol}^{-1}$ ; CH<sub>3</sub>Cl -19.6  $\text{kcal mol}^{-1}$ ; CH<sub>3</sub>Br, -8.5  $\text{kcal mol}^{-1}$ ; CH<sub>2</sub>F, -8 ± 2  $\text{kcal mol}^{-1}$ ; CH<sub>2</sub>Cl, 29 ± 2  $\text{kcal mol}^{-1}$ ; CH<sub>2</sub>Br, 40 ± 2  $\text{kcal mol}^{-1}$ ; HF, -65.14 ± 0.2  $\text{kcal mol}^{-1}$ ; F, 19.0 ± 0.1  $\text{kcal mol}^{-1}$ ).<sup>32</sup> For reactions CH<sub>3</sub>F + F → CH<sub>2</sub>F + HF, CH<sub>3</sub>Cl + F → CH<sub>2</sub>Cl + HF, and CH<sub>3</sub>Br + F → CH<sub>2</sub>Br + HF, the complexes CR<sub>1</sub>, CR<sub>2</sub>, and CR<sub>3</sub> are first formed with relative energies being -0.38, -0.32, and -1.52  $\text{kcal mol}^{-1}$  below the reactants of CH<sub>3</sub>F + F, CH<sub>3</sub>Cl + F, and CH<sub>3</sub>Br + F at the CCSD(T)//MP2 level with ZPE corrections, respectively. Then starting from the complex, each of the reaction passes through a reactant-like transition state to form another complex with relative energies of 0.70, 0.19, and 0.23  $\text{kcal mol}^{-1}$  below the products of CH<sub>2</sub>F + HF, CH<sub>2</sub>Cl + HF, and CH<sub>2</sub>Br + HF, respectively. It is seen that the energies of the complexes are very close to those of the reactants or products, so one question arises: do these complexes really exist or is it an artifact due to the theoretical methods? To test the stability of the complexes, the basis set superposition error (BSSE) correction is estimated using the counterpoise method. At the CCSD(T)/6-311++G(3df, 2pd) level, we obtain BSSEs of 0.28, 0.38, and 0.48  $\text{kcal mol}^{-1}$  for CR<sub>1</sub>, CR<sub>2</sub>, and CR<sub>3</sub>, respectively. Thus, the BSSE-corrected energies for three complexes are -0.11, 0.06, and -1.04  $\text{kcal mol}^{-1}$ , respectively. Similarly, the corresponding values for complexes



**Figure 2.** Classical potential energy curve ( $V_{\text{MEP}}$ ), ground-state vibrationally adiabatic energy curve ( $V_a^G$ ), and zero-point energy curve (ZPE) as functions of  $s$  ( $\text{amu}^{1/2}$  bohr) at the CCSD(T)/6-311++G(3df, 2pd)//MP2/6-311G(d, p) level with BSSE correction for the  $\text{CH}_3\text{F} + \text{F} \rightarrow \text{CH}_2\text{F} + \text{HF}$ .

$\text{CP}_1$ ,  $\text{CP}_2$ , and  $\text{CP}_3$  are  $-0.22$ ,  $0.30$ , and  $0.25$   $\text{kcal mol}^{-1}$  relative to the products, respectively. It can be found that complexes  $\text{CR}_2$ ,  $\text{CP}_2$  and  $\text{CP}_3$  disappear when BSSE correction is included. With respect to the barrier heights, the calculated results of three reactions obtained at the MP2 level are  $3.48$ ,  $3.86$ , and  $3.90$   $\text{kcal mol}^{-1}$ , respectively, while at the CCSD(T)//MP2 level, the corresponding values are  $4.02$ ,  $5.12$ , and  $5.16$   $\text{kcal mol}^{-1}$  below the reactants, respectively. When the BSSE is considered, as calculated in the complexes, the barrier heights are  $-2.27$ ,  $-3.18$ , and  $-3.30$   $\text{kcal mol}^{-1}$  at CCSD(T)//MP2 level, respectively. Moreover, for comparison we performed the optimization calculation for reaction R1 at higher level of electron correlation QCISD/6-311G(d, p) with the single-point energies for the stationary points at the same CCSD(T)/6-311++G(3df, 2pd) level (Table 2). The calculated enthalpies ( $\Delta H_{298}^\circ$ ) value is  $-34.81$   $\text{kcal mol}^{-1}$  at the CCSD(T)//QCISD level, which is in excellent agreement with both the experimental results and the theoretical ones obtained at the CCSD(T)//MP2 level. The barrier height of reaction R1 is  $-0.81$   $\text{kcal mol}^{-1}$  at CCSD(T)//QCISD level with ZPE and BSSE corrections. It can be found that the classical barrier heights obtained at two higher levels are very closed except that the value calculated at CCSD(T)//QCISD level is slightly higher than that obtained at CCSD(T)//MP2 level.

The potential profile is further refined by performing a series of single point calculations at the CCSD(T)/6-311++G(3df, 2pd)//MP2 level including the BSSE correction. The classical potential energy curve ( $V_{\text{MEP}}(s)$ ), the vibrationally adiabatic ground-state potential energy curve ( $V_a^G(s)$ ), and the zero-point energy (ZPE) curve of the reaction R1 as a function of the intrinsic reaction coordinate ( $s$ ) are plotted in Figure 2, where  $V_a^G(s) = V_{\text{MEP}}(s) + \text{ZPE}(s)$ . As can be seen, the  $V_{\text{MEP}}$  and  $V_a^G$  curves are similar in shape, and the ZPE is practically constant as  $s$  varies with only a gentle drop near the saddle point. It should be noted that the locations of maximum on the  $V_a^G(s)$  and  $V_{\text{MEP}}(s)$  energy curves shift in the  $s$  direction toward the reactants



**Figure 3.** Changes in the generalized normal-mode vibrational frequencies as functions of  $s$  ( $\text{amu}^{1/2}$  bohr) at the MP2/6-311G(d, p) level for the  $\text{CH}_3\text{F} + \text{F} \rightarrow \text{CH}_2\text{F} + \text{HF}$ .

and the maximum of the two curves are slightly higher than the reactants. This is the case that the saddle-point position of the dual-level is generally shifted. The same conclusion can be drawn from the other two reactions.

The variations of the generalized normal-mode vibrational frequencies along the MEP of reaction R1 is shown in Figure 3, and the similar figures for the reactions R2 and R3 are omitted for simplicity. In the negative limit of  $s$ , the frequencies are associated with the reactants  $\text{F} + \text{CH}_3\text{F}$ , and in the product region the frequencies correspond to the products  $\text{HF} + \text{CH}_2\text{F}$ . In Figure 3, all of the frequencies except for mode 1 do not change significantly on going from the reactants to the products. The “reactive mode” 1 relating to the breaking/forming bonds has a significant change from  $s = 0$  to  $1.0$  ( $\text{amu}^{1/2}$  bohr) on the MEP. These drops should cause considerable falls in the ZPE near the saddle point. On the other hand, the two lowest harmonic frequencies corresponding to free rotations and translations of the reactants evolve into vibrations and they present a maximum near the saddle point. The behavior of these transitional modes partially compensated the fall in the ZPE caused by the reactive mode, thus the ZPE shows small variations with  $s$ .

**2. Dynamics Calculation.** Dual-level (X/Y)<sup>16–18</sup> direct dynamics calculations are carried out for the three reactions using the variational transition-state theory. The BSSE-corrected and noncorrected PES information for each reaction obtained at the CCSD(T)/6-311++G(3df, 2pd)//MP2/6-311G(d, p) level is put into Polyrate 8.4.1 program<sup>19</sup> to calculate the VTST<sup>20,21</sup> rate constants over the temperature range from 189 to 2000 K. The forward rate constants are calculated by canonical variational transition-state theory (CVT)<sup>24</sup> with the small-curvature tunneling (SCT)<sup>25</sup> method.

The theoretical rate constants based on the BSSE-corrected PES and the available experimental values are shown in Figure 4a-c. For the reaction of  $\text{CH}_3\text{F} + \text{F}$  (see Figure 4a), our calculated results agree well with the experimental values<sup>5–9</sup> except the results reported by Nielsen et. al. ( $8.8 \times 10^{-11}$   $\text{cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ )<sup>4</sup> and the values given by Kowalczyk et. al. ( $(6.59 \pm 1.51) \times 10^{-12}$   $\text{cm}^3 \text{ molecule}^{-1}$



**Table 3.** Calculated Rate Constants (in  $\text{cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ ) for the Reaction (a)  $\text{CH}_3\text{F} + \text{F} \rightarrow \text{CH}_2\text{F} + \text{HF}$ , (b)  $\text{CH}_3\text{Cl} + \text{F} \rightarrow \text{CH}_2\text{Cl} + \text{HF}$ , (c)  $\text{CH}_3\text{Br} + \text{F} \rightarrow \text{CH}_2\text{Br} + \text{HF}$  in the Temperature Range from 189 to 2000 K and Available Experimental Values

| $T$ (K) | $k_1(10^{11})$ | $k_{\text{exptl.}}(10^{11})$   | $k_2(10^{11})$ | $k_{\text{exptl.}}(10^{11})$       | $k_3(10^{11})$ | $k_{\text{exptl.}}(10^{11})$                          |
|---------|----------------|--|----------------|------------------------------------|----------------|---|
| 189     | 0.90           | $1.30 \pm 0.10^a$  | 1.23           |                                    | 1.45           |   |
| 219     | 1.10           | $1.76 \pm 0.20^a$  | 1.43           |                                    | 1.70           |   |
| 225     | 1.13           |  | 1.47           |                                    | 1.75           |   |
| 251     | 1.31           | $2.17 \pm 0.20^a$  | 1.61           |                                    | 1.98           |   |
| 275     | 1.48           |  | 1.74           |                                    | 2.19           |   |
| 295     | 1.62           | $3.7 \pm 0.8^b$  | 1.88           | $2.4 \pm 0.5^h$<br>$3.3 \pm 0.7^b$ | 2.36           | $6.1 \pm 0.7^j$<br>$4.5 \pm 0.9^k$<br>$3.0 \pm 0.7^b$ |
| 296     | 1.63           |  | 1.89           |                                    | 2.37           | $4.5 \pm 0.2^l$                                       |
| 298     | 1.64           | $8.8^c$<br>$3.6 \pm 0.2^d$<br>$3.0^e$<br>$2.8 \pm 0.6^f$<br>$2.76 \pm 0.20^a$<br>$0.66 \pm 0.15^g$ | 1.90           | $2.4 \pm 0.7^i$                    | 2.39           |   |
| 300     | 1.66           |  | 1.92           |                                    | 2.41           |   |
| 400     | 2.46           |  | 2.69           |                                    | 3.31           |   |
| 500     | 3.41           |  | 3.60           |                                    | 4.35           |   |
| 600     | 4.56           |  | 4.66           |                                    | 5.65           |   |
| 900     | 8.94           |  | 8.73           |                                    | 10.5           |   |
| 1000    | 10.6           |  | 10.4           |                                    | 12.5           |   |
| 1200    | 14.2           |  | 13.9           |                                    | 16.8           |   |
| 1500    | 20.6           |  | 19.6           |                                    | 24.0           |   |
| 2000    | 33.0           |  | 31.3           |                                    | 37.8           |   |

<sup>a</sup> From ref 9. <sup>b</sup> From ref 7. <sup>c</sup> From ref 4. <sup>d</sup> From ref 5. <sup>e</sup> From ref 6. <sup>f</sup> From ref 8. <sup>g</sup> From ref 10. <sup>h</sup> From ref 12. <sup>i</sup> From ref 11. <sup>j</sup> From ref 13. <sup>k</sup> From ref 14. <sup>l</sup> From ref 15.

$\text{s}^{-1}$ ).<sup>10</sup> The deviation remains within a factor of about 1.4–2.0. Moreover, the Arrhenius expression of  $k_1 = 4.71 \times 10^{-11} \exp(-317.2/T) \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$  fitted by the CVT/SCT rate constant in the temperature range 189–298 K is in good accord with that reported by Persky,<sup>9</sup>  $k = (1.03 \pm 0.13) \times 10^{-10} \exp[-(390 \pm 60)/T] \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ . As to the hydrogen abstraction reaction of  $\text{CH}_3\text{Cl} + \text{F}$  (see Figure 4b), the agreement between the theoretical rate constants and experimental ones<sup>7,11–12</sup> is considerably good, where the factor of deviation is only 1.0–1.4. As can be seen from Figure 4c the experimental rate constant<sup>7,13–15</sup> for  $\text{CH}_3\text{Br} + \text{F}$  reaction determined by Iyer and Rowland<sup>13</sup> is slightly higher than the other data, and our calculated result is in better agreement with the latter. In addition, the theoretical rate constants calculated on the potential energy surface without a BSSE correction are shown in the Supporting Information. It is obvious that when the BSSE correction is included in the classical barrier heights increase, as a result, the values of rate constants are lowered about 1–3 times in the lower temperature range. However, the discrepancy between them becomes smaller and almost disappears with the temperature increasing.

To reflect the effect of halogen substitution on the reactivity of the C–H bond, Arrhenius expressions are fitted based on the calculated rate constants of the three reactions in the temperature range 225–600 K. The preexponential factors (A) and activation energies ( $E_a$ ) are given in Table 4. It is shown that there is a slight influence on both A and  $E_a$  for the halogen substitutions from F to Br. The activation energies decrease in the order of R1 (1.00) > R2 (0.84) >

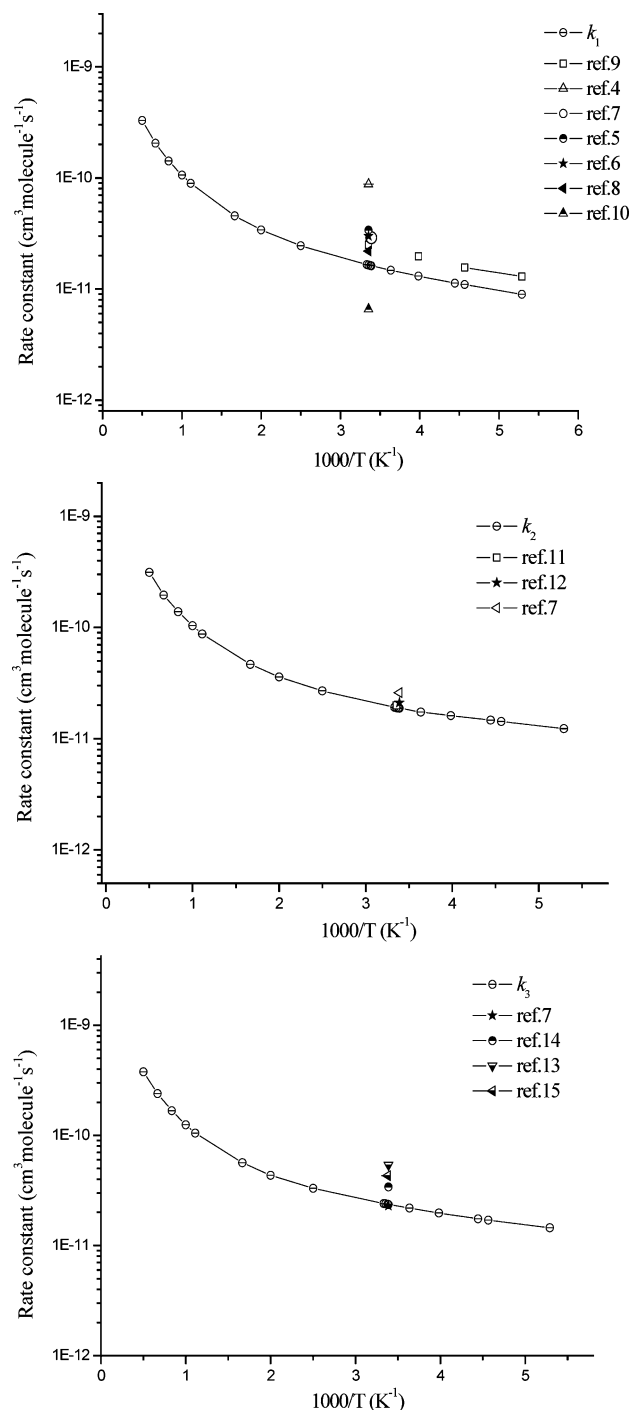
R3 (0.83), but the discrepancy among them is considerably small. Thus, it results in a small decrease of the rate constant from R3 to R1, i.e.,  $k_3 > k_2 > k_1$ . In addition, the activation energies of these reactions are correlated with the hardness of the halomethane molecules.<sup>33</sup> The hardness  $\eta$  is defined as  $\eta = (\text{IE} - \text{EA})/2$ , where IE and EA are the first vertical ionization energy and electron affinity of the molecule, respectively. The values of the hardness  $\eta$  of  $\text{CH}_3\text{F}$ ,  $\text{CH}_3\text{Cl}$ , and  $\text{CH}_3\text{Br}$  are 164.6, 139.4, and 130.6  $\text{kcal mol}^{-1}$  at the CCSD(T)//MP2 level, respectively, decreasing in the order of  $\text{CH}_3\text{F} > \text{CH}_3\text{Cl} > \text{CH}_3\text{Br}$ . Clearly, if the systems become softer, charge transfer between the reactants will be easier and the reaction will become more activity.

Owing to the good agreement between the theoretical and experimental values, it is reasonable to believe that our calculated results will provide a good estimate for the kinetics of the reactions in the high-temperature range. Note that although there have been some kinetic studies performed on the title reactions, most of the rate constants are measured around room temperature. Therefore, for convenience of future experimental measurements, the three-parameter fits for the CVT rate constants for the title reactions within 189–2000 K give the expressions as follows (in  $\text{cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ ):

$$k_1 = 5.38 \times 10^{-16} T^{1.75} \exp(109.7/T)$$

$$k_2 = 3.81 \times 10^{-16} T^{1.78} \exp(202.2/T)$$

$$k_3 = 7.06 \times 10^{-16} T^{1.72} \exp(176.3/T)$$



**Figure 4.** Plot of the CVT/SCT rate constants calculated at the CCSD(T)/6-311++G(3df, 2pd)//MP2/6-311G(d, p) level with BSSE correction and the available experimental values versus  $1000/T$  between 189 and 2000 K for the (a)  $\text{CH}_3\text{F} + \text{F} \rightarrow \text{CH}_2\text{F} + \text{HF}$ , (b)  $\text{CH}_3\text{Cl} + \text{F} \rightarrow \text{CH}_2\text{Cl} + \text{HF}$ , (c)  $\text{CH}_3\text{Br} + \text{F} \rightarrow \text{CH}_2\text{Br} + \text{HF}$ .

**Table 4.** Arrhenius Parameters<sup>a</sup> for the Title Reactions

| reaction                          | A ( $\text{cm}^3 \text{s}^{-1}$ ) | $E_a$ ( $\text{kcal mol}^{-1}$ ) |
|-----------------------------------|-----------------------------------|----------------------------------|
| $\text{F} + \text{CH}_3\text{F}$  | $9.27 \times 10^{-11}$            | 1.00                             |
| $\text{F} + \text{CH}_3\text{Cl}$ | $8.26 \times 10^{-11}$            | 0.84                             |
| $\text{F} + \text{CH}_3\text{Br}$ | $1.01 \times 10^{-10}$            | 0.83                             |

<sup>a</sup> In the temperature range of 225–600 K.

## Conclusion

In the present paper, we employ an ab initio direct dynamic method to study three hydrogen abstraction reactions of  $\text{CH}_3\text{X}$  ( $\text{X}=\text{F}$ ,  $\text{Cl}$ , and  $\text{Br}$ ) with  $\text{F}$  atoms. The potential energy surface information is obtained at the MP2/6-311G(d, p) level and higher level energies of the stationary points are calculated at the higher CCSD(T)/6-311++G(3df, 2pd) level with ZPE and BSSE corrections. The rate constant calculations are carried out using the variational transition state theory (VTST) at the CCSD(T)//MP2 level over a wide temperature range of 189–2000 K. The correlation among the activation energies, the rate constants, and the hardness of the halomethane molecules is discussed. For three reactions the calculated rate constants decrease in the order of  $k_3 > k_2 > k_1$ , which is just opposite of the orders of the theoretical activation energies and the hardness of halomethane molecules. Theoretical rate constants show good agreement with the available experimental values. The three-parameter expressions (in  $\text{cm}^3 \text{molecule}^{-1} \text{s}^{-1}$ ) for three reactions within 189–2000 K are  $k_1 = 5.38 \times 10^{-16} T^{1.75} \exp(109.7/T)$ ,  $k_2 = 3.81 \times 10^{-16} T^{1.78} \exp(202.2/T)$ , and  $k_3 = 7.06 \times 10^{-16} T^{1.72} \exp(176.3/T) \text{ cm}^3 \text{molecule}^{-1} \text{s}^{-1}$ . We hope the theoretical results may be useful for estimating the kinetics of the reactions over a wide temperature range where no experimental data are available.

**Acknowledgment.** We thank Professor Donald G. Truhlar for providing of the POLYRATE 8.4.1 program. This work was supported by the National Natural Science Foundation of China (20333050, 20303007), the Doctor Foundation by the Ministry of Education, the Foundation for University Key Teacher by the Ministry of Education, the Key Subject of Science and Technology by the Ministry of Education of China, and the Innovational Foundation by Jilin University.

**Supporting Information Available:** Plot of the CVT/SCT rate constants (Figure S1). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Atkinson, R. Kinetics and mechanisms of the gas-phase reactions of the hydroxyl radical with organic compounds under atmospheric conditions. *Chem. Rev.* **1986**, *86*, 69–201.
- (2) Paddison, S. J.; Tschuikow-Roux, E. Structures, Vibrational Frequencies, Thermodynamic Properties, and Bond Dissociation Energies of the Bromomethanes and Bromomethyl Radicals: an Ab Initio Study. *J. Phys. Chem. A* **1998**, *102*, 6191–6199.
- (3) Bilde, M.; Wallington, T. J.; Ferronato, C.; Orlando, J. J.; Tyndall, G. S.; Estupiñan, E.; Haberkorn, S. Atmospheric Chemistry of  $\text{CH}_2\text{BrCl}$ ,  $\text{CHBrCl}_2$ ,  $\text{CHBr}_2\text{Cl}$ ,  $\text{CF}_3\text{CHBrCl}$ , and  $\text{CBr}_2\text{Cl}_2$ . *J. Phys. Chem. A* **1998**, *102*, 1976–1986.
- (4) Pollock, T. L.; Jones, W. E. Gas-phase reactions of fluorine atoms. *Can. J. Chem.* **1973**, *51*, 2041–2046.
- (5) Manning, R. G.; Grant, E. R.; Merrill, J. C.; Parks, N. J.; Root, J. W. Hydrogen abstraction by fluorine atoms under conditions of thermal initiation: Hydrocarbons and fluorinated hydrocarbons. *Int. J. Chem. Kinet.* **1975**, *7*, 39–44.

- (6) Smith, D. J.; Setser, D. W.; Kim, K. C.; Bogan, D. J. HF infrared chemiluminescence. Relative rate constants for hydrogen abstraction from hydrocarbons, substituted methanes, and inorganic hydrides. *J. Phys. Chem.* **1977**, *81*, 898–905.
- (7) Wallington, T. J.; Hurley, M. D.; Shi, J.; Maricq, M. M.; Sehested, J.; Nielsen, O. J.; Ellermann, T. A kinetic study of the reaction of fluorine atoms with CH<sub>3</sub>F, CH<sub>3</sub>Cl, CH<sub>3</sub>-Br, CF<sub>2</sub>H<sub>2</sub>, CO, CF<sub>3</sub>H, CF<sub>3</sub>CHCl<sub>2</sub>, CF<sub>3</sub>CH<sub>2</sub>F, CHF<sub>2</sub>CHF<sub>2</sub>, CF<sub>2</sub>ClCH<sub>3</sub>, CHF<sub>2</sub>CH<sub>3</sub>, and CF<sub>3</sub>CF<sub>2</sub>H at 295 ± 2 K. *Int. J. Chem. Kinet.* **1993**, *25*, 651–665.
- (8) Moore, C.; Smith, I. W. M. Rate constants for the reactions of fluorine atoms with alkanes and hydrofluorocarbons at room temperature. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 3041–3044.
- (9) Persky, A. The temperature dependence of the rate constants for the reactions F + CH<sub>3</sub>F and F + CH<sub>2</sub>F<sub>2</sub>. *Chem. Phys. Lett.* **2003**, *376*, 181–187.
- (10) Kowalczyk, J.; Jowko, A.; Symanowicz, M. Kinetics of radical reactions in Freons. *J. Radioanal. Nucl. Chem.* **1998**, *232*, 75–78.
- (11) Clyne, M. A. A.; McKenney, D. J.; Walker, R. F. Reaction kinetics of ground-state fluorine, F(<sup>2</sup>P), atoms. I. Measurement of fluorine atom concentrations and the rates of reactions F + CHF<sub>3</sub> and F + Cl<sub>2</sub> using mass spectrometry. *Can. J. Chem.* **1973**, *51*, 3596–3604.
- (12) Wickramaratchi, M. A.; Setser, D. W.; Hildebrandt, H.; Korbitzer, B.; Heydtmann, H. Evaluation of HF product distributions deduced from infrared chemiluminescence. II. F atoms reactions. *Chem. Phys.* **1985**, *94*, 109–129.
- (13) Iyer, R. S.; Rowland, F. S. Atom-transfer reaction rates for thermal fluorine atoms with CH<sub>3</sub>X and CF<sub>3</sub>X (X = Br, I). *J. Phys. Chem.* **1981**, *85*, 2493–2497.
- (14) Nielsen, O. J.; Munk, J.; Locke, G.; Wallington, T. J. Ultraviolet absorption spectra and kinetics of the self-reaction of bromomethyl and peroxybromomethyl radicals in the gas phase at 298 K. *J. Phys. Chem.* **1991**, *95*, 8714–8719.
- (15) Sehested, J.; Bilde, M.; Mogelberg, T.; Wallington, T. J.; Nielsen, O. J. Kinetics and mechanism of the reaction of F atoms with CH<sub>3</sub>Br. *J. Phys. Chem.* **1996**, *100*, 10989–10998.
- (16) Truhlar, D. G. In *The Reaction Path in Chemistry: Current Approaches and Perspectives*; Heidrich, D. Ed.; Kluwer: Dordrecht The Netherlands, 1995, p 229.
- (17) Truhlar, D. G.; Garrent, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.
- (18) Hu, W. P.; Truhlar, D. G. Factors Affecting Competitive Ion–Molecule Reactions: ClO<sup>-</sup> + C<sub>2</sub>H<sub>5</sub>Cl and C<sub>2</sub>D<sub>5</sub>Cl via E2 and S<sub>N</sub>2 Channels. *J. Am. Chem. Soc.* **1996**, *118*, 860–869.
- (19) Chuang, Y.-Y.; Corchado, J. C.; Fast, P. L.; Villa, J.; Hu, W.-P.; Liu, Y.-P.; Lynch, G. C.; Jackels, C. F.; Nguyen, K. A.; Gu, M. Z.; Rossi, I.; Coitino, E. L.; Clayton, S.; Melissas, V. S.; Lynch, B. J.; Steckler, R.; Garrett, B. C.; Isaacson, A. D.; Truhlar, D. G. *Polyrate, version 8.4.1*, University of Minnesota, Minneapolis, MN, 2000.
- (20) Truhlar, D. G.; Garrett, B. C. Variational transition-state theory. *Acc. Chem. Res.* **1980**, *13*, 440–448.
- (21) Truhlar, D. G.; Isaacson, A. D.; Garrett, B. C. In *The Theory of Chemical Reaction Dynamics*; Baer, M., Ed.; CRC Press: Boca Raton, FL, 1985, p 65.
- (22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Boboul, A. G.; Stefnov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, L.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision X.; Gaussian Inc.: Pittsburgh, PA, 1998.
- (23) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
- (24) Garrett, B. C.; Truhlar, D. G. Criterion of minimum state density in the transition state theory of bimolecular reactions. *J. Chem. Phys.* **1979**, *70*, 1593–1598.
- (25) Liu, Y.-P.; Lynch, G. C.; Truong, T. N.; Lu, D.-h.; Truhlar, D. G.; Garrett, B. C. Molecular modeling of the kinetic isotope effect for the [1,5]-sigmatropic rearrangement of *cis*-1, 3-pentadiene. *J. Am. Chem. Soc.* **1993**, *115*, 2408–2415.
- (26) Corchado, J. C.; Truhlar, D. G.; Espinosa-Garcia, J. Potential energy surface, thermal, and state-selected rate coefficients, and kinetic isotope effects for Cl + CH<sub>4</sub> → HCl + CH<sub>3</sub>. *J. Chem. Phys.* **2000**, *112*, 9375–9388.
- (27) JANAF Thermochemical Tables: National Standard Reference Data Series; National Bureau of Standards: Washington, DC, 1986; Vol. 37, 2nd ed.
- (28) Kuchitsu, K. in *Structure of Free Polyatomic Molecules Basic Data* (Springer, Sakado), 1998, Vol. 104, p 94.
- (29) Linstrom, P. J.; Mallard, W. G. NIST Chemistry Webbook, <http://webbook.nist.gov/chemistry/> 1998.
- (30) Hammond, G. S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77*, 334–338.
- (31) Lide, D. R. In *CRC Handbook of Chemistry and Physics*, 73rd ed. (CRC, Boca Raton, 1992).
- (32) DeMore, W. B.; Sander, S. P.; Golden, S. P.; Howard, C. J.; Golden, D. M.; Kolb, C. E.; Hampson, R. F.; Molina, M. J. *Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling*, 1997.
- (33) Chandra, A. K.; Uchimaru, T.; Sugie, M.; Sekiya, A. Correlation between hardness and activation energies for reactions of OH radical with halomethanes. *Chem. Phys. Lett.* **2000**, *318*, 69–71.

## Hierarchical Numerical Solution of Smoluchowski Equations with Rough Potentials

Polina Banushkina and Markus Meuwly\*

Department of Chemistry, University of Basel, Klingelbergstrasse 80,  
4056 Basel, Switzerland

Received September 1, 2004

**Abstract:** We present an efficient and numerically robust algorithm to follow diffusive processes on rough potential energy surfaces. The hierarchical nature of the algorithm (hierarchical discrete approximation or HDA) fully explores the fine- and coarse-grained structure of the underlying interaction potential. The present approach does not impose any restriction on the topology of the potential. The hierarchical grid allows to capture the roughness of the potential and achieve significant reduction of computational time using fewer grid points compared to other DA methods. HDA is shown to be accurate and efficient by comparing with results from the conventional DA and from the “mean first passage time” (MFPT) method. Using potential-optimized grids HDA monotonically converges to results from an analytical treatment for a very rough interaction potential (107 minima). Contrary to MFPT the solution from HDA is numerically stable. Because of the hierarchical structure of the method HDA can be extended to multidimensional problems.

### I. Introduction

The notion of structured (rough) energy landscapes is an important concept in understanding dynamical processes in complex systems. Examples include, but are not limited to, the folding of proteins<sup>1,2</sup> or the reaction of two end-groups in a polymer or polypeptide chain.<sup>3–5</sup> For most systems exhibiting complex behavior the underlying potential or free energy surface has a hierarchical structure with many local, functionally relevant minima.<sup>6,7</sup> As an example, the motion of a peptide or a protein takes place on an energy landscape characterized by a large number of conformational substates (CS).<sup>8–11</sup> At sufficiently low temperatures the system can be trapped in one of the many, essentially isoenergetic substates. The dynamics on this landscape is governed by the heights of the barriers between the local minima which can be of the order of several  $kT$ . An important question concerns the time scales on which an initial population of CSs relaxes toward a steady state which, in the case of a protein, is the native state. For this, one or several progression coordinates are usually defined along which an initial population evolves in space and time.<sup>12</sup>

It is of considerable interest to follow the reaction kinetics governed by this complex, rough potential energy surface. One possibility is to characterize the temporal and spatial relaxation of an initial distribution  $p(x, 0)$  to the final, steady-state (equilibrium) distribution  $p_{\text{eq}}(x)$ . To this end the Smoluchowski equation (SE) is solved for the particular potential energy function  $V(x)$  or the multidimensional energy landscape (for an illustrative example see Figure 9 in ref 9). Since  $V(x)$  may be rough (which implies variations on a short spatial dimension), with functionally relevant local minima, it is important to resolve as many details of  $V(x)$  as possible. Under such circumstances the entire morphology of  $V(x)$  is important, and averaging over its structural features would alter the predicted dynamics and time scales.

There exist several methods to solve SEs. They include finite-difference schemes in  $x$  and  $t$ ,<sup>12</sup> finite-differences in  $x$  with time propagation based on the formal solution of the time-dependent part,<sup>13,14</sup> basis set expansions,<sup>15,16</sup> computer simulations, or path integrals.<sup>15</sup> In the first approach one follows the evolution of  $p(x, t)$  using finite difference approximations in  $x$  and  $t$ . Such an approach (implicit scheme) is based on spatial and temporal discretization with resolution  $\Delta x$  and  $\Delta t$ . Since finding the solution at a later time requires knowledge about the solution of the SE at the

\* Corresponding author phone: +41 61 267 3821; fax: +41 61 267 3855; e-mail: m.meuwly@unibas.ch.

previous time steps implicit methods are best suited to investigate fast processes or situations where interconversion barriers are low.<sup>12</sup> The second approach uses the closed form for the probability distribution. This requires the diagonalization of the rate constant matrix constructed on a discrete spatial grid with resolution  $\Delta x$ . With an increasing number of grid points the computational effort significantly increases due to storage requirements and execution time. Both grow exponentially with the number of degrees of freedom. After diagonalization of the rate matrix the time evolution of  $p(x, t)$  can be readily evaluated at any given  $t$ . One method belonging to the latter class is the discrete approximation (DA). Although the first two methods are conceptually simple and appealing, their applicability to problems with realistic potential energy curves is limited.<sup>17</sup> With an increasing number of intermediate states,  $\Delta x$  must decrease and the number of grid points required increases. Such a procedure becomes impractical for very fine grids since the size of the rate matrix increases with an increasing number of CSs. Basis set expansions suffer from similar limitations and the fact that the required eigenvalues are only known for very simple interaction potentials.<sup>16</sup> The major concern in applying methods based on computer simulations is the reduction of statistical errors. Finally, progress has been made in using path integral techniques although numerically robust results at a reasonable computational effort were limited to smooth one-dimensional potentials.<sup>18,19</sup> Other methods, primarily related to the motion on rough potential energy curves, include approximations based on analytical mean first passage times.<sup>14,17,20</sup> Unfortunately, these approaches are restricted to one-dimensional systems, and generalizations to multidimensional dynamics are difficult and in fact seem not to have been successful so far.<sup>17</sup>

In this work we present a numerically robust and computationally efficient algorithm to solve the SE for rough potentials. The method explores the hierarchical structure of complex energy landscapes for which averaging over the roughness may not be possible. We apply the method to smooth and rough potentials and, if available, compare the results with previously developed methods. The model potentials we investigate include a harmonic potential  $V(x) = 1/2 kx^2$  (H1), a double well potential  $V(x) = \cos(3.14(x - 3))$  (A1), a rough one-dimensional potential  $V(x) = kx^2/2 + \epsilon(\cos(167x) + \sin(73x))$  (R1) where  $k = 19.84$  (kcal/mol)/ $\text{\AA}^2$ ,  $\epsilon = 0.595$  kcal/mol (see Figure 1) and two-dimensional potentials  $V(x, y) = 1/2 kr^2$  (H2),  $V(x, y) = -2\cos(6r) + (\exp(r - 0.7))^2$  (A2),  $V(x, y) = 5r^2 + 0.5(\cos(45(r - 0.0655)) + \sin(30(r - 0.0655)))$  (R2), where  $r^2 = x^2 + y^2$  and the last is a rough potential.

Each of the potentials relates to different physical situations. H1 is a test potential used in previous work<sup>14</sup> for the motion in a harmonic potential and the problem can be solved analytically, and A1 involves barrier crossing which becomes important for example in chemical reaction dynamics. R1 is a harmonic potential with a rough background, which was previously used and for which an algebraic treatment exists.<sup>14,17</sup> The same physical picture applies to the potentials in two dimensions.

## II. Description of the Hierarchical DA Method

In the large-friction limit the Fokker–Planck equation leads to the Smoluchowski equation which, in one spatial dimension  $x$ , reads

$$\frac{\partial p(x, t)}{\partial t} = \frac{\partial}{\partial x} D(x) e^{-\beta V(x)} \frac{\partial}{\partial x} [e^{\beta V(x)} p(x, t)] \quad (1)$$

Here,  $D(x)$  is a space-dependent diffusion constant,  $V(x)$  is the interaction potential,  $\beta = 1/kT$  is the Boltzmann factor and  $p(x, t)$  is the space and time-dependent probability distribution. In the following  $D$  is taken to be constant, i.e.,  $D(x) = D = 0.03 \text{\AA}^2/\text{s}$ . Using finite differences to approximate the derivatives, equation (1) can be recast as a Master equation<sup>13</sup>

$$\frac{\partial p(x_n, t)}{\partial t} = l(n|n+1)p(x_{n+1}, t) + l(n|n-1)p(x_{n-1}, t) - (l(n+1|n) + l(n-1|n))p(x_n, t) \quad (2)$$

Here  $l(n|n \pm 1)$  are the rates to move to the left and to the right from the starting point  $x_n$ .

The time evolution of  $p(x, t)$  in Equation (1) can be followed numerically whereby the space coordinate is discretized such that the continuous variable  $x$  takes discrete values  $x_n$ ,  $n = 1, \dots, N + 1$ . The solution to (1) may be obtained by solving equations (2). In DA the rate coefficient  $l(m|n)$  for making a transition  $x_n \rightarrow x_m$  is given by

$$l(m|n) = \frac{D(n) + D(m)}{2d^2} \exp\left(-\frac{\beta(V(m) - V(n))}{2}\right)$$

with  $d = x_{n+1} - x_n$  (see ref 13).

For rough potentials  $V(x)$  the number of grid points  $n$  required to resolve the roughness increases rapidly. This leads to large matrices that have to be accurately diagonalized. In the present work we explore the possibility to solve the (SE) on a hierarchy of grids by defining a coarse grid ( $N$ ) and a subgrid ( $M$ ) by dividing each interval  $(x_n, x_{n+1})$  into  $j = 1, \dots, M + 1$  points. The mean passage time  $\tau(n+1|n)$  from  $x_n$  to  $x_{n+1}$  is given by<sup>21</sup>

$$\tau(n+1|n) = \int_0^\infty \sum_{j=1}^M p(x_j, t) dt = \int_0^\infty (1 - p(x_{M+1}, t)) dt \quad (3)$$

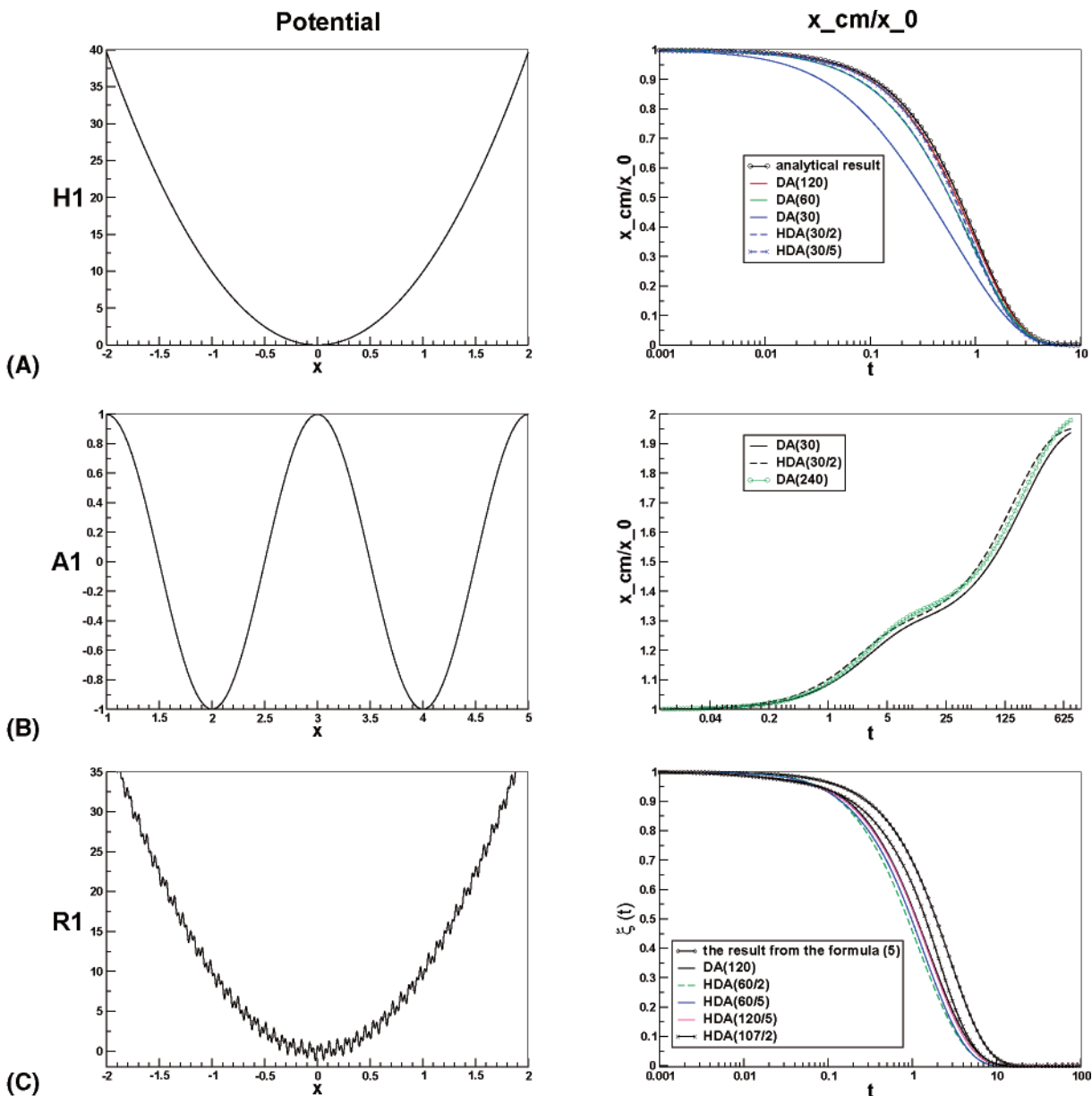
and the corresponding rate coefficients are  $l_{\text{HDA}}(n+1|n) = 1/\tau(n+1|n)$ . For finding  $\vec{p}_{\{M+1\}}(x, t)$  ( $\vec{p}_{\{M+1\}}(x, t)$  is the vector of probabilities in points  $x_1, \dots, x_{M+1}$  on each subinterval) we use the DA method, and the solution of (2) can be determined from

$$\vec{p}_{\{M+1\}}(x, t) = U \exp(\vec{\lambda}t) U^{-1} \vec{p}_{\{M+1\}}(x, 0)$$

The elements of matrix  $U$  and  $\vec{\lambda}$  are the eigenvectors and eigenvalues of the rate matrix, respectively. The boundary conditions on each subinterval are

$$l(1|0) = l(0|1) = 0,$$

$$l(M+2|M+1) = l(M+1|M+2) = 0,$$



**Figure 1.** The different 1-d potentials for which the Smoluchowski equation is solved together with the time evolution of the position of the center-of-mass coordinate  $x_{cm}(t)/x_0$  as a function of time. (A) For the harmonic potential HDA(30/M) with moderate values of M gives the known analytical result, while for conventional DA 120 points are required. (B) For the double well potential (A1) the initial distribution at  $x = 1.5$  relaxes to  $x_{cm} = 2$  which is halfway between the two minima of the symmetric potential. HDA(30/2) is found to yield dynamics very close to DA(240). (C) R1 has 107 minima with the global minimum at  $x = -0.06$ . All methods show decay to the same limiting  $\xi(t)$ , while their dynamics differs somewhat. Note that  $t$  is on a logarithmic scale throughout.

$$l(M|M+1) = 0$$

and the initial condition for the probability distribution is a  $\delta$ -function at the first point of each subinterval. Taking  $\vec{p}_{\{M+1\}}(x, 0) = \{\delta_{ip_1}\}$  the probability  $p(x_{M+1}, t)$  at grid point  $x_{M+1}$  (required for  $\tau(n+1|n)$ ) is

$$p(x_{M+1}, t) = \sum_k U_{M+1,k} \exp(\lambda_k t) U_{k,1}^{-1}$$

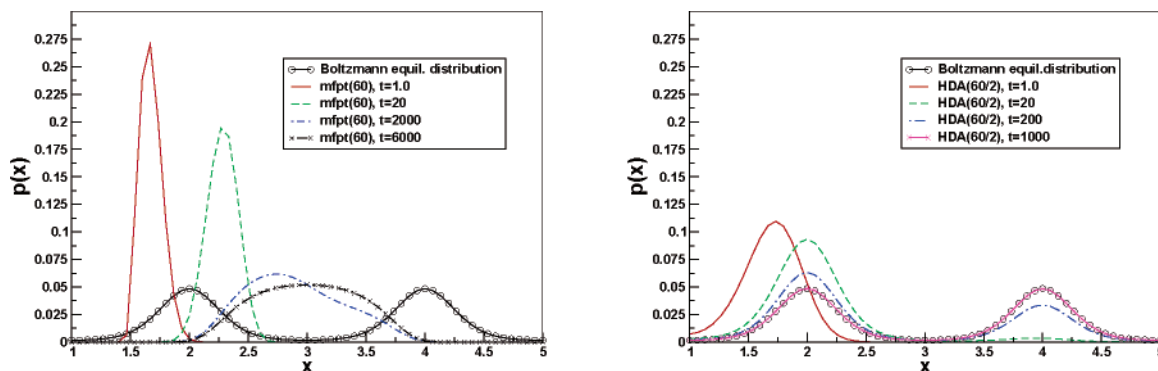
where  $U_{M+1,k}$  and  $U_{k,1}^{-1}$  are the  $(M+1, k)$ -th and  $(k, 1)$ -th elements of the matrices  $U$  and  $U^{-1}$ , respectively. Substituting this expression into equation (3) the integral can be evaluated analytically to yield

$$\tau(n+1|n) = \sum_k U_{M+1,k} \frac{1}{\lambda_k} U_{k,1}^{-1}$$

The coefficients  $l_{\text{HDA}}(n|n+1)$  are determined from the condition of detailed balance

$$l_{\text{HDA}}(n|n+1) = \frac{l_{\text{HDA}}(n+1|n)p_e(n)}{p_e(n+1)} \quad (4)$$

where  $p_e(x)$  is the Boltzmann equilibrium probability distribution for the potential  $V(x)$ .



**Figure 2.** Normalized probability distributions  $p(x, t)$  for different times  $t$  calculated by the MFPT method (ref 14) and HDA for the A1 potential. For both cases the stationary Boltzmann equilibrium distribution  $p_e(x)$  is also given. As shown  $p(x, t)$  from MFPT(60) does not converge to  $p_e(x)$ , while the distribution from HDA(60/2) does.

### III. Results

In all cases reflecting boundary conditions are used for the coarse grid  $N$ . The left and right boundaries are  $[-2, 2]$  for H1 and R1 and  $[1.5]$  for A1. For the 2-d calculations they are a square on  $[-1, 1]$  in both dimensions for H2 and R2 and  $[-1.5, 1.5]$  in both dimensions for A2. In the following we will use the notation DA( $N$ ) for a DA calculation with  $N$  spatial grid points and HDA( $N/M$ ) for an HDA calculation with  $N$  coarse grid points and  $M$  points in the subgrid.

**1-Dimensional Potentials.** A convenient measure for the relaxation from the initial distribution  $p(x, 0)$  is to follow the center-of-mass  $\xi(t) = x_{cm}(t)/x_0$  of the distribution, where  $p(x, 0)$  is the initial distribution which is a  $\delta$ -function located at the nearest grid point to  $x_0 = -1.5, 1.5$  and  $-1.0$  for cases H1, A1, and R1, respectively. In Figure 1 the time evolution of  $\xi(t)$  of the distribution function calculated by DA and HDA is shown for the three different 1-d potentials.

For the harmonic potential H1, DA(30) is far from the analytical result, while HDA(30/2) and HDA(30/5) rapidly approach the exact  $\xi(t)$ . Convergence toward the known result is much slower for conventional DA and only achieved for DA(120). The relative speedup of HDA compared to conventional DA is about a factor of 50.

For the anharmonic potential (A1) DA(30) yields a somewhat slower decay toward the minimum than HDA(30/2) which gives  $\xi(t)$  in close agreement to the one calculated with DA(240) (see Figure 1B). It is only in the very long time limit ( $t > 500$ s) that  $\xi(t)$  from HDA(30/2) and DA(240) start to show appreciable differences. Calculations with DA(1000) give  $\xi(t)$  marginally different from DA(240). Using MFPT(60), the method from ref 14,  $\xi(t)$  converged to the correct long-time limit ( $\xi(t) = 2$ ). However, the overall behavior of  $\xi(t)$  was different from the ones shown in Figure 1B using either DA or HDA. To understand this further  $p(x, t)$  at different time  $t$  was calculated and is shown in Figure 2. They were calculated with MFPT(60) (the method from ref 14) and HDA(60/2). Using the MFPT-method  $p(x, t)$  has reached the equilibrium distribution  $p_e(x)$  for  $t > 6000$  while  $p(x, t)$  from HDA is stationary after  $t > 1000$  which means that the two methods yield different dynamics. As is shown,  $p(x, t \rightarrow \infty)$  from MFPT does not converge to the known  $p_e(x)$  although  $\xi(t)$  reaches the correct long time limit. One possible explanation is that the rate coefficients used

in the MFPT method do not explicitly fulfill the condition for detailed balance (equation (4)).

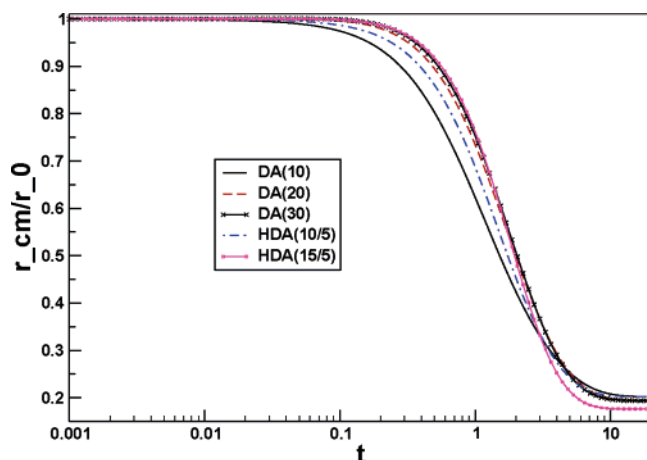
Besides the motion in a harmonic and bistable potential, energy profiles with roughness are of interest. Previous work<sup>14,17</sup> has established that the average position  $x_{cm}(t)$  of the particle diffusing on the R1 potential (roughness with a harmonic background) can be found from the analytical formula for H1 with a modified effective diffusion coefficient  $D'$

$$x_{cm}(t) = x_0 \exp(-k\beta D't) \quad (5)$$

where  $D' = D \exp[-(\epsilon/k\beta T)^2]$ . Generalizations to potentials with different than a harmonic background apparently have not yet been discussed in the literature. The R1 potential has 107 local minima on the interval  $[-2, 2]$ . In Figure 1C the results from DA, HDA and the result from formula (5) are shown. The latter result is taken as the reference. One possible (but somewhat arbitrary) measure to compare the different methods is the time  $\tau$  at which  $\xi(t)$  has reached its half time value  $\xi(t) = 0.5$ . For the result from formula (5)  $\tau = 1.9$  while for DA(120)  $\tau = 1.2$ . Using HDA(60/2) and HDA(60/5) gives faster kinetics (with  $\tau = 0.89$  and  $\tau = 1.07$ , respectively). It is interesting to note that with increasing  $M$  at constant  $N$  the solution from HDA systematically improves and converges. However, with equidistant points it was not possible to approach the result from formula (5).

With DA and HDA it is also possible to use nonequidistant grids. Locating the  $x_i$  at the local minima of R1, i.e.,  $N = 107$ , the  $\xi(t)$  from HDA(107/2) is closer to the  $\xi(t)$  calculated from formula (5) than  $\xi(t)$  from DA(120) and for HDA(120/5) and  $\xi(t) = 0.5$  was reached at  $\tau = 1.52$ . Previously it was shown that the MFPT method gives essentially exact results even with  $N = 30$ .<sup>14</sup> However, by increasing  $N$  to  $N = 35$  we found that MFPT is unstable. For the R1 potential  $\xi(t)$  should be strictly bounded ( $1 \leq \xi(t) \leq 0$ , see Figure 1C). However, for  $N > 30$   $\xi(t)$  from MFPT takes values outside this interval. Contrary to that, HDA monotonically converges for fixed  $N$  and increasing  $M$ .

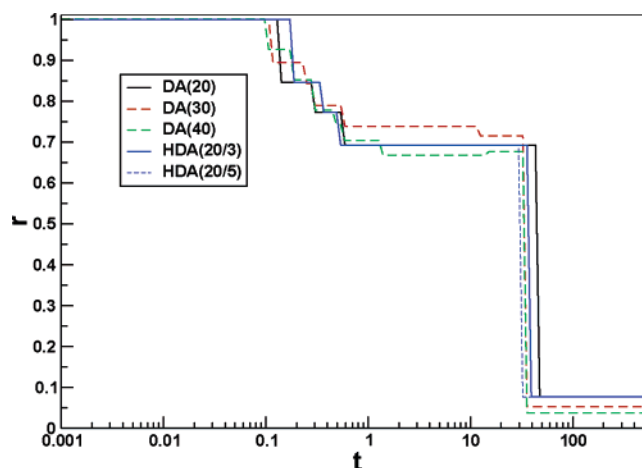
**2-Dimensional Potentials.** In 2 spatial dimensions the application of conventional DA already becomes prohibitive. The maximum number of intervals possible was  $N = 50$  in each dimension. Here we used the same number of intervals



**Figure 3.** The mean radius  $\xi(t) = (r_{cm}(t)/r_0)$  of  $p(x, y, t)$  is plotted as a function of time for the H2 potential. The limiting values for  $\xi(t \rightarrow \infty)$  are different because the locations of the grid points differ for  $N$  even and  $N$  odd. The dynamics for DA(30) and HDA(15/5) are virtually identical. Computing times, however, differ by a factor of 30 (see text). The time  $t$  is on a logarithmic scale.

in the  $x$  and  $y$  direction. However, depending upon the topology of  $V(x, y)$  this might not be the optimal choice. There are different possible measures how to compare the relaxation from  $p(x, y, 0)$  to  $p_e(x, y)$  ( $p(x, y, 0)$  is a  $\delta$ -function located at the nearest grid point to  $(x_0, y_0) = (-0.8, -0.8)$  for H2;  $(-1, -1)$  for A2 and  $(-0.9, -0.9)$  for R2, respectively). One possibility is to follow the mean value of the length of the radius vector of the distribution  $p(x, y, t)$ . The radius vector is measured from the point  $(0, 0)$  to the point  $(x_i, y_j)$ , and we define the mean radius as  $r_m(t) = \sum_{i,j=1}^{N+1} r(x_i, y_j) p(x_i, y_j, t)$ . To compare the results of the DA and HDA method on the H2 potential the time evolution of the value  $\xi(t) = r_m(t)/r_0$  is followed, where  $r_0 = \sqrt{x_0^2 + y_0^2}$ . Figure 3 shows that  $\xi(t)$  from HDA(10/5) is close to DA(20) and HDA(15/5) approaches the result from DA(30). With HDA(20/5) (results not shown)  $\xi(t)$  is virtually identical to the one from DA(30). The different long-time value for  $\xi(t \rightarrow \infty)$  for DA(30) and HDA(15/5) is a consequence of the different locations of the grid points. To calculate the decay from the initial distribution  $p(x, y, t)$  was followed for 80 time steps. The computing times were 0.3, 2, 5 and 77 min for HDA(10/5), HDA(15/5), DA(20) and DA(30), respectively, on an AMD 1666 MHz workstation.

For the A2 potential  $\xi(t) = r_m(t)/r_0$  turned out to be not a suitable measure. We found that the time dependence of  $\xi(t)$  depends on the location of the grid points. Even small differences in the values of the probability distribution calculated by DA and HDA at the same time lead to considerable changes in  $\xi(t)$ . It is more useful to monitor the position of the maximum  $p_{max}(x, y, t)$  relative to its initial position  $p(x_0, y_0, 0)$ , i.e., the time evolution of  $g(t) = r_{max}(t)/r_0$ , where  $r_{max}(t)$  and  $r_0$  are the distances of  $p_{max}(x, y, t)$  and  $p(x_0, y_0, 0)$  from the point  $(0, 0)$ , respectively. The function  $g(t)$  is not continuous since the maximum can be localized for extended periods of time on the same grid point. Initially,  $p_{max}(x, y, t)$  is in  $(x_0, y_0)$ . Although  $p(x, y, t)$  changes in time, the position of the maximum resides in  $(x_0, y_0)$  up to

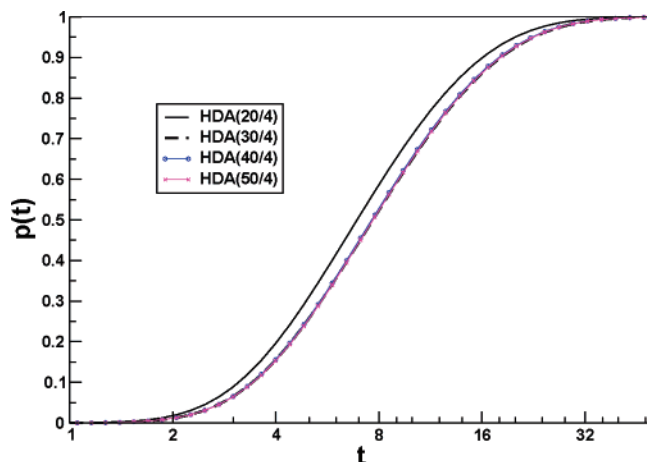


**Figure 4.** Position of the maximum  $p_{max}(x, y, t)$  relative to its initial position  $p(x_0, y_0, 0)$  as a function of time for the A2 potential. Results from calculations with different grid sizes for DA and HDA are shown. The long-time value of  $g(t)$  depends on the number of grid points.

$t = 0.1$ . This can be understood from the shape of A2 which is a rotated double well potential around  $(0, 0)$ . At early times  $p(x, y, t)$  spreads within the outer rim and up to  $t = 0.7$  the position of  $p_{max}(x, y, t)$  changes marginally. Between  $t = 1$  and  $t = 20$  the global minimum is populated, which, after about  $t = 30$  becomes more populated than any other region. For DA(20) the time  $t_{lim}$  after which  $g(t)$  stabilizes (see Figure 4) is reached significantly later ( $t_{lim} = 49$ ) than for DA(30) and DA(40) for which  $t_{lim} = 35$  (Figure 4). For HDA(20/3) and HDA(20/5)  $t_{lim}$  are 38 and 33.5, respectively, close to the results from DA(30) and DA(40). However, the CPU time for DA(30) is 110 min compared to 6 min for HDA(20/3). Again, the limiting value  $g(t \rightarrow \infty)$  depends on the grid spacing. This is why the curves in Figure 4 do not converge to the same value for  $g(t)$ . As in the case of (H1) and (A1) rapid convergence of the observables for increasing  $M$  at fixed  $N$  is found.

Since for A2 the distribution of the grid points influences the analysis of how  $p(x, y, t)$  relaxes this is even more so for the rough R2 potential. The R2 potential has about 160 local minima. Grids with  $N = 30, 40, 50$  have their points in different places, and the measure for how  $p(x, y, t)$  evolves in time will depend on this (see also discussion of Figure 4). To avoid this problem for R2 we follow the distribution of the first passage times for the system to reach the minimum starting from the initial distribution. The value  $p(t)$  is the probability that the first passage time to reach the minimum is  $t$ . The distribution  $p(t)$  was calculated for HDA(20/4), HDA(30/4), HDA(40/4), HDA(50/4). As is shown in Figure 5 HDA(30/4) gives the same result as HDA(50/4). This is in line with the observations for R1 where we also found that at fixed  $M$  the results from HDA rapidly converge for increasing  $N$  (compare HDA(60/5) and HDA(120/5) in Figure 1). However, for this potential we found that increasing the number of inner points  $M$  does not change the behavior of  $p(t)$ .





**Figure 5.** Temporal evolution of the distribution  $p(t)$  of the first passage times in the R2 potential. For given  $M = 4$  convergence of  $p(t)$  is reached at  $N = 30$ .

#### IV. Discussion and Conclusion

Numerical approaches to solve the SE for a general form of the (rough) potential energy surface are a useful complement to analytical formulas to treat barrier crossing and escape processes. This is mainly because analytical work gives exact results only for limiting cases such as high barriers, weak or large friction, or particular forms for the potential.<sup>22</sup> In this work a numerically stable and computationally efficient algorithm to solve the SE for rough potential energy surfaces, where averaging over the coarse- and fine-grained structure of the potential is not possible, has been developed. Such approaches become increasingly important since free energy simulations are possible routinely and will provide free energy surfaces in several dimensions.<sup>23,24</sup> To demonstrate its accuracy the SE was first solved for a number of test potentials for which other solution strategies have been presented in the literature. For smooth 1d potentials, results from analytical solutions and the MFPT method serve as comparison for the HDA method. Application of DA and HDA to an anharmonic and rough 1d potential showed that the MFPT method<sup>14</sup> has to be used with circumspection. For A1 MFPT does not converge to the known Boltzmann distribution at long times, and for R1 MFPT(30) yields essentially exact results while MFPT(35) is unstable. For some cases considered here, MFPT appears not always to converge properly upon increasing the number of grid points  $N$  as expected. It is interesting to note that a previous study<sup>16</sup> employed 301 grid points to numerically solve the Smoluchowski equation for a 1d, symmetric bistable potential. Here, a similar number of grid points was deemed sufficient (DA(240) for potential A1) to reach convergence, while HDA(30/2) yields dynamics very similar to DA(240) at much reduced computational cost (eigenvalues of a  $30 \times 30$  instead of a  $240 \times 240$  matrix). It is clear that in higher dimensions the DA formulation of this problem is computationally intractable (see discussion of A2 potential). An important advantage of HDA for rough potentials is that the method converges monotonically for given  $N$  and increasing  $M$ . HDA calculations on the very rough R1 potential (107 minima) using a nonuniform, potential-adapted grid leads to kinetics similar to the one based on the approach by Zwanzig where

the roughness is treated as a modified diffusion constant.<sup>17</sup> Further exploration of how to optimally place the grid points for more rapid convergence is worthwhile in particular for problems in more than 1 dimension.

Applying a hierarchical separation considerably speeds up calculations while retaining accuracy. In 2 dimensions HDA scales as  $(M)^3 \times (N_1)^2 + (N_1^2)^3$ , while conventional DA scales as  $(N_2)^3$  with  $N_1 \ll N_2$ . Because  $M$  is considerably smaller than  $N_1$  there is an appreciable net overall speedup. For example HDA(15/5) gives the same results at 30 times less CPU time than DA(30) for the 2-dimensional harmonic potential (H2). Exploring the hierarchical structure of the potential or free energy landscape is essential for cases where the many local minima correspond to functional states of a system. Smoothing such a potential is not possible (as would be for noise) since this eliminates important physically, chemically and biologically relevant information. Most numerical and analytical approaches currently available have not yet been applied to this class of potentials.

In summary, HDA is a computationally efficient and numerically robust method for rough potentials. Since HDA can be applied recursively still finer subgrids can be used. This together with more sophisticated ways to place the grid points open possibilities to follow the temporal evolution of the dynamics on rough potentials in more than 2 dimension. Given the impediments for other approaches such as storage limitations (basis set expansions), convergence issues (path integral methods) and reduction of statistical errors (computer simulations)<sup>16,18,19</sup> HDA provides an attractive alternative for approximate solutions of Smoluchowski equations for multidimensional systems involving rough interaction potentials. The applicability of HDA to more general formulations of diffusive motion is currently explored.

**Acknowledgment.** The authors acknowledge financial support from the Schweizerischer Nationalfonds. M.M. is grateful for the award of a Förderungsprofessur. Comments on the manuscript from Dr. S. Krivov, Prof. J. D. Doll, Prof. T. Kiefhaber, and Dr. R. Bemish are gratefully acknowledged.

#### References

- (1) Succi, N.; Onuchic, J.; Wolynes, P. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* **1996**, *104*, 5860.
- (2) Yang, W. Y.; Gruebele, M. Detection-dependent kinetics as a probe of folding landscape microstructure. *J. Am. Chem. Soc.* **2004**, *126*, 7758.
- (3) Wilemski, G.; Fixman, M. Diffusion-controlled interchain reactions of polymers. 1. *Theory J. Chem. Phys.* **1974**, *60*, 866.
- (4) Krieger, F.; Fierz, B.; Bieri, O.; Drewello, M.; Kiefhaber, T. Dynamics of Unfolded Polypeptide Chains as Model for the Earliest Steps in Protein Folding. *J. Mol. Biol.* **2003**, *332*, 265.
- (5) Wang, X. J.; Nau, W. Kinetics of end-to-end collision in short single-stranded nucleic acids. *J. Am. Chem. Soc.* **2004**, *126*, 808.

- (6) Jensen, M.; Park, S.; Tajkhorshid, E.; Schulten, K. Energetics of glycerol conduction through aquaporin GpF. *Proc. Natl. Acad. Sci.* **2002**, *99*, 6731.
- (7) Krivov, S.; Chekmarev, S.; Karplus, M. Potential energy surface and conformational transitions in biomolecules: a successive confinement approach applied to a solvated tetrapeptide. *Phys. Rev. Lett.* **2002**, *88*, 038101.
- (8) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. Dynamics of ligand-binding to myoglobin. *Biochem.* **1975**, *14*, 5355.
- (9) Dobson, C. M.; Sali, A.; Karplus, M. Protein Folding: A Perspective from Theory and Experiment. *Angew. Chem., Int. Ed.* **1998**, *37*, 868.
- (10) Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Phys. Chem.* **1997**, *106*, 1495.
- (11) Wagner, C.; Kiefhaber, T. Intermediates can accelerate protein folding. *Proc. Natl. Acad. Sci.* **1999**, *96*, 6716.
- (12) Jun, B.; Weaver, D. One-dimensional potential barrier model of protein folding with intermediates. *J. Chem. Phys.* **2002**, *116*, 418.
- (13) Bicout, D. J.; Szabo, A. Electron-transfer reaction dynamics in non-Debye solvents. *J. Chem. Phys.* **1998**, *109*, 2325.
- (14) Ansari, A. Mean first passage time solution of the Smoluchowski equation: Application to relaxation dynamics in myoglobin. *J. Chem. Phys.* **2000**, *112*, 2516.
- (15) Risken, H. *The Fokker–Planck equation*. Springer, Berlin, Heidelberg, New York, 1989.
- (16) Dunkel, R.; Ebeling, W.; Schimansky-Geier, L.; Hänggi, P. Kramers problem in evolutionary strategies. *Phys. Rev. E* **2003**, *67*, 061118.
- (17) Zwanzig, R. Diffusion in a rough potential. *Proc. Natl. Acad. Sci.* **1988**, *85*, 2029.
- (18) Drozdov, A. N.; Talkner, P. Path integrals for Fokker–Planck dynamics with singular diffusion: Accurate factorization for the time evolution operator. *J. Chem. Phys.* **1998**, *109*, 2080.
- (19) Drozdov, A. N.; Hayashi, S. Self-similar renormalization approach to barrier crossing processes. *Phys. Rev. E* **1999**, *60*, 3804.
- (20) Goychuk, I.; Hänggi, P. Ion channel gating: A first-passage time analysis of the Kramers type. *Proc. Natl. Acad. Sci.* **2002**, *99*, 3552.
- (21) Szabo, A.; Schulten, K.; Schulten, Z. 1st passage time approach to diffusion controlled reactions. *J. Chem. Phys.* **1980**, *72*, 4350.
- (22) Hänggi, P.; Talkner, P.; Borkovec, M. Reaction Rate Theory: 50 years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251.
- (23) Simonson, T.; Archontis, G.; Karplus, M. Free energy simulation come of age: Protein–ligand recognition. *Acc. Chem. Res.* **2002**, *35*, 430.
- (24) Ulmschneider, J. P.; Jorgensen, W. L. Polypeptide folding using Monte Carlo sampling, concerted rotation, and continuum solvation. *J. Am. Chem. Soc.* **2004**, *126*, 1849.

CT0499480

## A Theoretical Study on the Factors Influencing Cyanine Photoisomerization: The Case of Thiacyanine in Gas Phase and in Methanol

R. Improta<sup>†</sup> and F. Santoro<sup>\*‡</sup>

*Istituto di Biostrutture e Biommagini del CNR, via Mezzocannone 6, I-80134 Napoli, Italy, and Istituto per i Processi Chimico-Fisici del CNR, Area della Ricerca del CNR di Pisa, via Moruzzi 1, I-56124 Pisa, Italy*

Received November 4, 2004

**Abstract:** The effects influencing cyanine photoisomerization on the  $S_1$  surface in the condensed phase have been investigated by an integrated quantum mechanical approach, focused mainly on 3,3'-diEt-2,2'-thiacyanine. After excitation, a barrierless motion, involving the torsion coupled to bond skeletal deformation, leads to a slightly nonplanar local  $C_2$  minimum, which we propose to be the fluorescent state. Crossed a barrier of  $\approx 120 \text{ cm}^{-1}$ , a steeper path drives to a more stable  $C_1$  minimum  $S_1$ -Min, corresponding to a pseudoperpendicular twisted intramolecular charge transfer (TICT) state. CAS(6,6) optimization allows for locating the lowest energy  $S_1/S_0$  conical intersection in the isomerization path which is reached from  $S_1$ -Min by an increased asymmetry of the two rings and a marked pyramidalization at one N center. The  $S_1$  surface is rather flat in the Franck–Condon region and suggests that other paths can be competitive with the minimum energy one. The comparison among different cyanines shows how variation of the molecular scaffold and/or of its substituents modulate the dynamics of the photoisomerization. All the indications coming from our computations are in line with and provide an explanation to the available experimental results.

### 1. Introduction

In the past decades the continuous progress in ultrafast laser technology has allowed the rise of a new fascinating research field named femtochemistry,<sup>1,2</sup> devoted to the understanding and, more recently, to the control of the elementary excited-state reactive chemical processes.<sup>3</sup> An ultrafast pump pulse prepares the excited molecule in a coherent state, a matter wave packet, which explores the electronic surface in the nuclear coordinates space moving according to quantum mechanics. The complexity of the phenomena which can take place, due to the rich topologies of the molecular excited-state surfaces, requires both experimental and theoretical studies for their full understanding.<sup>2</sup> With the aim to get highly accurate results, most of the existing theoretical studies

on excited-state reactivity have been performed in the gas phase on relatively small model systems, in which the “photoactive moiety” is as small as possible. These kinds of studies have allowed a fundamental advance in the knowledge of the basic mechanisms and processes involved in the excited-state reactivity, just think for example about the clarification of the primary role played by conical intersections between electronic states in determining the fate of photochemical processes.<sup>4</sup>

Parallel studies are desirable to understand how these basic mechanisms and processes apply in relatively large molecules and how they are affected by the existence of bulky and/or aromatic substituents linked to the photoactive moiety, as it often happens in molecules of technological interest. While these studies have been traditionally tackled from a “physical” perspective, introducing a bath of intramolecular modes to simulate decoherence and dissipative effects,<sup>5</sup> new computational technologies and the increased computer speed

\* Corresponding author e-mail: f.santoro@ipcf.cnr.it.

<sup>†</sup> Istituto di Biostrutture e Biommagini del CNR.

<sup>‡</sup> Istituto per i Processi Chimico-Fisici del CNR.

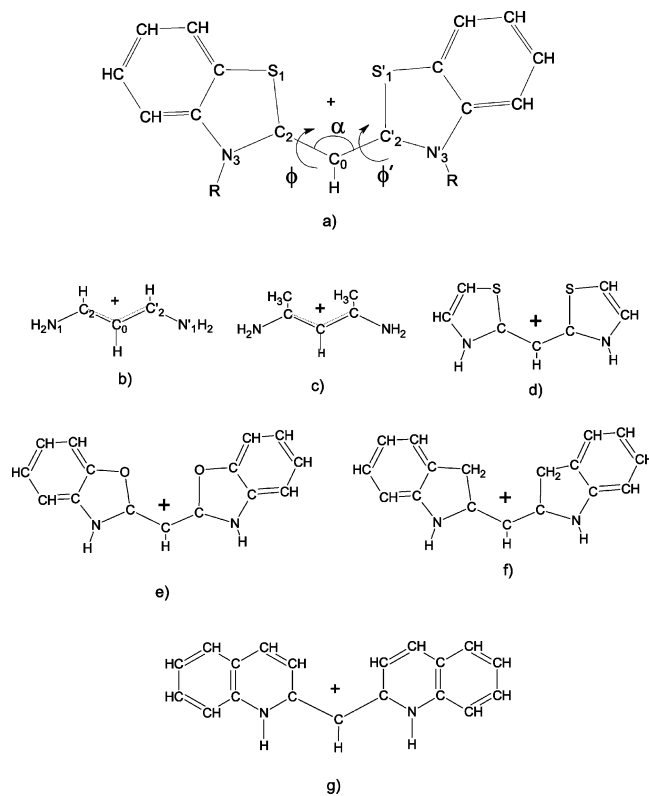
are making feasible a more satisfactorily approach, which takes into account the chemical specificity of the substituents including them explicitly in the electronic calculations. In this spirit, for example, an increasing number of “on the fly” quantum or semiclassical dynamical studies has been appearing in the literature.<sup>6</sup>

The hope and the ultimate goal of studies on the chemical effects of substituents on photochemical processes is to contribute to the establishment of some general and transferable propensity rules which allow the “factive” knowledge the chemists look for, enabling them to predict and, at the occurrence, to alter the dynamics and the fate of the process of interest. In this paper we deal with photoisomerization processes through twisting around a double bond. These reactions are widely studied for their biological and technological interest both in prototype molecules as ethylene and stilbene<sup>7</sup> and in molecules more directly involved in biological processes as retinal<sup>8</sup> or in important technological applications as cyanines.<sup>9–11</sup>

Besides exhibiting intriguing properties in their aggregated form,<sup>12,13</sup> cyanines, due their photophysical behavior, can be considered a prototypical case of study.<sup>14–29</sup> A trans–cis isomerization around one of the CC bonds of the polymethine chain is involved in the  $S_1 \rightarrow S_0$  deactivation process.<sup>25,26</sup> This feature explains why cyanines exhibit rather short excited-state lifetime and a low fluorescence quantum yield,<sup>19–22</sup> unless they are dissolved in very viscous solvents<sup>27</sup> or adsorbed on surfaces.<sup>23</sup> The effectiveness of the deactivation depends on general and basic physicochemical effects, such as the initial driving force to the twisting, the existence of a minimum on the excited surface, the location of the conical intersection relative to the minimum, the necessity for an IVR (intramolecular vibrational redistribution) during the dynamics, that are operative in several classes of compounds. This increases the general interest of understanding the factors that modulate the trans–cis isomerization, the deactivation mechanisms, and the lifetime of the excited state of cyanines. Therefore here we report a quantum mechanical study of several cyanines, sketched in Figure 1. This paper, by means of a series of static calculations (dynamics simulation will be the object of a forthcoming study), is aimed to get a close picture of the cyanine excited-state behavior, tackling the following questions:

(i) Do the substituents on the photoactive moiety change the character of the electronic transitions? (ii) Do the eventual differences influence their excited-state dynamics? (iii) If the answer to the previous points is positive, how can the effect of the nature of the atoms of the fused rings and/or of its substituents be rationalized? (iv) What is the role played by steric hindrances? (v) To what extent the photophysical behavior is modulated by environmental effects? (vi) Do the substituents have an understandable effect also on the location and the energetic of the conical intersection?

In our study we selected 3,3'-diEt-2,2'-thiacyanine iodide (Et-TCY according to ref 14) as the reference compound, since in the literature can be found a good amount of data concerning its stationary and time-resolved absorption and fluorescence spectra<sup>14,15</sup> and resonance Raman spectra.<sup>17</sup> Furthermore it has been utilized in the first coherent control



**Figure 1.** Schematic drawing of the different cyanine studied in the present paper: (a) Et-TCY (R = ethyl group) and H-TCY (R=H); (b) stCY; (c) CH<sub>3</sub>-stCY; (d) H-TCY-sm; (e) OCY; (f) CH<sub>2</sub>CY; (g) 2-Quino.

experiment of a photoisomerization process in the condensed phase,<sup>22</sup> and an exploration of the  $S_1$  reactive surface is necessary to get insights on the mechanism of the control induced by the shaped laser pulse.

To distinguish between conjugation and steric effects on the photoisomerization we also investigate a similar molecule (H-TCY, see Figure 1) in which ethyl groups are replaced by hydrogen atoms. Comparing Et-TCY and H-TCY with the simple stCY, with oxacyanine (OCY) and CH<sub>2</sub>CY (two analogues of H-TCY in which the sulfur atoms are substituted by oxygen atoms and CH<sub>2</sub> groups, respectively), and with 2-Quino (in which the nitrogen atom is inserted in a six-member ring) will help to highlight the effect of fused aromatic ring and of their chemical nature.

Our study has been performed both in the gas phase and in methanol, through the inclusion of explicit solvent molecules and the adoption of polarizable continuum models.<sup>30</sup> This represents a significant improvement with respect to gas-phase models, allowing to take into proper account, and hence to investigate, the electrostatic effects of the solvent at least on the initial part of the excited electronic surface, i.e., on the driving force to the bond twisting.

From the methodological point of view, the theoretical study of the excited state reactivity, in the condensed phase, of molecules containing more than 40 atoms is not a trivial task, that can be difficultly tackled by using a single method. We have thus resorted to an integrated computational approach, in which several quantum mechanical methods have been cross-utilized in order to couple the computational

feasibility with the maximum of accuracy attainable at the moment for studying such large compounds. On the other hand, we resort to the conceptually simple but powerful traditional tools of the theoretical chemistry, such as molecular orbitals and population analysis, trying to infer general and transferable information on the driving forces of the photoisomerization reaction and on the physical-chemical effects that can influence it.

## 2. Computational Details

Due to the large size of the system under investigation, an extended use of the CASSCF/MP2 approach (utilized in ref 29) for the characterization of the stationary points on the  $S_1$  surface and the minimum energy path is not feasible, we thus utilized different computational methods. Vertical transition energies have been calculated by Time-Dependent density functional (TD-DFT), employing the PBE0<sup>31</sup> functional. Despite the absence of adjustable parameters, when employed in TD-DFT calculations, PBE0 (TD-PBE0) has already provided excitation spectra in very good agreement with the available experimental results.<sup>32,33</sup> We have checked that an extension of the basis set beyond the 6-31+G(d,p) does not significantly change the computed spectra.

Geometry optimizations on the  $S_1$  surface have been performed at the CIS/6-31G(d) level, refining the energies by single-point TD-PBE0/6-31+G(d,p) calculations.

The lowest-energy  $S_0/S_1$  conical intersection (CI) in the isomerization path has been located by CASSCF/6-31G(d) and the method of Bearpark et al.<sup>34</sup> Unfortunately the inclusion of the full  $\pi$ -orbitals active space is computationally prohibitive for the system under investigation. We adopted SA-2-CAS(6,6), a state-averaged CASSCF with equal weights for  $S_0$  and  $S_1$ , with 6 electrons in 6 orbitals (6,6), the largest feasible active space for such a search. We checked the dynamical correlation effect by TD-DFT, which very recently<sup>35</sup> has been proven to be able to reproduce the CASPT2 results at CI points in a model of retinal, calculated with a full  $\pi$ -orbitals active space.

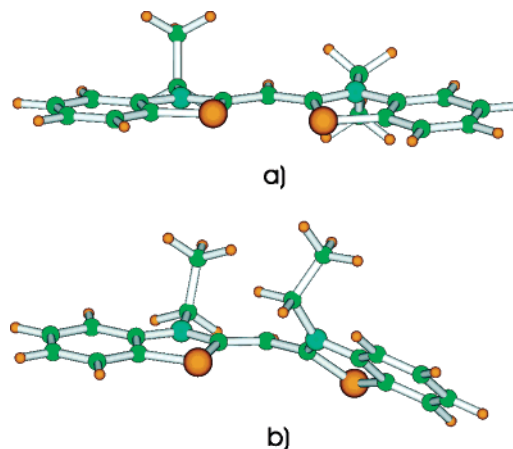
Bulk solvent effects on the ground and the excited states have been taken into account by means of the Polarizable Continuum Model (PCM).<sup>30</sup> In this model the molecule is embedded in a cavity surrounded by an infinite dielectric, with the dielectric constant of the solvent (for methanol we have used the value 32.64). The cavity of the solute is defined in terms of interlocking spheres centered on non-hydrogen atoms, whose radii are optimized according to the UAHF model.<sup>36</sup> PCM/TD-PBE0 calculations have been performed according to the procedure outlined in ref 37.

Atomic charges and Wiberg bond orders<sup>38</sup> have been calculated on the ground of the natural bond orbital (NBO) analysis.<sup>39</sup>

All the calculations have been performed by using the Gaussian03 package.<sup>40</sup>

## 3. Results

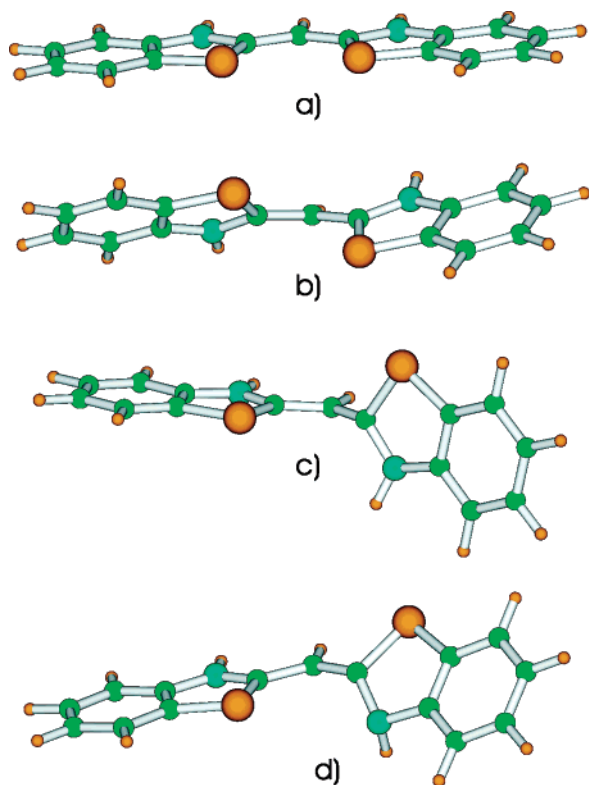
While important theoretical contributions have been given to the qualitative understanding of the cyanine's photoisomerization,<sup>28</sup> only one high-accuracy, full  $\pi$ -orbitals active space, CASSCF/MP2 extended theoretical investigation has



**Figure 2.** Schematic drawing of the PBE0/6-31G(d)  $S_0$  geometry minima of Et-TCY: (a) trans minimum and (b) cis minimum.

been reported in the literature by Galvez et al.,<sup>29</sup> on the simple trimethine streptocyanine (stCY) reported in Figure 1, and on its five- and seven-unit polymethine analogues. As a first step of our analysis we have tested the computational approach described above on stCY. The comparison of our results with those of ref 29 (see Supporting Information for a detailed discussion) is comforting and supports the reliability of our calculations. The qualitative description of the  $S_1$  PES obtained at the TD-PBE0/6-31G(d)//CIS/6-31G(d) level is very similar to that provided by CASMP2 calculation. Contrary to previous suggestion,<sup>26</sup> the initial relaxation of stCY on the  $S_1$  PES is barrierless and can be described by a two-mode model, a skeletal stretching strongly coupled to a torsional deformation of the carbon framework, which leads toward a local  $C_2$  minimum and then to the  $S_1$  minimum, where the two amino groups exhibit a pseudoperpendicular arrangement. This minimum, corresponding to a Twist Intramolecular Charge Transfer (TICT), is slightly more stable than an adjacent  $S_1/S_0$  conical intersection (CI) providing a path for the nonadiabatic evolution to  $S_0$ , either to the trans or the cis minimum. From the quantitative point of view, the closest the system is to planarity the more accurate are the results of our approach, that instead underestimates the  $S_1 - S_0$  gaps of the TICT pseudoperpendicular state, confirming the possible deficiencies of TD-DFT in describing charge-transfer transitions.<sup>41,42</sup> As we will show below, the effect of the different substituents to the central trimethine moiety is already operative in the initial steps of the reaction, when the molecule is still close to planarity, and consequently their existence and nature is predicted to influence significantly the dynamics of the reaction.

**3.1. Vertical  $S_0 \rightarrow S_1$  Electronic Transition.** Figures 2 and 3 show the trans and cis equilibrium structures of Et-TCY (Et-TCY and Et-TCY<sup>cis</sup>) and H-TCY (H-TCY and H-TCY<sup>cis</sup>), respectively, and Tables 1 and 2 report their main geometrical parameters optimized at the PBE0/6-31G(d) level of theory. Trans and cis isomers are defined with reference to the dihedral  $N'_3-C'_2-C_0-C_2 = \phi'$  (see Figure 1 for atom labeling). Both for H-TCY and Et-TCY the two moieties bonded to the central carbon atom  $C_0$  are identical. At



**Figure 3.** Schematic drawing of the different points on the  $S_1$  PES of H-TCY: (a) FC; (b)  $S_1$ -Min $C_2$ ; (c)  $S_1$ -Min; and (d)  $S_0$ -cis.

**Table 1.** Main Geometrical Parameters of the Equilibrium  $S_0$  Structures of Et-TCY and of the Two Minima on the  $S_1$  Surface<sup>e</sup>

|                 | $S_0$ -trans <sup>a</sup> | $S_1$ -Min $C_2$ <sup>b</sup> | $S_1$ -Min <sup>b</sup> | $S_0$ -cis <sup>a</sup> |
|-----------------|---------------------------|-------------------------------|-------------------------|-------------------------|
| $S_1-C_2$       | 1.742                     | 1.740                         | 1.719                   | 1.748                   |
| $C_2-N_3$       | 1.358                     | 1.369                         | 1.330                   | 1.359                   |
| $N_3-C_4$       | 1.395                     | 1.374                         | 1.394                   | 1.397                   |
| $C_4-C_9$       | 1.398                     | 1.398                         | 1.389                   | 1.399                   |
| $C_2-C_0$       | 1.397                     | 1.403                         | 1.414                   | 1.393                   |
| $S'_1-C'_2$     | 1.742                     | 1.740                         | 1.759                   | 1.740                   |
| $C'_2-N'_3$     | 1.358                     | 1.369                         | 1.390                   | 1.351                   |
| $N'_3-C'_4$     | 1.395                     | 1.374                         | 1.389                   | 1.401                   |
| $C'_4-C'_9$     | 1.398                     | 1.398                         | 1.390                   | 1.399                   |
| $C'_2-C_0$      | 1.397                     | 1.403                         | 1.429                   | 1.409                   |
| $S_1-C_2-C_0$   | 125.5                     | 122.9                         | 119.3                   | 124.5                   |
| $S'_1-C'_2-C_0$ | 125.5                     | 122.9                         | 117.9                   | 119.3                   |
| $C_2-C_0-C'_2$  | 127.6                     | 123.4                         | 120.5                   | 127.9                   |
| $\Omega C_2^c$  | 360.0                     | 360.0                         | 351.6                   | 360.0                   |
| $\Omega N_3^c$  | 360.0                     | 360.0                         | 359.2                   | 360.0                   |
| $\phi^d$        | 176.1                     | 153.0                         | 177.3                   | 179.5                   |
| $\phi'^d$       | 176.1                     | 153.0                         | 114.8                   | 41.7                    |

<sup>a</sup> Computed at PBE0/6-31G(d) level of theory. <sup>b</sup> Computed at CIS/6-31G(d) level of theory. <sup>c</sup> Sum of the bond angles with vertex at the indicated nucleus. <sup>d</sup>  $\phi = N_3-C_2-C_0-C'_2$ ,  $\phi' = N'_3-C'_2-C_0-C_2$ . <sup>e</sup> Distances in angstroms, angles in degrees.

variance with H-TCY, Et-TCY shows a small deviation from planarity ( $\phi$  and  $\phi'$  dihedral angles in Figure 1 are  $\approx 176^\circ$ ), allowing for the decrease in the steric hindrance between the ethyl substituents at  $N_3$  and  $N'_3$ . This result is in agreement with experiments<sup>14</sup> which reports an angle between the two rings smaller than  $15^\circ$  ( $\approx 8^\circ$  according to our

calculations). The other geometrical parameters are very similar in the trans isomers of the two molecules, confirming that the substitution of Et- with H- does not significantly alter their  $\pi$  electron system.

On the contrary, the equilibrium structures of H-TCY<sup>cis</sup> and Et-TCY<sup>cis</sup> are remarkably different, since Et-TCY<sup>cis</sup> exhibits a significant deviation from the planarity, with one ring about coplanar to the central  $C_0-H$  bond and the other rotated ( $\phi' \approx 42^\circ$ ) so as to decrease the steric repulsion between the ethyl substituents (see Figure 2). This effect is by far reduced when Et- is substituted by H-, and H-TCY<sup>cis</sup> can keep a pseudoplanar structure ( $\phi' = 5^\circ$ ), allowing a more effective  $\pi$ -electrons delocalization.

In Table 3 are collected the computed energies of the  $S_0 \rightarrow S_1$  electronic transition for the trans and cis isomer of Et-TCY and H-TCY.

TD-PBE0 computations predict a strong absorption band located at  $\approx 3.4$  eV for Et-TCY and  $\approx 3.2$  eV for Et-TCY<sup>cis</sup>. Those values are within 0.5 eV of their experimental counterpart (obtained in methanol solution), and the agreement is within 0.36 eV when solvent effect is taken into account by means of the PCM method (vide infra). This good result witnesses that TCY is a more favorable case for TD-DFT than the unsubstituted stCY (see Supporting Information).

When considering the relative behavior of cis and trans isomers of Et-TCY, TD-PBE0 results show a perfect agreement with experiments. In fact,  $S_0 \rightarrow S_1$  transition in Et-TCY is predicted more intense and blue-shifted (by 0.18 and 0.17 eV according to calculations and experiments, respectively) with respect to Et-TCY<sup>cis</sup>. Such a large difference is mainly due to the different degree of planarity of the two isomers, as testified by the much smaller difference found between the two isomers of H-TCY, both planar. These results show that TD-PBE0 calculations are able to reproduce the effect of deviation from the planarity on the absorption spectra of Et-TCY and H-TCY.

Apart from the differences mentioned above, the behavior of the excited states of H-TCY and Et-TCY is very similar. For both compounds, the  $S_0 \rightarrow S_1$  transition has a predominant HOMO  $\rightarrow$  LUMO character. Considering the central allyl moiety, the HOMO is mainly localized on the  $C_0$  atom (see Figure 4), as a "classical" bonding HOMO orbital of an allyl group (see Figure 5), whereas the LUMO has a nodal plane on  $C_0$ , and its density is moved on the two "external" atoms ( $C_2$  and  $C'_2$ ) of the allyl moiety. LUMO has also an antibonding character with respect to the  $C_2-S_1$  and  $C_2-N_3$  bonds, and benzenic rings remarkably contribute both to the HOMO and the LUMO. Considering the trans isomer, a natural bond order analysis (NBO) (at HF and CIS level of theory) confirms that both in  $S_0$  and  $S_1$  the electronic density is symmetric on the two side-moieties and most of the positive charge is carried by the 5-atoms rings adjacent to the central  $C_0$  atom (which in  $S_0$  is partially negatively charged) and in particular by their sulfur nuclei (whose charges are  $\approx +0.5$  au in both the states), revealing their important electron-donor function in thiocyanines. Upon the  $S_0 \rightarrow S_1$  transition there is a net transfer from the central  $C_0$

**Table 2.** Main Geometrical Parameters of the Equilibrium  $S_0$  Structures of H-TCY and Some Important Points on the  $S_1$  Surface<sup>f</sup>

|   | $S_0$ -trans <sup>a</sup> | $S_1$ -180 <sup>b</sup> | $S_1$ -MinC <sub>2</sub> <sup>b</sup> | $S_1$ -135 <sup>b</sup> | $S_1$ -Min <sup>b</sup> | CI-CAS <sup>c</sup> | $S_0$ -cis <sup>a</sup> |
|---|---------------------------|-------------------------|---------------------------------------|-------------------------|-------------------------|---------------------|-------------------------|
| S <sub>1</sub> -C <sub>2</sub>                    | 1.740                     | 1.744                   | 1.739                                 | 1.721                   | 1.714                   | 1.697               | 1.744                   |
| C <sub>2</sub> -N <sub>3</sub>                    | 1.352                     | 1.371                   | 1.365                                 | 1.334                   | 1.327                   | 1.311               | 1.352                   |
| N <sub>3</sub> -C <sub>4</sub>                    | 1.387                     | 1.359                   | 1.367                                 | 1.383                   | 1.385                   | 1.390               | 1.387                   |
| C <sub>4</sub> -C <sub>9</sub>                    | 1.398                     | 1.400                   | 1.398                                 | 1.390                   | 1.388                   | 1.386               | 1.398                   |
| C <sub>2</sub> -C <sub>0</sub>                    | 1.392                     | 1.395                   | 1.399                                 | 1.397                   | 1.411                   | 1.446               | 1.389                   |
| S <sub>1</sub> -C <sub>2</sub> '                  | 1.740                     | 1.744                   | 1.739                                 | 1.760                   | 1.758                   | 1.788               | 1.739                   |
| C <sub>2</sub> '-N <sub>3</sub> '                 | 1.352                     | 1.371                   | 1.365                                 | 1.383                   | 1.391                   | 1.426               | 1.349                   |
| N <sub>3</sub> '-C <sub>4</sub> '                 | 1.387                     | 1.359                   | 1.367                                 | 1.381                   | 1.389                   | 1.428               | 1.391                   |
| C <sub>4</sub> '-C <sub>9</sub> '                 | 1.398                     | 1.400                   | 1.398                                 | 1.389                   | 1.386                   | 1.370               | 1.398                   |
| C <sub>2</sub> '-C <sub>0</sub> '                 | 1.392                     | 1.395                   | 1.399                                 | 1.436                   | 1.425                   | 1.461               | 1.396                   |
| S <sub>1</sub> -C <sub>2</sub> -C <sub>0</sub>    | 127.3                     | 128.6                   | 125.6                                 | 122.5                   | 122.8                   | 123.2               | 127.2                   |
| S <sub>1</sub> -C <sub>2</sub> '-C <sub>0</sub> ' | 127.3                     | 128.6                   | 125.6                                 | 117.6                   | 119.9                   | 118.0               | 121.0                   |
| C <sub>2</sub> -C <sub>0</sub> -C <sub>2</sub> '  | 127.4                     | 130.6                   | 124.5                                 | 120.5                   | 121.4                   | 120.2               | 127.9                   |
| $\Omega$ C <sub>2</sub> <sup>d</sup>              | 360.0                     | 360.0                   | 360.0                                 | 348.0                   | 351.2                   | 346.6               | 360.0                   |
| $\Omega$ N <sub>3</sub> <sup>d</sup>              | 360.0                     | 360.0                   | 359.8                                 | 356.6                   | 354.1                   | 334.5               | 360.0                   |
| $\phi$ <sup>e</sup>                               | 180.0                     | 180.0                   | 155.4                                 | 171.0                   | 178.5                   | 180.5               | 176.7                   |
| $\phi'$ <sup>e</sup>                              | 180.0                     | 180.0                   | 155.4                                 | 135.0                   | 111.7                   | 109.2               | 5.3                     |

<sup>a</sup> Computed at PBE0/6-31G(d) level of theory. <sup>b</sup> Computed at CIS/6-31G(d) level of theory. <sup>c</sup> Computed at SA-2-CAS(6,6)/6-31G(d) level of theory. <sup>d</sup> Sum of the bond angles with vertex at the indicated nucleus. <sup>e</sup>  $\phi = \text{N}_3\text{-C}_2\text{-C}_0\text{-C}_2'$ ,  $\phi' = \text{N}_3'\text{-C}_2'\text{-C}_0\text{-C}_2'$ . <sup>f</sup> Distances in angstroms, angles in degrees.

**Table 3.** Vertical Excitation Energy (in eV) for the Cis and Trans Isomers of Et-TCY and H-TCY: TD-PBE0/6-31+G(d,p)//PBE0/6-31G(d) Calculations<sup>d</sup>

| Et-TCY     |                         |                  |                         |            |                  |
|------------|-------------------------|------------------|-------------------------|------------|------------------|
| trans      |                         |                  | cis                     |            |                  |
| gas phase  | solution                | exp <sup>c</sup> | gas phase               | solution   | exp <sup>c</sup> |
| 3.37(1.01) | 3.29(1.19)              | 2.93             | 3.18(0.81)              | 3.11(0.97) | 2.76             |
| H-TCY      |                         |                  |                         |            |                  |
| trans      |                         | cis              |                         |            |                  |
| gas phase  | solution <sup>a</sup>   | gas phase        | solution <sup>a</sup>   |            |                  |
| 3.42(1.08) | 3.35(1.29)              | 3.36(1.07)       | 3.29(1.27)              |            |                  |
|            | 3.34(1.29) <sup>b</sup> |                  | 3.28(1.28) <sup>b</sup> |            |                  |

<sup>a</sup> PCM calculations. <sup>b</sup> Geometry optimized in methanol solution by the PCM method. <sup>c</sup> From ref 14. <sup>d</sup> Oscillator strengths are given in parentheses.

to the side rings. In particular the density lost by C<sub>0</sub> moves to the adjacent C<sub>2</sub>/C<sub>2</sub>' centers.

Electron density of Et-TCY<sup>cis</sup> is instead partially asymmetric, due to the strong nonplanarity of its equilibrium geometry. In  $S_0$  the positive charge is slightly more localized (by  $\approx 0.1$  au) on the rotated ring (hereafter labeled as right), since the HOMO is slightly more localized on the planar moiety (left ring + C<sub>0</sub> atom). *Therefore in the case of Et-TCY<sup>cis</sup> the  $S_0 \rightarrow S_1$  transition has a weak character of charge transfer* (approximately 0.2 au from the planar to the rotated moiety). Nevertheless, this fact does not affect the reliability of our TD-PBE0 calculations, as shown by the nice comparison with the experimental vertical excitation energies (vide supra). The charge transfer character is remarkably smaller in H-TCY<sup>cis</sup>, whose structure is much less asymmetric.

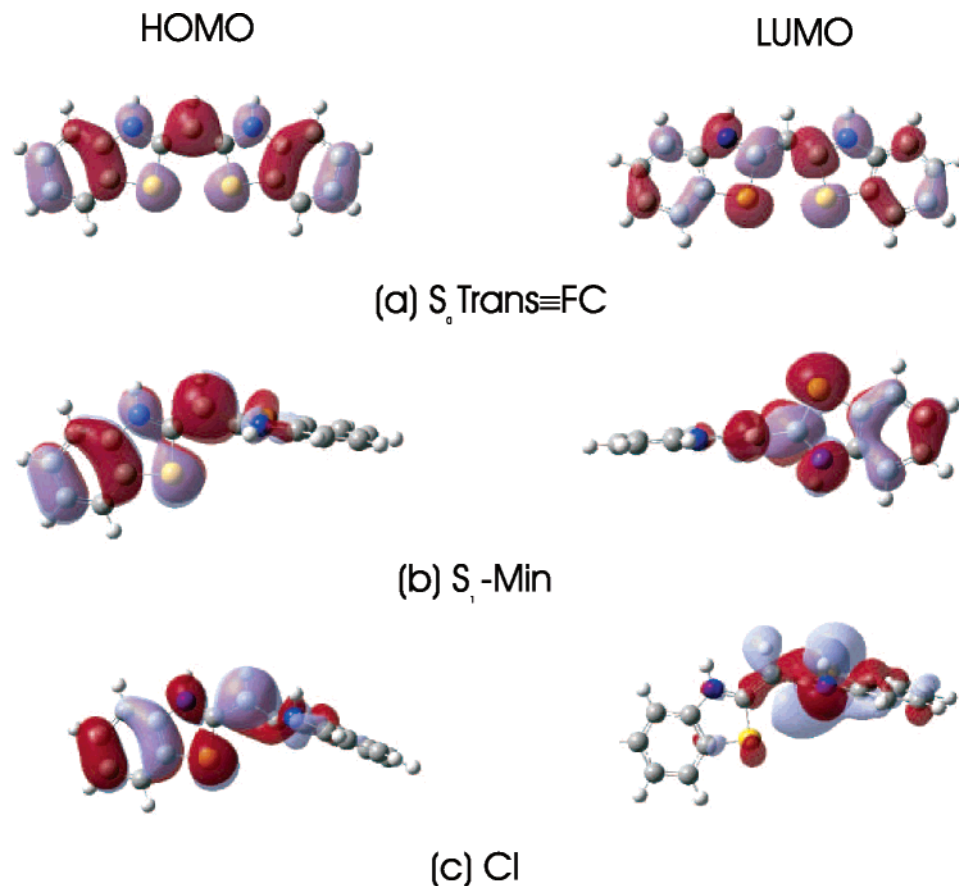
In both isomers, the more “delocalized” character of the LUMO can explain why  $S_1$  has a slightly larger dipole moment than  $S_0$  ( $\approx 4.1$  and 3.9 D, respectively, according to

CIS/6-31G(d) calculations in H-TCY) giving account of the red-shift predicted by PCM/TD-PBE0 calculations in methanol solution.

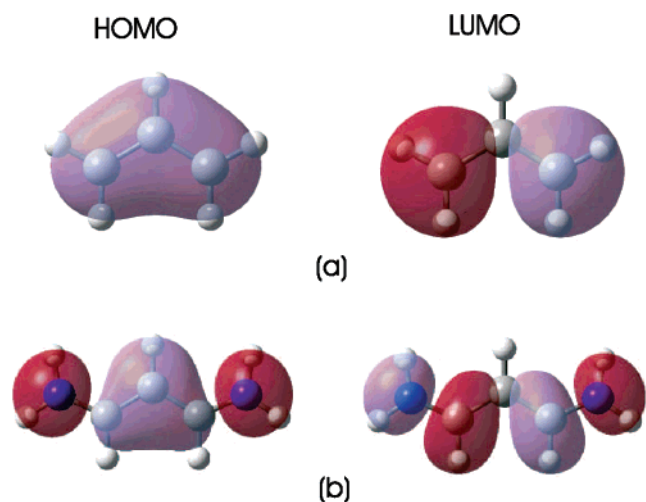
We have further examined the solvent effect by performing a test geometry optimization on H-TCY in methanol, finding that, due to the remarkable rigidity of the fused aromatic rings, the solvent has a negligible effect on the equilibrium geometries. In fact the vertical excitation energy changes indeed by only  $\approx -0.007$  eV when the geometry optimized in solution is used.

The possible effect on the absorption spectra of the formation of explicit interactions with methanol solvent molecules has been tested optimizing the coordination of two methanol molecules to H-TCY at the PBE0/6-31G(d) level. It is predicted the formation of hydrogen bonds between the NH groups (partially positively charged) and the oxygen atom of methanol, with a coordination energy to H-TCY  $\approx 7$  kcal/mol more exoergonic than a hydrogen bond between two solvent molecules (PCM/PBE0/6-31+G(d,p) calculations in methanol solution). However the two methanol molecules do not significantly affect the absorption spectra (red-shift of  $\approx 0.001$  eV). Furthermore, the ethyl substituents in Et-TCY are expected to weaken the coordination of methanol molecules, so that explicit solute–solvent interactions can be neglected when considering the optical properties of cyanine.

In summary, TD-PBE0 predictions on the vertical absorption maxima of cis and trans isomers of Et-TCY, and, especially, on their difference, are in good agreement with experiments, supporting the reliability of our computational approach for the study of cis–trans isomerization on  $S_1$  surface. Furthermore, H-TCY is a good model for Et-TCY, mostly when the characteristic and the dynamics of the trans isomer are taken into account. Solvent effect on the equilibrium structure is negligible so that in the following we always consider geometries optimized in the gas phase.



**Figure 4.** Schematic drawing of the HOMO and the LUMO of H-TCY, at **FC** (a),  **$S_1$ -Min** (b), and **CI** (c), computed at PBE0/6-31G(d) level. Notice that at TD-PBE0 level the description of the transition is only slightly improved by changing the basis set from 6-31G(d) to 6-31+G(d,p); this also means that the shape of the more involved orbitals, computed by the two different basis sets, is the same.



**Figure 5.** Schematic drawing of the HOMO and the LUMO of (a) an allyl group and (b) stCY, computed at PBE0/6-31G(d) level.

**3.2. Stationary Points on the  $S_1$  PES.** Minima on  $S_1$  have been computed both for H-TCY and Et-TCY, and their structures are very similar in the two molecules, as can be seen from Tables 1 and 2. As a consequence, the time-consuming optimization of the  $S_0/S_1$  conical intersection has been performed only for the smaller H-TCY. If not stated differently, in the following we discuss the results for H-TCY.

Starting from the Franck–Condon (**FC**) structure CIS/6-31G(d) geometry optimization predicts the existence of a shallow minimum for  $\phi = \phi' \approx 155^\circ$  (see Figure 3b), confirmed by a calculation of the Hessian matrix. The two rings keep an equivalent structure in this minimum (hereafter  **$S_1$ -Min $C_2$** ) exhibiting a  $C_2$  symmetry. The most relevant geometry changes with respect to the  $S_0$  *trans* minimum (hereafter  **$S_0$ -*trans*=FC**) concern the elongation of the  $C_2-N_3$  bonds and a shortening of the  $N_3-C_4$  ones. The above results can be explained by the inspection of Figure 4: the HOMO is nonbonding on  $C_2-N_3$  and antibonding on  $N_3-C_4$ , while the opposite occurs for the LUMO. The charge distribution of the  $S_0$  and  $S_1$  states of H-TCY at  **$S_1$ -Min $C_2$**  is very similar to the ones computed at  **$S_0$ -*trans***, confirming that no significant change occurs in these states till the symmetry is preserved.

CIS geometry optimizations predict also the existence of a strongly asymmetric absolute minimum ( **$S_1$ -Min**) where  $\phi' \approx 112^\circ$ , confirmed also through a calculation of the Hessian matrix. In  **$S_1$ -Min** the two rings are no more equivalent,  $\phi$  is  $\approx 180^\circ$  and the “left” ring of Figure 3c is coplanar with the central  $C_0-H$  bond. Furthermore,  $N_2-C_3$  and  $C_2-S_1$  are considerably shorter than in  **$S_0$ -*trans***, while the opposite occurs for  $N'_2-C'_3$  and  $C'_2-S'_1$ . These features can be rationalized by comparing the frontier orbitals of  **$S_0$ -*trans*** and  **$S_1$ -Min** reported in Figure 4. In  **$S_1$ -Min** the HOMO

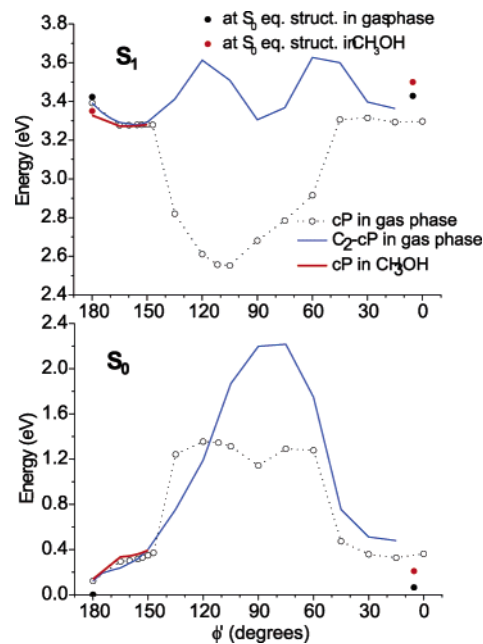


orbital is totally localized on the left ring and the central  $C_0$  atom, while the LUMO is localized on the right ring. At  $S_1$ -**Min**, according to TD-PBE0,  $S_0 \rightarrow S_1$  is mainly a HOMO  $\rightarrow$  LUMO transition and can therefore be described as an intramolecular charge transfer, confirming previous computational results on the small stCY.<sup>29</sup> Going from  $S_0$ -*trans* to  $S_1$ -**Min** on the  $S_1$  surface, the left ring loses the LUMO strong antibonding interactions on  $N_2-C_3$  and  $C_2-S_1$  thus inducing their shortening. The opposite geometry changes occur on the right ring, since here the  $N'_2-C'_3$  and  $C'_2-S'_1$  antibonding character of the LUMO is stronger at  $S_1$ -**Min** than at  $S_0$ -*trans*.

The TICT nature of the  $S_0 \rightarrow S_1$  is confirmed by a NBO population analysis on  $S_0$  and  $S_1$  (at HF and CIS level of theory, respectively), showing that in  $S_0$  the positive charge is almost totally localized on the right ring (+0.85 au) and in particular on the 5-atoms ring (0.64 au), while in  $S_1$  the right ring is almost neutral (+0.15 au). It is interesting to notice that within the planar moiety (left ring +  $C_0-H$ ), the most evident density change takes place on the central  $C_0-H$  (strongly contributing to the HOMO), whose charge increases by  $\approx +0.5$  au following  $S_0 \rightarrow S_1$  transition.

**3.3.  $S_1/S_0$  Conical Intersection.** The main geometrical parameters of the SA-2-CAS(6,6)  $S_1/S_0$  conical intersection (**CI**) are reported in Table 2, and its structure is depicted in Figure 3d. A calculation with a far larger active space SA-2-CAS(12,12) confirms that at this geometry the two states are close to degeneracy (gap 0.1 eV). The reliability of the **CI** structure here determined is further supported by its close analogy with that accurately located in ref 29 in stCY. In fact the **CI** structure differs from the  $S_1$ -**Min** for a asymmetric-skeletal motion (similar to the one driving  $S_1$ -**Min** $C_2$  to  $S_1$ -**Min**) and a marked pyramidalization of  $N'_3$  which pushes the H atom out-of-the average left-ring plane. Both these motions play an important role in reaching the **CI** from  $S_1$ -**Min**. The energy of  $S_0$  increases more steeply along the pyramidalization while the energy of  $S_1$  increases more steeply along the skeletal-stretching. At **CI** the TD-PBE0/6-31+G(d,p) calculation also predicts a very small  $S_0 - S_1$  energy gap (0.27 eV), confirming that the  $S_1/S_0$  conical intersection is reached from  $S_1$ -**Min** mainly distorting its structure along the asymmetric stretching and the  $N'_3$  pyramidalization. This conclusion is supported by a TD/PBE0 scan of the  $S_0$  and  $S_1$  surface as a function of the  $H'-N'_3-C'_2-S'_1$  torsion and of an asymmetric collective coordinate interpolating the  $S_1$ -**Min** and **CI** structure. There is a region around **CI** where  $S_0$  and  $S_1$  are nearly degenerate. The best TD-PBE0 estimate of the conical intersection ( $S_1 - S_0$  energy gap = 0.04 eV and  $E(S_1) = 3.08$  eV) slightly differentiates with respect to **CI** for a less pronounced pyramidalization at  $N'_3$  and a more pronounced asymmetry of the two rings.

Pyramidalization at nitrogen or carbon centers are frequently involved in the occurrence of a  $S_1/S_0$  conical intersection, see for example the case of ethylene,<sup>43</sup> stilbene,<sup>7</sup> and stCY.<sup>29</sup> For H-TCY **CI** is reached from  $S_1$ -**Min** by destabilizing strongly the  $S_0$  state ( $\approx 1.5$  eV) and weakly the  $S_1$  one ( $\approx 0.5$  eV). Both at  $S_1$ -**Min** and at **CI** the  $S_0 \rightarrow S_1$  transition is characterized by a charge transfer from the



**Figure 6.** Energy of the lowest energy states as a function of the  $\phi'$  dihedral for H-TCY. TD-PBE0/6-31G(d)//CIS/6-31G(d) calculations.

HOMO, fully localized on the left ring and the central  $C_0-H$  bond, to the LUMO on the right ring (see Figure 4). Further analysis shows that the pyramidalization strongly stabilizes the LUMO (by  $\approx 0.9$  eV), since as it appears in Figure 4, at **CI** the  $N'_3$  lone pair can take part in an extended  $\sigma$ -bonding orbital. This effect partially compensates the  $S_1$  energy destabilization due to geometry distortion from  $S_1$ -**Min** to **CI** and, since it is only operative on  $S_1$  (the LUMO is empty in  $S_0$ ), it explains why the two electronic surfaces get close in energy, making feasible the occurrence of a conical intersection.

## 4. Reactive Paths on the $S_1$ Surface

In this section we study the possible paths on  $S_1$  driving the system from the **FC** point ( $\equiv S_0$ -*trans*) to the conical intersection.

**4.1. From the FC Structure to the Perpendicular Minimum. One-Dimensional Path Analysis.** To characterize the steps of the preferential path followed by the wave packet (WP) from **FC** to  $S_1$ -**Min** we performed a relaxed PBE0/6-31+G(d,p)//CIS/6-31G(d) scan of the  $S_1$  surface as a function of the  $\phi'$  dihedral which is varied between  $180^\circ$  and  $0^\circ$  (see Figure 6).

The most significant structures of this constrained minima path (cP) are reported in Table 2. In the very initial dynamics at planar configurations (represented by  $S_1$ -180 structure) the system enlarges the central  $C_2-C_0-C'_2$  bond-angle (hereafter  $\alpha$ ) to decrease the  $C_2-C'_2$  antibonding character of the LUMO. Rotating  $\phi'$  the local minimum  $S_1$ -**Min** $C_2$  is encountered. At the TD-PBE0 level it is closer to planarity ( $\phi = \phi' \approx 165^\circ$ ) than it is at the CIS level, with an energy about 0.13 eV lower than the **FC** structure. The PBE0/6-31+G(d,p) computed fluorescence emission peak is  $\approx 2.96$  eV in the gas phase and  $\approx 2.91$  eV in methanol solution, i.e., with a Stokes shift of 0.46 and 0.44 eV, respectively

(for Et-TCY the same Stokes shifts are 0.48 and 0.46 eV, respectively).

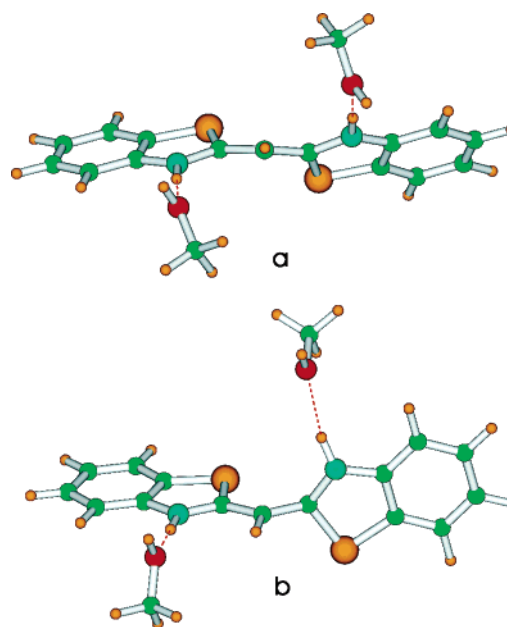
Up to  $\phi' = 150^\circ$  cP predicts that the molecule keeps approximately a  $C_2$  arrangement with  $\phi$  and  $\phi'$  moving in a conrotatory way. In fact cP is practically isoenergetic to a path  $C_2$ -cP (see Figure 6) in which the  $C_2$  symmetry has been imposed, and therefore no significant driving force exists to lower the symmetry. On the contrary, upon further rotation of the  $\phi'$  dihedral, a sudden and drastic geometry asymmetrization occurs accompanied by a corresponding remarkable stabilization of the  $S_1$  state. That process is almost accomplished at  $\phi' = 135^\circ$ , whose geometry is strongly asymmetric ( $S_1$ -135 in Table 2), and whose energy is about 0.5 eV lower than at  $\phi' = 150^\circ$ . Our computations thus predict the existence of an energy maximum on cP curve for  $\phi' \approx 150^\circ$ . The estimated energy barrier in H-TCY is very small, only  $\approx 30 \text{ cm}^{-1}$  at TD-PBE0/6-31+G(d,p) level in gas phase. However, it is significantly higher  $\approx 120 \text{ cm}^{-1}$  in Et-TCY, probably due to the larger steric hindrances of the two Et-groups and could influence the dynamics on the  $S_1$  surface, increasing the  $S_1$ - $\text{Min}C_2$  lifetime.

After  $S_1$ -135, significant but minor skeletal rearrangements drive the system to the  $S_1$ - $\text{Min}$  structure, where TD-PBE0/6-31+G(d,p) predicts a  $S_1 - S_0$  energy gap of 1.21 eV for H-TCY and 1.40 for Et-TCY. These values are probably underestimated for the known deficiencies of TD-DFT in reproducing charge-transfer transition-energies.<sup>41,42</sup> A SA-2-CAS(6,6)/6-31G(d) computation for H-TCY predicts a value of 1.31 eV (to be compared with 1.18 eV obtained by TD-PBE0 with the same basis set), and this value increases to 1.38 eV if a much larger active space is used (SA-2-CAS(12,12)). This value is still probably underestimated by about 0.8 eV as one can roughly estimate from the results in ref 29 (see also Supporting Information) showing that for stCY the  $S_1 - S_0$  CAS energy gap at  $S_1$ - $\text{Min}$  is  $\approx 65\%$  of the CASMP2 value.

The motion along the cP from  $\phi' = 150^\circ$  to  $S_1$ - $\text{Min}$  involves a different (asymmetric) skeletal motion with respect to the symmetric one along which the system moves from  $S_1$ -180 to  $S_1$ - $\text{Min}C_2$ ; in the same way the conrotatory motion of the two dihedral angles  $\phi$  and  $\phi'$  becomes disrotatory, and the left ring reassumes a planar arrangement with the  $C_0$ -H ( $\phi \rightarrow 180^\circ$ ). The necessity to acquire the right momenta along these modes can require some time, slowing down the reaction beyond what is done by the barrier along the cP.

Figure 6 shows that further rotating  $\phi'$  beyond the  $S_1$ - $\text{Min}C_2$  structure the  $C_2$ -cP shows two other minima: the first at  $\phi' \approx 90^\circ$ , the other corresponding to a double-cis isomer. On  $S_0$  this latter is slightly nonplanar in H-TCY ( $\phi = \phi' = 14^\circ$ ), and it is 0.23 eV less stable than the trans isomer; in Et-TCY it is still less planar  $\phi = \phi' = 28^\circ$  and less stable (0.46 eV) with respect to the trans isomer, due to the Et-groups effect. Anyway, Figure 6 shows that the high barrier makes the double-cis region inaccessible from FC, so that it can be ignored for the scopes of the present work.

**Solvent Effect.** Inspection of Figure 6 shows that including the solvent effect by means of the PCM method does not remarkably affect the first ( $C_2$ ) part of cP but for an almost



**Figure 7.** Schematic drawing of the adduct on the  $S_1$  surface between H-TCY and two methanol molecules: (a)  $S_1$ - $\text{Min}C_2$  and (b)  $S_1$ - $\text{Min}$ .

uniform stabilization of the  $S_1$  state. It is not possible at the moment to give a reliable quantitative estimate of solvent effect on the TICT process, i.e., when the two moieties become strongly asymmetric. From the physical point of view, the nonequilibrium solvation effect should disfavor sudden electronic density change, since there is a slower component of solvent polarization that cannot be equilibrated to the solute. On the other hand, a polar solvent could favor charge localization. However, CIS/6-31G(d) calculations show that the dipole moment of the  $S_1$  state does not dramatically increase when going from  $S_1$ - $\text{Min}C_2$  to  $S_1$ - $\text{Min}$  (see Figure S2 in the Supporting Information), even if the latter exhibits a much larger charge asymmetry, as testified by a dipole moment oriented along the *long* molecular axis.

As a matter of fact, also pseudoplanar arrangements of the cyanine rings, though characterized by a symmetric charge distribution, exhibit a large dipole moment oriented along the *short* molecular axis, due to the presence of two parallel strong  $S_1^+ - N_3^-$  local dipoles. This effect is lost for pseudoperpendicular geometry, as in  $S_1$ - $\text{Min}$ . Looking for a qualitative estimate of static solvent effect on the isomerization, we have calculated electrostatic solute-solvent interactions in the framework of PCM model, by using the excited-state atomic charges computed in methanol solution at the PCM/CIS/6-31G(d) level, according to the Merz-Kolmann<sup>44</sup> population analysis. The results of our model calculations (see Figure S2 in the Supporting Information) show that the presence of a polar solvent should favor the TICT  $S_1$ - $\text{Min}$  state over  $S_1$ - $\text{Min}C_2$  by only  $\approx 0.15$  eV.

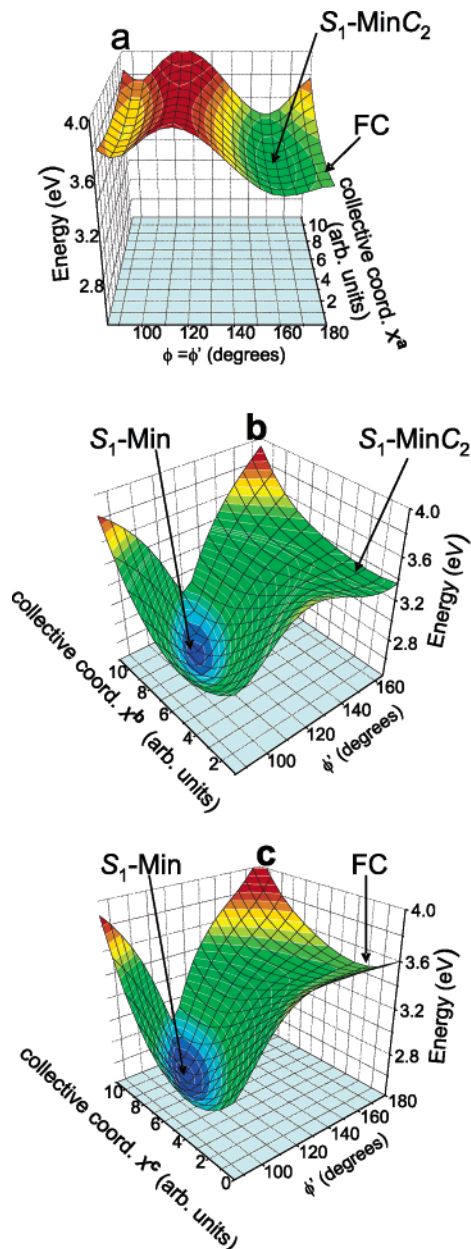
To get insights on the role of explicit solute-solvent interactions, we have then optimized at the CIS/6-31G(d) level the coordination of two methanol molecules to  $S_1$ - $\text{Min}C_2$  and  $S_1$ - $\text{Min}$  (see Figure 7). In the symmetric  $S_1$ - $\text{Min}C_2$  both methanol molecules have a hydrogen bond distance with the NH group of 1.90 Å, while in  $S_1$ - $\text{Min}$  the

hydrogen bond coordination is strongly asymmetric, with the  $\text{CH}_3\text{HO}\cdots\text{HN}_3$  and  $\text{CH}_3\text{HO}\cdots\text{HN}'_3$  hydrogen bond distances of 1.84 and 2.01 Å, respectively. Due to the TICT process the  $S_1$  positive charge is localized mainly on the “left” ring (vide supra), leading to a stronger hydrogen bond. On the other hand, when the hydrogen bond distance optimized for  $S_1\text{-Min}C_2$  is imposed, the energy of the  $2\text{CH}_3\text{OH}/S_1\text{-Min}$  rises by  $\approx 0.01$  eV only. Those results, together with those of the electrostatic model calculations and with the fact that ethyl substituent should make solute/solvent interactions less strong, suggest that static solvent effects do not significantly affect the isomerization process.

Unfortunately, our present treatment is not able to give a quantitative estimate of the influence of nonequilibrium effects on the dynamical motion on cP and on the height of the energy barrier between  $S_1\text{-Min}C_2$  and  $S_1\text{-Min}$ . An “ad hoc” study of the different relaxation time of solvent and solute degrees of freedom would indeed be necessary. On the other hand, our finding that total dipole moment changes its orientation during the motion on the  $S_1$  surface suggests that a solvent dynamical effect could play a non-negligible role in the isomerization process.

**Two-Dimensional Analysis.** In the region  $180^\circ < \phi' < 150^\circ$  the cP curve is very flat, suggesting the existence of a plateau on the  $S_1$  surface. In such a case, it is likely that the wave packet (WP) broadens and follows also other decay paths than the cP. To verify this hypothesis and gain a better picture of the dynamics, we computed some bidimensional (2D) maps of the  $S_1$  surface at TD-PBE0/6-31G(d) level. To build up each map we consider two structures chemically important for the reaction, say  $P_1$  and  $P_2$ , and explore the part of the  $S_1$  surface which connects them. Restricting the 3N-6 dimensional space to a 2D one is clearly arbitrary; the nature of the process under consideration suggests that an important coordinate is the torsion around  $\phi'$  dihedral; the second “collective” coordinate  $x$  is defined imposing that all the other internal coordinates move in a synchronous way. Say  $b$  any of these internal coordinates,  $b_1$  and  $b_2$  its values at the reference structures  $P_1$  and  $P_2$ , where the collective coordinate  $x$  is fixed to values  $x_1$  and  $x_2$ , respectively, then at each value  $\bar{x}$  of  $x$  the corresponding value  $\bar{b}$  of  $b$  is computed by linear interpolation  $\bar{b} = b_1 + (\bar{x} - x_1)(b_2 - b_1)/(x_2 - x_1)$ .

The first scan, reported in Figure 8a, describes a  $C_2$  motion connecting **FC** and  $S_1\text{-Min}C_2$  (i.e., the  $\phi$  and  $\phi'$  dihedral angles are kept equal, and, moving the collective coordinate  $x^a$ , the  $C_2$  symmetry is preserved). As expected by the cP calculation, there is not any significant barrier to the torsion. Moreover in the 2D map the conrotatory rotation of  $\phi$  and  $\phi'$  appears to be coupled with the symmetric skeletal rearrangement from the beginning of the motion, a feature that could not be appreciated by the simple inspection of cP. Figure 8a gives further support to the existence of a local minimum at  $\phi' \approx 160^\circ$  also at the TD-PBE0 level ( $165^\circ$  is the more accurate estimate in the previous section). For a larger deviation from planarity, the potential along this symmetric motion increases till a barrier at  $\phi = \phi' \approx 120^\circ$ , already seen in the  $C_2\text{-cP}$  path (Figure 6) too high to be crossed by WP.



**Figure 8.** 2D maps of the  $S_1$  surface as a function of the reactive  $\phi'$  torsion and a collective coordinate  $x^p$  ( $p = \mathbf{a,b,c}$ ) obtained interpolating linearly (see the text) all the other coordinates between **FC** and  $S_1\text{-Min}C_2$  (a),  $S_1\text{-Min}C_2$  and  $S_1\text{-Min}$  (b), **FC** and  $S_1\text{-Min}$  (c). Notice that (a) is a  $C_2$  map and hence  $\phi' = \phi$ .

Figure 8b reports a 2D surface connecting  $S_1\text{-Min}C_2$  and the  $S_1\text{-Min}$  structures; the first coordinate is the reactive  $\phi'$  torsion, while the second collective coordinate  $x^b$  linearly interpolates the remaining internal coordinates, as explained above. As predicted by cP, the two structures are divided by a small barrier; the map additionally shows that the  $S_1$  surface in the region of  $S_1\text{-Min}C_2$  and of the barrier is considerably flat, bringing new information and suggesting that in this region WP spreads sensibly. Momenta orientation and intramolecular vibration redistribution (IVR) certainly plays a fundamental role at this stage on the time needed to cross the barrier, and a full dynamical study would be required.

Finally, to inquire the possibility that WP does not initially follow the cP path toward  $S_1$ -**MinC**<sub>2</sub> moving instead in a direction that removes the  $C_2$  symmetry from the beginning, we performed another 2D scan of the  $S_1$  surface, choosing the  $\phi'$  coordinate and the collective  $x^c$  obtained interpolating linearly all the other coordinates (including  $\phi$ ) between **FC** and  $S_1$ -**Min** (Figure 8c). There is a large initial plateau till about  $\phi' = 150^\circ$ , without any barrier to break the symmetry of the two rings. The 2D map therefore evidences the existence of an alternative path to cP for the isomerization which does not involve  $S_1$ -**MinC**<sub>2</sub>. A quantitative estimate of its relative efficiency with respect to cP needs a full dynamical investigation. To get a very rough estimate we computed an average acceleration from **FC** to two structures at  $\phi' = 165^\circ$ , the first corresponding to the  $C_2$  minimum at the TD-PBE0 level ( $S_1$ -**MinC**<sub>2</sub>[TD]), the second ( $S_1$ -165as) corresponding to the minimum energy structure along the 2D map in Figure 8c. The time needed to reach  $S_1$ -165as is predicted to be comparable to the one needed to reach  $S_1$ -**MinC**<sub>2</sub>[TD]. Even if the energy gain at  $S_1$ -165as is much lower, it is indeed geometrically closer to **FC** than  $S_1$ -**MinC**<sub>2</sub>[TD] is, since this latter is more shifted along other coordinates.

## 5. Fluorescent State and Photoisomerization Dynamics

**Fluorescence.** The Et-TCY experimental fluorescence is peaked at  $\approx 480$  nm (2.58 eV),<sup>14,16</sup> i.e., with a Stokes shift of  $\approx 0.35$  eV with respect to the absorption maximum. Our results strongly suggest that this fluorescence arises from the  $S_1$ -**MinC**<sub>2</sub> intermediate. In fact, on one hand, its computed Stokes shift, 0.48 eV, is in good agreement with the experiments and, on the other hand, the alternative candidate  $S_1$ -**Min** has to be discarded since it is predicted to emit at far shorter frequencies (1.4 eV).

The lifetime of the Et-TCY fluorescence is expected to be on the picosecond time scale (an old experimental estimate<sup>15</sup> reports a value of 1.7 ps). This feature supports our assignment of  $S_1$ -**MinC**<sub>2</sub> as the fluorescent state. In fact a relatively long lifetime of  $S_1$ -**MinC**<sub>2</sub> can be explained on the grounds of the following arguments: (a)  $S_1$  is very flat in the **FC** region, and a small but not negligible barrier (about  $120\text{ cm}^{-1}$ ) separates the  $S_1$ -**MinC**<sub>2</sub> intermediate from  $S_1$ -**Min**; (b) once reached  $S_1$ -**MinC**<sub>2</sub> it is necessary to pump energies in asymmetric modes (disrotatory motion  $\phi$  and  $\phi' +$  asymmetric skeletal mode of the two rings) which are different from the initial  $C_2$  motions driving **FC** to  $S_1$ -**MinC**<sub>2</sub>; (c) the mass of the system and the presence of the solvent both slow the dynamics.

On the contrary, structures close to  $S_1$ -**Min** could be responsible for the low emission experimentally found when Et-TCY is adsorbed on aerosol-OT (AOT) surfactant ( $\approx 2.06$  eV).<sup>13</sup> This peak has been interpreted as coming from AOT<sub>6</sub>/Et-TCY<sub>6</sub> aggregates, in which the Et-TCY molecules exhibit a twisted geometry, and our results support this interpretation. In fact, as we discussed in a previous section, even if TD-DFT results predict a much lower frequency for the emission from  $S_1$ -**Min** ( $\approx 1.4$  eV in gas phase), CAS results on H-TCY and comparison with CAS/CASMP2 results for st-CY (see

Supporting Information) suggest that this value could be underestimated by about 0.8 eV. As expected for an ICT transition, the  $S_1 - S_0$  oscillator strength at  $S_1$ -**Min** predicted by TD-PBE0 is very low ( $\approx 0.01$ ). However, it must be noticed that also the experimental ICT band<sup>13</sup> is very weak and that its intensity is increased by aggregation phenomena on the surface, which we cannot take into account by our calculations.

**Photoisomerization Dynamics.** Our results suggest that two steps can be individuated in the  $S_1$  dynamics: (i) the motion from **FC** to  $S_1$ -**Min**, during which no  $S_1/S_0$  non-radiative transition is possible due to their large energy gap (see Figure 6) and (ii) the motion from  $S_1$ -**Min** to **CI**. As discussed above, it is likely that the first step occurs along different paths, i.e., removing the  $C_2$  symmetry from the beginning or passing through the  $S_1$ -**MinC**<sub>2</sub> intermediate.

Up-conversion fluorescence of a cyanine similar to Et-TCY (where the Et-groups are substituted by sulfopropyl-groups), recorded at frequencies close to the maximum of the fluorescence steady spectrum, shows a biexponential decay, with a fast component on the subpicosecond scale and a slow component of  $\approx 5$  ps.<sup>45</sup> The former component is assigned to a motion on the excited surface out of the **FC** region, while the latter is assigned to the  $S_1 \rightarrow S_0$  transition. Analogous assignments have been proposed for the similar time dependence of fluorescence, excited-state absorption, and ground-state recovery in 1144-C cyanine.<sup>46,47</sup> For this molecule it has been speculated<sup>47</sup> that the two-time decays is a sign of a multidimensional dynamics on  $S_1$  and hence of the involvement of different dynamical paths.

Our results on Et-TCY are consistent with this picture and suggest that the two different time scales could be due to the parts of the wave packet which go directly from **FC** to  $S_1$ -**Min**, escaping quickly from the **FC** region, or pass by  $S_1$ -**MinC**<sub>2</sub>, emitting for a longer time. A full dynamical study is planned to confirm this hypothesis.

For the case of 1144-C cyanine, different studies ascribe the rate-limiting step of the photoisomerization either to the initial bond twisting<sup>48</sup> (i.e. our step (i)) or to the nonadiabatic  $S_1 \rightarrow S_0$  transition<sup>47</sup> (i.e., the  $S_1$  motion from the perpendicular minimum to the conical intersection, our step (ii)). Our results for Et-TCY indicate that both the two steps are likely to be important for the overall reaction rate. Step (i) is expected to be considerably slowed, beyond what is done by the existence of the barrier on  $S_1$ , also by dynamical factors as an inertia to the torsion, the solvent friction (which clearly influences the dynamics of 1144-C),<sup>46,47</sup> and the necessity to pump energy on the right coordinates, at least for the path passing through  $S_1$ -**MinC**<sub>2</sub>. In fact, during this minimum energy motion on  $S_1$ , it is required to transfer energies from symmetric to asymmetric modes, and this process likely requires a nonvanishing time to be accomplished.

On the other hand, also step (ii) is expected to require a substantial amount of time because of the predicted energy gap on  $S_1$  between **CI** and  $S_1$ -**Min** which is ( $\approx 0.5$  eV), even if this value is probably overestimated in our calculations since no TD-PBE0 optimization of the **CI** structure is possible. From the dynamical point of view, it is interesting to notice that **CI** is shifted with respect to  $S_1$ -**Min** along an

**Table 4.** NBO Population Analysis of the CIS/6-31G(d) Density Matrix of the Lowest Energy States for a Series of Different Carbocyanine (See Figure 1)

|                       | atom-atom bond order           |                                |                                |                                | Wiberg bond order              |                                |                                |                                | extra diagonal Mulliken population |                                 |
|-----------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|------------------------------------|---------------------------------|
|                       | C <sub>0</sub> -C <sub>2</sub> | N <sub>3</sub> -C <sub>2</sub> | N <sub>3</sub> -C <sub>4</sub> | C <sub>2</sub> -X <sub>1</sub> | C <sub>0</sub> -C <sub>2</sub> | N <sub>3</sub> -C <sub>2</sub> | N <sub>3</sub> -C <sub>4</sub> | C <sub>2</sub> -X <sub>1</sub> | C <sub>0</sub> -C <sub>2</sub>     | C <sub>2</sub> -C' <sub>2</sub> |
|                       | <i>S</i> <sub>0</sub>          |                                |                                |                                |                                |                                |                                |                                |                                    |                                 |
| allyl                 | 1.167                          |                                |                                |                                | 1.508                          |                                |                                |                                | 0.466                              | -0.036                          |
| stCY                  | 1.135                          | 1.089                          |                                |                                | 1.397                          | 1.391                          |                                |                                | 0.534                              | -0.030                          |
| CH <sub>3</sub> -stCY | 1.133                          | 1.081                          |                                |                                | 1.366                          | 1.351                          |                                |                                | 0.540                              | -0.021                          |
| H-TCY                 | 1.131                          | 1.009                          | 0.884                          | 0.890                          | 1.353                          | 1.237                          | 1.039                          | 1.173                          | 0.514                              | -0.016                          |
| H-TCY-sm              | 1.130                          | 1.003                          | 0.897                          | 0.906                          | 1.348                          | 1.226                          | 1.073                          | 1.195                          | 0.517                              | -0.020                          |
| OCY                   | 1.125                          | 1.028                          | 0.862                          | 0.889                          | 1.333                          | 1.237                          | 1.023                          | 1.065                          | 0.543                              | -0.035                          |
| CH <sub>2</sub> CY    | 1.144                          | 1.043                          | 0.843                          | 0.865                          | 1.377                          | 1.302                          | 0.994                          | 1.015                          | 0.545                              | -0.018                          |
| 2-Quino               | 1.109                          | 0.983                          |                                | 1.024                          | 1.326                          | 1.199                          |                                | 1.185                          | 0.523                              | -0.028                          |
|                       | <i>S</i> <sub>1</sub>          |                                |                                |                                |                                |                                |                                |                                |                                    |                                 |
| allyl                 | 0.992                          |                                |                                |                                | 1.132                          |                                |                                |                                | 0.319                              | -0.104                          |
| stCY                  | 1.039                          | 1.028                          |                                |                                | 1.191                          | 1.283                          |                                |                                | 0.427                              | -0.092                          |
| CH <sub>3</sub> -stCY | 1.036                          | 1.024                          |                                |                                | 1.163                          | 1.252                          |                                |                                | 0.415                              | -0.078                          |
| H-TCY                 | 1.107                          | 0.951                          | 0.924                          | 0.849                          | 1.296                          | 1.125                          | 1.108                          | 1.092                          | 0.482                              | -0.051                          |
| H-TCY-sm              | 1.103                          | 0.943                          | 0.931                          | 0.851                          | 1.286                          | 1.116                          | 1.142                          | 1.089                          | 0.476                              | -0.057                          |
| OCY                   | 1.100                          | 0.971                          | 0.904                          | 0.846                          | 1.279                          | 1.131                          | 1.096                          | 0.990                          | 0.514                              | -0.067                          |
| CH <sub>2</sub> CY    | 1.115                          | 0.981                          | 0.896                          | 0.859                          | 1.311                          | 1.172                          | 1.086                          | 1.013                          | 0.516                              | -0.053                          |
| 2-Quino               | 1.082                          | 0.930                          |                                | 1.046                          | 1.267                          | 1.102                          |                                | 1.234                          | 0.488                              | -0.046                          |

<sup>a</sup> Calculations relative to vertical excitation on the C<sub>2</sub> S<sub>0</sub> energy minimum optimized at the PBE0/6-31G(d) level.

asymmetric skeletal mode and a pyramidalization motion at N<sub>3</sub>'. While it is clear that some time is necessary to populate the pyramidalization motion at N<sub>3</sub>', the asymmetric motion driving S<sub>1</sub>-Min to CI is similar to the one driving to S<sub>1</sub>-Min, so that the wave packet arriving to S<sub>1</sub>-Min should already have significant momentum along this coordinate.

## 6. Comparison with Different Cyanines

**Effect of the Fused Rings: TCY vs stCY.** The qualitative topology of the S<sub>1</sub> surface from FC to S<sub>1</sub>-Min discussed above for x-TCY (x = Et- or H-) is similar to the one reported in the literature for stCY.<sup>29</sup> Nevertheless there are significant quantitative differences. In fact in stCY the initial relaxation path occurs along a steep path dominated by C<sub>0</sub>-C<sub>2</sub> and C<sub>0</sub>-C'<sub>2</sub> stretchings strongly coupled with torsional deformation around the above bonds, and the S<sub>1</sub>-MinC<sub>2</sub> stationary structure (which is not a true minimum) already shows a pseudoperpendicular arrangement of the two side-chains. On the contrary for x-TCY, S<sub>1</sub>-MinC<sub>2</sub> is only slightly nonplanar and is a true minimum separated by an energy barrier from S<sub>1</sub>-Min.

Since the S<sub>0</sub> → S<sub>1</sub> transition has a predominant HOMO → LUMO character for both H-TCY and stCY and the shape of the frontier orbitals is qualitatively similar, how can we explain the significantly different shapes of the S<sub>1</sub> PES of H-TCY and stCY, with the consequent different dynamical behavior of the two molecules?

Analysis of the molecular orbitals can help rationalize these differences. Figure 4 shows that, due to the presence of the rings and of the sulfur substituents, the weight of the C<sub>2</sub>/C'<sub>2</sub> atomic orbitals in the LUMO and the consequent C<sub>2</sub>/C'<sub>2</sub> antibonding character is rather small, as testified by the small value of the extradiagonal negative Mulliken overlap population existing between those atoms (see Table 4). The

situation is different for stCY, where the contribution of the C<sub>2</sub>/C'<sub>2</sub> atoms to the LUMO and their antibonding character are larger (see Figure 5), and indeed, the extradiagonal negative C<sub>2</sub>/C'<sub>2</sub> S<sub>1</sub> Mulliken overlap population is almost double than in H-TCY. This results in a larger driving force to rotate the φ dihedral on S<sub>1</sub> for stCY, since a nonplanar arrangement of the two side-moieties decreases the C<sub>2</sub> - C'<sub>2</sub> antibonding character of the LUMO.

Significant differences are found in the HOMO as well. In stCY, C<sub>2</sub> and C'<sub>2</sub> noticeably contribute to the HOMO, that therefore has a substantial C<sub>0</sub>/C<sub>2</sub> weight and C<sub>0</sub>-C<sub>2</sub>/C'<sub>2</sub> bonding character. As a consequence the S<sub>0</sub> → S<sub>1</sub> transition should be associated with a remarkable reduction of the C<sub>0</sub>-C<sub>2</sub>/C'<sub>2</sub> bond orders. In H-TCY and Et-TCY the contribution of C<sub>2</sub> and C'<sub>2</sub> to the LUMO is negligible and, thus, both HOMO and LUMO can be considered nonbonding orbitals with respect to the C<sub>0</sub>-C<sub>2</sub>/C'<sub>2</sub> bonds. On this ground it is not surprising that the increase of the lengths of these bonds associated with the S<sub>0</sub> → S<sub>1</sub> transition is very small (see Tables 1 and 2).

To put all the above considerations on a more quantitative basis, we resorted to natural bond order (NBO)<sup>39</sup> and Wiberg bond order (WBO) analysis.<sup>38</sup> The inspection of Table 4 shows that the S<sub>0</sub> → S<sub>1</sub> transition leads the C<sub>0</sub>-C<sub>2</sub> bond order to decrease by ≈15% for stCY and only by ≈4% for H-TCY. Similar indications are provided by the overlap-weighted bond order based on the NBO analysis.

Analyzing the variation in the bond orders and in the equilibrium geometry suggests that, following S<sub>0</sub> → S<sub>1</sub> transition, the initial dynamics should involve the C<sub>2</sub>-C<sub>0</sub>-C'<sub>2</sub> bending and the C<sub>2</sub>-S<sub>1</sub>, C<sub>2</sub>-N<sub>3</sub>, and N<sub>3</sub>-C<sub>4</sub> stretching motions (together with the symmetric ones on the right ring). The excitation of symmetric normal modes comprising these motions is probably responsible for the shoulder observed

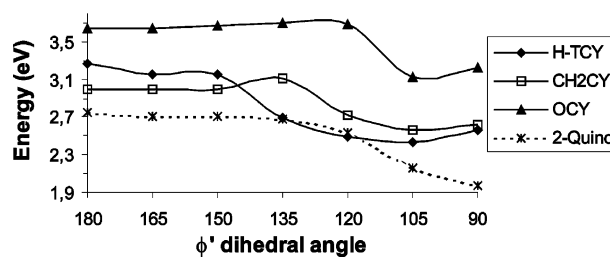
on the absorption spectrum at  $\approx 400$  nm<sup>14</sup> (i.e.  $\approx 1300$ – $1400$  cm<sup>-1</sup> shifted with respect to the maximum) and the strong Raman bands<sup>17</sup> at 1230 cm<sup>-1</sup> and at  $\approx 1450$ – $1500$  cm<sup>-1</sup> (a group of bands).<sup>49</sup> A Raman spectrum of Et-TCY, computed at PBE0/6-31G(d) level, nicely reproduces these features showing a strong band at 1270 cm<sup>-1</sup> and a group of strong bands in the region 1500–1550 cm<sup>-1</sup> and confirms that they correspond to modes with strong contributions by the motions indicated above. On the other hand, as anticipated above, for stCY a larger involvement of the C<sub>0</sub>–C<sub>2</sub> stretching modes is likely.

Shortly, for what concerns the C<sub>2</sub>–C<sub>0</sub>–C<sub>2</sub>' moiety, while for stCY the S<sub>0</sub> → S<sub>1</sub> transition has a predominant bonding/antibonding character, for H-TCY and Et-TCY it can be better described as nonbonding/nonbonding.

Both stCY and H-TCY show a S<sub>1</sub>/S<sub>0</sub> conical intersection which is reached from S<sub>1</sub>-Min, increasing the energies of both electronic states and moving along the asymmetric stretching and the pyramidalization at the nitrogen N<sub>3</sub>' of the rotated moiety. Though the qualitative picture is similar, there is a substantial quantitative difference inasmuch that the S<sub>1</sub>-Min and CI structures are much closer in stCY than in TCY, both from the energetic and the geometrical point of view. In fact in stCY the gap on the S<sub>1</sub> surface from S<sub>1</sub>-Min to CI is just 0.05 eV at the CAS level, while it rises to about 0.5 eV for H-TCY at the TD-PBE0 level. Even if, as discussed above, this value is an upper limit, a significant increase of this energetic gap from stCY to H-TCY is expected, due to the cost to distort the fused ring. In particular, while at S<sub>1</sub>-Min the hybridization of the N<sub>3</sub>' atom is sp<sup>2</sup> (sum of the nitrogen valence bond angles 354°) in x-TCY, it is almost sp<sup>3</sup> in stCY (sum of the nitrogen valence bond angles 336°). Notice that the sp<sup>2</sup> hybridization of N<sub>3</sub>' in x-TCY is not an artifact of the CIS optimization since this method correctly reproduces (apart from a slight underestimation) the N<sub>3</sub>' hybridization in stCY (see Supporting Information). Therefore, we are dealing with a physical effect: when the ring π-system geometrical and electronic constraints are absent, N<sub>3</sub>' pyramidalization is indeed much easier. These differences, together with those in the path from FC to S<sub>1</sub>-Min reported above, suggest that the isomerization is much faster in stCY than in TCY.

**Nature of the Fused Ring and of the Specific Heteroatom.** Once verified that the presence of the fused rings deeply affects cyanine photoisomerization, it is interesting to investigate what is the effect of variations in their chemical structure. The presence of the methyl substituent on C<sub>2</sub>/C<sub>2</sub>' atoms in CH<sub>3</sub>-stCY (see Figure 1) does not remarkably change the picture obtained for stCY (see Table 4).

This table also shows that, analogously, from the qualitative point of view, the behavior of the other fused ring cyanines (reported in Figure 1) is very similar to that of H-TCY. In particular, the bond order variation associated with the S<sub>0</sub> → S<sub>1</sub> transition is quite limited as well as the C<sub>2</sub>–C<sub>2</sub>' antibonding character of the LUMO. Nevertheless, it is worthy to notice that when the sulfur atoms are substituted by oxygen atoms (oxacyanine, OCY) or CH<sub>2</sub> groups (CH<sub>2</sub>CY) the C<sub>0</sub>–C<sub>2</sub>/C<sub>2</sub>' bonds in the S<sub>1</sub> state are predicted to be slightly stronger, suggesting that for those



**Figure 9.** Energy of the S<sub>1</sub> state as a function of the φ' dihedral for four different cyanine molecules. TD-PBE0/6-31+G(d,p)//CIS/6-31G(d) calculations.

species the distortion from the planarity is less favored than in H-TCY. Figure 9, comparing the TD-PBE0/6-31+G(d,p)//CIS/6-31G(d) cP path for H-TCY, OCY, and CH<sub>2</sub>CY, confirms this hypothesis. In fact also OCY and CH<sub>2</sub>CY exhibit two minima on the S<sub>1</sub> surfaces: the first, rather shallow, at φ' ≈ 180°, the second, the absolute minimum in the isomerization reaction path, with a pseudoperpendicular arrangement of the fused rings, corresponding to a TICT. Nevertheless, the energy barrier associated with the motion from the planar to the perpendicular minimum is remarkably larger, for OCY ( $\approx 500$  cm<sup>-1</sup>) and for CH<sub>2</sub>CY ( $\approx 950$  cm<sup>-1</sup>) than for H-TCY, suggesting that for OCY and CH<sub>2</sub>CY isomerization should be slower than for H-TCY and Et-TCY. This prediction nicely agrees with the experimental finding that the S<sub>1</sub> fluorescence lifetime for Et-OCY is longer than for Et-TCY (6.4 and 1.7 ps, respectively).<sup>15</sup>

A possible explanation of the lower barrier and the easier asymmetrization (which occurs at configurations closer to be planarity) in x-TCY is that the intramolecular charge-transfer process, and hence the asymmetrization of the molecule, can be assisted by the availability of the low-lying 3d orbital on the sulfur atoms for the formation of multiple bond with C<sub>2</sub> atoms.

Figure 9 reports also the cP curve of 2-Quino, which though being very flat up to 120°, does not exhibit any energy barrier before falling in the asymmetric TICT minimum at 90°. The larger steric hindrances of pyridine-like rings increases indeed the driving force to the torsion for 2-Quino, as testified by the fact that also the S<sub>0</sub> minimum energy geometry is not planar.

As a last remark we notice that a further support to the reliability of our calculations is the fact that the absorption maxima of OCY and of 2-Quino in methanol are predicted at 3.76 and 2.80 eV, respectively i.e.,  $\approx 0.4$  eV blue-shifted and  $\approx 0.55$  red-shifted with respect to H-TCY. Those values are in very nice agreement with the experimental shift between Et-TCY, Et-OCY, and Et-2-Quino (0.33 eV and  $-0.57$  eV, respectively).<sup>14</sup>

## 7. Discussion and Concluding Remarks

In this paper we have performed a thorough theoretical analysis of the S<sub>1</sub> excited-state surface of several *n* = 0<sup>14</sup> cyanines, with a special focus on two of them, namely Et-TCY and H-TCY. The absorption and the fluorescence spectra computed in solution compare nicely with the available experimental results, supporting the reliability of our conclusions. Generally, the comparison with experimental

data for TCY, and with CASMP2 results for stCY,<sup>51</sup> indicate that the computational methods adopted here give results quantitatively reliable for geometries close to planarity, while the TD-PBE0  $S_1$  energies must be considered affected by noticeable errors at strongly distorted geometries, where the  $S_0 \rightarrow S_1$  transition has a clear charge transfer character. This would be a serious problem in detailed dynamical calculations, but, for the scopes of the present paper, the results presented still furnish a qualitatively correct description of the  $S_1$  surface, as shown by the comparison with CASPM2 in the stCY case. The substituent effects here discussed are already operative for geometries close to planarity, where our results are reliable.

Due to their specific structure, cyanines show a rich dynamics with a competition between symmetric and asymmetric reactive paths and are a nice example of how the substituents can partially modify the nature itself of the optical transition, altering the initial driving force to the reaction.

Although the orbitals involved in the  $S_0 \rightarrow S_1$  electronic transition are similar, the presence of the fused aromatic rings produces indeed some significant difference in x-TCY with respect to the model compound stCY. In x-TCY the  $S_0 \rightarrow S_1$  transition is mainly nonbonding/nonbonding, and the  $S_1$  surface around the FC region is rather flat with a scarce driving force to torsion. A shallow slightly nonplanar local minimum  $S_1$ -MinC<sub>2</sub> exists on the  $S_1$  surface, separated by a small barrier ( $\approx 120$  cm<sup>-1</sup>) from the absolute asymmetrical minimum  $S_1$ -Min, corresponding, as in stCY, to a TICT state.

Our results strongly suggest that  $S_1$ -MinC<sub>2</sub> is the state responsible for the Stokes-shifted fluorescence and for the  $S_1 \rightarrow S_n$  transient absorption. Although a full dynamical study would be necessary to compute the lifetime of  $S_1$ -MinC<sub>2</sub>, the topology of the  $S_1$  surface and the existence of the barrier are consistent with a picosecond(s) lifetime.

A barrier on the  $S_1$  minimum energy path is predicted for several of the investigated cyanines (x-TCY, OCY, and CH<sub>2</sub>CY), and it is suggested that a key factor determining its existence is the degree of planarity of the  $S_0$  equilibrium geometry. When, due to steric hindrances, the  $S_0$  minimum is already strongly twisted,  $S_1$  PES could be barrierless (as in 2-Quino), and/or the FC excitation could occur in a region of the  $S_1$  surface where the electronic effects responsible for the barrier are no more active. The existence of barriers on  $S_1$  and hence local minima has been predicted<sup>50</sup> and is generally accepted in cases of carbocyanine ( $n > 0$ ).

Concerning the possible isomerization path, our results suggest that the cP passing through the  $S_1$ -MinC<sub>2</sub> intermediate and other paths which remove the symmetry already in the FC region are probably competitive. This possible competition is enhanced by the scarce initial driving force to the torsion which is clearly lower in Et-TCY than in the smaller stCY. The exciting possibility that in Et-TCY the wave packet can follow simultaneously different paths toward the minimum  $S_1$ -Min could play an important role in the coherent control of the Et-TCY photoisomerization which has been recently attained.<sup>22</sup> In fact, the parts of the wave packet which reach  $S_1$ -Min along these different paths could interfere depending on their relative phases, as in molecular

version of the Young double-slit interferential experiment, and this phenomenon would influence the later dynamics of the wave packet.

Apart from the insights on the dynamical behavior of Et-TCY, a synoptical analysis of the shape of the frontier orbitals of different cyanines not only explains the features depicted above but also allows us to sketch which are the most important interactions modulating the planarity of the cyanine moiety: (i) the C<sub>2</sub>-C<sub>0</sub>-C'<sub>2</sub> bonding character of the HOMO, favoring a planar conformation of the allyl moiety, (ii) the C<sub>2</sub>-C'<sub>2</sub> antibonding character of the LUMO resulting in a driving force to distortion from the planarity of the cyanine, which increases at the increase of the weight of those atoms in the LUMO, and (iii) steric repulsion between the substituents on C<sub>2</sub> and C'<sub>2</sub> atoms. The first two factors explain the different flatness with the respect the initial torsion around the  $\phi'$  dihedral of the  $S_1$  PES of x-TCY and of stCY. The larger steric hindrance of the ethyl substituents gives instead an account for the energy barrier higher in Et-TCY than in H-TCY. The inspection of the frontier orbitals also explains why the pyramidalization of the N<sub>3</sub>' center is required to go from the  $S_1$  minimum to the  $S_1/S_0$  conical intersection.

The photophysical behavior of cyanine depends on the chemical structure of the fused aromatic ring as well. The peculiar electronic structure of sulfur atoms can explain why in x-TCY the energy barrier connecting  $S_1$ -MinC<sub>2</sub> and  $S_1$ -Min is noticeably lower than in OCY and CH<sub>2</sub>CY, in agreement with experimental indications.

Our results confirm for Et-TCY the general description of the photoisomerization reaction predicted by Galvez et al.<sup>29</sup> on stCY, but they also point out significant quantitative differences showing that ad hoc modifications, even small, of the molecular scaffold and/or of its substituents can be used to modulate the dynamical behavior of  $n = 0$  cyanines, suggesting which physicochemical effects (either electronic or steric) to rely on in order to design cyanines with optimized excited-state properties.

As a last remark cyanine is a clear example where remarkable changes in the electronic structure along the path have a large purely dynamical effect on the reaction. In fact, in correspondence to the shift from a symmetric to an asymmetric electronic density distribution, there is a change in the nature itself of the nuclear modes involved in the motion. This effect is likely to slow the reaction dynamics, even if the surface does not exhibit any significant energy barrier, since an energy redistribution is necessary to proceed along the reactive path. Of course also the solvent is sensible to a sudden change in the solute electronic density so that, beyond what is done by frictional effects depending on the shape of the twisting moiety, also solvent nonequilibrium effects are expected to contribute to slow the reaction rate.

**Acknowledgment.** The authors thank G. Krampert, P. Niklaus, G. Vogt, and G. Gerber of the Universität Würzburg who stimulated this study for many valuable discussions. F.S. acknowledges the hospitality of the Physikalisches Institut, Universität Würzburg (Germany) and the support of the European Community through the COCOMO Network.

**Supporting Information Available:** Detailed comparison of the performance on stCY of TD-PBE0//CASSCF and TD-PBE0//CIS methods utilized in this paper with respect to CASMP2//CASSCF of ref 29, excited-state dipole moment and solvation energy, and tables with the Cartesian geometries of the main structures of H-TCY and Et-TCY discussed in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- Zewail A. H. *Femtochemistry: Ultrafast Dynamics of the Chemical Bond*; World Scientific Publishing Co. Pte. Ltd.: 1994; Vol. I and II.
- Femtochemistry and Femtobiology*; Douhal, A., Santamaria, J., Eds.; World Scientific Publishing Co. Pte. Ltd.: 2002.
- (a) Judson R. S.; Rabitz H. *Phys. Rev. Lett.* **1992**, *68*, 1500. (b) Assion A.; Baumert T.; Bergt M.; Brixner T.; Kiefer B.; Seyfried W.; Strehle M.; Gerber G. *Science* **1998**, *282*, 918. (c) Shapiro M.; Brumer P. *Principles of the Quantum Control of Molecular Processes*; J. Wiley & Sons Inc.: 2003.
- Conical Intersections: Electronic Structure, Dynamics & Spectroscopy*; Domcke, W., Yarkony, D. R., Köppel, H., Eds.; World Scientific Publishing Co. Pte. Ltd.: 2004.
- Domcke, W.; Stock, G. *Adv. Chem. Phys.* **1997**, *100*, 1.
- (a) Ciminelli, C.; Granucci, G.; Persico, M. *Chem. Eur. J.* **2004**, *10*, 2327. (b) Weigart, O.; Migani, A.; Olivucci, M.; Robb, M. A.; Buss, V.; Hunt, P. *J. Phys. Chem. A* **2004**, *108*, 4685. (c) Ko, C.; Levine, B.; Toniolo, A.; Manohar, L.; Olsen, S.; Werner, H.-J.; Martinez, T. J. *J. Am. Chem. Soc.* **2003**, *125*, 12710.
- Quenneville, J.; Martínez, T. J. *J. Phys. Chem. A* **2003**, *107*, 10076, see also references therein.
- González-Luque, R.; Garavelli, M.; Bernardi, F.; Merchán, M.; Robb, M. A.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9379, see also references therein.
- (a) Fabian, J.; Nakazumi, H.; Matsuoka, M. *Chem. Rev.* **1992**, *92*, 1197. (b) Czerney, P.; Graness, G.; Birckner, E.; Vollmer, F.; Rettig W. *J. Photochem. Photobiol. A* **1995**, *89*, 31. (c) Kaliteevskaya, E. N.; Razumova, T. K.; Tarnovskii, A. N. *Opt. Spectrosc.* **1999**, *1*, 126.
- (a) Naber, A.; Fischer, U. C.; Kirchner, S.; Dziomba, T.; Kollar, G.; Chi, L. F.; Fuchs, H. *J. Phys. Chem. B* **1999**, *103*, 2709. (b) Spittler, M. T.; Ehret, A.; Kietzmann, R.; Willig, F. *J. Phys. Chem. B* **1997**, *101*, 2552.
- (a) Seifert, J. L.; Connor, R. E.; Kushon, S. A.; Wang, M.; Armitage, B. A. *J. Am. Chem. Soc.* **1999**, *121*, 2987. (b) Nunnally, B. K.; He, H.; Li, L. C.; Tucker, S. A.; McGown, L. B. *Anal. Chem.* **1997**, *69*, 2392. (c) Rettig, W.; Rurack, K.; Szczepan M. In *New trends in Fluorescence Spectroscopy. Applications to Chemical and Life Sciences*; Valeur, B., Brochon, J. C., Eds.; Springer-Verlag: Berlin, 2000.
- (a) Jelly, E. E. *Nature* **1936**, *138*, 1009. (b) Scheibe, G. *Angew. Chem.* **1937**, *50*, 51; **1937**, *50*, 212. (c) Renge, I.; Wild, U. P. *J. Phys. Chem. A* **1997**, *101*, 7977.
- Mandal, A. K.; Pal, M. K. *Chem. Phys.* **2000**, *253*, 115.
- Meyer, Y. H.; Pittman, M.; Plaza, P. *J. Photochem. Photobiol. A* **1998**, *114*, 1.
- Roth, N. J. L.; Craig, A. C. *J. Phys. Chem.* **1974**, *78*, 1154.
- Sahyun, M. R.; Blair, J. T. *J. Photochem. Photobiol. A* **1997**, *104*, 179.
- Fujimoto, Y.; Katayama, N.; Ozaki, Y.; Yasui, S.; Iriyama, K. *J. Mol. Struct.* **1992**, *274*, 183.
- (a) Hofer, L. J. E.; Grabenstetter, R. J.; Wiig, E. O. *J. Am. Chem. Soc.* **1950**, *72*, 203. (b) Benson, R. C.; Kues, H. A. *J. Chem. Eng. Data* **1977**, *22*, 379.
- Tredwell, C. J.; Keary, C. M. *Chem. Phys.* **1979**, *43*, 307.
- Rentsch, S. K.; Danielius, R.; Gadonas, R. *Chem. Phys. Lett.* **1981**, *84*, 450.
- (a) Levitus, M.; Martin Negri, R.; Aramendia, P. F. *J. Phys. Chem.* **1995**, *99*, 14231. (b) Martini, I.; Hartland, G. V. *J. Phys. Chem.* **1996**, *100*, 19764. (c) Chibisov, A. K.; Zakharova, G. *J. Chem. Soc., Faraday Trans.* **1996**, *92*, 4917. (d) Murphy, S.; Sauerwein, B.; Drickamer, H. G.; Schuster, G. B. *J. Phys. Chem.* **1994**, *98*, 13476. (e) Aramendia, P. F.; Martin Negri, R.; San Roman, E. *J. Phys. Chem.* **1994**, *98*, 3165.
- Vogt, G.; Krampert, G.; Niklaus, P.; Gerber, G. personal communication.
- (a) Kawasaki, M.; Inokuma, H. *J. Phys. Chem. B* **1999**, *103*, 1233. (b) Owen, D. J.; VenDerveer, D.; Schuster, G. B. *J. Am. Chem. Soc.* **1998**, *120*, 1705. (c) Laguitton-Pasquier, H.; Van der Auweraer, M.; De Schryver, F. C. *Langmuir* **1998**, *14*, 5172. (d) Botelho do Rego, A. M.; Penedo Pereira, L.; Reis, M. J.; Oliveira, A. S.; Vieira Ferreira, L. F. *Langmuir* **1997**, *13*, 6787.
- (a) Mandal, A. M.; Pal, M. K. *Spectrochim. Acta A* **1999**, *55*, 1347. (b) Ghosh, J. K.; Mandal, A. K.; Pal, M. K. *Spectrochim. Acta A* **1999**, *55*, 1877.
- Aramendia, P. F.; Martin Negri, R.; San Roman, E. *J. Phys. Chem.* **1994**, *98*, 3165.
- Rulliere, C. *Chem. Phys. Lett.* **1976**, *43*, 303.
- (a) Kasatani, K.; Sato, H. *Bull. Chem. Soc. Jpn.* **1996**, *69*, 3455. (b) Eske, A. T.; Razi N. K. *Chem. Phys. Lett.* **1979**, *63*, 128. (c) De Rossi, U.; Daehne, S.; Reisfeld, R. *Chem. Phys. Lett.* **1996**, *251*, 259.
- Momicchioli, F.; Baraldi, I.; Berthier, G. *Chem. Phys.* **1988**, *123*, 103.
- Sanchez-Galvez A.; Hunt, P.; Robb, M. A.; Olivucci, M.; Vreven, T.; Schlegel, H. B. *J. Am. Chem. Soc.* **2000**, *122*, 2911.
- Cossi, M.; Rega, N.; Scalmani, M.; Barone, V. *J. Chem. Phys.* **2001**, *114*, 5691.
- (a) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158. (b) Enzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029.
- Adamo, C.; Scuseria, G. E.; Barone, V. *J. Chem. Phys.* **2000**, *111*, 2889.
- Improta, R.; Santoro, F.; Dietl, C.; Papastathopoulos, E.; Gerber, G. *Chem. Phys. Lett.* **2004**, *387*, 509.
- Bearpark, M. J.; Robb M. A.; Schlegel, H. B. *Chem. Phys. Lett.* **1994**, *223*, 269.
- Fantacci, S.; Migani A.; Olivucci M. *J. Phys. Chem. A* **2004**, *108*, 1208.
- Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.
- Cossi, M.; Barone, V. *J. Chem. Phys.* **2001**, *115*, 4708.



- (38) Wiberg, K. B. *Tetrahedron* **1968**, *24*, 1083.
- (39) (a) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211. (b) Reed, A.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066. (c) Glendening, E. D.; Weinhold, F. *J. Comput. Chem.* **1998**, *19*, 593.
- (40) Gaussian 03, Revision B.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 2003.
- (41) Wanko, M.; Garavelli, M.; Bernardi, F.; Niehaus, T. A.; Frauenheim, T.; Elstner, M. *J. Chem. Phys.* **2004**, *120*, 1674.
- (42) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 2943.
- (43) Ben-Nun, M.; Quenneville, J.; Martínez, T. J. *J. Phys. Chem. A* **2000**, *104*, 5161.
- (44) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
- (45) Zhang, T.; Chen, C.; Gong, Q.; Yan, W.; Wang, S.; Yang, H.; Jian, H.; Xu, G. *Chem. Phys. Lett.* **1998**, *105*, 236.
- (46) Åberg, U.; Åkesson, E.; Fedchenia, I.; Sundström, V. *Chem. Phys.* **1994**, *183*, 269.
- (47) Yartsev, A.; Alvarez, J.-L.; Åberg, U.; Sundström, V. *Chem. Phys. Lett.* **1995**, *243*, 281.
- (48) Xu, Q.-H.; Fleming, G. R. *J. Phys. Chem. A* **2001**, *105*, 10187.
- (49) Though the authors<sup>17</sup> speak of a Resonance Raman experiment, we notice that at the 480 nm exciting wavelength they used, the absorption spectrum is vanishingly small. Therefore it is unlikely that any intermediate state is significantly populated, and at least the existence of a nonresonant contribution should be considered in assignment of the bands.
- (50) Rodríguez, J.; Sherlis, D.; Aramendía, P. F.; Negri, R. M. *J. Phys. Chem. A* **1997**, *101*, 6998.
- (51) Concerning the stCY vertical excitation, the agreement with experiment is deeply improved if the computation is performed at CASPT2 level of theory. See Schreiber, M.; Buss, V.; Fülcher, M. *Phys. Chem. Chem. Phys.* **2001**, *3*, 3906.

CT049899R

# JCTC

Journal of Chemical Theory and Computation

## Ab Initio and DFT Conformational Studies of Propanal, 2-Butanone, and Analogous Imines and Enamines

Haizhen Zhong,<sup>†</sup> Eugene L. Stewart,<sup>‡</sup> Maria Kontoyianni,<sup>§</sup> and J. Phillip Bowen<sup>\*,†</sup>

*Center for Drug Design, Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, North Carolina 27402, Computational Center for Molecular Structure and Design, Department of Chemistry, University of Georgia, Athens, Georgia 30602, and Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599*

Received November 11, 2004

**Abstract:** The potential energy surfaces (PES) of 2-butanone, 2-butanamine, 1-butanamine, propanal, and propanimine have been explored with ab initio and DFT calculations at the RHF/6-311G\*\*, MP2/6-311G\*\*, and B3LYP/6-311G\*\* levels of theory. In agreement with previous experimental and computational results, the PES provides two minima for each of the above molecules with the exception of 2-butanone, which clearly shows three distinct minima. Factors influencing the conformational preferences are also elaborated. Our calculations suggest that for 2-butanone and propanal, the steric and the bond dipole interactions are primarily responsible for the conformational preferences of these compounds. Additional charge–charge interactions might also play an important role in determining the imine conformations. For enamines, however, steric interactions play a critical role, with bond dipole interactions exerting some influence. Our results also suggest that for imine formation from butanone and/or propanal, the imine is the predominant product, not the enamine, which is consistent with experimental observations. Therefore, these calculations should provide a better understanding of the ketone/aldehyde to imine and enamine transformations. This transformation may introduce an important imine moiety for the analogues of *trans-N*-methyl-4-(1-naphthylvinyl)pyridine (NVP), a choline acetyltransferase (ChAT) inhibitor.

### Introduction

Imine formation is a particularly important chemical reaction in many biological processes.<sup>1,2</sup> For example, the covalent binding of carbonyl-containing compounds to an enzyme usually involves the formation of an imine. The imine moiety is formed by condensation of a ketone (or aldehyde) carbonyl group with a primary or secondary amine. Equilibrium between imine and enamine may be established when at least one hydrogen atom on the imine nitrogen (Scheme 1). The

relative equilibrium of these species depends on the symmetry of the parent ketone and the substituents on the amine.

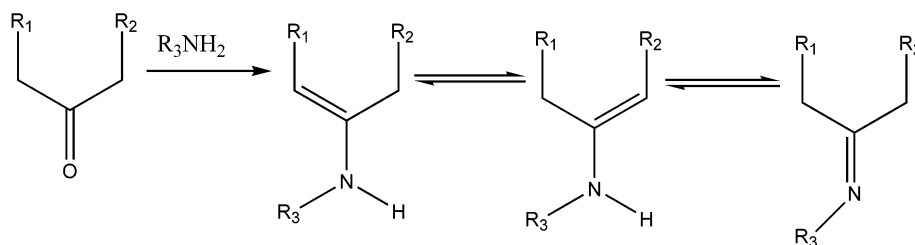
Since the imine bond is labile and readily hydrolyzed to the corresponding ketone, it is difficult to accurately monitor and determine which factors influence the extent of imine formation. For example, experiments using Raman and infrared spectroscopy to study the stretching frequencies of imines have been reported.<sup>3</sup> By studying the kinetics of imine formation from acetophenone and substituted aniline, Lee et al. were successful in monitoring the conversion from a ketone to an imine.<sup>3</sup> Egawa and Konaka<sup>4</sup> reported a study on the molecular structure of 2-butanamine, determined from gas-phase electron diffraction experiments as well as MP2 and DFT calculations. Their results indicate the (*E*)-sp con-

\* Corresponding author phone: (336)334-4257; fax: (336)334-5402; e-mail: jpbowen@uncg.edu.

<sup>†</sup> University of North Carolina at Greensboro.

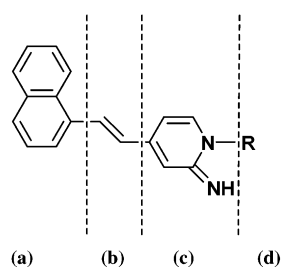
<sup>‡</sup> University of Georgia.

<sup>§</sup> University of North Carolina at Chapel Hill.

**Scheme 1.** Possible Reaction Pathway and Equilibria from Ketones to Enamines to Imines

formation as the most favorable for 2-butanamine. The use of an assumed dihedral angle ( $\varphi_{\text{NCCC}}$  at  $117.6^\circ$ ), however, introduces bias into the calculations. A number of additional computational studies focusing on ketones and/or imines have also been reported.<sup>5,6</sup> However, all of the aforementioned experimental and computational studies did not address imine-enamine tautomerism. In addition, the factors influencing the conformational preferences for the relevant ketone, imine and enamine have not been established.

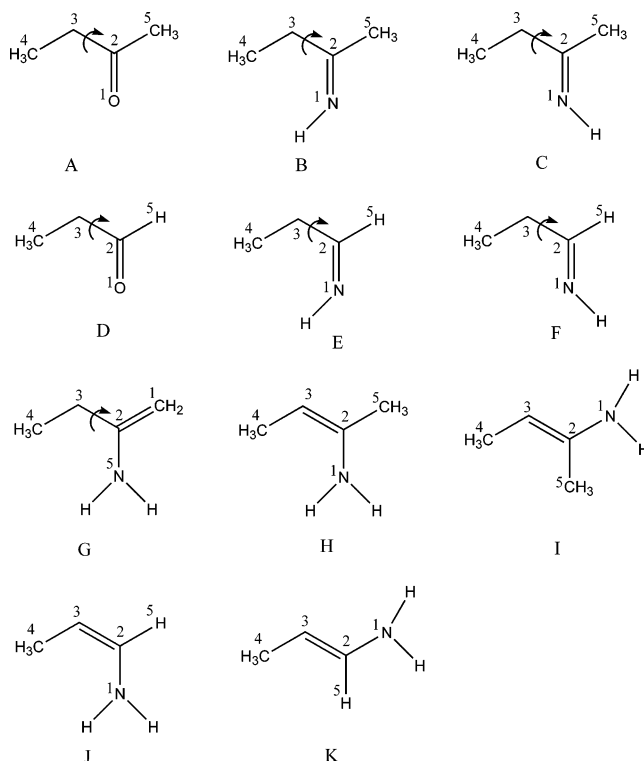
Our interest in the conformational preferences of imines stems in part from ongoing investigations into understanding the structural requirements for enzyme choline acetyltransferase (ChAT) inhibitors. Effective inhibitors of ChAT presumably will reduce the levels of acetylcholine. This leads to intriguing suggestion that ChAT inhibitors, alone or coupled with other agents, might be used as potential prophylactic protecting agents for those who might be exposed to nerve gases that block acetylcholine esterase as their mechanism of action.<sup>7</sup> Among the types of inhibitors of ChAT reported, *trans-N*-methyl-4-(1-naphthylvinyl)pyridine (NVP) analogues (Figure 1)<sup>8–10</sup> represent the most

**Figure 1.** *trans*-1-Alkyl-4-(1-naphthylvinyl)pyridinium (NVP) analogue.

extensively explored compounds for structure activity relationship (SAR) purposes using classical approaches. NVP analogues are divided into four regions labeled *a*, *b*, *c*, *d* (Figure 1). Region *a* contains an aromatic group that is attached via a *trans*-ethylene linkage, region *b*, to a pyridinium ring that comprises region *c*. Region *d* defines the various alkyl groups that can be bonded to the pyridinium nitrogen. In region *c*, 2-pyridone-imine analogues of NVP were also reported as a stronger base than pyridine.<sup>8</sup> Reports also show that an imine moiety  $-\text{N}=\text{CH}$  in region *b* enhances the activity.<sup>9</sup> All of these stimulated our interest in the fundamental conformational aspects of imines, how they differ from the corresponding carbonyl functional groups, and what significance this could have on inhibitor design.

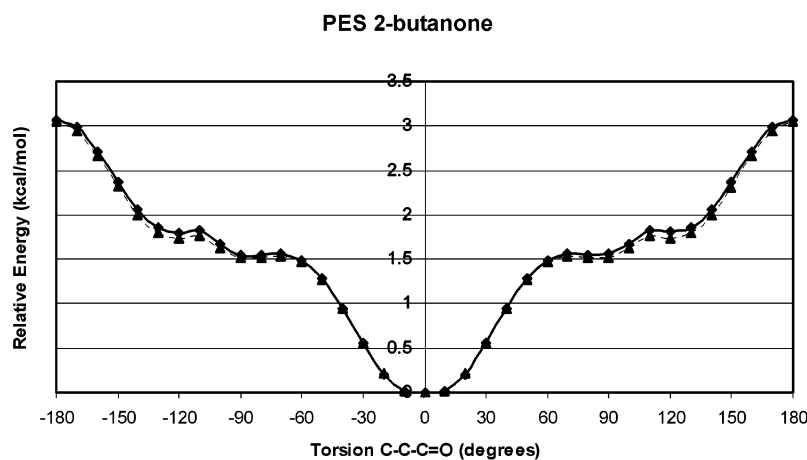
To provide insight into imine-enamine tautomerism as well as the factors that determine the conformational preferences

of ketones, imines, and enamines, we performed ab initio RHF/6-311G\*\*, MP2/6-311G\*\* and DFT B3LYP/6-311G\*\* calculations on various ketones and aldehydes and their corresponding imines and enamines: i.e., 2-butanone, *Z/E*-2-butanamine, 1-butenamine, and *Z/E*-2-butenamine, propanal, *Z/E*-propanimine, and *Z/E*-propenamine (Figure 2). We also hoped to investigate the effect of the *Z/E* configurations of these compounds on their conformations.

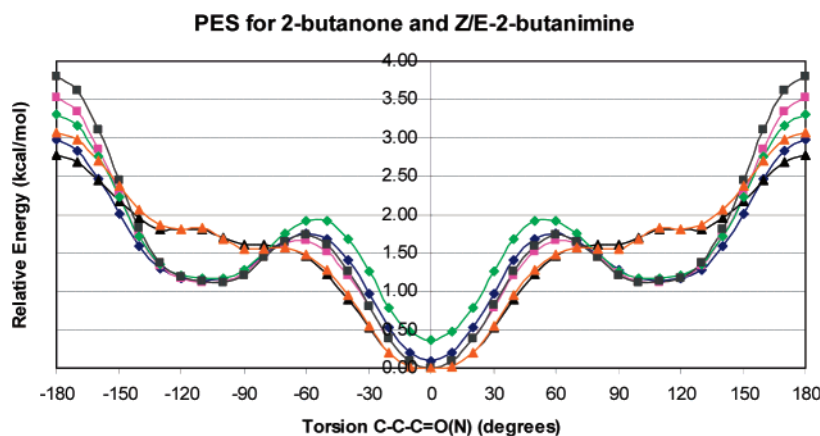
**Figure 2.** Chemical structures of 2-butanone (A); *Z*-2-butanamine (B); *E*-2-butanamine (C); propanal (D); *Z*-propanimine (E); *E*-propanimine (F); 1-butenamine (G); *Z*-2-butenamine (H); *E*-2-butenamine (I); *Z*-propenamine (J); and *E*-propenamine (K).

## Methods

In the present study, all ab initio and density functional theory (DFT) calculations were carried out using the Gaussian03 package.<sup>11</sup> The RHF and the Møller–Plesset perturbation theory (MP2)<sup>12</sup> of ab initio methods and Becke's three parameter exact exchange functional (B3)<sup>13</sup> combined with gradient corrected correlation functional of Lee–Yang–Parr (LYP)<sup>14</sup> of the DFT method have been employed to optimize the geometry by implementing the 6-311G\*\* basis set. During both SCF and geometry optimizations, the default convergence criteria were used. For seven molecules in



**Figure 3.** Potential energy surface for 2-butanone at MP2/6-311G\*\* (diamond, black) and at RHF/6-311G\*\* (grey, triangle).



**Figure 4.** Potential energy surface for 2-butanone and its analogues. The color codes are Z-2-butanimine (B3LYP, blue, diamond); E-2-butanimine (B3LYP, magenta, square); 2-butanone (B3LYP, black, triangle); Z-2-butanimine (MP2, green, diamond); E-2-butanimine (MP2, gray, square); and 2-butanone (MP2, orange, triangle), respectively.

Figure 2 (A to G), complete exploration of their conformational space was carried out along the coordinate of the dihedral angle as defined by atom1-atom2-atom3-atom4 over a range of  $360.0^\circ$  at an increment of  $10.0^\circ$ . All reported minima along the potential energy surface were subjected to full geometry optimizations. The minima were confirmed using frequency calculations. The infrared data are reported, and the respective infrared peak assignments were confirmed by Gaussview. Transition structures were located using the STQN method and were confirmed with frequency calculations. For molecules H to K in Figure 2, no dihedral scan was applied because of the coplanar nature of the C=C double bonds, and only optimization and frequency calculations were carried out. NBO charges were determined for each of the optimized stationary points. Solvent effects for the minima were investigated by performing single-point energy calculations using self-consistent reaction field (SCRFF) theory with the isodensity surface polarized continuum model (IPCM) method in Gaussian03 at the RHF/6-311G\*\*,<sup>15</sup> The SCRFF-IPCM calculations were carried out in chloroform, methanol and water at 298.15 K with solvent dielectric constants of 4.90, 32.63, and 78.39, respectively.

## Results and Discussion

The PES maps for the seven molecules in Figure 2 (A to G), derived from RHF/6-311G\*\*, MP2/6-311G\*\*, B3LYP/

6-311G\*\* levels of theory, are given in Figures 1S-7S (see Supporting Information). As shown in Figure 3, the PES trajectories from RHF/6-311G\*\* (grey dash line) and MP2/6-311G\*\* (black line) are very close. To simplify the comparisons, only data from MP2/6-311G\*\* and B3LYP/6-311G\*\* are reported in the PES figures herein. Systematic searching of the conformational space for 2-butanone from MP2/6-311G\*\* yielded five minima. As defined by the  $C_4-C_3-C_2=O$  torsional angles, these angles were approximately  $0.0^\circ$ ,  $90.0^\circ$ ,  $120.0^\circ$ ,  $-120.0^\circ$ ,  $-90.0^\circ$  (Figure 3). The symmetry of 2-butanone provides for a conformational degeneracy in the molecule and gives rise to only three distinct conformations. There are only two distinct minima for both Z- and E-2-butanimine (Figure 4), where Z- and E- designate the configuration of the N-H bond (cis and trans to the ethyl group, respectively). The gas-phase energies and characteristic torsion angles for all seven minima are reported in Table 1. The differences in energy for these seven minima are less than 4.00 kcal/mol. In any case, the cis conformation has the lowest gas-phase energy (Table 1 and Figures 3 and 4). The cis, skew, gauche, and trans conformations as defined by the  $C_4-C_3-C_2=O$  dihedral angle are  $0.0^\circ$ ,  $90.0^\circ$ ,  $120.0^\circ$ ,  $180.0^\circ$ , respectively. The activation energy barriers among each of these three minima are less than 2.0 kcal/mol, indicating that these conformations are interchangeable at room temperature.

**Table 1.** Ab Initio and DFT Gas-Phase Energies and Characteristics of the Conformers for 2-Butanone and Analogs Calculated at the 6-311G\*\* Level of Theory

| conformer | name                           | torsion <sup>a</sup> | HF energy (Hartree) | $\Delta E$ (HF) (kcal/mol) | MP2 energy (Hartree) | $\Delta E$ (MP2) (kcal/mol) | B3LYP energy (Hartree) | $\Delta E$ (B3LYP) (kcal/mol) |
|-----------|--------------------------------|----------------------|---------------------|----------------------------|----------------------|-----------------------------|------------------------|-------------------------------|
| 1         | <i>cis</i> -butanone           | 0.0                  | -231.058385417      | 0.00                       | -231.056291767       | 0.00                        | -232.538467751         | 0.00                          |
| 2         | <i>skew</i> -butanone          | 85.4                 | -231.055975740      | 1.51                       | -231.053829221       | 1.55                        | -232.535899977         | 1.61                          |
| 3         | <i>gauche</i> -butanone        | 120.8                | -231.055619900      | 1.74                       | -231.053401635       | 1.81                        | -232.535608393         | 1.79                          |
| 4         | <i>trans</i> -butanone-TS      | 180.0                | -231.053509290      | 3.06                       | -231.051406829       | 3.07                        | -232.534052150         | 2.77                          |
| 5         | <i>cis-Z</i> -butanimine       | 0.0                  | -211.206000593      | 0.39                       | -211.203861555       | 0.36                        | -212.650927530         | 0.09                          |
| 6         | <i>gauche-Z</i> -butanimine    | 104.1                | -211.204828851      | 1.12                       | -211.202506462       | 1.21                        | -212.649273550         | 1.13                          |
| 7         | <i>trans-Z</i> -butanimine-TS  | 180.0                | -211.201297364      | 3.34                       | -211.199159258       | 3.31                        | -212.646346026         | 2.97                          |
| 8         | <i>gauche-Z</i> -butanimine-TS | 60.4                 | -211.203570010      | 1.91                       | -211.201379150       | 1.92                        | -212.648286182         | 1.75                          |
| 9         | <i>cis-E</i> -butanimine       | 0.0                  | -211.206615733      | 0.00                       | -211.204438625       | 0.00                        | -212.651078141         | 0.00                          |
| 10        | <i>gauche-E</i> -butanimine    | 101.6                | -211.204920944      | 1.06                       | -211.202663787       | 1.11                        | -212.649302964         | 1.11                          |
| 11        | <i>trans-E</i> -butanimine-TS  | 180.0                | -211.200556444      | 3.80                       | -211.198371715       | 3.81                        | -212.645467645         | 3.52                          |
| 12        | <i>gauche-E</i> -butanimine-TS | 60.0                 | -211.203946883      | 1.67                       | -211.201683488       | 1.73                        | -212.648431749         | 1.66                          |
| 13        | 2-butyl- <i>E</i> -enamine     | -3.1                 | -211.197879766      | 5.48                       | -211.196080926       | 5.24                        | -212.646451815         | 2.90                          |
| 14        | 2-butyl-enamine                | 174.3                | -211.195768495      | 6.81                       | -211.194132716       | 6.47                        | -212.644203418         | 4.31                          |
| 15        | 1-butyl-enamine-0              | 177.4                | -211.196500492      | 6.35                       | -211.194955066       | 5.95                        | -212.643305187         | 4.88                          |
| 16        | 1-butyl-enamine-110            | -71.7                | -211.196617694      | 6.27                       | -211.195066736       | 5.88                        | -212.643408341         | 4.81                          |
| 17        | 1-butyl-enamine-n110           | 64.5                 | -211.197357797      | 5.81                       | -211.195763110       | 5.44                        | -212.643943377         | 4.48                          |

<sup>a</sup> Torsion angle as  $\Phi(\text{C}_4\text{-C}_3\text{-C}_2=\text{O}/\text{N})$  (degrees).

**Table 2.** Comparison of Calculated and Observed Infrared Frequencies ( $\text{cm}^{-1}$ ) and Vibrational Assignment for 2-Butanone from RHF/6-311G\*\*<sup>a</sup>

| observed infrared (gas) | calculated     |                    |                  | assignment                            |
|-------------------------|----------------|--------------------|------------------|---------------------------------------|
|                         | <i>cis</i> (0) | <i>gauche</i> (60) | <i>skew</i> (90) |                                       |
| 2902                    | 3252 (2904)    | 3250 (2902)        | 3250 (2902)      | CH <sub>3</sub> symmetric stretch     |
| 2892                    | 3238 (2891)    | 3235 (2889)        | 3239 (2892)      | CH <sub>2</sub> symmetric stretch     |
| 1739                    | 1993 (1780)    | 1991 (1778)        | 1993 (1780)      | C=O stretch                           |
| 1360                    | 1525 (1361)    | 1520 (1358)        | 1520 (1358)      | CH <sub>3</sub> symmetric deformation |
| 1186                    | 1287 (1149)    | 1328 (1186)        | 1318 (1177)      | CH <sub>3</sub> in-plane rock         |

<sup>a</sup> The frequency data within parentheses are scaled from the calculated data by a factor of 0.8929.

Ab initio calculations have previously been carried out for 2-butanone.<sup>5,16,17</sup> In previous calculations, however, full geometry optimizations failed to give three distinct conformations. By allowing the torsional coordinate to vary by 10.0° increments from 0.0° to 180.0° and constraining all other structural parameters to those obtained from the optimized *trans* conformation, Durig et al.<sup>5</sup> reported only two minima for 2-butanone, one in which the C<sub>4</sub>-C<sub>3</sub>-C<sub>2</sub>=O torsion angle was 0.0°, and the other at 100.0°. Wiberg and Martin<sup>16</sup> found that the *skew* conformer (C<sub>4</sub>-C<sub>3</sub>-C<sub>2</sub>=O at 110.0°) was 2.05 kcal/mol higher in energy than the eclipsed (*cis*) geometry. Our data show that the *gauche*-butanone (C<sub>4</sub>-C<sub>3</sub>-C<sub>2</sub>=O at 120.0°) is about 1.81 kcal/mol higher in energy than the eclipsed (*cis*) geometry (Table 1), which is in good agreement with Wiberg's paper. Both papers, however, reported an energetically flat region between 90.0° and 120.0°. The inability to locate a third minimum in previous studies probably results from the application of geometrical constraints or the optimization of the *trans* and *gauche* conformations without a torsional scan. Using MM2 calculations, Bowen et al. also reported an energetically flat region in the PES between 70.0° and 130.0°.<sup>17</sup> The nature of this region makes the minima at 90.0° and 120.0° indistinguishable. In addition, different model chemistries (basis sets) yield slightly different results. As shown in Figure 4 and Figure 1S, PES of 2-butanone from RHF is not as flat

as those from MP2 and B3LYP. The PES trajectory of 2-butanone from B3LYP is so flat between 90.0° and 120.0° that the minima region seems like a shoulder region. The full optimization of the geometries at 90.0° and 120.0° gave two minima, with energies 1.61 and 1.79 kcal/mol above the *cis* conformer.

To our knowledge, we are the first to locate three distinct minima along the PES of 2-butanone (Figures 3 and 4). The energy difference between the minima at 90.0° and 120.0°, derived from MP2, is approximately 0.26 kcal/mol. The activation energy barrier from *gauche* (120.0°) to *skew* (90.0°) and from *skew* (90.0°) to *cis* (0.0°) is less than 0.2 kcal/mol. These minor energy differences suggest that the *cis* and the *skew/gauche* conformations are readily interconvertible. The energy difference between the *cis* and *gauche/skew* forms and the preference of the *cis* conformation are in very good agreement with other experimental<sup>18</sup> and computational results.<sup>5,16,17</sup> All three minima of 2-butanone were geometrically optimized and subsequently confirmed by frequency calculations. The major peaks of the calculated infrared spectra of these three minima, which were derived from the RHF/6-311G\*\* torsional scan of 2-butanone, are listed in Table 2 and are compared with experimental data.<sup>5</sup> The excellent agreement between our calculated infrared frequencies and the experimental values demonstrates the calculation reliability.

**Table 3.** Comparison of Calculated and Observed Infrared Frequencies ( $\text{cm}^{-1}$ ) and Vibrational Assignment for 2-Butanimine from RHF/6-311G\*\*<sup>a</sup>

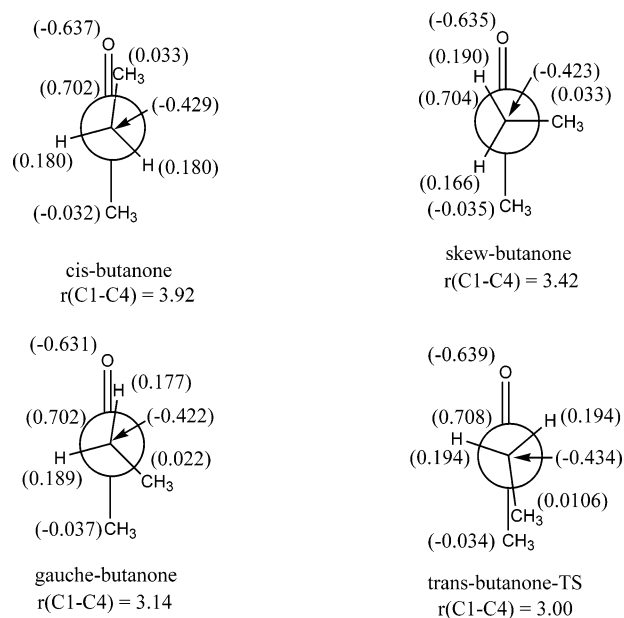
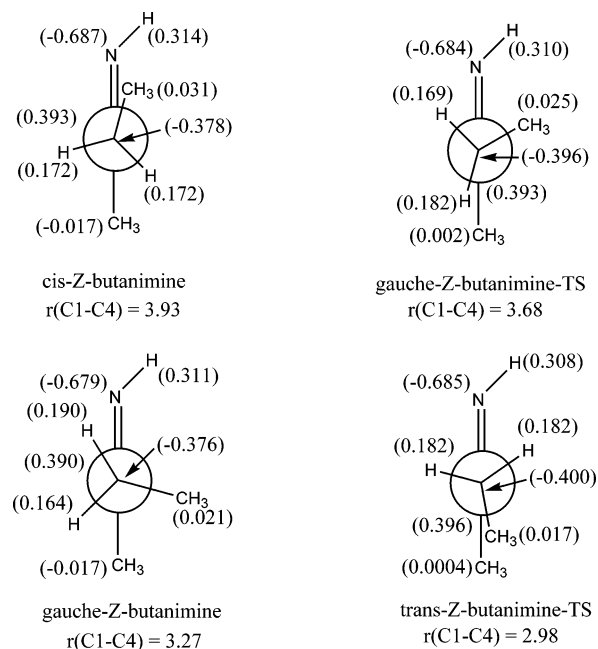
| observed infrared (gas) | calculated   |                 |              |                 | assignment                            |
|-------------------------|--------------|-----------------|--------------|-----------------|---------------------------------------|
|                         | <i>Z-cis</i> | <i>Z-gauche</i> | <i>E-cis</i> | <i>E-gauche</i> |                                       |
| 2894                    | 3231 (2884)  | 3232 (2885)     | 3236 (2889)  | 3233 (2886)     | CH <sub>3</sub> symmetric stretch     |
| 1652                    | 1897 (1694)  | 1896 (1693)     | 1899 (1696)  | 1897 (1694)     | C=N stretch                           |
| 1385                    | 1561 (1393)  | 1551 (1385)     | 1556 (1389)  | 1542 (1376)     | CH <sub>3</sub> symmetric deformation |
| 1286                    | 1426 (1273)  | 1459 (1302)     | 1432 (1278)  | 1451 (1295)     | CH <sub>2</sub> wag + CNH bend        |
| 878                     | 966 (862)    | 971 (869)       | 952 (850)    | 951 (849)       | C-C=N-H torsion                       |

<sup>a</sup> The frequency data within parentheses are scaled from the calculated data by a factor of 0.8929.

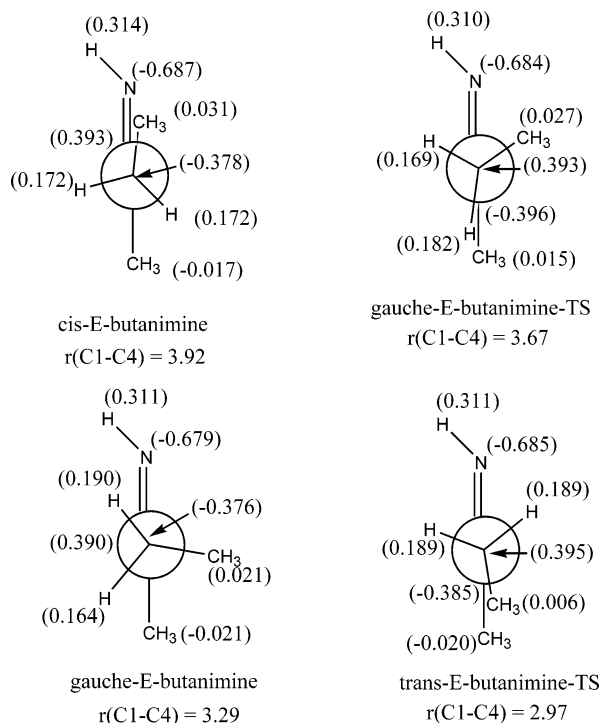
In contrast to 2-butanone, the torsional scan of 2-butanimine from all three models resulted in two minima: a *cis* and *gauche* conformation at  $0.0^\circ$  and  $110.0^\circ$ , respectively, as defined by the  $\text{C}_4\text{-C}_3\text{-C}_2\text{=N}$  dihedral (Figure 4). The energies for these minima as well as their transition states (with a “TS” tag) are listed in Table 1. Our data implies that the *Z/E* configuration of the  $\text{C}=\text{N}$  double bond does have an influence on the relative energy of the minima; the *E*-form (grey, square in Figure 4) is about 0.36 kcal/mol lower in energy than the *Z*-isomer (green, diamond in Figure 4). In Table 3, the calculated infrared frequencies of *Z/E*-butanimine derived from RHF/6-311G\*\* are listed and compared to those from experiments.<sup>19</sup> The scaled calculated  $\text{C}=\text{N}$  stretching frequency at  $1693\text{ cm}^{-1}$  and  $\text{C}_3\text{-C}_2\text{=N-H}$  torsional frequency at  $869\text{ cm}^{-1}$  are close to those observed experimentally ( $1652\text{ cm}^{-1}$  and  $878\text{ cm}^{-1}$ , respectively). The good agreement between the calculated and observed frequencies strongly suggests the accuracy of our calculations. Table 3 also shows that the *E*-configuration has a lower  $\text{C-C=N-H}$  torsional frequency than the *Z*-configuration. This observation further confirms that the configuration of the  $\text{C}=\text{N}$  bond does influence properties of different configurations. Among the four minima of 2-butanimine, our calculations show that the *cis*-(*E*) conformation has the lowest energy. This observation agrees very well with the computational results from other groups.<sup>4,6</sup>

Our results show a deep energy well for 2-butanimine in the region of  $90.0^\circ$  to  $120.0^\circ$ , whereas the corresponding region for 2-butanone gives a relatively flat energy well. The activation energy barrier from the *gauche*-butanimine to the *cis*-butanimine is about 0.6–0.8 kcal/mol and is much larger than that of the *gauche*- to *skew*- and the *skew*- to *cis*-butanone ( $<0.2$  kcal/mol). Such an energy difference contributes to a larger population of the imine *gauche* form and produces the deep well in the butanimine PES between  $60.0^\circ$  and  $90.0^\circ$ .

The Newman projections and natural charge populations for all minima and transition structures (with a “TS” tag) for 2-butanone and its imines are illustrated in Figures 5–7. Note that the charges of the terminal methyl groups are given as the sum of the charges on the methyl carbon atom and the three attached hydrogens. As shown in Figures 5–7, the stabilizing interaction leading to a preference for eclipsing of the methyl group with the carbonyl group or imine group cannot be explained by steric interaction alone. Although it is true that such a steric interaction poses a highly unfavorable hindrance for the transition structures *trans*-butanone and *trans*-*E/Z*-butanimine, such a steric interaction by itself

**Figure 5.** Newman projection structures and partial charges for atoms of 2-butanone.**Figure 6.** Newman projection structures and partial charges for atoms of Z-2-butanimine.

cannot explain why the *gauche-Z/E*-butanimine is an energy minimum, while *gauche-Z/E*-butanimine-TS is a transition structure due to the fact that the distance between two



**Figure 7.** Newman projection structures and partial charges for atoms of *E*-2-butanamine.

terminal carbons is 3.27 Å for the former and 3.68 Å for the latter (Figure 6). However, the bond dipole interaction between the C=O and the  $-\text{CH}_2-\text{CH}_3$  is maximized in *cis*-butanone because of the *anti*-parallel orientation of two bond dipoles. Though the dipole for the C=O group and the C-H bond in *gauche*-butanone is in the parallel orientation, the increasing steric interaction between two terminal methyl groups leads to a higher energy minimum for the *gauche*-butanone. Although there is an angle of roughly 20.0° between the opposite dipoles of the C=O and C-H bonds, the distance between the two terminal methyl groups in the *skew*-butanone is larger, and, thus the steric interaction is minimized. Each of these factors contributes to a minimum energetically more favorable than the *gauche*-butanone. As for the transition structure between the *gauche*- and the *skew*-butanone, the steric interaction between two terminal methyl groups should fall in the range defined by the *skew*- and the *gauche*-butanone. The angle between the bond dipoles along the C=O and  $-\text{CH}_2-\text{CH}_3$  bonds of the transition structure is similar to that of both the *skew*- and *gauche*-butanone minima. The combination of such steric and bond dipole interactions contributes to a transition structure that is structurally and energetically similar to the *skew* and *gauche* conformation and thus results in a flat region in the PES.

Combinations of steric and bond dipole interactions can help rationalize the higher energy of the *trans-Z/E*-butanimine-TS, the lower energy of the *gauche-Z/E*-butanimine, and the even lower energy of the *cis-Z/E*-butanimine. Although the transition structure of the *gauche-Z/E*-butanimine-TS has less steric interactions than the *gauche-Z/E*-butanimine (Figures 6 and 7), the stabilizing bond dipole interaction between the C=N group and the C-H group is less dominant in the transition structure than in the *gauche*-minimum which may be attributed to the larger magnitude

of the resultant bond dipole vector. More importantly, the combined charge of the terminal methyl becomes positive in the transition structure and subsequently introduces a highly destabilizing repulsion between the positively charged hydrogen atom of the imine and the positively charged methyl group. For *gauche-Z/E*-butanimine, the combined charge for the methyl group is negative and a stabilizing attraction between the negatively charged methyl group and the positively charged hydrogen atom on the imine exists. The combination of steric, bond dipole, and charge-charge interactions leads to a much higher energy for the transition structure of *gauche-Z/E*-butanimine-TS relative to the *gauche-Z/E*-butanimine. Therefore, a deep well appears in the region where the *gauche-Z/E*-butanimine conformation lies. The slight stability of the *E*- over the *Z*-configuration of butanimine, as shown in Figures 4–7, can also be explained by this bond dipole interaction. The orientation of the imine hydrogen atom in the *Z/E* form provides for a bond dipole that combines with the imine bond dipole resulting in a different composite dipole. This dipole then interacts with either the  $\text{CH}_2-\text{CH}_3$  or the C-H dipoles. It is obvious that the hydrogen atom of the *E*-form C=N-H is in a better position to produce a more favorable bond dipole and thus contribute to a lower energy minimum.

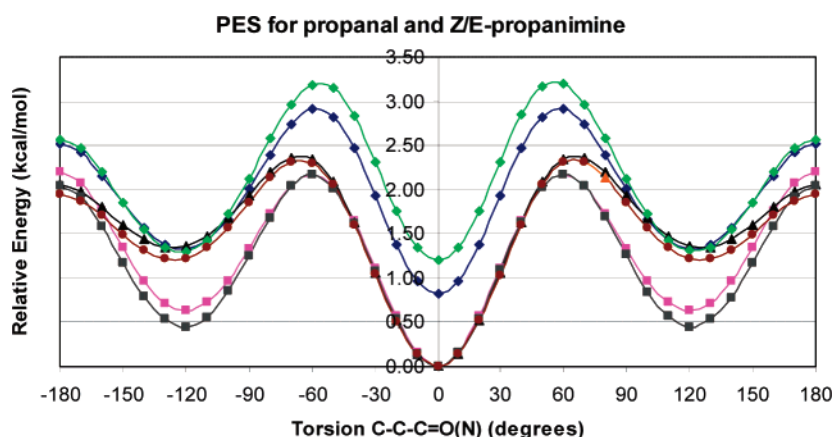
The steric, bond dipole, and charge-charge interactions also aid in understanding propanal and *Z/E*-propanimine conformations. Systematic torsional scans yield two minima (*cis* and *gauche*) for each of the three molecules from three different torsional scan methods (Figures 8 and 4S-6S). The energies for each of the minima and the transition structures are listed in Table 4. The Newman projections and the partial charges of propanal and *Z/E*-propanimine are shown in Figures 8S-10S. A favorable bond dipole interaction and a minimal steric interaction between the  $\text{CH}_2-\text{CH}_3$  and the C(O)-H groups result in the *cis* conformation being the most favorable. Note that the *gauche* minimum has a favorable bond dipole interaction between the C-H and the C=O (or C=N) bonds and that the *gauche* transition structure has an unfavorable electrostatic repulsion resulting from two closely interacting positively charged hydrogen atoms. As a result, the *gauche* transition structure of all three analogues is much higher in energy than the *gauche* minima. As shown in Figure 8 and Table 4, the *gauche* transition structures in propanal and *Z/E*-propanimine, derived from three different model chemistries, have the highest energy relative to their *cis* minima, approximately 2.1–3.2 kcal/mol higher in energy. This leads to a deep well around the *gauche* minimum. Again, the *Z/E* configuration plays an important role in *Z/E*-propanimines. For example, the *cis-Z*-propanimines from all three methods are 0.8–1.2 kcal/mol higher in energy than their *E* counterparts. Other research groups have observed similar PES for propanal.<sup>16,20</sup> The optimized minima were further verified by the frequency calculations. The calculated infrared spectra for propanal and both propanimines show similar frequencies for the C=N stretch and C-C=N-H torsions and demonstrate that these calculations successfully reproduce the data from experiments.<sup>20</sup>

As mentioned earlier, during the interconversion of the carbonyl group of a ketone or an aldehyde to an imine, there

**Table 4.** Ab Initio and DFT Gas-Phase Energies and Characteristics of the Conformers for Propanal and Propanimines Calculated at the 6-311G\*\* Level of Theory

| conformer | name                            | torsion <sup>a</sup> | HF energy (Hartree) | $\Delta E$ (HF) (kcal/mol) | MP2 energy (Hartree) | $\Delta E$ (MP2) (kcal/mol) | B3LYP energy (Hartree) | $\Delta E$ (B3LYP) (kcal/mol) |
|-----------|---------------------------------|----------------------|---------------------|----------------------------|----------------------|-----------------------------|------------------------|-------------------------------|
| 1         | <i>cis</i> -propanal            | 0.0                  | -192.003484947      | 0.00                       | -192.001350873       | 0.00                        | -193.202772379         | 0.00                          |
| 2         | <i>gauche</i> -propanal         | 125.2                | -192.001564766      | 1.20                       | -191.999408387       | 1.22                        | -193.200639060         | 1.34                          |
| 3         | <i>gauche</i> -propanal-TS      | 70.0                 | -191.999847100      | 2.28                       | -191.997661366       | 2.32                        | -193.199007296         | 2.36                          |
| 4         | <i>trans</i> -propanal-TS       | 180.0                | -192.000359676      | 1.96                       | -191.998256835       | 1.94                        | -193.199503493         | 2.05                          |
| 5         | <i>cis-E</i> -propanimine       | 0.0                  | -172.155841845      | 0.00                       | -172.153690609       | 0.00                        | -173.319150184         | 0.00                          |
| 6         | <i>gauche-E</i> -propanimine    | 120.0                | -172.155092712      | 0.47                       | -172.152939119       | 0.47                        | -173.318146664         | 0.63                          |
| 7         | <i>gauche-E</i> -propanimine-TS | 60.0                 | -172.152483243      | 2.11                       | -172.150229734       | 2.17                        | -173.315687045         | 2.17                          |
| 8         | <i>trans-E</i> -propanimine-TS  | 180.0                | -172.152509425      | 2.09                       | -172.150437311       | 2.04                        | -173.315646499         | 2.20                          |
| 9         | <i>cis-Z</i> -propanimine       | 0.0                  | -172.153906329      | 1.21                       | -172.151773840       | 1.20                        | -173.317832209         | 0.83                          |
| 10        | <i>gauche-Z</i> -propanimine    | 121.9                | -172.153743898      | 1.32                       | -172.151586161       | 1.32                        | -173.317037405         | 1.33                          |
| 11        | <i>gauche-Z</i> -propanimine-TS | 60.0                 | -172.150811089      | 3.16                       | -172.148594804       | 3.20                        | -173.314511103         | 2.91                          |
| 12        | <i>trans-Z</i> -propanimine-TS  | 180.0                | -172.151655370      | 2.63                       | -172.149591801       | 2.57                        | -173.315139277         | 2.52                          |
| 13        | Z-propyl-enamine                | 2.6                  | -172.148720083      | 4.47                       | -172.146764767       | 4.35                        | -173.316285220         | 1.80                          |
| 14        | E-propyl-enamine                | 180.0                | -172.146175156      | 6.07                       | -172.144875228       | 5.53                        | -173.314324062         | 3.03                          |

<sup>a</sup> Torsion angle as  $\Phi(C4-C3-C2=O/N)$  (degrees).



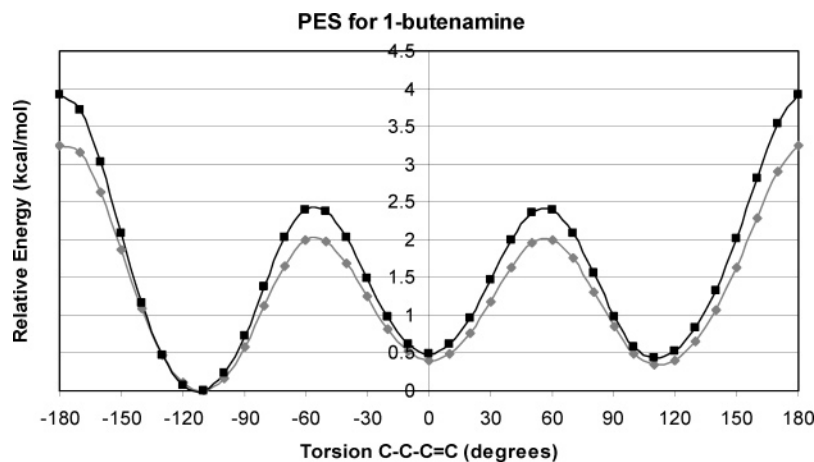
**Figure 8.** Potential energy surface for propanal and its analogues. The color codes are Z-propanimine (B3LYP, blue, diamond); E-propanimine (B3LYP, magenta, square); propanal (B3LYP, black, triangle); Z-propanimine (MP2, green, diamond); E-propanimine (MP2, gray, square); and propanal (MP2, orange, triangle), respectively.

exists some equilibrium between the enamine and imine structures (Scheme 1). Such an equilibrium has been reported for enamino-sulfones<sup>21</sup> and methyl 2-acetamidoacrylates.<sup>22</sup> The imine-enamine tautomerism is proposed as a mechanism of the proline-catalyzed direct aldol reaction between acetone and acetaldehyde.<sup>23</sup> Due to the geometry of the C=C double bond, only optimizations of 2-butyl-Z/E-enamine and Z/E-propyl-enamine were carried out. The energies of the optimized geometries are listed in Tables 1 and 4, respectively. For 1-butyl-enamine, there are three minima along the PES which was discovered from a torsional scan of the C-C-C=C dihedral angle (Figures 9 and 7S). The energies for these three minima are listed in Table 1. The relative energies of the enamines were determined based on the lowest energy of the corresponding isomeric imines (*cis-E*-butanimine). Through this comparison, the enamines are shown to be much higher in energy than their imine isomers, 5.2–6.8 kcal/mol higher than the *cis-E*-butanimine isomer in both RHF and MP2 methods. This finding suggests that the conversion of 2-butanone and propanal will predominantly yield the corresponding imine instead of enamine, which is consistent with experiments.<sup>22</sup> Furthermore, the

calculations also show that the eclipsed configuration, the Z form, has lower energy than the E-configuration (Tables 1 and 4). The energetic difference can be attributed mainly to minimal steric hindrance in the Z-configurations (Figure 11S, in Supporting Information). The stability of 1-butyl-enamine-110 is likely to be the result of less steric interaction between the C=CH<sub>2</sub> and the terminal CH<sub>3</sub> groups and better bond dipole interaction between the C-CH<sub>3</sub> and the C-NH<sub>2</sub> bonds. The small energy difference between the lowest energy minima for 1-butenamine and 2-butenamine suggests that if there is any equilibrium between imines and enamines, these two butenamines may have approximately equal populations.

The solvent effect on the optimized minima was calculated using the IPCM method and is shown in Table 5. For 2-butanone and propanal, the *cis* conformations not skew or *gauche*-conformations are still preferred in chloroform, methanol, and water. For propanimines, *cis-E*-propanimine has the lowest energy in the gas phase, while *gauche-E*-propanimine and *cis-Z*-propanimine are more favorable in more polar solvents, such as methanol and water. Similarly, *cis-Z*-butanimine is more favorable in more polar solvents, while





**Figure 9.** Potential energy surface for 1-butenamine (MP2: black square; B3LYP: gray triangle).

**Table 5.** Ab Initio SCRF-IPCM Energies of the Conformers for 2-Butanone, Propanal, and Their Analogs Calculated at the RHF/6-311G(2d, 2p) Level of Theory

| conformer | name                         | $\Delta E$ (gas)<br>torsion <sup>a</sup> | $\Delta E$ (gas)<br>(kcal/mol) | energy<br>(CHCl <sub>3</sub> , Hartree) | $\Delta E$ (CHCl <sub>3</sub> )<br>(kcal/mol) | energy<br>(MeOH, Hartree) | $\Delta E$ (MeOH)<br>(kcal/mol) | energy<br>(H <sub>2</sub> O, Hartree) | $\Delta E$ (H <sub>2</sub> O)<br>(kcal/mol) |
|-----------|------------------------------|--|--------------------------------|---|---|---------------------------|---------------------------------|---------------------------------------|---|
| 1         | <i>cis</i> -butanone         | 0.0                                      | 0.00                           | -231.0648250                            | 0.00  | -231.0677853              | 0.00                            | -231.0681391                          | 0.00  |
| 2         | <i>skew</i> -butanone        | 85.4                                     | 1.51                           | -231.0628746                            | 1.22  | -231.0658420              | 1.22                            | -231.0661906                          | 1.22  |
| 3         | <i>gauche</i> -butanone      | 120.8                                    | 1.74                           | -231.0622612                            | 1.61  | -231.0653047              | 1.56                            | -231.0656627                          | 1.55  |
| 4         | <i>cis-Z</i> -butanimine     | 0.0                                      | 0.39                           | -211.2115951                            | -0.20   | -211.2141008              | -0.29                           | -211.2143951                          | -0.29                                       |
| 5         | <i>gauche-Z</i> -butanimine  | 104.1                                    | 1.12                           | -211.2104819                            | 0.50  | -211.2130419              | 0.38                            | -211.2133429                          | 0.37  |
| 6         | <i>cis-E</i> -butanimine     | 0.0                                      | 0.00                           | -211.2112795                            | 0.00  | -211.2136461              | 0.00                            | -211.2139325                          | 0.00  |
| 7         | <i>gauche-E</i> -butanimine  | 101.6                                    | 1.06                           | -211.2104131                            | 0.54  | -211.2129336              | 0.45                            | -211.2132328                          | 0.44  |
| 8         | 2-butyl- <i>E</i> -enamine   | -3.1                                     | 5.48                           | -211.2010881                            | 6.40  | -211.2029120              | 6.74                            | -211.2031344                          | 6.78  |
| 9         | 2-butyl- <i>Z</i> -enamine   | 174.3                                    | 6.81                           | -211.1991754                            | 7.60  | -211.2010528              | 7.90                            | -211.2012799                          | 7.94  |
| 10        | 1-butyl-enamine-0            | 177.4                                    | 6.35                           | -211.2004321                            | 6.81  | -211.2024490              | 7.03                            | -211.2026928                          | 7.05  |
| 11        | 1-butyl-enamine-110          | -71.7                                    | 6.27                           | -211.2006143                            | 6.69  | -211.2026835              | 6.88                            | -211.2029344                          | 6.90  |
| 12        | 1-butyl-enamine-n110         | 64.5                                     | 5.81                           | -211.2009515                            | 6.48  | -211.2029256              | 6.73                            | -211.2031691                          | 6.75  |
| 13        | <i>cis</i> -propanal         | 0.0                                      | 0.00                           | -192.0094892                            | 0.00  | -192.0122541              | 0.00                            | -192.0125835                          | 0.00  |
| 14        | <i>gauche</i> -propanal      | 125.2                                    | 1.20                           | -192.0079538                            | 0.96  | -192.0107282              | 0.96                            | -192.0110528                          | 0.96  |
| 15        | <i>cis-E</i> -propanimine    | 0.0                                      | 0.00                           | -172.1600476                            | 0.00  | -172.1622079              | 0.00                            | -172.1624707                          | 0.00  |
| 16        | <i>gauche-E</i> -propanimine | 120.0                                    | 0.47                           | -172.1600621                            | -0.01   | -172.1623677              | -0.10                           | -172.1626419                          | -0.11                                       |
| 17        | <i>cis-Z</i> -propanimine    | 0.0                                      | 1.21                           | -172.1597472                            | 0.19  | -172.1622479              | -0.03                           | -172.1625386                          | -0.04                                       |
| 18        | <i>gauche-Z</i> -propanimine | 121.9                                    | 1.32                           | -172.1596196                            | 0.27  | -172.1621684              | 0.02                            | -172.1624661                          | 0.00  |
| 19        | <i>Z</i> -propyl-enamine     | 2.6                                      | 4.47                           | -172.1522905                            | 4.87  | -172.1542339              | 5.00                            | -172.1544685                          | 5.02  |
| 20        | <i>E</i> -propyl-enamine     | 180.0                                    | 6.07                           | -172.1494489                            | 6.65  | -172.1512576              | 6.87                            | -172.1514772                          | 6.90  |

<sup>a</sup> Torsion angle as  $\Phi(C4-C3-C2=O/N)$  (degrees).

*cis-E*-butanimine is preferred in the gas phase. The IPCM energy of each conformation of 2-butanone, propanal, and their imine analogues shows that the difference between the gas-phase energy and the IPCM energies from three different solvents decreases as the solvent polarity increases. This indicates that those conformations would be more stable in more polar solvents, e.g. methanol and/or water. For enamine, however, the energies become increasingly larger than the gas-phase energy as the solvent polarity increases. Therefore, the energy difference between enamine and the relevant imine becomes larger. Our calculations indicate that the polar solvent would stabilize the imines. These results are in very good agreement with the calculations from Rankin's paper, which reported enamine to be 27.3 kJ/mol lower than the imine in the gas phase but 9.8 kJ/mol higher than the imine in DMSO.<sup>23</sup>

All of our above discussions on imine linkage may help in understanding the conformational preferences for  $-N=CH$

and  $-CH=N$  moieties in NVP analogues, which might suggest a modification of the pharmacophore model.

## Conclusion

Our calculations have provided an explanation for the conformational behavior of 2-butanone, propanal, and their imines and enamine derivatives. Our data suggest that for 2-butanone and propanal, steric and bond dipole interactions are primarily responsible for the conformational preferences of ketones and propanals. Additionally, charge-charge interaction might also play an important role in determining the imine conformations. In contrast, for enamines, the steric interactions are the major energy contributions, although bond dipole interactions might also exert some influence. Our results also suggest that for the conversion of butanone and/or propanal, the imine would be the predominant product, not the enamine. Lastly, our results are in good agreement with previous experimental and computational results and

should provide a better understanding of the ketone (aldehyde), imine and enamine interconversion. Moreover, the calculations strongly suggest conformational preferences in NVP analogues where the *trans*-ethylene connecting group might be replaced by imine linkages. Whether or not this has conformational implications for ChAT inhibitor design remains to be answered and will be addressed in future work.

**Supporting Information Available:** PES maps for the conformational scan for seven molecules in Figure 2(A to G) (Figures 1S-7S) and Newman projection structures and partial charges for atoms of propanal (Figure 8S), *Z*-propanimine (Figure 9S), *E*-propanimine (Figure 10S), and *Z/E*-butenamine (Figure 11S). This material is available free of charge via the Internet at <http://pubs.acs.org>.

### References

- (1) Smith, S. O.; Pardeon, T. A.; Mulder, P. P. J.; Cuny, B.; Lugtenburg, J.; Mathies, R. *Biochemistry* **1983**, *22*, 6141–6148.
- (2) Vdovenko, S. I.; Gerus, I. I.; Wójcik, J. *J. Phys. Org. Chem.* **2001**, *14*, 533–542.
- (3) Lee, M.; Kim, H.; Rhee, H.; Choo, J. *Bull. Korean Chem. Soc.* **2003**, *24*, 205–208.
- (4) Egawa, T.; Konaka, S. *J. Phys. Chem. A* **2001**, *105*, 2085–2090.
- (5) Durig, J. R.; Feng, F. S.; Wang, A.; Phan, H. V. *Can. J. Chem.* **1991**, *69*, 1827–1844.
- (6) Wang, Y.; Poirier, R. A. *J. Phys. Chem. A* **1997**, *101*, 907–912.
- (7) Gray, A. P.; Platz, R. D.; Henderson, T. R.; Chang, T. C. P.; Takahashi, K.; Dretchen, K. L. *J. Med. Chem.* **1988**, *31*, 807–814.
- (8) Cavallito, C. J.; Yun, H. S.; Edwards, M. L.; Foldes, F. F. *J. Med. Chem.* **1971**, *14*, 130–133.
- (9) Kontoyianni, M.; McGaughey, G. B.; Stewart, E. L.; Cavallito, C. J.; Bowen, J. P. *J. Med. Chem.* **1994**, *37*, 3128–3131.
- (10) Chandrasekaran, V.; McGaughey, G. B.; Cavallito, C. J.; Bowen, J. P. *J. Mol. Graphics Modell.* **2004**, *23*, 69–76.
- (11) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 2003.
- (12) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.
- (13) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (14) Lee, C.; Wang, Y.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (15) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098–16104.
- (16) Wiberg, K. B.; Martin, E. *J. Am. Soc. Chem.* **1985**, *107*, 5035–5041.
- (17) Bowen, J. P.; Pathiaseril, A.; Profeta, P. S. Jr.; Allinger, N. L. *J. Org. Chem.* **1987**, *52*, 5162–5166.
- (18) Abe, M.; Kuchitsu, K.; Shimanouchi, T. *J. Mol. Struct.* **1969**, *4*, 245–253.
- (19) Egawa, T.; Ito, M.; Konaka, S. *J. Mol. Struct.* **2001**, *560*, 337–344.
- (20) Randell, J.; Hardy, J. A.; Cox, A. P. *J. Chem. Soc., Faraday Trans. 2* **1998**, *84*, 1199–1212.
- (21) Forzato, C.; Felluga, F.; Gombac, V.; Nitti, P.; Pitacco, G.; Valentin, E. *Arkivoc* **2003** (xiv), 210–224.
- (22) Cativiela, C.; García, J. I.; Mayoral, J. A.; Salvatella, L. *J. Mol. Struct. (THEOCHEM)* **1996**, *368*, 57–66.
- (23) Rankin, K. N.; Gauld, J. W.; Boyd, R. J. *J. Phys. Chem. A* **2002**, *106*, 5155–5159.

CT049890P

## Theoretical Surface Spectroscopy of NO on the Pt(111) Surface with the DAM (Dipped Adcluster Model) and the SAC-CI (Symmetry-Adapted-Cluster Configuration-Interaction) Method

Hiroshi Nakatsuji,<sup>\*,†,‡</sup> Norihiko Matsumune,<sup>†</sup> and Kei Kuramoto<sup>‡</sup>

*Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Nishikyo, Kyoto 615-8510, Japan, and Fukui Institute for Fundamental Chemistry, Kyoto University, Sakyo, Kyoto 606-8103, Japan*

Received September 20, 2004

**Abstract:** Theoretical surface spectroscopy for NO on the Pt(111) surface is carried out and combined with the experimentally known facts to elucidate the structures, absorption sites, absorption energies and K-shell binding energies of NO adsorbates on the surface. The electronic structures were studied by using the dipped adcluster model (DAM) for chemisorptions on a metal surface proposed previously and the symmetry-adapted-cluster configuration-interaction (SAC-CI) method, which is an established method for studying molecular spectroscopy. The natures of the two different adsorption species suggested experimentally have clearly been identified based on the studies on the geometries, vibrational frequencies, adsorption energies, and the N and O K-shell binding energies. The PES (potential energy surface) of NO on the metal surface was also calculated. The most stable adsorption species was on the fcc or the hcp hollow site, and the on-top one was less stable. The 2-fold bridge site did not have a minimum on the PES and therefore was only transient. The inter NO interactions at higher densities were shown to be rather weak. We examined the cluster model (CM) vs the DAM as a model of the surface adsorption on a metal surface. The CM was shown to be unable to describe the adsorption of NO on a metal surface, demonstrating the importance of the electron transfer between the NO and Pt surfaces included in the DAM. The DAM and the SAC-CI methods proved to be a useful tool for studying the nature, electronic structure, and the spectroscopic properties of the adsorbates on a metal surface.

### 1. Introduction

Nitric oxides (NO<sub>x</sub>) are the substances included in exhaust gases from automobiles, factories and industries. The nature of the adsorptions of NO<sub>x</sub> on metal surfaces has been widely investigated by using spectroscopic techniques<sup>1–15</sup> such as electron energy loss spectroscopy (EELS), infrared reflection absorption spectroscopy (IRAS), scanning tunneling microscopy (STM), low-energy electron diffraction (LEED), elec-

tron spectroscopy for chemical analysis (ESCA), and ultraviolet(U) or X-ray(X) photoemission spectroscopy (PS). These experiments have given the information about the electronic structures and the bonding of the adsorbates on the metal surfaces. We investigate here the nature of the adsorptions of the simplest nitric oxide, NO on the Pt(111) surface.

There are many examples of molecular and dissociative adsorptions of NO on metal surfaces.<sup>16</sup> On a Pt surface, it is generally agreed that NO adsorbs molecularly at low temperature with its N atom toward the surface. Due to EELS<sup>1,2,6,14</sup> and IRAS<sup>3,7,9,11</sup> experiments, an N–O stretching band lay at the range of 1476–1516 cm<sup>-1</sup> at low coverage,

\* Corresponding author e-mail: hiroshi@sbchem.kyoto-u.ac.jp.

† Department of Synthetic Chemistry and Biological Chemistry, Kyoto University.

‡ Fukui Institute for Fundamental Chemistry, Kyoto University.

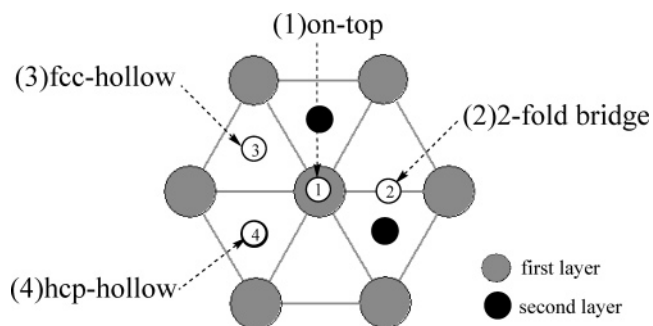
and as the coverage increased, a new band began to develop around 1700–1725  $\text{cm}^{-1}$  and at the same time, the first band decreased: the first peak nearby 1500  $\text{cm}^{-1}$  was converted into the new peak nearby 1700  $\text{cm}^{-1}$ . Hayden<sup>3</sup> considered that NO converts from the bridge site to the on-top site at near-saturation coverage. Sexton et al.<sup>2</sup> also reported the existence of the two types of adsorbates at low and high coverage. Various adsorption models were suggested by many experimental techniques such as LEED<sup>1,10,15</sup> and STM.<sup>13,14</sup> Materer et al.<sup>10</sup> reported that an fcc hollow site adsorption was observed by the  $p(2 \times 2)$  LEED pattern. Matsumoto et al.<sup>15</sup> also performed the dynamical LEED analysis and reported the adsorption sites of NO to be at the on-top site and at the fcc- and hcp-hollow sites.

The nature of the interaction between NO and a metal surface is interesting. NO is a radical in a gas phase and has one electron in the  $\pi^*$ -MO. When adsorbed on a metal surface, the NO stretching frequency shows a red shift from that of the gas phase: it is 1876  $\text{cm}^{-1}$  in a gas phase and around 1500  $\text{cm}^{-1}$  on Pt(111) at low coverage and is around 1600  $\text{cm}^{-1}$  on Pd(111) at the same condition. This indicates that the electron transfer occurs on a metal surface from the metal surface to the adsorbate NO. Between the donation and the back-donation interactions between NO and a metal surface, the back-donation interaction is more important.

In the previous theoretical studies, Ge et al.<sup>17</sup> reported that the 3-fold fcc hollow site is most stable and the 3-fold hcp hollow site is next by the density functional theory (DFT) calculations using a cluster model. Koper et al.<sup>18</sup> compared the adsorption energies of the on-top and hollow sites by the DFT calculations and explained that the hollow site is more stable. Aizawa et al.<sup>19</sup> calculated the adsorption energies for the three types of coverage using the periodic boundary model. All of these reports show that the 3-fold hollow site is the most stable.

When NO is adsorbed on a Pt surface, electrons are withdrawn from the metal to NO. When we use the cluster model (CM), the effects of the free electrons of the bulk metal are not well described. We therefore use the dipped adcluster model (DAM)<sup>20,21</sup> proposed by one of the authors to describe the chemisorptions and catalytic reactions on a metal surface. It naturally describes the effect of the electron transfer between adsorbates and the metal surface and the image force effects characteristic to the metal surface. We study the structures, vibrations, absorption sites, and surface PES (potential energy surface) for the NO on the Pt(111) surface with the DAM combined with the density functional theory (DFT). We also study the possibility of the interadsorbate interactions expected at high coverage adsorptions.

The K-shell binding energies of NO adsorbates are observed experimentally and show some interesting features. Recently, we have shown that the SAC (symmetry adapted cluster)<sup>22</sup>/SAC-CI (configuration interaction)<sup>23</sup> method describes well the K-shell ionizations of various molecules in excellent agreement with the experimental values.<sup>24</sup> We apply here the SAC-CI method combined with the DAM to elucidate the origin of the different K-shell binding energies observed for NO at different coverage situations.



**Figure 1.** Model and four adsorption sites on the Pt(111) surface: (1) on-top, (2) 2-fold bridge, (3) 3-fold fcc-hollow, and (4) 3-fold hcp-hollow.

In all the aspects of the present study of the NO adsorbed on the Pt(111) surface, the importance of the DAM has been stressed. We then compare the DAM with the CM directly for the adsorption energy and some other properties of NO on the Pt(111) surface and then enter into the detailed discussions on the nature of the chemisorptions.

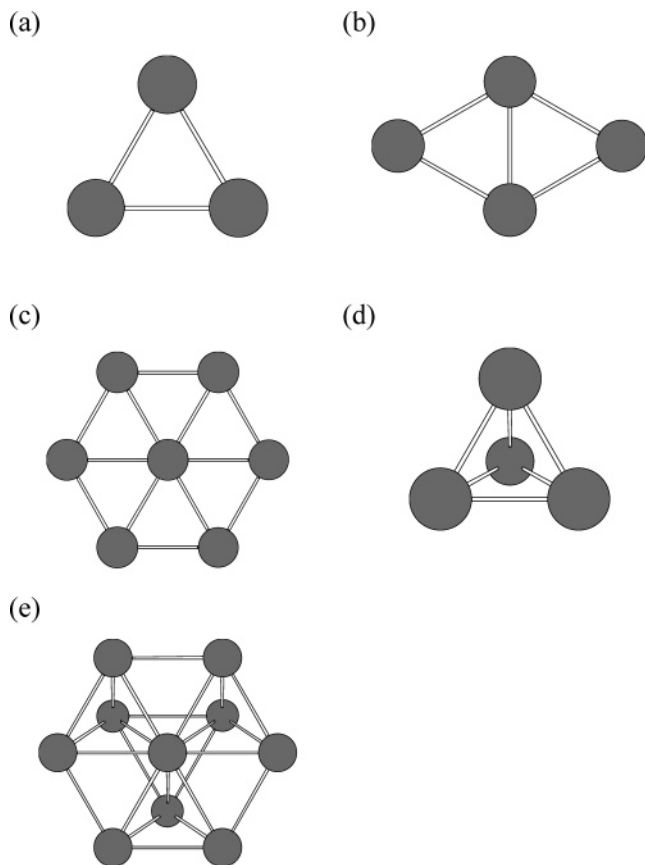
## 2. Computational Details

In this study, we used the DFT method with the B3LYP<sup>25,26</sup> potential and combined it with the dipped adcluster model (DAM).<sup>20,21</sup> The DAM was used to include the effect of the electron exchange between the adsorbates and the bulk metal, which is considered to be very important on the metal surface. The cluster model cannot describe such effect. We also included the energy correction due to the image force<sup>27</sup> on the metal surface. We used the highest spin-coupling model, and so one-electron transfer from the bulk metal to the adcluster was assumed.<sup>20,21</sup>

Four adsorption sites, (1) on-top, (2) 2-fold bridge, (3) 3-fold fcc-hollow, and (4) 3-fold hcp-hollow, shown in Figure 1, were investigated using the experimental geometry of the (111) surface. We used five different sizes of clusters shown in Figure 2 to mimic the (111) surface: (a) Pt<sub>3</sub>(3,0), (b) Pt<sub>4</sub>(4,0), and (c) Pt<sub>7</sub>(7,0) clusters, which contain respectively, three, four, and seven atoms only in the first layer, (d) Pt<sub>4</sub>(3,1) and (e) Pt<sub>10</sub>(7,3) clusters, which contain double layers.

The geometry of the adsorbed molecule on the surface was optimized using the DAM+DFT method. Materer et al. reported that the geometry relaxation of the metal–metal distance on the Pt(111) surface was very small; the change of the Pt–Pt bond length was less than 0.1 Å. Then, we used the fixed Pt–Pt bond length at its bulk lattice value of 2.77 Å.<sup>28</sup> The convergence criterion of the geometry optimization on the first derivative of the total energy was 0.0003 hartree/bohr (default value). After the optimizations, the vibrational analyses were carried out to evaluate the harmonic frequencies and the IR intensities.

All calculations were performed using the Gaussian 03 software package.<sup>29</sup> For the DFT calculations, the basis set for Pt atom was (3s3p3d)/[3s3p2d], and Xe core was replaced by the effective core potential.<sup>30</sup> For oxygen and nitrogen, we used the standard LANL2DZ plus diffuse and polarization functions.<sup>31</sup> On the other hand, for the calculations of the core electron binding energy of the adsorbed NO using the



**Figure 2.** Cluster models used in this work: (a)  $\text{Pt}_3(3,0)$ , (b)  $\text{Pt}_4(4,0)$ , (c)  $\text{Pt}_7(7,0)$ , (d)  $\text{Pt}_4(3,1)$ , and (e)  $\text{Pt}_{10}(7,3)$ .

SAC-CI general- $R$  method, we used the basis set well tested in the correlation calculations. We used the Ahlrich's core-valence triple- $\zeta$  basis<sup>32</sup> for N and O atoms and Christiansen's ECP basis for Pt.<sup>33</sup> The reference functions for the SAC-CI calculations were the Hartree–Fock wave function chosen for the DAM.

### 3. DAM vs CM

Table 1 shows a comparison between the DAM and CM for the adsorption energy and the charges of NO on the Pt(111) surface. The adsorption geometries were optimized independently for both CM and DAM. Referring to the adsorption energy, we see that the CM does not describe a stable adsorption of NO on the Pt(111) surface. The adsorption energy calculated by the CM is negative for all the adsorption sites, while the DAM gives positive (stable) adsorption energy for all the adsorption sites. This clearly shows the advantage of the DAM over the CM, since experimentally NO is preferably adsorbed on the Pt(111) surface. Furthermore, the stability order is also in contradiction with that of DAM and the experiment (as seen later): the 3-fold site is more repulsive than the on-top site. Detailed discussions on the relative stabilities, etc., will be given below.

We show in Table 2 the size dependence of the adsorption energy in the DAM and CM. Not only for the  $\text{Pt}_{10}(7,3)$  model given in Table 1, the CM gives negative adsorption energy in various cases, while the DAM gives always the positive adsorption energy. The cluster size dependence in the DAM will be discussed in the next section.

We see that the DAM describes larger electron transfer to NO than the CM. The NO charge calculated with the DAM is  $-0.4$  for the 3-fold site,  $-0.1$  for the 2-fold site, and  $-0.14$  for the on-top site, while that with the CM is  $-0.24$ ,  $+0.03$ , and  $0.0$ , respectively. This electron transfer from the Pt surface to the  $\pi^*$  orbital of NO seems to be important for describing the interaction between the chemisorbed NO on a metal surface.

We show in Table 1 the calculated stretching vibration frequency of NO ( $\nu_{\text{NO}}$ ) and the IR intensity both with the DAM and CM. As we will see from discussions given below, the DAM gives reasonable vibration frequency for both 3-fold and on-top geometries, but the frequency calculated with the CM is not well comparable with the experimental values.

The IR peak intensity of the higher frequency ( $1700\text{--}1725\text{ cm}^{-1}$ ) is larger than lower one ( $1476\text{--}1516\text{ cm}^{-1}$ ).<sup>16</sup> This means that the IR intensity is larger at the on-top site than at the 3-fold site. This is reproduced with the DAM but not with the CM.

The above comparative calculations show that the electron transfer from the bulk metal to the adsorbates is important for the adsorption of NO on the Pt surface. Then we use the DAM to study the chemisorption and spectroscopy of NO on the Pt(111) surface.

### 4. Structure, Vibrational Frequency, and Adsorption Energy – Cluster Size Dependence

The structure of NO on the Pt(111) surface was observed by spectroscopic techniques such as NEXAFS.<sup>12</sup> Two types of NO adsorbates with the bond lengths of 1.16 and 1.24 Å<sup>12</sup> were reported on the Pt(111) surface. Croci et al. reported experimentally the adsorption energy of NO on the Pt(111) surface to be 1.29 eV.<sup>8</sup> Only one adsorption energy was reported, and this energy seems to correspond to the most stable adsorption species of NO on the Pt(111) surface. This adsorption energy on the Pt(111) surface is smaller than those on other indices of the Pt surface reported by Yeo, Vattuone, and King, namely 1.98 eV on Pt(100)<sup>34</sup> and 1.68 eV on Pt(110),<sup>35</sup> representing the natures of the interactions between the surfaces and the adsorbates.

We examined here first the cluster size dependence of the DAM on the optimized geometry, vibrational frequency, and the adsorption energy by using five different types of clusters shown in Figure 2. The result of the calculation was summarized in Table 3. We see first small model size dependence on the optimized NO length. For the on-top site, the calculated length was 1.19, 1.19, 1.20, 1.17, and 1.20 Å for the  $\text{Pt}_3(3,0)$ ,  $\text{Pt}_4(4,0)$ ,  $\text{Pt}_7(7,0)$ ,  $\text{Pt}_4(3,1)$ , and  $\text{Pt}_{10}(7,3)$  models, respectively. This result indicates that the experimental bond length of 1.16 Å<sup>12</sup> would correspond to the on-top type NO. Similarly, we also examined the NO length for other adsorption sites, 2-fold bridge, 3-fold fcc hollow, and 3-fold hcp hollow sites. The calculated NO length for the 2-fold bridge site was in the range of 1.20–1.21 Å and that of the 3-fold hollow fcc and hcp sites was in the range of 1.22–1.24 Å. The experimentally proposed value for the bridge or hollow site was 1.24 Å.<sup>12</sup> From the present result,

**Table 1.** Adsorption Energies ( $E_{\text{ads}}$ ) and Charges of  $\text{Pt}_{10}(7,3)$  Model: Comparison of the DAM and the Neutral Cluster Model

| site model                             |        | on-top    |       | 2-fold |       | 3-fold fcc |       | 3-fold hcp |       |
|--|--------|-----------|-------|--------|-------|------------|-------|------------|-------|
|  |        | DAM       | CM    | DAM    | CM    | DAM        | CM    | DAM        | CM    |
| $E_{\text{ads}}$ (eV)                  | calc   | 0.37      | -1.41 | 0.50   | -1.31 | 1.25       | -1.53 | 1.26       | -1.60 |
|  | exptl. |           |       |        |       |            | 1.29  |            |       |
| $\nu_{\text{NO}}$ ( $\text{cm}^{-1}$ ) | calc   | 1706      | 1752  | 1626   | 1689  | 1482       | 1536  | 1475       | 1484  |
|  | exptl. | 1700–1725 |       |        |       | 1476–1516  |       |            |       |
| IR intensity                           |        | 639       | 432   | 522    | 496   | 396        | 362   | 497        | 492   |
| charge on N                            |        | -0.12     | -0.03 | -0.09  | -0.01 | -0.21      | -0.11 | -0.20      | -0.11 |
| charge on O                            |        | -0.02     | 0.03  | -0.01  | 0.04  | -0.19      | -0.13 | -0.22      | -0.12 |
| charge on $\text{Pt}^a$                |        | 0.57      | 1.00  | 0.21   | 0.37  | 0.10       | 0.18  | 0.17       | 0.12  |

<sup>a</sup> The averaged value for the nearest Pt atoms from NO adsorbate.

**Table 2.** Size Dependencies in the DAM and the CM on the Adsorption Energy ( $E_{\text{ads}}$ ) in eV Using Five Clusters Shown in Figure 2

| site model            | on-top |       | 2-fold |       | 3-fold fcc |       | 3-fold hcp |       |
|-----------------------|--------|-------|--------|-------|------------|-------|------------|-------|
|                       | DAM    | CM    | DAM    | CM    | DAM        | CM    | DAM        | CM    |
| $\text{Pt}_3(3,0)$    | 2.09   | 0.22  | 1.49   | -0.50 | 1.89       | -0.21 |            |       |
| $\text{Pt}_4(4,0)$    | 1.90   | -0.57 | 1.59   | -1.19 | 2.71       | -0.15 |            |       |
| $\text{Pt}_4(3,1)$    | 1.24   | -0.73 | 1.23   | -1.03 |            |       | 1.43       | -0.83 |
| $\text{Pt}_7(7,0)$    | 1.70   | 0.60  | 2.02   | -0.02 | 2.94       | 0.24  |            |       |
| $\text{Pt}_{10}(7,3)$ | 0.37   | -1.41 | 0.50   | -1.31 | 1.25       | -1.53 | 1.26       | -1.60 |
| exptl. <sup>a</sup>   |        |       |        |       |            | 1.29  |            |       |

<sup>a</sup> Reference 8.

**Table 3.** Cluster Size Dependences of the DAM on the Calculated NO Length, Frequency, and the Adsorption Energy ( $E_{\text{ads}}$ ) Using Five Clusters Shown in Figure 2

| site              | model                 | $R_{\text{N-O}}$ (Å) | $\nu_{\text{N-O}}$ ( $\text{cm}^{-1}$ ) | $E_{\text{ads}}$ (eV) |
|-------------------|-----------------------|----------------------|---|-----------------------|
| on-top            | $\text{Pt}_3(3,0)$    | 1.19                 | 1661                                    | 2.09                  |
|                   | $\text{Pt}_4(4,0)$    | 1.19                 | 1690                                    | 1.90                  |
|                   | $\text{Pt}_4(3,1)$    | 1.20                 | 1666                                    | 1.24                  |
|                   | $\text{Pt}_7(7,0)$    | 1.17                 | 1762                                    | 1.70                  |
|                   | $\text{Pt}_{10}(7,3)$ | 1.18                 | 1706                                    | 0.37                  |
| 2-fold            | $\text{Pt}_3(3,0)$    | 1.21                 | 1503                                    | 1.49                  |
|                   | $\text{Pt}_4(4,0)$    | 1.21                 | 1594                                    | 1.59                  |
|                   | $\text{Pt}_4(3,1)$    | 1.21                 | 1572                                    | 1.23                  |
|                   | $\text{Pt}_7(7,0)$    | 1.20                 | 1648                                    | 2.02                  |
|                   | $\text{Pt}_{10}(7,3)$ | 1.20                 | 1626                                    | 0.50                  |
| 3-fold fcc-hollow | $\text{Pt}_3(3,0)$    | 1.24                 | 1406                                    | 1.89                  |
|                   | $\text{Pt}_4(4,0)$    | 1.24                 | 1442                                    | 2.71                  |
|                   | $\text{Pt}_7(7,0)$    | 1.22                 | 1503                                    | 2.94                  |
|                   | $\text{Pt}_{10}(7,3)$ | 1.23                 | 1482                                    | 1.25                  |
| 3-fold hcp-hollow | $\text{Pt}_4(3,1)$    | 1.23                 | 1453                                    | 1.43                  |
|                   | $\text{Pt}_{10}(7,3)$ | 1.23                 | 1475                                    | 1.26                  |
| exptl.            |                       | 1.16 <sup>b</sup>    | 1700–1725 <sup>a</sup>                  | 1.29 <sup>c</sup>     |
|                   |                       | 1.24 <sup>b</sup>    | 1476–1516 <sup>a</sup>                  |                       |

<sup>a</sup> Reference 16. <sup>b</sup> Reference 12. <sup>c</sup> Reference 8.

we assigned the observed bond length of 1.24 Å to be due to the 3-fold hollow site. Thus, the small cluster size dependence of the calculated NO length and the reliability of the present DAM model have enabled us to make a unique assignment of the adsorption site.

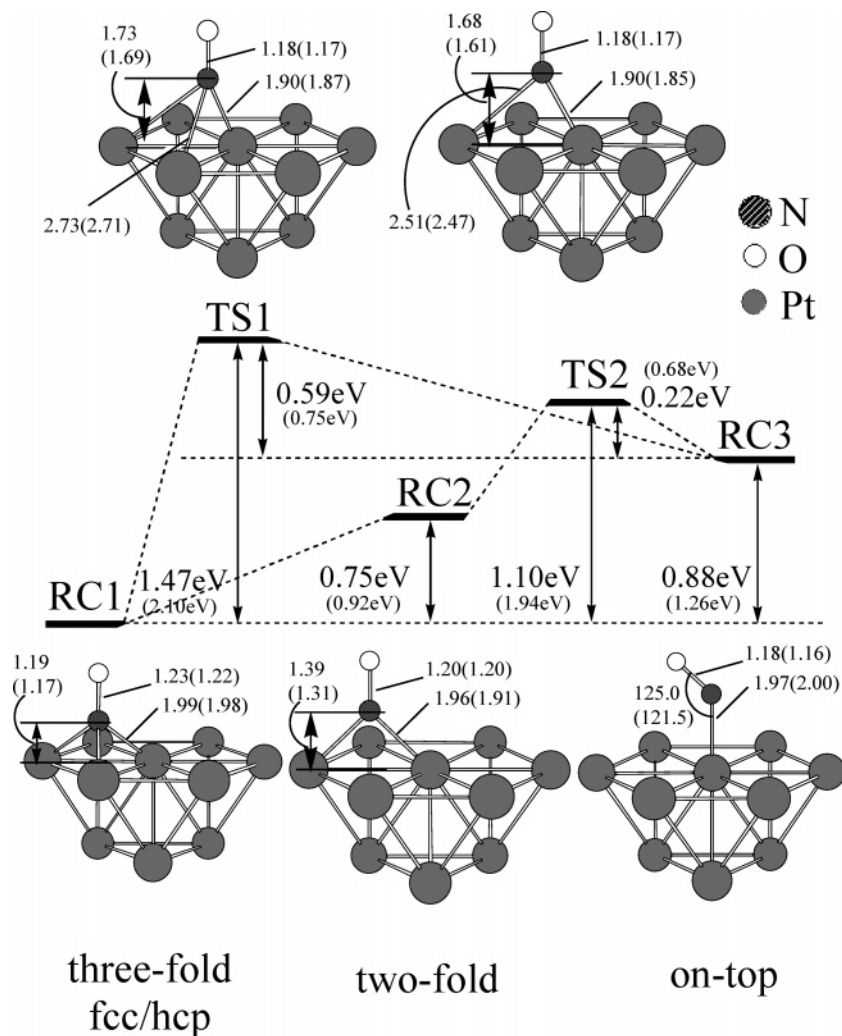
We next calculated the NO vibration frequency at the optimized geometry of the different size DAM. The calculated N–O stretching frequency for the on-top site was in the range of 1661–1762  $\text{cm}^{-1}$ , that for the 2-fold site was 1503–1648  $\text{cm}^{-1}$ , and that for the 3-fold hollow site was

1406–1503  $\text{cm}^{-1}$ . Two different stretching frequencies were reported by EELS and IRAS,<sup>16</sup> which would correspond to different absorption species of NO on the Pt(111) surface. The lower frequency was 1476–1516  $\text{cm}^{-1}$  and the higher one was 1700–1725  $\text{cm}^{-1}$ . Compared to our theoretical results, the stretching band of the lower frequency is assigned to the NO on the 3-fold hollow site and that of higher frequency is assigned to the on-top one.

Finally, we examine the adsorption energy ( $E_{\text{ads}}$ ), which is defined as the stabilization energy of NO adsorbed on the surface relative to the gas-phase;  $E_{\text{ads}} = E(\text{Pt}_n) + E(\text{NO}) - E(\text{Pt}_n\text{NO})$ . The calculated adsorption energy shows a relatively large model size dependence and a small size DAM tends to overestimate the adsorption energy. This is because the small cluster has a relatively large number of unsaturated bonds that are stabilized by forming  $\text{Pt}_n\text{NO}$ . The result of the largest DAM using  $\text{Pt}_{10}(7,3)$  cluster is 1.25 eV for the 3-fold fcc-hollow site and 1.26 eV for the hcp-hollow site, in good agreement with the experimental value, 1.29 eV. This indicates that the experimental value is due to the adsorption on the 3-fold hollow site. The comparison with the  $\text{Pt}_7(7,0)$  model for the 3-fold fcc-hollow site clearly shows an important role of the second Pt layer in the  $\text{Pt}_{10}(7,3)$  cluster.

## 5. Geometry and Potential Energy Diagram of NO on the Surface

We examined the potential energy diagram of NO on the Pt(111) surface using the  $\text{Pt}_{10}(7,3)$  model and showed the result in Figure 3. The values in parentheses are the results obtained with the  $\text{Pt}_7(7,0)$  model. The potential energy ( $E_p$ ) is defined similarly to the adsorption energy as  $E_p = E(\text{Pt}_n) + E(\text{NO}) - E(\text{Pt}_n\text{NO})$  and is a function of the position of NO on the surface. We also show in Figure 3 the optimized geometries (RC1–RC3) and the transition states (TS1, TS2). The energy difference between the 3-fold fcc and hcp hollow sites was very small, only 0.01 eV, and therefore, we use in the following discussions the result of the 3-fold fcc hollow site model. For the migrating from the on-top site to the 3-fold site, the transition state (TS1) was found and the barrier was 0.59 eV. We also found another TS (TS2) from the on-top site to the 2-fold site and the barrier was 0.22 eV. The on-top site (RC1) and the 3-fold hollow site (RC3) had true minima, but there was no true minimum at the 2-fold bridge site: it was a transient species that is not stable. This



**Figure 3.** Potential energy diagram and the geometries of the NO adsorbates on the Pt<sub>10</sub>(7,3) DAM. Values in parentheses are due to Pt<sub>7</sub>(7,0) DAM.

result indicates that, at low coverage, the NO adsorbed on the on-top site migrates to the most stable 3-fold site via the 2-fold site. The NO at the 3-fold (fcc or hcp) hollow site is strongly bound with a relatively high barrier of about 0.75 eV for the migration to another 3-fold (hcp or fcc) hollow site and therefore would not easily move to the other sites at low temperature.

Next, we examined the Pt–N–O angle at the on-top site using the Pt<sub>10</sub>(7,3) DAM. At each Pt–N–O angle, the Pt–N and N–O lengths were optimized. The adsorption energy of the linear form (180 degree) was  $-2.24$  eV and of the bent form (121.7 degree) was  $1.49$  eV, which was the local minimum. Namely, at the on-top site, NO takes a bent form with the Pt–N–O angle of around 120 degrees. Similar calculations were also done at the 3-fold fcc hollow site. There, the adsorption energy of the linear form was  $2.81$  eV and that of the bent form was  $-0.20$  eV. So, a linear form is preferred at this site. In summary, the two adsorption species are linear NO at the 3-fold fcc hollow site and the bent NO at the on-top site.

In Table 4, our theoretical results are compared with those of other experimental and theoretical studies on the on-top and the 3-fold hollow sites. When we use the Pt<sub>10</sub>(7,3) model, there was no apparent difference between the fcc and hcp

**Table 4.** Comparison of the Present and Past Results on the Adsorption Energy of NO on the Pt Surfaces

|                   | present<br>DAM+B3LYP<br>for Pt(111) <sup>a</sup> | other<br>theoretical<br>works         | exptl.   |
|-------------------|--|---------------------------------------|--|
| on-top            | 0.37   | 0.64 <sup>c</sup> , 1.61 <sup>b</sup> |  |
| 3-fold fcc-hollow | 1.25   | 1.75 <sup>c</sup> , 2.09 <sup>b</sup> | 1.29 [Pt(111)] <sup>d</sup>                                |
| hcp-hollow        | 1.25   | 1.92 <sup>b</sup>                     | 1.68 [Pt(100)] <sup>e</sup><br>1.98 [Pt(110)] <sup>f</sup> |

<sup>a</sup> Pt<sub>10</sub>(7,3) DAM. <sup>b</sup> Reference 19. <sup>c</sup> Reference 17. <sup>d</sup> Reference 8. <sup>e</sup> Reference 33. <sup>f</sup> Reference 32.

sites. The experimental adsorption energy on Pt(111) was  $1.29$  eV,<sup>8</sup> while the calculated adsorption energy was  $0.37$  eV for the on-top site and  $1.25$  eV for the 3-fold hollow site. So, the measured adsorbate would be NO on the 3-fold hollow site. The adsorption energies of the other index Pt(100)<sup>34</sup> and Pt(110)<sup>35</sup> were reported to be  $1.98$  and  $1.68$  eV, respectively. These differences reflect the nature of the interaction between NO and the surface.

## 6. Vibrational Analysis

In a gas phase, the N–O length is  $1.15077$  Å and the vibrational frequency is  $1876$  cm<sup>-1</sup>.<sup>36</sup> The corresponding

**Table 5.** Bond Length and Stretching Frequency of NO on the Pt(111) Surface and Atomic Charge of Each Atom

| adsorption site | $R_{\text{N-O}}$ (Å) | $\nu_{\text{N-O}}$ ( $\text{cm}^{-1}$ ) | atomic charge |       |       |
|-----------------|----------------------|---|---------------|-------|-------|
|                 |                      |   | N             | O     |       |
| on-top          | 1.18                 | 1706                                    | -0.12         | -0.02 |       |
| 3-fold          | fcc-hollow           | 1.23                                    | 1482          | -0.19 | -0.21 |
|                 | hcp-hollow           | 1.23                                    | 1475          | -0.20 | -0.22 |
|                 | 1.16 <sup>b</sup>    | 1476–1516 <sup>a</sup>                  |               |       |       |
| exptl.          | 1.24 <sup>b</sup>    | 1700–1725 <sup>a</sup>                  |               |       |       |

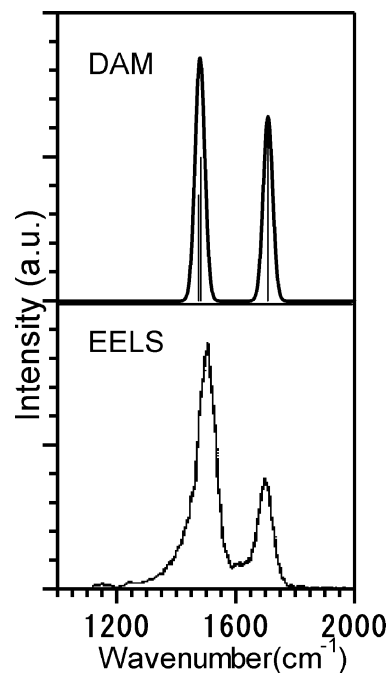
<sup>a</sup> Reference 16. <sup>b</sup> Reference 12.

values on the Pt(111) surface calculated with the Pt<sub>10</sub>(7,3) DAM are summarized in Table 5 together with the Mulliken charges on N and O atom. In the EELS experiment, the peak developed in 1476–1516  $\text{cm}^{-1}$  appears first at low coverage and at higher coverage a second new peak appears at around 1700  $\text{cm}^{-1}$  and the first peak is replaced with the second peak as the coverage is further increased. The present result shown in Table 5 clearly assigns the first peak as being due to the NO at the stable 3-fold hollow site and the second peak as being due to the on-top NO. Further, the experimental bond lengths of 1.24 and 1.16 Å due to Esch et al.<sup>12</sup> are also clearly assigned as being due to the NO's at the 3-fold hollow site and at the on-top site, respectively.

The dipole moment of the adsorbed NO on the 3-fold site was smaller than that of the on-top site, so that the vibrational peak intensity of the 3-fold site would be smaller than that of the on-top site, in agreement with the experimental spectrum.<sup>2</sup> We show in Figure 4 the theoretical and experimental vibrational spectra<sup>2</sup> of NO/Pt(111). The ratio of the adsorption molecules was assumed to be 1:1:1 (on-top:3-fold fcc:3-fold hcp) in the theoretical spectrum. The theoretical spectrum agrees well with the experimental one taken at the 0.50 ML coverage.

## 7. K-Shell Ionization Potential

Core electron binding energies (CEBEs) and the chemical shifts of the adsorbates are important information of analytical chemistry for surface adsorbates. We can calculate the CEBEs of various molecules in high accuracy using the SAC/SAC-CI general-*R* method.<sup>24,37–39</sup> The orbital reorganization due to the core–electron ionization is described by the general-*R* method together with the important correlation effect. Here, we apply this established method to the calculations of the N 1s and O 1s CEBEs of the NO molecule adsorbed on the Pt(111) surface. In this calculation we used Ahlrich's core-valence triple- $\zeta$  basis<sup>32</sup> set for N and O atoms



**Figure 4.** Theoretical vs experimental vibrational spectrum of NO/Pt(111). The experimental spectrum was cited from the 0.25 L exposure data of NO by Gland et al. (Figure 1 of ref 2).

and Christiansen's ECP for Pt atom.<sup>33</sup> These basis sets are well examined for the SAC-CI calculations of the K-shell ionization potential.<sup>24,37–39</sup>

The results of the calculations were shown in Table 6. The CEBEs of a free NO molecule were calculated to be 411.1 and 543.1 eV for N and O atoms, respectively, which were in good agreement with the experimental values, 410.8 and 543.6 eV,<sup>40</sup> respectively. For the NO adsorbed at the on-top, fcc-hollow, and hcp-hollow sites, the calculated O 1s CEBEs were 538.5, 537.5, and 537.2 eV, respectively. The experimental CEBEs are estimated to be 538.9 and 537.0 eV<sup>4</sup> using the work function  $\phi = 6.42$  eV,<sup>41</sup> and so the band of higher energy was assigned to the on-top site and the lower one to the fcc/hcp-hollow site. The theoretical energy difference from the free molecule to the adsorbate is 5.1, 6.1, and 6.4 eV, respectively, for the on-top, fcc-hollow, and hcp-hollow sites, which were proportional to the atomic charges, -0.02 (on-top), -0.21 (fcc-hollow), and -0.22 (hcp-hollow), which were also shown in Table 5.

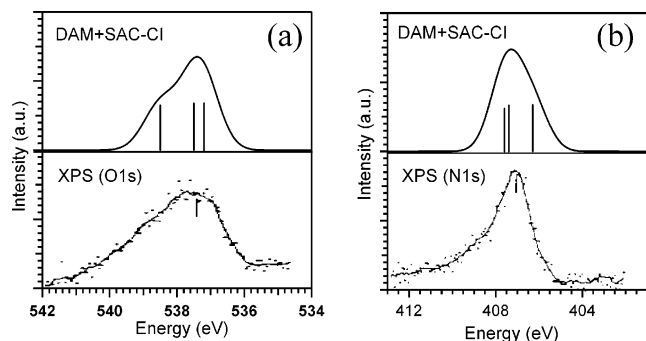
The theoretical values of N 1s CEBEs were 406.3, 407.4, and 407.6 eV for the on-top, fcc- and, hcp-hollow sites,

**Table 6.** Core Electron Binding Energies and Atomic Charges of Free and Adsorbed NO

| state         | N 1s               |            |             | O 1s               |            |             |
|---------------|--------------------|------------|-------------|--------------------|------------|-------------|
|               | exptl. (eV)        | DAM+SAC-CI |             | exptl. (eV)        | DAM+SAC-CI |             |
|               |                    | IP (eV)    | charge on N |                    | IP (eV)    | charge on O |
| on-top        | 407.0 <sup>a</sup> | 406.3      | 0.07        | 538.9 <sup>a</sup> | 538.5      | -0.24       |
| 2-fold bridge |                    | 405.8      | 0.02        |                    | 538.7      | -0.18       |
| 3-fold        | fcc-hollow         | 407.4      | 0.09        | 537.0 <sup>a</sup> | 537.5      |             |
|               | hcp-hollow         | 407.6      | 0.10        |                    | 537.2      | -0.38       |
| gas phase     | 410.8 <sup>b</sup> | 411.1      | 0.15        | 543.6 <sup>b</sup> | 543.2      | -0.15       |

<sup>a</sup> Reference 4. <sup>b</sup> Reference 38.





**Figure 5.** DAM+SAC-Cl theoretical vs experimental photoemission spectrum of NO/Pt(111): (a) O 1s hole spectrum and (b) N 1s hole spectrum. The experimental spectra was observed by Kiskinova et al. (Figure 3 of ref 4).

respectively, and the experimental value was only 407.0 eV<sup>4</sup>, which was near the average value of our results, 407.1 eV. We show the theoretical and experimental CEBE spectra of NO/Pt(111) in Figure 5. The ratio of the adsorbates was assumed to be 1:1:1 (on-top: 3-fold fcc: 3-fold hcp) in calculating the theoretical spectrum, which corresponds to the experimental condition of the higher or saturated coverage.

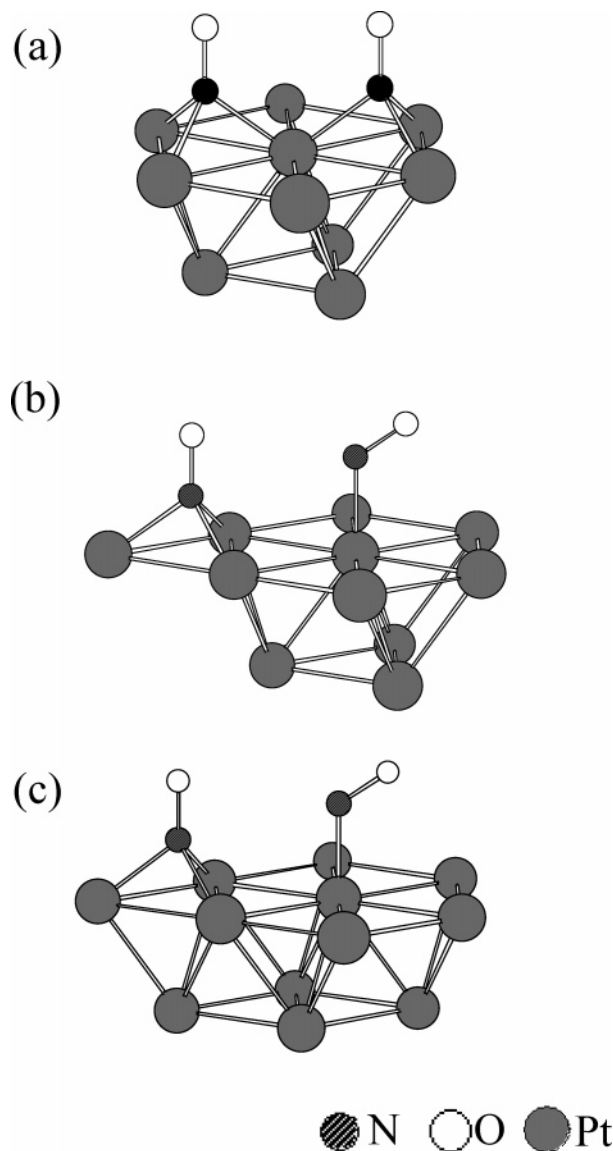
## 8. NO–NO Interaction at High Coverage

We examine here the NO–NO interaction that is expected in high coverage. For the on-top adsorption, the former model (one NO molecule on the surface) may not have been appropriate since it appears only at a high coverage situation. Therefore, we examine Pt<sub>x</sub>(NO)<sub>2</sub> ( $x=10,11,12$ ) DAM shown in Figure 6 for the high coverage model. When two NO molecules are put in the  $p(2 \times 2)$  unit cell at the coverage of 0.50 ML, we assume that one is the on-top site and another is the 3-fold site (model (b) and (c)), referring to the experimental LEED pattern<sup>15</sup> and STM image.<sup>13,14</sup> We also examined the interaction between the NO's on the 3-fold fcc and hcp sites (model (a)). The calculation method and the basis sets are the same as before.

Table 7 shows the results, which should be compared with the single NO results shown in Table 5. There were little interactions between the on-top site and the 3-fold fcc or hcp hollow sites. The NO distances were completely the same, and the frequencies were shifted only slightly. On the other hand, a small interaction was found between the fcc and the hcp hollow sites and the theoretical spectrum was split into two peaks, 1430 and 1485 cm<sup>-1</sup>. We assume that the vibrational spectrum in the region of 1400–1500 cm<sup>-1</sup> becomes broad at high coverage by such an interaction between the adsorbates.

## 9. Conclusion

The DAM has been applied to the NO on the Pt(111) surface. The spectroscopic properties such as the adsorption energy, the geometry and the vibrational frequency of the NO adsorbates were theoretically calculated and compared with the experimentally available data. We also calculated the potential energy surface of the NO adsorbate on the Pt(111) surface. We could show that only the 3-fold hollow site (fcc and hcp) and the on-top site are stable. We gave a definite



**Figure 6.** Optimized geometries of NO of Pt<sub>x</sub>(NO)<sub>2</sub> [ $x = 10–12$ ]: (a) fcc and hcp-hollow, (b) on-top and fcc-hollow, and (c) on-top and hcp-hollow.

**Table 7.** Bond Length and Frequency of Adsorbed NO and Atomic Charge of Each Atom: (a) the fcc and hcp-hollow Model, (b) the on-top and fcc-hollow Model, and (c) the on-top and hcp-hollow Model

| model                       | R <sub>N–O</sub><br>(Å) | ν <sub>N–O</sub><br>(cm <sup>-1</sup> ) | atomic charge |       |
|-----------------------------|-------------------------|---|---------------|-------|
|                             |                         |   | N             | O     |
| (a) fcc-hollow + hcp-hollow | 1.23                    | 1430                                    | -0.12         | -0.25 |
|                             | 1.23                    | 1485                                    | -0.17         | -0.25 |
| (b) on-top + fcc-hollow     | 1.18                    | 1708                                    | -0.08         | -0.01 |
|                             | 1.23                    | 1457                                    | -0.27         | -0.20 |
| (c) on-top + hcp-hollow     | 1.18                    | 1707                                    | -0.04         | -0.01 |
|                             | 1.23                    | 1468                                    | -0.23         | -0.20 |
| exptl.                      | 1.16 <sup>a</sup>       | 1700–1725 <sup>b</sup>                  |               |       |
|                             | 1.24 <sup>a</sup>       | 1476–1516 <sup>b</sup>                  |               |       |

<sup>a</sup> Reference 12. <sup>b</sup> Reference 16.

assignment of the observed spectra and clarified the nature and the electronic structure of the adsorbate on the Pt(111) surface.

The dependence on the cluster size used in the DAM was examined for the adsorption energy, the geometry and the vibrational frequency. The dependence was relatively large for the adsorption energy but small for the geometry and the vibrational frequency. Anyway, the largest model examined has given the results that agree best with the experimental values. Generally speaking, it is rather difficult to eliminate the edge effects of the model, but comparing with the experimental values, this effect seems to be minimized by using the largest adcluster model.

The inner-core electron binding energy of NO on the Pt surface was calculated by the SAC-CI general-R method using the MO's obtained with the DAM. Likewise the gas-phase case, the calculated results reproduced well the existing experimental values and predicted the values for nonexisting cases. Since the ESCA spectroscopy is frequently done for surface adsorbates, the SAC-CI + DAM method will provide a powerful method for identifying the surface species.

A comparison of the cluster model (CM) result with the DAM result has clearly shown the importance of the DAM for the adsorbates on a metal surface. These results may be considered as proof of the validity of the DAM.

In conclusion, by combining the experimental and theoretical surface spectroscopies, we can understand the chemistry of the surface adsorbates more deeply than doing only by one of them.

**Acknowledgment.** This study was supported by a Grant for Creative Scientific Research from the Ministry of Education, Science, Culture, and Sports of Japan and ITBL (IT-based laboratory project) of Japan.

### References

- (1) Ibach, H.; Lehwald, S. *Surf. Sci.* **1978**, *76*, 1–12.
- (2) Gland, J. L.; Sexton, B. A. *Surf. Sci.* **1980**, *94*, 355–368.
- (3) Hayden, B. E. *Surf. Sci.* **1983**, *131*, 419–432.
- (4) Kiskinova, M.; Pirug, G.; Bonzel, H. P. *Surf. Sci.* **1984**, *136*, 285–295.
- (5) Seebauer, E. G.; Kong, A. C. F.; Schmidt, L. D. *Surf. Sci.* **1986**, *176*, 134–156.
- (6) Bartram, M. E.; Koel, B. E. *Surf. Sci.* **1989**, *219*, 467–489.
- (7) Agrawal, V. K.; Trenary, M. *Surf. Sci.* **1991**, *259*, 116–128.
- (8) Croci, M.; Felix, C.; Vandoni, G.; Harbich, W.; Monot, R. *Surf. Sci.* **1994**, *307–309*, 460–464.
- (9) Song, M. B.; Suguri, M.; Fukutani, K.; Komori, F.; Murata, Y. *Appl. Surf. Sci.* **1994**, *79/80*, 25–33.
- (10) Materer, N.; Barbieri, A.; Gardin, D.; Starke, U.; Batteas, J. D.; Van Hove, M. A.; Somorjai, G. A. *Surf. Sci.* **1994**, *303*, 319–332.
- (11) Yoshinobu, J.; Kawai, M. *Chem. Lett.* **1995**, *1995*, 605–606.
- (12) Esch, F.; Greber, Th.; Kennou, S.; Siokou, A.; Ladas, S. *Catal. Lett.* **1996**, *38*, 165–170.
- (13) Matsumoto, M.; Tatsumi, N.; Fukutani, K.; Okano, T.; Yamada, T.; Miyake, K.; Hate, K.; Shigekawa, H. *J. Vac. Sci. Technol. A* **1999**, *17(4)*, 1577–1580.
- (14) Matsumoto, M.; Fukutani, K.; Okano, T.; Miyake, K.; Shigekawa, H.; Kato, H.; Okuyama, H.; Kawai, M. *Surf. Sci.* **2000**, *454–456*, 101–105.
- (15) Matsumoto, M.; Tatsumi, N.; Fukutani, K.; Okano, T. *Surf. Sci.* **2002**, *513*, 485–500.
- (16) Garin, F. *Appl. Catal. A* **2001**, *222*, 183–219.
- (17) Ge, Q.; King, D. A. *Chem. Phys. Lett.* **1998**, *285*, 15–20.
- (18) Koper, M. T. M.; van Santen, R. A.; Wasileski, S. A.; Weaver, M. J. *J. Chem. Phys.* **2000**, *113(10)*, 4392–4407.
- (19) Aizawa, H.; Morikawa, Y.; Tsuneyuki, S.; Fukutani, K.; Ohno, T. *Surf. Sci.* **2002**, *514*, 394–403.
- (20) Nakatsuji, H. *J. Chem. Phys.* **1987**, *87(8)*, 4995–5001.
- (21) Nakatsuji, H. *Prog. Surf. Sci.* **1997**, *54(1)*, 1–68.
- (22) Nakatsuji, H.; Hirao, K. *J. Chem. Phys.* **1978**, *68*, 2053–2065.
- (23) (a) Nakatsuji, H. *Chem. Phys. Lett.* **1978**, *59*, 362–364. (b) Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 329–333, 334–342.
- (24) (a) Kuramoto, K.; Ehara, M.; Nakatsuji, H. submitted for publication. (b) Kuramoto, K.; Ehara, M.; Nakatsuji, H.; Kitajima, M.; Tanaka, H.; Fanis, A. D.; Tanemori, Y.; Ueda, K. *J. Electron Spectrosc. Relat. Phenom.* **2004**, in press.
- (25) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (26) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (27) Nakatsuji, H.; Nakai, H.; Fukunishi, Y. *J. Chem. Phys.* **1991**, *95(1)*, 640–647.
- (28) Sutton, L. E. *Tables of Interatomic Distances and Configuration in Molecules and Ions*; Royal Society of Chemistry: London, 1965.
- (29) Gaussian 03, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh, PA, 2003.
- (30) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270–283.
- (31) Check, C. E.; Faust, T. O.; Bailey, J. M.; Wright, B. J.; Gilbert, T. M.; Sunderlin, L. S. *J. Phys. Chem. A* **2001**, *105*, 8111–8116.
- (32) Schafer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
- (33) Ross, R. B.; Ermler, W. C.; Christiansen, P. A. *J. Chem. Phys.* **1990**, *93*, 6654–6670.

- (34) Yeo, Y. Y.; Vattuone, L.; King, D. A. *J. Chem. Phys.* **1997**, *106*, 1990–1996.
- (35) Yeo, Y. Y.; Vattuone, L.; King, D. A. *J. Chem. Phys.* **1996**, *104*, 3810–3821.
- (36) Huber, K. P.; Herzberg, G. *Molecular Spectra and Molecular Structure, IV. Constant of Diatomic Molecules*; Van Nostrand: Princeton, NJ, 1979.
- (37) Nakatsuji, H. *Chem. Phys. Lett.* **1991**, *177*, 331–337.
- (38) Nakatsuji, H. *J. Chem. Phys.* **1985**, *83*, 713–722.
- (39) Nakatsuji, H. *J. Chem. Phys.* **1985**, *83*, 5743–5748.
- (40) Bakke, A. A.; Chen, A. W.; Jolly, W. L. *J. Electron Spectrosc.* **1980**, *20*, 333–366.
- (41) Wandelt, K. *Chemistry and Physics of Solid Surface VII*; Springer: 1990; p 289.

CT049938Z

# JCTC

Journal of Chemical Theory and Computation

## Pseudospectral Local Second-Order Møller–Plesset Methods for Computation of Hydrogen Bonding Energies of Molecular Pairs

George A. Kaminski\*

*Department of Chemistry, Central Michigan University, Mt. Pleasant, Michigan 48859*

Jon R. Maple, Robert B. Murphy, and Dale A. Braden

*Schrödinger, Inc., 120 West 45th Street, Tower 45, 32nd Floor,  
New York, New York 10036*

Richard A. Friesner

*Department of Chemistry, Columbia University, New York, New York 10027*

Received November 19, 2004

**Abstract:** We present a methodology for computing the binding energy of molecular dimers based on extrapolation of pseudospectral local second-order Møller–Plesset (MP2), or PS-LMP2, energies to the basis set limit. The extrapolation protocol is based on carrying out PS-LMP2 calculations with the Dunning cc-pVTZ (-f) and cc-pVQZ (-g) basis sets and then using a simple two-parameter function to compute the final basis set limit results. The function is parametrized to ultralarge basis set MP2 calculations for 5 molecular pairs taken from the literature and then tested by calculating results for a set of formamide dimers for which such calculations have also been carried out. The results agree to within ca. 0.2 kcal/mol with the conventional MP2 large basis set calculations. A specialized, but relatively simple, protocol is described for eliminating noise due to overcompleteness of the basis set. Timing results are presented for the LMP2 calculations, and comparisons are made with the LMP2 methodology of the QChem program. CPU time required by each of the methods scales as  $N^3$ , where  $N$  is the number of the basis functions, with the PS-LMP2 approach displaying a 2- to 3-fold advantage in the prefactor. We also discuss one set of test cases for which the PS-LMP2 results disagree with those obtained from an alternative type of MP2 calculation, *N*-methyl acetamide (NMA) dimers, and show that the results for liquid-state simulations using polarizable parameters derived by fitting to the PS-LMP2 binding energies appear to produce better results when compared with experimental data. The convergence issues associated with the alternative MP2 formulation remain to be investigated.

### I. Introduction

The computation of hydrogen bonding energies between pairs of molecular fragments is an important problem in theoretical chemistry. Hydrogen bonding strengths are crucial for a wide range of chemical processes in both the gas phase and in

condensed phase. Our own interests lie particularly in the areas of molecular mechanics force field development and structure based drug design; for these endeavors, errors on the order of 0.5 kcal/mol can be significant. At the same time, pharmaceutical compounds span a very broad range of chemical functionality, the great majority of which is not accessible to experimental characterization of the inter-

\* Corresponding author e-mail: kamin1ga@cmich.edu.

molecular interactions in a straightforward fashion. To address such problems, a computational approach is required that can be applied to hundreds or thousands of molecular pairs, many of significant size, while at the same time routinely providing accuracy to better than 0.5 kcal/mol. A description of such a method is the objective of the present paper.

Over the past 5 years, it has become apparent that second-order Moller–Plesset perturbation theory (MP2) provides answers of the quality specified above (at least for hydrogen bonded pairs of organic compounds, on which we shall focus our attention in the present paper) provided that the basis set limit of the calculation can be obtained.<sup>1</sup> The problem is that very large basis sets are required to achieve convergence, and this in turn necessitates the expenditure of large amounts of computation time if standard algorithms for MP2 calculations are employed. For treating pair hydrogen bond interactions between molecular fragments of the size relevant to force field development, a key determinant of efficiency is scaling with basis set size; in particular, the use of distance cutoffs is less relevant than, for example, if one is interested in methods for treating a linear chain of carbon atoms. At the same time, a high degree of precision in relative energies is required, which may pose problems for approximate methods, for example those relying on basis set expansions of product charge distributions.

The localized pseudospectral methods (PS-LMP2)<sup>2</sup> that we have developed and implemented in the Jaguar suite of ab initio electronic structure programs<sup>3</sup> are well suited to computations of this type. These calculations can be shown to scale as  $N_b^2$ , where  $N_b$  is the number of basis functions per atom and no cutoffs are assumed. The overall scaling of the method with system size for relatively small molecules is on the order of  $N_{\text{atom}}^3$ . These scaling behaviors imply that tractable CPU times can be obtained for the accurate hydrogen bonding calculations that are the focus of the present paper.

With a highly efficient method in hand for carrying out large basis set LMP2 calculations, one can employ extrapolation techniques to reach the basis set limit, many of which are now reported in the literature.<sup>5</sup> We use here a simple approach based on the Dunning correlation consistent basis sets<sup>6</sup> and a two-parameter fit to an exponential functional form. The extrapolation method is “trained” on a series of dimers for which converged, large basis set results have been reported by Tsuzuki and co-workers<sup>1a</sup> and then tested via calculation of the binding energies of a number of formamide dimers. The agreement for the test set, on the order of a few tenths of a kcal/mol in absolute binding energy for all five complexes, suggests that a reliable approach to the computation of hydrogen bonding energetics has been produced.

Over the past several years, we have extensively utilized these extrapolated LMP2 calculations to develop a polarizable molecular mechanics force field that is capable of predicting condensed phase properties accurately without the use of adjustable parameters. The development of parameters for a number of small molecules, and subsequent prediction of liquid state heats of vaporization and densities, is discussed in ref 8. Heats of vaporization are generally predicted with

an accuracy of  $\sim 0.5$  kcal/mol; this quantity is very sensitive to the hydrogen bonding energy (which controls the fitting of the short range van der Waals parameters of the polar atoms in the molecule), and hence the results provide a further confirmation of the accuracy of our LMP2 extrapolation protocol (as well as a validation of the proposition that extrapolated MP2 provides accurate hydrogen bonding energies). A case of particular interest is the NMA molecule, where our results differ substantially from those in ref 1b, computed by conventional quantum chemical methods, despite the fact that for the five formamide dimers mentioned above, the results of the two methods agree to within a few tenths of a kcal/mol. We discuss our construction of polarizable liquid-state models using both sets of hydrogen bonding energies and show that our LMP2 results yield excellent agreement with experimental data, whereas reliance on the ref 1b results produces significant overbinding in the condensed phase simulations.

The remainder of the paper is organized as follows. In Section II, we describe our computational methods, briefly reviewing the relevant aspects of the PS-LMP2 approach (which has been presented in detail in previous work) and discussing the extrapolation protocol. There are technical issues associated with overcompleteness of the basis set, which we have addressed and successfully surmounted, and these are presented in some detail. Section III presents results for the various systems that we have investigated, demonstrating that high accuracy can be achieved with PS-LMP2 methods. Timing results are also presented and compared with those obtained from the LMP2 methodology in the Q-Chem program.<sup>7</sup> Finally, in the conclusion, we summarize our results and briefly discuss directions of future research.

## II. Computational Methods

**A. Pseudospectral LMP2 Technique.** We have described PS methods for carrying out local second-order Moller Plesset (LMP2) perturbation theory calculations in previous papers.<sup>2,4</sup> Briefly, the use of a numerical grid at a key stage in the evaluation of two electron integrals allows the usual four index transform to be eliminated and replaced by an algorithmic step which, when localized orbitals are used, has superior scaling properties with system and basis set size. This key step involves the assembly of the exchange integrals via a sum over grid points ( $g$ )

$$K_{ij}^{pq} = \sum_g Q_i(g) R_p(g) A_{jq}(g) \quad (1)$$

over local occupied orbitals ( $ij$ ) and local virtual orbitals.  $Q_i(g)$  is the least-squares fitting operator for orbital  $i$ ,  $R_p(g)$  is the physical space representation of virtual orbital  $p$ , and  $A_{jq}(g)$  is the three center, one electron orbital over molecular orbitals  $j, q$ .

The scaling of this calculation is  $n_{\text{occ}}^2 N_b^2 N_g$ , where  $n_{\text{occ}}$  is the number of occupied orbitals in the molecule,  $N_b$  is the number of orbitals in the local space, and  $N_g$  is the number of grid points on the mesh.<sup>2</sup> Since  $N_b$  is in principle invariant to the size of the molecule, the scaling of this step with system size (roughly represented by the number of atoms  $N_a$ ) is proportional to  $N_a^3$ . Of equal importance for the present

work is the scaling with the size of the atomic basis set. For ultralarge basis sets, it is not necessary to continue to increase the size of the grid beyond a certain point. Then, the computational effort scales as  $N_b^2$  with the number of atomic basis functions in the localized space (in essence the size of the atomic basis sets). In contrast, the initial step of the traditional four index transform scales as  $N_b^4$ . Thus, we would expect the advantage of the PS-LMP2 methods, as compared with the standard canonical MP2 calculations, to increase as the basis set size is increased. There are other methods that have been developed to reduce the computational expense of MP2 methods (e.g., refs 9 and 10); the question of which approach yields the best performance for a given problem, in terms of speed and accuracy, requires direct comparison of performance for the same systems. We present below a comparison of this type with the LMP2 methodology in the Q-Chem program, where quantitative (although not qualitative) advantage in efficiency is demonstrated. However, the results presented below can be used by others if desired to make such comparisons themselves.

**B. Extrapolation Methods.** Initial studies suggested that localized perturbation methods might converge to the basis set limit more quickly than canonical MP2 methods (even when augmented by counterpoise corrections) due to the reduction of basis set superposition error arising from truncation of the virtual space for each electron pair.<sup>11</sup> However, over the past several years, it has been demonstrated via ultralarge basis set computations that extrapolation to the basis set limit is necessary, even if localized methods are employed, if accurate intermolecular pair energetic properties are to be computed.<sup>1,12</sup> On the other hand, these same studies also show that, for typical hydrogen bonded dimer structures, reaching the MP2 basis set limit provides results within a few tenths of a kcal/mol in accuracy; corrections from higher order correlation effects (e.g. at the CCSD (T) level) are very small due to cancellation of several different terms (note that this conclusion does not necessarily hold for other strong nonbonded interactions, such as aromatic ring  $\pi$  stacking interactions). Hence, our goal here is to develop an extrapolation approach based on PS-LMP2 methods, which agrees with the benchmark studies quoted above, while at the same time provides significant reductions in computational effort, as compared to alternative protocols for achieving this level of accuracy.

After considerable experimentation, we have settled upon the following simple methodology. PS-LMP2 calculations are carried out for the pair and for the separated monomers using the cc-pVTZ (-f) and cc-pVQZ (-g) basis sets. Counterpoise BSSE corrections are determined for the Hartree–Fock part of the energy only. This approach is motivated by the idea that LMP2 based approaches, of the type that we have implemented, have a relatively low level of BSSE present, due to the use of localized virtual spaces. Other researchers suggested extrapolation schemes, including the one with the two-parameter exponential functional form to be used for the basis set extrapolation parameters.<sup>13</sup> We obtain the final energy by using our data with the following extrapolation scheme, i.e.

$$E_{bind} = [a \cdot E_{bind}(cc - pvqz) - b \cdot E_{bind}(cc - pvtz)] / (a - b) \quad (2)$$

where  $E_{bind}(cc - pvqz)$  and  $E_{bind}(cc - pvtz)$  are the cc-pVQZ (-g) and cc-pVTZ (-f) counterpoise corrected binding energies, respectively. The values of the parameters  $a$  and  $b$  are determined by fitting to a “training set” of molecular dimers, for which benchmark results have been obtained using conventional MP2 methods and ultralarge basis sets. The extrapolation protocol is then tested by taking a new set of dimers and again comparing with high level calculations, this time without adjustment of  $a$  and  $b$ .

In the results presented below, we optimize the dimer geometries at the LMP2/cc-pVTZ (-f) level of theory, unless explicitly noted otherwise. Such optimizations are relatively expensive, and it is of interest to ask whether the use of smaller basis sets in the LMP2 optimization, or even optimization using DFT methods, would yield equivalent results. We do not consider this question in the present paper; it will be explored in future publications.

We note that the use of the specific basis sets suggested above (which are modified from their normal composition by the elimination of g functions from the QZ basis and f functions from the TZ basis) constitutes in and of itself a substantial increase in efficiency as compared to the use of larger basis sets for extrapolation of hydrogen bond energies. For the particular property of interest here, the higher angular momentum basis functions are not as important as when one is for example trying to compute covalent bond energies, although we have found that it was not possible to simply eliminate f functions completely from the calculations. Further reduction in the basis set size may be possible; we have not investigated this in the present paper.

**C. Problems Due to Basis Set Overcompleteness.** While Gaussian basis sets have many compelling advantages for electronic structure calculations, they suffer from one principal disadvantage, nonorthogonality of the basis set, leading in the limit of a large number of basis functions to overcompleteness and the associated numerical instabilities. This problem is particularly an issue when using approximate numerical methods, such as pseudospectral methods, to compute integrals. If one is not careful, small errors in the integrals can translate into large variation in the energy of the highly oscillatory orbitals generated from linear combinations of strongly overlapping basis functions. While we have spent a great deal of effort to ensure that the PS method correctly handles problems of this type, the use in the present calculations of QZ basis sets and the desire to achieve precision in the calculations on the order of a few tenths of a kcal/mol, renders this endeavor more challenging than in our previously published works. Note that this same problem could conceivably arise in other schemes involving numerical approximations to the two electron integrals, such as resolution of the identity (RI-MP2) approaches.

The simplest and most effective way of dealing with overcompleteness is singular value decomposition. The overlap matrix is diagonalized, and eigenvectors corresponding to small eigenvalues are eliminated from the basis set. This methodology directly removes the highly oscillatory

**Table 1.** Comparison of Five Small Molecule ab Initio Dimerization Energies in kcal/mol

| dimer                                | MP2 <sup>a</sup> | CCSD(T)<br>limit <sup>a</sup> | LMP2 <sup>b</sup> | LMP2/MP2<br>geometry <sup>c</sup> |
|--------------------------------------|------------------|-------------------------------|-------------------|-----------------------------------|
| H <sub>2</sub> O – MeOH              | 4.99             | 4.90                          | 4.95              | 4.93                              |
| H <sub>2</sub> O – Me <sub>2</sub> O | 5.70             | 5.51                          | 5.68              | 5.48                              |
| H <sub>2</sub> O – H <sub>2</sub> CO | 5.21             | 5.17                          | 5.21              | 5.10                              |
| MeOH – MeOH                          | 5.58             | 5.45                          | 5.79              | 5.46                              |
| HCOOH – HCOOH                        | 13.79            | 13.93                         | 16.18             | 13.90                             |

<sup>a</sup> Reference 1a. <sup>b</sup> Plus extrapolation. This work. Optimized geometry. <sup>c</sup> Plus extrapolation. This work. Geometries from reference 1a.

functions and hence avoids errors in the energy of the type discussed above. Standard thresholds for the smallest allowable eigenvalue have been developed and successfully used for many years.

This method is usable for the present problem, but there is one caveat; the number of functions deleted from the monomer and dimer calculations must be consistent, if an accurate energy difference is to be computed. This is a straightforward protocol to implement. First, a cutoff for the monomer eigenvalues is specified; after some experimentation on a wide range of dimers, we have settled on a value of  $3.5 \times 10^{-4}$ , which gives robust results for all of the cases tested to date. Then, the number of orbitals used in the monomer calculations is saved, and the number employed in the dimer calculations is forced to be equal to the sum of the numbers for the monomers. One can imagine that in some cases, a more elaborate method, in which correspondence between the orbitals deleted in the dimer and monomer calculations (in terms of spatial composition, energy, etc.) was checked, might be necessary; so far, however, the simple approach described above appears to be adequate. It is important to emphasize, however, that failure to use an appropriate cutoff, or allowing inconsistent numbers of orbitals between the dimer and monomer calculations, can lead to errors as large as 1–2 kcal/mol in the binding energy. The present approach, on the other hand, contributes deviations on the order of 0.1–0.2 kcal/mol, a crucial order of magnitude improvement.

### III. Results

**A. Training Set.** Tsuzuki et al. have studied five molecular dimers using conventional MP2 with counterpoise correction methods for basis sets up to cc-pV5Z and also at the CCSD (T) level, along with large basis set MP2 level geometry optimizations.<sup>1a</sup> Their final results for binding energies in the MP2 basis set limit and with the inclusion of higher levels of electron correlation are summarized in Table 1. As discussed above, the MP2 results are within a few tenths of a kcal/mol of CCSD (T) results. As the CCSD (T) methodology is at present clearly too computationally intensive for our ultimate objectives (e.g. generation of hundreds or thousands of molecular pair energies so as to design a molecular mechanics force field with broad coverage of chemical space), we limit our objectives to the achievement of accuracy on the scale of the MP2 basis set limit results—i.e. better than 0.25 kcal/mol with an average error on the order of 0.1 kcal/mol.

We have carried out extrapolated LMP2 calculation using our own LMP2/cc-pVTZ (-f) optimized geometries (LMP2 column in Table 1) and at the MP2/6-311G\*\* geometries supplied in ref 1a (we refer to the canonical MP2 calculations in ref 1a as “MP2” in what follows). These results are presented in Table 1. Agreement of the extrapolated LMP2 results with those of ref 1a at the geometries specified therein (these data are what are used to train our extrapolation method) is on the order of  $\sim 0.1$  kcal/mol.

Parameter values obtained from the fitting to the training set (defined in eq 2) are as follows:

$$a = \exp(-A), b = \exp(-B), A = 1.8, B = 2.7 \quad (3)$$

The uniformly small size of the deviations in Table 1 for every training set molecule, combined with the fact that there are five data points and only two adjustable parameters, implies that the fitting protocol is extremely stable and robust. Note also that the difference between the LMP2 and MP2 protocols is generally less than or comparable to the difference between the MP2 and CCSD (T) protocols and that in fact the standard deviation of the LMP2 and CCSD (T) results is no worse than that between MP2 and CCSD (T). While there is no guarantee that this would be true for every molecular pair, the results again suggest that our protocol is a more or less equivalent replacement for large basis set extrapolated MP2 calculations.

For four of the five dimers, our LMP2 optimized geometries and energies are very close to those reported in ref 1a. However, for the fifth, a carboxylic acid dimer, there is a substantial disagreement, with a significantly larger binding energy being computed at the LMP2 optimized geometry. Given the very close agreement of the energies at the common geometry, there are two possible explanations for this discrepancy:

(1) The geometry optimizations have located different local minima (or, possibly, that one of the reported geometries is not in fact a local minimum, but on a plateau), due to differences in the geometry optimization algorithms in the two programs used for the calculations.

(2) The potential surfaces on which the geometry optimizations are carried out are different, leading to different locations for the minimum. This is in fact possible because we employ a different basis set, and we are using LMP2 as opposed to canonical MP2 methods (so that, for example, a substantial fraction of the basis set superposition error is not present even in the geometry optimization, as opposed to the final energy evaluation when counterpoise corrections are incorporated).

We cannot distinguish between these two possibilities without access to the code used by Tsuzuki and co-worker to obtain the results in ref 1a, which we do not have. We can, however, compare the energies of the two geometries using a common method (our own LMP2 extrapolation protocol). When these values are compared, the total electronic energy of the LMP2-optimized dimer is 2.28 kcal/mol lower than that of the MP2-optimized dimer. This fact, together with the very close energy agreement between our LMP2 calculations and those obtained by Tsuzuki and co-workers when the same geometries are used, suggests that we have, in fact, managed to locate a lower energy minimum than the one used in ref 1a, and we conclude that the correct

**Table 2.** Comparison of Formamide ab Initio Dimerization Energies in kcal/mol

| dimer | CCSD(T) limit <sup>a</sup> | LMP2 <sup>b</sup> |
|-------|----------------------------|-------------------|
| fm_1  | 14.35                      | 14.23             |
| fm_2  | 9.70                       | 9.62              |
| fm_3  | 7.34                       | 7.05              |
| fm_4  | 6.76                       | 6.87              |
| fm_5  | 5.02                       | 5.23              |

<sup>a</sup> Reference 1b. <sup>b</sup> Plus extrapolation. This work. LMP2/cc-pVTZ(-f) optimized geometry.

**Table 3.** Wall Clock Timing Results (s) for LMP2 Calculations Using the Jaguar and QChem Programs<sup>a</sup>

| dimer       | basis fns | QChem  | Jaguar | ratio<br>(QChem/Jaguar) |
|-------------|-----------|--------|--------|-------------------------|
| ala2        | 936       | 113203 | 41675  | 2.72                    |
| gly2        | 820       | 61551  | 34198  | 1.80                    |
| nma-dimer2  | 496       | 15439  | 4467   | 3.46                    |
| nma-acetone | 452       | 10004  | 4699   | 2.13                    |
| nma-dimer1  | 496       | 14200  | 4656   | 3.05                    |
| nma-dme     | 422       | 8465   | 2999   | 2.82                    |
| nma-meoh    | 364       | 4835   | 1657   | 2.92                    |

<sup>a</sup> nma = *N*-methyl acetamide, dme = dimethyl ether, gly2 = glycine dimer, ala2 = alanine dimer, nma-dimer1 and nma-dimer2 are two different nma dimers.

equilibrium gas-phase electronic binding energy for the formic acid dimer is 16.18 kcal/mol.

**B. Test Set.** As a test set, we use sets of five formamide molecular dimers recently studied in ref 1b using conventional MP2, CCSD (T), and extrapolation techniques. It can be seen in Table 2 that the extrapolated LMP2 results again are in excellent agreement, this time without any fitting to the high level data. This test further confirms the validity of the protocol we have developed.

**C. Computational Effort.** Table 3 presents the computational effort for PS-LMP2 calculations using the cc-pVTZ basis set for a number of small molecule dimers as a function of molecular size. All calculations were carried out on a 1.4GHz Pentium III PC. It can be seen that the scaling of computational effort with the size of the molecule is approximately in the  $N^3$  range in accordance with previous studies.

Table 3 also compares PS-LMP2 timings with those obtained from the LMP2 methodology in the Q-Chem program; the PS-LMP2 results are 2–3 times faster in wall clock time than those from Q-Chem, version 2.0. The LMP2 method used in Q-Chem was in the TRIM (triatomics in molecules) formulation.<sup>14</sup> It is different in its details from that in Jaguar; one major difference is that the local virtual space is somewhat larger, leading to recovery of a higher fraction of the canonical MP2 correlation energy. It is far from clear, however, whether this leads to greater accuracy in computation of properties such as binding affinity. The key issue is convergence of energy differences to the extrapolated basis set limit. The present paper demonstrates unambiguously that our definition of the local virtual space<sup>2,4</sup> leads to highly accurate extrapolated results using relatively small basis set calculations as an input, at least across the range of test cases that we have considered here. The

**Table 4.** Comparison of NMA ab Initio Dimerization Energies in kcal/mol

| dimer | CCSD(T)<br>limit <sup>a</sup> | LMP2 <sup>b</sup> | dimer | CCSD(T)<br>limit <sup>a</sup> | LMP2 <sup>b</sup> |
|-------|-------------------------------|-------------------|-------|-------------------------------|-------------------|
| nm_6  | 17.18                         | 14.96             | nm_8  | 10.76                         | 8.61              |
| nm_7  | 12.37                         | 10.24             | nm_9  | 9.67                          | 19.61             |

<sup>a</sup> Reference 1b. <sup>b</sup> Plus extrapolation. This work. The same geometry as in reference 1b.

**Table 5.** NMA ab Initio Dimerization Energies Computed Using QChem, kcal/mol

| dimer | cc-pVTZ(-f) <sup>a</sup> | cc-pVQZ(-g) <sup>a</sup> |
|-------|--------------------------|--------------------------|
| nm_6  | 13.30                    | 15.36                    |
| nm_7  | 8.72                     | 10.72                    |
| nm_8  | 7.84                     | 9.43                     |
| nm_9  | 18.98                    | 20.95                    |

<sup>a</sup> The same geometry as in reference 1b.

extrapolation properties of alternative local space definitions remain to be investigated.

#### D. Interaction Energies for *N*-Methyl Acetamide Dimers.

The work described in ref 1b also included calculations for four *N*-methyl acetamide dimers, in a number of different hydrogen bonding geometries. These calculations were performed at the cc-pVTZ level, and a BSSE correction was then applied; the results are shown in Table 4. Unlike the case of formamide, it was not possible to more rigorously determine the basis set MP2 limit binding energy, due to the larger size of the molecules involved.

Table 4 also compares our extrapolated PS-LMP2 results for the NMA dimers with these results. It can be seen that there is a substantial discrepancy even when identical geometries are used to carry out the calculation (there are also some differences in the optimized geometries as well); in general, the PS-LMP2 binding affinities are  $\sim 2$  kcal/mol smaller. This result was unanticipated, because one would ordinarily expect the BSSE corrected results to exhibit a lower binding energy than the converged basis set limit calculations. Thus, there is a discrepancy between our PS-LMP2 results and the results of ref 1b, which cannot be explained by a simple difference in geometry.

To investigate the problem further, we ran the local MP2 methodology in the Q-Chem program, using the TRIM approach, which is supposed to more closely reproduce full MP2 calculations. Results for cc-pVTZ(-f) and cc-pVQZ(-g) calculations, including a BSSE correction, are given in Table 5. It can be seen that the TZ results are much smaller than those of ref 1b and that the QZ results are reasonably close to our extrapolated LMP2 results. Extrapolation of the TRIM results cannot be performed using our LMP2-specific formulation; there is no particular reason to believe that the formulas should have significant similarity. Thus, all that can be said at this point is that the extrapolated PS-LMP2 and TRIM results are not clearly in disagreement, based on the results reported here. However, the possibility that the TRIM results when extrapolated would exceed the PS-LMP2 extrapolated results cannot be ruled out, and if this is the case, the resulting discrepancy is one that also would require further investigation.



As was discussed in the Introduction, a central application of high accuracy MP2-based calculations of binding energies is in the development of molecular mechanics force fields. In ref 8, we describe a protocol for developing a polarizable force field based on calculated dimer binding energies. The protocol employs electrostatic and polarization parameters fit to large basis set DFT calculations, universal  $1/r^6$  dispersion parameters for each atom to describe the long-range component of the van der Waals interaction, and fitting of the short-range atom–atom pair potential to the binding energy of molecular dimers. We have carried out this protocol for NMA, using the two different sets of dimer binding energies for nm\_6 in Table 4, but employing otherwise identical values for the dispersion parameters (which were fit to simulations of other small molecules, as is discussed in detail in ref 8), and the electrostatic model, which is determined from large basis set DFT calculations independent of the dimer binding energies. Thus, the models are identical apart from the short-range atom–atom pair terms. We then perform liquid-state simulations in an NPT ensemble, using methods described in ref 8, and determine the heat of vaporization and the density of the model for the liquid.

The results obtained are as follows. For the LMP2-derived parameters, the predicted heat of vaporization is 13.2 kcal/mol as compared to the experimental value of 13.4 kcal/mol; the density is 0.876 g/mol as compared to the OPLS-AA value of 0.897 (the OPLS-AA result was used as a reference, as no experimental data was available, and the other OPLS-AA properties of liquid NMA are in close agreement with experiment<sup>15</sup>). These results are entirely consistent with the other data presented in ref 8 for 16 molecular liquids, in which the average errors in the heat of vaporization were on the order of 0.2–0.3 kcal/mol, and errors in the density were under 5%. Note that no parameters in the calculation are fit to experimental data on NMA, so the results constitute a true test of the predictive capability of the methodology in the condensed phase. In contrast, if the NMA dimer nm\_6 binding energy from ref 1b is used to fit the short-range nonbonded parameters, the simulations fail to converge and the heat of vaporization observed during the simulation period reaches ca. 40 kcal/mol, which is in qualitative disagreement with the experimental data. These condensed phase tests therefore provide further confirmation of the robustness of our PS-LMP2 extrapolation protocol. It is unclear at present what the explanation is of the problematic results reported in ref 1b. However, our liquid state results, in conjunction with the previously discussed small molecule comparisons, suggest that the PS-LMP2 protocol has produced an accurate binding affinity for NMA. Further exploration of the discrepancies noted above remains to be undertaken.

#### IV. Conclusion

In this paper we have shown that a PS-LMP2 based approach to the calculation of intermolecular interaction energy is both highly accurate and very efficient in terms of computational efficiency as compared to alternatives. Achieving both speed and high accuracy together is nontrivial. DFT methods, for

example, would be considerably faster but have not yet been shown to yield the level of precision demonstrated above. When used to develop polarizable force fields for liquid-state simulations, this level of accuracy is inadequate to predict heats of vaporization to the desired level, whereas the PS-LMP2 values predict this and other condensed phase properties with good precision and reliability.

We are presently using this methodology to develop a complete polarizable force field for proteins and also for a wide range of organic chemical functional groups. However, there are some limitations which need to be addressed. First, as has been shown by Sherrill and co-workers,<sup>16</sup> extrapolated LMP2 methods are not adequate to achieve high accuracy for stacked benzene dimers; here, the CCSD (T) level corrections are significant. Fortunately, it appears as though these corrections can be performed with relatively small basis sets. Experiments need to be done to determine what other types of interactions require these higher-level correction. Examples of interactions that need to be investigated include those involving charged groups, different types of ring structures, and interactions involving a transition metal containing species.

**Acknowledgment.** This work was partially funded by the National Institutes of Health under Grant GM52018 to R.A.F.

#### References

- (1) (a) Tsuzuki, S.; Uchimaru, T.; Matsumara, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906–11910. (b) Vargas, R.; Garza, J.; Friesner, R. A.; Stern, H.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem.* **2001**, *105*, 4963–4968 and references therein.
- (2) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *J. Chem. Phys.* **1995**, *103*, 1481–1490.
- (3) Jaguar v5.0, Schrödinger, Inc. Portland, OR, 2003.
- (4) Murphy, R. B.; Pollard, W. T.; Friesner, R. A. *J. Chem. Phys.* **1997**, *106*, 5073–5084.
- (5) See for example (a) Petersson, G. A.; Frisch, M. J. *J. Phys. Chem. A* **2000**, *104*, 2183–2190. (b) Park, S. Y.; Lee, J. S. *J. Chem. Phys.* **2002**, *116*, 5389–5394. (c) Hobza, P.; Spöner, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808. (d) Huh, S. B.; Lee, J. S. *J. Chem. Phys.* **2003**, *118*, 3035–3042.
- (6) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (7) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y. H.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W. M.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C. P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532–1548.
- (8) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621–627.
- (9) Werner, H. J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149–8160.

- (10) (a) Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 3660–3671. (b) Schütz, M.; Hetzer, G.; Werner, H. J. *J. Chem. Phys.* **1999**, *111*, 5691–5705. (c) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592–3601. (d) Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H. J. *J. Chem. Phys.* **2000**, *113*, 9443–9455. (e) Saebø, S.; Pulay, P. *J. Chem. Phys.* **2001**, *115*, 3975–3983. (f) Schütz, M. *J. Chem. Phys.* **2002**, *116*, 8772–8785. (g) Walter, D.; Szilva, K. N. A. B.; Carter, E. A. *J. Chem. Phys.* **2002**, *117*, 1982–1993.
- (11) Schütz, M.; Rauhut, G.; Werner, H. J. *J. Phys. Chem. A* **1998**, *102*, 5997–6003.
- (12) (a) Fomine, S.; Tlenkopatchev, M.; Martinez, S.; Fomina, L. *J. Phys. Chem. A* **2002**, *106*, 3941–3946. (b) Huang, N.; MacKerell, A. D. *J. Phys. Chem. A* **2002**, *106*, 7820–7827.
- (13) See for example (a) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104–6114. (b) Martin, J. M. L. *J. Chem. Phys.* **1994**, *100*, 8186–8193. (c) Wilson, A. K.; Dunning, T. H. *J. Chem. Phys.* **1997**, *106*, 8718–8726. (d) Valeev, E. F.; Allen, W. D.; Hernandez, R.; Sherrill, C. D.; Schaefer, H. F. *J. Chem. Phys.* **2003**, *118*, 8594–8610.
- (14) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592–3601.
- (15) Jorgensen, W. L.; Maxwell D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *113*, 11225–11236.
- (16) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893.

CT0498800

## Connecting Pauling and Mulliken Electronegativities

David R. Herrick\*

Department of Chemistry and Institute of Chemical Physics, University of Oregon,  
Eugene, Oregon 97403-1253

Received September 7, 2004

**Abstract:** A conversion formula between Pauling and Mulliken electronegativities is derived from a new analytic treatment of the classic valence bond model for a single bond A–B in terms of Mulliken-Jaffe “a and b” parameters for the atomic valence states. The new formula works better than the traditional linear and half-power empirical formulas. The results are consistent with an absolute Pauling scale in which both electronegativities approach zero simultaneously.

### 1. Introduction

An unresolved issue in the evolution of modern electronegativity<sup>1–4</sup> has been the possibility of a practical quantum formula connecting the scales invented by Pauling<sup>5,6</sup> and Mulliken.<sup>7,8</sup> Pauling electronegativities from standard heats of formation<sup>9</sup> are linked to the empirical bond energy formula

$$D(A-B) = \frac{D(A-A) + D(B-B)}{2} + K_P(X_A - X_B)^2 \quad (1)$$

representing nonpolar and polar contributions, respectively, with  $K_P = 1.00$  eV. Mulliken electronegativities for atoms in bonds are defined by

$$X = \frac{IE_v + EA_v}{2} \quad (2)$$

with effective ionization energies and electron affinities related to Mulliken-Jaffe “a and b” parameters<sup>1,2</sup> for common valence states. Although the quantum roots of electronegativity are evident in leading valence bond and molecular orbital approaches<sup>10,11</sup> and density functional theory<sup>12–15</sup> this has not identified a clear working formula connecting the Pauling and Mulliken scales. Traditional empirical conversion formulas have assumed a linear relationship. Early formulas based on limited data include Mulliken’s<sup>8</sup> original formula (1935, 11 atoms)

$$\Delta_P = 0.360\Delta_M \quad (3)$$

for differences  $\Delta = |X_A - X_B|$ , Skinner and Pritchard’s<sup>16</sup> absolute formula (1953, 15 atoms)

$$X_P = 0.317X_M \quad (4)$$

and Hinze, Whitehead, and Jaffe’s<sup>17–19</sup> shifted formula (1963, 11 atoms)

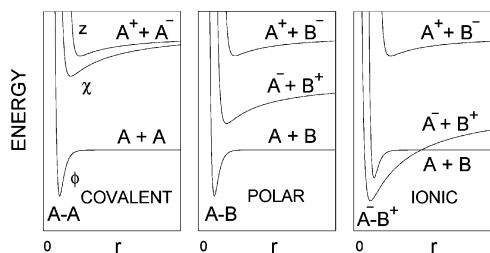
$$X_P = 0.336(X_M - 0.615) \quad (5)$$

More recently, Bratsch<sup>1</sup> has proposed an ad hoc half-power formula (1988, 36 atoms)

$$X_P = 1.35X_M^{1/2} - 1.37 \pm 0.14 \quad (6)$$

based on revised and extended electronegativities. While the theoretical basis for these formulas is not clear, Komorowski<sup>20</sup> has suggested an interesting possibility for the slope in the linear formula in terms of a constant  $(4\pi\epsilon_0/7)^{1/2} = 0.3151$  from a charge-transfer model for  $\Delta_M$ . Our present theoretical approach identifies new conversion formulas, derived here from an exact nonlinear  $2 \times 2$  reduction of the classic  $3 \times 3$  valence bond model. The model simplifies in two different bonding limits. For polar bonds it explains the basis, limitations, and modifications of the traditional linear conversion formula. In the ionic bond limit it yields a nontraditional half-power formula similar to Bratsch’s formula. In both cases the new formulas give more precise fits of empirical electronegativities.

\* Corresponding author phone: (541)346-2529; e-mail: dherrick@uoregon.edu.



**Figure 1.** Potential energy curves for covalent and ionic states in simple bonds ( $X_A > X_B$ ). Labels indicate the bonding at the minimum of each ground-state curve.

## 2. Theoretical Basis

The general strategy first casts the exact bond energy into nonpolar and polar parts

$$D(A-B) = D^n(A-B) + D^p(A-B) \quad (7)$$

and determines Mulliken's  $\Delta_M$  from a valence bond estimate of the polar term  $D^p(A-B)$ . The energy separation involves breaking the wave function into orthogonal pieces  $\psi = \phi + \chi$  that are symmetric ( $\phi$ ) or antisymmetric ( $\chi$ ) when orbitals are switched between the atoms. The nonpolar and polar terms in eq 7 are defined here by the matrix elements  $D^n(A-B) = -(\phi|H|\phi)$  and  $D^p(A-B) = -(\phi|H|\chi)$  when things are normalized to  $(\phi|\phi) = 1$ . The connection to Pauling's electronegativity is made directly by setting  $D^p(A-B) = K_p(\Delta_p)^2$ . One feature of this approach is that it does not depend on Pauling's arithmetic mean approximation. A similar definition of Pauling electronegativities has been made in terms of the charge-transfer affinity<sup>20</sup> in the corresponding partial charge model.

**2.1. Valence Bond Model.** The preceding strategy is applied to the valence bond wave function in terms of covalent and ionic structures ( $AB$ ), ( $A^-B^+$ ) and ( $A^+B^-$ ). Figure 1 illustrates the potential energy curves related to the Pauling and Mulliken pictures.

We start with symmetrized wave functions that have the general form

$$\begin{aligned} \phi &= c_1(A_1B_2 + B_1A_2) + c_2(A_1A_2 + B_1B_2) \\ \chi &= c_3(A_1A_2 - B_1B_2) \\ z &= c_4(A_1B_2 + B_1A_2) + c_5(A_1A_2 + B_1B_2) \end{aligned} \quad (8)$$

The covalent bond  $A-A$  is 100%  $\phi$ ; the first ionic state above it is 100%  $\chi$ , and the highest state is 100%  $z$ . The polar bond  $A-B$  is a blend including mostly  $\phi$  with a weaker admixture of  $\chi$ . The ionic bond  $A^-B^+$  involves a stronger mix that occurs when the ionic state crosses below the covalent curve. The coefficients  $c_1$  and  $c_2$  for each bond are initially selected to minimize the nonpolar ground-state expectation energy  $E_\phi = (\phi|H|\phi)$  with  $(\phi|\phi) = 1$ . The polar mixing of  $\phi$  and  $\chi$  then shifts the ground-state energy downward to a new value  $E$  described by the  $2 \times 2$  secular determinant

$$\begin{vmatrix} E_\phi - E & -2^{1/2}c_2\Delta \\ -2^{1/2}c_2\Delta & E_\phi - E + \frac{1}{2}\Sigma \end{vmatrix} = 0 \quad (9)$$

with energy parameters

$$\Delta = \frac{E(A^+B^-) - E(A^-B^+)}{2}$$

$$\Sigma = E(A^+B^-) + E(A^-B^+) - 2E_\phi(AB) \quad (10)$$

consistent with orthonormal orbitals and neglect of differential overlap. The parameter  $c_2$  in the off-diagonal term comes from the coupling between  $\chi$  and the ionic part of  $\phi$ . The exact ground state is obtained when  $c_2$  is readjusted to minimize  $E$ .<sup>21</sup> This nonlinear approach gives the following optimized formulas:

$$\begin{aligned} E &= E_\phi - \frac{2K\Sigma\delta^2}{1 + \sqrt{1 + 8K\delta^2}} \\ c_3 &= \frac{4c_2\delta}{1 + \sqrt{1 + 8K\delta^2}} \end{aligned} \quad (11)$$

with  $K = 4c_2^2$  and  $\delta = \Delta/\Sigma$ . The polar part of the bond energy is  $D^p(A-B) = E_\phi - E$ , and the valence bond estimate of the bond partial charge is

$$\delta_{VB} = \frac{4c_2c_3}{1 + 2c_3^2} = \frac{2K\delta}{\sqrt{1 + 8K\delta^2}} \quad (12)$$

Given the simplicity of the model, Pauling's formula, and uncertainties in the empirical electronegativity assignments we focus on the leading order scaling with the valence bond parameters. Perturbation theory gives

$$\begin{aligned} D^p(A-B) &= K\Sigma\delta^2 \\ \delta_{VB} &= 2K\delta \end{aligned} \quad (13)$$

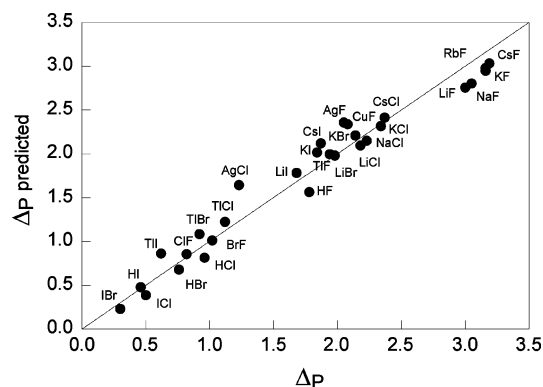
Similar scaling is expected with other wave functions. When overlap ( $s$ ) between  $A$  and  $B$  is included in the off-diagonal coupling in eq 9, for example, the Mulliken approximation<sup>22</sup> for differential overlap gives

$$\begin{aligned} c_2 &\rightarrow \frac{c_2 + sc_1}{\sqrt{1 - s^2}} \\ K &\rightarrow \frac{1}{1 - s^2} + 2(c_2^2 - c_1^2) \end{aligned} \quad (14)$$

**2.2. Electronegativities.** Replacing the molecular energies  $\Delta$  and  $\Sigma$  with equivalent atomic valence energies makes the connection to Mulliken's electronegativity. This gives working valence bond parameters  $\Delta_M$ ,  $\Sigma_M$ , and  $\delta_M$  defined by

$$\begin{aligned} \Delta_M &= \frac{(IE_v + EA_v)_A - (IE_v + EA_v)_B}{2} \\ \Sigma_M &= (IE_v - EA_v)_A + (IE_v - EA_v)_B \\ \delta_M &= \frac{\Delta_M}{\Sigma_M} \end{aligned} \quad (15)$$

which are related to Mulliken-Jaffe parameters<sup>1,2</sup> "a" (atom electronegativity) and "b" (atom charge coefficient) by  $\Delta_M$



**Figure 2.** Correlation of predicted electronegativity  $\Delta_P$  (eq 17,  $\alpha = 1.065$ ) with Pauling's  $\Delta_P$ . Line is predicted  $\Delta_P = \Delta_P$ .

$\leftrightarrow a_A - a_B, \Sigma_M \leftrightarrow b_A + b_B$  and  $\delta_M \leftrightarrow \delta_{MJ}$  (partial ionic charge parameter). The important point here is that eq 13 connects Pauling and Mulliken electronegativities through the formula

$$K_P(\Delta_P)^2 = K_M \Sigma_M \delta_M^2 \quad (16)$$

where  $K_M$  is treated initially as an empirical constant.

### 3. Theoretical Analysis

**3.1. Conversion Formulas.** Two alternative forms of eq 16 are the conversion formula

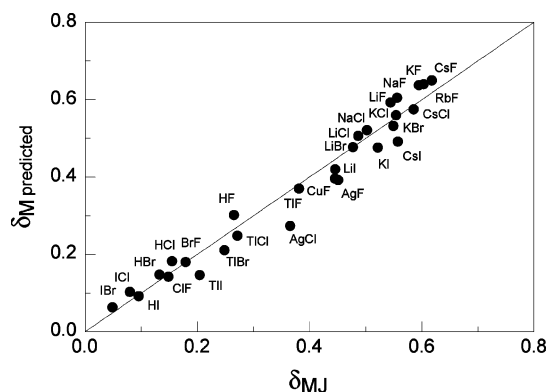
$$\Delta_P = \alpha \frac{\Delta_M}{\sqrt{\Sigma_M}} \quad (17)$$

with  $\alpha = (K_M/K_P)^{1/2}$  and the partial charge formula

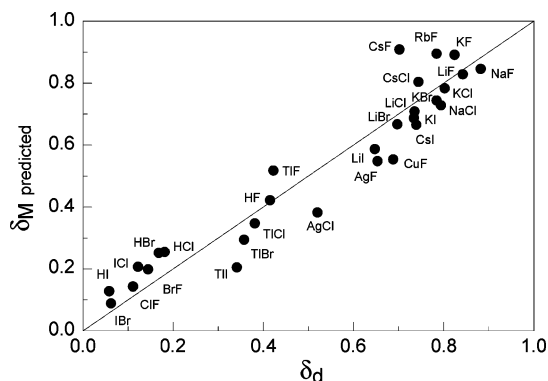
$$\delta_M = \beta \frac{\Delta_P}{\sqrt{\Sigma_M}} \quad (18)$$

with  $\beta = (K_P/K_M)^{1/2}$ . Note that interpretations with these formulas will depend on the weight factor  $1/(\Sigma_M)^{1/2}$  in addition to electronegativities in the conjugate scaling variables  $\Delta_M/(\Sigma_M)^{1/2}$  and  $\Delta_P/(\Sigma_M)^{1/2}$ . In particular, eq 17 predicts the traditional linear scaling law  $\Delta_P \propto \Delta_M$  only to the extent that  $\Sigma_M$  is constant. Figure 2 shows the empirical correlation between eq 17 and Pauling's  $\Delta_P$  for a sample of 29 diatomic bonds. All electronegativity data were taken from Bratsch's tables.<sup>1</sup> The standard deviation of the fit is 23% lower than that for the conventional linear fit  $\Delta_P \propto \Delta_M$ . According to eq 17 this improvement reflects a greater uncertainty in the linear fit due to the distribution of values of  $\Sigma_M$  (13–31, avg =  $20 \pm 5$ ) over the sample. Figure 3 shows the related correlation of the same data between eq 18 and the Mulliken-Jaffe partial charge. Figure 4 shows a similar, but weaker correlation between eq 18 and the historical degree of ionization<sup>10,11</sup>  $\delta_d = \mu/r_e$  from experimental dipole moments and bond lengths.<sup>23</sup> Although it is well known that the empirical values of  $\delta_d$  do not represent a simple dipole charge<sup>10,11</sup> the correlation with the current formula for  $\delta_M$  is nonetheless interesting.

**3.2. Dependence of K on Partial Ionic Charge.** Additional improvements can be made by taking into account variations of the valence bond coefficient  $K = 4c_2^2$  with bond



**Figure 3.** Correlation of predicted partial ionic charge  $\delta_M$  (eq 18,  $\beta = 0.939$ ) with Mulliken-Jaffe partial ionic charge  $\delta_{MJ}$ . Line is  $\delta_M = \delta_{MJ}$ .



**Figure 4.** Correlation of predicted partial ionic charge  $\delta_M$  (eq 18,  $\beta = 1.31$ ) with empirical "dipole charge"  $\delta_d = \mu/r_e$ . Line is  $\delta_M = \delta_d$ .

polarity. Minimizing  $E$  in eq 9 tends to shift  $K$  to higher values due to a negative slope  $dE/dc_2 < 0$  at the initial minimum for  $E_\phi$ . As a simple model, the polar molecular orbital wave function  $\lambda_1 \lambda_2$  with

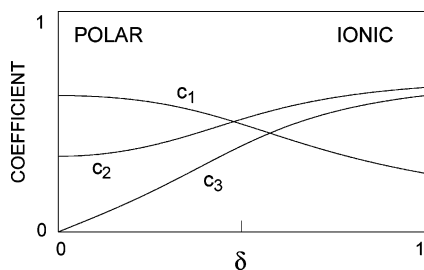
$$\lambda = \sqrt{\frac{1+\delta}{2}}A + \sqrt{\frac{1-\delta}{2}}B \quad (19)$$

gives  $K = 2/(2 - \delta^2)$ , which increases between  $K = 1$  for a nonpolar bond ( $\delta = 0$ ) and  $K = 2$  for an ionic bond ( $\delta = 1$ ). A model more consistent with the current valence bond approach is obtained from

$$\begin{bmatrix} -E & 2\beta & 0 \\ 2\beta & \frac{1}{2}\Sigma_o - E & -\Delta \\ 0 & -\Delta & \frac{1}{2}\Sigma_o - E \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = 0 \quad (20)$$

in terms of effective resonance and average excitation energy parameters  $\beta$  and  $\Sigma_o$ . Matching this to the energies  $E_\phi$  and  $\Sigma$  in eq 9 when  $\delta = 0$  gives the relationship  $\Sigma_o = \Sigma(1 - \gamma^2)$  with the reduced resonance parameter  $\gamma = 4\beta/\Sigma$ . Figure 5 illustrates general behavior of the mixing coefficients between the covalent and ionic regions when  $\gamma$  is constant.

A formula for  $K$  as a function of  $\delta$  is derived from an exact representation of eq 20 with a nonlinear  $2 \times 2$  matrix for  $\phi$  in which the coupling with  $\chi$  is folded in as an effective potential  $H_{\phi\chi}(E - H_{\chi\chi})^{-1}H_{\chi\phi}$ . In terms of reduced energies  $\epsilon$



**Figure 5.** Dependence of ground-state valence bond coefficients ( $\gamma^2 < 1$ ) on the partial charge parameter  $\delta$ . This illustrates a smooth increase of  $c_2$  and hence of  $K$ . According to eq 14 the coefficient  $K$  satisfies  $K > 1$  on the ionic side of the crossing point  $c_1 = c_2$ . At high  $\delta$  the coefficients approach  $c_1 = 0$  and  $c_2 = c_3 = 1/2^{1/2}$  with  $K = 2$  for the ionic limit  $A^-B^+$ . When  $\gamma^2 > 1$ , the order of the coefficients  $c_1$  and  $c_2$  at  $\delta = 0$  switches to  $c_2 > c_1$  with  $K > 1$ .

$= 2E/\Sigma$  and  $\sigma = \Sigma_o/\Sigma = 1 - \gamma^2$ , this projection gives

$$\begin{bmatrix} -\epsilon & \gamma \\ \gamma & \sigma - \epsilon - \omega \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0 \quad (21)$$

with the effective ionic interaction

$$\omega = \frac{4\delta^2}{\sigma - \epsilon} \quad (22)$$

ground-state energies from

$$\epsilon = \frac{1}{2}(\sigma - \omega) - \sqrt{\frac{1}{4}(\sigma - \omega)^2 + \gamma^2} \quad (23)$$

and

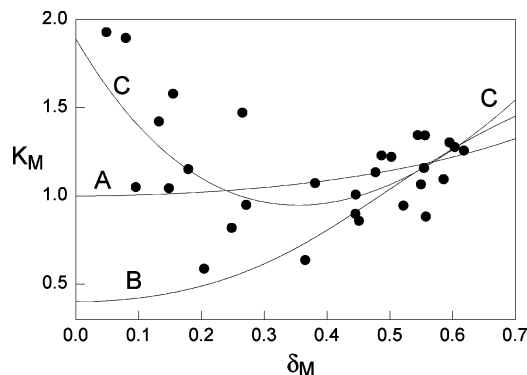
$$K = \frac{2}{1 + (\gamma/\epsilon)^2} \quad (24)$$

The curves for  $K$  as a function of  $\delta$  have a positive curvature at  $\delta = 0$ , and they approach the limit  $K = 2$  at high  $\delta$ . At  $\delta = 0$  eq 24 reduces to

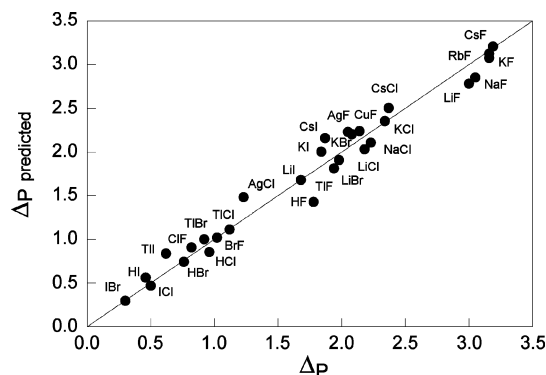
$$K = \frac{2\gamma^2}{1 + \gamma^2} \quad (25)$$

which gives a range of values  $K = 0-1$  when  $\gamma^2 < 1$  (mostly covalent nonpolar bond),  $K = 1$  when  $\gamma^2 = 1$  (50% covalent and 50% ionic nonpolar bond), and  $K = 1-2$  when  $\gamma^2 > 1$  (mostly ionic nonpolar bond). The value  $\gamma^2 = 1/3$  is special because it makes the leading valence bond partial charge in eq 13 equal to the Mulliken-Jaffe partial charge,  $\delta_{VB} = \delta_M$ . The crossover value  $\gamma^2 = 1$  gives an exact solution  $\omega = 4\delta^2/(1 + 4\delta^2)^{1/2}$  with  $K = (1 + 4\delta^2)/(1 + 2\delta^2)$ . At  $\delta = 0$  this reduces to the molecular orbital model in eq 18. For other values of  $\gamma^2$ , we found that using the empirical function  $\omega = 4\delta^2/(1 + 3\delta^2)^{1/2}$  directly in eq 23 gives surprisingly good estimates of  $K$  over  $\delta = 0-1$ .

**3.3. Empirical Values of  $K_M$ .** Figure 6 compares the models for  $K$  to empirical values of  $K_M$  from eq 16. The previous fits in Figures 2 and 3 correspond here to a constant value  $K_M = 1.13$ . Curve A shows the gradual rise of  $K$  predicted by the molecular orbital model. Curve B shows  $K$



**Figure 6.** Dependence of empirical values of  $K_M = K_P \Sigma_M (\Delta_P / \Delta_M)^2$  on partial charge  $\delta_M = \Delta_M / \Sigma_M$ . Curves: A) molecular orbital model; B) valence bond model; C) fit in eq 26.



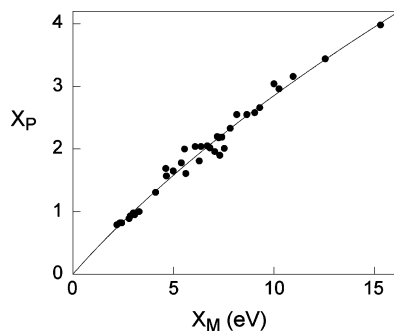
**Figure 7.** Improved correlation of predicted electronegativity  $\Delta_P$  (eq 17 using  $K$  from eq 26) with Pauling's  $\Delta_P$ . Line is predicted  $\Delta_P = \Delta_P$ .

from the valence bond model with fixed parameters  $\Sigma = 16$  eV and  $\beta = -2$  eV selected to represent the lower covalent region. Further improvements with these data for  $K_M$  are found with functions that have a minimum near  $\delta = 0.3-0.4$ . Curve C shows one empirical function

$$K_M = 1.89e^{-3.23\delta} + 2.75\delta^2 \quad (26)$$

which describes a minimum at  $\delta = 0.35$ . The comparison to Pauling's  $\Delta_P$  is shown in Figure 7, which includes noticeable improvements over Figure 2. The standard deviation of the fit is 31% lower than that for the linear fit  $\Delta_P \propto \Delta_M$ . This improves to 35% when the function in eq 26 is optimized in terms of the scaling variable  $\Delta_M/(\Sigma_M)^{1/2}$  instead of  $\delta_M$ .

The apparent rise of the empirical values of  $K_M$  near  $\delta_M = 0$  in Figure 6 corresponds to a higher slope in eq 17 for the bonds  $I\text{Br}$ ,  $I\text{Cl}$ ,  $\text{HBr}$ ,  $\text{HCl}$  and  $\text{HF}$ , possibly due in part to systematic deviations between  $\Delta_P$  and  $\Delta_M$ . Higher values of  $K$  for weakly polar bonds would also be consistent with shifts of the parameter  $\gamma^2$  to higher values with  $\gamma^2 > 1$ . Because of these possibilities it is not clear to what extent the empirical behavior of  $K_M$  might reflect higher order contributions from the exact valence bond energy in eq 11. This would have  $K_M$  approach zero at high  $\delta$  and give a negative curvature at  $\delta = 0$  when  $\gamma^2 > 1$ .



**Figure 8.** Correspondence between Pauling and Mulliken electronegativities from ref 1. Curve is the ionic bond half-power formula in eq 29 which gives a zero intercept at  $X_M = 0$ . Comparable curve is described by the polar bond half-power formula in eq 30.

#### 4. Half-Power Scaling

The valence bond formulas also suggest a partial basis for Bratsch's empirical half-power formula in eq 6. For comparison we first fitted the same data for 36 atoms with eq 17 relative to fluorine. This gave a better fit

$$X_P = 3.98 + \alpha \frac{X_M - 15.30}{\sqrt{b_{MJ} + 17.81}} \pm 0.11 \quad (27)$$

with a single adjustable parameter  $\alpha = 1.137$ . While this improvement over eq 6 suggests half powers may not be necessary, it is interesting that the valence bond model does in fact predict half-power scaling under certain conditions.

One approach involves the ionic bond limit of the exact solution of the secular determinant

$$K_P(\Delta_P)^2 \propto \frac{2K\Sigma\delta^2}{1 + \sqrt{1 + 8K\delta^2}} \quad (28)$$

For small  $\delta$  this reduces to the polar bond formula in eq 16. At high values of  $\delta$  it gives the half-power formula  $\Delta_P \propto (\Delta_M)^{1/2}$ , including Bratsch's formula as a special case when the reference state in  $\Delta_M$  is assumed to have  $X_M = 0$ . Relaxing this condition gives a better global fit

$$X_P = 1.934\sqrt{X_M + 7.02} - 5.12 \pm 0.12 \quad (29)$$

which is shown in Figure 8.

In a second approach, we note that the polar bond formula eq 27 predicts a half-power directly for low values of  $\delta$  when the charge coefficient  $b_{MJ}$  is a linear function of  $X_M$ . This is not a bad approximation,<sup>1</sup> and a direct fit with  $b_{MJ} = cX_M + d$  gives  $c = 1.04 \pm 0.08$  and  $d = 0.93 \pm 0.57$ . Refitting eq 27 with a similar linear function gives the half-power formula

$$X_P = 3.98 + 1.154 \frac{X_M - 15.30}{\sqrt{X_M + 19.68}} \pm 0.12 \quad (30)$$

which gives results similar to eq 29. It also reduces to a simple half-power formula at high  $X_M$ .

An interesting feature of the independent half-power fits in eqs 29 and 30 is that both functions extrapolate to a zero

intercept  $X_P = 0.00$  at  $X_M = 0$ . This empirical result supports a possible absolute Pauling scale as a nonlinear generalization of the Skinner and Pritchard formula in eq 4. Both eqs 29 and 30 give the same linear formula  $X_P = 0.36X_M$  near the intercept, including a slope consistent with Mulliken's original empirical formula in eq 3.

#### 5. Conclusion

A method for polar contributions to the bond energy described herein was applied to the valence bond wave function to derive formulas and results which appear to clarify and improve the connection between Pauling and Mulliken electronegativities. In effect the relationship in eq 17 (and its application such as eq 27) defines a semiempirical Pauling scale for valence states consistent with Mulliken-Jaffe parameters. And eq 18 defines a semiempirical partial charge consistent with Pauling's electronegativity. These conjugate formulas provide a self-consistent basis for further studies and possible refinements of the electronegativity scales.

**Acknowledgment.** The author thanks R. Hoffmann and W. B. Jensen for helpful comments.

#### References

- Bratsch, S. G. Revised Mulliken Electronegativities. *J. Chem. Educ.* **1988**, *65*, 34–41; 223–227.
- Huheey, J. E.; Keiter, E. A.; Keiter, R. L. *Inorganic Chemistry: Principles of Structure and Reactivity*, 4th ed.; Harper Collins: New York, 1993; pp 182–199.
- Sanderson, R. T. Principles of Electronegativity. *J. Chem. Educ.* **1988**, *65*, 112–118; 227–231.
- Jensen, W. B. Electronegativity from Avogadro to Pauling. *J. Chem. Educ.* **1996**, *73*, 11–20; **2003**, *80*, 279–287.
- Pauling, L. The Nature of the Chemical Bond. IV. The Energy of Single Bonds and the Relative Electronegativity of Atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570–3582.
- Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University: Ithaca, NY, 1960; pp 64–97.
- Mulliken, R. S. A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J. Chem. Phys.* **1934**, *2*, 782–793.
- Mulliken, R. S. Electronic Structures of Molecules. XI. Electroaffinity, Molecular Orbitals and Dipole Moments. *J. Chem. Phys.* **1935**, *3*, 573–585.
- Allred, A. L. Electronegativity Values from Thermochemical Data. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215–221.
- Pritchard, H. O.; Skinner, H. A. The Concept of Electronegativity. *Chem. Rev.* **1955**, *55*, 745–786.
- Ferreira, R. Electronegativity and Chemical Bonding. *Adv. Chem. Phys.* **1967**, *13*, 55–84.
- Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. Electronegativity: The Density Functional Viewpoint. *J. Chem. Phys.* **1978**, *68*, 3801–3807.
- Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989; pp 70–98.

- (14) Chermette, H. Chemical Reactivity Indexes in Density Functional Theory. *J. Comput. Chem.* **1999**, *20*, 129–154.
- (15) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103*, 1793–1873.
- (16) Skinner, H. A.; Pritchard, H. O. The Measure of Electronegativity. *Trans. Faraday Soc.* **1953**, *49*, 1254–1262.
- (17) Hinze, J.; Jaffe, H. H. Electronegativity. I. Orbital Electronegativity of Neutral Atoms. *J. Am. Chem. Soc.* **1962**, *84*, 540–546.
- (18) Hinze, J.; Whitehead, M. A.; Jaffe, H. H. Electronegativity. II. Bond and Orbital Electronegativities. *J. Am. Chem. Soc.* **1963**, *85*, 148–154.
- (19) Hinze, J.; Jaffe, H. H. Electronegativity. IV. Orbital Electronegativities of the Neutral Atoms of the Periods Three A and Four A and of the Positive Ions of Periods One and Two. *J. Phys. Chem.* **1963**, *67*, 1501–1506.
- (20) Komorowski, L. Chemical Hardness and L. Pauling's Scale of Electronegativity. *Z. Naturforsch.* **1987**, *42a*, 767–773.
- (21) This exact treatment differs from the usual linear  $2 \times 2$  approximation which neglects the highest structure ( $A^+B^-$ ) at the outset by setting  $c_2 = c_3$ .
- (22) Mulliken, R. S. Quelques Aspects De La Théorie des Orbitales Moléculaires. *J. Chim. Phys.* **1949**, *46*, 497–542; 675–713.
- (23) NIST Diatomic Spectral Database (accessed Feb 2004): <http://physics.nist.gov/PhysRefData/MolSpec/Diatomic/index.html>  
CT049942A



## An Examination of Basis Set Superposition Error at the Correlated Level: Illuminating the Role of the Exchange Repulsion

Clinton S. Nash\*

*Department of Chemistry and Physics, University of New England,  
11 Hills Beach Road, Biddeford, Maine 04005*

Received November 24, 2004

**Abstract:** A reformulated ‘virtuals-only’ counterpoise procedure for the correction of basis set superposition error is introduced and applied to an examination of the interaction energy of the helium dimer. This method allows a direct calculation of the exchange (Pauli) repulsion at the correlated level. Our findings suggest that at the correlated level the Pauli repulsion may be interpreted as a decrease in the monomer correlation energy calculated using the dimer virtual space relative to that with a monomer-only virtual space. This observation provides further theoretical justification for the Full Counterpoise procedure in preference to any virtuals-only counterpoise procedure.

### Introduction

The use of various methods to remove basis set superposition error (BSSE) from ab initio quantum chemical calculations has become standard practice. The recognition that the application of such a correction is crucial to a quantitative description of bonding, particularly those in weakly interacting systems, has focused attention on those methods. The most important of these are the *a priori* Chemical Hamiltonian Approach (CHA) and symmetry-adapted perturbation theory (SAPT) and the *a posteriori* Function Counterpoise Correction (FCP) of Boys and Bernardi.<sup>1–3</sup> Because of its conceptual simplicity and ease of implementation, the latter is much more commonly applied than either of the former and is the focus of this contribution. That BSSE is an unphysical computational artifact brought about by the artificial extension of the variational space available to a monomer, A, in an interacting system AB is in fact demonstrated by application of the FCP procedure. Namely, the energy of the monomer A calculated at some level of theory in the combined basis set of the dimer, AB, is lowered relative to its energy found using only its own basis set. In the function counterpoise procedure this energy difference is simply subtracted from the energy of the dimer (along with difference between the energy of B calculated in the

dimer basis and in only its own) evidently leaving a total energy devoid of basis set superposition error.

Almost from its original introduction, however, objections were raised that the Bernardi and Boys full counterpoise correction overestimates the BSSE.<sup>4–6</sup> The rationale for these objections, particularly for correlated calculations, is based on the Pauli exclusion principle and a perceived imbalance in the treatment of a dimer on one hand and the individual monomers on the other. In a correlated counterpoise calculation on a monomer (i.e. using the basis set of the dimer), excitations to (amplitudes of) orbitals corresponding to occupied MOs of B are allowed, whereas in the dimer they are not. Accordingly, contributions from these ‘spurious’ excitations would lead to an FCP monomer energy that is too low, an overall counterpoise correction that is too large, and therefore a potential energy surface that is too shallow. Although various ad hoc procedures with varying degrees of theoretical justification have been advanced to compensate for that anticipated overcorrection, most are unsatisfactory.<sup>7</sup>

To address the issue directly Daudey et al. devised a ‘virtuals-only’ counterpoise procedure (VCP) in which only the virtual SCF orbitals of the ‘ghost’ atom B are used to augment the molecular orbital basis set of monomer A in a correlated CP calculation on A.<sup>4</sup> In principle, this should lead to a smaller monomer counterpoise correction than the full counterpoise procedure and ultimately a less shallow interaction potential energy surface. In subsequent years, several

\* Corresponding author e-mail: cnash@une.edu.

studies have been conducted comparing efficacy of the full and virtuals-only counterpoise procedures.<sup>8–11</sup> Though the issue remains somewhat contentious, most of the recent work seems to endorse the use of the FCP method, especially when larger, more complete basis sets are used.<sup>12,13</sup> At the same time, some authors, notably Gutowski et al., have advanced formal arguments supporting the correctness of the FCP and have argued that excluding the ‘to be occupied’ orbitals from the FCP virtual space leads to an underestimation of the exchange repulsion in the AB supermolecule.<sup>10,14</sup>

We have reexamined this issue in the case of the helium dimer and focused on the role of the Pauli repulsion using a new virtuals-only counterpoise procedure that is similar in philosophy to but substantially different from that proposed by Daudey. As the weakest possible interaction, this choice for a molecular system provides a sensitive test of the various counterpoise correction prescriptions and allows comparisons with an extensive body of existing literature. Our new approach raises new questions about the counterpoise procedure and suggests that a new interpretation of the role that the ‘spurious’ orbitals play may be in order. In particular, we have found it appropriate to reinterpret the exchange (Pauli) repulsion at the correlated level as a disruption of the correlation energy of each monomer in a supermolecular aggregate by the presence of occupied orbitals located on its partner(s).

## Theoretical Methods

For all of the *ab initio* calculations done in our examination of BSSE in the helium dimer we have used Dunning’s aug\_cc-pVXZ ( $X=T, Q, 5$ ) basis sets and a slightly modified version of the GAMESS quantum chemistry package.<sup>16–18</sup> Our variant of the virtuals-only counterpoise procedure (vVCP), to be described below, was conducted using a custom program designed for that purpose. The various counterpoise corrections were carried out according to the standard formulas (eq 1 below). The energy of the interacting dimer AB (here,  $A=B=He$ ) determined at some level of theory (SCF, MP2, CCSD, etc.) is modified by the difference between the energy of each of the monomers A and B calculated in the union of their basis sets  $(A \cup B)_{XCP}$  and their energies calculated using their individual basis sets.

$$E_{AB}^{XCP} = E_{AB} - (E_A^{XCP} + E_B^{XCP}); X = F, vV, V$$

$$E_A^{XCP} = E_A(A \cup B)_{XCP} - E_A(A)$$

$$E_B^{XCP} = E_B(A \cup B)_{XCP} - E_B(B) \quad (1)$$

We have calculated MP2, CCSD, and CCSD(T) interaction energies for the helium dimer at a series of 11 internuclear distances between 4.5 and 14 Bohr using each of the correlation-consistent basis sets. These surfaces were corrected using the full counterpoise procedure (FCP) as well as our variant of the virtuals-only counterpoise procedure (vVCP). For comparison, Daudey’s virtuals-only counterpoise procedure (VCP) was also applied to the  $He_2$  PES calculated under aug\_cc-pVTZ atomic orbital basis set. The various molecular orbital basis sets symbolized by  $(A \cup B)_{XCP}$

have been constructed according to the requirements of each counterpoise correction scheme, X ( $X=F, vV, V$ ).

In Daudey’s original formulation of the virtuals-only counterpoise procedure (VCP), the basis set of a monomer, A, is augmented by only the virtual SCF orbitals of the ghost monomer B.<sup>4</sup> These ghost virtuals are projected onto the orthogonal subspace of the molecular orbital basis set of A and reorthogonalized to it using Löwdin’s symmetric orthogonalization procedure.<sup>19</sup> The result is an orthonormal molecular orbital basis set useful in correlated VCP calculations which omits orbitals occupied in the AB dimer.

In contrast, our variant VCP formulation (vVCP) is more closely related to the standard counterpoise procedure and in fact reduces to it at the SCF level. Rather than projecting the virtual ghost orbitals of B onto the orthogonal subspace of the monomer A, we perform a typical full counterpoise SCF calculation (i.e. SCF on monomer A using the  $A \cup B$  atomic orbital basis set) and project out of the resulting virtual space functions corresponding to the  $n_B$  ( $=1$  for helium) occupied orbitals of the ghost atom B. These have been generated using a separate fragment (atomic) SCF calculation under the same atomic basis set used for B in both the FCP and AB supermolecule. Prior to the application of the projection operator, these occupied ‘ghost’ functions are orthogonalized to the occupied orbitals of A. The linear dependency that results in the supermolecular MO basis is resolved by deleting the  $n_{B,occ}$  functions identified as being most similar to the original occupied ghost orbital(s) of B. The remaining molecular orbital basis set of dimension  $N_A + N_B - n_{B,occ}$  (where  $N_A$  and  $N_B$  are the numbers of atomic basis functions for fragment A and B, respectively) is then renormalized and orthogonalized using Löwdin’s procedure in order to minimize the differences between the original FCP MO basis set and our vVCP MO basis set. The result of this procedure is a virtual space having the orbital structure of a hypothetical noninteracting AB dimer.

In our view, this vVCP method has several features that distinguish it from Daudey’s approach. First, the ghost orbitals of B to be removed from CP virtual space are constrained to be orthogonal to the occupied orbitals of A as they must in the supermolecule. Second, it is a simple matter to return the excised functions to the virtual space and thereby produce a MO basis that spans the same space as the original FCP MO basis. In addition, there is no ambiguity about what constitutes the ‘missing’ orbital—it is identically the SCF function corresponding to the occupied ghost orbital(s) in the supermolecule. Finally, the vVCP virtual space is more similar to the FCP virtual space than that produced using the standard VCP method.

## Results

As indicated in Table 1 and Figures 1A, 2A, and 3A, both our uncorrected and FCP-corrected results for the helium dimer are consistent with those of numerous previous studies. The current state-of-the-art theoretical and available experimental results indicate a  $He_2$  interaction energy of  $\sim 11$  K at an internuclear distance of  $5.6 a_0$ , values we approach using the more extensive basis sets.<sup>20–31</sup> Indeed, for the aug\_cc-pVXZ ( $X=T, Q, 5$ ) basis sets used here, the uncorrected as

**Table 1.** He<sub>2</sub> Interaction Energies (in K) Calculated Using aug\_cc-pVXZ (X=D,T,Q,5) Basis Sets with No BSSE Correction Scheme, the Full and Virtuals-Only Counterpoise Correction (FCP, VCP), and the Variant Virtuals-Only Counterpoise Correction (vVCP) Described Herein<sup>a</sup>

| basis set | MP2                           | CCSD                          | CCSD(T)                       |
|-----------|-------------------------------|-------------------------------|-------------------------------|
| <b>DZ</b> |                               |                               |                               |
| no CP     | 11.45 K @ 5.74 a <sub>0</sub> | 12.56 K @ 5.69 a <sub>0</sub> | 13.37 K @ 5.69 a <sub>0</sub> |
| FCP       | 4.01 K @ 6.18 a <sub>0</sub>  | 5.30 K @ 6.07 a <sub>0</sub>  | 5.86 K @ 6.01 a <sub>0</sub>  |
| vVCP      | n/a                           | 7.68 K @ 5.69 a <sub>0</sub>  | 8.51 @ 5.65 a <sub>0</sub>    |
| <b>TZ</b> |                               |                               |                               |
| no CP     | 6.99 K @ 5.84 a <sub>0</sub>  | 8.82 K @ 5.74 a <sub>0</sub>  | 9.92 K @ 5.72 a <sub>0</sub>  |
| FCP       | 5.62 K @ 5.89 a <sub>0</sub>  | 7.42 K @ 5.79 a <sub>0</sub>  | 8.45 K @ 5.74 a <sub>0</sub>  |
| VCP       |                               | 8.17 K @ 5.68 a <sub>0</sub>  | 9.48 K @ 5.62 a <sub>0</sub>  |
| vVCP      | n/a                           | 11.60 K @ 5.42 a <sub>0</sub> | 13.19 K @ 5.37 a <sub>0</sub> |
| <b>QZ</b> |                               |                               |                               |
| no CP     | 6.85 K @ 5.84 a <sub>0</sub>  | 8.78 K @ 5.72 a <sub>0</sub>  | 10.03 K @ 5.67 a <sub>0</sub> |
| FCP       | 6.25 K @ 5.84 a <sub>0</sub>  | 8.20 K @ 5.72 a <sub>0</sub>  | 9.44 K @ 5.67 a <sub>0</sub>  |
| vVCP      | n/a                           | 13.11 K @ 5.37 a <sub>0</sub> | 14.99 K @ 5.32 a <sub>0</sub> |
| <b>5Z</b> |                               |                               |                               |
| no CP     | 6.89 K @ 5.79 a <sub>0</sub>  | 8.88 K @ 5.72 a <sub>0</sub>  | 10.20 K @ 5.67 a <sub>0</sub> |
| FCP       | 6.62 K @ 5.79 a <sub>0</sub>  | 8.68 K @ 5.72 a <sub>0</sub>  | 10.00 K @ 5.67 a <sub>0</sub> |
| vVCP      | n/a                           | 13.84 K @ 5.32 a <sub>0</sub> | 15.97 K @ 5.27 a <sub>0</sub> |

<sup>a</sup> The data reported correspond to the minima for each basis set and computational method among the He–He distances sampled. Note: by construction, the vVCP scf energy is identical to the FCP scf energy.

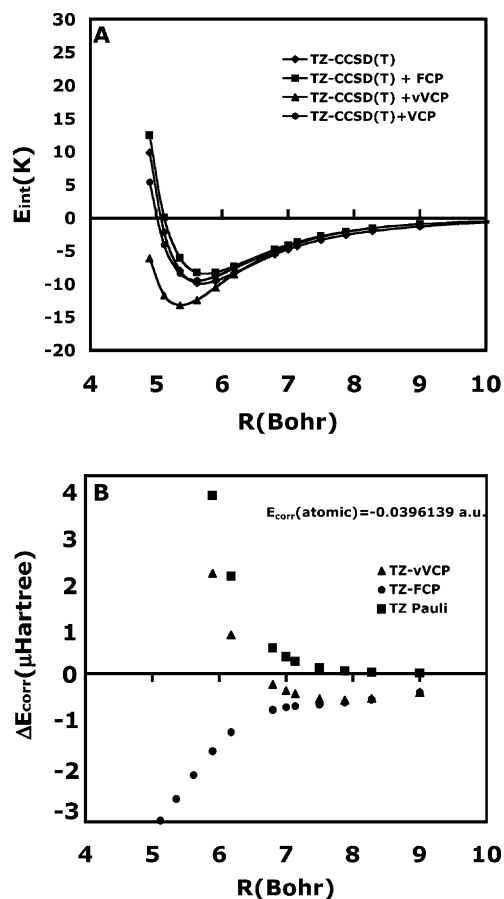
well as the FCP-corrected (and VCP-corrected) well depths agree reasonably well with these available data despite the fact that even the aug\_cc-pV5Z basis set is evidently not sufficient to fully capture the exquisitely weak interaction in the helium dimer. Numerous excellent studies of the helium dimer potential energy surface have been published which employ basis sets of aug\_cc-pV6Z quality or better several of which include midbond functions. Though results of such studies and especially recent quantum Monte Carlo calculations<sup>24,26,27</sup> more closely approximate the basis set limit and the experimental result than do ours, our focus is to examine the role of the virtual orbitals and their relationship to BSSE and the Pauli repulsion, and therefore these differences are immaterial.<sup>20–31</sup>

As expected, use of the FCP procedure to correct BSSE leads to shallower potential energy surfaces compared to uncorrected calculations, though the room for improvement and therefore the degree of that improvement is much smaller for the larger basis sets. And, as expected, the He<sub>2</sub> PES corrected using Daudey's VCP method is deeper than the corresponding FCP surface and shallower than the uncorrected one. This indicates that notwithstanding questions of its usefulness, the smaller VCP orbital space does attenuate the counterpoise correction. What is striking is the remarkable difference that the vVCP results give with respect to those of other counterpoise methods. Not only are the vVCP potential energy surfaces in each case deeper by as much as half (e.g., 15.9 K vs 11 K) than the current best estimates but also the minima occur at much shorter internuclear distances (~5.4 a<sub>0</sub> vs 5.6 a<sub>0</sub>).<sup>20</sup>

Figures 1B, 2B, and 3B illustrate the variation of helium CCSD(=CISD) FCP and vVCP counterpoise corrections to the correlation energy relative to the pure atomic result as a function of internuclear separation (strictly, monomer-ghost separation) for, respectively, the augmented TZ, QZ, and 5Z basis sets. This analysis reveals the source of the unantici-

pated vVCP well deepening. As expected and in keeping with the variational principle, the FCP correction calculated for each basis set is negative at all monomer-ghost distances indicating some degree of improvement in the description of the monomer electronic structure by the presence of virtual orbitals located on the ghost. Also according to expectations, the magnitude of the FCP correction at each internuclear separation decreases with increasing quality of atomic basis set, and at the same time the calculated correlation energy asymptotically approaches the pure atomic result as the value of R increases. All of this explains why the FCP corrected potential energy surfaces for all basis sets are to varying degrees shallower than the uncorrected ones. However, particularly for the aug\_cc-pV5Z basis, the correction is minimal, and the uncorrected and FCP-corrected potential energy surfaces are nearly coincident.

In contrast, at many atom-ghost distances the vVCP monomer correlation energies are actually *higher* (less negative) than their corresponding pure atomic results, a finding that is actually exacerbated for larger basis sets which are presumably closer to completeness. While at first this may seem a violation of the variation principle (which applies here because of the equivalence of CCSD and CISD for two-electron systems) it is not. Unlike Daudey's VCP procedure, the vVCP virtual space is not simply an augmentation of the set of monomer by virtual ghost orbitals but rather consists of the standard FCP virtual space with one function, identical to the ghost He 1s orbital of the noninteracting partner, projected out of it. So, the structure of the vVCP virtual space more closely resembles that of the dimer in that orthogonality of the 'core' 1s electrons on each monomer (or, if you will monomer and ghost) is preserved. Therefore, from the perspective of an atomic calculation, though the (AUB)<sub>vVCP</sub> supermolecular orbital basis set may be 'larger' than the pure AO basis set, it is not necessarily 'better' than the monomer basis in its capacity to minimize calculated

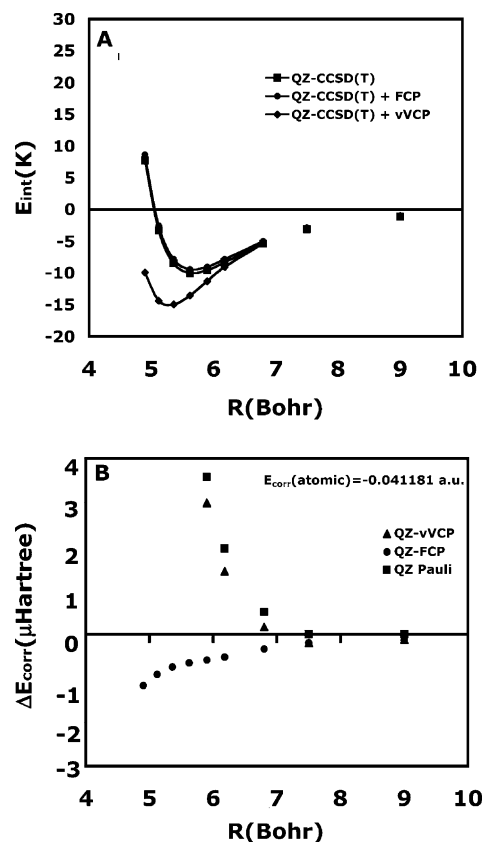


**Figure 1.** (A) CCSD(T) potential energy surfaces for the Helium dimer calculated with the aug\_cc-pVTZ basis sets using no BSSE correction; and the full (FCP), virtuals-only (VCP), and variant virtuals-only (vVCP) counterpoise procedures. (B) FCP and vVCP counterpoise corrections relative to the atomic He CCSD(=CISD) correlation energy calculated using the aug\_cc-pVTZ basis set.  $\Delta E_{\text{corr}} = E_{\text{XCP}} - E_{\text{atom}}$ .

total atomic energies. In any case, this ‘negative’ counterpoise correction calculated with the vVCP procedure at short distances tends to deepen and contract the calculated potential energy surfaces.

Some additional points merit mention here. First, we note that as the monomer-ghost distance increases the vVCP correction energy does change sign and once again becomes positive, though less so than for the FCP correction because of the smaller virtual space. This is best illustrated in Figure 1B though it is qualitatively similar in Figures 2B and 3B (the CP correction here is small compared to the scale of the graph). And as it should, the asymptotic behavior of the vVCP correction resembles and becomes identical to the FCP results as the monomer-ghost distance increases. Finally, in all cases we have found that if the ‘excised’ orbital(s) in the vVCP case is simply reappended to the virtual space the corresponding FCP result is fully recovered.

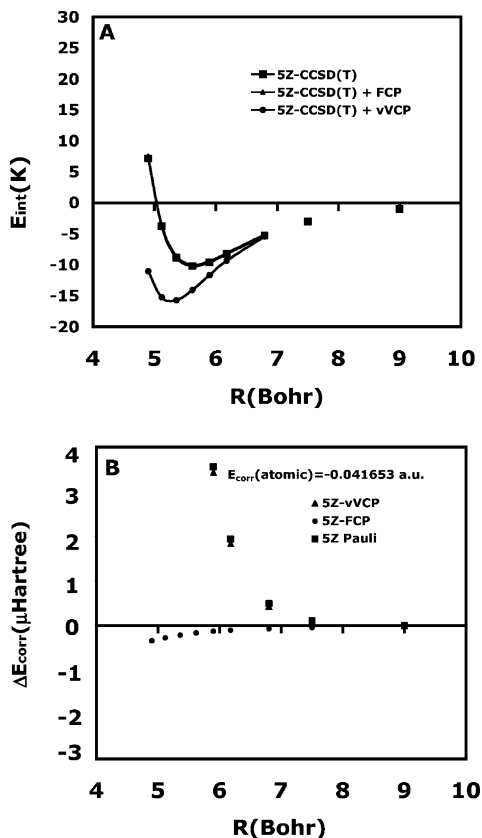
It is clear that the anomalous counterpoise correction calculated at the CCSD(=CISD) level using the vVCP method cannot be reasonably construed as the result of a basis set superposition error. Instead, the data suggest that the degradation of monomer electron correlation in a dimer, independent of any electrostatic or bonding interaction, may



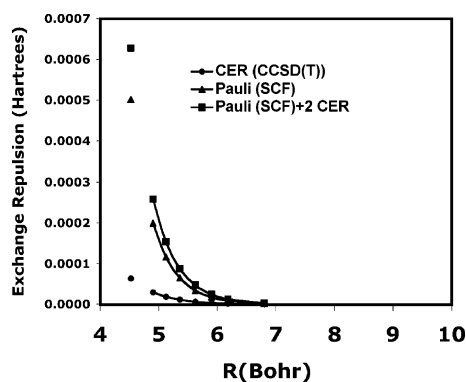
**Figure 2.** (A) CCSD(T) potential energy surfaces for the helium dimer calculated with the aug\_cc-pVQZ basis sets using no BSSE correction; and the full (FCP), and variant virtuals-only (vVCP) counterpoise procedures. (B) FCP and vVCP counterpoise corrections relative to the atomic He CCSD(=CISD) correlation energy calculated using the aug\_cc-pVQZ basis set.  $\Delta E_{\text{corr}} = E_{\text{XCP}} - E_{\text{atom}}$ .

be interpreted as an effective repulsion which we conclude corresponds to the exchange (Pauli) repulsion calculated at the correlated level (the correlation exchange repulsion, CER). This interpretation follows from the fact that the disruption of monomer correlation is the result of the mere presence of ‘occupied’ (i.e. excluded) orbitals on the ghost partner. It is clear that this repulsive energy component is also present in a true dimer calculation, and, therefore, the removal of this important repulsive energy component according to the vVCP method produces the anomalously deep and compact potential energy well. Any basis set superposition error would tend to mitigate the effect of removing this correlation exchange repulsion, which explains why our seemingly anomalous vVCP results become more pronounced with more complete basis sets. This monomer correlation exchange repulsion is also depicted in Figures 1B, 2B, and 3B for each basis set. The curves labeled XZ Pauli (X=T,Q,5) are generated by correcting the corresponding vVCP curve for BSSE by subtracting the FCP counterpoise correction.

Figure 4 depicts a comparison of the He<sub>2</sub> SCF exchange repulsion calculated with the calculated monomer correlation exchange repulsion (the latter corrected for BSSE). The total exchange repulsion is taken as the sum of the SCF exchange repulsion and two times the monomer correlation exchange repulsion. Importantly, the correlation exchange repulsion



**Figure 3.** (A) CCSD(T) potential energy surfaces for the helium dimer calculated with the aug\_cc-pV5Z basis sets using no BSSE correction; and the full (FCP), and variant virtuals-only (vVCP) counterpoise procedures. (B) FCP and vVCP counterpoise corrections relative to the atomic He CCSD(=CISD) correlation energy calculated using the aug\_cc-pV5Z basis set.  $\Delta E_{\text{corr}} = E_{\text{XCP}} - E_{\text{atom}}$ .



**Figure 4.** The helium dimer exchange repulsion at the SCF level, Pauli (SCF), and the effective monomer correlation exchange repulsion (CER) calculated at the CCSD(=CISD) level. Also depicted is the total exchange repulsion. The data correspond to values calculated using the aug\_cc-pVQZ basis set.

exhibits qualitatively the same dependence on monomer-ghost distance as the SCF exchange repulsion. Finally, the fact that our observation of this phenomenon, which is not per se a basis set superposition error, does not decrease but rather increases with larger basis sets suggests that basis set incompleteness is not the issue here.

## Conclusions

The formal arguments advanced by Gutowski et al., to support the use of the full counterpoise procedure hold among other things that it, in contrast to virtuals-only approaches, permits a proper description of the electron density in the internuclear (intermonomer) region. Because of this, use of any virtuals-only counterpoise method necessarily degrades a description of exchange repulsion.<sup>10–14</sup> While our results corroborate this finding, we should emphasize that our interpretation of the reason for this underestimation is somewhat different. From our perspective, exchange repulsion at the correlated level is manifested as a disruption of the correlation energy of the monomer calculated in the dimer (supermolecule) molecular orbital basis set relative to that calculated in only the monomer's own basis set. By performing a correlated monomer calculation according to the vVCP prescription and subtracting the result from the supermolecular PES, we actually remove this essential component of the interaction energy. Daudey's VCP method, or any other technique that merely supplements the monomer virtual space with additional functions, approximates this effect and therefore does lead to an understatement of the exchange repulsion. Finally, use of the vVCP method directly illuminates the role that the 'occupied' partner orbitals play and permits a full evaluation of the exchange repulsion at the correlated level.

## References

- (1) Mayer, I. *Int. J. Quantum Chem.* **1993**, *23*, 341–363.
- (2) Jeziorski, B.; Moszynski, B.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.
- (3) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- (4) Daudey, J. P.; Claverie, P.; Malrieu, J. P. *Int. J. Quantum Chem.* **1974**, *8*, 1–15.
- (5) Johansson, A.; Kollman, P.; Rothenberg, S. *Theor. Chim. Acta* **1973**, *29*, 167.
- (6) Cook, D. B.; Sordo, J. A.; Sordo, T. L. *Int. J. Quantum Chem.* **1993**, *48*, 375–384.
- (7) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *Chem. Rev.* **1994**, *94*, 1887.
- (8) Collins, J. R.; Gallup, G. A. *Chem. Phys. Lett.* **1986**, *123*, 56–61.
- (9) Collins, J. R.; Gallup, G. A. *Chem. Phys. Lett.* **1986**, *129*, 329–330.
- (10) Gutowski, M.; van Lenthe, J. H.; Verbeek, J.; van Duijneveldt, F. B.; Chalinski, G. *Chem. Phys. Lett.* **1986**, *124*, 370–375.
- (11) Gutowski, M.; van Duijneveldt, F. B.; Chalasinski, G.; Piela, L. *Chem. Phys. Lett.* **1986**, *129*, 325–328.
- (12) van Mourik, T.; Wilson, A. K.; Peterson, K. A.; Woon, D. E.; Dunning, T. H. *Adv. Quantum Chem.* **1999**, *31*, 105–135.
- (13) Gutowski, M.; Szczśniak, M. M.; Chalasinski, G. *Chem. Phys. Lett.* **1995**, *241*, 140–145.
- (14) Gutowski, M.; Chalasinski, G. *J. Chem. Phys.* **1993**, *98*, 5540–5555.
- (15) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873.

- (16) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J.; Koseki, S.; Masunaga, N.; Nguyen, K.A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (17) Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 02/25/04, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, Washington 9932, USA, and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multiprogram laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RLO 1830. Contact David Feller for Karen Schuchardt for further information.
- (18) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 2975–2988.
- (19) Löwdin, P. O. *J. Chem. Phys.* **1950**, *18*, 435.
- (20) Tang, K. T.; Toennis, J. P. *J. Chem. Phys.* **2003**, *118*, 4976–4983.
- (21) Giese, T. J.; Audette, V. M.; York, D. M. *J. Chem. Phys.* **2003**, *119*, 2618–2622.
- (22) Gdanitz, R. *J. Mol. Phys.* **2001**, *99*, 923–930.
- (23) Gdanitz, R. *J. Mol. Phys.* **1999**, *96*, 1423–1434.
- (24) Anderson, J. B.; Traynor, C. A.; Boghosian, B. M. *J. Chem. Phys.* **1993**, *99*, 345–351.
- (25) Mella, M.; Anderson, J. B. *J. Chem. Phys.* **2003**, *119*, 8225–8228.
- (26) Anderson, J. B. *J. Chem. Phys.* **2001**, *115*, 4546–4548.
- (27) Anderson, J. B. *J. Chem. Phys.* **2004**, *120*, 9886–9887.
- (28) van Mourik, T.; Dunning, T. H. *J. Chem. Phys.* **1999**, *111*, 9248–9258.
- (29) Laschuk, E. F.; Martins, M. M.; Evangelisti, S. *Int. J. Quantum Chem.* **2003**, *95*, 303–312.
- (30) Korona, T.; Williams, H. L.; Bukowski, R.; Jeziorski, B.; Szalewicz, K. Helium Dimer Potential from Symmetry-Adapted Perturbation Theory Calculations Using Large Gaussian Geminal and Orbital Basis Sets. *J. Chem. Phys.* **1997**, *106*, 5109–5122.
- (31) Janzen, A. R.; Aziz, R. A. *J. Chem. Phys.* **1997**, *107*, 914–919.

CT049871F

## Functional Group Basis Sets

Benjamin G. Janesko and David Yaron\*

*Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213*

Received December 14, 2004

**Abstract:** The electronic properties of molecular functional groups (methyl, benzyl, etc.) are generally conserved across different molecules. However, atomic basis sets use the same electronic variational space on an element regardless of the functional group it is in. The computational expense of ab initio calculations on molecules may be reduced by constructing functional-group-specific basis sets. Our functional group (FG) basis functions are contractions over a “parent” atomic basis on all atoms in the functional group of interest. Contraction coefficients are obtained by calculating the functional group’s electron density in a set of representative molecules and performing principle component analysis of the results. When the functional group’s chemical identity is maintained (e.g. no bonds are broken in the group), FG basis sets provide accuracies similar to the parent basis while using a smaller number of basis functions. We demonstrate our methods by parametrizing FG basis sets for the –OH and –CH– functional groups and testing them in DFT calculations on several molecules. The results suggest that FG basis sets may be useful in many contexts, especially for treating spectator groups in mixed basis calculations.

### 1. Introduction

Ab initio quantum chemistry is practical because of basis sets, which turn the integro-differential Schrodinger’s equation into a linear algebra problem over a finite variational space.<sup>1</sup>

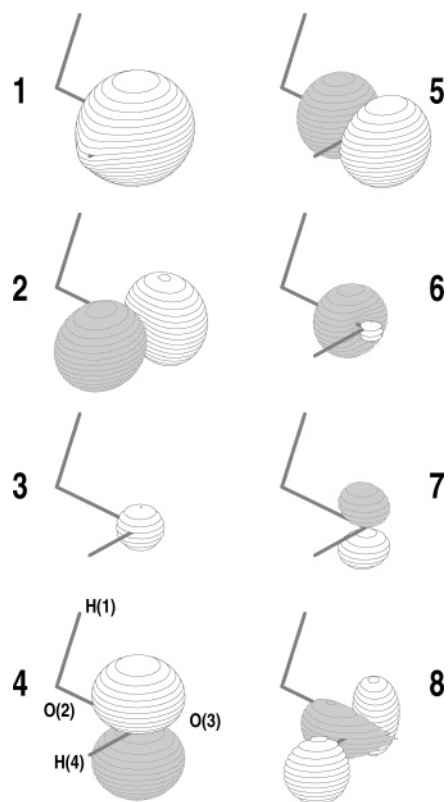
A primary challenge to modern theoretical chemistry is accurately treating large systems. Accuracy requires both a large variational space and high-level approximations for effects such as electron correlation. Unfortunately, large variational spaces are computationally expensive. Canonical ab initio calculations in an  $N$ -orbital variational space scale from  $\mathcal{O}(N^2) - \mathcal{O}(N^3)$  for Hartree–Fock and DFT methods<sup>2</sup> to  $\mathcal{O}(e^N)$  for full inclusion of electron correlation. Several authors have developed physically motivated  $\mathcal{O}(N)$  approximations in order to reduce this expense.<sup>2–5</sup> Here we develop a complimentary approach: physically motivated reductions of the variational space itself.

We suggest that exploiting molecular similarity can reduce the variational spaces needed for molecular calculations. “Molecular similarity” is the familiar assumption that the

properties of molecular functional groups (methyl, benzyl, hydroxy, etc.) are similar in different molecules. It is a natural consequence of the “nearsightedness” of electronic structure. Nearsightedness, which underlies many of the  $\mathcal{O}(N)$  approximations cited above, is the observation that distant regions of a large molecule have a weak influence on each other.<sup>6</sup> Molecular similarity implies that electrons in a particular functional group will explore a well-defined variational space. The relevant variational spaces will differ in different functional groups. Molecular similarity is often a good approximation and underlies the success of molecular mechanics methods.<sup>7</sup> However, existing atomic orbital (AO) basis sets do not use this approximation. Such basis sets treat each element with the same variational space, independent of what functional group it is in.

We parametrize basis sets to describe the most important components of a particular functional group’s electronic variational space. In systems where the assumption of molecular similarity holds, these functional group (FG) basis sets should provide accuracies comparable to AO basis sets, while leading to a significant dimensional reduction. They will not work where molecular similarity breaks down, e.g. where a bond breaks in a functional group.

\* Corresponding author e-mail: yaron@cmu.edu.



**Figure 1.** Surface plots of the first eight  $-OH$  group basis functions obtained from the FG(HOOH,HOH) training set (Table 1). Contour is at 0.15 au. The basis functions are shown on one OH group of an  $H_2O_2$  molecule (gray lines).

Functional group basis functions are obtained as contractions over a “parent” atomic orbital basis. Let  $\{\phi_i^A\}$  be the set of parent AO basis functions located on atom  $A$  and indexed by  $i$ . FG basis functions for a group  $F$  ( $\Phi_j^F$ ) are obtained from the parent basis functions on all atoms  $A$  in  $F$ :

$$\Phi_j^F = \sum_{A \in F} \sum_i C_{j,i}^{F,A} \phi_i^A \quad (1)$$

For example, our  $-OH$  group basis functions (section 3) are linear combinations of the 42 6-311G++(3d,p) AO basis functions on oxygen and hydrogen. Figure 1 plots<sup>8</sup> the first eight functions from one of our  $-OH$  group basis sets.

In quantum chemistry, basis functions are typically constructed from the eigenfunctions of simple systems such as isolated atoms. We use a very different approach to parametrize FG basis sets. This approach entails calculating the electronic structure of a functional group in many different environments and extracting the electronic degrees of freedom (basis functions) that are most important in these environments. This process is referred to as feature extraction.<sup>9</sup> To construct an FG basis for functional group  $F$ , we build a training set of  $F$ -containing molecules and perform ab initio calculations in the parent basis set on each molecule. These molecules are designed to (a) contain sufficient diversity to explore the behavior of  $F$  in all molecules of interest and (b) be small enough to permit ab initio calculations in the parent basis set. For example, an FG basis set for amino acid  $A$  in proteins might be trained on  $R -$

$A - R'$  trimers, with  $R$  and  $R'$  selected from the 20 biologically important amino acids. We add diversity to the training sets by applying random geometric and electrostatic perturbations to each molecule.

We next use principle component analysis (PCA) to extract the most important electronic degrees of freedom from the training set calculations. Consider a set of data vectors  $\{\nu_i\}$  that are defined in a variational space of dimension  $N_0$ . PCA of  $\{\nu_i\}$  returns  $N_0$  basis vectors, ordered according to their importance in describing the training set.<sup>9</sup> In our case,  $N_0$  is the number of parent AO basis functions on a group  $F$  and  $\{\nu_i\}$  come from  $F$ 's calculated electron density in the training set molecules (see section 7 for details). The first  $N \leq N_0$  vectors returned by PCA can be used as an  $N$ -orbital FG basis set. For example, Figure 1 shows the eight most important  $-OH$  group basis functions obtained from PCA of 42 parent basis functions. We can use the first 6 of these functions to give an  $-OH$  basis set of minimal-basis size, the first 11 functions to give an  $-OH$  basis of valence-double- $\zeta$  size, and so forth.

The remainder of this paper is as follows. Section 2 presents a discussion of related work. Section 3 presents information on the parametrization and testing of FG basis sets for two functional groups,  $-OH$  and  $-CH-$ . The results from section 3 are discussed in section 4, with a focus on the benefits and limitations of the approach. Section 5 demonstrates the potential utility of this approach for hybrid electronic structure calculations. Section 6 gives the conclusions, and section 7 presents details of the method and implementation.

## 2. Related Work

To our knowledge, the current work is unique in constructing a transferable set of functional-group-specific basis functions by analyzing variation across a set of molecules. Some related methods attempt to transfer functional groups' electronic structures (electron densities<sup>10–13</sup> or localized, occupied molecular orbitals<sup>14–19</sup>) between different calculations. One difficulty faced by these methods is that the transferred electronic structures should be both localized to a functional group and orthogonal to the rest of the system.<sup>20</sup> Ab initio valence bond calculations<sup>21</sup> can use nonorthogonal occupied orbitals, but the complexity of the resulting Hamiltonians limits this method's generality. Approximately orthonormal monomer orbitals can be used for calculations<sup>22–24</sup> and interpretations<sup>25–27</sup> of intermolecular interactions. Orthogonality can also be treated by transferring occupied orbital energies<sup>28–30</sup> or localized properties<sup>31</sup> rather than the orbitals themselves or by constructing transferable models of functional groups from ab initio densities.<sup>32,33</sup> Finally, some methods treat orthogonality issues by using the transferred orbitals as a basis set for SCF calculations.<sup>19,24</sup> Our methods differ from these in that we focus entirely on transferable variational spaces and in our parametrization to a set of representative molecules. In our view, this is a promising alternative to the complexities inherent in the wholesale transfer of electronic structures between calculations. Other relevant reduced-variational-space methods include dual basis set methods,<sup>34,35</sup> polarized atomic orbit-



**Table 1:** Data Sets Used To Train the  $-\text{OH}$  and  $-\text{CH}-$  Functional Group Basis Sets

| data set             | molecule(s)                                     | groups <sup>a</sup> | perturbations     |
|----------------------|---|---------------------|-------------------|
| FG(HOH)              | H <sub>2</sub> O                                | 320                 | Rand <sup>b</sup> |
| FG(HOOH)             | H <sub>2</sub> O <sub>2</sub>                   | 320                 | Rand              |
| FG(HOOH,HOH)         | H <sub>2</sub> O, H <sub>2</sub> O <sub>2</sub> | 320                 | Rand              |
| FG(ROH) <sup>c</sup> | R-OH  | 5                   | none              |
| propene              | propene   | 200                 |                   |

<sup>a</sup> Number of  $-\text{OH}$  or  $-\text{CH}-$  functional groups in the training set. Both  $-\text{OH}$  groups are used from each H<sub>2</sub>O<sub>2</sub> molecule. <sup>b</sup> A unique set of random electrostatic and geometric perturbations is applied to each molecule. Geometric perturbations vary bond lengths, angles, and dihedrals by  $\leq 0.2$  Å,  $\leq 3$  degrees, and  $\leq 360$  degrees, respectively. Electrostatic perturbations randomly place 10 fractional charges in a 14.0 Å square box around the molecule, with  $\geq 0.7$  Å charge-atom separations. <sup>c</sup> Trained on R-OH molecules with R = H, F, OH, COH, O<sup>-</sup>, and Li; at their RHF/6-31G optimized geometries.

als,<sup>36,37</sup> molecule intrinsic minimal basis sets,<sup>38</sup> and mixed basis sets (see section 5).

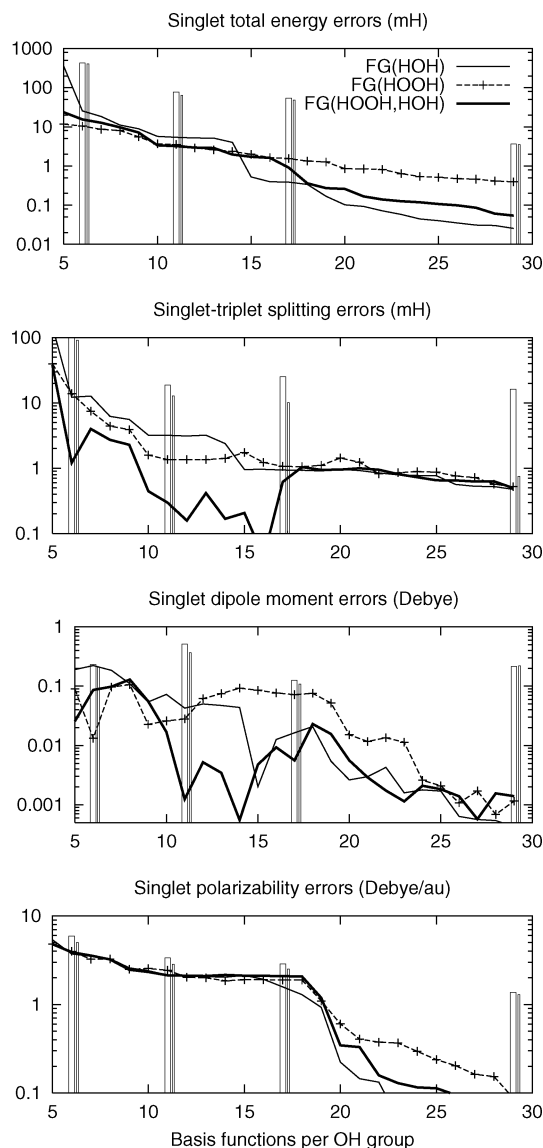
Our PCA-based parametrization differs considerably from the standard methods for generating basis sets. As mentioned above, electronic basis functions are typically constructed from the eigenfunctions of simpler systems such as isolated atoms.<sup>39-43</sup> Exceptions include molecular fragment<sup>44</sup> and atomic natural orbital<sup>45</sup> AO basis sets.

Other uses of PCA in computational chemistry include PCA of ab initio calculations to predict the crystal structures of binary metallic alloys<sup>46</sup> and PCA of molecular dynamics trajectories to determine condensed-phase vibrational spectra.<sup>47,48</sup> PCA has also been extensively applied in analysis of experimental data.<sup>49,50</sup> We have previously used PCA of calculated electron densities in constructing a functional group based model for electron correlation.<sup>51</sup>

### 3. Results

**3.1. Parametrization.** We parametrized functional group basis sets for the  $-\text{OH}$  functional group to singlet and triplet B3LYP/6-311G++(3d,p) calculations on the first four training sets in Table 1. The 6-311G++(3d,p) parent basis has 42 basis functions per  $-\text{OH}$  group. FG basis sets for  $-\text{CH}-$  were parametrized to singlet and triplet B3LYP/6-31G(d) calculations on the last training set in Table 1. The 6-31G(d) parent basis has 17 basis functions per  $-\text{CH}-$  group.  $-\text{CH}-$  training sets included both orientations of the  $-\text{CH}-$  group ( $\text{CH}_2 = \text{CH} - \text{CH}_3$  and  $\text{CH}_3 - \text{CH} = \text{CH}_2$ ) to avoid bias in treating the substituents. Details of the parametrization procedure are discussed in section 7. Unperturbed geometries are from experimental data.<sup>52</sup> Systematic analysis of FG basis set properties as a function of the parametrization variables (training set size and composition, thresholds in selecting vectors for PCA, etc.) is left for later work.

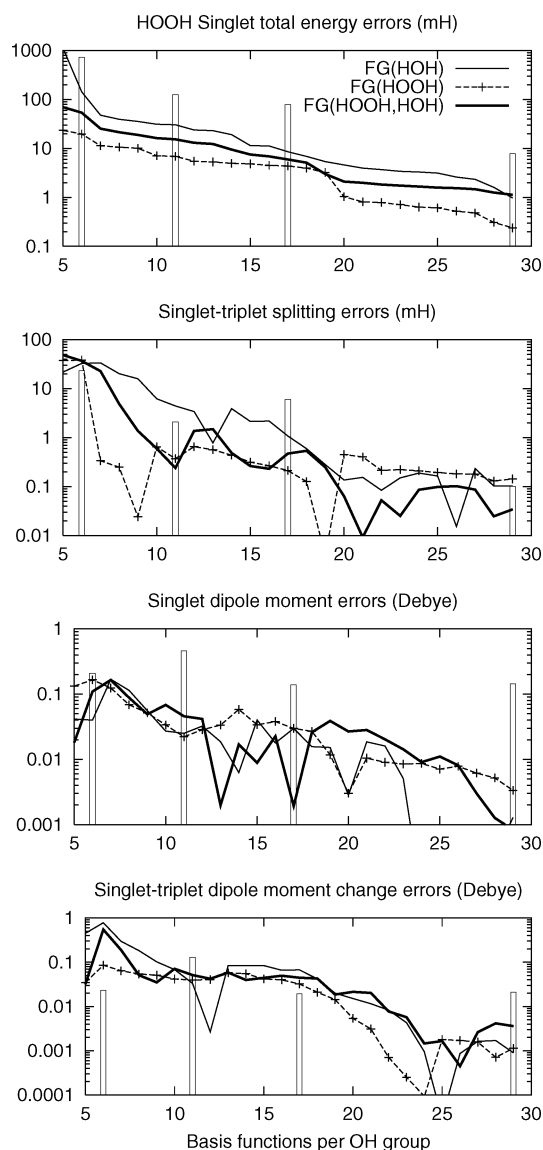
**3.2. Testing.** The  $-\text{OH}$  and  $-\text{CH}-$  functional group basis sets were tested by performing B3LYP/(FG basis) calculations on various molecules and comparing the results to B3LYP in the parent basis. Calculations for functional group  $F$  ( $-\text{OH}$  or  $-\text{CH}-$ ) used the first  $N$  FG basis functions on each  $F$  group and the parent basis on the remainder of the molecule. (For example, the "FG(HOH)" points in Figure 2 are B3LYP calculations on H<sub>2</sub>O where one  $-\text{OH}$  group is



**Figure 2.** Errors relative to B3LYP/6-311G++(3d,p) in the energy and properties of unperturbed H<sub>2</sub>O. Results are plotted vs the number of basis functions on the  $-\text{OH}$  group. Lines are the results of B3LYP calculations that use an FG basis on the  $-\text{OH}$  group and the parent basis on the remainder (see section 3.2). Columns are the results of B3LYP calculations that use small AO basis sets on the  $-\text{OH}$  group: 6-311G+(d,p) at 29, 6-31G(d) at 17, 6-31G at 11, and STO-6G at 6 basis functions per  $-\text{OH}$  group. Narrow and wide columns correspond to treating the rest of the molecule with the small AO basis (uniform basis set) or the parent basis (mixed basis set), respectively. The FG basis sets are obtained from the training sets in Table 1.

treated using the first 5, 6, 7, etc.  $-\text{OH}$  basis functions obtained from the FG(HOH) training set and the remaining H atom is treated using 6-311G++(3d,p).) The results are compared to calculations in smaller AO basis sets as well as hybrid-basis calculations that use a small AO basis on all  $F$  groups and the parent basis on the remainder. The results of these tests are listed below and discussed in section 4.

The  $-\text{OH}$  basis sets were tested on H<sub>2</sub>O and H<sub>2</sub>O<sub>2</sub> in their equilibrium geometries (Figures 2 and 3), on the dihedral rotational barrier of H<sub>2</sub>O<sub>2</sub> (Table 2 and Figure 6), and on 90



**Figure 3.** Errors in the energy and properties of unperturbed  $\text{H}_2\text{O}_2$ , plotted as in Figure 2. The same basis set (FG or small AO) is used for both  $-\text{OH}$  groups.

$\text{H}_2\text{O}_2$  molecules with perturbations as in Table 1 (Figure 4). The  $\text{H}_2\text{O}_2$  molecules in Figure 4 were not included in the FG(HOOH) training set. The results in Figure 4 are similar to those for perturbed  $\text{H}_2\text{O}$  (data not shown). Basis functions from the FG(HOOH,HOH) and FG(ROH) training sets were tested in calculations on a set of eight R-OH molecules (Table 3). FG(HOOH,HOH) was also tested on the potential energy surfaces for the water O-H bond stretch (Figure 7). Figure 5 shows the highest occupied molecular orbitals (MOs) from FG and AO basis calculations on triplet  $\text{H}_2\text{O}$ . All errors are calculated versus B3LYP/6-311G++(3d,p) and plotted vs the number of basis functions per  $-\text{OH}$  group.

The  $-\text{CH}-$  basis sets were tested on 43 propene molecules with perturbations as in Table 1 (Figure 8) and on unperturbed benzene (Figure 9). The molecules in Figure 8 were not in the propene training set. Errors are calculated relative to B3LYP/6-31G(d) and plotted vs the number of basis functions per  $-\text{CH}-$  group.

**Table 2:** B3LYP Rotational Barriers and Minimum-Energy Dihedral Angle for the Dihedral Rotation of Singlet  $\text{H}_2\text{O}_2$

| $N^a$ | basis set      | rotational barrier (mH) | minimum dihedral (deg) <sup>b</sup> |
|-------|----------------|-------------------------|-------------------------------------|
| 42    | 6-311G++(3d,p) | 17.19                   | 121                                 |
| 29    | 6-311G+(d,p)   | 18.83                   | 131                                 |
| 29    | FG(HOH)        | 16.96                   | 121                                 |
| 29    | FG(HOOH)       | 17.22                   | 121                                 |
| 17    | 6-31G(d)       | 19.86                   | 131                                 |
| 17    | FG(HOH)        | 17.06                   | 121                                 |
| 17    | FG(HOOH)       | 17.40                   | 127                                 |
| 11    | 6-31G          | 27.65                   | 178                                 |
| 11    | FG(HOH)        | 16.89                   | 127                                 |
| 11    | FG(HOOH)       | 17.10                   | 140                                 |
| 6     | STO-6G         | 19.03                   | 180                                 |
| 6     | FG(HOH)        | 16.95                   | 127                                 |
| 6     | FG(HOOH)       | 18.27                   | 131                                 |

<sup>a</sup> Number of basis functions per  $-\text{OH}$  group. <sup>b</sup> Sampled in 1-degree increments, with other geometry parameters frozen at equilibrium values.

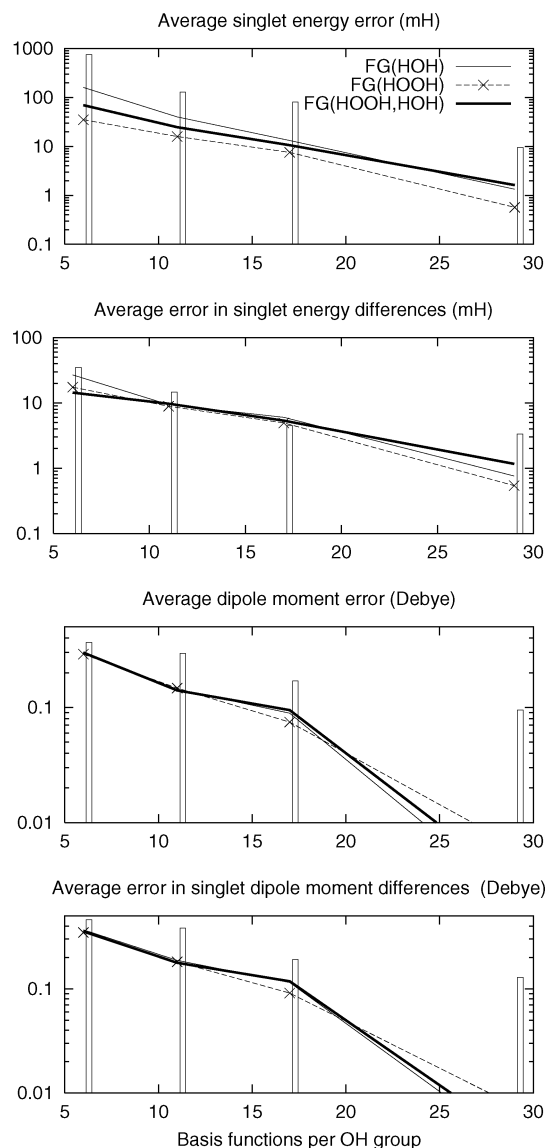
An alternative to extracting basis functions from molecules in different environments is to extract them from single molecules. Table 3 contains results from two such basis sets. The “One” results are an  $-\text{OH}$  functional group basis set that was parametrized to B3LYP/6-311G++(3d,p) singlet and triplet calculations on a single unperturbed  $\text{H}_2\text{O}$  molecule (section 7.1). The “Orb” results use  $-\text{OH}$  basis functions obtained from the Kohn–Sham orbitals of a B3LYP/6-311G++(3d,p) calculation on singlet  $\text{H}_2\text{O}$ . Here, the orbitals are sorted by orbital energy, projected onto one OH group of  $\text{H}_2\text{O}$ , and renormalized before use. Results from the latter set of basis functions are included in Figure 5 as “ $\text{H}_2\text{O}$  Orb”.

## 4. Discussion

This section considers the results in section 3 and argues that the functional group basis sets can treat a wide range of systems at useful levels of theory. The results highlight the potential utility of FG basis sets and suggest further work.

The most important result in section 3 is that small functional group basis sets can reliably reproduce the parent basis results. Calculations using 29 of the 42  $-\text{OH}$  group basis functions always return small errors in total energies, singlet–triplet splittings, rotational barriers, and dipole moments (energy errors  $< 1.5$  mH, dipole errors  $< 0.01$  D for results in Figures 2, 3, 4 and 6 and Tables 2 and 3). Calculations using 12 of the 17  $-\text{CH}-$  group basis functions also return small errors (Figure 8).

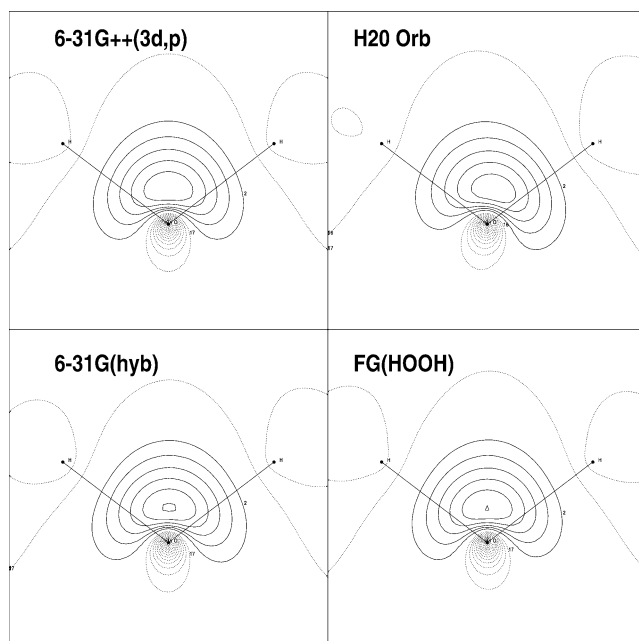
Another important result is that the first  $N$  functional group basis functions from a particular training set are generally better than an  $N$ -orbital AO basis set at reproducing the parent basis results. The FG basis sets are better than AO basis sets at reproducing the parent basis total energy (upper panels of Figures 2, 3, 4, 8, and 9, and first section of Table 3). The FG basis sets are also better than AO basis sets at treating most of the properties tested. Notable examples include the polarizability (Figure 2) of unperturbed  $\text{H}_2\text{O}$ , the energies and properties of R-OH molecules (Table 3), and the average properties of perturbed molecules (Figures 4 and



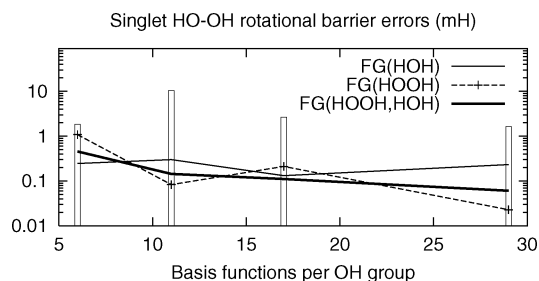
**Figure 4.** Average errors, relative to B3LYP/6-31G++-(3d,p), in the energies and properties of 90 perturbed singlet  $\text{H}_2\text{O}_2$  molecules. Perturbations are as in Table 1 and results are plotted as in Figure 2. “Differences” correspond to the average error in the differences between the energies and properties of pairs of the  $\text{H}_2\text{O}_2$  molecules (panels 2 and 4).

8). The only tested properties that are not well-treated by the FG basis sets are the singlet–triplet energy splitting and dipole moment changes of unperturbed  $\text{H}_2\text{O}_2$  (Figure 3).

Another important result is that functional group basis sets can extrapolate to systems that are outside of their training sets. The results in Figures 4 and 8 show that the FG basis sets can extrapolate to different geometries and electrostatic environments. Table 3 shows extrapolations to entirely novel systems: a chemically diverse set of R–OH molecules. The FG basis sets in Table 3 clearly outperform equivalently sized atomic basis sets. Average errors in energy, singlet–triplet splitting, and dipole moment are generally smaller for these functional group basis sets than for the AO or hybrid AO basis sets. The functional group basis sets also usually return better values for the *difference* between the properties of a pair of R–OH molecules. The results are especially notable



**Figure 5.** B3LYP HOMO of unperturbed triplet  $\text{H}_2\text{O}$ . Contours are 0.05 au, 0.10 au, 0.15 au, etc. All results other than “6-31G++(3d,p)” use the parent basis on the left H atom and eleven basis functions on the remaining OH group. The “6-31G(hyb)” and “FG(HOOH)” results treat this OH group with the 6-31G AO basis set and the first 11 –OH basis functions from the FG(HOOH) training set, respectively. The “H2O Orb” results treat this OH group with the first 11 –OH basis functions obtained from the Kohn–Sham orbitals of singlet  $\text{H}_2\text{O}$  (section 3.2).



**Figure 6.** Errors in the dihedral rotational barrier of singlet  $\text{H}_2\text{O}_2$ , plotted as in Figure 2. Selected rotational barrier values are in Table 2.

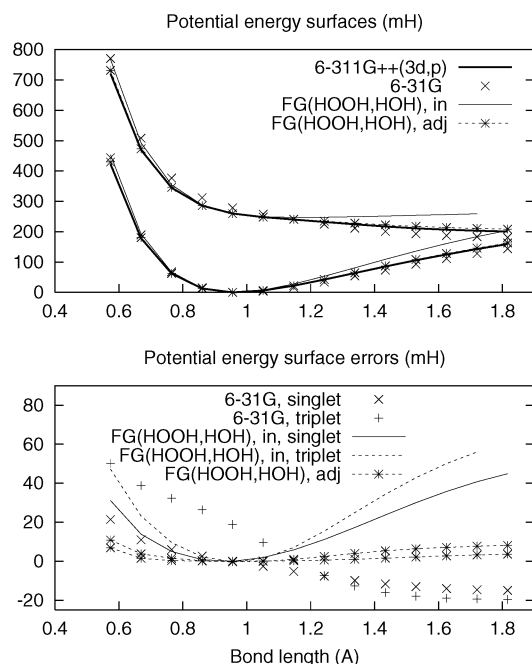
in that none of the R–OH groups in Table 3 were used in the FG basis training sets.

Table 3 contains results from two basis sets (“One” and “Orb”) that were extracted from a single molecule rather than molecules in different environments. These basis sets are parametrized to the electron density and Kohn–Sham orbitals of unperturbed  $\text{H}_2\text{O}$  (section 3.2). While these basis sets are generally better than similarly sized AO basis sets, they are not as good as the large-training-set basis sets. The “Orb” and “One” basis sets are comparable to FG(HOOH,HOH) and generally worse than FG(ROH). This is especially true for larger basis set sizes: while the “Orb” results give the best minimal –OH basis set, the FG(ROH) basis generally gives the best results for larger basis sizes. This result is not surprising, as only a chemically diverse training

**Table 3:** B3LYP Energy and Property Errors for R–OH Molecules, R = COO<sup>−</sup>, Na, OCH<sub>3</sub>, Cl, CN, CH<sub>3</sub>, CFH<sub>2</sub>, NH<sub>3</sub><sup>+</sup>

| Prop <sup>c</sup> | N <sup>d</sup> | average error <sup>a</sup> |                  |                  |                  |                 |                      | average difference error <sup>b</sup> |       |       |       |       |         |
|-------------------|----------------|----------------------------|------------------|------------------|------------------|-----------------|----------------------|---------------------------------------|-------|-------|-------|-------|---------|
|                   |                | AO <sup>e</sup>            | Hyb <sup>f</sup> | One <sup>g</sup> | Orb <sup>h</sup> | FG <sup>i</sup> | FG(ROH) <sup>j</sup> | AO                                    | Hyb   | One   | Orb   | FG    | FG(ROH) |
| E                 | 6              | 1028.                      | 297.5            | 17.17            | 20.39            | 23.81           | 21.98                | 510.0                                 | 11.48 | 18.97 | 24.39 | 23.94 | 19.97   |
|                   | 11             | 132.1                      | 51.90            | 11.66            | 14.07            | 14.55           | 8.70                 | 47.20                                 | 2.55  | 14.85 | 16.69 | 15.25 | 4.04    |
|                   | 17             | 83.06                      | 35.90            | 8.41             | 8.69             | 8.02            | 4.86                 | 26.09                                 | 1.27  | 10.29 | 11.24 | 9.34  | 0.77    |
|                   | 29             | 8.36                       | 2.92             | 3.76             | 1.94             | 1.09            | 2.12                 | 2.20                                  | 1.26  | 3.97  | 2.92  | 1.06  | 0.27    |
| ΔE                | 6              | 49.08                      | 11.66            | 8.94             | 12.12            | 6.05            | 10.27                | 61.39                                 | 10.18 | 5.72  | 6.43  | 15.32 | 9.93    |
|                   | 11             | 20.19                      | 4.47             | 4.84             | 5.37             | 2.30            | 1.90                 | 27.47                                 | 6.32  | 3.92  | 3.86  | 5.75  | 1.88    |
|                   | 17             | 22.39                      | 3.69             | 3.61             | 1.74             | 0.66            | 0.57                 | 28.89                                 | 5.38  | 3.15  | 1.75  | 1.59  | 0.76    |
|                   | 29             | 4.85                       | 2.77             | 2.60             | 0.61             | 0.35            | 0.25                 | 7.47                                  | 3.95  | 2.33  | 0.72  | 1.07  | 0.28    |
| μ                 | 6              | 0.802                      | 0.419            | 0.168            | 0.174            | 0.202           | 0.354                | 0.927                                 | 0.486 | 0.239 | 0.262 | 0.277 | 0.474   |
|                   | 11             | 0.525                      | 0.283            | 0.076            | 0.128            | 0.111           | 0.071                | 1.262                                 | 0.321 | 0.111 | 0.200 | 0.163 | 0.082   |
|                   | 17             | 0.326                      | 0.130            | 0.062            | 0.072            | 0.079           | 0.033                | 1.055                                 | 0.174 | 0.091 | 0.108 | 0.117 | 0.040   |
|                   | 29             | 0.096                      | 0.043            | 0.017            | 0.009            | 0.007           | 0.009                | 0.265                                 | 0.039 | 0.022 | 0.014 | 0.011 | 0.012   |

<sup>a</sup> Errors are relative to B3LYP/6-311G++(3d,p), all calculations except “AO” use 6-311G++(3d,p) on the R group and a small AO or FG basis on the OH group. <sup>b</sup> Average error in the difference between the energies and properties of pairs of different R–OH molecules. <sup>c</sup> Properties tested: E and ΔE are errors in total energy and singlet–triplet splitting (mH), μ is error in total singlet dipole (Debye). <sup>d</sup> Number of basis functions on the OH group. <sup>e</sup> Small AO basis set (STO-6G, 6-31G, 6-31G(d), or 6-311G+(d,p) for 6, 11, 17, and 29 basis functions per OH, respectively) used on the entire molecule. <sup>f</sup> OH basis is a small AO basis. <sup>g</sup> OH basis is an FG basis parametrized to unperturbed H<sub>2</sub>O. <sup>h</sup> OH basis is the first N orbitals from a B3LYP/6-311G++(3d,p) calculation on H<sub>2</sub>O (section 3.2). <sup>i</sup> OH basis is FG(HOOH, HOH) (Table 1). <sup>j</sup> OH basis is FG(ROH) (Table 1).



**Figure 7.** Testing –OH basis sets on bond dissociation. Results are singlet and triplet B3LYP potential energy surfaces (energy in mH vs bond length in Å) for stretching an O–H bond in H<sub>2</sub>O, with the other O–H bond and the H–O–H angle frozen at their equilibrium values. Upper and lower panels are the potential energy surfaces and the energy errors (mH) relative to B3LYP/6-311G++(3d,p). Calculations in the functional group basis used the first 11 –OH basis functions from the FG(HOOH,HOH) training set (Table 1) on one OH group and the parent basis on the remaining H atom. These calculations have the dissociating H–OH bond either within (“in”) or adjacent to (“adj”) the OH group treated with the FG basis. Potential energy surfaces are normalized such that the singlet energy at 0.956 Å is zero.

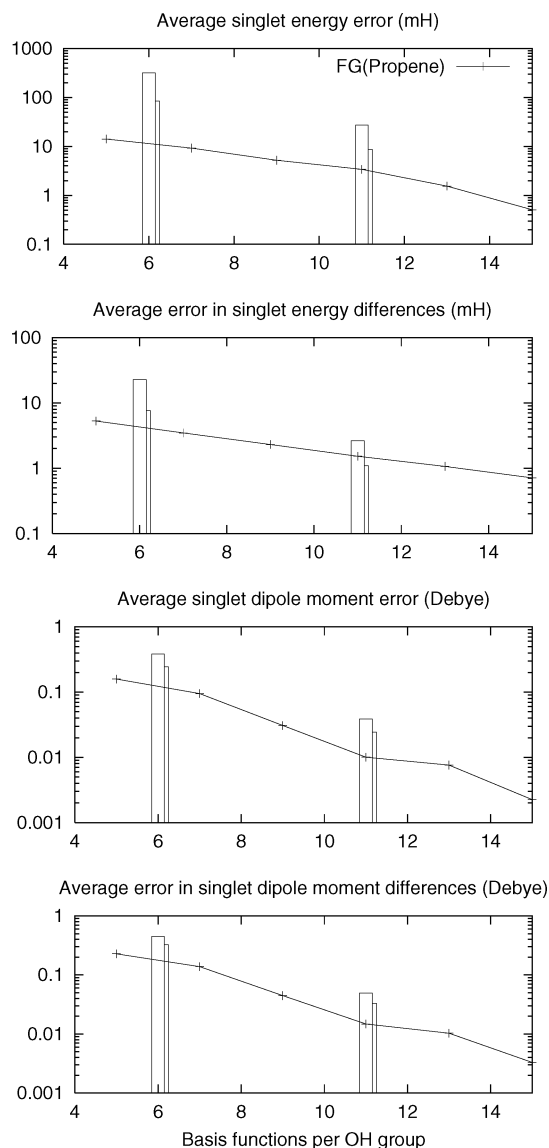
set will contain information about which basis functions are most important for describing the response to different R groups.

FG basis sets from molecules in different environments also generally have better symmetry properties than those obtained from single molecules. For example, Figure 5 contains results from the “Orb” basis set discussed above (“H<sub>2</sub>O Orb”). The highest-occupied Kohn–Sham orbital calculated with this basis set is asymmetric, unlike the one obtained from the FG(HOOH) basis. These results are representative, both for other basis sizes and for the “One” basis in Table 3 (data not shown).

FG basis sets have the useful property of converging smoothly with basis size. Smooth convergence can be seen, for example, in using the first 5, 6, 7, etc. –OH group basis functions from the FG(HOOH,HOH) training set to describe unperturbed H<sub>2</sub>O<sub>2</sub> (top panel of Figure 3). Smooth convergence is desirable as it enables extrapolation of large-basis results from a series of small-basis calculations. Several atomic basis sets, including the correlation-consistent,<sup>53</sup> polarization-consistent,<sup>54</sup> and correlation-consistent pseudo-potential<sup>55</sup> basis sets, have been explicitly designed to give smooth convergence. We find that FG basis set convergence is smoothest for total energies and averaged properties (Figures 4 and 8 and Table 3). Convergence is generally smoother for FG basis sets than for the small AO basis sets tested (see e.g. H<sub>2</sub>O<sub>2</sub> rotational barrier errors in Figure 6).

Some of the functional group basis set results in Figures 2 and 3 do not appear to converge smoothly with basis size. This is partly an artifact of the logarithmic scale. For example, the large apparent variations in the FG(HOOH,HOH) singlet–triplet splittings are all < 1.5 mH when using 10 or more FG basis functions. The rough convergence is also partly due to symmetry effects. The individual functional group basis functions do not (and cannot) all have the same symmetry as the functional group (see Figure 1). We are currently experimenting with partitioning the FG basis functions into properly symmetrized groups.

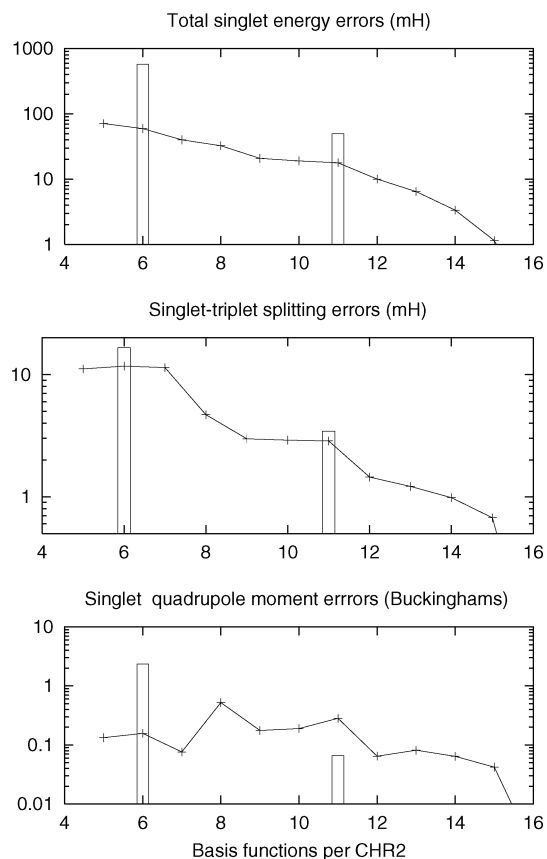
It is perhaps counterintuitive that functional group basis functions can work for multiple group geometries (see Table



**Figure 8.** Average errors, relative to B3LYP/6-31G(d), in the energies and properties of 43 perturbed singlet propene molecules. Perturbations are as in Table 1, and results are plotted as in Figure 2. AO-basis results use 6-31G and STO-6G (11 and 6 basis functions per  $-\text{CH}-$  group, respectively). “Differences” correspond to the average error in the differences between the energies and properties of pairs of the propene molecules (panels 2 and 4).

3 and Figures 4 and 8). This is possible because our FG basis functions are contractions over atomic basis functions and can therefore change shape with molecular geometry. Consider a simple example: an FG basis set for  $\text{H}_2$ , where the parent AO basis has one  $s$  orbital on each atom. The FG basis functions will be the bonding and antibonding combination of the two atomic orbitals. The first (bonding) FG basis function will be the singlet RHF ground state for all H–H bond lengths.

Figure 7 presents a detailed test of geometry variations in a functional group basis set. As expected, the FG basis set gives large errors for large bond variations. However, it does better than an equivalently sized AO basis over the range of bond lengths included in its training set ( $0.956 \pm 0.3 \text{ \AA}$ , see Table 1). In this regime, the 6-31G energy errors change



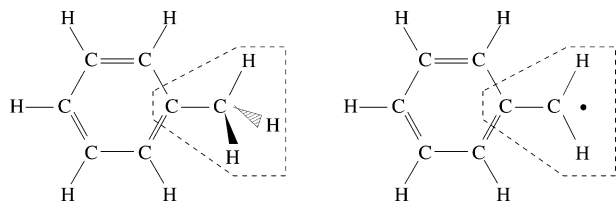
**Figure 9.** Errors relative to B3LYP/6-31G(d) in the energy and properties of unperturbed benzene.  $-\text{CH}-$  basis sets are from the training set in Table 1, and results are plotted as in Figure 8. Quadrupole moments are measured perpendicular to the plane of the benzene ring.

dramatically with bond length, while the  $-\text{OH}$  basis errors are approximately constant. These results are representative of the other basis set sizes tested (data not shown). We also consider a situation where the parent basis is used on the dissociating H atom and the FG basis is used on the remaining  $-\text{OH}$  fragment. Since the functional group retains its identity in this situation, we expect that the errors will be small. The “FG(HOOH,HOH), adj” results in Figure 7 show that this is the case.

The results in section 3 show that functional group basis sets work for useful levels of theory (B3LYP density functional theory). These results are representative for self-consistent field methods, including Hartree–Fock theory and DFT with other functionals (data not shown). While the electron density implicitly describes electron correlation via the Hohenberg–Kohn theorem,<sup>56</sup> we expect that FG basis sets for explicitly correlated methods will need to be parametrized to explicit information about electron–electron interactions. We are currently using ideas from our electron-pair-density methods<sup>51</sup> to develop such basis sets.

## 5. Hybrid Electronic Structure Methods

Functional group basis sets may be particularly useful for hybrid electronic structure calculations. Many large systems, such as protein catalysts and solvated molecules, have a relatively small region of chemical interest. Such systems



**Figure 10.** Reactant (left) and product (right) of the toluene PhCH<sub>2</sub>–H bond dissociation. The active region, which requires treatment with a large variational space, is enclosed with dotted lines.

can be treated using hybrid methods. These methods combine a high-level treatment of the active region with a lower-level treatment of the remaining spectator groups.<sup>57–60</sup> Semiempirical or molecular-mechanics treatments of spectator groups is problematic for systems with complex interactions between the active and spectator regions.<sup>60–63</sup> The PhCH<sub>2</sub>–H bond dissociation of toluene (Figure 10) is one such system, as the product PhCH<sub>2</sub>\* radical can delocalize onto the C<sub>5</sub>H<sub>5</sub> spectator group.

Systems with complex interactions between active and spectator groups may be treated with mixed basis set methods. This hybrid method involves an ab initio calculation on the entire system, with a large basis set on the active region and a smaller basis elsewhere.<sup>64–68</sup> Like all hybrid methods, mixed basis sets require care in treating the high-level/low-level interface. Inappropriate choice of basis sets can produce inaccuracies due to basis set imbalance. A simple example of basis set imbalance is a homonuclear diatomic with a minimal basis on one atom and a large basis on the other. Electron density will tend to move onto the atom with the large basis, giving the molecule a spurious dipole moment.

We have tested FG basis sets vs atomic basis sets in a series of hybrid calculations. These tests, detailed below, use the parent AO basis on the active region and a functional group or small atomic basis on the remainder. We find that FG basis sets have smaller basis set imbalance effects than equivalently sized atomic basis sets. We believe that this is because the FG basis functions are contracted over the parent basis set and so are likely to be more compatible with the full parent basis.

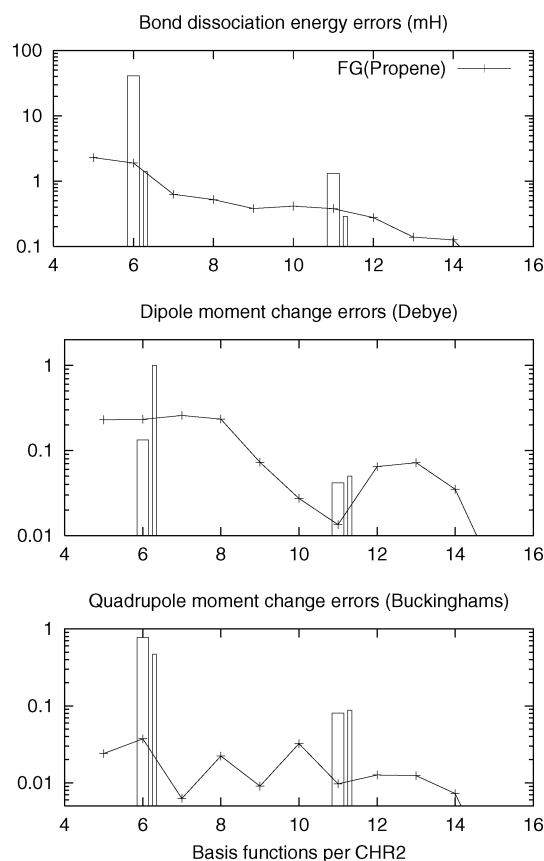
Table 3 in section 3 gives errors for mixed basis calculations on several R–OH molecules. Here, R is treated with the parent basis, and OH is treated with either a small AO basis (“Hyb”) or an –OH functional group basis (“FG”, “FG-(ROH)”). The average errors in energy, singlet–triplet splittings, dipole moment, and the difference between these properties for pairs of R–OH molecules, are almost always lower for the functional group basis sets.

Table 4 presents calculated dipole moments for trans H<sub>2</sub>O<sub>2</sub>. This molecule is analogous to a homonuclear diatomic and has zero dipole moment when treated with a balanced basis set. Here, one OH group is treated with either a FG basis set or a small AO basis. The other group is treated with the parent AO basis. The FG basis sets have much lower dipole moments than the equivalently sized AO basis sets, indicating less basis set imbalance.

**Table 4:** Hybrid-Basis Dipole Moments (Debye) for trans H<sub>2</sub>O<sub>2</sub>

| N <sup>a</sup> | AO <sup>b</sup> | FG <sup>c</sup> |
|----------------|-----------------|-----------------|
| 6              | 0.938           | 0.204           |
| 11             | 0.258           | 0.118           |
| 17             | 0.179           | 0.072           |
| 29             | 0.075           | 0.004           |

<sup>a</sup> Number of basis functions on the second OH group. <sup>b</sup> The second OH group is treated with a small AO basis (STO-6G, 6-31G, 6-31G(d), or 6-311G+(d,p) for 6, 11, 17, and 29 basis functions per OH, respectively). <sup>c</sup> The second OH group is treated with the first N basis functions from the FG(HOOH,HOH) training set (Table 1).



**Figure 11.** Errors in PhCH<sub>2</sub>–H bond dissociation energy and PhCH<sub>3</sub>→PhCH<sub>2</sub>\* dipole and out-of-plane quadrupole moment change, relative to B3LYP/6-31G(d), plotted vs number of basis functions per –CH– group. Lines use 6-311G(d) on the active region (Figure 10) and the –CH– functional group basis sets from section 3 on the remainder. Narrow columns use 6-31G(d) on the active region and a small AO basis set (STO-6G at 6 and 6-31G at 11 basis functions per –CH–, respectively) on the remainder. Wide columns use a small AO basis set on the entire molecule.

Finally, we consider mixed basis calculations on the PhCH<sub>2</sub>–H bond dissociation shown in Figure 10. Figure 11 presents errors in the calculated dissociation energy and PhCH<sub>2</sub>–H→PhCH<sub>2</sub>\* change in total dipole moment and out-of-plane quadrupole moment, relative to B3LYP/6-31G(d). The active region (see Figure 10) is treated with the parent 6-31G(d) basis, and the C<sub>5</sub>H<sub>5</sub> spectator group is treated with either a small AO basis set or a –CH– FG basis set. The functional group basis sets perform about the same as equivalently sized atomic basis sets for the dissociation

energy and somewhat better for the change in multipole moments. Figure 11 also presents results obtained using a small atomic basis on the entire molecule (wide columns). The use of a balanced but small basis reduces the error in the multipole moment changes, at the expense of a substantially increased error in the dissociation energy. Hybrid-basis calculations with the  $-\text{CH}-$  FG basis have the advantage of giving reasonable results for both dissociation energy and multipole moments.

## 6. Conclusions

This paper presents a new approach to reducing the variational space needed for ab initio calculations on molecules. The electronic structures of molecular functional groups are largely conserved across different molecules. However, existing atomic orbital basis sets treat each element with the same variational space, regardless of what functional group it is in. We generate basis sets for functional groups in molecules ( $-\text{OH}$  groups,  $-\text{CH}-$  groups, and so forth). The functional group (FG) basis sets are designed to describe the electronic variational space that is most important for electrons in the group of interest. FG basis functions for a functional group  $F$  are generated by contracting over a “parent” AO basis on all atoms in  $F$ . The contraction coefficients are determined via feature extraction (principle component analysis)<sup>9</sup> of the calculated electronic structures of a “training set” of  $F$ -containing molecules. This approach differs considerably from the standard methods of constructing basis sets.

In the current work, we generate basis sets for the  $-\text{OH}$  and  $-\text{CH}-$  functional groups. Our results demonstrate that these basis sets enable substantial dimensional reduction with minimal loss of accuracy relative to the parent basis. The FG basis sets always give lower energy errors, and usually give lower property errors, than comparably sized atomic basis sets. They are able to extrapolate to chemically diverse sets of molecules. In addition, their convergence with basis size is generally fairly smooth, allowing one to extrapolate more readily to the large-basis limit. The FG basis sets are especially suited for mixed-basis calculations, due to the reduced effects of basis set imbalance.

One interesting aspect of this method is that we can tailor the dimensional reduction to specific instances. For example, one might parametrize an FG basis set for electronic polarizability by applying electric fields to the training set molecules and doing PCA on the difference between the calculated electron densities in the presence and absence of applied fields. We intend to explore such methods in future work.

The results presented here suggest that FG basis sets may be useful for ab initio calculations on large systems, especially for treating spectator groups in mixed-basis calculations. However, a number of aspects are left to future work. These include FG basis sets for explicitly correlated methods and a complete implementation into a production quality electronic structure package.

## 7. Implementation Details

**7.1. Parametrization Method.** Basis functions for a functional group  $F$  are parametrized as follows. First, we define

a set of local coordinates for  $F$  (section 7.2). We select a parent atomic orbital basis set and a set of small  $R$  groups and build a training set of molecules from  $\{R\}$  and  $F$  (e.g.  $R - \text{OH}$  for  $-\text{OH}$  groups and  $R - \text{CH} - R'$  for  $-\text{CH}-$  groups). We usually add diversity to the training set by generating  $\mathcal{O}(100)$  conformers for each molecule, each with a different set of geometric and electrostatic perturbations (“Rand” in Table 1). Geometric perturbations are generated by randomly changing all bond lengths, angles, and dihedrals. Electrostatic perturbations are generated by randomly placing point charges in a box around the molecule and discarding those that are too close to the nuclei.

Next we perform ab initio calculations, in the parent basis set, on all molecules in the training set. Each calculation yields a one-electron density matrix in the parent AO basis ( ${}^1D$ ) and the overlap matrix of the parent basis functions ( $S$ ). We extract the sub-blocks that correspond to functional group  $F$ : all matrix elements  ${}^1D_{ij}$  and  $S_{ij}$  where AO basis functions  $i$  and  $j$  are on atoms in  $F$ . Each density matrix sub-block is rotated into  $F$ 's local coordinates (section 7.2) and diagonalized. The eigenvectors, referred to here as “group natural orbitals”, provide a representation of how the electrons in  $F$  are distributed across the parent basis functions. We save all group natural orbitals with eigenvalues (occupancies)  $> 0.1$  electrons and perform PCA on the saved group natural orbitals from the entire training set. The vectors returned by PCA are the functional group basis functions in the parent AO basis. FG basis sets that are parametrized to very small training sets (e.g. results labeled “One” in Table 3) have the input group natural orbitals weighted by their occupancies before doing the PCA. We have experimented with using this technique for large training sets but have not found it to improve the results.

Principle component analysis on vectors in an orthonormal basis returns a set of orthogonal vectors. We minimize the overlap of the FG basis functions by performing the PCA in an approximately orthogonal basis set. This basis is obtained by averaging the saved overlap matrix sub-blocks from the entire training set and performing Schmidt orthogonalization on the result. The saved group natural orbitals are projected into this basis before PCA, and the resulting PCA vectors are projected back into the parent AO basis.

**7.2. Orienting Functional Group Basis Functions.** Since functional groups are generally not spherically symmetric, our FG basis functions must be oriented with respect to the geometry of each molecule of interest. Consider, for example, the  $\text{H}_2\text{O}_2$  molecules in Figure 1. Here, the FG basis functions are plotted on the  $-\text{OH}$  group containing atoms O(3) and H(4). Let the  $z$  axis be vertical, such that basis function 4 is mostly composed of the  $p_z$  orbitals on O(3). If the basis functions were transferred without rotation to the other  $-\text{OH}$  group, basis function 4 would be composed of the  $p_z$  orbitals on O(2) and would have the wrong orientation with respect to the H(1)–O(2) bond. Rotation ensures that this basis function is composed of the linear combinations of p orbitals on O(2) that are perpendicular to the plane defined by atoms 1–3.

We orient our FG basis functions by defining them relative to an internal coordinate system. For example, the internal

coordinates of the R–OH group have the  $x$ -axis along the R–O bond and the O–H group in the  $x$ – $y$  plane. The basis functions are oriented by calculating the Cartesian rotation matrix between a group's local coordinates and its orientation in the molecule of interest. This rotation matrix is used to rotate each p- and d-orbital shell from the local orientation to the correct global orientation.<sup>69,70</sup>

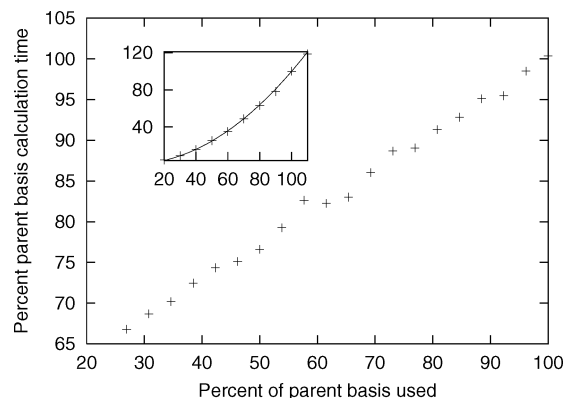
**7.3. GAMESS Implementation and Timing Tests.** A complete implementation of functional group basis sets would calculate the one- and two-electron operators in the parent AO basis, rotate each group's FG basis functions into the correct orientation (section 7.2), project the operators into the FG basis, and use the transformed operators in the ab initio calculation. The time savings would be similar to those for contracted vs uncontracted AO basis sets.<sup>39</sup> Initial one- and two-integral evaluations would scale as the number of Gaussian primitives, and all subsequent steps (dimension of Fock matrix, number of possible determinants) would scale as the number of FG basis functions.

The results presented here use a preliminary implementation of FG basis sets built into GAMESS.<sup>71</sup> This implementation builds the Fock matrices in the parent AO basis and diagonalizes them in the FG basis. Transformations into the FG basis are performed by exploiting the methods in GAMESS for removing linearly dependent combinations of basis functions. Our initial implementation used GAMESS for the parent-basis operator evaluations and home-built code for ab initio calculations in the FG basis. Test calculations showed that the two implementations return identical results (data not shown).

GAMESS removes linearly dependent combinations of basis functions as follows.<sup>72</sup> Consider a system with an  $N$ -orbital parent AO basis. Let  $S$  be the overlap matrix of the AO basis functions. SCF calculations in the parent basis diagonalize the transformed Fock matrix  $Q^\dagger F Q$ , where  $F$  is the AO-basis Fock matrix and  $Q$  is an orthonormal basis set generated by diagonalizing  $S$  ( $Q^\dagger S Q = 1$ ). Linear dependencies are removed by discarding eigenvectors of  $S$  with negligible eigenvalues, such that a basis with  $N' < N$  linearly independent vectors yields a rectangular  $Q$  matrix and a transformed Fock matrix of dimension  $N' \times N'$ .

We implement FG basis functions as follows. Consider calculations using  $N_{\text{used}} < N$  FG basis functions  $U$ , where  $U_{ji}$  is the projection of the  $i$ th FG basis function onto the  $j$ th parent AO basis function. We replace  $S$  with the  $N \times N$  matrix  $R^\dagger S R$  in the GAMESS routine for calculating  $Q$ . Matrix  $R$  zeros out the degrees of freedom that are outside of the FG basis, such that  $R^\dagger S R$  has  $N - N_{\text{used}}$  negligible eigenvalues.  $R$  is defined as  $U \sigma^{-1} U^\dagger S$  where  $\sigma = U^\dagger S U$  is the FG basis functions' overlap matrix. Finally, we replace  $Q$  with  $RQ$ , such that SCF calculations diagonalize the  $N_{\text{used}} \times N_{\text{used}}$  matrix given by  $Q^\dagger R^\dagger F R Q$ . The FG basis functions are rotated from their internal coordinates outside of GAMESS (section 7.2).

Figure 12 presents a timing test of our GAMESS implementation of functional group basis sets. The figure plots the total calculation times for RHF/6-311G++(3d,p) calculations on 10 noninteracting H<sub>2</sub> molecules. The FG basis set is the RHF/6-311G++(3p) molecular orbitals of a single H<sub>2</sub>,



**Figure 12.** Timing tests for functional group basis calculations on a model system of 10 noninteracting H<sub>2</sub> molecules. X axis is basis size as a percent of the parent AO basis. Y axis is RHF calculation time as a percent of the time for a calculation in the parent AO basis. Inset is calculation time vs total basis size for  $N$  H<sub>2</sub> molecules, percentages relative to  $N=10$ , and a quadratic fit to the data.

and all FG basis calculations return the parent-basis wave function. The FG basis sets have a clear computational savings and a relatively small overhead (calculation time 100.4% of parent basis at basis size 100%). The computational savings of our FG basis implementation will depend on the linear-scaling approximations used in the calculation. In the systems tested here, we use a fast multipole moment approximation for the two-electron integrals. The total calculation time scales quadratically (Figure 12 inset) and the time savings scales linearly.

**Acknowledgment.** This work was supported by the National Science Foundation. BGJ thanks the NSF for additional support.

## References

- (1) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- (2) Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*(25), 4782–4790.
- (3) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*(4), 1085–1123.
- (4) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*(18), 8330–8343.
- (5) Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*(2), 661–680.
- (6) Kohn, W. *Rev. Mod. Phys.* **1999**, *71*(5), 1253–1266.
- (7) Machida, K. *Principles of molecular mechanics*; Wiley: New York, 1999.
- (8) Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Design* **2000**, *14*, 123–134.
- (9) Cherkassky, V.; Mulier, F. *Learning from Data: Concepts, Theory, and Methods*; Wiley-Interscience: 1998.
- (10) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050–8053.
- (11) Lee, T.-s.; Yang, W. *Int. J. Quantum Chem.* **1998**, *69*, 397–404.
- (12) Ermolaeva, M. D.; van der Vaart, A.; Merz, Jr., K. M. *J. Phys. Chem. A* **1999**, *103*, 1868–1875.



- (13) Bader, R. *Atoms in Molecules – A Quantum Theory*; Oxford University Press: Oxford, 1990.
- (14) Trindle, C.; Sinanoglu, O. *J. Chem. Phys.* **1968**, *49*, 65–71.
- (15) Naray-Szabo, G.; Kramer, G.; Nagy, P.; Kugler, S. *J. Comput. Chem.* **1987**, *8*, 555–561.
- (16) Lichtenberger, D. L.; Fenske, R. F. *J. Chem. Phys.* **1976**, *64*, 4247–4264.
- (17) Carpenter, J. E.; Weinhold, F. *J. Am. Chem. Soc.* **1988**, *110*, 368–372.
- (18) Hierse, W.; Stechel, E. B. *Phys. Rev. B* **1994**, *50*(24), 17811–17819.
- (19) Hierse, W.; Stechel, E. B. *Phys. Rev. B* **1996**, *54*(23), 16515–16522.
- (20) England, W.; Gordon, M. S.; Reudenberg, K. *Theor. Chim. Acta* **1975**, *37*, 177–216.
- (21) Gallup, G. A. *Valence Bond Methods: Theory and Applications*; Cambridge University Press: 2002.
- (22) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon: Oxford, 1996.
- (23) Flurry Jr., R. L. *Theor. Chim. Acta* **1971**, *23*, 1–11.
- (24) von Niessen, W. *J. Chem. Phys.* **1971**, *55*, 1948–1957.
- (25) Morokuma, K.; Kitaura, K. In *Chemical applications of atomic and molecular electrostatic potentials*; Politzer, L., Truhlar, D. G., Eds.; Plenum Press: New York, 1981.
- (26) Stevens, W. J.; Fink, W. H. *Chem. Phys. Lett.* **1987**, *139*, 15–22.
- (27) Dewar, M. J. S.; Dougherty, R. C. *The PMO Theory of Organic Chemistry*; Plenum Press: New York, 1975.
- (28) Rauhut, G.; Boughton, J. W.; Pulay, P. *J. Chem. Phys.* **1995**, *103*(13), 5662–5673.
- (29) Kapuy, E.; Kozmutza, C. *J. Chem. Phys.* **1991**, *94*, 5565–5573.
- (30) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935–10944.
- (31) Gagliardi, L.; Lindh, R.; Karlstrom, G. *J. Chem. Phys.* **2004**, *121*, 4494–4500.
- (32) Poteau, R.; Ortega, I.; Alary, F.; Solis, A. R.; Barthelat, J.-C.; Daudey, J.-P. *J. Phys. Chem. A* **2001**, *105*, 198–205.
- (33) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- (34) Liang, W.; Head-Gordon, M. *J. Phys. Chem. A* **2004**, *108*, 3206–3210.
- (35) Wolinski, K.; Pulay, P. *J. Chem. Phys.* **2003**, *118*, 9497–9502.
- (36) Lee, M. S.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *107*, 9085–9095.
- (37) Berghold, G.; Parrinello, M.; Hutter, J. *J. Chem. Phys.* **2002**, *116*, 1800–1810.
- (38) Lu, W. C.; Wang, Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Reudenberg, K. *J. Chem. Phys.* **2004**, *120*, 2629–2637.
- (39) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- (40) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (41) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (42) Frisch, M. J.; Binkley, J. S.; Pople, J. A. *J. Chem. Phys.* **1984**, *80*, 3265–3269.
- (43) Huzinaga, S. *J. Chem. Phys.* **1965**, *42*, 1293–1302.
- (44) Spangler, D.; Christoffersen, R. E. *Int. J. Quantum Chem.* **1980**, *17*, 1075–1097.
- (45) Widmark, P.-O.; Malmqvist, P.-A.; Roos, B. O. *Theor. Chim. Acta* **1990**, *77*, 291–306.
- (46) Curtarolo, S.; Morgan, D.; Persson, K.; Rodgers, J.; Ceder, G. *Phys. Rev. Lett.* **2003**, *91*(13), 135503.
- (47) Wheeler, R. A.; Dong, H.; Boesch, S. E. *ChemPhysChem* **2003**, *4*, 382–384.
- (48) Razeghifard, M. R.; Kim, S.; Patzlaff, J.; Hutchison, Ronald, S.; Krick, T.; Ayala, I.; Steenhuis, J. J.; Boesch, S. E.; Wheeler, R. A.; Barry, Bridgette, A. *J. Phys. Chem. B* **1999**, *103*, 9790–9800.
- (49) Otto, M. *Chemometrics: statistics and computer application in analytical chemistry*; Wiley-VCH: New York, 1999.
- (50) Malinowski, E. R. *Factor analysis in chemistry*; Wiley: New York, 2002.
- (51) Janesko, B. G.; Yaron, D. *J. Chem. Phys.* **2004**, *121*, 5635–5645.
- (52) NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101 Release 10, May 2004, Editor: Russell D. Johnson III <http://srdata.nist.gov/cccbdb>.
- (53) Dunning, Jr., T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (54) Jensen, F. *J. Chem. Phys.* **2001**, *115*, 9113–9125.
- (55) Peterson, K. A. *J. Chem. Phys.* **2003**, *119*, 11099–11112.
- (56) Hohenburg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, b864-b871.
- (57) Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* **1972**, *94*, 5612–5625.
- (58) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- (59) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (60) Cui, Q.; Guo, H.; Karplus, M. *J. Chem. Phys.* **2002**, *117*(12), 5617–5631.
- (61) Reuter, N.; Dejaegere, A.; Maignet, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720–1735.
- (62) Gogonea, V.; Westerhoff, L. M.; Merz, Jr., K. M. *J. Chem. Phys.* **2000**, *113*(14), 5604–5613.
- (63) Govind, N.; Wang, Y. A.; Carter, E. A. *J. Chem. Phys.* **1999**, *110*(16), 7677–7688.
- (64) Jensen, J. H.; Gordon, M. S. *J. Comput. Chem.* **1991**, *12*, 421–426.
- (65) Chesnut, D. B.; Moore, K. D. *J. Comput. Chem.* **1989**, *10*, 648–659.
- (66) Hinton, J. F.; Guthrie, P.; Pulay, P.; Wolinski, K. *J. Am. Chem. Soc.* **1992**, *114*, 1604–1605.
- (67) Chesnut, D. B.; Byrd, E. F. C. *Chem. Phys.* **1996**, *213*, 153–158.
- (68) Montoya, A.; Mondragon, F.; Truong, T. N. *Carbon* **2002**, *40*, 1863–1872.

- (69) Ivanec, J.; Reudenberg, K. *J. Phys. Chem.* **1996**, *100*, 6342–6347.
- (70) Schlegel, H. B.; Frisch, M. J. *Int. J. Quantum Chem.* **1995**, *54*, 83–87.
- (71) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery Jr., J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (72) Hollauer, E.; Dupis, M. *J. Chem. Phys.* **1992**, *96*, 5220–5228.

CT049853Y

## The Perfluoroadamantyl Radicals C<sub>10</sub>F<sub>15</sub> and Their Anions

Xue-jun Feng and Qian-shu Li\*

Department of Chemistry, School of Science, Beijing Institute of Technology,  
Beijing 100081, P. R. China

Yaoming Xie and Henry F. Schaefer III\*

Center for Computational Chemistry, University of Georgia, Athens, Georgia 30602

Received November 22, 2004

**Abstract:** The optimized geometries, electron affinities, and harmonic vibrational frequencies of perfluoroadamantyl radicals (C<sub>10</sub>F<sub>15</sub>) have been obtained using four carefully calibrated density functional theory methods in conjunction with diffuse function augmented double- $\zeta$  plus polarization (DZP++) basis sets. There are two C<sub>10</sub>F<sub>15</sub> isomers with close energies. With the DZP++ B3LYP method, the C<sub>3v</sub> isomer (1-C<sub>10</sub>F<sub>15</sub>) lies energetically *above* the C<sub>s</sub> isomer (2-C<sub>10</sub>F<sub>15</sub>) by 0.086 eV (2.0 kcal/mol), while the anionic 1-C<sub>10</sub>F<sub>15</sub><sup>-</sup> isomer is predicted to lie *below* 2-C<sub>10</sub>F<sub>15</sub><sup>-</sup> anion by 1.00 eV (23.0 kcal/mol). The DZP++ B3LYP method predicts the ZPVE-corrected adiabatic electron affinity for the C<sub>3v</sub> isomer (1-C<sub>10</sub>F<sub>15</sub>) to be 4.16 eV, and that for the C<sub>s</sub> isomer (2-C<sub>10</sub>F<sub>15</sub>) is 3.10 eV. These EA<sub>ad</sub> values are significantly larger than that (1.31 eV) of the parent molecule perfluoroadamantane (C<sub>10</sub>F<sub>16</sub>). For the 1-C<sub>10</sub>F<sub>15</sub> radical, the C\*–C bond length is shortened by 0.043 Å upon removal of F from the C<sub>10</sub>F<sub>16</sub> molecule. The C\*–C bond distance for the 1-C<sub>10</sub>F<sub>15</sub><sup>-</sup> anion is 0.068 Å shorter than that for C<sub>10</sub>F<sub>16</sub>. Similarly, for 2-C<sub>10</sub>F<sub>15</sub> the C\*–C distance is 0.053 Å shorter than for C<sub>10</sub>F<sub>16</sub>, while r<sub>e</sub>(C\*–C) for the anion is 0.061 Å shorter than for C<sub>10</sub>F<sub>16</sub>.

### Introduction

Both adamantane (C<sub>10</sub>H<sub>16</sub>) and, surprisingly, the two adamantyl radicals (C<sub>10</sub>H<sub>15</sub>) have been predicted to have negative adiabatic electron affinities (EAs) with carefully calibrated density functional theory (DFT) methods.<sup>1</sup> That is, these systems do not readily attract an electron.<sup>2</sup> However, in the same theoretical study, it was suggested that suitably substituted adamantyl radicals would be expected to have positive electron affinities.<sup>1</sup> Perfluorohydrocarbons have generated recent attention because of their ability to bind an additional electron.<sup>3–7</sup> The perfluoroadamantane (C<sub>10</sub>F<sub>16</sub>) molecule is predicted to have a substantial EA<sub>ad</sub> value (1.31 eV).<sup>7</sup> One would expect the perfluoroadamantyl radicals to have even larger electron affinities.

For the larger perfluorohydrocarbons reliable experimental electron affinities (EAs) are lacking. Theoretical studies can often aid in the interpretation of experimental photoelectron spectra and usually provide accurate predictions of electron affinities.<sup>8</sup> DFT methods have been systematically applied to large molecules, and the average error for the B3LYP results is about 0.15 eV compared to experiment.<sup>8,9</sup> In the present study we examine the two isomers of the perfluoroadamantyl radical, namely 1-C<sub>10</sub>F<sub>15</sub> and 2-C<sub>10</sub>F<sub>15</sub>, and their anions using four selected DFT methods in conjunction with DZP++ basis sets, with the goal of establishing reliable theoretical predictions for the unknown EAs. We also compare these radical EAs with that for the C<sub>10</sub>H<sub>15</sub> radicals and those for their parent molecule C<sub>10</sub>F<sub>16</sub>.

Three neutral-anion energy separations, i.e., the adiabatic electron affinity (EA<sub>ad</sub>), the vertical electron affinity (EA<sub>vert</sub>), and the vertical detachment energy (VDE), have been

\* Corresponding author e-mail: hfs@arches.uga.edu.

predicted as differences in total energies according in the following manner:

$$EA_{\text{ad}} = E(\text{optimized neutral}) - E(\text{optimized anion})$$

$$EA_{\text{vert}} = E(\text{optimized neutral}) - E(\text{anion at the optimized neutral geometry})$$

$$VDE = E(\text{neutral at the optimized anion geometry}) - E(\text{optimized anion})$$

## Theoretical Methods

Similar to previous studies,<sup>6,7</sup> the four DFT methods employed in the present research have been denoted BHLYP, B3LYP, BP86 and BLYP. The functional named BHLYP by the developers of the GAUSSIAN programs combines a modification of Becke's half and half exchange functional<sup>10</sup> with the Lee, Yang, and Parr (LYP) correlation functional.<sup>11</sup> The B3LYP method is a hybrid Hartree–Fock/density functional (HF/DFT) method using Becke's three-parameter hybrid functional<sup>12</sup> with the LYP correlation functional. The BP86 approach is Becke's 1988 exchange functional<sup>13</sup> in conjunction with Perdew's 1986 correlation functional.<sup>14</sup> The BLYP functional combines Becke's 1988 exchange functional<sup>13</sup> with the LYP correlation functional.

Basis sets of double- $\zeta$  quality plus polarization functions augmented with diffuse functions (DZP++) were used in the present work. They are the Huzinaga–Dunning<sup>15,16</sup> contracted Gaussian double- $\zeta$  functions appended with one set of five d-type polarization functions, plus a set of *sp* diffuse functions based on an “even-tempered” formula.<sup>17</sup> For fluorine, the orbital exponents of the polarization function are  $\alpha_d(\text{F}) = 1.00$  and the exponents of the diffuse functions are  $\alpha_s(\text{F}) = 0.10490$  and  $\alpha_p(\text{F}) = 0.08260$ . For carbon, the exponents are  $\alpha_d(\text{C}) = 0.75$ ,  $\alpha_s(\text{C}) = 0.04302$  and  $\alpha_p(\text{C}) = 0.03629$ . Thus the final basis sets may be described as C, F(10s6p1d/5s3p1d). All computations for open-shell systems were done in a spin-unrestricted formalism.

Absolute total energies, optimized geometries, and harmonic vibrational frequencies for each structure were predicted using the above four DFT methods. Zero-point vibrational energies (ZPVEs) were evaluated at each level. The ZPVE differences between the neutrals and the corresponding anions are then used for the correction of the electron affinities. Our computations were performed with the Gaussian94 programs.<sup>18</sup> The default integration grid (75,302) of Gaussian94 was applied, and the tight SCF convergence was required. Cartesian coordinates for all structures are reported in the Supporting Information, Table S1. Total energies are reported in Table 1.

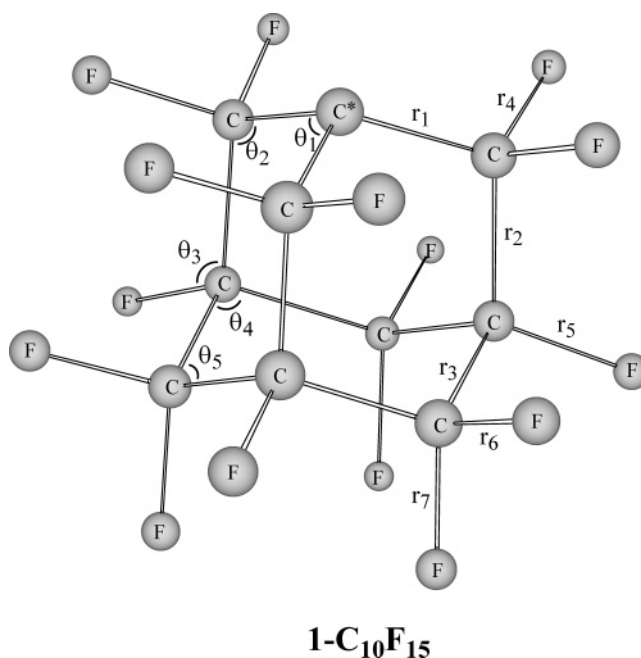
## Results and Discussion

**A. 1-C<sub>10</sub>F<sub>15</sub> and 1-C<sub>10</sub>F<sub>15</sub><sup>-</sup>.** There are two structural types of carbon atoms forming the perfluoroadamantane (C<sub>10</sub>F<sub>16</sub>, *T<sub>d</sub>*) skeleton: four vertex carbon atoms, each connected to other three carbon atoms, and six bridging carbons, each bonding to other two carbon atoms.

The first isomer of perfluoroadamantyl radical (1-C<sub>10</sub>F<sub>15</sub>) is the structure in which a fluorine atom is removed from a vertex carbon atom, and this radical has *C<sub>3v</sub>* symmetry. The

**Table 1.** Total Energies (in Hartree) for the C<sub>10</sub>F<sub>16</sub>/C<sub>10</sub>F<sub>16</sub><sup>-</sup>, 1-C<sub>10</sub>F<sub>15</sub>/1-C<sub>10</sub>F<sub>15</sub><sup>-</sup>, and 2-C<sub>10</sub>F<sub>15</sub>/2-C<sub>10</sub>F<sub>15</sub><sup>-</sup> Systems

| compound                                       | BHLYP       | B3LYP       | BP86        | BLYP        |
|--|-------------|-------------|-------------|-------------|
| C <sub>10</sub> F <sub>16</sub>                | -1978.23483 | -1978.98955 | -1978.99646 | -1978.75110 |
| C <sub>10</sub> F <sub>16</sub> <sup>-</sup>   | -1978.24388 | -1979.02862 | -1979.05271 | -1978.80430 |
| 1-C <sub>10</sub> F <sub>15</sub>              | -1878.34193 | -1879.06063 | -1879.06834 | -1878.82976 |
| 1-C <sub>10</sub> F <sub>15</sub> <sup>-</sup> | -1878.48337 | -1879.21185 | -1879.22118 | -1878.97875 |
| 2-C <sub>10</sub> F <sub>15</sub>              | -1878.34358 | -1879.06379 | -1879.07106 | -1878.83350 |
| 2-C <sub>10</sub> F <sub>15</sub> <sup>-</sup> | -1878.44503 | -1879.17528 | -1879.18482 | -1878.94441 |



**Figure 1.** Sketch of the *C<sub>3v</sub>* structures for the 1-C<sub>10</sub>F<sub>15</sub> radical and the 1-C<sub>10</sub>F<sub>15</sub><sup>-</sup> anion. The optimized geometrical parameters for these structures are reported in Tables 2 and 3, respectively.

neutral 1-C<sub>10</sub>F<sub>15</sub> radical has a <sup>2</sup>A<sub>1</sub> ground electronic state, and the corresponding anionic 1-C<sub>10</sub>F<sub>15</sub><sup>-</sup> has a closed-shell <sup>1</sup>A<sub>1</sub> ground state. Vibrational frequency analyses show that both *C<sub>3v</sub>* neutral and anion structures are genuine minima, with all real harmonic frequencies predicted by the four DFT methods. Our optimized structures for the neutral 1-C<sub>10</sub>F<sub>15</sub> and the anionic 1-C<sub>10</sub>F<sub>15</sub><sup>-</sup> are displayed in Figure 1, and their geometrical parameters are summarized in Tables 2 and 3 and Figure 2. The geometry predictions by the four DFT methods are in reasonably good agreement with each other, and the trend for the bond distances predicted by the different functionals is BHLYP < B3LYP < BP86 < BLYP.

Compared with the closed-shell perfluoroadamantane molecule (C<sub>10</sub>F<sub>16</sub>),<sup>7</sup> the changes for the geometry of the 1-C<sub>10</sub>F<sub>15</sub> radical are substantial, especially for those bond lengths and bond angles related to the radical carbon atom (Table 2). At the DZP++ B3LYP level, the C\*–C (C\* indicating the radical site C atom) bond distance (*r*<sub>1</sub>) decreases by 0.043 Å, whereas the C–C bond (*r*<sub>2</sub>) increases by 0.013 Å. The C–C\*–C angle (*θ*<sub>1</sub>) increases by 4.1° upon the removal of one F atom, while the C\*–C–C angle (*θ*<sub>2</sub>) decreases by 4.0°.

**Table 2.** Optimized Geometries for the 1-C<sub>10</sub>F<sub>15</sub> Radical (C<sub>3v</sub>)<sup>a</sup>

|                       | BHLYP  | B3LYP  | BP86   | BLYP   |
|-----------------------|--------|--------|--------|--------|
| r <sub>1</sub> (C*–C) | 1.516  | 1.525  | 1.528  | 1.534  |
| r <sub>2</sub>        | 1.561  | 1.581  | 1.591  | 1.602  |
| r <sub>3</sub>        | 1.555  | 1.572  | 1.580  | 1.588  |
| r <sub>4</sub>        | 1.332  | 1.350  | 1.363  | 1.370  |
| r <sub>5</sub>        | 1.340  | 1.357  | 1.368  | 1.375  |
| r <sub>6</sub>        | 1.331  | 1.350  | 1.352  | 1.368  |
| r <sub>7</sub>        | 1.331  | 1.350  | 1.362  | 1.369  |
| θ <sub>1</sub>        | 113.2° | 113.5° | 113.8° | 113.7° |
| θ <sub>2</sub>        | 105.8° | 105.6° | 105.4° | 105.6° |
| θ <sub>3</sub>        | 109.0° | 108.9° | 108.8° | 108.8° |
| θ <sub>4</sub>        | 109.3° | 109.3° | 109.7° | 109.3° |
| θ <sub>5</sub>        | 110.2° | 110.2° | 110.2° | 110.1° |

<sup>a</sup> Bond lengths are in Å and bond angles are in degrees. Geometrical parameters correspond to those identified in Figure 1.

**Table 3.** Optimized Geometries for the 1-C<sub>10</sub>F<sub>15</sub><sup>−</sup> Anion (C<sub>3v</sub>)<sup>a</sup>

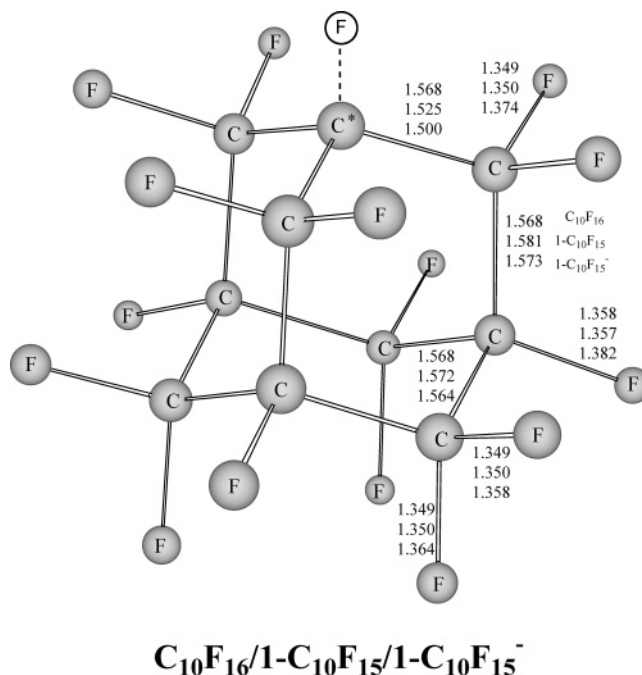
|                       | BHLYP  | B3LYP  | BP86   | BLYP   |
|-----------------------|--------|--------|--------|--------|
| r <sub>1</sub> (C*–C) | 1.494  | 1.500  | 1.503  | 1.509  |
| r <sub>2</sub>        | 1.558  | 1.573  | 1.577  | 1.587  |
| r <sub>3</sub>        | 1.548  | 1.564  | 1.572  | 1.581  |
| r <sub>4</sub>        | 1.353  | 1.374  | 1.387  | 1.395  |
| r <sub>5</sub>        | 1.360  | 1.382  | 1.399  | 1.406  |
| r <sub>6</sub>        | 1.338  | 1.358  | 1.371  | 1.378  |
| r <sub>7</sub>        | 1.345  | 1.364  | 1.376  | 1.384  |
| θ <sub>1</sub>        | 110.3° | 110.8° | 111.4° | 111.3° |
| θ <sub>2</sub>        | 110.6° | 110.1° | 109.5° | 109.8° |
| θ <sub>3</sub>        | 108.4° | 108.3° | 108.2° | 108.2° |
| θ <sub>4</sub>        | 108.9° | 108.9° | 108.8° | 108.8° |
| θ <sub>5</sub>        | 110.3° | 110.3° | 110.6° | 110.4° |

<sup>a</sup> Bond lengths are in Å and bond angles are in degrees. Geometrical parameters correspond to those identified in Figure 1.

For the 1-C<sub>10</sub>F<sub>15</sub><sup>−</sup> anion (Table 3), the geometry also differs considerably from that<sup>6</sup> of the radical anion C<sub>10</sub>F<sub>16</sub><sup>−</sup>. At the DZP++ B3LYP level, the C\*–C bond (r<sub>1</sub>) decreases by 0.051 Å, while the C–C bond distance (r<sub>2</sub>) increases by 0.022 Å. The C–C\*–C angle (θ<sub>1</sub>) increases by 1.0°, while the C\*–C–C angle (θ<sub>2</sub>) increases by 1.3°.

The differences in the bond distances between the neutral 1-C<sub>10</sub>F<sub>15</sub> radical and anionic 1-C<sub>10</sub>F<sub>15</sub><sup>−</sup> are small. The C–C bonds for the neutral are longer than those for the anion. The largest difference is for the C\*–C bonds (0.025 Å with the DZP++ B3LYP method), while the differences for other C–C bonds are only 0.008 Å (B3LYP, see Tables 2 and 3). In contrast, the C–F distances in the neutral are shorter (by 0.024, 0.025, 0.008, and 0.014 Å, respectively) than those for the anion. For the bond angles, only those close to the C\* atom display significant differences between the neutral and anion. Angle θ<sub>1</sub> decreases by 2.7° from the neutral to the anion at the DZP++ B3LYP level, and angle θ<sub>2</sub> increases by 4.5°. This may be attributed to the “last” electron, which increases the electron density on the radical carbon atom, yielding a larger C–C\*–C angle (θ<sub>1</sub>) to provide more space for the anion lone pair.

Table 4 lists the DFT predicted EA<sub>ad</sub>, EA<sub>vert</sub>, and VDE values for the title radicals. For the 1-C<sub>10</sub>F<sub>15</sub> radical, the adiabatic EA ranges from 3.85 eV (BHLYP) to 4.16 eV

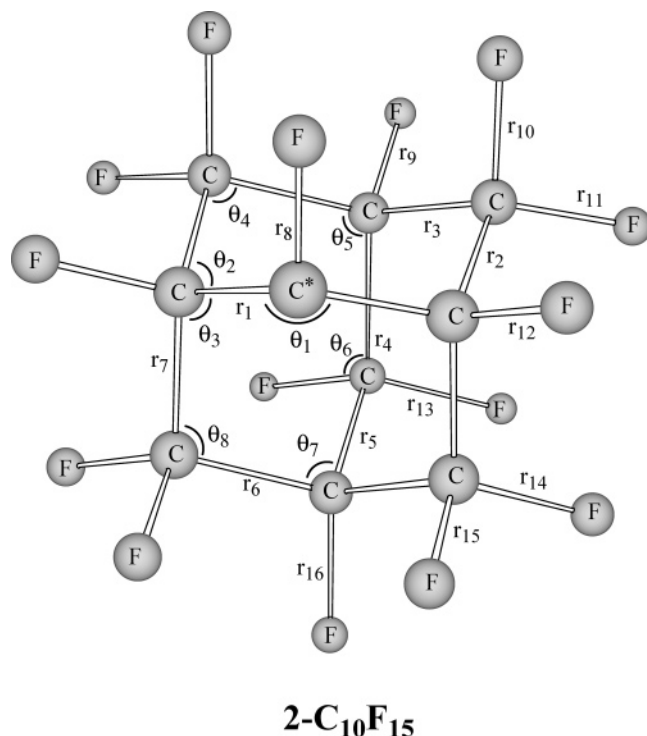
**Figure 2.** Bond distances for the 1-C<sub>10</sub>F<sub>15</sub> radical and the 1-C<sub>10</sub>F<sub>15</sub><sup>−</sup> anion. The bond distances for perfluoroadamantane C<sub>10</sub>F<sub>16</sub> are also shown for comparison.**Table 4.** Adiabatic Electron Affinities (EA<sub>ad</sub>) and Vertical Electron Affinities (EA<sub>vert</sub>) for the C<sub>10</sub>F<sub>15</sub> Radicals and Vertical Detachment Energies (VDE) for the C<sub>10</sub>F<sub>15</sub><sup>−</sup> Anions in eV (kcal/mol in parentheses)<sup>a</sup>

| compound                          | method | EA <sub>ad</sub> | EA <sub>vert</sub> | VDE          |
|-----------------------------------|--------|------------------|--------------------|--------------|
| 1-C <sub>10</sub> F <sub>15</sub> | BHLYP  | 3.85 (88.8)      | 3.49 (80.4)        | 4.22 (97.3)  |
|                                   | B3LYP  | 4.11 (94.9)      | 3.76 (86.8)        | 4.48 (103.2) |
|                                   | BP86   | 4.16 (95.9)      | 3.82 (88.0)        | 4.51 (104.0) |
|                                   | BLYP   | 4.05 (93.5)      | 3.71 (85.5)        | 4.40 (101.6) |
| 2-C <sub>10</sub> F <sub>15</sub> | BHLYP  | 2.76 (63.7)      | 2.03 (46.8)        | 3.46 (79.9)  |
|                                   | B3LYP  | 3.03 (70.0)      | 2.34 (54.0)        | 3.68 (85.0)  |
|                                   | BP86   | 3.10 (71.4)      | 2.44 (56.3)        | 3.69 (85.1)  |
|                                   | BLYP   | 3.02 (69.6)      | 2.37 (54.6)        | 3.62 (83.4)  |

<sup>a</sup> Values are not corrected for ZPVE and were obtained with the DZP++ basis sets.

(BP86). The trend of the theoretical EAs is in the order of BP86 > B3LYP > BLYP > BHLYP. The B3LYP EA<sub>ad</sub> prediction of 4.11 eV is expected to be the most reliable. The value of EA<sub>vert</sub> is predicted to be 3.76 eV (B3LYP), while the VDE is 4.48 eV (B3LYP). Yan, Brinkmann, and Schaefer<sup>1</sup> predicted the EA<sub>ad</sub> for the two C<sub>10</sub>H<sub>15</sub> radicals to be negative (−0.13 eV with the B3LYP functional). Our results show that the perfluorination of adamantyl radical dramatically improves the ability to bind an electron.

It should also be noted that the predicted electron affinity of the 1-C<sub>10</sub>F<sub>15</sub> radical is also much larger than that for the closed-shell C<sub>10</sub>F<sub>16</sub> molecule (1.06 eV, B3LYP),<sup>7</sup> demonstrating that the radical has a much stronger tendency than the corresponding closed-shell molecule to bind an additional electron. This is, of course, because the addition of an electron to the radical makes it a stable closed-shell system, while an extra electron added to the stable closed-shell C<sub>10</sub>F<sub>16</sub> molecule created a radical anion. Nevertheless, the magnitude



**Figure 3.** Sketch of the  $C_s$  structures for the 2- $C_{10}F_{15}$  radical and the 2- $C_{10}F_{15}^-$  anion. The optimized geometrical parameters for these structures are reported in Tables 5 and 6, respectively.

of the increase (3.05 eV) in EA between  $C_{10}F_{16}$  and 1- $C_{10}F_{15}$  is large.

**B. 2- $C_{10}F_{15}$  and 2- $C_{10}F_{15}^-$ .** The second isomer 2- $C_{10}F_{15}$  has a fluorine atom removed from one bridging carbon atom of  $C_{10}F_{16}$ , and this radical possesses  $C_s$  symmetry. The structures predicted for the neutral 2- $C_{10}F_{15}$  and the analogous anionic 2- $C_{10}F_{15}^-$  are displayed in Figure 3, and the corresponding geometrical parameters are summarized in Tables 5 and 6 and Figure 4. The neutral radical 2- $C_{10}F_{15}$  has a  $^2A'$  ground electronic state, while the closed-shell anionic 2- $C_{10}F_{15}^-$  has a  $^1A'$  ground state. Vibrational analyses show that both the neutral 2- $C_{10}F_{15}$  and the anion 2- $C_{10}F_{15}^-$  are genuine minima.

Compared with the closed-shell  $C_{10}F_{16}$  molecule, the  $C^*-C$  bond distances of the 2- $C_{10}F_{15}$  ( $r_1$ ) radical decrease by 0.053 Å, and the other  $C-C$  bonds ( $r_2$  to  $r_7$ ) have very small changes (with differences less than 0.004 Å, Table 5). The  $C-C^*-C$  angle ( $\theta_1$ ) increases by 5.7° at the DZP++ B3LYP level, while the changes for other angles ( $\theta_2$ - $\theta_8$ ) are relatively small (Table 5).

For the closed-shell anionic 2- $C_{10}F_{15}^-$ , the geometry changes with respect to  $C_{10}F_{16}$  are similar to that predicted for 1- $C_{10}F_{15}^-$ . At the DZP++ B3LYP level, the  $C^*-C$  bonds ( $r_1$ ) decrease by 0.044 Å, while the other  $C-C$  bonds increase by smaller amounts, i.e., by 0.017 Å ( $r_2$ ), 0.005 Å ( $r_3$ ), -0.005 Å ( $r_4$ ), -0.006 Å ( $r_5$ ), -0.014 Å ( $r_6$ ) and -0.011 Å ( $r_7$ ). The  $C-C^*-C$  angle ( $\theta_1$ ) and the  $C^*-C-C$  angle ( $\theta_2$ ) increase by 1.0° and 3.5°, respectively, while the changes for other angles are insignificant (Table 6).

The  $C-C$  bond distances do not reveal significant differences between the neutral radical 2- $C_{10}F_{15}$  and its anion

**Table 5.** Optimized Geometries for the 2- $C_{10}F_{15}$  Radical ( $C_s$ )<sup>a</sup>

|                     | BHLYP  | B3LYP  | BP86   | BLYP   |
|---------------------|--------|--------|--------|--------|
| $r_1(C^*-C)$        | 1.507  | 1.515  | 1.517  | 1.524  |
| $r_2$               | 1.551  | 1.570  | 1.579  | 1.589  |
| $r_3$               | 1.555  | 1.571  | 1.580  | 1.588  |
| $r_4$               | 1.553  | 1.570  | 1.579  | 1.588  |
| $r_5$               | 1.553  | 1.570  | 1.578  | 1.587  |
| $r_6$               | 1.551  | 1.567  | 1.574  | 1.583  |
| $r_7$               | 1.546  | 1.564  | 1.573  | 1.582  |
| $r_8(C^*-F)$        | 1.308  | 1.324  | 1.334  | 1.341  |
| $r_9$               | 1.341  | 1.359  | 1.369  | 1.376  |
| $r_{10}$            | 1.334  | 1.353  | 1.365  | 1.371  |
| $r_{11}$            | 1.330  | 1.348  | 1.360  | 1.367  |
| $r_{12}$            | 1.342  | 1.360  | 1.371  | 1.379  |
| $r_{13}$            | 1.331  | 1.349  | 1.361  | 1.368  |
| $r_{14}$            | 1.330  | 1.348  | 1.360  | 1.366  |
| $r_{15}$            | 1.333  | 1.352  | 1.364  | 1.371  |
| $r_{16}$            | 1.341  | 1.359  | 1.370  | 1.377  |
| $\theta_1(C-C^*-C)$ | 114.8° | 115.3° | 115.6° | 115.6° |
| $\theta_2$          | 108.4° | 108.4° | 108.3° | 108.4° |
| $\theta_3$          | 107.5° | 107.5° | 107.5° | 107.6° |
| $\theta_4$          | 109.0° | 108.9° | 108.8° | 108.8° |
| $\theta_5$          | 109.3° | 109.3° | 109.2° | 109.2° |
| $\theta_6$          | 110.0° | 110.0° | 110.0° | 110.1° |
| $\theta_7$          | 109.2° | 109.2° | 109.2° | 109.2° |
| $\theta_8$          | 109.5° | 109.5° | 109.5° | 109.5° |

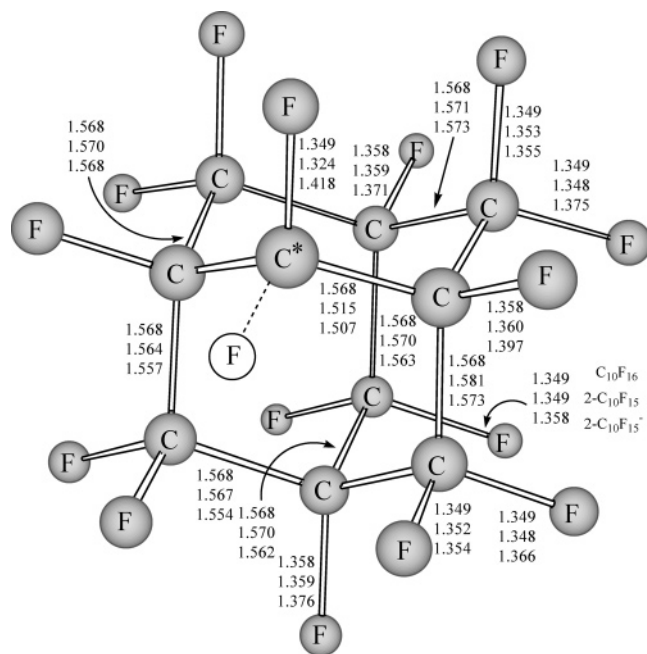
<sup>a</sup> Bond lengths are in Å and bond angles are in degrees. Geometrical parameters correspond to those identified in Figure 3.

**Table 6.** Optimized Geometries for the 2- $C_{10}F_{15}^-$  Anion ( $C_s$ )<sup>a</sup>

| method              | BHLYP  | B3LYP  | BP86   | BLYP   |
|---------------------|--------|--------|--------|--------|
| $r_1(C^*-C)$        | 1.505  | 1.507  | 1.507  | 1.512  |
| $r_2$               | 1.552  | 1.568  | 1.571  | 1.581  |
| $r_3$               | 1.554  | 1.573  | 1.584  | 1.593  |
| $r_4$               | 1.547  | 1.563  | 1.571  | 1.580  |
| $r_5$               | 1.546  | 1.562  | 1.567  | 1.576  |
| $r_6$               | 1.541  | 1.554  | 1.559  | 1.567  |
| $r_7$               | 1.543  | 1.557  | 1.564  | 1.572  |
| $r_8(C^*-F)$        | 1.405  | 1.418  | 1.420  | 1.431  |
| $r_9$               | 1.353  | 1.371  | 1.384  | 1.391  |
| $r_{10}$            | 1.336  | 1.355  | 1.368  | 1.374  |
| $r_{11}$            | 1.352  | 1.375  | 1.393  | 1.401  |
| $r_{12}$            | 1.371  | 1.397  | 1.418  | 1.427  |
| $r_{13}$            | 1.339  | 1.358  | 1.371  | 1.378  |
| $r_{14}$            | 1.347  | 1.366  | 1.378  | 1.385  |
| $r_{15}$            | 1.334  | 1.354  | 1.367  | 1.374  |
| $r_{16}$            | 1.354  | 1.376  | 1.393  | 1.401  |
| $\theta_1(C-C^*-C)$ | 109.1° | 109.8° | 110.6° | 110.6° |
| $\theta_2$          | 113.6° | 113.3° | 112.6° | 112.8° |
| $\theta_3$          | 108.1° | 108.4° | 108.6° | 108.6° |
| $\theta_4$          | 107.9° | 107.5° | 107.1° | 107.1° |
| $\theta_5$          | 109.2° | 109.2° | 109.2° | 109.2° |
| $\theta_6$          | 110.5° | 110.5° | 110.4° | 110.5° |
| $\theta_7$          | 109.2° | 109.1° | 109.0° | 109.1° |
| $\theta_8$          | 109.3° | 109.3° | 109.5° | 109.4° |

<sup>a</sup> Bond lengths are in Å and bond angles are in degrees. Geometrical parameters correspond to those identified in Figure 3.

2- $C_{10}F_{15}^-$ . With the DZP++ B3LYP method, the  $C-C$  bond lengths for the neutral 2- $C_{10}F_{15}$  are predicted to be 1.515–1.571 Å, and those for the anionic species are 1.507–1.573



**C<sub>10</sub>F<sub>16</sub>/2-C<sub>10</sub>F<sub>15</sub>/2-C<sub>10</sub>F<sub>15</sub><sup>-</sup>**

**Figure 4.** Bond distances for the 2-C<sub>10</sub>F<sub>15</sub> radical and the 2-C<sub>10</sub>F<sub>15</sub><sup>-</sup> anion. The bond distances for perfluoroadamantane C<sub>10</sub>F<sub>16</sub> are also shown for comparison.

Å, with the C–C bonds for the neutral being slightly longer than those for the anion. For most C–F bonds ( $r_9$  to  $r_{16}$ ), the differences between the neutral and the anion are quite small, with those in the neutral being slightly shorter than those in anion. However, the C\*–F bond ( $r_8$ ) predicted for the neutral is 0.094 Å shorter than that in the anion. This is the largest structural difference predicted for these adamantane systems.

The trend for the predicted C–C and C–F bond lengths is in the order of BHLYP < B3LYP < BP86 < BLYP. The C–C\*–C angle ( $\theta_1$ ) in the neutral is larger than that in the anion by 5.5°, while the C\*–C–C angles ( $\theta_2$ ) in the neutral are smaller by 4.9° (B3LYP). The differences in other angles between the neutral and the anion are quite small. It is also found that the four methods predict very similar values for the bond angles (Tables 5 and 6).

Energetically, the four DFT methods predict that the neutral 2-C<sub>10</sub>F<sub>15</sub> radical lies close to the neutral 1-C<sub>10</sub>F<sub>15</sub> isomer. (Total energies are shown in Table 1.) For instance, the B3LYP method predicts the radical structure 2-C<sub>10</sub>F<sub>15</sub> lies below 1-C<sub>10</sub>F<sub>15</sub> by only 2.0 kcal/mol. However, the anionic 2-C<sub>10</sub>F<sub>15</sub><sup>-</sup> isomer lies much higher than the anionic 1-C<sub>10</sub>F<sub>15</sub><sup>-</sup> isomer (e.g., by 23.0 kcal/mol at B3LYP). This is because the orbital energy for the singly occupied molecular orbital ( $a_1$ ) for 1-C<sub>10</sub>F<sub>15</sub> (–0.206 hartree at B3LYP) is much lower than that ( $a'$ ) for 2-C<sub>10</sub>F<sub>15</sub> (–0.157 hartree at B3LYP). Consequently, the neutral-anion energy separations (including  $EA_{ad}$ ) of the isomer 2-C<sub>10</sub>F<sub>15</sub> will be much smaller than those of the isomer 1-C<sub>10</sub>F<sub>15</sub>. The  $EA_{ad}$  and  $EA_{vert}$  values for the 2-C<sub>10</sub>F<sub>15</sub> radical and the VDE for its 2-C<sub>10</sub>F<sub>15</sub><sup>-</sup> anion are reported in Table 4 with four DFT methods. The BHLYP method predicts the lowest  $EA_{ad}$  value of 2.76 eV, followed by BLYP (3.02 eV) and B3LYP (3.03 eV), while the BP86

**Table 7.** Harmonic Vibrational Frequencies (in cm<sup>-1</sup>) for the 1-C<sub>10</sub>F<sub>15</sub> Radical and the 1-C<sub>10</sub>F<sub>15</sub><sup>-</sup> Anion with the B3LYP Method<sup>a</sup>

| sym            | neutral        | anion      |         |
|----------------|----------------|------------|---------|
| a <sub>1</sub> | 192 (0)        | 194 (0)    |         |
|                | 255 (1)        | 250 (5)    |         |
|                | 294 (4)        | 308 (0)    |         |
|                | 364 (0)        | 308 (5)    |         |
|                | 457 (17)       | 416 (3)    |         |
|                | 614 (15)       | 438 (5)    |         |
|                | 654 (6)        | 593 (4)    |         |
|                | 717 (3)        | 638 (10)   |         |
|                | 981 (273)      | 684 (2)    |         |
|                | 1069 (3)       | 954 (498)  |         |
|                | 1236 (2)       | 1178 (26)  |         |
|                | 1270 (1)       | 1210 (1)   |         |
|                | 1282 (435)     | 1249 (421) |         |
|                | 1311 (24)      | 1282 (30)  |         |
|                | a <sub>2</sub> | 129 (0)    | 129 (0) |
|                |                | 223 (0)    | 234 (0) |
|                |                | 244 (0)    | 257 (0) |
|                |                | 269 (0)    | 277 (0) |
|                |                | 385 (0)    | 385 (0) |
| 427 (1)        |                | 840 (0)    |         |
| 847 (0)        |                | 1037 (0)   |         |
| 1082 (0)       |                | 1054 (35)  |         |
| 1154 (0)       |                | 1141 (0)   |         |
| e              |                | 136 (0)    | 138 (0) |
|                |                | 194 (0)    | 195 (0) |
|                |                | 204 (0)    | 213 (0) |
|                |                | 242 (0)    | 247 (0) |
|                |                | 252 (0)    | 256 (0) |
|                |                | 262 (0)    | 264 (0) |
|                |                | 294 (8)    | 298 (2) |
|                |                | 296 (0)    | 300 (2) |
|                |                | 379 (2)    | 383 (0) |
|                |                | 399 (10)   | 409 (8) |
|                | 431 (6)        | 431 (0)    |         |
|                | 599 (0)        | 589 (0)    |         |
|                | 642 (36)       | 636 (30)   |         |
|                | 779 (0)        | 779 (16)   |         |
|                | 947 (400)      | 925 (620)  |         |
|                | 995 (60)       | 980 (44)   |         |
|                | 1061 (0)       | 1021 (2)   |         |
|                | 1070 (78)      | 1055 (12)  |         |
|                | 1154 (28)      | 1146 (124) |         |
| 1224 (12)      | 1170 (0)       |            |         |
| 1245 (0)       | 1216 (410)     |            |         |
| 1277 (532)     | 1241 (128)     |            |         |
| 1304 (648)     | 1286 (934)     |            |         |

<sup>a</sup> The infrared intensities in km/mol are listed in parentheses.

gives the highest (3.10 eV). Among these we conclude that the B3LYP result 3.03 eV is the most reliable, being 1.08 eV smaller than that for 1-C<sub>10</sub>F<sub>15</sub>. The B3LYP method also predicts the  $EA_{vert}$  value to be 2.34 eV and the VDE value for the anion to be 3.68 eV (Table 4). Although the neutral-anion energy separations for 2-C<sub>10</sub>F<sub>15</sub>/C<sub>10</sub>F<sub>15</sub><sup>-</sup> are smaller than those for 1-C<sub>10</sub>F<sub>15</sub>/C<sub>10</sub>F<sub>15</sub><sup>-</sup>, they are indeed much larger than those for the C<sub>10</sub>F<sub>16</sub> molecule.<sup>7</sup>

**C. Vibrational Frequencies and ZPVE Corrections.** The B3LYP predicted harmonic vibrational frequencies for the 1-C<sub>10</sub>F<sub>15</sub>/C<sub>10</sub>F<sub>15</sub><sup>-</sup> systems are reported in Table 7, and those

**Table 8.** Harmonic Vibrational Frequencies (in  $\text{cm}^{-1}$ ) for the 2- $\text{C}_{10}\text{F}_{15}$  Radical and the 2- $\text{C}_{10}\text{F}_{15}^-$  Anion with the B3LYP Method<sup>a</sup>

| sym | neutral    | anion      |
|-----|------------|------------|
| a'  | 113 (0)    | 137 (0)    |
|     | 160 (0)    | 184 (0)    |
|     | 177 (0)    | 193 (1)    |
|     | 209 (0)    | 206 (1)    |
|     | 246 (0)    | 239 (1)    |
|     | 250 (0)    | 250 (1)    |
|     | 252 (1)    | 252 (1)    |
|     | 258 (1)    | 266 (1)    |
|     | 270 (0)    | 272 (0)    |
|     | 294 (4)    | 295 (1)    |
|     | 298 (1)    | 300 (2)    |
|     | 320 (2)    | 318 (0)    |
|     | 369 (2)    | 371 (3)    |
|     | 391 (1)    | 391 (1)    |
|     | 406 (3)    | 401 (11)   |
|     | 411 (9)    | 409 (2)    |
|     | 464 (8)    | 448 (8)    |
|     | 540 (5)    | 546 (12)   |
|     | 567 (1)    | 549 (0)    |
|     | 628 (8)    | 623 (6)    |
|     | 658 (9)    | 656 (10)   |
|     | 691 (0)    | 678 (1)    |
|     | 848 (0)    | 832 (12)   |
|     | 961 (276)  | 918 (245)  |
|     | 978 (298)  | 942 (344)  |
|     | 1009 (21)  | 966 (164)  |
|     | 1055 (8)   | 1018 (5)   |
|     | 1064 (1)   | 1034 (7)   |
|     | 1090 (4)   | 1078 (2)   |
|     | 1206 (29)  | 1126 (63)  |
|     | 1239 (19)  | 1130 (5)   |
|     | 1254 (169) | 1178 (41)  |
|     | 1265 (174) | 1205 (48)  |
|     | 1277 (196) | 1243 (82)  |
|     | 1290 (74)  | 1265 (87)  |
|     | 1294 (8)   | 1267 (265) |
|     | 1313 (96)  | 1277 (229) |
|     | 1390 (134) | 1288 (40)  |
| a'' | 107 (0)    | 124 (0)    |
|     | 115 (0)    | 130 (0)    |
|     | 182 (0)    | 192 (0)    |
|     | 200 (0)    | 206 (0)    |
|     | 232 (0)    | 226 (0)    |
|     | 246 (0)    | 242 (0)    |
|     | 249 (0)    | 248 (0)    |
|     | 257 (1)    | 262 (0)    |
|     | 264 (1)    | 263 (1)    |
|     | 268 (0)    | 270 (0)    |
|     | 279 (3)    | 280 (0)    |
|     | 298 (4)    | 302 (2)    |
|     | 384 (0)    | 378 (1)    |
|     | 385 (0)    | 382 (0)    |
|     | 406 (4)    | 407 (1)    |
|     | 434 (9)    | 433 (3)    |
|     | 591 (1)    | 593 (2)    |
|     | 645 (15)   | 638 (15)   |
|     | 795 (1)    | 757 (6)    |
|     | 845 (0)    | 838 (0)    |
|     | 948 (215)  | 913 (311)  |
|     | 991 (36)   | 971 (13)   |
|     | 1063 (4)   | 1010 (26)  |
|     | 1083 (6)   | 1046 (49)  |
|     | 1119 (0)   | 1085 (10)  |
|     | 1148 (4)   | 1126 (38)  |
|     | 1170 (0)   | 1157 (58)  |
|     | 1228 (4)   | 1197 (343) |
|     | 1244 (22)  | 1215 (19)  |
|     | 1287 (533) | 1234 (1)   |
|     | 1313 (1)   | 1271 (212) |

<sup>a</sup> The infrared intensities in  $\text{km/mol}$  are listed in parentheses.**Table 9.** Zero-Point Vibrational Energies (ZPVE) and ZPVE-Corrected Adiabatic Electron Affinities  $\text{EA}_{\text{ad}}(\text{ZPVE})$  in eV ( $\text{kcal/mol}$  in parentheses) for the  $\text{C}_{10}\text{F}_{15}$  Radicals

| compound                             | BHLYP       | B3LYP       | BP86        | BLYP        |
|--------------------------------------|-------------|-------------|-------------|-------------|
| 1- $\text{C}_{10}\text{F}_{15}$      | 3.06 (70.6) | 2.87 (66.1) | 2.74 (63.1) | 2.69 (62.0) |
| 1- $\text{C}_{10}\text{F}_{15}^-$    | 3.02 (69.6) | 2.82 (65.0) | 2.69 (62.0) | 2.64 (60.8) |
| $\Delta(\text{ZPVE})$                | 0.04 (1.0)  | 0.05 (1.1)  | 0.05 (1.1)  | 0.05 (1.2)  |
| $\text{EA}_{\text{ad}}(\text{ZPVE})$ | 3.89 (89.8) | 4.16 (96.0) | 4.21 (97.0) | 4.10 (94.7) |
| 2- $\text{C}_{10}\text{F}_{15}$      | 3.06 (70.6) | 2.87 (66.2) | 2.74 (63.2) | 2.70 (62.2) |
| 2- $\text{C}_{10}\text{F}_{15}^-$    | 3.00 (69.2) | 2.80 (64.6) | 2.66 (61.4) | 2.62 (60.3) |
| $\Delta(\text{ZPVE})$                | 0.06 (1.4)  | 0.07 (1.6)  | 0.08 (1.8)  | 0.08 (1.9)  |
| $\text{EA}_{\text{ad}}(\text{ZPVE})$ | 2.82 (65.1) | 3.10 (71.6) | 3.18 (73.2) | 3.10 (71.5) |

for the 2- $\text{C}_{10}\text{F}_{15}/\text{C}_{10}\text{F}_{15}^-$  species are listed in Table 8. All of the vibrational frequencies are real, suggesting that all these systems (neutral and anion) are genuine minima on their potential hypersurfaces. Although no experimental results are available for these systems, the B3LYP predictions should be qualitatively reliable.

When molecular systems under study become larger, the ZPVE corrections to electron affinities are sometimes substantial. In a study of polycyclic aromatic hydrocarbon (PAH) EAs,<sup>9</sup> the ZPVE corrections were found to be several tenths of an electronvolt and to account for a significant portion (as much as 40% for coronene) of the overall EAs. In research on the unsubstituted hydrocarbon  $\text{C}_{10}\text{H}_{15}$  radicals,<sup>1</sup> the ZPVE corrections were found to be 0.13 and 0.10 eV, respectively, corresponding to the same magnitude as the  $\text{EA}_{\text{ad}}$  values.

Zero-point vibrational energies (ZPVEs) for the perfluoroadamantyl radicals and the related anions are presented in Table 9. The ZPVE differences between the neutrals and the anions are in the range 0.04–0.08 eV, which are then used for the correction of the electron affinities. After the ZPVE corrections the  $\text{EA}_{\text{ad}}$  values (B3LYP) increase to 4.16 eV for 1- $\text{C}_{10}\text{F}_{15}$  and 3.10 eV for 2- $\text{C}_{10}\text{F}_{15}$  (Table 9). Compared with the aromatic radicals<sup>5</sup>  $\text{C}_6\text{F}_5$ ,  $\text{C}_{10}\text{F}_7$  and  $\text{C}_{14}\text{F}_9$ , the ZPVE corrections for the perfluoroadamantyl radicals are relatively small, i.e., 0.05 eV (1.2%) for 1- $\text{C}_{10}\text{F}_{15}$  and 0.07 eV (2.3%) for 2- $\text{C}_{10}\text{F}_{15}$  at the B3LYP level.

## Concluding Remarks

There are two forms of perfluoroadamantyl radicals ( $\text{C}_{10}\text{F}_{15}$ ): 1- $\text{C}_{10}\text{F}_{15}$  with  $C_{3v}$  symmetry and 2- $\text{C}_{10}\text{F}_{15}$  with  $C_s$  symmetry. The neutral 1- $\text{C}_{10}\text{F}_{15}$  isomer has a slightly higher energy (by 0.086 eV or 1.98 kcal/mol, at the B3LYP level) than the 2- $\text{C}_{10}\text{F}_{15}$  radical, while the anionic 1- $\text{C}_{10}\text{F}_{15}^-$  isomer lies below the 2- $\text{C}_{10}\text{F}_{15}^-$  anion by 1.00 eV (23.0 kcal/mol). Both isomers are predicted to have large positive adiabatic electron affinities. The reliable predicted adiabatic electron affinities with the DZP++ B3LYP method are 4.11 and 3.03 eV for the two perfluoroadamantyl radicals, respectively. It is shown that the perfluoroadamantyl radicals have considerable ability to bind an additional electron and may thus be useful in the study of new materials and new chemical reactions. The ZPVE corrections for the  $\text{EA}_{\text{ad}}$  values are not substantial, i.e., 0.05 eV (1.2%) for 1- $\text{C}_{10}\text{F}_{15}$  and 0.07 eV (2.3%) for 2- $\text{C}_{10}\text{F}_{15}$  at the DZP++ B3LYP level, and the ZPVE-corrected  $\text{EA}_{\text{ad}}$  are 4.16 and 3.10 eV, respectively. It



is hoped that the present theoretical predictions will stimulate further experimental studies on these and related systems. In light of the very large electron affinity of the 1-C<sub>10</sub>F<sub>15</sub> radical, it would be of special interest to examine the perfluorinated superadamantyl systems<sup>19</sup> C<sub>35</sub>F<sub>36</sub>, C<sub>35</sub>F<sub>35</sub>, and C<sub>35</sub>F<sub>35</sub><sup>-</sup>.

**Acknowledgment.** We appreciate the support of the National Science Foundation of China. U.S. National Science Foundation Grant (CHE-0136186) at the University of Georgia is similarly acknowledged.

**Supporting Information Available:** Cartesian coordinates for all molecular structures predicted by the B3LYP method (Table S1). This material is available free of charge via the Internet at <http://pubs.acs.org>.

### References

- (1) Yan, G.; Brinkmann, N. R.; Schaefer, H. F. *J. Phys. Chem. A* **2003**, *107*, 9479.
- (2) In contrast, the methyl radical does have a small electron affinity  $0.08 \pm 0.03$  eV, see: Ellison, G. B.; Engelking, P. C.; Lineberger, W. C. *J. Am. Chem. Soc.* **1978**, *100*, 2556. The present B3LYP and BLYP methods also predict EA-(CH<sub>3</sub>) to be positive, namely 0.03 and 0.05 eV, respectively.
- (3) Beck, C. M.; Burdeniuc, J.; Crabtree, R. H.; Rheingold, A. L.; Yap, G. P. *Inorg. Chem. Acta* **1998**, *270*, 559.
- (4) Alkorta, I.; Rozas, I.; Elguero, J. *J. Am. Chem. Soc.* **2002**, *124*, 8593.
- (5) Xie, Y.; Schaefer, H. F.; Cotton, F. A. *Chem. Commun.* **2003**, 102.
- (6) Li, Q.; Feng, X.; Xie, Y.; Schaefer, H. F. *J. Phys. Chem. A* **2004**, *108*, 7071.

- (7) Li, Q.; Feng, X.; Xie, Y.; Schaefer, H. F. *J. Phys. Chem. A* (in press).
- (8) Rienstra-Kiracofe, J. C.; Tschumper, G. S.; Schaefer, H. F.; Nandi, S.; Ellison, G. B. *Chem. Rev.* **2002**, *102*, 231.
- (9) Rienstra-Kiracofe, J. C.; Barden, C. J.; Brown, S. T.; Schaefer, H. F. *J. Phys. Chem. A* **2001**, *105*, 524.
- (10) The BHandHLYP method implemented in the Gaussian programs has the formula,  $0.5*Ex(LSDA)+0.5*Ex(HF)+0.5*Delta-Ex(B88)+Ec(LYP)$ , which is *not* precisely the formulation proposed by A. D. Becke in his paper. *J. Chem. Phys.* **1993**, *98*, 1372.
- (11) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785.
- (12) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (13) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- (14) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822; **1986**, *34*, 7406.
- (15) Huzinaga, S. *J. Chem. Phys.* **1965**, *42*, 1293.
- (16) Dunning, T. H. *J. Chem. Phys.* **1970**, *53*, 2823.
- (17) Lee, T. J.; Schaefer, H. F. *J. Chem. Phys.* **1985**, *83*, 1784.
- (18) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian Inc.: Pittsburgh, PA, 1995.
- (19) Shen, M.; Schaefer, H. F.; Liang, C.; Lii, J.-H.; Allinger, N. L.; Schleyer, P. R. *J. Am. Chem. Soc.* **1992**, *114*, 497.

CT049876C

## Activation of H–H, C–H, C–C and C–Cl Bonds by Pd and PdCl<sup>−</sup>. Understanding Anion Assistance in C–X Bond Activation

Axel Diefenbach,<sup>†</sup> G. Theodoor de Jong,<sup>‡</sup> and F. Matthias Bickelhaupt<sup>\*‡</sup>

*Fachbereich Chemie der Philipps-Universität Marburg, Hans-Meerwein-Strasse, D-35032 Marburg, Germany, and Afdeling Theoretische Chemie, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands*

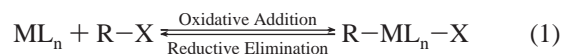
Received September 2, 2004

**Abstract:** To understand the mechanism of anion assistance in palladium-catalyzed H–H, C–H, C–C and C–Cl bond activation, several mechanistic pathways for oxidative addition of Pd and PdCl<sup>−</sup> to H<sub>2</sub> (H–H), CH<sub>4</sub> (C–H), C<sub>2</sub>H<sub>6</sub> (C–C and C–H) and CH<sub>3</sub>Cl (C–Cl) were studied uniformly at the ZORA-BP86/TZ(2)P level of relativistic nonlocal density functional theory (DFT). Oxidative addition of the neutral, uncoordinated Pd atom proceeds, as reported earlier, via direct oxidative insertion ( $\Delta H^\ddagger_{298}$  is −22 to 10 kcal/mol), whereas straight S<sub>N</sub>2 substitution (yielding, e.g., PdCH<sub>3</sub><sup>+</sup> + X<sup>−</sup>) is highly endothermic (144–237 kcal/mol) and thus not competitive. Anion assistance (i.e., going from Pd to PdCl<sup>−</sup>) lowers all activation barriers and increases the exothermicity of all model reactions studied. The effect is however selective: it favors the highly endothermic S<sub>N</sub>2 mechanism over direct oxidative insertion (OxIn). Activation enthalpies  $\Delta H^\ddagger_{298}$  for oxidative insertion of PdCl<sup>−</sup> increase along C–H (−14.0 and −13.5 kcal/mol for CH<sub>4</sub> and C<sub>2</sub>H<sub>6</sub>)  $\approx$  C–Cl (−11.2 kcal/mol) < C–C (6.4 kcal/mol), i.e., essentially in the same order as for neutral Pd. Interestingly, in case of PdCl<sup>−</sup> + CH<sub>3</sub>Cl, the two-step mechanism of S<sub>N</sub>2 substitution followed by leaving-group rearrangement becomes the preferred mechanism for oxidative addition. The highest overall barrier of this pathway (−20.2 kcal/mol) drops below the barrier for direct oxidative insertion (−11.2 kcal/mol). The effect of anion assistance is analyzed using the Activation Strain model in which activation energies  $\Delta E^\ddagger$  are decomposed into the activation strain  $\Delta E^\ddagger_{\text{strain}}$  and the stabilizing transition state (TS) interaction  $\Delta E^\ddagger_{\text{int}}$  between the reactants in the activated complex:  $\Delta E^\ddagger = \Delta E^\ddagger_{\text{strain}} + \Delta E^\ddagger_{\text{int}}$ . For each type of activated bond and reaction mechanism, the activation strain  $\Delta E^\ddagger_{\text{strain}}$  adopts characteristic values which differ only moderately, within a relatively narrow range, between corresponding reactions of Pd and PdCl<sup>−</sup>. The lowering of activation barriers through anion assistance is caused by the TS interaction  $\Delta E^\ddagger_{\text{int}}$  becoming more stabilizing.

### 1. Introduction

Oxidative addition (eq 1) is a key step in many catalytic reactions<sup>1</sup> and has been intensively investigated both

experimentally<sup>2–7</sup> and theoretically.<sup>5,7–12</sup>



In the present study, we aim at understanding how anion assistance, i.e., the introduction of an additional, negatively charged ligand, influences a catalyst's electronic structure and how this in turn affects its reactivity toward the archetypical H–H, C–H, C–C and C–Cl bonds in H<sub>2</sub>, CH<sub>4</sub>,

\* Corresponding author fax: +31–20–59 87629; e-mail: FM.Bickelhaupt@few.vu.nl.

<sup>†</sup> Fachbereich Chemie der Philipps-Universität Marburg. Present address: Sercon GmbH.

<sup>‡</sup> Scheikundig Laboratorium der Vrije Universiteit.



**Table 1.** Reaction Profiles for the Oxidative Insertion (OxIn), S<sub>N</sub>2 Substitution and S<sub>N</sub>2/Cl-Rearrangement Reactions of Pd (from Ref 12a) and PdCl<sup>-</sup> (This Work) with H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub> and CH<sub>3</sub>Cl, Respectively: 298 K Enthalpies (in kcal/mol) Relative to Reactants<sup>a</sup>

| activated bond                    | reactants   | reactant complex            | transition state            | product                                  |
|-----------------------------------|---|-----------------------------|-----------------------------|--|
| Oxidative Insertion               |   |                             |                             |  |
| H–H                               | Pd + H <sub>2</sub>                               | -24.4 (-24.1 <sup>b</sup> ) | -21.7 (-21.7 <sup>b</sup> ) | -21.1 <sup>c</sup> (-22.2 <sup>b</sup> ) |
|                                   | PdCl <sup>-</sup> + H <sub>2</sub>                | -36.3                       | <sup>d</sup>                | <sup>d</sup>                             |
| C–H                               | Pd + CH <sub>4</sub>                              | -11.4                       | -5.0                        | -9.7                                     |
|                                   | PdCl <sup>-</sup> + CH <sub>4</sub>               | -17.5                       | -14.0                       | -15.5                                    |
|                                   | Pd + C <sub>2</sub> H <sub>6</sub>                | -11.6                       | -4.1                        | -11.6 <sup>e</sup>                       |
| C–C                               | PdCl <sup>-</sup> + C <sub>2</sub> H <sub>6</sub> | -17.7                       | -13.5                       | -16.1 <sup>f</sup>                       |
|                                   | Pd + C <sub>2</sub> H <sub>6</sub>                | -11.6                       | 9.6                         | -14.1                                    |
|                                   | PdCl <sup>-</sup> + C <sub>2</sub> H <sub>6</sub> | -17.7                       | 6.4                         | -16.0                                    |
| C–Cl                              | Pd + CH <sub>3</sub> Cl                           | -15.6                       | -6.0                        | -35.7                                    |
|                                   | PdCl <sup>-</sup> + CH <sub>3</sub> Cl            | -20.7                       | -11.2                       | -58.7                                    |
| S <sub>N</sub> 2 Substitution     |   |                             |                             |  |
| H–H                               | Pd + H <sub>2</sub>                               | -24.4 <sup>g</sup>          | <sup>h</sup>                | 237.0 <sup>i</sup>                       |
|                                   | PdCl <sup>-</sup> + H <sub>2</sub>                | -36.3 <sup>g</sup>          | <sup>h</sup>                | 78.5 <sup>i</sup>                        |
| C–H                               | Pd + CH <sub>4</sub>                              | -11.4                       | <sup>h</sup>                | 228.7 <sup>i</sup>                       |
|                                   | PdCl <sup>-</sup> + CH <sub>4</sub>               | -17.5                       | <sup>h</sup>                | 88.2 <sup>i</sup>                        |
| C–C                               | Pd + C <sub>2</sub> H <sub>6</sub>                | -11.6                       | <sup>h</sup>                | 228.6 <sup>i</sup>                       |
|                                   | PdCl <sup>-</sup> + C <sub>2</sub> H <sub>6</sub> | -17.7                       | <sup>h</sup>                | 88.1 <sup>i</sup>                        |
| C–Cl                              | Pd + CH <sub>3</sub> Cl                           | -10.1                       | <sup>h</sup>                | 143.5 <sup>i</sup>                       |
|                                   | PdCl <sup>-</sup> + CH <sub>3</sub> Cl            | -23.9                       | -22.8 <sup>j</sup>          | (-24.1) <sup>k</sup> 3.0 <sup>i</sup>    |
| S <sub>N</sub> 2/Cl-Rearrangement |   |                             |                             |  |
| C–Cl                              | Pd + CH <sub>3</sub> Cl                           | -10.1                       | 21.2                        | -35.7                                    |
|                                   | PdCl <sup>-</sup> + CH <sub>3</sub> Cl            | -23.9                       | -22.8 <sup>j</sup>          | (-24.1) <sup>k</sup>                     |
|                                   |   |                             | -20.2 <sup>l</sup>          | -58.7 <sup>m</sup>                       |

<sup>a</sup> Computed at BP86/TZ(2)P (see also Figures 1–6). <sup>b</sup> Electronic energies. <sup>c</sup> No reverse activation enthalpy. <sup>d</sup> No stable product. <sup>e</sup> Primary product (P) of insertion. Rearrangement via second TS (at -11.0 kcal/mol) to more stable conformation (at -12.0 kcal/mol). <sup>f</sup> Primary product (P) of insertion. Rearrangement via second TS (at -15.6 kcal/mol) to more stable conformation (at -16.3 kcal/mol). <sup>g</sup> Linear, C<sub>∞v</sub> symmetric Pd–H–H (at -8.8 kcal/mol) or <sup>-</sup>ClPd–H–H structures (at -12.6 kcal/mol) are second-order saddle points. <sup>h</sup> No reverse activation barrier. <sup>i</sup> Products (P) of straight S<sub>N</sub>2 substitution. <sup>j</sup> TS of straight S<sub>N</sub>2 substitution. <sup>k</sup> Product complex (PC) of straight S<sub>N</sub>2 substitution. <sup>l</sup> TS of leaving-group rearrangement, proceeding from PC of straight S<sub>N</sub>2 reaction of PdCl<sup>-</sup> + CH<sub>3</sub>Cl (see footnote i). <sup>m</sup> Oxidative addition product.

density and to represent the Coulomb and exchange potentials accurately in each SCF cycle.<sup>17i</sup>

Geometries and energies were calculated using the generalized gradient approximation (GGA). Exchange is described by Slater's X $\alpha$  potential,<sup>17j</sup> with nonlocal corrections owing to Becke.<sup>17k,1</sup> Correlation is treated in the Vosko-Wilk-Nusair (VWN) parametrization using formula V,<sup>17m</sup> with nonlocal corrections due to Perdew.<sup>17n</sup> Relativistic effects were taken into account by the zeroth-order regular approximation (ZORA).<sup>18</sup> Prior investigations showed that relativistic effects are significant for our systems and that the ZORA formalism is well suited for describing them.<sup>14</sup>

All energy minima and transition state<sup>17q</sup> structures were verified by frequency calculations:<sup>17r</sup> for minima all normal modes have real frequencies, whereas transition states have one normal mode with an imaginary frequency. The character of the normal mode associated with the imaginary frequency was analyzed to ensure that the correct transition state was found.

Bond enthalpies at 298.15 K and 1 atm ( $\Delta H_{298}$ ) were calculated from electronic bond energies ( $\Delta E$ ) according to eq 9, assuming an ideal gas.<sup>19</sup>

$$\Delta H_{298} = \Delta E + \Delta E_{\text{trans},298} + \Delta E_{\text{rot},298} + \Delta E_{\text{vib},0} + \Delta(\Delta E_{\text{vib},0})_{298} + \Delta(pV) \quad (9)$$

Here,  $\Delta E_{\text{trans},298}$ ,  $\Delta E_{\text{rot},298}$  and  $\Delta E_{\text{vib},0}$  are the differences between products and reactants in translational, rotational

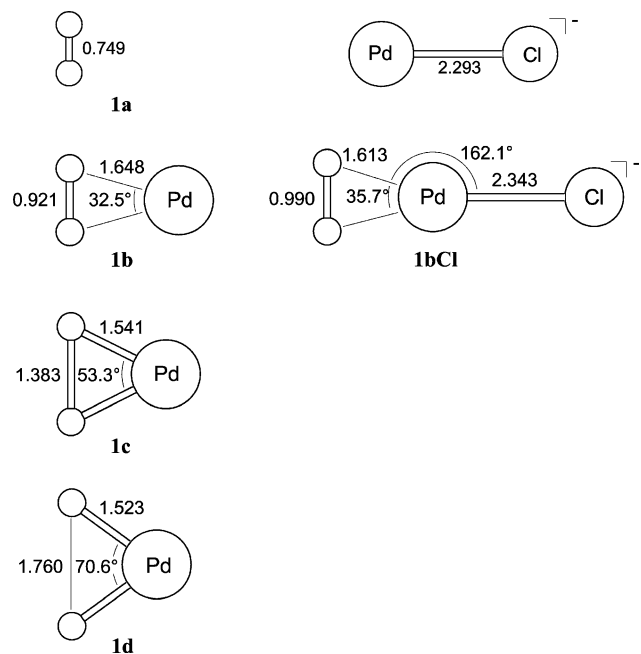
and zero point vibrational energy, respectively;  $\Delta(\Delta E_{\text{vib}})_{298}$  is the change in the vibrational energy difference as one goes from 0 to 298.15 K. The vibrational energy corrections are based on the frequency calculations. The molar work term  $\Delta(pV)$  is  $(\Delta n)RT$ ;  $\Delta n = -1$  for two reactants (e.g. PdCl<sup>-</sup> and CH<sub>3</sub>X) combining to one species. Thermal corrections for the electronic energy are neglected.

### 3. Results and Discussion

**3.1. Reaction Profiles and Geometries.** In this section, we discuss the potential energy surfaces (PES) of the various oxidative insertion and S<sub>N</sub>2 reactions as well as the geometries of the transition states of the various reactions. The results are summarized in Figures 1–6 (geometries) and Tables 1–3 (thermochemistry and geometries) and 4 (Activation Strain analyses).

Here, we focus on the reactions of the anionic, mono-coordinated PdCl<sup>-</sup> and how these differ from the reactions of the neutral, uncoordinated Pd atom. For a full account of the neutral reactions, the reader is referred to ref 12a.

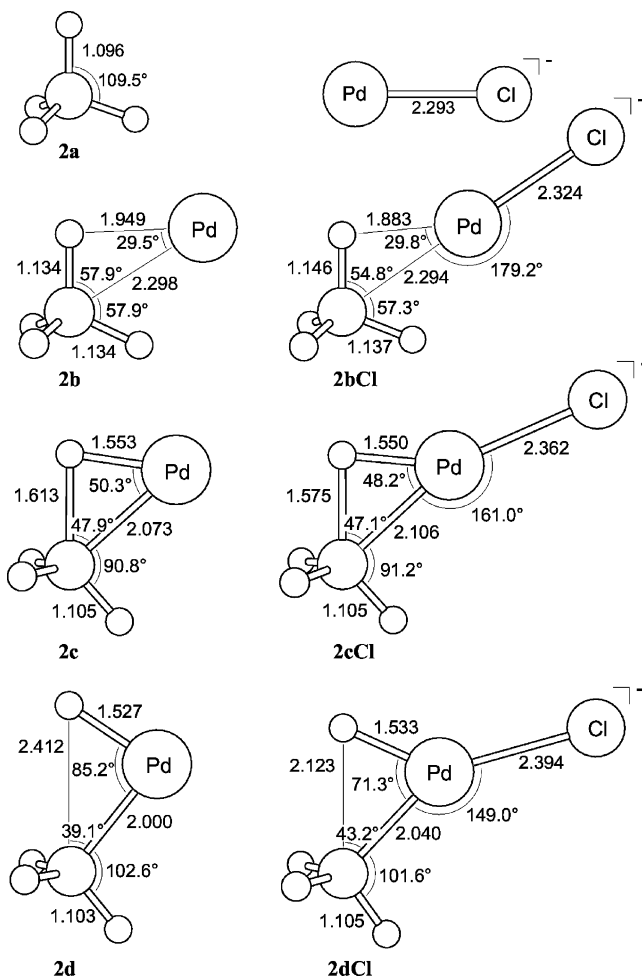
**Direct Oxidative Insertion (OxIn).** First, the direct oxidative insertion (OxIn) reactions are examined. For both, Pd and PdCl<sup>-</sup>, and for all substrates, they proceed from a reactant complex via a TS to a product, see Figures 1–5. The effect of anion assistance, i.e., going from Pd to PdCl<sup>-</sup>, is a substantial stabilization relative to the separate reactants of all stationary points along the PES (see Table 1): reactant



**Figure 1.** Geometries (in Å, deg) at ZORA-BP86/TZ(2)P of stationary points along the potential energy surface for oxidative insertion (OxIn) reaction of Pd (from ref 12a) and PdCl<sup>-</sup> (this work) into the H–H bond of H<sub>2</sub>.

complexes gain an additional stabilization of 5–12 kcal/mol, reaction barriers are lowered by 3–14 kcal/mol, and the reactions become overall 2–23 kcal/mol more exothermic. Whereas the oxidative insertion of the uncoordinated Pd + H<sub>2</sub> occurs with a minimal reverse activation energy of only half a kcal/mol, the reverse barrier vanishes completely for the oxidative insertion of PdCl<sup>-</sup> + H<sub>2</sub>, see Table 1. In other words, there is no stable product and the only stationary point left is that of the reactant complex **1bCl** at –36.3 kcal/mol (see Table 1 and Figure 1).

The PdCl<sup>-</sup> induced C–H activation is investigated using two different substrates, CH<sub>4</sub> (**2a**) and C<sub>2</sub>H<sub>6</sub> (**3a**). In this way, we can reveal if, e.g., the activation strain (see Section 3.2) adopts a characteristic value for a particular type of bond (here, the C–H bond) in different substrates. Geometries for both reactions are shown in Figures 2 and 3, structures **2aCl**–**2dCl** and **3aCl**–**3dCl**. Again, one finds relatively stable reactant complexes, **2bCl** and **3bCl**, respectively, in which PdCl<sup>-</sup> binds via palladium symmetrically, in an η<sup>2</sup> fashion, to two C–H bonds of the substrate, similar to the situation found for Pd in **2b** and **3b** (see Figures 3 and 4). Note that in fact all substrates but CH<sub>3</sub>Cl form reactant complexes that are of about the same thermodynamic stability as the corresponding product of oxidative addition to PdCl<sup>-</sup> (or Pd). For PdCl<sup>-</sup> + CH<sub>4</sub>, eq 3b, the activation enthalpy is –14.0 kcal/mol relative to the reactants. In the TS, **2cCl**, the C–H bond has been stretched by 0.479 Å (44%) and amounts to 1.575 Å, somewhat less than in the corresponding neutral TS, **2c**. The C–H activation of PdCl<sup>-</sup> + C<sub>2</sub>H<sub>6</sub>, eq 4b, proceeds via a transition state **3cCl** at –13.5 kcal/mol relative to the reactants, to a product **3dCl** at –16.1 kcal/mol. The C–H bond in TS **3cCl** is lengthened by 0.561 Å (51%) and amounts to 1.660 Å (see Table 2 in which, for the convenience of the reader, we have collected the data on

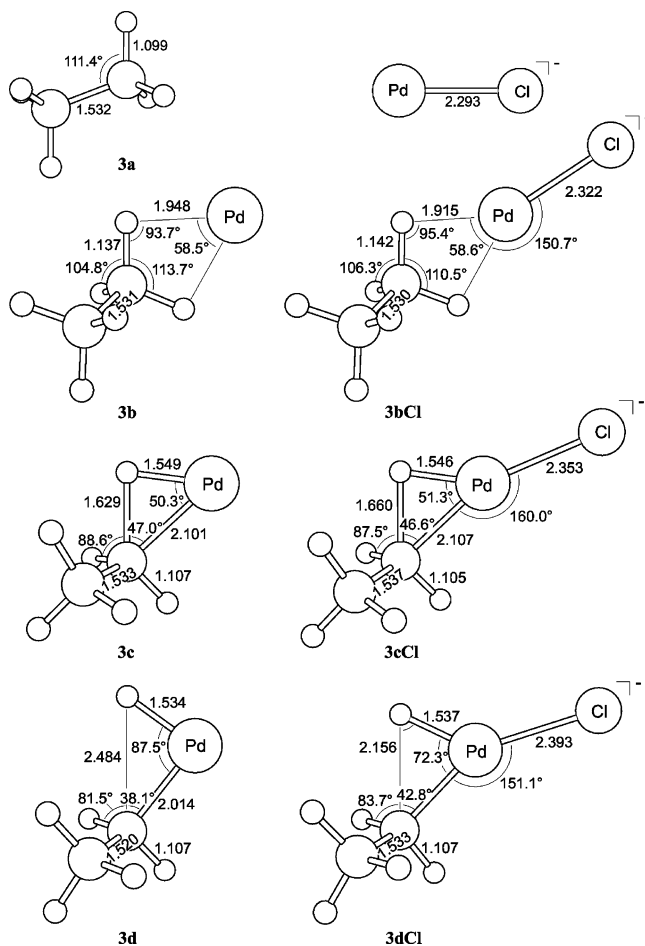


**Figure 2.** Geometries (in Å, deg) at ZORA-BP86/TZ(2)P of stationary points along the potential energy surface for oxidative insertion (OxIn) reaction of Pd (from ref 12a) and PdCl<sup>-</sup> (this work) into the C–H bond of CH<sub>4</sub>.

C–X and H–H bond lengths and stretching in the TS). This is very similar to the C–H activation of PdCl<sup>-</sup> + CH<sub>4</sub>. The reaction may further proceed via a second transition state (0.5 kcal/mol above **3dCl** and corresponding to an internal rotation around the Pd–C bond) to a conformation that is 0.2 kcal/mol more stable than the initial product complex **3dCl**.

The OxIn activation of the C–C bond of ethane by PdCl<sup>-</sup> (eq 2d) starts from the reactant complex **4bCl**, which is identical to the reactant complex **3bCl** for C–H activation in the same substrate. The barrier associated with the transition state, **4cCl**, is with 6.4 kcal/mol significantly higher than in case of C–H activation. Yet, in this TS, the C–C bond is only stretched by 0.362 Å (24%) and amounts to 1.894 Å; the relative stretching is less than in the case of C–H activation.

The activation of CH<sub>3</sub>Cl **5a** by direct oxidative insertion of PdCl<sup>-</sup> is connected with a slight lengthening of the C–Cl bond by 0.240 Å (13%) in the TS, where this bond has a length of 2.043 Å. The activation barrier of –11.2 kcal/mol is lower than for C–H and C–C activation. It is also lower than the barrier of –6.0 kcal/mol associated with activation of the same bond by direct oxidative insertion of uncoordinated Pd. The formation of product **5dCl** is the most



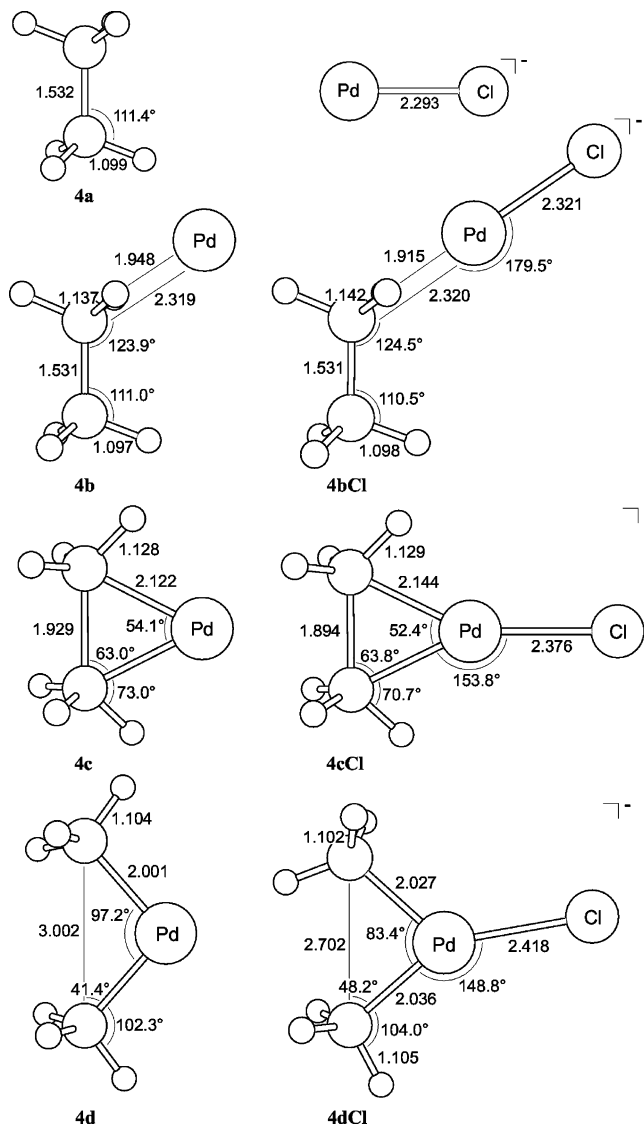
**Figure 3.** Geometries (in Å, deg) at ZORA-BP86/TZ(2)P of stationary points along the potential energy surface for oxidative insertion (OxIn) reaction of Pd (from ref 12a) and PdCl<sup>-</sup> (this work) into the C-H bond of C<sub>2</sub>H<sub>6</sub>.

exothermic oxidative addition among the model reactions studied, with a reaction enthalpy of  $-58.7$  kcal/mol.

Thus, the 298 K activation enthalpies (relative to the separate reactants) of the anion-assisted reactions increase more or less in the same order as those induced by the bare Pd atom, namely, in the order C-H ( $-14.0$  and  $-13.5$  kcal/mol for CH<sub>4</sub> and C<sub>2</sub>H<sub>6</sub>)  $\approx$  C-Cl ( $-11.2$  kcal/mol)  $<$  C-C ( $6.4$  kcal/mol).<sup>20</sup> (As pointed out above, for PdCl<sup>-</sup> + H-H, there is no stable product and thus no TS for oxidative insertion.) The higher barrier for C-C than for C-H activation by uncoordinated Pd originates from a much less favorable TS interaction<sup>12a</sup> with the model catalyst in the former bond and not, as thought previously,<sup>11in</sup> from a higher strain associated with the tilting of two methyl groups for C-C activation versus only one methyl group for C-H activation. As will become clear through the Activation Strain analysis in Section 3.3, the same holds true for the present PdCl<sup>-</sup> induced C-C and C-H activation processes.

#### S<sub>N</sub>2 Substitution versus Direct Oxidative Insertion.

Next, we discuss the S<sub>N</sub>2-type reactions and how they compete with direct oxidative insertion. For both, Pd and PdCl<sup>-</sup>, and for all substrates but CH<sub>3</sub>Cl, the actual substitution process of the straight S<sub>N</sub>2 reaction, i.e., formation of the new Pd-C (or Pd-H) bond and breaking of the old C-X (or H-H) bond, proceeds from the same reactant complexes



**Figure 4.** Geometries (in Å, deg) at ZORA-BP86/TZ(2)P of stationary points along the potential energy surface for oxidative insertion (OxIn) reaction of Pd (from ref 12a) and PdCl<sup>-</sup> (this work) into the C-C bond of C<sub>2</sub>H<sub>6</sub>.

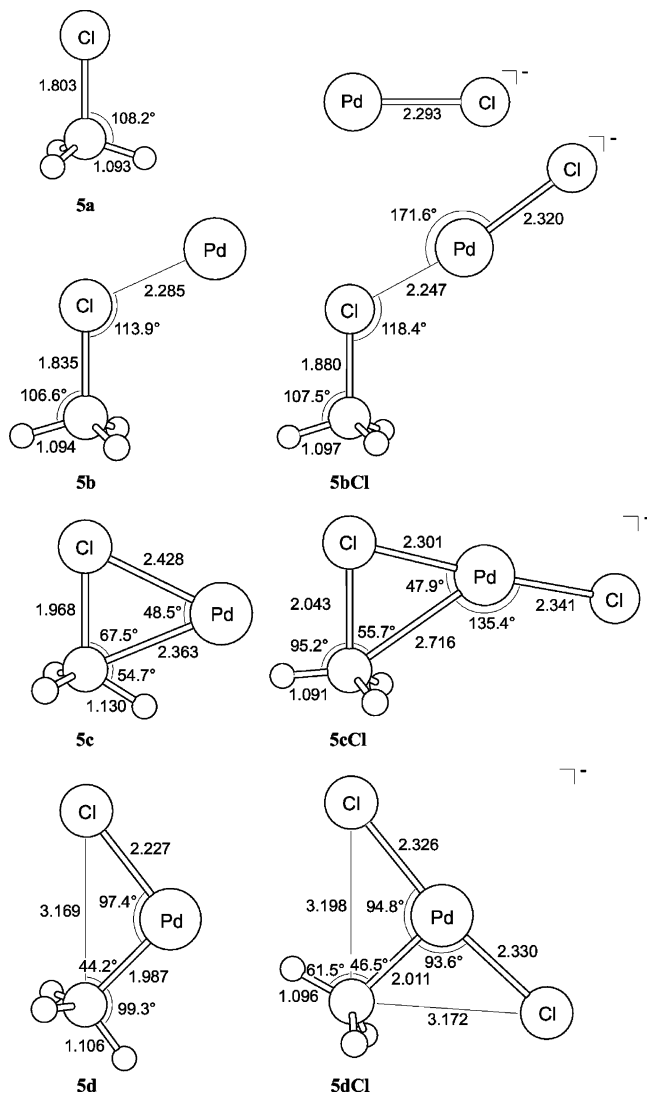
as the competing direct oxidative insertion (OxIn) reactions discussed above (see Figures 1–5). The effect of introducing a chloride ligand on palladium, i.e., going from Pd to PdCl<sup>-</sup>, is a dramatic reduction, by 140 kcal/mol or more, of the endothermicity of the straight S<sub>N</sub>2 reaction (see Table 1). The reason for this is that the energetically highly unfavorable charge separation, which occurs in the products of Pd (e.g., PdCH<sub>3</sub><sup>+</sup> + X<sup>-</sup>), disappears in those of PdCl<sup>-</sup> (e.g., PdClCH<sub>3</sub> + X<sup>-</sup>). Still, in all cases except for the reaction of PdCl<sup>-</sup> + CH<sub>3</sub>Cl, the straight S<sub>N</sub>2 reactions are highly endothermic (79–88 kcal/mol) and proceed without a reverse barrier, i.e., without a regular S<sub>N</sub>2 transition state. In other words, C-H and C-C bond activation occurs through direct oxidative insertion (OxIn) for both model catalysts, Pd and PdCl<sup>-</sup>.

The situation is strikingly different for C-Cl activation. For PdCl<sup>-</sup> + CH<sub>3</sub>Cl, a regular S<sub>N</sub>2 transition state **6cCl** was found, at  $-22.8$  kcal/mol relative to the reactants, that separates the reactant complex **6bCl** (at  $-23.9$  kcal/mol)

**Table 2.** Length and Stretching in the TS of the Activated Bond for Direct Oxidative Insertion (OxIn)<sup>a</sup>

| activated bond | reactants                          | length in substr (Å) | length in TS (in Å) |                   | stretching in TS (in Å) |                   | stretching in TS (in %) |                   |
|----------------|------------------------------------|----------------------|---------------------|-------------------|-------------------------|-------------------|-------------------------|-------------------|
|                |                                    |                      | Pd                  | PdCl <sup>−</sup> | Pd                      | PdCl <sup>−</sup> | Pd                      | PdCl <sup>−</sup> |
| H–H            | Pd + H <sub>2</sub>                | 0.749                | 1.383               | <i>b</i>          | 0.634                   | <i>b</i>          | 85                      | <i>b</i>          |
| C–H            | Pd + CH <sub>4</sub>               | 1.096                | 1.613               | 1.575             | 0.517                   | 0.479             | 47                      | 44                |
|                | Pd + C <sub>2</sub> H <sub>6</sub> | 1.099                | 1.629               | 1.660             | 0.530                   | 0.561             | 48                      | 51                |
| C–C            | Pd + C <sub>2</sub> H <sub>6</sub> | 1.532                | 1.929               | 1.894             | 0.397                   | 0.362             | 26                      | 24                |
| C–Cl           | Pd + CH <sub>3</sub> Cl            | 1.803                | 1.968               | 2.043             | 0.165                   | 0.240             | 9                       | 13                |

<sup>a</sup> Computed at ZORA-BP86/TZ(2)P. <sup>b</sup> No stable product.



**Figure 5.** Geometries (in Å, deg) at ZORA-BP86/TZ(2)P of stationary points along the potential energy surface for oxidative insertion (OxIn) reaction of Pd (from ref 12a) and PdCl<sup>−</sup> (this work) into the C–Cl bond of CH<sub>3</sub>Cl.

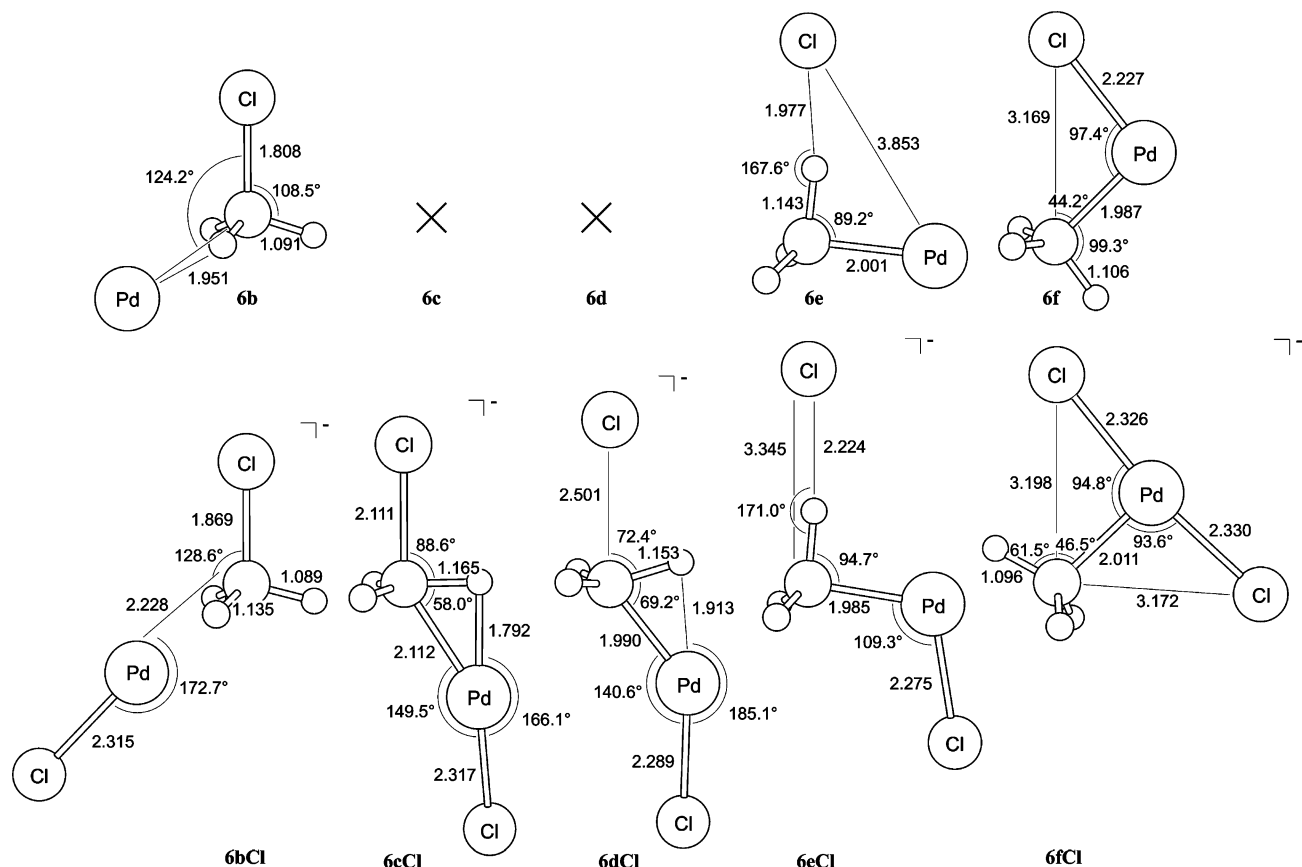
from an S<sub>N</sub>2 product complex **6dCl** (at −24.1 kcal/mol), see Figure 6. Dissociation of this complex leads to the products of the straight S<sub>N</sub>2 substitution at 3.0 kcal/mol relative to reactants (Table 1). In addition, a second transition state **6eCl** was found at −20.2 kcal/mol, which is reached from **6dCl** and corresponds to a rearrangement of the leaving group Cl<sup>−</sup> (Cl-ra) yielding the net oxidative addition product **6fCl** at an energy of −58.7 kcal/mol. This product is identical to structure **5dCl** resulting from direct insertion. We recall that

for Pd + CH<sub>3</sub>Cl a similar pathway (S<sub>N</sub>2/Cl-ra) was found which, however, proceeds in one elementary reaction step.<sup>12</sup> This pathway involves a transition state **6e**, 21.2 kcal/mol above the reactants, which has already proceeded beyond the actual substitution stage and has mainly the character of Cl<sup>−</sup> rearrangement. Thus, the effect of anion assistance is a lowering of the overall barrier of the S<sub>N</sub>2 pathway by 42.4 kcal/mol (from 21.2 kcal/mol in case of **6e** to −20.2 kcal/mol in case of **6eCl**).<sup>21</sup> For the other substrates, a TS associated with an S<sub>N</sub>2/Cl-ra mechanism could not be found.

Interestingly, through anion assistance, we have achieved a switch of mechanistic pathway for C–Cl bond activation. While for Pd + CH<sub>3</sub>Cl the overall barrier for the pathway of S<sub>N</sub>2 substitution followed by Cl<sup>−</sup> rearrangement (21.2 kcal/mol) is much higher than that for oxidative insertion (−6.0 kcal/mol), it drops in case of PdCl<sup>−</sup> + CH<sub>3</sub>Cl to −20.2 kcal/mol, i.e., well below the oxidative insertion barrier of −11.2 kcal/mol. Hence, oxidative addition of the C–Cl bond proceeds through direct oxidative insertion for Pd and through the S<sub>N</sub>2 pathway for PdCl<sup>−</sup> + CH<sub>3</sub>Cl. The OxIn mechanism is associated with retention of configuration at the carbon atom of the activated C–X bond, whereas the S<sub>N</sub>2/Cl-ra mechanism goes with inversion of configuration. For the present, rather simple model substrate, CH<sub>3</sub>Cl, this leads to identical products. However, the two pathways yield different enantiomers as product in case of more complex substrates that involve an asymmetric carbon atom in the activated C–X bond. Thus, by switching anion assistance on or off, we can steer the stereochemical course of the bond activation process. In Section 3.3, we discuss why anion assistance has the effect of shifting the preference from OxIn to S<sub>N</sub>2 pathway.

**Entropy Effects on Barriers.** Entropy effects at 298 K are important in the sense that they increase the magnitude of the activation free energy ΔG<sup>‡</sup> by a few kcal/mol, but, as can be seen in Table 3, they do not discriminate much between the various bond activation reactions and pathways considered.

The 298 K activation entropies ΔS<sup>‡</sup> for direct oxidative insertion (OxIn) of PdCl<sup>−</sup> are somewhat more negative than those of Pd and amount to −29.8 (C–H in methane), −32.2 (C–H in ethane), −32.6 (C–C) and −30.6 cal/mol K (C–Cl). Likewise, the activation entropies for the two steps of the S<sub>N</sub>2 pathway to C–Cl bond activation of PdCl<sup>−</sup> are more negative than that of the corresponding one-step mechanism of Pd and amount to −31.6 (S<sub>N</sub>2 step) and −29.8 cal/mol K (Cl rearrangement). This translates into −TΔS<sup>‡</sup> values of 8.9–9.7 kcal/mol for the reactions involving PdCl<sup>−</sup> and 5.4–



**Figure 6.** Geometries (in Å, deg) at ZORA-BP86/TZ(2)P of stationary points along the potential energy surface for the  $S_N2/Cl^-$ -rearrangement reaction of Pd (from ref 12a) and  $PdCl^-$  (this work) with  $CH_3Cl$ .

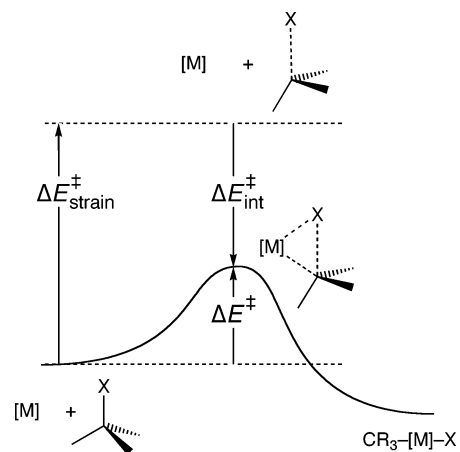
**Table 3.** Entropy Effects at 298 K on Activation Barriers (Relative to Reactants) for the Oxidative Insertion (OxIn) and  $S_N2/Cl^-$ -Rearrangement Reactions of Pd and  $PdCl^-$  with  $H_2$ ,  $CH_4$ ,  $C_2H_6$  and  $CH_3Cl$ , Respectively<sup>a</sup>

| activated bond             | reactants           | $\Delta H^\ddagger$ | $\Delta S^\ddagger$ <sup>a</sup> | $-\Delta \Delta S^\ddagger$ <sup>a</sup> | $\Delta G^\ddagger$ <sup>a</sup> |
|----------------------------|---------------------|---------------------|----------------------------------|--|----------------------------------|
| Oxidative Insertion        |                     |                     |                                  |  |                                  |
| H-H                        | Pd + $H_2$          | -21.7               | -18.1                            | 5.4                                      | -16.3                            |
|                            | $PdCl^-$ + $H_2$    | <sup>b</sup>        | <sup>b</sup>                     | <sup>b</sup>                             | <sup>b</sup>                     |
| C-H                        | Pd + $CH_4$         | -5.0                | -22.5                            | 6.7                                      | 1.7                              |
|                            | $PdCl^-$ + $CH_4$   | -14.0               | -29.8                            | 8.9                                      | -5.1                             |
|                            | Pd + $C_2H_6$       | -4.1                | -24.3                            | 7.3                                      | 3.2                              |
|                            | $PdCl^-$ + $C_2H_6$ | -13.5               | -32.2                            | 9.6                                      | -3.9                             |
| C-C                        | Pd + $C_2H_6$       | 9.6                 | -26.1                            | 7.8                                      | 17.4                             |
|                            | $PdCl^-$ + $C_2H_6$ | 6.4                 | -32.6                            | 9.7                                      | 16.1                             |
| C-Cl                       | Pd + $CH_3Cl$       | -6.0                | -23.4                            | 7.0                                      | 1.0                              |
|                            | $PdCl^-$ + $CH_3Cl$ | -11.2               | -30.6                            | 9.1                                      | -2.1                             |
| $S_N2/Cl^-$ -Rearrangement |                     |                     |                                  |  |                                  |
| C-Cl                       | Pd + $CH_3Cl$       | 21.2 <sup>c</sup>   | -21.6 <sup>c</sup>               | 6.5 <sup>c</sup>                         | 27.7 <sup>c</sup>                |
|                            | $PdCl^-$ + $CH_3Cl$ | -22.8 <sup>d</sup>  | -31.6 <sup>d</sup>               | 9.4 <sup>d</sup>                         | -13.4 <sup>d</sup>               |
|                            |                     | -20.2 <sup>e</sup>  | -29.8 <sup>e</sup>               | 8.9 <sup>e</sup>                         | -11.3 <sup>e</sup>               |

<sup>a</sup> Computed at BP86/TZ(2)P (see also Table 1). <sup>b</sup> No stable product. <sup>c</sup> Concerted  $S_N2/Cl^-$ -rearrangement. <sup>d</sup> Separate  $S_N2$  substitution. <sup>e</sup> Separate  $Cl^-$ -rearrangement.

7.8 kcal/mol for those involving Pd. Thus, in terms of relative activation free energies  $\Delta G^\ddagger$ , this does essentially not change the picture that emerges from the relative heights of activation energies  $\Delta E^\ddagger$  of the various reactions. Therefore, in the following, we focus on further analyzing the origin of and difference between the energy barriers  $\Delta E^\ddagger$  of the reactions.

**3.2. Activation Strain Model.** To gain insight into how anion assistance affects the activation barriers of the different oxidative insertion reactions, i.e., insight into how this effect



**Figure 7.** Illustration of the Activation Strain model in case of C-X bond activation by a transition metal system [M]. The activation energy  $\Delta E^\ddagger$  is decomposed into the activation strain  $\Delta E^\ddagger_{\text{strain}}$  of and the stabilizing TS interaction  $\Delta E^\ddagger_{\text{int}}$  between the reactants in the transition state.

depends on the nature of the concomitant geometrical deformation and electronic structure of catalyst and substrate, they were analyzed using the Activation Strain model of chemical reactivity.<sup>10,12</sup> In this model, the activation energy  $\Delta E^\ddagger$  is decomposed into the activation strain  $\Delta E^\ddagger_{\text{strain}}$  and the transition state (TS) interaction  $\Delta E^\ddagger_{\text{int}}$  (see eq 10 and Figure 7):

$$\Delta E^\ddagger = \Delta E^\ddagger_{\text{strain}} + \Delta E^\ddagger_{\text{int}} \quad (10)$$



**Table 4.** Analysis of the Activation Energies for Pd and PdCl<sup>−</sup> Induced Activation of the Indicated Bonds of H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub> and CH<sub>3</sub>Cl through Direct Oxidative Insertion and, for C–Cl, Also S<sub>N</sub>2 Substitution in Terms of the Activation Strain Model

|  | H <sub>2</sub> (H–H) |                    | CH <sub>4</sub> (C–H) |                   | C <sub>2</sub> H <sub>6</sub> (C–H) |                   | C <sub>2</sub> H <sub>6</sub> (C–C) |                   | CH <sub>3</sub> Cl (C–Cl) |                   | CH <sub>3</sub> Cl (C–Cl) via S <sub>N</sub> 2 |                    |
|--|----------------------|--------------------|-----------------------|-------------------|-------------------------------------|-------------------|-------------------------------------|-------------------|---------------------------|-------------------|--|--------------------|
|  | Pd                   | PdCl <sup>−a</sup> | Pd                    | PdCl <sup>−</sup> | Pd                                  | PdCl <sup>−</sup> | Pd                                  | PdCl <sup>−</sup> | Pd                        | PdCl <sup>−</sup> | Pd   | PdCl <sup>−</sup>  |
| Energy Decomposition (in kcal/mol)                         |                      |                    |                       |                   |                                     |                   |                                     |                   |                           |                   |  |                    |
| ΔE <sup>‡</sup>  | −21.7                | −35.3              | −1.6                  | −11.1             | −0.7                                | −10.3             | 12.6                                | 9.1               | −4.3                      | −10.3             | 24.5   | −18.5              |
| ΔE <sup>‡</sup> <sub>strain</sub>                          | 55.6                 | 56.1               | 53.5                  | 49.7              | 54.7                                | 58.1              | 39.4                                | 36.3              | 8.8                       | 9.6               | 87.7   | 91.8               |
| ΔE <sup>‡</sup> <sub>int</sub>                             | −77.3                | −91.4              | −55.1                 | −60.8             | −55.4                               | −68.4             | −26.8                               | −27.2             | −13.1                     | −19.9             | −63.2  | −110.3             |
| ΔE <sub>Pauli</sub>  | 208.7                | 176.3              | 211.1                 | 169.4             | 209.8                               | 184.1             | 192.6                               | 180.7             | 112.3                     | 89.3              | 112.5  | 162.2              |
| ΔV <sub>elst</sub>   | −183.7               | −173.6             | −170.4                | −143.3            | −171.9                              | −159.2            | −139.5                              | −126.1            | −76.7                     | −66.9             | −74.2  | −117.7             |
| ΔE <sub>oi</sub>   | −102.3               | −94.1              | −95.8                 | −86.9             | −93.3                               | −93.3             | −79.9                               | −81.8             | −48.7                     | −42.3             | −101.4   | −154.7             |
| Fragment Orbital Overlap (Catalyst Substrate) <sup>b</sup> |                      |                    |                       |                   |                                     |                   |                                     |                   |                           |                   |  |                    |
| (HOMO HOMO)  | 0.256                | 0.126              | 0.260                 | 0.153             | 0.306                               | 0.206             | 0.158                               | 0.063             | 0.180 <sup>c</sup>        | 0.117             | 0.146 <sup>d</sup>                             | 0.275 <sup>d</sup> |
| (HOMO LUMO)  | 0.300                | 0.343              | 0.327                 | 0.496             | 0.450                               | 0.539             | 0.136                               | 0.093             | 0.082                     | 0.087             | 0.228  | 0.403              |
| (LUMO HOMO)  | 0.566                | 0.709              | 0.401                 | 0.425             | 0.359                               | 0.396             | 0.213                               | 0.212             | 0.144 <sup>c</sup>        | 0.229             | 0.174 <sup>d</sup>                             | 0.070 <sup>d</sup> |
| Fragment Orbital Population (in Electrons) <sup>b</sup>    |                      |                    |                       |                   |                                     |                   |                                     |                   |                           |                   |  |                    |
| cat. LUMO  | 0.45                 | 0.21               | 0.38                  | 0.10              | 0.38                                | 0.11              | 0.22                                | 0.04              | 0.18                      | 0.11              | 0.22   | 0.10               |
| cat. HOMO  | 9.28                 | 9.32               | 9.32                  | 9.34              | 9.31                                | 9.27              | 9.42                                | 9.32              | 9.59                      | 9.57              | 9.26   | 8.97               |
| substr. LUMO   | 0.43                 | 0.57               | 0.36                  | 0.46              | 0.36                                | 0.51              | 0.25                                | 0.31              | 0.17                      | 0.31              | 0.64   | 0.98               |
| substr. HOMO   | 1.73                 | 1.89               | 1.71                  | 1.85              | 1.74                                | 1.86              | 1.83                                | 1.89              | 1.91 <sup>c</sup>         | 1.90              | 1.78 <sup>d</sup>                              | 1.83 <sup>d</sup>  |

<sup>a</sup> No stable product. Analysis based on fictitious TS structure corresponding to the TS structure for Pd + H<sub>2</sub> frozen with Cl<sup>−</sup> added and an optimized Pd–Cl distance of 2.370 Å. <sup>b</sup> For Pd, HOMO refers to the five degenerated 4d AOs and their combined overlaps or populations. For PdCl<sup>−</sup>, HOMO refers to the 5s, two 3π and two 1δ orbitals and their combined overlaps or populations. The Pd LUMO is the 5s, the PdCl<sup>−</sup> LUMO is the 6s (see also Scheme 1). <sup>c</sup> Values for substrate HOMO-1 instead of HOMO are given. In the TS for OxIn of Pd into the C–Cl bond of CH<sub>3</sub>Cl, the lone pair on Cl in A' symmetry that overlaps and interacts with the model catalyst's HOMO or LUMO is the HOMO-1 not the HOMO. <sup>d</sup> Values for substrate HOMO-2 instead of HOMO are given. In the TS for S<sub>N</sub>2/Cl-ira of Pd or PdCl<sup>−</sup> + CH<sub>3</sub>Cl, it is the substrate HOMO-2, i.e., the bonding σ<sub>CCl</sub> (not the HOMO) that overlaps and interacts most strongly with the model catalyst's HOMO or LUMO.

The activation strain ΔE<sup>‡</sup><sub>strain</sub> is the strain energy associated with deforming the reactants from their equilibrium geometry to the geometry they acquire in the activated complex (Figure 7). The TS interaction ΔE<sup>‡</sup><sub>int</sub> is the actual interaction energy between the deformed reactants in the transition state. In the present study, one of the reactants is either the neutral, uncoordinated Pd-*d*<sup>10</sup> atom or the anionic model complex PdCl<sup>−</sup> and the other reactant is one of the substrates H<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, and CH<sub>3</sub>Cl.

The TS interaction ΔE<sup>‡</sup><sub>int</sub> between the strained reactants is further analyzed in the conceptual framework provided by the Kohn–Sham molecular orbital (KS-MO) model.<sup>16</sup> To this end, it is further decomposed into three physically meaningful terms (eq 11) using the extended transition state (ETS) method<sup>22</sup> developed by Ziegler and Rauk.

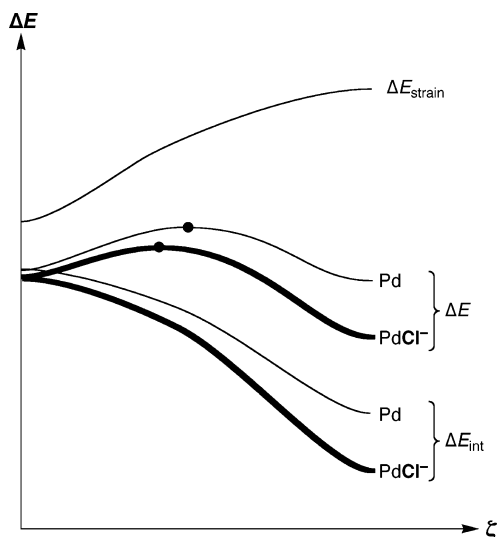
$$\Delta E_{\text{int}}^{\ddagger} = \Delta V_{\text{elst}} + \Delta E_{\text{Pauli}} + \Delta E_{\text{oi}} \quad (11)$$

The term ΔV<sub>elst</sub> corresponds to the classical electrostatic interaction between the unperturbed charge distributions of the deformed reactants and is usually attractive. The Pauli repulsion ΔE<sub>Pauli</sub> comprises the destabilizing interactions between occupied orbitals and is responsible for the steric repulsion. The orbital interaction ΔE<sub>oi</sub> accounts for charge transfer (interaction between occupied orbitals on one moiety with unoccupied orbitals of the other, including the HOMO–LUMO interactions) and polarization (empty-occupied orbital mixing on one fragment due to the presence of another fragment).

**3.3. Analysis of the Reaction Barriers for Bond Activation.** The results of the Activation Strain analysis are listed in Table 4. Anion assistance, as mentioned above, lowers the reaction barrier for all bonds and for both types of

mechanisms, i.e., direct oxidative insertion (OxIn) and the S<sub>N</sub>2 pathway. The effect of anion assistance is major if we consider the competition between the latter two pathways for C–Cl bond activation, which shifts from OxIn (retention of configuration at C in C–Cl) to S<sub>N</sub>2 pathway (inversion of configuration). As will become clear below, a quite straightforward and comprehensible physical mechanism hides behind this effect. But first we focus on understanding the somewhat more subtle mechanism of anion assistance in case of direct oxidative insertion into H–H, C–H, C–C and C–Cl bonds. Note that in the case of PdCl<sup>−</sup> + H–H, the reverse barrier for oxidative insertion actually disappears and that therefore no stable product or TS exist. To nevertheless allow for a comparison with the situation of the corresponding neutral reaction of Pd + H–H and how this is affected by anion assistance, we have analyzed a fictitious TS for oxidative insertion of PdCl<sup>−</sup> + H–H. This fictitious TS has been obtained by using the geometry of corresponding TS 1c for the insertion of uncoordinated Pd into H<sub>2</sub> and adding a Cl<sup>−</sup> ligand to Pd, keeping the PdH<sub>2</sub> portion frozen and only optimizing the Pd–Cl distance of the C<sub>2v</sub> symmetric species, yielding Pd–Cl = 2.370 Å (see Table 4).

**Activation Strain Analysis of OxIn Pathways.** The activation energy ΔE<sup>‡</sup> varies, not unexpectedly, from one model reaction to another. Interestingly, the activation strain ΔE<sup>‡</sup><sub>strain</sub> fluctuates much less and, in fact, adopts values that for each type of bond are in a characteristic range (see Table 4). The lowering of the activation barriers through anion assistance is in most cases caused by the TS interaction ΔE<sup>‡</sup><sub>int</sub> becoming more stabilizing. This is so especially if the effects become relatively large, as for H–H activation, whereas the



**Figure 8.** Schematic representation of typical reaction profiles provided by the energy  $\Delta E$  of the reaction system for direct oxidative insertion of Pd (thin lines) and  $\text{PdCl}^-$  (bold lines) into an H–H or C–X bond, and its decomposition into the strain energy  $\Delta E_{\text{strain}}$  of and interaction energy  $\Delta E_{\text{int}}$  between substrate and catalyst. Note that anion assistance stabilizes the TS (bullet) and shifts it back along the reaction coordinate  $\zeta$  (the extent of H–H or C–X bond stretching) toward the reactant side at the left.

situation becomes less clear-cut in case of marginal anion assistance, as for C–C activation. For example, going from Pd to  $\text{PdCl}^-$ , the barrier of the H–H bond activation is reduced from  $-21.7$  kcal/mol to  $-35.3$  kcal/mol (note that the latter value refers to a fictitious TS, vide supra), the activation strain remains in the same order, ca. 56 kcal/mol, and the strength of the TS interaction increases from  $-77.3$  kcal/mol to  $-91.4$  kcal/mol. Overall, for H–H, C–H (in methane and ethane) and C–Cl bond activation, anion assistance lowers barriers by 6–14 kcal/mol, which stems mainly from the strengthening of the TS interaction by 6–13 kcal/mol, whereas the activation strain changes only by 1–4 kcal/mol. Note that in the case of H–H, the energy of what can be conceived as the range of the PES where the TS might occur is stabilized more than the range of the PES where the product might occur. As a result, anion assistance makes the reverse barrier for insertion of  $\text{PdCl}^-$  into the H–H bond disappear and there is thus no longer a stable product. Anion assistance has hardly any effect on C–C bond activation: both the reduction of the barrier  $\Delta E^\ddagger$  (by 3.5 kcal/mol) and the changes in its components  $\Delta E_{\text{strain}}^\ddagger$  and  $\Delta E_{\text{int}}^\ddagger$  (3.1 and 0.4 kcal/mol) are marginal compared to the situation for the other bonds. The activation strain appears to be related to the bond strength of the activated bond and with the percentage-wise extent of bond stretching in the transition state. Typical strengths and lengths of the bonds under investigation are as follows: 104 (H–H), 99 (C–H), 83 (C–C) and 78 kcal/mol (C–Cl), and  $\sim 0.7$  (H–H),  $\sim 1.1$  (C–H),  $\sim 1.5$  (C–C) and  $\sim 1.8$  Å (C–Cl).<sup>23</sup> Furthermore, we recall that the percentage-wise extent of bond stretching in the TS for oxidative insertion is 85% (H–H), 44–51% (C–H), 24–26% (C–C) and 9–13% (C–Cl, see Table 2). Thus, both the bond strength and the percentage-wise bond

elongation in the TS decrease in the order H–H > C–H > C–C > C–Cl. This correlates well with the activation strain, which decreases in approximately the same order, namely, along H–H  $\approx$  C–H (50–58 kcal/mol) > C–C (36–39 kcal/mol) > C–Cl ( $\sim 9$  kcal/mol, see Table 4).<sup>24</sup>

Note that anion assistance causes reduction in activation strain (Table 4) and a contraction of the activated bond in the TS (Table 2) in case of methane C–H and ethane C–C bond activation. One might at first expect the opposite, namely, that the stronger catalyst-substrate interaction in case of  $\text{PdCl}^-$  causes the palladium atom to be further inserted into the activated bond in the TS. An understanding of these (and other) more subtle variations in activation strain  $\Delta E_{\text{strain}}^\ddagger$  requires that we take into account that the precise location of the TS (and thus the values of  $\Delta E_{\text{strain}}^\ddagger$  and  $\Delta E_{\text{int}}^\ddagger$ ) is the result of achieving, along the reaction coordinate  $\zeta$ , a (labile) balance between reactants strain  $\Delta E_{\text{strain}}(\zeta)$  and catalyst-substrate interaction  $\Delta E_{\text{int}}(\zeta)$ . While such detailed analyses are beyond the scope of the present work, we wish to report here preliminary results that provide a simple way of understanding the above-mentioned reduction in  $\Delta E_{\text{strain}}^\ddagger$  and contraction of activated bonds in the TS, caused by anion assistance. Figure 8 shows a schematic representation of a typical reaction profile. First, we examine the energy profile ( $\Delta E$ ) for the Pd-induced reaction, which is decomposed into the strain energy  $\Delta E_{\text{strain}}(\zeta)$  of the reactants plus their mutual interaction energy  $\Delta E_{\text{int}}(\zeta)$ . Along the reaction coordinate  $\zeta$ ,  $\Delta E_{\text{strain}}(\zeta)$  increases because the C–X bond of the substrate is stretched, while the Pd-substrate interaction  $\Delta E_{\text{int}}(\zeta)$  becomes more stabilizing due to the decreasing HOMO–LUMO gap of the deformed substrate. The net result is the reaction profile of  $\Delta E$  with the transition state indicated by a bullet (thin lines in Figure 8). Note that the reaction coordinate can be represented as the extent of stretching of the H–H or C–X bond. Next, we switch on anion assistance. This has not much effect on the  $\Delta E_{\text{strain}}(\zeta)$  curve, which closely resembles the C–X bond dissociation profile in both the Pd and  $\text{PdCl}^-$  induced reactions. The catalyst-substrate interaction, on the other hand, is substantially stabilized due to anion assistance and this strengthening becomes larger as the reaction proceeds (see bold lines in Figure 8). This is so simply because the inherent strength of the Pd-substrate interaction also increases along this direction (vide supra). Thus, the  $\Delta E_{\text{int}}(\zeta)$  curve is stabilized and it becomes steeper. Therefore, also the energy ( $\Delta E$ ) profile and, thus, all stationary points of the  $\text{PdCl}^-$  induced reaction are stabilized and, because of the steeper descent of  $\Delta E_{\text{int}}(\zeta)$ , the maximum shifts to the left, i.e., the TS becomes more reactant-like. This implies that the activated bond is less stretched and the activation strain somewhat reduced, compared to the corresponding Pd-induced reaction. Note that the observed behavior, i.e., the shift of the TS geometry toward the educt side in case of a more exothermic reaction, is reminiscent of the Hammond postulate<sup>25</sup> and that the origin and working of this postulate emerges naturally in terms of the Activation strain model as being the result of the interplay between the strain  $\Delta E_{\text{strain}}(\zeta)$ , which remains constant, and interaction  $\Delta E_{\text{int}}(\zeta)$ , which increases here as we go from Pd to  $\text{PdCl}^-$ . We plan to extend this type of analyses also to the ethane

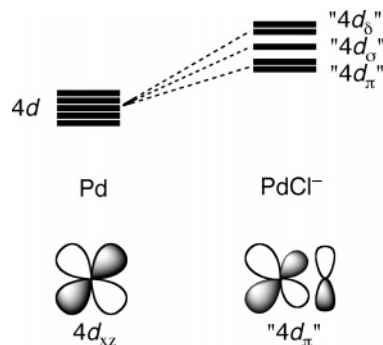
C–H and the C–Cl bond which show another, yet unexplained behavior.

Whereas the activation strain  $\Delta E_{\text{strain}}^{\ddagger}$  appears to correlate well with the strength and length of the activated bond,<sup>24</sup> there is no such straightforward relationship at all for the activation barrier  $\Delta E^{\ddagger}$  which shows a more complex behavior. This is so, of course, because  $\Delta E^{\ddagger}$  arises as the sum of  $\Delta E_{\text{strain}}^{\ddagger}$  and  $\Delta E_{\text{int}}^{\ddagger}$ , and it mirrors the interplay of the trends, often mutually counteracting, in *both* of these two quantities. The TS interaction  $\Delta E_{\text{int}}^{\ddagger}$  becomes less stabilizing in the order H–H > C–H > C–C > C–Cl. We have shown previously<sup>12a</sup> for the Pd-induced reactions and find here again for the PdCl<sup>−</sup>-induced reactions that this trend is to an important extent determined by the donor–acceptor orbital interactions between occupied Pd 4d AOs and the empty  $\sigma_{\text{C-X}}^*$  (or  $\sigma_{\text{H-H}}^*$ ) acceptor orbital associated with the bond to be activated in the substrate (see Table 4). The orbital interactions decrease along H–H, C–H and C–C because the  $\sigma^*$  acceptor orbital of the substrate in the TS goes up in energy leading to a larger HOMO–LUMO gap with the Pd 4d AOs. They further decrease from C–C to C–Cl because of a relatively small Pd 4d–CH<sub>3</sub>Cl  $\sigma_{\text{C-Cl}}^*$  overlap. This trend is reinforced by the electrostatic attraction  $\Delta V_{\text{elst}}$ , which also decreases along H–H, C–H, C–C and C–Cl as the Pd–substrate distance increases along this series (see Table 4 and Figures 1–5), and by the donor–acceptor orbital interactions between substrate HOMO and catalyst LUMO, which is however less important, in particular for the negatively charged and thus poorly electron-accepting PdCl<sup>−</sup> (see, e.g., the much smaller population of the PdCl<sup>−</sup> versus the Pd LUMO in Table 4).

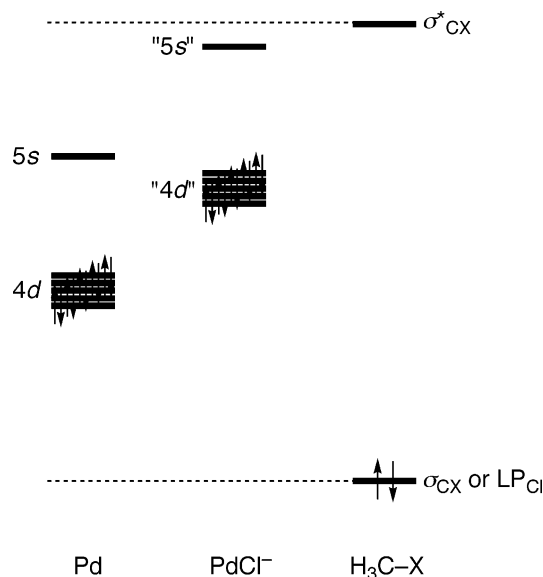
Here, we address the question *why* anion assistance leads to an increased, more stabilizing TS interaction  $\Delta E_{\text{int}}^{\ddagger}$ . Our Kohn–Sham orbital analyses show that the bonding frontier-orbital interactions are provided by the following: (i) back-donation from the occupied metal 4d (or metal–ligand 4d hybrid) orbitals to the unoccupied C–X (or H–H)  $\sigma^*$  antibonding orbital of the substrate and (ii) donation from the occupied C–X (or H–H)  $\sigma$  bonding orbital of the substrate to the unoccupied metal 5s (or metal–ligand 5s hybrid) orbital (see also ref 12a). In addition, there is an important Pauli-repulsive (i.e., occupied–occupied) interaction between the metal 4d (or metal–ligand 4d hybrid) orbitals and the C–X (or H–H)  $\sigma$  bonding orbital. In the case of C–Cl, the most important occupied–occupied repulsion is that with the lone-pair on chlorine pointing toward the metal; this lone-pair orbital is the HOMO of chloromethane (see also ref 12a). The effect of anion assistance can now be understood in terms of how these bonding and repulsive orbital interactions are affected as the Cl<sup>−</sup> ligand coordinates to the metal. The main effect of this coordination is that the palladium 4d orbitals are pushed up in energy and split into  $4d_{\pi}$ ,  $4d_{\sigma}$  and  $4d_{\delta}$  (this splitting is however marginal, i.e., ca. 0.1 eV), as shown in Scheme 1.

Also the shape of the 4d AOs changes: they are Pd–Cl antibonding combinations with a small bonding admixture of Pd 5p; this is illustrated in Scheme 1 for the Pd  $4d_{xz}$  and PdCl<sup>−</sup> “ $4d_{\pi}$ ” orbitals. These changes in the catalyst’s electronic structure reduce the metal “4d”–substrate  $\sigma^*$

**Scheme 1.** Cl<sup>−</sup> Ligand Pushes up and Splits Pd 4d Orbitals



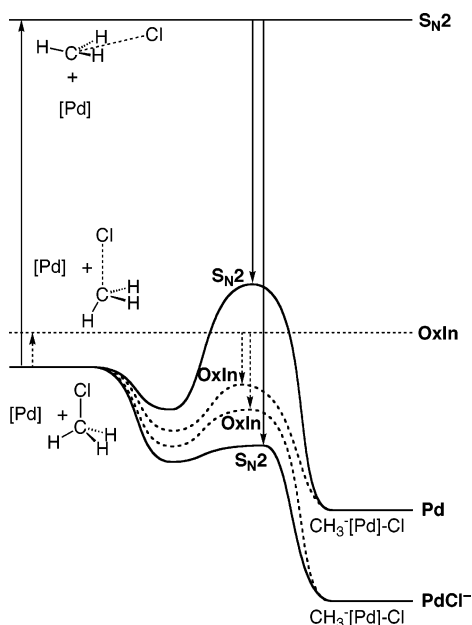
**Scheme 2.** Frontier Orbitals of Pd, PdCl<sup>−</sup> and H<sub>3</sub>C–X



orbital-energy gap, whereas they increase the substrate  $\sigma$  (or Cl lone pair)–metal “5s” orbital energy gap (as illustrated in Scheme 2), which causes two counteracting effects: back-donation is reinforced and donation is weakened.

Thus, overall, only moderate changes in the bonding orbital interactions are expected. The Pauli repulsive orbital interactions, on the other hand, are expected to become less repulsive in the case of anion assistance as the HOMO–HOMO energy gap between metal “4d” and substrate  $\sigma$  (or, for CH<sub>3</sub>Cl, LP<sub>Cl</sub>) increases and, as shown in Table 4, the corresponding overlap decreases. The quantitative decomposition of the TS interactions  $\Delta E_{\text{int}}^{\ddagger}$  indeed confirms these qualitative expectations (see Table 4). It shows that the increased stabilization of  $\Delta E_{\text{int}}^{\ddagger}$  in the case of PdCl<sup>−</sup> stems primarily from a reduction in Pauli repulsion  $\Delta E_{\text{Pauli}}$  (by −12 to −42 kcal/mol), whereas the bonding orbital interactions  $\Delta E_{\text{oi}}$  change much less (by −2 to 9 kcal/mol) and the electrostatic attraction  $\Delta V_{\text{elst}}$  becomes somewhat less stabilizing (by 10 to 27 kcal/mol).

**OxIn versus S<sub>N</sub>2 Pathways to C–X Bond Activation.** Finally, we aim at a simple, physical understanding of *why* anion assistance causes a switch of mechanistic pathway for C–Cl bond activation from direct oxidative insertion (barriers for Pd: −6.0 kcal/mol for OxIn versus 21.2 kcal/mol for S<sub>N</sub>2) to S<sub>N</sub>2 substitution (barriers for PdCl<sup>−</sup>: −11.2 kcal/



**Figure 9.** Fragment-oriented Design of a Catalyst (FDC) for stereoselective C–Cl bond activation based on the Activation Strain model: reaction profiles and Activation Strain analyses ( $\Delta E_{\text{strain}}^{\ddagger}$ : arrows up,  $\Delta E_{\text{int}}^{\ddagger}$ : arrows down) for oxidative addition of Pd or PdCl<sup>−</sup> to the H<sub>3</sub>C–Cl bond through oxidative insertion (OxIn, dashed lines) and S<sub>N</sub>2 substitution pathways (straight lines).

mol for OxIn versus  $-20.2$  kcal/mol for S<sub>N</sub>2) and, thus, from retention of configuration at the carbon atom to inversion of this configuration (see Table 1 and Section 3.1). Again, it appears from the activation strain analysis that the activation strain  $\Delta E_{\text{strain}}^{\ddagger}$  is relatively constant and, thus, a characteristic for a particular mechanism: it is low (8.8 kcal/mol for Pd and 9.6 kcal/mol for PdCl<sup>−</sup>; see Table 4) for the OxIn pathway in which the C–Cl bond is only slightly elongated, and it is high (87.7 kcal/mol for Pd and 91.8 kcal/mol for PdCl<sup>−</sup>; see Table 4) for the S<sub>N</sub>2 pathway in which chloromethane must undergo a major deformation as the Cl<sup>−</sup> leaving group migrates around CH<sub>3</sub> toward the palladium atom. It is the TS interaction  $\Delta E_{\text{int}}^{\ddagger}$  that is responsible for both the fact that the barriers for both pathways decrease through anion assistance and the fact that this decrease is much stronger for the S<sub>N</sub>2 pathway which, therefore, becomes the preferred mechanism. This is illustrated in Figure 9, which shows the potential energy surfaces (PES) of the OxIn and S<sub>N</sub>2 pathways for C–Cl activation by Pd and PdCl<sup>−</sup>.<sup>21</sup>

Now, a simple picture emerges of how anion assistance works and how we can rationally design a catalyst's selectivity for, in this example, the stereochemical course of C–X bond activation. The two pathways for this process are associated with very different geometrical reorganizations of the substrate and, thus, pronouncedly different and characteristic values for the activation strain: the latter is high for the S<sub>N</sub>2 pathway and low for direct insertion. The strong distortions that are responsible for the high activation strain for S<sub>N</sub>2 also cause the TS interaction for this pathway to be more stabilizing than for OxIn. This is because going from the overall transition state for OxIn to that for the S<sub>N</sub>2

pathway, the C–Cl antibonding  $\sigma^*$  acceptor orbital drops by ca. 4.6 eV in energy (not shown in Table 4) as the C–Cl bond is more elongated and, thus, becomes a better partner in the HOMO–LUMO interaction with the HOMO of a given catalyst. If this catalyst is one that interacts relatively weakly with the substrate, such as the uncoordinated Pd atom, the TS interaction cannot reverse the trend in activation strains and, as a result, direct oxidative insertion (OxIn) with its low activation strain remains the preferred mechanism, leading to C\*–Cl bond activation with retention of configuration at C\*. However, if we make the catalyst a sufficiently good electron donor, e.g., by pushing up the energy of the 4d orbitals with a Cl<sup>−</sup> ligand, as in our model catalyst PdCl<sup>−</sup>, the TS interaction can become strong enough to take over the role of trend setter, that is, to reverse the trend and make S<sub>N</sub>2 substitution, despite its high activation strain, the dominant pathway, leading to C\*–Cl bond activation with inversion of configuration at C\*.

#### 4. Conclusions

Anion assistance, i.e., coordination of a Cl<sup>−</sup> ligand to the metal atom, lowers the activation barriers of the Pd(0)-catalyzed activation of prototypical C–H, C–C and C–Cl bonds. Activation enthalpies  $\Delta H_{298}^{\ddagger}$  for oxidative insertion (OxIn) of PdCl<sup>−</sup> into the various types of bonds (computed at ZORA-BP86/TZ(2)P) are lower than the corresponding ones for Pd, but they increase essentially in the same order, namely, along C–H  $\approx$  C–Cl < C–C.

The effect of anion assistance is selective: it favors the highly endothermic S<sub>N</sub>2 mechanism over direct oxidative insertion (OxIn). Interestingly, in the case of C\*–Cl bond activation in CH<sub>3</sub>Cl, this leads to a shift in mechanism and stereochemistry, namely, from the OxIn pathway that goes with retention of configuration at C\* to the S<sub>N</sub>2 pathway that goes with inversion of configuration at C\*. This is of practical relevance for substrates in which C\* is asymmetric (which is obviously not the case in our simple model system).

To obtain a physical understanding of how anion assistance works, we have analyzed the various model reactions using the Activation Strain model in which the activation energy  $\Delta E^{\ddagger}$  is decomposed into the activation strain  $\Delta E_{\text{strain}}^{\ddagger}$  of and the stabilizing transition state (TS) interaction  $\Delta E_{\text{int}}^{\ddagger}$  between the reactants in the activated complex:  $\Delta E^{\ddagger} = \Delta E_{\text{strain}}^{\ddagger} + \Delta E_{\text{int}}^{\ddagger}$ . Interestingly, the activation strain  $\Delta E_{\text{strain}}^{\ddagger}$  adopts characteristic values for a particular type of bond and reaction mechanism, e.g., low for C–Cl activation through OxIn and high for C–Cl activation through S<sub>N</sub>2. The lowering of activation barriers upon anion assistance is primarily caused by a stronger, more stabilizing TS interaction  $\Delta E_{\text{int}}^{\ddagger}$ . This increase in  $\Delta E_{\text{int}}^{\ddagger}$  can be ascribed to the raise in Pd-4d derived orbitals in PdCl<sup>−</sup> which translates, among others, into more stabilizing donor–acceptor orbital interactions between the metal and the substrate.

The case of C–Cl bond activation exemplifies how a catalyst's selectivity regarding retention or inversion of configuration of the carbon atom in the activated bond can be tuned by simply increasing or decreasing the TS interaction. Eventually, we wish to apply this approach to rationally tuning, through a clever choice of ligands, a

catalyst's selectivity for a particular bond in a substrate, e.g., C–H versus C–C.

**Acknowledgment.** We thank the Fonds der Chemischen Industrie (FCI) for a doctoral stipendium for A.D. and The Netherlands Organization for Scientific Research (NWO-CW) for financial support.

### References

- (1) (a) Collman, J. P.; Hegedus, L. S.; Norton, J. R.; Finke, R. G. *Principles and Applications of Organotransition Metal Chemistry*; University Science Books: Mill Valley, CA, 1987. (b) Elschenbroich, Ch.; Salzer, A. *Organometallics. A Concise Introduction*, 2nd ed.; VCH: Weinheim, Germany, 1992. (c) Amatore, C.; Jutand, A. *Acc. Chem. Res.* **2000**, *33*, 314. (d) Yang, H.; Kotz, K. T.; Asplund, M. C.; Wilkens, M. J.; Harris, C. B. *Acc. Chem. Res.* **1999**, *32*, 551.
- (2) Experimental studies on reactions of metal complexes in the condensed phase: (a) Luh, T.-Y.; Leung, M.-k.; Wong, K.-T. *Chem. Rev.* **2000**, *100*, 3187. (b) Stürmer, R. *Angew. Chem.* **1999**, *111*, 3509. (c) Hau, L.-B.; Tanaka, M. *Chem. Commun.* **1999**, *5*, 395. (d) Casado, A. L.; Espinet, P. *Organometallics* **1998**, *17*, 954. (e) Kayser, B.; Missling, C.; Knizek, J.; Noeth, H.; Beck, W. *Eur. J. Inorg. Chem.* **1998**, *3*, 375. (f) Guillevic, M.-A.; Rocaboy, C.; Arif, A. M.; Horvath, I. T.; Gladysz, J. A. *Organometallics* **1998**, *17*, 707. (g) Edelbach, B. L.; Lachicotte, R. J.; Jones, W. D. *J. Am. Chem. Soc.* **1998**, *120*, 2843. (h) Crabtree, R. H. *Chem. Rev.* **1995**, *95*, 987. (i) Grushin, V. V.; Alper, H. *Chem. Rev.* **1994**, *94*, 1047. (j) Ellis, P. R.; Pearson, J. M.; Haynes, A.; Adams, H.; Bailey, N. A.; Maitlis, P. M. *Organometallics* **1994**, *13*, 3215. (k) Wright, M. W.; Smalley, T. L.; Welker, M. E.; Rheingold, A. L. *J. Am. Chem. Soc.* **1994**, *116*, 6777. (l) Sakakaurka, T.; Sodeyama, T.; Sasaki, K.; Wada, K.; Tanaka, M. *J. Am. Chem. Soc.* **1990**, *112*, 7221. (m) Casalnuovo, A. L.; Calabrese, J. C.; Milstein, D. *J. Am. Chem. Soc.* **1988**, *110*, 6738. (n) Janowics, A. H.; Bergman, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 3929. (o) Jones, W. D.; Feher, F. J. *J. Am. Chem. Soc.* **1982**, *104*, 4240.
- (3) (a) Hickey, C. E.; Maitlis, P. M. *J. Chem. Soc., Chem. Commun.* **1984**, 1609. (b) Forster, D. *Adv. Organomet. Chem.* **1979**, *17*, 255. (c) Forster, D. *J. Am. Chem. Soc.* **1975**, *97*, 951.
- (4) Experimental studies on reactions of ionic metal atoms and complexes in the gas phase: (a) Brönstrup, M.; Schröder, D.; Schwarz, H. *Organometallics* **1999**, *18*, 1939. (b) Aschi, M.; Brönstrup, M.; Diefenbach, M.; Harvey, J. N.; Schröder, D.; Schwarz, H. *Angew. Chem.* **1998**, *110*, 858. (c) Freiser, B. S. *J. Mass Spectrom.* **1996**, *31*, 703. (d) van Koppen, P. A. M.; Kemper, P. R.; Bushnell, J. E.; Bowers, M. T. *J. Am. Chem. Soc.* **1995**, *117*, 2098. (e) Wesendrup, R.; Schröder, D.; Schwarz, H. *Angew. Chem.* **1994**, *105*, 1232. (f) Chen, Y.-M.; Clemmer, D. E.; Armentrout, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7815. (g) van den Berg, K. J.; Ingemann, S.; Nibbering, N. M. M.; Gregor, I. K. *Rapid. Commun. Mass Spectrom.* **1993**, *7*, 769. (h) Chowdhury, A. K.; Wilkins, C. L. *J. Am. Chem. Soc.* **1987**, *109*, 5336. (i) Weil, D. A.; Wilkins, C. L. *J. Am. Chem. Soc.* **1985**, *107*, 7316. (j) Jones, R. W.; Staley, R. H. *J. Phys. Chem.* **1982**, *86*, 1669. (k) Jones, R. W.; Staley, R. H. *J. Am. Chem. Soc.* **1980**, *102*, 3794.
- (5) Combined experimental and theoretical studies on reactions of ionic metal atoms and complexes in the gas phase: (a) Yi, S. S.; Reichert, E. L.; Holthausen, M. C.; Koch, W.; Weisshaar, J. C. *Chem. Eur. J.* **2000**, *6*, 2232. (b) Blomberg, M.; Yi, S. S.; Noll, R. J.; Weisshaar, J. C. *J. Phys. Chem. A* **1999**, *103*, 7254. (c) Diefenbach, M.; Brönstrup, M.; Aschi, M.; Schröder, D.; Schwarz, H. *J. Am. Chem. Soc.* **1999**, *121*, 10614. (d) Schwarz, J.; Schröder, D.; Schwarz, H.; Heine-mann, C.; Hrusák, J. *Helv. Chim. Acta* **1996**, *79*, 1110.
- (6) Experimental studies on reactions of neutral metal atoms in the gas phase: (a) Wen Y.; Porembski, M.; Ferrett, T. A.; Weisshaar, J. C. *J. Phys. Chem. A* **1998**, *102*, 8362. (b) Wen, Y.; Yethiraj, A.; Weisshaar, J. C. *J. Chem. Phys.* **1997**, *106*, 5509. (c) Carroll, J. J.; Weisshaar, J. C. *J. Phys. Chem.* **1996**, *100*, 12355. (d) Chertihin, G. V.; Andrews, L. *J. Am. Chem. Soc.* **1994**, *116*, 8322. (e) Carroll, J. J.; Haug, K. L.; Weisshaar, J. C. *J. Am. Chem. Soc.* **1993**, *115*, 6962. (f) Carroll, J. J.; Weisshaar, J. C. *J. Am. Chem. Soc.* **1993**, *115*, 800. (g) Ritter, D.; Carroll, J. J.; Weisshaar, J. C. *J. Phys. Chem.* **1992**, *96*, 10636. (h) Mitchell, S. A.; Hackett, P. A. *J. Chem. Phys.* **1990**, *93*, 7822. (i) Ritter, D.; Weisshaar, J. C. *J. Am. Chem. Soc.* **1990**, *112*, 6425. (j) Fayet, P.; Kaldor, A.; Cox, D. M. *J. Chem. Phys.* **1990**, *92*, 254.
- (7) Combined experimental and theoretical studies on reactions of neutral metal atoms in the gas phase: (a) Porembski, M.; Weisshaar, J. C. *J. Phys. Chem. A* **2000**, *104*, 1524. (b) Carroll, J. J.; Haug, K. L.; Weisshaar, J. C.; Blomberg, M. R. A.; Siegbahn, P. E. M.; Svensson, M. *J. Phys. Chem.* **1995**, *99*, 13955. (c) Carroll, J. J.; Weisshaar, J. C.; Siegbahn, P. E. M.; Wittborn, A. M. C.; Blomberg, M. R. A. *J. Phys. Chem.* **1995**, *99*, 14388. (d) Mitchell, S.; Blitz, M. A.; Siegbahn, P. E. M.; Svensson, M. *J. Chem. Phys.* **1994**, *100*, 423. (e) Weisshaar, J. C. *Acc. Chem. Res.* **1993**, *26*, 213.
- (8) Theoretical studies on reactions of metal complexes: (a) Dedieu, A. *Chem. Rev.* **2000**, *100*, 543. (b) Torrent, M.; Solà, M.; Frenking, G. *Chem. Rev.* **2000**, *100*, 439. (c) Griffin, T. R.; Cook, D. B.; Haynes, A.; Pearson, J. M.; Monti, D.; Morris, G. E. *J. Am. Chem. Soc.* **1996**, *118*, 3029. (d) Aullón, G.; Alvarez, S. *Inorg. Chem.* **1996**, *35*, 3137. (e) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651. (f) Koga, N.; Morokuma, K. *Chem. Rev.* **1991**, *91*, 823. (g) Bickelhaupt, F. M.; Baerends, E. J.; Ravenek, W. *Inorg. Chem.* **1990**, *29*, 350.
- (9) Bickelhaupt, F. M.; Ziegler, T.; von Ragué Schleyer, P. *Organometallics* **1995**, *14*, 2288.
- (10) Bickelhaupt, F. M. *J. Comput. Chem.* **1999**, *20*, 114.
- (11) Theoretical studies on reactions of neutral metal atoms: (a) Maseras, F.; Lledós, A.; Clot, E.; Eisenstein, O. *Chem. Rev.* **2000**, *100*, 601. (b) Cui, Q.; Musaeu, D. G.; Morokuma, K. *J. Chem. Phys.* **1998**, *108*, 8418. (c) Wittborn, A. M. C.; Costas, M.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Chem. Phys.* **1997**, *107*, 4318. (d) Siegbahn, P. E. M. *J. Am. Chem. Soc.* **1994**, *116*, 7722. (e) Siegbahn, P. E. M. *Organometallics* **1994**, *13*, 2833. (f) Perry, J. K.; Ohanessian, G.; Goddard, W. A., III *Organometallics* **1994**, *13*, 1870. (g) Blomberg, M. R. A.; Siegbahn, P. E. M.; Svensson, M. *Inorg. Chem.* **1993**, *32*, 4218. (h) Siegbahn, P. E. M.; Blomberg, M. R. A.; Svensson, M. *J. Phys. Chem.* **1993**, *97*, 2564. (i) Siegbahn, P. E. M.; Blomberg, M. R. A.; Svensson, M. *J. Am. Chem. Soc.* **1993**, *115*, 1952. (j) Siegbahn, P. E. M.; Blomberg, M. R. A.; Svensson, M. *J. Am. Chem. Soc.* **1993**, *115*, 4191. (k) Siegbahn, P. E. M.; Blomberg, M. R. A. *J. Am. Chem. Soc.* **1992**, *114*, 10548. (l) Blomberg, M. R. A.; Siegbahn, P. E. M.; Svensson, M. *J. Am. Chem. Soc.* **1992**, *114*, 6095. (m) Svensson, M.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Am. Chem. Soc.* **1991**, *113*, 3, 7076. (n) Novaro, O.; Jarque, C. *Theor. Chim. Acta* **1991**, *80*, 19. (o) Blomberg, M. R. A.; Siegbahn, P. E. M.; Nagashima, U.; Wennerberg, J. *J. Am. Chem. Soc.* **1991**, *113*, 424. (p) Carter, E. A.;

- Goddard, W. A., III *J. Phys. Chem.* **1988**, *92*, 5679. (q) Nakatsuji, H.; Hada, M.; Yonezawa, T. *J. Am. Chem. Soc.* **1987**, *109*, 1902. (r) Low, J. J.; Goddard, W. A., III *Organometallics* **1986**, *5*, 609. (s) Koga, N.; Obara, S.; Kitaura, K.; Morokuma, K. *J. Am. Chem. Soc.* **1985**, *107*, 7109. (t) Low, J. J.; Goddard, W. A., III *J. Am. Chem. Soc.* **1984**, *106*, 8321.
- (12) (a) Diefenbach, A.; Bickelhaupt, F. M. *J. Phys. Chem. A* **2004**, *108*, 8460. (b) Diefenbach, A.; de Jong, G. Th.; Bickelhaupt, F. M. *Mol. Phys.* **2005**, *103*, in press.
- (13) (a) Basile, A.; Fasson, S.; Vitulli, G.; Drioli, E. *Stud. Surf. Sci. Catal.* **1998**, *119*, 453. (b) Malleron, J.-L.; Fiaud, J.-C.; Legros, J.-Y. *Handbook of Palladium Catalyzed Organic Reactions*; Academic Press: 1997. (c) Cornils, R.; Herrmann, W. A. *Applied Homogeneous Catalysis with Organometallic Compounds. Vol. 1*; VCH: Weinheim, 1996; p 394.
- (14) Diefenbach, A.; Bickelhaupt, F. M. *J. Chem. Phys.* **2001**, *115*, 4030.
- (15) Density functional theory (DFT): (a) Dreizler, R. M.; Gross, E. K. U. *Density Functional Theory. An approach to the Quantum Many-Body Problem*; Springer: Berlin, 1990. (b) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
- (16) Kohn–Sham MO model in DFT: (a) Bickelhaupt, F. M.; Baerends, E. J. In *Reviews in Computational Chemistry*, Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2000; Vol. 15, Chapter 1. (b) Baerends, E. J.; Gritsenko, O. V. *J. Phys. Chem. A* **1997**, *101*, 5383.
- (17) Amsterdam Density Functional (ADF) program: (a) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; van Gisbergen, S. J. A.; Fonseca Guerra, C.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931. (b) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41. (c) Baerends, E. J.; Ros, P. *Chem. Phys.* **1975**, *8*, 412. (d) Baerends, E. J.; Ros, P. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1978**, *S12*, 169. (e) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391. (f) Boerrigter, P. M.; te Velde, G.; Baerends, E. J. *Int. J. Quantum Chem.* **1988**, *33*, 87. (g) te Velde, G.; Baerends, E. J. *J. Comput. Phys.* **1992**, *99*, 84. (h) Snijders, J. G.; Baerends, E. J.; Vernooijs, P. *At. Nucl. Data Tables* **1982**, *26*, 483. (i) Krijn, J.; Baerends, E. J. *Fit-Functions in the HFS-Method; Internal Report* (in Dutch); Vrije Universiteit: Amsterdam, 1984. (j) Slater, J. C. *Quantum Theory of Molecules and Solids Vol. 4*; McGraw-Hill: New York, 1974. (k) Becke, A. D. *J. Chem. Phys.* **1986**, *84*, 4524. (l) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098. (m) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200. (n) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822 (Erratum: *Phys. Rev. B* **1986**, *34*, 7406). (o) Fan, L.; Ziegler, T. *J. Chem. Phys.* **1991**, *94*, 6057. (p) Versluis, L.; Ziegler, T. *J. Chem. Phys.* **1988**, *88*, 322. (q) Fan, L.; Ziegler, T. *J. Chem. Phys.* **1990**, *92*, 3645. (r) Fan, L.; Versluis, L.; Ziegler, T.; Baerends, E. J.; Ravenek, W. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1988**, *S22*, 173.
- (18) ZORA approach: (a) Chang, C.; Pelissier, M.; Durand, P. *Phys. Scr.* **1986**, *34*, 394. (b) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1993**, *99*, 4597. (c) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1994**, *101*, 9783. (d) van Lenthe, E.; van Leeuwen, R.; Baerends, E. J.; Snijders, J. G. *Int. J. Quantum Chem.* **1996**, *57*, 281.
- (19) Atkins, P. W. *Physical Chemistry*; Oxford University Press: Oxford, 1998.
- (20) A negative activation energy does not imply the complete absence of any barrier. The reaction is still hampered by a statistical or entropic bottleneck that is associated with the decrease in the number of available quantum states (e.g., of translation, rotation, and vibration) as one goes from the separate, unbound reactants to the tightly bound transition state.
- (21) Note that the [Pd]-CH<sub>3</sub><sup>+</sup>···Cl<sup>-</sup> structure, which occurs between the stages of S<sub>N</sub>2 substitution and Cl-rearrangement, is labile (i.e., a shoulder on the PES instead of a local minimum) for [Pd] = Pd, while it becomes stable (i.e., a local minimum) in case of [Pd] = PdCl<sup>-</sup> (see Table 1 for computed energies). One way of viewing this is that the [Pd]-CH<sub>3</sub><sup>+</sup>···Cl<sup>-</sup> structure benefits relatively more, in terms of relative stability, from anion assistance than the transition-state structures corresponding to the S<sub>N</sub>2 and Cl-rearrangement stages of the mechanism. An alternative but equivalent view is that PdCl<sup>-</sup> (with its high-energy HOMO) is a strong enough nucleophile for preventing its spontaneous expulsion by Cl<sup>-</sup> while Pd (with its low-energy HOMO) is not.
- (22) Bond energy decomposition: (a) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1558. (b) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1755. (c) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1977**, *16*, 1. (d) Bickelhaupt, F. M.; Nibbering, N. M. M.; van Wezenbeek, E. M.; Baerends, E. J. *J. Phys. Chem.* **1992**, *96*, 4864.
- (23) Holleman, A. F.; Wiberg, N. *Lehrbuch der Anorganischen Chemie*; de Gruyter: Berlin, 1985.
- (24) The present model catalysts contribute only marginal (PdCl<sup>-</sup>) or not at all (Pd) to the activation strain. We anticipate that introducing more than one ligand into the catalytically active transition metal system brings into play the phenomenon of ligand reorganization as an additional important factor that affects the activation strain of the reaction.
- (25) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334.

CT0499478

## Melting-like Transition in a Ternary Alkali Nanoalloy: $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$

Andrés Aguado\* and José M. López

*Departamento de Física Teórica, Universidad de Valladolid, Valladolid 47011, Spain*

Received November 9, 2004

**Abstract:** A theoretical analysis of the equilibrium geometry and thermal behavior of the ternary  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$  alkali nanoalloy is presented. The calculations are based on the orbital-free approach to density functional theory and the classical Newtonian equations to deal with the electronic and atomic subsystems, respectively. An onion-like polyicosahedral structure is found to have the lowest energy, with a core shell formed by Li atoms and an external (incomplete) surface shell formed by Cs atoms, the remaining Na atoms forming an intermediate shell. In a narrow range of 10 meV/atom above the ground-state energy, we identify several other isomers, with varying compositional and structural disorder, but all of them based on a polyicosahedral growing pattern. The most important result extracted from an analysis of thermal properties is that diffusion of Cs atoms at the surface starts at  $\approx 140$  K, which is 50 K above typical surface melting temperatures of homogeneous Cs clusters. Thus we conclude that alloying may be useful in enlarging the thermodynamic stability of solid surfaces of clusters beyond its homogeneous limit. As the chemical reactivity of a cluster is known to be highly structure dependent, this observation may be especially relevant to heterogeneous catalysis and related applications. We also analyze the dynamical melting behavior of one of the higher-energy isomers and compare it to that of the ground-state structure.

### I. Introduction

Metal nanoparticles have been known for some time to exhibit enhanced catalytic activity as compared to the bulk phase<sup>1</sup> as well as other chemical and physical “anomalies”. By analogy with the bulk situation, it may be expected that the properties of nanoalloys will strongly depend on the segregation/mixing tendencies and geometric structure.

Heterogeneous clusters show a much richer structure in their isomer energy spectra than homogeneous clusters. Following López et al.<sup>2</sup> and Jellinek et al.<sup>3</sup> we will indistinctly use the terms permutational isomer and homotop to designate the isomers obtained from a given topological structure by a permutation of atomic sites. In a homogeneous cluster, such a permutation does not lead to a new isomer. In a nanoalloy, on the contrary, those permutations involving exchanges between atoms of different species will result in

a different isomer. Similar to what is found in bulk alloys, the local structure is not exactly the same for different homotops, due to structural relaxations induced by differences in size, bonding, etc., of the atomic species involved. In the case of ternary nanoalloys, the number of homotops increases with the number of atoms  $N$  as  $3^N$ . The different atomic volumes, bond strengths and surface tensions of the species involved will conjointly determine the structure of the lowest-energy isomer and its thermal properties for each composition.<sup>4–6</sup>

Most of the theoretical work on heterogeneous metal clusters has been devoted to structural properties. López et al.<sup>7</sup> studied the segregation properties of Na–Cs and Na–Li nanoalloys of several compositions through static calculations. Similar studies were carried out by Bol et al.<sup>8</sup> on binary Na–K and ternary Na–K–Cs clusters. Ab initio calculations, like those of Deshpande et al.<sup>9</sup> on Na–Li, Joshi and Kanhere<sup>10</sup> on Li–Sn and Chacko et al.<sup>11</sup> on Al–Li clusters, are restricted to small sizes, while consideration of structural

\* Corresponding author phone: 00 +34 983 423147; e-mail: aguado@metodos.fam.cie.uva.es.

preferences in larger bimetallic clusters is usually performed by modeling with phenomenological potentials.<sup>12,13</sup> There are also a few molecular dynamics (MD) simulations on bimetallic clusters which consider their freezing<sup>14</sup> and melting<sup>2,10,15,16</sup> transitions and the influence of temperature on segregation.<sup>17</sup> Very recently, Aguado et al.<sup>4–6</sup> have performed MD simulations of melting in the impurity-doped alkali clusters  $A_1Na_{54}$ , with  $A=Li, K, Rb$  and  $Cs$ , and in the binary nanoalloys  $Na_{43}Cs_{12}$ ,  $Li_{13}Na_{42}$  and  $Na_{13}Cs_{42}$ . In this paper, we report the results of extensive MD simulations of the melting-like transition in the ternary  $Li_{13}Na_{30}Cs_{12}$  nanoalloy. Although clusters composed only of alkali elements are not especially interesting for specific applications, homogeneous alkali clusters have been thoroughly studied in the past because they are simple metallic systems amenable to both experimentation and computer modeling. They have represented an ideal test-bed for the development of theoretical and experimental methods in cluster physics, and consideration of their properties has helped to clarify very interesting issues such as the evolution of structural, dynamic (thermal) and electronic properties from the atomic to the bulk limit or the discovery of so-called magic numbers. Alkali clusters are thus very important from a fundamental point of view. While it is true that studies on binary and ternary alkali nanoalloys are still scarce, they are expected to play a similar role, contributing to the general understanding of cluster properties specific to multicomponent systems, which could be qualitatively shared by more complex metallic nanoalloys. It is within this wider perspective that consideration of the thermal properties of  $Li_{13}Na_{30}Cs_{12}$  is pertinent. We will show that the ground-state structure of  $Li_{13}Na_{30}Cs_{12}$ , as those of  $Li_{13}Na_{42}$  and  $Na_{13}Cs_{42}$ ,<sup>6</sup> is based on core–shell polyicosahedral growing. The different atomic species are distributed in onion-like radial shells, whose radii increase with decreasing surface tension. Contrary to what we found in  $Li_{13}Na_{42}$  and  $Na_{13}Cs_{42}$ , however, the surface of  $Li_{13}Na_{30}Cs_{12}$  remains solid up to temperatures which are significantly higher than typical surface melting temperatures of homogeneous  $Cs$  clusters.<sup>18</sup> We thus point out the possibility of enlarging the thermal stability range of solid surfaces of clusters by alloying, an observation relevant to applied fields such as heterogeneous catalysis.

The rest of the paper is structured as follows: section II presents a brief summary of our theoretical method, a full account of which can be found in our recent publications.<sup>5</sup> Section III describes our (necessarily approximate) strategy to locate minimum-energy isomers and an extensive analysis of the MD simulations of cluster melting. Finally, section IV offers some concluding remarks.

## II. Theory

For a given spatial configuration of atoms, we evaluate the energy of the cluster and the force acting on each atom by employing density functional theory (DFT) in its Hohenberg–Kohn (HK)<sup>19</sup> representation where the valence electron density stands as the basic variable, thus avoiding employment of auxiliary one-particle orbitals as in its Kohn–Sham (KS)<sup>20</sup> representation. The details of our implementation of this so-called orbital-free DFT scheme have been described

in previous work,<sup>4,5,21–23</sup> so we just present briefly the main technical issues. The electronic kinetic energy functional of the electron density is approximated by the gradient expansion around the homogeneous limit through second order.<sup>19,24–26</sup> This means that we keep the local Thomas–Fermi term and the lowest order density gradient correction. The local density approximation is used for exchange and correlation.<sup>27,28</sup> The ionic field acting on the electrons is represented by the local pseudopotential of Fiolhais et al.<sup>29</sup> We have shown in recent publications,<sup>4,5</sup> by explicit comparison of HK- and KS-DFT calculations, that the orbital-free level of theory is adequate to study alkali clusters. This conclusion may not apply to more complex metallic elements, for which an extension either of the electronic kinetic energy functional or the local pseudopotential might be needed.

The cluster under study is placed in a unit cell of a cubic superlattice with edge 62 au and the set of plane waves periodic in that superlattice, up to an energy cutoff of 20 Ryd, is used as a basis set to expand the valence electron density. Following Car and Parrinello,<sup>30</sup> the coefficients of that expansion are regarded as generalized coordinates of a set of fictitious classical particles, and the corresponding Lagrange equations of motion for the electron density distribution are solved in order to determine the optimal electron density for each atomic configuration, as described in ref 5. Forces on atoms are then evaluated by using Hellmann–Feynman’s theorem. Thus the dynamics of ions is not Car–Parrinello but Born–Oppenheimer. Fourier transforms are calculated on a  $144 \times 144 \times 144$  mesh. The equations of motion are integrated using the Verlet algorithm<sup>31</sup> for both electrons and ions, with time steps of  $1 \times 10^{-4}$  atu and  $3 \times 10^{-3}$  atu for the electronic and ionic motions, respectively. These choices resulted in a conservation of the total energy better than 0.1%. Several MD runs at different constant energies were performed in order to obtain the caloric curve for each cluster. Previous to each constant-energy run, isokinetic thermalization runs were performed to fix the average value of the temperature. The total simulation time was at least 100 ps for each run at constant energy, but for those energies close to the melting-like transition, some runs longer than 200 ps were performed. The total simulated time for each isomer was close to 2 ns.

The theoretical indicators employed to locate the melting-like transition are as follows: (a) the caloric and specific heat curves as a function of the internal cluster temperature, which is defined through the equipartition theorem for the ionic kinetic energy; (b) the root-mean-squared bond-length fluctuation parameter  $\delta$ ; (c) the diffusion coefficient, obtained from the long time behavior of the mean square displacement; (d) short-time averages of the “atomic equivalence indexes”<sup>32</sup>

$$\sigma_i(t) = \sum_j |\vec{R}_i(t) - \vec{R}_j(t)| \quad (1)$$

where  $\vec{R}_i(t)$  gives the position of atom  $i$ . These indexes take different values for geometrically inequivalent atoms, and thus are very useful indicators of isomerization and/or



melting transitions; (e) the microcanonical average of the atomic distribution function, defined by

$$dN_{at}(r) = g(r)dr \quad (2)$$

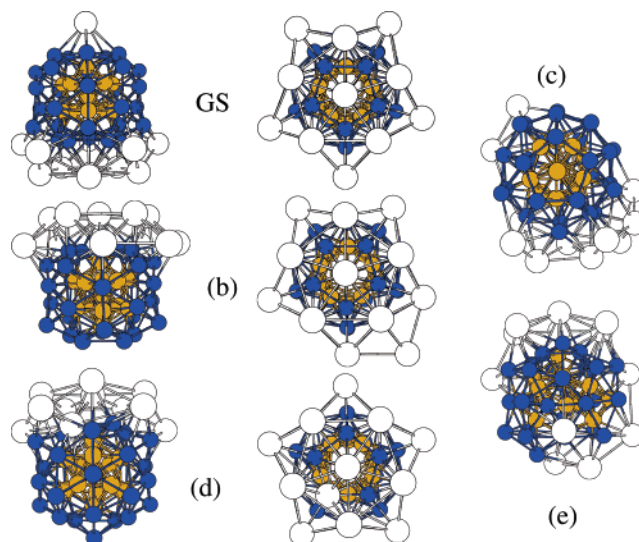
where  $dN_{at}(r)$  is the number of atoms at distances from the center of mass between  $r$  and  $r + dr$ . By including just the number of atoms of a given species in the previous expression, we may obtain the partial contributions of each atomic type to the total  $g(r)$ . Explicit expressions for the rest of indicators can be found in previous publications.<sup>4,5,21–23</sup>

### III. Results

**A. Lowest Energy Isomers.** We have not tried to extensively sample the energy landscape of  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$  by using global optimization techniques such as genetic<sup>12,33</sup> or basin hopping<sup>34</sup> algorithms, as this is a prohibitive task for the energy model we are employing. Instead, we have adopted an approximate sampling scheme, based on dynamical simulated annealing runs and the direct generation of physically motivated structures. This procedure will lead to isomers which are at least reasonably close to the real ground-state structure. Specifically, simulated annealing runs were performed starting from a liquid cluster equilibrated at 200 K, at a cooling rate of 0.2 K/ps, which means a simulation length of 1 ns takes the cluster to 0 K. In practice, we rather stopped the annealing simulation at 5 K and then performed a conjugate gradients optimization of the resulting structure. We also considered isomers constructed by hand with icosahedral, decahedral and cuboctahedral symmetries, which are between the expected topological structures for metal clusters. For each of these isomers, we additionally performed a mild annealing simulation, by heating the cluster to approximately 100 K (in any case, a temperature lower than the melting point) and cooling it down at a rate of 0.4 K/ps. In many cases, this has the effect of locating an isomer of the same symmetry as the original one but with a slightly lower energy.

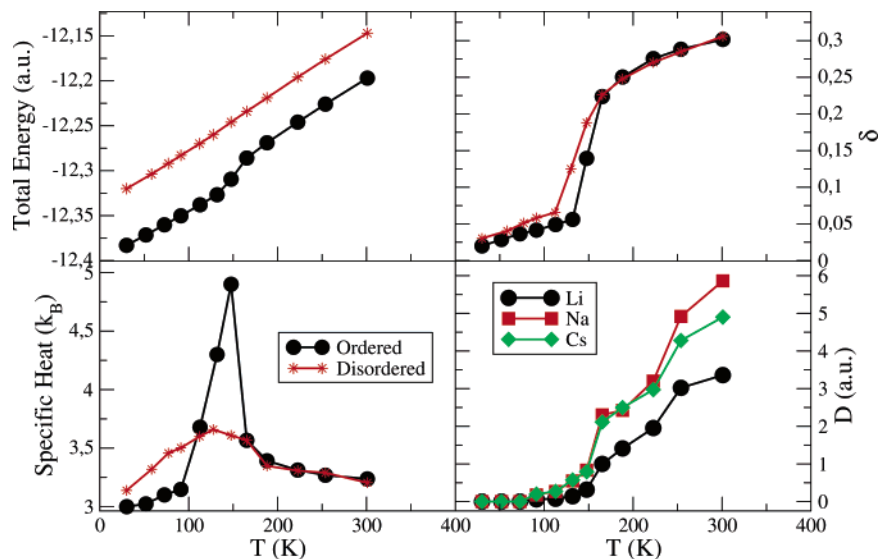
In the construction of isomers with icosahedral, decahedral and cuboctahedral structures, we took advantage of the knowledge gathered in our previous work on binary Li–Na and Na–Cs clusters.<sup>6</sup> There we found that segregation to the cluster surface of the atomic species with lower surface tension (and, in the case of alkalis, larger atomic radius) was always favored. Thus, we start from an inner core with 13 Li atoms and study what is the best way of adding Na atoms to its surface and then Cs atoms to the resulting structure. Figure 1 shows some of the minimum energy isomers found through this manual procedure as well as from unbiased simulated annealing runs. The isomers obtained from simulated annealing show reduced structural order but competitive energies, and all low-energy isomers are based on poly-icosahedral (or anti-Mackay) growing (even though this is less obvious by visual inspection for isomers (c) and (e), for reasons explained below).

In the ground-state isomer as well as in isomers (b) and (d) twenty Na atoms sit on the faces of the  $\text{Li}_{13}$  icosahedral core and ten Na atoms on top of vertex sites of  $\text{Li}_{13}$ . The differences between these three isomers (and also some higher-energy isomers not explicitly shown) are mostly the



**Figure 1.** Lowest energy structure (GS) and low-lying isomers of  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$ . Side and top views are offered for isomers GS, (b) and (d), which show a higher structural order, while a single view is offered of isomers (c) and (e), which are representative of the results of simulated annealing runs. Small light balls represent Li atoms, medium-size dark balls represent Na atoms, and large white balls represent Cs atoms. The energy differences with respect to the ground-state isomer (GS), in meV/atom, are 0.49 (b), 2.22 (c), 2.34 (d) and 6.7 (e).

allocation of the remaining Cs atoms. In the GS isomer, two of them complete the anti-Mackay 32-atom shell by sitting on top of the uncapped vertex sites of  $\text{Li}_{13}$ , while the others form a single umbrella on top of the  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_2$  structure. The lonely Cs atom in the GS isomer moves close to the rest of Cs atoms in isomer (b), so that one of the vertex sites of  $\text{Li}_{13}$  remains uncapped. In isomer (d), this vertex site is capped by a Na atom, and one of the Cs atoms sits on top of a face position of  $\text{Li}_{13}$ . Isomers (c) and (e) are representative of the kind of isomers found from simulated annealing runs. Isomer (c) contains an icosahedral  $\text{Li}_{12}\text{Na}$  core, which shows that slight mixing of Li and Na species has not a big energy penalty. The core shell of isomer (e) is a  $\text{Li}_9\text{Na}_4$  icosahedron. Due to the larger number of Na atoms in the inner shell, the second shell is formed by a mixture of Li, Na, and also Cs atoms. There are also some Na atoms in the most external surface shell, so mixing is much more pronounced in this isomer. Growing is still strictly poly-icosahedral, so that, apart from the outermost surface shell, all isomers in Figure 1 are just homotops of the same inherent structure. The reason that isomers (c) and (e) are more structurally disordered is just the local distortions appearing due to the different atomic species. In fact, we did not find any isomer with a type of growing other than anti-Mackay in any of the simulated annealing runs, which demonstrates that this is really the preferred structural motif when different shells are preferentially formed by alkalis of different size. However, the simulated annealing runs, at least at the cooling rates employed here, are not able to find the optimal compositional order.



**Figure 2.** Caloric and specific heat curves (left side) and rms bond length fluctuation and diffusion constants (right side) of  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$ , taking the average internal cluster temperature as the independent variable. The diffusion coefficients are shown just for the ground-state isomer, while the rest of indicators are also shown for the isomer (e) in Figure 1. The caloric curve for this last isomer has been vertically displaced for clarity.

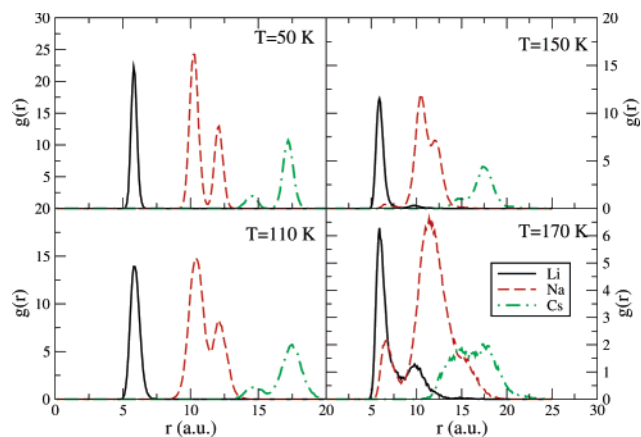
Polyicosahedral structures in binary metal clusters have been independently identified by Rossi et al.<sup>13</sup> for the case of noble metals and by Aguado and López<sup>6</sup> for the case of alkali metals. In homogeneous clusters, this growing pattern accumulates a very large strain in the inner 13-atom icosahedron and thus is not usually observed. In a binary AB nanoalloy, on the contrary, the atomic species with the shortest equilibrium interatomic distance and/or higher surface tension (A) may form the inner  $A_{13}$  icosahedral core with little or no accumulated strain. At the same time, A–B interatomic distances and packing of B-atoms can both be optimized by anti-Mackay growing because only 20 faces (as opposed to 30 edges) need to be capped. As Rossi et al. point out, particularly stable polyicosahedral structures are expected whenever the differences in surface energies and atomic sizes of the species involved are sufficiently large. In this work, we have shown this to be generalizable to the case of ternary Li–Na–Cs nanoalloys. However, there is an important difference between  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$  and the binary clusters considered in ref 6: as an anti-Mackay overlayer is completed at composition  $A_{13}B_{32}$ , the 10 Cs atoms forming the outermost shell in  $\text{Na}_{13}\text{Cs}_{42}$  sit on top of an inner Cs shell, and these external Cs atoms do not strictly conform to polyicosahedral order. In  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$ , each shell is mostly formed by a different type of atom, which enforces strict polyicosahedral order. We will see that the higher structural order confers the cluster with enhanced thermal stability.

To close this section, we would like to emphasize that the energetic ordering of the isomers presented here is preserved by orbital-based KS-DFT calculations performed with the SIESTA code,<sup>35</sup> under the same approximation for exchange-correlation effects and with core electrons substituted by norm-conserving pseudopotentials<sup>36</sup> in their fully nonlocal form.<sup>37</sup> This is a general result that we have found in all our previous studies on alkali clusters, which shows that the energy landscape generated by the orbital-free technique is realistic for these materials. Binding energies,

measuring the energy required to dissociate the cluster into infinitely separated atoms, are usually much less accurate than energy differences between isomers,<sup>21</sup> because the present orbital-free technique is not well suited to the description of an isolated atom. This means that the evaporation limit cannot be properly described by our method, but the solid-to-liquid transition of clusters of these sizes can be safely addressed. Average interatomic distances are of the order of 1–2% longer than corresponding ab initio values, so melting points may be somewhat underestimated compared to Kohn–Sham results. We are presently considering the possibility that an adaptation of the local pseudopotentials (generated in a bulk environment) to a cluster environment may help to solve this problem, providing results in even better agreement with KS calculations. Finally, the ground-state structure shown in Figure 1 is stable against mild annealing from a temperature of 100 K, and no isomers with significantly lower energy were found in the heating runs reported in the next sections. Thus, the structure proposed is really a good candidate for the minimum energy structure.

**B. Melting-like Transition.** In this section, we analyze by constant-energy MD simulations the melting process for two  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$  isomers, namely the ground state and isomer (e). As it is well-known, the melting transition may be intrinsically isomer-dependent in these simulations if several basins in the energy landscape are separated by high energy barriers. We thus explicitly consider the melting transition of isomer (e) in order to have an estimate for the effect of compositional and structural order on melting properties.

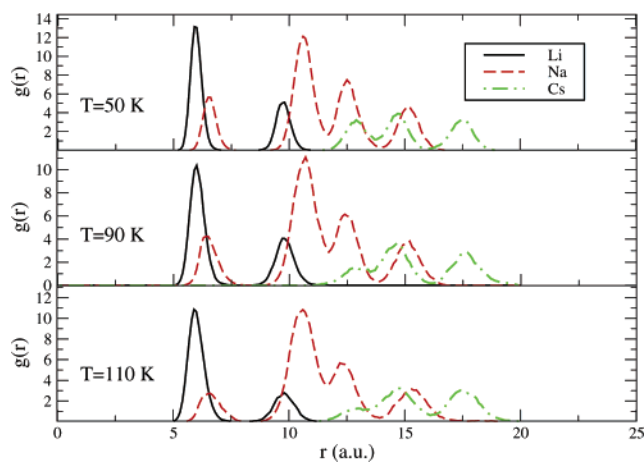
Figure 2 shows the caloric curve as well as the temperature evolution of the specific heat, rms bond length fluctuation and diffusion constants. The most noteworthy result from this figure is that the caloric and specific heat curves of both isomers are very different, even though the location of the melting point itself, given the difficulty of obtaining con-



**Figure 3.** Partial contribution of each atomic species to the time averaged radial atomic density distribution of the ground-state isomer of  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$ , at some representative temperatures. The  $g(r)$  distributions are calculated with respect to the innermost Li of the 13-atom icosahedral core for both isomers. Due to the nonspherical shape of the cluster, employing the center of mass (as was done in our previous publications) does not permit to distinguish the atomic shell structure. At  $T = 170$  K (lower-right panel), that innermost Li atom also interchanges its position with other Li atoms. In this case, the  $g(r)$  at each time step was evaluated with respect to that Li atom which instantaneously occupies the innermost position.

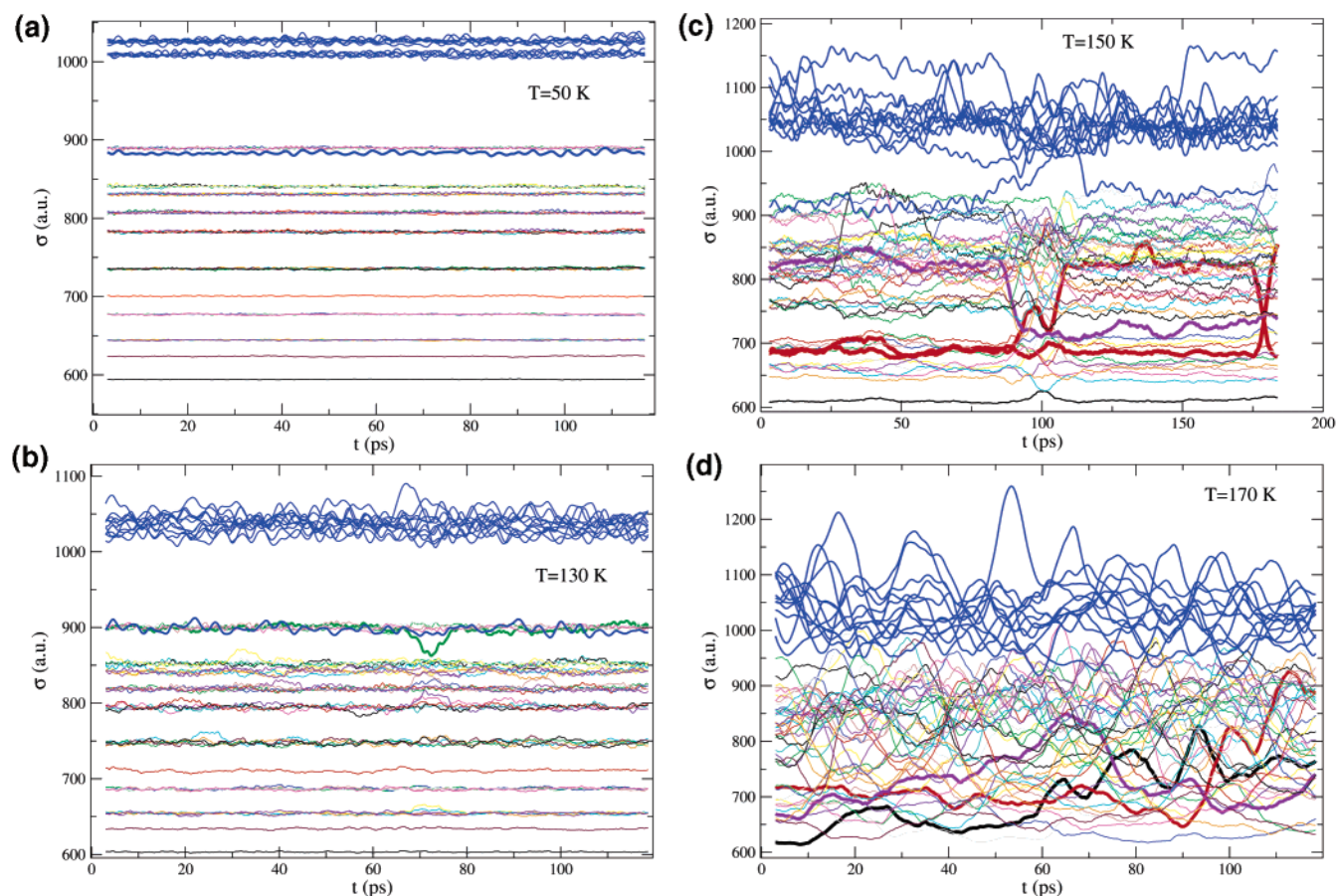
verged statistical averages in the transition region, is about the same in both cases. It is important to remark here that in a fully ergodic simulation all isomers compatible with a given constant energy  $E$  are visited, and thus the differences in specific heat curves can only reflect the nonergodicity of MD simulations. Put in different terms, the averages shown in Figure 2 are not strictly thermodynamic. Nevertheless, what is really important is to understand which situation (if any) is realized in a given experimental determination of melting properties. In the calorimetric experiments of Haberland's group, for example,<sup>38</sup> mass-selected clusters are first thermalized by direct contact with a macroscopic heat bath and then transferred to a high-vacuum region, where each cluster evolves at constant energy. The lower the temperature of the heat bath, the narrower the energy distribution of isomers obtained in the thermalization step, and thus the obtained caloric curve is expected to approach that calculated from constant-energy MD simulations of melting of the GS isomer alone. On the contrary, higher initial temperatures will lead to experimental caloric curves which should be compared to some weighted average of those obtained from constant-energy MD simulations of melting of several isomers.

All indicators in Figure 2 predict that  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$  melts at an approximate temperature  $T_m$  of 140 K. The main difference between both isomers is that the GS isomer is much more resistant to premelting (isomerization) effects, as evidenced by the different values of the specific heat at low temperatures. To better understand the origin of the differences, Figures 3 and 4 show partial contributions to the radial atomic density distributions of each isomer, and Figures 5 and 6 short-time averages of the atomic equivalence

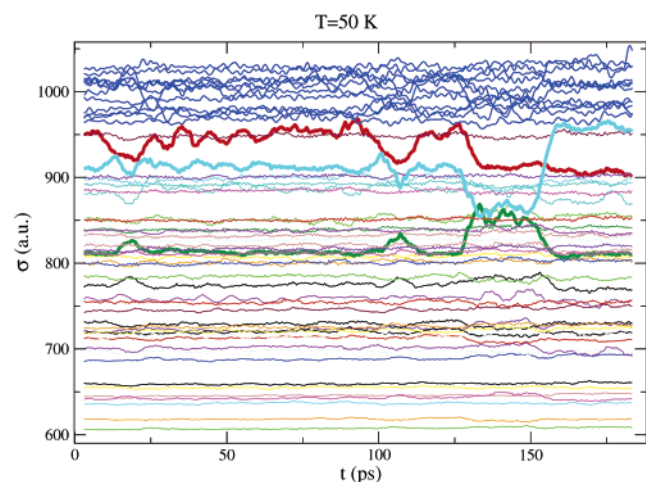


**Figure 4.** Same as Figure 3, but for isomer (e) in Figure 1.

indexes, for several average temperatures. Regarding the GS isomer, the partial contributions to  $g(r)$  hardly overlap up to the melting point. At  $T=150$  K (slightly above  $T_m$ ), we observe the first interchange between Li and Na atoms, which means that the basin of isomer (c) starts to be visited. At a higher temperature of 170 K, Li–Na mixing is enhanced, and some overlap of Na and Cs contributions is also appreciated at the surface. In this region, isomers similar to (e) are routinely visited. Figure 4 shows that the cooling rate computationally affordable in the simulated annealing runs is still too fast to find the correct distribution of atoms within the polyicosahedral shells. In fact, as far as the partial contributions to  $g(r)$  are concerned, the structure of isomer (e) is very similar to an average liquid structure. Thus, the distribution of atoms in shells does not appreciably change upon heating. Figures 5a and 6 show that radial structural order is much higher in the GS structure, as the several  $\sigma$  lines are separated by well defined gaps, while just a loose differentiation between core and surface shells is apparent for isomer (e). Nevertheless, both isomers have approximately the same volume, despite the strong local relaxations in isomer (e). As stated in the previous section, this important difference in the atomic equivalence indexes is just driven by local structural relaxations, as both isomers are homotops of the same topological isomer. Figure 5b shows that at a temperature as high as 130 K, the GS isomer is still a hot solid, while structural isomerizations are observed in isomer (e) even at 50 K (Figure 6), implying that we are in a region of configurational space with a high density of isomeric states, which are separated by sufficiently low barriers (this is a feature typical of the energy landscapes of amorphouslike materials). Figure 4 shows that these low-T isomerizations try to drive the cluster structure toward that of the GS isomer: for example, Na–Cs mixing is reduced upon heating to 90 K, and the same is true of Li–Na mixing at 110 K (note the heights of the several  $g(r)$  peaks). This observation suggests that a much slower cooling rate in simulated annealing runs would result in better approximations to the GS structure, as expected. For the GS isomer, the liquid structure is not yet well developed even at 150 K (Figure 5c). For example, the innermost Li-atom stays vibrating about its equilibrium position for the whole simulation, which is close to 200 ps long. The cluster is clearly liquidlike at 170



**Figure 5.** Time evolution of atomic equivalence indexes of the ground-state isomer of  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$ , averaged over time intervals of 1000 steps, at four representative temperatures.  $\sigma$  curves corresponding to Cs atoms as well as those of other species involved in isomerization transitions and diffusion are represented by bold lines in order to help visualization.



**Figure 6.** Same as Figure 5, but for the disordered  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$  isomer at 50 K.

K, however (Figure 5d), in correspondence with the steplike increase in diffusion appreciated in Figure 2.

#### IV. Summary and Discussion

In this paper, orbital-free DFT molecular dynamics simulations have been employed in order to analyze the structure and melting mechanism in a ternary alkali nanoalloy, namely  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$ . The orbital-free DFT method is chosen because

it provides a good compromise between statistical accuracy and realistic interatomic forces.

The lowest-energy structure is dictated by the relief of core strain and the tendency of surface bonds to contract. When the atomic species have significantly different sizes and surface energies, these “rules” lead to polyicosahedral clusters (with segregation of the lower surface energy component to the cluster surface) in a natural way. No significant strain is accumulated in the core shell of these structures as the smaller size species has a shorter equilibrium bond distance. At the same time, optimal distances for the surface bonds may be obtained by growing of the large size species on the faces of the inner icosahedron (growing on edge sites would result in surface bonds which are too short). Anti-Mackay growing is so energetically favorable that none of the simulated annealing runs has found even one isomer with a different symmetry. Simulated annealing is nevertheless not able to find the correct compositional order for computationally affordable cooling rates.

The most important finding is that the melting transition in the GS isomer of  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$  is located at a temperature which is approximately 50 K higher than typical surface melting temperatures of pure Cs clusters obtained with the same method.<sup>18</sup> On the contrary, premelting effects are noticeable for  $\text{Li}_{13}\text{Na}_{42}$  and  $\text{Na}_{13}\text{Cs}_{42}$  at considerably lower temperatures as compared to homogeneous Na and Cs

clusters, respectively.<sup>6</sup> The main difference between both cases is that the GS structure of  $\text{Li}_{13}\text{Na}_{30}\text{Cs}_{12}$  is strictly polyicosahedral, with atoms of different species mostly occupying different radial shells; on the contrary, the outermost surface atoms of  $\text{Li}_{13}\text{Na}_{42}$  and  $\text{Na}_{13}\text{Cs}_{42}$  grow on top of an inner shell formed by ions of the same type, and the growing pattern is not polyicosahedral in that external region. It thus seems that strict polyicosahedral packing of approximately spherical layers, each formed by alkali atoms of different types lead to enhanced thermal stability of the solid phase. The net result is that diffusion of surface Cs atoms grown on a Cs “substrate” is easier than the corresponding diffusion on a substrate formed by alkali atoms of smaller size. In the future, we plan to undertake MD simulations of a larger number of binary and ternary alkali clusters with geometrically compact structures in order to determine more unambiguously the physical reason for this melting temperature increase. It is important to get such a sound understanding because it might be used to build material surfaces which remain solid up to significantly higher temperatures than expected.

We have also explicitly simulated the thermal behavior of a low-lying isomer of the same cluster, obtained from simulated annealing runs. This isomer also has a perfect polyicosahedral structure, but each radial shell is formed by a mixture of atoms of different types. As a result of this partial compositional disorder, local structural distortions appear in much the same way as in a bulk alloy, and the resulting distribution of atoms in radial shells is reminiscent of that expected for an amorphous-like structure. Premelting effects, in this case realized as structural isomerizations, are then observed even at very low temperatures. This seems to be the most general behavior of alkali nanoalloys, while the enhanced melting temperatures are observed only in structures with high topological (polyicosahedral in the case at hand) and compositional orders.

Finally, we have emphasized that, even though the premelting effects are highly isomer-dependent, the homogeneous melting temperature itself is the same for both isomers within the statistical accuracy of our simulations. The caloric curves are different for both isomers simply because constant-energy MD simulations, started from a given isomer, can only sample the region of phase space which is dynamically accessible to that isomer. At low energies, this represents an essential loss of ergodicity (one which cannot be restored just by increasing the length of the simulation) if basins of different isomers are separated by high energy barriers. We stress this point because we believe that our MD runs for the GS structure are long enough as to obtain nearly-converged results with respect to simulation time, except possibly for the transition region. The simulations presented for the low-lying isomer, where relatively *high energies* (and, correspondingly, phase-space regions with high density of states) are sampled at *low temperatures*, might suffer more from convergence problems at the lowest temperatures. In any case, we have indicated that care must be exercised in comparing the MD results to a hypothetical experimental determination of the caloric curve.

**Acknowledgment.** This work was supported by Junta de Castilla y León (Project VA073/02) and DGES (Project MAT2002-04393-C02-01). Thanks are due to J. M. Soler and his team for providing us with a copy of the SIESTA code and to M. J. López for helping us with the construction of icosahedral, cuboctahedral and decahedral geometries. A.A. also acknowledges financial support from the Spanish Ministry of Science and Technology, under the Ramón y Cajal program.

## References

- (1) Haruta, M. *Catal. Today* **1997**, *36*, 153–166.
- (2) López, M. J.; Marcos, P. A.; Alonso, J. A. *J. Chem. Phys.* **1996**, *104*, 1056–1066.
- (3) Jellinek, J.; Krissinel, E. B. *Chem. Phys. Lett.* **1996**, *258*, 283–292. Krissinel, E. B.; Jellinek, J. *Chem. Phys. Lett.* **1997**, *272*, 301–312.
- (4) Aguado, A.; González, L. E.; López, J. M. *J. Phys. Chem. B* **2004**, *108*, 11722–11731; Aguado, A.; Núñez, S.; López, J. M. *Comput. Mater. Sci.* in press.
- (5) Aguado, A.; González, D. J.; González, L. E.; López, J. M. in *Progress in Chemical Physics Research*; Columbus, F., Ed.; Nova Science Publishers: New York, 2005, in press.
- (6) Aguado, A.; López, J. M. *Phys. Rev. B* in press.
- (7) López, M. J.; Mañanes, A.; Alonso, J. A.; Íñiguez, M. P. *Z. Phys. D* **1989**, *12*, 237–239; López, M. J.; Íñiguez, M. P.; Alonso, J. A. *Phys. Rev. B* **1990**, *41*, 5636–5642; Mañanes, A.; Íñiguez, M. P.; López, M. J.; Alonso, J. A. *Phys. Rev. B* **1990**, *42*, 5000–5008.
- (8) Bol, A.; Alonso, J. A.; López, J. M.; Mañanes, A. *Z. Phys. D* **1994**, *30*, 349–356. Bol, A.; Martín, G.; López, J. M.; Alonso, J. A. *Z. Phys. D* **1993**, *28*, 311–319.
- (9) Deshpande, M. D.; Kanhere, D. G.; Panat, P. V.; Vasiliev, I.; Martin, R. M. *Phys. Rev. A* **2002**, *65*, 053204(1)–053204(5). Deshpande, M. D.; Kanhere, D. G.; Vasiliev, I.; Martin, R. M. *Phys. Rev. A* **2002**, *65*, 033202(1)–033202(6).
- (10) Joshi, K.; Kanhere, D. G. *Phys. Rev. A* **2002**, *65*, 043203(1)–043203(7).
- (11) Chacko, S.; Kanhere, D. G.; Paranjape, V. V. *Phys. Rev. A* **2004**, *70*, 023204(1)–023204(9).
- (12) Darby, S.; Mortimer-Jones, T. V.; Johnston, R. L.; Roberts, C. *J. Chem. Phys.* **2002**, *116*, 1536–1550. Bailey, M. S.; Wilson, N. T.; Roberts, C.; Johnston, R. L. *Eur. Phys. J. D* **2003**, *25*, 41–55.
- (13) Rossi, G.; Rapallo, A.; Mottet, C.; Fortunelli, A.; Baletto, F.; Ferrando, R. *Phys. Rev. Lett.* **2004**, *93*, 105503(1)–105503(4).
- (14) Chushak, Y. G.; Bartell, L. S. *J. Phys. Chem. B* **2003**, *107*, 3747–3751.
- (15) Huang, S.; Balbuena, P. B. *J. Phys. Chem. B* **2002**, *106*, 7225–7236.
- (16) Joshi, K.; Kanhere, D. G. *J. Chem. Phys.* **2003**, *119*, 12301–12307.
- (17) Mainardi, D. S.; Balbuena, P. B. *Int. J. Quantum Chem.* **2001**, *85*, 580–591.
- (18) Aguado, A. *Phys. Rev. B* **2001**, *63*, 115404(1)–115404(9).
- (19) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.
- (20) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.

- (21) Aguado, A.; López, J. M.; Alonso, J. A.; Stott, M. J. *J. Chem. Phys.* **1999**, *111*, 6026–6035.
- (22) Aguado, A.; López, J. M.; Alonso, J. A.; Stott, M. J. *J. Phys. Chem. B* **2001**, *105*, 2386–2392.
- (23) Aguado, A.; Molina, L. M.; López, J. M.; Alonso, J. A. *Eur. Phys. J. D* **2001**, *15*, 221–227.
- (24) *Theory of the inhomogeneous electron gas*; Lundqvist, S.; March, N. H., Eds.; Plenum Press: New York, 1983.
- (25) Yang, W. *Phys. Rev. A* **1986**, *34*, 4575–4585.
- (26) Perdew, J. P. *Phys. Lett. A* **1992**, *165*, 79–82.
- (27) Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048–5079.
- (28) Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, *45*, 566–569.
- (29) Fiolhais, C.; Perdew, J. P.; Armster, S. Q.; McLaren, J. M.; Brajczewska, H. *Phys. Rev. B* **1995**, *51*, 14001–14011. Fiolhais, C.; Perdew, J. P.; Armster, S. Q.; McLaren, J. M.; Brajczewska, H. *Phys. Rev. B* **1996**, *53*, 13193–13193.
- (30) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474. Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. *Rev. Mod. Phys.* **1992**, *64*, 1045–1097.
- (31) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103. Swope, W. C.; Andersen, H. C. *J. Chem. Phys.* **1982**, *76*, 637–649.
- (32) Bonacić-Koutecký, V.; Jellinek, J.; Wiechert, M.; Fantucci, P. *J. Chem. Phys.* **1997**, *107*, 6321–6334. Reichardt, D.; Bonacić-Koutecký, V.; Fantucci, P.; Jellinek, J. *Chem. Phys. Lett.* **1997**, *279*, 129–139.
- (33) Soler, J. M.; Garzón, I. L.; Joannopoulos, J. D. *Solid State Comm.* **2001**, *117*, 621–625, and references therein.
- (34) Doye, J. P. K. *Phys. Rev. B* **2003**, *68*, 195418(1)–195418(11).
- (35) Soler, J. M.; Artacho, E.; Gale, J. D.; Garcia, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745–2779.
- (36) Hamann, D. R.; Schlüter, M.; Chiang, C. *Phys. Rev. Lett.* **1979**, *43*, 1494–1497.
- (37) Kleinman, L.; Bylander, D. M. *Phys. Rev. Lett.* **1982**, *48*, 1425–1428.
- (38) Schmidt, M.; Kusche, R.; Kronmüller, W.; von Issendorff, B.; Haberland, H. *Phys. Rev. Lett.* **1997**, *79*, 99–102. Schmidt, M.; Kusche, R.; von Issendorff, B.; Haberland, H. *Nature* **1998**, *393*, 238–240. Kusche, R.; Hippler, Th.; Schmidt, M.; von Issendorff, B.; Haberland, H. *Eur. Phys. J. D* **1999**, *9*, 1–6. Schmidt, M.; Haberland, H. *Compt. Rend. Physique* **2002**, *3*, 327–340.

CT049892+

## Linear and Nonlinear Optics Properties of Polyphosphazene/Polynitrile Alternating Copolymers

Denis Jacquemin\*

*Laboratoire de Chimie Théorique Appliquée, Facultés Universitaires Notre-Dame de la Paix, rue de Bruxelles, 61, B-5000 Namur, Belgium*

Received November 18, 2004

**Abstract:** The linear and nonlinear optical properties of polyphosphazene/polynitrile alternating copolymers,  $-(\text{PH}_2 = \text{N} - \text{CH} = \text{N})_N$ , are studied, at the MP2/6-31G(d) level of approximation, by using an oligomeric approach. We report the evolution with the chain length of the geometry, charges, dipole moments, polarizabilities, and first hyperpolarizabilities of two conformers (trans-transoid and trans-cisoid). Comparisons with the polyphosphazene/polyacetylene and poly(thiophosphazene) structures are performed. It turns out that the polyphosphazene/polynitrile copolymers present dipole moment and polarizabilities comparable to polyphosphazene but larger (+50%) first hyperpolarizabilities. The relative response of these copolymers is comparable to the one of standard push–pull systems.

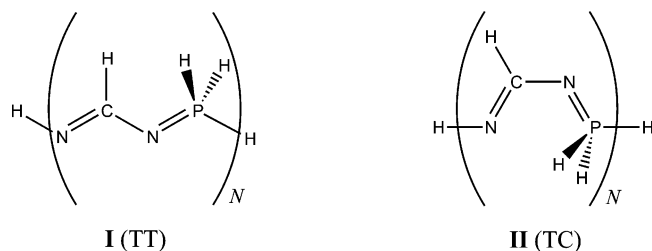
### I. Introduction

In the quest for large first hyperpolarizabilities ( $\beta$ ), different strategies have been set up to obtain organic materials showing large nonlinear optics (NLO) responses.<sup>1</sup> To reach this goal, one of the difficulties is to combine delocalizable electrons which are necessary in order to get large NLO responses and the asymmetry required because  $\beta$  is an odd term in the dipole moment expansion. Recently, we have been interested in AB oligomers<sup>2–7</sup> which show asymmetric unit-cells (two different nuclei and two different bonds) and possess mobile electrons. On the contrary to push–pull chains,<sup>8</sup> a  $\beta$  response may indeed be obtained for any AB chain length, and the  $\beta$  value of the polymer may be nonzero. Two typical examples of AB polymers are polyphosphazene and polynitrile. Polyphosphazene [PP,  $-(\text{P}(\text{R}, \text{R}') = \text{N})_N$  where  $N$  is the number of unit cells] and its derivatives is an archetype of a successful inorganic polymer. Many applications have been found for PP, from flame retardant materials, to proton-exchange membrane in fuel cells. We refer the reader to ref 9 for a list of applications and references. The  $\beta$  of PP has been measured experimentally<sup>10</sup> and studied theoretically at semiempirical<sup>11</sup> and ab initio<sup>2</sup> levels of approximation. In ref 10, it has been found that the response of the PP backbone is negligible, whereas, by adding

chromophore side groups, small nonlinear optics (NLO) responses could be observed. Although the results from semiempirical<sup>11</sup> and ab initio<sup>2</sup> studies differ quantitatively, both found that the  $\beta$  of unsubstituted (i.e.  $\text{R} = \text{R}' = \text{H}$ ) PP should be quite small, about one-third of that of classical push–pull systems. Polynitrile [PN, also named polycarbonitrile or polymethineimine,  $-(\text{C}(\text{R}) = \text{N})_N$ ] has been synthesized by Whörle in the 1970s<sup>12,13</sup> and more recently by Komatsu<sup>14</sup> by ring-opening polymerization (ROP) of triazine. To our knowledge, the NLO properties of PN have not been studied experimentally, but they have been the subject of several theoretical calculations, reviewed in ref 3. It turns out that for PN,  $\beta$  reported to the size of the chain ( $\beta/N$ ) can be extremely large; as large as in the best  $\alpha,\omega$ -nitro,amino-polyacetylene ( $\alpha,\omega$ -nitro,amino-PA) chains, i.e., much larger than in classical push–pull compounds.<sup>15,16</sup>

In 1989, Manners, Allcock, Renner and Nuyken<sup>17</sup> presented the first synthesis of PP/PN copolymers, so-called poly(carbophosphazene). Their procedure is based on a ROP of the six-member cycle  $-\text{C}(\text{Cl})=\text{N}-\text{P}(\text{Cl}_2)=\text{N}-\text{P}(\text{Cl}_2)=\text{N}-$  which leads after 4 h of heating to a polymer with a CNPNPN backbone sequence. By further reaction with NaOPh (PhNH<sub>2</sub>) more stable aryloxy (amino) polymers have been obtained.<sup>17</sup> The aryloxy chains were further characterized by the same group in 1991,<sup>18</sup> and it was found that the poly(carbophosphazene) skeleton is much less flexible than

\* Corresponding author e-mail: denis.jacquemin@fundp.ac.be, URL: <http://www.fundp.ac.be/~jacquemd>.



**Figure 1.** Schematic representation of trans-transoid and trans-cisoid PP/PN alternating copolymers.

PP, probably due to the inherent differences in the  $\pi$ -bonding nature of C=N and P=N double bonds. Following these initial works, additional structures, like various aminocarbo-phosphazenes, have been synthesized.<sup>19</sup> In addition, closely related copolymers, polythiophosphazenes<sup>20</sup> and polythionylphosphazene,<sup>21</sup> both presenting the SNPNP backbone sequence have been obtained by ROP. We refer the reader to ref 22 for a detailed review of these inorganic polymers. If, six-member CNPNCN cycles are well characterized structures,<sup>23,24</sup> there are also a few examples of the corresponding eight member cycles showing the CNPNCNPN sequence.<sup>25,26</sup> In a recent contribution Rivard<sup>26</sup> reports the synthesis of various linear structures including the first component of the alternating PP/PN copolymer, i.e.,  $P(R_2R')=N-C(R)=N-R''$  which has been characterized by X-ray diffraction. Such CNPN chains would certainly be an opportunity for NLO applications if the large NLO response of PN and the stability of PP could be combined.

In this paper, we investigate the geometry, charges, dipole moment, polarizability and first hyperpolarizability of increasingly long PP/PN alternating copolymers showing the CNPN sequence (Structures **I** and **II** in Figure 1). To our knowledge, no previous *ab initio* investigation has been performed on these systems, though a study of the six member cycles with the CNPNCN and SNPNP patterns has been performed by Jaeger and co-workers.<sup>24</sup> The same group evaluated the structure, conformation and flexibility of polythionylphosphazene.<sup>27,28</sup> The present study aims to rationalize the impact of copolymerization on  $\beta$ . Such rationalization is particularly welcome as, in this case, the nature of the two “parent” polymers are very different (organic/inorganic, with second/third row atoms, presenting a delocalized electron cloud/island- $\pi$  bonding, ...).

## II. Computational Details

We have selected the MP2/6-31G(d) level for both the geometry optimization and the calculation of electronic properties. With this approach, one takes into account both the indirect (on the ground-state geometry) and direct (on the wave function) dynamic electron correlation (EC) contributions. This choice matches the computational schemes that have been demonstrated to be adequate for both PP and PN. Indeed, for  $\beta$  calculations on PP, it has been found that the MP2/6-31G(d)/HF/6-31G(d) method is sufficient, i.e., direct/indirect EC effects are significant/negligible.<sup>2</sup> The use of polarization functions being mandatory as expected for compounds presenting third-row atoms. For PN, the impact of EC on the geometry is larger, and the MP2/6-31G//MP2/

6-31G technique provides a semiquantitative estimate of the NLO properties.<sup>16</sup> If the use of polarization (or diffuse) functions modifies the  $\beta$  responses of short PN oligomers,<sup>16,29</sup> it has a small effect on the corresponding polymeric property. For instance, the MP2/6-31G//MP2/6-31G polymeric first hyperpolarizability per unit cell reaches  $13 \times 10^4$  au in PN,<sup>16</sup> whereas the MP2/6-31G(d)//MP2/6-31G(d) value is  $14 \times 10^4$ .<sup>30</sup> A similar MP2/6-31G(d) approach has been selected by Jaeger and co-workers for their investigations of the six-membered cycles.<sup>24</sup> The calculations have been performed with the Gaussian03 program,<sup>31</sup> by using the following procedure:

1. The ground-state geometry of each oligomer has been determined by the optimization of its structural parameters. In this first investigation, two conformations have been chosen: planar trans-transoid (TT) and trans-cisoid (TC). The unit cells of the chains are defined according to Figure 1. In addition to planarity, we have also imposed the linearity of the chain, so that the longitudinal axis could be properly defined. To avoid bent chains, all the backbone angles of TT chains have been set equal, whereas in TC chains only two different backbone angles have been used ( $X=N-Y$  and  $N=X-N$ ;  $X, Y = P$  or  $C$ ).<sup>32</sup> After the geometry optimization, each oligomer has been oriented in the Cartesian frame so that the longitudinal axis runs through the center of the first and the last double bonds.

2. In addition to the dipole component parallel to the longitudinal axis ( $\mu_L$ ), the partial atomic charges have been computed for the optimized geometries using the Merz–Kollman (MK)<sup>33</sup> approaches within the MP2/6-31G(d) approximation.

3. Static electronic polarizabilities ( $\alpha$ ) and first hyperpolarizabilities ( $\beta$ ) have been evaluated for the optimized geometries. In quasilinear chains, the longitudinal components of  $\alpha$  and  $\beta$  tensors ( $\alpha_L$  and  $\beta_L$ ) often dominate the total response for sufficiently long chains. For instance, at the HF/6-31G(d)/MP2/6-31G(d) level, the  $\beta_L$  of the TC octamer ( $N=8$ ; structure **II**) is 5.7 times larger than the next larger component. For this reason, we focus on the longitudinal components in this paper. Because their practical determination at EC levels remains cpu-costly for extended oligomers, the vibrational contributions to  $\alpha$  and  $\beta$  ( $\alpha^v$  and  $\beta^v$ ) have been neglected, although, they could make important contributions to the total static values in conjugated systems.<sup>34–36</sup> Static  $\alpha_L$  and  $\beta_L$  have been evaluated by using the numerical finite-field procedure based on the differentiation of the energies computed under several electric field amplitudes. We refer the reader to ref 4 for a complete description of this procedure. It has been found that the accuracy of the procedure could be slightly improved by performing the finite field (FF) procedure on the MP2 energy corrections (rather than the MP2 total energies) and adding the results to the fully analytic coupled-perturbed Hartree–Fock (CPHF) results computed with the same basis set on the same geometry. By doing so, the final accuracy on the MP2  $\alpha_L$  values is 0.1 au, whereas the accuracy on  $\beta_L$  is estimated to be  $\sim 1\%$ . In this paper we adopt the usual sign convention for  $\beta_L$ : positive when orientated in the same direction as the dipole moment, negative otherwise.



**Table 1.** Evolution with Chain Length of the Central Bond Lengths and Bond Angles of Structures **I** and **II**<sup>a</sup>

| N | TT, <b>I</b> |           |           |           |                                   | TC, <b>II</b> |           |           |           |                  |                  | $\Delta E$ |
|---|--------------|-----------|-----------|-----------|-----------------------------------|---------------|-----------|-----------|-----------|------------------|------------------|------------|
|   | $d_{N=C}$    | $d_{C=N}$ | $d_{N=P}$ | $d_{P=N}$ | $\alpha_{X=N-Y} = \alpha_{N=X-N}$ | $d_{N=C}$     | $d_{C=N}$ | $d_{N=P}$ | $d_{P=N}$ | $\alpha_{N=X-N}$ | $\alpha_{X=N-Y}$ |            |
| 1 | 1.285        | 1.391     | 1.582     |           | 120.5                             | 1.298         | 1.374     | 1.616     |           | 121.7            | 108.6            | +13        |
| 2 | 1.291        | 1.372     | 1.573     | 1.645     | 118.2                             | 1.320         | 1.356     | 1.596     | 1.692     | 119.5            | 114.4            | +16        |
| 3 | 1.298        | 1.360     | 1.592     | 1.664     | 117.7                             | 1.322         | 1.351     | 1.609     | 1.680     | 119.2            | 115.6            | +17        |
| 4 | 1.305        | 1.351     | 1.598     | 1.655     | 117.5                             | 1.328         | 1.344     | 1.612     | 1.672     | 119.1            | 116.2            | +17        |
| 5 | 1.309        | 1.346     | 1.606     | 1.649     | 117.4                             | 1.330         | 1.340     | 1.617     | 1.666     | 119.1            | 116.6            | +17        |
| 6 | 1.313        | 1.341     | 1.607     | 1.644     | 117.3                             | 1.332         | 1.337     | 1.619     | 1.662     | 119.2            | 116.9            | +17        |
| 7 | 1.316        | 1.338     | 1.614     | 1.640     | 117.3                             | 1.333         | 1.336     | 1.622     | 1.660     | 119.2            | 117.1            | +17        |
| 8 | 1.318        | 1.336     | 1.617     | 1.638     | 117.2                             | 1.335         | 1.334     | 1.623     | 1.658     | 119.2            | 117.2            | +17        |

<sup>a</sup> All results have been obtained at the MP2/6-31G(d) level. Bond lengths are given in Å, angles in degrees. At the extreme right of the table, the energetic difference between TT and TC conformers ( $\Delta E$ ) is reported in kcal/mol per N.

4. The polymeric responses have been obtained by extrapolating the oligomeric values. To carry out the extrapolations, we define the  $\beta_L$  (and  $\mu_L$  and  $\alpha_L$ ) per unit cell as  $\Delta\beta_L(N) = [\beta_L(N) - \beta_L(N-1)]$ . This definition removes most of the chain end effects and leads to a fast convergence toward the asymptotic limit ( $N \rightarrow \infty$ ). Our fitting procedure allows to obtain  $\Delta\mu_L(\infty)$ ,  $\Delta\alpha_L(\infty)$  and  $\Delta\beta_L(\infty)$  as well as estimates of the corresponding standard deviations. We refer the reader to ref 37 for extra details. To allow direct comparison with previous works, we have also defined the polymeric  $\beta$  per backbone atom:  $\Delta\beta_L^{\text{at}}(\infty) = \Delta\beta_L(\infty)/4$  (and similarly for  $\mu$  and  $\alpha$ ).

### III. Results

**A. Geometries and Charges.** The backbone geometry and relative stability of the TT and TC conformers are given in Table 1. First, we note that the TT structures are not favored, the TC structures being more stable by 17 kcal/mol per cell, provided  $N \geq 2$ . This could have been expected because the TC oligomers are the most stable in PP<sup>2,38</sup> and the TT conformers are also the least stable in PN<sup>3,39</sup> which favors off-planar structures. For the smallest chain of **II**, the bond lengths are 1.30, 1.37, and 1.62 Å, in reasonable agreement with the experimental values obtained for a similar structure (**5** of ref 26): 1.28, 1.40, and 1.54 Å, with packing effects and side substitutions by large groups (Cl, Ph and *t* Bu). The C=N and P=N bond lengths are also in good agreement with the MP2/6-31G(d) values obtained for the  $-C(\text{Cl})=N-P(\text{Cl}_2)=N-P(\text{Cl}_2)=N-$  cycle by Jaeger et al.: 1.33 and 1.60 Å, respectively.<sup>24</sup> As TT chain length is increased, the double bonds become longer and the single bonds shorter as expected in conjugated chains. For the longest chains, the bond length alternation,  $\Delta r^{\text{TT}} = d_+ - d_-$  is 0.018 Å for the PN segment and 0.021 Å for the PP segment. In TC chains, the qualitative behavior is similar, but the two CN bonds are almost equal for medium chains; it even seems that the double/single bond ordering is reversed for  $N \geq 8$ . For  $N = 8$ ,  $\Delta r^{\text{TC}}$  attains  $-0.001$  Å for the PN part and 0.035 Å for the PP part. The negative  $\Delta r$  means that the PN component of long oligomers would favor a cis-transoid (with the double bonds being parallel rather than perpendicular to the longitudinal axis) over a trans-cisoid conformation. These results can be compared to the MP2/6-31G(d)  $\Delta r$  obtained for the 16-unit of TT PN (0.096 Å), TC PN (0.111 Å) and TC PP (0.011 Å) oligomers. Consequently, the copolymer-

**Table 2.** Evolution with Chain Length of the Charges Borne by the Central Atoms<sup>a</sup>

| N | TT, <b>I</b>     |                |                  |                | TC, <b>II</b>    |                |                  |                |
|---|------------------|----------------|------------------|----------------|------------------|----------------|------------------|----------------|
|   | $q^{\text{N}^1}$ | $q^{\text{C}}$ | $q^{\text{N}^2}$ | $q^{\text{P}}$ | $q^{\text{N}^1}$ | $q^{\text{C}}$ | $q^{\text{N}^2}$ | $q^{\text{P}}$ |
| 1 | -0.68            | 0.48           | -0.67            | 0.68           | -0.76            | 0.50           | -0.75            | 0.75           |
| 2 | -0.60            | 0.54           | -0.71            | 1.06           | -0.73            | 0.71           | -0.76            | 0.86           |
| 3 | -0.59            | 0.53           | -0.72            | 1.12           | -0.80            | 0.73           | -0.73            | 0.93           |
| 4 | -0.51            | 0.51           | -0.72            | 1.10           | -0.73            | 0.75           | -0.84            | 1.02           |
| 5 | -0.61            | 0.51           | -0.69            | 1.11           | -0.86            | 0.75           | -0.85            | 1.05           |
| 6 | -0.62            | 0.51           | -0.69            | 1.11           | -0.87            | 0.75           | -0.86            | 1.08           |
| 7 | -0.63            | 0.52           | -0.67            | 1.10           | -0.88            | 0.75           | -0.86            | 1.09           |
| 8 | -0.63            | 0.51           | -0.67            | 1.10           | -0.89            | 0.75           | -0.86            | 1.11           |

<sup>a</sup> All values are in  $e$  and have been obtained within the MK approach at the MP2/6-31G(d)/MP2/6-31G(d) level of theory. N<sup>1</sup> (N<sup>2</sup>) is the nitrogen atom forming the double bond with C (P).

ization clearly results in a  $\Delta r$  decrease for PN (the largest  $\Delta r$ ) and an increase of the  $\Delta r$  of PP (the smallest  $\Delta r$ ). Therefore, it is difficult to predict, at this stage, if the copolymer is more or less delocalizable than its constituents taken separately.

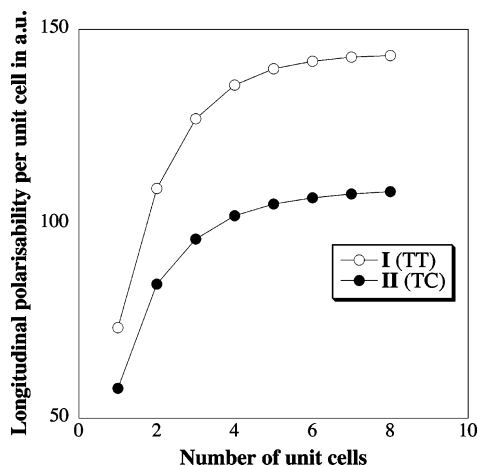
In addition to the geometry of the chains, another parameter is crucial for assessing the delocalization and the asymmetry along the backbone: the charge distribution. Table 2 presents the MP2/6-31G(d) MK charges borne by the central atoms of increasingly long oligomers. For every atoms, the charges are quite constant with chain length. For the most stable conformers (**II**), the charges are  $-0.9e$ ,  $0.8e$ , and  $1.1e$  for N, C and P, respectively. This means that phosphorus and nitrogen on one hand and carbon and nitrogen on the other hand exchange (almost) one electron. In TC PP the same behavior was found,<sup>38</sup> whereas in TC PN the exchange attains  $1.5 e$  (compared to  $0.8 + 0.9 = 1.7 e$  here), leading to the conclusion that the copolymerization has little effect on the partial atomic charges. Note that the C and N charges are smaller in the TT chains, i.e., the exchange of electrons is less pronounced in TT than in TC. Using a Natural Population Analysis rather than the MK approach, Jaeger and co-workers got a similar charge pattern (positive P and C, negative N) for the  $-C(\text{Cl})=N-P(\text{Cl}_2)=N-P(\text{Cl}_2)=N-$  cycle, although an even stronger charge exchange was predicted, maybe due to the presence of Cl atoms.<sup>24</sup>

**B. Dipole Moments and (Hyper)polarizabilities.** Table 3 gives the  $\mu_L$ ,  $\alpha_L$  and  $\beta_L$  of structures **I** and **II**, whereas Figures 2 and 3 depict the evolution with chain length of

**Table 3.** Longitudinal Dipole Moment, Static Polarizability and Static First Hyperpolarizability (au) of Structures **I** and **II**<sup>a</sup>

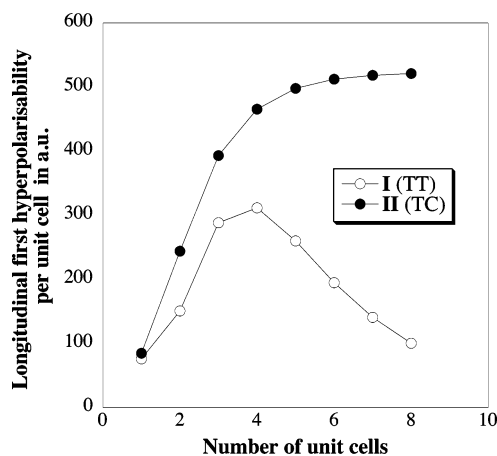
| <i>N</i>            | TT, <b>I</b> |            |           | TC, <b>II</b> |            |           |
|---------------------|--------------|------------|-----------|---------------|------------|-----------|
|                     | $\mu_L$      | $\alpha_L$ | $\beta_L$ | $\mu_L$       | $\alpha_L$ | $\beta_L$ |
| 1                   | 1.64         | 73.3       | 76        | 0.39          | 57.7       | 85        |
| 2                   | 4.37         | 182.4      | 227       | 2.09          | 142.2      | 329       |
| 3                   | 7.86         | 309.4      | 515       | 4.17          | 238.3      | 722       |
| 4                   | 11.80        | 445.1      | 827       | 6.53          | 340.4      | 1188      |
| 5                   | 16.02        | 585.0      | 1087      | 9.04          | 445.5      | 1686      |
| 6                   | 20.43        | 726.8      | 1282      | 11.63         | 552.2      | 2199      |
| 7                   | 24.94        | 869.7      | 1423      | 14.29         | 659.9      | 2717      |
| 8                   | 29.53        | 1013.0     | 1524      | 16.98         | 768.2      | 3238      |
| $\infty^b$          | 4.9          | 144        |           | 2.9           | 110        | 524       |
| $\Delta_{\infty}^b$ | 0.2          | 1          |           | 0.2           | 1          | 2         |

<sup>a</sup> All results have been obtained with the MP2/6-31G(d)//MP2/6-31G(d) approach. At the bottom of the table, the extrapolated polymeric values are given (see the text for more details on the procedure used to obtain these values). 1 au of  $\mu = 2.5418$  Debyes. 1 au of  $\alpha = 1.6488 \times 10^{-41} \text{ C}^2 \text{ m}^2 \text{ J}^{-1} = 0.14818 \text{ \AA}^3$ . 1 au of  $\beta = 3.2063 \times 10^{-53} \text{ C}^3 \text{ m}^3 \text{ J}^{-2} = 8.641 \times 10^{-33} \text{ esu}$ . <sup>b</sup>  $\infty$  gives the extrapolated values whereas  $\Delta_{\infty}$  is the estimated extrapolation error; i.e. polymeric values are given by  $\infty \pm \Delta_{\infty}$ .

**Figure 2.** Evolution with chain length of the MP2/6-31G(d)//MP2/6-31G(d) longitudinal polarizability per unit cell,  $\Delta\alpha_L(N)$ , of structures **I** and **II**.

the static  $\Delta\alpha_L$  and  $\Delta\beta_L$ . The  $\Delta\mu_L$  evolution with chain length of both systems presents a standard shape: it increases for small oligomers, then enters a saturation regime where it converges toward the polymeric value.  $\Delta\mu_L^{\text{at}(\infty)}$  of **I** and **II** are 3.1 and 1.9 D, respectively. The larger longitudinal dipole moment for TT structures was awaited from previous calculations on AB systems.<sup>5</sup> These values may be compared to their TT/TC PP counterpart 3.3/2.2 D<sup>40</sup> and to 0.9/0.5 D for TT/TC PN,<sup>30,41</sup> meaning that it is PP that eventually guides the total dipole moment. As a first approximation,  $\mu_L$  is mainly connected to the asymmetry rather than to the delocalization,<sup>42</sup> so we forecast the asymmetry of the copolymers to be closer to that of PP than PN.

As expected for increasingly long compounds,<sup>42–46</sup> the  $\Delta\alpha_L$  rapidly increases with chain length for short oligomers, then enters a saturation regime where it tends toward the asymptotic value characterizing the infinite polymer (Figure 2). The  $\Delta\alpha_L^{\text{at}(\infty)}$  is 36 au (28 au) for **I** (**II**). The PP(TT/TC)

**Figure 3.** Evolution with chain length of the MP2/6-31G(d)//MP2/6-31G(d) longitudinal first hyperpolarizability per unit cell,  $\Delta\beta_L(N)$ , of structures **I** and **II**.

has a MP2/6-31G(d)//HF/6-31G(d)  $\Delta\alpha_L^{\text{at}(\infty)}$  of 25/23 au and PN(TT/TC) presents a MP2/6-31G(d)//MP2/6-31G(d)  $\Delta\alpha_L^{\text{at}(\infty)}$  of 74/55 au. Therefore it is the less polarizable system (PP) that mainly guides the (quite small)  $\Delta\alpha$  amplitude of the copolymer. For comparison, PA<sup>47</sup> and polysilane<sup>48</sup> present  $\Delta\alpha_L^{\text{at}(\infty)}$  close to 65 au. As  $\alpha$  does not depend on the asymmetry but only on the delocalizability, this emphasizes that PP/PN is more delocalizable than PP, although the electron mobility is still limited compared to conjugated compounds. These results are consistent with the island delocalization model proposed for PP.<sup>49</sup> In this island model, delocalization of valence electrons takes place only for three atoms (PNP). By adding CN segments, one could consider that the electron mobility occurs over five atoms (PNCNP), which explains the limited increase of  $\alpha_L$  (+22% for **II**). On the other hand, in PN, the delocalization takes place over a much larger number of atoms, similarly to PA.

For PP/PN,  $\beta_L$  is always positive for all chain lengths, but the evolution with the chain length of  $\Delta\beta_L$  is very different for the two conformers. Indeed, for **I**,  $\Delta\beta_L$  first increases, reaches a maximum for  $N = 4$ , and then decreases toward a small polymeric limit. This is the typical shape for push–pull systems.<sup>8,15</sup> In **II**, the  $\Delta\beta_L$  versus  $N$  curve is similar to a “polarizability” curve: first an increase due to the delocalization increase, followed by the saturation toward the polymeric limit. As for the dipole moment and polarizability, it seems that PP guides the behavior of the copolymer. Indeed, in PN, both the TT and TC curves are “polarizability”-like, whereas in PP it is only the case for TC and not for TT (push–pull like curve).<sup>50</sup> To rationalize these shapes, one can split  $\beta$  into chain-ends and unit cells contributions.<sup>2–6</sup> Indeed, these components are the two parts responsible for the asymmetry. The contribution of each component to  $\beta$  is first increasing with chain length, due to the improvement of electron mobility. For extended oligomers, the chain-end contribution to  $\beta$  becomes constant, while the unit-cell contribution is proportional to the length of the oligomer. For **I**, the evolution of  $\Delta\beta_L$  (depicted in Figure 3) can be interpreted as a consequence of a negligible unit cell contribution and a significant chain-end component: (i) for short oligomers, the chain-end contribution (which is posi-

**Table 4.** Longitudinal Static First Hyperpolarizability (au) of Structures I and II<sup>a</sup>

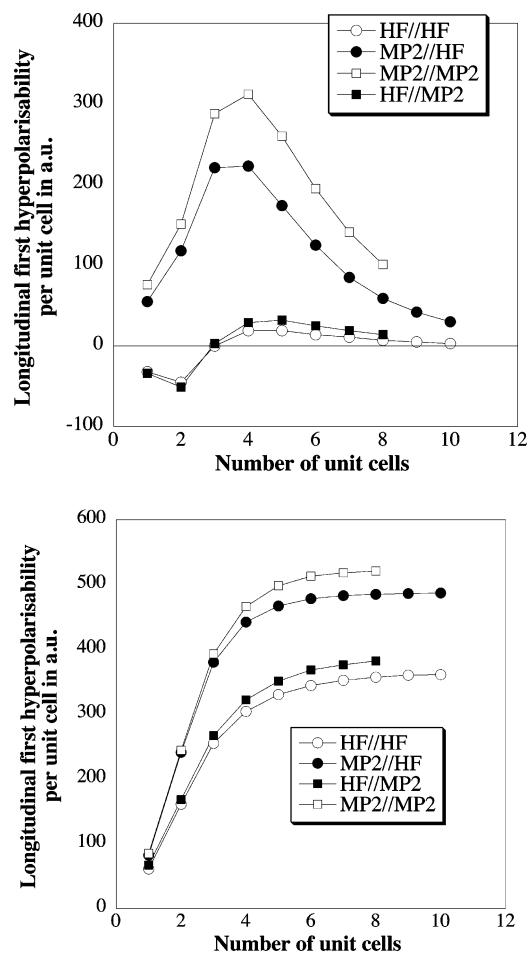
| N  | TT, I         |                |               |                | TC, II        |                |               |                |
|----|---------------|----------------|---------------|----------------|---------------|----------------|---------------|----------------|
|    | HF geometry   |                | MP2 geometry  |                | HF geometry   |                | MP2 geometry  |                |
|    | $\beta_L[HF]$ | $\beta_L[MP2]$ | $\beta_L[HF]$ | $\beta_L[MP2]$ | $\beta_L[HF]$ | $\beta_L[MP2]$ | $\beta_L[HF]$ | $\beta_L[MP2]$ |
| 1  | -32           | 55             | -34           | 76             | 61            | 82             | 67            | 85             |
| 2  | -77           | 173            | -85           | 227            | 222           | 323            | 235           | 329            |
| 3  | -77           | 394            | -82           | 515            | 477           | 703            | 502           | 722            |
| 4  | -58           | 617            | -53           | 827            | 781           | 1145           | 824           | 1188           |
| 5  | -39           | 791            | -21           | 1087           | 1111          | 1612           | 1175          | 1686           |
| 6  | -25           | 916            | 4             | 1282           | 1455          | 2090           | 1543          | 2199           |
| 7  | -14           | 1001           | 23            | 1423           | 1807          | 2573           | 1919          | 2717           |
| 8  | -7            | 1060           | 37            | 1524           | 2164          | 3058           | 2301          | 3238           |
| 9  | -2            | 1102           |               |                | 2524          | 3544           |               |                |
| 10 | 1             | 1132           |               |                | 2885          | 4031           |               |                |

<sup>a</sup> All results have been obtained with the 6-31G(d) basis set.

tive) dominates the total response, (ii) due to the increase of the electronic delocalization, the amplitude of chain-end contribution increases as the chain lengthens, and (iii) for long chains, the relative impact of the chain-end on  $\Delta\beta$  becomes more and more diluted,  $\beta$  becomes constant and  $\Delta\beta$  falls down to zero. For **II**, the reverse is true: a large unit-cell contribution combined to a small (and parallel) chain-end contribution. For long chains, each unit cell brings the same contribution to  $\beta$  and  $\Delta\beta_L$  becomes constant. For **II**, the  $\Delta\beta_L^{\text{at}(\infty)}$  attains 131 au. These results can be compared to the values obtained, using the same conformation for PP (85)<sup>2</sup> and PN (> 2500)<sup>50</sup> at the MP2/6-31G(d)//HF/6-31G(d) and MP2/6-31G(d)//MP2/6-31G(d) level of theory, respectively. Similarly to most other properties, the  $\Delta\beta_L^{\text{at}(\infty)}$  is mainly guided by PP although there is a significant improvement of the response (+50%). In PP, the  $\Delta\beta_L(\infty)$  are limited by delocalization rather than asymmetry,<sup>2</sup> meaning that if one increases the delocalization (decreases the asymmetry), the  $\Delta\beta_L(\infty)$  amplitude is improved. The values obtained for the copolymer are consistent with these findings: smaller  $\Delta\mu_L(\infty)$  and larger  $\Delta\alpha_L(\infty)$  correspond to larger  $\Delta\beta_L(\infty)$ . As a consequence, one could expect the TC CNPN chains to have a  $\Delta\beta_L^{\text{at}(\infty)}$  response between the PP and the **II** values.

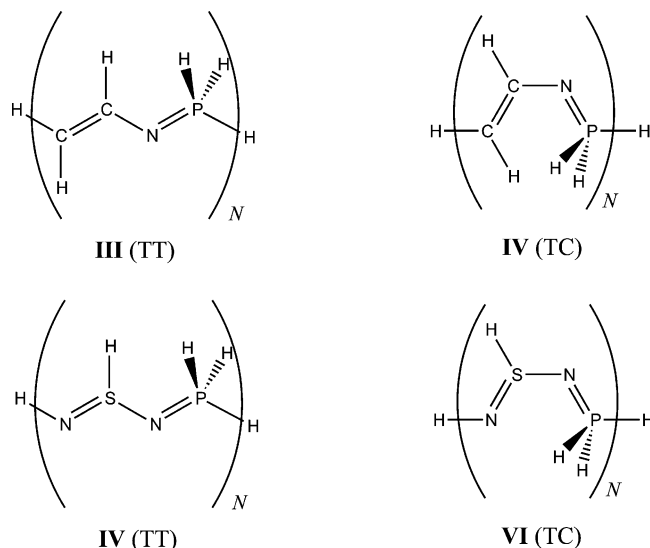
**C. Electron Correlation Effects on the First Hyperpolarizability.** We have studied the EC effects on the copolymer **I** and **II** because these effects strongly differ in PP and PN. Indeed, the direct and indirect impact of EC are very large in PN(TT). Especially, the direct EC contributions changes the sign of  $\beta_L$  and  $\Delta\beta_L$  for short chains, whereas it modifies the amplitude of  $\Delta\beta_L(\infty)$  by a factor of 6.4 (the indirect effect leads to an increase of  $\Delta\beta_L(\infty)$  by a factor of 1.5).<sup>30</sup> For PP(TC), the direct impact is much smaller (1.8), whereas the indirect shift is negligible.<sup>2</sup> For **I** and **II** the HF/MP2  $\beta_L$  are given in Table 4, whereas Figure 4a and b depicts the corresponding evolution with  $N$  of  $\Delta\beta_L(N)$ .

Quite surprisingly, the TT copolymers whose NLO responses are dominated by the PP part react more like PN for what concerns EC effects. Indeed, at the HF//HF and HF//MP2 levels,  $\Delta\beta_L(N)$  is first negative, reaches a minimum, goes through zero and slowly saturates toward a small asymptotic limit. This means that, at the HF level the chain-end contribution to  $\beta_L$  of **I** is, at least, overshoot and probably



**Figure 4.** Comparisons between the HF/6-31G(d) and MP2/6-31G(d) evolution with chain length of the longitudinal first hyperpolarizability per unit cell,  $\Delta\beta_L(N)$ , of structures **I** (Figure 4a, top) and **II** (Figure 4b, bottom).

pointing in the incorrect direction. The (HF) chain-end contribution is mainly dominated by the PN component of the copolymer. In addition, this means that the incorrect HF behavior in PN is probably due to the N–H terminal bond rather than to the CH<sub>2</sub> end group. The direct EC ratio does not evolve monotonically with  $N$ . The indirect EC ratio (MP2//MP2/MP2//HF) presents a smooth shape but is still changing relatively quickly wrt  $N$ , from 1.31 ( $N = 3$ ) to 1.44



**Figure 5.** Schematic representation of trans-transoid and trans-cisoid PP/PA and poly(thiophosphazene) alternating copolymers.

**Table 5.** Longitudinal Dipole Moment, Static Polarizability and Static First Hyperpolarizability (au) of Structures **III** and **IV**<sup>a</sup>

| <i>N</i>            | TT, <b>III</b> |            |           | TC, <b>IV</b> |            |           |
|---------------------|----------------|------------|-----------|---------------|------------|-----------|
|                     | $\mu_L$        | $\alpha_L$ | $\beta_L$ | $\mu_L$       | $\alpha_L$ | $\beta_L$ |
| 1                   | 1.24           | 81.1       | 68        | 0.81          | 65.2       | 120       |
| 2                   | 3.54           | 198.6      | 262       | 2.64          | 158.8      | 538       |
| 3                   | 6.38           | 337.0      | 941       | 4.86          | 266.7      | 1274      |
| 4                   | 9.51           | 486.2      | 2053      | 7.32          | 382.0      | 2224      |
| 5                   | 12.81          | 641.5      | 3431      | 9.90          | 501.5      | 3305      |
| 6                   | 16.21          | 800.5      | 4966      | 12.55         | 623.3      | 4465      |
| 7                   | 19.67          | 961.8      | 6599      | 15.25         | 746.6      | 5674      |
| 8                   | 23.17          | 1124.4     | 8289      | 17.98         | 870.7      | 6913      |
| $\infty^b$          | 3.6            | 168        | 1869      | 2.8           | 127        | 1440      |
| $\Delta_{\infty}^b$ | 0.1            | 2          | 166       | 0.1           | 2          | 178       |

<sup>a</sup> All results have been obtained with the MP2/6-31G(d)//MP2/6-31G(d) approach. At the bottom of the table, the extrapolated polymeric values are given. <sup>b</sup>  $\infty$  gives the extrapolated values whereas  $\Delta_{\infty}$  is the estimated extrapolation error; i.e. polymeric values are given by  $\infty \pm \Delta_{\infty}$ .

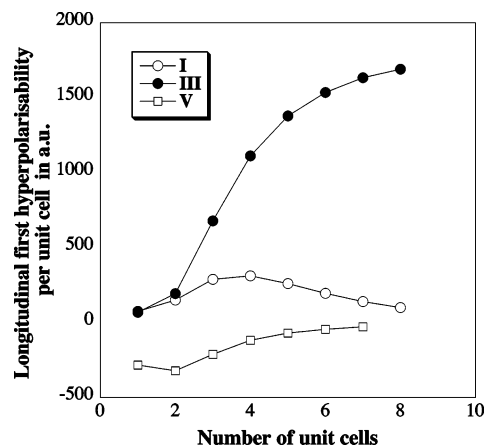
( $N = 8$ ). These ratios are slightly smaller than for PN chains of comparable size.<sup>30</sup>

For TC copolymers, one obtains the classical increase of NLO values when the EC corrections are taken into account, while the qualitative evolution with chain length is conserved. Consequently, the ratio between EC and HF  $\beta$  are much less sensitive to chain length than  $\beta$  itself. Provided  $N \geq 6$ , the direct EC ratio (MP2//HF//HF//HF) is decreasing with  $N$  (from for 1.47  $N = 3$  to 1.40 for  $N = 10$ ), whereas the indirect contribution is almost constant when the chain lengthens (from 1.05 for  $N = 3$  to 1.06 for  $N = 8$ ). Actually, if the ground-state geometry of **II** would have been optimized at the HF level, the error on the NLO properties would have been as limited as in PP. The direct EC ratio for the octamer of **II** is 1.41 and can be compared to the ratio obtained for the  $N = 16$  TC PP oligomers (1.63)<sup>2</sup> and TC PN chains (4.1).<sup>50</sup> The direct EC effects are thus smaller than in the “parent” polymers. The EC contributions seem almost

**Table 6.** Longitudinal Dipole Moment, Static Polarizability and Static First Hyperpolarizability (au) of Structures **V** and **VI**<sup>a</sup>

| <i>N</i>            | TT, <b>V</b> |            |           | TC, <b>VI</b> |            |           |
|---------------------|--------------|------------|-----------|---------------|------------|-----------|
|                     | $\mu_L$      | $\alpha_L$ | $\beta_L$ | $\mu_L$       | $\alpha_L$ | $\beta_L$ |
| 1                   | 2.38         | 76.8       | -284      | 0.48          | 58.5       | -56       |
| 2                   | 6.44         | 176.3      | -605      | 2.14          | 138.1      | -67       |
| 3                   | 11.09        | 282.6      | -816      | 4.22          | 226.9      | 71        |
| 4                   | 15.98        | 391.6      | -934      | 6.49          | 319.4      | 336       |
| 5                   | 20.98        | 501.8      | -1004     | 8.87          | 413.8      | 683       |
| 6                   | 26.03        | 612.7      | -1048     | 11.29         | 509.1      | 1076      |
| 7                   | 31.11        | 724.0      | -1076     | 13.75         | 604.9      | 1498      |
| $\infty^b$          | 5.1          | 112        |           | 2.6           | 97         | 557       |
| $\Delta_{\infty}^b$ | 0.1          | 1          |           | 0.1           | 1          | 86        |

<sup>a</sup> All results have been obtained with the MP2/6-31G(d)//MP2/6-31G(d) approach. At the bottom of the table, the extrapolated polymeric values are given. <sup>b</sup>  $\infty$  gives the extrapolated values whereas  $\Delta_{\infty}$  is the estimated extrapolation error; i.e. polymeric values are given by  $\infty \pm \Delta_{\infty}$ .

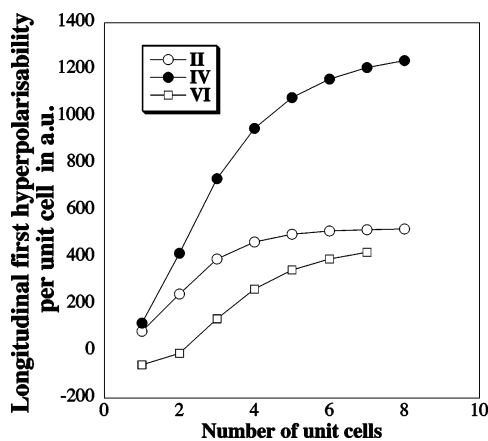


**Figure 6.** Evolution with chain length of the MP2/6-31G(d)//MP2/6-31G(d) longitudinal first hyperpolarizability per unit cell,  $\Delta\beta_L(N)$ , of structures **I**, **III** and **V**.

independent from each other. Indeed, for  $N = 8$ , we have (starting with the HF//HF  $\beta_L$ ) the following:  $2164 \times (3058/2164) \times (2301/2164) = 2164 \times 1.41 \times 1.06 = 3251$  au which is in perfect agreement with the MP2//MP2 amplitude of 3238 au (less than 1% of error).

**D. Comparison with Other Polyphosphazene-Based Copolymers.** In addition to PP/PN copolymers, we have also investigated two other polyphosphazene-based copolymers. They are displayed in Figure 5. These structures have not yet been synthesized, but polythiophosphazenes with SNP-NPN structures have been reported in the literature.<sup>20</sup> The MP2/6-31G(d)  $\mu_L$ ,  $\alpha_L$  and  $\beta_L$  of structures **III** and **IV** are given in Table 5. Table 6 provides the corresponding information for compounds **V** and **VI**. The evolution with chain length of  $\Delta\beta_L(N)$  is given in Figures 6 and 7, for TT and TC chains, respectively.

There are striking differences between the PP/PA and PP/PN chains. First, for the two conformers, the evolution with chain length of the  $\Delta\beta_L(N)$  curves are similar: first a fast increase and then a saturation toward the (nonzero) infinite chain limit(s). These limits are higher than in PP/PN. A possible explanation could be that the replacement of CH=



**Figure 7.** Evolution with chain length of the MP2/6-31G(d)//MP2/6-31G(d) longitudinal first hyperpolarizability per unit cell,  $\Delta\beta_L(N)$ , of structures **II**, **IV** and **VI**.

N segments by CH=CH leads to less asymmetric and more delocalizable copolymers. Consequently,  $\Delta\mu_L(\infty)$  decreases while  $\Delta\alpha_L(\infty)$  increases. As  $\beta$  is limited by delocalization, one can explain the larger  $\Delta\beta_L(\infty)$  for PP/PA than for PP/PN. Nevertheless, the increase (1440 au versus 524 au for TC chains) is larger than what could be expected from the small changes in  $\Delta\mu_L(\infty)$  and the 15% improvement of  $\Delta\alpha_L(\infty)$ . Our hypothesis is that in PP/PN, there is some “destructive” interactions between the two (PP and PN) components that are not observed in PP/PA because PA is perfectly symmetric. The fact that the TT PP/PA polymer presents a nonzero  $\Delta\beta_L(\infty)$  is more difficult to rationalize because both PA and PP(TT) chains actually show a zero response! At first, one could simply state that the zero  $\Delta r$  noted in long PP(TT) chains is broken in the PP/PA macromolecule (different bond lengths), explaining the nonzero  $\Delta\beta_L(\infty)$ . If this appears a valid explanation, it seems therefore difficult to explain the zero  $\Delta\beta_L(\infty)$  in the TT PP/PN copolymer. Of course, there is the nonprobable possibility that the TT PP/PA chains present a shape like TT PP/PN, with a maximum occurring for much longer chains lengths ( $N \geq 8$ ).<sup>51</sup>

The SNPN oligomers directly parallel the CNPN structures, with a “dromedary-back-type” curve for the TT structure and a “polarizability-like” evolution for the TC structure. However, the response of TT chains is always negative indicating a negative chain-end contribution. The TC SNPN structure, **VI**, seems less asymmetric ( $\Delta\mu_L(\infty)$  reaches 2.6 au instead of 2.9 au) and less polarizable than **II** ( $\Delta\alpha_L(\infty)$  of 97 au versus 110 au). With one favorable factor (the decrease of the asymmetry) against one unfavorable factor (the decrease of the delocalization), the amplitude of  $\Delta\beta_L(\infty)$  stays mainly constant (557 au versus 524 au).

#### IV. Conclusions, Comparisons and Outlook

We have investigated the copolymerization effects upon the geometry, charges, dipole moments, polarizability and first hyperpolarizability of polycarbophosphazene and some of its derivatives. It turned out that the alternating PP/PN copolymer presents dipole moment, polarizability and first

hyperpolarizability that are closer from PP than from PN, indicating that the least delocalizable structure mainly determines the NLO properties of the whole system, although the conformation of the chain has a crucial impact. The EC effects are extremely large for TT but remain limited for TC. However, the TC PP/PN copolymers show a significant improvement over the PP (first hyperpolarizability +50%). These results are consistent with the island- $\pi$  bonding model in PP: adding CH=N groups increases the size the conjugated areas but does not break the island pattern. If one uses PA instead of PN in the structure, the improvement is much larger (+300%). In addition the PP/PA copolymers present a nonzero NLO response for both selected conformations. On the other hand, polythiophosphazene does not constitute an improvement over poly(carbophosphazene) for NLO.

The static  $|\beta_L|/W$  (first hyperpolarizability per unit of weight) of the TC copolymers here investigated can easily be estimated from the corresponding  $|\Delta\beta_L(\infty)|$ :  $0.06 \times 10^{-30} \text{ cm}^5 \text{ esu}^{-1} \text{ g}^{-1} \text{ mol}$  for **II**,  $0.17 \times 10^{-30} \text{ cm}^5 \text{ esu}^{-1} \text{ g}^{-1} \text{ mol}$  for **IV** and  $0.05 \times 10^{-30} \text{ cm}^5 \text{ esu}^{-1} \text{ g}^{-1} \text{ mol}$  for **VI**. One might compare these values to  $0.03 \times 10^{-30}$  for TC PP<sup>2</sup> and  $4.2 \times 10^{-30}$  for TT PN<sup>16</sup> but also to the  $0.02 \times 10^{-30}$  value reported for polyphosphinoborane (PPB),<sup>6</sup>  $0.78 \times 10^{-30}$  reported for polysilaacetylene (PSA),<sup>5</sup>  $0.10 \times 10^{-30}$  for 3-methyl-4-nitroaniline (MNA) monomer,<sup>52</sup>  $0.06 \times 10^{-30}$  for N-(4-nitrophenyl)-(L)-prolinol (NPP),<sup>53</sup> and 0.66 for  $\alpha,\omega$ -nitro,amino-transhexatriene.<sup>54</sup> At this point, we can conclude that the copolymers could be interesting for NLO applications if noncentrosymmetric crystal are obtained.

Our research is now focused in determining the origin of the nonzero response for TT PP/PA chains (and nonzero in TT PP/PN), a surprising result as both TT polymers display a zero  $\Delta\beta_L(\infty)$ . The study of orbital interactions could help in assessing the origin of such differences. In addition, we also investigate the first hyperpolarizability of promising fully inorganic copolymer.

**Acknowledgment.** D.J. thanks the Belgian National Fund for Scientific Research (FNRS) for his Research Associate position. D.J. is indebted to Prof. J. M. André and Dr. E. A. Perpète of the CTA lab (FUNDP, Namur) for their help. Most calculations have been performed at the Inter-university Scientific Computing Facility (ISCF), installed at the Facultés Universitaires Notre-Dame de la Paix (Namur, Belgium), for which the author gratefully acknowledges the financial support of the FNRS-FRFC and the “Loterie Nationale” for the convention number 2.4578.02 and of the FUNDP.

#### References

- (1) Kanis, D. R.; Ratner, M. A.; Marks, T. J. *Chem. Rev.* **1994**, *94*, 195–242.
- (2) Jacquemin, D.; Quinet, O.; Champagne, B.; André, J. M. *J. Chem. Phys.* **2004**, *120*, 9401–9409.
- (3) Jacquemin, D.; Perpète, E. A.; Champagne, B.; André, J. M.; Kirtman, B. Recent Research Developments in Physical Chemistry. In *Transworld Research Network*, Trivandrum, India, 2002; Vol. 6.
- (4) Jacquemin, D.; Perpète, E. A.; Champagne, B. *Phys. Chem. Chem. Phys.* **2002**, *4*, 432–440.

- (5) Jacquemin, D.; Perpète, E. A.; André, J. M. *J. Chem. Phys.* **2004**, *120*, 10317–10327.
- (6) Jacquemin, D. *J. Phys. Chem. A* **2004**, *108*, 500–506.
- (7) Jacquemin, D. *J. Phys. Chem. A* **2004**, *108*, 9260–9266.
- (8) Morley, J. O.; Docherty, V. J.; Pugh, D. *J. Chem. Soc., Perkin Trans. 2* **1987**, 1351–1355.
- (9) Allcock, H. R. *Chemistry and Applications of Polyphosphazenes*; Wiley: New York, 2002.
- (10) Allcock, H. R.; Ravikiran, R.; Olshavsky, M. A. *Macromolecules* **1998**, *31*, 5206–5214.
- (11) Chandra Jha, P.; Krishnan, A.; Das, P. K.; Ramasesha, S. *J. Chem. Phys.* **2002**, *117*, 2873–2881.
- (12) Wöhrle, D. *Tetrahedron Lett.* **1971**, *22*, 1969.
- (13) Wöhrle, D. *Makromol. Chem.* **1974**, *175*, 1751–1760.
- (14) Komatsu, T. *Inorg. Chem.* **1998**, *8*, 2475–2478.
- (15) Jacquemin, D.; Champagne, B.; Perpète, E. A.; Luis, J.; Kirtman, B. *J. Phys. Chem. A* **2001**, *105*, 9748–9755.
- (16) Jacquemin, D.; Champagne, B.; André, J. M. *Chem. Phys. Lett.* **1998**, *284*, 24–30.
- (17) Manners, I.; Allcock, H. R.; Renner, G.; Nuyken, O. *J. Am. Chem. Soc.* **1989**, *111*, 5478–5480.
- (18) Allcock, H. R.; Coley, S. M.; Manners, I.; Nuyken, O.; Renner, G. *Macromolecules* **1991**, *24*, 2024–2028.
- (19) Allcock, H. R.; Coley, S. M.; Morrissey, C. T. *Macromolecules* **1994**, *27*, 2904–2911.
- (20) Allcock, H. R.; Dodge, J. A.; Manners, I. *Macromolecules* **1993**, *26*, 11–16.
- (21) Liang, M.; Manners, I. *J. Am. Chem. Soc.* **1991**, *113*, 4044–4045.
- (22) Manners, I. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1602–1621.
- (23) Allcock, H. R.; Coley, S. M.; Manners, I.; Visscher, K. B.; Parvez, M.; Nuyken, O.; Renner, G. *Inorg. Chem.* **1993**, *32*, 5088–5094.
- (24) Jaeger, R.; Debowski, M.; Manners, I.; Vancso, G. J. *Inorg. Chem.* **1999**, *38*, 1153–1159.
- (25) Chandrasekhar, V.; Chivers, T.; Kumaravel, S.; Meetsma, A.; van de Grampel, J. C. *Inorg. Chem.* **1991**, *30*, 3402–3407.
- (26) Rivard, E.; Lough, A. J.; Chivers, T.; Manners, I. *Inorg. Chem.* **2004**, *43*, 802–811.
- (27) Jaeger, R.; Lagowski, J. B.; Manners, I.; Vancso, G. J. *Macromolecules* **1995**, *28*, 539–546.
- (28) Jaeger, R.; Vancso, G. J.; Gates, N. Y.; Manners, I. *Macromolecules* **1997**, *30*, 6869–6872.
- (29) Jacquemin, D.; Champagne, B.; André, J. M. *J. Mol. Struct. (THEOCHEM)* **1998**, *425*, 69–79.
- (30) Jacquemin, D.; André, J. M.; Perpète, E. A. *J. Chem. Phys.* **2004**, *121*, 4389–4396.
- (31) Frisch, M. J. et al. “Gaussian 03, Revision B.04”, Gaussian, Inc., Wallingford, CT, 2004.
- (32) For short chains, using optimized, i.e., bent, TT chains leads to small differences for the NLO properties. For instance, for  $N=2$  of **I** (see Table 3), the polarizability is only modified by 1 au.
- (33) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431–439.
- (34) Champagne, B.; Kirtman, B. *Chem. Phys.* **1999**, *245*, 213–226.
- (35) Kirtman, B.; Champagne, B.; Bishop, D. M. *J. Am. Chem. Soc.* **2000**, *122*, 8007–8012.
- (36) Torrent-Sucarrat, M.; Solà, M.; Duran, M.; Luis, J. M.; Kirtman, B. *J. Chem. Phys.* **2003**, *118*, 711–718.
- (37) Champagne, B.; Jacquemin, D.; André, J. M.; Kirtman, B. *J. Phys. Chem. A* **1997**, *101*, 3158–3165.
- (38) Sun, H. *J. Am. Chem. Soc.* **1997**, *119*, 3611–3618.
- (39) Perpète, E. A.; Champagne, B.; Jacquemin, D. *J. Mol. Struct. (THEOCHEM)* **2000**, *529*, 65–71.
- (40) From MP2/6-31G(d)//HF/6-31G(d) calculations performed on the long PP chains.
- (41) From MP2/6-31G(d)//MP2/6-31G(d) calculations performed on the long PN chains.
- (42) Jacquemin, D.; Champagne, B.; Kirtman, B. *J. Chem. Phys.* **1997**, *107*, 5076–5087.
- (43) Hurst, G. J. B.; Dupuis, M.; Clementi, E. *J. Chem. Phys.* **1988**, *89*, 385–395.
- (44) Kirtman, B. *Chem. Phys. Lett.* **1988**, *143*, 81–83.
- (45) Champagne, B.; Mosley, D. H.; André, J. M. *J. Chem. Phys.* **1994**, *100*, 2034–2043.
- (46) Toto, J. L.; Toto, T. T.; de Melo, C. P.; Kirtman, B.; Robins, K. A. *J. Chem. Phys.* **1996**, *104*, 8586–8592.
- (47) Toto, T. T.; Toto, J. L.; de Melo, C. P.; Hasan, M.; Kirtman, B. *Chem. Phys. Lett.* **1995**, *244*, 59–64.
- (48) Kirtman, B.; Hasan, M. *J. Chem. Phys.* **1992**, *96*, 470.
- (49) Allcock, H. R. *Chem. Rev.* **1972**, *72*, 315–356.
- (50) Perpète, E. A.; personal communication.
- (51) To check this possibility, we have performed MP2/6-31G(d)//HF/6-31G(d) calculations up to  $N = 10$ , but no indication of a maximum could be found. Actually it seems that  $\Delta\beta_L(N)$  is converging to a nonzero constant close to 1600 au: 1469 au ( $N = 7$ ), 1520 au ( $N = 8$ ), 1551 au ( $N = 9$ ) and 1571 au ( $N = 10$ ).
- (52) Castet, F.; Champagne, B. *J. Phys. Chem. A* **2001**, *105*, 1366–1370.
- (53) Champagne, B.; Perpète, E. A.; Legrand, T.; Jacquemin, D.; André, J. M. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 1547–1553.
- (54) Jacquemin, D.; Champagne, B.; Hättig, C. *Chem. Phys. Lett.* **2000**, *319*, 327–334.

CT049884T

## Molecular Dynamics Simulations of HIV-1 Protease Suggest Different Mechanisms Contributing to Drug Resistance

Florian Wartha, Anselm H. C. Horn, Heike Meiselbach, and Heinrich Sticht\*

*Abteilung Bioinformatik, Institut für Biochemie, Emil-Fischer-Zentrum,  
Friedrich-Alexander-Universität Erlangen-Nürnberg, Fahrstrasse 17,  
D-91054 Erlangen, Germany*

Received November 26, 2004

**Abstract:** A major problem in the antiretroviral treatment of HIV-infections with protease-inhibitors is the emergence of resistance, resulting from the occurrence of distinct mutations within the protease molecule. In the present work molecular dynamics simulations of an active-site mutation (D30N) and a nonactive-site mutation (N88S) of HIV-1 protease that both directly confer resistance to the protease inhibitor Nelfinavir but not to Amprenavir were performed and compared to wild-type HIV-protease. A decreased interaction energy between protease and Nelfinavir was observed for the D30N mutant giving a plausible explanation for resistance, while the N88S mutation did not significantly affect the interaction energies in the bound form. Structural analysis including both ligand-bound and unliganded HIV-1 proteases revealed that the free N88S mutant protease shows significant differences in its hydrogen bonding pattern compared to free or Nelfinavir-bound wild-type protease. In particular, Asp30 forms more frequently a hydrogen bond with Ser88 in the unbound N88S mutant thus interfering with the Asp30-Nelfinavir interaction. These findings suggest that different molecular mechanisms contribute to resistance in active-site and nonactive-site mutants and propose a mechanism for the N88S mutant that is based on a shift of the conformational equilibrium of the unbound protease.

### Introduction

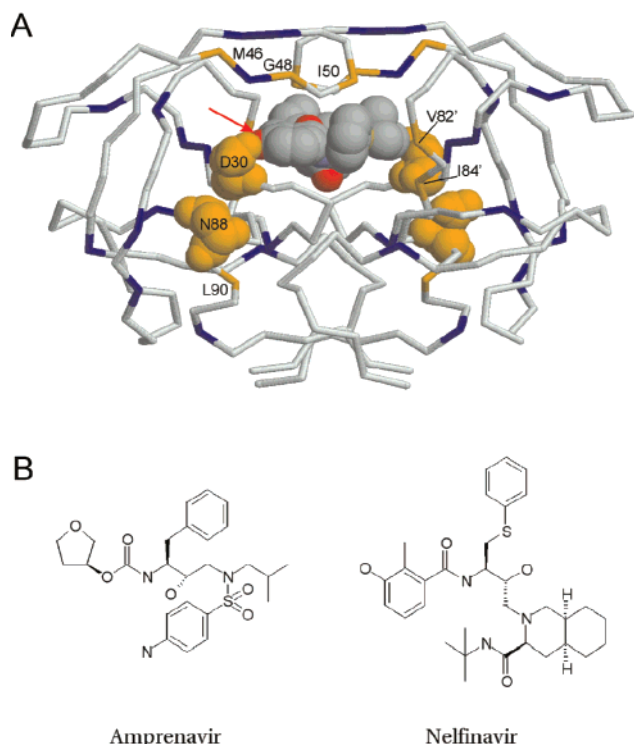
The global spread of the human immunodeficiency virus (HIV) causing the acquired immune deficiency syndrome (AIDS) has evolved into an immense health problem with total estimated infection numbers ranging from 34 to 46 million people.<sup>1</sup> The HIV-1 protease is essential for replication and assembly of the virus, and the inactivation of the HIV-1 protease leads to the production of noninfectious viral particles.<sup>2</sup> The idea of inhibiting viral replication by disturbing the protease function has led to the development of a class of drugs known as protease inhibitors (PI).<sup>3</sup> Modern HIV combination therapies, referred to as “Highly Active Anti-Retroviral Therapy” (HAART), attack the virus with a combination of one protease inhibitor and two reverse

transcriptase (RT) inhibitors.<sup>4</sup> The extensive use of antiretrovirals, however, has led to the emergence of resistant virus variants that possess various degrees of cross-resistance due to mutations.<sup>5</sup>

Mutations can either occur at active-site or nonactive-site locations in HIV-1 protease (Figure 1A) and can also confer different levels of resistance. Primary mutations directly confer resistance to one or more protease inhibitors, whereas secondary mutations only contribute to resistance and often occur together with primary ones or in synergistic form with other secondary mutations. Active-site mutations are exclusively primary ones, but not all primary mutations are necessarily limited to the active-site (e.g. the nonactive site mutations at sequence positions 46, 88 and 90 can also directly confer resistance<sup>6</sup>).

The mechanism of active-site mutations can frequently be rationalized in structural terms in which resistance is directly

\* Corresponding author phone: +49–9131–8524614; fax: +49–9131–8522485; e-mail: h.sticht@biochem.uni-erlangen.de.



**Figure 1.** A) Structure of the dimeric HIV-1 protease with bound Nelfinavir showing the most common primary and secondary HIV-1 protease drug-resistance causing mutations. Only the backbone-atoms of a wild-type HIV protease are depicted, whereas the protease inhibitor Nelfinavir is represented in space-filled representation with standard cpk coloring. Primary mutations (30, 46, 48, 50, 82, 84, 88 and 90) are drawn in orange, secondary mutations (10, 20, 24, 32, 33, 36, 47, 53, 54, 63, 71, 73, 77 and 93) in blue. Primary mutations are labeled in one of the subunits, and the two mutations studied here (D30 and N88) are shown in a space-filled representation. A red arrow indicates the hydrogen bond between the Asp30 side chain of one subunit and the phenyl oxygen of Nelfinavir which is discussed in the text. B) HIV-1 protease inhibitors used in this study. The chemical formulas are depicted in Lewis representation. Amprenavir susceptibility is maintained after D30N and N88S mutation, whereas both mutations confer resistance to Nelfinavir.

associated with a change of the contacts and thus the interaction energy between drug and target.<sup>7</sup> In contrast, the mechanism of nonactive-site mutations, influencing binding from distal locations, is not satisfyingly understood. These mutations are able to affect the enzymatic activity via mechanisms that have for example been related to differences in the conformational flexibility<sup>8</sup> or to the alteration of the binding site geometry through the accumulation of mutations within the core of the protease, suggesting that nonactive-site mutations can even represent the primary source of resistance.<sup>9</sup>

Resistance can either be assessed from genotypic assays based on the detection of mutations associated with the resistance or from phenotypic assays by measurement of the virus sensibility to increasing concentration of antiretroviral substances.<sup>10</sup> The correlation between genotypic mutations and phenotypic resistance, however, is not fully under-

stood,<sup>11,12</sup> and therefore genotypic testing needs to rely on comprehensive statistical data of known drug-resistance mutations.<sup>13</sup> In contrast, phenotypic assays directly give information about resistance, but they are experimentally demanding and difficult to standardize.<sup>14</sup> One disadvantage of both genotypic and phenotypic assays is that they do not provide a molecular mechanistical explanation for the emergence of resistance which would be extremely useful for developing more effective and longer lasting treatment strategies. In this context, computational studies of the dynamics and energetics of HIV-1 protease and protease-inhibitor complexes can give valuable insight into the molecular mechanism of mutations in conferring drug resistance.<sup>8,15,16</sup>

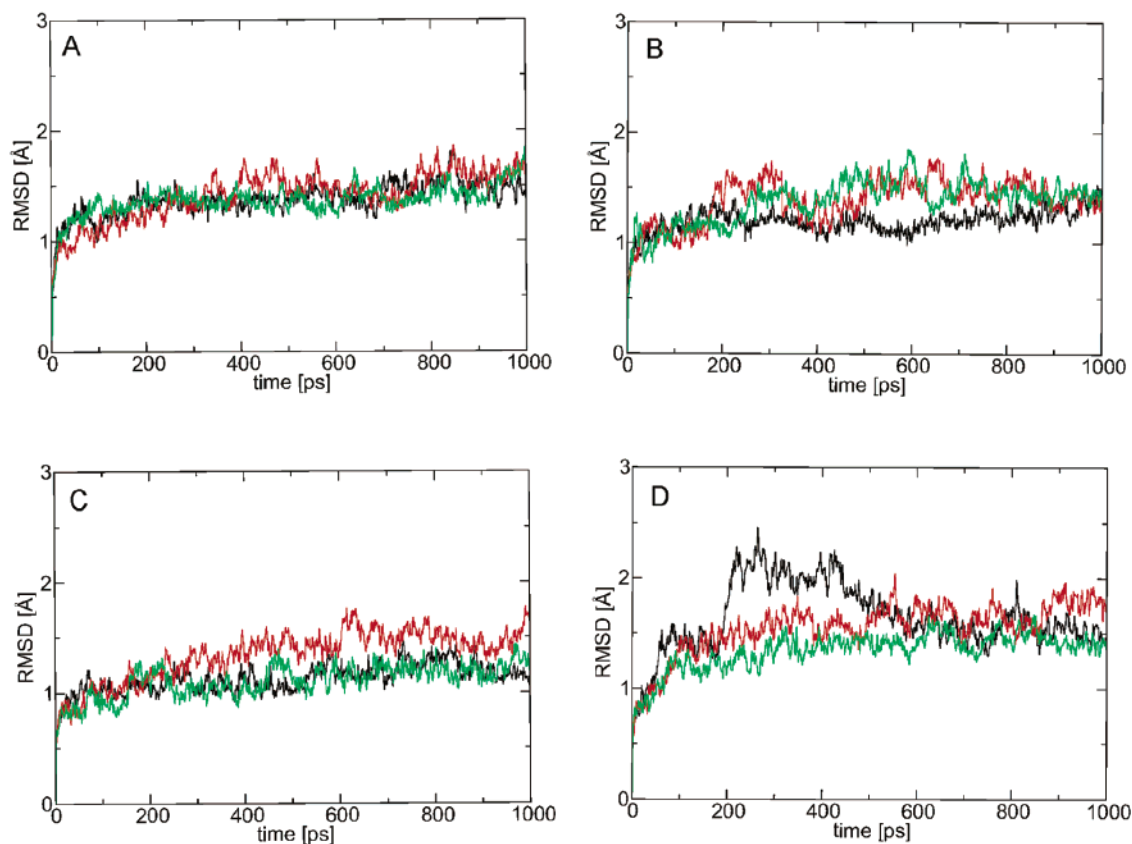
In the present study, two primary mutations (D30N, N88S) which often occur during treatment of HIV-infected patients<sup>17–19</sup> were analyzed using molecular dynamics simulations and subsequent structural and energetic analysis. D30N constitutes an active-site mutation, that reduces Nelfinavir (NLF) susceptibility by 5–20-fold<sup>6</sup> and is able to do so without any further major mutations.<sup>20</sup> Although N88S is a nonactive-site mutation, it has the remarkable ability to cause NLF-resistance even in the absence of active-site mutations.<sup>6</sup> Similar to D30N, the N88S mutation also causes resistance selectively for NLF (Figure 1B). Therefore studies of wild-type (WT) and mutant proteases in complex with other inhibitors (e.g. Amprenavir, APV) can serve as a control in both cases. To take into account that the mechanisms underlying resistance caused by the D30N and N88S mutation might differ substantially, the dynamics of the unliganded wild-type and mutant proteases were characterized here as well.

## Methods

**Preparation of Starting Structures.** The starting structures of the protease-drug and protease-substrate complexes were taken from following entries available in the Protein Data Bank (PDB<sup>21</sup>): 1HPV for protease complexed with APV,<sup>22</sup> 1OHR for protease + NLF<sup>23</sup>, and 4HVP for the protease + substrate-analogue.<sup>24</sup> Since no high-resolution crystal structure of an unliganded HIV-1 protease is available, the starting structures of the unbound forms were generated by removing an inhibitor from the APV-bound form, thus following the common strategy applied in previous studies.<sup>15,25</sup>

The protease sequence of the 1HPV and 1OHR-entries was taken as a wild-type sequence. The protease of the 4HVP file, containing one secondary mutation (L63P) and four polymorphisms (K14R, S37N, R41K, I64V), was modified by mutating the affected residues to the WT-sequence as described previously<sup>26,27</sup> using Sybyl 6.9.<sup>28</sup> N88S and D30N mutant structures were also generated using Sybyl. According to the strategy outlined by Piana et al.,<sup>8</sup> the structure of the substrate was generated from a crystal structure (4HVP) containing a substrate analogue (MVT-101) by replacing the uncleavable CH<sub>2</sub>NH linkage between Nle-Nle by a Met-Met peptide bond, leading to a substrate (Thr-Ile-Met-Met-Gln-Arg) that contains a natural cleavage site for the HIV-1 protease, the so-called p2/NC site.<sup>29</sup>





**Figure 2.** Root-mean-square deviation (RMSD) to the starting structure as a function of simulation time. Proteases liganded with A) substrate, B) APV, C) NLF, D) unliganded. Black: WT-protease; red: D30N; green: N88S. Calculations were performed for the backbone atoms of the respective structure versus the backbone atoms of the respective simulation's starting structure (excluding drug/substrate).

The simulations were performed under physiological pH conditions which required ensuring the correct protonation state of ionizable groups. In particular, one of the aspartates (D25) of the catalytic site exhibits an increased  $pK_a$  value of 5.2 in the inhibitor-bound protease,<sup>30</sup> while no increased  $pK_a$  was reported for the free form of the protease ( $pK_a = 4.5$ <sup>31</sup>). Therefore, a fully deprotonated active site should be prevalent at a physiological pH, and thus this protonation state was used throughout the simulations. However, to ensure that the results of the simulations do not critically depend on the protonation state, a 1 ns control simulation of a NLF-bound HIV-1 protease with a protonated carboxylate oxygen of Asp25 was also performed.

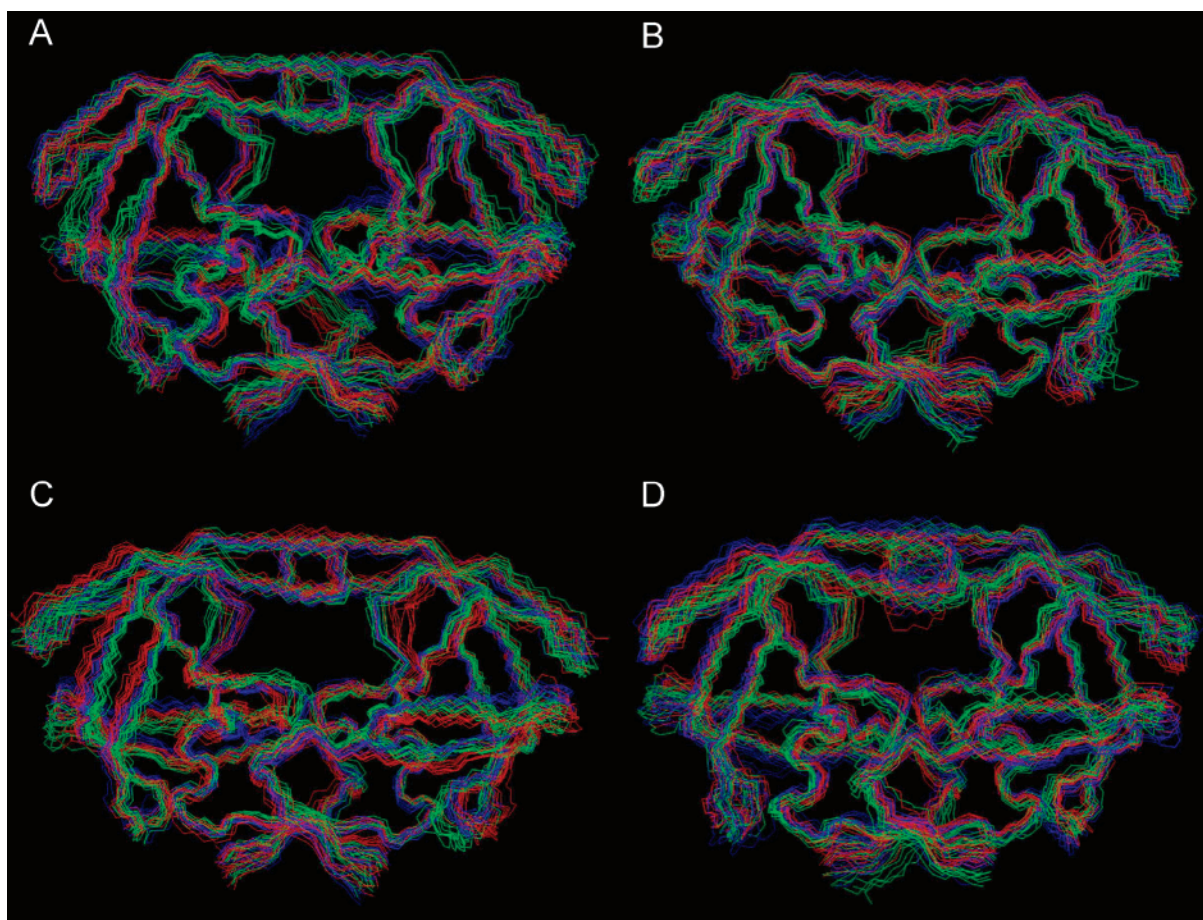
**Parameter Generation for Inhibitors.** The initial coordinates of APV and NLF were extracted from the 1OHR and 1HPV pdb files. ArgusLab<sup>32</sup> was utilized to add missing hydrogens, ensure correct protonation states and perform an initial geometry optimization using the Universal Force Field method.<sup>33</sup> The resulting structures were then further geometry optimized using the semiempirical molecular orbital Hamiltonian AM1<sup>34</sup> as implemented in Vamp.<sup>35</sup> The structures thus obtained were subjected to three consecutive geometry optimizations with Gaussian98<sup>36</sup> using the ab initio methods HF/MIDI!, B3LYP/6-31G(d) and HF/6-31G(d). For all four quantum mechanical geometry optimizations the stationary points found were ensured to be true minima by the calculation of the vibrational frequencies. For the two final structures of APV and NLF the atomic charges were then

obtained following the established procedure<sup>37,38</sup> by fitting the charges to the HF/6-31G(d) computed electrostatic potential using the “antechamber” tool from the AMBER program suite.

**Molecular Dynamics Simulations and Analysis.** All MD simulations presented in this work were performed by using the AMBER 7.0 suite of programs<sup>39</sup> with the Cornell et al. force field (ff99).<sup>40,41</sup> For the organic compounds APV and NLF the general AMBER force field (gaff)<sup>42</sup> was used. An appropriate number of  $Cl^-$  counterions was added to neutralize the system, and afterward the molecules were solvated in a box of water, using the TIP3P water model<sup>43</sup> with at least 10.0 Å of water around every atom of the solute.

All structures were minimized first by using the sander module in AMBER with an atom-based 10 Å cutoff on electrostatic interactions, using a constant dielectric for the electrostatic interactions,  $\epsilon = 1$ . Particle Mesh Ewald summation<sup>44</sup> was used to calculate the long-range electrostatic interactions during minimization and during molecular dynamics.

The minimization procedure was split into three different parts. First the solvent was allowed to relax while restraining the protein atoms to their original position with a force constant of 500 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Afterward additional relaxation of the protein side chains was allowed by restraining only the backbone atoms before in a final minimization all restraints were removed. The minimizations



**Figure 3.** Fitted backbone overlay of proteases. Eleven structures of WT and both mutant proteases were taken in 100 ps intervals from the trajectory of the proteases liganded with A) substrate, B) APV, C) NLF, D) unliganded. WT: blue; D30N: red and N88S: green. All structures are backbone-fitted on the first structure of the wild-type simulations. The ligands are not shown for clarity.

consisted of 250 steps of steepest descent followed by 250 or 7250 steps of conjugate gradient minimization, respectively.

MD simulations were carried out thereafter, using a dielectric constant for the electrostatic interactions of  $\epsilon = 1$  and a time-step of 1 fs. The temperature of the system was raised gradually from 50 to 298 K in 10 ps. Subsequently, 1 ns MD simulations with standard NPT conditions were performed for data collection, resulting in a total of 1000 snapshots. The SHAKE procedure<sup>45</sup> was used to constrain all bonds involving hydrogens. An 8.5 Å cutoff was used for the nonbonded interactions which were updated every 15 steps. For the visualization and structural analysis of the programs RasMol,<sup>46</sup> Sybyl 6.9,<sup>28</sup> IsisDraw,<sup>47</sup> AMBER,<sup>39</sup> and X-PLOR<sup>48</sup> were used.

## Results and Discussion

**Stability of the Trajectories and Properties of the Global Structure.** A qualitative examination of the trajectories obtained from 12 MD simulations performed in the presence of different mutants and inhibitors (Figure 2) shows that all systems deviated to a quite similar extent from their starting structures resulting in a backbone RMSD of approximately 1.2–1.8 Å after 1 ns. The magnitude of the fluctuations is in the same range as those observed for the same time scale

in recent simulations of HIV-1 protease.<sup>15,16,49</sup> Thus, one can conclude that the simulation runs produced stable trajectories which should provide a suitable basis for the subsequent analyses.

The structure and conformational variability of the ensembles obtained over the simulation time (Figure 3) reveals that both the structures and the magnitude of the fluctuations are quite similar for all 12 systems studied. The observation that neither mutations nor different ligands lead to larger structural changes is consistent with the findings of Zoete et al.<sup>50</sup> who showed that 73 crystal structures of HIV-1 protease do not differ significantly in their geometry, despite a large variety in the position and number of mutations and types of ligands bound. Since subsequent comparison focused on the analysis of protein–ligand interaction energies and on local structural changes induced by the different mutations, a simulation time of 1 ns was considered sufficient. As shown in previous simulations of HIV-1 protease, the respective data can even be obtained from simulations on a sub-ns time scale.<sup>50,51</sup>

**Protease-Ligand Interaction Energies.** The van der Waals interaction energy between protease and ligand provides a suitable parameter to obtain first information about the role of particular sequence positions for the emergence of drug resistance.<sup>51</sup> A decrease of the van der Waals

**Table 1:** Protein-Ligand van der Waals Interaction Energies and Corresponding Standard Deviations<sup>a</sup>

| ligand | protease | energy [kcal mol <sup>-1</sup> ] |
|--------|----------|----------------------------------|
| SUB    | WT       | -83.5 ± 3.9                      |
| SUB    | D30N     | -79.3 ± 5.0                      |
| SUB    | N88S     | -78.4 ± 4.0                      |
| APV    | WT       | -60.7 ± 3.7                      |
| APV    | D30N     | -61.0 ± 3.8                      |
| APV    | N88S     | -61.4 ± 4.0                      |
| NLF    | WT       | -68.2 ± 3.2                      |
| NLF    | D30N     | -58.9 ± 3.5                      |
| NLF    | N88S     | -66.9 ± 3.7                      |

<sup>a</sup> The energy was calculated using the AMBER parm99 force field and a 99 Å cutoff for all pairwise interactions. Water molecules and ions were not included. Mean energy values and standard deviations were based on a set of 1000 structures collected every ps over the entire simulation.

interaction energy in mutant proteases is indicative for a loss of the tight packing of the inhibitor in the active site resulting either from steric clashes or from the loss of favorable attractive interactions (e.g. hydrogen bonds, salt bridges). The overall values of these interaction energies were first confirmed to be not critically affected by the protonation state of the active site as shown by a control simulation for the Nelfinavir-bound wild-type protease, which yielded an interaction energy of  $-66.4 \pm 3.6$  kcal mol<sup>-1</sup> compared to  $-68.2 \pm 3.2$  kcal mol<sup>-1</sup> for the simulation using a fully deprotonated active site.

Assuming that a phenotypically observed decreased susceptibility is correlated with a decreased interaction energy, one would expect that the D30N and N88S mutations do not affect the interaction energies for the substrate- and Amprenavir-liganded proteases. Indeed, no significant decrease of the van der Waals interaction energy upon both mutations was observed for the substrate and Amprenavir (Table 1). For Nelfinavir, a significant decrease in binding-energy from  $-68.2 \pm 3.2$  kcal mol<sup>-1</sup> to  $-58.9 \pm 3.5$  kcal mol<sup>-1</sup> was observed for the D30N mutation, while the N88S mutation did not result in significant changes (Table 1).

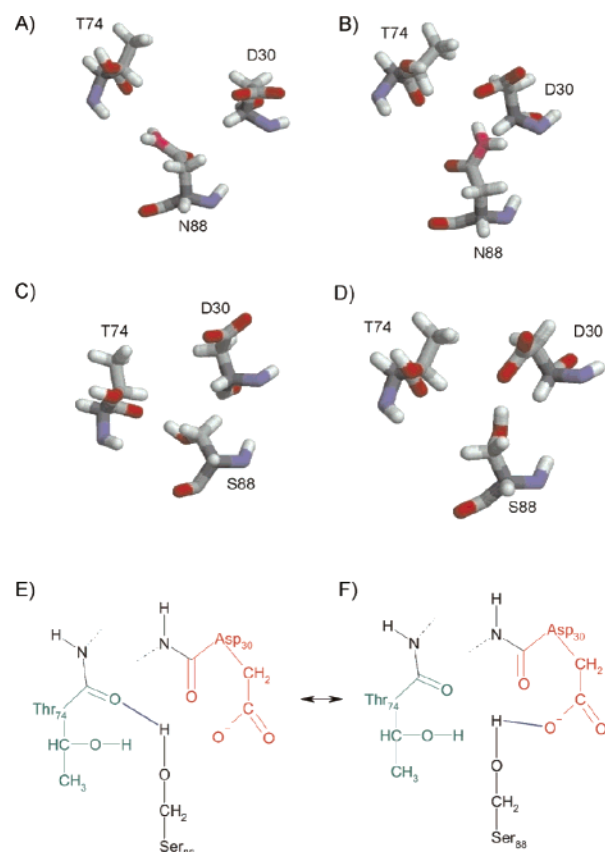
The effect of the D30N mutation can be rationalized by inspection of the crystal structure<sup>23</sup> in which the carboxyl group of the D30 side chain of one subunit is critically involved in a hydrogen bond with the hydroxyl group of the aromatic ring of NLF (Figure 1; red arrow). This mechanism for resistance is supported by previous data of Wang and Kollman,<sup>51</sup> who also reported a decreased interaction energy for other primary, active-site mutations.

In contrast, the nonactive-site mutation N88S investigated here did not lead to a significant decrease of van der Waals interaction energy for all ligands studied (Table 1) showing that the NLF resistance caused by this mutation cannot easily be explained by simple energetic considerations of the ligand-bound proteases. Similar observations have been made in a very recent study by Chen et al.<sup>16</sup> in which 14 HIV-1 protease mutants complexes with the inhibitor Indinavir were characterized by MD simulations and subsequent energetic analysis. In particular for the nonactive site mutants L24I, G73S, N88D, and L90M a prediction of resistance on the

**Table 2:** Hydrogen Bonds Observed at the Site of the Mutation<sup>a</sup>

| protease  | N/S88 – T74 (%) | N/S88 – D30 (%) |
|-----------|-----------------|-----------------|
| WT-NLF    | 94.3            | 0.1             |
| WT-free   | 66.1            | 19.0            |
| N88S-free | 25.7            | 28.7            |

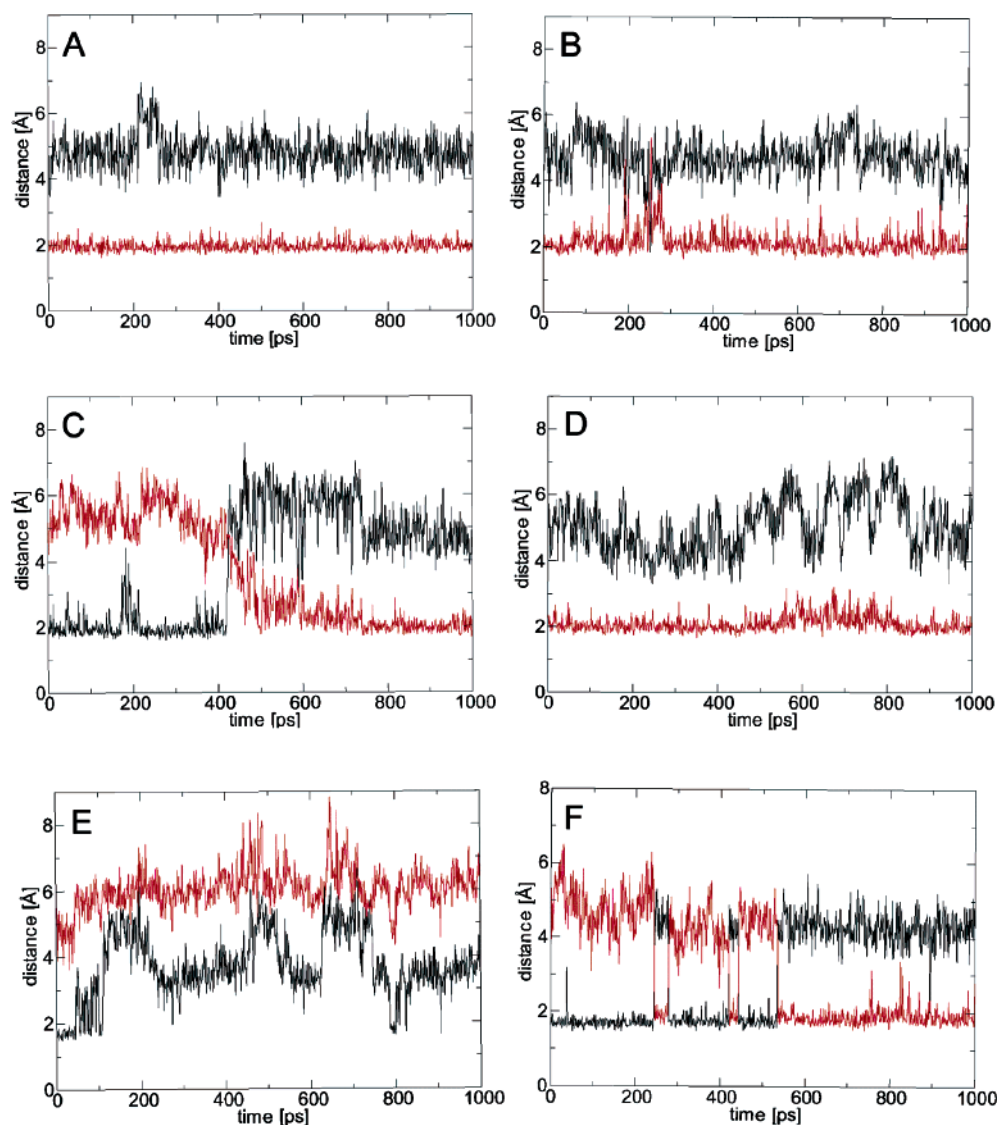
<sup>a</sup> Hydrogen bond donors are the side chain amide group of Asn88 and the side chain hydroxyl group of Ser88. Acceptor atoms are the backbone O of Thr74 and the side chain OD1 of Asp30 (Figure 4E, F). The presence of a hydrogen bond was inferred from a proton-acceptor distance of < 2.6 Å and a donor-proton-acceptor angle of > 120°. Values are given in percent and represent averages over both subunits and the 1 ns simulation time.



**Figure 4.** Alternative hydrogen bond pairs 88–30 and 88–74 in the unliganded WT and N88S mutant structures. Only the three amino acids involved are depicted in stick representation. A+B) unliganded wild-type protease, C+D) unliganded N88S mutant. The snapshots of the WT were taken at 700 and 150 ps, and for the N88S protease at  $t = 250$  and 500 ps representing situations in which either the 88–74 (A+C) or the 88–30 (B+D) hydrogen bond is present. E+F) Schematic representation of the hydrogen bonds observed during MD simulation in the unliganded N88S protease. Hydrogen bonds are drawn in blue.

basis of the dynamic properties and interaction energies proved to be difficult.<sup>16</sup>

**Comparison of Free and Ligand-Bound Structures of HIV Protease.** To find an explanation for the NLF resistance of the N88S mutant, structural analysis was extended to the unbound forms of HIV-1 protease. The rationale behind this strategy is the observation that conformational sampling of



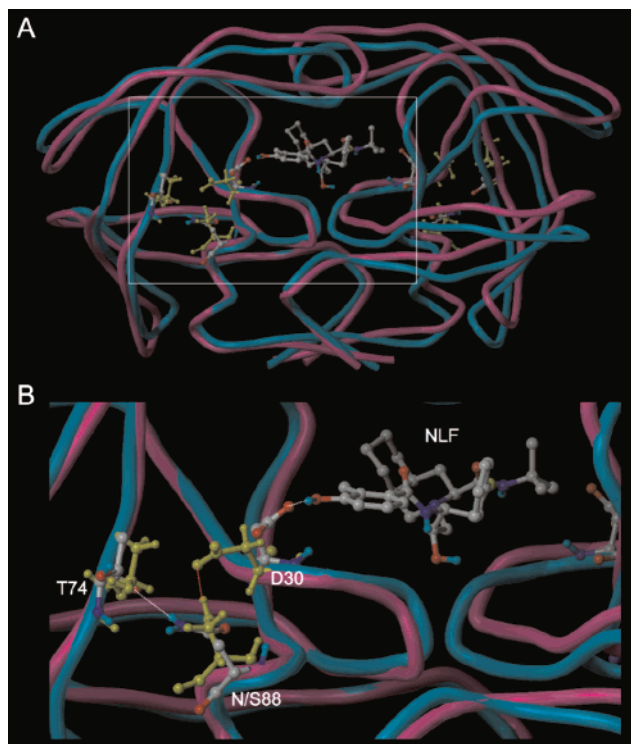
**Figure 5.** Distances between hydrogen bond donor and acceptor atoms of Asp30, Thr74 and Asn/Ser88 as a function of simulation time. Hydrogen bond donor atoms are side chain amide proton of Asn88 and side chain hydroxyl proton of Ser88. Acceptor atoms are backbone O of Thr 74 (red line) and side chain OD1 of Asp30 (black line). Subunit 1 (left panel), subunit 2 (right panel). A+B) WT protease liganded with NLF, C+D) unliganded WT protease, E+F) unliganded N88S protease. Analysis of the hydrogen bonding pattern was performed separately for both subunits in order to take into account the asymmetric structure of NLF. Distances of  $\sim 2$  Å indicate that the groups are sufficiently close for the formation of a hydrogen bond.

proteins is frequently critically affected by mutations<sup>52–54</sup> and thus may also influence ligand binding affinity due to the presence of different amounts of “binding-competent” conformations.<sup>55,56</sup> The present analysis therefore focused on the Nelfinavir-bound wild-type protease as well as the free wild-type and free N88S mutant. Based on the considerations outlined above, one would expect that binding of the ligand is facilitated if the “binding-competent” conformation is already predefined in the unbound form of the protease.

Analysis of the structure in the region of the mutation revealed that a N88-T74 hydrogen bond exists over 94.3% of the simulation time for the Nelfinavir-bound wild-type and may thus be considered as feature of a “binding-competent” conformation. In the free wild-type and N88S protease the portion of structures exhibiting this hydrogen bond is generally lower than in the Nelfinavir-bound form

but differs considerably between both free proteases (66.1% in the WT vs 25.7% in the mutant; Table 2). This difference in the population of this hydrogen bond by a factor of  $\sim 2.5$  between the free wild-type and free N88S protease indicates that this mutation can actually shift the conformational equilibrium of the free protease.

Analysis of those conformations that lack the 88–74 hydrogen bond revealed that the predominant alternative hydrogen bond of residue 88 is formed to the Asp30 side chain which plays an important role for NLF binding. This role is evidenced by the fact that a replacement of Asp by Asn in the active-site D30N mutant leads to NLF resistance.<sup>6</sup> Analysis of the geometry of the corresponding region in the wild-type and N88S mutant (Figure 4) suggests that the 88–74 and 88–30 hydrogen bonds should be mutually exclusive. This is confirmed by monitoring the donor–acceptor distance over the simulation time (Figure 5) showing that both



**Figure 6.** A) Overlay of the crystal structure of NLF-bound protease (cyan) with the structure of the free N88S mutation obtained after 300 ps in the MD simulation (magenta). Residues D30, N/S88, and T74, that are part of the hydrogen bonding network, and the inhibitor are shown in ball-and-stick representation. Residues of the NLF-bound protease are colored in cpk and those of the free N88S mutant in yellow. The part of the structure marked by the white square is shown as enlargement in B) and the alternative hydrogen bonds are indicated. White dotted lines indicate the N88-T74 and D30-NLF hydrogen bonds present in the NLF-bound wild-type, whereas a red dotted line indicates the alternative S88-D30 hydrogen bond present in the free mutant. For the latter interaction a significant rotation of the D30 side chain is observed, hampering an interaction with the inhibitor.

hydrogen bonds never exist at the same time, and a switching between both conformations is observed instead (Figure 5C, F).

Interestingly, the alternative 88–30 hydrogen bond is observed more frequently in the free N88S mutant (28.7%; Table 2) compared to the free wild-type (19.0%) offering an explanation for the resistance caused by the N88S mutation. In addition to this increase of  $\sim 50\%$  in the occurrence of this hydrogen bond, the N88S mutation also affects the  $\chi_1$  side chain rotamer distribution of Asp30. While in the simulation of the free wild-type only the  $-60^\circ$  rotamer is observed, an alternative rotamer ( $\chi_1 = \sim -150^\circ$ ) is detected in the simulation of the N88S mutant when the 88–30 hydrogen bond is formed. This deviation from this  $\chi_1 = -60^\circ$  rotamer, that is also present in the Nelfinavir-bound state, has the consequence that the D30 side chain is rotated away from the ligand binding pocket in the N88S mutant (Figure 6) thus no longer providing a suitable interaction partner for NLF. The fact that such alternative rotamers are not observed in the free wild-type even when a N88-D30

hydrogen bond is formed can most likely be attributed to differences in the length and geometry of the Asp88 and Ser88 side chain. Taken together, there are three structural properties of the free N88S mutant that clearly differ from the free wild-type protease thus offering an explanation for resistance: A decrease in the frequency of the 88–74 hydrogen bond characteristic for the Nelfinavir-bound form by more than 60%, an increase of the alternative 88–30 hydrogen bond by  $\sim 50\%$ , and the observation of an alternative rotamer for D30 which is critically involved in NLF binding.

For the following reasons it can be excluded that these structural differences are an artifact resulting from the simulation or the choice of the starting structures: First, both simulations of wild-type and mutant unbound protease started from the same crystal structure and thus differ only by the presence of the N88S mutation. Second, all starting structures of the unliganded proteases were generated by removal of APV and not of NLF from a ligand-bound protease. Therefore, no information about particular structural properties of the Nelfinavir-bound form, which could result in a bias of the corresponding simulations, was “transferred” to the unbound starting structures. Third, the switching between different hydrogen bonds shows that the simulation time was sufficiently long to allow a sampling of these different conformations rendering an analysis of their relative population valid.

Thus, a conformational equilibrium exists for the HIV-1 protease, prior to binding of drug or substrate, between two populations that differ in their hydrogen bond pattern and side chain rotamers of Asp30. Due to the N88S mutation, the equilibrium is shifted toward a conformation that hampers NLF binding. This interpretation also offers an explanation for the fact that the effects of the N88S mutation could not be detected by an energetic analysis of the ligand bound form and shows that the approach of predicting resistance from analysis of ligand-bound structures alone cannot generally be applied.

The influence of a conformational-equilibrium on binding kinetics has already been described in the literature for several systems such as thrombin,<sup>57</sup> cAMP-dependent protein kinase,<sup>58</sup> and the oestrogen receptor<sup>59</sup> as well as aldose reductase.<sup>55,60</sup> Another well-studied example is the binding of SH3 domains to putative peptide ligands.<sup>61</sup> For this system, manifold conformations exist prior to binding, which are significantly reduced upon binding and are thus coupled to a cooperative conformational change. Recently, Perryman et al.<sup>15</sup> also observed a conformational equilibrium in the unliganded HIV-1 protease that is affected by mutation. In the V82F/I84V mutant differences in the curling of the protease’s flap region compared to the wild-type were detected that resulted in a shift of the equilibrium between closed and semiopen conformation, thus offering a plausible explanation for the drug resistance mechanism of this mutant. Hence, this study is in accordance with the present findings suggesting that conformational equilibria might generally play a role for ligand binding and emergence of resistance for HIV-1 protease.

The mechanism suggested by the present study in which alternative hydrogen bonding patterns play an important role may also apply to other nonactive site mutations, for example for L90M which also directly confers resistance although it does not directly contact the ligand. Analysis of protease crystal structures containing the L90M mutation<sup>62,63</sup> reveals that the sulfur atom of the methionine is in close spatial proximity to the active-site residue D25 which is critically involved in ligand binding. Therefore, one could envisage a mode of action similar to the N88S mutant in which the sulfur of M90, that has no equivalent in the wild-type L90, could function as a novel hydrogen bond acceptor and thus interfere with the hydrogen bonding network present in wild-type protease.

More, and possibly also longer MD simulations, will be necessary for a comprehensive investigation of the role of unbound conformations in the emergence of resistance caused by nonactive-site mutations and thus will help to design novel and more effective drugs, e.g. by targeting different residues or by developing allosteric inhibitors that are capable of regulating protease dynamics.

**Acknowledgment.** We thank the Computer-Chemie-Centrum Erlangen for support with the Gaussian calculations.

### References

- (1) UNAIDS/WHO 2004 Report on the global AIDS epidemic; **2004**.
- (2) Kohl, N. E.; Emmi, E. A.; Schleif, W. A.; Davies, L. J.; Heimbach, J. C.; Dixon, R. A. F. Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci.* **1988**, *85*, 686–690.
- (3) Katoh, I.; Yasunaga, T.; Ikawa, Y.; Yoshinaka, Y. Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor. *Nature* **1987**, *329*, (6140), 654–6.
- (4) Stephenson, J. The art of ‘HAART’: researchers probe the potential and limits of aggressive HIV treatments. *Jama* **1997**, *277*, (8), 614–6.
- (5) Voelker, R. The world in medicine: HIV drug resistance. *Jama* **2000**, *284*, (2), 169.
- (6) Shafer, R. W. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clin. Microbiol. Rev.* **2002**, *15*, (2), 247–77.
- (7) Draghici, S.; Potter, R. B. Predicting HIV drug resistance with neural networks. *Bioinformatics* **2003**, *19*, (1), 98–107.
- (8) Piana, S.; Carloni, P.; Röthlisberger, U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. *Protein Sci.* **2002**, *11*, (10), 2393–402.
- (9) Muzammil, S.; Ross, P.; Freire, E. A major role for a set of nonactive site mutations in the development of HIV-1 protease drug resistance. *Biochemistry* **2003**, *42*, (3), 631–8.
- (10) Walter, H.; Schmidt, B.; Korn, K.; Vandamme, A. M.; Harrer, T.; Uberla, K. Rapid, phenotypic HIV-1 drug sensitivity assay for protease and reverse transcriptase inhibitors. *J. Clin. Virol.* **1999**, *13*, (1–2), 71–80.
- (11) Martinez-Picado, J.; Savara, A. V.; Shi, L.; Sutton, L.; D’Aquila, R. T. Fitness of human immunodeficiency virus type 1 protease inhibitor-selected single mutants. *Virology* **2000**, *275*, (2), 318–22.
- (12) Schmidt, B.; Walter, H.; Moschik, B.; Paatz, C.; van Vaerenbergh, K.; Vandamme, A. M.; Schmitt, M.; Harrer, T.; Uberla, K.; Korn, K. Simple algorithm derived from a geno-/phenotypic database to predict HIV-1 protease inhibitor resistance. *Aids* **2000**, *14*, (12), 1731–8.
- (13) Schinazi, R. F.; Larder, B.; Mellors, J. W. Mutations in retroviral genes associated with drug resistance: 2000–2001 update. *Int. Antivir. News* **2000**, *8*, 65–92.
- (14) Blaise, P.; Clevenbergh, P.; Vaira, D.; Moutschen, M.; Dellamonica, P. HIV resistance to antiretroviral drugs: mechanisms, genotypic and phenotypic resistance testing in clinical practice. *Acta Clin. Belg.* **2002**, *57*, (4), 191–201.
- (15) Perryman, A. L.; Lin, J. H.; McCammon, J. A. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **2004**, *13*, (4), 1108–23.
- (16) Chen, X.; Weber, I. T.; Harrison, R. W. Molecular dynamics simulations of 14 HIV protease mutants in complexes with indinavir. *J. Mol. Model. (Online)* **2004**, *10*, (5–6), 373–81.
- (17) Markowitz, M.; Conant, M.; Hurley, A.; Schluger, R.; Duran, M.; Peterkin, J.; Chapman, S.; Patick, A.; Hendricks, A.; Yuen, G. J.; Hoskins, W.; Clendeninn, N.; Ho, D. D. A preliminary evaluation of nelfinavir mesylate, an inhibitor of human immunodeficiency virus (HIV)-1 protease, to treat HIV infection. *J. Infect. Dis.* **1998**, *177*, (6), 1533–40.
- (18) Petropoulos, C. J.; Parkin, N. T.; Limoli, K. L.; Lie, Y. S.; Wrin, T.; Huang, W.; Tian, H.; Smith, D.; Winslow, G. A.; Capon, D. J.; Whitcomb, J. M. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **2000**, *44*, (4), 920–8.
- (19) Ziermann, R.; Limoli, K.; Das, K.; Arnold, E.; Petropoulos, C. J.; Parkin, N. T. A mutation in human immunodeficiency virus type 1 protease, N88S, that causes in vitro hypersensitivity to amprenavir. *J. Virol.* **2000**, *74*, (9), 4414–9.
- (20) Patick, A. K.; Duran, M.; Cao, Y.; Shugarts, D.; Keller, M. R.; Mazabel, E.; Knowles, M.; Chapman, S.; Kuritzkes, D. R.; Markowitz, M. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob. Agents Chemother.* **1998**, *42*, (10), 2637–44.
- (21) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, (1), 235–42.
- (22) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal Structure of HIV-1 Protease in Complex with VX-478, a potent and orally bioavailable Inhibitor of the Enzyme. *J. Am. Chem. Soc.* **1995**, *117*, 1181–1182.
- (23) Kaldor, S. W.; Kalish, V. J.; Davies, J. F., II; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **1997**, *40*, (24), 3979–85.

- (24) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B.; Wlodawer, A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* **1989**, *246*, (4934), 1149–52.
- (25) Kurt, N.; Scott, W. R.; Schiffer, C. A.; Haliloglu, T. Cooperative fluctuations of unliganded and substrate-bound HIV-1 protease: a structure-based analysis on a variety of conformations from crystallography and molecular dynamics simulations. *Proteins* **2003**, *51*, (3), 409–22.
- (26) Dolezal, O.; De Gori, R.; Walter, M.; Doughty, L.; Hattarki, M.; Hudson, P. J.; Korrt, A. A. Single-chain Fv multimers of the anti-neuraminidase antibody NC10: the residue at position 15 in the V(L) domain of the scFv-0 (V(L)-V(H)) molecule is primarily responsible for formation of a tetramer-trimer equilibrium. *Protein Eng.* **2003**, *16*, (1), 47–56.
- (27) Chen, Y. Z.; Gu, X. L.; Cao, Z. W. Can an optimization/scoring procedure in ligand-protein docking be employed to probe drug-resistant mutations in proteins? *J. Mol. Graph. Model.* **2001**, *19*, 450–570.
- (28) Tripos Sybyl 6.9, Release 7.0A., St. Louis, Missouri, USA, 1991–2002.
- (29) Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K. J.; Starcich, B.; Josephs, S. F.; Doran, E. R.; Rafalski, J. A.; Whitehorn, E. A.; Baumeister, K.; et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **1985**, *313*, (6000), 277–84.
- (30) Tie, Y.; Boross, P. I.; Wang, Y. F.; Gaddis, L.; Hussain, A. K.; Leshchenko, S.; Ghosh, A. K.; Louis, J. M.; Harrison, R. W.; Weber, I. T. High-resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J. Mol. Biol.* **2004**, *338*, (2), 341–52.
- (31) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Biol.* **1996**, *3*, (11), 946–50.
- (32) Thomson, M. A. *Arguslab 3.1*, Planaria Software LLC: Seattle, WA, 2004.
- (33) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A. I.; Skiff, W. M. UFF, a Rule-Based Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (34) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (35) Clark, T.; Alex, A.; Beck, B.; Burkhardt, F.; Chandrasekhar, J.; Gedeck, P.; Horn, A. H. C.; Hutter, M.; Martin, B.; Rauhut, G.; Sauer, W.; Schindler, T.; Steinke, T. *Vamp8.1 B33*, Erlangen, 2000–2004.
- (36) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revision A.11.3*, Gaussian Inc.: Pittsburgh, PA, 1998.
- (37) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (38) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.
- (39) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal-mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (40) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (41) Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, (4), 845–62.
- (42) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, (9), 1157–74.
- (43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (44) Darden, T. A.; York, D. M.; Pedersen, L. G. Particle mesh Ewald. An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (45) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (46) Sayle, R. A.; Milner-White, E. J. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **1995**, *20*, (9), 374.
- (47) MDL *ISIS/Draw 2.3.*, San Leandro, CA, 1990–2000.
- (48) Brünger, A. T. *XPLOR*, version 3.1. A system for X-ray crystallography and NMR. *Yale University Press: New Haven, CT* **1992**.
- (49) Lepsik, M.; Kriz, Z.; Havlas, Z. Efficiency of a second-generation HIV-1 protease inhibitor studied by molecular dynamics and absolute binding free energy calculations. *Proteins* **2004**, *57*, (2), 279–93.
- (50) Zoete, V.; Michielin, O.; Karplus, M. Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. *J. Mol. Biol.* **2002**, *315*, (1), 21–52.
- (51) Wang, W.; Kollman, P. A. Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, (26), 14937–42.

- (52) Tsai, C. J.; Ma, B.; Nussinov, R. Folding and binding cascades: shifts in energy landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, (18), 9970–2.
- (53) Luque, I.; Freire, E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. *Proteins* **2000**, *Suppl 4*, 63–71.
- (54) Sinha, N.; Nussinov, R. Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, (6), 3139–44.
- (55) Davis, A.; Teague, S. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 736–749.
- (56) Rauh, D.; Klebe, G.; Stubbs, M. T. Understanding protein–ligand interactions: the price of protein flexibility. *J. Mol. Biol.* **2004**, *335*, (5), 1325–41.
- (57) Engh, R. A.; Brandstetter, H.; Sucher, G.; Eichinger, A.; Baumann, U.; Bode, W.; Huber, R.; Poll, T.; Rudolph, R.; von der Saal, W. Enzyme flexibility, solvent and ‘weak’ interactions characterize thrombin–ligand interactions: implications for drug design. *Structure* **1996**, *4*, (11), 1353–62.
- (58) Prade, L.; Engh, R. A.; Girod, A.; Kinzel, V.; Huber, R.; Bossemeyer, D. Staurosporine-induced conformational changes of cAMP-dependent protein kinase catalytic subunit explain inhibitory potential. *Structure* **1997**, *5*, (12), 1627–37.
- (59) Brzozowski, A. M.; Pike, A. C.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engstrom, O.; Ohman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M., Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, *389*, (6652), 753–8.
- (60) Urzhumtsev, A.; Tete-Favier, F.; Mitschler, A.; Barbanton, J.; Barth, P.; Urzhumtseva, L.; Biellmann, J. F.; Podjarny, A.; Moras, D. A ‘specificity’ pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil. *Structure* **1997**, *5*, (5), 601–12.
- (61) Ferreon, J. C.; Hilser, V. J. Ligand-induced changes in dynamics in the RT loop of the C-terminal SH3 domain of Sem-5 indicate cooperative conformational coupling. *Protein Sci.* **2003**, *12*, (5), 982–96.
- (62) Clemente, J. C.; Moose, R. E.; Hemrajani, R.; Whitford, L. R.; Govindasamy, L.; Reutzel, R.; McKenna, R.; Agbandje-McKenna, M.; Goodenow, M. M.; Dunn, B. M. Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease. *Biochemistry* **2004**, *43*, (38), 12141–51.
- (63) Mahalingam, B.; Wang, Y. F.; Boross, P. I.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004**, *271*, (8), 1516–24.

CT0498690



## Electrostatic Polymer Condensation and the A/B Polymorphism in DNA: Sequence Effects

Alexy K. Mazur\*

*CNRS UPR9080, Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, Paris 75005, France*

Received October 1, 2004

**Abstract:** Dynamics of the polymorphic A $\leftrightarrow$ B transitions in DNA is compared for two polypurine sequences, poly(dA).poly(dT) and poly(dG).poly(dC), long known to exhibit contrasting properties in experiments. In free molecular dynamics simulations reversible transitions are induced by changing the size of a water drop around DNA neutralized by sodium ions. In poly(dG).poly(dC) the B $\leftrightarrow$ A transitions are easy, smooth and perfectly reversible. In contrast, a B $\rightarrow$ A transition in a poly(dA).poly(dT) dodecamer fragment could not be obtained even though its A-form is stable under low hydration. An intermediate range of hydration numbers is identified where opposite transitions are observed, namely, A $\rightarrow$ B in poly(dA).poly(dT) and B $\rightarrow$ A in poly(dG).poly(dC). The two sequences exhibit qualitatively different counterion distributions, with a characteristic accumulation of sodium in the major groove of poly(dG).poly(dC) and the B $\rightarrow$ A transition driven by the electrostatic condensation mechanism. The resistance of the poly(dA).poly(dT) sequence to adopting the A-form is traced to the specific steric interactions of thymine methyl groups in the major groove. With these methyls replaced by hydrogens, reversible B $\leftrightarrow$ A transitions become possible and the difference between the two molecules is significantly reduced. The good overall agreement with experimental data corroborates the general role of the electrostatic condensation mechanism in the A/B polymorphism in DNA.

### Introduction

The double helical DNA may adopt a number of different structural forms.<sup>1</sup> Forms A and B are the most interesting because they are observed in vivo. The B-form is the dominant biological conformation, whereas the A-form is sometimes found in protein-DNA complexes and is considered as a high energy state adopted temporarily.<sup>2,3</sup> Reversible B $\leftrightarrow$ A transitions probably represent one of the modes for governing protein-DNA interactions. The B-helix is long and narrow, with its core formed by stacked base pairs. In contrast, the A-helix is much shorter, and its core represents a 6 Å solvent accessible hole, with strongly inclined base pairs wrapping around it. Despite these very different shapes, both forms are right-handed helical duplexes with identical topologies and hydrogen bonding; transitions between them

do not require base pair opening or destacking and involve relatively small conformational barriers.<sup>4</sup>

The B $\leftrightarrow$ A transitions can be also induced in vitro by changing the DNA environment.<sup>5–8</sup> In condensed preparations, that is, in crystalline and amorphous fibers as well as in films, DNA always adopts the B-form under high relative humidity, but it can be reversibly driven to the A-form by placing the samples under relative humidity below 80%.<sup>5,7,8</sup> In aqueous solutions, single DNA molecules exhibit reversible B $\leftrightarrow$ A transitions when certain organic solvents are added.<sup>2,6</sup> In both cases, the transition occurs at about the same water activity suggesting that the B $\leftrightarrow$ A conformational switch is driven by the hydration state of the double helix.<sup>9</sup> The most known molecular mechanism proposed by Saenger et al.<sup>10</sup> postulates that, in A-DNA, water molecules form bridges between consecutive phosphate groups, which is not possible for the B-form where the interphosphate distances are too long. Therefore, under low water activity, hydration

\* Corresponding author e-mail: alexey@ibpc.fr.

forces impose a more “economical” A-form.<sup>10</sup> This elegant mechanism contradicts, however, some long-known observations,<sup>11</sup> and it does not explain why, in both experiment and simulations, the B form generally remains stable in high salt where water activity is also reduced.<sup>12–14</sup> Some alternative mechanisms assume that the B→A equilibrium is governed by sequence dependent base pair stacking<sup>15</sup> as well as the hydrophobic effect.<sup>16,17</sup> Interbase interactions should be somehow involved because B↔A transitions are strongly sequence dependent. Notably, poly(dA).poly(dT) never goes to the A-form in condensed state<sup>7</sup> and is strongly A-phobic in solution.<sup>18</sup> In contrast, the poly(dG).poly(dC) duplex reportedly exhibits A-like features even under normal environment conditions.<sup>19–21</sup> Better understanding of the driving forces of the in vitro B↔A transition is essential for interpretation of its biological manifestations.

A number of theoretical approaches were applied in this field during its long history.<sup>12,15,17,22,23</sup> In the recent years, significant progress has been achieved by using all atom molecular dynamics (MD) simulations.<sup>24–31</sup> Spontaneous A→B transitions were shown to occur after less than a nanosecond of conventional simulations with periodical boundaries.<sup>24,30</sup> However, simulations of A-DNA under similar conditions encountered serious difficulties. In a homogeneous media composed of experimental concentrations of water, ions, and EtOH the A-DNA duplex was found to be unstable.<sup>26,30</sup> Its dynamics could be observed during a few nanoseconds under biphasic solvation, with only water and ions in the inner shell. Transitions from B- to A-DNA were obtained only with multivalent macroions initially placed at specific positions around DNA.<sup>27</sup> The energetic balance between the two forms in solution is very delicate and force field dependent.<sup>25,26,32</sup> An important recurrent feature revealed in these simulations consisted of the accumulation of free metal cations in the major groove of A-DNA suggesting that the electrostatic interactions across the major groove are essential for stabilization of the A-form<sup>26,27,30,33</sup> and that they can cause the B→A transition at least in some cases.

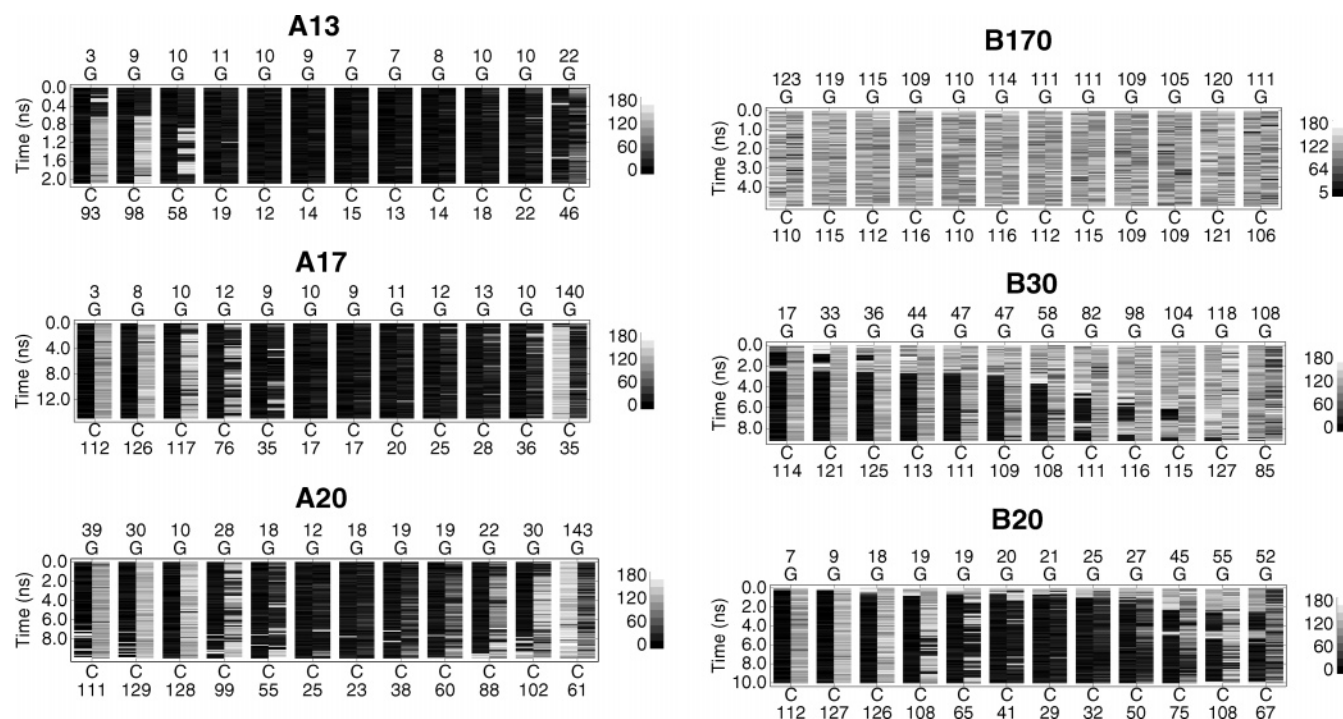
Stable dynamics of A-DNA and reversible B↔A transitions can be studied in MD simulations with vacuum boundaries.<sup>31</sup> In this approach, a DNA fragment is placed in a water drop with the desired number of free ions.<sup>34</sup> The drop size is maintained during simulations by periodical artificial return of the evaporated water molecules. The relative stability of A and B forms is modulated by changing the size of the drop, with spontaneous transitions observed in both directions and reversed, if necessary. The estimated critical hydration in the transition point was reasonably close to experiment without any additional fitting of the force field,<sup>31</sup> suggesting that the physics of the transitions is reasonably well reproduced even though water drop conditions do not correspond to any earlier experiments. It was shown that the major driving force of the B→A transition comes from accumulation of free metal cations in the opening of the major groove that results in inversion of electrostatic interactions between the phosphates of opposite backbone strands. This transition, therefore, is similar to the well-known phenomenon of free-ion-mediated electrostatic con-

denation of polyelectrolytes, but here it occurs inside the DNA duplex. With certain assumptions, this mechanism can be applied to experimental B↔A transitions, notably, it offers a detailed interpretation of the long known cooperative effects.<sup>31</sup> However, the relevance of MD simulation results to experimental B↔A transitions remains an open issue, and some authors consider that in reality this process should be too slow to be modeled in a computer.<sup>35</sup>

To get further insight in the A/B polymorphism in DNA it is necessary to rationalize the mechanisms of its sequence effects. A number of such effects were earlier discovered and studied in detail experimentally.<sup>2,7,18,36–38</sup> The relative A/B propensities of different base pair steps appear such that they tend to compensate one another, and generic DNA sequences commonly are neither A- nor B-philic.<sup>18</sup> There are, however, a few sequence motives characterized by very different A/B propensities. Among them stretches of consecutive guanines and adenines (G-tracts and A-tracts, respectively) represent the most known example. Disclosing the physical origin of these differences is an important challenge because these sequences also play special roles in vivo. They are overrepresented in eukariotic genomes as well as prokaryotes and archaeobacteria<sup>39,40</sup> and have a long record of involvement in various biological processes. Contrasting physical properties of A- and G-tract DNA duplexes are well characterized.<sup>7,20,36,41–54</sup> A systematic review of these data exists in the literature.<sup>55</sup> In G-tracts, the B→A transition is very easy. The poly(dG).poly(dC) is generally prone to adopt the A-DNA conformation in conditions where random sequences stay firmly in the B-form.<sup>7,45,48</sup> In contrast, in A-tracts, the B→A transition is difficult,<sup>48,54</sup> and it was never observed for poly(dA).poly(dT) in condensed phase.<sup>7,36</sup>

The mechanism by which the base pair sequence can affect A↔B transitions is not clear. Early calculations<sup>22</sup> demonstrated that stacking in different base pair steps can partially account for their A- or B-philicity. A-tracts may be specifically stabilized in the B-form by specific minor groove hydration patterns<sup>56</sup> as well as cross-strand hydrogen bonding.<sup>57</sup> There are also evidences suggesting that the exceptional resistance of poly(dA).poly(dT) to the B→A transition is related to the thymine methyl groups that form a continuous hydrophobic cluster in the major DNA groove and provide several other stereochemical factors that may specifically stabilize B- versus A-form.<sup>18,58</sup>

Here we study the aforementioned sequence effects in B↔A transitions by using all atom MD simulations with free vacuum boundaries. DNA fragments with A- and G-tract motives are found to be distinctly different as regards their relative preferences toward B- and A-forms in good qualitative agreement with experiment. In poly(dG).poly(dC) dodecamer, the B→A transition is easy, smooth and perfectly reversible. In contrast, only A→B transitions could be observed in a similar poly(dA).poly(dT) fragment, whereas attempts of the B→A transition always lead to denaturation. An intermediate range of hydration numbers is found where the B-form poly(dA).poly(dT) dodecamer is stable and the A→B transition is observed, while the poly(dG).poly(dC) duplex is stable in A-form and makes the opposite transition. The denaturation of poly(dA).poly(dT) in intermediate



**Figure 1.** A $\leftrightarrow$ B transitions in C12 shown by the dynamics of sugar pucker pseudorotation.<sup>92</sup> The boundary pucker values (ca. 0° and 180°) are assigned the black and white colors, respectively, with intermediate values mapped linearly to the gray scale levels. Therefore, the C3'-endo conformation of the A-form is seen as a nearly black field, whereas typical B-form C2'-endo and C1'-exo conformations correspond to nearly white and light gray, respectively. Each base pair step is characterized by a column consisting of two subcolumns, with the left subcolumns referring to the sequence written above in 5'-3' direction from left to right and the time averaged phases given on top. The right subcolumns refer to the complementary sequence shown below together with the corresponding time averaged phases. The starting A- and B-DNA states and the corresponding hydration numbers ( $\Gamma$ ) are included in the labels.

conditions can be prevented by flanking A-tracts with GC pairs. In this case the B $\rightarrow$ A transition becomes possible in long duplexes, but dodecamer fragments stay in B-form down to very low hydration and then collapse without denaturation, which makes this process reversible and allows testing the role of different factors, notably, the thymine methyls. It appears that with methyls replaced by hydrogens, i.e., thymine residues substituted by uracils, the B $\rightarrow$ A transitions become possible, and the difference in the relative A/B philicity between A-tracts and G-tracts is drastically reduced.

## Methods

A series of MD simulations was carried out for double helical dodecamer fragments poly(dA).poly(dT) and poly(dG).poly(dC) referred to as T12 and C12, respectively, as well as a few derivatives of these sequences. The sizes of the water drop were chosen to provide a hydration number ( $\Gamma$ ) of 13, 17, 20, 25, 30, 40, 60, and 80 water molecules per nucleotide. Additional simulations for  $\Gamma \approx 170$  were carried out by using periodical boundary conditions. The simulations either started from standard A- and B-DNA states prepared as explained below or continued from a final state of a previous trajectory with water molecules added or removed to reach a desired hydration level. Similar results are obtained with both strategies. In this way a quasi-static pattern of B $\leftrightarrow$ A transitions can be reproduced that mimics in vitro titration experiments.

The simulation protocols were similar to the earlier water drop simulations.<sup>34,31</sup> We use the internal coordinate molecular dynamics (ICMD) method<sup>59,60</sup> adapted for DNA<sup>61,62</sup> with the time step of 0.01 ps. In this approach, the DNA molecule has all bond lengths and almost all bond angles fixed at their standard values. The only variable bond angles are those centered at the sugar C1',...,C4', and O4' atoms, which ensures the flexibility of the furanose rings. In contrast, bases, thymine methyls, and phosphate groups move as articulated rigid bodies, with only rotations around single bonds allowed. The highest frequencies in thus obtained models are additionally balanced by increasing rotational inertia of the lightest rigid bodies as described earlier.<sup>61,63</sup> The possible physical effects of the above modifications have been discussed elsewhere.<sup>60,64</sup> The electrostatic interactions are treated with the SPME method<sup>65</sup> adapted for infinite vacuum boundary conditions.<sup>34</sup> The common values of Ewald parameters were used, that is 9 Å truncation for the real space sum and  $\beta \approx 0.35$ . Reference simulations with periodical boundaries were carried out as described before,<sup>34</sup> with the standard SPME method in NVT ensemble conditions with a rectangular unit cell of 45 × 45 × 65 Å under normal water density.

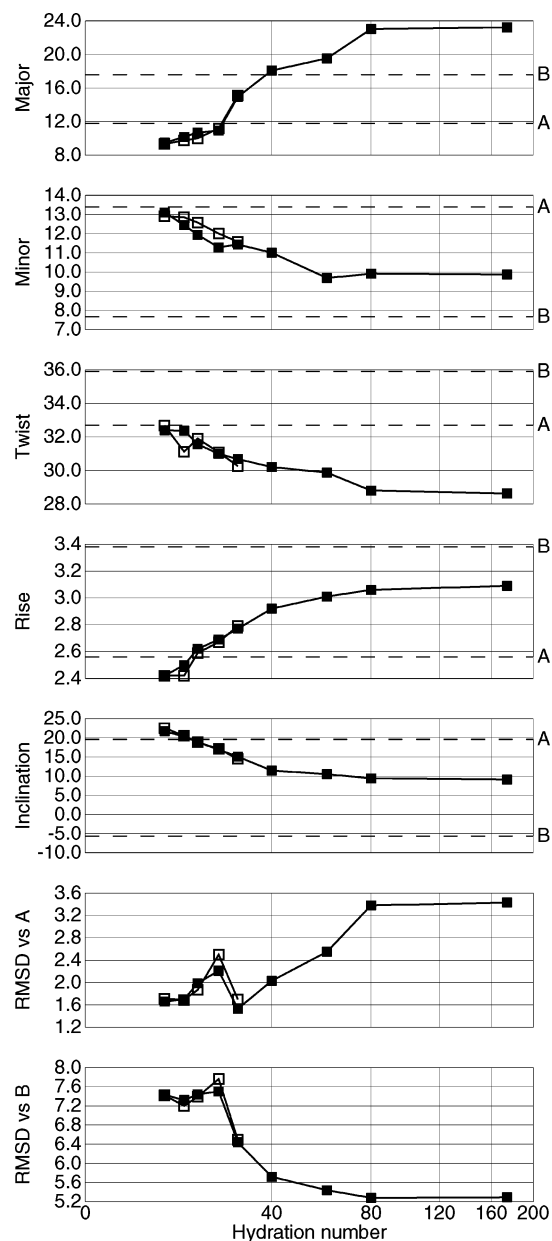
The standard initial states were prepared as follows. Either A- or B-DNA canonical structures<sup>66</sup> were first immersed in a large rectangular TIP3P<sup>67</sup> water box and next external solvent molecules were removed by using a spherical distance cutoff from DNA atoms. The cutoff radius was

adjusted to obtain the desired number of water molecules remaining. The drop was neutralized by randomly placing  $\text{Na}^+$  ions at water positions selected so that their distances from DNA were 5 Å or larger. The initial counterion distribution was preequilibrated by running 1 ns dynamics in water drops of 500 molecules for A-DNA and 800 molecules for B-DNA, with DNA atoms restrained to their initial positions. The final drop size was adjusted by adding or removing water from the surface. The described procedure was intended to ensure the start of dynamics from closely similar states regardless of the drop size.

Every system was energy minimized first with the solute held rigid and then with all degrees of freedom. Dynamics were initiated with the Maxwell distribution of generalized momenta at low temperature. The system was next slowly heated to 250 K and equilibrated during several picoseconds. Production trajectories were computed with the temperature bound to 293 K by the Berendsen algorithm<sup>68</sup> with a relaxation time of 10 ps. For better comparison with earlier simulations of A $\leftrightarrow$ B transitions, the original Cornell et al. force field<sup>69</sup> was used. Duration of production runs varied from 2 to 25 ns depending upon the observed character of dynamics. The conformations were saved with a 2.5 ps interval. The results were analyzed with in-house routines and the Curves program.<sup>70</sup>

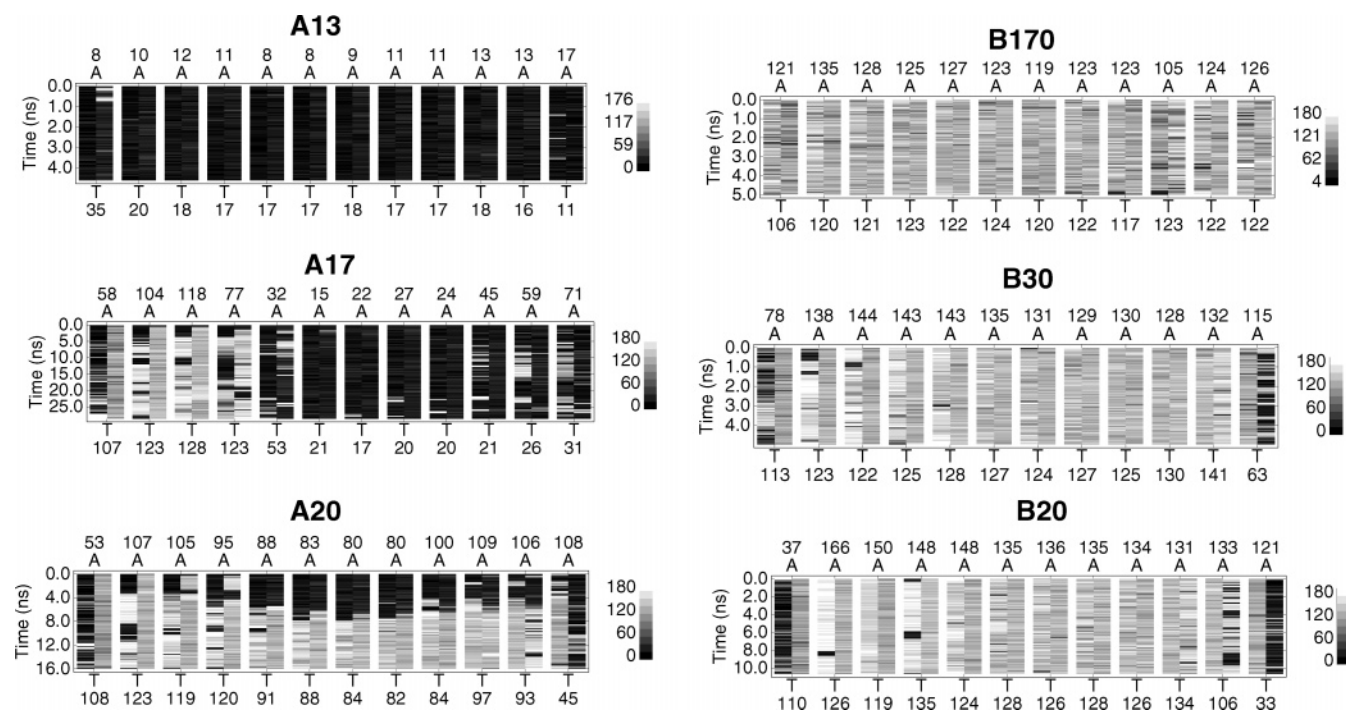
## Results

**Comparative Dynamics of A $\leftrightarrow$ B Transitions.** The B $\leftrightarrow$ A transition dynamics for C12 are shown in Figure 1 and Figure 2. Simulations starting from A- and B-DNA converged to identical ensembles of structures. This was checked for several  $\Gamma$  values covering the whole range of interest. The A-form loses stability with  $\Gamma \geq 20$ , which is similar to earlier results for the dodecamer CGCGAATTCGCG.<sup>31</sup> In contrast, the low limit of the B-form is shifted to higher  $\Gamma$  values. With  $\Gamma \approx 30$ , sugars in the purine strand concertedly switch to C3'-endo, and for  $20 < \Gamma < 30$  an intermediate state is observed, with one strand in nearly A-DNA conformation and the other closer to the B-form. As a result, the B $\leftrightarrow$ A transition looks easier and much smoother than for CGCGAATTCGCG.<sup>31</sup> The transition is complete in the center, and the A-DNA zone gradually spreads toward the termini as long as  $\Gamma$  is reduced. From the experiment, the B-form is known to be particularly resistant near the termini.<sup>71</sup> For the CGCGAATTCGCG sequence this B-philicity involved three base pairs, and it did not change with the duplex length increased.<sup>31</sup> Here this effect is much smaller. All this agrees with the known A-philicity of the poly(dG).poly(dC).<sup>72</sup> Figure 2 shows a quasi-static titration-like pattern of these transitions as monitored by different structural parameters. All traces exhibit S- or Z-shaped profiles sometimes with a very distinct transition zone.<sup>73</sup> The A-DNA structures obtained under low hydration are very close to the canonical conformation, with the final RMSD values below 2 Å. In contrast, under the highest hydration, an underwound B-DNA is observed with a strong bias toward A-form. Similar deviations were reported earlier by others,<sup>21,74,75</sup> and they are partially due to a known force field bias.<sup>76</sup>



**Figure 2.** A $\leftrightarrow$ B transitions in C12 monitored by different parameters. The data were obtained by averaging over the last 1 ns of trajectories computed with different hydration numbers. Open and closed squares show the results obtained from trajectories starting from A- and B-DNA, respectively. The two top plates show the average groove widths measured as described elsewhere.<sup>31,93</sup> The helical parameters are computed with program Curves.<sup>70</sup> All distances are in angstroms and angles in degrees. The dotted lines mark canonical A- and B-DNA levels.

Similar results for T12 are shown in the next two figures. In this case we failed to obtain a B $\rightarrow$ A transition despite many attempts. Trajectories starting from B-DNA with  $\Gamma < 20$  lead to deformed structures with some base pairs opened. The corresponding A-form trajectories were stable and showed no structural trends during reasonable simulation time (up to 30 ns). For  $\Gamma \geq 20$  normal A $\rightarrow$ B transitions were observed, and trajectories converged to similar structures starting from A- and B-DNA states. Because of the



**Figure 3.** A $\leftrightarrow$ B transitions in T12 shown by the dynamics of sugar pucker pseudorotation. The notation is as in Figure 1.

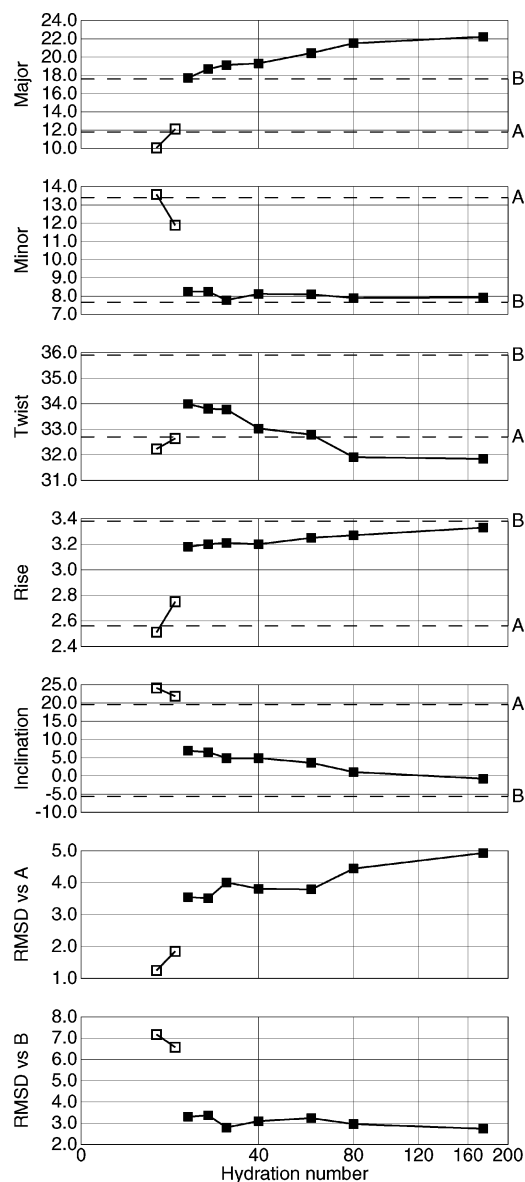
divergence of trajectories one cannot say which of the final states is more probable below  $\Gamma \approx 20$ . However, comparison of plates A20 and B20 in Figure 1 and Figure 3 shows that with  $\Gamma \geq 20$  opposite transitions are observed in C12 and T12. In this range of hydration, trajectories starting from A- and B-DNA were convergent for both duplexes; therefore, we can conclude that the experimental difference between poly(dG).poly(dC) and poly(dA).poly(dT) is qualitatively reproduced. Figure 5 shows the snapshots for T12 and C12 at the end of dynamics with  $\Gamma = 25$ . For T12 this is a B-DNA with a narrow minor groove near both ends and a widening in the middle. A significant number of the  $\text{Na}^+$  ions are seen in front of the minor groove. In contrast, C12 gives a typical A-DNA conformation with a layer of  $\text{Na}^+$  sandwiched between the opposite phosphate groups in the narrow major groove. This “electrostatic sandwich” provides the main driving force of the B  $\rightarrow$  A transition according to the free counterion driven condensation mechanism.<sup>26,33,31</sup>

The time averaged pattern of the transitions for T12 is shown in Figure 4. Compared to that in Figure 2 it shows similarities as well as differences. The low hydration A-form structures are very close to those for C12 and the canonical A-DNA. Unlike for C12, however, the computed high hydration structures represent rather good B-DNA. Only the twist remains clearly underestimated in agreement with earlier reports.<sup>76</sup> The traces in Figure 4 are interrupted to indicate that they perhaps do not sample from a single continuous transition pathway. Note that when  $\Gamma$  changes from 60 to 30 the structures show no trend toward the A-form. Instead some parameters are nearly stable and other even exhibit an opposite trend. A shift toward the A-form is observed under lower hydration, but it is relatively small and inconsistent in different parameters. This behavior is very different from C12 as well as the CGCGAATTCGCG dodecamer,<sup>31</sup> suggesting that the T12 structure may be

trapped in a pathway that leads away from the A-form and the B  $\rightarrow$  A transition would never occur if the corresponding trajectories were continued.

The major driving force of the B  $\rightarrow$  A transitions is arguably due to accumulation of free solvent ions in the major DNA groove.<sup>31</sup> Therefore, the difference between T12 and C12 should be also visible in the behavior of counterions. The corresponding cylindrical radial distribution functions are compared in Figure 6. These patterns provide useful information about the structure-specific interactions between DNA and free solvent cations. Because the duplexes are entirely covered by water the radial distributions of water oxygens effectively show the free space around DNA. All this space is, in principle, accessible for  $\text{Na}^+$  ions since their size is not very different from that of water. In the absence of specific DNA-ion interactions, the solid and dotted traces in Figure 6 should have similar peak positions. In contrast, sites of strong  $\text{Na}^+$ -DNA interactions produce separate peaks. Note also that the height of each peak in a cylindrical distribution should be multiplied by the corresponding distance when their relative weights are estimated.

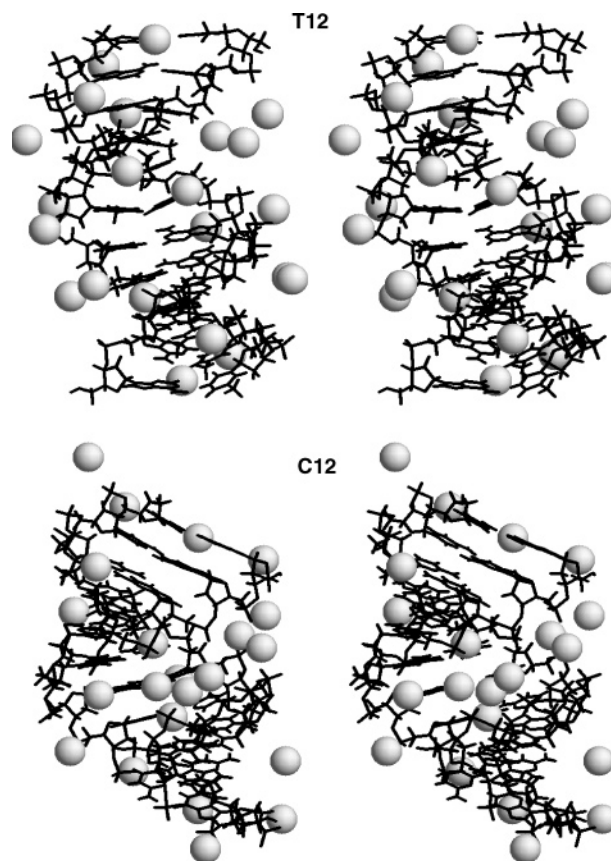
The characteristic distributions for B-DNA are best seen in the top three plates of T12. The three broad water peaks at approximately 4, 8 and 12.5 Å correspond, respectively, to the first water layer in the major groove, the next few layers in both major and minor grooves and the bulk water outside DNA radius which is about 10 Å for both A- and B-DNA. The characteristic A-DNA distributions can be seen in the bottom two plates of T12 as well as in several plates of C12. In this case the water radial distributions exhibit two broad maxima, one in the interior and another in the exterior of DNA. Each of the many peaks in the  $\text{Na}^+$  traces commonly involve contributions from more than one type of ion-DNA contacts. For instance, the prominent peak at 5 Å from the center of B-DNA results from direct as well as



**Figure 4.** A↔B transitions in T12 monitored by different parameters. Solid squares show results for  $\Gamma \geq 20$  that were similar for trajectories starting from A- as well as B-DNA. Open squares show results obtained from trajectories starting from A-DNA only. Other notation is as in Figure 2.

water-mediated contacts in depth of both major and minor DNA grooves. Most interesting for our purposes are the outer peaks that contain the major part of ions. For B-DNA this is the broad peak at about 12.5 Å which forms the helical axis. It corresponds to highly mobile ions involved in nonspecific phosphate screening. In A-DNA, the major  $\text{Na}^+$  peak is found at 10 Å, and it corresponds to the ions sandwiched in the opening of the major groove as seen in Figure 5.

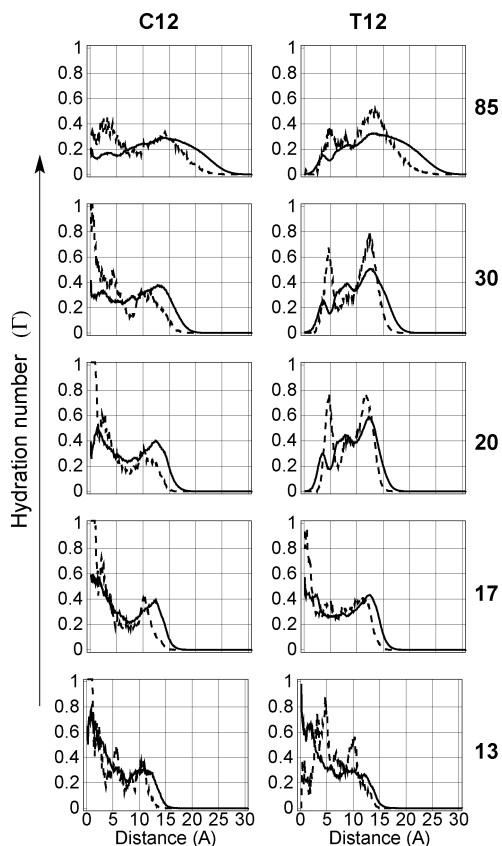
For C12 even under high hydration the  $\text{Na}^+$  distribution is rather different from typical B-DNA. These structures have a strong negative  $X_{\text{disp}}$ ; therefore, the helical axis is shifted to the major groove and is accessible to water and ions. Nevertheless, the large majority of counterions are found in the single broad outer peak characteristic of B-form. This peak does not change when the amount of water is further increased. In fact, ion distributions obtained in these condi-



**Figure 5.** Stereo snapshots of the convergent final states of T12 and C12 in a water drop corresponding to  $\Gamma = 25$ . The  $\text{Na}^+$  ions are shown by spheres.

tions are identical to those observed in infinite media with periodic boundaries.<sup>34</sup> In contrast, with reduced  $\Gamma$ , these counterions are progressively pushed closer to DNA, and, eventually, they accumulate in the characteristic A-DNA peak at 10 Å. This peak is predominant with  $\Gamma \leq 30$ . The foregoing scenario corresponds to the relatively smooth B→A transition shown in Figure 1 and Figure 2. The T12 patterns in the right column are radically different. In this case both water and  $\text{Na}^+$  distributions retain the characteristic B-DNA shapes even with  $\Gamma \approx 20$ . The ions are not pushed inside DNA when the outer water shell is reduced, and the  $\text{Na}^+$  peak at 12.5 Å remains dominant. As shown in Figure 5 many of these ions are aligned along the minor groove between the two phosphate strands. Such ions are essentially immobilized because there is not enough water for them to diffuse elsewhere.

**The Origin of the Resistance of poly(dA).poly(dT) to Adopting the A-Form.** The resistance of T12 to the B→A transition turned out to be only relative because in longer poly(dA).poly(dT) fragments such transitions were reproduced without major difficulties (the results not shown). The tendency of terminal base pairs to break was overcome by adding GC pairs at both duplex termini. Nevertheless, the special case of T12 deserves attention because its distinction from C12 is in clear correspondence with long-known experimental data, and it is interesting to clarify the origin of such behavior at least in simulations. To this end, we carried out a series of additional calculations with varied



**Figure 6.** Cylindrical radial distribution functions for water oxygens (solid lines) and  $\text{Na}^+$  ions (dashed lines) around C12 and T12. DNA structures saved in the last nanosecond of dynamics together with surrounding water and counterions were superimposed with the canonical B-DNA with the global coordinate OZ direction as the common helical axis. The  $\text{Na}^+$  ions and water oxygens were counted in coaxial 0.1 Å thick cylinders. The distributions are volume normalized, that is scaled with a factor of  $1/r$ , and the final plots were area normalized to fit the same scale. The corresponding hydration numbers are shown on the right.

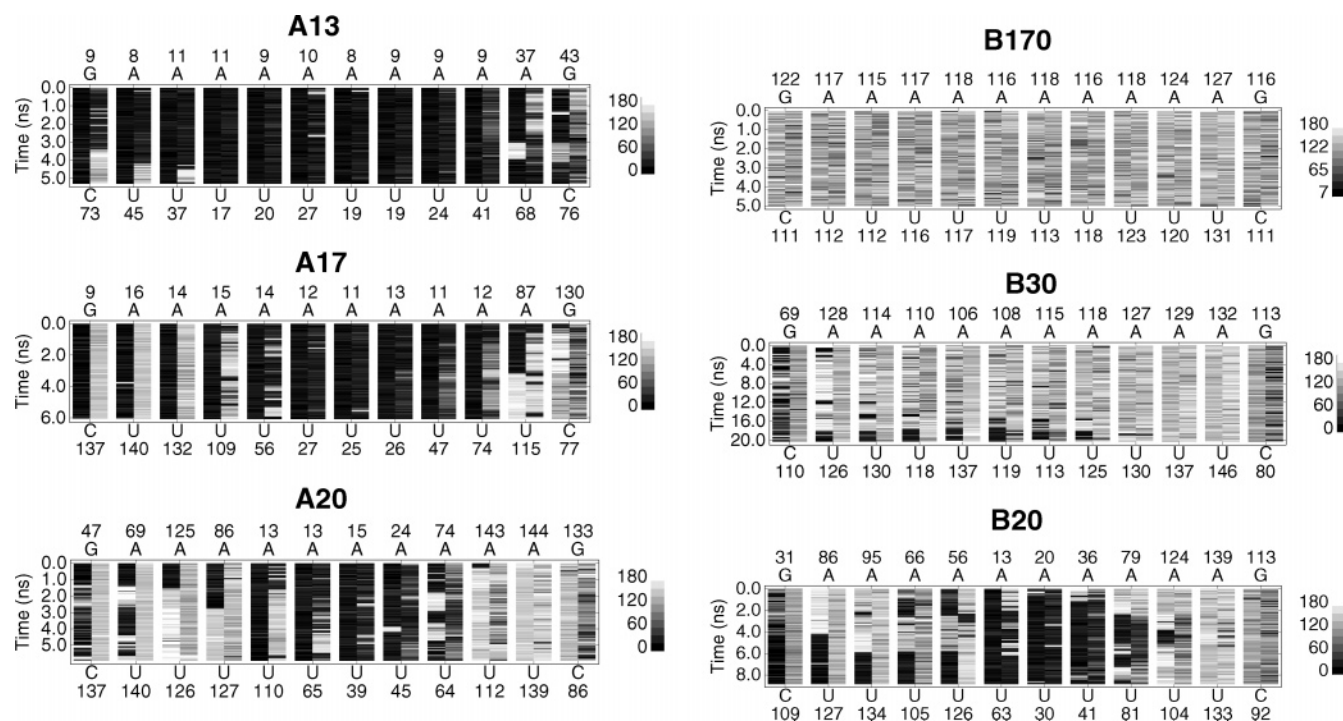
conditions. (i) It was checked that the A-form of poly(dA).poly(dT) is sufficiently stable with respect to the base pair opening. The forced breaking of one base pair did not perturb the overall structure even in very long simulations under intermediate hydration. (ii) The collapse of T12 under low hydration is reproduced when the size of the water drop is reduced gradually in several steps by removing each time only outer molecules not in contact with ions. This proves that the observed onset of denaturation is not caused by the initial conditions when the B-form is immediately placed under low hydration. (iii) The B $\rightarrow$ A transition still does not occur when the base pair opening is prevented by placing GC pairs at both termini (CT10C dodecamer). Starting from the B-form this duplex remained stable even under  $\Gamma = 17$ . With  $\Gamma = 13$  water started to leave the minor groove and the structure exhibited irregular deformations that drove it to a shrunk collapsed conformation but not to the A-form. At the same time, the CT10C duplex exhibited higher stability in the A-form, with AB transitions observed only beyond  $\Gamma = 25$ .



**Figure 7.** Stable conformations of C12 and CT10C under  $\Gamma = 13$ .

Figure 7 compares the conformations of C12 and CT10C observed in simulations starting from the canonical B-DNA under  $\Gamma = 13$ . For C12 this is an A-form structure with a very straight helical axis. In contrast, the CT10C structure represents a strongly curved B-like duplex with the minor groove closed near one end. The strong bend was evidently due to several  $\text{Na}^+$  ions sandwiched between the opposite phosphates in the locally narrowed major groove also seen in Figure 7. In this region, the major groove width is similar to that in A-DNA; therefore, the main driving factor of the B $\rightarrow$ A transition is present and it is sufficiently strong to bring the opposite phosphates together, but the overall duplex structure prefers to deform rather than go to the A-form. The CT10C transition from straight B-DNA to the shrunk structure in Figure 7 is easily reversible. With a few outer water shells added, the  $\text{Na}^+$  ions rapidly leave the major groove, the helical axis straightens up, and water invades the collapsed region of the minor groove.

Earlier it was noticed that the thymine methyl groups affect a number of different interactions involved in B $\leftrightarrow$ A transitions.<sup>18,58</sup> The array of thymine methyls in the major groove of poly(dA).poly(dT) strongly stabilize the B-DNA conformation,<sup>58,77</sup> both mechanically and via the hydrophobic effect, which would hinder the B $\rightarrow$ A transition. The distinct behavior of the CT10C duplex gives an opportunity to check this hypothesis. To this end we carried out a series of additional simulations for an analogous dodecamer duplex with thymine methyls replaced by hydrogens (CU10C). To avoid the possible additional effect of the charge redistribution inside pyrimidine rings all atoms in the newly obtained uracil residues had parameters identical to those in thymine, whereas the H6 hydrogens carried a residual charge identical to that of the thymine methyl group. The results obtained appear in Figures 8–10. It is readily seen that they agree with the above hypothesis. With C6 pyrimidine methyls replaced by hydrogens the B $\rightarrow$ A transition becomes possible, and reversible B $\leftrightarrow$ A transition are observed in the same range of hydration numbers as in C12 and earlier in CGCGAAT-TCGCG.<sup>31</sup> The transition patterns in Figure 8 resemble those in Figure 1 in that the sugars in the purine strand switch from C2'-endo to C3'-endo early in the B $\rightarrow$ A transition. The B $\rightarrow$ A transition occurs with somewhat lower  $\Gamma$  values suggesting that the CU10C duplex is less A-philic, which is also seen in Figure 9. Unlike for C12, the ion and water radial distributions obtained in large drops as well as with periodical boundaries exhibit clear B-DNA patterns with the



**Figure 8.** A $\leftrightarrow$ B transitions in CU10C shown by the dynamics of sugar pucker pseudorotation. The notation is as in Figure 1.

center of the duplex inaccessible for the solvent (see Figure 10). Overall this behavior resembles that of C12 as well as CGCGAATTCGCG,<sup>31</sup> but it is radically different from T12 and CT10C in agreement with an earlier suggested key role of the thymine methyls in the poly(dA).poly(dT) resistance to the B $\rightarrow$ A transformation. These results show also that the thymine methyls affect the B-DNA structure under high hydration and not just the B $\leftrightarrow$ A transition state.

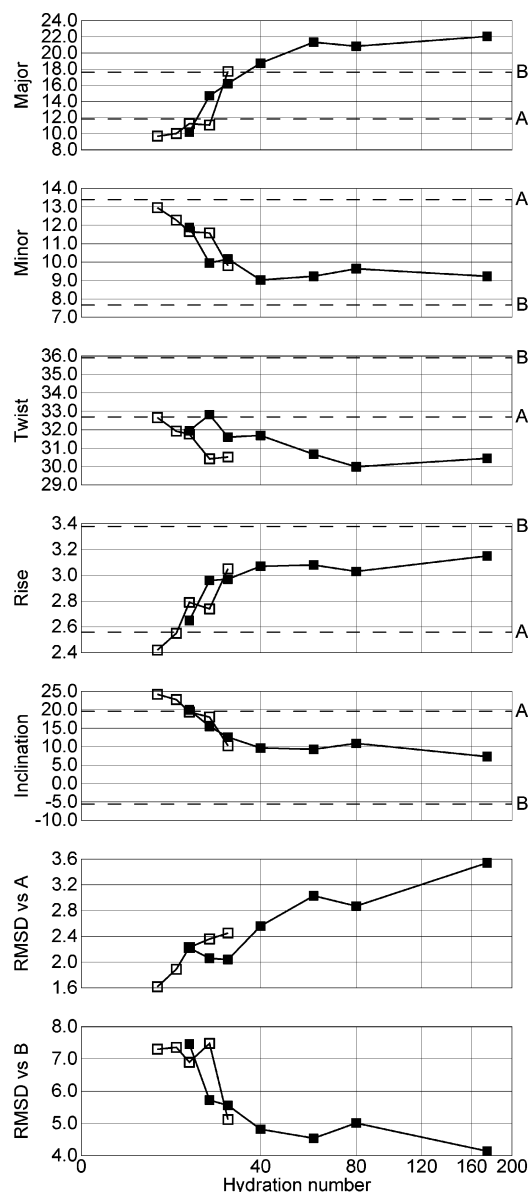
## Discussion

The striking contrast in the A/B polymorphism of poly(dA).poly(dT) and poly(dG).poly(dC) is a very long-known unexplained phenomenon.<sup>78</sup> Owing to the recent progress in the force field development<sup>69,79</sup> it is now possible to reproduce this effect in atomistic detail in a computer. The DNA structures computed in the course of this study are reasonably close to experiment, and several important earlier findings are qualitatively reproduced. (i) There is a range of hydration numbers where poly(dA).poly(dT) and poly(dG).poly(dC) are stable in B- and A-forms, respectively, with reproducible opposite transitions observed for the two sequences. Even with no additional force field fitting applied, the characteristic hydration values are close to experiment. (ii) The C12 dodecamer exhibits a very easy and smooth B $\leftrightarrow$ A transition under relatively high humidity, with its B-form featuring a strong A-DNA bias. (iii) The T12 dodecamer refuses to take the B-to-A transition pathway and is prone to denaturation. At the same time, longer poly(dA).poly(dT) fragments with GC termini can reversibly transform. These features resemble the experimental behavior of poly(dA).poly(dT) which is B-philic and tends to collapse instead of a B $\rightarrow$ A transition but still can go to the A-form in long A-tracts in solution.<sup>80</sup>

Under low humidity, all trajectories converge to similar A-form conformations that seem to be independent of the sequence and are always very close to the fiber canonical A-DNA. In contrast, a much more significant and sequence dependent divergence from experimental structures is observed for B-DNA. This result is in a surprising correspondence with experimental data. The high regularity and the absence of sequence effects for A-form was noticed long ago for X-ray fiber diffraction patterns<sup>7</sup> and later confirmed in the ensemble of single-crystal A-DNA structures.<sup>81</sup> Thus, in both experiment and simulations, the A-form of DNA is virtually insensitive to the base pair sequence. One should note, in addition, that it is much less sensitive to the accuracy of the force field than the B-form. Really, despite specific fitting of empirical potentials, at present, B-DNA structures obtained in free simulations only rarely approach experimental conformations closer than 2.5–3.0 Å RMSD for dodecamer duplexes. This is significantly higher than the RMSD numbers observed for A-DNA without any additional parameter fitting. All these observations are consistently explained if we assume that the A-form is dominated by the ion/phosphate “electrostatic sandwich” in the major groove. The strong interactions of phosphate groups with metal cations effectively impose geometric constraints upon the interphosphate distances and suppress all other factors that might affect the overall structure. The same interactions can maintain A-DNA in protein complexes. Interestingly, in all such structures refined until now the major DNA groove is exposed to solvent, with protein-DNA contacts limited to the minor groove only.

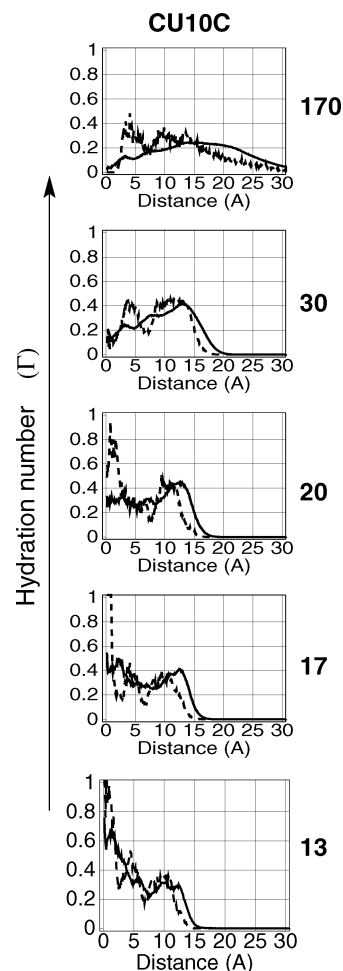
The observed systematic bias in the computed B-DNA conformations is similar to earlier reports.<sup>21,74,75</sup> The Cornell et al. parameters<sup>69</sup> tend to underestimate the helical twist in B-DNA,<sup>76</sup> and for the sequences studied here this bias is





**Figure 9.** A $\leftrightarrow$ B transitions in CU10C monitored by different parameters. Open and closed squares show the results obtained from trajectories starting from A- and B-DNA, respectively. Other notation is as in Figure 2.

perhaps the largest. The computed poly(dA).poly(dT) conformations are relatively close to the canonical B-DNA structure although, as earlier,<sup>82</sup> it is underwound by ca. 4° with respect to experiment in solution.<sup>41–43</sup> For poly(dG).poly(dC) a similar bias results in structures with very A-like helical parameters although the sugar pseudorotation dynamics features mainly south/east phases characteristic of B-DNA (see Figure 1). Earlier such structures were considered as strongly underwound B-DNA or as A-DNA with B-like backbone.<sup>21,74,75</sup> It is difficult to tell exactly how strong these deviations are with respect to experimental data. Compared to free G-tracts in crystals the computed structures deviate toward the B-form. However, these X-ray data are hardly appropriate for comparison because they all are obtained for a unique A-DNA specific packing characterized by very special DNA-DNA interactions in the minor groove.<sup>83</sup> In contrast, compared to G-tracts in protein-DNA



**Figure 10.** Cylindrical radial distribution functions for water oxygens (solid lines) and Na<sup>+</sup> ions (dashed lines) around CU10C. See details in the legend to Figure 6.

complexes, the computed structures deviate toward the A-form. However, it is also not clear if these data are relevant for comparison. For instance, there is a G5-tract in the complex of the refined structure of DNA polymerase Y.<sup>84</sup> This G-tract is protruding from the enzyme active center and makes no contacts with the protein. Its structure features a normal B-DNA conformation with south sugar pucker and the average twist of 34.2°, i.e., rather close to that of generic B-DNA. This value, however, diverges from experiments in solution (ca. 32.7° for long G-tracts<sup>41</sup>) suggesting that the structure may still be affected by the crystal environment.

The foregoing arguments show how challenging is the problem of modeling of sequence effects in DNA. On the other hand, NMR and circular dichroism (CD) studies in solution qualitatively agree with the character of G-tract structures obtained here and earlier by others with the same force field. The corresponding CD data suggest that the high hydration conformation of poly(dG).poly(dC) already has an A-like base pair stacking.<sup>21,74</sup> Nevertheless, these structures cannot be assigned to the A-form because the sugar phases determined by NMR are all in the south and because a cooperative transition to a genuine A-form is still distinguishable in solution as well as in crystalline fibers.<sup>19</sup> Since the corresponding spectral changes are relatively small, they can well result from a transition pattern similar to that in

Figure 2, with sugar phases switching to the north and grooves changing their width, but relatively minor shifts in helical parameters.

A number of earlier observations indicate that, during the B $\leftrightarrow$ A transition, the double helical DNA temporarily loses its stability. With reduced water content in solution the melting temperature of B-DNA first decreases and then goes up, with a minimum passed before the B $\rightarrow$ A transition occurs.<sup>85</sup> In some sequences even a temporary denaturation takes place followed by renaturation in the A-form when the intermediate range of hydration is passed.<sup>86</sup> The base pair opening in the AT-alternating sequence may also explain the reportedly very slow relaxation kinetics of B $\leftrightarrow$ A transitions in this duplex.<sup>35</sup> Therefore, the onset of denaturation in T12 in our simulations is not completely surprising. This DNA fragment as well as CT10C may represent examples of duplexes that are stable in A- as well as B-forms but cannot sustain the intermediate conditions. Similar experimental examples exist in the literature.<sup>86</sup> It is understood that, if the transition pathway passes through a denaturated state, it could not be reproduced in calculations.

A number of possible mechanisms were earlier discussed in relation to the resistance of poly(dA).poly(dT) to the B $\rightarrow$ A transition. This sequence effect can originate from specific base stacking as well as DNA-solvent interactions, but all such factors are mixed and difficult to separate.<sup>15,22,17</sup> For instance, the distinctive narrow minor groove of the B'-form of poly(dA).poly(dT)<sup>87</sup> is probably due to internal interactions, but it is also sealed by a very stable water spine hydration structure in the minor groove.<sup>88</sup> Moreover, the narrow groove should produce a 2-fold effect upon the counterion distribution around DNA.<sup>33,55</sup> The free positive ions should tend to accumulate in front of the narrow minor groove, and, simultaneously, their concentration in the major groove is reduced. All these features should stabilize the B-form and hinder the B $\rightarrow$ A transition. Different authors earlier pointed to thymine methyls as the probable principal cause of this effect.<sup>17,18,31</sup> We are not aware of published experiments on B $\leftrightarrow$ A transitions in poly(dA).poly(dU), but a generally strong effect of C6 substitutions in the uracil residue is well-known.<sup>77,86</sup> Here this hypothesis has been confirmed in realistic atom-level simulations. The thymine methyls generally do not hinder B $\leftrightarrow$ A transitions in water drop simulations with sequences other than poly(dA).poly(dT), for instance, poly(dAT) (unpublished results of the author). In contrast, for short fragments of poly(dA).poly(dT) their effect is very strong, which allowed us to set up a direct "yes or no" test.

Three different mechanisms of the thymine methyl effect were proposed earlier. (i) These groups may hinder a negative slide movement involved in the B $\rightarrow$ A transition.<sup>17</sup> (ii) They form a continuous nonpolar cluster in the major groove of B-DNA that should provide additional hydrophobic stabilization.<sup>18</sup> (iii) They reduce the accessible volume of the major groove, which may prevent accumulation of free solvent cations.<sup>31</sup> Our results agree with the first two explanations better than with the last one. Really, it turns out that, with removed methyl groups, the properties of B-DNA under high hydration are strongly affected and not just the B $\rightarrow$ A

transition state. It should be noted that all electrostatic properties of uracil bases here were nearly identical to those of thymines. Therefore the effect cannot be due to different ring stacking angles and should be due to either major groove hydration or some steric factors related to the arrays of methyl groups in tracts of stacked thymines.

Many years ago, the contrasting propensities of poly(dA).poly(dT) and poly(dG).poly(dC) to adopt A- and B-forms presented the first experimental demonstration of sequence dependent properties of the double helical DNA structure.<sup>78</sup> Since then the repertory of reported sequence effects has many times increased, and yet the exact physical origin of this particular difference remains controversial. It is shown here that A $\leftrightarrow$ B transitions observed in water drop simulations exhibit clear trends qualitatively similar to the long known experimental observations. These results corroborate the putative general role of the intraduplex electrostatic condensation mechanism for A $\leftrightarrow$ B transitions in DNA in vitro and suggest that future studies in the same direction can give more definite answers to the issues discussed here.

## References

- (1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (2) Ivanov, V. I.; Minchenkova, L. E. *Mol. Biol.* **1995**, *28*, 780–788.
- (3) Lu, X. J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300*, 819–840.
- (4) Foloppe, N.; Nilsson, L.; MacKerell, A. E., Jr. *Biopolymers* **1999**, *61*, 61–76.
- (5) Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *171*, 740–741.
- (6) Tunis-Schneider, M. J.; Maestre, M. F. *J. Mol. Biol.* **1970**, *52*, 521–541.
- (7) Leslie, A. G. W.; Arnott, S.; Chandrasekaran, R.; Ratliff, R. L. *J. Mol. Biol.* **1980**, *143*, 49–72.
- (8) Piskur, J.; Rupprecht, A. *FEBS Lett.* **1995**, *375*, 174–178.
- (9) Malenkov, G.; Minchenkova, L.; Minyat, E.; Schyolkina, A.; Ivanov, V. *FEBS Lett.* **1975**, *51*, 38–42.
- (10) Saenger, W.; Hunter, W. N.; Kennard, O. *Nature* **1986**, *324*, 385–388.
- (11) Fuller, W.; Mahensdasingam, A.; Forsyth, V. T. *Nature* **1988**, *335*, 596.
- (12) Ivanov, V. I.; Minchenkova, L. E.; Minyat, E. E.; Frank-Kametetskii, M. D.; Schyolkina, A. K. *J. Mol. Biol.* **1974**, *87*, 817–833.
- (13) Zimmerman, S. B.; Pfeiffer, B. H. *J. Mol. Biol.* **1980**, *142*, 315–330.
- (14) Cheatham, T. E., III.; Kollman, P. A. in *Interactions and Expression of Biological Macromolecules. Proceedings of the 10th Conversation, State University of New York, Albany, N. Y. 1998*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: New York, 1998; pp 99–116.
- (15) Calladine, C. R.; Drew, H. R. *J. Mol. Biol.* **1984**, *178*, 773–782.
- (16) Ivanov, V. I.; Minchenkova, L. E.; Schyolkina, A. K.; Poletaev, A. I. *Biopolymers* **1973**, *12*, 89–110.
- (17) Hunter, C. A. *J. Mol. Biol.* **1993**, *230*, 1025–1054.

- (18) Tolstorukov, M. Y.; Ivanov, V. I.; Malenkov, G. G.; Jernigan, R. L.; Zhurkin, V. B. *Biophys. J.* **2001**, *81*, 3409–3421.
- (19) Arnott, S.; Selsing, E. *J. Mol. Biol.* **1974**, *88*, 551–2.
- (20) Sarma, M. H.; Gupta, G.; Sarma, R. H. *Biochemistry* **1986**, *25*, 3659–3665.
- (21) Trantirek, L.; Stefl, R.; Vorlickova, M.; Koca, J.; Sklenar, V.; Kypr, J. *J. Mol. Biol.* **2000**, *297*, 907–22.
- (22) Mazur, J.; Sarai, A.; Jernigan, R. L. *Biopolymers* **1989**, *28*, 1223–1233.
- (23) Ivanov, V. I.; Zhurkin, V. B.; Zavriev, S. K.; Lysov, Y. P.; Minchenkova, L. E.; Minyat, E. E.; Frank-Kamenetskii, M. D.; Schyolkina, A. K. *Int. J. Quantum Chem.* **1979**, *16*, 189–201.
- (24) Cheatham, T. E., III.; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434–444.
- (25) Yang, L.; Pettitt, B. M. *J. Phys. Chem. B* **1996**, *100*, 2564–2566.
- (26) Cheatham, T. E., III.; Crowley, M. F.; Fox, T.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9626–9630.
- (27) Cheatham, T. E., III.; Kollman, P. A. *Structure* **1997**, *5*, 1297–1311.
- (28) Cieplak, P.; Cheatham, T. E., III.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 6722–6730.
- (29) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Am. Chem. Soc.* **1998**, *120*, 10629–10633.
- (30) Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 4658–4667.
- (31) Mazur, A. K. *J. Am. Chem. Soc.* **2003**, *125*, 7849–7859.
- (32) Feig, M.; Pettitt, B. M. *Biophys. J.* **1998**, *75*, 134–149.
- (33) Rouzina, I.; Bloomfield, V. A. *Biophys. J.* **1998**, *74*, 3152–3164.
- (34) Mazur, A. K. *J. Am. Chem. Soc.* **2002**, *124*, 14707–14715.
- (35) Jose, D.; Porschke, D. *Nucleic Acids. Res.* **2004**, *32*, 2251–2258.
- (36) Zimmerman, S. B. *Annu. Rev. Biochem.* **1983**, *51*, 395–427.
- (37) Becker, M. M.; Wang, Z. *J. Biol. Chem.* **1989**, *265*, 4163–4167.
- (38) Basham, B.; Schroth, G. P.; Ho, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6464–6468.
- (39) Dechering, K. J.; Cuelenaere, K.; Konings, R. N.; Leunissen, J. A. *Nucleic Acids Res.* **1998**, *26*, 4056–4062.
- (40) Vashakidze, R. P.; Prangishvili, D. A. *FEBS Lett.* **1987**, *216*, 217–220.
- (41) Peck, L. J.; Wang, J. C. *Nature* **1981**, *292*, 375–378.
- (42) Rhodes, D.; Klug, A. *Nature* **1981**, *292*, 378–380.
- (43) Strauss, F.; Gaillard, C.; Prunell, A. *Eur. J. Biochem.* **1981**, *118*, 215–222.
- (44) Hogan, M.; LeGrange, J.; Austin, B. *Nature* **1983**, *304*, 752–754.
- (45) Nishimura, Y.; Torigoe, C.; Tsuboi, M. *Nucleic Acids Res.* **1986**, *14*, 2737–2749.
- (46) Behling, R. W.; Kearns, D. R. *Biochemistry* **1986**, *25*, 3335–3346.
- (47) Benevides, J. M.; Wang, A. H.; Rich, A.; Kyogoku, Y.; van der Marel, G. A.; van Boom, J. H.; Thomas, G. J., Jr. *Biochemistry* **1986**, *25*, 41–50.
- (48) Peticolas, W. L.; Wang, Y.; Thomas, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2579–2583.
- (49) Buckin, V. A.; Kankiya, B. I.; Bulichov, N. V.; Lebedev, A. V.; Gukovsky, I. Y.; Chuprina, V. P.; Sarvazyan, A. P.; Williams, A. R. *Nature* **1989**, *340*, 321–322.
- (50) Buckin, V. A.; Kankiya, B. I.; Sarvazyan, A. P.; Uedaira, H. *Nucleic Acids Res.* **1989**, *17*, 4189–4203.
- (51) Brahms, S.; Fritsch, V.; Brahms, J. G.; Westhof, E. *J. Mol. Biol.* **1992**, *223*, 455–476.
- (52) Chalikian, T. V.; Plum, G. E.; Sarvazyan, A. P.; Breslauer, K. J. *Biochemistry* **1994**, *26*, 8629–8640.
- (53) Vorlickova, M.; Subriana, J. A.; Chladkova, J.; Tejralova, I.; Huynh-Dinh, T.; Arnold, L.; Kypr, J. *Biophys. J.* **1986**, *71*, 1530–1538.
- (54) Ivanov, V. I.; Minchenkova, L. E.; Burckhardt, G.; Birch-Hirschfeld, E.; Fritzsche, H.; Zimmer, C. *Biophys. J.* **1996**, *71*, 3344–3349.
- (55) Hud, N. V.; Plavec, J. *Biopolymers* **2003**, *69*, 144–159.
- (56) Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535–556.
- (57) Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug, A. *Nature* **1987**, *330*, 221–226.
- (58) Umezawa, Y.; Nishio, M. *Nucleic Acids. Res.* **2002**, *30*, 2183–2192.
- (59) Mazur, A. K. *J. Comput. Chem.* **1997**, *18*, 1354–1364.
- (60) Mazur, A. K. In *Computational Biochemistry and Biophysics*; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker: New York, 2001; pp 115–131.
- (61) Mazur, A. K. *J. Am. Chem. Soc.* **1998**, *120*, 10928–10937.
- (62) Mazur, A. K. *J. Chem. Phys.* **1999**, *111*, 1407–1414.
- (63) Mazur, A. K. *J. Phys. Chem. B* **1998**, *102*, 473–479.
- (64) Mazur, A. K.; Sumpter, B. G.; Noid, D. W. *Comput. Theor. Polym. Sci.* **2001**, *11*, 35–47.
- (65) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (66) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504–1509.
- (67) Jorgensen, W. L.; Chandreskhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (68) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (69) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (70) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63–91.
- (71) Minchenkova, L. E.; Schyolkina, A. K.; Chernov, B. K.; Ivanov, V. I. *J. Biomol. Struct. Dyn.* **1986**, *4*, 463–476.

- (72) The current AMBER parameters reportedly overestimate the stability of cytosine C2'-endo sugar pucker with respect to the C3'-endo.<sup>76,89</sup> This force field defect is perhaps responsible for a delayed transition of cytosine sugars to C3'-endo, and it may somewhat reduce the A-philicity of poly-(dG).poly(dC) DNA.
- (73) The quasi-static profiles in our calculations should be distinguished from familiar S-shaped experimental plots of cooperative A $\leftrightarrow$ B transitions.<sup>12</sup> In the latter case they result from two different contributions: (i) the shift in relative populations of A- and B-forms in the ensemble of DNA conformations and (ii) small deformations of the A- and B-conformations under varied hydration. Our simulations probe only the second contribution. Its relative weight in experiments is not known well. The Ising model of cooperative A $\leftrightarrow$ B transitions considers only the first contribution.<sup>12</sup> The second one is effectively neglected by assuming that in small DNA fragments the transition is described by a step function. It is known, however, that, beyond the transition zone of hydration numbers, conformations of both B- and A-DNA change in a smooth and noncooperative manner and that these changes are rather significant.<sup>90,91</sup> Evaluation of equilibrium populations of A- and B-forms from these types of simulations rests beyond the current possibilities; nevertheless, reversion of A $\leftrightarrow$ B transitions in a single MD trajectory can be observed in very long simulations under intermediate hydration (unpublished results).
- (74) Stefl, R.; Trantirek, L.; Vorlickova, M.; Koca, J.; Sklenar, V.; Kypr, J. *J. Mol. Biol.* **2001**, *307*, 513–24.
- (75) Lankas, F.; Cheatham, T. E., III.; Spaskova, N.; Hobza, P.; Langowski, J.; Sponer, J. *Biophys. J.* **2002**, *82*, 2592–2609.
- (76) Cheatham, T. E., III.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.
- (77) Wang, S.; Kool, E. T. *Biochemistry* **1995**, *34*, 4125–4132.
- (78) Pilet, J.; Blicharski, J.; Brahms, J. *Biochemistry* **1975**, *14*, 1869–1876.
- (79) MacKerell, A. D., Jr.; Wiórkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.
- (80) Ivanov, V. I.; Krylov, D. Y. *Methods. Enzymol.* **1992**, *211*, 111–27.
- (81) Suzuki, M.; Amano, N.; Kakinuma, J.; Tateno, M. *J. Mol. Biol.* **1997**, *274*, 421–35.
- (82) McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2001**, *314*, 23–40.
- (83) Wahl, M. C.; Sundaralingam, M. In *Oxford Handbook of Nucleic Acid Structure*; Neidle, S., Ed.; Oxford University Press: New York, 1999; pp 117–144.
- (84) Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. *Cell* **2001**, *107*, 91–102.
- (85) Ivanov, V. I.; Krylov, D. Y.; Minyat, E. E. *J. Biomol. Struct. Dyn.* **1985**, *3*, 43–55.
- (86) Vorlickova, M.; Sagi, J.; Hejtmankova, I.; Kypr, J. *J. Biomol. Struct. Dyn.* **1991**, *9*, 571–578.
- (87) Alexeev, D. G.; Lipanov, A. A.; Skuratovskii, I. Y. *Nature* **1987**, *325*, 821–823.
- (88) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Kopka, M. L. *Science* **1982**, *216*, 475–485.
- (89) Olson, W. K.; Zhurkin, V. B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 286–297.
- (90) Lee, C.-H.; Mizusawa, H.; Kakefuda, T. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2838–2842.
- (91) Harmouchi, M.; Albiser, G.; Premilat, S. *Eur. Biophys. J.* **1990**, *19*, 87–92.
- (92) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.
- (93) Mazur, A. K. *J. Mol. Biol.* **1999**, *290*, 373–377.

CT049926D

## Theoretical Study of the Antioxidant Activity of Vitamin E: Reactions of $\alpha$ -Tocopherol with the Hydroperoxy Radical

Marta Navarrete, Cipriano Rangel, Joaquín Espinosa-García, and José C. Corchado\*

*Departamento de Química Física, Universidad de Extremadura, 06071 Badajoz, Spain*

Received November 9, 2004

**Abstract:** The reactivity of the hydroperoxy radical with  $\alpha$ -tocopherol—a prototype of the chemical reactions involved in biological antioxidant actions—was studied theoretically. Two pathways were analyzed: hydrogen abstraction from the phenolic hydrogen and hydroperoxy addition to the aromatic ring. The reaction paths for the two mechanisms were traced independently, and the respective thermal rate constants were calculated using variational transition-state theory with multidimensional small-curvature tunneling. The reactivity of the hydroperoxy radical was found to be dominated by the hydrogen abstraction mechanism on  $\alpha$ -tocopherol, with a rate constant of  $1.5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  at 298 K. It was also found that the mechanism of the reaction is not direct but passes through two intermediates, one of which may have a significant role in preventing the pro-oxidant effects of  $\alpha$ -tocopherol.

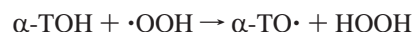
### 1. Introduction

Vitamin E is the most important lipid-soluble peroxy radical trapping antioxidant, retarding the oxidative degradation of lipids,<sup>1,2</sup> being located in the lipophilic domains of membranes and lipoproteins.<sup>3</sup> Its role is to react with the peroxy radicals present in the cytosol, preventing the chain reactions that lead to lipid peroxidation of the lipidic parts of membranes and lipoproteins.<sup>3,4</sup> The most active form of vitamin E is  $\alpha$ -tocopherol ( $\alpha$ -TOH).

Considerable experimental work has been devoted to the study of the activity of free-radical-chain-breaking antioxidants on biological systems.<sup>2,5,6</sup>  $\alpha$ -TOH and coenzyme Q (CoQ) partition in the lipid bilayer and act as natural radical-trapping antioxidants, avoiding or at least significantly reducing free radical reaction damage in the cytosol.<sup>7–10</sup> From these studies it can be concluded that the antioxidant activity of  $\alpha$ -tocopherol depends largely on its location. Thus, its scavenging activity is more pronounced in solution than in membranes and micelle systems,<sup>11–13</sup> and, when in solution, it significantly depends on the solvent.<sup>14</sup> This dependence has been partially ascribed to the possibility of hydrogen bonding with the solvent.<sup>14</sup> It has been found that the relative antioxidant activities of  $\alpha$ -tocopherol versus coenzyme Q

are as follows:  $\text{CoQ} > \alpha\text{-TOH}$  in LDL;  $\text{CoQ} < \alpha\text{-TOH}$  in homogeneous solution; and  $\text{CoQ} \approx \alpha\text{-TOH}$  in aqueous lipid dispersions.<sup>15</sup>

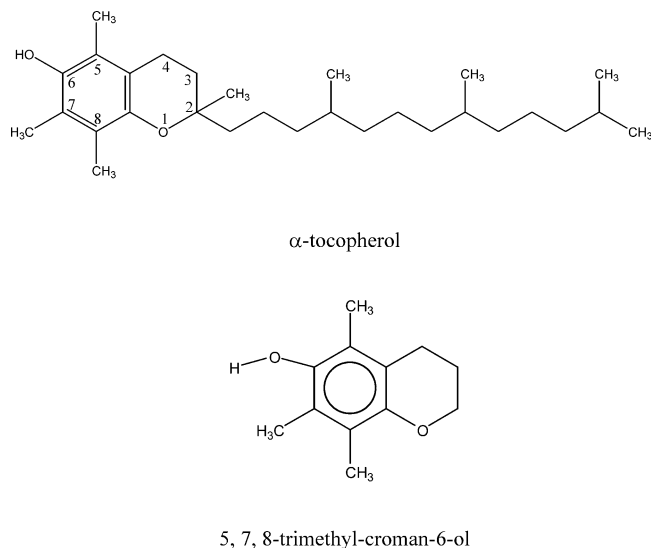
It is also well-known that  $\alpha$ -tocopherol can act as a pro-oxidant molecule.<sup>6</sup> This is a result of the relatively long life of the  $\alpha$ -tocopheryl radical ( $\alpha\text{-TO}\cdot$ ) formed after the proton-transfer reaction



takes place. This  $\alpha$ -tocopheryl radical can react with lipids present in the nearest environment (membranes or lipoproteins) rather than reacting with other molecules that would regenerate the  $\alpha$ -tocopherol, such as vitamin C or CoQ. Fortunately, the rate of the hydrogen abstraction reaction from lipids is very slow. Hence, even though lipids are in close contact with the recently formed  $\alpha$ -tocopheryl radical, it is rather unlikely that in the usual *in vivo* conditions the hydrogen abstraction reaction from lipidic chains should occur. Thus, the  $\alpha$ -tocopheryl radical will rather remain unmodified until a proper vitamin E regenerator reacts with it, transferring it a proton to form  $\alpha$ -tocopherol.

In previous papers we studied for the first time the mechanism and kinetics of reactions between CoQ and the OH radical (theoretically and experimentally)<sup>16</sup> and between CoQ and the OOH radical (only theoretically).<sup>17</sup> In this last

\* Corresponding author e-mail: corchado@unex.es.



**Figure 1.** Numbering scheme for α-tocopherol and its model system employed in the present calculation.

case, directly related with the aim of the present work, we found that the OOH free radical attacks CoQ via two mechanisms: a hydrogen abstraction reaction from the phenolic hydrogen on the reduced form of CoQ (ubiquinol) and addition on the oxidized form (ubiquinone). We found that the reactivity of the OOH radical is dominated by the hydrogen abstraction reaction, with a rate constant of  $5.3 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$  at 298 K.

In the general context of the study of the antioxidant activity, and as a continuation of our previous studies with CoQ, in the present work we study the gas-phase reactivity of α-tocopherol with the hydroperoxy radical ( $\text{HOO}^\bullet$ ) as a model of the reactions against oxidative stress that take place in biological systems without including environmental effects. The aim is 3-fold: first, to propose a mechanism to account for the attack of the hydroperoxy radical on α-tocopherol, i.e., to analyze theoretically the possible pathways; second, to obtain theoretical kinetics information; and third, to compare the antioxidant activity of the two potent natural radical scavengers, α-TOH and CoQ. In Section 2, we describe the theoretical methods and computational details used in the work; results and a discussion are given in Section 3; and conclusions are presented in Section 4.

## 2. Methods and Computational Details

**Modeling.** Because of the large size for our molecular system and the great number of calculations performed in the reaction path constructions, the real biological reaction was modeled in the following way. First, we replaced the trimethyltridecyl and methyl chains attached to carbon 2 in the α-tocopherol molecule by hydrogen atoms (see Figure 1). These chains have a prominent role in anchoring the tocopherol molecule to the membrane and restricting its mobility. However, it is very unlikely they have any major influence on the reactivity of the system, since the reactive part of tocopherols is the chromanol ring, which has been kept unchanged in our model system. Moreover, these chains are protected by the lipidic environment against attack from

cytosolic radicals; only part of the chromanol ring projects out from the membrane,<sup>3</sup> and this is the only point where reactions with polar molecules solvated by cytosolic water can take place. Therefore, we modeled the α-tocopherol molecule by the 5,7,8-trimethylcroman-6-ol molecule (Figure 1). A similar model reducing the size of the real system had already been used with success in our previous studies of CoQ,<sup>16,17</sup> and it was consistent with the conclusions of Foti et al.<sup>18</sup> that the hydrogen abstraction reaction from CoQ is independent of the size of the chain attached to the aromatic ring.

Second, while the theoretical study was performed in the gas-phase, given the nonpolar character of the natural environment (lipid bilayer), one can reasonably assume that the conclusions will be roughly the same in both environments.

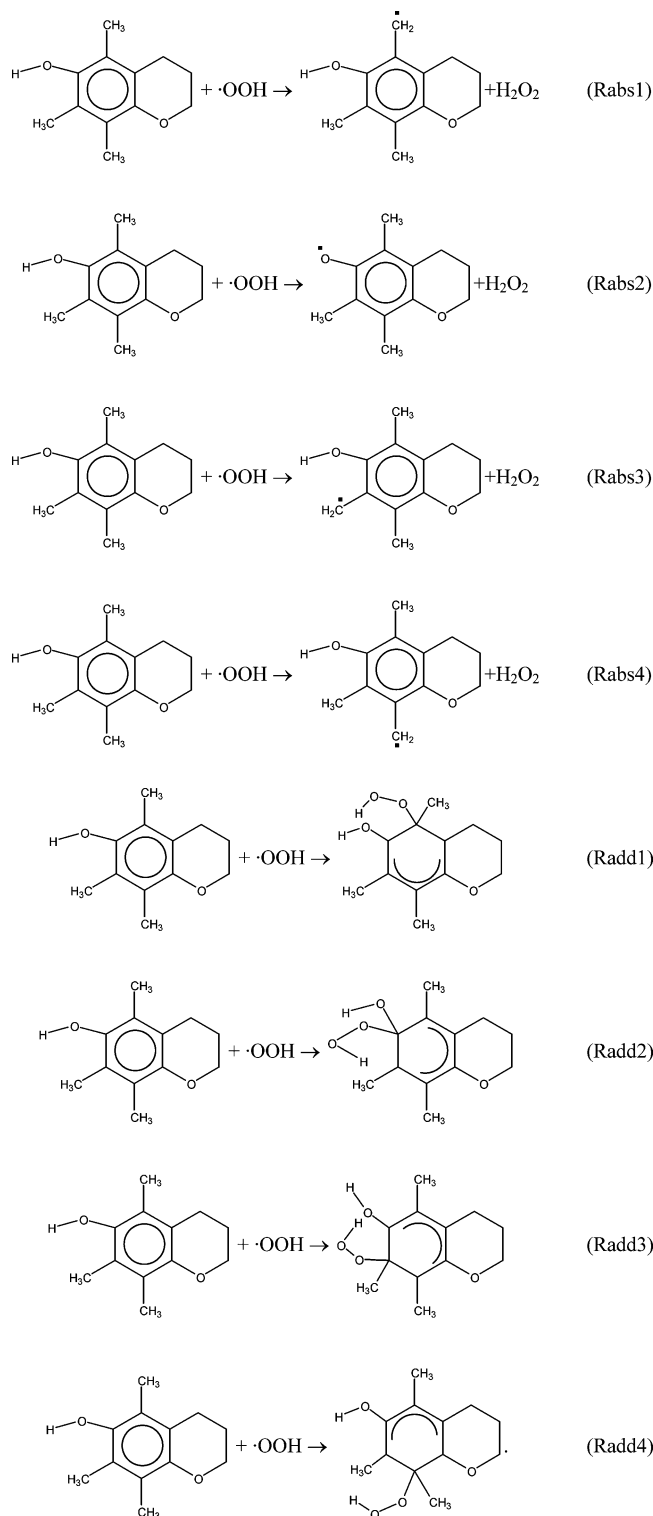
To study all the possible side reactions that can take place between α-tocopherol and the hydroperoxy radical, we took two mechanisms into account: the hydrogen abstraction reaction and the reaction of addition to the aromatic ring. With respect to the hydrogen abstraction mechanism, we considered that the hydroperoxy radical can abstract either the phenolic hydrogen (attached to the oxygen on carbon 6) or any of the methylic hydrogens (from the methyl groups attached to carbon atoms 5, 7, and 8). With respect to the addition reaction, we also take into account four addition centers, namely carbons 5, 6, 7, and 8. Therefore, we took eight possible reactions into account, four hydrogen abstraction reactions and four addition reactions, as shown in Figure 2.

**Electronic Structure Calculations.** The geometries of the reactants (α-TOH and  $\cdot\text{OOH}$ ) and products (four abstraction products plus HOOH, as well as four addition products) were optimized using hybrid density functional theory (DFT) as implemented in the Gaussian 98<sup>19</sup> suite of programs. Exchange and correlation were treated by the BHandHLYP method, which is based on Becke's half-and-half method<sup>20</sup> and the gradient-corrected correlation functional of Lee, Yang and Parr,<sup>21</sup> using the 6-31G basis set.<sup>22</sup> We will denote this level by its usual abbreviation:

- BHandHLYP/6-31G (hereafter called Level 0).

Vibrational frequencies were calculated using Level 0 in order to check that these geometries correspond to true minima on the potential energy surface, verifying that all the vibrational frequencies are real. Tables listing the geometries, energies, and vibrational frequencies of all the stationary points are given as Supporting Information. For the most favorable reaction channels (see Section 3) we located their saddle points and calculated reaction barrier heights using Level 0, checking the nature of the saddle points by verifying that they possess a single imaginary frequency. Starting from a saddle point, we followed the reaction path both toward reactants and products,<sup>23</sup> calculating vibrational frequencies after projecting out the motion along the reaction path using redundant internal coordinates.<sup>24</sup>

When following the reaction path for the hydrogen abstraction reaction from the phenolic hydrogen, we found a minimum on the reactant side and another minimum on the product side. The two minima, which we will denote as



**Figure 2.** Reaction mechanisms taken into account for the  $\alpha$ -tocopherol +  $\cdot\text{OOH}$  reaction.

reactant well and product well, were further optimized and their vibrational frequencies were calculated using Level 0. Both were true minima (all the vibrational frequencies were real). Geometries, energies, and vibrational frequencies of the computed saddle points and complexes are also given as Supporting Information.

**Dynamics Calculations.** The kinetics study of the selected reaction mechanisms of  $\alpha$ -tocopherol with the hydroperoxy radical were carried out using the direct dynamics approach,<sup>25</sup>

by using the information on reaction paths described above and a mapping interpolation procedure,<sup>26</sup> in order to minimize the errors caused by the limited information we calculated. Rate constants were estimated using canonical variational transition state theory (CVT).<sup>27</sup> Quantum effects on motions transversal to the reaction path were included using quantum-mechanical vibrational partition functions in the harmonic oscillator approach, while quantum effects on the motion along the reaction path were included using a semiclassical multidimensional method for tunneling, namely the small-curvature tunneling method (SCT).<sup>27</sup> Kinetics calculations were performed using the Polyrate<sup>28</sup> and Gaussrate<sup>29</sup> computer codes.

It has been found that DFT methods usually underestimate barrier heights for hydrogen abstraction reactions, and an alternative DFT method has been devised for the accurate description of these barrier heights, denoted MPW1K.<sup>30</sup> Moreover, it is clear that the results also depend on the basis set employed in the calculations. To check the accuracy of our calculations, we performed selected single-point calculations using both BHandHLYP and MPW1K functionals and several extended basis sets, namely 6-311G(2d,p),<sup>31</sup> 6-311+G(d,p),<sup>31</sup> 6-311G+(2d,p),<sup>31</sup> and MG3.<sup>32</sup> These single-point calculations will be denoted by the usual double-slash notation:

- MPW1K/6-311G(2d,p)//Level 0 (hereafter called Level 1),
- BHandHLYP/6-311+G(d,p)//Level 0 (Level 2),
- BHandHLYP/6-311G(2d,p)//Level 0 (Level 3),
- BHandHLYP/6-311+G(2d,p)//Level 0 (Level 4),
- MPW1K/MG3//Level 0 (Level 5).

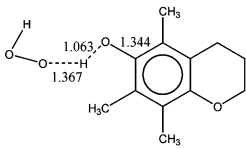
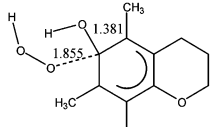
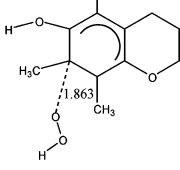
Using Level 1 we calculated single-point energies at selected points along the reaction path, while Levels 2 through 5 were used for single-point calculations only at the most relevant stationary points (reactants, products, and saddle point). This information was conveniently interpolated using the methods of ref 33 in order to perform dual-level kinetics calculations.<sup>34</sup> The dual-level kinetics calculations will be denoted using the recommended triple-slash notation, X//Y///Z, where the term to the right of the triple-slash, Z, denotes the lower level (Level 0), while the term to its left, X//Y, denotes the higher level (Levels 1 through 5).

### 3. Results and Discussion

**Thermochemistry.** As we noted in Section 2, we considered eight possible reactions: four abstraction reactions, namely Rabs1, Rabs2, Rabs3, and Rabs4, and four addition reactions, namely Radd1, Radd2, Radd3, and Radd4, as shown in Figure 2.

A priori, abstraction reactions seem to be more plausible, since addition reactions involve the breaking of the aromaticity of the system. Moreover, taking into account that the bond dissociation energies (BDE's) of C–H bonds are usually higher than those of phenolic O–H bonds,<sup>35</sup> from simple inspection of the products, chemical intuition leads us to predict that the most exothermic abstraction reaction will be Rabs2. Table 1 lists the calculated changes in the classical (Born–Oppenheimer) energy for the four abstrac-

**Table 1.** Energy Changes (in kcal/mol) for the Eight Reactions and Selected Saddle Points Calculated Using Level 0 (Distances are in Å)

| Reaction | $\Delta E$ | Saddle point  | $\Delta E^\ddagger$ |
|----------|------------|---|---------------------|
| Rabs1    | 7.9        |   |                     |
| Rabs2    | -7.5       |  | 4.2                 |
| Rabs3    | 8.7        |   |                     |
| Rabs4    | 8.2        |   |                     |
| Radd1    | 8.9        |   |                     |
| Radd2    | 5.5        |  | 6.8                 |
| Radd3    | 3.3        |  | 10.2                |
| Radd4    | 6.7        |   |                     |

tion reactions. Only Rabs2 is an exoergic reaction. It can also be seen that Rabs1, Rabs3, and Rabs5 have similar endoergicities, which can be attributed to the fact that the bond being dissociated is the same (C–H bond) and their environments are quite similar. Since the differences in exoergicities between Rabs2 and the other abstraction reactions are greater than 15 kcal/mol, one can expect that the only significant abstraction reaction will be Rabs2, and this is the only abstraction reaction that we will study further. With respect to the addition reactions (Table 1), it seems that Radd2 and Radd3 are the most plausible reactions. Although they are endoergic, addition reactions sometimes show low activation energies, and in principle they could compete kinetically with the abstraction reaction Rabs2. Therefore we will continue with the study of Radd2 and Radd3 despite the differences in energy changes with respect to Rabs2 (more than 10 kcal/mol).

The geometry and energy (measured with respect to the reactants) of the saddle points along the most favorable reaction paths, namely, Rabs2, Radd2 and Radd3, are also listed in Table 1. It is also interesting to look at the changes of some geometrical features as the reactions proceed. Thus, in the addition reactions, the bonds being formed are significantly stretched at the saddle points for Radd2 and Radd3, being somewhat larger than a single C–O bond (1.435 Å in methanol, for example) and slightly shorter for Radd2, in agreement with the fact that this reaction shows a lower barrier if we assume that the reason is that the new bond is stronger. With respect to the abstraction reaction, the bond being broken is about 11% larger than in the

**Table 2.** Rate Constants and Transmission Coefficients for the Rabs2 and Radd2 Reactions (in  $M^{-1} s^{-1}$ ) at Level 0

| T (K) | Rabs2                 |                |          | Radd2    |                |          |
|-------|-----------------------|----------------|----------|----------|----------------|----------|
|       | CVT                   | $\kappa$ (SCT) | CVT-SCT  | CVT      | $\kappa$ (SCT) | CVT-SCT  |
| 250   | 9.8(+03) <sup>a</sup> | 9.1            | 8.9(+04) | 3.0(-02) | 0.9            | 2.7(-02) |
| 275   | 1.9(+04)              | 6.4            | 1.2(+05) | 1.3(-01) | 0.9            | 1.2(-01) |
| 298   | 3.1(+04)              | 4.9            | 1.5(+05) | 4.3(-01) | 0.9            | 3.7(-01) |
| 300   | 3.2(+04)              | 4.8            | 1.5(+05) | 4.7(-01) | 0.9            | 4.1(-01) |
| 325   | 5.1(+04)              | 3.8            | 2.0(+05) | 1.4(+00) | 0.8            | 1.2(+00) |
| 350   | 7.8(+04)              | 3.1            | 2.5(+05) | 3.6(+00) | 0.8            | 3.0(+00) |
| 400   | 1.6(+05)              | 2.4            | 3.8(+05) | 1.7(+01) | 0.8            | 1.4(+01) |
| 500   | 4.7(+05)              | 1.6            | 7.6(+05) | 1.7(+02) | 0.8            | 1.4(+02) |
| 600   | 1.0(+06)              | 1.3            | 1.4(+06) | 8.3(+02) | 0.3            | 6.9(+02) |

<sup>a</sup> 9.8(+03) stands for  $9.8 \times 10^{+03}$ .

reactants (1.063 versus 0.960 Å), while the bond being formed is 41% larger than in the products (1.367 versus 0.968 Å), i.e., the saddle point resembles reactants rather than products (it is an early transition state). Note also that the bond between carbon 6 and the phenolic oxygen is shortened. The reason is that the radical being formed is starting to take on part of the aromatic structure, and the C–O bond is starting to take on some double-bond character.

**Kinetics.** As the starting point in our kinetics studies, we located the saddle points along the reaction paths for the three selected reactions, namely Rabs2, Radd2, and Radd3 (Table 1). There are several features we would like to note. First, the reaction barrier is lower for the Rabs2 case. Second, the most endoergic reaction, Radd2, shows a lower barrier height than Radd3. Third, while the saddle point for Rabs2 maintains the aromaticity of the ring, the saddle points for Radd2 and Radd3 lose part of this aromaticity, as is manifested by a deviation of the planarity of the carbon to which OOH is being bonded with respect to the remaining carbons. The latter observation agrees with the fact that addition reactions show a higher barrier. Since Radd3 has a barrier about 3.5 kcal/mol higher than Radd2, in the rest of the paper we will consider that the only plausible path for an addition reaction is the addition to carbon 6, and therefore only take into account the competition between the abstraction Rabs2 and the addition Radd2 reactions.

After calculating the saddle points, we followed the reaction paths for both reactions and used this information to calculate the CVT rate constants and SCT transmission coefficients. The resulting CVT/SCT rate constants are listed in Table 2 for the temperature range 250–600 K, using Level 0. Several features merit discussion.

The rate constants for the abstraction reaction are several orders of magnitude greater than the rate constants for the addition reaction. Therefore, we can consider that the reaction between  $\alpha$ -tocopherol and the hydroperoxy radical occurs exclusively by hydrogen abstraction of the phenolic hydrogen. There are three reasons why this is the only reaction that takes place. First, from an energy point of view, both the saddle point for the abstraction reaction and the products keep the aromaticity of the chromanol ring, while the addition reaction involves losing this aromaticity. Thus, the energy barrier against the reaction is about 2.5 kcal/mol lower for the abstraction case.



**Table 3.** Energy ( $\Delta E$ ) and Enthalpy ( $\Delta H$ ) at 298 K of Reaction (R) and Activation ( $\neq$ ) for Rabs2 Using Different Levels of Calculation (in kcal mol<sup>-1</sup>)

|         | $\Delta E_R$ | $\Delta H_R$ | $\Delta E^\neq$ | $\Delta H^\neq$ |
|---------|--------------|--------------|-----------------|-----------------|
| Level 0 | -7.7         | -8.0         | 4.2             | 2.6             |
| Level 1 | -5.2         | -5.5         | 6.3             | 4.7             |
| Level 2 | -7.3         | -7.6         | 9.0             | 7.4             |
| Level 3 | -6.7         | -7.0         | 8.5             | 6.8             |
| Level 4 | -6.9         | -7.2         | 9.3             | 7.7             |
| Level 5 | -5.3         | -5.6         | 7.5             | 5.9             |

Second, from an entropy point of view, the abstraction reaction is also favored. In both cases, as the reaction proceeds the relative motion of the two reactants is hindered, and some entropy is lost. However, in the abstraction reaction there is a cleavage of an O–H bond, while a new O–H bond is being formed. This leads to a larger amplitude motion of the hydrogen being transferred, increasing the entropy of this part of the system with respect to the reactants. This effect does not occur in the addition reaction, where there is a more subtle rearrangement of bonds and electronic structure, with a much smaller effect on the entropy of the system. Therefore, from an entropy point of view, the abstraction reaction is also favored. It can be analyzed quantitatively in a simple way by comparing the vibrational zero-point energies of the saddle points of the two reactions. Thus, while the zero-point energy of the saddle point of Rabbs2 is 175.5 kcal/mol, for Radd2 it is 178.3 kcal/mol, almost 3 kcal/mol higher.

Third, the transmission coefficient, which mostly takes into account quantum effects on the motion along the reaction path, increases the abstraction reaction by about a factor of 5 at 298 K, while it slightly diminishes the addition reaction.<sup>38</sup> This behavior was to be expected, since the abstraction reaction involves the motion of a light particle (a hydrogen atom) that can easily tunnel through the reaction barrier, while the addition reaction involves the movement of heavy atoms. In fact, at 298 K, about 80% of the abstraction reaction is due to tunneling.

Thus, the effective classical barrier height is expected to be around 6 kcal/mol lower for the abstraction reaction, and quantum effects are expected to lower this barrier much more than the addition reaction barrier. Addition can therefore not compete with abstraction.

These results agree with those obtained for the CoQ + OOH reaction,<sup>17</sup> where the hydrogen abstraction reaction is also the dominant mechanism, although in that case only 50% of the reaction at 298 K is due to tunneling.

**Influence of the Level of Calculation.** The computational requirements for the thermochemical calculations of eight reactions and two converged rate constants for a system of this size are extremely high. This is why we selected a quite small basis set for our calculations (Level 0). Although it has been pointed out that DFT methods are not as sensitive as ab initio methods to the size of the basis set,<sup>40</sup> we decided to check the accuracy of our calculations against experimental values and higher-level calculations.

Table 3 lists the energy and enthalpy changes at 298 K of reaction and activation (estimated at the saddle point)

**Table 4.** Bond Dissociation Energies at 298 K and Enthalpy of the Rabbs2 Reaction Calculated Using Different Computational Levels (in kcal mol<sup>-1</sup>)

|         | $\alpha$ -tocopherol    | H <sub>2</sub> O <sub>2</sub> | $\Delta H_R$ |
|---------|-------------------------|-------------------------------|--------------|
| Level 0 | 70.0                    | 78.0                          | -8.0         |
| Level 2 | 73.3                    | 80.9                          | -7.6         |
| Level 3 | 72.7                    | 79.7                          | -7.0         |
| Level 4 | 73.0                    | 80.2                          | -7.2         |
| Exp.    | 77.3 ± 1.0 <sup>a</sup> | 88.2 ± 1.0 <sup>b</sup>       | -10.9 ± 2.0  |

<sup>a</sup> Reference 36. <sup>b</sup> Reference 41.

calculated for Rabs2 using the different levels listed in Section 2. We shall begin by analyzing the reaction enthalpy. The experimental value of the enthalpy of reaction at 298 K as predicted by the differences in bond dissociation energies of the O–H bond in  $\alpha$ -tocopherol (77.3 ± 1.0 kcal/mol)<sup>36</sup> and hydrogen peroxide (88.2 ± 1.0 kcal/mol)<sup>41</sup> is -10.9 ± 2.0 kcal/mol. Therefore, all of the methods predict a reaction less exothermic than the experimental reference values, Level 0 being the one that shows closest agreement taking into account the experimental error bar, and the MPW1K-based levels being the ones that deviate the most. The reason the smallest basis set gives the value closest to experiment is that errors in the bond dissociation energies compensate each other. Table 4 lists the BDE's of  $\alpha$ -tocopherol and hydrogen peroxide computed using the BHandHLYP based levels. As one increases the basis set, both BDE's come closer to the experimental values, but the  $\alpha$ -tocopherol BDE is improved more than the hydrogen peroxide BDE. As a result, increasing the basis set leads to poorer results, and Level 0 seems to give a more balanced description of the reaction. This seems to be a general behavior for these systems, because a similar result was found in our previous study of the CoQ + OOH reaction.<sup>17</sup>

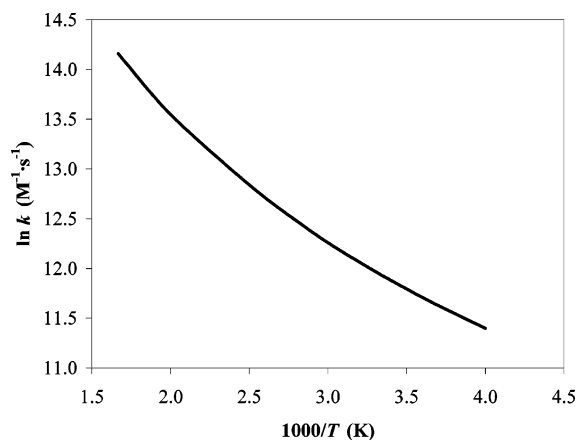
With respect to the barrier height, direct comparison with experimental or theoretical works is unfortunately not possible because such information is unavailable. We found that, leaving aside Level 0, the levels based on the MPW1K functionals (Levels 1 and 5) predict lower barrier heights than the BHandHLYP-based methods (Levels 2,3,4). In both cases, changes in the basis sets modify the computed magnitudes by about 1 kcal/mol for methods that use the same functional, confirming the conclusion of Koch and Holthausen<sup>40</sup> that DFT methods are fairly insensitive to the size of the basis set. In sum, Level 0 predicts the most exothermic reaction and the lowest barrier height, and based on its best agreement with the only experimental magnitude available, the enthalpy of reaction at 298 K, it will be the level of calculation used in the remainder of the paper.

To complete the check of the levels and basis sets, Table 5 lists the dual-level rate constants for Rabs2 at 298 K, using different functionals and basis sets. This table shows that increasing the level of calculation leads to a severe underestimate of the rate constant as compared with the value of the reaction between  $\alpha$ -tocopherol and the *tert*-butylperoxyl radical in a nonpolar solvent, cyclopentane, which is the system closest to our reaction that has been measured experimentally.<sup>42</sup> The reason is the higher barrier to the reaction predicted by higher-level methods, consistent with the fact that higher-level methods predict a less exothermic

**Table 5.** Dual-Level Rate Constants at 298 K for the Rabs2 Reaction (in  $M^{-1} s^{-1}$ )

|                   | $k^{CVT/SCT}$ |
|-------------------|---------------|
| Level 0           | $1.5(+05)^a$  |
| Level 1///Level 0 | $2.6(+03)$    |
| Level 2///Level 0 | $4.1(+02)$    |
| Level 3///Level 0 | $1.7(+02)$    |
| Level 4///Level 0 | $6.9(+01)$    |
| Level 5///Level 0 | $6.3(+02)$    |
| Exp.              | $2.6(+06)^b$  |
|                   | $2.0(+05)^c$  |

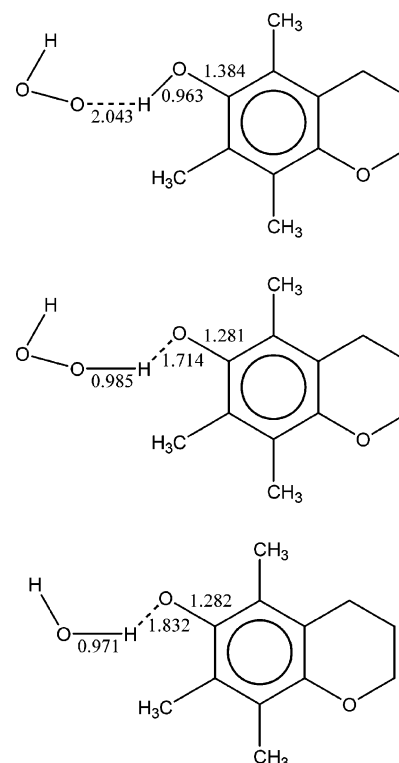
<sup>a</sup>  $1.5(+05)$  stands for  $1.5 \times 10^{+05}$ . <sup>b</sup> Reference 42, in cyclopentane solvent. <sup>c</sup> Reference 43, in water solvent.

**Figure 3.** Arrhenius plot of the rate constant for the  $\alpha$ -tocopherol +  $\cdot$ OOH  $\rightarrow$   $\alpha$ -tocopheryl +  $H_2O_2$  reaction at Level 0.

reaction. The modest Level 0 presents the best agreement with experiment, which is consistent with the behavior of this level in reproducing the experimental enthalpy of reaction at 298 K. However, the rate constant at 298 K is still lower than experiment, which leads us to think that the Level 0 barrier height is still slightly overestimated. However, since the system studied is different from the experimentally studied systems, it is not possible to know to which extent the Level 0 results are erroneous. In fact, changing the solvent leads to a decrease of the rate constant ( $2.0 \times 10^{-5} M^{-1} s^{-1}$  in water,<sup>43</sup> for example). Therefore, we shall accept the Level 0 results as our best prediction for the gas-phase reaction between  $\alpha$ -tocopherol and the hydroperoxy radical without further corrections to the theoretical method.

It is interesting to note that, although increasing the calculation level reduces the rate constant for Rabs2, no method predicts the Radd2 reaction to be competitive with Rabs2. Thus, as an example, the Level 1///Level 0 dual-level rate constants for Radd2 are 7 orders of magnitude lower than the Rabs2 rate constants calculated using the same level. Therefore, our conclusions hold no matter which level of calculation we use.

In brief, our best predictions for the overall rate constants are those given in Table 3 for the Rabs2 reaction, since Level 0 provides the best agreement with the available experimental measurements of the exothermicity and rate constants. These rate constants are shown in an Arrhenius plot in Figure 3. The curvature of the Arrhenius plot is an indication of the importance of tunneling. In fact at 298 K, about 80% of the

**Figure 4.** Geometry of the reactant well and product well for the  $\alpha$ -TOH +  $\cdot$ OOH reaction and the product well for the  $\alpha$ -TOH + OH reaction (distances in Å).

reaction takes place by tunneling, rather than by classical over-the-barrier processes. A three-term Arrhenius fit to these data gave the expression

$$k(T) = 4.5 \times 10^{-2} T^{2.75} \exp(213/T)$$

for the rate constant in  $M^{-1} s^{-1}$  in the interval 250–600 K. The activation energy can be obtained from the total rate constants through the usual definition

$$E_a(T) = R \frac{d(\ln k)}{d\left\{\frac{1}{T}\right\}}$$

with  $R$  being the gas constant, which is equivalent to determining the slope of the Arrhenius plot. At 298 K, our best estimate is 1.8 kcal/mol, very similar to our best estimate of 2.1 kcal/mol for the CoQ+OOH reaction.<sup>17</sup>

**Intermediate Complexes.** When following the Rabs2 reaction path at Level 0 starting from the saddle point and going downhill to reactants and products, we found that the reaction does not proceed directly but through the formation of two complexes. The first, that we will denote the *reactant well*, is a complex where the hydrogen bonded to the phenolic oxygen is weakly bonded to the oxygen in the hydroperoxy radical and is located on the reactant side of the reaction path (before the saddle point). The second, that will be denoted the *product well*, appears on the product side of the reaction path and shows that the hydrogen of the  $H_2O_2$  product is bonded to the oxygen of the  $\alpha$ -tocopheryl radical. The two wells are depicted in Figure 4. The energies and enthalpies of the reactant and product wells (measured with respect to reactants and products, respectively) are listed in Table 6.

**Table 6.** Energies and Enthalpies (at 298 K, in kcal mol<sup>-1</sup>) of the Reactant Well and Product Well, Measured with Respect to Reactants and Products, Respectively

|         | reactant well |            | product well |            |
|---------|---------------|------------|--------------|------------|
|         | $\Delta E$    | $\Delta H$ | $\Delta E$   | $\Delta H$ |
| Level 0 | -4.6          | -2.5       | -11.7        | -9.2       |
| Level 1 | -0.3          | +1.8       | -7.3         | -4.8       |
| Level 2 | -3.3          | -1.2       | -9.2         | -6.7       |

We have to note that the reactant complex is weakly bound, and its stability decreases as we increase the calculation level. This can be taken as an indication of the artificial stability afforded by the larger basis set superposition error of Level 0. Thus, the complex disappears when we include the thermal effects in using Level 1 energies, and its enthalpy stability is reduced to 1.3 kcal/mol with respect to reactants when we use Level 2 energies. Therefore, we can expect this well to have little influence on the mechanism of the reaction. However, the product complex is significantly more stable than the products. Even if we assume that Level 0 overestimates its stability due to the basis set superposition error, the complex is at least about 5 kcal/mol more stable than the products. Therefore, it might have some effect on the mechanism of the reaction.

In fact, the high stability of the product well can suggest that, after the reaction takes place, the H<sub>2</sub>O<sub>2</sub> product remains bonded to the  $\alpha$ -tocopheryl radical for a certain length of time. This is a positive effect for the cell. The partial protection provided by the H<sub>2</sub>O<sub>2</sub> product to the most reactive center of the  $\alpha$ -tocopheryl radical avoids or at least reduces its further attack on lipids in the membrane where it is anchored. In sum, this effect decreases the pro-oxidant action of the newly formed  $\alpha$ -tocopheryl radical.

To check the stability of this complex, we also calculated the energy of a similar complex between water and the  $\alpha$ -tocopheryl radical (Figure 4). Using Level 0 energies, this water- $\alpha$ -tocopheryl complex is 10.1 kcal/mol more stable than water and  $\alpha$ -tocopheryl at infinite separation, and about 1.6 kcal/mol less stable than the complex between H<sub>2</sub>O<sub>2</sub> and  $\alpha$ -tocopheryl. Therefore, the H<sub>2</sub>O<sub>2</sub> complex will have a longer life, hindering further reactions of the resulting radical better than cytosolic water. Thus, the presence of this complex could have a beneficial effect in avoiding the undesired pro-oxidant effect of  $\alpha$ -tocopherol.

#### 4. Concluding Remarks

There has been considerable experimental effort directed at understanding the role of antioxidants in biological processes and cell survival, in contrast with the paucity of theoretical studies. As a continuation of our previous studies on CoQ, we here studied the kinetics and dynamics of this antioxidant process using as a prototype reaction that of  $\alpha$ -tocopherol with the hydroperoxy radical in the gas phase.

The hydroperoxy radical can attack  $\alpha$ -tocopherol by two different pathways. We found that the most favorable mechanism is the hydrogen abstraction reaction from the phenolic hydrogen. This mechanism is favored by a lower transition state as a consequence of the fact that aromaticity

is conserved along the reaction path, as well as by a higher entropy of the transition state and a large probability of tunneling below the classical barrier. We found that this hydrogen abstraction reaction on  $\alpha$ -tocopherol is responsible for the overall rate constant,  $1.5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ , with nearly 80% of the reactivity at 298 K being given by quantum-mechanical tunneling.

These results for  $\alpha$ -TOH agree with the behavior observed for CoQ, where the rate constant at 298 K is  $5.3 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ , and the tunneling contribution is about 50%. This similarity of the rate constants indicates that the two natural radical-trapping antioxidants provide similar antioxidant protection to the cell.

The most favorable abstraction reaction proceeds through two intermediates, one in the reactant channel with little or no influence on the dynamics of the reaction and another in the product channel. The high stabilization of the latter complex, formed by the H<sub>2</sub>O<sub>2</sub> product bonded to the resulting  $\alpha$ -tocopheryl radical, could hinder further reactions of the radical that can cause the observed pro-oxidant effects of  $\alpha$ -tocopherol.

**Acknowledgment.** This work was partially supported by the Junta de Extremadura (Projects No. 2PR01A002 and 2PR03A075) and by the Dirección General de Investigación Científica y Técnica of Spain (Project BQU2003-04448)

**Supporting Information Available:** Cartesian coordinates, energies, and vibrational frequencies for the optimized stationary points computed at Level 0. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### References

- (1) (a) Niki, E. *Free Radical Res.* **2000**, *33*, 693–704. (b) Pryor, W. A. *Free Radical Biol. Med.* **2000**, *28*, 141–164.
- (2) Burton, G. W.; Ingold, K. U. *Acc. Chem. Res.* **1986**, *19*, 194–201.
- (3) Wang, X.; Quinn, P. J. *Prog. Lipid Res.* **1999**, *38*, 309–336.
- (4) Bowry, V. W.; Ingold, K. E. *Acc. Chem. Res.* **1999**, *32*, 27–34.
- (5) Niki, E. *Chem. Phys. Lipids* **1987**, *44*, 227–253.
- (6) For a recent review see Niki, E.; Nogichu, N. *Acc. Chem. Res.* **2004**, *37*, 45–51, and references therein.
- (7) Stocker, R. *Trends Biochem. Sci.* **1999**, *24*, 219–223.
- (8) Ernster, L.; Dallner, G. *Biochim. Biophys. Acta* **1995**, *1271*, 195–204.
- (9) Forsmark-Andrée, P.; Dallner, G.; Ernster, L. *Free Radical Biol. Med.* **1995**, *19*, 749–757.
- (10) Wolf, G. *Nutr. Rev.* **1997**, *55*, 376–378.
- (11) Barclay, L. R. C. *Can. J. Chem.* **1993**, *71*, 1–16.
- (12) Castle, L.; Perkins, M. J. *J. Am. Chem. Soc.* **1986**, *108*, 6381–6382.
- (13) Pryor, W. A.; Stickland, T.; Church, D. F. *J. Am. Chem. Soc.* **1988**, *110*, 2224–2229.

- (14) Valgimigli, L.; Banks, J. T.; Lusztyk, J.; Ingold, K. U. *J. Org. Chem.* **1999**, *64*, 3381–3383.
- (15) Bowry, V. W.; Stocker, R.; Walling, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 45–49.
- (16) Espinosa-García, J.; Gutierrez-Merino, C. *J. Phys. Chem. A* **2003**, *107*, 9712–9723.
- (17) Espinosa-García, J. *J. Am. Chem. Soc.* **2004**, *126*, 920–927.
- (18) Foti, M.; Ingold, K. U.; Lusztyk, J. *J. Am. Chem. Soc.* **1994**, *116*, 9440–9447.
- (19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratman, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuk, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, J. J.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; González, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; GAUSSIAN98 program, Revision A.7, Gaussian Inc., Pittsburgh, PA 1998.
- (20) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (21) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (22) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (23) Fast, P. L.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 3721–3729.
- (24) Chuang, Y.-Y.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 242–247.
- (25) See, for example, Truhlar, D. G.; Gordon, M. S. *Science* **1990**, *249*, 491–498, and references therein.
- (26) Corchado, J. C.; Coitiño, E. L.; Chuang, Y.-Y.; Fast, P. L.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 2424–2438.
- (27) For a recent review on these methods see Truhlar, D. G.; Gao, J.; García-Viloca, M.; Alhambra, C.; Corchado, J. C.; Sánchez, M. L.; Poulsen, T. D. *Int. J. Quantum Chem.* **2004**, *100*, 1136–1152.
- (28) Corchado, J. C.; Chuang, Y.-Y.; Fast, P. L.; Villà, J.; Hu, W.-P.; Liu, Y.-P.; Lynch, G. C.; Nguyen, K. A.; Jackels, C. F.; Melissas, V. S.; Lynch, B. J.; Rossi, I.; Coitiño, E. L.; Fernandez-Ramos, A.; Pu, J.; Albu, T. V.; Steckler, R.; Garrett, B. C.; Isaacson, A. D.; Truhlar, D. G. POLYRATE-version 9.0, University of Minnesota, Minneapolis, 2002.
- (29) Corchado, J. C.; Coitiño, E. L.; Chuang, Y.-Y.; Truhlar, D. G. GAUSSRATE-version 9.0, University of Minnesota, Minneapolis, 2002.
- (30) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811–4815.
- (31) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio Molecular Orbital Theory*, Wiley: New York, 1987.
- (32) Fast, P. L.; Sánchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407–410.
- (33) Chuang, Y.-Y.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 1140–1149.
- (34) Corchado, J. C.; Espinosa-García, J.; Hu, W.-P.; Rossi, I.; Truhlar, D. G. *J. Phys. Chem.* **1995**, *99*, 687–694.
- (35) As an example, Wayner et al. (ref 36) list an O–H bond dissociation energy for  $\alpha$ -tocopherol of 77.3 kcal/mol, while Berkowitz et al. (ref 37) give a C–H bond dissociation energy in toluene of 88.5 kcal/mol.
- (36) Wayner, D. D. M.; Lusztyk, E.; Ingold, K. U.; Mulder, P. J. *Org. Chem.* **1996**, *61*, 6430–6433.
- (37) Berkowitz, J.; Ellison, G. B.; Gutman, D. *J. Phys. Chem.* **1994**, *98*, 2744–2765.
- (38) Note that the transmission coefficient  $\kappa$  is not only a tunneling correction but also takes into account three effects: tunneling, which tends to make  $\kappa > 1$ , nonclassical reflection of trajectories that classically would overcome the barrier, which tends to make  $\kappa < 1$ ; and a better treatment of the reaction threshold than the canonical variational transition state theory, which also makes  $\kappa < 1$ . The net effect in the abstraction reaction is  $\kappa > 1$ , while in the addition reaction it is  $\kappa < 1$ . For a detailed description of the calculation of transmission coefficients and threshold corrections see ref 39.
- (39) Truhlar, D. G.; Isaacson, A. D.; Garrett, B. C. In *Theory of Chemical Reaction Dynamics*, Vol. 4, Baer, M., Ed. CRC Press: Boca Raton, FL, 1985; pp 65–137.
- (40) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd edition. Wiley-VCH: Weinheim (Germany), 2000.
- (41) Shum, L. G. S.; Benson, S. W. *Int. J. Chem. Kinet.* **1983**, *15*, 323–339.
- (42) Burton, G. W.; Doba, T.; Gabe, E. J.; Hughes, L.; Lee, F. L.; Prasad, L.; Ingold, K. U. *J. Am. Chem. Soc.* **1985**, *107*, 7053–7065.
- (43) Arudi, R. L.; Sutherland, M. W.; Bielski, B. H. J. *Oxy Radicals and Their Scavenger Systems*, Vol. 1, Cohen, G.; Greenwald, R. A. Eds. Elsevier: Amsterdam (The Netherlands), 1983.

CT0498932