JCTC Journal of Chemical Theory and Computation

Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria

William L. Jorgensen* and Laura L. Thomas

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Received January 8, 2008

Background

The calculation of free energy changes is fundamental for the thermodynamic characterization of reaction pathways and chemical equilibria. The conventional approaches, free energy perturbation (FEP) and thermodynamic integration (TI), were predicated on Kirkwood's introduction of the continuous coupling parameter in his integral equation studies of fluid systems.[1,2] Later work by Zwanzig further advanced the mathematical formalism of free energy perturbation theory,[3] and this, in conjunction with molecular dynamics and Monte Carlo sampling techniques, has become a principal approach for calculating free-energy differences.[2,4,5] Some applications have included computation of relative free energies of solvation, relative pK_a values, medium-effects on conformational equilibria, host−guest binding affinities, and free-energy surfaces for organic and biochemical reactions.

Free energy perturbation theory, as presented by Zwanzig in his 1954 paper,[3] relates the free energy difference between an initial (reference) and a final (target) state of a system to an average of a function of their energy difference evaluated by sampling for the initial state (eq 1).

$$\Delta F = F_1 - F_0 = -kT \ln\left\langle \exp\left[\frac{-(E_1 - E_0)}{kT}\right]\right\rangle_0 \qquad (1)$$

Zwanzig further went on to derive an expression for the free energy change as a power series. Defining the energy of the target state, E_1, as the sum of the energy of a reference state, E_0, and a small perturbing potential, V (eq 2), eq 3 was obtained by expanding eq 1 to second order.

$$E_1 = E_0 + V \qquad (2)$$

$$\Delta F = F_1 - F_0 = \langle V\rangle - \frac{1}{2kT}\left(\langle V^2\rangle - \langle V\rangle^2\right) \qquad (3)$$

Here, the free energy difference between the reference and target states is estimated to first order by the reference ensemble average of the perturbing potential and to second order by including the average squared deviation of the perturbing potential from its mean, or the variance.[2,3] This approximation for calculating free energy differences has been recognized as especially relevant for systems with fluctuations governed by a Gaussian probability distribution,[2] and, for example, related protocols have been applied to the charging of ions in solution,[6] the estimation of intrinsic pK_as and pK_a shifts in proteins,[7] and the calculation of solvent contributions to the excited states of tryptophan.[8]

Landau, Lifshitz, and Peierls.[9–12] Perturbation theory in its various forms has long been of interest to the theoretical physics community.[2] Specifically, a section on "Thermodynamic Perturbation Theory" appeared in the 1951 Russian edition of Statistical Physics by Landau and Lifshitz,[13] the fifth volume in their encyclopedic Course of Theoretical Physics. This section begins with a definition of the partition function and its expansion in powers of V to second-order as given by eq 4 (their eq 32.2). Here F is the target free energy (F_1 above) and analogously $V = E - E_0$.

$$\exp[-F/kT] = \int \exp[-(E_0(p,q) + V(p,q))/kT]\ \mathrm{d}\Gamma \cong \int \exp[-E_0/kT]\left(1 - \frac{V}{kT} + \frac{V^2}{2(kT)^2}\right)\mathrm{d}\Gamma \qquad (4)$$

Then "taking the logarithm and again expanding in series to the same approximation" they obtain eq 5 (their eq 32.3). Eq 5 is identical to eq 3 using the equality, $\overline{V^2} - \overline{V}^2 = \overline{(V - \overline{V})^2}$, which they note.

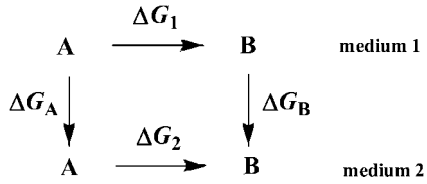
$$F = F_0 + \overline{V} - \frac{1}{2kT}\overline{(V - \overline{V})^2} \qquad (5)$$

While few citations appear in the 1951 edition of the book, due in part to Landau's preference for his own derivations, a rare attribution to Landau's colleague, R. E. Peierls, appears on p 93 at the beginning of the section on "Thermodynamic Perturbation Theory" in the 1958 English edition: "it is possible to develop a kind of 'perturbation theory' to

* Corresponding author e-mail: william.jorgensen@yale.edu.

10.1021/ct800011m CCC: $40.75
Published on Web 05/09/2008

Scheme 1



calculate thermodynamic quantities (R. E. Peierls, 1933)".[14] Except for the "R. E. Peierls, 1933", the 1951 Russian and 1958 English versions of this section are essentially identical. The year was changed from 1933 to 1932 in later editions.[15]

This citation likely refers to Sir Rudolf E. Peierls' paper "*On the Theory of the Diamagnetism of Conduction Electrons*",[16] submitted in 1932 and published in the original German in 1933. In this work, Peierls investigates the diamagnetic susceptibility of free electrons, and at the outset he provides an outline of thermodynamic perturbation theory. The paper includes a perturbative expansion of the partition function that encompasses eq 4 (eq 7 in ref 16). The connection between the material in Peierls's 1933 paper and the section on "Thermodynamic Perturbation Theory" in the book by Landau and Lifshitz may have been strengthened by interactions between Landau and Peierls, as both were affiliated with Wolfgang Pauli's laboratory around 1930; Landau was a visiting student, and Peierls was Pauli's assistant at the ETH during 1929−1932. It is also possible that the "R. E. Peierls, 1933" reference was added by the translators for the 1958 English edition, E. Peierls and R. F. Peierls.[14] Although their full appellations do not appear in the text, presumably the former is R. E. Peierls's wife, Eugenia, who was also a classmate and friend of Landau's at the Leningrad Physicotechnical Institute, and the latter is the Peierls' son, Ronald, who received a Ph.D. in Physics from Cornell University in 1959. R. E. Peierls was a visiting lecturer in Leningrad in 1931, worked on the Manhattan Project in the early 1940s, and eventually became the Wykeman Professor of Theoretical Physics at Oxford during 1963−1974.[12] His contributions to the development of thermodynamic perturbation theory appear to be underappreciated; it would be reasonable to refer to eq 5 and equivalents such as eq 3 as the Peierls equation.

Sampling Methods. For applications of eq 1 to chemical equilibria, normally, states 0 and 1 are different molecules, A and B, and a change in medium is investigated by comparing the free energy change for the conversion of A to B in two environments.[5] Thus, a thermodynamic cycle is considered as in Scheme 1, where ΔG_A and ΔG_B are the free energies of transfer of A and B from medium 1 to medium 2. ΔG_1 and ΔG_2 are computed, and the medium effect is given by $\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_B - \Delta G_A$. Gibbs free energies are shown here, which correspond to standard experimental conditions and to use of the isothermal−isobaric ensemble in the statistical mechanics simulations. The energetics for A and B can be represented by molecular mechanics (MM) or quantum mechanics (QM).[17] The former has been the most used as it allows the treatment of large systems and the creation and deletion of atoms. When MM is used for the solutes, the solute−solvent and solvent−solvent interactions are also represented with MM. With MM, raw energy changes between two different molecules, even isomers, are normally not meaningful by themselves, and it is essential to perform two perturbations, such as ΔG_1 and ΔG_2, to cancel out artifacts of the arbitrary energy reference with force fields and conceptual omissions such as resonance effects.

The first FEP computations of the type represented by Scheme 1 were not carried out for a molecular system until 1985. The initial application was the computation of the relative free energies of hydration of ethane (A) and methanol (B) such that medium 1 was the gas phase and medium 2 was water; the aqueous FEP calculations were carried out in a periodic cube containing 125 TIP4P water molecules.[5] This paper also introduced a straightforward scheme for the conversion of one molecule to another, which later became known as single-topology FEP calculations.[2] Specifically, a coupling parameter λ_i was used to linearly scale all geometrical and force-field parameters χ from those for A to those for B (eq 6). Owing to poor convergence for

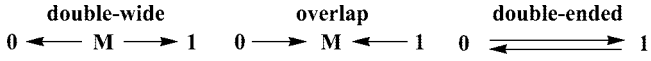
$$\chi_i = \lambda_i\chi_B + (1 - \lambda_i)\chi_A \tag{6}$$

eq 1 when the initial and final states are too different, it is necessary to carry out a series of simulations for multiple values of λ_i spanning between A ($\lambda_i = 0$) and B ($\lambda_i = 1$). The individual simulations are called "windows" or "stages", and choices have to be made for the individual λ_i for a particular FEP series. The computed free-energy changes are increasingly precise as more windows are used; however, the total cost of a FEP calculation increases linearly with the number of windows. The resulting computational demands were sufficient that calculations like the ethane to methanol example were not attempted before 1985, and they only became routine with the advent of multiprocessor computer systems.

In order to minimize the computational demands, the 1985 paper also introduced "double-wide" sampling.[5] In principle, the FEP calculations could be performed in the forward or reverse directions. E.g., if a constant $\Delta\lambda = 0.25$ were used, four windows could be executed to perturb $\lambda = 0 \rightarrow 0.25 \rightarrow 0.50 \rightarrow 0.75 \rightarrow 1.0$ or in reverse, $\lambda = 1 \rightarrow 0.75 \rightarrow 0.50 \rightarrow 0.25 \rightarrow 0.0$. This is sometimes referred to as "direct sampling". However, it was recognized that only two simulations were actually required at $\lambda = 0.25$ and 0.75, if forward and reverse perturbations were evaluated simultaneously using "double-wide" sampling, $0 \leftarrow 0.25 \rightarrow 0.50$ and $0.50 \leftarrow 0.75 \rightarrow 1.0$. If one does the FEP calculations on-the-fly using all configurations in the averaging rather than postprocessing saved configurations, more bookkeeping is required. Thus, in the *BOSS* program,[18] coordinates for the reference (e.g., $\lambda = 0.25$) and two perturbed solutes (e.g., $\lambda = 0.0$ and 0.50) are maintained to evaluate both free-energy changes on-the-fly. Furthermore, it is expected that with use of a small number of windows, systematic errors arise such that the cumulative free-energy change for the forward FEP series is not the same as the negative of the result from the reverse series. This was shown to be the case for the ethane to methanol example and that double-wide sampling provides a more accurate estimate of the true free-energy change.

Alternatively, "double-ended sampling"[19] can be employed, which refers to performing both the $\lambda_i \rightarrow \lambda_j$ perturbation and its reverse, $\lambda_j \rightarrow \lambda_i$, and taking $\Delta G(\lambda_i \rightarrow \lambda_j) = (\Delta G(\lambda_i \rightarrow \lambda_j) - \Delta G(\lambda_j \rightarrow \lambda_i))/2$; however, this again requires twice as many simulations as double-wide sampling and has no obvious advantages for accuracy.

Scheme 2



Overlap Sampling. A variant of these sampling choices is known as "overlap sampling". It is similar to double-ended sampling but to an intermediate point *M* rather than the end points 0 and 1 (Scheme 2).[2,20,21] From the Zwanzig expression (eq 1), the free-energy change for $0 \rightarrow 1$ from overlap sampling (OS) is then given by eq 7, while double-wide (DW) and double-ended (DE) sampling yield eqs 8 and 9.

$$\Delta G^{OS} = -kT \ln[\langle \exp(-(E_M - E_0)/kT)\rangle_0 / \langle \exp(-(E_M - E_1)/kT)\rangle_1] \quad (7)$$

$$\Delta G^{DW} = kT \ln[\langle \exp(-(E_0 - E_M)/kT)\rangle_M / \langle \exp(-(E_1 - E_M)/kT)\rangle_M] \quad (8)$$

$$\Delta G^{DE} = -{}^1\!/_2 kT \ln[\langle \exp(-(E_1 - E_0)/kT)\rangle_0 / \langle \exp(-(E_0 - E_1)/kT)\rangle_1] \quad (9)$$

There appears to be little difference between eqs 7 and 8 except that the sampling is performed at *M* for DW and at the end points for OS and that only one simulation is required for DW versus two for OS. However, DW requires a geometrical definition of point *M* at, e.g., $\lambda = 0.5$, while *M* can be more conceptual for OS. Specifically, *M* can be defined by a parametric expression such as eq 10, and λ_M could be chosen to minimize the statistical noise (random

$$E_M = \lambda_M E_1 + (1 - \lambda_M) E_0 \quad (10)$$

error) in the results from application of eq 7. It is desirable for the energy distributions for *M* and the end points to overlap as much as possible, so a reasonable choice is $\lambda_M = 0.5$ and then $E_M = (E_0 + E_1)/2$. With this assignment, eq 7 becomes eq 11.

$$\Delta G^{OS} = -kT \ln[\langle \exp(-(E_1 - E_0)/2kT)\rangle_0 / \langle \exp(-(E_0 - E_1)/2kT)\rangle_1] \quad (11)$$

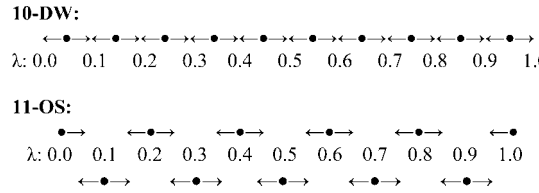
The advantage of eq 11 over DE (eq 9) is clear since the energy differences inside the averages are now divided by two. The advantage over eq 8 depends on the geometrical choice of *M* for the DW calculations; it is likely that the usual, convenient choice of $\lambda = 0.5$ is less optimal than the choice of $\lambda_M = 0.5$ in eq 10. Indeed, Lu et al. refer to use of eq 11 as "simple overlap sampling"[20,21] and claims have been made that it is "usually markedly more accurate than simple sampling",[2] i.e., direct averaging, that DW is just a "time-saving implementation of direct averaging",[21] and that "the direct averaging method can thus be seen as not reliable".[21] Though convergence of eq 11 (OS) is expected to be better than for eq 9 (DE), the case versus DW is less clear, particularly when a similar number of simulations is run. This has not been well tested previously, so results for a large set of comparisons are reported here. Specifically, FEP calculations have been performed with DW and OS alternatives for 30 conversions of substituted benzenes (PhX $\rightarrow$ PhY) in the gas phase and in water, which yield relative free energies of hydration via Scheme 1.[5] These are fundamental conversions that are used routinely, for example, in FEP-guided design of enzyme inhibitors,[22] so establishment of optimal protocols is particularly important. In addition, the performance of the expansion given by the Peierls equation has also been tested to first and second order.

Table 1. Increments $\Delta\lambda$ Between Simulations[a]

method	$\Delta\lambda$	method	$\Delta\lambda$
10–DW	0.10	6–OS	0.20
14–DW[b]	0.10	11–OS	0.10
25–DW	0.04	15–OS[b]	0.10
5–OS	0.25	21–OS	0.05

[a] *n*-DW and *n*-OS are double-wide and overlap sampling using *n* simulations. For DW and OS, the first simulations occur at $\lambda = \Delta\lambda/2$ and $\lambda = 0$, respectively, and the spacing between simulations is $\Delta\lambda$. [b] $\Delta\lambda = 0.05$ is used for $\lambda = 0$ to 0.20 and $\lambda = 0.80$ to 1.0.

Scheme 3



Computational Details

The substituted benzenes were represented by the OPLS-AA force field[23] and the water molecules by the TIP4P model.[24] Metropolis Monte Carlo statistical mechanics (MC) calculations were performed for a single solute in a periodic cube with 500 water molecules at 25 °C and 1 atm in the isothermal−isobaric ensemble using the *BOSS* program.[18] The water−water cutoff was at 10 Å based on the OO distance, and the solute−water interactions were included if any non-hydrogen atom of the solute was within 10 Å of the water O. The interactions were quadratically smoothed to zero within 0.5 Å of the cutoff. For each FEP window, a MC simulation was performed with equilibration for 3 million configurations and averaging for a subsequent 10 million configurations; this requires ca. 35 min on a 3.0 GHz Pentium D processor. Our experience is that runs of this length are necessary to allow adequate sampling for different hydrogen-bonded states for a solute. For example, alcohols typically participate in one to three hydrogen bonds with water molecules, and these states do not rapidly interconvert. If the MC or molecular dynamics run is not long enough to sample well the different states, then significant errors can arise in computed free-energy changes independent of the sampling details, e.g., DW or OS.

The TIP4P water molecules underwent only rigid-body translations and rotations, while the sampling of the solutes

included all internal degrees of freedom as well as the total translations and rotations. Solute and volume moves were attempted every 120 and 3125 configurations, respectively. Acceptance rates of 30−50% for new configurations were obtained by use of ranges for translations and rotations of ±0.06 Å and ±6° for the solutes and ±0.15 Å and ±15° for the water molecules. For the gas-phase MC calculations, each window consisted of an equilibration phase of 0.15 million configurations followed by averaging over an additional 1 million configurations. Mutations involving different numbers of atoms for PhX and PhY were performed using dummy atoms that have no nonbonded parameters (Lennard-Jones and Coulomb); the r_0 values for bond lengths to dummy atoms are taken as 0.3 Å. The reported uncertainties in the averages ($\pm 1\sigma$) for the free energy changes were obtained from the fluctuation in separate averages over batches of 2 million configurations.[4b,25] Eq 12 is used where m is the number of batches, θ_i is the average of property θ for the ith batch, and $<\theta>$ is the overall average for θ. These are all standard procedures that have been followed in prior computations of relative free energies of hydration.[26] It may be noted that there are no "end-state problems when $\lambda = 0$ or 1"[20] with the MC simulations since calculations of forces are not required.

$$\sigma^2 = \sum_i^m (\theta_i - \langle\theta\rangle)^2/m(m-1) \qquad (12)$$

Eight FEP protocols were examined as summarized in Table 1. DW calculations were executed using 10, 14, and 25 windows. For the 10-window calculations, the first window is run at $\lambda = 0.05$ with perturbations to $\lambda = 0.0$ and 0.10; the second window is at $\lambda = 0.15$, etc. For the 25-window case, the first simulation is run at $\lambda = 0.02$, and the spacing is also even with $\Delta\lambda = 0.04$. The 14-window alternative has become a standard one in our inhibitor design work;[22] this is the same as the 10-window case except that the spacing is reduced to $\Delta\lambda = 0.05$ for $\lambda = 0$ to 0.2 and 0.8 to 1.0. I.e., simulations are run at $\lambda = 0.025$, 0.075, 0.125, 0.175, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.825, 0.875, 0.925, and 0.975. This choice reflects the fact that the free-energy changes are most rapid near the end points as atoms are created or annihilated or atomic charges are changed when eq 6 is applied. OS calculations were executed using eq 11 and 5, 6, 11, 15, and 21 windows. The λ-spacing was uniform for all cases except 15-OS, which is the same as 11-OS with the spacing halved for $\lambda = 0$ to 0.2 and 0.8 to 1.0 as for 14-DW. For example, the simulations for 11-OS were run at $\lambda = 0.0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0, and two free-energy increments were computed at each λ as for DW (Scheme 3). In each case, the individual free-energy changes are summed to yield the total result.

An example of the distributions of total solute−water interaction energies from such a series of simulations is presented in Figure 1. There is clearly substantial overlap in the distributions for simulations at adjacent λ values even for a case like the illustrated one where the hydrogen-bonding abilities of the limiting solutes are significantly different. The illustrated example is, in fact, the worst case for the perturbations in Table 1 in that the average solute−water interaction energy shifts the most from $\lambda = 0$ to $\lambda = 1$. For cases that do not involve large differences in hydrogen bonding, the centers of the distributions for $\lambda = 0$ and $\lambda = 1$ are only shifted by a few kcal/mol.

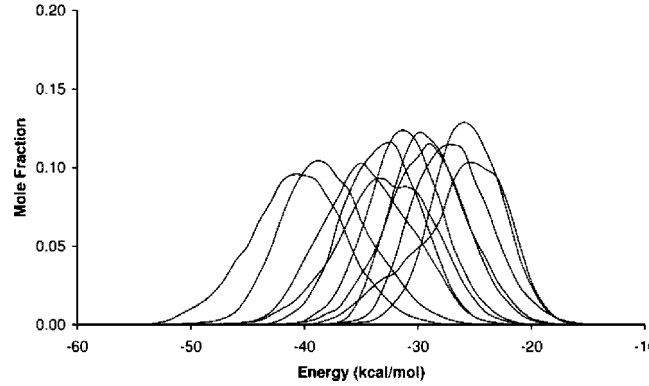


Figure 1. Distributions for the total solute−water interaction energies for the 11 windows in the 11-OS calculations for the PhCOMe → $PhCONH_2$ conversion. The distributions progress from that for $\lambda = 0.0$ (acetophenone) on the far right to $\lambda = 1.0$ (benzamide) on the far left. Units for the *y*-axis are mole fraction per kcal/mol.

FEP Results

FEP calculations were executed for the 28 PhX → PhY changes listed in Table 2. The perturbations involve common substituents, and no more than one non-hydrogen atom is deleted or created in each case. Two forward and backward cases are considered for PhCl ⟷ PhH and PhOH ⟷ PhOMe; the results are not expected to be identical in magnitude because in the present protocol the simulations were executed sequentially starting from $\lambda = 0$ (PhX) and the last configuration from a completed window is used as the starting point for equilibration of the next window. The results in Table 2 are the differences in free energies of hydration for PhX and PhY, $\Delta\Delta G_{hyd}$, which are computed from the difference in the FEP results in water and in the gas phase, ΔG(aq) − ΔG(gas). There is essentially no statistical uncertainty in the gas-phase results; the uncertainties discussed below arise predominantly from the simulations in water.

Two degenerate cases were also studied in depth; PhCl was converted to itself by replacing the chlorine at C1 with hydrogen and replacing the hydrogen at C4 with chlorine. Similarly, phenol was converted to itself by analogous shrinking and growing of the hydroxyl group. The exact result is obviously $\Delta G = 0$ in these degenerate cases. The results for ΔG(aq) are listed in Table 3 along with the statistical uncertainty ($\pm 1\sigma$) for the ΔG(aq) as computed from eq 12. Each protocol was executed at least twice with small changes to the ranges for solute translations and rotations to yield independent estimates of the ΔG values. As first indications of the precision of the alternative protocols, the average absolute value of the ΔG results in Table 3 are listed in Table 4. The average statistical uncertainty for each method from the 28 results in Table 2 is also recorded in Table 4. Some salient points follow. (1)

Table 2. Computed and Experimental Differences in Free Energies of Hydration (kcal/mol) for Substituted Benzenes[a]

X →	Y	10-DW	14-DW	25-DW	6-OS	11-OS	15-OS	21-OS	exptl[b]
Br	Cl	0.60	0.61	0.57	0.63	0.59	0.59	0.57	0.33
CH_3	CF_3	1.20	0.99	1.45	1.35	1.27	1.36	1.33	0.64
Cl	F	−0.13	−0.18	−0.10	0.00	0.00	−0.01	−0.12	0.31
Cl	H	−0.38	−0.20	−0.41	−0.77	−0.57	−0.58	−0.41	0.26
H	Cl	0.53	0.32	0.31	0.65	0.48	0.23	0.39	−0.26
CN	F	2.70	2.68	2.66	2.56	2.43	2.77	2.52	3.41
COMe	$CONH_2$	−7.58	−6.45	−6.72	−7.31	−6.53	−6.42	−6.55	−6.42
COMe	NO_2	0.78	0.80	0.48	0.89	0.72	0.66	0.95	0.46
Et	Me	−0.42	−0.50	−0.35	−0.35	−0.27	−0.32	−0.29	−0.10
Et	OMe	−0.42	−0.58	−0.58	−0.67	−0.84	−0.18	−0.51	−1.67
F	H	−0.23	−0.26	−0.40	−0.40	−0.21	−0.34	−0.36	−0.05
i-Pr	Et	−0.36	−0.13	−0.27	−0.28	−0.38	−0.59	−0.12	−0.49
Me	Cl	1.03	0.37	0.63	0.69	0.69	0.95	0.86	−0.23
Me	H	0.42	0.48	0.34	0.45	0.23	0.44	0.22	0.03
Me	NH_2	−4.08	−4.48	−4.28	−4.09	−4.40	−4.32	−4.32	−4.60
$MeC{=}CH_2$	COMe	−4.51	−3.55	−4.28	−4.16	−4.17	−3.64	−4.19	−3.34
NH_2	OH	−0.54	−0.57	−0.48	−0.41	−0.18	−0.70	−0.38	−1.13
NHMe	NH_2	−0.52	−0.92	−0.99	−0.91	−1.25	−1.12	−0.88	−0.80
NMe_2	NHMe	−1.80	−0.85	−0.95	−0.70	−1.10	−1.22	−0.99	−1.24
NMe_2	NO_2	0.44	−0.06	0.46	0.33	0.77	0.32	0.13	−0.67
OH	H	5.45	5.22	5.35	5.20	5.03	5.29	4.97	5.76
OH	F	5.22	5.66	5.35	4.89	5.51	5.38	5.54	5.81
OH	OMe	4.41	4.36	4.82	4.55	4.46	4.70	4.29	4.16
OMe	OH	−4.67	−4.15	−4.70	−4.46	−4.35	−4.77	−4.77	−4.16
Pr	Et	−0.18	−0.65	−0.30	−0.43	−0.41	−0.48	−0.64	−0.49
SH	OH	−5.03	−5.22	−4.93	−5.06	−5.10	−5.20	−5.18	−4.07
SMe	OMe	0.30	0.32	0.05	0.03	0.03	0.07	0.09	0.27
Pr	CH_2OMe	−2.39	−2.44	−2.48	−2.44	−2.41	−2.64	−2.44	

[a] Results for PhX → PhY; *n*-DW and *n*-SOS are FEP results using double-wide sampling and overlap sampling with *n* simulations. [b] Reference 27.

Table 3. Computed Free Energy Changes and Their Uncertainties[a] in Water for Degenerate Mutations (kcal/mol)

	PhCl → ClPh		PhOH → HOPh	
	ΔG	σ	ΔG	σ
10-DW-1	0.11	0.09	−0.86	0.15
10-DW-2	−0.24	0.07	0.38	0.16
14-DW-1	0.13	0.08	−0.29	0.16
14-DW-2	−0.06	0.09	−0.11	0.14
25-DW-1	0.04	0.05	−0.40	0.09
25-DW-2	0.12	0.05	−0.04	0.10
5-OS-1	0.20	0.16	1.12	0.23
5-OS-2	−0.35	0.13	−0.44	0.25
6-OS-1	−0.16	0.10	−0.56	0.27
6-OS-2	−0.22	0.12	0.12	0.24
11-OS-1	−0.07	0.09	−0.18	0.20
11-OS-2	0.01	0.09	−0.27	0.21
11-OS-3	0.19	0.10	0.31	0.18
11-OS-4	−0.06	0.08	−0.56	0.24
15-OS-1	−0.32	0.09	0.11	0.15
15-OS-2	−0.03	0.09	0.04	0.16
21-OS-1	0.13	0.14	−0.15	0.07
21-OS-2	0.00	0.07	0.27	0.11

[a] From eq 12.

Table 4. Average Absolute ΔG for the Degenerate Mutations in Table 3 and Statistical Uncertainties for the 28 FEP Calculations in Table 2 (kcal/mol)

method	ΔG	σ	method	ΔG	σ
10-DW	0.40	0.12	6-OS	0.27	0.22
14-DW	0.15	0.12	11-OS	0.21	0.13
25-DW	0.15	0.07	15-OS	0.13	0.11
5-OS	0.53	0.25	21-OS	0.14	0.09

Table 5. Average Absolute and RMS Deviations for Computed and Experimental Free Energy Changes (kcal/mol)[a]

	10-DW	14-DW	25-DW	6-OS	11-OS	15-OS	21-OS	exptl
10-DW	0	0.30	**0.21**	0.21	0.27	0.26	**0.24**	0.58
14-DW	0.43	0	**0.23**	0.25	0.24	0.23	**0.18**	0.39
25-DW	**0.30**	**0.29**	**0**	**0.16**	**0.18**	**0.17**	**0.15**	**0.51**
6-OS	0.28	0.34	**0.23**	0	0.18	0.23	**0.18**	0.58
11-OS	0.35	0.30	**0.20**	0.27	0	0.20	**0.15**	0.50
15-OS	0.36	0.28	**0.22**	0.32	0.27	0	**0.18**	0.48
21-OS	**0.33**	**0.24**	**0.21**	**0.26**	**0.22**	**0.24**	**0**	**0.51**
exptl	0.67	0.50	**0.61**	0.68	0.64	0.61	**0.63**	0

[a] Using the data in Table 2. Average absolute deviations in upper triangle, rms deviations in lower triangle. Deviations with 25-DW and 21-OS are in bold.

The trends in the results are reasonable with the random error decreasing with the number of windows for both DW and OS. (2) The errors are larger for phenol since the removal/growth of a hydrogen-bonding group is a larger perturbation than interconversion of chlorine and hydrogen. (3) The uncertainty based on the fluctuations in the averages, σ, is somewhat less than the uncertainties indicated by the deviations of the ΔG values in Tables 3 and 4 from zero. (4) There is little difference in the quality of results from 14-DW, 15-OS, 21-OS, and 25-DW. The average error of ca. 0.15 kcal/mol in Table 4 appears to reflect the limit for the simulations of the present length, independent of the number of windows. (5) The errors with 5-OS and 10-DW are likely too large for these procedures to be generally useful, while 6-OS warrants further consideration.

Turning back to Table 2, the $\Delta\Delta G_{hyd}$ results are provided for all methods except 5-OS. Table 5 summarizes the results by providing the average unsigned and rms deviations

Table 6. Average Absolute and RMS Deviations between Computed Free Energy Changes (kcal/mol)[a]

	6-OS-1	6-OS-2	11-OS-1	11-OS-2	11-OS-3	11-OS-4	25-DW	21-OS
6-OS-1	0	0.30	0.18	0.18	0.23	0.20	0.16	0.18
6-OS-2	0.36	0	0.26	0.30	0.30	0.24	0.26	0.27
11-OS-1	0.27	0.32	0	0.16	0.18	0.17	0.18	0.15
11-OS-2	0.26	0.35	0.22	0	0.16	0.19	0.17	0.13
11-OS-3	0.31	0.37	0.23	0.20	0	0.18	0.16	0.14
11-OS-4	0.29	0.30	0.22	0.22	0.24	0	0.16	0.15
25-DW	0.23	0.32	0.20	0.21	0.19	0.22	0	0.15
21-OS	0.26	0.33	0.22	0.16	0.18	0.18	0.21	0

[a] *m*-OS-*n* refers to *n* independent *m*-OS evaluations for the 28 PhX → PhY FEP calculations in Table 2. Average absolute deviations in upper triangle, rms deviations in lower triangle.

between the seven sets of computed results and the experimental data. Agreement with experiment is not important for the present methodological testing, though it is noted that the average error of 0.5 kcal/mol for $\Delta\Delta G_{hyd}$ is typical with OPLS-AA[26] and it was also found previously in an extensive study by Price and Brooks on the performance of OPLS-AA for the properties of liquid substituted benzenes and their free energies of hydration.[28] Again, there are many points to highlight. (1) The average deviation between the 21-OS and 25-DW results is the smallest, 0.15 kcal/mol, which reinforces this figure as the precision limit for the present methods. The smallest rms deviation for 21-OS is also with 25-DW, 0.21 kcal/mol. Based on the results in Tables 2, 3, 4, and 5, 21-OS and 25-DW can be assumed to be the most accurate procedures. (2) The average deviations between 21-OS and 25-DW and the other procedures (10-DW, 14-DW, 6-OS, 11-OS, 15-OS) are all below 0.25 kcal/mol. The rms deviations are all below 0.30 kcal/mol except with 10-DW, which appears to be the least accurate of these alternatives. (3) The results in the two cases where the PhX to PhY and PhY to PhX FEP calculations were performed, chlorobenzene to benzene and phenol to anisole, are also near the precision limit with average discrepancies of 0.1−0.2 kcal/mol. (4) The largest single discrepancies with the 21-OS results are 1.03 (10-DW), 0.63 (14-DW), 0.53 (25-DW), 0.76 (6-OS), 0.65 (11-OS), and 0.55 kcal/mol (15-OS). The perturbations that show the largest oscillations in the results are acetophenone to benzamide, α-methylstyrene to acetophenone, and *N,N*-dimethylaniline to *N*-methylaniline; all involve substantial changes in hydrogen-bonding character and deletion of atoms. (5) The average deviations between the results with 5-OS (not shown) and 25-DW and 21-OS are 0.29 and 0.31 kcal/mol, and the rms deviations are 0.42 and 0.46 kcal/mol. The largest discrepancy with the 21-OS result is now 1.67 kcal/mol for the phenol to anisole conversion. Thus, 5-OS is reinforced as unacceptable for such FEP calculations. 10-DW is the next least acceptable. 6-OS performs notably well, while 14-DW, 11-OS, and 15-OS are comparably more reliable.

As a further check for consistency, the full set of 28 FEP calculations was repeated with the 11-OS procedure three more times with changes to the ranges for solute moves. The average unsigned and rms deviations for the four 11-OS data sets and the 25-DW and 21-OS results are presented in Table 6. The consistency is excellent with all rms values being close to 0.20 kcal/mol. The 6-OS calculations were also repeated; the rms discrepancies for this second set of the 28 results increased to 0.32 and 0.33 kcal/mol in comparison to the results from 25-DW and 21-OS. Overall, in our view, 11-OS, 14-DW, and 15-OS are all viable procedures for routine FEP calculations of relevance to molecular design. The difference in performance between DW and OS is insignificant for the present typical FEP calculations with the number of windows in this range. Furthermore, the slightly greater accuracy obtainable with 21-OS or 25-DW does not justify the increased costs, while 6-OS may be useful for preliminary exploration of large numbers of mutations. An 8-OS alternative with the first and last window divided in half was also tried for the 28 perturbations and yielded rms discrepancies of 0.29 and 0.33 kcal/mol with the 25-DW and 21-OS results.

It may be noted that the present limit of precision of about 0.2 kcal/mol with 11-OS, 14-DW, and 15-OS is adequate in many situations. E.g., for computation of relative free energies of binding for two inhibitors, FEP calculations are required for the inhibitors by themselves in water and bound to the biomolecule. The total uncertainty would then be $(0.2^2 + 0.2^2)^{1/2} = 0.3$ kcal/mol corresponding to a factor of 1.6 in inhibition constants, which is at the limit of typical experimental assays. If improved precision is needed, then the length of the simulation for each window should be increased with the usual expectation that a factor of 2 improvement in precision requires at least a factor of 4 increase in sampling.[2,4]

Results with the Peierls Equation

Application of eq 3 was also considered, though there is no practical advantage over use of the exact eq 1 since the MC or MD simulations still need to be run and the additional cost of averaging the quantities in either equation is insignificant. However, the performance of the expansion is of fundamental interest and has not been examined previously for a series of typical A → B FEP calculations. Thus, for each window in the 10-DW and 14-DW calculations, the energy changes between the reference and perturbed states and their squares were also averaged. The resulting differences between the aqueous- and gas-phase calculations are summarized in Table 7 for the 14-DW case; the comparisons are with the $\Delta\Delta G_{hyd}$ values calculated using 14-DW and eq 1.

The results from just the first-order term, the average energy changes, are very good with an average unsigned deviation of only 0.15 kcal/mol to the free-energy results from eq 1. However, in the case of phenol to anisole and its reverse the favorability of the hydration of phenol is

Table 7. Differences in Free Energies of Hydration (kcal/mol) for Substituted Benzenes Using Eq 1 and Its Expansion in Eq 3[a]

X →	Y	ΔE	$\Delta\Delta G$ (eq 3)	$\Delta\Delta G$ (eq 1)
Br	Cl	0.61	0.61	0.61
CH_3	CF_3	1.00	0.98	0.99
Cl	F	−0.17	−0.16	−0.18
Cl	H	−0.22	−0.16	−0.20
H	Cl	0.34	0.28	0.32
CN	F	2.13	3.77	2.68
COMe	$CONH_2$	−6.46	−6.42	−6.45
COMe	NO_2	0.77	0.86	0.80
Et	Me	−0.59	−0.41	−0.50
Et	OMe	−0.61	−0.56	−0.58
F	H	−0.27	−0.26	−0.26
i-Pr	Et	−0.21	−0.05	−0.13
Me	Cl	0.35	0.38	0.37
Me	H	0.40	0.58	0.48
Me	NH_2	−4.53	−4.42	−4.48
$MeC{=}CH_2$	COMe	−3.57	−3.55	−3.55
NH_2	OH	−0.57	−0.58	−0.57
NHMe	NH_2	−1.33	−0.06	−0.92
NMe_2	NHMe	−1.34	0.85	−0.85
NMe_2	NO_2	−0.15	0.09	−0.06
OH	H	5.20	5.22	5.22
OH	F	5.65	5.65	5.66
OH	OMe	5.40	3.94	4.36
OMe	OH	−5.22	−1.57	−4.15
Pr	Et	−0.73	−0.57	−0.65
SH	OH	−5.21	−5.20	−5.22
SMe	OMe	0.32	0.33	0.32
Pr	CH_2OMe	−2.46	−2.42	−2.44
	aad[b]	0.15	0.27	0
	rms	0.30	0.63	0

[a] Results for PhX → PhY from the MC simulations with 14 windows. $\Delta\Delta G$ (eq 1) is the exact result using eq 1, while $\Delta E = \langle V \rangle$ in eq 3. [b] Average absolute deviation.

underestimated by 1.1 kcal/mol. The results get significantly worse when the full second-order treatment (eq 3) is used owing to poor convergence of the fluctuation term. The latter is basically a $T\Delta C_V$ correction that only changes for successful solute moves. Thus, the convergence is expected to be significantly worse than for the computation of a pure liquid's heat capacity, which is well-known to be slow.[4b,29] There are now more cases with serious problems including PhCN → PhF, $PhNMe_2$ → PhNHMe, and PhOMe → PhOH; and, the rms deviation rises from 0.30 kcal/mol with the first-order estimate to 0.63 kcal/mol including both terms in eq 3. The situation further deteriorates when only 10 windows are used with the full eq 3; e.g., a 10 kcal/mol error was obtained for the $PhNMe_2$ → PhNHMe case, though the average error with only the first-order term remains acceptable at 0.19 kcal/mol. On the other hand, if changes that involve OH, NHR, and CN groups are excluded, it is apparent from Table 7 that there is negligible difference between the results from the Zwanzig equation (eq 1) and those from the first- or second-order expansion in the Peierls equation. The greater ranges of solute−water interactions for strongly hydrogen-bonding solutes are problematic for application of eq 3.

Conclusion

An overview has been provided on the history of free-energy perturbation (FEP) calculations and their application to computation of relative free energies of solvation. In view of the expanding role of FEP calculations for pursuing a wide range of problems in chemistry and biochemistry including molecular design, it is important to establish reliable protocols for their execution. For 30 interconversions of substituted benzenes, which represent prototypical cases for molecular design, alternative sampling procedures and numbers of FEP windows were considered. Simple overlap sampling with 11 or 15 windows and double-wide sampling with 14 windows emerged as viable procedures for such routine FEP calculations; the present protocols with these sampling methods yield statistical uncertainties of 0.2 kcal/mol for free-energy differences. The slightly greater accuracy obtainable with 21 to 25 windows does not clearly justify the increased costs, while overlap sampling with 6 or 8 windows may be useful for preliminary exploration of large numbers of molecular changes. Application of the power-series expansion of the Zwanzig equation was also tested. Use of the first-order term is generally reliable, while inclusion of the slowly convergent, second-order fluctuation term causes deterioration in the results for strongly hydrogen-bonded solutes.

Acknowledgment. Gratitude is expressed to the National Science Foundation (CHE-0446920) and the National Institutes of Health (GM032136) for support of this work and to Drs. A. Alexandrova and B. Albrecht for assistance with translation of text.

References

(1) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.

(2) For a recent review, see: Chipot, C.; Pohorille, A. In *Springer Series in Chemical Physics, Vol. 86: Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C.,Pohorille, A., Eds.; Springer-Verlag: Berlin, 2007; pp 33−75.

(3) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(4) (a) Jorgensen, W. L. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; Wiley: New York, 1998; Vol. 2, pp 1061−1070. (b) Jorgensen, W. L. In Encyclopedia of Computational Chemistry; Schleyer, P. v. R., Allinger, N. L., Clark, T., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; Wiley: New York, 1998; Vol. 2, pp 1754−1763. (c) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184–189.

(5) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050–3054.

(6) Levy, R. M.; Belhadj, M.; Kitchen, D. B. *J. Chem. Phys.* **1991**, *95*, 3627–3633.

(7) (a) Del Buono, G. S.; Figueirido, F. E.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 85–97. (b) Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.

(8) Simonson, T.; Wong, C. F.; Brünger, A. T. *J. Phys. Chem. A* **1997**, *101*, 1935–1945.

(9) For biographies of Landau and Peierls, see ref. 10−12.

(10) Livanova, A. *Landau: A Great Physicist and Teacher*, translated by Sykes, J. B.; Pergamon Press: Elmsford, NY, 1980.

(11) Khalatnikov, I. M. *Landau: The Physicist and the Man*, translated by Sykes, J. B.; Pergamon Press: Elmsford, NY, 1989.

(12) Peierls, R. *Bird of Passage: Recollections of a Physicist*; Princeton University Press: Princeton, NJ, 1985.

(13) Landau, L. D.; Lifshitz, E. M. *Statisticheskaia fizika, klassicheskaia I kvantovaia*; Gos. izd. tekhniko-teoret. litry.; Moskva, 1951; pp 112−115.

(14) Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics, Vol. 5: Statistical Physics*; translated from Russian by Peierls, E., Peierls, R. F.; Pergamon Press: London, 1958; pp 93−95.

(15) Landau, L. D.; Lifshitz, E. M., *Course of Theoretical Physics, Vol. 5: Statistical Physics*; translated from Russian by Sykes, J. B., Kearsley, M. J.; Pergamon Press: Elmsford, NY, 1969; pp 90−93.

(16) (a) Peierls, R. Z. *Phys.* **1933**, *80*, 763–791. (b) *Selected Scientific Papers of Sir Rudolf Peierls: With Commentary (Series in 20th Century Physics)*; translated from German by Dalitz, R. H.; Peierls, R. E., Ed.; World Scientific Publishing: Singapore, 1997.

(17) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **205**, *102*, 6665–6670.

(18) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689–1700.

(19) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(20) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Comput. Chem.* **2004**, *25*, 28–39.

(21) Lu, N.; Woolf, T. B. In *Springer Series in Chemical Physics, Vol 86: Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer-Verlag: Berlin, 2007; pp 199−247.

(22) (a) Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 663–667. (b) Thakur, V. V.; Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5664–5667. (c) Barreiro, G.; Kim, J. T.; Guimaraes, C. R. W.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. *J. Med. Chem.* **2007**, *47*, 2416–2428.

(23) (a) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236. (b) Jorgensen, W. L.; Laird, E. R.; Nguyen, T. B.; Tirado-Rives, J. *J. Comput. Chem.* **1993**, *14*, 206–215. (c) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. *J. Phys. Chem. B* **2004**, *108*, 16264–16270.

(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(25) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Clarendon: Oxford, 1987; pp 191−198

(26) (a) Jorgensen, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993**, *14*, 195–205. (b) Udier-Blagovic, M.; Morales de Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 1322–1332.

(27) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.

(28) Price, D. J.; Brooks, C. L., III *J. Comput. Chem.* **2005**, *26*, 1529–1541.

(29) Jorgensen, W. L. *Chem. Phys. Lett.* **1982**, *92*, 405–410.

CT800011M

JCTC Journal of Chemical Theory and Computation

Perspective on Foundations of Solvation Modeling: The Electrostatic Contribution to the Free Energy of Solvation

Aleksandr V. Marenich, Christopher J. Cramer,* and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received January 29, 2008

1. Introduction

Solvation effects are essential components of all liquid-state chemistry, and it is impossible to understand liquid-phase organic, biological, or inorganic chemistry without including them. The Nobel-Prize-winning gas-phase quantum mechanical electronic structure methods of Pople[1] and Kohn et al.[2] require the additional inclusion of solvent for reliably addressing problems in liquid-phase chemistry. Methods that include the solvent implicitly are especially powerful because they allow one to retain the minimal representation of the solute, thereby facilitating progress with quantum mechanical calculations at the same high levels as those used in the gas phase,[3,4] and because they allow one to model the solvent with the correct bulk permittivity. Reliable calculations of solutes in solution must take account of electrostatics, cavitation, dispersion, and solvent structure, but solvation effects are frequently dominated by electrostatics. Therefore, achieving a solid understanding of electrostatic solvation effects is an excellent starting point for understanding solvation and improving solvation models.

Thermodynamically, there is no unique way to separate electrostatic from nonelectrostatic contributions to the free energy of solvation;[5] only their sum is a state function.[6] Nevertheless, one might try to *define* the electrostatic contribution in a way that most practitioners would find reasonable and definitive and to use that as a starting point for physical analysis. It is already known, however, that various successful solvation models make this partition differently.[7] This is sometimes attributed to their use of different parameters, especially the atomic radii. If the atomic radii were the only issue, one might try to define them in some unambiguous way, for example, from radial distribution functions,[8] and thereby establish a useful working definition of the electrostatic contribution to equilibrium solvation effects in terms of the static relative permittivity and systematically defined atomic radii.

The first subject of the present contribution is to examine popular implicit solvation models with a view to understanding if the above procedure has merit. After that, we draw some general conclusions about solvation modeling.

In section 2, we summarize the relevant theoretical background. In section 3, we summarize five methods for estimating the electrostatic contribution to the free energy of solvation, and in section 4, we compare calculations (especially designed for this perspective) carried out with these methods. Section 5 contains discussion, and section 6 gives the principal conclusions.

2. Theory

In the electrostatic theory of dielectric media (i.e., nonconducting media), the medium is associated with a relative permittivity ϵ, which is a scalar constant for isotropic homogeneous media and a scalar function of position for isotropic nonhomogeneous media. The charges of the medium (called bound charges) do not appear explicitly, and the charge density per unit volume of explicitly treated objects in the medium is called the free charge density ρ_f. In a linear isotropic homogeneous medium, the scalar electric potential Φ satisfies Poisson's equation:

$$\varepsilon\nabla^2\Phi = -4\pi\,\rho_f \tag{1}$$

For a nonhomogeneous medium, ε depends on position. For such a medium eq 1 is replaced by[9]

$$\nabla\cdot(\varepsilon\nabla\Phi) = -4\pi\,\rho_f \tag{2}$$

which will be called the nonhomogeneous Poisson equation (NPE). In the application considered in this article, ρ_f is the solute charge density. Implicit solvation models (sometimes called continuum solvation models or implicit or continuum

* Corresponding author e-mail: cramer@umn.edu (C.J.C.); truhlar@umn.edu (D.G.T.).

10.1021/ct800029c CCC: $40.75
Published on Web 05/08/2008

solvent models) represent a solvated molecule at an atomic level of detail inside a molecule-sized (and usually molecule-shaped) cavity surrounded by a dielectric medium that represents the solvent. The relative permittivity is equated to the bulk static solvent value ϵ_s outside the cavity and to a smaller value inside the cavity. The smaller value is usually taken as unity (the relative permittivity of empty space), which is appropriate if the solute polarization is treated explicitly. Although models that treat the solute by unpolarized molecular mechanics are sometimes used, the rest of this article is concerned with models that treat the solute as polarizable and quantum mechanical. One then solves the NPE coupled to the quantal description of the solute to obtain the electric potential due to the polarized dielectric continuum and the polarization of the solute; this potential, called the reaction field φ, equals the total potential Φ minus the electrostatic potential[10] $\Phi^{(0)}$ of the gas-phase molecule.

From the reaction field, one calculates the free-energy change corresponding to the solvation process. If we approximate the solute as rigid, the electrostatic contribution to the free energy of solvation is given by[11]

$$\Delta G_{EP} = \langle \Psi | H^{(0)} - \frac{e}{2}\varphi | \Psi \rangle + \frac{e}{2}\sum_k Z_k \varphi_k - \langle \Psi^{(0)} | H^{(0)} | \Psi^{(0)} \rangle \tag{3}$$

where e is the atomic unit of charge, φ_k is the reaction field evaluated at atom k, Z_k is the atomic number of atom k, $H^{(0)}$ and $\Psi^{(0)}$ are the solute electronic Hamiltonian and electronic wave function, respectively, in the gas phase, and Ψ is the polarized solute electronic wave function in solution. This equation includes the polarization of the solvent by the solute and the distortion of the solute that is induced by this polarization effect. If the charge density (the sum of the electronic charge density $e|\Psi|^2$ and the nuclear charges) is represented by a set of nuclear-centered partial charges q_k, then

$$-\langle \Psi | e\varphi | \Psi \rangle + e\sum_k Z_k \varphi_k = \sum_k q_k \varphi_k \tag{4}$$

which is always negative (favorable to the solvation process) if the model is accurate.

A key issue in all implicit solvation models is the boundary between the solute cavity where $\epsilon < \epsilon_s$ and the solvent continuum where $\epsilon = \epsilon_s$. In modern models, the boundary is defined to enclose a superposition of atom-centered spheres with radii ρ_k, which are called intrinsic Coulomb radii. The boundary that encloses this superposition of spheres may be defined to precisely enclose these spheres and nothing else,[12,13] which is called the solvent-accessible surface (SAS), or it may also enclose the regions not touched by a certain-size ball (whose radius is called the solvent radius or probe radius) rolling on the SAS;[14] the latter is called the solvent-excluding surface or Connolly's molecular surface. The intrinsic Coulomb radii may be the van der Waals radii of the atoms, scaled van der Waals radii, van der Waals radii augmented by a solvent radius, or freely adjustable parameters. (Some models use spheres, ellipsoids, or isodensity surfaces of the solute electron density as the boundary, but such cavity boundaries are either unrealistic or numerically troublesome and will not be considered here.)

Other implicit solvent models do not start with the NPE, and it is instructive to classify implicit solvation models into two groups, those that solve the NPE numerically, for example, by finite differences[15] or by boundary element methods based on integral equations,[16] and those that employ a starting point based on Coulomb's law, for example, single-center multipolar expansions[17,18] or the generalized Born (GB) approximation.[19–27] The Coulomb's law models represent the solute as a collection of point charges[19–27] (the distributed monopole approximation), almost always located at the nuclei, or as a set of point multipoles at the center of the molecule.[17,18] Most NPE solvers use the continuous electron density of the solute, without approximating it by charges or multipoles. One may label such solvers as density-based NPE solvers. However, at least one NPE solver, which is called the Poisson−Boltzmann solver (PBS),[28] uses an iterative two-step approach in which the first step of each iteration (the calculation of an apparent surface charge that represents the field due to the polarized dielectric medium) uses a distributed monopole representation $\{q_k\}$ of the solute charge density, and the second step (the calculation of the interaction of the apparent surface charge with the solute electron density) uses the continuous density distribution of the solute.

In practice the NPE solvers approximate the reaction field at an arbitrary position $\mathbf{r}_K$ within the cavity as

$$\varphi_K = \sum_m \frac{q_m}{|\mathbf{r}_K - \mathbf{r}_m|} \tag{5}$$

where $\mathbf{r}_m$ is the coordinate of an element m of surface area on the solute−solvent boundary (these elements are called tessarae), and q_m is the apparent surface charge on element m. In contrast, the GB approximation is equivalent to approximating the reaction field distribution as

$$\varphi_k = \sum_{k'} \frac{q_{k'}}{|\mathbf{r}_k - \mathbf{r}_{k'}|} f_{kk'} \tag{6}$$

where $\mathbf{r}_k$ and $\mathbf{r}_{k'}$ are evaluated only at atomic positions, and $f_{kk'}$ is a function to be specified. The value of a single term in eq 6 (excluding $q_{k'}$) is called a Coulomb integral. The most successful function $f_{kk'}$ for approximating the Coulomb integrals is the dielectric descreening approximation,[26] which yields

$$f_{kk'} = -\left(1 - \frac{1}{\varepsilon_s}\right)\frac{r_{kk'}}{\sqrt{r_{kk'}^2 + \alpha_k \alpha_{k'} \exp(-r_{kk'}^2/d\alpha_k\alpha_{k'})}} \tag{7}$$

where

$$r_{kk'} \equiv |\mathbf{r}_k - \mathbf{r}_{k'}| \tag{8}$$

and d is a parameter and α_k is the shielded atomic radius of atom k; α_k represents an appropriately weighted average distance of atom k from the solvent. The number of elements m in eq 5 is in principle increased to convergence, whereas the number of terms k' in eq 6 is equal to the number of atoms in the solute. A relation connecting eqs 5–8 is that[4,29]

$$\sum_m q_m = -\left(1-\frac{1}{\varepsilon_s}\right)\sum_{k'} q_{k'} \quad (9)$$

In sections 3 and 4, we consider calculations with various NPE solvers and compare them to GB calculations. (We do not include calculations with one-center multipole expansions among our test calculations in this article because for large molecules this method is slowly convergent with respect to the number of multipole moments.)

3. Computational Methods

The NPE coupled to a quantum simulation of a rigid solute was solved using four implicit solvent models implemented in four popular quantum chemical program packages:

(1) The Integral Equation Formalism Polarizable Continuum Model of *Gaussian 03*,[30] namely, IEFPCM[29,31–33]
(2) The conductor-like PCM model in the *GAMESS* computer package,[34–36] namely, C-PCM[37–43]
(3) The PBS algorithm[28,44] in the *Jaguar* computer package[45]
(4) The Generalized Conductor-like Screening Model (GCOSMO) as implemented in *NWChem*.[46]

There are various ways to implement a conductor-like model, and the various later implementations[4,38,40,47–50] should not be confused with the original[37] COSMO method. In this article, GCOSMO refers to the default implementation of the conductor-like method in the NWChem, version 4.7, computer package.[46] This implementation uses the GCOSMO dielectric screening factor[38] for the conductor-like apparent surface charge.

Except for intrinsic Coulomb radii, solvent radii, and ϵ_s, which are specified in section 4, all NPE calculations were carried out with the default numerical parameters and cavity definitions of the respective programs. We varied tesserae surface areas from 0.01 to 0.4 Å^2 and found no significant dependence of the solvation free energy on this parameter (see the Supporting Information for more details). In the case of large intrinsic Coulomb radii (>5 Å) with IEFPCM/*Gaussian 03*, in order to improve the convergence of the self-consistent reaction field, we set the average area of the tesserae generated on each sphere in the cavity surface to 0.4 Å^2 (instead of the default value of 0.2 Å^2).

The GB calculations were carried out with a locally modified version[51] of *Gaussian 03*.[30] The locally modified version is based on algorithms described previously.[13,52–55] The solute is again considered to be rigid. The Coulomb integrals in the electrostatic polarization formula are approximated with the dielectric descreening approximation of eq 7, with the parameter d set equal to 3.7, which is the value used in the SM6[56] and SM8[57] solvation models. The shielded Coulomb radius of atom k is computed by a three-dimensional integration over the entire region outside the atomic sphere defined by the intrinsic Coulomb radius ρ_k of that atom.[13,26] (A pairwise approximation that has been used in some past work[58] is not used in the present article.)

In all five methods, the calculations are iterated to self-consistency of the solute electron density with the reaction field due to the solvent, and reported electrostatic solvation energies include the reorganization costs of distorting the solute and polarizing the solvent. (For four of the methods, the solute distortion cost is broken out in tables presented in the Supporting Information.)

In the example calculations presented in this article, in all five methods, the solute cavity is determined from the same intrinsic Coulomb radii ρ_k that depend only on the atomic numbers Z_k of the atoms; these parameters are called R_Z.

All solvation free energies, ΔG_S, in this paper are the electrostatic contribution only. That is, the cavity, dispersion, solvent structure, repulsion, and liberation components are not included because our goal in this perspective is to compare alternative approaches to the electrostatic contribution. We note that to better illustrate the physics we present results for solute cavities of various sizes, some quite different from those that would be produced using the default atomic radii for specific models. We adopt this strategy because examining the behavior of the models over a large range of parameter space, including intentionally unphysical values, provides insight into the way they work. Analysis of the accuracy of these models with their default settings has been presented elsewhere.[7,56,57]

4. Results

Nine atomic species ($Ne^{0,\pm}$, $Na^{0,\pm}$, $Cl^{0,\pm}$) and nine molecular species ($C_2H_4^{0,\pm}$, C_2H_6, CH_3OH, $CH_3OH_2^+$, CH_3O^-, CH_2CHO, and CH_2CF_2) were studied as model systems with the solvent relative permittivity ϵ_s set equal to the aqueous value of 78.4. Electrostatic calculations were performed with the solvent radius set to zero, in which case the solvent-excluded surface reduces to the SAS. The solute was treated by Hartree−Fock theory with the 6-31G(d)[59] basis set. In the case of open-shell electronic systems ($Ne^{\pm}$, Na, Cl, and $C_2H_4^{\pm}$) we employed either restricted open-shell Hartree−Fock (GCOSMO/*NWChem*) or spin-unrestricted Hartree−Fock wave functions (for all other models). We use the Hartree−Fock method in calculations of electrostatic energies of solvation because it is available in all four of the tested programs. We also tested the M05-2X density functional[60] available in *Gaussian 03*,[30] *Jaguar*,[45] and *NWChem*.[46] The results obtained by density functional theory were nearly identical to those obtained by Hartree−Fock theory (see the Supporting Information for such a comparison). Electrostatic contributions to the free energies of solvation of molecular species were computed at gas-phase geometries optimized with the M06-2X density functional[60] and the 6-31+G(d,p) basis set[59] except that, for the $C_2H_4^{\pm}$ calculations, we used the M06-2X/6-31+G(d,p) geometry of the neutral molecule.

Tables 1–3 give results for atoms and monatomic ions. In each case, we considered several sets of atomic radii. For atoms and monatomic ions, the GB result reduces to the original Born formula:[61]

$$\Delta G_{EP} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon_s}\right)\frac{q_k^2}{R_Z} \quad (10)$$

which is the exact classical result for the case where the solute is a conducting sphere of radius R_Z. Tables 4–6 give results for molecules and molecular ions. In these cases, we

Table 1. Electrostatic Free Energies of Solvation for Ne, Ne^+, and Ne^-

	ΔG_S (kcal/mol)				
R_{Ne} (Å)	GB	C-PCM	GCOSMO	IEFPCM	PBS
			Ne		
0.85	0.0	−3.8	0.0	−3.8	0.0
1.00	0.0	−0.5	0.0	−0.5	0.0
1.15	0.0	−0.1	0.0	−0.1	0.0
≥1.30	0.0	0.0	0.0	0.0	0.0
			Ne^+		
0.85	−192.8	−231.7	−213.0	−231.3	−215.0
1.00	−163.9	−177.4	−171.6	−177.2	−173.1
1.15	−142.5	−147.2	−145.4	−147.2	−146.7
1.30	−126.1	−127.7	−127.2	−127.6	−128.5
1.54[a]	−106.4	−106.8	−106.8	−106.7	−107.9
1.848[b]	−88.7	−88.8	−88.8	−88.8	−89.9
2.00	−82.0	−82.0	−82.0	−82.0	−83.0
4.00	−41.0	−41.0	−41.0	−41.0	−41.5
10.00	−16.4	−16.4	−16.4	−16.4	−16.6
			Ne^-		
0.85	−192.8	−107.9	−148.0	−110.5	−138.6
1.00	−163.9	−129.8	−151.1	−134.3	−144.2
1.15	−142.5	−130.6	−140.8	−134.7	−136.6
1.30	−126.1	−122.5	−127.2	−125.2	−125.3
1.54[a]	−106.4	−106.0	−107.6	−107.3	−107.5
1.848[b]	−88.7	−88.7	−89.3	−89.2	−89.8
2.00	−82.0	−82.0	−82.3	−82.3	−83.0
4.00	−41.0	−41.0	−41.0	−41.0	−41.5
10.00	−16.4	−16.4	−16.4	−16.4	−16.6

[a] van der Waals radius according to Bondi. [b] 1.2 × Bondi's radius.

considered several sets of radii obtained by scaling the radii used in the SM6[56] and SM8[57] aqueous solvation models by the formula

$$R_Z = MR_Z^{SM8} \tag{11}$$

where M is a scaling factor. The other two sets of R_Z values correspond to using the van der Waals radii of Bondi[62] and the same Bondi radii scaled by a factor of 1.2. We chose 1.2 × Bondi because previous recommendations based on COSMO-RS calculations suggested that the optimal radii are about 1.17−1.2 times the Bondi radii.[63,64] The last set of R_Z values is to set all R_Z = 10 Å.

In studying the tables, it is instructive to have an idea of what radii would be most physical for practical calculations. The optimum scaling factor for the radii depends to some extent on how they are optimized, but the following comments might be useful. The scale factor of 1.0 is by definition optimal for the GB algorithm when optimized against the SM6 aqueous training set.[56] Optimizing M with IEFPCM against all 2821 data points in the SM8 aqueous and nonaqueous training set[57] yields M = 1.08. If, instead of scaling the aqueous SM8 radii (which are the same as the SM6 radii), we scale the Bondi radii and optimize the scale factor against the 2821 data points in the SM8 aqueous and nonaqueous training set, the optimum is found to be 1.08 × Bondi. (In each of these optimizations, we simultaneously optimized a set of SM8-like atomic surface tensions.) When one takes account of these optimizations and when one compares the SM8 radii to the Bondi radii, one finds that the physical region of the tables is between M = 1 and M = 1.3 and between the Bondi row and the 1.2 × Bondi row.

Table 2. Electrostatic Free Energies of Solvation for Na, Na^+, and Na^-

	ΔG_S (kcal/mol)				
R_{Na} (Å)	GB	C-PCM	GCOSMO	IEFPCM	PBS
			Na		
1.00	0.0	−51.1	0.0	−51.0	0.0
1.15	0.0	−31.3	0.0	−31.3	0.0
1.30	0.0	−19.4	0.0	−19.3	0.0
1.65	0.0	−6.1	0.0	−6.1	0.0
2.00	0.0	−1.8	0.0	−1.8	0.0
2.27[a]	0.0	−0.7	0.0	−0.7	0.0
2.724[b]	0.0	−0.1	0.0	−0.1	0.0
≥3.00	0.0	0.0	0.0	0.0	0.0
			Na^+		
1.00	−163.9	−170.8	−167.2	−170.6	−170.2
1.15	−142.5	−144.6	−143.5	−144.6	−145.7
1.30	−126.1	−126.7	−126.4	−126.5	−128.1
1.65	−99.3	−99.4	−99.4	−99.4	−100.7
2.00	−82.0	−82.0	−82.0	−81.9	−83.0
2.27[a]	−72.2	−72.2	−72.2	−72.2	−73.1
2.724[b]	−60.2	−60.2	−60.2	−60.2	−60.9
4.00	−41.0	−41.0	−41.0	−41.0	−41.5
10.00	−16.4	−16.4	−16.4	−16.4	−16.6
			Na^-		
1.00	−163.9	−5.1	17.7	−5.1	23.1
1.15	−142.5	0.0	−3.8	0.0	0.5
1.30	−126.1	−2.4	−18.2	−2.4	−14.6
1.65	−99.3	−17.3	−37.0	−17.3	−34.5
2.00	−82.0	−29.9	−45.1	−29.8	−43.4
2.27[a]	−72.2	−35.8	−47.5	−35.8	−46.2
2.724[b]	−60.2	−40.4	−47.6	−40.4	−46.9
3.00	−54.6	−41.2	−46.4	−41.2	−46.0
4.00	−41.0	−38.2	−39.5	−38.2	−39.7
5.00	−32.8	−32.4	−32.6	−32.4	−33.0
10.00	−16.4	−16.4	−16.4	−16.4	−16.6

[a] van der Waals radius according to Bondi. [b] 1.2 × Bondi's radius.

When M is increased to greater than ∼1.4, the cavity becomes more and more unphysical, but calculations with such M values are included to show how the algorithms respond when the cavity encloses larger and larger fractions of the solute charge density. Similarly, when M gets significantly smaller than 1.0, the cavity again becomes unrealistic, and calculations with such values of M show what happens when the amount of charge outside the cavity gets even larger than it is for realistic cavities.

5. Discussion

The tables show a number of interesting features, especially when one considers that they all correspond in some sense to the same electrostatic problem. Consider first the monatomic species in Tables 1–3. The C-PCM and IEFPCM algorithms yield very similar results for all neutral atoms and cations with all radii examined, except for very small (unphysical) radii. For large atomic radii, these two methods and the GCOSMO method agree precisely with the Born formula for ions, which is comforting; however, the PBS result is always larger by a factor of 1.0182 because *Jaguar*, for reasons unexplained in the original articles,[28,44] premultiplies all radii by 0.9821. For anions with small radii, the GCOSMO and PBS algorithms sometimes yield unphysical positive results. Even for the Bondi radii, the deviations of the various algorithms from one another are as large as 36.4

Table 3. Electrostatic Free Energies of Solvation for Cl, Cl^+, and Cl^-

	ΔG_S (kcal/mol)				
R_{Cl} (Å)	GB	C-PCM	GCOSMO	IEFPCM	PBS
			Cl		
1.00	0.0	−66.9	−8.1	−66.7	0.0
1.15	0.0	−23.2	−6.4	−23.2	0.0
1.30	0.0	−9.7	−4.9	−9.7	0.0
1.65	0.0	−2.5	−2.3	−2.6	0.0
1.75[a]	0.0	−1.9	−1.9	−1.9	0.0
2.10[b]	0.0	−0.9	−0.9	−0.9	0.0
≥4.00	0.0	0.0	0.0	0.0	0.0
			Cl^+		
1.00	−163.9	−367.0	−258.3	−365.7	−235.6
1.15	−142.5	−239.2	−197.0	−239.2	−178.6
1.30	−126.1	−175.8	−158.1	−175.6	−144.8
1.65	−99.3	−112.3	−109.5	−112.3	−104.2
1.75[a]	−93.7	−102.8	−101.2	−102.8	−97.2
2.10[b]	−78.1	−81.2	−80.9	−81.2	−79.5
4.00	−41.0	−41.1	−41.1	−41.1	−41.5
10.00	−16.4	−16.4	−16.4	−16.4	−16.6
			Cl^-		
1.00	−163.9	−6.0	−29.6	−5.9	−17.1
1.15	−142.5	−36.0	−69.4	−36.0	−61.9
1.30	−126.1	−60.5	−86.1	−60.5	−81.9
1.65	−99.3	−81.2	−89.7	−81.2	−89.0
1.75[a]	−93.7	−81.5	−87.3	−81.4	−87.1
2.10[b]	−78.1	−75.3	−76.7	−75.3	−77.3
4.00	−41.0	−41.0	−41.0	−41.0	−41.5
10.00	−16.4	−16.4	−16.4	−16.4	−16.6

[a] van der Waals radius according to Bondi. [b] 1.2 × Bondi's radius.

kcal/mol, and for radii smaller than the Bondi radii, the deviations are very unsystematic. One should notice the distinction between the model problem of a cavity with a well-defined boundary, where ϵ is discontinuous and there is no solute charge outside the cavity, and a real atom with a nebulous boundary and a solute charge distribution that decreases exponentially to zero far from the nucleus rather than exactly to zero at a boundary. In some methods, the wave function has been constrained to go to zero at the cavity boundary,[65] and the radius was then variationally optimized. We do not consider such methods here, in part because the great majority of calculations in current use do not employ this constraint, and in part because this procedure does not give realistic results for anionic solutes.[65]

For molecules (Tables 4–6), the clearest trend is that, for a given set of radii, the GB method yields much smaller (less-negative) free energies of solvation than the NPE solvers for neutral molecules (as also found previously[7]), but the GB free energy of solvation is not necessarily less negative for ions. This cannot be attributed to radii, since the radii are the same, and they do not yield this kind of difference for atomic species. For Bondi radii, the various methods differ by up to 5.2 kcal/mol for neutrals, 10.2 kcal/mol for cations, and 5.5 kcal/mol for anions. Even for radii 15% larger than the SM6−SM8 radii, the deviations are up to 4.9 kcal/mol for neutrals, 8.3 kcal/mol for cations, and 4.4 kcal/mol for anions. Even with $M = 4$, some deviations remain as large as 1.2 kcal/mol.

Whereas the GB approximation must give 0 for a neutral atom, it need not give 0 for a neutral molecule. The experimental solvation free energies of C_2H_4, C_2H_6, and CH_3OH in water are[66] +1.3, +1.8, and −5.1 kcal/mol, respectively. If, just for the sake of argument, we approximate the cavity, dispersion, and solvent structure contributions as being approximately equal for C_2H_4 and C_2H_6, then ΔG_{EP} should be about 0.5 kcal/mol more negative for C_2H_4 than for C_2H_6. For $M \leq 1.30$, Tables 4 and 5 show much larger differences for the NPE solvers. Similarly, ΔG_{EP} should probably be at most 3−4 kcal/mol more negative for CH_3OH than for C_2H_6, but Table 5 shows much greater differences for the NPE solvers for $M \leq 1.15$ and for GB for $M \leq 0.85$.

A large part of the complication in comparing different approaches to the electrostatic problem must be attributed to outlying charge effects.[67] These arise from the fact that the Poisson solvers other than IEFPCM (and methods equivalent to it—see next paragraph) assume that the entire solute charge distribution lies within the solute cavity, but the electron density used in the NPE solvers has tails that lie outside the cavity. The GB algorithm and the first step of each iteration in the PBS algorithm assume nuclear-centered charges and therefore have no outlying charge, but there is a related error due to approximating the real charge distribution by distributed monopoles that lie entirely within the cavity. The Born formula itself is not valid for real atomic ions; one reason is that their charge is not confined to a finite cavity.

The IEFPCM method includes an approximate treatment of outlying charge that is equivalent[28,68] to the approximate SS(V)PE method[69] of Chipman, but it seems unlikely that this accounts for the entire difference from the GCOSMO algorithm. For example, it has been noted that numerical aspects of the solution of the Poisson equation lead to outlying charge artifacts even for nuclear-centered charges;[29] perhaps for this reason it is sometimes called "escaped charge."

The comparison of calculated solvation energies for $C_2H_4^+$ and $C_2H_4^-$ for a given set of radii is very complex. For example, for $M = 1.30$, $C_2H_4^-$ is better solvated, whereas for $M \leq 1.00$, $C_2H_4^+$ is better solvated with NPE methods, but $C_2H_4^-$ is better solvated in the GB approximation. Note that a loss of outlying charge decreases the intracavity (negative) charge of an anion and *increases* the intracavity (positive) charge of a cation, which rationalizes the behavior of the NPE solvers. In the GB case, the unit nature of the total charge is unaffected by the choice of cavity size. However, the individual partial atomic charges appearing in eq 6 need not be the same for the anion and the cation, so that different solvation free energies are possible, effectively recognizing higher moments than the molecular monopole by use of an atom-centered monopole expansion of the charge distribution.

We note that GB and NPE methods have been compared previously for the same sets of radii,[7,70,71] but some of the key issues *vis-a-vis* the GB method that are evident in Tables 1–6 were not brought out in those studies, and the comparison with more than one NPE solver also brings new insight. Furthermore, we note that we do not consider the GB approximation to be an approximation to the model problem solved by the NPE solvers (with its unrealistic or arbitrary

Table 4. Electrostatic Free Energies of Solvation for C_2H_4, $C_2H_4^+$, and $C_2H_4^-$

			ΔG_S (kcal/mol)				
M	R_H (Å)	R_C (Å)	GB	C-PCM	GCOSMO	IEFPCM	PBS
			C_2H_4				
0.85	0.87	1.33	−0.7	−5.6	−5.1	−5.6	−3.8
1.00	1.02	1.57	−0.3	−3.6	−3.7	−3.6	−3.0
1.15	1.17	1.81	−0.2	−2.3	−2.3	−2.3	−2.1
Bondi	1.20	1.70	−0.2	−2.2	−2.3	−2.2	−2.0
1.30	1.33	2.04	−0.1	−1.5	−1.7	−1.4	−1.2
1.2 × Bondi	1.44	2.04	−0.1	−1.2	−1.2	−1.2	−1.0
1.65	1.68	2.59	0.0	−0.5	−0.4	−0.5	−0.4
2.00	2.04	3.14	0.0	−0.2	−0.2	−0.2	−0.2
3.00	3.06	4.71	0.0	0.0	0.0	0.0	0.0
n.a.	10.00	10.00	0.0	0.0	0.0	0.0	0.0
			$C_2H_4^+$				
0.85	0.87	1.33	−88.7	−109.0	−101.9	−109.0	−104.8
1.00	1.02	1.57	−80.1	−88.8	−88.6	−88.8	−87.7
1.15	1.17	1.81	−73.0	−77.2	−78.1	−77.1	−77.5
Bondi	1.20	1.70	−74.3	−79.3	−81.1	−79.7	−78.9
1.30	1.33	2.04	−66.9	−69.3	−71.7	−69.3	−68.3
1.2 × Bondi	1.44	2.04	−65.6	−67.7	−69.9	−67.7	−67.8
1.65	1.68	2.59	−55.8	−56.6	−56.5	−56.7	−56.3
2.00	2.04	3.14	−47.2	−47.6	−47.9	−47.8	−48.6
3.00	3.06	4.71	−32.5	−32.7	−33.0	−32.7	−33.1
4.00	4.08	6.28	−24.8	−24.9	−24.8	−24.9	−23.7
5.00	5.10	7.85	−20.1	−20.1	−20.0	−20.1	−20.3
n.a.	10.00	10.00	−14.9	−14.9	−14.8	−14.9	
			$C_2H_4^-$				
0.85	0.87	1.33	−96.1	−74.6	−89.4	−74.5	−85.3
1.00	1.02	1.57	−84.2	−78.8	−86.5	−78.7	−80.8
1.15	1.17	1.81	−75.5	−75.5	−79.9	−75.3	−77.0
Bondi	1.20	1.70	−77.9	−77.6	−82.9	−77.6	−80.3
1.30	1.33	2.04	−68.5	−69.8	−73.0	−69.8	−69.0
1.2 × Bondi	1.44	2.04	−67.5	−69.7	−72.9	−69.7	−69.9
1.65	1.68	2.59	−56.4	−57.1	−56.8	−57.4	−56.7
2.00	2.04	3.14	−47.6	−47.9	−48.2	−48.1	−48.5
3.00	3.06	4.71	−32.6	−32.8	−33.1	−32.8	−33.1
4.00	4.08	6.28	−24.9	−24.9	−24.8	−24.9	−23.7
5.00	5.10	7.85	−20.1	−20.1	−20.0	−20.1	−20.3
n.a.	10.00	10.00	−14.9	−14.9	−14.8	−14.9	

solute−solvent boundary) but rather to be a separate approximation to the physical problem of a molecule in a real solution.

For a given atomic radius, the amount of outlying charge is larger for an anion than a cation. This explains why there are larger deviations between NPE solvers and the GB equation for large R for anions than for cations in Tables 1–3. But it is unlikely that the NPE result is meaningful when there is outlying charge because charge transfers to the solvent, if present, should be treated quantum mechanically, not by dielectric continuum models.

From one point of view, the NPE solvers and the multipole expansion methods may seem more fundamental than the generalized Born approximation because the NPE is equivalent to one of Maxwell's four equations and is exact in classical electrostatics, and multipole expansions are exact when converged. However, there are also contrary considerations:

1. As illustrated above, the NPE solvers in current use in molecular quantum chemistry codes give quite different results for many realistic situations, even when employed with the same atomic radii as parameters, and even when the continuous electron density is not replaced by a point charge representation. Hence, they do not necessarily represent the correct solution even to the idealized nonhomogeneous dielectric problem. We have attributed a significant part of this deviation to outlying charge. In this regard, we should also note that the interaction of a test charge with a charge distribution (such as a the charge distribution of a solute) can be precisely represented by a multipole expansion only when the test charge lies outside the charge distribution, but due to atomic tails, this is never actually satisfied; therefore, the multipole expansion methods also suffer from outlying charge inconsistencies at the step where they represent the solute charge distribution as a multipole expansion on a finite cavity surface. An additional contributor to the differences between the various NPE solutions is that the default tesserae in the implementation of GCOSMO in *NWChem* are farther from convergence than those employed in the default implementation of IEFPCM in *Gaussian 03* or the default implementation of C-PCM in *GAMESS*.

2. Even if the Poisson equation were solved correctly for the situation of ϵ equals unity in an interior region and ϵ equals the bulk relative permittivity in an exterior region representing the solvent, this is not an exact model of a real solution. Actually, the permittivity changes gradually over a fluctuating region with a width of at least a few tenths of an Ångstrom and maybe even a width of one or two solvent

Table 5. Electrostatic Free Energies of Solvation for C_2H_6, CH_3OH, $CH_3OH_2^+$, and CH_3O^-

				ΔG_S (kcal/mol)				
M	R_H (Å)	R_C (Å)	R_O (Å)	GB	C-PCM	GCOSMO	IEFPCM	PBS
				C_2H_6				
0.85	0.87	1.33		−0.4	−2.6	−3.2	−2.6	0.2
1.00	1.02	1.57		−0.1	−1.1	−0.7	−1.1	0.0
1.15	1.17	1.81		−0.1	−0.5	−0.5	−0.5	0.0
Bondi	1.20	1.70		−0.1	−0.4	−0.4	−0.4	0.0
1.30	1.33	2.04		0.0	−0.3	−0.5	−0.3	0.0
1.2 × Bondi	1.44	2.04		0.0	−0.2	−0.3	−0.2	0.0
1.65	1.68	2.59		0.0	−0.1	−0.1	−0.1	0.0
≥2.00	2.04	3.14		0.0	0.0	0.0	0.0	0.0
				CH_3OH				
0.85	0.87	1.33	1.29	−6.2	−18.0	−17.7	−18.0	−14.8
1.00	1.02	1.57	1.52	−3.5	−10.9	−11.3	−10.9	−9.4
1.15	1.17	1.81	1.75	−2.2	−6.8	−7.1	−6.8	−6.0
Bondi	1.20	1.70	1.52	−2.9	−8.1	−8.1	−8.0	−7.4
1.30	1.33	2.04	1.98	−1.5	−4.5	−4.9	−4.5	−4.0
1.2 × Bondi	1.44	2.04	1.82	−1.6	−4.7	−4.7	−4.6	−4.3
1.65	1.68	2.59	2.51	−0.7	−2.0	−2.0	−2.0	−1.8
2.00	2.04	3.14	3.04	−0.4	−1.0	−1.1	−1.0	−1.0
3.00	3.06	4.71	4.56	−0.1	−0.3	−0.3	−0.3	−0.3
4.00	4.08	6.28	6.08	0.0	−0.1	−0.1	−0.1	−0.1
5.00	5.10	7.85	7.60	0.0	−0.1	−0.1	−0.1	−0.1
n.a.	10.00	10.00	10.00	0.0	0.0	0.0	0.0	0.0
				$CH_3OH_2^+$				
0.85	0.87	1.33	1.29	−92.9	−116.6	−116.4	−116.5	−106.9
1.00	1.02	1.57	1.52	−82.8	−95.5	−95.7	−95.5	−90.8
1.15	1.17	1.81	1.75	−75.1	−82.7	−83.4	−82.7	−79.8
Bondi	1.20	1.70	1.52	−76.6	−84.5	−86.8	−84.4	−80.8
1.30	1.33	2.04	1.98	−68.7	−73.7	−76.3	−73.7	−72.5
1.2 × Bondi	1.44	2.04	1.82	−67.6	−71.9	−72.9	−71.8	−70.5
1.65	1.68	2.59	2.51	−57.0	−59.4	−60.6	−59.5	−58.8
2.00	2.04	3.14	3.04	−48.1	−49.5	−50.4	−49.5	−49.8
3.00	3.06	4.71	4.56	−33.0	−33.4	−33.4	−33.4	−32.8
4.00	4.08	6.28	6.08	−25.1	−25.3	−25.3	−25.3	−24.3
5.00	5.10	7.85	7.60	−20.3	−20.4	−20.4	−20.4	
n.a.	10.00	10.00	10.00	−14.8	−14.8	−14.8	−14.8	
				CH_3O^-				
0.85	0.87	1.33	1.29	−103.2	−94.5	−106.5	−94.5	−101.3
1.00	1.02	1.57	1.52	−88.7	−87.8	−93.4	−87.7	−90.9
1.15	1.17	1.81	1.75	−78.5	−79.4	−82.9	−79.3	−80.6
Bondi	1.20	1.70	1.52	−86.3	−88.1	−91.8	−88.2	−89.8
1.30	1.33	2.04	1.98	−70.7	−71.9	−73.6	−71.9	−72.7
1.2 × Bondi	1.44	2.04	1.82	−73.2	−76.3	−78.4	−76.2	−77.0
1.65	1.68	2.59	2.51	−57.6	−58.5	−59.9	−58.6	−57.4
2.00	2.04	3.14	3.04	−48.5	−49.1	−50.2	−49.2	−49.6
3.00	3.06	4.71	4.56	−33.2	−33.5	−33.5	−33.5	−33.9
4.00	4.08	6.28	6.08	−25.3	−25.4	−25.4	−25.4	−24.9
5.00	5.10	7.85	7.60	−20.4	−20.5	−20.4	−20.5	−20.7
n.a.	10.00	10.00	10.00	−15.2	−15.2	−15.3	−15.2	−15.3

Table 6. Electrostatic Free Energies of Solvation for CH_2CHCHO and CH_2CF_2

			ΔG_S (kcal/mol)				
M	R_H (Å)	R_C (Å)	GB	C-PCM	GCOSMO	IEFPCM	PBS
			CH_2CHCHO				
Bondi	1.20	1.70	−4.3	−10.3	−10.6	−10.4	−9.3
1.2 × Bondi	1.44	2.04	−2.7	−6.0	−6.1	−6.0	−5.6
			CH_2CF_2				
Bondi	1.20	1.70	−0.6	−3.3	−3.9	−3.3	−2.7
1.2 × Bondi	1.44	2.04	−0.4	−1.5	−1.9	−1.5	−1.2

shells. (Some workers have used a three-parameter smooth function instead of a discontinuous one for the permittivity,[72,73] but the uncertainty in the shape and parameters remains.) The shape and size of the region where the permittivity differs from the bulk value are not known, and the correct value or function to use for the permittivity in this region is also not known, and these uncertainties make the electrostatic contribution quite uncertain. It is not clear that it is worthwhile to expend the effort to solve a partial differential equation accurately when there are significant uncertainties

in the parameters of the differential equation and the results depend strongly on these parameters.

3. Even if the parameters governing the spatial dependence of the solvent permittivity were known, and even if the Poisson equation were easy to solve, it is not strictly valid to use an equation of macroscopic electrostatics on an atomic scale. Just as macroscopic thermodynamics breaks down at the nanoscale, so does macroscopic electrostatics.

In contrast to the NPE solvers, the generalized Born approximation simplifies the treatment of charge distributions by replacing the continuous charge density of the solute by a set of atom-centered partial charges for all stages of the calculation. This eliminates the incorrect treatment of outlying charge but at the cost of incorrectly eliminating the outlying charge itself. Furthermore, the GB method introduces errors due to the fact that certain types of charge distributions may be poorly described by atom-centered monopoles, for example, the local charge distributions about atoms with lone pairs.[74]

One way to try to ascertain the correct partition of solvation effects into electrostatic and nonelectrostatic terms is to consider the change in observable or potentially observable properties of solutes. For example, the change in solute dipole moment is mainly a response to the electrostatic interaction with the solvent, so that if two models give different electrostatic contributions to the solvation process, the one that predicts a more accurate solute dipole moment in solution might (other factors being equal) be judged to be the more reasonable one (although actual calculations of this response property show more complex relationships than one might have expected between predicted electrostatic contributions to free energy of solvation and predicted changes in dipole moments[7]). Unfortunately, the molecular dipole moment is not well-defined for a molecule immersed in a solvent, even in the absence of charge transfer (although considerable effort has gone into methods for obtaining physically reasonable values[75–77]), and charge transfer to and from the solvent complicates this further.[78] Perhaps there are observables, for example, in vibrational spectroscopy or in NMR spectroscopy, that will eventually give useful information of this type, but so far the consideration of molecular properties in solution has not yielded any firm conclusions about the best way to approximate the electrostatic effect.

Some studies have shown good agreement between explicit and implicit solvent models for free energies of solvation,[79–86] and other studies have found differences that were interpreted in terms of deficiencies in the GB type of treatment.[87] Thus, one might consider defining the electrostatic contribution statistical-mechanically from explicit calculations, but there are difficulties with this as well. In addition to the fundamental difficulty[5,6] mentioned in the Introduction, there are uncertainties due to the molecular mechanics representation of the solvent. Furthermore, the partial charges that control electrostatic terms in molecular mechanics are sometimes not optimized separately from the short-range van der Waals terms. Thus, the electrostatic and nonelectrostatic terms are not separately meaningful in a quantitative sense.

In light of the intrinsic uncertainties in apportioning the solvation free energy into electrostatic and nonelectrostatic terms, either thermodynamically or by well-defined model problems, another strategy may be considered. First, recognizing that there is some arbitrariness, one defines the bulk electrostatic contribution in a physically reasonable way. Since, for any physical definition of the electrostatic contribution, electrostatics will dominate the solvation energy of ions, the reasonableness of this definition may be judged by its success for ionic solutes. If this definition is based on algorithms such as the NPE or GB ones, it will fully include the long-range interaction of the solute with bulk solvent, for which it is reasonable to use the bulk permittivity. All remaining contributions to the solvation free energy are then of short-range, and they include cavitation, dispersion (the $|\mathbf{r} - \mathbf{r}_k|^{-6}$ dependence of dispersion is of shorter range than the $|\mathbf{r} - \mathbf{r}_k|^{-4}$ dependence of electrostatic polarization), and solvent structural considerations that cause the short-range response of the solvent to differ from the defined bulk electrostatic contributions. The solvent structural component has two kinds of contributions, which one may label as short-range (or nonbulk) electrostatic contributions and nonelectrostatic contributions. The nonelectrostatic contributions include effects such as the solvent entropic components of the hydrophobic effect and the partial covalent character of hydrogen bonds. The short-range electrostatic contributions include all of the deviations of the electrostatics from the assumed bulk model, such as the inexactness of the solute charge model (whether the solute charge distributions are modeled by wave functions or partial charges, there is in practice a residual inaccuracy, for example due to basis sets or fixed solute geometries, and this may have a systematic component) and the inexactness of the solvent permittivity model (including assumed values for intrinsic Coulomb radii). Since the short-range effects are primarily associated with physical effects in the first solvation shell and the inexactness and uncertainties of the bulk electrostatic model in the first solvation shell, they can be modeled in terms of short-range analytic forms such as surface tensions associated with the SAS. Such a strategy has already been used for solvation modeling, and it has led to more accurate models than those for which the electrostatic and nonelectrostatic terms were determined separately.[56,57]

6. Conclusions

There is no unique way to define the electrostatic contribution to the free energy of solvation, and methods currently in use can give very different answers for the same idealized problem of a molecular solute in a continuum dielectric medium. Although the electrostatic component of the free energy of solvation is not well-defined in thermodynamics, one might hope to define it by statistical mechanics. However, current methods based on the nonhomogeneous Poisson equation have uncertainties due to the definition, size, and shape of the solute cavity; the assumed way in which the permittivity changes at and near the solute−solvent boundary; and the portion of the solute charge that lies outside the cavity. The generalized Born approximation has different but equally serious approximations.

The electrostatic contribution to solvation, like the electrostatic contribution to any molecular modeling problem, cannot be considered in isolation from the whole model. Therefore, the "electrostatic" and "nonelectrostatic" terms should be considered as "bulk-electrostatic" and "non-bulk-electrostatic" terms, and the latter should account not only for cavity, dispersion, solvent structure, repulsion, and liberation, but also for deviations of the "true" electrostatics from those corresponding to the bulk model assumed in the bulk-electrostatic terms. This deviation is sensitive to the solvent structure in the first solvation shell, and it can be considered to be part of the solvent structure contribution to the nonbulk-electrostatic terms. The validity of a model can be judged by the usefulness of the whole model in predicting and correlating experimental observables, but not by any supposed rigor in the electrostatic part of the formulation.

Acknowledgment. The authors are grateful to Masahiro Higashi for helpful comments. This work was supported in part by the Office of Naval Research under award number N00014-05-1-0538 and by the National Science Foundation (grants CHE06-10183 and CHE07-04974). Computer resources were provided by the Minnesota Supercomputing Institute and by a Molecular Science Computing Facility Computational Grand Challenge grant (gc20893) for Computational Chemical Dynamics of Complex Systems at the Environmental Molecular Sciences Laboratory of Pacific Northwest National Laboratory.

Supporting Information Available: Intramolecular and intermolecular contributions to the free energy of solvation for the case of a rigid solute; electrostatic components of the free energies of solvation for $CH_3OH_2^+$ and CH_3O^- calculated using the Hartree−Fock method and using the M05-2X density functional; electrostatic components of the free energies of solvation for $CH_3OH_2^+$ calculated by IEFPCM/*Gaussian 03* using tesserae of different sizes. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.

(2) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.

(3) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(4) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

(5) Pethica, B. A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6253.

(6) (a) Smith, P. E.; Van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *98*, 13735. (b) Boresch, S.; Archontis, G.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 25.

(7) Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-López, M. F.; Rinaldi, D.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2003**, *24*, 284.

(8) (a) Luque, F. J.; Negre, M. J.; Orozco, M. *J. Phys. Chem.* **1993**, *97*, 4386. (b) Ten-no, S.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1994**, *100*, 7443. (c) Casanova, D.; Gusarov, S.; Kovalenko, A.; Ziegler, T. *J. Chem. Theory Comput.* **2007**, *3*, 458.

(9) Wangsness, R. K. *Electromagnetic Fields*; Wiley: New York, 1979; p 179.

(10) *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Politzer, P., Truhar, D. G., Ed.; Plenum: New York, 1981.

(11) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer; Dordrecht, The Netherlands, 1996; Understanding Chem., Reactivity Series 17, p. 1.

(12) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(13) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422.

(14) (a) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151. (b) Connolly, M. L. *Science* **1983**, *221*, 709.

(15) (a) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671. (b) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.

(16) (a) Zauhar, R. J.; Morgan, R. S. *J. Mol. Biol.* **1985**, *186*, 815. (b) Vorobjev, Y. N.; Grant, J. A.; Scheraga, H. A. *J. Am. Chem. Soc.* **1992**, *114*, 3189. (c) Cammi, R.; Tomasi, J. *J. Comput. Chem.* **1995**, *16*, 1449.

(17) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098.

(18) Rivail, J. L.; Rinaldi, D. In *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 1996; Comp. Chem. Rev. Curr. Trends Series 1, Vol. 1, p 139.

(19) Hoijtink, G. J.; de Boer, E.; Van der Meij, P. H.; Weigland, W. P. *Rec. Trav. Chim. Pays-Bas Belg.* **1956**, *75*, 487.

(20) Peradejordi, *Cah. Phys.* **1963**, *17*, 393.

(21) Jano, I. *Compt. Rend. Acad. Sci. Paris* **1965**, *261*, 103.

(22) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Viellard, A., Eds.; Wiley: London, 1981; Vol. 2, p 25.

(23) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1.

(24) (a) Kozaki, T.; Morihashi, K.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547. (b) Takahashi, O.; Sawahata, H.; Ogawa, Y.; Kikuchi, O. *THEOCHEM* **1997**, *393*, 141.

(25) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164.

(26) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(27) (a) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305. (b) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117.

(28) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875.

(29) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.

(30) 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.;

Bakken, V.; 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 03*, revisions C.01, C.02, and D.02; Gaussian, Inc.: Wallingford, CT, 2004.

(31) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151.

(32) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253.

(33) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43.

(34) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(35) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; p 1167.

(36) *GAMESS computer package*, version 7 SEP 2006 (release R6); Iowa State University: Ames, IA, 2006. http://www.msg.ameslab.gov/GAMESS/GAMESS.html (accessed Feb 2007).

(37) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799.

(38) Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1995**, *240*, 253.

(39) Baldridge, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622.

(40) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.

(41) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.

(42) Li, H.; Pomelli, C. S.; Jensen, J. H. *Theor. Chem. Acc.* **2003**, *109*, 71.

(43) Li, H.; Jensen, J. H. *J. Comput. Chem.* **2004**, *25*, 1449.

(44) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.

(45) *Jaguar 6.5*, release 112; Schrödinger, Inc.: Portland, OR, 2005.

(46) Bylaska, E. J.; de Jong, W. A.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Aprà, E.; Windus, T. L.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Früchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, R. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. *NWChem, A Computational Chemistry Package for Parallel Computers*, version 5.0; Pacific Northwest National Laboratory: Richland, WA, 2006.

(47) Truong, T. N.; Stefanovich, E. V. *J. Phys. Chem.* **1995**, *99*, 14700.

(48) Truong, T. N.; Stefanovich, E. V. *J. Chem. Phys.* **1995**, *103*, 3709.

(49) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2000**, *21*, 340.

(50) Pomelli, C. S.; Tomasi, J.; Barone, V. *Theor. Chem. Acc.* **2001**, *105*, 446.

(51) Olson, R. M.; Marenich, A. V.; Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Winget, P.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM*, version 2007-beta; University of Minnesota: Minneapolis, MN, 2007.

(52) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057.

(53) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293.

(54) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 5503.

(55) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.

(56) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.

(57) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G *J. Chem. Theory Comput.* **2007**, *3*, 2011.

(58) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.

(59) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley-Interscience: New York, 1986.

(60) (a) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103. (b) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, [Online] http://dx.doi.org/10.1007/s00214-007-0310-x.

(61) Born, M. *Z. Phys.* **1920**, *1*, 45.

(62) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

(63) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074.

(64) Klamt, A.; Eckert, F. *Fluid Phase Equilib.* **2000**, *172*, 43.

(65) Hsu, C. P.; Head-Gordon, M.; Head-Gordon, T. *J. Chem. Phys.* **1999**, *111*, 9700.

(66) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.

(67) For further discussion of the outlying charge effect, see refs 31 and 39 of the present paper and refs 1, 2, 28, 29, 31, and 32 of ref 39.

(68) (a) Cancès, E.; Mennucci, B. *J. Chem. Phys.* **2001**, *114*, 4744. (b) Chipman, D. *J. Chem. Phys.* **2006**, *124*, 224111.

(69) Chipman, D. M. *J. Chem. Phys.* **2000**, *112*, 5558.

(70) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 265.

(71) Mongan, J.; Svrcek-Seiler, W. A.; Onufriev, A. *J. Chem. Phys.* **2007**, *127*, 185101.

(72) Fattebert, J.-L.; Gygi, F. *Int. J. Quantum Chem.* **2003**, *93*, 139.

(73) Scherlis, D. A.; Fattebert, J.-L.; Gygi, F.; Cococcioni, M.; Marzari, N. *J. Chem. Phys.* **2006**, *124*, 74103.

(74) (a) Stone, A. J.; Price, S. L. *J. Phys. Chem.* **1988**, *92*, 3325. (b) Tschampel, S. M.; Kennerty, M. R.; Woods, R. J. *J. Chem. Theory Comput.* **2007**, *3*, 1721.

(75) Smith, J. W. *Trans. Faraday Soc.* **1952**, *48*, 802.

(76) Pilpel, N. *J. Am. Chem. Soc.* **1955**, *77*, 2949.

(77) Costa Cabral, B. J.; Guedes, R. C.; Pai-Panandiker, R. S.; Nieto de Castro, C. A. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4200.

(78) Marenich, A. V.; Olson, R. M.; Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2055.

(79) Jean-Charles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 1454.

(80) Orozco, M.; Jorgensen, W. L.; Luque, F. J. *J. Comput. Chem.* **1993**, *14*, 1498.

(81) Jorgensen, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993**, *14*, 195.

(82) Ösapay, K.; Young, W. S.; Bashford, D.; Brooks, C. L., III; Case, D. A. *J. Phys. Chem.* **1996**, *100*, 2698.

(83) Luque, F. J.; Zhang, Y.; Alemán, C.; Bachs, M.; Gao, J.; Orozco, M. *J. Phys. Chem.* **1996**, *100*, 4269.

(84) Best, S. A.; Merz, K. M., Jr.; Reynolds, C. H. *J. Phys. Chem. B* **1999**, *103*, 714.

(85) Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 1722.

(86) Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167.

(87) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 1846.

CT800029C

JCTC Journal of Chemical Theory and Computation

Improved Description of Stereoelectronic Effects in Hydrocarbons Using Semilocal Density Functional Theory

Gábor I. Csonka,*,† Adrienn Ruzsinszky,‡ John P. Perdew,‡ and Stefan Grimme§

Department of Inorganic and Analytical Chemistry, Budapest University of Technology and Economics, H-1521 Budapest, Hungary, Department of Physics and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118, and Theoretical Organic Chemistry, Organic Chemistry Institute, University of Münster, Corrensstrasse 40, 48149 Münster, Germany

Received January 3, 2008

Abstract: Serious and systematic errors with popular density functionals occur for isodesmic stabilization energies of n-alkanes, isomerization, and dimerization energies of hydrocarbons and geometries of sterically overcrowded aromatic systems. These functionals are too biased toward the correct description of free atoms. Changing two parameters within the Perdew-Burke-Ernzerhof approximation leads to a new nonempirical functional, PBEsol, that improves the description of large organic systems.

Proper description of stereoelectronic (SE) effects[1] is desirable for any theoretical method to be used in organic chemistry. The SE design rules are frequently used in synthetic organic chemistry to design and explain new reactions by electron donating and withdrawing effects or steric interactions. It is shown in this communication that a new, nonempirical semilocal Kohn−Sham (KS) density functional theory (DFT)[2] approximation, called PBEsol,[3] describes more correctly the SE effects in many hydrocarbons than do the semilocal DFT functionals tested in previous reports.[4,5]

KS DFT approximations for the exchange-correlation energy can be classified into two major groups, empirically fitted or nonempirical. The nonempirical functionals form a hierarchy of semilocal approximations with increasing complexity and (usually) accuracy. The simplest approximation is the local spin density approximation (LSDA), while the next two levels are the generalized gradient approximation or GGA (e.g., PBE[6]) and the meta-GGA (e.g., TPSS[7]). The hybrid functionals define a fourth level that is fully nonlocal and semiempirical. In the assessment of density functional approximations, great weight is usually given to the accuracy of molecular atomization energies or the enthalpies of formation constructed from calculated atomization energies. Earlier studies have shown[8] that GGA and global hybrid GGA functionals (e.g., B3LYP,[9] or B3PW91[10]), that were parametrized for the enthalpies of formation of the relatively small molecules, can fail seriously for larger molecules. Due to the computational expense of the exact exchange, calculations using hybrid functionals can be very expensive compared to the pure GGA or meta-GGA functionals. Solid-state calculations are also quite inefficient with hybrid functionals, and thus development of good quality, reliable semilocal functionals is important.

We have shown[11] that the TPSS meta-GGA[7] achieves remarkably accurate atomization energies for 50 large hydrocarbons and substituted hydrocarbons. We also presented strong evidence that most of the error of previous nonempirical functionals resides in the energy of the free atom and so cancels out of typical reaction energies. Finally, we suggested that enthalpies of formation, calculated without any reference to the free atoms, would provide a fairer assessment of the performance of approximate density functionals. (See also refs 12 and 13.)

Recently examples of stereoelectronic effects in alkane isomers have been summarized.[4] Independent works have shown serious and systematic errors of several popular DFT methods (e.g., B3LYP or to a lesser extent PBE) for isodesmic stabilization energies of n-alkanes[5] and octane or $(CH)_{12}$ hydrocarbon isomer energy differences.[14]

More recently it has been argued that popular exchange-correlation functionals are biased toward the correct description of free atoms[3,15] by their gradient coefficients for exchange and thus often fail to improve upon LSDA for solids. These functionals also fail for larger organic molecules. It has been shown that the exact second-order gradient expansion for exchange is relevant for realistic densities of densely packed solids, while the similar expansion for correlation is less relevant. This suggests using the exact second-order gradient expansion coefficient ($\mu_{GE} = 10/81$) within the PBE exchange enhancement factor, $F_x(s)$ where s is a dimensionless density gradient, for such densities. Also adjusting the correlation functional to satisfy another condition leads to a new nonempirical GGA, PBEsol, that performs well for solids and their surfaces and could improve the description of large organic systems and reactions.

* Corresponding author e-mail: csonka@web.inc.bme.hu.
† Budapest University of Technology and Economics.
‡ Tulane University.
§ University of Münster.

10.1021/ct800003n CCC: $40.75
Published on Web 04/22/2008

With increasing size, the electron densities of molecules become more similar to solid-state densities.

We note that PBE and PBEsol are both first-principles GGAs, but they are based upon different selections of exact constraints to satisfy. At the GGA level, but not at the higher meta-GGA level, one can at most satisfy two out of the following three constraints exactly: (I) second-order gradient expansion for exchange, (II) second-order gradient expansion for correlation, (III) LSD-like linear density response of a uniform electron gas. PBE satifies (II) and (III) but not (I). PBEsol satisfies (I) and compromises between (II) and (III). Any PBE GGA subroutine can be converted to PBEsol simply by changing two gradient coefficients[3] (μ from 0.21951 to 10/81 and β from 0.0667 to 0.0460). Further details of the new PBEsol functional and its performance can be found in ref 3.

In order to estimate the long-range dispersion energy we use the well established and tested DFT-D method,[16] employing damped atom-pairwise $-C_6/R^6$ terms. (For a recent review see 17.) For PBEsol-D we take the (asymptotically correct) s_6 scaling factor of unity (for PBE $s_6 = 0.75$) and merely readjust the R_{scal} scaling factor of the atomic van der Waals radii from the original value of 1.1 to 1.42 to fit the computed interaction energies of a few typical van der Waals complexes (S22 set, see the Supporting Information). The r_0 value used in the damping function is the sum of the two scaled atomic van der Waals radii.[16] The larger R_{scal} accounts for the more attractive character of the interatomic PBEsol interactions in the medium-range correlation regime (making the dispersion correction more long-ranged). While no GGA without a dispersion correction can account for long-range correlation, the PBEsol (after the error cancelation between exchange and correlation expected for local and semilocal approximations) can apparently account for medium-range interaction better than PBE. Moreover, for Ne_2 and Ar_2, PBE is reasonably good,[18,19] but PBEsol is even better[20] for the repulsive part of the binding energy curve.

Following Zhao and Truhlar,[21] we have used a small test set composed of the following components: experimentally derived zero-point exclusive chemical reaction energies of

$$n\text{-}C_6H_{14} + 4CH_4 \rightarrow 5C_2H_6 \quad (1)$$

and

$$n\text{-}C_8H_{18} + 6CH_4 \rightarrow 7C_2H_6 \quad (2)$$

experimentally derived[22] zero-point exclusive relative energies of 2,2,3,3-tetramethylbutane **3a** and *n*-octane **3b** and relative energies of three $(CH)_{12}$ isomers **4a**, **4b**, and **4c**. We added to the test set the anthracene dimerization reaction energy (**5a** and **5b**). The structures are shown in Figure 1.

We note that the serious discrepancies between computed and experimental isodesmic reaction energies of (1) and (2) have been attributed to the so-called "protobranching" effect, defined as the stabilizing interaction of geminal methyl ($-CH_3$) or methylene ($-CH_2-$) groups. It has been shown[4] that many popular exchange-correlation functionals fail for this effect. Later it was shown that the heavily parametrized M05-2X global hybrid functional (for organic chemistry only) performs well on these tests.[21] This was attributed to an improved correlation functional for the medium-range electron correlation. While we concur with the importance of medium-range correlation in these

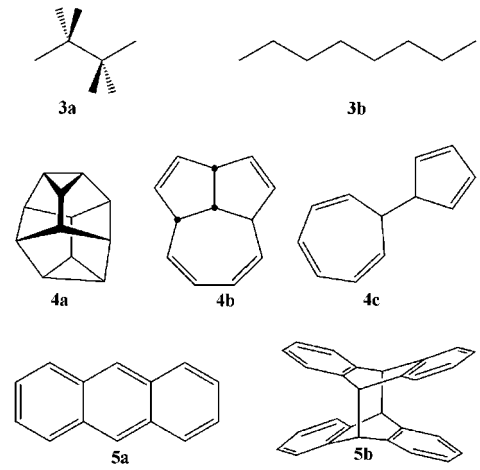


Figure 1. Structures of 2,2,3,3-tetramethylbutane **3a**, *n*-octane **3b**, and the $(CH)_{12}$ isomers, where **4a**, **4b**, and **4c** correspond to the structures 1, 22, and 31 in ref 14. Anthracene is **5a** and its dimer is **5b**.

Table 1. Zero-Point Exclusive Energies (kcal/mol) of Hexane ($\Delta E(1)$) and Octane ($\Delta E(2)$) Isodesmic Stabilization Energies of Reactions 1 and 2 and Relative Energies of Two Octane Isomers

method	ΔE (1) hexane	ΔE (2) octane	E(**3b**)-E(**3a**) octane isomers
expt	9.3[a]	14.0[a]	1.9[a]
CCSD(T)	8.6[b]		
SCS-MP2	8.0[c]	12.0[c]	1.9[c]
MP2	9.7[c]	14.6[c]	5.1[c]
M05-2X	8.2[d]	12.2[d]	2.0[d]
B3PW91	4.7[d]	7.0[d]	−7.0[d]
B3LYP	4.0[b]	5.9[b]	−8.4[d]
TPSS	3.9[e]	5.7[e]	−5.6[e]
PBE	5.6[e]	8.3[e]	−4.6[e]
PBEsol	7.4[e]	11.0[e]	−1.3[e]
PBEsol-D	8.3[e]	12.5[e]	0.3[e]
LSDA	9.9[e]	14.8[e]	3.7[e]

[a] For experimental values see ref 22. [b] Reference 9. [c] Present work, single point frozen core SCS-MP2 or MP2/TZV(*2df,2pd*)//PBEh/TZV(*2d,p*) results. [d] Reference 21. [e] Present work, single point 6-311+G(*2d,2p*)//M05-2X/6-311+G(*2df,2p*) results. PBEsol-D: an estimated dispersion correction was added to PBEsol values.

systems, it is interesting to note that the LSDA that is the most local functional performs extremely well for these problems, as observed by Wodrich et al.[5] and shown in Table 1. These good LSDA results can be attributed to the serious overbinding error of the Slater exchange that is imperfectly balanced by the LSDA correlation.[19]

The results in Table 1 show the relatively good performance of M05-2X, PBEsol, and LSDA for the isodesmic stabilization energies of hexane and octane and for the isomer energy difference of octane. Good performance of PBEsol for $(CH)_{12}$-(E(**4c**)-E(**4b**)) can also be observed in Table 2 (taking CCSD(T) as the standard for **4**; see details below). In contrast, the 8.3 kcal/mol CCSD(T) energy difference between the **4c** and **4b** isomers is seriously underestimated by the PBE, TPSS, and especially the B3LYP functionals. (The B3LYP functional gives very poor results for structures with single bonds only and for bicyclic hydrocarbons, as pointed out in ref 14.) However, E(**4b**)-E(**4a**), involving a large change in the number of multiple bonds, is strongly overestimated by LSDA and PBEsol, while

Table 2. Relative Energies of Three $(CH)_{12}$ Isomers[e]

method	E(**4c**)-E(**4b**)	E(**4b**)-E(**4a**)
CCSD(T)	8.3[a]	12.2[a]
SCS-MP2	8.1[a]	17.3[a]
MP2	7.3[a]	24.4[a]
M05-2X	7.4[b]	14.0[b]
B3PW91	6.7[c]	16.8[c]
B3LYP	1.3[c]	−1.7[c]
TPSS	3.4[d]	10.8[d]
PBE	5.2[d]	13.2[d]
PBEsol	9.4[d]	25.9[d]
PBEsol-D	10.0[d]	26.5[d]
LSDA	11.4[d]	29.3[d]

[a] Present work, single point frozen core CCSD(T)/complete basis set TQ extrapolated// PBEh/TZV(*d,p*) and SCS-MP2 or MP2/TZV(*2df,2pd*)// PBEh/TZV(*2d,p*) results. [b] Reference 21. [c] Reference 14. [d] Present work, single point 6-311+G(*2d,2p*)//M05-2X/6-311+G(*2df,2p*) results. PBEsol-D: an estimated dispersion correction was added to PBEsol values. [e] The number of multiple bonds is zero for **4a**, four for **4b**, and five for **4c**.

PBE, TPSS, and M05-2X give energy differences close to that of CCSD(T). The M06-L functional[23] gives slightly worse results[24] for these problems than the M05-2X functional.

The photodimerization of anthracene (**5a**) is a reversible [4 + 4] cycloaddition which yields the covalently bound polycyclic dimer (**5b**). The condensed phase dimer is labile at elevated temperatures and exothermally dissociates to monomers. The dissociation energy of the gas phase dimer is well studied, and the highest levels of calculations give $D_e = 9 \pm 3$ kcal/mol (endothermic).[25] It can be observed that HF, B3LYP, PBE, and dispersion corrected PBE-D fail considerably for D_e, giving −19.6, −24.6, −13.2, and 3.2 kcal/mol, respectively (using TZV(*2p,d*) basis set and MP2/TZV(*d,p*) geometries; see also ref 26). The LSDA/6-311+G(*2d,2p*) D_e is too positive (13 kcal/mol). Our new PBEsol and dispersion corrected PBEsol-D/TZV(*2p,d*) single point energies are considerably better, giving for D_e −0.5 and 9.4 kcal/mol, respectively. (See the Supporting Information).

Longstanding problems for density functionals are the geometries of cyclophanes and related sterically overcrowded aromatic systems.[27] As analyzed in detail in ref 27, typical density functionals properly account for the Pauli-exchange repulsion between the clamped aromatic rings but do not describe accurately the electron correlation effects between the π-systems. These are at distances below the van der Waals radius of carbon, showing that this is not a typical dispersion problem. In effect this leads to too long inter-ring spacings and too strong stretching of the bridging single bonds (cf. Figure 2 for PBE results and ref 27 for even worse B3LYP data).

As can be seen from Figure 2, this problem is more or less solved with PBEsol, which yields almost perfect agreement with experiment for all important structural parameters. This holds in particular for the (chirality inducing) torsion angle ϕ which is too small with PBE and even zero with B3LYP but very accurate with PBEsol. The differences between the PBE and PBEsol geometries are much larger than the expected packing effects. PBEsol geometry is much closer to the SCS-MP2 or experimental geometry[27] than PBE is.

The presented results suggest that a diminished gradient dependence makes PBEsol better than PBE and other standard semilocal functionals not only for solids (moderately varying densities with metallic or single bonds) but also for single-bonded hydrocarbon molecules and for E(**4c**)-E(**4b**), while PBE and TPSS remain better for energy differences between some multiply and singly bonded isomers (e.g., E(**4b**)-E(**4a**) or E(**4c**)-E(**4a**)). PBEsol is clearly better than PBE for the gas phase dissociation energy of the anthracene dimer (**5a** and **5b**). For **6**, where the medium-range electron correlation is important, the new PBEsol functional gives excellent agreement with the experimental geometry. It can be expected that PBEsol, which provides good lattice constants for solids, will also give good geometries for large molecules. Further studies are needed to confirm and refine these tentative conclusions. The stronger enhancement of exchange in PBE GGA and other standard semilocal density functionals is needed for free atoms, for hydrogen bonds, and perhaps for some multiple bonds. In independent work, Vela[28] has also found improvement for octane isomerization energies from a GGA that uses the exact gradient coefficient for exchange over a wide range of density gradients.

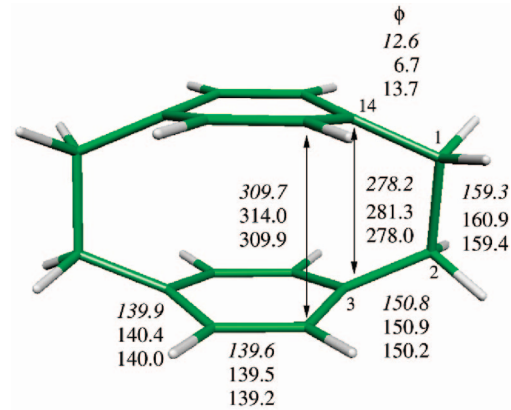


Figure 2. Optimized structure (PBEsol/TZV(*2df,2p*)) of [2.2]paracyclophane **6** (D_2-symmetry) and important geometrical parameters (bond lengths in pm, torsion angle ϕ(3−2−1−14) in °; top: experimental X-ray data,[27] middle: PBE, bottom: PBEsol).

In summary, the simple PBEsol GGA, which was developed nonempirically for densely packed solids, is also useful for large organic molecules (in the absence of free atoms). It can be useful for quick evaluations of geometries and frequencies, to be followed possibly by single-point energy calculations at higher levels of approximation. Moreover, it may be possible to build the PBEsol construction principle (recovery of the gradient expansion for exchange over a wide range of slowly- or moderately varying densities) into meta-GGAs, which can be more widely useful than GGAs and are not much more expensive.

Acknowledgment. We thank Y. Zhao and D. G. Truhlar for the M06-L results. Support is acknowledged from OTKA, (PD-050014), NSF (DMR-0501588), and NSF-MTA-OTKA (98).

Appendix: Technical Details

The experimental values used in Table 1 were derived from experimental standard enthalpies of formation.[22] The values were corrected to 0 K, and an experimental zero-point vibration energy correction was subtracted. The error bars

for the experimental values are about 0.5 kcal/mol. The frozen core CCSD(T) isomerization energies for **4** were obtained from complete basis set extrapolation using SCS-MP2/cc-pVXZ (X = T, Q) energies. Smaller, double-ζ quality basis sets used in ref 14 give 2−3 kcal/mol basis set error for the energy differences of the $(CH)_{12}$ isomers shown in Table 1. We have calculated the effect of the core-correlation on the relative energies of **4**, and it is comparable to the complete valence basis set extrapolation error (0.3−0.6 kcal/mol). Reference 14 shows that single-point CCSD(T) calculations that use DFT and MP2 geometries give similar isomerization energies within the range of 1 kcal/mol. The M05-2X/6-311+G(*2df,2p*), B3PW91 and B3LYP results obtained with triple-ζ quality basis sets are taken from the literature.[4,8,14,21] The M05-2X/cQZV3P//MP2/TZV(*d,p*) model gives slightly different values for Δ*E* (1), Δ*E* (2), and E(**3b**)-E(**3a**) (within 0.6 kcal/mol). For the TPSS, PBE, PBEsol, and LSDA (with the SVWN5 functional for LSDA) calculations, we use 6-311+G(*2d,2p*), cc-pVTZ(*-f*), and TZV(2*d*,*p*) or TZV(2*df*,2*p*) basis sets of Ahlrichs et al.[29] and PBE/6-311+G(*d,p*), M05-2X/6-311+G(*2df,2p*) and MP2/TZV(*d,p*) geometries. Using the complete cc-PVTZ basis set (adding *f* functions) changes the Δ*E* (1) by 0.2 kcal/mol, so using *f* functions has a small influence on the calculated results. The DFT calculations were performed with the modified GAUSSIAN 03 program[30] and with the slightly modified versions of the Turbomole suite of programs.[31] It was observed that the geometry differences result in a maximum of 0.6 kcal/mol change in the energies of Δ*E* (1), Δ*E* (2), and E(**3b**)-E(**3a**). The basis set effect is about 1 kcal/mol. We confirm the suggestion of ref 31: Triple-ζ quality basis sets are needed for such DFT calculations, and an increase of the polarization part to (*2d*,*2p*) is advantageous.

Supporting Information Available: Geometries and/or total energies of compounds **4**, **5**, and **6** and dissociation energies and statistical data for the S22 benchmark set. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) *Stereoelectronic Effects in Organic Chemistry: P. Delongchampsin, Organic Chemistry Series*; Baldwin, J. E., Ed.; Pergamon: Oxford, 1983; Vol. *1,* pp 5−20.

(2) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.

(3) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L.; Zhou, X.; Burke, K. *Phys. Rev. Lett.* **2008**, *100*, 136406.

(4) Grimme, S. *Angew. Chem., Int. Ed.* **2006**, *45*, 4460.

(5) Wodrich, M. D.; Corminboeuf, C.; Schleyer, P. v. R. *Org. Lett.* **2006**, *8*, 3631.

(6) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(7) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(8) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374.

(9) Stephens, P. J.; Devlin, F. J.; Chablowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(10) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(11) Csonka, G. I.; Ruzsinszky, A.; Tao, J.; Perdew, J. P. *Int. J. Quantum Chem.* **2005**, *101*, 506.

(12) Grimme, S. *J. Phys. Chem. A* **2005**, *109*, 3067.

(13) Brothers, E. N.; Scuseria, G. E. *J. Chem. Theory Comput.* **2006**, *2*, 1045.

(14) Schreiner, P. R.; Fokin, A. A.; Pascal, R. A., Jr; de Meijere, A *Org. Lett.* **2006**, *8*, 3635.

(15) Perdew, J. P.; Constantin, L. A.; Sagvolden, E.; Burke, K. *Phys. Rev. Lett.* **2006**, *97*, 223002.

(16) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.

(17) Grimme, S.; Antony, J.; Schwabe, T.; Mück-Lichtenfeld, C. *Org. Biomol. Chem.* **2007**, *5*, 741.

(18) Tao, J. M.; Perdew, J. P *J. Chem. Phys.* **2005**, *122*, 114102.

(19) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. *J. Phys. Chem. A.* **2005**, *109*, 11015.

(20) Csonka, G. I., private communication.

(21) Zhao, Y.; Truhlar, D. G. *Org. Lett.* **2006**, *8*, 5753.

(22) NIST Standard Reference Database. See http:// webbook.nist.gov/chemistry/ (accessed Feb 1, 2008).

(23) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(24) Zhao, Y.; Truhlar, D. G., private communication.

(25) Grimme, S.; Diedrich, C.; Korth, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 625.

(26) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(27) Grimme, S. *Chem. Eur. J.* **2004**, *10*, 3423.

(28) Vela, A., private communication.

(29) For standard basis sets see http://www.emsl.pnl.gov/forms/basisform.html (accessed Feb 19, 2008). The basis sets for Turbomole are available from the Turbomole homepage via the FTP server button (in the subdirectories basen, jbasen, and cbasen). See http://www.turbomole.com (accessed Feb 19, 2008).

(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria G. E. Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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.; Bakken, V.; 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 03, Revision D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(31) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169. Turbomole (Version 5.9): R. Ahlrichs et al., Universität Karlsruhe, 2003. See also: http://www.turbomole.com (accessed Feb 1, 2008)

(32) Grimme, S.; Steinmetz, M.; Korth, M. *J. Org. Chem.* **2007**, *72*, 2118.

CT800003N

JCTC Journal of Chemical Theory and Computation

Symmetry-Adapted Perturbation-Theory Interaction-Energy Decomposition for Hydrogen-Bonded and Stacking Structures

Hubert Cybulski and Joanna Sadlej*

Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland

Received February 28, 2008

Abstract: This letter reports the computational ab initio studies on the stacked and hydrogen-bonded geometries of the uracil dimer and pyrimidine···*p*-benzoquinone complex with a special regard to the ratios of different interaction-energy terms calculated by means of the symmetry-adapted perturbation theory (SAPT). In the hydrogen-bonded systems the absolute value of the dispersion term constitutes approximately half of the absolute value of the total SAPT0 interaction energy, while in the stacking complexes the ratio of the dispersion to the total interaction energy is much larger, ca. 1.2−2.0. Our SAPT results are compared with the DFT-SAPT results published recently by the Hobza group (*J. Chem. Phys.* **2007**, *127*, 075104), and the role of the dispersion contribution in stacking and hydrogen-bonded arrangements is discussed. The methodological part of this letter presents the influence of counterpoise corrections in the optimization procedure on the geometries of the systems and the calculated SAPT contributions.

Noncovalent interactions make an important contribution to the stabilization of the structure of various forms of molecular systems, among them DNA and RNA. There are two main types of noncovalent interactions: hydrogen bonds and stacking interactions. It is well-known that the helical structure of DNA is stabilized not only by hydrogen bonds in Watson−Crick adenine−thymine (AT) and guanine−cytosine (GC) pairs but also by stacking interactions between pyrimidine and purine bases along the DNA helical backbone. It was also shown that stacking interactions play an important role in the long-distance DNA radiation-induced damage repair.[1]

The origin of the stabilization of hydrogen bonding and stacking interaction is different.[2] It is widely and intuitively accepted that the former bonding originates mainly in electrostatic interactions, while the stacking bonding is mainly due to the dispersion energy.[3–8] For a long time they were believed to be much weaker than the hydrogen bonding. However, the recently calculated interaction energies have shown that the stacked pairs are about as stable as the AT planar pair, the weakest hydrogen-bonding interaction in DNA.[6,9] This finding changes the widely spread opinion about stabilization of a DNA double helix where a dominant contribution was expected to originate in the hydrogen bonding.[6,9]

Despite the great importance of these molecular systems from a biological point of view, relatively little attention has been devoted to the discussion of the nature of interactions binding them. Inspiration for this study has been the recent paper published by Hobza and co-workers.[10] They authors have found that dispersion energy in stacked and hydrogen-bonded complexes is of similar magnitude. A partial explanation has been drawn from comparison of the results with dispersion energies calculated by the empirical C_6/r^6 formula. The DFT-SAPT and empirical dispersion energies agreed well, and the authors argued that "our intuitive perception of the dispersion interaction is not accurate and very close contacts in the hydrogen-bonded complexes can bring as much dispersion stabilization as is found in the stacking molecules".[10]

The supermolecular ab initio calculations have become very popular tools for the investigation of intermolecular interactions. However, this method does not provide the information on the character of these interactions. To get more insight into the physical nature of the interaction one can use the interaction-energy decomposition scheme[11–15] or the perturbation scheme known as the symmetry-adapted perturbation-theory (SAPT) method.[16] The first step in such an investigation is to find the optimal structure on the intermolecular potential energy surface (IPES) by a supermolecular method, while the second stage is to perform the SAPT calculations for the minimal configuration.

The basis-set superposition error (BSSE) arising from the use of an incomplete basis set can strongly influence calculated binding energies of weakly bound complexes. The Boys−Bernardi counterpoise (CP) correction is a simple procedure for estimating the size of the basis-set superposition error in the interaction energy calculations.[17] The effectiveness of the CP procedure tested on a broad range of interacting systems supports this approach for hydrogen-bonded complexes.[18]

The counterpoise correction in the geometry optimization procedure increases the H-bond distance.[19,20] BSSE introduces a nonphysical attraction between the two units. Thus, the counterpoise correction generally makes intermolecular complexes less stable with longer intermolecular distances than

* Corresponding author e-mail: sadlej@chem.uw.edu.pl.

10.1021/ct800067m CCC: $40.75
Published on Web 05/15/2008

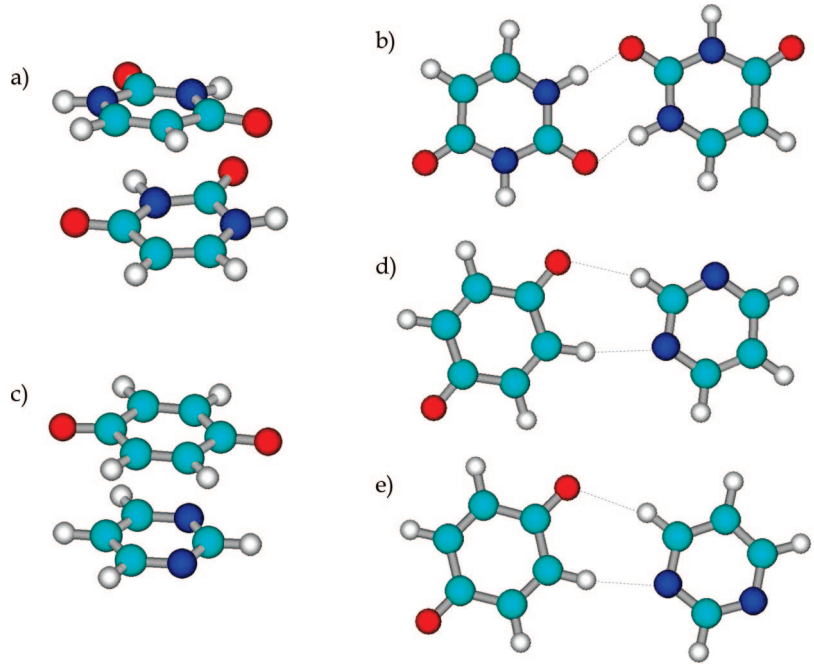


Figure 1. The structures of the uracil dimer: stacked (a) and hydrogen-bonded (b) and of the pyrimidine···*p*-benzoquinone complex: stacked (c) and hydrogen-bonded: planar A (d) and planar B (e) are optimized at the MP2/aug-cc-pVDZ level.

apparent from the normally optimized structure.[21,22] Optimization on the standard IPES using medium-size basis sets leads sometimes to a completely wrong geometric structure, while optimization on a counterpoise-corrected IPES yields the correct structure.[20] Counterpoise-corrected surfaces lead to noticeable improvements in the geometrically optimized transition states obtained with relatively small basis sets.[23] The geometries and associated activation energies become closer to those obtained with larger basis sets.

The object of this Letter is 2-fold. The first goal is to investigate the systems with hydrogen bonds and stacking arrangement in terms of physically meaningful contributions to interaction energy with special regards to the ratios of electrostatic and dispersion terms. The second purpose, a methodological one, is to study how counterpoise correction influences the geometries of the interacting systems and the calculated SAPT decomposition of the total interaction energy.

As model systems we have chosen two dimers which can form hydrogen-bonded as well as stacking conformers. Since all the calculations are time-consuming, we have selected two relatively small systems: the uracil dimer and the pyrimidine···*p*-benzoquinone complex, Figure 1. The uracil dimer is well-known to form several different structures among them the hydrogen-bonded and *face-to-face* stacked are the most studied.[3,24–28] The *face-to-face* stacked structure is lower in energy than the second possible stacked structure of the dimer, the *face-to-back*.[26,28] The pyrimidine···*p*-benzoquinone complex was studied previously[29] in a search for a system forming a stable stacked complex. Quinone was selected as one of the components in the complex because it can form only weak C—H···O and C—H···N hydrogen bond interactions with pyrimidine in the planar dimers which are presumably not strong competitors with the stacking interactions. The symmetrical quinone molecule was also chosen since it can form a very limited number of different dimer conformers with pyrimidine.[29] However, a limited analysis of harmonic frequencies with an experimental infrared data has not confirmed undoubtedly the presence of a stacking structure in low-temperature argon matrices since the energy differences between all conformers are rather small.[29]

The structures of the uracil dimer and the pyrimidine···*p*-benzoquinone dimer were optimized by means of the frozen-core Møller—Plesset (MP2) perturbation theory with the aug-cc-pVDZ basis set.[30,31] We employed two optimization procedures: with counterpoise corrections (CP) and without (no CP) for the basis-set superposition error. At each stationary point vibrational frequencies were calculated in order to confirm the nature of the stationary points on the potential energy surface. Geometry optimization and vibrational analysis were carried out with the Gaussian 03 program package.[32]

The interaction energy components were calculated by means of the SAPT method[16,33] implemented in the SAPT2002 program.[34] In this study we have employed the SAPT0 approach which is recommended for large systems since it takes much less time than the full SAPT calculation.

In this work we employed the following approximation to the intermolecular interaction energy

$$E_{\mathrm{int}}^{\mathrm{SAPT0}} = E_{\mathrm{elst}}^{(10)} + E_{\mathrm{exch}}^{(10)} + E_{\mathrm{ind,\,resp}}^{(20)} + E_{\mathrm{exch-ind,}\,r}^{(20)} + E_{\mathrm{disp}}^{(20)} + E_{\mathrm{exch-disp}}^{(20)} \quad (1)$$

where $E_{\mathrm{elst}}^{(10)}$ is the classical (Coulombic) electrostatic energy, $E_{\mathrm{exch}}^{(10)}$ is the exchange term that results from the antisymmetrization (symmetry adaptation) of the wave function, $E_{\mathrm{ind,resp}}^{(20)}$ denotes the induction (with response) energy, $E_{\mathrm{exch\text{-}ind},r}^{(20)}$ is the second-order exchange-induction (with response) energy term, $E_{\mathrm{disp}}^{(20)}$ is the dispersion energy, and $E_{\mathrm{exch\text{-}disp}}^{(20)}$ denotes the exchange-dispersion contribution. This level of theory may introduce about 20−30% errors with respect to the exact interaction energies, but we believe such an accuracy can be acceptable for large systems.

Table 1. Calculated SAPT and SAPT(DFT) Decomposition (See Definitions in Text) of Interaction Energy (kJ/mol) for Stacked and Hydrogen-Bonded Systems[a,b]

	stacked			hydrogen-bonded		
	SAPT0		SAPT(DFT)	SAPT0		SAPT(DFT)
uracil dimer	CP	no CP	CP	CP	no CP	CP
$E_{elst}^{(10)}$	−39.99	−53.64	−37.55	−128.00	−142.02	−119.34
$E_{exch}^{(10)}$	46.89	79.27	49.07	109.20	133.88	123.81
$E_{ind,resp}^{(20)}$	−19.72	−34.79	−20.93	−57.68	−69.68	−67.62
$E_{exch\text{-}ind,r}^{(20)}$	14.83	28.15	15.83	26.86	33.58	32.67
$E_{disp}^{(20)}$	−55.98	−75.01	−48.92	−38.52	−43.56	−40.39
$E_{exch\text{-}disp}^{(20)}$	5.59	9.19	4.74	6.56	7.80	7.10
E_{int}^{SAPT}	−48.39	−46.87	−37.76	−81.58	−80.00	−63.77
E_{int}^{MP2}	−42.21	−39.80		−79.70	−80.68	

	stacked			hydrogen-bonded					
				planar A			planar B		
	SAPT0		SAPT(DFT)	SAPT0		SAPT(DFT)	SAPT0		SAPT(DFT)
pyrimidine···*p*-benzoquinone dimer	CP	no CP	CP	CP	no CP	CP	CP	no CP	CP
$E_{elst}^{(10)}$	−18.22	−30.56	−15.07	−24.52	−29.30	−23.19	−28.91	−33.84	−27.21
$E_{exch}^{(10)}$	37.39	68.11	33.36	20.90	29.56	25.09	23.18	32.27	27.61
$E_{ind,resp}^{(20)}$	−16.23	−31.32	−13.38	−7.38	−10.08	−8.95	−8.40	−11.34	−10.07
$E_{exch\text{-}ind,r}^{(20)}$	14.65	28.78	11.89	3.65	5.43	5.10	4.07	5.93	5.48
$E_{disp}^{(20)}$	−47.10	−64.93	−36.23	−15.73	−18.93	−16.34	−16.83	−20.16	−17.44
$E_{exch\text{-}disp}^{(20)}$	5.36	9.01	3.54	1.73	2.36	1.95	1.92	2.55	2.11
E_{int}^{SAPT}	−24.13	−20.93	−15.88	−21.32	−20.98	−16.35	−25.02	−24.60	−19.52
E_{int}^{MP2}	−24.35	−22.07		−17.35	−16.92		−20.64	−20.20	

[a] The CP and no CP abbreviations refer to the counterpoise corrected and standard optimized geometries, respectively. [b] The total MP2 interaction energies (E_{int}^{MP2}) are counterpoise corrected.

To compare our results with previously published literature[6–8] we performed the SAPT(DFT)[35] (accordingly to the original notation) calculations where the total interaction energy is defined as a sum:

$$E_{int}^{SAPT(DFT)} = E_{elst}^{(1)}(KS) + E_{exch}^{(1)}(KS) + E_{ind}^{(2)}(CKS) + \tilde{E}_{exch-ind}^{(2)}(CKS) + E_{disp}^{(2)}(CKS) + \tilde{E}_{exch-disp}^{(2)}(CKS) \quad (2)$$

The corrections $E_{elst}^{(1)}(KS)$ and $E_{exch}^{(1)}(KS)$ are obtained by using Kohn−Sham orbitals in the expressions for $E_{elst}^{(10)}$ and $E_{exch}^{(10)}$, respectively. The second-order induction and dispersion energies, $E_{ind}^{(2)}(CKS)$ and $E_{disp}^{(2)}(CKS)$, are calculated in the coupled Kohn−Sham approximation. The exchange-induction energy is estimated by scaling the SAPT(KS) result $E_{exch-ind}^{(2)}(KS)$ using $E_{ind}^{(2)}(UCKS)$ calculated in the uncoupled KS approximation:

$$\tilde{E}_{exch-ind}^{(2)}(CKS) = E_{exch-ind}^{(2)}(KS)\frac{E_{ind}^{(2)}(CKS)}{E_{ind}^{(2)}(UCKS)} \quad (3)$$

The exchange-dispersion term, $E_{exch-disp}^{(2)}(CKS)$, is estimated similarly as the exchange-induction energy.

Because of the system size the SAPT and SAPT(DFT) calculations were carried out in aug-cc-pVDZ basis set. This basis set is large enough to provide a reliable estimation of the energy terms.[36] The dispersion term is underestimated by about 10−15% in this basis but should serve well enough for a qualitative purpose. Also our previous calculations for dihydrogen-bonded systems[37] proved using of this basis set to recognize the correct proportion between the most important contributions to the total interaction energy.

The hydrogen-bonded structures of both dimers are minima on the standard and CP-corrected PESs. The stacked structure of the uracil dimer is also a minimum, while the stacked structure of the pyrimidine···*p*-benzoquinone dimer turned out to be a transition state with one imaginary frequency (−51 cm^{-1} for the CP-corrected and −68 cm^{-1} for the standard PES).

The differences between intramolecular geometric parameters (tables in the Supporting Information) for the CP-corrected and standard optimized structures are small, especially, for bond lengths. The intermolecular distances differ significantly, and, as expected,[19,20,22] the CP-corrected optimized are longer than the standard ones. The largest deviations are ca. +0.20 Å for the stacking structures, and the differences in hydrogen bond lengths are smaller.

The calculated decomposition of the interaction energy is presented in Table 1 along with the all-electron MP2 counterpoise-corrected supermolecular energies for the complexes. The SAPT0 interaction energies agree reasonably with the MP2 ones. However, at this level of SAPT one cannot expect accurate agreement of the results, and the SAPT0 energies are mostly lower than the respective total MP2 calculated energies. The following discussion shall be given with respect to the results for the CP-optimized structures.

The interaction-energy decomposition for hydrogen-bonded structures is similar, and the dominant attraction energy originates in the electrostatic term which is larger for the uracil dimer than for the pyrimidine···*p*-benzoquinone complex. The exchange repulsion compensates the attractive electrostatic term, but the first-order energy is still attractive and for the hydrogen-bonded uracil dimer it arises to ca. −18.8 kJ/mol. This sum in the pyrimidine···*p*-benzoquinone complexes is much smaller with values of ca. −3.6 and −5.7 kJ/mol. The second-order induction energy is attractive, and in the case of the uracil dimer

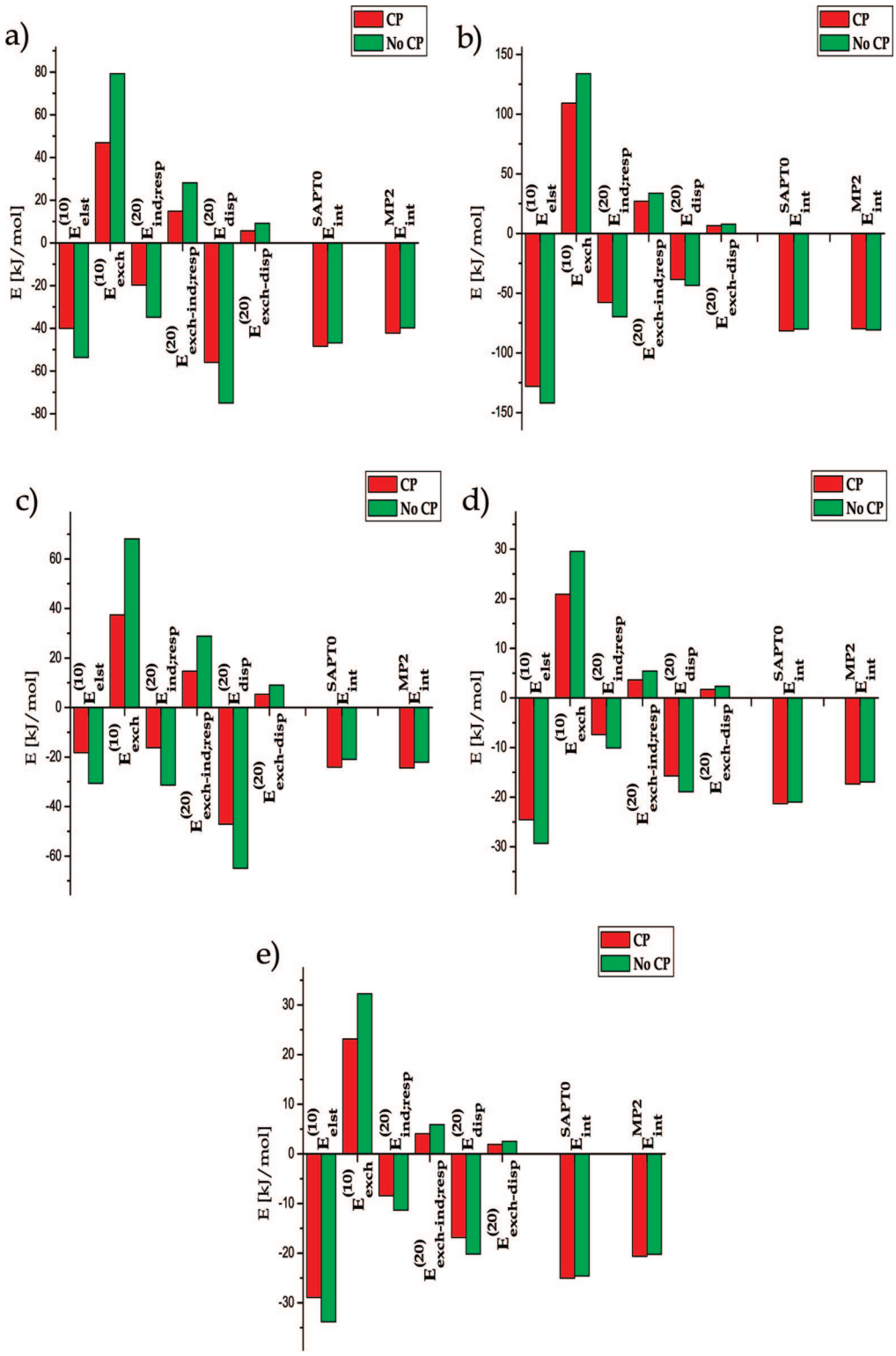


Figure 2. The SAPT0 decomposition of the total interaction energy for the uracil dimer: stacked (a) and hydrogen-bonded (b) and for the pyrimidine···*p*-benzoquinone complex: stacked (c) and hydrogen-bonded: planar A (d) and planar B (e) level.

this attractive contribution to the total interaction energy is approximately twice as large as that for both the pyrimidine···*p*-benzoquinone complexes (71% vs ca. 35%). However, this contribution is reduced by the repulsive $E^{(20)}_{\text{exch-ind},r}$ term which is approximately half of the $E^{(20)}_{\text{ind,resp}}$ absolute value in both systems. The dispersion effect, as it is usual for the hydrogen-bonded systems,[13] constitutes ca. 50% of the total interaction energy (the ratio of $E^{(20)}_{\text{disp}}/E^{\text{SAPT0}}_{\text{int}}$ is 0.5 and 0.7 for the uracil dimer and the pyrimidine···*p*-benzoquinone complex, respectively).

Now, let us discuss the stacked systems. Electrostatic terms are attractive here, but, in contrary to the hydrogen-bonded structures, they are not dominant and the exchange contribution is much larger. Thus, the first-order energies are repulsive, and the inclusion of the second-order terms makes the systems stable.

Comparing results for the two stacked structures the first-order energy for the weaker pyrimidine···*p*-benzoquinone complex is almost three times larger than for the uracil dimer. The largest attraction comes from dispersion, and this term is significantly larger in absolute value than for the hydrogen-bonded complexes. The dispersion energy is about twice as much as the electrostatic ($E_{disp}^{(20)}/E_{elst}^{(10)} = 2.6$) or total interaction energy ($E_{disp}^{(20)}/E_{int}^{SAPT0} = 2.0$) for the pyrimidine···*p*-benzoquinone complex. This is the most noticeable difference between the hydrogen-bonded and the stacking structures (it can be easily identified from Figure 2). $E_{ind,resp}^{(20)}$, the third attractive term, despite being quite substantial is roughly compensated by the repulsive $E_{exch\text{-}ind,r}^{(20)}$ contribution.

Our second goal has been the methodological one. The supermolecular counterpoise-corrected MP2 interaction energies for geometries obtained for the CP-corrected structures are systematically lower than for the noncorrected (no CP) ones. The one exception is the hydrogen-bonded uracil dimer; however, the energy differences result from slight differences between both geometries (vide infra). The difference between each pair of E_{int}^{SAPT0} for CP-corrected and standard optimized structures is also small, and the calculated interaction energies are close to each other. The most noticeable differences between the geometries obtained within the counterpoise-corrected and standard procedures are observed for intermolecular bond lengths. Thus, the changes in energetic terms arise mainly due to a variation in a separation between two subunits. These discrepancies are greater for stacked structures, and it is also reflected in the values of SAPT terms.

Similarly, all the energetic terms for the CP-optimized structures are smaller than for these optimized without counterpoise corrections (see Figure 2). The differences are much larger for the stacking structures than for the hydrogen-bonded structures. The largest terms remain dominant, but the order of smaller terms can vary in pairs (CP vs no CP) and the contributions of similar magnitude can change their places in the sequence of increasing energy (for instance, the $E_{elst}^{(10)}$ and $E_{ind,resp}^{(20)}$ terms in the stacked pyrimidine···*p*-benzoquinone complex). On the other hand, relations between some terms are maintained: the ratios $E_{elst}^{(10)}/E_{disp}^{(20)}$, $E_{exch}^{(10)}/E_{ind,resp}^{(20)}$, $E_{exch}^{(10)}/E_{exch\text{-}disp}^{(20)}$, and $E_{ind,resp}^{(20)}/E_{exch\text{-}disp}^{(20)}$ are approximately constant and do not depend on the optimization procedure (CP or no CP).

In conclusion, we would like to stress several points. First, previous studies[7,10] have shown that for Watson—Crick (WC) structures of DNA pairs (for instance, mAmT WC and mCmG WC) the DFT-SAPT calculated sum of first-order electrostatic and exchange energies is repulsive. Our observations are in contrast with these findings since the electrostatic term usually overcompensates exchange energy and this first-order sum is attractive for hydrogen-bonded complexes. The SAPT(DFT) results for our model systems presented in Table 1 show that the first-order energy for these hydrogen-bonded geometries is repulsive. Intramonomer correlation contained in the SAPT(DFT) (or DFT-SAPT) approach affects the values of interaction-energy contributions. Thus, one cannot compare directly the terms calculated by means of these two methods. The electrostatic-exchange sum in SAPT(DFT) (or DFT-SAPT) is not the Heitler—London energy in its pure form as in the original SAPT approach. The value of the Heitler—London energy tells us whether a system is bonded at an uncorrelated level (by electrostatic interaction) or an inclusion of correlation terms makes it stable (among others, by dispersion). The electrostatic-exchange sum along with the induction terms ($E_{ind,resp}^{(20)}$ and $E_{exch\text{-}ind,r}^{(20)}$) provides a good approximation to the supermolecular Hartree—Fock interaction energy.[38] In the SAPT-DFT (or DFT-SAPT) approximation such a picture is indistinct because of the intramonomer correlation integrated within the monomers' electron density.

Second, the comparison of the percentage of dispersion contribution to the total interaction energy makes clear a distinction between hydrogen-bonded and stacking interactions. In the hydrogen-bonded systems the absolute value of the dispersion term constitutes approximately half of the absolute value of the total SAPT0 interaction energy, E_{int}, while in the stacking complexes the absolute value of the dispersion energy is larger than the absolute value of E_{int}. Similar ratios of the dispersion energy to the total interaction energy one can find for the components presented in Table 1 of ref 10, even though the absolute values of the dispersion terms for both hydrogen-bonded and stacking complexes are similar. This observation is also in line with previous works on stacked interactions in the DNA pairs.[7] Thus, the statement in the paper "the similarity of magnitudes of the dispersion energy in stacked and H-bonded complexes is indeed puzzling" seems to be unjustified. The dispersion energies (in absolute values) in our model systems are larger for stacked than for hydrogen-bonded arrangements. However, this cannot be a general trend since the total interaction energies can differ largely (stacking vs hydrogen bonded), and the values of individual terms maintain their relative proportion. The ratio of these energy terms to the total interaction energy seems to be the valid criterion for a discrimination between hydrogen-bonded and stacking systems.

Finally, we would like to compare the SAPT terms calculated for two sets of optimized geometries. The absolute values of the SAPT terms differ for geometries obtained by means of two optimization methods (with CP or without CP optimized geometry). It is especially meaningful for weak systems where the difference between the calculated bond lengths by means of standard and counterpoise-corrected optimization procedures becomes substantial and the interaction-energy terms are well-known to be distance sensitive.[16,27] However, the sequence of individual interaction-energy terms is approximately maintained. The comparison between two sets of differently obtained geometries[7,10] (taken from geometry optimization[7] and in experimental crystal geometries)[10] indicates that despite differences in the total interaction energy which are greater for stacked structures, the sequence of the individual SAPT terms is still maintained.

Acknowledgment. The authors acknowledge the computational Grant G18-4 from the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) of the University of Warsaw. The computational part of this work was also done using the computer cluster at the Computing Center of Faculty of Chemistry, University of Warsaw.

Supporting Information Available: The MP2/aug-cc-pVDZ optimized bond lengths and bond angles for the studied

systems. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dandliker, P. J.; Holmlin, R. E.; Barton, J. K. *Science* **1997**, *275*, 1465.

(2) Hobza, P.; Šponer, J. *J. Chem. Rev.* **1999**, *99*, 3247.

(3) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802.

(4) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104.

(5) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *J. Chem. Phys.* **2002**, *120*, 647.

(6) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.

(7) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Am. Chem. Soc.* **2006**, *128*, 11730.

(8) Fiethen, A.; Jansen, G.; Hesselmann, A.; Schütz, M. *J. Am. Chem. Soc.* **2008**, *130*, 1802.

(9) Rutledge, L. R.; Campbell-Verduyn, L. S.; Wetmore, S. D. *Chem. Phys. Lett.* **2007**, *444*, 167.

(10) Sedlák, R.; Jurečka, P.; Hobza, P. *J. Chem. Phys.* **2007**, *127*, 075104.

(11) Morokuma, K. *J. Chem. Phys.* **1971**, *55*, 1236.

(12) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325.

(13) Chałasiński, G.; Szczęśniak, M. *Chem. Rev.* **1994**, *94*, 1723.

(14) Sokalski, W. A.; Roszak, S.; Pecul, K. *Chem. Phys. Lett.* **1988**, *153*, 153.

(15) Cybulski, S. M.; Chałasiński, G.; Moszyński, R. *J. Chem. Phys.* **1990**, *92*, 4357.

(16) Jeziorski, B.; Moszyński, R.; Szalewicz, K *Chem. Rev* **1994**, *94*, 1887.

(17) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H *Chem. Rev.* **1994**, *94*, 1873.

(18) Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157.

(19) Simon, S.; Bertran, J.; Sodupe, M. *J. Phys. Chem. A* **2001**, *105*, 4359.

(20) Hobza, P.; Havlas, Z. *Theor. Chem. Acc.* **1998**, *99*, 372.

(21) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Phys. Chem. A* **1999**, *103*, 1640.

(22) Hobza, P.; Bludský, O.; Suhai, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3073.

(23) Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2001**, *105*, 1944.

(24) Šponer, J.; Hobza, P.; Leszczynski, J. *J. Phys. Chem.* **1996**, *100*, 5590.

(25) Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7.

(26) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. *J. Phys. Chem. A* **2002**, *106*, 3850.

(27) Czyżnikowska, Ż.; Zaleśny, R.; Ziółkowski, M.; Gora, R. W.; Cysewski, P. *Chem. Phys. Lett.* **2007**, *450*, 132.

(28) Frey, J. A.; Müller, A.; Losada, M.; Leutwyler, S. *J. Phys. Chem. B* **2007**, *111*, 3534.

(29) McCarthy, W.; Plokhotnichenko, A. M.; Radchenko, E. D.; Smets, J.; Smith, D. M. A.; Stepanian, S. G.; Adamowicz, L. *J. Phys. Chem. A* **1997**, *101*, 7208.

(30) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(31) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T. J.; 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 03*, *Revision C.02;* 2004.

(33) Jeziorski, B.; Moszyński, R.; Ratkiewicz, A.; Rybak, S.; Szalewicz, K.; Williams, H. L. *Methods and Techniques in Computational Chemistry: METECC-94*; Clementi,E., Ed.; STEF: Cagliari, 1993; Vol. *B*.

(34) Bukowski, R.; Cenek, W.; Jankowski, P.; Jeziorski, B.; Jeziorska, M.; Kucharski, S. A.; Misquitta, A. J.; Moszyński, R.; Patkowski, K.; Rybak, S.; Szalewicz, K.; Williams, H. L.; Wormer, P. E. S. *SAPT2002: an ab initio program for many—body symmetry—adapted perturbation theory calculations of intermolecular interaction energies. Sequential and parallel versions*; University of Delaware and University of Warsaw, 2003.

(35) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*, 214103.

(36) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.

(37) Cybulski, H.; Pecul, M.; Sadlej, J.; Helgaker, T. *J. Chem. Phys.* **2003**, *119*, 5094.

(38) Jeziorska, M.; Jeziorski, B.; Čižek, J. *Int. J. Quantum Chem.* **1987**, *32*, 149.

CT800067M

JCTC Journal of Chemical Theory and Computation

Microscopic Picture of the Aqueous Solvation of Glutamic Acid

Elske J. M. Leenders, Peter G. Bolhuis, and Evert Jan Meijer*

Van 't Hoff Institute for Molecular Sciences, Universiteit van Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

Received December 19, 2007

Abstract: We present molecular dynamics simulations of glutamic acid and glutamate solvated in water, using both density functional theory (DFT) and the Gromos96 force field. We focus on the microscopic aspects of the solvation—particularly on the hydrogen bond structures and dynamics—and investigate the influence of the protonation state and of the simulation method. Radial distribution functions show that the hydrogen bonds are longer in the force field systems. We find that the partial charges of the solutes in the force field simulations are lower than the localized electron densities for the quantum simulations. This lower polarization decreases the hydrogen bond strength. Protonation of the carboxylate group renders glutamic acid a very strong and stable hydrogen bond donor. The donated hydrogen bond is shorter and lives longer than any of the other hydrogen bonds. The solute molecules simulated by the force field accept on average three more hydrogen bonds than their quantum counterparts. The life times of these bonds show the opposite result: the residence times are much longer (up to a factor 4) in the ab initio simulations.

1. Introduction

Glutamic acid, or its deprotonated equivalent glutamate, is one of the 20 natural amino acids. Solvated in water, it exists in zwitterionic form, see Figure 1. Next to being a building block in almost all proteins, glutamate also serves as a ligand in receptor proteins.[1] Our interest in this molecule arose as a result of our previous simulation research on the protonation reaction in the photocycle of the photoactive yellow protein (PYP).[2] This signaling protein has a chromophore[3,4] that responds to UV light and isomerizes from the cis to the trans state. We studied the solvation of this chromophore in detail.[5] After the isomerization, the chromophore is protonated, most likely by the nearby glutamic acid. When glutamic acid loses a proton, it takes up a negative charge. We showed that the reactive event involves the enhanced stabilization of glutamate by hydrogen bonds.[2] In the ground state of the protein, glutamic acid takes part in only one hydrogen bond: the one it donates to the negatively charged chromophore. When deprotonated, glutamate is only stable when it accepts at least three hydrogen bonds. This implies that upon proton transfer, hydrogen bonds in the environment of glutamate have to rearrange. They can be donated not only by surrounding amino acids or the chromophore but also by water molecules that penetrate the chromophore binding pocket of the protein.[6] We found that the hydrogen bond rearrangements occur easier and faster in force field molecular dynamics simulations of the protein, compared to QMMM (quantum mechanics/molecular mechanics) simulations that simulate glutamate and some of its possible hydrogen bond donors with quantum methods and the environment (protein and water molecules) with a force field.

These findings underline the importance of a proper description of the hydrogen bonding of glutamate and glutamic acid (both abbreviated as Glu) in the modeling of PYP. We therefore study the solvation of Glu in water, the prototype hydrogen bonding environment. Sun et al.[7] studied Glu in the gas phase with different quantum methods. Prabhakar[8] simulated a small part of the Glu side chain in a model system with three water molecules in the gas phase. Both systems are too small to see hydrogen bond exchanges. Experimental and quantum simulation studies on glutamic

* Corresponding author phone: +31-20-525-5265; fax: +31-20-525-5604; e-mail: ejmeijer@science.uva.nl.

10.1021/ct700344f CCC: $40.75
Published on Web 05/15/2008

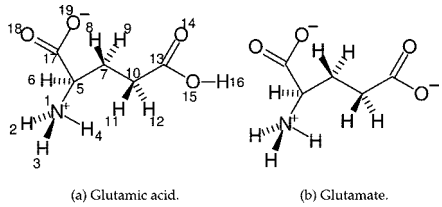


Figure 1. Glutamic acid (a) and glutamate (b). The two molecules differ in the protonation state of oxygen 15. At neutral pH, the amino acid is a zwitterion and the side chain is deprotonated, as its pK_a value is 4.07.

acid and glutamate solvated in water focus on spectroscopic properties.[1,9] Nowadays, classical simulations on simple systems as glutamic acid aim to optimize the parametrization of the force field.[10–14] They focus on the force field, not on glutamic acid itself, and they reproduce macroscopic thermodynamic quantities such as the density and the free enthalpies of solvation. To our knowledge, their is no previous work that investigates the microscopic and dynamic properties of glutamic acid and glutamate in solution.

We present in this paper a study on both protonation states of Glu in water, simulated with either a quantum or a force field method. Besides giving us a detailed picture of the hydrogen bonds and their dynamics in solution, this helps us to better interpret force field and QMMM results on PYP in previous and future work. The intention of this work is 1) to provide a microscopic picture of solvated Glu, focusing on aspects that have not or cannot be measured in experiments and 2) to understand what the influence of the simulation method is, so that we can interpret simulations of Glu correctly (not only in solvent but also in a protein environment).

In the next section, we explain how we equilibrated our systems and which simulation methods we used for our production runs. In the Results section, we compare various aspects of the simulations in different subsections and discuss them. In the concluding section, we summarize these results and show their impact on other simulations, such as our calculations on the photoactive yellow protein.

2. Computational Methods. As input structures, we equilibrated glutamic acid and glutamate in 135 SPC waters using the Gromos96 force field[15] (version G43a1) with the Gromacs software package.[16] The waters could form more than one solvation shell even when the molecule was extended. For all our force field simulations we used a time step of 2 fs and a Nosé-Hoover thermostat[17] with $\tau = 0.1$ ps, keeping the temperature at 300 K. The LINCS[18] and SETTLE[19] algorithms constrained the bond lengths and angles. For both systems we started with an NPT simulation of 2 ns, using the isotropic Berendsen barostat[20] with $\tau = 1$ ps and a reference pressure of 1 bar. Using the final configuration as a starting point, we performed NVT simulations of 2 ps with iteratively chosen box sizes. We took the box with the pressure nearest to 1 bar as a starting point for our production runs. For glutamic acid, the cubic box had a size of 16.3395 Å; for glutamate this was 16.2949 Å. Our force field production runs started from here, simulating NVT for 10 ns. We also performed one force field simulation of glutamate in a box of size 37.2350 Å containing 1638 water molecules to check for finite size effects.

Table 1. Four Systems Described in This Paper

system	method	protonation	charge	time step (fs)	simulation time (ps)
QMH	Quickstep	protonated (glutamic acid)	0	0.5	15.632
QM-	Quickstep	deprotonated (glutamate)	−1	0.5	15.593
FFH	force field	protonated (glutamic acid)	0	2	10000
FF-	force field	deprotonated (glutamate)	−1	2	10000

From the same starting structures, we performed quantum molecular dynamics simulations based on density functional theory (DFT) using Quickstep,[21] which is part of the CP2K program package.[22] It uses a Born−Oppenheimer molecular dynamics (BOMD) algorithm, meaning that it calculates the ground-state electron density every time step and from that the forces on the ions. It employs a hybrid Gaussian and plane waves (GPW) basis set,[23] which makes efficient and accurate density functional calculations of large systems (up to 1000 atoms) possible. We used Goedecker-Teter-Hutter (GTH) pseudopotentials[24,25] and the TZV2P basis set.[21] As it serves well for simulating liquid water,[26] we used the BLYP functional.[27,28] The density cutoff for the plane-wave basis set was 280 Ry and the time step 0.5 fs. A Nosé-Hoover chain thermostat[29] with a chain length of 3 and a time constant of 1000 fs fixed the temperature at 300 K. Every SCF step, the energy was converged up to 1.10^{-5} Hartree. Every time step, we recorded the positions, velocities, and Mulliken charges of all the atoms. We simulated for almost 18 ps; the first 2 ps were considered to be equilibration time and not included to calculate the properties. Izvekov et al.[30] showed that there is no velocity autocorrelation anymore after 1 ps in a quantum simulation of 64 water molecules. We checked the radial distribution functions and Mulliken charges in our simulations as well, and they are converged within 2 ps.

We studied glutamic acid and glutamate in water. For every protonation state of this molecule, one simulation was performed using the force field, another using the quantum mechanical method. In Table 1, we add symbols to these four systems, which we will use throughout this paper.

3. Results

In this section we will discuss the differences between various parameters calculated in the four different systems. First we will discuss the charges localized on the atoms. Then we will compare the hydrogen bond structures around the amino acid molecules and finally the dynamics of these bonds.

3.1. Charges. One of the main differences between force field and ab initio simulations is that the former uses fixed (partial) charges on atoms, while the latter calculates the electron density and hence the charge distribution every time step on the fly. In this section we check the difference between these fixed charges and the calculated ones, on the

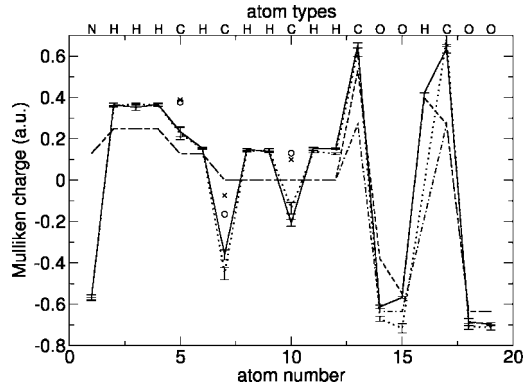


Figure 2. Mulliken charges on glutamic acid (system QMH, solid line) and glutamate (system QM-, dotted) and fixed force field charges for the systems FFH (dashed) and FF- (dashed-dotted). The numbers of the atoms agree with Figure 1. We also summed the Mulliken charges on the CH and CH_2 groups, to enable comparison with the force field, in which these hydrogens are implicit. For system QMH, the charge on these groups is represented by crosses, for system QM- by circles. The error bars give 1.96 times the standard deviation of the average of ten bins.

amino acid atoms and the water atoms. We distinguish between waters in the first solvation shell and bulk waters.

Figure 2 shows the average Mulliken charges on the atoms of the ab initio molecules, compared to the fixed charges of the force field. The force field does not include nonpolar hydrogens; the charge given is the charge of the complete CH or CH_2 group. Noticeable is that the two quantum simulations hardly differ in their partial charges. Only the charge on the two oxygens in the side chain carboxylate group differs, which can be expected as this is the group that is protonated in QMH and deprotonated in QM-. The differences however are very small. For most atoms, the force field charges are closer to zero than their quantum equivalents. This lower degree of polarization can lead to weaker hydrogen bonds in the force field simulations. Note that Mulliken charges are only one way to attribute charge densities to specific atoms; other methods might give somewhat different results. Although the force field partial charges are to some extent based on ab initio simulations, they have been adjusted to fit thermodynamic properties,[31] not to represent the charge density as good as possible. This being said, other researchers also saw the problem of the low polarization. Villa et al.[11] found too low values for the hydration free energies of many neutral amino acids with the Gromos96 force field. In that case, the results were in better agreement with experiments when all partial charges were multiplied with a factor 1.1.

Furthermore, while the N atom is very negatively charged in the DFT simulations (almost −0.6), it has a small positive charge of 0.13 in the force field. Although the hydrogen-bonding properties of the amino group are not the subject of this paper, this charge difference must influence these properties to some extent.

Table 2 shows the total charge on the Glu molecules. It shows that glutamic acid in system QMH has a small but significant positive charge. In the QM- case, the negative charge is not completely on the glutamate molecule but

Table 2. Total Charge on Glu in the Four Different Systems[a]

system	charge
QMH	0.0795 ± 0.0090
QM-	-0.643 ± 0.015
FFH	0
FF-	−1

[a] The error for the quantum systems is 1.96 times the standard deviation of the average of ten bins; the force field imposes an integer charge on the molecules.

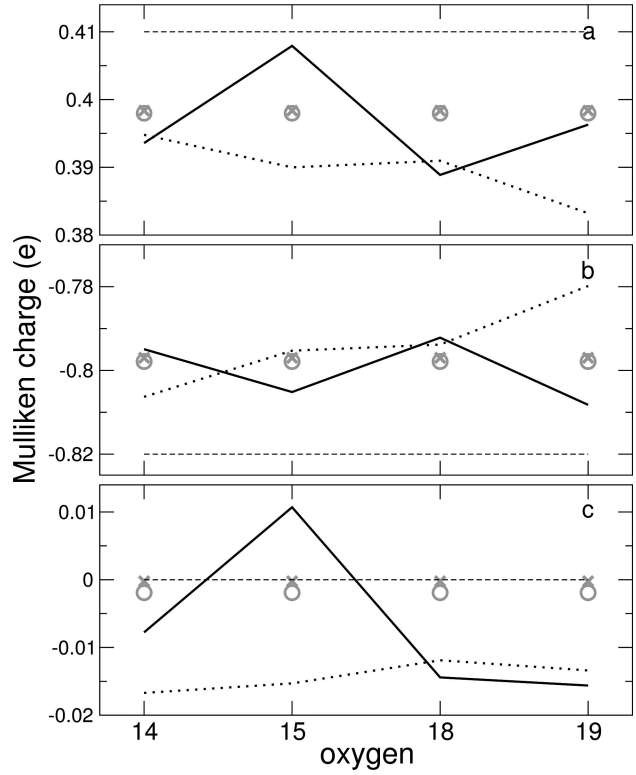


Figure 3. (Mulliken) charges on the water hydrogens (a) and the water oxygens (b) and the total charge per molecule (c). The dashed line is the constant charge from the force field simulations. The solid line represents the first solvation shell waters of QMH, the dotted line of QM-. The charges on bulk water molecules are given by the grey crosses (QMH) and circles (QM-). Note that the scales on the *y* axes differ. The indices on the *x* axis refer to the oxygen to which the waters hydrogen bond.

spread over the waters as well. Figure 3 is a visualization of the 'charge transfer' to the water molecules. It shows the (Mulliken) charges for the water molecules in the force field simulations and the quantum simulations, in the bulk and in the first solvation shell around the amino acid oxygens. Both the bulk and the first solvation shell waters of the QM-system are negatively charged, as could be expected from Table 2. The force field hydrogens in water are more positive than their quantum counterparts, the water oxygens more negative. Hence, force field water molecules are more polarized. This is in contrast with the solute molecules, that are more polar in the ab initio simulations.

Oxygen 15 is the protonated one in system QMH. The OH group influences the Mulliken charges of water molecules in its first solvation shell: only these molecules have

a positive charge, which is mainly caused by a larger positive charge on the water hydrogens (the oxygen is even more negative than in the bulk waters). This suggests that when glutamic acid donates a hydrogen bond, it transfers a part of the charge on the very positive hydrogen 16 to the accepting water molecule. This water molecule shows a higher polarization than bulk waters in the same simulation.

Another remarkable point is that oxygens in the same carboxylate group do not always show similar behavior. If we look at oxygen 18 and 19, the total charge of the water molecules in their solvation shell is almost the same, but the atomic charges vary. In QMH, we see that the water molecules around oxygen 18 are less polarized than bulk waters, but they are more polarized around oxygen 19. In QM-, it is exactly the other way around. This cannot be explained by the Mulliken charges on the oxygen atoms in Figure 2, as they are very similar. It is most likely an effect of different solvatation, perhaps because one of the groups is more sterically hindered than the other. Or, as the effect is opposite in the two different simulations, it can be accidental: a long-lasting hydrogen bond happens to point toward one of the oxygens, resulting in less space for possible hydrogen bond donors pointing toward the other. If this is the case, this is an unphysical effect caused by the short simulation times reachable with dynamical DFT simulations. Oxygens 14 and 15 also affect their first solvation shell in different ways in QM-, although less pronounced; we cannot compare this with the QMH simulation, as the protonation of oxygen 15 has a major influence. We will come back to this issue in section 3.2.2.

3.2. Structure. We compare the water structures around the Glu oxygens through the radial distribution functions of water oxygens and hydrogens. We also count the coordination numbers of water oxygens around the Glu oxygens, averaged over all frames. We saw no finite size effects in the structure calculations. The radial distribution functions from the small and large box are positioned exactly on top of one another; the maximum difference in the coordination numbers is 1%.

3.2.1. Radial Distribution Functions. In the FF- simulations, the structures around oxygens 14 and 15 are the same. The same holds for oxygens 18 and 19. For this reason, we present only the oxygen 14 and 18 results in Figure 4. The peaks in the radial distribution function for both carboxylate groups appear at the same distance but are more pronounced for the side chain oxygens. For the protonated case, oxygen 18 and 19 behave the same; their radial distribution functions are almost equal to the FF- results. However, the structures around oxygens 14 and 15 differ considerably (from each other and from the deprotonated results), see Figure 5. Both first hydrogen peaks are hardly present; the oxygen peaks are less pronounced than the oxygen peaks around oxygen 18 and 19 but still very clear. The oxygen−oxygen distance around oxygen 15 is much smaller than around oxygen 14. This indicates a very strong hydrogen bond between the protonated oxygen 15−acting as a hydrogen bond donor−and an accepting water molecule. It is likely that this strong bond disrupts the preferred water structure around oxygen 14, resulting in a less stable water network around this atom.

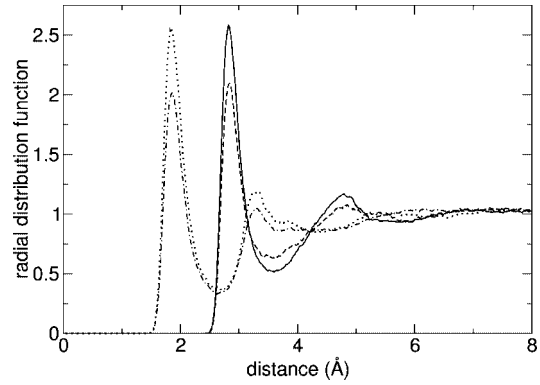


Figure 4. Radial distribution functions of water around Glu oxygens in system FF-. In both carboxylate groups, the two oxygens have the same distribution function. The solid line represents water oxygens around oxygen 14 (and 15), the dotted line water hydrogens around this oxygen. The dashed line shows water oxygens around oxygen 18 (and 19), the dashed-dotted lines the water hydrogens. All radial distribution functions in this paper are calculated over 1000 bins, presenting the results as running averages over 10 data points to smoothen the graphs.

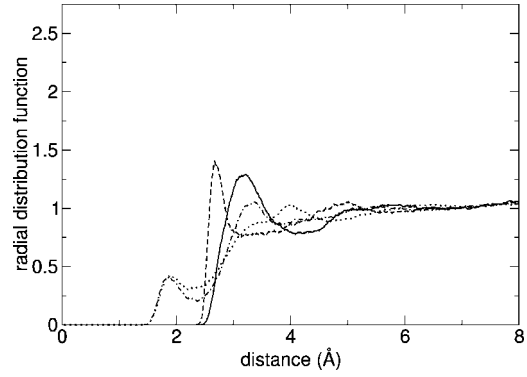


Figure 5. Radial distribution functions of water around the Glu oxygens in system FFH. The solid line represents water oxygens around oxygen 14, the dotted line water hydrogens around this oxygen. The dashed line shows water oxygens around oxygen 15, the dashed-dotted lines the water hydrogens. Around oxygen 18 and 19, the radial distribution function is exactly the same as for system FF- in Figure 4.

For the quantum simulations, we present the radial distribution functions in a different format: Figure 6 shows the water structure around oxygen 14 for both systems, Figure 7 around oxygen 15. For both graphs, the peaks are much more pronounced in system QM-. Around the nonprotonated oxygen 14, both simulations show their first peaks at the same distances. But for the QMH simulation, the first peaks are much lower and the second peaks are hardly visible, while they are clearly there for the QM- system. Apparently, the negative charge on the carboxylate group in system QM- stabilizes the water structure around it to a great extent. The graphs around oxygen 15 differ even more. The radial distribution functions around oxygen 14 and 15 are very similar in system QM-, although they are all slightly more prominent in the latter case. The first hydrogen peak around oxygen 15 is completely missing in system QMH (it is there in system FFH, though very small). This means that there are no water molecules that donate a hydrogen bond

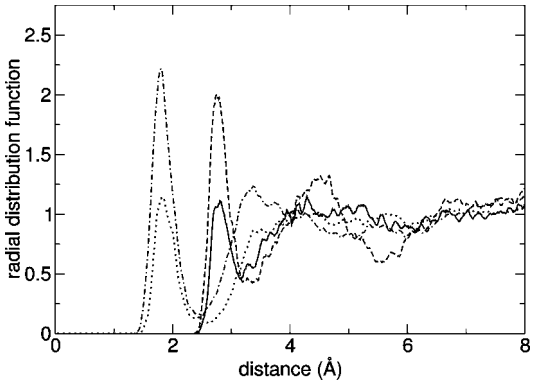


Figure 6. Radial distribution functions of water around oxygen 14 in both quantum simulations. The solid line shows the water oxygens and the dotted line the hydrogens around oxygen 14 in QMH. The dashed and dashed-dotted lines respectively show the same for QM-.

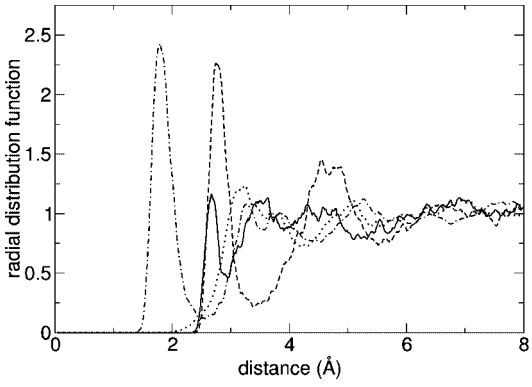


Figure 7. Radial distribution functions of water around oxygen 15 in both quantum simulations. The solid line shows the water oxygens and the dotted line the hydrogens around oxygen 15 in QMH. The dashed and dashed-dotted lines respectively show the same for QM-.

to it; the existence of a small oxygen peak shows that the glutamic acid OH group does take part in a hydrogen bond but as a donor. Because it can only donate one hydrogen bond, the surface below the peak is small. We will come back to this when we discuss the coordination numbers in section 3.2.2.

As the radial distribution functions around oxygen 18 and 19 in the quantum simulations are almost all the same, we present them in one figure, Figure 8. For the first peaks, both for water hydrogen and oxygen, the lines are very similar. Some dissimilarity occurs for the second solvation shell and further.

In the final radial distribution we present here, we compare the distribution of water around proton 16 in QMH and FFH (Figure 9). We see that the oxygen graphs show the same first peak, but the second peak is much clearer in the quantum simulation. The hydrogen peak is also sharper, but somewhat smaller, in this simulation compared to the force field system. The charge transfer of hydrogen 16 to water (see Figure 3) enhances the polarization of the water molecules in its solvation shell and hence strengthens the hydrogen bonds with the second solvation shell. Together with Figure 7, that lacks a peak for hydrogen around oxygen 15, we can conclude that the quantum oxygen 15 is chiefly a hydrogen

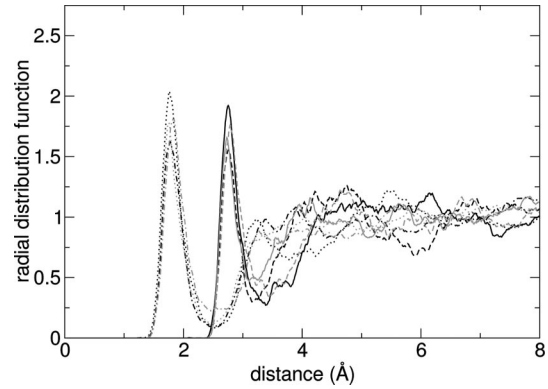


Figure 8. Radial distribution functions around oxygens 18 and 19 in the quantum simulations. The black lines refer to system QMH, the grey lines to system QM-. The solid lines show water oxygens and the dotted lines water hydrogens around oxygen 18. The dashed and dashed-dotted lines respectively show the same for oxygen 19.

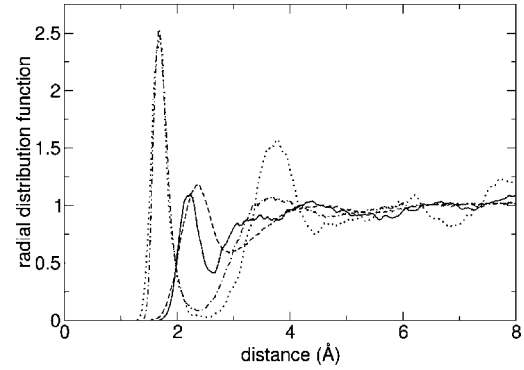


Figure 9. Radial distribution functions around hydrogen 16. The solid line shows the water hydrogen and the dotted line the water oxygen distribution in system QMH. The dashed and dashed-dotted lines respectively show the same in system FFH.

Table 3. Oxygen−Oxygen Distances for All Hydrogen Bonds, Taken from the Radial Distribution Functions[a]

atom	QM-	QMH	FF-	FFH
O_{14}	2.75 (3.22)	2.78 (3.17)	2.82 (3.59)	3.19 (4.09)
O_{15}	2.77 (3.39)	2.67 (2.95)	2.82 (3.59)	2.67 (3.14)
O_{18}	2.73 (3.11)	2.76 (3.39)	2.83 (3.56)	2.83 (3.51)
O_{19}	2.79 (3.44)	2.76 (3.17)	2.83 (3.56)	2.83 (3.51)
H_{16}		1.68 (2.48)		1.68 (2.37)

[a] All values are in Å. First we give the first maximum of the distribution function and after that between brackets the distance at the first minimum. Only in the case of proton 16, we give an oxygen−hydrogen distance (from the water oxygen to the Glu proton).

bond donor when protonated and hardly accepts any hydrogen bonds from water.

Table 3 gives an overview of the first peaks in all radial distribution functions. Overall, almost all hydrogen bond distances are slightly shorter in the quantum simulations. Also, the peaks are wider in the force field simulations. This is particularly the case for oxygen 14 in the protonated simulations. Only the donated hydrogen bond from oxygen 15 and proton 16 to water has the same length with both methods. This is in accordance with the lower polarization

shown in section3.1 for the force field, which results in weaker hydrogen bonds. This effect is corrected to a small extent by the somewhat higher polarization in the force field water molecules (see Figure 3). Note however that the *y* axis in Figure 2 has a much larger scale than in Figure 3: the polarization changes in the solute are a factor 3−10 times larger than in the solvent. It is likely that the newest parameter sets of the Gromos96 force field (53A5 and particularly 53A6) will show somewhat improved results for the hydrogen bond lengths. For these parameters, special attention is paid to fit hydration and solvation better, resulting in a remarkable improvements on the free enthalpies of solvation in water. For glutamic acid for example, this enthalpy is only 0.2 kJ/mol off the experimental value of −27.0 kJ/mol; version 43A2 gave a value of −16.2 kJ/mol.[10]

Unfortunately, we cannot compare these radial distribution numbers to experiments, as they have not been measured for water around Glu. What we can do−and we will make this comparison later in this paper as well−is look at what pure water simulations and experiments teach us. Pure water has been studied extensively with experiments, force field and quantum simulations; radial distribution functions are available for all three methods. The first peak is found at a distance of 2.73 Å with two different experimental methods.[32,33] With quantum methods this distance is between 2.69 Å and 2.78 Å,[33−36] for force field simulations between 2.69 Å and 2.86 Å.[33,34,37] The simulation methods closest to ours predict 2.75−2.78 Å for SPC water simulations[33,37] and 2.75 Å for a BOMD-BLYP simulation.[36] Most of the calculated values are close to the experimental ones; force field distances are on average somewhat more overestimated. Although the location of the peaks is predicted quite well, in many simulations the peaks are a bit too sharp and the first minima too deep. This overstructuring is more common in ab initio calculations than in force field tests. If we translate this result to our simulations−although we cannot be sure that solvation around Glu shows the same trends−we can conclude that the hydrogen bond distances are slightly better predicted by the QM simulations and the peak shape is better fitted by the force field.

3.2.2. Coordination Numbers. By calculating the coordination numbers, we know how many water molecules form a first solvation shell of a Glu oxygen. For every frame of our simulations, we count the number of water oxygens within the first solvation shell around each oxygen. For this purpose, we use a distance criterion taken from the radial distribution functions: the minimum after the first oxygen peak, with a maximum of 3.5 Å. To make sure that we count waters close to both oxygens of a carboxylate group only once, we assign it to the nearest Glu oxygen. Hence the coordination number for a certain oxygen is roughly the same parameter as the number of waters hydrogen bonding to it.

Figure 10 gives an overview of all average coordination numbers. Here we see that the total number of hydrogen bonds in the quantum simulations is always about 3 lower than in the corresponding force field simulation. The reason might be that the formation of hydrogen bonds in the force field is only induced by the (negative) charge on the oxygens. In the quantum simulation, the orbitals involved in the

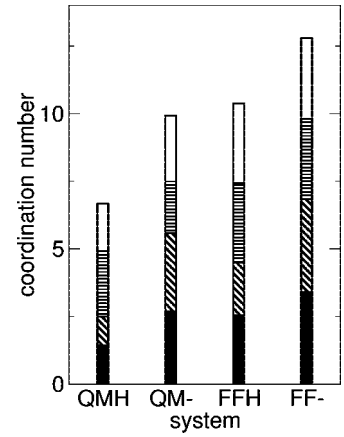


Figure 10. The coordination numbers for all systems and all Glu oxygens. The black bars give the average coordination number around oxygen 14, the diagonally striped bars around oxygen 15, the horizontally striped bars around oxygen 18, and the white bars around oxygen 19. The bars add up to give the total coordination number per amino acid.

hydrogen bonding have a distinct spatial distribution; hence, water cannot approach from any direction. Particularly, a carboxylate group has a resonance structure, with three lone pairs on one oxygen and two on the other. Because the force field cannot distinguish between the two, each oxygen can accept three hydrogen bonds. In a quantum simulation, the two states (one oxygen with a single bond to the carbon and one with a double bond) are distinguishable. On average, the electronic structure should be symmetric, but in a single time step it is not. We see this in our simulations when we look at the coordination and the C−O distance in each frame: the carboxylate oxygen with the lowest coordination number is usually the one with the shortest C−O distance and vice versa. This shows that the oxygen with the longer single bond to the carbon atom has more lone pairs. There is a similar pattern for the Mulliken charges on the oxygens: the oxygen with the higher coordination number has a more negative charge. The charge difference is small however, less than 0.1 e.

The average electronic structure is not perfectly symmetrical: the bars in Figure 10 do not have the same length for oxygens in the same carboxylate group, due to the fact that the exact electronic structure depends on the surrounding water structure as well. Our quantum simulations are not long enough for the water structure to switch numerous times between a coordination number of two and three around each oxygen.

The coordination around proton 16 is not shown in this picture (as it is included in the coordination around oxygen 15): it is 1.0 for both FFH and QMH, indicating that it is practically always donating the one hydrogen bond it can donate. However, the total coordination number of oxygen 15 equals 1 as well for QMH, while it is 2 in FFH. That means that oxygen 15 is an hydrogen bond acceptor as well in the force field simulation, but not in QMH. The protonation of oxygen 15 has more influence in QMH than in FFH, on the coordination of both oxygen 14 and oxygen 15.

The coordination numbers around oxygens 18 and 19 in the force field simulations are all very similar and close to 3. This was found as well by Speranskiy et al.,[1] who did stationary quantum simulations on solvated glutamate equilibrated by classical molecular dynamics, and by Alagona et al.,[38] who performed force field Monte Carlo on an acetate anion. In both these papers, the structure around the solute oxygens was calculated using a force field. The other carboxylate group, with oxygens 14 and 15, behaves in the same way in system FF-. However, for the quantum simulations the total coordination number per carboxylate groups is much less than 6. This is understandable, as the hydrogen bonds to the quantum oxygens are donated to five lone pairs at maximum. A coordination number of 4 per carboxylate group is already less common in the ab initio than in force field simulations. In order to reach high coordination numbers, the water molecules need to find a perfect position, adapting themselves not only to the Glu orbitals but also to the bulk water structure. When only the value of the charge matters, and not its spatiality—as is the case in the force field simulations—high coordination numbers are more common.

Villa et al.[11] found that the force field underestimates the free energy of hydration, as we discussed in section3.1. Our results show that this is not due to the number of hydrogen bonds: all force field simulations show an overcoordination. That means that it is most likely the strength of the bonds that is underestimated; this is in accordance with the longer hydrogen bonds in Table 3 and the lower polarization in Figure 2 for the force field solute. This effect is only compensated to a very small extent by the (much smaller) polarization increase of the force field water molecules shown in Figure 3.

Although the quantum simulations underestimate the coordination numbers compared to the force field, they overestimate the sharpness of the radial distribution peaks, as we showed in section3.2.1. We should actually identify two aspects of structure: 1) the distribution of hydrogen bond lengths (peak shape in the radial distribution function) and 2) the number of hydrogen bonds. Overstructuring of the first type is reported by ab initio studies on pure water, and our simulations show a similar behavior for water around Glu (see Table 3). In contrast, we find that for the second aspect the force field simulations overstructure compared to the ab initio calculations. This result is corroborated by the pure water results of Fernández et al.,[39] who also see a larger coordination number in force field compared to quantum simulations. We did not find any other pure water studies in literature that compare coordination numbers.

3.3. Dynamics. To compare the dynamical aspects of the hydrogen bond networks around glutamic acid and glutamate, we look at the residence times of the hydrogen bonds. Figure 11 shows the residence time of the water molecules around the oxygens of the chromophore. We defined a water molecule to be hydrogen bonded to a chromophore oxygen when the O−O distance was less than 3.5 Å and the O−H−O angle was more than 150°.[40] When a hydrogen bond deviated from this definition for less than 0.5 ps, the hydrogen bond was considered unbroken; when it existed for less than 0.5 ps, we did not take it into account when calculating the average residence time. In the quantum simulations, some residence times are of the same order of magnitude as the simulation times. Because we cannot say how much longer a water molecule would have donated or accepted a hydrogen bond after stopping the simulation, this implies that we can only provide a lower bound for the average hydrogen bond lifetime. The statistics of the force field residence times is much better, resulting in very similar values for oxygens in the same carboxylate group, as one would expect. In the quantum simulations, a single long residence time can influence the average significantly. Steric effects and the orbital structure play a role in the latter simulation too, so at least some part of the difference is physical.

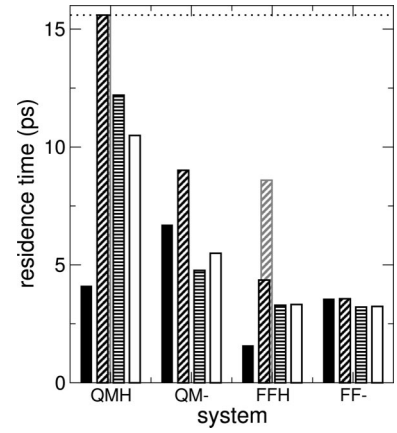


Figure 11. Residence times of water molecules hydrogen bonding to Glu oxygens. The horizontal dotted line indicates the total simulation time for the quantum simulations. The black bars correspond to oxygen 14, the diagonally striped bars to oxygen 15, the horizontally striped bars to oxygen 18, and the white bars to oxygen 19. When proton 16 was present, we represented the residence times of its donated hydrogen bond with the grey version of the oxygen 15 bar (as it is included in that residence time as well). In the QMH simulation, the residence times for oxygens 15 and proton 16 were exactly the same and therefore overlapping. In system FFH, the hydrogen bond donated by proton 16 lived about twice as long as the average hydrogen bond involving oxygen 15.

In spite of the fact that the quantum residence times might be somewhat underestimated, all of them are (much) longer than their force field counterparts. Especially in the protonated simulations, the difference is up to a factor 4. The tallest bar in Figure 11 is the result of one water molecule accepting a hydrogen bond from oxygen 15 and proton 16 during the complete QMH simulation. The special character of the hydrogen bond donated by oxygen 15 is exemplified by the fact that it is accepted by the only water molecule in the Glu solvation shell with a net positive charge (see Figure 3). The residence time could even be much longer than this, if we could simulate for longer times. Note however that the statistics here are poor: the 'average' residence time is calculated on the basis of only one hydrogen bond. The difference between the oxygen 15 residence times of system QM- and FF- is smaller but still large. In simulations of glutamate embedded in the photoactive yellow protein, we

saw too that the residence times of hydrogen bonds donated to Glu were shorter in force field simulations than in QMMM simulations.[2]

When force field water molecules move from one hydrogen bond acceptor to the other (for example from one carboxylate oxygen to the other or from glutamate to water), they move their positively charged hydrogen from one region of negative charge to the other. As the charge has no directionality, the proton will be somewhat stabilized by both acceptors when it is exactly in the middle of the transition. A quantum water in a quantum environment feels less electron density during the transition, as the orbital structure of the two acceptors only stabilizes the proton's positive charge when it is pointing in the right direction. This will make the barrier for hydrogen bond switching higher and the residence times longer.

Both the force field and the quantum simulations show the same trend on protonation: the residence time on oxygen 15 increases, while the hydrogen bonds donated to oxygen 14 have a shorter lifetime. Particularly in QMH we see that oxygen 15 is a very stable hydrogen bond donor when protonated. In FFH too, the hydrogen bonds donated by oxygen 15 last longer than the ones accepted by it. We saw this phenomena as well in our simulations of p-coumaric acid in water:[5] when the solute has an OH bond, which is more flexible than a C═O bond, it can more easily adapt to the water structure around it, and, hence, donated hydrogen bonds tend to be more stable than accepted ones on the oxygen in question. This flexibility effect compensates for the diminished stability due to fact that glutamic acid is now neutral and oxygens 14 and 15 are not as negative as in glutamate. For oxygen 14, only the latter destabilizing effect is present and hence it stabilizes its hydrogen bonds somewhat less.

When checking for finite size effects, we found that the life times of all the hydrogen bonds in simulation FF- are a factor 1.3 too long compared to the force field simulation in the large box. We do not expect that the size of the larger simulation box still has a substantial effect, because the box contains more than 6 solvation shells around Glu in each direction. Although we could not test the finite size effect for the quantum simulations, we expect that these effects are the same for the first principles simulations. In case of QMMM simulations of a full protein in water, this overestimation should be negligible, as the box sizes are many times larger than the ones in this work. So although we can still compare the life times of the hydrogen bonds presented here, they will most likely deviate from the 'real' values.

In studies of pure water, we did not find measurements of 'individual' hydrogen bond life times. We can look however at the calculated self-diffusion constants. Comparison of ab initio simulations of pure water with experiments and force field simulations[30,34–36,39,41–46] has shown that self-diffusion is difficult to calculate (separate simulations give results that differ up to 2 orders of magnitude) and often underestimated, especially when simulated ab initio. Some papers[34,35] state that this diminished mobility is due to an overcoordination of water (sharper peaks in the radial distribution function). We think it is not proven that this is the only or even the most important factor. The box sizes in these simulations were only 32−64 water molecules, sometimes not even comprising two full solvation shells in a cubic box. We have shown that the box size is an important factor, especially for dynamical properties.

4. Conclusions

When simulating the solvatation of glutamic acid and glutamate in water, it does matter which method one uses: based on a force field (Gromos96) or on DFT. With DFT-BLYP we see a larger variation in the strength and number of hydrogen bonds. This should be attributed to the fact that in the DFT method the electronic distribution in the molecules can adapt to the changing environment giving rise to changes in polarization, whereas in the force field method the partial charges on the atoms are fixed. This is not only of importance for the Glu atoms but also for the water molecules, particularly the ones in the first solvation shell.

Force field simulations tend to overcoordinate water molecules with about three extra hydrogen bonds per Glu molecule compared to quantum systems. Particularly the negatively charged carboxylate groups in the force field systems are very strong acceptors, accepting three hydrogen bonds per oxygen on average. In the quantum description the resonance structure of the carboxylate group has only five lone pairs, which are not always occupied by hydrogen bonds, as steric hindrance by the bulk water or glutamate itself does not allow it. This substantially decreases the average coordination number.

In the case of glutamic acid, the differences are most pronounced. The protonation decreases the coordination number around oxygen 15 drastically to 1 in system QMH, while it is almost 2 in FFH. The difference is similar on oxygen 14:1.5 vs 2.5. In the quantum simulation, we see a transfer of positive charge from proton 16 to the nearest water molecule. The hydrogens of this water become more positive and the oxygen more negative. This stronger polarization increases the lifetime of the bond significantly, up to values longer than the simulation time, that is at least 15.6 ps. In system FFH, the residence time of the average water molecule around oxygen 15 is only 4.4 ps. The accepting water molecule stays longer but only 8.6 ps on average. All the other residence times are much shorter in the force field than in the quantum simulations too.

The structural differences between the two methods are mainly that more waters hydrogen bond in force field simulations. The hydrogen bonds itself are not that much different; they are slightly longer in the force field description but within a range of 4%. Only one hydrogen bond differed more than that, almost 15%. The largest dissimilarity however is in the residence times. These times are always underestimated by the force field, sometimes up to a factor 4. Our results suggest that for a proper force field description of the aqueous solvation of glutamic acid and glutamate, the force field should allow for a varying and asymmetric charge distribution of the carboxylate group. This would accommodate an asymmetric coordination of the water molecules around the Glu oxygens as seen in our DFT-based simula-

tions. A stronger polarization (that is, a larger absolute value of the partial charges) will result in shorter and stronger bonds.

Fernández et al.[39] show a correlation between the self-diffusion coefficient of water and the number of defects. This effect is present in 'real' and simulated water, and it is not an unphysical result of the simulation method. The overcoordination in Glu in our force field simulations will induce more defects in the water structure than the ab initio simulation does, and this will make the water molecules more mobile. Although the diffusion is coupled to the lifetime of hydrogen bonds, it is not a one-to-one relation. Hydrogen bond switching is a more local phenomenon than self-diffusion, and estimates of hydrogen bond life times are likely to be better than diffusivity predictions. To properly estimate self-diffusion, a molecule should travel a significant distance, but in ab initio simulation times water molecules can barely move 1 Å. Summarizing, we should be well aware that we cannot draw strong conclusions on which simulation compares better to real Glu solvated in water: the force field or the ab initio simulation. Our errors are most likely less severe than in most of the pure water simulations that can be related to experiments, because of our larger box sizes and the focus on local processes. With ab initio simulations, one includes electron density spatiality, yielding a detailed picture of related properties such as coordination numbers and hydrogen bond switching events. But the time scale for changes in these properties is of the same order as the simulation times for quantum simulations, resulting in better sampling in the force field simulations. From the simulations presented in this paper we can understand, and to some extent quantify, the difference between force field and quantum simulations, in order to recognize unphysical effects around the border of QM and MM parts in combined simulations.

In relation to the simulation of the proton transfer in the photoactive yellow protein, it is interesting to know that the residence time changes upon protonation are much larger for the quantum system than for the force field system. The changes are also larger than the residence time differences upon protonation of the phenolic oxygen of p-coumaric acid in our previous paper.[5] This means that deprotonation of glutamic acid has a larger residence time decreasing effect than the opposite effect on the chromophore and, hence, that it is likely that the hydrogen bond between the two molecules in the protein breaks easier after the proton transfer. This indicates that a direct proton transfer from Glu to p-coumaric acid can be a starting point for the breaking of hydrogen bonds and ultimately unfolding of the protein. We also saw that upon deprotonation of Glu, glutamate attracts three extra hydrogen bonds to the now negatively charged carboxylate group−independent of the simulation method. In force field simulations of glutamate in the protein as well,[2] the number of hydrogen bonds increased from 1 to 4 in less than 10 ps. The need to stabilize the negative charge on Glu makes hydrogen bond rearrangements easier. Exactly these rearrangements are part of the process of unfolding the protein into its signaling state.

Acknowledgment. This work was sponsored by the Stichting Nationale Computerfaciliteiten (National Computing Facilities Foundation, NCF) for the use of the supercomputer facilities. We would like to thank the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organization for Scientific Research, NWO) for funding.

References

(1) Speranskiy, K.; Kurnikova, M. *J. Chem. Phys.* **2004**, *121*, 1516–1524.

(2) Leenders, E. J. M.; Guidoni, L.; Röthlisberger, U.; Vreede, J.; Bolhuis, P. G.; Meijer, E. J. *J. Phys. Chem. B* **2007**, *111*, 3765–3773.

(3) Hoff, W. D.; Düx, P.; Hård, K.; Devreese, B.; Nugteren-Roodzant, I. M.; Crielaard, W.; Boelens, R.; Kaptein, R.; Van Beeumen, J.; Hellingwerf, K. J. *Biochemistry* **1994**, *33*, 13959–13962.

(4) Baca, M.; Borgstahl, G. E. O.; Boissinot, M.; Burke, P. M.; Williams, D. R.; Slater, K. E.; Getzoff, E. D. *Biochemistry* **1994**, *34*, 14369–14377.

(5) Leenders, E. J. M.; VandeVondele, J.; Bolhuis, P. G.; Meijer, E. J. *J. Phys. Chem. B* **2007**, *111*, 13591–13599.

(6) Kamiya, M.; Saito, S.; Ohmine, I. *J. Phys. Chem. B* **2007**, *111*, 2948–2956.

(7) Sun, W.; Kinsel, G. R.; Marynick, D. S. *J. Phys. Chem. A* **1999**, *103*, 4113–4117.

(8) Prabhakar, R.; Blomberg, M. R. A.; Siegbahn, P. E. M. *Theor. Chem. Acc.* **2000**, *104*, 461–470.

(9) Jayaraman, V.; Keesey, R.; Madden, D. R. *Biochemistry* **2000**, *39*, 8693–8697.

(10) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(11) Villa, A.; Mark, A. E. *J. Comput. Chem.* **2002**, *23*, 548–553.

(12) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(13) 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.

(14) Smith, B. J. *J. Comput. Chem.* **1999**, *20*, 428–442.

(15) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; Van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.

(16) Van der Spoel, D.; Lindahl, E.; Hess, B.; Van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; Van Drunen, R.; Berendsen, H. J. C. Gromacs user manual, version 3.3; 2005. Gromacs Web site http://www.gromacs.org (accessed Dec 7, 2007).

(17) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(18) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(19) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.

(20) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(21) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(22) CP2K developers home page. http://cp2k.berlios.de (accessed Dec 7, 2007).

(23) Lippert, G.; Hutter, J.; Parrinello, M. *Mol. Phys.* **1997**, *92*, 477–487.

(24) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(25) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641–3662.

(26) Sprik, M.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1996**, *105*, 1142–1152.

(27) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(28) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(29) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(30) Izvekov, S.; Voth, G. *J. Chem. Phys.* **2002**, *116*, 10372–10376.

(31) Hermans, J.; Berendsen, H. J. C.; Van Gunsteren, W. F.; Postma, J. P. M. *Biopolymers* **1984**, *23*, 1513–1518.

(32) Soper, A. K.; Bruni, F.; Ricci, M. A. *J. Chem. Phys.* **1997**, *106*, 247–254.

(33) Sorenson, J. M.; Hura, G.; Glaeser, R. M.; Head-Gordon, T. *J. Chem. Phys.* **2000**, *113*, 9149–9161.

(34) Schwegler, E.; Grossman, J. C.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *121*, 5400–5409.

(35) Grossman, J. C; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G *J. Chem. Phys.* **2004**, *120*, 300–311.

(36) Kuo, I. W.; Mundy, C. J.; McGrath, M. J.; Siepmann, J. I. *J. Chem. Theory Comput.* **2006**, *2*, 1274–1281.

(37) Glattli, A.; Daura, X.; Van Gunsteren, W. F. *J. Chem. Phys.* **2002**, *116*, 9811–9828.

(38) Alagona, G. *J. Am. Chem. Soc.* **1986**, *108*, 185–191.

(39) Fernández-Serra, M. V.; Artacho, E. *J. Chem. Phys.* **2004**, *121*, 11136–11144.

(40) Luzar, A.; Chandler, D. *Phys. Rev. Lett.* **1996**, *76*, 928–931.

(41) Mills, R. *J. Phys. Chem.* **1973**, *77*, 685–688.

(42) Watanabe, K.; Klein, M. L. *Chem. Phys.* **1989**, *131*, 157167..

(43) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2001**, *114*, 363–366.

(44) Kuo, I. W.; Mundy, C. J.; McGrath, M. J.; Siepmann, J. I.; VandeVondele, J.; Sprik, M.; Hutter, J.; Chen, B.; Klein, M. L.; Mohamed, F.; Krack, M.; Parrinello, M. *J. Phys. Chem. B* **2004**, *108*, 12990–12998.

(45) Asthagiri, D.; Pratt, L. R.; Kress, J. D. *Phys. Rev. E* **2003**, *68*, 041505.

(46) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 3572–3580.

CT700344F

JCTC Journal of Chemical Theory and Computation

All-Electron Scalar Relativistic Basis Sets for Third-Row Transition Metal Atoms

Dimitrios A. Pantazis,[†] Xian-Yang Chen,[‡] Clark R. Landis,[‡] and Frank Neese*,[†]

Lehrstuhl für Theoretische Chemie, Institut für Physikalische und Theoretische Chemie, Universität Bonn, Wegelerstrasse 12, D-53115 Bonn, Germany, and Department of Chemistry, University of Wisconsin—Madison, 1101 University Avenue, Madison, Wisconsin 53706

Received February 15, 2008

Abstract: A family of segmented all-electron relativistically contracted (SARC) basis sets for the elements Hf−Hg is constructed for use in conjunction with the Douglas−Kroll−Hess (DKH) and zeroth-order regular approximation (ZORA) scalar relativistic Hamiltonians. The SARC basis sets are loosely contracted and thus offer computational advantages compared to generally contracted relativistic basis sets, while their sufficiently small size allows them to be used in place of effective core potentials (ECPs) for routine studies of molecules. Practical assessments of the SARC basis sets in DFT calculations of atomic (ionization energies) as well as molecular properties (geometries and bond dissociation energies for MH_n complexes) confirm that the basis sets yield accurate and reliable results, providing a balanced description of core and valence electron densities. CCSD(T) calculations on a series of gold diatomic compounds also demonstrate the applicability of the basis sets to correlated methods. The SARC basis sets will be of most utility in calculating molecular properties for which the core electrons cannot be neglected, such as studies of electron paramagnetic resonance, Mössbauer and X-ray absorption spectra, and topological analysis of electron densities.

Introduction

Third-row transition metals play important roles in various branches of chemistry. Foremost, they are used in a wide range of catalytic processes including olefin polymerization and numerous forms of hydrocarbon functionalization. In a biological context, tungsten is the only naturally occurring third-row transition element. It is found in a number of enzyme active sites, where it usually catalyzes redox reactions similar to those performed by the analogous molybdenum enzymes.[1,2] The biologically relevant redox states are W(IV), W(V), and W(VI) where only W(V) is amenable to electron paramagnetic resonance (EPR) studies owing to the $5d^1$ electron configuration. However, other third-row transition metals like osmium, gold, platinum, and rhenium are sometimes also used in a biological context, the long-established medicinal applications of platinum and gold compounds being prominent examples.[3]

In recent years, it has become commonplace to supplement experimental studies by quantum chemical calculations that are designed to assist in the interpretation of experimental data or to provide information where experiments cannot be performed with realistic effort. The emphasis of most calculations, commonly performed at some level of density functional theory (DFT),[4,5] is on the relative energies of reactants, products, transition states, isomers, or conformers. A potential problem in calculations on third-row transition elements is the relatively high computational effort required to treat all the inner shell electrons (1s−4f) together with the importance of relativistic effects that are certainly not negligible in the third transition row.[6] As long as only energetics is concerned it has therefore been found to be convenient to employ effective core potentials (ECPs).[7–17] Properly designed ECPs together with suitable valence basis sets fulfill the double purpose of decreasing the computational

* Corresponding author. E-mail: neese@thch.uni-bonn.de.
[†] Universität Bonn.
[‡] University of Wisconsin—Madison.

10.1021/ct800047t CCC: $40.75
Published on Web 05/08/2008

effort—since only valence electrons are explicitly considered—as well as incorporating scalar relativistic effects. The calculations then proceed as in the nonrelativistic case.[18]

In various assessments and practical applications, it has been established that ECPs provide reliable approximations to all-electron scalar relativistic calculations as far as geometries and relative energies are concerned.[19] By contrast, obvious limitations arise when properties of the inner shells are being probed, as is the case in EPR or X-ray absorption experiments. At the very least for such cases, there should be basis sets available that allow scalar relativistic calculations with realistic effort, being adapted to the popular formulations of scalar relativistic Hamiltonian operators, such as the zeroth-order regular approximation[20–23] (ZORA), the infinite-order regular approximation[24] (IORA), and the Douglas−Kroll−Hess[25–29] (DKH) approach. It is common practice in basis set design to use a single contracted basis function for each core orbital since these orbitals undergo very limited changes upon bond formation. Unfortunately, different scalar relativistic approximations lead to very different shapes of the core orbitals and hence require contractions that are quite different from each other and from the nonrelativistic case. Therefore, standard contracted basis sets generally lack the flexibility in the core region for scalar relativistic methods and although uncontracted basis sets provide greater flexibility, their large size makes routine molecular calculations too slow to be practical.

Scalar relativistic basis sets for the ZORA approximation have been developed by the Amsterdam group for Slater functions and are available in the Amsterdam Density Functional code.[30,31] Van Wüllen has performed scalar relativistic contractions for the ZORA approximation, but to the best of our knowledge, his basis sets are not publicly available.[32] Various workers have recently reported basis sets for transition metal atoms that are contracted to be consistent with the second- or third-order DKH (DKH2, DKH3) procedures.[33–40] Representative examples are the atomic natural orbital (ANO) basis sets by the Lund group,[33] the correlation consistent basis sets for the first transition row developed by Peterson and co-workers[34] and the comprehensive basis sets proposed by Nakajima and Hirao[35] and by Koga and co-workers.[36–38] These basis sets are generally contracted and hence, as long as the integral generator of a given quantum chemical program does not take advantage of the general contraction, lead to fairly expensive calculations. This is not so much an issue in correlated ab initio calculations for which these basis sets were designed. It nevertheless becomes a serious efficiency issue for DFT calculations where the generation of two-electron integrals over basis functions dominates the computational effort.

Hence, we feel that it is important to have basis sets available that are not generally contracted but that can be used in scalar relativistic calculations. In this paper we propose such segmented all-electron relativistically contracted (SARC) basis sets, which are constructed for treatments of third-row transition metal systems in conjunction with scalar relativistic Hamiltonians (DKH or ZORA), yet are sufficiently small to be used in place of effective core potentials (ECPs) for routine studies of molecules. Exponents of the Gaussian primitives are derived from relatively simple empirical rules and contraction coefficients were determined in atomic complete active space self-consistent field (CASSCF) calculations in both the ZORA and second-order DKH (DKH2) schemes. In order to arrive at consistent sets of all-electron scalar relativistic basis sets covering most of the periodic table, we have furthermore recontracted the basis sets developed by the Karlsruhe group[41–43] along the same lines. Thus, basis sets of split-valence (SV), triple-ζ-valence (TZV), and quadruple-ζ-valence (QZV) quality are available in ZORA and DKH recontractions and can be combined with the SARC basis sets for third-row transition metals. All the above basis sets are now part of the freely available ORCA program package.[44] Scalar relativistic ZORA and DKH2 calculations based on the SARC basis sets are only slightly more expensive than standard nonrelativistic calculations and are not grossly less efficient than calculations that employ ECPs as long as the number of heavy atoms in the system is not too large. The performance of the SARC basis sets is assessed for both atomic properties (ionization energies) and molecular properties (structures and bond dissociation energies for MH_n complexes with less than 12 valence electrons).

Relativistic Hamiltonians

In this section, the well-known[45,46] underlying relativistic formalisms are briefly described. The reasoning starts with the four-component Dirac−Coulomb many electron Hamiltonian that consists of the one-electron Dirac Hamiltonian supplemented by the standard nonrelativistic electron−electron interaction terms. Calculations that are directly based on this Hamiltonian are feasible—as implemented in the Dirac code[47]—but are still computationally expensive. Much effort has been devoted to develop theories that decouple the large and small components of the four-component Dirac spinors. The earliest of these, the Breit−Pauli expansion, is known to face the variational collapse problem[45,46] and has therefore fallen into disuse in the framework of variational calculations.

The zeroth-order regular approximation (ZORA) is an effective Hamiltonian approach that was first discussed by Heully and co-workers[48,49] and later elaborated by van Lenthe et al.[20–22] Various modifications have been discussed by van Wüllen[23] and Filatov.[50–54] The ZORA method amounts to an expansion of the Dirac equation in powers of $E/(2c^2 - V)$ where V is the "molecular potential" and E is the Dirac energy. To leading order, the energy dependence does not appear. In a Kohn−Sham framework, the ZORA Hamiltonian is of the form $\{(c\boldsymbol{\sigma}\mathbf{p})[1/(2c^2 - V)](c\boldsymbol{\sigma}\mathbf{p}) + V\}\psi_i = \varepsilon_i\psi_i$, with $\boldsymbol{\sigma}$ representing the vector of Pauli spin-matrices[45] and $\mathbf{p}$ being the momentum operator.

The Douglas−Kroll−Hess (DKH) method represents an alternative way for decoupling the Dirac Hamiltonian. It is based on a series of unitary transformations, the first of which is the free particle Fouldy−Wouthuysen transformation.[55] The series apparently converges very fast and is usually truncated at second-order, which yields good results. Reiher and Wolf[55–61] have extensively developed and discussed higher-order terms in the DKH series. Other discussions of third-[62] and sixth-order[63] terms have also appeared in the literature. The equations are fairly complex and will not be

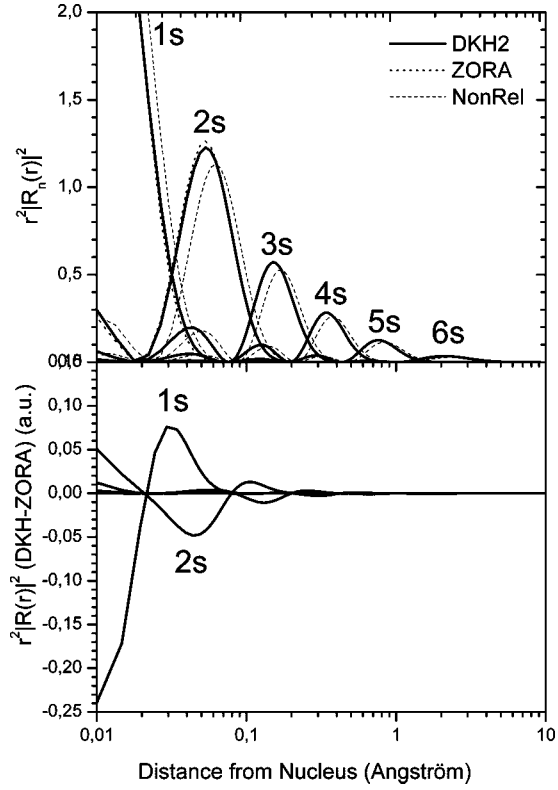


Figure 1. (top) Radial distribution functions of the neutral Hg atom (^{1}S) in the DKH2 (full line), ZORA (dotted line), and nonrelativistic (dashed line) approximations. (bottom) Difference between the DKH and ZORA radial distribution functions. All calculations were done at the RHF level with the decontracted WTBS basis set.

provided here. In programmable form, they have as examples been given in refs 56 and 59. More recently, high-order and even exact decoupling procedures have been discussed by various authors[57,60,64–68] and will likely find their way into major quantum chemistry codes in the future.

Once the reduction of the Dirac equation to two components has been achieved, it can be conveniently (but not uniquely)[69] divided into a spin-dependent and a spin-independent part using the well-known Dirac relation $(\boldsymbol{\sigma}\mathbf{u})(\boldsymbol{\sigma}\mathbf{v}) = \mathbf{uv} + i\boldsymbol{\sigma}(\mathbf{u} \times \mathbf{v})$ that holds for any two vectors $\mathbf{u}$, $\mathbf{v}$ that are independent of $\boldsymbol{\sigma}$. If the spin-dependent part (the spin−orbit coupling) is dropped, or taken into account perturbationally at a later stage, one arrives at a spin-free effective one-component method that typically only features a modified one-electron part of the Hamiltonian. This is the approach taken in this work. However, the effective one-electron Hamiltonians differ according to the chosen method of decoupling. Since the effective potentials provided by different decoupling procedures are quite different, it is necessary to use different contraction schemes for different methods. Hence, it is necessary to determine the all-electron contractions for each scalar-relativistic Hamiltonian separately.

The difference in radial functions is illustrated in Figure 1, where the radial functions of the s orbitals of the neutral Hg atom (^{1}S) are displayed. As is well-known, the relativistic and nonrelativistic orbitals are significantly different and all radial maxima of the nonrelativistic calculation appear at considerably too long distances. Compared to this major difference, the ZORA and DKH2 orbitals are relatively similar. The difference plot reveals that the only significant differences occur in the 1s and 2s orbitals. The ZORA 1s orbital is more compact than the DKH2 orbital and consequently shifted inward. The same effect is observed for the 2s orbital where the dominant radial maximum occurs at slightly shorter distances in the ZORA calculation, but in the region close to the nucleus, the DKH 2s orbital is larger. This is related to the orthogonality constraint of the 2s and 1s orbitals. The behavior of the DKH2 and ZORA orbitals is understood from the well-known fact that the deep core-orbital energies are much too low in the ZORA method.[20] Thus, the ZORA potential is too attractive close to the nucleus. This will ultimately have influence on deep core properties, but the valence region is described very similarly by both relativistic methods.

Table 1. Radial Expectation Values (in Bohr) Determined from CASSCF Calculations

	$\langle r_s \rangle$	$\langle r_p \rangle$	$\langle r_d \rangle$	$\langle r_f \rangle$
Hf	0.021044	0.075387	0.190489	0.638176
Ta	0.020753	0.074278	0.187201	0.607539
W	0.020470	0.073201	0.184024	0.580906
Re	0.020195	0.072155	0.180953	0.557585
Os	0.019927	0.071138	0.177983	0.536501
Ir	0.019666	0.070149	0.175109	0.517408
Pt	0.019411	0.069188	0.172325	0.500012
Au	0.019164	0.068253	0.169628	0.484067
Hg	0.018922	0.067342	0.167014	0.469201

Construction of Basis Sets

CASSCF calculations (or restricted Hartree−Fock (RHF) for d^{10} ^{1}S states) were first carried out for each atom in order to obtain the innermost radial expectation values to be used in the generation of the new primitives. These calculations employed Huzinaga's well-tempered basis sets (WTBS)[70,71] in completely uncontracted form (28s21p18d12f), resulting in a total of 265 basis functions per atom except for Hg, for which the basis set is (29s21p19d13f) with a total of 278 basis functions. Each atom was considered in its ground state, with the exceptions of rhenium, for which we averaged with equal weights the lowest sextet state arising from each of $5d^56s^2$ and $5d^66s^1$ configurations, and osmium, for which we used the lowest state of the $5d^76s^1$ configuration. Specifically, the states of the atoms were: Hf (^{3}F), Ta (^{4}F), W (^{5}D), Re (^{6}S, ^{6}D), Os (^{5}F), Ir (^{4}F), Pt (^{3}D), Au (^{2}S), and Hg (^{1}S).

On the basis of the innermost radial expectation values $\langle r_l \rangle$ determined from the above calculations (Table 1), the maximal exponents per angular momentum α_l (l = s, p, d, f), i.e. the exponents of the tightest functions were then determined according to the formula:

$$\alpha_l = k_l \frac{2f_l^2}{\pi \langle r_l \rangle^2} \tag{1}$$

where f_l assumes the values of 1, 4/3, 8/5, and 64/35, and κ_l is a scaling factor, whose optimal values were determined to be 1000, 100, 33, and 10 for s, p, d, and f functions, respectively. The results are summarized in Table 2. Having

Table 2. Maximum Exponents Per Angular Momentum (in Bohr^{-2}) Used in the SARC Basis Sets

	α_s	α_p	α_d	α_f
Hf	1437551.912320	19914.282978	1482.157051	52.266455
Ta	1478149.466003	20513.378492	1534.679527	57.670758
W	1519303.147809	21121.442157	1588.126510	63.080083
Re	1560962.277531	21738.257763	1642.488936	68.467082
Os	1603231.663005	22364.247712	1697.762680	73.954211
Ir	1646069.065505	22999.300491	1753.949530	79.512920
Pt	1689601.567586	23642.643991	1811.079246	85.141859
Au	1733435.947505	24294.843503	1869.127577	90.843326
Hg	1778058.504339	24956.609026	1928.094310	96.691016

determined the maximum exponents, series of descending primitives were then generated as $\alpha_i\chi^{-i}$ (i is a positive integer), χ being 2.25, 2.50, 2.75, and 3.00, for s, p, d, and f. Extrapolation factors were chosen empirically in an attempt to produce the smallest possible set of primitives that does not compromise appreciably the accuracy in the valence region. Each series was terminated when the exponent became smaller than 0.05 for s, p, and d functions or 0.5 for f functions; this procedure yielded (22s15p11d6f) sets with a total of 164 primitives. From these, the innermost 6s, 5p, 4d, and 5 or 4f primitives were contracted in the final basis sets, leaving the remaining primitives uncontracted. Contraction coefficients were obtained through scalar relativistic CASSCF calculations using the same states for each atom as described above, except for Re and Os that were now also considered in their respective ground states, ^{6}S and ^{5}D. Distinct sets of contraction coefficients were determined depending on the treatment of scalar relativistic effects: the basis sets therefore exist in two forms, one optimized for the DKH2 Hamiltonian and one optimized for the ZORA approximation.

When the SARC basis sets are used in conjunction with SV basis sets, the f primitives are contracted in a [51] pattern, whereas a [411] pattern is used when combined with TZV basis sets. Finally, for use with more extensively polarized TZVPP basis sets or for correlated methods, the SARC basis sets are supplemented with an additional single g function taken from a def2-TZVPP basis.[42] Thus, the final contraction patterns and total basis functions for the SARC basis sets in the SV/SVP, TZV/TZVP, and TZVPP forms are [17s11p8d2f] (104 functions), [17s11p8d3f] (111 functions), and [17s11p8d3f1g] (120 functions), respectively. Complete listings of the basis sets are included in the Supporting Information.

As mentioned in the Introduction, we also produced DKH and ZORA relativistically contracted variants of the Karlsruhe SV, TZV (H−Xe), and QZV (H−Kr) basis sets,[41–43] in order to ensure consistency when these are combined with the SARC basis sets for third-row transition metals. A much simpler procedure was followed in this case as the original exponents were not altered. However, new contraction coefficients were determined following CASSCF calculations with completely uncontracted basis sets. It is emphasized that only the innermost primitives per angular momentum were recontracted, hence the relativistic SV, TZV, and QZV variants are slightly larger than the original nonrelativistic basis sets. These basis sets are automatically loaded in place of the SV, TZV, or QZV basis sets when the use of a scalar relativistic Hamiltonian is detected.

Table 3. Estimated Incompleteness Errors (E_h) from Comparison of the SARC and Uncontracted WTBS Basis Sets[a]

	unc WTBS	SARC	ΔE
Hf	−15064.0138878	−15059.7917292	4.2221586
Ta	−15589.9018509	−15585.3767151	4.5251358
W	−16127.3776871	−16122.5137047	4.8639824
Re	−16676.6130872	−16671.3706773	5.2424099
Os	−17237.6115677	−17231.9529489	5.6586188
Ir	−17810.6802821	−17804.5642806	6.1160015
Pt	−18395.9799302	−18389.3624406	6.6174896
Au	−18993.6960534	−18986.5266065	7.1694469
Hg	−19603.8631097	−19596.1093860	7.7537237

[a] Calculations were done at the CASSCF level of theory with the DKH2 Hamiltonian.

Preliminary Considerations

The particular choice of the contraction pattern in the SARC basis sets was guided by the need to strike a balance between reduced basis set size and reliable results. An estimation of the contraction error is therefore necessary in order to establish whether the contraction pattern is uniformly good or leads to specific failures. The contraction error is obtained as the difference between total CASSCF energies for the same atomic states calculated with the contracted and fully uncontracted versions of the basis sets. Specifically, DKH2 calculations with the SARC [17s11p8d3f] basis sets indicate that the contraction error is of the order of 0.1 E_h and shows no irregular deviations: it increases monotonically from a minimum of 79 mE_h for hafnium to a maximum of 121 mE_h for mercury, while 70−80% of the energy difference between the contracted and uncontracted basis sets was found to originate in the p contraction. Essentially, identical trends and numerical values are obtained from the corresponding ZORA energies. These results confirm the uniform validity of the chosen contraction pattern across the third transition metal row.

To put the numbers mentioned above into perspective, we note that the contraction error is insignificant compared to the inherent incompleteness error relative to the basis set limit. This is shown clearly in Table 3, where total electronic energies obtained with SARC are compared to those obtained with the fully uncontracted WTBS, the latter being a good approximation to an essentially complete basis set. A rising trend in deviations is observed again as we move from the lighter to the heavier atoms, the incompleteness error ranging from a little over 4 E_h for hafnium to 7.75 E_h for mercury. Crucially, however, these relatively large absolute differences are not expected to have any adverse effects on the prediction of molecular properties other than total energies. This point is more meaningfully demonstrated by comparison of the valence shell orbital energies obtained with the two basis sets: regardless of the incompleteness error, the energies of the 5d and 6s orbitals are shifted by only a few hundredths of an electronvolt relative to the uncontracted WTBS energies (average 0.066 eV) and they never exceed 0.1 eV for any metal. Moreover, the lack of periodic trends in this case

Table 4. Ionization Energies (eV) Computed with the B3LYP Functional and the Fully Uncontracted WTBS Basis Set, with and without Relativistic Corrections, Compared to Experimental Values

		nonrelativistic		DKH2		ZORA	
	exp[a]	IE	ΔE	IE	ΔE	IE	ΔE
Hf	6.825	6.714	−0.111	6.721	−0.104	6.726	−0.099
Ta	7.550	6.633	−0.916	7.472	−0.077	7.489	−0.061
W	7.864	5.977	−1.887	7.791	−0.073	7.808	−0.056
Re	7.834	6.925	−0.909	7.916	0.083	7.936	0.102
Os	8.438	8.998	0.560	8.476	0.038	8.490	0.052
Ir	8.967	10.101	1.134	8.895	−0.072	8.901	−0.066
Pt	8.959	7.290	−1.669	9.163	0.204	9.193	0.234
Au	9.226	7.321	−1.904	9.347	0.121	9.380	0.155
Hg	10.438	8.419	−2.019	10.319	−0.118	10.354	−0.083
MUE			1.234		0.099		0.101
rms			1.387		0.109		0.115

[a] Reference 72.

confirms that the basis sets yield a valence description of consistent quality across the transition row. For the chemically less important 4f and 5s orbitals, the energy differences are slightly more pronounced, but still they only exceed 0.2 eV for the three heaviest elements. Overall, the mean absolute deviation for the 4f−6s orbital energies averaged over all metals amounts to 0.106 eV. Thus, despite the considerable size reduction in the SARC basis sets (111 instead of 265 functions), the description of the valence region in orbital energetic terms is essentially unaltered; the importance of this will be demonstrated in the following sections detailing practical assessments for both atomic and molecular systems.

Ionization Energies. Reliable prediction of atomic ionization energies is a typical requirement for computational methods both because it is an obvious measure of the balanced treatment of different electronic configurations and because it often reflects on the performance of the method for molecular environments. With the exception of hafnium, the first ionization energy originates in the removal of an electron from the 6s orbital of the neutral atom, as might be anticipated. The resulting electronic configurations and ground states of the cations are then as follows: Hf^+ (d^1s^2, ^2D), Ta^+ (d^3s^1, ^5F), W^+ (d^4s^1, ^6D), Re^+ (d^5s^1, ^7S), Os^+ (d^6s^1, ^6D), Ir^+ (d^7s^1, ^5F), Pt^+ (d^9s^0, ^2D), Au^+ ($d^{10}s^0$, ^1S), and Hg^+ ($d^{10}s^1$, ^2S). Since the SARC basis sets were developed with the intention to be used in routine molecular studies where DFT currently dominates or is the only cost-effective option, we chose to carry out the present assessment using the popular B3LYP functional.

Complete neglect of relativity leads in general to severe underestimation of ionization energies, with a root-mean-squared (rms) error of 1.4 eV and nonsystematic individual errors that can reach 2 eV (Table 4). Inclusion of scalar relativistic effects reduces these deviations dramatically, bringing the rms error down to 0.11 eV. Pt presents the greatest error, its ionization energy being overestimated by more than 0.2 eV. The ZORA and DKH results are essentially of the same quality. Most importantly, the transition from the uncontracted WTBS to the much more compact SARC (Table 5) has minimal impact on the accuracy of the computed values, the rms error increasing by a mere 0.01 eV. The graphical comparison of experimental and calculated ionization energies in Figure 2 shows vividly how closely the predicted values follow experiment, with Pt producing again the only significant deviation. This outcome demonstrates not only the quality of the SARC basis sets, but also the excellent performance of the B3LYP functional: the rms error of 0.12 eV for the third transition metal series matches that reported by Roos et al. for CASPT2 with a much more extended relativistic ANO basis set.[33]

Table 5. Ionization Energies (eV) Computed with the B3LYP Functional and the SARC Basis Sets, Compared to Experimental Values

		DKH2		ZORA	
	exp	IE	ΔE	IE	ΔE
Hf	6.825	6.711	−0.115	6.716	−0.109
Ta	7.550	7.466	−0.084	7.481	−0.069
W	7.864	7.791	−0.073	7.807	−0.057
Re	7.834	7.915	0.081	7.933	0.099
Os	8.438	8.483	0.045	8.503	0.065
Ir	8.967	8.896	−0.071	8.904	−0.063
Pt	8.959	9.197	0.238	9.223	0.264
Au	9.226	9.382	0.157	9.412	0.186
Hg	10.438	10.342	−0.096	10.373	−0.065
MUE			0.107		0.108
rms			0.120		0.128

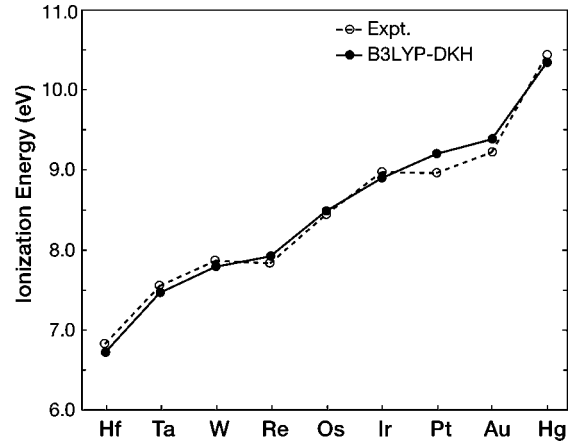


Figure 2. Comparison between experimental and computed (B3LYP-DKH2) first ionization energies.

Molecular Tests

The calculation of molecular properties such as geometries and bond dissociation enthalpies is an essential application of modern electronic structure theory. Third-row transition metal hydrides having the formula MH_n, for which the total valence electron count does not exceed 12,[73–76] constitute a computationally accessible set for testing the influence of relativistic effects on molecular properties. Such complexes have been examined recently by Landis and co-workers[77] using effective core potentials and their associated contracted valence basis sets, providing a useful comparison set for the basis sets described herein. These metal hydride complexes have electronic structures that are compatible with simple Lewis-like formulations,[73] and natural bond orbital (NBO) analysis[73,78–84] of their electronic structures using both SARC and uncontracted basis sets with different scalar relativistic treatments probes the robustness of simple Lewis-like

Table 6. Geometric Data and Total Energies of MH_n and MH_{n-1} Compounds at the B3LYP Level with Uncontracted (UC) and Contracted (C) SARC Basis Sets, Using the DKH and ZORA Hamiltonians

MH_n	point group	basis set	M−H bond (Å)	H−M−H angle (deg)	total energies (E_h)
HfH_4	T_d	UC-SARC-DKH	1.835	109.56	−15067.61713715
		C-SARC-DKH	1.841	109.56	−15067.52054279
		UC-SARC-ZORA	1.832	109.47	−15650.32180781
		C-SARC-ZORA	1.837	109.47	−15650.20652157
	C_{4v}	UC-SARC-DKH	1.814	78.42	−15067.55910933
		C-SARC-DKH	1.820	78.29	−15067.46219894
		UC-SARC-ZORA	1.813	78.58	−15650.26167290
		C-SARC-ZORA	1.819	78.42	−15650.14619041
HfH_3	C_{3v}	UC-SARC-DKH	1.848	119.13	−15066.98814663
		C-SARC-DKH	1.853	119.11	−15066.89160658
		UC-SARC-ZORA	1.845	119.15	−15649.69245223
		C-SARC-ZORA	1.850	119.13	−15649.57717761
TaH_5	C_{4v}	UC-SARC-DKH	1.744, 1.792	77.64/117.55	−15593.84558554
		C-SARC-DKH	1.751, 1.797	77.66/117.53	−15593.74525223
		UC-SARC-ZORA	1.746, 1.781	77.79/117.38	−16213.94501662
		C-SARC-ZORA	1.752, 1.788	77.69/117.50	−16213.82508781
	C_{2v}	UC-SARC-DKH	1.712, 1.746,1.782	63.82, 117.90	−15593.84302285
		C-SARC-DKH	1.718, 1.752, 1.787	63.75, 117.93	−15593.74260863
		UC-SARC-ZORA	1.716, 1.747,1.773	63.7, 117.61	−16213.94250329
		C-SARC-ZORA	1.722, 1.753,1.779	63.71, 117.70	−16213.82243700
TaH_4	T_d	UC-SARC-DKH	1.766	109.54	−15593.23874364
		C-SARC-DKH	1.772	109.54	−15593.13848434
		UC-SARC-ZORA	1.762	109.50	−16213.33985941
		C-SARC-ZORA	1.768	109.50	−16213.21979555
	C_{4v}	UC-SARC-DKH	1.748	78.13	−15593.21134922
		C-SARC-DKH	1.755	78.12	−15593.11108567
		UC-SARC-ZORA	1.750	78.33	−16213.31044183
		C-SARC-ZORA	1.756	78.31	−16213.19054552
WH_6	C_{3v}	UC-SARC-DKH	1.652, 1.708	63.24, 113.5	−16131.64727768
		C-SARC-DKH	1.657, 1.713	63.16, 113.6	−16131.54346908
		UC-SARC-ZORA	1.658, 1.705	63.38, 113.3	−16791.11504169
		C-SARC-ZORA	1.663, 1.709	63.34, 113.3	−16790.99057155
WH_5	C_{2v}	UC-SARC-DKH	1.665, 1.681,1.720	62.36, 116.44	−16131.03550734
		C-SARC-DKH	1.671, 1.687,1.724	62.31, 116.40	−16130.93176895
		UC-SARC-ZORA	1.670, 1.684,1.712	62.26, 116.29	−16790.50733939
		C-SARC-ZORA	1.676, 1.689,1.717	62.24, 116.28	−16790.38274965
	C_{4v}	UC-SARC-DKH	1.744, 1.677	76.54/118.91	−16131.03530403
		C-SARC-DKH	1.748, 1.683	76.50/118.95	−16130.93144660
		UC-SARC-ZORA	1.732, 1.681	76.61/118.72	−16790.50659133
		C-SARC-ZORA	1.736, 1.686	76.60/118.75	−16790.38195088
	C_{5v}	UC-SARC-DKH	1.672	64.86	−16131.01858750
		C-SARC-DKH	1.677	64.77	−16130.91477084
		UC-SARC-ZORA	1.673	64.82	−16790.48867413
		C-SARC-ZORA	1.678	64.81	−16790.36407530
ReH_5	C_{5v}	UC-SARC-DKH	1.620	62.99	−16679.90974069
		C-SARC-DKH	1.625	62.97	−16679.80333331
		UC-SARC-ZORA	1.623	63.17	−17380.81762186
		C-SARC-ZORA	1.627	63.19	−17380.68863645
	C_{4v}	UC-SARC-DKH	1.695, 1.616	76.12/119.33	−16679.90681889
		C-SARC-DKH	1.699, 1.621	76.09/119.43	−16679.79954812
		UC-SARC-ZORA	1.682, 1.621	76.23/119.21	−17380.81524275
		C-SARC-ZORA	1.686, 1.626	76.20/119.23	−17380.68542780
	C_{2v}	UC-SARC-DKH	1.640, 1.612,1.654	116.99, 56.95, 65.81	−16679.90890282
		>C-SARC-DKH	1.647, 1.616,1.658	116.53, 56.55, 65.93	−16679.80201210
		UC-SARC-ZORA	1.643, 1.617,1.650	116.47, 56.14, 66.32	−17380.81861597
		C-SARC-ZORA	1.648, 1.621,1.655	116.28, 55.72, 66.50	−17380.68910645
ReH_4	T_d	UC-SARC-DKH	1.644	109.13	−16679.30977871
		C-SARC-DKH	1.648	109.70	−16679.20326767
		UC-SARC-ZORA	1.642	109.47	−17380.22240021
		C-SARC-ZORA	1.647	109.47	−17380.09310021
	C_{4v}	UC-SARC-DKH	1.617	68.49/76.73	−16679.29524772
		C-SARC-DKH	1.622	68.38/76.94	−16679.18850066
		UC-SARC-ZORA	1.622	68.40/79.13	−17380.20339418
		C-SARC-ZORA	1.627	68.43/79.17	−17380.07408378
OsH_4	T_d	UC-SARC-DKH	1.592	109.47	−17240.05827094
		C-SARC-DKH	1.597	109.47	−17239.94858367
		UC-SARC-ZORA	1.591	109.47	−17984.56862903
		C-SARC-ZORA	1.595	109.47	−17984.43496057

Table 6. Continued

MH_n	point group	basis set	M−H bond (Å)	H−M−H angle (deg)	total energies (E_h)
	C_{4v}	UC-SARC-DKH	1.562	71.04	−17240.05287730
		C-SARC-DKH	1.567	71.09	−17239.94267012
		UC-SARC-ZORA	1.569	72.25	−17984.55665962
		C-SARC-ZORA	1.573	72.23	−17984.42277243
OsH_3	C_{3v}	UC-SARC-DKH	1.588	104.98	−17239.43763702
		C-SARC-DKH	1.593	105.15	−17239.32799126
		UC-SARC-ZORA	1.589	105.24	−17983.94753645
		C-SARC-ZORA	1.593	105.37	−17983.81389535
IrH_3	C_{3v}	UC-SARC-DKH	1.534	87.67	−17812.20363055
		C-SARC-DKH	1.538	88.04	−17812.09019536
		UC-SARC-ZORA	1.539	91.47	−18602.56588914
		C-SARC-ZORA	1.543	91.49	−18602.42677137
IrH_2		UC-SARC-DKH	1.542	88.11	−17811.57595938
		C-SARC-DKH	1.546	88.51	−17811.46295422
		UC-SARC-ZORA	1.547	92.81	−18601.93950386
		C-SARC-ZORA	1.550	92.92	−18601.80067284
PtH_2		UC-SARC-DKH	1.510	82.56	−18396.49795977
		C-SARC-DKH	1.514	82.75	−18396.38120805
		UC-SARC-ZORA	1.515	86.80	−19235.07546132
		C-SARC-ZORA	1.518	86.77	−19234.93083937
PtH		UC-SARC-DKH	1.527		−18395.86957063
		C-SARC-DKH	1.530		−18395.75344984
		UC-SARC-ZORA	1.530		−19234.44728022
		C-SARC-ZORA	1.533		−19234.30312472
AuH		UC-SARC-DKH	1.537		−18993.10537990
		C-SARC-DKH	1.541		−18992.98480346
		UC-SARC-ZORA	1.538		−19882.38086053
		C-SARC-ZORA	1.539		19882.23043055
Au		UC-SARC-DKH			−18992.49444792
		C-SARC-DKH			−18992.37453488
		UC-SARC-ZORA			−19881.76785321
		C-SARC-ZORA			−19881.61798680
H		UC-SARC-DKH			−0.49876441
		C-SARC-DKH			−0.49875809
		UC-SARC-ZORA			−0.49877893
		C-SARC-ZORA			−0.49877160

structures to more complete descriptions of the electron density distributions. The comprehensive comparison of the computed X−MH_n bond dissociation energies that was performed using B3LYP and CCSD(T) methods with relativistic effective core potentials[77] is also of great relevance for our present assessment. In the following, we assess the performance of the all-electron SARC basis sets in both DKH and ZORA scalar relativistic approaches comparing the results with this wide array of important molecular properties such as structure and bond dissociation enthalpies along with chemically meaningful interpretations such as charge distributions and orbital hybridizations.

MH_n Computational Details. All calculations were performed using either ORCA[44] or ADF 2005.[31,85–88] In ORCA calculations, a full geometry optimization for each structure was performed using the hybrid density functional B3LYP method.[89,90] Scalar relativistic corrections were included using the Douglas−Kroll−Hess (DKH2) and zeroth-order regular approximation (ZORA) methods with the option "onecenter true" enabled. The all-electron SARC basis sets were used in both contracted (17s11p8d2f) and uncontracted (22s15p11d6f) form for all transition metal elements. The SCF convergence threshold was set to VeryTight and the natural population anaysis (NPA)[80,81,83] was performed using the interface of ORCA to the GenNBO program version 5.0 on the geometries optimized at the B3LYP/uncontracted-SARC level. For ADF 2005 calculations, the density functional used was based on Vosko−Wilk−Nusair (VWN)[91] local-spin-density correlated potential with gradient corrections of the exchange correlation due to Perdew and Wang (PW91).[92] The frozen-core approximation was adopted for M shells ($1s^2 \rightarrow 4d^{10}$). The core electrons were calculated by the accurate relativistic Dirac−Slater method[93] and then transferred unchanged into the molecules. The valence orbitals of H and transition metal

Table 7. Natural Population Analysis (NPA) Charges on M (Q_M) and H (Q_H) Computed for the Most Stable MH_n and MH_{n-1} Species at the B3LYP Level with the SARC Basis Sets and the DKH2 Hamiltonian

compounds	point group	Q_M	Q_H
HfH_4	T_d	1.774	−0.444
HfH_3	C_{3v}	1.453	−0.484
TaH_5	C_{4v}	1.357	−0.348, −0.254
TaH_4	T_d	1.317	−0.329
WH_6	C_{3v}	0.481	−0.002, −0.158
WH_5	C_{2v}	0.748	−0.056, −0.214, −0.133
ReH_5	C_{5v}	−0.069	0.014
ReH_4	T_d	0.460	−0.115
OsH_4	T_d	0.088	−0.022
OsH_3	C_{3v}	0.125	−0.042
IrH_3	C_{3v}	−0.176	0.058
IrH_2	C_{2v}	−0.007	0.004
PtH_2	C_{2v}	−0.152	0.076
PtH		0.024	−0.024
AuH		0.047	−0.047

Table 8. Computed Bond Dissociation Energies (kcal/mol) at Various Levels of Theory for the Most Stable Neutral, Valence-Saturated MH_n Compounds at Geometries Optimized for Each Method

	B3LYP					CCSD(T)	PW91
	UC-SARC DKH	C-SARC DKH	UC-SARC ZORA	C-SARC ZORA	lacv3p++** [a]	lacv3p++** [a]	TZ2P-ZORA
HfH_4	81.72	81.69	81.94	81.93	81.59	73.90	79.11
TaH_5	67.82	67.78	66.75	66.84	70.06	66.6	66.6
WH_6	70.91	70.87	68.35	68.43	67.78	66.15	72.01
ReH_5	63.50	63.57	60.52	60.72	62.22	62.22	67.67
OsH_4	76.47	76.45	76.75	76.74	77.9	78.67	77.22
IrH_3	80.89	80.62	80.07	80.08	80.29	80.29	84.45
PtH_2	81.34	80.95	81.20	80.91	80.92	83.81	85.12
AuH	70.38	69.97	71.68	71.33	71.00	–	77.82

[a] Previously reported.[77]

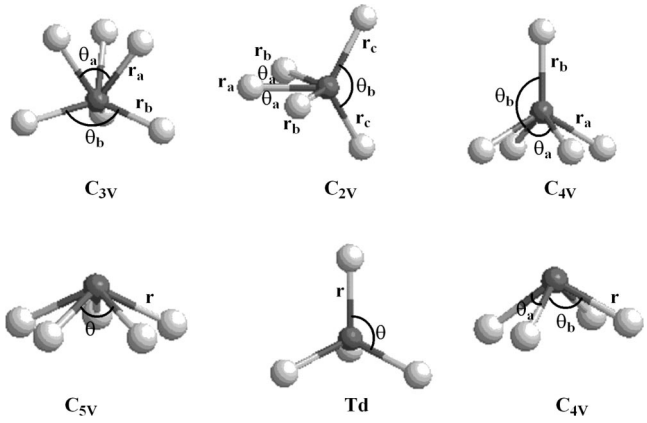


Figure 3. Illustrations of some energy minimized structures for MH_n complexes ($n = 4, 5, 6$).

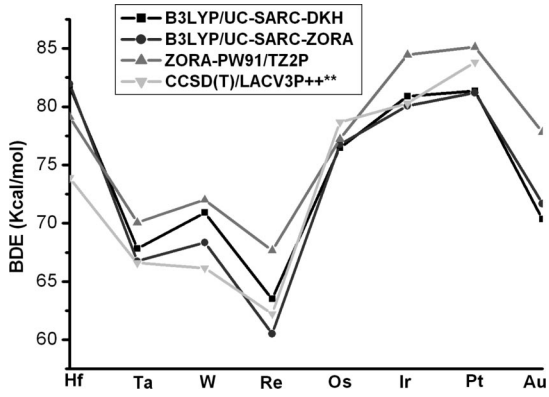


Figure 4. Single bond dissociation energies (BDE, kcal/mol) of $H-MH_{n-1}$ complexes along the third-row transitional metal at the B3LYP level with SARC-DKH and SARC-ZORA basis sets and ZORA-PW91/TZ2P and CCSD(T)/LACV3P++** methods.

atoms used triple-ζ Slater-type orbital (STO) basis sets with two p polarization functions (TZ2P).[94] Relativistic corrections employed the zeroth-order regular approximation (ZORA) method, and the integration level was set to 6.0.

Previously reported M−H bond energies are compared with the new results reported herein.[77] Prior post-Hartree−Fock (CCSD(T))[95] calculations were performed using Gaussian98 and the LACV3P++** relativistic effective core potential and the valence basis set of triple-ζ quality including a set of polarization and diffuse functions. Prior DFT calculations used the Jaguar program and the hybrid density functional B3LYP in association with the built-in LACV3P++** basis set. For consistency with previously reported bond dissociation energies,[77] the reported values are not corrected for zero-point energies or estimated basis set superposition errors.

Geometries of MH_n Complexes. As revealed by the geometric features shown in Table 6, the optimized geometries of MH_n molecules using the SARC all-electron basis sets and DKH or ZORA Hamiltonians conform to those computed previously[89,90] and expected based on a simple hybridization and Lewis-like considerations.[73,74] For example, the structure of WH_6 displays strong deviation from the octahedron predicted by valence shell electron pair repulsion (VSEPR) in favor of C_{3v} point group symmetry with 63° and 117° bond angles as predicted by idealized sd^5 hybridization (Figure 3 and Table 6). All-electron computations with scalar relativistic treatments exhibit M−H bond lengths and charge distributions that are similar to prior computations with relativistic effective core potentials. For example, tetrahedral MH_4 fragments exhibit the following M−H average bond lengths: M = Hf, $R_{Hf-H} = 1.835$ Å and M = Os, $R_{Os-H} = 1.593$ Å for all-electron scalar relativistic treatments compared with values of $R_{Hf-H} = 1.833$ Å and $R_{Os-H} = 1.601$ Å. Similarly, NPA based charges, $Q_{Hf} = +1.77$ and $Q_{Os} = +0.08$, for all-electron computations (Table 7), closely match those, $Q_{Hf} = +1.76$ and $Q_{Os} = +0.12$, determined with effective core potentials. It is interesting to note that different relativistic treatments, DKH and ZORA, have no significant effects on important molecular properties such as molecular geometries, relative energies of different stationary states, charge distributions, and bond dissociation energies. The contraction error of the SARC basis is estimated to be ~0.005 Å for M−H bond distances.

MH_n Bond Dissociation Energies. Simple M−H bond dissociation energies computed exhibit remarkable consistency over the range of computational methods and basis sets reported herein. Within the all-electron calculations using the B3LYP functional, the greatest range of computed bond dissociation energies, 3.05 kcal/mol, occurs for ReH_5. Among seven different calculations of M−H bond dissociation energies for eight metals, the rmsd is 1.81 kcal/mol.

It is interesting to note that relativistic core potential representations and all-electron treatments with scalar relativistic calculations using the B3LYP functional yield similar bond dissociation energies (see Table 8 and Figure 4). This is not surprising in that bond dissociation energies are primarily

Table 9. Equilibrium Geometries (Å), Dissociation Energies (eV), and Spectroscopic Constants (cm^{-1}) of Au Diatomic Molecules[a]

	R_e	D_e	ω_e	$\omega_e x_e$	ref
	AuH				
CCSD(T)/SARC-DKH	1.521	3.31	2307	49.9	
CCSD(T)/SARC-ZORA	1.520	3.33	2312	49.7	
CCSD(T)/AE	1.525	2.92	2288	42	96, 107
CCSD(T)/PP	1.510	3.28	2324		106
DFT-BDF	1.537	3.34	2259		97
DFT-ZORA(MP)	1.537	3.33	2258		97
expt	1.524	3.36	2305		109
	AuF				
CCSD(T)/SARC-DKH	1.932	2.63	551	3.6	
CCSD(T)/SARC-ZORA	1.932	2.63	551	3.6	
CCSD(T)/PP-aVTZ	1.942	2.89	541	2.8	98
CCSD(T)/PP-CBS	1.935	2.97	548	2.8	98
DFT-BDF	1.941	3.48	531		97
DFT-ZORA(MP)	1.948	3.39	526		97
expt	1.918	3.01	564	3.3	99
	AuCl				
CCSD(T)/SARC-DKH	2.226	3.10	371	1.4	
CCSD(T)/SARC-ZORA	2.226	3.08	371	1.4	
CCSD(T)/PP-aVTZ	2.230	2.80	367	1.3	98
CCSD(T)/PP-CBS	2.213	2.96	375	1.3	98
DFT-BDF	2.228	2.96	359		97
DFT-ZORA(MP)	2.247	2.91	352		97
DFT(BP)/PP	2.243	3.00	356		104
expt	2.199	3.13	384	1.5	100
	AuBr				
CCSD(T)/SARC-DKH	2.344	2.75	255	0.7	
CCSD(T)/SARC-ZORA	2.344	2.74	255	0.7	
CCSD(T)/PP-aVTZ	2.346	2.61	256	0.7	98
CCSD(T)/PP-CBS	2.334	2.75	260	0.6	98
DFT-BDF	2.351	2.59	250		97
DFT-ZORA(MP)	2.366	2.70	244		97
expt	2.318	2.96	264	0.7	100
	AuI				
CCSD(T)/SARC-DKH	2.474	3.10	220	0.5	
CCSD(T)/SARC-ZORA	2.474	3.09	220	0.5	
CCSD(T)/PP-aVTZ	2.500	2.48	209	0.5	98
CCSD(T)/PP-CBS	2.487	2.62	211	0.5	98
expt	2.471	2.86	216	0.5	101
	Au_2				
CCSD(T)/SARC-DKH	2.501	2.38	183	0.4	
CCSD(T)/SARC-ZORA	2.500	2.37	183	0.4	
CCSD(T)/PP	2.494	2.18	196		106
cp-CCSD(T)/AE	2.488	2.19	187		108
CCSD(T)/AE-DK3 point-charge	2.484	2.25	192		102
CCSD(T)/AE-DK3 finite-nucleus	2.480	2.31	194		35
DFT(B3LYP)/AE	2.541	2.03	169		104
DFT-BDF	2.513	2.22	183		97
DFT-ZORA(MP)	2.521	2.25	175		97
expt	2.472	2.31	191		109

[a] Spectroscopic constants were obtained in this work by a least-squares fit to a Morse potential.

determined by valence shell interactions. The new contracted basis sets for scalar relativistic calculations retain much of the computational efficiency of the effective core potential calculations but should also prove particularly useful in computing properties that are not accessible with effective potentials, such as hyperfine and quadrupole coupling parameters.

Coupled Cluster Calculations of Au Compounds. In the preceding sections, we established that the SARC basis sets perform excellently in DFT calculations of third-row transition metal species, in terms of both structures and energetics. We now address the question whether the applicability of the SARC basis sets is limited to DFT or whether they can be used in more demanding correlated ab initio approaches. Since coupled-cluster theory at the CCSD(T) level (coupled-cluster with single, double, and perturbative triple excitations) is currently regarded as one of the most reliable and efficient correlated methods in the field of small-molecule applications, we employ it for our study of six diatomic compounds of gold: AuH, AuF, AuCl, AuBr, AuI, and Au_2. CCSD(T) calculations using both DKH2 and ZORA scalar relativistic Hamiltonians are performed on this test set by correlating all valence electrons. Note that since the SARC basis sets were not explicitly designed for correlated calculations, it is advisable to exclude the core and outer-core (4f) electrons from correlation treatments. Otherwise some inaccuracies are introduced that we attribute to the effects of basis set superposition error.

The basis sets used are of TZVPP quality and for Au correspond to the [17s11p8d3f1g] contraction, whereas for the other elements to the appropriate relativistically recontracted Karlsruhe TZVPP basis sets (H (5s2p1d) → [3s2p1d], F (11s6p2d1f) → [6s3p2d1f], Cl (14s9p2d1f) → [8s4p2d1f], Br(17s13p8d1f)→[10s8p4d1f],I(19s15p9d2f)→[12s10p5d2f]). All these molecules have been the subject of numerous experimental and theoretical studies in the past.[35,96–108] The wealth of available data and the volume of existing publications are such that extensive comparisons are prohibitive for our present purposes; thus, we necessarily limit the comparison of the SARC data to only a representative selection of most recent results (Table 9) and direct the reader to the cited literature, especially to the review by Pyykkö,[96] for a comprehensive coverage of the subject.

A striking feature is the remarkable agreement between the DKH and ZORA approaches; this gives us confidence that the SARC basis sets are well-adapted to the individual formalisms and that there are no imbalances in the basis set construction. Looking more closely at the actual data and making comparisons both with previous CCSD(T) results and with experiment, it becomes evident that the present CCSD(T)/SARC results are consistently accurate for all molecules considered. They are clearly superior to the available DFT results, yielding in general bond lengths that are significantly shorter and thus closer to experiment. Crucially, all predicted equilibrium parameters also compare favorably with previous CCSD(T) calculations that employ all-electron basis sets or—more commonly—pseudopotentials, often surpassing them in accuracy. Of special interest for comparison purposes are the gold halides, which have been recently studied by Puzzarini and Peterson using CCSD(T) with systematic sequences of correlation consistent pseudopotential basis sets (the augmented VTZ and projected CBS results are included in Table 9).[98] It is not our intention to establish whether the present CCSD(T)/SARC results are inferior or superior, but from the differences observed between the two approaches, we can safely conclude that the SARC basis

sets perform reliably in the context of scalar relativistic CCSD(T) calculations.

Summary

The contracted basis sets reported herein are constructed for accurate but yet affordable all-electron treatments of third-row transition metal systems in conjunction with scalar relativistic Hamiltonians (DKH or ZORA). In fact, they are sufficiently small to be used in place of effective core potentials (ECPs) for routine studies of molecules. Various metrics, including atomic ionization potentials, molecular geometries, and bond dissociation energies, indicate that the basis sets provide a balanced description of core and valence electron densities. Overall, the level of agreement between ECP based, DKH, and ZORA results for a variety of third-row transition metal species is striking, as is the success of the B3LYP functional upon comparison with either experiment or high level CCSD(T) calculations. Moreover, the excellent results obtained with the SARC basis sets in a series of CCSD(T) calculations of gold compounds confirm that the applicability of the SARC basis sets can be extended to correlated methods as long as a valence-only correlation strategy is followed.

The SARC basis sets are only loosely contracted, and hence, they are computationally more efficient than the generally contracted relativistic basis sets reported by other workers.[33–40] The main field of application for the new basis sets is expected to be calculation of molecular properties for which the core electrons cannot be neglected; for example, in studies of electron paramagnetic resonance, Mössbauer and X-ray absorption spectra. In addition, the SARC basis sets are ideally suited for the derivation of electron densities that will be subsequently subjected to topological analysis: as Frenking has pointed out,[110] total electron densities derived from ECP calculations may lead to artifacts in the topological analysis and therefore scalar relativistic all-electron calculations are to be preferred. It is interesting to note in this context that NBO derived properties appear to be less critical in this respect and showed a remarkable consistency between ECP and scalar relativistic all-electron calculation schemes.

Acknowledgment. D.A.P. and F.N. gratefully acknowledge financial support from the DFG priority program 1137 "Molecular Magnetism".

Supporting Information Available: Full listings of the SARC basis sets. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kletzin, A.; Adams, M. W. W. *FEMS Microbiol. Rev.* **1996**, *18*, 5.

(2) Hagen, W. R.; Arendsen, A. F. *Struct. Bonding (Berlin)* **1998**, *90*, 161.

(3) Farrell, N. *Uses of Inorganic Chemistry in Medicine*; Royal Society of Chemistry: Cambridge, 1999; p 208.

(4) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH: Weinheim, 2002; p 528.

(5) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989; p 352.

(6) Pyykkö, P. *Chem. Rev.* **1988**, *88*, 563.

(7) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270.

(8) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.

(9) Cundari, T. R.; Stevens, W. J. *J. Chem. Phys.* **1993**, *98*, 5555.

(10) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612.

(11) Hurley, M. M.; Pacios, L. F.; Christiansen, P. A.; Ross, R. B.; Ermler, W. C. *J. Chem. Phys.* **1986**, *84*, 6840.

(12) Lajohn, L. A.; Christiansen, P. A.; Ross, R. B.; Atashroo, T.; Ermler, W. C. *J. Chem. Phys.* **1987**, *87*, 2812.

(13) Ross, R. B.; Powers, J. M.; Atashroo, T.; Ermler, W. C.; Lajohn, L. A.; Christiansen, P. A. *J. Chem. Phys.* **1990**, *93*, 6654.

(14) Andrae, D.; Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1991**, *78*, 247.

(15) Andrae, D.; Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123.

(16) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866.

(17) Stoll, H.; Metz, B.; Dolg, M. *J. Comput. Chem.* **2002**, *23*, 767.

(18) Dolg, M. Effective core potentials. In *Modern Methods and Algorithms of Quantum Chemistry*, 2 ed.; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, 2000; Vol. 3, pp 507−540.

(19) Frenking, G.; Antes, I.; Böhme, M.; Dapprich, S.; Ehlers, A. W.; Jonas, V.; Neuhaus, A.; Otto, M.; Stegmann, R.; Veldkamp, A.; Vyboishchikov, S. F. *Rev. Comp. Chem.* **1996**, *8*, 63.

(20) van Lenthe, E.; Snijders, J. G.; Baerends, E. J. *J. Chem. Phys.* **1993**, *99*, 4597.

(21) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1994**, *101*, 9783.

(22) van Lenthe, E.; Snijders, J. G.; Baerends, E. J. *J. Chem. Phys.* **1996**, *105*, 6505.

(23) van Wüllen, C. *J. Chem. Phys.* **1998**, *109*, 392.

(24) Dyall, K. G.; van Lenthe, E. *J. Chem. Phys.* **1999**, *111*, 1366.

(25) Douglas, M.; Kroll, N. M. *Ann. Phys.* **1974**, *82*, 89.

(26) Hess, B. A. *Phys. Rev. A* **1985**, *32*, 756.

(27) Hess, B. A. *Phys. Rev. A* **1986**, *33*, 3742.

(28) Jansen, G.; Hess, B. A. *Phys. Rev. A* **1989**, *39*, 6016.

(29) Wolf, A.; Reiher, M.; Hess, B. A. *J. Chem. Phys.* **2002**, *117*, 9215.

(30) Amsterdam Density Functional (ADF); SCM, Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands. http://www.scm.com (accessed Feb 14, 2008).

(31) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Guerra, C. F.; Van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931.

(32) van Wüllen, C. *J. Comput. Chem.* **1999**, *20*, 51.

(33) Roos, B. O.; Lindh, R.; Malmqvist, P. A.; Veryazov, V.; Widmark, P. O. *J. Phys. Chem. A* **2005**, *109*, 6575.

(34) Balabanov, N. B.; Peterson, K. A. *J. Chem. Phys.* **2005**, *123*, 064107.

(35) Nakajima, T.; Hirao, K. *J. Chem. Phys.* **2002**, *116*, 8270.

(36) Watanabe, Y.; Tatewaki, H.; Koga, T.; Matsuoka, O. *J. Comput. Chem.* **2006**, *27*, 48.

(37) Sekiya, M.; Noro, T.; Miyoshi, E.; Osanai, Y.; Koga, T. *J. Comput. Chem.* **2006**, *27*, 463.

(38) Noro, T.; Sekiya, M.; Osanai, Y.; Koga, T.; Matsuyama, H. *J. Comput. Chem.* **2007**, *28*, 2511.

(39) Faegri, K. *Theor. Chem. Acc.* **2001**, *105*, 252.

(40) Faegri, K. *Chem. Phys.* **2005**, *311*, 25.

(41) Ahlrichs, R.; May, K. *Phys. Chem. Chem. Phys.* **2000**, *2*, 943.

(42) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

(43) Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *119*, 12753.

(44) Neese, F. *ORCA—an ab initio, Density Functional and Semiempirical Program Package*, 2.6−35; Universität Bonn: Bonn, Germany, 2008.

(45) Strange, P. *Relativistic Quantum Mechanics*; Cambridge University Press: Cambridge, 1998; p 610.

(46) Hess, B. A., Marian, C. M., Jensen, P. , Bunker, P. R. , Eds. Relativistic effects in the calculation of electronic energies. In *Computational Molecular Spectroscopy*; John Wiley & sons: New York, 2000; pp 169−220.

(47) Jensen, H. J. A.; Saue, T.; Visscher, L. Dirac, a relativistic ab initio electronic structure program. http://dirac.chem.sdu.dk (accessed Feb 14, 2008).

(48) Heully, J. L.; Lindgren, I.; Lindroth, E.; Martenssonpendrill, A. M. *Phys. Rev. A* **1986**, *33*, 4426.

(49) Lindroth, E.; Heully, J. L.; Lindgren, I.; Martenssonpendrill, A. M. *J. Phys. B.* **1987**, *20*, 1679.

(50) Filatov, M. *Chem. Phys. Lett.* **2002**, *365*, 222.

(51) Filatov, M.; Cremer, D. *Mol. Phys.* **2003**, *101*, 2295.

(52) Filatov, M.; Cremer, D. *J. Chem. Phys.* **2003**, *118*, 6741.

(53) Filatov, M.; Cremer, D. *Chem. Phys. Lett.* **2003**, *370*, 647.

(54) Filatov, M.; Cremer, D. *J. Chem. Phys.* **2005**, *122*, 044104.

(55) Reiher, M. *Theor. Chem. Acc.* **2006**, *116*, 241.

(56) Wolf, A.; Reiher, M.; Hess, B. A. *J. Chem. Phys.* **2002**, *117*, 9215.

(57) Reiher, M.; Wolf, A. *J. Chem. Phys.* **2004**, *121*, 10945.

(58) Wolf, A.; Reiher, M.; Hess, B. A. *J. Chem. Phys.* **2004**, *120*, 8624.

(59) Neese, F.; Wolf, A.; Fleig, T.; Reiher, M.; Hess, B. A. *J. Chem. Phys.* **2005**, *122*, 204107.

(60) Wolf, A.; Reiher, M. *J. Chem. Phys.* **2006**, *124*, 064103.

(61) Reiher, M.; Wolf, A. *Phys. Lett. A* **2007**, *360*, 603.

(62) Hirata, S.; Yanai, T.; de Jong, W. A.; Nakajima, T.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 3297.

(63) van Wüllen, C. *Chem. Phys.* **2005**, *311*, 105.

(64) Kutzelnigg, W.; Liu, W. J. *J. Chem. Phys.* **2005**, *123*, 241102.

(65) Kutzelnigg, W.; Liu, W. J. *J. Chem. Phys.* **2006**, *125*, 107102.

(66) Kutzelnigg, W.; Liu, W. J. *Mol. Phys.* **2006**, *104*, 2225.

(67) Reiher, M.; Wolf, A. *J. Chem. Phys.* **2004**, *121*, 2037.

(68) Wolf, A.; Reiher, M. *J. Chem. Phys.* **2006**, *124*, 064102.

(69) Visscher, L.; van Lenthe, E. *Chem. Phys. Lett.* **1999**, *306*, 357.

(70) Huzinaga, S.; Kolbukowski, M. *Chem. Phys. Lett.* **1993**, *212*, 260.

(71) Huzinaga, S.; Miguel, B. *Chem. Phys. Lett.* **1990**, *175*, 289.

(72) Linstrom, P. J.; Mallard, W. G. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; National Institute of Standards and Technology: Gaithersburg, MD, 2005.

(73) Weinhold, F.; Landis, C. R. *Valency and Bonding: A Natural Orbital Donor:Acceptor Perspective*; Cambridge University Press: New York, 2005; p 760.

(74) Landis, C. R.; Weinhold, F. *J. Comput. Chem.* **2007**, *28*, 198.

(75) Landis, C. R.; Cleveland, T.; Firman, T. K. *J. Am. Chem. Soc.* **1995**, *117*, 1859.

(76) Landis, C. R.; Firman, T. K.; Root, D. M.; Cleveland, T. *J. Am. Chem. Soc.* **1998**, *120*, 1842.

(77) Uddin, J.; Morales, C. M.; Maynard, J. H.; Landis, C. R. *Organometallics* **2006**, *25*, 5566.

(78) Brunck, T. K.; Weinhold, F. *J. Am. Chem. Soc.* **1979**, *101*, 1700.

(79) Carpenter, J. E.; Weinhold, F. *J. Am. Chem. Soc.* **1988**, *110*, 368.

(80) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.

(81) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.

(82) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736.

(83) Glendening, E. D. ; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Weinhold, F.; Bohmann, J. A.; Morales, C. M. *NBO 5.0*; Theoretical Chemistry Institute, University of Wisconsin: Madison, WI, 2001; http://www.chem.wisc.edu/∼nbo5 (accessed Feb 14, 2008).

(84) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(85) *Amsterdam Density Functional (ADF)*; 2005.01, SCM, Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands, 2005. http://www.scm.com (accessed Feb 14, 2008).

(86) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41.

(87) te Velde, G.; Baerends, E. J. *J. Comput. Phys.* **1992**, *99*, 84.

(88) Ziegler, T.; Rauk, A.; Baerends, E. J. *Theor. Chim. Acta* **1997**, *43*, 261.

(89) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(90) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(91) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(92) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.

(93) Rosen, A.; Lindgren, I. *Phys. Rev.* **1968**, *176*, 114.

(94) Snijders, J. G.; Baerends, E. J.; Vernooijs, P. *Atomic Nucl. Data Tables* **1982**, *26*, 483.

(95) Rittby, M.; Bartlett, R. J. *J. Phys. Chem.* **1988**, *92*, 3033.

(96) Pyykkö, P. *Angew. Chem., Int. Ed.* **2004**, *43*, 4412.

(97) Liu, W.; van Wüllen, C. *J. Chem. Phys.* **1999**, *110*, 3730.

(98) Puzzarini, C.; Peterson, K. A. *Chem. Phys.* **2005**, *311*, 177.

(99) Okabayashi, T.; Nakaoka, Y.; Yamazaki, E.; Tanimoto, M. *Chem. Phys. Lett.* **2002**, *366*, 406.

(100) Evans, C. J.; Gerry, M. C. L. *J. Mol. Spectrosc.* **2000**, *203*, 105.

(101) Reynard, L. M.; Evans, C. J.; Gerry, M. C. L. *J. Mol. Spectrosc.* **2001**, *205*, 344.

(102) Tsuchiya, T.; Abe, M.; Nakajima, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 4463.

(103) O'Brien, L. C.; Elliott, A. L.; Dulick, M. *J. Mol. Spectrosc.* **1999**, *194*, 124.

(104) Han, Y.-K.; Hirao, K. *Chem. Phys. Lett.* **2000**, *324*, 453.

(105) Guichemerre, M.; Chambaud, G.; Stoll, H. *Chem. Phys.* **2002**, *280*, 71.

(106) Lee, H.-S.; Han, Y.-K.; Kim, M. C.; Bae, C.; Lee, Y. S. *Chem. Phys. Lett.* **1998**, *293*, 97.

(107) Kaldor, U.; Hess, B. A. *Chem. Phys. Lett.* **1994**, *230*, 1.

(108) Hess, B. A.; Kaldor, U. *J. Chem. Phys.* **2000**, *112*, 1809.

(109) Huber, H. P.; Herzberg, G. *Constants of diatomic molecules*; Van Nostrand: New York, 1979; p 565

(110) Vyboishchikov, S. F.; Sierraalta, A.; Frenking, G. *J. Comput. Chem.* **1997**, *18*, 416.

CT800047T

JCTC Journal of Chemical Theory and Computation

Reaction Coordinates and the Transition-Vector Approximation to the IRC

Willem-Jan van Zeist, Anton H. Koers, Lando P. Wolters, and F. Matthias Bickelhaupt*

Department of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands,

Received August 24, 2007

Abstract: The appearance of a reaction profile or potential energy surface (PES) associated with the reaction path (defined as the path of steepest descent from the saddle point) depends on the choice of reaction coordinate onto which the intrinsic reaction coordinate is projected. This provides one with the freedom, but also the problem, of choosing the optimal perspective (i.e., the optimal reaction coordinate) for revealing what is essential for understanding the reaction. Here, we address this issue by analyzing a number of different reaction coordinates for the same set of model reactions, namely, prototypical oxidative addition reactions of C−X bonds to palladium. We show how different choices affect the appearance of the PES, and we discuss which qualities make a particular reaction coordinate most suitable for comparing and analyzing the reactions. Furthermore, we show how the transition vector (i.e., the normal mode associated with a negative force constant that leads from the saddle point to the steepest descent paths) can serve as a useful and computationally much more efficient approximation (designated TV-IRC) for full IRC computations, in the decisive region around the transition state.

1. Introduction

The terms "reaction profile" or "potential energy surface" are habitually used when chemical reactions are discussed. The first step toward understanding a reaction mechanism is locating the stationary points on the system's potential energy surface (PES).[1] On the entire PES, three stationary points are required to characterize the core reactivity parameters of a typical reaction, namely, the reactants, transition state (TS), and products. These stationary points provide one with the reaction energy and the activation barrier.

For a deeper understanding of the reaction, it is desirable to find connecting pathways between these points on the PES. This enables one to examine how exactly the reactants transform into the TS and then into the products. A reaction path is also an important ingredient, for example, in the extended activation strain model in which trends in activation barriers are described and understood in terms of how the original reactants affect each other along such a path.[2,3]

The most common approach is finding the intrinsic reaction path (IRP) between a saddle point (transition state) and the minimum in an adjacent valley (reactants or products) on the PES. This path is defined as the path of steepest descent from the TS to the minima and is found, in mass-weighted Cartesian coordinates, through intrinsic reaction coordinate (IRC) calculations.[4–7] The IRP has been successfully used in many studies as the reaction path.[8] The reaction profile or PES of the reaction can now be obtained, for example, by plotting the system's energy as a function of the IRC. Now that IRC calculations for larger systems become feasible, analysis of the IRP becomes an ever increasingly important topic.[9]

Once the reaction path (i.e., path of steepest descent) has been computed, we can in principle plot the energy along this path. This yields the reaction profile or PES of the reaction. However, the appearance of a reaction profile or PES associated with the reaction path depends on the choice

* Corresponding author. fax: +31-20-5987629, e-mail: FM.Bickelhaupt@few.vu.nl.

10.1021/ct700214v CCC: $40.75
Published on Web 05/02/2008

of reaction coordinate onto which the IRC, which defines the reaction path, is projected. This provides one with the freedom, but also the problem, of choosing the optimal perspective, that is, the optimal reaction coordinate, for revealing what is essential for understanding the reaction. The reaction coordinate is a means for a chemist to refer to the extent of progress of a reaction using one parameter. This may be chosen on intuitive grounds, and often it is not even explicitly quantified but used in a qualitative fashion. In computational studies, still, the IRC itself features as the reaction coordinate for plotting the energy or other properties of a system along the IRP. Indeed, the IRC is a well-defined and quantitative choice of a reaction coordinate since it precisely defines all geometrical changes along the IRP. However, the IRC does not discriminate between those deformations that constitute the core geometrical transformation associated with the reaction and other changes in geometry that may be large but not critical. Furthermore, the IRC is multidimensional in terms of simple geometry parameters and therefore less transparent.

In the present paper, prompted by earlier investigations on reactivity,[2,3,10,11] we address the issue of how to choose the reaction coordinate such that it is optimal for monitoring and recognizing those events and phenomena that determine a reaction, in particular the barrier height. Note, once more, the difference between the reaction *path* (i.e., the uniquely defined IRP from an IRC calculation) and the reaction *coordinate*, which serves to optimally represent properties along the reaction path, such as the energy (i.e., the reaction *profile*). Here, we elaborate on the issue of how this reaction coordinate is to be chosen. In practice, understanding differences between reactions (e.g., trends in barriers) is even more important, and a clever choice of reaction coordinate is key to understanding the observed differences between these reactions. Here, we anticipate that indeed certain choices of reaction coordinates reveal features that are associated with critical moments along the IRP, while others hide them.

Thus, we have investigated and compared the IRPs of four series of organic and organometallic reactions: (i) the oxidative addition of the C−C bonds of ethane, propane, methylpropane, and dimethylpropane to palladium via direct oxidative insertion (OxIn); (ii) the oxidative addition of the methane C−H, chloromethane C−Cl, fluoromethane C−F, and silane Si−H bonds to palladium via the OxIn mechanism; (iii) the oxidative addition of the methane C−H, chloromethane C−Cl, and silane Si−H bonds via an alternative S_N2-type pathway; (iv) nucleophilic substitution of Cl^- with chloro*methane* and with chloro*ethane*. This is done through IRC calculations with the Amsterdam Density Functional (ADF) program, using density functional theory at BLYP/TZ2P.[12,13]

Furthermore, we have explored the validity of the so-called transition-vector approximation to the intrinsic reaction coordinate (TV-IRC). The TV-IRC approximation consists of using the transition vector (i.e., the one and only normal mode with a negative eigenvalue that characterizes the transition state) instead of the full IRC in the region "before and after" the transition state. The TV-IRC approximation leads to an enormous reduction in computational cost of analyses along the IRP, in particular for larger, more realistic model reactions. But, of course, its validity depends on how well and, in particular, over which range the transition vector is really a good approximation to the IRC. We discuss both successful applications as well as situations that lead to a breakdown of the TV-IRC approximation.

2. Computational Details

The IRPs were generated through IRC calculations, with the ADF program.[12,13] The BLYP[14] functional was used in combination with a TZ2P basis set. The latter is of triple-ζ quality and has been augmented with two sets of polarization functions: 2p and 3d on hydrogen; 3d and 4f on carbon, chlorine, oxygen, fluorine, and phosphorus; and 5p and 4f on palladium.[15] The core shells of carbon (1s), fluorine (1s), chlorine (up to 2p), silicon (up to 2p), and palladium (up to 3d) were treated by the frozen-core approximation.[15] Scalar relativistic effects were taken into account by the zeroth-order regular approximation.[16] This approach has been shown to give accurate results for the systems under consideration.[17] The reaction profiles, and analysis thereof, were generated using the PyFrag program.[18]

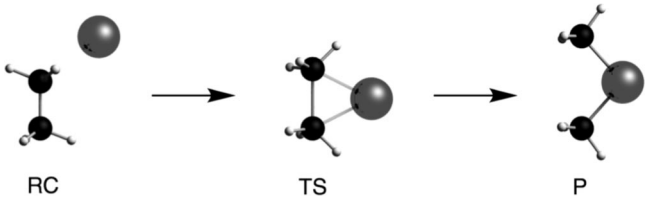
The IRC is constituted by the trajectory in mass-weighted coordinates, orthogonal to the equipotential contours of a PES, that connects two energy minima through a common TS from which it slopes downward along the steepest descent lines in $3N - 6$ configurational space (N = number of atoms).[5] The IRC method in ADF is based on the work by Fukui[5] and has been implemented in the ADF package by Deng and Ziegler[4] using mass-weighted coordinates and the constrained search method of Gonzalez and Schlegel.[6] In Cartesian coordinates, the mass-weighted coordinates are defined for each atom A as $X_A = m_A^{1/2} x_A$, $Y_A = m_A^{1/2} y_A$, and $Z_A = m_A^{1/2} z_A$ with x_A, y_A, and z_A the nonmass-weighted Cartesian coordinates.

The electron density distribution is analyzed using the Voronoi deformation density (VDD) method.[19] The VDD charge Q_A is computed as the (numerical) integral of the deformation density $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \sum_B \rho_B(\mathbf{r})$ associated with the formation of the molecule from its atoms in the volume of the Voronoi cell of atom A (eq 1). The Voronoi cell of atom A is defined as the compartment of space bounded by the bond midplanes on and perpendicular to all bond axes between nucleus A and its neighboring nuclei.

$$Q_A = \int_{\text{Voronoi-cell-A}} \left[\rho(r) - \sum\nolimits_B \rho_B(r)\right] \mathrm{d}r \qquad (1)$$

Here, $\rho(\mathbf{r})$ is the electron density of the molecule and $\sum_B \rho_B(\mathbf{r})$ the superposition of atomic densities ρ_B of a fictitious promolecule without chemical interactions that is associated with the situation in which all atoms are neutral. The interpretation of the VDD charge Q_A is rather straightforward and transparent. Instead of measuring the amount of charge associated with a particular atom A, Q_A directly monitors how much charge flows, due to chemical interactions, out of ($Q_A > 0$) or into ($Q_A < 0$) the Voronoi cell of atom A, that is, the region of space that is closer to nucleus A than to any other nucleus.

Scheme 1. Stationary Points in the Oxidative Insertion of Pd + Ethane



3. Choosing a Suitable Reaction Coordinate

To illustrate our thoughts, we will use four very similar varieties of an oxidative insertion reaction. Starting off with the oxidative insertion of palladium into the C−C bond of ethane, we consider the reactions where the hydrogens on one side of ethane are subsequently substituted by methyl groups. We thereby get a series of oxidative insertions of palladium into the C−C bonds of ethane, propane, methylpropane, and dimethylpropane. These reactions proceed from a reaction complex (RC) at ca. −6 kcal/mol relative to reactants, via a TS at around 20 up to some 30 kcal/mol depending on the alkane, toward the product (P) at ca. −10 kcal/mol (see Scheme 1). In terms of geometrical changes, this reaction is in the first place defined by the breaking of the C−C bond into which the metal inserts, as can be seen in Scheme 1.

In the following, we examine three perspectives on the reaction profiles that differ in the choice of the reaction coordinate onto which the IRC has been projected (see Figure 1): (a) the IRC itself and (b, c) two different projections of the IRC on simple geometry parameters. In all cases, the reaction profiles run from the reactant complex at the left to the product at the right.

In Figure 1a the reaction profiles are plotted as functions of the IRC itself as the reaction coordinate. Note that we display the progress of the reaction relative to the RC and not, as is often done, relative to the transition state, which technically is the starting point of an IRC calculation. The reason is that, for our purposes, namely, understanding the progress of a reaction and the height of the barrier in terms of the reactants, it is essential to take these reactants or the precursor complex as the point of reference. For completeness, we show all IRC graphs with the TS in the center in the Supporting Information. The IRC is the accumulative distance between the IRC points in mass-weighted Cartesian coordinates (see section 2). Note that the reaction profiles based on the IRC as the reaction coordinate vary rather chaotically from one reaction to the other. This behavior can be explained by the geometry changes along the reaction path. For example, in the case of propane, the TS peak is shifted toward the product side, that is, to a higher value of the reaction coordinate. The origin of this apparent "irregularity" is that one methyl group rotates freely over large parts of the reaction, thus creating a large coordinate distance in the IRC while influencing the total energy only very little. This can be seen in Scheme 2, which shows snapshots along the IRC of the reaction system starting from the RC until just before the TS is reached. This also illustrates the main problem: the IRC traces *all* movement of all of the nuclei in

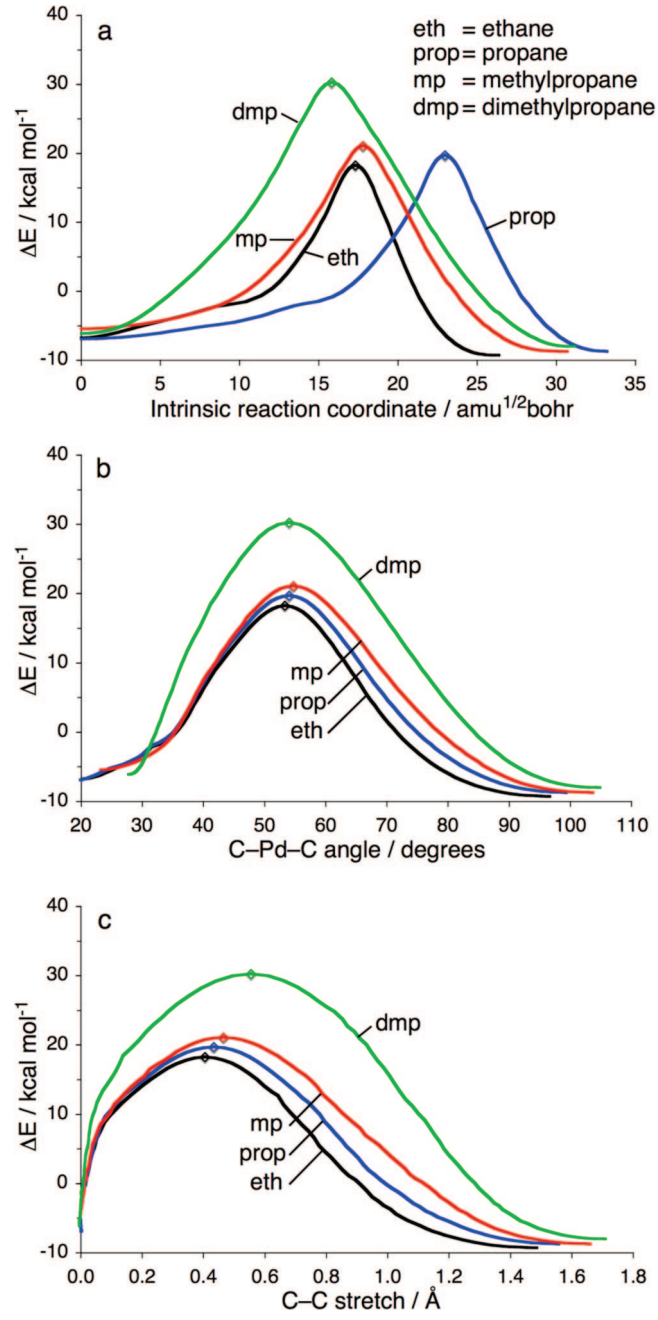
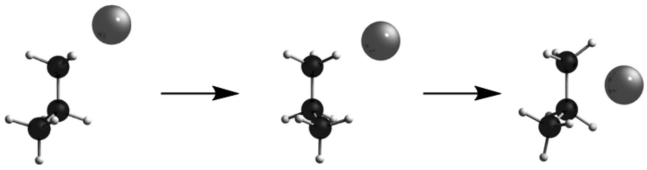


Figure 1. Comparison of three reaction coordinates for representing the PES of the oxidative insertion (OxIn) of Pd into the central C−C bond of ethane (black, eth), propane (blue, prop), methylpropane (red, mp), and dimethylpropane (green, dmp): (a) IRC distance (mass-weighted au) starting from the reactant complex, (b) bond angle C−Pd−C (in deg), (c) C−C bond stretch relative to substrate (in Å).

the reaction whether this is intrinsic to the actual process of bond breaking and insertion or just a derivative of the latter. Although there is some information to be gathered from these graphs, they are not really suitable to catch (the trends in) the essence of the various reactions. A more detailed discussion of this issue can be found in ref 20.

Next, in our quest for reaction coordinates that uncover the systematic trends in the oxidative insertion process, we examine projections of the IRC onto simple geometry parameters. We recall that the reaction coordinate should have significant amplitude in (and correlate with) the IRC-

Scheme 2. Snapshots along the IRC of the Oxidative Insertion of Pd + Propane, from the RC till just before the TS



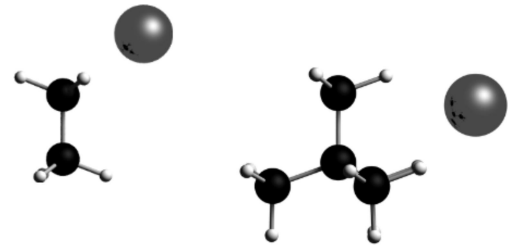
defined reaction path, such that it is a reliable measure of the progress of the reaction. Also, the reaction coordinate should provide us with insight into when and how the underlying features in the electronic structure (e.g., metal−substrate donation and backdonation orbital interactions) are active and decisive in determining the shape of the PES, in particular, the geometry of the TS and its energy. To this end, it is best to examine reaction profiles obtained with various choices of reaction coordinates which, in ADF 2007.01,[12] can be straightforwardly done with the ADFmovie tool. It can also be instructive to plot various geometrical parameters against each other. This can give nice insight into how the various geometrical features play a role, compared to each other, along the reaction path.

Two geometry parameters emerge as good candidates for a reaction coordinate: (i) the C−Pd−C angle, which increases as palladium approaches and the C−C bond expands, and (ii) (the stretch in) the C−C distance. This is of course not entirely unexpected in view of the fact that (i) the C−C bond breaking is an essential geometrical deformation which defines the oxidative insertion reaction and (ii) the C−Pd−C angle strongly correlates with the C−C distance on a large part of the reaction path.

Figure 1b shows the reaction profile as a function of the C−Pd−C angle as the reaction coordinate. It is immediately clear that the resulting reaction profiles (Figure 1b) behave much more systematically than in the previous representation (Figure 1a). From ethane to propane to methylpropane, there is a systematic and gradual change in the reaction profiles: they all start more or less at the same point but become higher in energy along this series. Furthermore, the TS shifts stepwise to the right, that is, to the product side. But the reaction profile for dimethylpropane is different, in particular at lower bond angles or, in other words, near the reactant complex. This is because palladium coordinates in a η^3 manner to the C−H bonds of three different methyl groups of dimethylpropane, as compared to the η^2 coordination with ethane (see Scheme 3), such that it is from the beginning closer to the C−C bond into which it inserts. Therefore, it has to reorient (and "travel") less in order to reach the TS, as can be seen in IRC movies (not shown) and also in the reaction profile based directly on the IRC as the reaction coordinate (see Figure 1a). The C−Pd−C bond angle is a reaction coordinate that magnifies differences in the reactant-complex region, that is, in early stages of the reaction. The height of the barrier is however determined beyond the RC, at a more advanced stage of the reaction.[2,3]

Figure 1c shows the reaction profile as a function of the C−C bond stretch (relative to the reactants) as the reaction coordinate. The reaction profiles vary very systematically

Scheme 3. η^2 and η^3 Coordination in Reactant Complexes of Pd + Ethane (left) and Pd + Dimethylpropane (right)



from one reaction to another. Now, they all start at the same point (in the reactants, the C−C bond is not yet stretched), and there is both a systematic increase in barrier height and a systematic shift to the right of the transition states. The systematic increase in energy and position along the C−C reaction coordinate is intimately connected with how the strain energy of the substrate varies and how the bonding capabilities of the substrate with the metal evolve. This has been previously pointed out in an activation strain analysis of C−H and C−C bond activation.[3] This explains the very systematic and smooth change in reaction profiles along the four reactions. A more detailed examination of the geometrical changes along the C−C reaction coordinate shows that this reaction coordinate magnifies the region along the IRC reaction path where the TS is located, that is, where the height of the barrier is determined, whereas the very early stage near the RC is more compressed. This is due to the fact that, during the first part of the reaction, the C−C bond does not change much, while the geometry of the reactant complex can change significantly due to migration of the metal from the optimal coordination site toward the point where it starts to insert into the C−C bond.

Next, we compare the oxidative insertion reactions of palladium into the C−H, C−C, C−F, and C−Cl bonds in methane, ethane, fluoromethane, and chloromethane, respectively.[3] The resulting PESs are again displayed as a function of the IRC as well as projections thereof onto the C−Pd−X angle and onto the C−X bond stretch relative to the reactant complex (see Figure 2). In addition, to the energy of the reaction system (i.e., the PES), we plot here also the VDD atomic charge of palladium and the population of the substrate's C−X antibonding σ^*_{C-X} acceptor orbital that becomes occupied in the course of the reaction.

The two main features in the metal−substrate bonding mechanism are, as has been briefly mentioned above, (i) the *backdonation* of charge from the palladium 4d orbital into the σ^*_{C-X} antibonding orbital of the C−X bond that is being broken and (ii) the donation of the σ_{C-X} orbital into the 5s orbital of palladium. Charge transfer and orbital populations are consequently quantities that are associated with the extent of progress of the oxidative-insertion reaction on the level of the electronic structure. The increasing population of the antibonding orbital shows in essence the bond-breaking process. All of these quantities are in the first place dependent on the (stretch in) C−X distance in the substrate. This is nicely illustrated in Figure 2, which shows plots of palladium VDD atomic charges[19] and σ^*_{C-X} antibonding orbital populations for each of the reaction coordinates. It is easily

seen that the C−X bond stretch allows us to represent the bond breaking process in the most smooth and consistent manner (see Figure 2g−i). Choosing the IRC on the x axis again leads to more chaotic variation along the different C−X bonds because the progress of the reaction, that is, the way in which the energy, atomic charge, and orbital population change, is directly linked to the C−X stretch and not to all of the other geometry parameters that contribute to the IRC (see Figure 2a−c). The C−Pd−X bond angle yields again a more smooth description (see Figure 2d−f), but the systematics of the bonding mechanisms underlying the reaction in progress come out best in the plots that are based on the C−X stretch.

Furthermore, the C−X bond stretch is the main (but not the only!) determinant for the strain energy $\Delta E_{\text{Strain}}(\zeta)$ of the, along the reaction path, increasingly deformed substrate, which together with the metal−substrate interaction $\Delta E_{\text{int}}(\zeta)$ determines the reaction profile or potential energy surface $\Delta E(\zeta) = \Delta E_{\text{Strain}}(\zeta) + \Delta E_{\text{int}}(\zeta)$ along the reaction coordinate ζ (cf. activation strain model, refs 2 and 3). Thus, the C−X stretch emerges as the optimal choice for representing the reaction profile and underlying features in the electronic structure for oxidative insertion reactions.

In the next example, we compare the oxidative addition of the methane C−H and chloromethane C−Cl bonds, each via two competing reaction mechanisms: direct OxIn and an alternative S_N2 mechanism (see Figure 3).[3,11] The S_N2 mechanism of Pd + CH_4 actually leads to the dehydrogenation of methane and yields molecular hydrogen plus a palladium−carbene complex (see Scheme 4). In the case of Pd + CH_3Cl, the actual substitution process is followed directly (i.e., without the occurrence of an intermediate S_N2 product complex) by a rearrangement of the expelled chloride leaving group toward the palladium atom to which it eventually coordinates (see also the snapshots in Figure 5d). The TS of the overall process occurs in the second stage, that is, during the migration of Cl^- toward Pd. Here, we see that both the IRC and the C−Pd−X bond angle yield a relatively disorderly picture in which no clear trends become apparent (see Figure 3). The C−X bond stretch, on the other hand, results in a more systematic appearance of the reaction profiles as well as the atomic charge and orbital population analyses (see Figure 3). The latter two show again that this reaction coordinate captures the essence of the reactions, namely, activation and rupture of the C−X bond through its reduction (i.e., through population of the σ^*_{C-X} orbital of the substrate). Note that this holds in particular also for the S_N2 pathway. An interesting phenomenon that occurs in the case of the S_N2 processes at CH_3Cl is the curling back of the curves near the end, that is, to the right in the representation as a function of the C−X stretch (see Figure 3g−i). This is a reflection of the fact that after the TS has been passed the S_N2-rearrangement pathway merges into the OxIn pathway. This is accompanied by a reduction of the

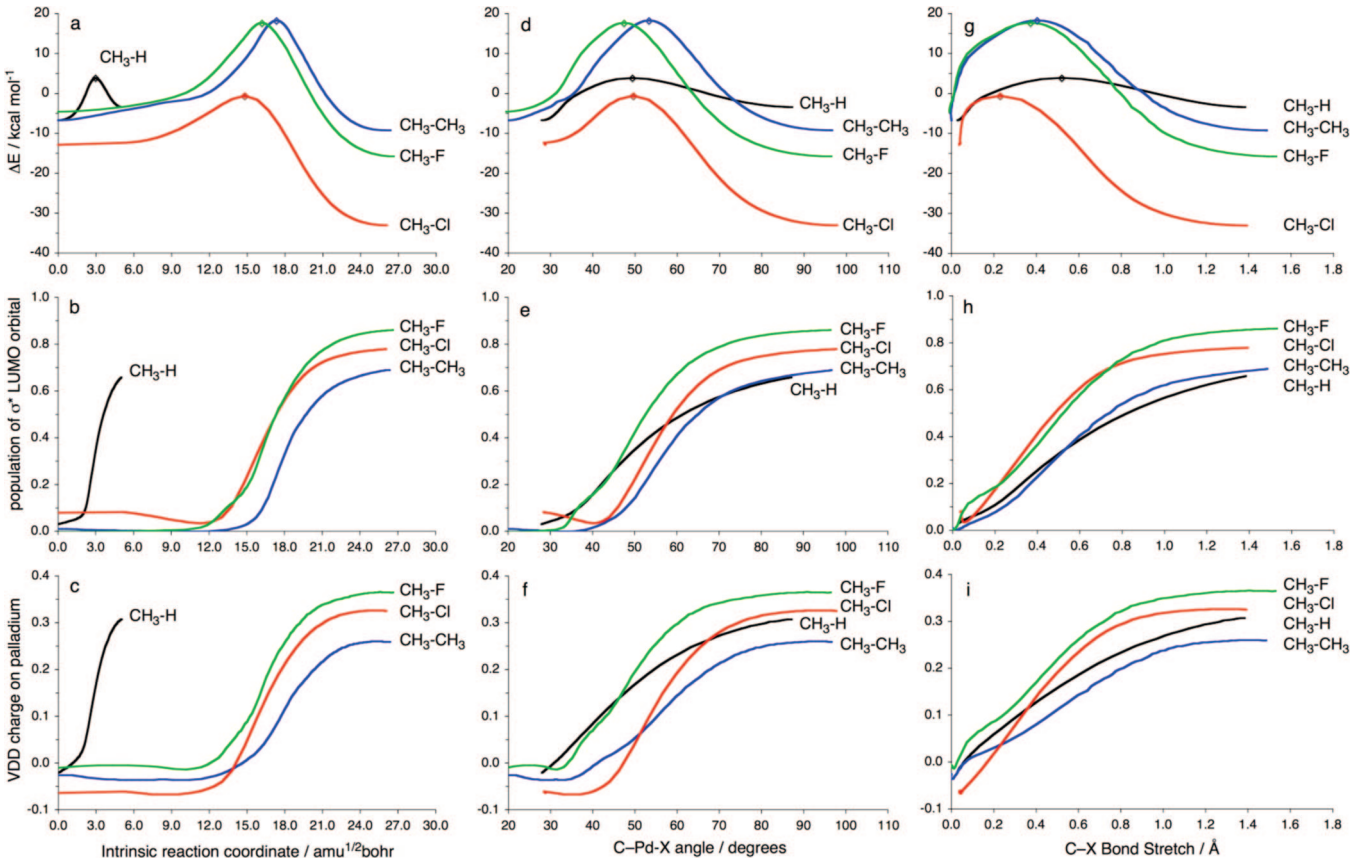


Figure 2. Comparison of three reaction coordinates for representing the PES and other properties of the oxidative insertion (OxIn) of Pd into the methane C−H (black), ethane C−C (blue), fluoromethane C−F (green), and chloromethane C−Cl bonds (red): (a, b, c) IRC distance (mass-weighted au) starting from the reactant complex, (d, e, f) bond angle C−Pd−X (in deg), and (g, h, i) C−X bond stretch relative to substrate (in Å). The plots show (a, d, g) the PES, (b, e, h) the population of the C−X σ^* antibonding LUMO, and (c, f, i) the VDD charge of the Pd atom.

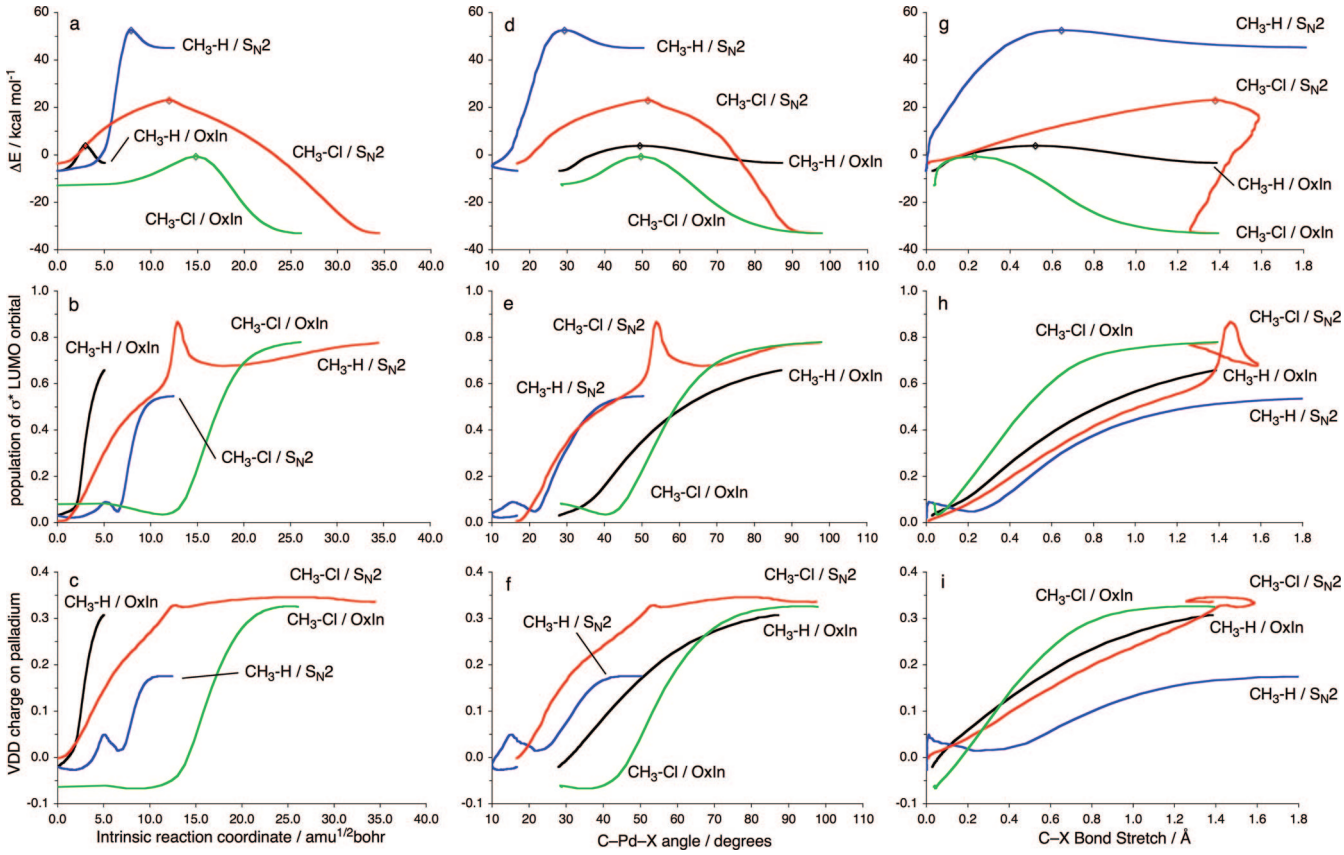
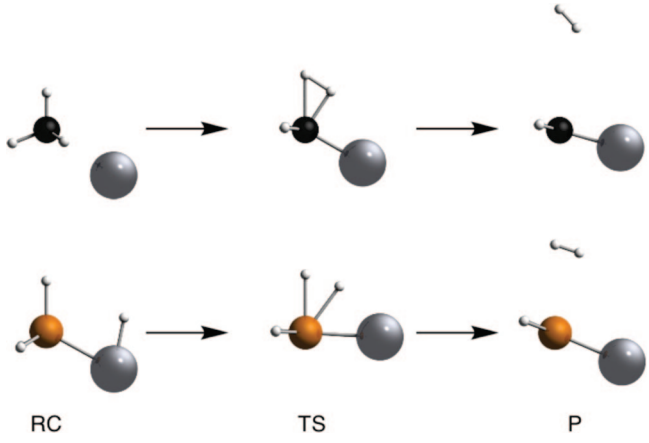


Figure 3. Comparison of three reaction coordinates for representing the PES and other properties of the oxidative insertion (OxIn) of Pd into methane C−H (black) and chloromethane C−Cl (green) as well as oxidative addition via S_N2 substitution of Pd with methane C−H (blue) and chloromethane C−Cl bonds (red): (a, b, c) IRC distance (mass-weighted au) starting from the reactant complex, (d, e, f) bond angle C−Pd−X (in deg), and (g, h, i) C−X bond stretch relative to the substrate (in Å). The plots show (a, d, g) the PES, (b, e, h) the population of the C−X σ^* antibonding LUMO, and (c, f, i) the VDD charge of the Pd atom.

Scheme 4. Stationary Points in the Pd-Induced Dehydrogenation of CH_4 (upper) and SiH_4 (lower)



C−Cl distance during the migration of the Cl^- leaving group (i.e., the curling back). Note that, after the migration stage is finished, the curve of the S_N2 pathway is superimposed on that of the OxIn pathway. The sudden increase of the σ^*_{C-X} orbital population takes places around the transition state point, where the C−Cl bond is also formally, that is, electronically, broken in the sense that we have a configuration switch toward a reduced C−Cl bond. For a detailed analysis of these curves, see ref 3.

Finally, we compare the activation of the methane C−H and silane Si−H bonds, again for each bond via the two competing OxIn and S_N2 pathways.[8] As noted in the previous example, the S_N2 reaction of Pd + CH_4 leads to the dehydrogenation of methane (see Scheme 4). The oxidative insertion proceeds without a barrier for Pd + SiH_4. Therefore, in this case, the IRC computations were started from an artificial reactant complex in which Pd coordinates end-on to a Si−H bond yielding a C_{3v} symmetric species Pd···H−SiH_3. This species is a second-order saddle point with two degenerate imaginary frequencies, both corresponding to a mode that leads to the OxIn product. Thus, the overall S_N2 reaction proceeds from the reactants, via the product of OxIn that plays the role of a precursor complex, toward an S_N2-type TS that, just as in the case of methane, eventually leads to dehydrogenation (see Scheme 4). Figure 4 shows the reaction profiles corresponding to the four pathways plotted as a function of the IRC, A−Pd−H angle, and A−H stretch. Similar to previous series of reactions, trends come out in the most systematic manner if PESs are plotted as a function of the A−H stretch. Activation strain analyses have shown that the higher energies of the reaction profiles involving methane are, among others, due to the fact that dissociating the C−H bond is inherently connected with a higher strain because this bond is stronger than the Si−H bond.[8]

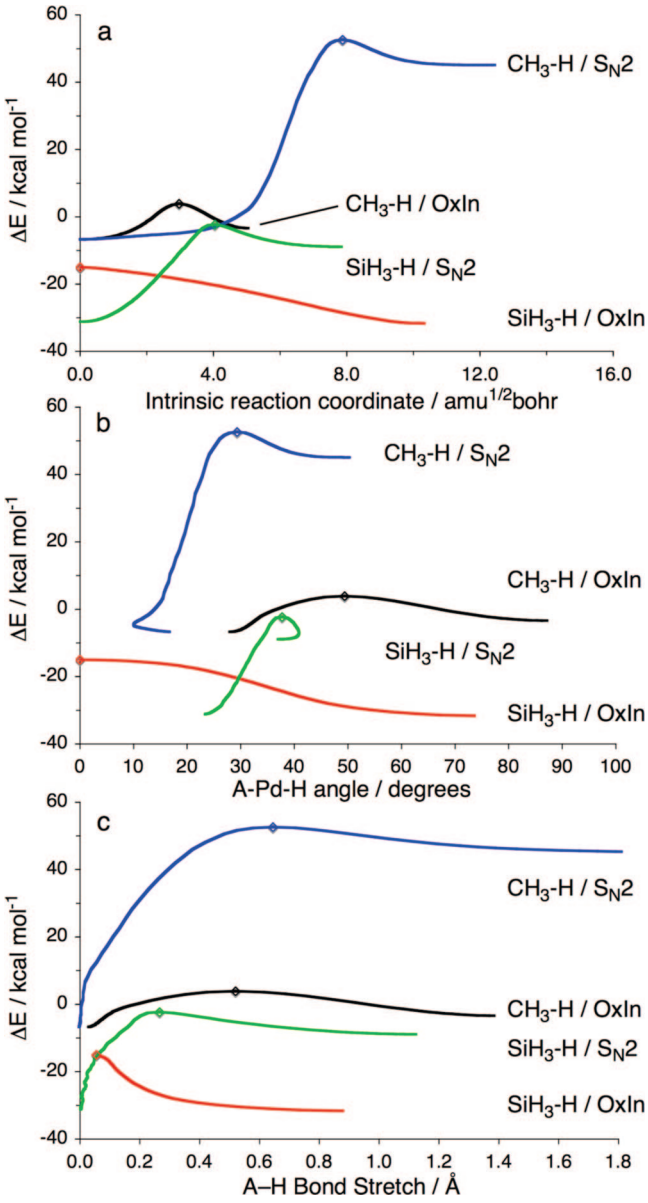


Figure 4. Comparison of three reaction coordinates for representing the PES of the oxidative insertion (OxIn) of Pd into the methane C−H (black) and silane Si−H (red) as well as the S_N2-type reactions of Pd with the methane C−H (blue) and silane Si−H bonds (green): (a) IRC distance (mass-weighted au) starting from the reactant complex, (b) bond angle C−Pd−X (in deg), and (c) C−X bond stretch relative to the substrate (in Å). See Scheme 4 for snapshots of stationary points.

4. Transition Vector as Approximation to IRC

Vibrational analysis reveals that C−C stretching also has a large amplitude in the transition vector of our oxidative insertion reactions, that is, the normal mode associated with a negative force constant that leads from the saddle point to the steepest descent paths. Thus, for very large model reaction systems for which IRC calculations become prohibitively expensive (or just impossible), following the path defined by the TV may be used as an approximation of that IRC in the region "around the TS". In fact, we find that the TV path approximates the IRC path very well over a sufficiently long interval to be useful for representing and analyzing the reaction profile of our oxidative insertion reactions. Numerical experiments shows that the reaction profile of oxidative insertion reactions (as a function of the C−C stretch and the C−Pd−C angle as the reaction coordinates) generated on the basis of the TV essentially coincides with that generated on the basis of the full IRC over an interval of about 0.5 Å of the C−C reaction coordinate around the TS (see Figure 5a and b). We designate this procedure as the TV-IRC approximation. Since TV-IRC calculations typically require around 10 single-point calculations only, the computational cost is dramatically decreased as compared to the corresponding partial IRC (let alone a full IRC), which would then require several constrained geometry optimizations in a stage of the reaction that goes with relatively strong structural reorganization on a shallow saddle region of the PES.

Interestingly, the TV-IRC reaction profile agrees over a longer interval with the IRC reaction profile if it is represented as a function of the C−C stretch (agreement over roughly one-third of the entire reaction interval!) than if it is represented as a function of the C−Pd−C angle (agreement over roughly one-fifth of the entire reaction interval, compare part a of Figure 5 with part b). This is consistent with the fact that the C−C stretch plays a more important role near the TS and that the reaction profile as a function of this reaction coordinate zooms in on the region around the TS.

We have also tested the TV-IRC approximation for other classes of reactions. In Figure 5c, we show the reaction profiles for the S_N2 reactions of $Cl^- + CH_3Cl$ and $Cl^- + CH_3CH_2Cl$ as functions of the C−Cl (i.e., carbon-leaving group) stretch. Again, the TV-IRC-based reaction profiles coincide with the IRC-based ones over a range of some 0.5 Å of the reaction coordinate, that is, roughly one-third of the entire reaction interval.

Finally, we also wish to stress the situations in which the TV-IRC approximation breaks down. This happens whenever the character of the geometrical deformations changes quickly or drastically along the reaction path. Typically, this happens when different elementary steps merge into one reaction step. An example is the alternative S_N2 mechanism for the oxidative addition of $Pd + CH_3Cl$. For the oxidative insertion mechanism, as can be seen in Figure 5d, the TV-IRC-based reaction profile again nicely coincides with the IRC-based one. At variance, for the alternative S_N2 mechanism, the TV-IRC-based reaction profile is valid only in very narrow interval around the TS. The reason is that the TV is mainly the migratory movement of the expelled chloride leaving group that is hydrogen-bonding to one of the C−H bonds.[3] But just before this TS, there is a real S_N2 stage in which the IRC has a large component of the characteristic Pd−C−Cl asymmetric stretch in combination with the methyl umbrella mode, while directly after the TS, the S_N2 reaction path merges into the regular oxidative insertion path.

5. Conclusions

The appearance of a reaction profile associated with the reaction path (defined as the path of steepest descent from

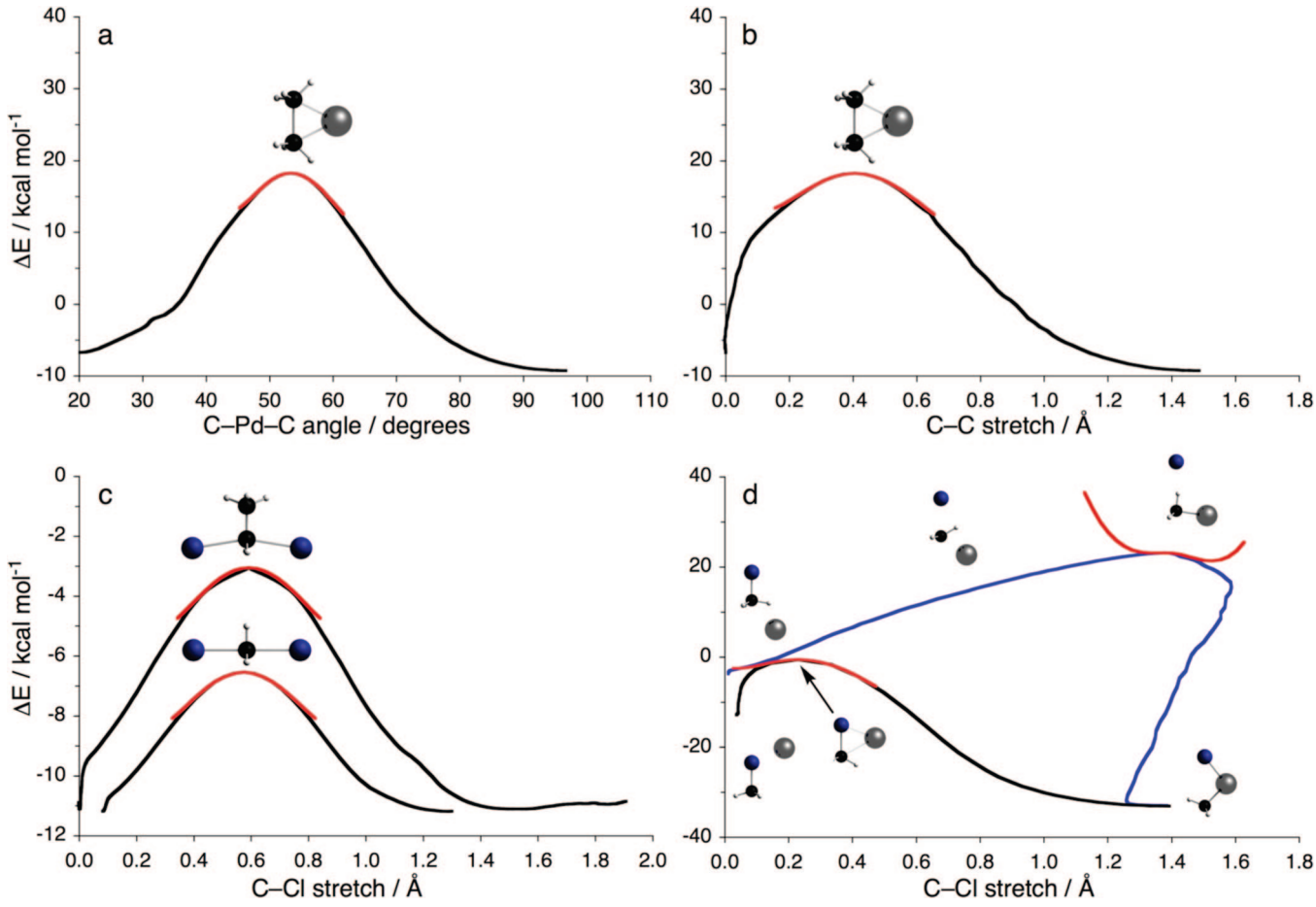


Figure 5. Comparison of reaction profiles based on the IRC (black/blue, full curves) and TV-IRC (red, partial curves): (a) oxidative insertion of Pd + ethane with the C–Pd–C angle as a reaction coordinate, (b) oxidative insertion of Pd + ethane with the C–C stretch as a reaction coordinate, (c) S_N2 reactions of Cl^- + CH_3Cl and Cl^- + CH_3CH_2Cl with the C–Cl stretch as a reaction coordinate, and (d) oxidative insertion versus the S_N2 mechanism for oxidative addition of Pd + CH_3Cl (note the breakdown of the TV-IRC approximation in the case of the latter S_N2 mechanism; see text).

the saddle point) depends on the choice of reaction coordinate onto which the IRC is projected. Here, we have shown that the choice of reaction coordinate is critical for zooming in on the important stage of the reaction and for revealing the origin of trends along series of reactions. Some criteria for a good reaction coordinate are (i) a large amplitude in coordinates that define the overall reaction, for example, the C−X bond into which a metal oxidatively inserts, (ii) a large amplitude in the transition vector, that is, the normal mode associated with a negative force constant that leads from the TS to the steepest descent paths, and (iii) preservation of this amplitude over a sufficiently long interval along the reaction path before and after the TS.

Furthermore, we have shown that the transition vector can serve as a useful and computationally much more efficient approximation (designated TV-IRC approximation) for full IRC computations if condition iii is satisfied. We are planning to use this approximation for extended activation-strain analyses[3] of computationally demanding oxidative addition reactions (large substrates and catalyst complexes) through direct insertion of the metal center into the activated bond. We have also shown that the TV-IRC approximation breaks down and can thus not be used for the competing S_N2 pathway of oxidative addition because here condition iii is not satisfied.

Acknowledgment. We thank the National Research School Combination-Catalysis (NRSC-C) and The Netherlands Organization for Scientific Research (NWO-CW) for financial support.

Supporting Information Available: Graphs of energies as a function of the IRC with the position of the transition state set to zero. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Ziegler, T.; Autschbach, J. *Chem. Rev.* **2005**, *105*, 2695–2722.

(2) (a) Bickelhaupt, F. M. *J. Comput. Chem.* **1999**, *20*, 114–128. (b) Diefenbach, A.; de Jong, G. T.; Bickelhaupt, F. M. *Mol. Phys.* **2005**, *103*, 995–998. (c) Bickelhaupt, F. M.; Baerends, E. J.; Nibbering, N. M. M. *Chem.—Eur. J.* **1996**, *2*, 196–207.

(3) de Jong, G. T.; Bickelhaupt, F. M. *ChemPhysChem* **2007**, *8*, 1170–1181.

(4) (a) Deng, L. Q.; Ziegler, T. *Int. J. Quantum Chem.* **1994**, *52*, 731–765. (b) Deng, L. Q.; Ziegler, T.; Fan, L. Y. *J. Chem. Phys.* **1993**, *99*, 3823–3835.

(5) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363–368.

(6) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523–5527.

(7) (a) Ishida, K.; Morokuma, K.; Komornicki, A. *J. Chem. Phys.* **1977**, *66*, 2153–2156. (b) Fukui, K. *J. Phys. Chem.* **1970**, *74*, 4161–4163.

(8) Heidrich, D. *The Reaction Path in Chemistry: Current Approaches and Perspectives*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995.

(9) (a) Yang, X. Y.; Wang, Y. C.; Geng, Z. Y.; Liu, Z. Y.; Wang, H. Q. *THEOCHEM* **2007**, *807*, 49–54. (b) Sakai, S. *J. Phys. Chem. A* **2006**, *110*, 12891–12899. (c) Lu, L. L.; Liu, X. W.; Wang, Y. C.; Wang, H. Q. *THEOCHEM* **2006**, *774*, 59–65. (d) Wang, Y. C.; Yang, X. Y.; Geng, Z. Y.; Liu, Z. Y. *Chem. Phys. Lett.* **2006**, *431*, 39–44. (e) Wang, Y. C.; Liu, Z. Y.; Geng, Z. Y.; Yang, X. Y. *Chem. Phys. Lett.* **2006**, *427*, 271–275. (f) Wang, Y. C.; Liu, Z. Y.; Geng, Z. Y.; Yang, X. Y.; Gao, L. G.; Chen, X. X. *THEOCHEM* **2006**, *765*, 27–34. (g) Domingo, L. R.; Picher, M. T.; Arroyo, P. *Eur. J. Org. Chem.* **2006**, 2570–2580. (h) Xie, J.; Feng, D. C.; Feng, S. Y.; Zhang, J. *THEOCHEM* **2005**, *755*, 55–63. (i) Shiota, Y.; Yasunaga, M.; Naka, A.; Ishikawa, M.; Yoshizawa, K. *Organometallics* **2004**, *23*, 4744–4749. (j) Shiota, Y.; Yoshizawa, K. *J. Chem. Phys.* **2003**, *118*, 5872–5879. (k) Sakai, S. *Int. J. Quantum Chem.* **2000**, *80*, 1099–1106. (l) Sakai, S. *J. Phys. Chem. A* **2000**, *104*, 922–927. (m) Nguyen, L. T.; Le, T. N.; De Proft, F.; Chandra, A. K.; Langenaeker, W.; Nguyen, M. T.; Geerlings, P. *J. Am. Chem. Soc.* **1999**, *121*, 5992–6001. (n) Yoshizawa, K.; Shiota, Y.; Yamabe, T. *J. Chem. Phys.* **1999**, *111*, 538–545. (o) Su, M. D.; Chu, S. Y. *J. Am. Chem. Soc.* **1999**, *121*, 1045–1058. (p) Sakai, S. *Int. J. Quantum Chem.* **1998**, *70*, 291–302. (q) Wong, C. K.; Li, W. K.; Baker, J. *THEOCHEM* **1995**, *357*, 75–86. (r) Deng, L. Q.; Branchadell, V.; Ziegler, T. *J. Am. Chem. Soc.* **1994**, *116*, 10645–10656. (s) Cundari, T. R. *J. Am. Chem. Soc.* **1994**, *116*, 340–347. (t) Taketsugu, T.; Hirano, T. *J. Chem. Phys.* **1993**, *99*, 9806–9814. (u) Yamabe, S.; Kawajiri, S.; Minato, T.; Machiguchi, T. *J. Org. Chem.* **1993**, *58*, 1122–1127.

(10) (a) Bento, A. P.; Bickelhaupt, F. M. *J. Org. Chem.* **2007**, *72*, 2201–2207. (b) van Bochove, M. A.; Swart, M.; Bickelhaupt, F. M. *J. Am. Chem. Soc.* **2006**, *128*, 10738–10744.

(11) van Stralen, J. N. P.; Bickelhaupt, F. M. *Organometallics* **2006**, *25*, 4260–4268.

(12) Baerends, E. J.; Autschbach, J. A.; Bérces, A.; Berger, J. A.; Bickelhaupt, F. M.; Bo, C.; de Boeij, P. L.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Harris, F. E.; van den Hoek, P.; Jacobsen, H.; Jensen, L.; Kadantsev, E. S.; van Kessel, G.; Klooster, R.; Kootstra, F.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Neugebauer, J.; Nicu, V. P.; Osinga, V. P.; Patchkovskii, S.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Romaniello, P.; Ros, P.; Schipper, P. R. T.; Schreckenbach, G.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E. M.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; Yakovlev, A.; Ziegler, T. *ADF2007.01*; SCM: Amsterdam, 2007.

(13) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.

(14) (a) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098. (b) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.

(15) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41.

(16) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1994**, *101*, 9783–9792.

(17) (a) de Jong, G. T.; Bickelhaupt, F. M. *J. Chem. Theory Comput.* **2006**, *2*, 322–335. (b) de Jong, G. T.; Geerke, D. P.; Diefenbach, A.; Solà, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2005**, *26*, 1006–1020.

(18) van Zeist, W. J.; Fonseca Guerra, C.; Bickelhaupt, F. M. *J. Comput. Chem.* **2008**, *29*, 312–315.

(19) (a) Fonseca Guerra, C.; Handgraaf, J. W.; Baerends, E. J.; Bickelhaupt, F. M. *J. Comput. Chem.* **2004**, *25*, 189–210. (b) Bickelhaupt, F. M.; van Eikema Hommes, N. J. R.; Fonseca Guerra, C.; Baerends, E. J. *Organometallics* **1996**, *15*, 2923–2931.

(20) Taichibana, A.; Fukui, K. *Theor. Chim. Acta* **1979**, *51*, 189–206.

CT700214V

JCTC Journal of Chemical Theory and Computation

E2 and S_N2 Reactions of $X^- + CH_3CH_2X$ (X = F, Cl); an *ab Initio* and DFT Benchmark Study

A. Patrícia Bento,[†] Miquel Solà,[‡] and F. Matthias Bickelhaupt*,[†]

Department of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands, and Institut de Química Computacional, Universitat de Girona, Campus Montilivi, E-17071 Girona, Catalonia, Spain

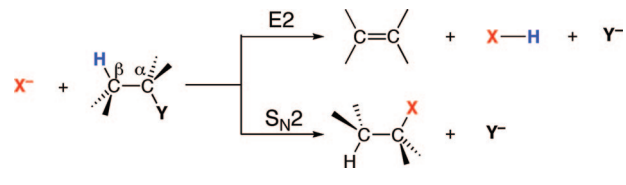
Received November 20, 2007

Abstract: We have computed consistent benchmark potential energy surfaces (PESs) for the *anti*-E2, *syn*-E2, and S_N2 pathways of $X^- + CH_3CH_2X$ with X = F and Cl. This benchmark has been used to evaluate the performance of 31 popular density functionals, covering local-density approximation, generalized gradient approximation (GGA), meta-GGA, and hybrid density-functional theory (DFT). The *ab initio* benchmark has been obtained by exploring the PESs using a hierarchical series of *ab initio* methods [up to CCSD(T)] in combination with a hierarchical series of Gaussian-type basis sets (up to aug-cc-pVQZ). Our best CCSD(T) estimates show that the overall barriers for the various pathways increase in the order *anti*-E2 (X = F) < S_N2 (X = F) < S_N2 (X = Cl) ∼ *syn*-E2 (X = F) < *anti*-E2 (X = Cl) < *syn*-E2 (X = Cl). Thus, *anti*-E2 dominates for $F^- + CH_3CH_2F$, and S_N2 dominates for $Cl^- + CH_3CH_2Cl$, while *syn*-E2 is in all cases the least favorable pathway. Best overall agreement with our *ab initio* benchmark is obtained by representatives from each of the three categories of functionals, GGA, meta-GGA, and hybrid DFT, with mean absolute errors in, for example, central barriers of 4.3 (OPBE), 2.2 (M06-L), and 2.0 kcal/mol (M06), respectively. Importantly, the hybrid functional BHandH and the meta-GGA M06-L yield incorrect trends and qualitative features of the PESs (in particular, an erroneous preference for S_N2 over the *anti*-E2 in the case of $F^- + CH_3CH_2F$) even though they are among the best functionals as measured by their small mean absolute errors of 3.3 and 2.2 kcal/mol in reaction barriers. OLYP and B3LYP have somewhat higher mean absolute errors in central barriers (5.6 and 4.8 kcal/mol, respectively), but the error distribution is somewhat more uniform, and as a consequence, the correct trends are reproduced.

1. Introduction

Base-induced elimination (E2) and nucleophilic substitution (S_N2) constitute two fundamental types of chemical reactions that play an important role in organic synthesis.[1] E2 elimination is, in principle, always in competition with S_N2 substitution, and the two pathways may occur as unwanted side reactions of each other (see Scheme 1). Gas-phase experiments have enabled the study of the intrinsic reactivity of reaction systems without the interference of solvent molecules. The resulting insights, in turn, can also shed light on the nature of the E2 and S_N2 reactions in solution, in

Scheme 1. E2 and S_N2 Reactions



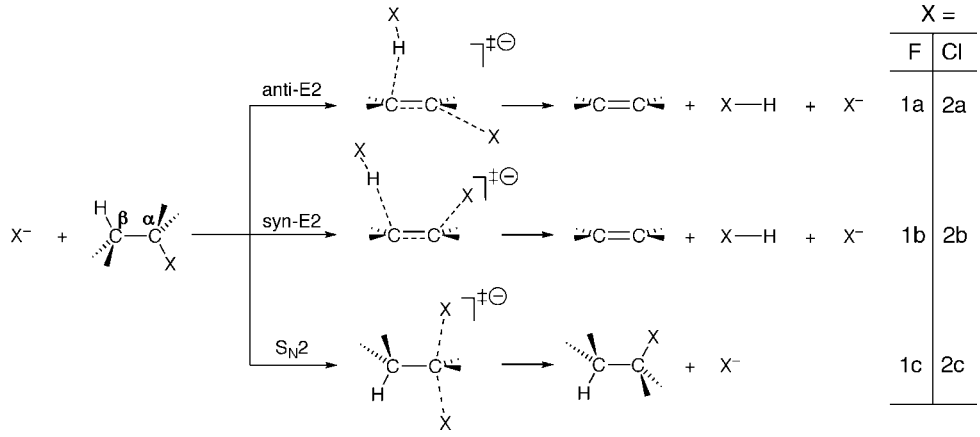
* Corresponding author fax: +31-20-59 87629; e-mail: FM.Bickelhaupt@few.vu.nl.

[†] Scheikundig Laboratorium der Vrije Universiteit.

[‡] Universitat de Girona.

10.1021/ct700318e CCC: $40.75
Published on Web 05/24/2008

Scheme 2. E2 and S_N2 Pathways for $X^- + CH_3CH_2X$



particular the effect of the solvent, by comparing the gas-phase[2] results with those of condensed-phase[3] experiments. The various experimental investigations have over the years been augmented by an increasing number of theoretical studies, which provide a detailed description of the stationary points and the potential energy surfaces (PESs) that determine the feasibility of the various competing E2 and S_N2 reaction channels.[4]

The purpose of the present study is 2-fold. First, we wish to obtain reliable benchmarks for the PESs of the E2 and S_N2 reactions of $F^- + CH_3CH_2F$ as well as $Cl^- + CH_3CH_2Cl$ (see reactions 1 and 2 in Scheme 2). Note that E2 eliminations can in principle proceed via two stereochemical, different pathways, namely, with the base and the β-proton anti- (*anti*-E2) and syn-periplanar (*syn*-E2) with respect to the leaving group (compare reactions a and b, respectively, in Scheme 2). This is done by exploring for both reaction systems the PESs of each of the three reaction mechanisms with a hierarchical series of *ab initio* methods [HF, MP2, MP4, CCSD, and CCSD(T)] in combination with a hierarchical series of Gaussian-type basis sets of increasing flexibility [up to quadruple-ζ + diffuse functions for reactions involving F and up to (triple+*d*)-ζ + diffuse functions for reactions involving Cl]. Our purpose is to provide a consistent set of *ab initio* PES data for accurately estimating trends associated with going from $F^- + CH_3CH_2F$ to $Cl^- + CH_3CH_2Cl$ as well as along *anti*-E2, *syn*-E2, and S_N2 pathways.

A second purpose is to evaluate and validate the performance of several popular density functionals for describing the above elimination and nucleophilic substitution reactions (see Scheme 2) against our *ab initio* benchmark PESs for the six model reactions. Although the *ab initio* approach is satisfactory in terms of accuracy and reliability, it is at the same time prohibitively expensive if one wishes to study more realistic model reactions involving larger nucleophiles and substrates. Thus, a survey of density functionals serves to validate one or more of these density functional theory (DFT) approaches as a computationally more efficient alternative to high-level *ab initio* theory in future investigations. A general concern associated with the application of DFT to the investigation of chemical reactions is its notorious tendency to underestimate activation energies.[5] Thus, we arrive at a ranking of density functional approaches in terms of the accuracy with which they describe the PES of our model reaction, in particular, the activation energy. We focus on the overall activation energy, that is, the difference in energy between the TS and the separate reactants,[6] as well as the central barrier, that is, the difference in energy between the TS and the reactant complex. Previous studies have shown that S_N2 reaction profiles obtained with OLYP and B3LYP agree satisfactorily with highly correlated *ab initio* benchmarks.[5c,7] Merrill et al.[4g] have shown that B3LYP in combination with the aug-cc-pVDZ basis set performs reasonably well for the E2 and S_N2 reactions of $F^- + CH_3CH_2F$ with deviations from G2+ of up to 3.5 kcal/mol but that it fails in locating the transition state associated with the *anti*-E2 elimination. Guner et al.[7g] have also shown that OLYP and O3LYP give comparable results to B3LYP and that these functionals work well for organic reactions. Very recently, Truhlar and co-worker[7h] have carried out an exhaustive performance analysis of various density functionals for describing barrier heights which shows that, for closed-shell S_N2 reactions, M06 and M06-2X perform best, followed by PBEh and M05-2X. B3LYP is also found to work reasonably well.

2. Methods

2.1. DFT Geometries and Potential Energy Surfaces. All DFT calculations were done with the Amsterdam Density Functional (ADF) program developed by Baerends and others.[8] Geometry optimizations have been carried out with the OLYP[9] density functional, which yields robust and accurate geometries.[7a] This density functional was used in combination with the TZ2P basis set, in which the molecular orbitals were expanded in a large uncontracted set of Slater-type orbitals (STOs) containing diffuse functions, and is of triple-ζ quality, being augmented with two sets of polarization functions: 2p and 3d on hydrogen and 3d and 4f on carbon, fluorine, and chlorine. The core shells of carbon (1s), fluorine (1s), and chlorine (1s2s2p) were treated by the frozen-core approximation.[8b,10] An auxiliary set of s, p, d, f, and g STOs was used to fit the molecular density and to represent the Coulomb and exchange potentials accurately in each self-consistent field (SCF) cycle. All stationary points

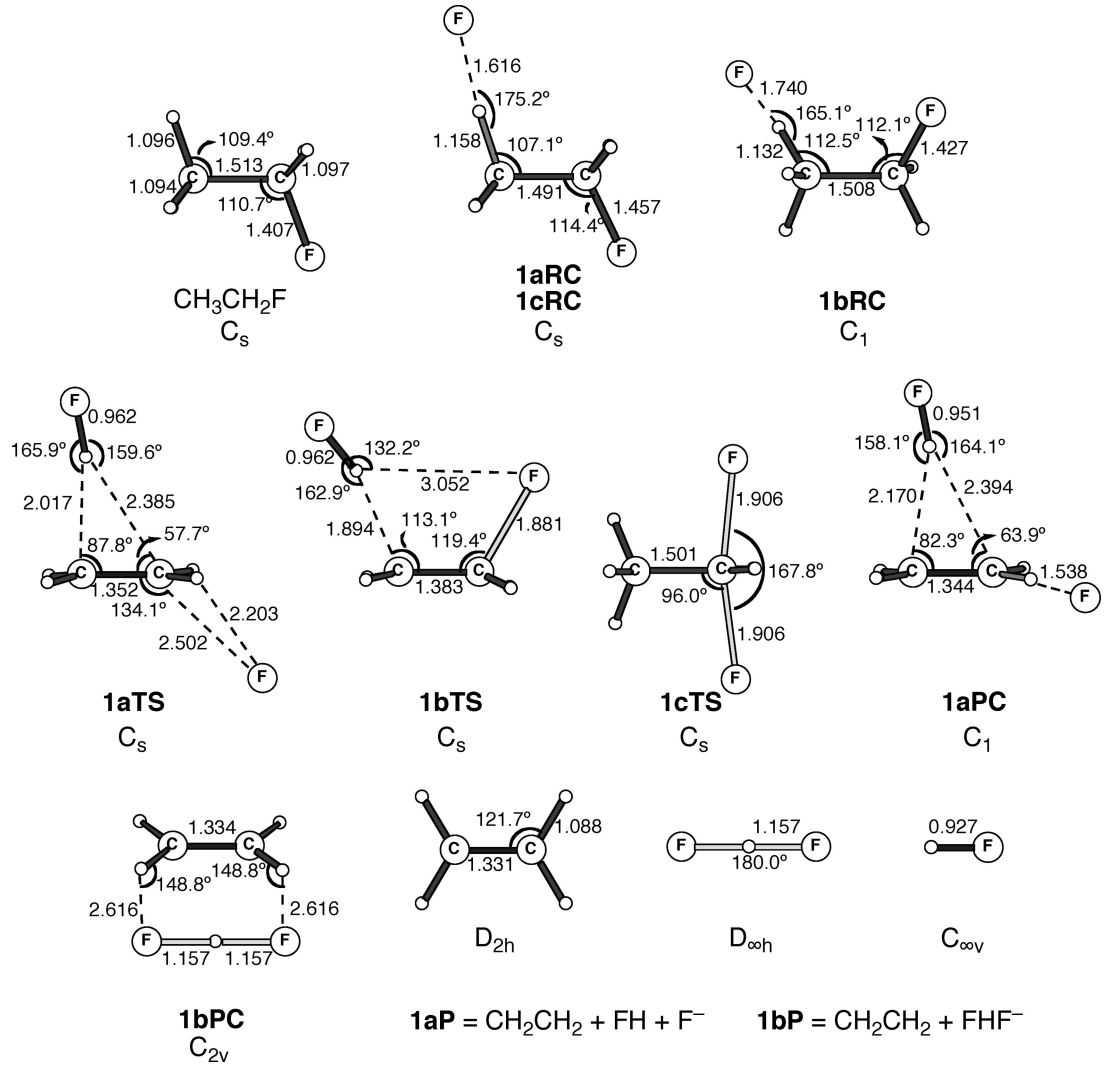


Figure 1. Geometries (in Å, deg) of stationary points along the potential energy surfaces for the *anti*-E2, *syn*-E2, and S_N2 reactions of $F^- + CH_3CH_2F$ (reaction 1), computed at OLYP/TZ2P.

were confirmed to be equilibrium structures (no imaginary frequencies) or a transition state[11] (one imaginary frequency) through vibrational analysis.[12]

In addition, based on OLYP/TZ2P geometries, we have computed the relative energies of stationary points along the PES for several density functionals: the local-density approximation (LDA) functional VWN;[13] the generalized gradient approximation (GGA) functionals BP86,[14] BLYP,[9b,14a] PW91,[15] PBE,[16] RPBE,[17] revPBE,[18] FT97,[19] HCTH/93,[20] HCTH/120,[21] HCTH/147,[21] HCTH/407,[22] BOP,[14a,23] and OPBE;[9a,16] the meta-GGA functionals PKZB,[24] VS98,[25] BLAP3,[26] OLAP3,[9a,26] TPSS,[27] and M06-L;[28] and the hybrid functionals B3LYP,[9b,29] O3LYP,[30] KMLYP,[31] BHandH,[32] mPBE0KCIS,[33] mPW1K,[34] M05,[35] M05-2X,[36] M06,[7h,37] and M06-2X.[7h,37] For technical reasons (i.e., frozen-core approximation and potentials in ADF are not available for all functionals), the energies obtained with these functionals were computed with an *all-electron* TZ2P basis set (ae-TZ2P) and in a post-SCF manner, that is, using the electron density obtained at OLYP/ae-TZ2P. This approximation has been extensively tested and has been shown to introduce an error in the computed energies of only a few tenths of a kilocalorie per mole.[38]

2.2. ***Ab Initio*** **Potential Energy Surfaces.** On the basis of the OLYP/TZ2P geometries, energies of the stationary points were computed in a series of single-point calculations with the program package Gaussian[39] using the following hierarchy of quantum chemical methods: Hartree−Fock (HF), Møller−Plesset perturbation theory[40] through the second order (MP2) and fourth order (MP4),[41] and couple-cluster theory[42] with single and double excitations (CCSD)[43] and triple excitations treated perturbatively [CCSD(T)].[44] At each level of theory, we used Dunning's[45] augmented correlation consistent polarized valence basis sets of double-, triple-, and quadruple-ζ quality, that is, aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ for the reactions involving F, and the modified second-row basis sets aug-cc-pV(D+*d*)Z and aug-cc-pV(T+*d*)Z for the reactions involving Cl (limitations of our computational resources prevented us from carrying out calculations with the aug-cc-pV(Q+*d*)Z basis set for the latter reactions). Furthermore, using eq 7 of ref 46, we have extrapolated the CCSD(T) energies to the complete basis set (CBS) values CBS-23 (i.e., based on aug-cc-pVDZ and aug-cc-pVTZ values for reactions involving F and aug-cc-pV(D+*d*)Z and aug-cc-pV(T+*d*)Z values for reactions involving Cl) and CBS-34 (i.e., based on aug-cc-

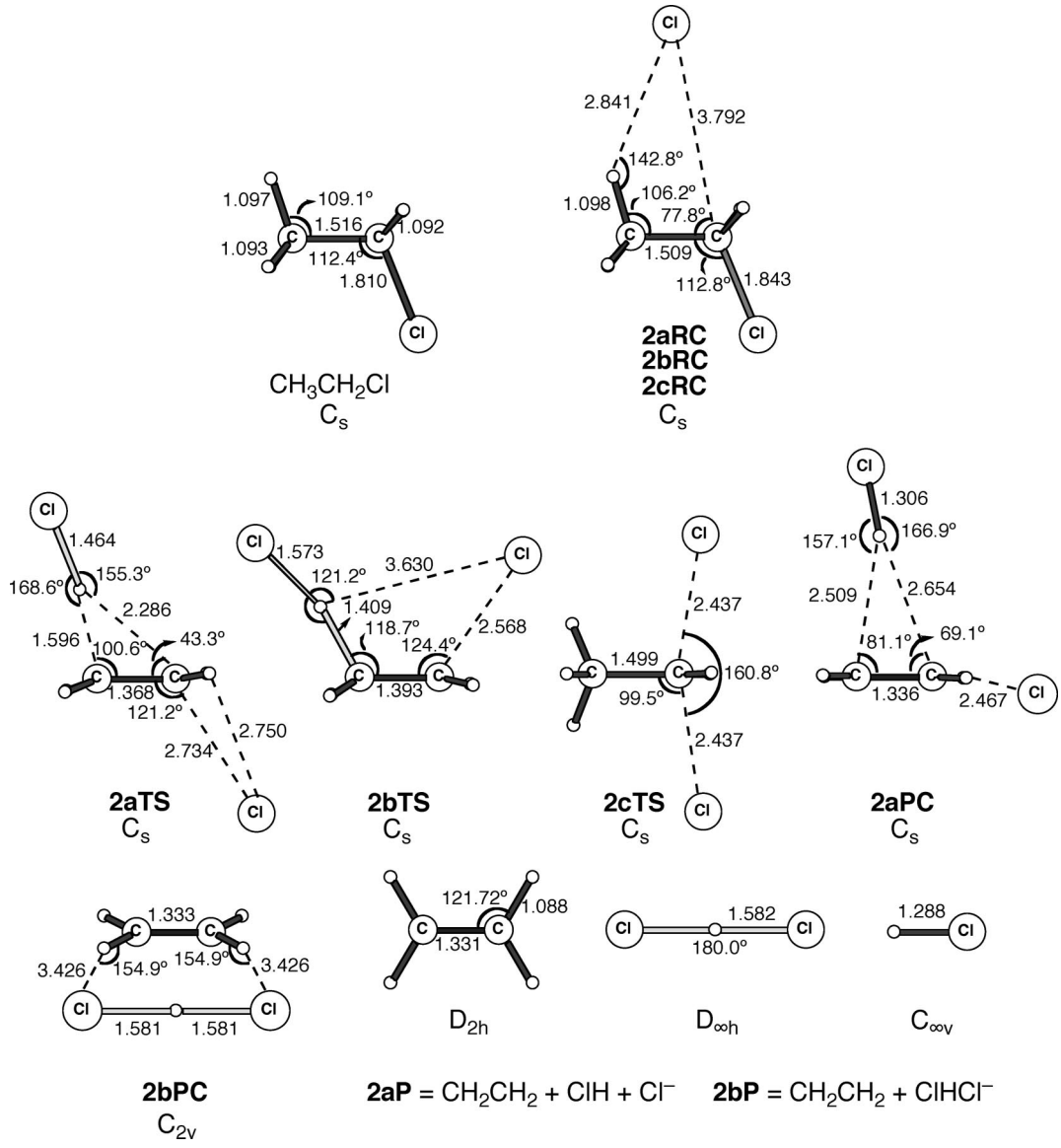


Figure 2. Geometries (in Å, deg) of stationary points along the potential energy surfaces for the *anti*-E2, *syn*-E2, and S_N2 reactions of $Cl^- + CH_3CH_2Cl$ (reaction 2), computed at OLYP/TZ2P.

pVTZ and aug-cc-pVQZ values, only for the reactions involving F).

3. Results and Discussion

3.1. Geometries of Stationary Points and Reaction Paths. First, we examine the geometries of stationary points along the reaction coordinate of *anti*-E2, *syn*-E2, and S_N2 reactions of $F^- + CH_3CH_2F$ and $Cl^- + CH_3CH_2Cl$. Previous studies have shown that the GGA functional OLYP is numerically robust and agrees well with available experimental and CCSD(T) geometries.[7a] Therefore, we choose OLYP in combination with the TZ2P basis set, to compute the geometries of the stationary points of our model reactions 1 and 2 (see Scheme 2). The resulting geometry parameters are collected in Figures 1 and 2, respectively (for full structural details, see Cartesian coordinates in Table S1 of the Supporting Information).

For both $F^- + CH_3CH_2F$ and $Cl^- + CH_3CH_2Cl$, the *anti*-E2, *syn*-E2, and S_N2 reactions proceed from the reactants via formation of a reactant complex (RC) toward the transition state (TS) and, finally, a product complex (PC). In the *anti*-E2 reactant complex, the base X^- binds to the $C^\beta-H$ bond that is anti to $C^\alpha-X$ with $X^-\cdots H^\beta$ distances of 1.616 and 2.841 Å in **1aRC** and **2aRC**, respectively (see Figures 1 and 2). The $C^\beta-H$ bond that participates in the hydrogen bond with the halide anion expands by 0.062 Å in **1aRC** (from 1.096 to 1.158 Å) and only very slightly, that is, by 0.001 Å, in **2aRC** (from 1.097 to 1.098) if compared to the isolated substrates CH_3CH_2F and CH_3CH_2Cl, respectively. In the *anti*-E2 transition states **1aTS** and **2aTS**, the elongation of the $C^\beta-H$ bonds further increases to 0.921 and 0.499 Å, respectively, again relative to the isolated substrates. The resulting product complexes **1aPC** and **2aPC** are composed of three rigid fragments, the conjugate acid HX, the olefin CH_2CH_2, and the leaving group X^-, which may eventually separate into products (**1aP** and **2aP**).

The *syn*-E2 elimination proceeds only in the case of $F^- + CH_3CH_2F$ via a separate reactant complex **1bRC** (see

Table 1. Relative Energies (in kcal/mol) of Stationary Points along the Reaction Coordinate for the *anti*-E2, *syn*-E2, and S_N2 Reactions of $F^- + CH_3CH_2F$, Computed at Several Levels of the *ab Initio* Theory

	anti-E2				*syn*-E2				S_N2	
method	**1aRC**	**1aTS**	**1aPC**	**1aP**	**1bRC**	**1bTS**	**1bPC**	**1bP**	**1cRC**	**1cTS**
					aug-cc-pVDZ					
HF	−10.49	4.81	0.12	16.77	−7.62	18.08	−28.00	−23.55	−10.49	8.71
MP2	−15.23	−1.77	−6.90	15.96	−11.01	4.86	−33.50	−27.40	−15.23	1.03
MP4	−15.64	−1.44	−6.03	17.10	−11.49	5.00	−31.68	−25.30	−15.64	−2.74
CCSD	−14.76	−0.30	−5.43	16.42	−10.92	8.38	−32.65	−26.70	−14.76	1.80
CCSD(T)	−15.81	−2.03	−7.16	16.11	−11.71	5.16	−33.88	−27.53	−15.81	−1.06
					aug-cc-pVTZ					
HF	−9.63	5.22	0.72	16.63	−7.05	18.42	−28.46	−24.25	−9.63	11.35
MP2	−14.69	−1.07	−6.06	16.48	−10.69	5.07	−33.81	−27.99	−14.69	3.56
MP4	−15.02	−0.88	−5.35	17.29	−11.08	5.05	−32.33	−26.27	−15.02	−0.20
CCSD	−14.13	0.56	−4.47	17.01	−10.54	8.89	−32.96	−27.29	−14.13	4.58
CCSD(T)	−15.17	−1.31	−6.28	16.51	−11.30	5.47	−34.30	−28.28	−15.17	1.56
					CBS[a]					
CCSD(T)	−15.27	−1.19	−6.17	16.74	−11.37	5.46	−34.28	−28.29	−15.27	1.55
					aug-cc-pVQZ					
HF	−9.58	5.12	0.57	16.31	−7.03	18.30	−28.43	−24.31	−9.58	11.50
MP2	−14.61	−1.25	−6.33	15.88	−10.60	4.93	−33.92	−28.30	−14.61	3.81
CCSD	−14.00	0.50	−4.61	16.45	−10.41	8.91	−32.98	−27.53	−14.00	4.97
CCSD(T)	−14.99	−1.33	−6.39	15.95	−11.12	5.54	−34.27	−28.49	−14.99	1.99
					CBS[b]					
CCSD(T)	−14.89	−1.27	−6.35	15.77	−11.00	5.68	−37.39	−28.60	−14.89	2.20

[a] These values were obtained from two-point fits (aug-cc-pVDZ and aug-cc-pVTZ) to eq 7 of ref 46. [b] These values were obtained from two-point fits (aug-cc-pVTZ and aug-cc-pVQZ) to eq 7 of ref 46.

Table 2. Relative Energies (in kcal/mol) of Stationary Points along the Reaction Coordinate for the *anti*-E2, *syn*-E2, and S_N2 Reactions of $Cl^- + CH_3CH_2Cl$, Computed at Several Levels of the *ab Initio* Theory

	anti-E2				*syn*-E2				S_N2	
method	**2aRC**	**2aTS**	**2aPC**	**2aP**	**2bRC**	**2bTS**	**2bPC**	**2bP**	**2cRC**	**2cTS**
					aug-cc-pV(D + d)Z					
HF	−9.33	26.88	10.03	17.93	−9.33	43.57	0.43	2.61	−9.33	9.06
MP2	−11.33	16.22	9.19	22.67	−11.33	29.51	−5.21	−1.50	−11.33	6.67
MP4	−11.45	16.22	8.08	21.30	−11.45	29.12	−5.01	−1.27	−11.45	4.39
CCSD	−10.98	18.95	8.30	20.40	−10.98	33.10	−4.12	−0.67	−10.98	6.43
CCSD(T)	−11.43	16.14	7.50	20.52	−11.43	29.17	−5.57	−1.86	−11.43	4.12
					aug-cc-pV(T + d)Z					
HF	−9.06	28.04	10.27	17.61	−9.06	44.99	0.99	3.04	−9.06	10.38
MP2	−11.06	17.90	10.80	23.86	−11.06	31.06	−4.42	−0.90	−11.06	8.22
CCSD	−10.64	20.97	9.90	21.38	−10.64	35.11	−3.15	0.07	−10.64	8.15
CCSD(T)	−11.10	17.92	9.17	21.58	−11.10	30.82	−4.90	−1.42	−11.10	5.70
					CBS[a]					
CCSD(T)	−11.07	18.18	9.77	22.16	−11.07	30.92	−4.85	−1.42	−11.07	5.81

[a] These values were obtained from two-point fits [aug-cc-pV(D + d)Z and aug-cc-pV(T+*d*)Z] to eq 7 of ref 46.

Figure 1). For $Cl^- + CH_3CH_2Cl$, all three elementary reactions (*anti*-E2, *syn*-E2, and S_N2) go via one and the same reactant complex, that is, **2aRC** = **2bRC** = **2cRC** (see Figure 2). In the *syn*-E2 transition states **1bTS** and **2bTS**, the C^β−H bonds are elongated by 0.798 and 0.312 Å and are oriented syn with respect to the C^α−X bond (see Figures 1 and 2). At variance with the *anti*-E2 pathway, the *syn*-E2 pathway leads to product complexes, **1bPC** and **2bPC**, that are composed of *two* rigid fragments: the leaving group microsolvated by the conjugate acid, XHX^-, and the olefin, CH_2CH_2. These product complexes are predestined to dissociate into the products $CH_2CH_2 + XHX^-$ (**1bP** and **2bP**).

S_N2 substitution proceeds for both $F^- + CH_3CH_2F$ and $Cl^- + CH_3CH_2Cl$, from the same reactant complex as the *anti*-E2 elimination (i.e., aRC = cRC). But now, the halide anion approaches to the backside of the α-methyl group of the substrate, which leads to the S_N2 transition states **1cTS** and **2cTS** in which a new X−C^α bond has been partially formed while simultaneously the old C^α−X bond has been elongated (see Figures 1 and 2). Note that, in our symmetric S_N2 model reactions, the nucleophile−C^α and C^α−leaving-group bonds are of the same length, namely, 1.906 and 2.437 Å in **1cTS** and **2cTS** (see Figures 1 and 2), and that the product complexes and products are identical to the corresponding reactant complexes and reactants.

3.2. *Ab Initio* Benchmark Potential Energy Surfaces. On the basis of the above OLYP/TZ2P geometries, we have computed our *ab initio* benchmark potential energy surfaces, which are summarized as relative energies in Tables 1 and 2 for reactions 1 and 2, respectively. The extrapolated CBS CCSD(T) values are also listed therein.

Table 3. Relative Energies (in kcal/mol) of Stationary Points along the Reaction Coordinate for the *anti*-E2, *syn*-E2, and S_N2 Reactions of $F^- + CH_3CH_2F$, Computed at Several Levels of the Density Functional Theory

	anti-E2				*syn*-E2				S_N2	
method	**1aRC**	**1aTS**	**1aPC**	**1aP**	**1bRC**	**1bTS**	**1bPC**	**1bP**	**1cRC**	**1cTS**
					LDA					
VWN	−28.23	−8.54	−13.84	26.01	−22.02	−12.50	−42.70	−35.13	−28.23	−13.67
					GGAs					
BP86	−22.19	−8.27	−12.55	16.97	−16.77	−7.51	−40.68	−35.87	−22.19	−9.33
BLYP	−22.23	−11.55	−15.08	13.55	−17.02	−8.66	−43.95	−38.71	−22.23	−11.27
PW91	−24.12	−9.58	−13.67	18.85	−18.66	−9.29	−42.15	−35.61	−24.12	−11.39
PBE	−23.79	−9.36	−13.49	18.48	−18.37	−8.98	−41.73	−35.43	−23.79	−10.73
RPBE	−21.79	−9.61	−13.49	14.89	−16.74	−7.39	−41.71	−35.87	−21.79	−8.56
revPBE	−21.37	−8.83	−12.85	15.00	−16.26	−6.85	−40.94	−35.73	−21.37	−7.91
FT97	−19.78	−6.54	−11.07	13.27	−14.08	−4.80	−37.86	−35.53	−19.78	−7.19
HCTH/93	−18.75	−7.32	−11.79	12.26	−13.98	−3.73	−40.42	−36.26	−18.75	−2.52
HCTH/120	−21.90	−9.85	−14.05	15.07	−16.92	−7.17	−42.93	−36.60	−21.90	−7.37
HCTH/147	−21.07	−8.93	−13.26	14.58	−16.10	−6.23	−42.13	−36.45	−21.07	−6.14
HCTH/407	−21.52	−10.56	−14.71	13.39	−16.74	−6.60	−44.06	−37.30	−21.52	−5.78
BOP	−19.67	−9.98	−13.56	11.28	−14.68	−6.01	−42.21	−38.16	−19.67	−7.66
OPBE	−18.68	−3.26	−8.45	15.78	−13.79	−2.07	−36.62	−32.69	−18.68	−0.22
OLYP	−20.01	−7.95	−12.49	12.85	−15.20	−4.93	−41.40	−36.41	−20.01	−4.16
					meta-GGAs					
PKZB	−19.16	−6.56	−9.65	14.74	−14.55	−3.93	−38.36	−32.85	−19.16	−7.27
VS98	−20.80	−13.42	−15.04	11.99	−16.25	−7.04	−43.88	−35.97	−20.80	−14.06
BLAP3	−18.54	−8.58	−12.38	12.58	−14.01	−2.47	−41.65	−36.58	−18.54	−4.88
OLAP3	−16.23	−4.62	−9.31	12.24	−12.13	1.57	−38.65	−33.88	−16.23	2.25
TPSS	−21.38	−5.26	−8.94	19.81	−16.28	−4.16	−37.83	−32.52	−21.38	−10.03
M06-L	−20.04	−1.23	−5.44	20.54	−15.33	1.68	−32.57	−27.78	−20.04	−2.95
					Hybrid Functionals					
B3LYP	−19.30	−5.38	−10.66	15.90	−14.50	−2.00	−40.32	−35.34	−19.30	−4.01
O3LYP	−18.12	−2.55	−7.97	16.52	−13.46	0.40	−38.06	−33.35	−18.12	0.24
KMLYP	−16.14	6.09	−2.78	23.69	−11.77	8.28	−33.82	−28.53	−16.14	7.54
BHandH	−19.68	3.90	−4.81	26.52	−14.87	3.86	−35.53	−29.15	−19.68	2.76
mPBE0KCIS	−19.57	−4.44	−10.63	16.89	−14.77	−1.77	−39.63	−33.94	−19.57	−1.15
mPW1K	−15.32	4.26	−3.38	20.37	−10.96	7.06	−33.31	−29.07	−15.32	6.24
M05	−18.68	−3.51	−8.73	18.54	−14.64	0.81	−38.03	−32.01	−18.68	−0.81
M05-2X	−14.53	0.99	−5.72	18.33	−10.30	3.85	−39.28	−34.27	−14.53	3.97
M06	−18.21	−2.21	−7.47	17.88	−13.96	1.14	−35.19	−30.59	−18.21	−0.35
M06-2X	−15.67	1.49	−5.62	18.37	−11.47	4.03	−37.77	−32.90	−15.67	5.82

First, we examine the PES obtained for the *anti*-E2 elimination of $F^- + CH_3CH_2F$. The energy of the respective reactant complex, **1aRC**, computed with our best basis set (aug-cc-pVQZ) ranges from −9.58 to −14.61 to −14.00 to −14.99 kcal/mol for HF, MP2, CCSD, and CCSD(T). Note that, due to large space requirements, full MP4 calculations for the QZ basis set were not possible. The three highest-level values are equal to each other within 1.0 kcal/mol (see Table 1). Similarly, the energy of the transition state, **1aTS**, computed again with our best basis set (aug-cc-pVQZ) varies from +5.12 to −1.25 to +0.50 to −1.33 kcal/mol for HF, MP2, CCSD, and CCSD(T), respectively. Thus, not unexpectedly, HF significantly overestimates the overall barrier, which is significantly reduced by the incorporation of Coulomb correlation into theoretical treatment. The inclusion of the triple excitations within the CCSD method further reduces the overall barrier by 1.8 kcal/mol. The three highest-level values are within a range of 1.8 kcal/mol. Furthermore, the CCSD(T) values are converged to the basis-set size (at aug-cc-pVQZ) to within a few hundreds of a kilocalorie per mole for the RC and the TS (see Table 1). Note that CBS CCSD(T) values do not differ much from the best pure values [CCSD(T)].

For the *anti*-E2 elimination of $Cl^- + CH_3CH_2Cl$, the energy of the reactant complex, **2aRC**, computed with our best basis set [now, with aug-cc-pV(T+*d*)Z] varies relatively little along the range of methods, that is ca. 2 kcal/mol, from −9.06 to −11.06 to −10.64 to −11.10 for HF, MP2, CCSD and CCSD(T), respectively (see Table 2). Now, our three highest-level values are equal to each other within 0.5 kcal/mol. At variance, the energy of the transition state, **2aTS**, depends more delicately on the level at which correlation is treated. This TS energy computed again with aug-cc-pV(T+*d*)Z varies from 28.04 to 17.90 to 20.97 to 17.92 kcal/mol along HF, MP2, CCSD, and CCSD(T), respectively. Note how HF dramatically overestimates the overall barrier, that is, by ca. 10 kcal/mol! Also note the substantial impact of including triple excitations in the CCSD approach, which reduced the overall barrier by an additional 3.0 kcal/mol. The three highest-level values are now distributed over a range of 3.1 kcal/mol (see Table 2).

Next, we examine the PES of the *syn*-E2 elimination of $F^- + CH_3CH_2F$. The energy of reactant complex **1bRC** computed with our best basis set (aug-cc-pVQZ) shows a similar behavior as that of the *anti*-E2 elimination. The energy of this RC varies from −7.03 to −10.60 to −10.41 to −11.12 kcal/mol for HF, MP2, CCSD, and CCSD(T), respectively, and the three highest-level values are within a range of less than a kcal/mol (see Table 1). In turn, the energy of the TS is more sensitive to the level at which correlation

Table 4. Relative Energies (in kcal/mol) of Stationary Points along the Reaction Coordinate for the *anti*-E2, *syn*-E2, and S_N2 Reactions of $Cl^- + CH_3CH_2Cl$, Computed at Several Levels of the Density Functional Theory

	anti-E2				*syn*-E2				S_N2	
method	**2aRC**	**2aTS**	**2aPC**	**2aP**	**2bRC**	**2bTS**	**2bPC**	**2bP**	**2cRC**	**2cTS**
					LDA					
VWN	−13.32	5.00	12.19	29.51	−13.32	11.45	−10.74	−7.11	−13.32	−4.64
					GGAs					
BP86	−10.66	7.21	8.80	20.08	−10.66	15.35	−11.23	−9.24	−10.66	−1.92
BLYP	−11.08	5.28	4.43	15.33	−11.08	14.04	−14.17	−11.93	−11.08	−3.69
PW91	−12.23	6.38	8.42	22.36	−12.23	14.22	−11.98	−8.56	−12.23	−3.24
PBE	−11.91	6.85	8.62	22.10	−11.91	14.75	−11.62	−8.49	−11.91	−2.43
RPBE	−11.20	7.78	6.85	18.45	−11.20	16.27	−12.47	−9.38	−11.20	−0.67
revPBE	−10.69	8.13	7.59	18.55	−10.69	16.59	−11.89	−9.29	−10.69	−0.20
FT97	−7.86	10.09	11.41	17.68	−7.86	19.37	−7.85	−6.76	−7.86	−0.04
HCTH/93	−9.37	10.25	6.99	15.53	−9.37	19.56	−11.80	−9.74	−9.37	3.82
HCTH/120	−11.60	7.46	6.03	18.16	−11.60	16.42	−13.03	−9.56	−11.60	−0.49
HCTH/147	−10.89	8.14	6.62	17.62	−10.89	17.15	−12.62	−9.54	−10.89	0.50
HCTH/407	−11.71	8.19	4.55	16.45	−11.71	17.82	−13.88	−9.93	−11.71	1.99
BOP	−9.91	6.83	4.60	13.45	−9.91	15.96	−13.74	−12.01	−9.91	−1.20
OPBE	−8.64	13.99	12.33	20.78	−8.64	22.32	−8.34	−6.24	−8.64	7.56
OLYP	−9.66	10.68	7.45	16.33	−9.66	19.58	−11.81	−9.28	−9.66	4.04
					meta-GGAs					
PKZB	−10.93	11.36	8.73	17.81	−10.93	20.17	−10.07	−7.06	−10.93	1.23
VS98	−14.96	8.52	2.10	14.62	−14.96	17.05	−12.73	−8.46	−14.96	−6.44
BLAP3	−11.24	8.51	3.28	14.65	−11.24	18.94	−13.92	−10.77	−11.24	0.08
OLAP3	−9.92	14.00	6.20	15.82	−9.92	24.53	−11.55	−8.08	−9.92	7.72
TPSS	−10.99	9.34	8.95	20.88	−10.99	17.58	−9.89	−7.01	−10.99	−3.22
M06-L	−14.02	12.92	8.99	25.92	−14.02	22.77	−4.73	−2.25	−14.02	2.63
					Hybrid Functionals					
B3LYP	−10.60	11.00	7.03	17.83	−10.60	21.22	−10.78	−8.53	−10.60	0.92
O3LYP	−9.63	14.78	10.26	19.98	−9.63	24.43	−8.70	−6.19	−9.63	6.39
KMLYP	−10.49	20.85	13.68	26.27	−10.49	32.97	−3.27	−1.21	−10.49	8.45
BHandH	−11.59	18.31	14.43	29.18	−11.59	29.33	−4.38	−1.25	−11.59	5.60
mPBE0KCIS	−10.94	13.10	8.92	21.03	−10.94	23.21	−9.44	−6.45	−10.94	4.26
mPW1K	−9.55	19.65	13.23	23.78	−9.55	31.18	−4.58	−2.41	−9.55	8.20
M05	−11.99	19.83	3.73	21.95	−11.99	23.34	−8.86	−3.70	−11.99	4.64
M05-2X	−8.93	12.58	16.60	18.97	−8.93	28.46	−7.75	−5.67	−8.93	6.84
M06	−12.68	17.33	6.17	22.96	−12.68	23.67	−6.49	−2.92	−12.68	3.36
M06-2X	−12.49	10.65	14.98	22.49	−12.49	30.29	−5.85	−4.74	−12.49	10.73

is treated. This TS energy computed again with our best basis set, aug-cc-pVQZ, varies from 18.30 to 4.93 to 8.91 to 5.54 kcal/mol along HF, MP2, CCSD, and CCSD(T), respectively. Note again that HF clearly overestimates the barrier by 9 kcal/mol (see Table 1). Moreover, the CCSD(T) values are converged as a function of the basis-set size (at aug-cc-pVQZ) to within less than half a kilocalorie per mole (see Table 1).

The *syn*-E2 elimination of $Cl^- + CH_3CH_2Cl$ proceeds via the same reactant complex as the *anti*-E2 elimination, which has been already examined above. The energy of the *syn*-E2 transition state computed at aug-cc-pV(T+*d*)Z is again sensitive to the level at which correlation is treated. It ranges from 44.99 to 31.06 to 35.11 to 30.82 along the series of *ab initio* methods (see Table 2). The CCSD(T) values change by less than 2 kcal/mol going from the aug-cc-pV(D+*d*)Z to the aug-cc-pV(T+*d*)Z basis set (see Table 2) and again do not differ much from the CBS energies.

The S_N2 transition states for reactions 1c and 2c are also found to be quite sensitive to the level at which correlation is treated. Thus, at the HF level, at which Coulomb correlation is not included, the energies of the transition states **1cTS** and **2cTS** computed with our best basis set (aug-cc-pVQZ for X = F and aug-cc-pV(T+*d*)Z for X = Cl) amount to 11.50 and 10.38 kcal/mol, respectively (see Table 1 and 2). Introducing Coulomb correlation into the theoretical treatment substantially lowers the barrier. Thus, along HF, MP2, CCSD, and CCSD(T), the energy of **1cTS** ranges from 11.50 to 3.81 to 4.97 to 1.99 kcal/mol and that of **2cTS** from 10.38 to 8.22 to 8.15 to 5.70 kcal/mol, respectively (see Table 1 and 2). Thus, HF significantly overestimates the overall barriers by some 10 and 5 kcal/mol, respectively. Note again how including the triple excitations in the CCSD calculations reduces the overall barrier by 3.0 and 2.4 kcal/mol, respectively. The three highest-level values are within a range of 3.0 and 2.5 kcal/mol for reactions 1c and 2c, respectively, Furthermore, the CCSD(T) values for **1cTS** are converged as a function of the basis-set size to within 0.4 kcal/mol and again do not differ much from the CBS extrapolated CCSD(T) values.

In conclusion, our best CCSD(T) estimate leads to a relative order in overall barriers (i.e., TS energy relative to reactants) of *anti*-E2 (X = F: −1.33 kcal/mol) < S_N2 (X = F: +1.99 kcal/mol) < *syn*-E2 (X = F: +5.54 kcal/mol) ∼ S_N2 (X = Cl: +5.70 kcal/mol) < *anti*-E2 (X = Cl: +17.92 kcal/mol) < *syn*-E2 (X = Cl: +30.82 kcal/mol). The change in preference from *anti*-E2 for X = F to S_N2 for X = Cl is also recovered in the trend of the *central* barriers. Our benchmark consolidates the G2+ values for the relative energies of **1aRC**, **1aTS**, **1bTS**, and **1cTS** on the PES of

Table 5. Errors in Overall and Central Barriers (in kcal/mol) for Various Density Functionals for the *anti*-E2, *syn*-E2, and S_N2 reactions of $X^- + CH_3CH_2X$ (X = F, Cl) Compared to CCSD(T)[a]

	anti-E2				*syn*-E2				S_N2			
	err. in barr. rel. to R		err. in barr. rel. to RC		err. in barr. rel. to R		err. in barr. rel. to RC		err. in barr. rel. to R		err. in barr. rel. to RC	
method	F	Cl	F	Cl	F	Cl	F	Cl	F	Cl	F	Cl
						LDA						
VWN	−7.21	−12.92	6.03	−10.70	−18.04	−19.37	−7.14	−17.15	−15.66	−10.34	−2.42	−8.12
						GGAs						
BP86	−6.94	−10.71	0.26	−11.15	−13.05	−15.47	−7.40	−15.91	−11.32	−7.62	−4.12	−8.06
BLYP	−10.22	−12.64	−2.98	−12.66	−14.20	−16.78	−8.30	−16.80	−13.26	−9.39	−6.02	−9.41
PW91	−8.25	−11.54	0.88	−10.41	−14.83	−16.60	−7.29	−15.47	−13.38	−8.94	−4.25	−7.81
PBE	−8.03	−11.07	0.77	−10.26	−14.52	−16.07	−7.27	−15.26	−12.72	−8.13	−3.92	−7.32
RPBE	−8.28	−10.14	−1.48	−10.04	−12.93	−14.55	−7.31	−14.45	−10.55	−6.37	−3.75	−6.27
revPBE	−7.50	−9.79	−1.12	−10.20	−12.39	−14.23	−7.25	−14.64	−9.90	−5.90	−3.52	−6.31
FT97	−5.21	−7.83	−0.42	−11.07	−10.34	−11.45	−7.38	−14.69	−9.18	−5.74	−4.39	−8.98
HCTH/93	−5.99	−7.67	−2.23	−9.40	−9.27	−11.26	−6.41	−12.99	−4.51	−1.88	−0.75	−3.61
HCTH/120	−8.52	−10.46	−1.61	−9.96	−12.71	−14.40	−6.91	−13.90	−9.36	−6.19	−2.45	−5.69
HCTH/147	−7.60	−9.78	−1.52	−9.99	−11.77	−13.67	−6.79	−13.88	−8.13	−5.20	−2.05	−5.41
HCTH/407	−9.23	−9.73	−2.70	−9.12	−12.14	−13.00	−6.52	−12.39	−7.77	−3.71	−1.24	−3.10
BOP	−8.65	−11.09	−3.97	−12.28	−11.55	−14.86	−7.99	−16.05	−9.65	−6.90	−4.97	−8.09
OPBE	−1.93	−3.93	1.76	−6.39	−7.61	−8.50	−4.94	−10.96	−2.40	1.86	1.29	−0.60
OLYP	−6.62	−7.24	−1.60	−8.68	−10.47	−11.24	−6.39	−12.68	−6.15	−1.66	−1.13	−3.10
						meta-GGAs						
PKZB	−5.23	−6.56	−1.06	−6.73	−9.47	−10.65	−6.04	−10.82	−9.26	−4.47	−5.09	−4.64
VS98	−12.09	−9.40	−6.28	−5.54	−12.58	−13.77	−7.45	−9.91	−16.05	−12.14	−10.24	−8.28
BLAP3	−7.25	−9.41	−3.70	−9.27	−8.01	−11.88	−5.12	−11.74	−6.87	−5.62	−3.32	−5.48
OLAP3	−3.29	−3.92	−2.05	−5.10	−3.97	−6.29	−2.96	−7.47	0.26	2.02	1.50	0.84
TPSS	−3.93	−8.58	2.46	−8.69	−9.70	−13.24	−4.54	−13.35	−12.02	−8.92	−5.63	−9.03
M06-L	0.10	−5.00	5.15	−2.08	−3.86	−8.05	0.35	−5.13	−4.94	−3.07	0.11	−0.15
						Hybrid Functionals						
B3LYP	−4.05	−6.92	0.26	−7.42	−7.54	−9.60	−4.16	−10.10	−6.00	−4.78	−1.69	−5.28
O3LYP	−1.22	−3.14	1.91	−4.61	−5.14	−6.39	−2.80	−7.86	−1.75	0.69	1.38	−0.78
KMLYP	7.42	2.93	8.57	2.32	2.74	2.15	3.39	1.54	5.55	2.75	6.70	2.14
BHandH	5.23	0.39	9.92	0.88	−1.68	−1.49	2.07	−1.00	0.77	−0.10	5.46	0.39
mPBE0KCIS	−3.11	−4.82	1.47	−4.98	−7.31	−7.61	−3.66	−7.77	−3.14	−1.44	1.44	−1.60
mPW1K	5.59	1.73	5.92	0.18	1.52	0.36	1.36	−1.19	4.25	2.50	4.58	0.95
M05	−2.18	1.91	1.51	2.71	−4.73	−7.48	−1.21	−6.68	−2.80	−1.06	0.89	−0.26
M05-2X	2.32	−5.34	1.86	−7.51	−1.69	−2.36	−2.51	−4.53	1.98	1.14	1.52	−1.03
M06	−0.88	−0.59	2.34	0.99	−4.40	−7.15	−1.56	−5.57	−2.34	−2.34	0.88	−0.76
M06-2X	2.82	−7.27	3.50	−5.88	−1.51	−0.53	−1.16	0.86	3.83	5.03	4.51	6.42

[a] Relative to CCSD(T)/aug-cc-pVQZ benchmark for reactions involving F and relative to CCSD(T)/aug-cc-pV(T + d)Z benchmark for reactions involving Cl. R = reactants, RC = reactant complex.

$F^- + CH_3CH_2F$ computed by Gronert and co-workers,[4h] which agree within 2.3 kcal/mol with our best CCSD(T) estimates.

3.3. Validation of DFT: Mean Absolute Error. Next, we examine the relative energies of stationary points computed with (i) the LDA functional VWN; (ii) the GGA functionals BP86, BLYP, PW91, PBE, RPBE, revPBE, FT97, HCTH/93, HCTH/120, HCTH/147, HCTH/407, BOP, OPBE, and OLYP; (iii) the meta-GGA functionals PKZB, VS98, BLAP3, OLAP3, TPSS, and M06-L; and (iv) the hybrid functionals B3LYP, O3LYP, KMLYP, BHandH, mPBE0KCIS, mPW1K, M05, M05-2X, M06, and M06-2X using the following procedure: (i) all functionals except OLYP are evaluated using the OLYP/ae-TZ2P density computed at the OLYP/TZ2P geometries; (ii) the OLYP functional is evaluated using the OLYP/TZ2P density computed at the OLYP/TZ2P geometries (see Methods section). Extensive previous validation studies have shown that the use of the all-electron ae-TZ2P versus the frozen-core TZ2P basis set leads to differences in relative energies of less than half a kilocalorie per mole.[38] The DFT relative energies for reactions 1 and 2 are collected in Tables 3 and 4, respectively (see Table S2 in the Supporting Information for an overview of *overall barriers together with central barriers* for all of the *anti*-E2, *syn*-E2, and S_N2 reactions of $F^- + CH_3CH_2F$ and $Cl^- + CH_3CH_2Cl$ computed with all our 31 functionals).

Here, we focus on the overall barrier, that is, the difference in energy between the TS and the separate reactants (R), and the central barrier, that is, the difference in energy between the TS and the reactant complex (RC). The overall barrier is decisive for the rate of chemical reactions in the gas phase, in particular, if they occur under low-pressure conditions,[2b,6] whereas the central barrier becomes decisive in the high-pressure regime, when termolecular collisions are sufficiently efficient to cool the otherwise rovibrationally hot reactant complex, causing it to be in thermal equilibrium with the environment.[2b,6]

The performance of the various density functional approaches is assessed by a systematic comparison of the resulting PESs with our CCSD(T)/aug-cc-pVQZ benchmark in the case of reaction 1 (Table 1) and the CCSD(T)/aug-

Table 6. Mean Absolute Errors (MAE) in Overall and Central Barriers (in kcal/mol) for Various Density Functionals for the *anti*-E2, *syn*-E2, and S_N2 reactions of $X^- + CH_3CH_2X$ (X = F, Cl) compared to CCSD(T)[a]

	MAE in *anti*-E2 barr.		MAE in *syn*-E2 barr.		MAE in S_N2 barr.		MAE in barr. X = F		MAE in barr. X = Cl		MAE	
method	rel. to R	rel. to RC	rel. to R	rel. to RC	rel. to R	rel. to RC	rel. to R	rel. to RC	rel. to R	rel. to RC	rel. to R	rel. to RC
						LDA						
VWN	10.07	8.37	18.71	12.15	13.00	5.27	13.64	5.20	14.21	11.99	13.92	8.59
						GGAs						
BP86	8.83	5.71	14.26	11.66	9.47	6.09	10.44	3.93	11.27	11.71	10.85	7.82
BLYP	11.43	7.82	15.49	12.55	11.33	7.72	12.56	5.77	12.94	12.96	12.75	9.36
PW91	9.90	5.65	15.72	11.38	11.16	6.03	12.15	4.14	12.36	11.23	12.26	7.69
PBE	9.55	5.52	15.30	11.27	10.43	5.62	11.76	3.99	11.76	10.95	11.76	7.47
RPBE	9.21	5.76	13.74	10.88	8.46	5.01	10.59	4.18	10.35	10.25	10.47	7.22
revPBE	8.65	5.66	13.31	10.95	7.90	4.92	9.93	3.96	9.97	10.38	9.95	7.17
FT97	6.52	5.75	10.90	11.04	7.46	6.69	8.24	4.06	8.34	11.58	8.29	7.82
HCTH/93	6.83	5.82	10.27	9.70	3.20	2.18	6.59	3.13	6.94	8.67	6.76	5.90
HCTH/120	9.49	5.79	13.56	10.41	7.78	4.07	10.20	3.66	10.35	9.85	10.27	6.75
HCTH/147	8.69	5.76	12.72	10.34	6.67	3.73	9.17	3.45	9.55	9.76	9.36	6.61
HCTH/407	9.48	5.91	12.57	9.46	5.74	2.17	9.71	3.49	8.81	8.20	9.26	5.85
BOP	9.87	8.13	13.21	12.02	8.28	6.53	9.95	5.64	10.95	12.14	10.45	8.89
OPBE	2.93	4.08	8.06	7.95	2.13	0.95	3.98	2.66	4.76	5.98	4.37	4.32
OLYP	6.93	5.14	10.86	9.54	3.91	2.12	7.75	3.04	6.71	8.15	7.23	5.60
						meta-GGAs						
PKZB	5.90	3.90	10.06	8.43	6.87	4.87	7.99	4.06	7.23	7.40	7.61	5.73
VS98	10.75	5.91	13.18	8.68	14.10	9.26	13.57	7.99	11.77	7.91	12.67	7.95
BLAP3	8.33	6.49	9.95	8.43	6.25	4.40	7.38	4.05	8.97	8.83	8.17	6.44
OLAP3	3.61	3.58	5.13	5.22	1.14	1.17	2.51	2.17	4.08	4.47	3.29	3.32
TPSS	6.26	5.58	11.47	8.95	10.47	7.33	8.55	4.21	10.25	10.36	9.40	7.28
M06-L	2.55	3.62	5.96	2.74	4.01	0.13	2.97	1.87	5.37	2.45	4.17	2.16
						Hybrid Functionals						
B3LYP	5.49	3.84	8.57	7.13	5.39	3.49	5.86	2.04	7.10	7.60	6.48	4.82
O3LYP	2.18	3.26	5.77	5.33	1.22	1.08	2.70	2.03	3.41	4.42	3.06	3.22
KMLYP	5.18	5.45	2.45	2.47	4.15	4.42	5.24	6.22	2.61	2.00	3.92	4.11
BHandH	2.81	5.40	1.59	1.54	0.44	2.93	2.56	5.82	0.66	0.76	1.61	3.29
mPBE0KCIS	3.97	3.23	7.46	5.72	2.29	1.52	4.52	2.19	4.62	4.78	4.57	3.49
mPW1K	3.66	3.05	0.94	1.28	3.38	2.77	3.79	3.95	1.53	0.77	2.66	2.36
M05	2.05	2.11	6.11	3.95	1.93	0.58	3.24	1.20	3.48	3.22	3.36	2.21
M05-2X	3.83	4.69	2.03	3.52	1.56	1.28	2.00	1.96	2.95	4.36	2.47	3.16
M06	0.74	1.67	5.78	3.57	2.34	0.82	2.54	1.59	3.36	2.44	2.95	2.02
M06-2X	5.05	4.69	1.02	1.01	4.43	5.47	2.72	3.06	4.28	4.39	3.50	3.72

[a] Relative to CCSD(T)/aug-cc-pVQZ benchmark for reactions involving F and relative to CCSD(T)/aug-cc-pV(T + d)Z benchmark for reactions involving Cl. R = reactants, RC = reactant complex.

cc-pV(T+*d*)Z benchmark in the case of reaction 2 (Table 2). Note that our best CCSD(T) results do not differ much from the CBS extrapolated CCSD(T) values. Thus, they were used (instead of the CBS values) as our benchmark since we prefer to have as little as possible empirical extrapolations in the benchmark reference values. For all 31 functionals, we have computed the errors in the overall and central barriers (see Table 5) and the corresponding mean absolute errors (MAE) relative to the CCSD(T) benchmarks for all model reactions together as well as for certain categories thereof (see Table 6).

It is clear from Tables 5 and 6 that LDA suffers from its notorious overbinding: it yields too-low barriers and too-exothermic complexation and reaction energies (see also Tables 3 and 4). But also many of the GGA (e.g., BLYP, BOP, BP86, PW91, and PBE) and some meta-GGA functionals (VS98 and TPSS) perform more or less equally poorly as LDA: together, these poorly performing functionals have MAE values, for all reactions together, in the range 7−9 kcal/mol for central and 9−14 kcal/mol for overall barriers (see Table 6).

Best overall agreement with our *ab initio* benchmark barriers is obtained by representatives from each of the three categories of functionals, GGA, meta-GGA, and hybrid DFT, with MAEs in central barriers of 4.3 (OPBE), 2.2 (M06-L), and 2.0 kcal/mol (M06), respectively, and MAEs in overall barriers of 4.4 (OPBE), 3.3 (OLAP3), and 1.6 kcal/mol (BHandH), respectively (see Table 6). The top three best functionals is constituted for the central barriers of M06, M06-L, and M05 with MAE values, for all reactions together, of 2.0, 2.2, and 2.2 kcal/mol, respectively, and for the overall barriers of BHandH, M05-2X, and mPW1K with MAE values, for all reactions together, of 1.6, 2.5, and 2.7 kcal/mol, respectively (see Table 6). An important point to note is that the OPBE functional is, not only for all reactions together but also for each individual category of reactions (e.g., *anti*-E2 reactions or reactions with X = F, etc.), in the top regions of performance (MAE in a category typically 1−6 kcal/mol, only for *syn*-E2 it reaches 8.1 kcal/mol) of all functionals studied, and it is the best of all *GGA* functionals. OLYP (7.2 and 5.6 kcal/mol relative to R and RC) and B3LYP (6.5 and 4.8 kcal/mol relative to R and RC) are of comparable quality, and both have somewhat larger MAE values for all reactions together than OPBE (4.4 and 4.3 kcal/mol relative to R and RC; see Table 6). OLYP (MAE for S_N2: 3.9 and 2.1 kcal/mol relative to R and RC)

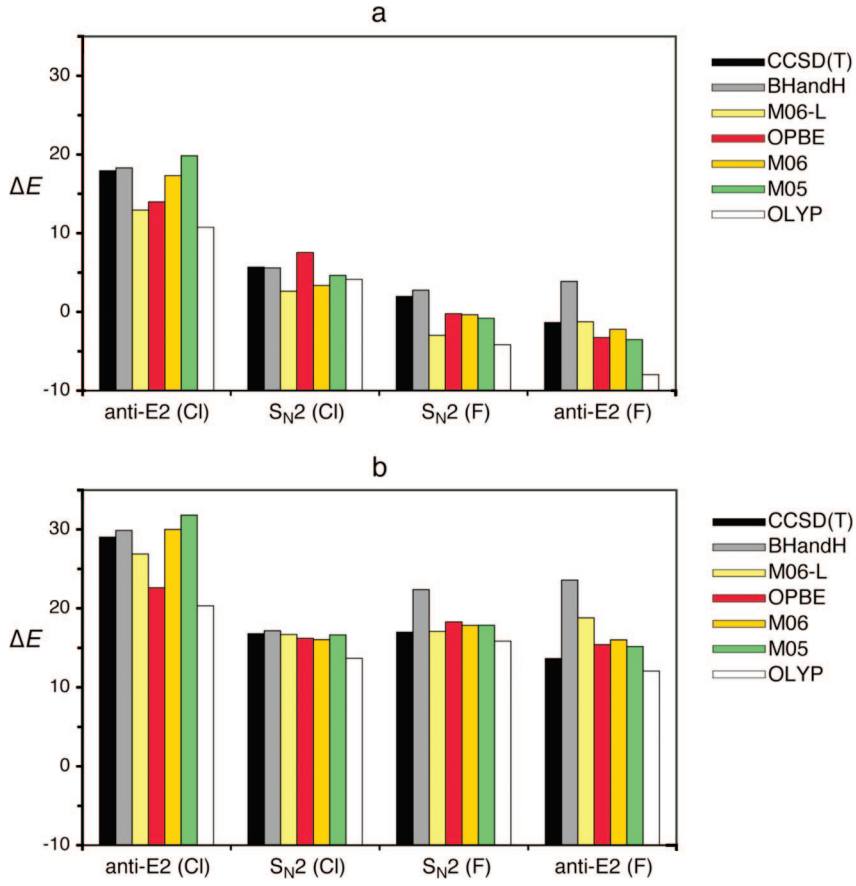


Figure 3. Overall (a) and central (b) barriers (in kcal/mol) for the *anti*-E2 and S_N2 reactions of $X^- + CH_3CH_2X$ (X = F, Cl), computed with CCSD(T) and selected density functionals.

is however slightly better than B3LYP (MAE for S_N2: 5.4 and 3.5 kcal/mol relative to R and RC) for the category of S_N2 reactions (see Table 6), in agreement with previous work.[7a]

Finally, complexation energies of the reactant complexes relative to reactants as well as reaction energies of our model reactions appear to be, in general, somewhat less dependent on the level of both ab initio (see Tables 1 and 2) and density functional theory (see Tables 3–6) if compared with the relative energies of the transition states discussed above. The density functionals that perform best for reaction barriers in terms of MAE, namely, BHandH, M06, M06-L, M05, M05-2X, and mPW1K, also show satisfactory agreement with the CCSD(T) benchmark regarding these complexation and reaction energies, with MAEs in the range of 0.7−4.8 kcal/mol (values not shown in Table 6). OPBE and B3LYP also achieve MAE values within this range, whereas OLYP has MAE values of 3.5 and 6.0 kcal/mol for complexation and reaction energies, respectively.

3.4. Validation of DFT: Trends. So far, we have concentrated on the MAE, which leads to a certain ranking of density functionals regarding their performance in computing overall or central barriers for the six model reaction pathways (see Scheme 2). Interestingly (and importantly), such an MAE-based ranking does not necessarily say something about the performance for reproducing the right trends in reactivity.

For example, according to the MAE criterion, BHandH and M06-L belong to the best functionals. Yet, they erroneously predict that, for $F^- + CH_3CH_2F$, the *anti*-E2 reaction has both a higher overall and central barrier than the S_N2 reaction, as can be seen in Figure 3a and b, respectively (see also Tables 3 and 4). For comparison, both OPBE and OLYP do reproduce the correct trend (see Figure 3), in spite of the fact that the MAE is larger than that for BHandH or M06-L (see Table 6). In the latter two functionals, the error is apparently somewhat less uniformly distributed. This is an interesting phenomenon, but it should also not be overrated because the energy differences concerned are rather small.

M06 and M05 are good both in terms of one of the smallest MAE values (see Table 6) and a correct trend in reactivity (see Figure 3 and Tables 3 and 4). On the other hand, they are computationally somewhat more expensive than OPBE and OLYP. And, at variance with the latter, M06 and M05 are (in ADF) evaluated post-SCF with the density of another potential (e.g., OPBE or OLYP).

4. Conclusions

We have computed *ab initio* benchmarks for the archetypal competing E2 and S_N2 reactions of fluoride + fluoroethane and chloride + chloroethane. These benchmarks derive from hierarchical series of methods up to CCSD(T)/aug-cc-pVQZ (up to CCSD(T)/aug-cc-pV(T+*d*)Z for chloride + chloro-

ethane), which are converged with respect to the basis-set size within less than half a kilocalorie per mole. The resulting reaction profiles show that *anti*-E2 dominates for $F^- + CH_3CH_2F$ while S_N2 dominates for $Cl^- + CH_3CH_2Cl$. This change in preference is reflected by both overall and central barriers. On the other hand, *syn*-E2 is in both reaction systems the least favorable pathway.

Our *ab initio* benchmark is used to evaluate the performance of 31 density functionals for describing the above *anti*-E2, *syn*-E2, and S_N2 reactions. The best overall agreement regarding central reaction barriers with our *ab initio* benchmark is obtained by representatives from each of the three categories of functionals, GGA, meta-GGA, and hybrid DFT, with mean absolute errors of 4.3 (OPBE), 2.2 (M06-L), and 2.0 kcal/mol (M06), respectively.

Importantly, the hybrid functional BHandH and the meta-GGA M06-L yield incorrect trends and qualitative features of the PESs (in particular, an erroneous preference for S_N2 over the *anti*-E2 in the case of $F^- + CH_3CH_2F$) even though they are among the best functionals as measured by their small mean absolute errors of 3.3 and 2.2 kcal/mol in reaction barriers. OLYP and B3LYP have somewhat higher mean absolute errors in central barriers (5.6 and 4.8 kcal/mol, respectively), but the error distribution is somewhat more uniform, and as a consequence, the correct trends are reproduced.

Acknowledgment. We thank The Netherlands Organization for Scientific Research (NWO-CW), the European HPC-Europa program, the Spanish Ministerio de Educación y Cultura (MEC), and the Catalan DURSI (Generalitat de Catalunya) for financial support. Excellent service by the Stichting Academisch Rekencentrum Amsterdam (SARA) and the Centre de Supercomputació de Catalunya (CESCA) is gratefully acknowledged.

Supporting Information Available: Cartesian coordinates of all stationary points. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Smith, M. B.; March, J. *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*; Wiley: New York, 2007. (b) Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry, Part A*; Plenum Press: New York, 1984. (c) Ingold, C. K. *Structure and Mechanism in Organic Chemistry*; Cornell University Press: Ithaca, NY, 1969. (d) Lowry, T. H.; Richardson, K. S. *Mechanism and Theory in Organic Chemistry*, 3rd ed.; Harper and Row: New York, 1987.

(2) (a) Gronert, S. *Chem. Rev.* **2001**, *101*, 329. (b) Bickelhaupt, F. M. *Mass Spectrom. Rev.* **2001**, *20*, 347. (c) Gronert, S.; Pratt, L. M.; Mogali, S. *J. Am. Chem. Soc.* **2001**, *123*, 3081. (d) Gronert, S. *Acc. Chem. Res.* **2003**, *36*, 848. (e) Bickelhaupt, F. M.; de Koning, L. J.; Nibbering, N. M. M. *J. Org. Chem.* **1993**, *58*, 2436. (f) Uggerud, E.; Bache-Andreassen, L. *Chem.—Eur. J.* **1999**, *5*, 1917. (g) Gronert, S.; Fagin, A. E.; Okamoto, K.; Mogali, S.; Pratt, L. M. *J. Am. Chem. Soc.* **2004**, *126*, 12977. (h) Villano, S. M.; Kato, S.; Bierbaum, V. M. *J. Am. Chem. Soc.* **2006**, *128*, 736. (i) Bickelhaupt, F. M.; Buisman, G. J. H.; de Koning, L. J.; Nibbering, N. M. M.; Baerends, E. J. *J. Am. Chem. Soc.* **1995**, *117*, 9889. (j) Flores, A. E.; Gronert, S. *J. Am. Chem. Soc.* **1999**, *121*, 2627. (k) DePuy, C. H.; Gronert, S.; Mullin, A.; Bierbaum, V. M. *J. Am. Chem. Soc.* **1990**, *112*, 8650.

(3) (a) Orita, A.; Otera, J. *Chem. Rev.* **2006**, *106*, 5387. (b) Alunni, S.; De Angelis, F.; Ottavi, L.; Papavasileiou, M.; Tarantelli, F. *J. Am. Chem. Soc.* **2005**, *127*, 15151. (c) Meng, Q.; Thibblin, A. *J. Am. Chem. Soc.* **1995**, *117*, 9399. (d) Pirinccioglu, N.; Thibblin, A. *J. Am. Chem. Soc.* **1998**, *120*, 6512.

(4) (a) Bickelhaupt, F. M. *J. Comput. Chem.* **1999**, *20*, 114. (b) Ensing, B.; Laio, A.; Gervasio, F. L.; Parrinello, M.; Klein, M. L. *J. Am. Chem. Soc.* **2004**, *126*, 9492. (c) De Angelis, F.; Tarantelli, F.; Alunni, S. *J. Phys. Chem. B* **2006**, *110*, 11014. (d) Almerindo, G. I.; Pliego, R., Jr *Org. Lett.* **2005**, *7*, 1821. (e) Ensing, B.; Klein, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6755. (f) Chung, D. S.; Kim, C. K.; Lee, B.-S.; Lee, I. *J. Phys. Chem. A* **1997**, *101*, 9097. (g) Merrill, G. N.; Gronert, S.; Kass, S. R. *J. Phys. Chem. A* **1997**, *101*, 208. (h) Gronert, S. *J. Org. Chem.* **1995**, *60*, 488. (i) Minato, T.; Yamabe, S. *J. Am. Chem. Soc.* **1985**, *107*, 4621. (j) Minato, T.; Yamabe, S. *J. Am. Chem. Soc.* **1988**, *110*, 4586. (k) Gronert, S. *J. Am. Chem. Soc.* **1993**, *115*, 652. (l) Gronert, S. *J. Am. Chem. Soc.* **1991**, *113*, 6041.

(5) (a) Bach, R. D.; Glukhovtsev, M. N.; Gonzales, C. *J. Am. Chem. Soc.* **1998**, *120*, 9902. (b) Baker, J.; Muir, M.; Andzelm, J. *J. Chem. Phys.* **1995**, *102*, 2063. (c) Baker, J.; Pulay, P. *J. Chem. Phys.* **2002**, *117*, 1441. (d) Barone, V.; Adamo, C. *J. Chem. Phys.* **1996**, *105*, 11007. (e) Gritsenko, O. V.; Ensing, B.; Schippers, P. R. T.; Baerends, E. J. *J. Phys. Chem. A* **2000**, *104*, 8558. (f) Poater, J.; Solà, M.; Duran, M.; Robles, J. *Phys. Chem. Chem. Phys.* **2002**, *4*, 722. (g) Thümmel, H. T.; Bauschlicher, C. W. *J. Phys. Chem. A* **1997**, *101*, 1188. (h) Diefenbach, A.; Bickelhaupt, F. M. *J. Chem. Phys.* **2001**, *115*, 4030.

(6) (a) Nibbering, N. M. M. *Adv. Phys. Org. Chem.* **1988**, *24*, 1. (b) Nibbering, N. M. M. *Acc. Chem. Res.* **1990**, *23*, 279.

(7) (a) Bento, A. P.; Solà, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2005**, *26*, 1497. (b) Swart, M.; Solà, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2007**, *28*, 1551. (c) Swart, M.; Ehlers, A. W.; Lammertsma, K. *Mol. Phys.* **2004**, *102*, 2467. (d) Xu, X.; Goddard, W. A., III *J. Phys.Chem. A* **2004**, *108*, 8495. (e) Gonzales, J. M.; Allen, W. D.; Schaefer, H. F., III *J. Phys. Chem. A* **2005**, *109*, 10613. (f) Gonzales, J. M.; Pak, C.; Cox, R. S.; Allen, W. D.; Schaefer, H. F., III; Császár, A. G.; Tarczay, G. *Chem.—Eur. J.* **2003**, *9*, 2173. (g) Guner, V. A.; Khuong, K. S.; Houk, K. N.; Chuma, A.; Pulay, P. *J. Phys. Chem. A* **2004**, *108*, 2959. (h) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(8) (a) Baerends, E. J.; Autschbach, J.; Bérces, A.; Berger, J. A.; Bickelhaupt, F. M.; Bo, C.; de Boeij, P. L.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; van Faassen, M.; Fan, T.; Fischer, T. H.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Harris, F. E.; van den Hoek, P.; Jacob, C. R.; Jacobsen, H.; Jensen, L.; Kadantsev, E. S.; van Kessel, G.; Klooster, R.; Kootstra, F.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Neugebauer, J.; Nicu, V. P.; Osinga, V. P.; Patchkovskii, S.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Romaniello, P.; Ros, P.; Schipper, P. R. T.; Schreckenbach, G.; Snijders, J.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E. M.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; Yakovlev, A. L.; Ziegler, T. *ADF200701*; SCM: Amsterdam, The Netherlands. (b) te Velde, G.; Bickelhaupt,

F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931. (c) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391.

(9) (a) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(10) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41.

(11) Fan, L.; Ziegler, T. *J. Chem. Phys.* **1990**, *92*, 3645.

(12) Fan, L.; Versluis, L.; Ziegler, T.; Baerends, E. J.; Ravenek, W. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1988**, *S22*, 173.

(13) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(14) (a) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098. (b) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822.

(15) (a) Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991. (b) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *46*, 6671. (Erratum: *Ibid.* **1993**, *48*, 4978).

(16) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865. (Erratum: *Ibid.* **1997**, *78*, 1396)

(17) Hammer, B.; Hansen, L. B.; Nørskov, J. K. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 7413.

(18) Zhang, Y.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 890.

(19) Filatov, M.; Thiel, W. *Mol. Phys.* **1997**, *91*, 847.

(20) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264.

(21) Boese, A. D.; Doltsinis, N. L.; Handy, N. C.; Sprik, M. *J. Chem. Phys.* **2000**, *112*, 1670.

(22) Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2001**, *114*, 5497.

(23) Tsuneda, T.; Suzumura, T.; Hirao, K. *J. Chem. Phys.* **1999**, *110*, 10664.

(24) (a) Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. *Phys. Rev. Lett.* **1999**, *82*, 2544. (Erratum: *Ibid.* **1999**, *82*, 5179)

(25) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.

(26) Proynov, E. I.; Sirois, S.; Salahub, D. R. *Int. J. Quantum Chem.* **1997**, *64*, 427.

(27) (a) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401. (b) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129. (Erratum: *Ibid.* **2004**, *121*, 11507)

(28) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(29) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(30) Cohen, A. J.; Handy, N. C. *Mol. Phys.* **2001**, *99*, 607.

(31) Kang, J. K.; Musgrave, C. B. *J. Chem. Phys.* **2001**, *115*, 11040.

(32) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(33) Toulouse, J.; Adamo, C. *Chem. Phys. Lett.* **2002**, *362*, 72.

(34) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.

(35) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

(36) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput* **2006**, *2*, 364.

(37) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215. (Erratum: *Ibid.* **2008**, *119*, 525).

(38) (a) Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. *J. Phys. Chem. A* **2004**, *108*, 5479. (b) de Jong, G. Th.; Bickelhaupt, F. M. *J. Chem. Theory Comput.* **2006**, *2*, 322. (c) de Jong, G. Th.; Bickelhaupt, F. M. *J. Phys. Chem. A* **2005**, *109*, 9685. (d) de Jong, G. Th.; Geerke, D. P.; Diefenbach, A.; Solà, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2005**, *26*, 1006. (e) de Jong, G. Th.; Geerke, D. P.; Diefenbach, A.; Bickelhaupt, F. M. *Chem. Phys.* **2005**, *313*, 261.

(39) 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.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gromperts, 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, 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.; Cheng, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian Inc.: Pittsburgh, PA, 2003.

(40) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(41) (a) Krishnan, R.; Pople, J. A. *Int. J. Quantum Chem.* **1978**, *14*, 91. (b) Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 4244.

(42) Cizek, J. *J. Chem. Phys.* **1966**, *45*, 4256.

(43) Purvis, G. D., III; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.

(44) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

(45) (a) Dunning, T. H., Jr *J. Chem. Phys.* **1989**, *90*, 1007. (b) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796. (c) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244. (d) The aug-cc-pV(n + d)Z basis set was obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 02/02/06, 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, WA 99352, and funded by the U.S. Department of Energy.

(46) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.

CT700318E

JCTC Journal of Chemical Theory and Computation

Mechanisms for the Reactions of Hydroxyl Radicals with Acrolein: A Theoretical Study

Santiago Olivella*,† and Albert Solé‡

Institut d'Investigacions Químiques i Ambientals de Barcelona, CSIC, Jordi Girona 18-26, 08034-Barcelona, Catalonia, Spain, and Departament de Química Física i Institut de Recerca en Química Teòrica i Computacional, Universitat de Barcelona, Martí i Franquès 1, 08028-Barcelona, Catalonia, Spain

Received March 6, 2008

Abstract: Three low-energy pathways for the reaction of HO• with acrolein, a key reaction in atmospheric environments, have been investigated by means of quantum-mechanical electronic structure methods (UQCISD and RQCISD(T)). The first step of all the reaction pathways studied involves the barrierless formation of a prereaction loosely bound complex in the entrance channel, lying a few kcal/mol below the energy of the reactants. The lowest-energy barrier pathway at 0 K is found to be the HO• abstraction of the aldehydic H-atom through a transition-state structure lying 1.1 kcal/mol below the energy of the reactants. The addition of HO• to the terminal carbon atom of the C═C double bond proceeds via a transition-state structure lying 0.7 kcal/mol below the energy of reactants at 0 K, whereas the HO• addition to the central carbon atom takes place via a transition-state structure lying 0.8 kcal/mol above the energy of the reactants at 0 K. On the basis of conventional transition-state theory calculations at 298 K, it is predicted that 74.5% of the HO• reaction with acrolein proceeds via abstraction of the aldehydic H-atom, 24.2% via HO• addition to the terminal carbon atom of the double bond, and 1.3% via HO• addition to the central carbon atom of the double bond. These results are in close agreement with available experimental data.

1. Introduction

Acrolein (CH_2═CH-CHO) is an unsaturated aldehyde present in the atmosphere as the result of direct anthropogenic emissions (e.g., from combustion sources) or as the result of the atmospheric hydroxyl radical (HO•)- and ozone (O_3)-initiated oxidation of 1,3-dienes.[1–4] In several large cities, ambient concentrations up to 9 ppb of acrolein have been reported.[5] Similarly to other unsaturated carbonyl compounds, the main degradation process of acrolein under atmospheric conditions is via reaction with HO•, with photolysis and reaction with O_3 and nitrate radical (NO_3•) playing at most a minor role.

The reported rate constants for HO• reaction with acrolein[6–10] are in the range $(1.83-2.66) \times 10^{-11}$ cm^3 molecule^{-1} s^{-1}, and only three product studies on this reaction have been published so far.[10–12] Magneron et al.[10] published data which suggested that ≈20−25% of the reaction of HO• with acrolein occurs via addition to the C═C double bond. In a subsequent work, Orlando and Tyndall[12] determined that about 68% of the HO• reaction with acrolein proceeds via abstraction of the aldehydic H-atom (i.e., reaction of eq 1), with the remainder occurring via addition to the C═C double bond (i.e., reactions of eqs 2 and 3). These results are in moderate

$$CH_2=CHCHO + HO\cdot \rightarrow CH_2=CHC\cdot O + H_2O \quad (1)$$

$$CH_2=CHCHO + HO\cdot \rightarrow HOCH_2C\cdot HCHO \quad (2)$$

$$CH_2=CHCHO + HO\cdot \rightarrow \cdot CH_2CH(OH)CHO \quad (3)$$

agreement with those of the study by Magneron et al.[10] An additional unresolved question concerns the actual site of the HO• addition (i.e., on the terminal carbon atom (reaction

* Corresponding author e-mail: sonqtc@iiqab.csic.es.
† Institut d'Investigacions Químiques i Ambientals de Barcelona.
‡ Universitat de Barcelona.

10.1021/ct8000798 CCC: $40.75
Published on Web 05/21/2008

of eq 2) or on the central carbon atom (reaction of eq 3)). On the basis of their product study, Orlando and Tyndall[12] estimated that at least 80% of the HO• addition takes place on the terminal carbon atom.

We feel that the results of the experimental studies of Magneron et al.[10] and Orlando and Tyndall[12] merit a theoretical study to analyze the subtle balance between the different pathways in the HO• reaction with acrolein. With this aim, herein we report the results of high level quantum-mechanical electronic structure calculations on the low-energy reaction pathways of eqs 1–3[13] The energetic, structural, and vibrational results furnished by these calculations are subsequently used to perform conventional transition-state computations to predict the rate coefficients and the branching ratios of the competing addition and abstraction reactions.

2. Computational Details

2.1. Electronic Structure Calculations. The geometries of the relevant stationary points (minima and first-order saddle points) on the lowest-energy potential energy surface (PES) of each reaction system were optimized by using the spin-unrestricted quadratic configuration-interaction method with all single and double excitations,[15] denoted as UQCISD, with core−electrons excluded (frozen core approximation), employing Dunning's augmented correlation-consistent polarized valence double-ζ (aug-cc-pVDZ) basis set.[16] The harmonic vibrational frequencies of these stationary points were computed at the latter level of theory. Connections of the transition-state structures between designated minima were confirmed in each case by intrinsic reaction coordinate (IRC)[17] calculations using the second-order algorithm of Gonzalez and Schlegel.[18]

Since energy barriers affect the calculated rate coefficients exponentially, it is crucial to compute accurately the energies of the transition-state structures relative to those of the reactants. A special difficulty is encountered in the case of the transition-state structures located for the competing reactions of eqs 1–3 because we found a significant difference in the degree of "spin contamination" shown by the spin-unrestricted Hartree−Fock (UHF) wave function underlying the UQCISD calculations. In fact, the expected values of the spin-squared operator S^2 for the UHF/aug-cc-pVDZ wave function of the transition-state structures calculated for these reactions were found to be 0.7654, 1.0952, and 1.0815, respectively. Therefore, all the energies were refined by performing single point energy calculations on the UQCISD geometries using the (frozen core) UQCISD method with a perturbative estimation of all connected triple excitations,[15] denoted as UQCISD(T). Finally, energies were also evaluated from partially spin-adapted QCISD(T) calculations based on a restricted open-shell Hartree−Fock reference determinant,[19] denoted as RQCISD(T), to accomplish the spin contamination in spin-unrestricted quadratic configuration-interaction wave functions.[20] The related spin-restricted coupled-cluster method[21] including all single and double excitations with a perturbative estimation of all connected triple excitations,[22] denoted as RCCSD(T), has been shown[23] to achieve "chemical accuracy" even in situations where spin contamination would normally be a problem. However, Senosiain et al.[14,24] have recently concluded that the RQCISD(T) method performs slightly better than RCCSD(T) in the calculation of a series of well-known adiabatic energy barriers.[25] Both the UQCISD(T) and RQCISD(T) calculations were carried out with Dunning's augmented correlation-consistent polarized valence triple-ζ (aug-cc-pVTZ) basis set.[16]

Zero-point vibrational energies (ZPVEs) were determined from unscaled harmonic vibrational frequencies. Thermal corrections to enthalpy and Gibbs energy values were obtained assuming ideal gas behavior from the unscaled harmonic frequencies and moments of inertia by conventional methods.[26] A standard pressure of 1 atm was taken in the absolute entropies calculations.

For the hydrogen-bonded complexes found in this work, the basis set superposition error (BSSE) was calculated at the UQCISD(T)/aug-cc-pVTZ level by using the counterpoise method of Boys and Bernardi.[27]

To examine the characteristics of the bonding and interactions in the most relevant structures we have also performed an analysis of the electron density within the framework of the topological theory of atom in molecules (AIM).[28] The net atomic charges were examined by means of the Mulliken population analysis.[29] The Z density matrix obtained from UQCISD gradient calculations with the aug-cc-pVDZ basis set, an effective correlated density matrix,[30] was used in both analyses.

All the UQCISD and UQCISD(T) calculations were carried out by using the Gaussian 03 program package,[31] whereas the MOLPRO 98 program package[32] was employed for the RQCISD(T) computations. The PROAIM and EXTREME programs of Bader et al.[33] were used to perform the AIM analysis of the electronic density.

2.2. Rate Coefficient Calculations. With the main purpose of evaluating the overall rate coefficient of the HO• reaction with acrolein, conventional transition-state theory calculations were carried out for reactions of eqs 1–3. As it will be shown in section 4, all these reaction pathways consist of a reversible first step involving the barrierless formation of a prereaction loosely bound complex in the entrance channel, followed by the irreversible formation of products. Therefore, each reaction pathway is a two-step process as described by eq 4, where the corresponding complex is in equilibrium with the reactants

$$CH_2=CHCHO^{\bullet} + HO \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{Complex} \xrightarrow{k_2} \text{Products} \qquad (4)$$

If k_1 and k_{-1} are the rate constants for the first step and k_2 is the rate constant for the second step, then a steady-state analysis leads[34] to an overall rate constant for the reaction pathway under consideration, denoted as k_{RP}, which can be approximated as

$$k_{RP} = \frac{k_1}{k_{-1}} k_2 = K_{eq} k_2 \qquad (5)$$

where K_{eq} stands for the equilibrium constant in the first step, which can be written as

$$K_{eq} = \frac{Q_{CX}}{Q_{CH2=CHCHO} Q_{HO}} e^{\frac{-(E_C - E_R)}{RT}} \quad (6)$$

where the various Qs are the partition functions of the reactants ($Q_{CH2=CHCHO}$ and Q_{HO}) and prereaction complex (Q_{CX}); E_R and E_{CX} are the total electronic energy plus the ZPVE of the reactants and prereaction complex, respectively; R is the ideal gas constant; and T is the absolute temperature. The rate constant k_2 can be evaluated using the conventional transition-state theory equation[35]

$$k_2 = \Gamma \frac{k_b T}{h} \frac{Q_{TS}}{Q_{CX}} e^{\frac{-(E_{TS} - E_{CX})}{RT}} \quad (7)$$

where Q_{TS} and E_{TS} are the partition function and the total electronic energy plus ZPVE, respectively, of the transition state, and Γ is the tunneling factor.

According to the standard formulas,[26] the Qs were evaluated using the UQCISD/aug-cc-pVDZ geometries and harmonic vibrational frequencies, while the Es were taken as the ZPVE-corrected QCISD(T)/aug-cc-pVTZ energies. The Γs were evaluated by zero-order approximation to the vibrationally adiabatic PES model with zero curvature.[36] In this approximation the tunneling is assumed to occur along a unidimensional minimum energy path. The potential energy curve is approximated by an unsymmetrical Eckart potential energy barrier[37] that is required to go through the ZPVE corrected energy (denoted as E) of the reactants, transition state, and products. The equations that describe the Eckart potential energy function were adapted from Truong and Truhlar.[36] Solving the Schroedinger equation for the Eckart function yields the transmission probability, $\kappa(E)$. Then Γ is obtained by integrating the respective $\kappa(E)$ over all possible energies:

$$\Gamma(T) = \frac{1}{k_b T} e^{\frac{E_{TS} - E_{CX}}{RT}} \int_0^\infty e^{\frac{-E_{TS}}{RT}} \kappa(E) dE \quad (8)$$

3. Preliminary Test for the Aldehydic H-Atom Abstraction from Acetaldehyde by an HO Radical

To confirm the reliability of the theoretical methods described above in predicting the rate coefficient for H-atom abstraction from aldehydes by hydroxyl radicals, we tested these methods on the reaction of acetaldehyde (CH_3CHO) with HO$^\bullet$ (eq 9).

$$CH_3CHO + HO^\bullet \rightarrow CH_3C^\bullet O + H_2O \quad (9)$$

Previous theoretical work on the reaction of eq 9 by Vivier-Bunge et al.,[38] using geometries optimized at the UMP2 level with the 6−311++G(d,p) basis set[39] and single point energy calculations at the UCCSD(T) level employing the same basis set, has shown that the reaction is not elemental. It consists of a reversible first step involving the formation of a prereaction hydrogen-bonded complex, followed by the irreversible formation of a loosely bound complex between the products $CH_3C^\bullet O$ and H_2O. Selected geometrical parameters of the prereaction hydrogen-bonded complex (labeled as **CXR**), transition-state structure (labeled as **TS**), and the loosely bound complex between the products

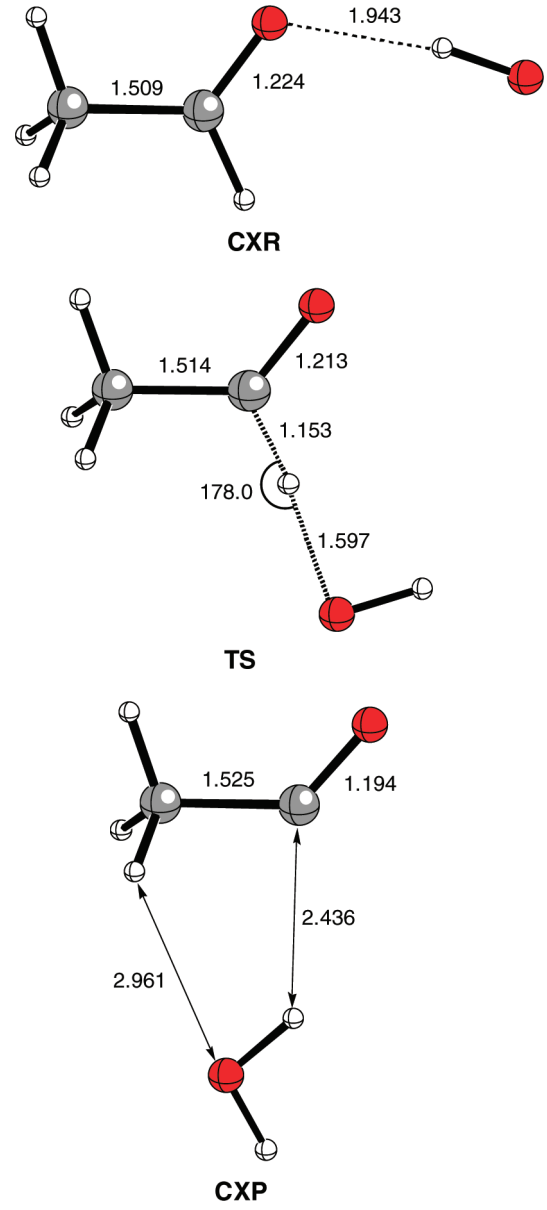


Figure 1. Selected geometrical parameters of the most relevant structures of the reaction pathway for the aldehydic H-atom abstraction from acetaldehyde by HO$^\bullet$. Distances are given in Å and angles in deg.

(labeled as **CXP**) calculated at the UQCISD/aug-cc-pVDZ level for the reaction of eq 9 are shown in Figure 1. The Cartesian coordinates of all structures reported in this article are available as Supporting Information. Total electronic energies computed at the different levels of theory as well as the ZPVEs, thermal corrections to enthalpy and Gibbs energy, for the structures concerning the reaction of eq 9 are collected in Table S1 (Supporting Information). Table 1 gives the relative electronic energies (designated by ΔU) as well as the relative energies at 0 K (designated by ΔE(0 K)) and the relative enthalpies (designated by ΔH(298 K)) and Gibbs energies (designated by ΔG(298 K)) at 298 K, for these structures calculated at both the UQCISD(T) and RQCISD(T) levels of theory with the aug-cc-pVTZ basis set. The equilibrium constant (K_{eq}) of the first step, the tunneling factor (Γ) and the rate coefficient (k_2) of the second

Table 1. Relative Energies (kcal/mol) of the Most Relevant Stationary Points on the Ground-State Potential Energy Surface for the Aldehydic H-Atom Abstraction from Acetaldehyde by HO•[a]

stationary point[b]	ΔU	ΔE (0 K)	ΔH (298 K)	ΔG (298 K)
$CH_3CHO + HO^•$	0.0	0.0	0.0	0.0
CXR	−6.1 (−6.1)	−4.0 (−4.0)	−4.7 (−4.7)	2.1 (2.1)
TS	−1.2 (−1.7)	−1.1 (−1.6)	−1.7 (−2.2)	5.7 (5.2)
CXP	−32.7 (−32.9)	−31.0 (−31.2)	−30.6 (−30.8)	−25.5 (−25.7)
$CH_3CO^• + H_2O$	−29.3 (−29.2)	−29.0 (−28.9)	−28.6 (−28.5)	−29.8 (−29.7)

[a] Calculated at the RQCISD(T) level of theory with the aug-cc-pVTZ basis set. The values calculated at the UQCISD(T) level with the same basis set are given in parentheses. [b] See Figure 1.

Table 2. Equilibrium Constant (K_{eq} in molecule^{-1} cm^3; See Eq *6*) of the First Step, Tunneling Factor (Γ) and Rate Coefficient (k_2 in molecule^{-1} cm^3 s^{-1}; See Eq *7*) of the Second Step, and Overall Rate Coefficient (k_{RP} in molecule^{-1} cm^3 s^{-1}; See Eq *5*) at 298 K of the Reaction Pathway for the Aldehydic H-Atom Abstraction from Acetaldehyde by HO•

method	K_{eq}	Γ	k_2	k_{RP}
UQCISD(T)	1.1894×10^{-21}	1.0300	3.2768×10^{10}	3.8974×10^{-11}
RQCISD(T)	1.1894×10^{-21}	1.0320	1.4119×10^{10}	1.6793×10^{-11}
exp[a]				1.6×10^{-11}

[a] Reference 38.

step, and the overall rate coefficient (k_{RP}) at 298 K for the reaction of eq 9 are summarized in Table 2.

The geometries calculated for **CXR**, **TS**, and **CXP** (Figure 1) compare well with those computed at the UMP2 level with the 6−311++G(d,p) basis set by Vivier-Bunge et al.[38] However, on the basis of the bond lengths of the breaking C−H and forming H−O bonds in **TS**, it turns out that the present UQCISD/aug-cc-pVDZ calculations predict a transition-state structure which is somewhat more reactant-like than that calculated at the UMP2/6−311++G(d,p) level. The activation energy at 0 K (designated by $\Delta E^‡$(0 K)) of −1.6 kcal/mol calculated at the UQCISD(T)/aug-cc-pVTZ+ZPVE level (see Table 1) is in good agreement with the $\Delta E^‡$(0 K) of −1.71 kcal/mol obtained by Vivier-Bunge et al.[38] from UCCSD(T)/6−311++G(d,p)+ZPVE calculations, whereas the $\Delta E^‡$(0 K) of −1.1 kcal/mol computed at the RQCISD(T)/aug-cc-pVTZ+ZPVE level is 0.6 kcal/mol higher than the latter value. Nevertheless, the energy barrier (designated by $\Delta U^‡$) of −1.2 kcal/mol evaluated from the RQCISD(T) calculations (see Table 1) leads to an the overall rate coefficient at 298 K of 1.6793×10^{-11} molecule^{-1} cm^3 s^{-1} (see Table 2), which is in excellent agreement with the experimental[40] value of 1.6×10^{-11} molecule^{-1} cm^3 s^{-1}. In contrast, the $\Delta U^‡$ of −1.7 kcal/mol calculated at the UQCISD(T) level leads to an overall rate coefficient at 298 K of 3.8974×10^{-11} molecule^{-1} cm^3 s^{-1}, which is too large by a factor of 2.4 as compared to the experimental result. Therefore, it appears that the RQCISD(T) method performs better than UQCISD(T) in the calculation of the rate coefficient for H-atom abstraction from aldehydes by hydroxyl radicals, This finding may be partially explained by the fact that spin contamination is eliminated in the RQCISD(T) calculation of the energy barrier.

4. Results and Discussion

Figure 2 displays a schematic energy profile showing the most relevant structures concerning the main pathways on the lowest-energy PES for the reaction of HO• with acrolein. Figures 3 and 4 show selected geometrical parameters, and Table 3 gives the values of ΔU, ΔE(0 K), ΔH(298 K), and ΔG(298 K) calculated for these structures. Their total electronic energies computed at the different levels of theory as well as the ZPVEs, thermal corrections to enthalpy and Gibbs energy, are collected in Table S1 (Supporting Information). Finally, the values of K_{eq}, Γ, k_2, and the overall rate coefficient (k_{RP}) at 298 K for reactions of eqs 1–3 are summarized in Table 4.

4.1. Aldehydic H-Atom Abstraction from Acrolein by HO Radical. Two conformers for acrolein exist, due to internal rotation around the C−C single bond joining the vinyl and the aldehyde moieties of the molecule. Depending on whether the C═O and the C═C double bonds appear on the same or opposite side with respect to the C−C single bond, the conformer is called synperiplanar (labeled as **1-sp**) or antiperiplanar (labeled as **1-ap**). The QCISD/aug-cc-pVDZ minimum-energy structures corresponding to these two conformomers are shown in Figure S1 (Supporting Information). At the QCISD(T)/aug-cc-pVTZ level the **1-ap** conformer turns out to be 1.83 kcal/mol more stable than the **1-sp** conformer in terms of Gibbs energy at 298 K. This result provides an equilibrium constant $K_p°$ of 0.045, which gives an **1-ap**:1-**sp** population ratio of 95.66:4.34. Thus, it seems that just the **1-ap** conformer of acrolein has significant weight in its reactivity at room temperature. Therefore, we have only taken into account the reactions of HO• with the **1-ap** conformer.

As for acetaldehyde and in many gas-phase reactions of interest in atmospheric chemistry, Figure 2 shows that the aldehydic H-atom abstraction from acrolein by HO• begins with the barrierless formation of a prereaction complex in the entrance channel. The optimized geometry of this complex, labeled as **CXR1** (Figure 3), has C_s symmetry and was characterized as a true local minimum on the PES. The AIM topological analysis of the electron density in **CXR1** revealed the presence of a bond critical point between the oxygen atom of acrolein and the hydrogen atom of HO•, indicating that there is a bonding interaction between these atom pairs. The low value of the electron density (0.0262 e bohr^{-3}), the positive value of the its Laplacian (0.0892 e bohr^{-5}), and the positive value of the local energy density[41]

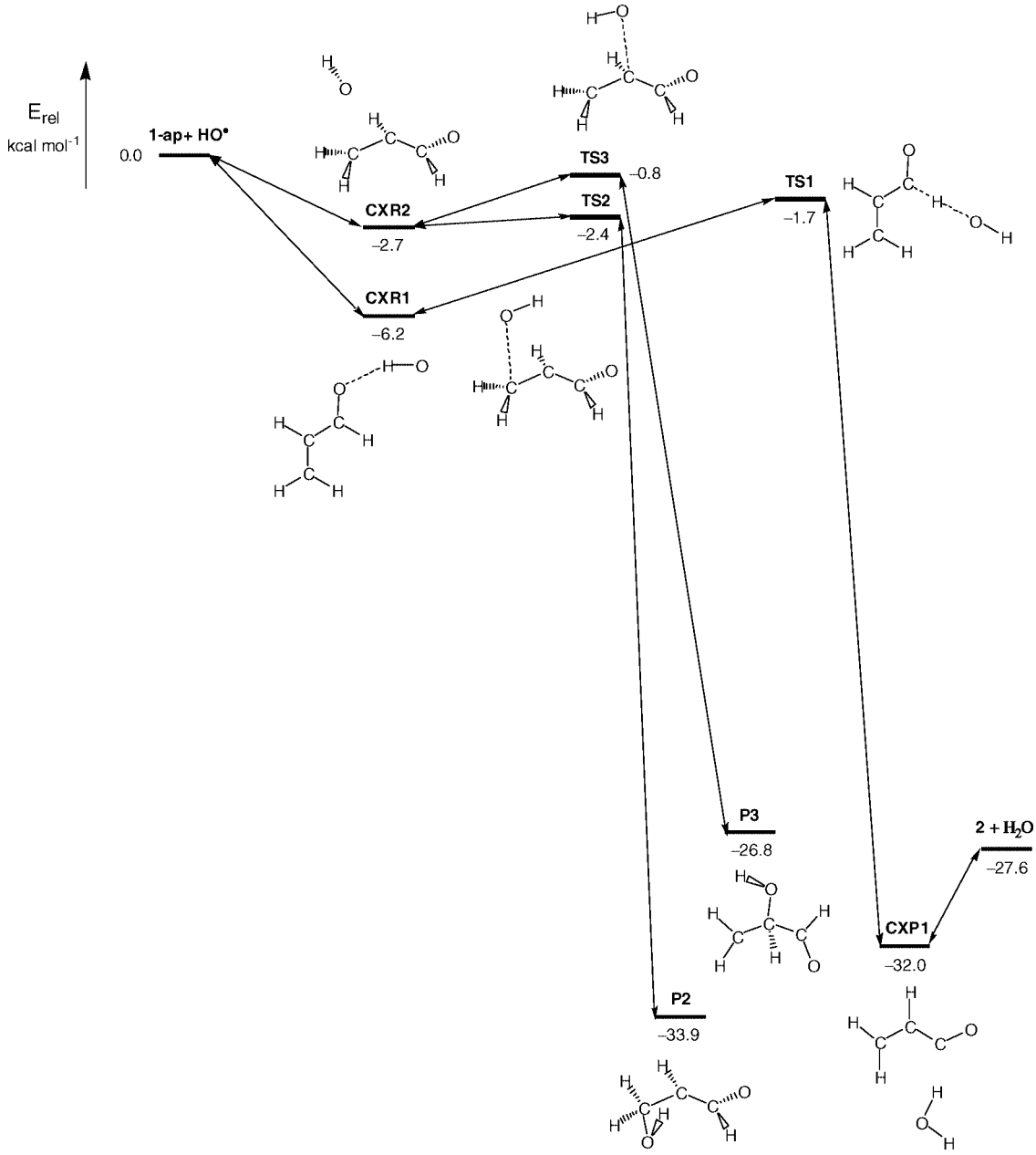


Figure 2. Schematic energy profiles showing the most relevant structures of the reaction pathways on the ground-state potential energy surface for the reaction of HO• with acrolein. Relative energy values calculated at the RQCISD(T) level of theory with the aug-cc-pVTZ basis set.

(0.0011 hartree bohr^{-3}) calculated for this bond critical point is typically associated with hydrogen-bond-like interactions. Therefore, it turns out that **CXR1** is a hydrogen-bonded complex.

The hydrogen-bond distance in **CXR1** is nearly identical to that calculated for the prereaction hydrogen-bonded complex **CXR** found in the H-atom abstraction of acetaldehyde (see Figure 1). In addition, the calculated ΔU (see Tables 1 and 3) indicate that **CXR1** and **CXR** lie 6.2 and 6.1 kcal/mol, respectively, below the energy of the isolated reactants. Therefore, the stabilization energy of these hydrogen-bonded complexes is also nearly identical. Inclusion of the correction for the BSSE leads to a stabilization energy of **CXR1** toward decomposition into their components of 5.7 kcal/mol.

After forming the prereaction hydrogen-bonded complex **CXR1**, the aldehydic H-atom is transferred from the acrolein to the HO• moiety through the transition-state structure labeled as **TS1**, displayed in Figure 3, which has C_s symmetry. A comparison between the geometries calculated for **TS1** and **TS** reveals that the bond lengths of the breaking C−H and forming H−O bonds as well as the C−H−O bond angle of both transition-state structures are very similar. The ΔU data listed in Table 3 show that **TS1** lies 1.7 kcal/mol below the energy of the reactants and 4.5 kcal/mol above the energy of **CXR1**. Inclusion of ZPVE corrections to energy leads to a $\Delta E^{\ddagger}$(0 K) of −1.1 kcal/mol for the aldehydic H-atom abstraction from acrolein by HO•. This value is identical to the $\Delta E^{\ddagger}$(0 K) calculated for acetaldehyde (see Table 1). However, the ΔG(298 K) data listed in Tables

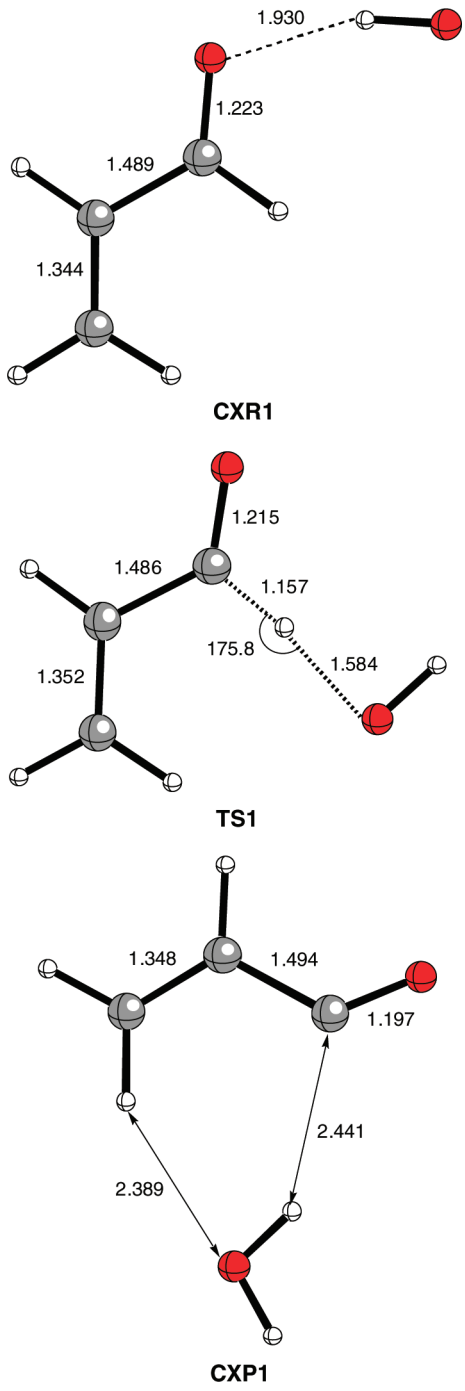


Figure 3. Selected geometrical parameters of the most relevant structures of the reaction pathway for the aldehydic H-atom abstraction from acrolein by HO•. Distances are given in Å and angles in deg.

1 and 3 show that the Gibbs energies of activation at 298 K (designated by $\Delta G^{\ddagger}$(298 K)) for the aldehydic H-atom abstraction from acrolein and acetaldehyde by HO• are calculated to be 6.3 and 5.7 kcal/mol, respectively.

The IRC calculations showed that **TS1** goes backward to **CXR1** and goes forward to give a product complex in which the already formed water is loosely bound to the CH_2=CHCO radical fragment. The optimized geometry of this complex, labeled as **CXP1** (Figure 3), was characterized as a true local minimum on the PES. The ΔU data listed in Table 3 show that **CXP1** lies 4.4 kcal/mol below the energy of the isolated products H_2O and CH_2=CHC•O (designated by **2**). Inclusion of the correction for the BSSE leads to a stabilization energy of **CXP1** toward decomposition into H_2O and **2** of 3.9 kcal/mol. We did not find an energetic barrier other than that imposed by the endoergicity for the **CXP1** complex to break apart to form the latter products. At this point we note that the aldehydic H-atom abstraction from acrolein by HO• is predicted to be exoergic. Thus, according to Table 3 the energy of reaction at 0 K (designated by ΔE_r(0 K)) is calculated to be −27.1 kcal/mol. This exoergicity is 1.9 kcal/mol lower than that calculated for acetaldehyde (see Table 1).

Concerning the results of the conventional transition-state theory calculations for the reaction of eq 1 given in Table 4, first we note that the value of 1.0345 calculated for the tunneling factor Γ indicates that the tunneling effect in the reaction of eq 1 is negligible. This feature is contrary to common belief that for an H-atom transfer process the tunneling effect should be important. However, it is worth noticing that for the related reaction of eq 9 the tunneling effect is also unimportant because the value of Γ is calculated to be 1.0320 (see Table 2). These unexpected results are ascribed to the fact that the energy barriers of reactions of eqs 1 and 9 are broad, as suggested by the small value of the imaginary vibrational frequency of the corresponding transition-state structure. Thus, the imaginary frequencies calculated at the UQCISD/aug-cc-pVDZ level for **TS** and **TS1** were found to be 207.9*i* and 221.9*i* cm^{-1}, respectively. Second, we note that the rate coefficient at 298 K of 5.7590 $\times 10^{-12}$ $molecule^{-1}$ cm^{-3} s^{-1} predicted for the reaction of eq 1 is about 1 order of magnitude smaller than the value of 1.6793 $\times 10^{-11}$ $molecule^{-1}$ cm^{-3} s^{-1} calculated for the reaction of eq 9 (see Tables 2 and 4). Since in both reactions the tunneling effect is negligible, these results may be partially explained by the fact that the $\Delta G^{\ddagger}$(298 K) calculated for the aldehydic H-atom abstraction from acrolein (6.3 kcal/mol) is 0.6 kcal/mol higher than that calculated for acetaldehyde (5.7 kcal/mol).

4.2. HO• Addition to the Acroleine C=C Double Bond. As for the aldehydic H-atom abstraction, the HO• addition to the C=C double bond of acrolein begins with the barrierless formation of a prereaction complex in the entrance channel (see Figure 2). The optimized geometry of this complex, labeled as **CXR2** (Figure 4), was characterized as a true local minimum on the PES. The net atomic charges evaluated by using the Mulliken population analysis of the effective UQCISD/aug-cc-pVDZ electron density showed an electron charge transfer in **CXR2** of 0.004 e in the direction acrolein → HO• due to a weak delocalization of the π electron density of the C=C double bond into the antibonding σ*(HO) orbital of the HO•. This result indicates that the dominant attractive interactions holding in association the acrolein and HO• partners in the **CXR2** complex arise mainly from dispersion forces. Therefore, **CXR2** is a van der Waals loosely bound complex.

As shown in Figure 4, the distance between the oxygen atom of the HO• moiety and the C=C double bond carbon atoms of the acrolein moiety are very long (i.e., 2.830 and 2.934 Å). Therefore, the structural perturbation of the two

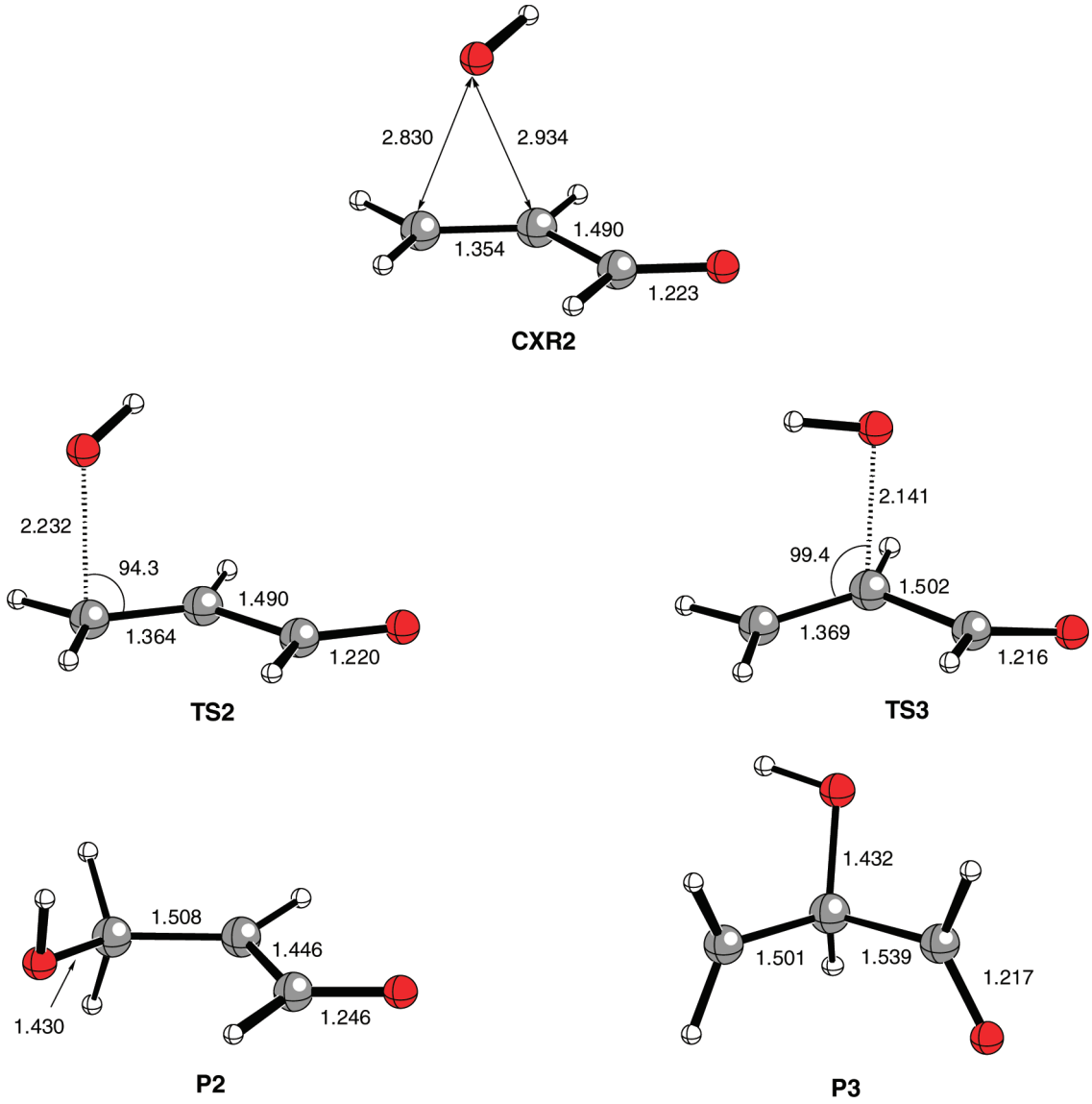


Figure 4. Selected geometrical parameters of the most relevant structures of the reaction pathway for the HO• addition to the C═C double bond of acrolein. Distances are given in Å and angles in deg.

Table 3. Relative Energies (kcal/mol) of the Most Relevant Stationary Points on the Ground-State Potential Energy Surface for the Aldehydic H-Atom Abstraction by HO• and the HO• Addition to the C═C Double Bond of Acrolein[a]

stationary point[b]	ΔU	ΔE (0 K)	ΔH (298 K)	ΔG (298 K)
1-ap + HO•	0.0	0.0	0.0	0.0
CXR1	−6.2	−4.0	−4.4	2.9
TS1	−1.7	−1.1	−1.8	6.3
CXP1	−32.0	−30.0	−29.7	−23.7
2 + H_2O	−27.6	−27.1	−26.8	−27.9
CXR2	−2.7	−1.7	−1.6	4.7
TS2	−2.4	−0.7	−1.5	7.1
TS3	−0.8	0.8	0.0	8.8
P2	−33.9	−29.7	−30.8	−21.3
P3	−26.8	−23.8	−24.7	−15.4

[a] Calculated at the RQCISD(T) level of theory with the aug-cc-pVTZ basis set. [b] See Figure 2.

partners in the **CXR2** complex is insignificant as compared with the separated species. The calculated ΔU (see Table 3) indicate that **CXR2** lies 2.7 kcal/mol below the energy of the isolated reactants. Inclusion of the correction for the BSSE leads to a stabilization energy of **CXR2** toward decomposition into their components of 2.1 kcal/mol.

After forming the prereaction complex **CXR2**, the reaction bifurcates into two different pathways. The lowest-energy barrier reaction pathway is the addition of HO• to the terminal carbon atom of the C═C double bond through the transition-state structure labeled as **TS2**, displayed in Figure 4. **TS2** shows a conserved character of the C═C double bond, planarity of the π system, and a long distance between the carbon and oxygen atoms (2.232 Å). Therefore, **TS2** is a reactant-like transition-state structure. The ΔU data listed in Table 3 show that **TS2** lies 2.4 kcal/mol below the energy of the reactants and 0.3 kcal/mol above the energy of **CXR2**. Inclusion of ZPVE corrections to energy leads to a $\Delta E^{\ddagger}$(0 K) of −0.7 kcal/mol for the HO• addition to the terminal carbon atom of the C═C double bond of acrolein. This value is 0.4 kcal/mol higher than the $\Delta E^{\ddagger}$(0 K) calculated for the aldehydic H-atom abstraction by HO•. However, the ΔG(298 K) data listed in Table 3 show that the $\Delta G^{\ddagger}$(298 K) for the HO• addition to the terminal carbon atom of the C═C double

Table 4. Equilibrium Constant (K_{eq} in molecule^{-1} cm^3; See Eq *6*) of the First Step, Tunneling Factor (Γ) and Rate Coefficient (k_2 in molecule^{-1} cm^3 s^{-1}; See Eq *7*) of the Second Step, and Rate Coefficient (k_{RP} in molecule^{-1} cm^3 s^{-1}; See Eq *5*) at 298 K of the Reaction Pathways for the Aldehydic H-Atom Abstraction by HO$^\bullet$ and the HO$^\bullet$ Addition to the C=C Double bond of Acrolein

reaction pathway	K_{eq}	Γ	k_2	k_{RP}
reaction eq 1	2.8906×10^{-22}	1.0345	1.9923×10^{10}	5.7590×10^{-12}
reaction eq 2	1.4950×10^{-23}	1.1229	1.2534×10^{11}	1.8739×10^{-12}
reaction eq 3	1.4950×10^{-23}	1.1013	6.5421×10^{9}	9.7535×10^{-14}
all[a]				0.77304×10^{-11}
exp[b]				2.0×10^{-11}

[a] Sum of the k_{RP} calculated for the three reaction pathways. [b] Reference 10.

bond of acrolein is 0.8 kcal/mol higher than the $\Delta G^{\ddagger}$(298 K) calculated for the aldehydic H-atom abstraction by HO$^\bullet$.

The IRC calculations showed that **TS2** goes backward to **CXR2** and goes forward to give the reaction product of reaction of eq 2. The optimized geometry of this adduct, labeled as **P2** (Figure 4), was characterized as a true local minimum on the PES. According to Table 3, the ΔE_r(0 K) for the reaction of eq 2 is calculated to be −29.7 kcal/mol. This exoergicity is 2.6 kcal/mol larger than that calculated for the reaction of eq 1.

As can be observed in Figure 2, the addition of HO$^\bullet$ to the central carbon atom of the C=C double bond takes place through the transition-state structure labeled as **TS3**, displayed in Figure 4. Like as found for **TS2**, the geometries of the reactants are only slightly perturbed in **TS3**, preserving the planarity of the C=C double bond π system. The distance between the carbon and oxygen atoms is long (2.141 Å). Therefore, **TS3** is also a reactant-like transition-state structure. The ΔU data listed in Table 3 show that **TS3** lies 0.8 kcal/mol below the energy of the reactants and 1.9 kcal/mol above the energy of **CXR2**. Inclusion of ZPVE corrections to energy leads to a $\Delta E^{\ddagger}$(0 K) of 0.8 kcal/mol for the HO$^\bullet$ addition to the central carbon atom of the C=C double bond. This value is 1.5 kcal/mol higher than the $\Delta E^{\ddagger}$(0 K) calculated for the for the HO$^\bullet$ addition to the terminal carbon atom of the C=C double bond.

The IRC calculations showed that **TS3** goes backward to **CXR2** and goes forward to give the reaction product of reaction of eq 3. The optimized geometry of this adduct, labeled as **P3** (Figure 4), was characterized as a true local minimum on the PES. According to Table 3, the ΔE_r(0 K) for the reaction of eq 3 is calculated to be −23.8 kcal/mol. This exoergicity is 5.9 kcal/mol lower than that calculated for the reaction of eq 2.

Our calculations show that the HO$^\bullet$ addition to the C=C double bond of acrolein is a very regioselective reaction. In fact, the difference of 1.7 kcal/mol between the $\Delta G^{\ddagger}$(298 K) calculated for reactions of eqs 2 and 3 (see Table 3) clearly indicates that HO$^\bullet$ attacks preferentially the less substituted carbon atom of the C=C double bond. Furthermore, on comparing the calculated overall rate coefficients in Table 4, it can be seen that at 298 K the HO$^\bullet$ addition to the terminal carbon atom of the C=C double bond is about 19 times faster than the addition to the central carbon atom. Specifically, the overall rate coefficients at 298 K calculated for reactions of eqs 2 and 3 lead to a branching ratio of 95.1:4.9, which is in good agreement with experimental data indicating that at least 80% of the HO$^\bullet$ addition takes place on the terminal carbon atom.[12]

The overall rate coefficients at 298 K listed in Table 4 show that the rate coefficient predicted for the reaction of eq 1 (0.57590×10^{-11} molecule^{-1} cm^{-3} s^{-1}) is about 3 times as large as that of the reaction of eq 2 (0.18739×10^{-11} molecule^{-1} cm^{-3} s^{-1}) and about 59 times larger than that of reaction of eq 3 (0.00975×10^{-11} molecule^{-1} cm^{-3} s^{-1}). As a consequence, the sum of the rate coefficients at 298 K calculated for the reactions of eqs 1 and 2 (i.e., 0.76329×10^{-11} molecule^{-1} cm^{-3} s^{-1}) is nearly identical to the value (0.77304×10^{-11} molecule^{-1} cm^{-3} s^{-1}) of the global rate coefficient at 298 K estimated for the reaction of HO$^\bullet$ with acrolein (see Table 4). At this point it is worth noting that the latter value is roughly in reasonable agreement with the reported experimental rate coefficients[6–10] ranging from 1.83 to 2.66×10^{-11} cm^3 molecule^{-1} s^{-1}. Furthermore, the overall rate coefficients at 298 K calculated for reactions of eqs 1–3 lead to a branching ratio of 74.5:24.2:1.3, which is in good agreement with experimental data indicating that about 68% of the HO$^\bullet$ reaction with acrolein proceeds via abstraction of the aldehydic H-atom, with the remainder occurring via addition to the C=C double bond.[12]

5. Summary and Conclusions

High level quantum-mechanical electronic structure calculations (UQCISD and RQCISD(T)) were carried out to analyze the subtle balance between the different pathways involved in the HO$^\bullet$ reaction with acrolein under atmospheric conditions. The energetic, structural, and vibrational results furnished by these calculations were subsequently used to perform conventional transition-state computations to predict the rate coefficients and the branching ratios of the competing abstraction and addition reactions. From the analysis of the results, the following main points emerge.

(1) The first step of all the reaction pathways studied involves the barrierless formation of a prereaction loosely bound complex in the entrance channel, lying a few kcal/mol below the energy of the reactants.

(2) The lowest-energy barrier pathway at 0 K is found to be the abstraction of the aldehydic H-atom through a transition-state structure lying 1.1 kcal/mol below the energy of the reactants. A tunneling factor of 1.0345 at 298 K is calculated for this H-atom transfer reaction. On the basis of the small value of the imaginary frequency calculated for the transition-state structure, this unexpected result is ascribed to the broad energy barrier of the reaction.

(3) After forming the prereaction complex in the $HO^{\bullet}$ addition entrance channel, the reaction bifurcates into two different pathways. The lowest-energy barrier pathway at 0 K is the addition of $HO^{\bullet}$ to the terminal carbon atom of the C═C double bond through a transition-state structure lying 0.7 kcal/mol below the energy of the reactants. The other pathway leads to the $HO^{\bullet}$ addition to the central carbon atom of the C═C double bond via a transition-state structure lying 0.8 kcal/mol above the energy of the reactants. In good agreement with available experimental data, the rate coefficients at 298 K calculated for these $HO^{\bullet}$ addition reactions lead to a branching ratio of about 95:5.

(4) The sum of the rate coefficients at 298 K calculated for the three reaction pathways studied is found to be 0.77304×10^{-11} molecule^{-1} cm^{-3} s^{-1}. This value is roughly in reasonable agreement with the reported experimental rate coefficients. Furthermore, the rate coefficients at 298 K calculated for the three reaction pathways lead to a branching ratio of about 75:24:1, which is in good agreement with available experimental data.

Acknowledgment. This research was supported by the Spanish MEC (Grant CTQ2005-07790-C02-01). Additional support came from the Catalonian AGAUR (Grant 2005SGR00111 and 2005PEIR0051/69). The larger calculations described in this work were performed at the Centre de Supercomputació de Catalunya (CESCA).

Supporting Information Available: Table summarizing total energies, zero-point vibrational energies, and thermal corrections to enthalpy and Gibbs energy as well as the Cartesian coordinates of all structures reported in this paper. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Grosjean, D. J. *Air Waste Manage. Assoc.* **1990**, *40*, 1664–1668.

(2) Tuazon, E. C.; Alvarado, A.; Aschmann, S. M.; Atkinson, R.; Arey, J. *Environ. Sci. Technol.* **1999**, *33*, 3586.

(3) Liu, X.; Jefries, H. E.; Sexton, K. G. *Atmos. Environ.* **1999**, *33*, 3005.

(4) Berndt, T.; Böge, O. *J. Phys. Chem. A* **2007**, *111*, 12099–12105.

(5) Agency for Toxic Substances and Disease Registry (ATSDR). Toxicological Profile for Acrolein, U.S. Public Health Service; U.S. Department Health and Human Services: Atlanta, GA, 1989.

(6) Maldotti, A.; Chiorboli, C.; Bignozzi, C. A.; Bartocci, C.; Carassiti, V. *Int. J. Chem. Kinet.* **1980**, *12*, 905–913.

(7) Kerr, J. A.; Sheppard, D. W. *Environ. Sci. Technol.* **1981**, *15*, 960–963.

(8) Atkinson, R.; Aschmann, S. M.; Pitts, J. N. *Int. J. Chem. Kinet.* **1983**, *15*, 75–81.

(9) Edney, E. O.; Kleindienst, T. E.; Corse, E. W. *Int. J. Chem. Kinet.* **1986**, *18*, 1355–1371.

(10) Magneron, I.; Thévenet, R.; Mellouki, A.; Le Bras, G.; Moortgat, G. K.; Wirtz, K. *J. Phys. Chem. A* **2002**, *106*, 2526–2537.

(11) Grosjean, E.; Williams, E. L.; Grosjean, D. *Total Environ.* **1994**, *153*, 195–202.

(12) Orlando, J. J.; Tyndall, G. S. *J. Phys. Chem. A* **2002**, *106*, 12252–12259.

(13) In addition to the low-energy pathways associated with reactions of eqs 1–3, which occur under atmospheric conditions, high-energy pathways could take place at high temperatures. For instance, the direct $HO^{\bullet}$ abstraction of olefinic H-atoms of acrolein leading to different substituted vinyl radicals plus H_2O might play an important role in combustion processes. Although there are not experimental data on these hypothetical reaction pathways, it is worthy to note that the energy barrier for H-atom abstraction from ethylene is 4.9 kcal/mol.[14] Since the energy barriers found for the pathways associated with reactions of eqs 1–3 range between −1.1 and 0.8 kcal/mol, it can be concluded that the reaction pathways for direct $HO^{\bullet}$ abstraction of olefinic H-atoms of acrolein are unimportant under atmospheric conditions.

(14) Senosiain, J. P.; Klippenstein, S. J.; Miller, J. A. *J. Phys. Chem. A* **2006**, *110*, 6960–6970.

(15) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(16) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(17) (a) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363. (b) Ishida, K.; Morokuma, K.; Kormornicki, A. *J. Chem. Phys.* **1977**, *66*, 2153.

(18) (a) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154. (b) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523.

(19) Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1993**, *99*, 5219.

(20) (a) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910. (b) Hampel, C.; Peterson, K. A.; Werner, H.-J. *Chem. Phys. Lett.* **1992**, *190*, 1. (c) Deegan, M. J. O.; Knowles, P. J. *Chem. Phys. Lett.* **1994**, *227*, 321.

(21) For a review, see: Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 1967.

(22) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

(23) Mayer, P. M.; Parkinson, C. J.; Smith, D. M.; Radom, L. *J. Chem. Phys.* **1998**, *108*, 604–615.

(24) Senosiain, J. P.; Miller, J. A. *J. Phys. Chem. A* **2007**, *111*, 3740–3747.

(25) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *108*, 604–615.

(26) See, e.g., McQuarrie, D. *Statistical Mechanics*; Harper and Row: New York, 1986.

(27) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(28) (a) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory;* Clarendon: Oxford, U.K., 1990. (b) Matta C. F.; Boyd R. J. An Introduction to the Quantum Theory of Atoms in Molecules. In *The Quantum Theory of Atoms in Molecules,* 1st ed.; Matta, C. F., Boyd, R. J., Eds.; Wiley-VCH: Weinheim, Germany, 2007; pp 1−34.

(29) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(30) See e.g. Wiberg, K. B.; Hadad, C. M.; LePage, T.; Breneman, C. M.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 671.

(31) 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 03 (Revision C.02)*; Gaussian, Inc.: Wallingford, CT, 2004.

(32) Werner, H.-J.; Knowles, P. J.; Almlöf, J.; Amos, R. D.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, S. T.; Hampel, C.; Leininger, C.; Lindh, R.; Lloyd, A. W.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri P.; Peterson, K. A.; Pitzer, R.; Pulay, P.; Rauhaut, G.; Schütz, M.; Stoll, H.; Stone, A. J.; Thorsteinsson, T. *MOLPRO, version 98.1*; University of Stuttgart: Germany, 1998.

(33) (a) Biegler-König, F. W.; Bader, R. F. W.; Tang, T.-H. *J. Comput. Chem.* **1982**, *3*, 317. (b) Bader, R. F. W.; Tang, T.-H.; Tal, Y.; Biegler-König, F. W. *J. Am. Chem. Soc.* **1982**, *104*, 946.

(34) Singleton, D. L.; Cvetanovic, R. J. *J. Am. Chem. Soc.* **1976**, *98*, 6812.

(35) Eyring, H. *J. Chem. Phys.* **1935**, *107*, 107.

(36) Truong, T. N.; Truhlar, D. G. *J. Chem. Phys.* **1990**, *93*, 1761.

(37) Eckart, C. *Phys. Rev.* **1930**, *35*, 1303.

(38) Alvarez-Idaboy, J. R.; Mora-Diez, N.; Boyd, R. J.; Vivier-Bunge, A. *J. Am. Chem. Soc.* **2001**, *123*, 2018.

(39) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley: NY, 1986; pp 86−87.

(40) Atkinson, R.; Baulch, D. L.; Cox, R. A.; Hampson, R. F.; Kerr, J. A.; Rossi, M. J.; Troe, J. *J. Phys. Chem. Ref. Data* **1999**, *28*, 191.

(41) (a) Cremer, D. *Croat. Chem. Acta* **1984**, *57*, 1259. (b) Cremer, D. *Angew. Chem., Int. Ed. Engl.* **1984**, *23*, 627.

CT8000798

JCTC Journal of Chemical Theory and Computation

Benchmarking pK_a Prediction Methods for Residues in Proteins

Courtney L. Stanton and Kendall N. Houk*

Department of Chemistry and Biochemistry, University of California Los Angeles, 607 Charles E. Young Drive East, Los Angeles, California 90095

Received January 2, 2008

Abstract: Methods for estimation of pK_a values of residues in proteins were tested on a set of benchmark proteins with experimentally known pK_a values. The benchmark set includes 80 different residues (20 each for Asp, Glu, Lys, and His), half of which consists of significantly variant cases (Δp$K_a \geq 1$ pK_a unit from the amino acid in solution). The method introduced by Case and co-workers [*J. Am. Chem. Soc.* **2004**, *126*, 4167−4180], referred to as the molecular dynamics/generalized-Born/thermodynamic integration (MD/GB/TI) technique, gives a root-mean-square deviation (rmsd) of 1.4 pK_a units on the benchmark set. The use of explicit waters in the immediate region surrounding the residue was shown to generally reduce high errors for this method. Longer simulation time was also shown to increase the accuracy of this method. The empirical approach developed by Jensen and co-workers [*Proteins* **2005**, *61*, 704−721], PROPKA, also gives an overall rmsd of 1.4 pK_a units and is more or less accurate based on residue type—the method does very well for Lys and Glu, but less so for Asp and His. Likewise, the absolute deviation is quite similar for the two methods—5.2 for PROPKA and 5.1 for MD/GB/TI. A comparison of these results with several prediction methods from the literature is presented. The error in pK_a prediction is analyzed as a function of variation of the pK_a from that in water and the solvent accessible surface area (SASA) of the residue. A case study of the catalytic lysine residue in 2-deoxyribose-5-phosphate aldolase (DERA) is also presented.

I. Introduction

Ionizable residues play a critical role in many of the important physical and chemical properties of proteins including folding and stability,[1–3] protein−protein interactions,[4] substrate binding,[5] and enzymatic reaction mechanisms.[6] Consequently, accurate pK_a prediction methods are of great interest for understanding pH-dependent properties of proteins and in the fields of rational drug and protein design.[7]

The pK_a value of an ionizable group can vary significantly from its value in solution due to the altered environment of the interior of the protein. These variant cases are not only the most difficult to predict, but are often the most interesting. One example can be found in the enzyme 2-deoxyribose-5-phosphate aldolase (DERA), which catalyzes the reaction shown in Figure 1.[8,9] The first step of the reaction involves nucleophilic attack by unprotonated Lys167. Lysine in solution is protonated at neutral pH, with a pK_a of 10.5. This value is perturbed to around 7 in the active site of the enzyme, allowing the reaction to occur. The environment of Lys167 is quite complicated, making it difficult to predict the pK_a.

The free energy profile of proton binding is dominated by electrostatic contributions from intraprotein interactions and protein−solvent interactions.[10] Explicit treatment of electrostatic interactions for every pair of charges in a fully atomistic model of both protein and solvent is computationally very expensive even with a classical force-field and was indeed completely infeasible before recent advances in computer power and electrostatic treatments such as the particle mesh Ewald procedure.[11] Therefore, most of the current developments in pK_a prediction have focused on implicit electrostatic treatments, especially solutions to the

* Corresponding author. E-mail: houk@chem.ucla.edu.

10.1021/ct8000014 CCC: $40.75
Published on Web 05/09/2008

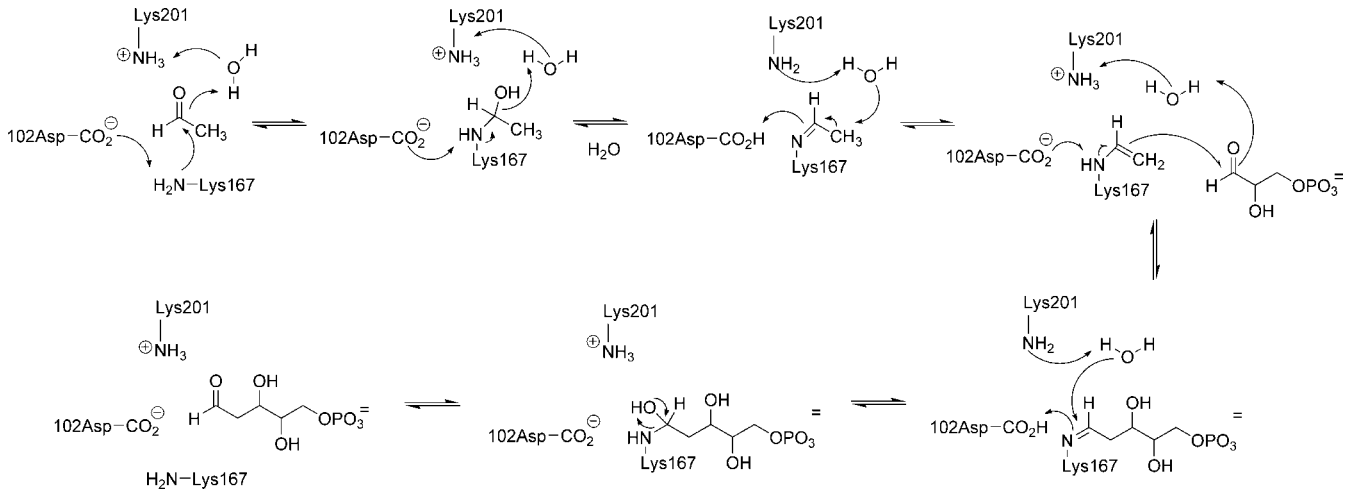


Figure 1. Mechanism of 2-deoxyribose-5-phosphate aldolase as proposed by Heine et al.[8,9]

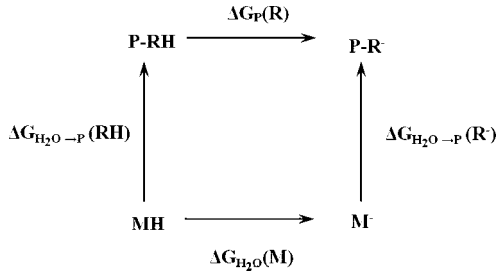


Figure 2. Thermodynamic cycle used to calculate pK_a shifts. MH refers to the model compound in aqueous solution, P-RH refers to the residue in the protein environment. p$K_{a,R}$ is the pK_a value for the residue in the protein, and p$K_{a,M}$ is the pK_a of the model compound in aqueous solution. The change in pK_a is calculated as shown in eq 1.

Poisson−Boltzmann equation (PBE),[12–16] which will be discussed in further detail in the Theoretical Background section.

This study was designed to test and compare recent promising pK_a prediction methods for the four ionizable residues Asp, Glu, Lys, and His. We present calculations from the literature and new calculations using two different methods: one method introduced by Case and co-workers[17] based on molecular dynamics and thermodynamic integration and another introduced by Jensen and co-workers,[18] which is purely empirical. The benchmark set includes equal numbers of two different groups of residues: (1) residues that have an experimental pK_a that does not vary significantly from the amino acid in solution (actually a model compound in solution; see theTheoretical Background section) with a ΔpK_a of <1, referred to as low variants, and (2) residues that vary significantly with a ΔpK_a of ≥ 1, referred to as high variants. A brief theoretical summary of pK_a prediction of residues in proteins is presented, followed by a more detailed description of the various methods.

II. Theoretical Background

Generally, pK_a prediction methods are based on the thermodynamic cycle shown in Figure 2.[19] The pK_a value of a residue in a protein (p$K_{a,R}$) is calculated relative to the experimentally determined pK_a of a model compound in aqueous solution (p$K_{a,M}$). The model compound is typically the amino acid side chain with neutral blocking groups meant to account for the backbone substituent effect in proteins, which decreases the pK_a of titratable residues[20] (see Figure 3 in the Computational Methods section). According to the cycle, the relative pK_a can be determined by calculating either (1) the difference in the free energy change of proton loss from the residue in the protein, $\Delta G_P(R)$, compared to proton loss of the model compound in solution, $\Delta G_{H_2O}(M)$, or (2) the difference in the free energy change of the protonated residue being transferred from the aqueous environment to the protein environment, $\Delta G_{H_2O} \rightarrow P(RH)$ compared to the unprotonated residue being transferred from the solvent to the protein, $\Delta G_{H_2O} \rightarrow P(R^-)$. Both strategies have been used to predict pK_a values.

$$\Delta \mathrm{p}K_a = (\mathrm{p}K_{a,R} - \mathrm{p}K_{a,M}) = \frac{1}{2.3}RT\,[\Delta G_P(R) - \Delta G_{H_2O}(M)] = \frac{1}{2.3}RT\,[\Delta G_{H_2O\rightarrow P}(R^-) - \Delta G_{H_2O\rightarrow P}(RH)] \quad (1)$$

Most studies report a root-mean-square deviation (rmsd) from experiment of ≤ 1 pK_a unit. However, this is somewhat misleading considering that most of the data are dominated by residues that are on the surface of the protein or residues that do not have strong neighboring intramolecular interactions and, therefore, do not generally vary significantly from the pK_a value of the model compound in solution. Residues that do vary substantially from p$K_{a,M}$ typically have more complex interactions and are therefore more difficult to model accurately.

As mentioned in the Introduction, solving the electrostatics for a fully atomistic model of a macromolecule such as a protein in solution is computationally quite expensive, due to the long-range nature of electrostatic interactions. Most methods have focused on decreasing this cost by introducing approximations to the full electrostatic treatment. The free energy changes in eq 1 are generally treated as purely electrostatic and can be broken down and evaluated in a number of ways. For example, Demchuk and Wade[21] describe the free energy change of transferring the protonated (or unprotonated) residue from the aqueous to the protein environment in this way:

$$\Delta\Delta G\,(\mathrm{MH} \rightarrow \mathrm{P\text{-}RH}) = \Delta\Delta G_{\mathrm{Born}} + \Delta\Delta G_{\mathrm{Boltzmann}} + \Delta\Delta G_{\mathrm{dipole}} + \Delta\Delta G_{\mathrm{charge}} \quad (2)$$

where ΔG_{Born} is the free energy to transfer a charge from deionized solvent to a neutral cavity with no permanent dipoles, $\Delta G_{\mathrm{Boltzmann}}$ represents the charge−charge interaction with the solvent−ions, $\Delta G_{\mathrm{dipole}}$ is the charge−charge interaction with the permanent dipoles, and $\Delta G_{\mathrm{charge}}$ is the charge−charge interaction in the protein.

In order to accurately calculate these free energies, one must account for a number of effects in both the solvent and in the protein, including induced dipoles,[22,23] structural relaxation,[24] Debye−Hückel screening,[25] and hydrogen bonding.[26] There are two ways to deal with these effects, either implicitly, treating them as an average macroscopic property, or explicitly, treating them as a microscopic property. The most popular methods follow a hybrid approach by approximating some or all of these effects implicitly and some explicitly. Many of these methods solve the linearized Poisson−Boltzmann equation (LPBE)[27,28] using numeric finite difference techniques.[29,30] In this framework, the protein is modeled as a low dielectric cavity with an assigned "protein dielectric" constant and is surrounded by a high dielectric medium, such as water, in which the distribution of counterions is described by a Boltzmann distribution. The electrostatic potential of the protein and solvent are calculated and the interaction energy is obtained by assigning fixed atomic charges in the protein and calculating the interaction with the protein and solvent potentials. These calculations are typically done using a molecular mechanics force-field, while the boundary between protein and solvent is determined by the atomic coordinates of the protein or model compound.

The Poisson−Boltzmann (PB) methods vary in which effects are modeled explicitly and which are approximated by adjusting the protein dielectric parameter. Assignment of the protein dielectric remains controversial, and values between 2 and 80 have been reported.[10,28,31–33] Presumably, a large protein dielectric can account for the protein relaxation and screening of electrostatic interactions;[24,34–36] this works well for solvent exposed residues and residues without significant protein charge−charge interactions, but less so for residues with more complex interactions. Theoretically, the more microscopic detail included in the model, the smaller the protein dielectric should be. If all interactions are treated explicitly, a protein dielectric equal to 1 (vacuum) should be used.

Demchuk and Wade[21] evaluated the effect of varying the protein dielectric and found that a homogeneous dielectric is not sufficient to account for the loss of microscopic detail for all residues and that the appropriate dielectric constant depends on the extent of solvent exposure. Recent methods have begun to account for the heterogeneity of the protein reaction field by introducing varying degrees of microscopic detail. Some methods retain the macroscopic continuum approach but modify the potentials used to calculate the interaction energy. Other methods account for structural reorganization by incorporating conformational sampling techniques such as molecular dynamics or Monte Carlo sampling. There are also examples of fully atomistic models that treat all the interactions explicitly. Specific examples are discussed in the Description of Methods section.

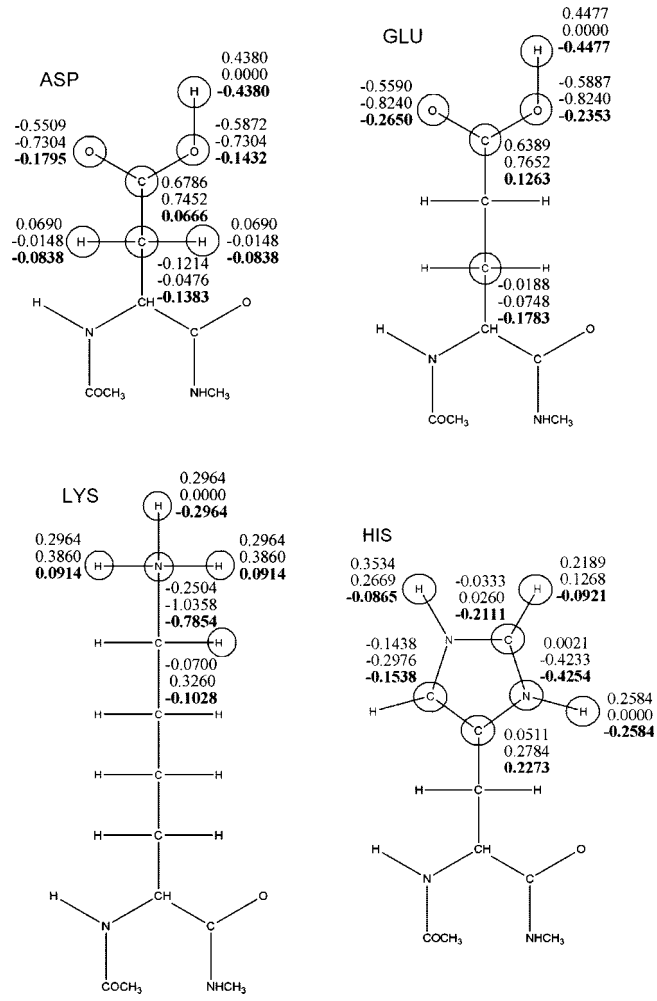


Figure 3. Model systems used for Asp, Glu, Lys, and His. The circled atoms were perturbed in the thermodynamics integration calculations for both the model system in water and the side chain in protein. The top number is the partial charge of the atom in the protonated state, the middle number is the charge in the unprotonated state, and the bottom number is the difference between the two (with some modifications to make sure that the total change was −1). Atoms were perturbed if the charge difference between the charged and uncharged state was >0.05.

The importance of choosing an appropriate benchmark set has been previously discussed.[37] If a benchmark study is dominated by surface residues (typically low variants), any model that employs a high protein dielectric constant will appear to give accurate results, since the pK_a of the model compound is also modeled in a high dielectric medium, such as water. For example, if a benchmark set includes a majority of low variants, even a model that predicts every Asp residue to have a pK_a of 4 (ΔpK_a = 0) will appear to be a good model regardless of whether it models more complex interactions properly. Presumably, it is important for a benchmark to include a significant number of high variants in order to test the ability of the model to predict unscreened charge−charge interactions.

III. Description of Methods

This section provides a description of some of the recent most promising approaches to calculating pK_a values. The methods that are later compared in the Results and Discussion section are introduced, as well as several other methods of interest.

A. Methods Addressing Protein Dielectric Heterogeneity. Mehler and Guarnieri developed a potential that analyzes the unique microenvironment surrounding each residue.[38] Specifically, they characterize the hydrophilicity or hydrophobicity around each titratable residue and use this information to modify the electrostatic potential at each site. Their method is based on sigmoidally screened Coulomb potentials (SCP) and requires less computational effort than solving the full LPBE. An rmsd of 0.5 pK_a units is reported for the method for a benchmark set of 103 experimental values in seven proteins.

Wisz and Hellinga addressed the issue of dielectric heterogeneity by introducing geometry-dependent dielectric constants for each pairwise interaction.[39] They also fit their model to an extensive set of experimental values to determine empirical parameters that take into account local structure. The benchmark set includes 260 ionizable residues in 41 different protein crystal structures. An rmsd of 0.95 pK_a units is reported.

B. Methods Addressing Conformational Sampling. Simonson and co-workers recently used a combined PB/linear response approximation (LRA) approach using molecular dynamics simulations to account for structural relaxation.[40] They predict the pK_a value for three different residues, two of which are high variants, using a small protein dielectric of 1 and 2. The average value from two calculations (with the dielectric equal to 1 or 2) gives an rmsd of 1.5 pK_a units. Pokala and Handel introduced the EGAD program as a method for protein design, and tested their electrostatic calculations against a benchmark set of experimental pK_a values.[41] A generalized-Born (GB) continuum model was used,[42] which is a much faster approximation to the PBE. This method is combined with a self-consistent mean field (SCMF) approach for rotamer optimization to account for side-chain relaxation. The reported rmsd is 0.92 pK_a units for 200 ionizable groups from 15 proteins. Gunner and co-workers published a method which combines Monte Carlo sampling with continuum electrostatics using a protein dielectric of 4 to give a reported rmsd of 0.83 pK_a units for 166 residues in 12 proteins.[43]

C. Methods Avoiding Protein Dielectric. Other methods attempt to avoid the errors incurred by using the protein dielectric as an approximation to the dielectric response. Merz used molecular dynamics/free energy perturbation simulations to calculate pK_a values with somewhat limited success, giving an rmsd of 2.8 pK_a units for 2 residues.[44] Molina and co-workers recently developed an accurate quantum mechanics/molecular mechanics (QM/MM) method, which calculates the reaction path of proton loss using quantum mechanics to model the active residue while the rest of the system is treated with a molecular mechanics force-field.[45] They report an rmsd of 0.3 pK_a units for five experimental values. This is also the most expensive method that we have discussed due to the inclusion of quantum mechanical electronic structure calculations. The development of QM/MM techniques using free energy perturbation to predict pK_a values has been pursued by others, including Cui and co-workers.[46] Most of these have met limited success or have not been tested on a benchmark set of significant size.

D. PROPKA and MD/GB/TI, Two Extremes in Computational Prediction of pK_a's. Two methods were tested here for our complete benchmark set, and these were also compared to values computed in the literature by other methods.

PROPKA. The PROPKA method is an empirical approach to calculating pK_a values developed by Jensen and co-workers.[18] It involves standard parameters for adjustments to the pK_a by residues in the vicinity of the ionizable group. The method is extremely fast and has a reported rmsd of 0.89 pK_a units. The benchmark set used to test the method includes experimental values for 314 residues in 44 proteins. The pK_a value is calculated by adding "environmental perturbations" to the pK_a value of the model residue in solution (p$K_{a,model}$) (these perturbations are referred to as ∂pK_a, not to be confused with the ΔpK_a that has been discussed previously as the change in pK_a from solution to protein):

$$\mathrm{p}K_a = pK_{a,\mathrm{model}} + \partial\mathrm{p}K_a \tag{3}$$

where ∂pK_a is the sum of the individual perturbations. Perturbations are calculated for three environmental factors: hydrogen bonding, desolvation effects, and charge−charge interactions. Each hydrogen bond is assigned a perturbation value (∂pK_a) described by a simple distance/angle function multiplied by an empirically determined parameter. The perturbation value for desolvation effects is determined by assessing how many protein atoms are within a given distance of the ionizable residue and multiplying by a parameter. Charge−charge interactions between buried pairs of residues are incorporated in a similar way—a perturbation value is assigned for each charged residue within a given distance and multiplied by an empirically determined parameter. The final pK_a is calculated by adding all of the perturbations to the p$K_{a,model}$ value. The parameters for each perturbation type were optimized empirically. This method is especially attractive for high throughput applications such as protein design.

MD/GB/TI. Case and co-workers recently published a method of pK_a prediction for protein residues using molecular dynamics free energy calculations to simulate a fully atomistic description of the entire protein.[17] While the protein electrostatics and other nonbonded forces are explicitly modeled in all of the simulations with a molecular mechanics force-field, the solvent is modeled either explicitly, as a periodic water box using the TIP3P water model,[47] or implicitly, using the generalized-Born (GB) continuum water model.[48] The GB model is an approximation to solving the PB equation. Free energies are calculated with the thermodynamic integration (TI) technique (see the Computational Methods section for a description of TI). This method will be referred to as the molecular dynamics/generalized-Born/thermodynamic integration (MD/GB/TI) technique. The authors report an rmsd of approximately 1 pK_a unit using

the MD/GB/TI method for three aspartic acid residues, two of which were previously shown to be very difficult to predict with other methods, including the LRA study discussed above in the Methods Addressing Conformational Sampling section.[40] Here, we explore the performance of this method more generally.

IV. Computational Methods

A. Preparation of Protein Atomic Coordinates. All atomic coordinates were downloaded from the PDB database, and the files were manually stripped of any solvent molecules, cofactors, metal ions, or inhibitors. The PDB identifiers and protein names in our benchmark set are the following, with corresponding references to the experimental pK_a determination: 4LZT hen egg-white lysozyme,[49,50] 2RN2 bacterial RNase H,[51,52] 1PPF turkey ovomucoid inhibitor,[53,54] 1BEO fungal beta-cryptogein,[55] 1PGA bacterial protein G B1 domain,[56] 3RN3 bovine RNase A,[20,57] 1DE3 fungal RNase alpha-sarcin,[58] 2TRX bacterial thioredoxin (oxidized),[59] 1A2P bacterial barnase,[60] 1ANS sea anemone neurotoxin III,[61] 1RGA fungal RNase T1,[62,63] 1HNG rat CD2,[64] 1XNB bacterial xylanase,[65] 2SNM bacterial nuclease mutant,[66] 1BTL bacterial beta-lactamase,[67] 1MUT bacterial MutT,[68] 1NFN human Apo E3,[69] 1FEZ bacterial phosphonoacetaldehyde hydrolase,[70] 1GS9 human Apo E4,[71] 1LE2 human Apo E2,[70] 1NZP human DNA polymerase lambda lyase domain,[72] 2BCA bovine calbindin D9K,[73] 2EBX snake erabutoxin b,[74] 3SSI bacterial proteinase inhibitor Ssi,[75] 1STN bacterial nuclease,[76] 1ERT human thioredoxin (reduced),[77] 1DG9 bovine PTPase,[78] 1L54 phage T4 lysozyme mutant,[79] 2LZM phage T4 lysozyme.[80]

B. PROPKA Calculations. PROPKA is a freely accessible program provided by the Jensen group at the University of Copenhagen.[18] Here, we used the web-based version PROPKA1.0.1.[81] Atomic coordinates can be retrieved from the PDB database by entering the PDB code or uploaded manually in the PDB format directly from the browser. Structures are automatically stripped of all nonprotein molecules, and the program calculates the pK_a for every ionizable residue (Asp, Glu, His, Lys, Tyr, Arg) in the protein within seconds. Here, the PDB files were edited manually to remove all nonprotein atoms, and, in the case of NMR structures or structures with more than one conformation for any residue, an average structure was submitted to the site. The program output reports the pK_a values and the various environmental perturbations used in the calculation.

C. MD/GB/TI Calculations. *Protein Calculations.* Protein crystal structures were downloaded from the PDB database, and the initial protonation states for each of the ionizable residues were assigned before running the molecular dynamics (MD) simulations. PROPKA, which predicts the pK_a values of Asp, Glu, Lys, His, Tyr, and Arg, was used for this purpose (generally speaking Tyr and Arg do not change protonation states in proteins; Asp, Glu, and Lys change protonation states rarely; and His is found quite frequently in either protonation state due to its "normal" pK_a of 6.3). The ionization state was determined from the PROPKA output: Asp, Glu, or His with predicted pK_a > 6.8 were protonated while Lys, Tyr, or Arg with values <7 were unprotonated in the MD simulations.

The MD simulations were performed with AMBER 8.0,[82] which includes a thermodynamic integration utility. Addition of some explicit waters improved results in certain cases (see the Results and Discussion section). Waters were added to the crystal structures using the AMBER module XLeap. The solvatecap command was used to add waters (model WAT) to within an 8 Å radius around the given residue with a van der Waals closeness parameter of 0.4 Å. This generally resulted in addition of 10−20 water molecules.

The OBC version (named for the authors: Onufriev, Bashford, and Case)[83] of the GB implicit solvent model was used. Two GB radii were tested—mbondi and mbondi2 as defined in the AMBER 8.0 literature.[82] Two AMBER forcefields were tested—ff99 and ff03. On the basis of the results of preliminary tests, ff03 and mbondi2 were used for the calculations reported here.

The thermodynamic integration utility was used to calculate the free energy change on going from the protonated state of the residue in question ($\lambda = 0$) to the unprotonated state ($\lambda = 1$). The integral describing the free energy change shown in eq 4 can be solved numerically as shown in eq 5. Dynamics were performed for three values of λ (0.11270, 0.5000, 0.88729). Equation 5 was solved with the corresponding weights (ω) of 0.27777, 0.44444, 0.27777, respectively, for each λ value. These values of λ and the weights have been determined to be optimal for calculating the free energy difference in a thermodynamic integration scheme.[84] Equation 6 describes the dependence of the potential function on λ, where V_0 is the Hamiltonian in the original protonated state and V_1 is the Hamilitonian in the unprotonated state and is referred to as linear mixing.[82]

$$\Delta G = G(\lambda = 1) - G(\lambda = 0) = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle \tag{4}$$

$$\Delta G \approx \sum_{i=1}^{n} \omega_i \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_i \tag{5}$$

$$V(\lambda) = (1 - \lambda)V_0 + \lambda V_1 \tag{6}$$

Equilibration dynamics were run for 200 ps, and production dynamics were run for an additional 200 ps. No restraints were used and the proteins were free to fluctuate. All protein structures were examined at the end of the simulations for structural integrity, which was maintained in all cases when the GB solvent model was used. There were rare instances when the protein structure denatured with the addition of some internal explicit water molecules, and these cases were not used in the benchmark set.

The change in protonation state is represented as a total change in charge of −1 (change in the van der Waals contribution from the disappearing proton is ignored). The change in charge is confined to the residue in question by altering the partial charges of specific atoms in the residue. Figure 3 shows the partial charges for each residue type for both the protonated and unprotonated states (the figure shows the model compound, but the same charges were used for both the model compound in solution and the residue in the protein).

Model Calculations. The model compounds are shown in Figure 3 and include the titratable amino acid with —CONHCH$_3$ and —NHCOCH$_3$ blocking groups. The change in free energy required to deprotonate the model compound in solution (ΔG_{H_2O}(M) in eq 1) for each residue type is determined using the same MD/GB/TI methodology as described above for protein residues. The value for ΔG_{H_2O}(M) for each residue type is calculated once and used as a reference to calculate the ΔpK_a for each protein residue in the benchmark set, using the following intrinsic pK_a values for each residue type (p$K_{a,M}$ in eq 1): Asp 4.0, Glu 4.4, Lys 10.5, and His 6.3.[28]

SASA Calculations. The percent of solvent accessible surface area (SASA) for each residue in the benchmark set was calculated with the program GETAREA 1.1.[85] GETAREA is a freely accessible web-based program provided by the Sealy Center for Structural Biology at the University of Texas.[86] Protein atomic coordinates must be supplied in PDB format and uploaded to the site for calculation. The same PDB files that were used for the PROPKA calculations were uploaded to the GETAREA Web site, and the SASA per residue was calculated simultaneously for every residue in the protein. The percent SASA is reported as the ratio of the side-chain surface area to the "random coil" surface area for that residue type. The random coil value of residue type X is defined as the average surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations. The values are listed in the GETAREA manual.[87]

V. Results and Discussion

A. Guide to Table 1. The calculated ΔpK_a values for 80 residues in 30 different proteins are presented in Table 1. The table is split into 4 subtables for residue types Asp, Glu, Lys, and His. For each type, there are 10 residues that are experimentally known to be high variants ($\Delta pK_a \geq 1$), and 10 residues that are low variants ($\Delta pK_a < 1$). The high variants are shown in bold face font. The data from columns labeled "MD/GB/TI w/waters", "MD/GB/TI w/out waters", and PROPKA were obtained in this work. The other columns represent values taken from the literature. Columns labeled "err" are the difference between predicted and experimental ΔpK_a values. The root-mean-square deviation (rmsd), mean absolute deviation (MAD), and maximum absolute deviation (MAX) from the experiment are shown in red in the bottom three rows and are calculated first according to residue type, and also as a total for all residues in blue at the end of the four tables. The reported values are ΔpK_a values, where $\Delta pK_a = pK_{a,R} - pK_{a,M}$. The MD/GB/TI and PROPKA methods were tested on the entire benchmark set. The other methods listed in Table 1 did not have data available for every residue in the benchmark set used here. The number of values used to calculate the rmsd and MAD are noted in parentheses. The first column is color-coded to show the extent of solvent accessibility. The percent SASA for each residue was calculated as described in Computational Methods: purple indicates <20% accessibility, green is >50% accessibility, and blue is ≤50% or ≥20%.

B. Methods. *MD/GB/TI.* The total rmsd of 1.4 pK_a units for the MD/GB/TI method using only the GB solvent model is relatively high compared to the PB methods. The GB approximation accounts for the solvent in a continuum way and does not always accurately model microscopic interactions such as hydrogen bonding. The worst predictions (noted with an asterisk in Table 1), with an error ≥1.5 pK_a units, were repeated with some explicit waters around the residue in question, as described in Computational Methods. The new values where the explicit water calculations were performed are entered in the column labeled "MD/GB/TI w/waters" in Table 1, but the entire benchmark set was not rerun with explicit waters due to the additional computational cost. Inclusion of explicit waters generally improved the predictions. The rmsd value for the 21 residues that were repeated was reduced from 2.4 to 1.9 pK_a units, without and with waters, respectively. The maximum absolute deviation was also significantly reduced from 5.1 to 2.5. In several cases, inclusion of explicit waters did not change the prediction significantly. When the predictions did change significantly, they were generally improved: 1DE3 Glu96, 1FEZ Lys53, 1NZP Lys312, 1STN His121, 4LZT His15, 3RN3 His48, 1DE3 His104. Exceptions to this include 4LZT Glu7 and 1DE3 His137, which were predicted somewhat less accurately with explicit waters. Inclusion of some explicit waters in the interior of the protein was used previously with a PB method.[88] The results from that study showed that experimental pK_a values could be predicted with a smaller protein dielectric when several explicit waters were used, indicating a more accurate microscopic model.

Figure 4 shows a plot of the data for the MD/GB/TI method without explicit waters. The experimental $\Delta pK_{a,exp}$ are plotted against the predicted $\Delta pK_{a,predict}$ values. The least-squares line through the origin is shown in solid black. The dashed line has a slope of 1 and is shown for comparison.[89] The R^2 value is 0.48, indicating that the two variables share 48% of their variability in common. Most of the data points fall above the dashed line in both of the plots. This indicates that the MD/GB/TI method generally overestimates the pK_a regardless of residue type. For example, Lys66 of PDB entry 2SNM was predicted to have a change in pK_a of −3.1, while the experimental value is −4.1. This means the actual pK_a value is predicted to be 7.4, which is higher than the experimental value of 6.4. The same is true for the majority of the MD/GB/TI predictions and may be the consequence of a systematic error in the method.

The method followed here involved a 200 ps equilibration and 200 ps production simulation for each value of λ. The issue of convergence was explored for several residues by extending the production run to 1 ns for each value of λ. The residues were the following: 1A2P Glu60, 1L54 Lys102, 4LZT His15, 3RN3 Asp14, and 3RN3 His48. For two of the cases, Glu60 and Asp14, the calculated ΔpK_a did not change significantly (<0.5 pK_a units). The other three predictions, Lys102, His15, and His48 were significantly improved by 1, 1.6, and 1.2 pK_a units, respectively. This indicates that many of the runs may not be fully converged and longer simulations may further reduce the error associated with this method. The entire benchmark set was not

Table 1. Comparison of Experimental pK$_a$ Values with Several Different Prediction Methods

PDB Code (residue)	Exp. ΔpK$_a$	MD/GB/TI w/ waters (This work, Simonson et al.) ΔpK$_a$		MD/GB/TI w/out waters* (This work, Simonson et al.) ΔpK$_a$		PROPKA (This work, Jensen et al.) ΔpK$_a$		Geom dep dielectric (Hellinga et al.) ΔpK$_a$		Microenv SCP (Guarnieri et al.) ΔpK$_a$		EGAD (Handel et al.) ΔpK$_a$		MCCE (Gunner et al.) ΔpK$_a$		QM/MM (Molina et al.) ΔpK$_a$	
ASP		calc	err[a]	calc	err	calc	err	calc	err	calc	err	calc	err	calc	err	calc	err
3RN3 (asp14)*	-2.2	-1.0	1.2	-0.3	1.9	-2.6	-0.4	-1.6	0.4	-1.4	0.6			-3.4	-1.2		
4LZT (asp87)	-1.9			-0.8	1.1	-1.5	0.4	-1.1	0.9	-0.9	1.1	-1.1	0.8	-2.8	-0.9		
1PPF (asp27)*	-1.8	0.2	2.0	0.6	2.4	-1.6	0.2	-0.6	1.2	-0.1	1.7	-1.0	0.8	-0.7	1.1	-2.1	-0.3
1XNB (asp11)	-1.5			-0.1	1.4	-2.0	-0.5	-0.8	0.7			-0.4	1.1				
1BEO (asp21)	-1.5			-1.1	0.4	-2.6	-1.1							1.1	2.6	-1.5	0.0
4LZT (asp18)	-1.3			-0.8	0.5	-1.2	0.1	-0.2	1.1	-0.5	0.8	-0.5	0.8	-1.0	0.3		
1XNB (asp106)	-1.3			-1.2	0.1	-1.0	0.3	-0.5	0.8			-0.5	0.8				
1PGA (asp22)	-1.1			-0.4	0.7	-1.8	-0.7	0.2	1.3	-1.1	0.0	-0.5	0.6	-1.8	-0.7		
3RN3 (asp121)*	-0.9	1.0	1.9	1.0	1.9	-0.3	0.6	-1.8	-0.9	-0.1	0.8			-0.8	0.1		
1A2P (asp75)	-0.9			0.3	1.2	-5.3	-4.4	-1.6	-0.7			2.3	3.2	0.5	1.4		
2RN2 (asp94)*	-0.8			0.3	1.1	-1.3	-0.5	-1.3	-0.5	-0.5	0.3			-0.2	0.6		
1PGA (asp47)	-0.6			-0.1	0.5	-1.3	-0.7	-0.1	0.5	-1.4	-0.8	-0.7	-0.1	-1.7	-1.1		
3RN3 (asp53)	-0.3			1.0	1.3	-0.6	-0.3	-0.5	-0.2	0.0	0.3			-0.3	0.0		
4LZT (asp52)[c]*	-0.3	2.1	2.4	2.0	2.3	-1.0	-0.7	0.8	1.1	-0.5	-0.2	-0.4	-0.1	-0.1	0.2		
1PGA (asp36)	-0.2			0.6	0.8	-0.1	0.1	-0.1	0.1	0.4	0.6	0.5	0.7	1.0	1.2		
2TRX (asp20)	-0.2			0.0	0.2	-1.5	-1.3										
1DE3 (asp59)	0.1			-0.1	-0.2	-0.8	-0.9	0.8	0.7			-1.6	-1.7				
1DE3 (asp57)	0.3			-0.1	-0.4	0.0	-0.3	-0.5	-0.8			-0.6	-0.9				
2RN2 (asp10)	2.1			2.7	0.6	3.0	0.9			1.9	-0.2			6.4	4.3		
2TRX (asp26)	4.1			3.7	-0.4	1.2	-2.9	1.9	-2.2								
RMSD (N)[b]		1.9 (4)		1.2 (20)		1.3 (20)		1.0 (17)		0.8 (12)		1.2 (13)		1.6 (14)		0.2 (2)	
MAD		1.9		1.0		0.9		0.8		0.6		1.0		1.0		0.3	
MAX		2.4		2.4		4.4		2.2		1.7		3.2		4.3		0.3	

Table 1. Continued

PDB Code (residue)	Exp. ΔpK_a	MD/GB/TI w/ waters		MD/GB/TI w/out waters		PROPKA		Geom dep dielectric		Microenv SCP		EGAD		MCCE		QM/MM	
GLU		calc	err	calc	err	calc	err	calc	err	calc	err	calc	err	calc	err	calc	Err
3RN3 (glu2)	-1.8			-2.0	-0.2	-1.7	0.1	-0.5	1.3	-0.6	1.2			-3.1	-1.3		
4LZT (glu7)*	-1.5	0.6	2.1	0.0	1.5	-0.7	0.8	-1.1	0.4	-0.9	0.6	-1.8	-0.3	-0.9	0.6	-1.7	-0.2
1PPF (glu19)	-1.2			-0.3	0.9	-0.5	0.7	-0.2	1.0	-0.3	0.9	-0.7	0.5	-2.8	-1.6	-1.7	-0.5
2RN2 (glu57)	-1.2			0.2	1.4	-1.8	-0.6	0.6	1.8	-1.7	-0.5			-1.9	-0.7		
1A2P (glu60)*	-1.2	0.4	1.6	0.7	1.9	-0.6	0.6	-1.3	-0.1			-1.2	0.0	-2.6	-1.4		
3RN3 (glu111)	-0.9			0.5	1.4	0.2	1.1	-0.4	0.5	0.0	0.9			-0.5	0.4		
2RN2 (glu129)	-0.8			-0.5	0.3	-0.9	-0.1	-1.0	-0.2	-1.4	-0.6			-1.6	-0.8		
2RN2 (glu61)	-0.5			0.5	1.0	-0.8	-0.3	-0.9	-0.4	-0.8	-0.3			-1.5	-1.0		
3RN3 (glu9)*	-0.4	1.6	2.0	1.3	1.7	0.2	0.6	-0.6	-0.2	0.2	0.6			1.0	1.4		
1BCA (glu26)	-0.3			0.4	0.7	0.4	0.7							-1.7	-1.4		
1PPF (glu10)	-0.3			0.7	1.0	0.0	0.3	-0.6	-0.3	-0.1	0.2	-0.3	0.0	-0.9	-0.6	-0.1	0.2
2RN2 (glu119)*	-0.3	1.3	1.6	1.6	1.9	-0.9	-0.6	-1.0	-0.7	-0.6	-0.3			-1.3	-1.0		
1PGA (glu27)	0.1			1.3	1.2	-1.2	-1.3	-0.7	-0.8	-1.3	-1.4	-0.1	-0.2	-0.6	-0.7		
1PPF (glu43)	0.4			0.3	-0.1	0.1	-0.3	-0.1	-0.5	0.0	-0.4	1.0	0.6	0.1	-0.3	0.1	-0.3
1DE3 (glu96)[c]*	0.7	0.7	0.0	2.4	1.7	0.6	-0.1	-0.1	-0.8			-0.3	-1.0				
1ANS (glu20)	1.0			0.7	-0.3	0.1	-0.9	-0.1	-1.1								
1RGA (glu28)	1.5			1.2	-0.3	-0.3	-1.8	-0.1	-1.6			1.3	-0.2				
4LZT (glu35)[c]	1.8			1.8	0.0	0.6	-1.2	0.8	-1.0	1.9	0.1	1.8	0.0	1.8	0.0		
1HNG (glu41)	2.3			2.3	0.0	0.3	-2.0					-1.0	-3.3	1.4	-0.9		
1XNB (glu172)[c]	2.3			2.7	0.4	2.9	0.6	0.3	-2.0			3.2	0.9				
RMSD (N)		1.6 (5)		1.1 (20)		0.9 (20)		1.0 (18)		0.7 (13)		1.1 (11)		1.0 (16)		0.3 (4)	
MAD		1.5		0.9		0.7		0.8		0.6		0.6		0.9		0.3	
MAX		2.1		1.9		2.0		2.0		1.4		3.3		1.6		0.5	

Table 1. Continued

PDB Code (residue)	Exp. ΔpK$_a$	MD/GB/TI w/ waters		MD/GB/TI w/out waters		PROPKA		Geom dep dielectric		Microenv SCP		EGAD		MCCE		QM/MM	
LYS		calc	err	calc	err	calc	err	calc	err	calc	err	calc	err	calc	err	calc	Err
2SNM (lys66)	-4.1			-3.1	1.0	-2.6	1.5	-2.6	1.5								
1L54 (lys102)*	-3.9	-1.4	2.5	-0.9	3.0	-2.6	1.3	-2.1	1.8								
1MUT (lys39)[c]*	-2.1	0.4	2.5	0.3	2.4	0.0	2.1										
1NFN (lys146)	-1.3			-0.8	0.5	0.0	1.3							-1.1	0.2		
1FEZ (lys53)[c]*	-1.2	0.2	1.4	1.8	3.0	-2.4	-1.2										
1GS9 (lys146)	-1.1			-0.8	0.3	0.0	1.1										
1LE2 (lys143)	-1.1			-0.2	0.9	-0.5	0.6										
1NFN (lys143)	-1.0			-0.3	0.7	0.0	1.0							-2.4	-1.4		
1NZP (lys312)[c]*	-1.0	0.0	1.0	0.7	1.7	-0.2	0.8										
1GS9 (lys143)	-0.6			-0.3	0.3	0.0	0.6										
1LE2 (lys146)	-0.6			-0.5	0.1	-0.1	0.5										
1PPF (lys34)	-0.4			-0.2	0.2	-0.3	0.1			0.5	0.9			-3.3	-2.9		
4LZT (lys33)	-0.1			-0.4	-0.3	-0.2	-0.1	-0.1	0.0	0.9	1.0			-0.7	-0.6		
2BCA (lys41)	0.3			0.2	-0.1	-0.1	-0.4	0.5	0.2	0.0	-0.3			0.1	-0.2		
4LZT (lys96)	0.3			0.1	-0.2	-0.3	-0.6	0.1	-0.2	0.2	-0.1			0.8	0.5		
1PGA (lys28)	0.4			0.2	-0.2	-0.6	-1.0	0.7	0.3	0.9	0.5			1.2	0.8		
2BCA (lys16)	0.4			0.7	0.3	-0.6	-1.0	1.0	0.6	0.2	-0.2			0.7	0.3		
1PPF (lys55)	0.6			0.2	-0.4	0.0	-0.6			0.0	-0.6			-0.3	-0.9		
2BCA (lys7)	0.7			0.9	0.2	0.0	-0.7	0.8	0.1	0.4	-0.3			0.4	-0.3		
2BCA (lys55)	1.3			1.3	0.0	0.0	-1.3	0.9	-0.4	0.7	-0.6			1.2	-0.1		
RMSD (N)		2.0 (4)		1.2 (20)		1.0 (20)		0.8 (10)		0.6 (9)		(0)		1.1 (11)		(0)	
MAD		1.9		0.8		0.9		0.6		0.5				0.7			
MAX		2.5		3.0		2.1		1.8		1.0				2.9			

Table 1. Continued

PDB Code (residue)	Exp. ΔpK_a	MD/GB/TI w/ waters		MD/GB/TI w/out waters		PROPKA		Geom dep dielectric		Microenv SCP		EGAD		MCCE		QM/MM	
HIS		**calc**	**err**	**calc**	**err**	**calc**	**err**	**calc**	**err**	**calc**	**err**	**calc**	**err**	**calc**	**err**	**calc**	**Err**
3EBX (his6)*	**-3.5**			-4.6	-1.1	0.0	3.5	-0.3	3.2								
3SSI (his43)	**-3.1**			-3.1	0.0	-1.4	1.7	-0.6	2.5								
1STN (his121)*	**-1.0**	0.9	1.9	1.4	2.4	-1.0	0.0	1.2	2.2			1.8	2.8				
4LZT (his15)*	-0.9	1.3	2.2	1.9	2.8	1.0	1.9	-0.2	0.7	-0.6	0.3	0.4	1.3	0.2	1.1		
1ERT (his43)	-0.8			0.2	1.0	0.0	0.8	0.4	1.2								
1DE3 (his137)[c]*	-0.5	2.1	2.6	1.3	1.8	-4.3	-3.8	-1.3	-0.8			0.7	1.2				
3RN3 (his48)*	-0.2	0.9	1.1	4.9	5.1	-3.1	-2.9	0.1	0.3	0.1	0.3			2.5	2.7		
3RN3 (his119)[c]	0.2			-0.5	-0.7	0.2	0.0	1.1	0.9	-0.1	-0.3			-0.9	-1.1		
3RN3 (his12)[c]	-0.3			0.5	0.8	-4.5	-4.2	-0.3	0.0	-0.5	-0.2			-2.1	-1.8		
1DE3 (his104)*	0.2	0.9	0.7	2.3	2.1	0.1	-0.1	-0.4	-0.6			1.3	1.1				
1DE3 (his36)	0.5			0.7	0.2	0.2	-0.3	0.3	-0.2			1.6	1.1				
2RN2 (his62)	0.7			0.7	0.0	0.7	0.0	0.6	-0.1	0.7	0.0	0.7	0.0	0.4	-0.3		
2RN2 (his124)*	0.8	-0.8	-1.6	-1.1	-1.9	0.1	-0.7	0.3	-0.5	-0.5	-1.3			-1.8	-2.6		
1DE3 (his50)*	**1.4**	3.5	2.1	3.1	1.7	-3.8	-5.2	0.4	-1.0			2.1	0.7				
1RGA (his92)[c]	**1.5**			1.2	-0.3	-0.4	-1.9	0.4	-1.1	1.1	-0.4			0.8	-0.7		
1RGA (his40)[c]	**1.6**			1.6	0.0	2.5	0.9	0.1	-1.5	1.1	-0.5			2.7	1.1		
2RN2 (his127)	**1.6**			1.6	0.0	0.8	-0.8	1.1	-0.5	1.3	-0.3			0.7	-0.9		
1DG9 (his66)	**2.0**			1.3	-0.7	1.3	-0.7	1.1	-0.9								
2LZM (his31)*	**2.8**	1.5	-1.3	0.9	-1.9	1.3	-1.5	0.9	-1.9			3.9	1.1				
1DG9 (his72)	**2.9**			3.2	0.3	1.6	-1.3	0.7	-2.2								
RMSD (N)		1.9 (8)		1.8 (20)		2.2 (20)		1.4 (20)		0.5 (9)		1.4 (8)		1.6 (9)		(0)	
Mean Abs Dev		1.8		1.3		1.6		1.1		0.4		1.2		1.4			
Max Abs Dev		2.4		5.1		5.2		3.2		1.3		2.8		2.7			
Total RMSD (N)		1.9 (21)		1.4 (80)		1.4 (80)		1.1 (65)		0.7 (43)		1.2 (32)		1.4 (50)		0.3 (6)	
Total MAD		1.8		1.0		1.0		0.8		0.5		0.9		1.0		0.3	
Total MAX		2.5		5.1		5.2		3.2		1.7		3.3		4.3		0.5	

[a] Root mean squared deviations (rmsd), mean absolute deviation (MAD), maximum absolute deviation (MAX) for predicted pK_a values are shown in red for each residue type. [b] Numbers in parentheses represent the number of values used to calculate rmsd, MAD, MAX. $\Delta pK_a = pK_{a,R} - pK_{a,M}$ from eq 1, where $pK_{a,R}$ is the residue in the protein and $pK_{a,M}$ is the model compound in solution. The values for $pK_{a,M}$ that were used for the model compound for each residue type were the following: Asp 4.0, Glu 4.4, Lys 10.5, His 6.3. Experimental ΔpK_a values in bold indicate residues whose experimental pK_a values vary more than 1 pK_a unit from their "normal" value in solution. PDB codes are color coded as follows: purple, buried residues (accessibility <20%); green, surface residues (accessibility >50%); blue, intermediate residues (20−50%).[17,18,38,39,41,43,45] [c] Catalytic residues. * Included several explicit water molecules as described in the Computational Methods section.

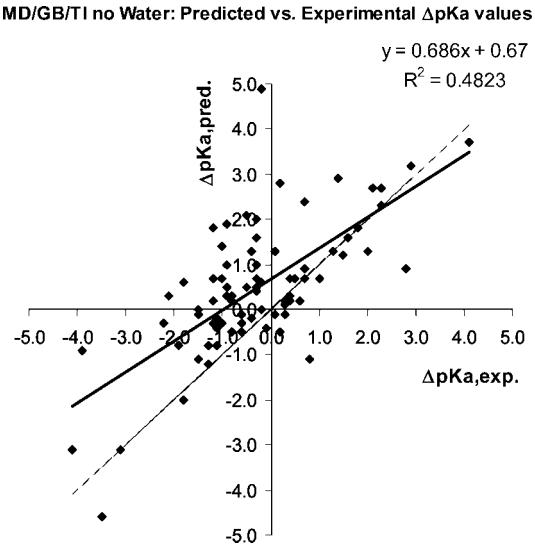


Figure 4. Plot of the predicted ΔpK_a versus experimental ΔpK_a values from Table 1 for the MD/GB/TI method with no explicit waters (under heading "MD/GB/TI w/out water"). ΔpK_a is defined as $\Delta pK_a = pK_{a,R} - pK_{a,M}$ from eq 1. The values used for $pK_{a,M}$ are Asp 4.0, Glu 4.4, Lys 10.5, and His 6.3. The plot contains 80 data points that were fitted linearly with a trendline. The resulting equation is shown with the corresponding R^2 value. The dashed line has a slope of 1 for comparison.

rerun due to the added computational cost. The path dependence of the change in free energy (i.e., forward reaction versus backward reaction) was also explored for two residues: 1A2P Glu60 and 1XNB Asp11. The free energy change was not shown to be significantly path-dependent in either case, with a difference in the free energy changes of <0.2 kcal/mol.

PROPKA. The PROPKA method has a relatively high total rmsd of 1.4 pK_a units compared to the PB methods, but has an equivalent rmsd as compared to the MD/GB/TI method. The breakdown per residue, however, shows that the method works better for Glu and Lys residues with an rmsd of close to 1.0 for both of these residue types. The maximum absolute deviation is also relatively low for Glu and Lys, 2.0 and 2.1 pK_a units, respectively. It fares somewhat worse for Asp with an rmsd of 1.3 and MAX of 4.4 and fares much worse for His with an rmsd of 2.2 and MAX of 5.2.

Figure 5 shows the same type of plot for PROPKA as is shown in Figure 4 for MD/GB/TI. The first plot shows the data for all residue types. The correlation is weaker than for MD/GB/TI, with an R^2 value of 0.28, indicating worse predictive power by PROPKA. The difference between means of the PROPKA data set and the MD/GB/TI data set is statistically significant at the 0.05 level (p-value < 0.0001).

A look at the same data plotted per residue type shows differences in predictions for the different residue types. Glu has the best correlation (R^2 is 0.49), followed by Lys (R^2 is 0.48) and Asp (R^2 is 0.48), and His (R^2 is 0.12), respectively. It appears that, for many of the His cases, the "local desolvation" effects that reflect the degree of burial in an area 4−5 Å surrounding the residue are significantly overestimated and reduce the pK_a considerably more than any of the other contributors that PROPKA estimates (i.e., Coulomb interactions and hydrogen bonding with nearby residues). One of the worst His predictions (His12 of RNase A, PDB code 3RN3), with an error of −4.2 pK_a units, is discussed in the Jensen paper.[18] The pK_a for this residue is known experimentally to be dependent on salt concentration.[56] The lack of explicit interactions between the residue and ions in the solvent is blamed for the poor prediction. The MCCE method also had trouble with this residue, giving an error of −2.0. However, three other methods were able to predict the change in pK_a within 1 unit, so it is unlikely that explicit consideration of salt−residue interactions is required for this residue.

The worst PROPKA prediction for Asp in this benchmark set is Asp75 of barnase (PDB 1A2P), with an error of −4.4 pK_a units. It is the worst Asp prediction in the Jensen study as well. The error in this case is blamed on particularly strong interactions with two nearby Arg residues. It is probably due to the double-counting of interactions with both Arg residues. The EGAD method also had a problem predicting the pK_a of this residue. However, three of the methods were able to predict the change in pK_a to within 1 unit.

A closer look at the plot for Lys predictions reveals that PROPKA tends to predict no change in pK_a for lysine. In fact, 40% of the Lys residues were predicted to have zero change, despite the fact that none of them actually had an experimental change of zero. This tendency can be explained by the observation that PROPKA predicted the majority of lysines to be surface residues. Charge−charge interactions for surface residues are not calculated in the PROPKA method, and few of the lysines in the benchmark set had hydrogen bonding interactions, leading to an unchanged pK_a. This might suggest that the criterion for a lysine residue to be considered buried is too stringent, or perhaps a third category between surface and buried would be useful.

The poor performance in some cases is probably due to lack of sampling, and incorporation of some type of sampling technique would most likely improve PROPKA predictions. The method shows great promise considering it is entirely empirical and the calculations take only seconds. The performance is quite comparable to the MD/GB/TI method, which took between 24 and 48 h per residue on a single Pentium III processor depending on protein size. In this study, PROPKA was used to assign initial protonation states before the MD/GB/TI calculations were performed, since the correct direction of pK_a change was predicted for 82% of the benchmark set (80% for His, which has the highest rmsd).

Other Methods. The geometry-dependent dielectric method of Hellinga and co-workers[39] has a low total rmsd of 1.1 for this benchmark set. It also has the most reported values in common with this benchmark set—65 out of 80. For this reason, a comparison between this method and the PROPKA and MD/GB/TI methods is the most valid. The maximum absolute deviation, 3.2 pK_a units, is comparable to that for MD/GB/TI method (with explicit waters), 2.6, and much lower than for PROPKA, 5.2.

The microenvironment SCP method of Mehler and Guarnieri[38] has the lowest rmsd of the PB methods, 0.7 pK_a units, calculated with 43 of the 80 benchmark residues. It was also the only method that predicted His residues with a better rmsd

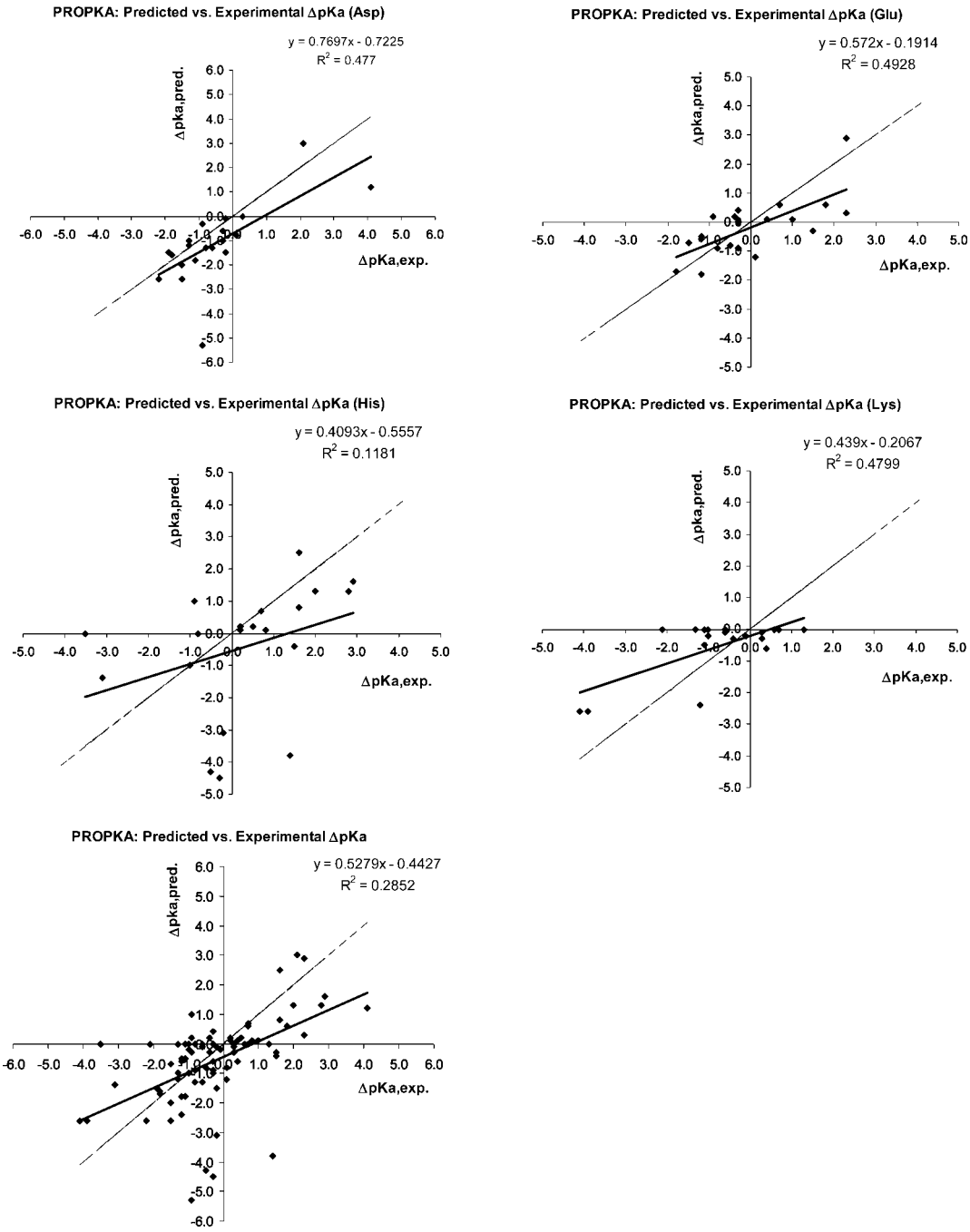


Figure 5. Plots of the predicted ΔpK_a versus experimental ΔpK_a values from Table 1. ΔpK_a is defined as $\Delta pK_a = pK_{a,R} - pK_{a,M}$ from eq 1. The last plot includes all 80 data points, while the remaining plots show the 20 data points for each residue type. In each case, the data were fitted linearly with a trendline. The resulting equation is shown for each graph with the corresponding R^2 value. The dashed line has a slope of 1.

(0.5) and MAX (1.3) than its overall values. The EGAD method and the MCCE method have a comparable total rmsd of 1.2 and 1.4, respectively. These values are slightly higher than their reported values of 0.92 and 0.86 pK_a units. The discrepancy is most likely due to the demanding nature of this benchmark set, which contains a high ratio of buried versus surface residues (21/32). The maximum absolute deviations for EGAD, 3.3, and MCCE, 4.3, are slightly higher than the other PB methods. The QM/MM method has the lowest rmsd, 0.3, and the lowest maximum absolute deviation, 0.5. However, there are only seven entries in Table 1 for this very expensive method, and none of which are buried residues. It would be interesting to see if QM/MM methods can maintain an extremely low rmsd even with buried residues.

C. Variance and SASA. The variance referred to here is the absolute change in pK_a on going from solution to protein. Special care was taken to choose a benchmark set that contains equal numbers of low and high variants, as mentioned in the Introduction. The assumption is that predicting high variants is more difficult, due to the complexity of interactions that cause a large change in pK_a. However, the data from this benchmark test do not show any such relationship. The error was plotted versus the variance (data not shown) giving a very small R^2 value of 0.05.

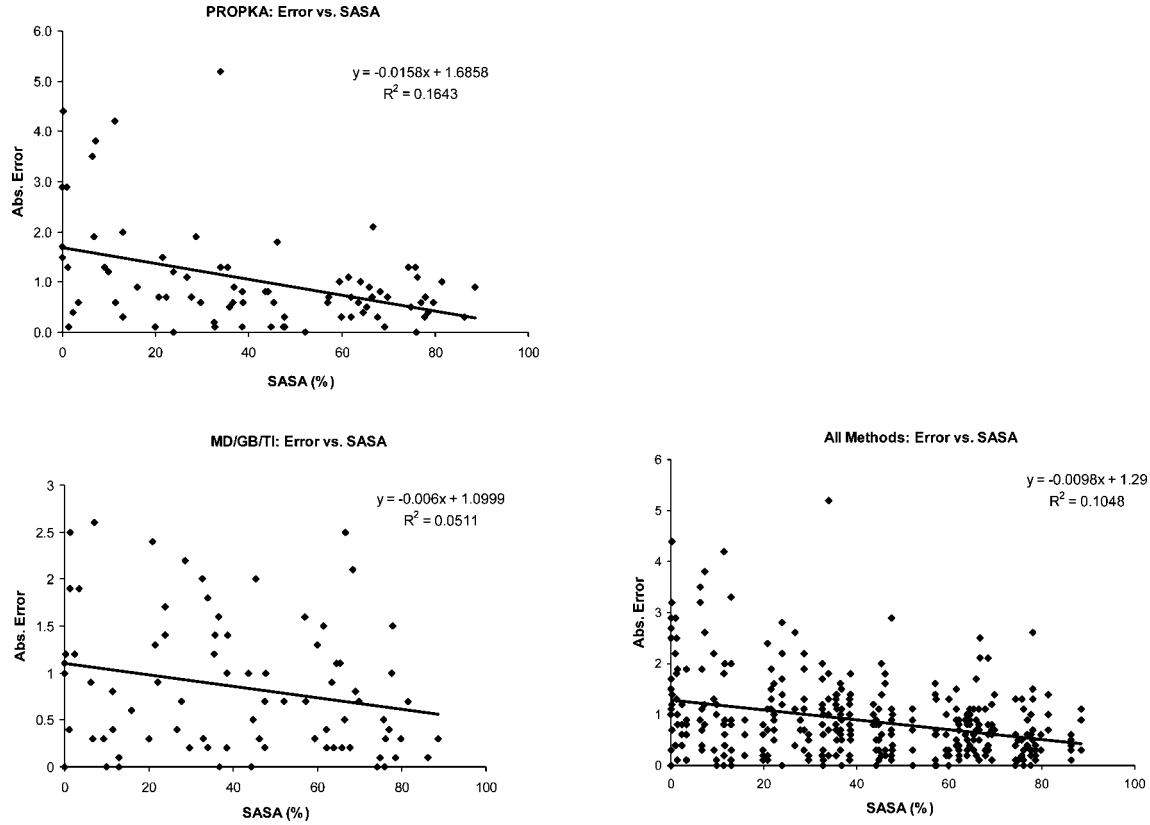


Figure 6. Plots of the calculated error versus the percent of solvent accessible surface area (SASA). (top) Results for the benchmark set including all the methods from Table 1. (bottom, left) Results for the benchmark set from only PROPKA calculations with 80 data points. (bottom, right) Results for the benchmark set from only MD/GB/TI calculations with 80 data points.

Another property that was explored here is the solvent accessible surface area (SASA). The importance of the SASA in determining the transfer energy associated with moving a compound from polar solvent to nonpolar solvent has been well established.[90–92] The thermodynamic cycle in Figure 2 shows why SASA is important for the calculation of pK_a's. As noted in the Introduction, the relative pK_a can be calculated by comparing the change in energy of transferring the protonated and unprotonated residue from solvent to protein. The relationship between error and the percent SASA was explored in Figure 6. It was expected that the error would decrease with increasing SASA, since buried residues have very different interactions than the model compound in solution.

The MD/GB/TI and PROPKA methods were plotted separately, and a third plot contains all the data from Table 1. The plot of all the data shows the expected trend—as the residues are more exposed to solvent (going from left to right on the *x*-axis), the error becomes closer to zero and predicting the change in pK_a becomes easier. The correlation is statistically significant (p-value < 0.0001; i.e., assuming the correlation is just chance, the probability of getting the results we did is less than 0.01%). Despite the statistical significance, the relationship is surprisingly weak, as indicated by a small R^2 value of 0.11. The PROPKA data show a stronger relationship with an R^2 value of 0.16 (p-value < 0.0001), while the MD/GB/TI data shows a significantly weaker relationship with an R^2 value of 0.05 (p-value is 0.02). The fact that SASA is a better predictor of error for PROPKA than for MD/GB/TI makes sense, since the latter method is not parametrized with a set of data that is largely dominated by surface residues, as is the case for PROPKA. When the points corresponding to MD/GB/TI method are removed from the all data plot in Figure 6, the R^2 value is increased to 0.12, showing a stronger relationship between error and SASA. However, the other methods do not collectively show as strong a relationship as PROPKA.

Ionic Strength and pK_a Dependence. One source of error in pK_a prediction is the ability to model accurately the effect of the surrounding ionic strength in solution. While most pK_a's are not highly dependent on ionic strength, one residue in particular that was used in this benchmark is known experimentally to be sensitive to salt concentration. The pK_a of Glu10 in turkey ovomucoid third domain is known to increase by 0.8 pK_a units on going from 1 M KCl to 1 mM KCl.[52] This relationship was probed using the MD/GB/TI method. The pK_a was predicted to increase by 0.7 pK_a units, indicating that this method can accurately model the effects of ionic screening, despite the implicit representation of ions in solution. However, in this case, it is known experimentally that conformational changes and not direct ionic interactions are responsible for the change in pK_a. For the latter interaction, an explicit representation is probably desirable. There is also evidence that the effect of salt concentration on pK_a's cannot be entirely accounted for by an ionic screening model.[93] However, this simple model was sufficient in this example.

D. Case Study: 2-Deoxyribose-5-phosphate Aldolase. Aldolases catalyze stereoselective reactions that involve carbon−carbon bond formation. This characteristic has been

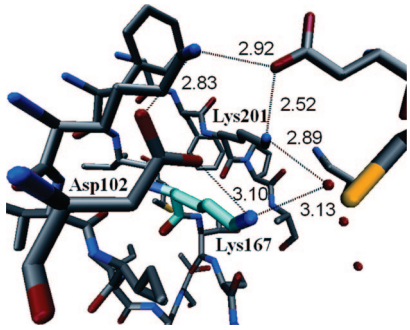


Figure 7. Active site of 2-deoxyribose-5-phosphate aldolase (PDB code: 1P1X). The catalytic lysine (Lys167) is shown in blue. There is an intricate hydrogen bonding network that involves two other lysines, two aspartic acids, and a crystal water.

exploited in the biocatalysis of useful products.[94,95] One well-studied aldolase is 2-deoxyribose-5-phosphate aldolase (DERA). As mentioned in the Introduction (see Figure 1), DERA catalyzes the aldol reaction of acetaldehyde and D-glyceraldehyde 3-phosphate to form 2-deoxyribose-5-phosphate. The first step of the reaction involves nucleophilic attack by an unprotonated lysine. Experimental isolation of the carbinolamine and Schiff base intermediates has proven that Lys167 is the catalytic active site residue.[8] Though the pK_a of Lys167 has not yet been determined experimentally, the pK_a must be close to 7 in order for the reaction to occur. The pK_a was calculated with the MD/GB/TI method to be 6.9 (ΔpK_a = −3.5). PROPKA predicted the pK_a of Lys167 to be 7.7. Figure 7 shows the active site of DERA, with Lys167 highlighted in blue. There is an intricate network of hydrogen bonds involving two aspartic acids, three lysines, and a crystal water. In the crystal structure, Lys167 is hydrogen bonded to one of the aspartates and the water. A neutral lysine would be favored in that position since the nearby aspartates are already ion-paired with other lysines. This case is an example where the MD/GB/TI and PROPKA methods can be used to identify possible catalytic residues.

VI. Conclusions

The benchmark set was chosen to include equal numbers of high (Δp$K_a \geq 1$ from solution to enzyme) and low (Δp$K_a < 1$ from solution to enzyme) variants, with the assumption that it is more difficult to predict high variants since the interactions that produce a change in pK_a are difficult to model. However, the variance was not shown to be a good predictor of the error in pK_a calculations. The SASA proved to be a somewhat better predictor and did show the expected trend of decreasing error with increasing solvent accessibility for all methods except for MD/GB/TI, which did not have a significant correlation.

Most recent methods for predicting pK_a values report an rmsd near 1 pK_a unit. The methods studied here and their corresponding total rmsd values for this benchmark set are the following: the MD/GB/TI method of Simonson et al.[17]—rmsd 1.4 with 80 values; the PROPKA method of Jensen and co-workers[18]—rmsd 1.4 with 80 values; the geometry-dependent dielectric method of Wisz and Hellinga[39]—rmsd 1.1 with 65 values; the microenvironment SCP method of Mehler and co-workers[26]—rmsd 0.7 with 43 values; the EGAD method of Pokala and Handel[41]—rmsd 1.2 with 32 values; the MCCE method of Georgescu et al.[43]—rmsd 1.4 with 50 values; and the QM/MM method of Molina and co-workers[45]—rmsd 0.3 with 6 values. Most of the methods were shown to produce fairly consistent results regardless of residue type, with the exception of His. Most methods fared somewhat worse for this residue type, except for the microenvironment SCP method. This is most likely due to the tautomerism of His, which is not explicitly modeled in any of the methods. A similar increase in rmsd for His residues was found in another pK_a benchmark study.[96]

A much higher MAX than rmsd was found for all methods (except in the case of the QMMM method). The highest MAX value of 5.2 pK_a units was a prediction from the PROPKA method. However, the other more sophisticated PB methods did only marginally better giving MAX values of 3.2,[39] 3.3,[41] and 4.3.[35] The fully atomistic MD/GB/TI method was shown to be equally unimpressive in this measure, giving MAX values of 2.5 and 5.1, with and without explicit waters, respectively (with longer simulation time, the MAX was reduced to 3.9 without explicit waters). The only method that gave a MAX less than 2 was the microenvironment SCP, which seems to be the most promising method. However, considering the high absolute errors, the challenge remains to find a highly accurate method that can be done with minimal computational cost. Improvements may include constant longer simulation times, pH simulations in which the pK_a of ionizable residues can change,[97] incorporation of cofactors and metal ions, improved solvent models and inclusion of some explicit water, and improved force fields.

Acknowledgment. We are grateful for the comments, criticisms, and helpful discussions contributed by David Case, Ernest Mehler, Navin Pokala, Tracy Handel, and especially Jan Jensen during the evolution and progress of this work. We are grateful to the Defense Advanced Research Projects Agency (DARPA) for financial support of this research. Part of this work was performed with funding from a University of California Lawrence Livermore National Laboratory (LLNL) graduate fellowship to C.S. under Contract No. B558556.

References

(1) Hendsch, Z. S.; Jonsson, T.; Sauer, R. T.; Tidor, B. *Biochemistry* **1996**, *35*, 7621–7625.

(2) Elcock, A. H.; McCammon, J. A. *J. Mol. Biol.* **1998**, *280*, 731–748.

(3) Schaefer, M.; Sommer, M.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 1663–1683.

(4) Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153–159.

(5) Warshel, A. *Acc. Chem. Res.* **1981**, *14*, 284–290.

(6) Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.

(7) Marshall, S. A.; Morgan, C. S.; Mayo, S. L. *J. Mol. Biol.* **2002**, *316*, 189–199.

(8) Heine, A.; Luz, J. G.; Wong, C. H.; Wilson, I. A. *J. Mol. Bio.* **2004**, *343*, 1019–1034.

(9) Heine, A.; DeSantis, G.; Luz, J. G.; Mitchell, M.; Wong, C.-H.; Wilson, I. A. *Science* **2001**, *294*, 369–374.

(10) Antosiewicz, J.; McCammon, J.; Gilson, M. *Biochemistry* **1996**, *35*, 7819–7833.

(11) York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345–8348.

(12) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509–521.

(13) Gilson, M. K.; Rashin, A.; Fine, R.; Honig, B. *J. Mol. Biol.* **1985**, *184*, 503–516.

(14) Warwicker, J.; Watson, H. J. *Mol. Biol.* **1982**, *157*, 671–679.

(15) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219–10225.

(16) Raquet, X.; Lounnas, V.; Lamotte-Brasseur, J.; Frere, J.; Wade, R. *Biophys. J.* **1997**, *73*, 2416–2426.

(17) Simonson, T.; Carlsson, J.; Case, A. D. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.

(18) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704–721.

(19) Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368–8372.

(20) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415–436.

(21) Demchuk, E.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 17373–17387.

(22) Warshel, A. *Nature* **1987**, *330*, 15–16.

(23) Warshel, A.; Russell, S. T.; Churg, A. K. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 4785–4789.

(24) Sham, Y. Y.; Muegge, I.; Warshel, A. *Biophys. J.* **1998**, *74*, 1744–1753.

(25) Garcia-Moreno, E. B. *Methods Enzymol.* **1994**, *240*, 645–667.

(26) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem. B* **2000**, *104*, 6490–6498.

(27) You, T. J.; Bashford, D. *Biophys. J.* **1995**, *69*, 1721–1733.

(28) Gilson, M. K.; Honig, B. H. *Nature* **1987**, *330*, 84–86.

(29) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Chem.* **1990**, *19*, 301–332.

(30) Yang, A.-S.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins* **2004**, *15*, 252–265.

(31) Pethig, R. *Annu. Rev. Phys. Chem.* **1992**, *43*, 177–205.

(32) King, G.; Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366–4377.

(33) Sternberg, M. J.; Hayes, F. R.; Russell, A. J.; Thomas, P. G.; Fersht, A. R. *Nature* **1987**, *330*, 86–88.

(34) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243–252.

(35) Gunner, M. R.; Alexov, E. *Biochim. Biophys. Acta* **2000**, *1458*, 63–87.

(36) Karp, D. A.; Gittis, A. G.; Stahley, M. R.; Fitch, C. A.; Stites, W. E.; Garcia-Moreno, B. E *Biophys. J.* **2007**, *92*, 2041–2053.

(37) Schutz, C.; Warshel, A. *Proteins* **2001**, *44*, 400–417.

(38) Mehler, E.; Guarnieri, F. *Biophys. J.* **1999**, *77*, 3–22.

(39) Wisz, M. S.; Hellinga, H. W. *Proteins* **2003**, *51*, 360–377.

(40) Archontis, G.; Simonson, T. *Biophys. J.* **2005**, *88*, 3888 3904.

(41) Pokala, N.; Handel, T. M. *Protein Sci.* **2004**, *13*, 925–936.

(42) Still, W.; Tempczyk, A.; Hawley, R.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(43) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophys. J.* **2002**, *83*, 1731–1748.

(44) Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 3572–3575.

(45) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. *J. Phys. Chem. A* **2005**, *109*, 6634–6643.

(46) Riccardi, D.; Schaefer, P.; Yang, Y.; Haibo, Y.; Ghosh, N.; Prat-Resina, X.; Koenig, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458–6469.

(47) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926–935.

(48) Bashford, D.; Case, D. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.

(49) Bartik, K.; Redfield, C.; Dobson, C. M. *Biophys. J.* **1994**, *66*, 1180–1184.

(50) Kuramitsu, S.; Hamaguchi, K. *J. Biochem.* **1980**, *87*, 1215–1219.

(51) Oda, Y.; Yoshida, M.; Kanaya, S. *J. Biol. Chem.* **1993**, *268*, 88–92.

(52) Oda, Y.; Yamazaki, T.; Nagayama, K.; Kanaya, S.; Kuroda, Y.; Nakamura, H. *Biochemistry* **1994**, *33*, 5275–5284.

(53) Scaller, W.; Robertson, A. D. *Biochemistry* **1995**, *34*, 4714–4723.

(54) Swint-Kruse, L.; Robertson, A. D. *Biochemistry* **1995**, *34*, 4724–4732.

(55) Gooley, P. R.; Keniry, M. A.; Dimitrov, R. A.; Marsh, D. E.; Gayler, K. R.; Grant, B. R. *J. Biol. NMR* **1998**, *12*, 523–534.

(56) Khare, D.; Alexander, P.; Antosiewicz, J.; Bryan, P.; Gilson, M.; Orban, J. *Biochemistry* **1997**, *36*, 3580–3589.

(57) Baker, W.; Kintanar, A. *Arch. Biochem. Biophys.* **1996**, *327*, 189–199.

(58) Perez-Canadillas, J. M.; Campos-Olivas, R.; Lacadena, J.; del Pozo, A. M.; Gavilanes, J. G.; Santoro, J.; Rico, M.; Bruix, M. *Biochemistry* **1998**, *37*, 15865–15876.

(59) Qin, J.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1996**, *35*, 7–13.

(60) Oliveberg, M.; Arcus, V. L.; Fersht, A. R. *Biochemistry* **1995**, *34*, 9424–9433.

(61) Norton, R. S.; Cross, K.; Braach-Maksvytis, V.; Wachter, E. *Biochem. J.* **1993**, *93*, 45–551.

(62) Giletto, A.; Pace, C. N. *Biochemistry* **1999**, *38*, 13379–13384.

(63) Inagaki, F.; Kawano, Y.; Shimada, I.; Takahashi, K.; Miyazawa, T. *J. Biochem.* **1981**, *89*, 1185–1195.

(64) Chen, H. A.; Pfuhl, M.; McAlister, M. S. B.; Driscoll, P. C. *Biochemistry* **2000**, *39*, 6814–6824.

(65) Joshi, M. D.; Hedberg, A.; McIntosh, L. P. *Protein Sci.* **1997**, *6*, 2667–2670.

(66) Garcia-Moreno, E. B.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. E.; Spencer, D. S.; Stites, W. E. *Biophys. Chem.* **1997**, *64*, 211–224.

(67) Damblon, C.; Raquet, X.; Lian, L.-Y.; Lamotte-Brasseur, J.; Fonze, E.; Charlier, P.; Roberts, G. C. K.; Frere, J. M. *Proc. Natl. Acad. Sci.* **1996**, *93*, 1747–1752.

(68) Harris, T. K.; Wu, G.; Massiah, M. A.; Mildvan, A. S. *Biochemistry* **2000**, *39*, 1655–1674.

(69) Lund-Katz, S.; Mohamed, Z.; Wehrli, S.; Dhanasekaran, P.; Baldwin, F.; Weisgraber, K. H.; Phillips, M. C. *J. Biol. Chem.* **2000**, *275*, 34459–34464.

(70) Zhang, G.; Mazurkie, A. S.; Dunaway-Mariano, D.; Allen, K. N. *Biochemistry* **2002**, *41*, 13370–13377.

(71) Lund-Katz, S.; Wehrli, S.; Zaiou, M.; Newhouse, Y.; Weisegraber, K. H.; Phillips, M. C. *J. Lipid Res.* **2001**, *42*, 984–901.

(72) Guanghua, G.; DeRose, E. F.; Kirby, T. W.; London, R. E. *Biochemistry* **2006**, *45*, 1785–1794.

(73) Kesvatera, T.; Jonsson, B.; Thulin, E.; Linse, S. *J. Mol. Biol.* **1996**, *259*, 828–839.

(74) Inagaki, F.; Miyazawa, T.; Hori, H.; Tamiya, N. *Eur. J. Biochem.* **1978**, *89*, 433–442.

(75) Fujii, S.; Akasaka, K.; Hatano, H. *J. Biochem.* **1980**, *88*, 798–796.

(76) Lee, K. K.; Fitch, C. A.; Garcia-Moreno, B. *Protein Sci.* **2002**, *11*, 1004–1016.

(77) Foreman-Kay, J. D.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1992**, *31*, 3442–3452.

(78) Dillet, V.; Van Etten, R. L.; Bashford, D. *J. Phys. Chem. B* **2000**, *104*, 11321–11333.

(79) Dao-pin, S.; Anderson, D. E.; Baase, W. A.; Dahlquist, F. W. I.; Matthews, B. W. *Biochemistry* **1991**, *30*, 11521–11529.

(80) Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 2403–2408.

(81) PROPKA can be accessed via the web at http://propka.ki.ku.dk.

(82) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, 2004.

(83) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.

(84) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004–2010.

(85) Fraczkiewicz, R.; Braun, W. *J. Comput. Chem.* **1998**, *19*, 319–333.

(86) GETAREA can be accessed via the web at http://pauli.utmb.edu/cgi-bin/get_a_form.tcl.

(87) The GETAREA manual can be accessed via the web at http://pauli.utmb.edu/getarea/area_man.html.

(88) Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; Garcia-Moreno, B. E. *Biophys. J.* **2002**, *82*, 3289–3304.

(89) All linear regression and correlation was done with Microsoft Excel.

(90) Street, A. G.; Mayo, S. L. *Fold Des.* **1998**, *3*, 253–258.

(91) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(92) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199–203.

(93) Kao, Y.-H.; Fitch, C. A.; Bhattacharya, S.; Sarkisian, C. J.; Lecomte, J. T. J.; Garcia-Moreno, B. E. *Biophys. J.* **2000**, *79*, 1637–1654.

(94) Gijsen, H. J. M.; Qiao, L.; Fitz, W.; Wong, C. H. *Chem. Rev.* **1996**, *96*, 443–473.

(95) Seebeck, F. P.; Guainazzi, A.; Amoreira, C.; Baldridge, K. K.; Hilvert, D. *Angew. Chem., Int. Ed.* **2006**, *45*, 6824–6826.

(96) Davies, M. N.; Toseland, C. P.; Moss, D. S.; Flower, D. R. *BMC Biochem.* **2006**, *7*, 18–30.

(97) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038–2048.

CT8000014

JCTC Journal of Chemical Theory and Computation

A Systematic Comparison of Second-Order Polarization Propagator Approximation (SOPPA) and Equation-of-Motion Coupled Cluster Singles and Doubles (EOM−CCSD) Spin−Spin Coupling Constants for Selected Singly Bonded Molecules, and the Hydrides NH_3, H_2O, and HF and Their Protonated and Deprotonated Ions and Hydrogen-Bonded Complexes

Janet E. Del Bene,*,† Ibon Alkorta,‡ and José Elguero‡

Department of Chemistry, Youngstown State University, Youngstown, Ohio 44555, and Instituto de Química Médica, CSIC, Juan de la Cierva, 3, E-28006 Madrid, Spain

Received April 1, 2008

Abstract: Second-order polarization propagator approximation (SOPPA) and equation-of-motion coupled cluster singles and doubles (EOM−CCSD) methods have been employed for the calculation of one-bond spin−spin coupling constants in series of small molecules and ions, and of one- and two-bond coupling constants across X−H···Y hydrogen bonds. For isolated molecules, one-bond SOPPA coupling constants ^1J(X-Y) involving ^{13}C, ^{15}N, ^{17}O, and ^{19}F have larger absolute values than corresponding EOM−CCSD coupling constants, with the EOM−CCSD values being in significantly better agreement with available experimental data. The difference between SOPPA and EOM−CCSD tends to increase as the number of nonbonding electrons on the coupled atoms increases, and the SOPPA values for O−F coupling are significantly in error. Similarly, the absolute values of SOPPA one-bond coupling constants ^1J(X−H) for the hydrides NH_3, H_2O, and FH and their protonated and deprotonated ions are greater than EOM−CCSD values, with the largest differences occurring for F−H coupling. One- and two-bond coupling constants ^1J(X-H), ^{1h}J(H−Y), and ^{2h}J(X−Y) across X−H···Y hydrogen bonds in neutral, protonated, and deprotonated complexes formed from the hydrides are similar at SOPPA and EOM−CCSD, with the largest differences again found for ^1J(F−H) in complexes with F−H as the proton donor, and ^{2h}J(F−F) for $(FHF)^-$. The signs of ^1J(X−H), ^{1h}J(H−Y), and ^{2h}J(X−Y) are the same at both levels of theory, as is their variation across the proton-transfer coordinate in F−H···NH_3. SOPPA would appear to provide a reliable and more cost-effective alternative approach for computing coupling constants across hydrogen bonds, although couplings involving F may be problematic.

Introduction

In an early paper on coupling constants, Enevoldsen et al. noted the critical dependence of computed coupling constants (J) on the quality of the wave function used to evaluate this NMR property[1] and employed the second-order polarization propagator approximation (SOPPA)[1−5] to obtain indirect spin−spin coupling constants. More recently, Barone and co-workers compared SOPPA and density functional theory (DFT) methods for two- and three-bond F−F coupling constants[6] and found the former to be in better agreement with experimental data, due in part to the treatment of electron-correlation effects in SOPPA. Indeed, it has been

* Corresponding author e-mail: jedelbene@ysu.edu.
† Youngstown State University.
‡ Instituto de Química Médica.

10.1021/ct800111j CCC: $40.75
Published on Web 05/14/2008

recognized that the quality of DFT coupling constants often decreases as the number of nonbonding pairs of electrons on the coupled atoms increases and electron correlation effects assume increased importance.[7]

Coupling constants may also be evaluated at an even higher level of electron correlation treatment using the equation-of-motion coupled cluster singles and doubles (EOM−CCSD) method.[8–11] This method is generally viewed as the benchmark method, but unfortunately, the computational cost of EOM−CCSD limits it applicability to relatively small systems or systems with high computational symmetry. Since SOPPA is a computationally much less expensive method than EOM−CCSD, we decided to undertake a systematic comparison of one-bond spin−spin coupling constants (J) obtained using these two methods on a set of small molecules H_mX-YH_n with X−Y single bonds, where X and Y = C, N, O, and F; selected fluorine derivatives of these molecules; and the hydrides NH_3, H_2O, and HF and their protonated and deprotonated ions, along with CH_4. In addition, one-bond ^1J(X−H) and ^{1h}J(H−Y) and two-bond ^{2h}J(X−Y) coupling constants across the X−H−Y hydrogen bonds in the neutral and charged hydrogen-bonded complexes formed from these hydrides and their ions, and the changes in these coupling constants along the proton-transfer coordinate for $F-H\cdots NH_3$, have been evaluated at both levels of theory. For this study, we have used the same geometries and the same basis set for both SOPPA and EOM−CCSD calculations, so that the only difference in coupling constants is a consequence of differences in the wave function model. In this paper, we compare the SOPPA and EOM−CCSD coupling constants and evaluate the performance of these methods by comparison with experimental data whenever possible. Previous comparisons of computed coupling constants obtained using SOPPA and CCSD-based methods have been reported,[12] but these are on a much more limited scope.

Methods

Computed optimized geometries for all monomers and complexes were obtained at second-order Møller−Plesset perturbation theory (MP2)[13–16] with the 6-31+G(d,p) basis set.[17–20] Vibrational frequencies were computed to verify that the computed geometries are equilibrium structures on the potential energy surfaces. These geometries were used for calculations of coupling constants, except for a few cases specifically identified below.

Spin−spin coupling constants involving ^{13}C, ^{15}N, ^{17}O, ^{19}F, and ^1H in all monomers and complexes were computed using the SOPPA[1–5] and EOM−CCSD methods in the configuration-interaction-like approximation,[8–11] with all electrons correlated. For these calculations, the Ahlrichs[21] qzp basis set was placed on nonhydrogen atoms and the qz2p basis set on all H atoms. In the Ramsey approximation, the total coupling constant (J) is a sum of four contributions: the paramagnetic spin−orbit (PSO), diamagnetic spin−orbit (DSO), Fermi-contact (FC), and spin-dipole (SD). All terms have been computed for all systems. Geometry optimizations were carried out with the Gaussian 03 suite of programs;[22] SOPPA calculations were performed using Dalton-2[23] at the CSIC Computing Center, and the EOM−CCSD calculations were done with ACES II[24] on the Itanium Cluster at the Ohio Supercomputer Center.

Table 1. SOPPA and EOM−CCSD One-Bond Heavy-Atom Coupling Constants ^1J(X−Y) (Hz) for Selected Small Molecules with X−Y Single Bonds[a]

molecule	SOPPA	EOM−CCSD	exptl.
H_3C-CH_3[b]	37.7	34.6	33.4[c]
H_3C-NH_2	−6.3	−5.7	−4.5[d]
H_3C-OH ^1J(C−O)	13.7	13.4	
H_3C-OH ^1J(O−H)	−80.0	−77.5	−85[e]
H_3C-F	−168.2	−161.9	
CF_4	−315.1	−289.4	−257[f]
H_2N-NH_2	−5.9	−5.7	
F_2N-NF_2 ^1J(N−N)	−18.4	−17.3	
F_2N-NF_2 ^1J(N−F)	196.0	183.7	164[g]
F_2N-NF_2 ^1J(N−F)	169.4	162.8	
H_2N-OH	−1.5	−2.5	
H_2N-F	−8.0	−4.1	
NF_3[h]	236.4	224.7	217[i]
HO−OH	26.2	22.2	
FO−OF[j] ^1J(O−F)	−690.0	−445.7	−424[k]
FO−OF ^1J(O−O)	−88.1	−79.4	
HO−F	−727.8	−576.1	
FO−F[l]	−382.0	−278.6	−300 ± −30[k]

[a] Calculations were carried out on molecules at optimized MP2/6-31+G(d,p) geometries, except as noted. If the sign of the experimental coupling constant was not reported, the computed sign was given. [b] Geometry from ref 34a. [c] See ref 25a. [d] See ref 25b. [e] See ref 26a. [f] See ref 25c. [g] See ref 26b. [h] Geometry from ref 31. [i] See ref 29. [j] Geometry from ref 33. [k] See ref 26a. [l] Geometry from ref 32.

Results and Discussion

Small Molecules with X−Y Single Bonds. Table 1 lists SOPPA, EOM−CCSD, and experimental[25–30] one-bond spin−spin coupling constants for the set of small molecules H_mX-YH_n, with X−Y being a single-bond. To this set, CF_4, F_2N-NF_2, NF_3, FOF, and FOOF have been added since experimental one-bond couplings are available for these molecules. When experimental coupling constants are available, the experimental geometry has been used if it is available,[31–34] since this circumvents to some extent the effect of zero-point motion on computed coupling constants.[35] The molecule F_2 has been omitted from this study since it has a very large CCSD t2 amplitude.

From Table 1, it can be seen that the signs of the computed coupling constants at both levels of theory are the same, and with one exception, namely, H_2N-OH, the absolute value of the SOPPA coupling constant is greater than the EOM−CCSD value. Table S1 of the Supporting Information reports the PSO, DSO, FC, and SD components of J. From Table S1, it can be seen that the DSO terms computed by both methods are relatively small and do not influence the comparisons. What is most interesting is that the absolute values of SOPPA PSO, FC, and SD terms are also greater than the corresponding EOM−CCSD values, with the only exceptions being the FC term for H_2N-OH and the PSO term for NF_3.

Table 1 lists the computed SOPPA and EOM−CCSD one-bond couplings and the available experimental data. The listing in Table 1 for the molecules H_mX-YH_n is in the order

of increasing electronegativity and increasing number of electron pairs on the coupled atoms. Thus, the first set of molecules consists of H_3C-CH_3, H_3C-NH_2, H_3C-OH, and H_3C-F and its derivative CF_4. The next set has H_2N-NH_2 and its derivative F_2N-NF_2, H_2N-OH, and H_2N-F and its derivative NF_3. For all of these molecules except H_2N-OH, SOPPA coupling constants have larger absolute values than EOM−CCSD. In the case of H_2N-OH, the computed coupling constants are very small at −1.5 and −2.5 Hz, respectively. For the series of C-containing molecules H_3C-CH_3, H_3C-NH_2, H_3C-OH, and H_3C-F, the SOPPA and EOM−CCSD coupling constants are similar. It is interesting to note that the EOM−CCSD value of ^1J(C−C) for H_3C-CH_3 at its experimental geometry differs from the experimental coupling constant by only 1 Hz, while the SOPPA value is 4 Hz too high. For H_3C-NH_2, the EOM−CCSD coupling constant ^1J(C−N) of −5.7 Hz is closer to the experimental coupling constant of −4.5 Hz,[25b] for which the SOPPA value is −6.3 Hz. The only significant difference found for coupling constants in the subset of C-containing molecules is ^1J(C−F) for CF_4, for which SOPPA and EOM−CCSD give −315.1 and −289.4 Hz, respectively, at the equilibrium geometry of this molecule. These overestimate the experimental value of −257 Hz,[25c] but part of this difference may be attributed to the use of an equilibrium geometry. Nevertheless, the EOM−CCSD value is closer to the experimental value.

SOPPA and EOM−CCSD coupling constants ^1J(N−N) are similar for both H_2N-NH_2 and F_2N-NF_2. There are two unique one-bond N−F coupling constants for F_2N-NF_2 in this molecule of C_2 symmetry. The average SOPPA and EOM−CCSD values are 182 and 173 Hz, respectively, with the EOM−CCSD average being closer to the experimental coupling constant of 164 Hz.[26b] The two remaining N-containing molecules H_2N-OH and H_2N-F have small coupling constants. For these, the SOPPA and EOM−CCSD values are −1.5 and −2.5 Hz and −8.0 and −4.1 Hz, respectively. Finally, SOPPA and EOM−CCSD values of ^1J(N−F) for NF_3 at its experimental geometry are 236.4 and 224.7 Hz, respectively. Once again, the EOM−CCSD value is closer to the experimental coupling constant of 217 Hz.[29] Thus, for the subset of N-containing molecules, the difference between the SOPPA and EOM−CCSD values is greatest when N is coupled to O and F.

SOPPA and EOM−CCSD coupling constants ^1J(O−O) for HO−OH and FO−OF are 26.2 and 22.2 Hz and −88.1 and −79.4 Hz, respectively. However, significantly larger differences between SOPPA and EOM−CCSD coupling constants are found for ^1J(O−F) for FO−OF, HO−F, and FO−F. At the experimental geometry of FO−OF, the SOPPA ^1J(O−F) is −690.0 Hz, compared to the EOM−CCSD value of −445.7 Hz. The EOM−CCSD value agrees with the experimental value of −424 Hz.[26a] However, it should be noted that, in the original paper,[28] this value was measured using ^{17}O NMR for a sample of ^{17}O-enriched FOOF and incorrectly assigned to a triplet. Hence, this value should be viewed with caution. ^1J(O−F) for HO−F is significantly overestimated by SOPPA at −727.8 Hz, compared to the EOM−CCSD value of −576.1 Hz, but no experimental value is available for comparison. Finally, ^1J(O−F) for FO−F at its experimental geometry is −382.0 Hz for SOPPA and −278.6 Hz for EOM−CCSD, the latter being within the uncertainty of the experimental value of −300 ± 30 Hz.[26a]

Table 2. SOPPA and EOM−CCSD One-Bond Coupling Constants ^1J(X−H) for NH_3, H_2O, and HF and Their Protonated and Deprotonated Ions[a]

species	SOPPA	EOM−CCSD	exptl.
CH_4	123.7	116.4	125.3[b]
NH_2^-	−44.2	−43.1	
NH_3	−63.9	−61.5	−61.2[c]
NH_4^+	−78.4	−75.0	−73.3[c]
OH^-	−65.7	−62.8	
H_2O[d]	−81.8	−79.7	−79[e]
OH_3^+	−124.3	−120.7	
FH[d]	531.4	520.9	529[f]
FH_2^+	648.5	631.5	

[a] Calculations were carried out on molecules at optimized MP2/6-31+G(d,p) geometries, except as noted. If the sign of the experimental coupling constant was not reported, the computed sign was given. [b] See ref 25d. [c] See ref 26c. [d] Geometry from ref 34b. [e] See ref 26a. [f] See ref 26d.

Since the above data suggest that the SOPPA method fails to describe O−F coupling properly, it is reasonable to ask whether this description could be improved by employing the SOPPA(CCSD) method, in which CCSD amplitudes are used instead of MP2.[1] The SOPPA(CCSD) values of ^1J(O−F) for FOF and FOOF are −391 and −777 Hz, respectively, both greater in absolute value than the corresponding SOPPA values and, therefore, further removed from the EOM−CCSD and experimental values. Therefore, no additional SOPPA(CCSD) calculations were done.

NH_3, H_2O, FH, and Their Protonated and Deprotonated Ions. Table 2 lists values of ^1J(X−H) for CH_4, and for the hydrides NH_3, H_2O, and FH, and their protonated and deprotonated ions. Table S2 (Supporting Information) lists the individual components of these coupling constants. A familiar pattern emerges, insofar as the absolute values of the SOPPA total coupling constants and their major components are greater than the corresponding EOM−CCSD values. Moreover, there is an overall increase in the difference between SOPPA and EOM−CCSD as the number of nonbonding electrons increases. The largest differences are found for ^1J(F−H) for FH and FH_2^+, but in these two cases, the differences are relatively small given the magnitude of ^1J(F−H).

Experimental values of ^1J(X−H) are available for CH_4,[25d] NH_4^+,[26c] NH_3,[26c] H_2O,[26a] CH_3OH,[26a] and FH.[26d] The SOPPA ^1J(C−H) value of 123.7 Hz is closer to the experimental value of 125.3 Hz than is the EOM−CCSD value of 116.4 Hz. The SOPPA ^1J(N−H) value for NH_4^+ is −78.4 Hz compared to −75.0 Hz for EOM−CCSD and −73.3 Hz for the experimental value. Similarly, the EOM−CCSD value of ^1J(O−H) for H_2O of −79.7 Hz is closer to the experimental value of −79 Hz than is the SOPPA value of −81.8 Hz, both evaluated at the experimental geometry of H_2O. At the equilibrium geometry of CH_3OH, both the computed SOPPA and EOM−CCSD values of ^1J(O−H) underestimate the experimental value. What is most interesting is the SOPPA ^1J(F−H) value for FH, which is 531.4 Hz at the

experimental geometry, and in closer agreement with the experimental value of 529 Hz than the EOM−CCSD value of 520.9 Hz.

Statistical Analysis of SOPPA and EOM−CCSD versus Experimental Results. A quantitative comparison of the performance of the SOPPA and EOM−CCSD methods versus experimental results can be achieved through a correlation analysis. Included in this analysis are two-bond H−H coupling constants for H_2O, NH_3, NH_4^+, and CH_4, which are known experimentally[30] and which are reported in the Supporting Information. Ideally, a perfect correlation between theory and experiment in a regression analysis yields a slope of 1.00, an intercept of 0.00 Hz, and a correlation coefficient of 1.00. The regression analysis for the SOPPA coupling constants versus the experimental ones yields a slope of 0.80 ± 0.04, an intercept of 13.1 ± 10.0 Hz, and a correlation coefficient of 0.962. For the computed EOM−CCSD coupling constants versus experimental results, the slope is 0.99 ± 0.01, the intercept is 2.0 ± 2.6 Hz, and the correlation coefficient is 0.997. Thus, the superiority of EOM−CCSD is manifest. Using reduced coupling constants (K) for the analysis yields similar results. It was noted above that SOPPA values of ^1J(F−O) coupling constants for F−O−F and F−O−O−F are in poor agreement with experimental values. If these two molecules are removed from the regression analysis, the performance of SOPPA significantly improves, as indicated by a slope of 0.95 ± 0.02, an intercept of 4.0 ± 3.1 Hz, and a correlation coefficient of 0.995. Eliminating these two molecules from the comparison of EOM−CCSD versus experimental results does not significantly change the statistics, yielding a slope of 0.99 ± 0.01, an intercept of 2.7 ± 2.1, and a correlation coefficient of 0.995. However, the agreement between computed and experimental coupling constants is still better for EOM−CCSD compared to SOPPA.

Some of the limitations of the treatment of the experimental data in the above analysis should be noted. If more than one experimental value of the coupling constant has been reported, the one judged to be the most reliable has been used. There has been no adjustment made for uncertainties in the experimental assignments and error bars given for the experimental data. Finally, there has been no attempt to take into account the fact that the experimental coupling constants have been measured under different conditions, some in the gas phase and others in solution. In the latter circumstance, no adjustments have been made to account for possible interactions between the solute and the solvent.

Hydrogen-Bonded Complexes. One-bond ^1J(X−H) and ^{1h}J(H−Y) and two-bond ^{2h}J(X−Y) spin−spin coupling constants across X−H···Y hydrogen bonds for $N_2H_5^-$, $(NH_3)_2$, $N_2H_7^+$, $H_2NH{:}OH^-$, $HOH{:}NH_3$, $H_3NH^+{:}OH_2$, $H_2NH{:}F^-$, $FH{:}NH_3$, $H_3NH^+{:}FH$, $O_2H_3^-$, $(H_2O)_2$, $O_2H_5^+$, $HOH{:}F^-$, $FH{:}OH_2$, $H_2OH^+{:}FH$, $(FHF)^-$, $(FH)_2$, and $F_2H_3^+$ are reported in Table 3, and the components of J are listed in Table S3 (Supporting Information). For heterodimers, the species listed first is the proton donor. The first important observation is that the signs of the coupling constants computed at these two levels of theory are identical. As a result, the generalizations made previously about the signs of coupling constants and their significance which were based on EOM−CCSD coupling constants can be applied to SOPPA coupling constants. In particular, since the magnetogyric ratios of ^{15}N and ^{17}O are negative and that of ^{19}F is positive, all two-bond reduced coupling constants ^{2h}K(X−Y) are positive[36] except for $(HF)_2$. Since the magnetogyric ratio of ^1H is also positive, all reduced one-bond coupling constants ^1K(X−H) are positive.[37] Finally, a negative value of ^{1h}K(H−Y) indicates that the hydrogen bond is traditional, while a positive value is indicative of the proton-shared character of the hydrogen bond.[38]

Table 3. SOPPA and EOM−CCSD Coupling Constants ^1J(X−H), ^{1h}J(H−Y), and ^{2h}J(X−Y) Across X−H···Y Hydrogen Bonds in Neutral, Anionic, and Protonated Complexes Formed from NH_3, H_2O, and FH[a]

	^1J(X−H)		^{1h}J(H−X)		^{2h}J(X−Y)	
complex	SOPPA	EOM	SOPPA	EOM	SOPPA	EOM
$N_2H_5^-$	−67.0	−64.4	6.4	6.3	10.5	10.2
$(NH_3)_2$	−67.3	−64.7	1.6	1.6	1.9	1.9
$N_2H_7^+$	−64.4	−61.3	−0.4	−0.1	13.1	12.9
$H_2NH{:}OH^-$	−66.0	−63.5	9.2	9.0	6.9	6.8
$HOH{:}NH_3$	−83.1	−80.8	3.3	3.4	3.4	3.4
$H_3NH^+{:}OH_2$	−76.1	−73.0	9.7	9.7	12.0	11.7
$H_2NH{:}F^-$	−66.3	−63.8	−55.6	−54.6	−24.1	−24.4
$FH{:}NH_3$	440.3	430.8	2.4	2.6	−42.5	−43.1
$H_3NH^+{:}FH$	−79.1	−75.8	−42.7	−41.5	−29.0	−28.1
$O_2H_3^-$	−53.2	−51.4	0.8	1.4	13.6	13.8
$(H_2O)_2$	−83.3	−81.0	4.8	4.8	1.3	1.3
$O_2H_5^+$	−49.3	−47.4	−49.5	−47.4	39.9	39.5
$HOH{:}F^-$	−68.6	−66.8	−65.2	−66.0	−36.5	−38.5
$FH{:}OH_2$	492.4	482.5	7.0	7.1	−18.0	−18.2
$H_2OH^+{:}FH$	−113.6	−110.3	−54.0	−55.0	−72.0	−71.1
$(FHF)^-$ [b]	109.2	106.0	109.2	106.0	228.6	254.4
$(FH)_2$	507.9	498.0	−26.6	−26.6	−40.7	−39.5
$F_2H_3^+$	203.4	195.4	203.4	195.4	612.0	609.8

[a] Calculations carried out at optimized MP2/6-31+G(d,p) geometries, except where noted. [b] Calculation carried out at the optimized CCSD(T)/aug′-cc-pVTZ geometry.

As observed above, the SOPPA coupling constants ^1J(X-H) for the isolated monomers have larger absolute values than the corresponding EOM−CCSD coupling constants. This same pattern is seen when X-H acts as the proton donor in hydrogen-bonded complexes. Moreover, the largest differences in the complexes are about 10 Hz for ^1J(F−H) in $FH{:}NH_3$, $FH{:}OH_2$, $(FH)_2$, and even $F_2H_3^+$. This difference reflects the monomer difference of 10.5 Hz.

Only small differences are found between SOPPA and EOM−CCSD ^{1h}J(H−Y) values, except for ^{1h}J(H−F) for the symmetric hydrogen bond in $F_2H_3^+$, in which case the SOPPA value of 203.4 Hz is 8.0 Hz greater than the EOM−CCSD value. Moreover, the two-bond coupling constants ^{2h}J(X−Y) are similar at the two levels of theory, except for $(FHF)^-$, in which case the SOPPA value is 228.6 Hz and less than the EOM−CCSD value of 254.4 Hz. The geometry used for this ion is the CCSD(T)/aug′-cc-pVTZ geometry, since these coupling constants are extremely sensitive to distance, and vibrationally averaged coupling constants for this system have been reported.[39] Zero-point vibrational averaging decreases ^{2h}J(F−F) from its EOM−CCSD equilibrium value of 254 Hz to 210 Hz and brings it into good agreement with the estimated experimental value of about 220 Hz.[40,41] Since the effect of vibrational averaging

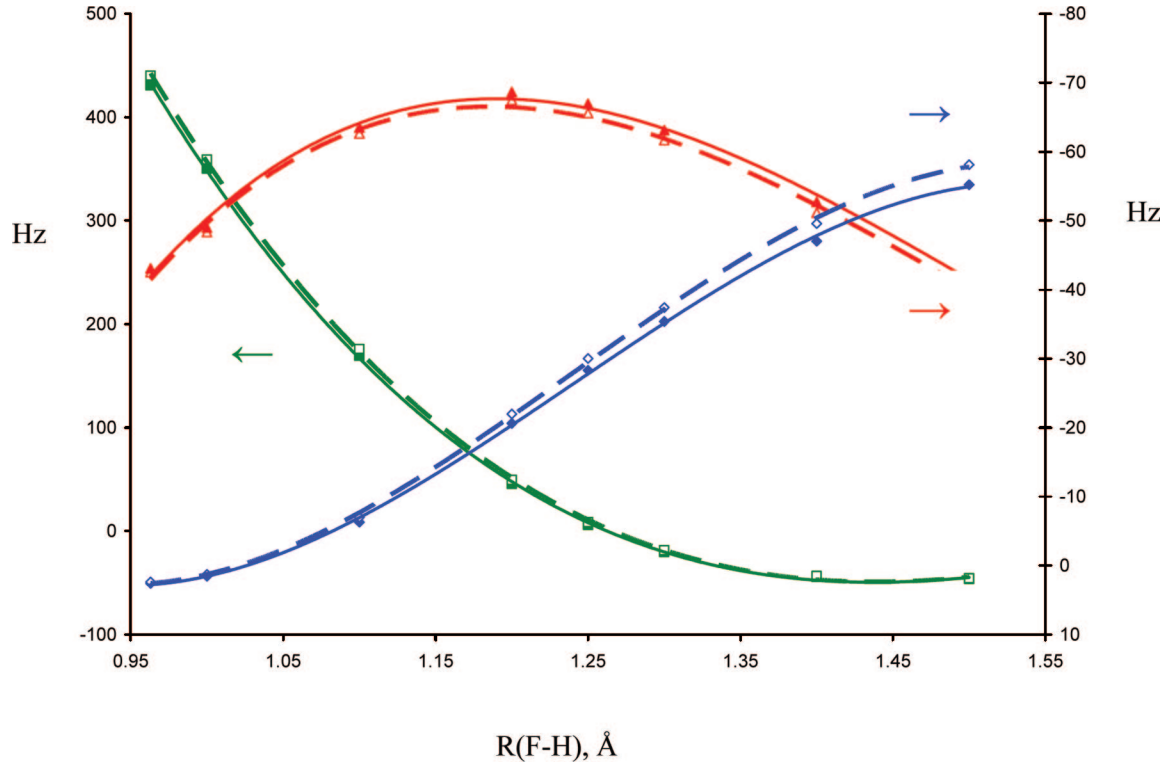


Figure 1. ^{1}J(F−H) (■), ^{1h}J(H−N) (♦), and ^{2h}J(F−N) (▲) across the proton-transfer coordinate for FH···NH_3. SOPPA (open symbols and dashed lines); EOM−CCSD (solid symbols and lines).

is to reduce this coupling constant by about 40 Hz, the reduced SOPPA value would be about 190 Hz, significantly less than the experimental value.

Changes in Coupling Constants along the Proton-Transfer Coordinate. Changes in one- and two-bond coupling constants for the FH:collidine complex as a function of temperature have been measured experimentally and attributed to proton transfer from F to N due to solvent ordering.[42,43] Such changes have been reproduced by EOM−CCSD calculations for proton transfer across F−H···N hydrogen bonds.[44] As part of the present investigation, SOPPA coupling constants along the proton transfer coordinate of F−H···NH_3 have also been evaluated, and these are compared with the EOM−CCSD coupling constants in Figure 1. The similarities between the two sets of data are striking. Both show that ^{2h}J(F−N) has its maximum absolute value in the proton-shared region of this coordinate; ^{1}J(F−H) decreases rapidly and changes sign, as observed experimentally, and ^{1h}J(H−N) asymptotically approaches the value for NH_4^+. Thus, the SOPPA and EOM−CCSD characteristics of coupling constants for proton transfer across the F−H···N hydrogen bond for F−H···NH_3 are very similar.

Conclusions

This study reports an evaluation of the performance of SOPPA and EOM−CCSD methods for determining spin−spin coupling constants. In particular, one-bond coupling constants in a series of small molecules and ions, and one- and two-bond coupling constants across hydrogen bonds in a set of neutral, cationic, and anionic complexes, have been compared. The following statements are supported by the computed results.

1. For isolated molecules, the one-bond SOPPA coupling constants ^{1}J(X−Y) involving C, N, O, and F have larger absolute values than the EOM−CCSD coupling constants, with the latter being in better agreement with available experimental data. The difference between SOPPA and EOM−CCSD tends to increase as the number of nonbonding electrons on the coupled atoms increases. SOPPA values for O−F coupling are significantly in error.

2. The absolute values of one-bond coupling constants ^{1}J(X−H) for the hydrides NH_3, H_2O, and FH and their protonated and deprotonated ions are also greater at SOPPA compared to EOM−CCSD. The difference between them is largest for F−H coupling.

3. One- and two-bond coupling constants ^{1}J(X−H), ^{1h}J(H−Y), and ^{2h}J(X−Y) across X−H···Y hydrogen bonds in neutral, protonated, and deprotonated complexes formed from the hydrides are similar at SOPPA and EOM−CCSD, with the largest differences occurring for ^{1}J(F−H) in complexes with F−H as the proton donor, and ^{2h}J(F−F) for $(FHF)^-$. Moreover, the signs of ^{1}J(X−H), ^{1h}J(H−Y), and ^{2h}J(X−Y) are the same at both levels of theory, as are their variations across the proton-transfer coordinate in F−H···NH_3.

On the basis of this study, it appears that SOPPA one-bond X−Y coupling constants, particularly those involving electronegative atoms, may not be reliable. However, SOPPA would appear to provide a reliable and more cost-effective approach for computing coupling constants across hydrogen bonds, although couplings involving F may be problematic. Further comparisons in other larger, more complicated systems should be made.

Acknowledgment. This work was carried out with financial support from the Spanish Ministerio de Educación y Ciencia (Project No. CTQ2007-61901/BQU) and Comunidad Autónoma de Madrid (Project MADRISOLAR, ref S-0505/PPQ/0225). This support and the continued support of the Ohio Supercomputer Center are gratefully acknowledged.

Supporting Information Available: SOPPA and EOM−CCSD PSO, DSO, FC, and SD components of J for all molecules, ions, and complexes investigated in this study. Also included are the full citations for refs 22 and 24. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Enevoldsen, T.; Oddershede, J.; Sauer, S. P. A. Correlated calculations of indirect nuclear spin-spin coupling constants using second-order polarization propagator approximations: SOPPA and SOPPA(CCSD). *Theor. Chem. Acc.* **1998**, *100*, 275.

(2) (a) Geertsen, J.; Oddershede, J.; Scuseria, G. E. Spin-spin coupling constants of CO and N_2. *J. Chem. Phys.* **1987**, *87*, 2138. (b) Oddershede, J.; Geertsen, J.; Scuseria, G. E. Nuclear spin-spin coupling constant of HD. *J. Phys. Chem.* **1988**, *92*, 3056.

(3) (a) Nielsen, E. S.; Jørgensen, P.; Oddershede, J. Transition moments and dynamic polarizabilities in a second order polarization propagator approach. *J. Chem. Phys.* **1980**, *73*, 6238. (b) Oddershede, J.; Jørgensen, P.; Yeager, D. L. *Comp. Phys. Rep.* **1984**, *2*, 33.

(4) Packer, M. J.; Dalskov, E. K.; Enevoldsen, T.; Jensen, H. J. Aa.; Oddershede, J. A new implementation of the second-order polarization propagator approximation (SOPPA): The excitation spectra of benzene and naphthalene. *J. Chem. Phys.* **1996**, *105*, 5886.

(5) Dalskov, E. K.; Sauer, S. P. A. Correlated, Static and Dynamic Polarizabilities of Small Molecules. Comparison of Four "Black Box" Methods. *J. Phys. Chem. A* **1998**, *102*, 5269.

(6) Barone, V.; Provasi, P. F.; Peralta, J. E.; Snyder, J. P.; Sauer, S. P. A.; Contreras, R. H. Substituent Effects on Scalar $^2J(^{19}F, ^{19}F)$ and $^3J(^{19}F,^{19}F)$ NMR Couplings: A Comparison of SOPPA and DFT Methods. *J. Phys. Chem. A* **2003**, *107*, 4748.

(7) Krivdin, L. B.; Contreras, R. H. Recent Advances in Theoretical Calculations of Indirect Spin-Spin Coupling Constants. *Annu. Rep. NMR Spectrosc.* **2007**, *61*, 133.

(8) Perera, S. A.; Sekino, H.; Bartlett, R. J. Coupled-cluster calculations of indirect nuclear coupling constants: The importance of non-Fermi contact contributions. *J. Chem. Phys.* **1994**, *101*, 2186.

(9) Perera, S. A.; Nooijen, M.; Bartlett, R. J. Electron correlation effects on the theoretical calculation of nuclear magnetic resonance spin-spin coupling constants. *J. Chem. Phys.* **1996**, *104*, 3290.

(10) Perera, S. A.; Bartlett, R. J. Predicted NMR Spectra for Ethyl Carbocations: A Fingerprint for Nonclassical Hydrogen-Bridged Structures. *J. Am. Chem. Soc.* **1995**, *117*, 8476.

(11) Perera, S. A.; Bartlett, R. J. Structure and NMR Spectra of the 2-Norbornyl Carbocation: Prediction of $^1J(^{13}C^{13}C)$ for the Bridged, Pentacoordinate Carbon Atom. *J. Am. Chem. Soc.* **1996**, *118*, 7849.

(12) (a) Provasi, P. F.; Aucar, G. A.; Sauer, S. P. A. The use of locally dense basis sets in the calculation of indirect nuclear spin-spin coupling constants: The vicinal coupling constants in H_3C-CH_2X (X = H, F, Cl, Br, I). *J. Chem. Phys.* **2000**, *112*, 6201. (b) Helgaker, T.; Watson, M.; Handy, N. C. Analytical calculation of nuclear magnetic resonance indirect spin-spin coupling constants at the generalized gradient approximation and hybrid levels of density-functional theory. *J. Chem. Phys.* **2000**, *113*, 9402. (c) Provasi, G. A.; Aucar, G. A.; Sauer, S. P. A. The effect of lone pairs and electronegativity on the indirect nuclear spin-spin coupling constants in CH_2X (X = CH_2, NH, O, S): Ab initio calculations using optimized contracted basis sets. *J. Chem. Phys.* **2001**, *115*, 1324. (d) Lantto, P.; Vaara, J.; Helgaker, T. Spin-spin coupling tensors by density-functional linear response theory. *J. Chem. Phys.* **2002**, *117*, 5998. (e) Bryce, D. L.; Wasylishen, R. E. Ab initio characterization of through-space indirect nuclear spin-spin coupling tensors for fluorine-X (X = F, C, H) spin pairs. *J. Mol. Struct.* **2002**, *602−603*, 463. (f) Filatov, M.; Cremer, D. Calculation of indirect nuclear spin-spin coupling constants within the regular approximation for relativistic effects. *J. Chem. Phys.* **2004**, *120*, 11407. (g) Ruden, T. A.; Helgaker, T.; Jaszuski, M. The NMR indirect nuclear spin-spin coupling constants for some small rigid hydrocarbons: molecular equilibrium values and vibrational corrections. *Chem. Phys.* **2004**, *296*, 53.

(13) Pople, J. A.; Binkley, J. S.; Seeger, R. Theoretical models incorporating electron correlation. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1976**, *10*, 1.

(14) Krishnan, R.; Pople, J. A. Approximate fourth-order perturbation theory of the electron correlation energy. *Int. J. Quantum Chem.* **1978**, *14*, 91.

(15) Bartlett, R. J.; Silver, D. M. Many-body perturbation theory applied to electron pair correlation energies. I. Closed-shell first-row diatomic hydrides. *J. Chem. Phys.* **1975**, *62*, 3258.

(16) Bartlett, R. J.; Purvis, G. D. Many-body perturbation theory, coupled-pair many-electron theory, and the importance of quadruple excitations for the correlation problem. *Int. J. Quantum Chem.* **1978**, *14*, 561.

(17) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self−Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian−Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1982**, *56*, 2257.

(18) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213.

(19) Spitznagel, G. W.; Clark, T.; Chandrasekhar, J.; Schleyer, P. v. R. Stabilization of methyl anions by first-row substituents. The superiority of diffuse function-augmented basis sets for anion calculations. *J. Comput. Chem.* **1982**, *3*, 363.

(20) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3−21+G basis set for first-row elements, Li−F. *J. Comput. Chem.* **1983**, *4*, 294.

(21) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571.

(22) Frisch, M. J. et al. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.

(23) Dalton-2, a molecular electronic structure program, Release 2.0, 2005. http://www.kjemi.uio.no/software/dalton/dalton.html (accessed Apr 2008).

(24) Stanton, J. F. et al. *ACES II, a program product of the Quantum Theory Project*; University of Florida: Gainesville, FL.

(25) (a) Kalinowski, H.-O.; Berger, S.; Braun, S. *Carbon-13 NMR Spectroscopy*; John Wiley and Sons: Chichester, U.K., 1988; p 549. (b) Kalinowski, H.-O.; Berger, S.; Braun, S. *Carbon-13 NMR Spectroscopy*; John Wiley and Sons: Chichester, U.K., 1988; p 570. (c) Kalinowski, H.-O.; Berger, S.; Braun, S. *Carbon-13 NMR Spectroscopy*; John Wiley and Sons: Chichester, U.K., 1988; p 577. (d) Kalinowski, H.-O.; Berger, S.; Braun, S. *Carbon-13 NMR Spectroscopy*; John Wiley and Sons: Chichester, U.K., 1988; p 496.

(26) (a) Berger, S.; Braun, S.; Kalinowski, H.-O. *NMR Spectroscopy of the Non-metallic Elements*; John Wiley and Sons: Chichester, U.K., 1997; p 386. (b) Berger, S.; Braun, S.; Kalinowski, H.-O. *NMR Spectroscopy of the Non-metallic Elements*; John Wiley and Sons: Chichester, U.K., 1997; p 277. (c) Berger, S.; Braun, S.; Kalinowski, H.-O. *NMR Spectroscopy of the Non-metallic Elements*; John Wiley and Sons: Chichester, U.K., 1997; p 245. (d) Berger, S.; Braun, S.; Kalinowski, H.-O. *NMR Spectroscopy of the Non-metallic Elements*; John Wiley and Sons: Chichester, U.K., 1997; p 587.

(27) Frankiss, S. G. Nuclear Magnetic Resonance Spectra of Some Substituted Methanes. *J. Phys. Chem.* **1963**, *67*, 752.

(28) Soloman, I. J.; Raney, J. K.; Kacmarek, A. J.; Maguire, R. G.; Noble, G. A. An Oxygen-17 and Fluorine-19 Nuclear Magnetic Resonance Study of the Oxygen Fluorides. *J. Am. Chem. Soc.* **1967**, *89*, 2015.

(29) Moy, D.; Yong, A. R. The Preparation of Fluorodiazonium Hexafluoroarsenate ($N_2F^+AsF^-$) from *cis*-Difluorodiazine. *J. Am. Chem. Soc.* **1965**, *87*, 1889.

(30) Alkorta, I.; Provasi, P. F.; Aucar, G. A.; Elguero, J. Spin-spin coupling constants in simple hydrides of the second and third periods. *Magn. Reson. Chem.* **2008**, *46*, 356.

(31) Otake, M.; Matsumura, C.; Morino, Y. Microwave spectra of nitrogen trifluoride in the excited vibrational states: Equilibrium structure. *J. Mol. Spectrosc.* **1968**, *28*, 316.

(32) Morino, Y.; Saito, S. Microwave spectrum of oxygen difluoride in vibrationally excited states; $\nu_1 - 2\nu_2$ Fermi resonance and equilibrium structure. *J. Mol. Spectrosc.* **1966**, *19*, 435.

(33) Jackson, R. H. The microwave spectrum, structure, and dipole moment of dioxygen difluoride. *J. Chem. Soc.* **1962**, 4585.

(34) (a) Herzberg, G. *Molecular Spectra and Molecular Structure*; D. Van Nostrand Co. Inc.: Princeton, NJ, 1967; p 585. (b) Herzberg, G. *Molecular Spectra and Molecular Structure*; D. Van Nostrand Co. Inc.: Princeton, NJ, 1967; p 619.

(35) (a) Ruden, T. A.; Lutnæs, O. B.; Helgaker, T.; Ruud, K. Vibrational corrections to indirect nuclear spin-spin coupling constants calculated by density-functional theory. *J. Chem. Phys.* **2003**, *118*, 9572. (b) Helgaker, T.; Jaszuński, M.; Pecul, M. The quantum-chemical calculation of NMR indirect spin-spin coupling constants. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, DOI: 10.1016/j.pnmrs.2008.02.002.

(36) Del Bene, J. E.; Elguero, J. Predicted signs of reduced two-bond spin-spin coupling constants ($^{2h}K_{X-Y}$) across X−H−Y hydrogen bonds. *Magn. Reson. Chem.* **2004**, *42*, 421.

(37) Del Bene, J. E.; Elguero, J. One-Bond Spin-Spin Coupling Constants of X−^{1}H Proton Donors in Complexes with X−H−Y Hydrogen Bonds, for X = ^{13}C, ^{15}N, ^{17}O, and ^{19}F: Predictions, Comparisons, and Relationships among $^1J_{X\text{-}H}$, $^1K_{X\text{-}H}$, and X-H Distances. *J. Am. Chem. Soc.* **2004**, *126*, 15624.

(38) Del Bene, J. E.; Elguero, J. Predicted Signs of One-Bond Spin-Spin Coupling Constants ($^{1h}J_{H\text{-}Y}$) across X-H-Y Hydrogen Bonds for Complexes with Y = ^{15}N, ^{17}O, and ^{19}F. *J. Phys. Chem. A* **2004**, *108*, 11762.

(39) Del Bene, J. E.; Jordan, M. J. T.; Perera, S. A.; Bartlett, R. J. Vibrational Effects on the F-F Spin-Spin Coupling Constant ($^{2h}J_{F\text{-}F}$) in FHF^- and FDF^-. *J. Phys. Chem. A* **2001**, *105*, 8399.

(40) Shenderovich, I. G.; Smirnov, S. N.; Denisov, G. S.; Gindin, V. A.; Golubev, N. S.; Dunger, A.; Reibke, R.; Kirpekar, S.; Malkina, O. L.; Limbach, H.-H. Nuclear Magnetic Resonance of Hydrogen Bonded Clusters Between F^- and $(HF)_n$: Experiment and Theory. *Ber. Bunsengens. Phys. Chem.* **1998**, *102*, 422.

(41) Benedict, H.; Shenderovich, I. G.; Malkina, O. L.; Malkin, V. G.; Denisov, G. S.; Golubev, N. S.; Limbach, H.-H. Nuclear Scalar Spin-Spin Couplings and Geometries of Hydrogen Bonds. *J. Am. Chem. Soc.* **2000**, *122*, 1979.

(42) Shenderovich, I. G.; Burtsev, A. P.; Denisov, G. S.; Golubev, N. S.; Limbach, H.-H. Influence of the temperature-dependent dielectric constant on the H/D isotope effects on the NMR chemical shifts and the hydrogen bond geometry of the collidine-HF complex in $CDF_3/CDClF_2$ solution. *Magn. Reson. Chem.* **2001**, *39*, S91.

(43) Golubev, N. S.; Shenderovich, I. G.; Smirnov, S. N.; Denisov, G. S.; Limbach, H.-H. Nuclear Scalar Spin-Spin Coupling Reveals Novel Properties of Low-Barrier Hydrogen Bonds in a Polar Environment. *Chem.—Eur. J.* **1999**, *5*, 492.

(44) Del Bene, J. E.; Bartlett, R. J.; Elguero, J. Interpreting ^{2h}J(F,N), ^{1h}J(H,N) and ^1J(F,H) in the hydrogen-bonded FH-collidine complex. *Magn. Reson. Chem.* **2002**, *40*, 767.

CT800111J

JCTC Journal of Chemical Theory and Computation

pH Dependence of a 3_{10}-Helix versus a Turn in the M-Loop Region of PDE4: Observations on PDB Entries and an Electronic Structure Study

Dandamudi Usharani,[†,‡] Palakuri Srivani,[§] G. Narahari Sastry,[§] and Eluvathingal D. Jemmis*[,‡]

School of Chemistry, University of Hyderabad, Hyderabad 500046 India, Department of Inorganic and Physical Chemistry, Indian Institute of Science, Bangalore 560012, India, and Molecular Modeling Group, Organic Chemical Sciences, Indian Institute of Chemical Technology, Tarnaka, Hyderabad 500007, India

Received October 6, 2007

Abstract: Available X-ray crystal structures of phosphodiesterase 4 (PDE 4) are classified into two groups based on a secondary structure difference of a 3_{10}-helix versus a turn in the M-loop region. The only variable that was discernible between these two sets is the pH at the crystallization conditions. Assuming that at lower pH there is a possibility of protonation, thermodynamics of protonation and deprotonation of the aspartic acid, cysteine side chains, and amide bonds are calculated. The models in the gas phase and in the explicit solvent using the ONIOM method are calculated at the B3LYP/6−31+G* and B3LYP/6−31+G*:UFF levels of theory, respectively. The molecular dynamics (MD) simulations are also performed on the M-loop region of a 3_{10}-helix and a turn with explicit water for 10 ns under NPT conditions. The isodesmic equations of the various protonation states show that the turn containing structure is thermodynamically more stable when proline or cysteine is protonated. The preference for the turn structure on protonation (pH = 6.5−7.5) is due to an increase in the number of the hydrogen bonding and electrostatic interactions gained by the surrounding environment such as adjacent residues and solvent molecules.

Introduction

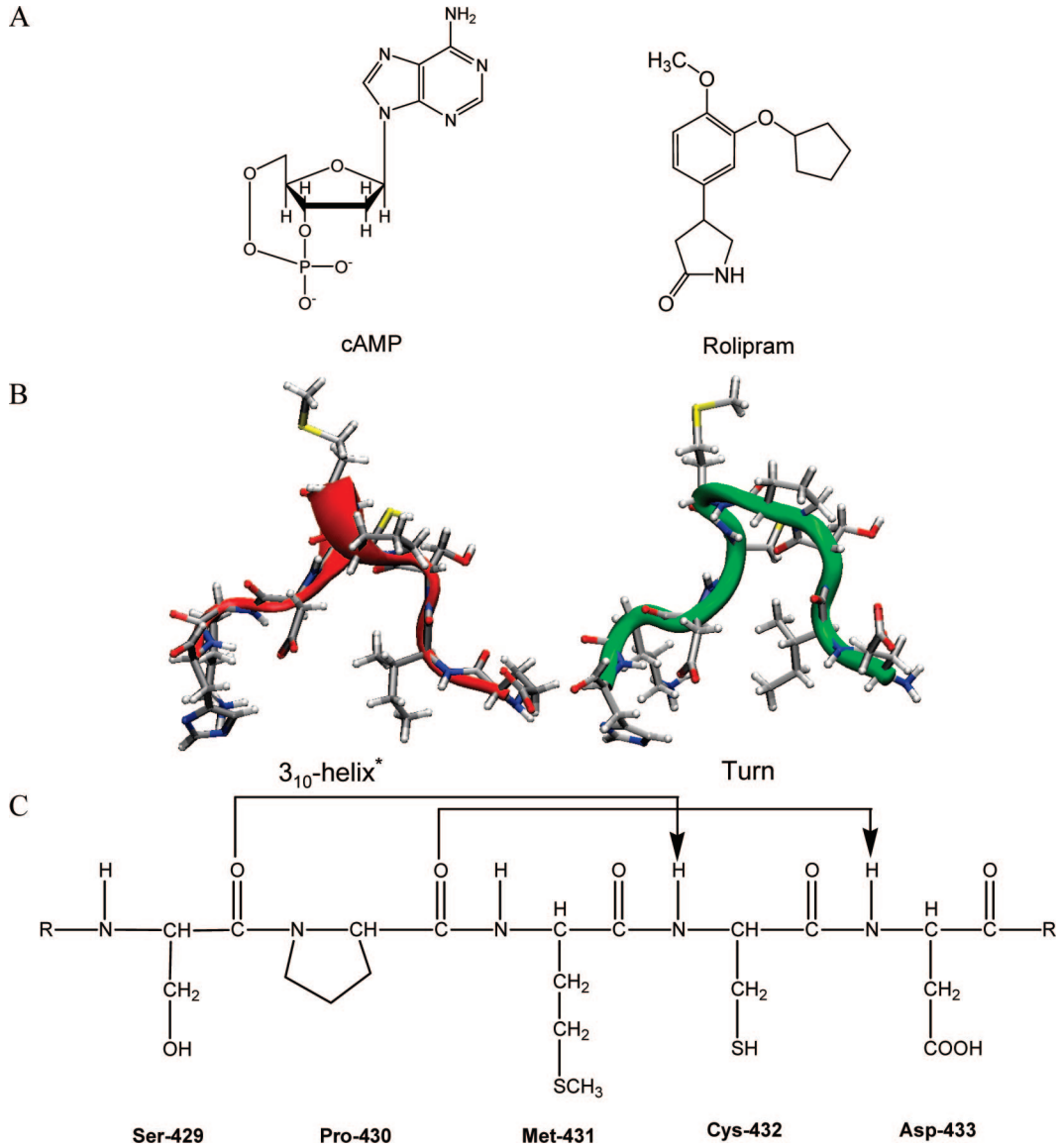
Phosphodiesterases (PDE) are metalloenzymes which hydrolyze the phosphodiesterase bond of cyclic adenosine 3′,5′-monophosphate (cAMP) and guanosine 3′,5′-monophosphate (cGMP) into the corresponding 5′-nucleotides (AMP and GMP) in various cells.[1a,b] The second messenger (cAMP and cGMP) concentration affects the specific protein phosphorylation cascades. Hence, these isozymes play a vital role in the regulation of various physiological functions like visual response, smooth muscle relaxation, platelet aggregation, immune response, cardiac contractibility, etc. They are the therapeutic targets for cardiovascular, inflammatory, and erectile dysfunction diseases.[1c,d] Among the 12 PDE isozymes, PDE4 and PDE5 have received much attention in the recent years.[1–3] The PDE4 catalyze specifically cAMP (Scheme 1A). The abundance of PDE4 in various inflammatory cells such as eosinophils, T cells, B cells, and neutrophils made it a keen target for the inflammatory diseases such as asthma, chronic obstructive pulmonary disease (COPD), allergic rhinitis, type II diabetics, and rheumatoid arthritis.[2b,c]

There are various classes of novel orally active PDE4 inhibitors discovered in this decade. Based on the structural motifs, PDE4 inhibitors can be broadly classified into three categories as xanthines, catechol ethers, and hetrocyclics (nitraquanzone, benzofurans, indoles, isoquinoline, pyridopyrimidinones, pyrazolepyridines, etc). Rolipram that belongs to catechol ether type has been reported as the first selective PDE4 inhibitor (Scheme 1A). The first-generation PDE4 inhibitors rolipram, RO-20-1724 (mesopram), and

* Corresponding author e-mail: jemmis@ipc.iisc.ernet.in.
[†] University of Hyderabad.
[‡] Indian Institute of Science.
[§] Indian Institute of Chemical Technology.

10.1021/ct700261b CCC: $40.75
Published on Web 05/15/2008

Scheme 1. A) Structures of Substrate cAMP and Inhibitor Rolipram of the PDE4, B) Secondary Structure of PDE4 in the M-Loop Region Existing as a 3_{10}-Helix (Red Color) and a Turn (Green Color),[a] and C) Schematic 2-D Representation of the 3_{10}-Helix Conformation[b,c]



[a] These are shown along with the amino acids represented as a capped stick model. [b] The hydrogen bonds are shown as lines and residue numbering is given for PDE4B. [c] The 3_{10}-helix is formed when the N−H group of an amino acid forms a hydrogen bond with the C=O group of the amino acid three residues earlier ($i+3 \rightarrow i$ hydrogen bonding).[13]

zardaverine are discarded from clinical trials due to the side effects like nausea, vomiting, dyspepsia, headache, and emesis.[2d] The second-generation rolipram analogs (cilomilast, roflumilast, and most potent piclamilast) that are now in clinical trials are also suffering from low therapeutic ratio.[2e] Among the various strategies for overcoming the side effects, subtype selectivity of PDE4 is a focus of investigation.[1,2]

Lack of the 3-dimensional structure of a target protein is usually a bottleneck for rational drug design. In the last couple of years there has been a dramatic increase in the number of crystal structures of PDE4 class of enzymes in the protein data bank (PDB). Despite the availability of 39 X-ray structures along with cocrystals for PDE4, progress toward drugs without undesired side effects has been marginal. Our interest in the modeling of the subtype selective PDE4 drug candidates brought us to a close examination of these crystal structures. We found that there is a characteristic secondary structural conformation difference in the PDE4 crystal structures.[4–10] Out of the 35 PDE4B and 4D crystal structures 12 of them have a 3_{10}-helix and the rest have a turn in the M-loop region[6b] which is near to the active site pocket. The four crystal structures, one for each of the PDE4 subtypes are yet to be deposited.[11] The distinct pattern of a secondary structure in the highly flexible M-loop region is prominent in these crystal structures (Scheme 1B). According to the standard protein structural rules, three and a half residues are required per turn of an α-helix. The hydrogen of the amide group of the first residue forms a hydrogen bond with the oxygen of the amino-terminal peptide bond of the fourth residue in an α-helix, whereas in a 3_{10}-helix it forms with the third residue.[12,13] Here we would like to know the factor responsible for this secondary structural change in the M-loop region. The methionine of the M-loop region has a hydrophobic interac-

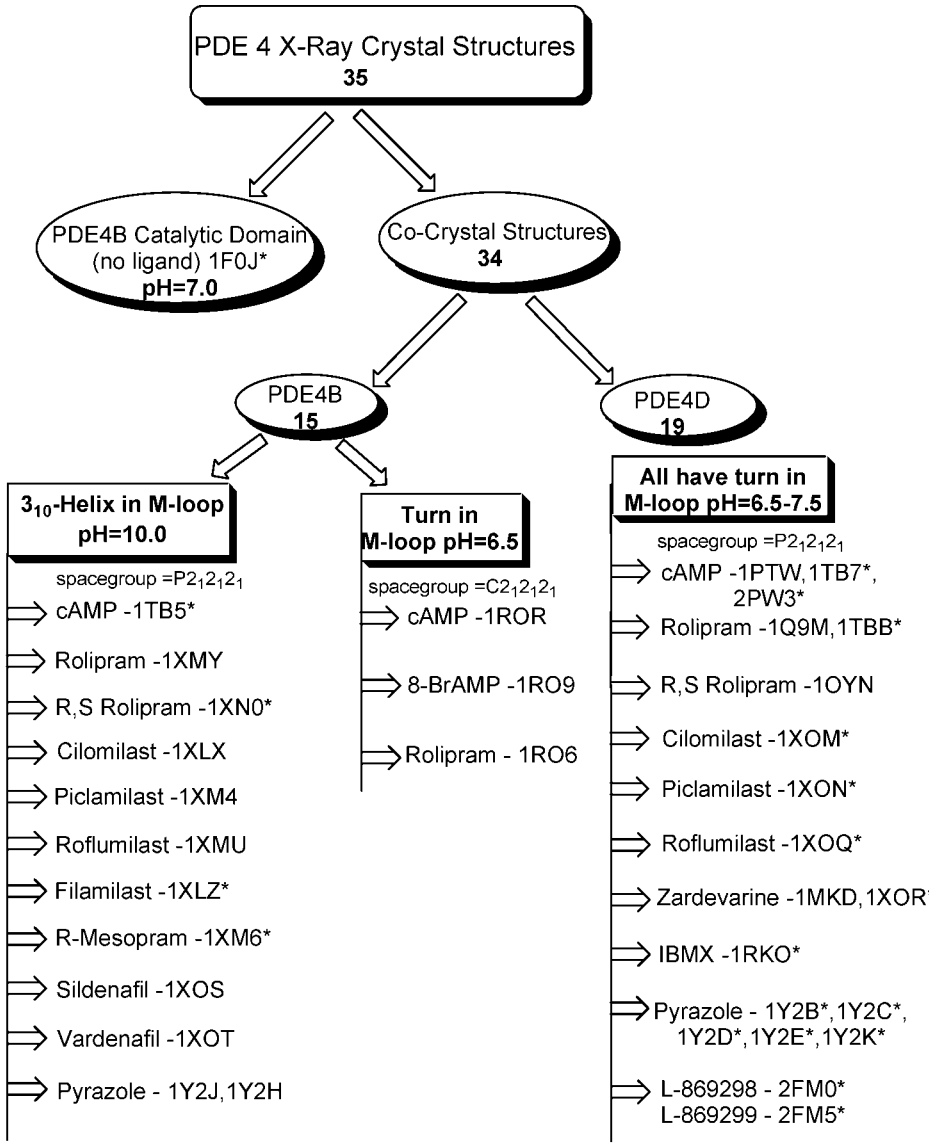


Figure 1. Details of the available X-ray crystallographic structures of PDE4 with various ligands deposited in the Protein Data Bank.

tion with inhibitors such as rolipram analogs[6a,8b] and exists in various conformations with the NVP inhibitor.[11] Is this structural change a determining factor for the subtype selectivity? We identify that an experimental variable pH seems to be controlling this structural variation rather than the sequential differences in the PDE4. The influence of pH is studied by calculating the various protonation states of side chains or amide bonds of these structures in the gas and explicit solvent phases. An electronic structure explanation is given for this observation.

Computational Details

Root Mean Square Deviation (rmsd). Residue-by-residue rmsd is calculated using a FORTRAN program.[14] RMSDs are calculated for the backbone residues of all 4B and 4D crystal structures having similar ligands. A molecular operating environment (MOE) homology module[15] is used for structural alignment of 4B and 4D structures. The rmsd values are calculated using the following equation where d_i is the distance between the ith atoms and n is the number of such distances.

$$\mathrm{RMSD} = \sqrt{\sum_{i}^{n} d_i^2 / n}$$

Electronic Structure Calculation. We have used ab initio calculations on the model structures, generated from the fragment SPMCD (Ser, Pro, Met, Cys, Asp) residue sequence of PDE4B and PDE4D PDB coordinates. The resulting dangling valencies are saturated by appropriately placing hydrogens. For uniformity and to average out the specific differences in the geometries arising from the process of formation of the crystals, only the hydrogen atom positions in each of the structures are optimized. The dihedral angles are kept constant during this process. The protonated and

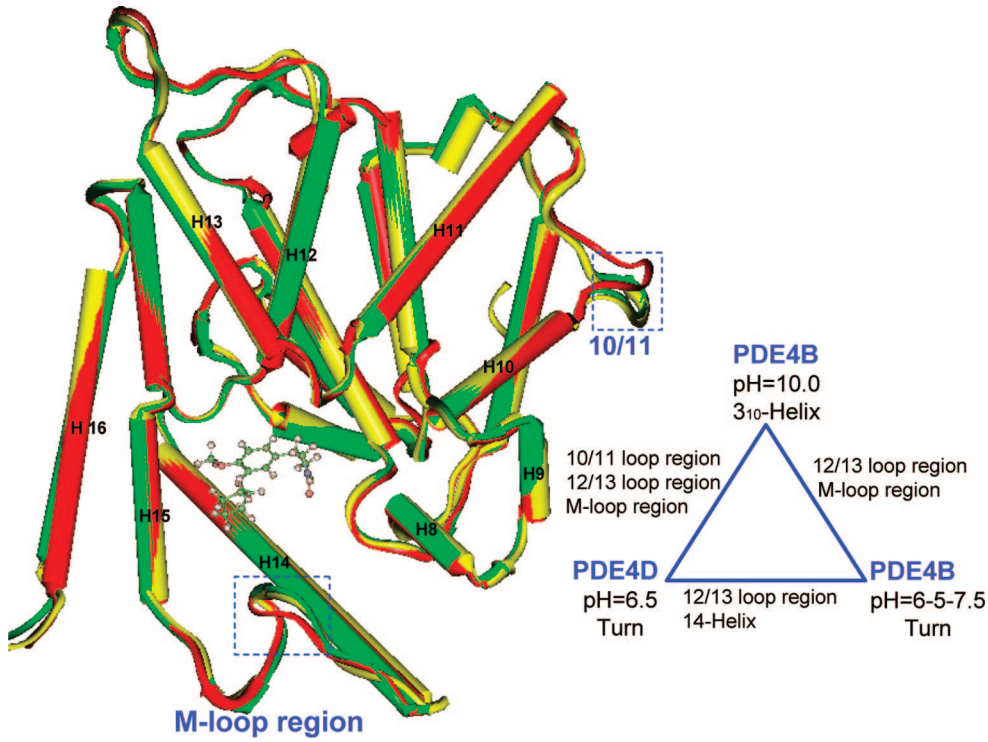


Figure 2. The superimposition of the α-carbon of the turn having PDE4B and PDE4D and a 3_{10}-helix having PDE4B structures with rolipram are shown with the yellow, green, and red color secondary structure at pH = 6.5, 7.0, and 10.0, respectively (left side). The ball and stick model represents the rolipram ligand in the active site. A schematic representation of structural differences with all the other similar ligands is also shown (right side).

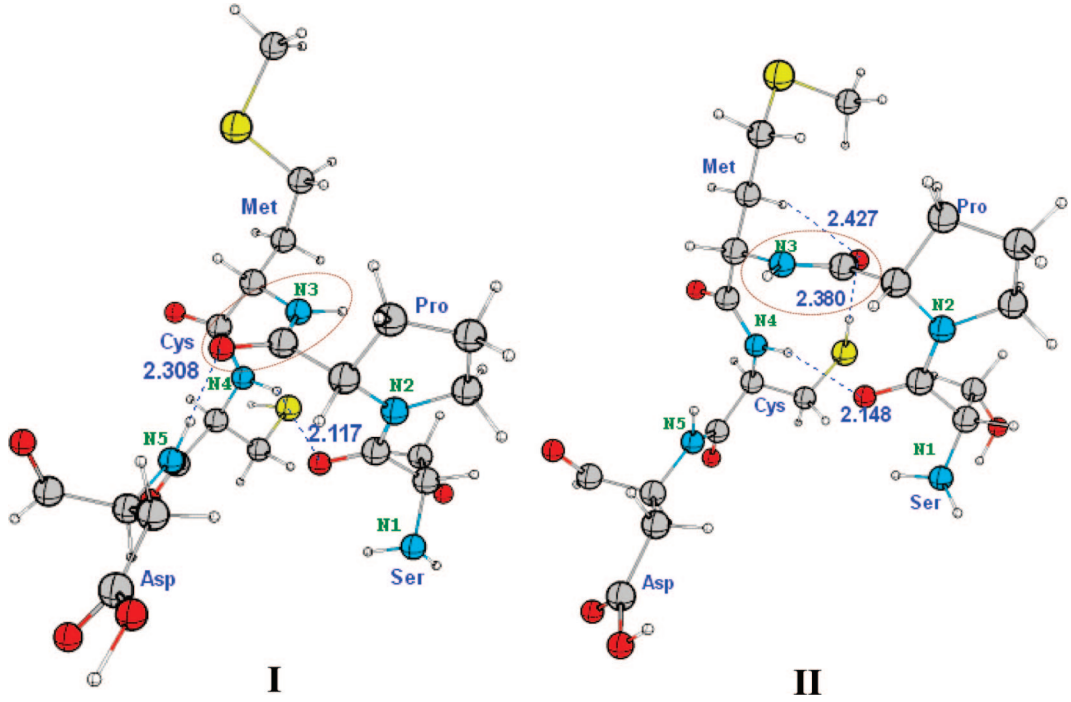


Figure 3. The optimized conformations, **I** (3_{10}-helix) and **II** (turn), of SPMCD residues at pH = 10.0 and pH = 6.5−7.5, respectively.

deprotonated states of aspartic acid and cysteine and amide bonds are calculated for seven sets of ligands. All electronic structure calculations are performed at the B3LYP/6−31+G* level of theory.[16]

The gas phase optimized model structures 1TB5 having a 3_{10}-helix and 1TB7 having a turn are further taken, and a layer of 5 Å thickness of water molecules is added using the "soak" option in InsightII.[17] The total number of water molecules present around the pentapeptide−protein structure is 119. The water molecules of the proteins are generally treated as low layer with molecular mechanics in QM/MM calculations to reduce the computational cost. The ONIOM[18,19]

Table 1. Reaction Energies in kcal/mol for Eq 1

ligand	I	II	ΔH
cAMP	1TB5	1ROR	−4.00
	1TB5	1TB7	−8.75
rolipram	1XMY	1RO6	2.78
	1XMY	1TBB	−0.13
R,S-rolipram	1XN0	1OYN	−9.68
cilomilast	1XLX	1XOM	−10.78
roflumilast	1XMU	1XOQ	−14.30
piclamilast	1XM4	1XON	−9.82
pyrazole	1Y2J	1Y2K	−4.48

Table 2. Reaction Energies in kcal/mol for Eq 2

ligand	I	II	ΔH
cAMP	1TB5	1ROR	−20.80
	1TB5	1TB7	−10.87
rolipram	1XMY	1RO6	−39.32
	1XMY	1TBB	−24.68
R,S-rolipram	1XN0	1OYN	−23.36
cilomilast	1XLX	1XOM	−19.79
roflumilast	1XMU	1XOQ	−23.80
piclamilast	1XM4	1XON	−15.21
pyrazole	1Y2J	1Y2K	−16.72

calculation is implemented for this system to analyze the explicit solvent effect on the structures. A two layer ONIOM calculation is adopted for the system, where the pentapeptide sequence SPMCD is defined quantum mechanically (high layer) with the B3LYP/6−31+G* level of theory and the water molecules are defined as a low layer molecular mechanics part with a universal force field.[20] The ONIOM energy of the system is given as

$$E^{\mathrm{ONIOM(QM:MM)}} = E^{\mathrm{QM}}_{\mathrm{model}} + E^{\mathrm{MM}}_{\mathrm{real}} - E^{\mathrm{MM}}_{\mathrm{model}} = E^{\mathrm{high}}_{\mathrm{model}} + E^{\mathrm{low}}_{\mathrm{real}} - E^{\mathrm{low}}_{\mathrm{model}}$$

These calculations are done using the Gaussian03 program package.[21]

Molecular Dynamics Simulations. The simulations are performed using the PMEMD module of AMBER9 version[22] in the IBM BLUEGENE/L machine. The deposited PDB structures 1TB5 and 1TB7 are taken for the molecular simulations. The coordinates of the M-loop region (423−440) of these protein structures are taken. The FF03 force field[23] is employed. The LEAP module is employed to construct a truncated octahedron solvate box (bcc) of cell length 56.46 Å with TIP3P water molecules around the protein. The solvation shell around the protein is 12 Å. The total number of solute and solvent atoms present around the 1TB5 and 1TB7 are 12578 and 12932, respectively. The protein−solvent system is minimized with 1000 steps of steepest descent and 2000 steps of conjugate gradient method. The protein is fixed during minimization using a harmonic constraint with a force constant of 500 kcal/mol/Å^2. This is done to remove any close contacts existing in the water shell with respect to protein. The whole system is also minimized after this for 200 steps steepest descent and 800 steps conjugate gradient method, so that the hydrogens of protein are optimized. The minimized structure obtained is then gradually heated up (over 40 ps) from 0 K to 300 K with harmonic constraints on solute using the SHAKE method. The dynamics is maintained under constant pressure-constant temperature (NPT) conditions using Berendsen (weak-coupling) temperature with a time step of 2 fs.[24] Finally, a production dynamics of 10 ns with protein is fixed under NPT conditions with a time step of 1 fs. The long-range interactions are calculated by the particle mesh Ewald (PME) method.[25] The grid size of 60 × 60 × 60 with grid spacing of 1.0 Å with a direct sum tolerance of 0.00001 and 9.0 Å cutoff is used. The cubic B-spline of fourth order is interpolated in PME. The resulting trajectories of molecular simulations are analyzed using the PTRAJ module of AMBER 8.0.

Results and Discussion

(a) Crystal Structure Analysis. There are about 60 PDE crystal structures in the PDB and 35 of these belong to the PDE4 class with various substrates such as cyclic adenosine monophosphate (cAMP), AMP, and 8-BrAMP and inhibitors such as rolipram analogs (rolipram, cilomilast, piclamilast, roflumilast, mesopram, and filamilast) and pyrazole analogs (Figure S1).[4−10] PDE4 is coded by four genes named as A, B, C, and D and are called subtypes. Among the four genes coding for PDE4, about 16 4B and 19 4D structures are crystallized (Figure 1). There is one each for PDE4 subtype with the NVP ({4-[8-(3-nitrophenyl)-[1,7]napthyridin-6-yl]benzoic acid}) inhibitor containing structures which are yet to be deposited.[11] The catalytic domain of the first crystal structure available for PDE4B contains 376 residues and explains the general architecture of PDE4.[4] The PDE4 has basically three subdomains comprising 17 α helices and a β hairpin. The catalytic pocket contains two divalent metal ions (Zn and Mg) that are crucial in hydrolysis of the substrate. Zn in the active site pocket has distorted trigonal bipyramid geometry, coordinating with Asp^{275}, Asp^{392}, His^{233}, His^{274}, and a water molecule of the protein.[13] On the other hand Asp^{275} of the protein and five water molecules surround the Mg atom. Furthermore the cocrystals of PDE4 with various ligands also have similar three subdomains comprising 16 α helices and a β hairpin (Figure 2).[5−10] Other than the conservation of helices and hairpin we noticed a surprising characteristic feature in these 35 structures: 12 structures have a 3_{10}-helix in the M-loop region. The remaining 23 of them have a turn in the M-loop region,[6b] 19 of these 23 belong to the PDE4D, and 4 belong to PDE4B (Figure 1).

(b) Root Mean Square Deviation (rmsd). We superimposed all structures having the 3_{10}- helix and the turn with seven similar sets of ligands, and the rmsd of residue by residue for the backbone is calculated.[14] There are two structures each for 4B and 4D with cAMP, three each with rolipram, and one each with cilomilast, piclamilast, roflumilast, and pyrazole analogs (Figure 1). The superimposition of 4B containing a 3_{10}-helix and 4D with similar ligand structures showed a major difference in three loop regions i.e., 10/11, 12/13, and the M-loop region (Figure 2). The residues with varying rmsd are Glu^{317} and Glu^{318} of helix 10/11 (316−319) which interacts with upstream conserved region (UCR)[1] UCR1 and UCR2 present in the regulatory domain (N-terminal) of the protein. The residues Ser^{368} and Ser^{369} of helix 12/13 (367−377) belonging to an extracellular signal-regulated kinases (ERK) docking site[1] that has

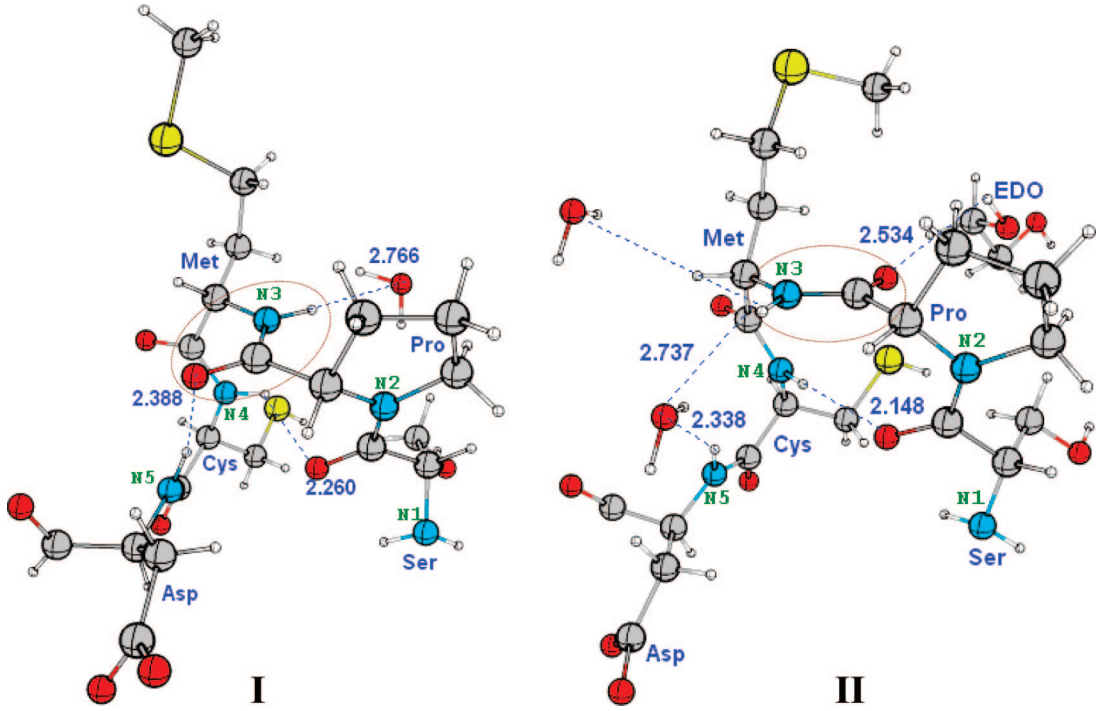


Figure 4. The X-ray crystal structural conformations of **I** (3_{10}-helix 1TB5) and **II** (turn 1TB7) at pH = 10.0 and pH = 6.5−7.5, respectively. In the SPMCD residue sequence the hydrogen bonding interaction of the amide bond with solvent molecules such as water and EDO is also shown.

protein−protein interactions is also altered. In the case of the M-loop region[6b] (424−437) Pro^{430} and Met^{431} residues are altered in the backbone of protein structure (Figures S2−S4). The extent of deviation in the ERK docking site varies slightly with the ligands.

On the other hand the superimposition of the 4B and 4D structures containing a turn with cAMP and rolipram shows a difference only in the 12/13 loop region (ERK docking site) and in helix 14, particularly from the residues Lys^{422} to Glu^{427} that are oriented toward the solvent region (Figures S2 and S3). The superimposition of all the 4B with cAMP and rolipram structures has a difference in 10/11 and the M-loop region (Figure 2). In spite of the major homology sequence similarity between 4B and 4D subtypes the common differences are found in the 12/13-loop region and either at the starting of the M-loop region or at the Pro^{430} residue. Thus the PDE4 structures can be divided into the two groups: one with the 3_{10}-helix and the other with the turn varying at 10/11 and in the M-loop region (Figure 2).

(c) Importance of Secondary Structural Change and Its Cause. In view of the important role of the secondary structural alterations in specific enzymatic functions of PDE5,[26] the M-loop region in PDE4 is scrutinized because the methionine residue here has hydrophobic interactions with all rolipram analogs.[6a,8b] The inhibitor selectivity of rolipram analogs to PDE4 is due to the hydrophobic interactions at the Q2 pocket, where methionine of the M-loop region is a constituent. The PDE 4A, 4B, 4C, and 4D with NVP, a subtype selective inhibitor also have shown large conformational changes of methionine having relatively high B-factors.[11] The conformational changes of methionine thus can have an impact on inhibitor binding. The analysis of these variations in PDE4 may help to understand the influence of the M-loop region in inhibitor selectivity.

The secondary structural difference of PDE4 crystals cannot be explained by the ligand binding or crystal packing or mutated residues. This is because there is a structural variation of a 3_{10}-helix and a turn in the M-loop region within PDE4B structures with the similar ligands and the same space group (Figure 1). The SPMCD residue sequence is conserved in all subtypes and hence cannot account for this structural difference. After an extensive search, the only difference we could find between these two sets of structures is the pH maintained during crystallization. Without any exception it was found that all the PDE4B structures having a 3_{10}-helix are crystallized at pH=10.0 from the polar solvents such as aqueous $LiSO_4$ and NH_4SO_4 solution. In contrast, all the turn containing 4B and 4D structures are obtained at the pH = 6.5−7.5 range with polyethylene glycol (PEG) conditions (Figure 1). Though we do not know the reason for the use of different pHs in these experiments, the results are surprising because in these examples the pH appears to dominate over the differences in sequences of PDE4B and 4D.

(d) Influence of pH on the M-Loop Region. As a first approximation we assume that the pH dependence on the different conformations is caused by the loss or gain of a proton that leads to ionizations of carboxylates and amines.[27] The changes in the electrostatic environment of PDE4B and 4D may alter the conformations and finally affect the secondary structure of the protein. The influence of the pH on the structure is known for lysozymes, plastocyanin, γ-chymotrypsin, azurin, and insulin.[28] Another example is of cubic insulin crystals that manifest alterations in the

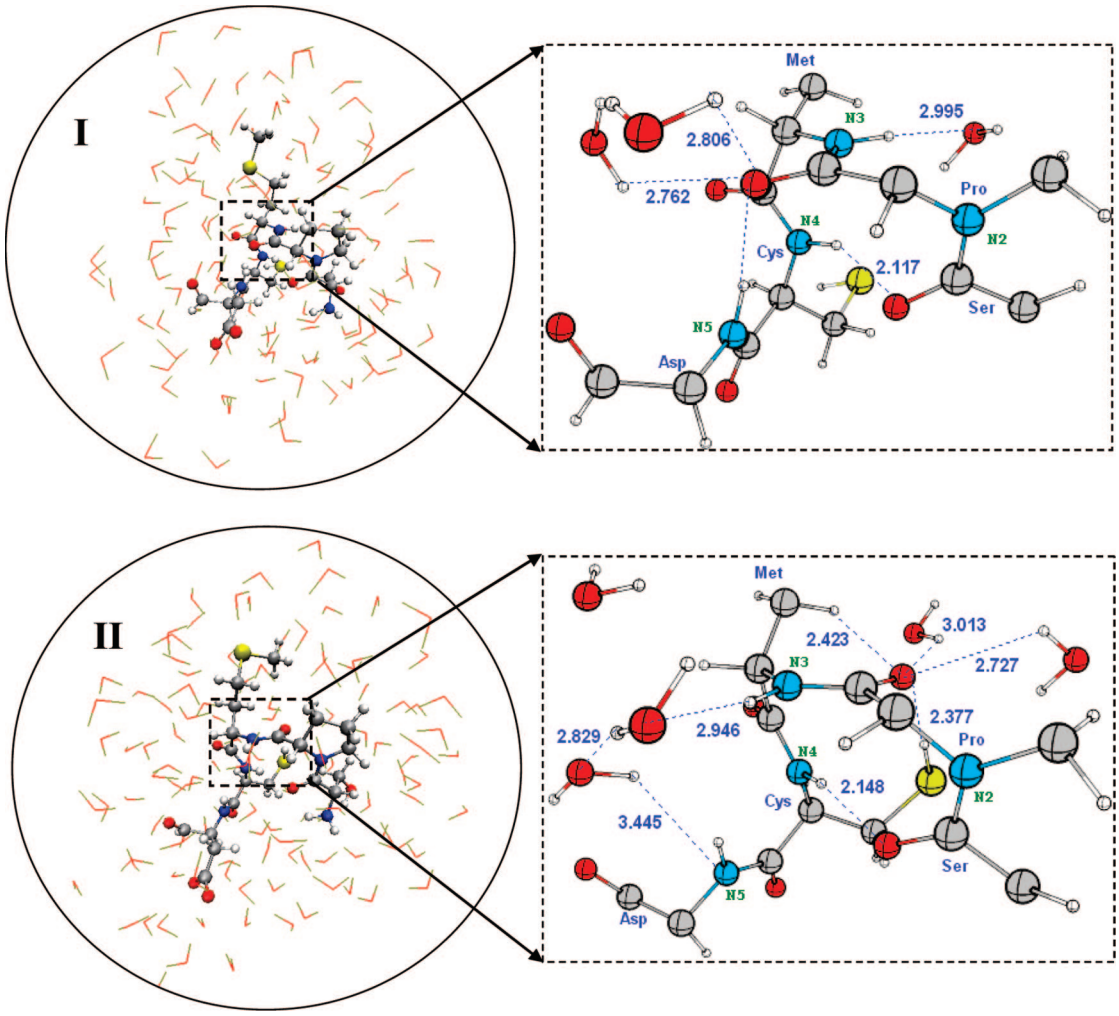


Figure 5. The ball and stick model represents the pentapeptide (SPMCD) that is considered the high layer QM region, and the shell with water molecules (lines) is considered as the low layer MM region. The optimized geometry of **I** and **II** and their hydrogen bonding network with the solvent−water molecules are also shown.

Table 3. Summary of a Number of Water Molecules Present Around PDE4B (1TB5) and PDE4D (1TB7) Structures

	no. of water molecules near to CO group of proline			no. of water molecules near to NH group of methionine		
structure	crystal	ONIOM	MD(10 ns)	crystal	ONIOM	MD(10 ns)
1TB5 (PDE4B)	0	2	1−2	1	2	0−2
1TB7 (PDE4D)	1 EDO	2	1−2	2	2−3	2−4

conformations of side chains of acidic residues as a function of the pH ranging from 7 to 11.[28] Such an alteration in the protonation states of acidic residues influencing the enzymatic role is well-known for the cytochrome P-450s.[29] The difference in the conformation around the 10/11 loop region of PDE4 as a function of the pH might arise from the repulsion between the side chains of two glutamic acid residues that are adjacent to each other.

In the M-loop region the difference in the secondary structure is observed at the backbone of the SPMCD residue sequence that has conformation **I** in the 3_{10}-helix of PDE4B and **II** in all the turn containing PDE4B and 4D structures (Figure 3). Structures **I** and **II** have a conserved hydrogen bond between the CO of serine and the NH of cysteine. These two structures have a difference in the orientation of the carbonyl group of the proline in the amide bond. This difference of the carboxy group orientation of the proline helps to gain extra hydrogen bonding with the NH group of the aspartic acid resulting in a twist and forming a 3_{10}-helix structure in **I** (Figure 3, Scheme 1C). Generally the constrained geometry of the proline initiates the turning of a protein strand, while here it brings a 3_{10}-helix structure of the conformation I.

The influence of pH can lead to the existence of various protonation states of side chains or amide bonds in structures **I** and **II**. The aspartic acid and cysteine side chains are generally known to be protonated in various protein structures.[28−30] There are some examples of backbone amide bonds that undergo protonation when they are involved in interactions with a metal as in belomycin.[29c] So, we anticipate that in the SPMCD residue sequence depending on pH, aspartic acid or cysteine side chains or amide bonds can be

protonated. We simulate different pH conditions by adding or removing a proton of side chains or amide bonds. The energies of the protonated and deprotonated states of aspartic acid, cysteine and amide bonds in all seven sets of similar ligands having a "3_{10}-helix" and a "turn" containing structures are calculated. The protonation and deprotonation states of aspartic acid residue do not explain the existing structural variations (S1). The isodesmic equation 1 shows the energetics of the protonation and the deprotonation of the cysteine residue. Except in rolipram ligand containing structures, the formation of the protonation state of cysteine in structure **II** and the deprotonated state in structure **I** are exothermic in the gas phase (Table 1).

$$\mathrm{I} + \mathrm{II}_{\mathrm{deprocys}} \rightarrow \mathrm{I}_{\mathrm{deprocys}} + \mathrm{II} \qquad (1)$$

The exothermicity of the reaction is due to the cysteine side chain (SH) in the protonated state gaining a weak hydrogen bond with the CO group of the proline in structure **II**, whereas in the deprotonated state it is missing (Figure 3). On the other hand in structure **I,** the deprotonated state of the cysteine side chain can have weak electrostatic interactions with the NH group of methionine.[31] The protonated state of the cysteine side chain in **II** and the deprotonated state in **I** have gained a noncovalent interaction each justifying their thermodynamic stabilities.

The feasibility of protonation of backbone amide bonds is also studied. Even though the differences of pH by three units may not lead to the protonation and deprotonation of amide bonds of PDE4 under experimental conditions, we anticipate that the protonation and deprotonation of N2 and N3 (Figure 3) of structures **I** and **II** would be an indicator of the influence of the pH. The deprotonation and protonation are done at N3 (NH of methionine) and N2 (N of proline) of the amide bonds, respectively, where the change of conformation is observed (Figure 3). Reaction energies of the isodesmic eq 2 for seven sets of similar ligands in their protonated and deprotonated states indicate that **II** is more stable in the protonated form and **I** in the deprotonated form (Table 2). The isodesmic eq 5 shows that all turn containing PDE4 (**II**) on protonation at N2 are stable over 3_{10}-helix containing structures (**I**), and in isodesmic eq 6 the reverse is observed for deprotonation at N3 (Tables S2 and S3).

The order of stability of the protonated state in all **II** and deprotonated form in **I** is attributed to the increase in the number of hydrogen bonding interactions. In structure **II,** the CO of the proline gains a 1,4 interaction with N2 on protonation and a weak interaction with the CH and SH of methionine and cysteine side chain residues.[31,32] On the other hand the deprotonation of N3 leads to a weak C−H−N3 interaction[33] in **I** similar to the C−H−O interaction seen in **II** (Figure 3). The thermodynamic stability of the protonated and the deprotonated states in structures **I** and **II** depends on the gain in noncovalent interactions such as hydrogen bonds and electrostatic interaction with the surrounding environment. The exothermicity of the backbone amide bond protonation is quite higher than cysteine protonation.

$$\mathbf{I}_{\mathrm{pro}} + \mathbf{II}_{\mathrm{depro}} \rightarrow \mathbf{I}_{\mathrm{depro}} + \mathbf{II}_{\mathrm{pro}} \qquad (2)$$

The extent of noncovalent interactions in structures **I** and **II** can also differ with respect to the surrounding solvent molecules. The 35 crystal structures are thus thoroughly examined near the M-loop region to know the influence of the water molecules. In all the crystal structures with respect to the methionine of the M-loop region a sphere radius of 10.0 Å is analyzed. There are 22 turn containing structures crystallized in PEG conditions. We found that the turn containing (**II**) PDE4D structures with resolution lower than 2.10 Å have ethane diol (EDO) and water molecules in this region (Table S4). The carbonyl group of the proline of structure **II** has a hydrogen bonding interaction with the EDO or water molecule. The amide group of the methionine of structure **II** has a hydrogen bonding network with the NH of aspartic acid through two water molecules (Figures 4 and S5). These additional interactions with the solvent thus stabilize structure **II**. The presence of EDO near the carbonyl group of proline in these structures may be due to PEG crystallization conditions, while such a type of EDO or water molecules are not found at the carbonyl group of proline in the M-loop region of four PDE4B (1F0J, 1ROR, 1RO6, and 1RO9) and in high resolution PDE4D turn containing structures, even though they have similar crystallization conditions (Table S4). Even the number of water molecules varies from 1−2 in the two chains (A and B) of the same crystal. This is also true in other crystals considered here (Table S4). It is possible to identify a water molecule even in the 12 low resolution structures with a 3_{10}-helix showing hydrogen bonding interaction to the amide group of methionine and carbonyl group of proline (Figures 4 and S5). The number of solvent molecules present around the M-loop region in these crystal structures is varied (Table S4).

The irregularity in the number and position of the water and the EDO molecules in the crystal structures led us to study the influence of the solvent molecules on the structures. We model this by constructing a layer of 5 Å thickness of water molecules around the pentapeptide SPMCD residue sequence of structures **I** and **II**. This model system with various protonation states of the cysteine and the amide bonds of 1TB5 and 1TB7 are studied using the QM/MM ONIOM method. In this model the pentapeptide is defined as the QM region and water molecules in the MM region (Figure 5). The ONIOM energies of the protonated and the deprotonated states of the cysteine and the amide bond (N2 and N3) in structures **I** and **II** along with a water layer are given in isodesmic eqs 3 and 4, respectively.

$$\mathbf{I} + \mathbf{II}_{\mathrm{deprocys}} \rightarrow \mathbf{I}_{\mathrm{depro\ cys}} + \mathbf{II} \quad \Delta \mathrm{E}^{\mathrm{ONIOM}} = -6.55\ \mathrm{kcal/mol} \qquad (3)$$

$$\mathbf{I}_{\mathrm{pro}} + \mathbf{II}_{\mathrm{depro}} \rightarrow \mathbf{I}_{\mathrm{depro}} + \mathbf{II}_{\mathrm{pro}} \quad \Delta E^{\mathrm{ONIOM}} = -10.06\ \mathrm{kcal/mol}$$
$$(\text{where } \mathbf{I} = 1\mathrm{TB5} \text{ and } \mathbf{II} = 1\mathrm{TB7}) \qquad (4)$$

The ONIOM calculation in explicit solvent shows the reaction is exothermic and similar as in the gas phase reaction (isodesmic eqs 1 and 2). The water molecules form a hydrogen bond with the NH group of methionine and the CO group of proline in structures **I** and **II**. The gain in the hydrogen bonding interaction of the amide bond in structure **II** makes it more stable than **I**. The hydrogen bonding

network of water with the residues in the crystal structures is reproduced by ONIOM calculations (Figures 4, S5, and 5).

To reinvestigate the average number of water molecules that can exist near the amide bond, the CO group of the proline, and the NH group of the methionine, MD simulation on M-loop region of a 3_{10}-helix (1TB5) and a turn (1TB7) containing structures are performed for 10 ns in explicit solvent under NPT conditions. The simulation result also shows similar hydrogen bonding interactions of water molecules around the CO group of the proline and the NH group of the methionine. The number of water molecules present around the NH group of the methionine and the CO group of the proline are retained as in ONIOM model systems (Figures S6 and S7 and Table 3). The gas phase, ONIOM, and MD calculations of the model systems have shown noncovalent interactions with adjacent residues, and solvent molecules in structures **I** and **II** are different.

Electronic structure calculations thus explain the structural anomaly that at pH = 10.0 conformation **I** is seen in all 3_{10}-helix PDE4B structures and at pH = 6.5−7.5 prefer conformation **II** as observed in all turn containing crystal structures. The pH variation thus can have different protonation states of cysteine side chain or backbone amide bonds. The different protonated states are stabilized by the noncovalent interactions of the amide bond with the surrounding environment such as adjacent residues and solvent molecules. Thus the orientation of the amide bond between proline and methionine is the deciding factor for the secondary structural change in the **II** or **I**. The pH dependence of the structures shows that there will be potential differences in the interaction at different parts of the cells in the body at different pH for the same enzyme−substrate combination. The secondary structure difference presented here has similar sequences and cannot account for subtype selectivity of PDE4 because at physiological pH it could have a turn structure. However, the variation in the sequence of the M-loop region may change the interactions with the inhibitor that can help in subtype selectivity. The exploration of the PDE4 inhibitor that can accommodate the M-loop region apart from the metal[8b,9] and the solvent pocket[34] of the active site can lead to better subtype selectivity.

Conclusions

The 35 X-ray crystal structures of the two subtypes of PDE4 show an influence of pH of the crystallization medium as an experimental variable on the secondary structural change in the M-loop region. The pattern of secondary structural change in the highly flexible M-loop region as a function of pH is striking in these crystal structures. Electronic structure calculations on the model sequence of the SPMCD residue of PDE4 explain the thermodynamic preferences for the secondary structure in the pattern of a 3_{10}-helix versus a turn in the M-loop region. The protonated states of the cysteine side chain and amide bond of turn containing structures are exothermic according to the isodesmic equations. It is due to the increase in the number of hydrogen bonding interactions with a surrounding environment such as adjacent residues and solvent molecules. Thus the variations in local (physiological) pH at the point of interaction can be an added parameter in optimizing the specificity of PDE4 inhibitors.

Acknowledgment. We thank CMSD and HPCF of the University of Hyderabad, MHPCC at Hawaii, and SERC of IISc for computing facilities. We thank Prabal Maiti for helping in the installation of AMBER 9.0 on a BLUE GENE/L machine. We thank DBT and DST for financial assistance. D.U.R. and P.S. also thank CSIR for fellowships.

Supporting Information Available: Structures of ligand in the cocrystals of various PDBs and the root-mean-square deviation graphs with various structures, isodesmic equations, total energies and Cartesian coordinates for optimized geometries, and graphs of water shell around the amide group of 1TB5 and 1TB7 in a 10 ns MD simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Bender, A. T.; Baevo, A. J. Cyclic Nucleotide Phosphodiesterases: Molecular Regulation to Clinical Use. *Pharmacol Rev.* **2006**, *58*, 488–520. (b) Houslay, M. D. PDE4-cAMP-specific Phosphodiesterases. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *69*, 249–315. (c) Chung, K. F. Phosphodiesterase Inhibitors in Airways Disease. *Eur. J. Pharmacol.* **2006**, *533*, 110–117. (d) Boswell-Smith, B.; Spina, D.; Page, C. P. Phosphodiesterase inhibitors. *Br. J. Pharmacol.* **2006**, *147*, S252−S257.

(2) (a) Castro, A.; Jerez, M. J.; Martinez, A.; Gil, C. Cyclic nucleotide Phosphodiesterases and their Role in Immunomodulatory Responses: Advances in the Development of Specific Phosphodiesterase Inhibitors. *Med. Res. Rev.* **2005**, *25*, 229–244. (b) Zhang, K. Y. J.; Ibrahim, P. N.; Gillette, S.; Bollag, G. Phosphodiesterase 4 as a Potential Drug Target. *Exp. Opin. Ther. Drugs* **2005**, *9*, 1283–1304. (c) Houslay, M. D.; Schafer, P.; Zhang, K. Y. J. Phosphodiesterase-4 as Therauptic target. *Drug Discovery Today* **2005**, *10*, 1503–1519.

(3) (a) Jeon, J. H.; Heo, Y.-S.; Kim, C. M.; Hyun, Y.-L.; Lee, T. G.; Ro, S.; Cho, J. M. Phosphodiestrase: Overview of Protein Structures, Potential Therapeutic Applications and Recent progress in Drug Development. *Cell. Mol. Life Sci.* **2005**, *62*, 1198–1220. (b) Srivani, P.; Srinivas, E.; Raghu, R.; Sastry, G. N. Molecular Docking Studies of Pyridopurinone Derivatives - Potential Phosphodiesterase5 Inhibitor. *J. Mol. Graphics Modell.* **2007**, *26*, 378–390.

(4) Xu, R. X.; Hassell, A. M.; Vanderwall, D.; Lambert, M. H.; Holmes, W. D.; Luthe, M. A.; Rocque, W. J.; Milburn, M. V.; Zhao, Y.; Ke, H.; Nolte, R. T. Atomic structure of PDE4: Insight into Phosphodiesterase Mechanism and Specificity. *Science* **2000**, *288*, 1822–1825.

(5) Lee, M. E.; Markowitz, J.; Lee, J.-O.; Lee, H. Crystal structure of Phosphodiesterase 4D and Inhibitor complex. *FEBS Lett.* **2002**, *530*, 53–58.

(6) (a) Huai, Q.; Wang, H.; Sun, Y.; Kim, H.-Y.; Liu, Y.; Ke, H. Three-dimensional Structures of PDE4D in Complex with Roliprams and Implication on Inhibitor selectivity. *Structure* **2003**, *11*, 865–873. (b) Huai, Q.; Liu, Y.; Francis, S. H.; Corbin, J. D.; Ke, H. Crystal structures of Phosphodiesterases 4 and 5 in Complex with Inhibitor 3-isobutyl-1-Methyl Xanthine suggest a Conformation Determinant of Inhibitor Selectivity. *J. Biol. Chem.* **2004**, *279*, 13095–13101.

(7) (a) Huai, Q.; Colicelli, J.; Ke, H. The crystal structure of AMP-bound PDE4 Suggests a Mechanism for Phosphodiesterase Catalysis. *Biochemistry* **2003**, *42*, 13220–13226. (b) Xu, R. X.; Rocque, W. J.; Lambert, M. H.; Vanderwall, D. E.; Luther, M. A.; Nolte, R. T. Crystal structures of the Catalytic Domain of Phosphodiesterase 4B complexed with AMP, 8-Br-AMP, and Rolipram. *J. Mol. Biol.* **2004**, *337*, 355–365.

(8) (a) Zhang, K. Y. J.; Card, G. L.; Suzuki, Y.; Artis, D. R.; Fong, D.; Gillette, S.; Hsieh, D.; Neiman, J.; West, B. L.; Zhang, C.; Milburn, M. V.; Kim, S.-H.; Schlessinger, J.; Bollag, G. A. Glutamine Switch Mechanism for Nucleotide Selectivity by Phosphodiesterases. *Mol. Cell* **2004**, *15*, 279–286. (b) Card, G. L.; England, B. P.; Suzuki, Y.; Fong, D.; Powell, B.; Lee, B.; Luu, C.; Tabrizizad, M.; Gillette, S.; Ibrahim, P. N.; Artis, D. R.; Bollag, G.; Milburn, M. V.; Kim, S.-H.; Schlessinger, J.; Zhang, K. Y. J. Structural Basis for the Activity of Drugs that Inhibit Phosphodiesterases. *Structure* **2004**, *12*, 2233–2247.

(9) Card, G. L.; Blasdel, L.; England, B. P.; Zhang, C.; Suzuki, Y.; Gillette, S.; Fong, D.; Ibrahim, P. N.; Artis, D. R.; Bollag, G.; Milburn, M. V.; Kim, S.-H.; Schlessinger, J.; Zhang, K. Y. J. A Family of Phosphodiesterase Inhibitors Discovered by Cocrystallography and Scaffold-based drug design. *Nat. Biotechnol.* **2005**, *23*, 201–207.

(10) Huai, Q.; Sun, Y.; Wang, H.; Macdonald, D.; Aspiotis, R.; Robinson, H.; Huang, Z.; Ke, H. Enantiomer discrimination Illustrated by the High Resolution Crystal Structures of Type 4 Phosphodiesterase. *J. Med. Chem.* **2006**, *49*, 1867–1873.

(11) Wang, H.; Peng, M.; Chen, Y.; Geng, J.; Robinson, H.; Houslay, M. D.; Cai, J.; Ke, H. Structures of the Four subfamilies of Phosphodiesterase−4 provide Insight into the Selectivity of their Inhibitors. *Biochem. J.* **2007**, *408*, 193–201.

(12) Nelson, D. L.; Cox, M. M. *Amino acids, peptides and proteins. Lehninger Principles of Biochemistry,* 4th ed.; W. H. Freman and Company: 41 Madison Avenue, New York, 2006.

(13) (a) Millhauser, G. L. Views of Helical Peptides: A Proposal for the Position of 3_{10}-Helix along the Thermodynamic Folding Pathway. *Biochemistry* **1995**, *34*, 3873–77. (b) Wu, Y.-D.; Zhao, Y.-L. A Theoretical Study on the Origin of Cooperativity in the Formation of 3_{10}- and α-Helices. *J. Am. Chem. Soc.* **2001**, *123*, 5313–19. (c) Chin, W.; Piuzzi, F.; Dognon, J.-P.; Dimicoli, I.; Tardivel, B.; Mons, M. Gas Phase Formation of a 3_{10}-Helix in a Three-Residue Peptide Chain: Role of Side Chain-Backbone Interactions as Evidenced by IR-UV Double Resonance Experiments. *J. Am. Chem. Soc.* **2005**, *127*, 11900–01.

(14) Bindu, P. H.; Sastry, G. M.; Murty, U. S. N.; Sastry, G. N. Structural and Conformational Changes Concomitant with the E1-E2 Transition in H+K+-ATPase: A Comparative Protein Modeling Study. *Biochem. Biophys. Res. Commun.* **2004**, *319*, 312–320.

(15) MOE version 2006; Chemical Computing Group, 954, First Floor, 16th Main, BTM Layout 2nd Stage, Bangalore, India 560 076.

(16) (a) Becke, A. D. Density-functional Thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. (b) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(17) Insight II Version 2000; Molecular Modelling System 2000 Molecular Simulations: 9685 Scrantan Road, San Deigo, CA.

(18) Dapprich, S.; Komáromi, I.; Suzie Byun, K.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives. *J. Mol. Struct. (Theochem)* **1999**, *462*, 1–21.

(19) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. Geometry Optimization with QM/MM, ONIOM, and Other Combined Methods. I. Microiterations and Constraints. *J. Comput. Chem.* **2003**, *24*, 760–769.

(20) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. UFF, A Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–35.

(21) *Gaussian 03, Revision B.03*, 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.; Bakken, V.; 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. B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Wallingford, CT, 2003.

(22) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Wang, J.; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; et al. *AMBER9*, 9th ed.;University of California: San Francisco, CA.

(23) Duan, Y.; Wu, C.; Chowdhury, S.; Le, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(24) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to An External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–90.

(25) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N\cdot\log(N)$ Method For Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–92.

(26) Wang, H.; Liu, Y.; Huai, Q.; Cai, J.; Zoraghi, R.; Francis, S. H.; Corbin, J. D.; Robinson, H.; Xin, Z.; Lin, G.; Ke, H. Multiple Conformations Of Phosphodiesterase5 Implications For Enzyme Function and Drug Development. *J. Biol. Chem.* **2006**, *281*, 21469–21479.

(27) (a) Alexov, E. Numerical Calculations of the pH of Maximal Protein Stability. *Eur. J. Biochem.* **2004**, *271*, 173–185. (b) Schaefer, M.; Sommer, M.; Karplus, M. pH Dependence of Protein Stability: Absolute Electrostatic Free Energy Differences between Conformations. *J. Phys. Chem.* **1997**, *101*, 1663–1683. (c) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediciton of pH-Dependent Properties of Proteins. *J. Mol. Biol.* **1994**, *238*, 415–436. (d) Pujato, M.; Navarro, A.; Versace, R.; Mancusso, R.; Ghose, R.; Tasayco, M. L. The

pH-dependence of Amide Chemical Shift of Asp/Glu Reflects its pk_a in Intrinsically Disordered Proteins with Local Interactions. *Biochim. Biophys. Acta* **2006**, *1764*, 1227–1233. (e) Thomas, P. G.; Russell, A. J.; Fersht, A. R. Tailoring the pH Dependence of Enzyme Catalysis using Protein Engineering. *Nature* **1985**, *318*, 375–376.

(28) Gursky, O.; Badger, J.; Li, Y.; Caspar, D. L. Conformational Changes in Cubic Insulin Crystals in the pH Range 7−11. *Biophys. J.* **1992**, *63*, 1210–1220.

(29) (a) Murthy, J. N.; Nagaraju, M.; Sastry, G. M.; Rao, A. R.; Sastry, G. N. Acitve site acidic residues and structural analysis of modeled human aromatase. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 857–870. (b) Altun, A.; Guallar, V.; Friesner, R. A.; Shaik, S.; Thiel, W. The Effect of Heme Environment on the Hydrogen Abstraction Reaction of Camphor in $P450_{cam}$ Catalysis: A QM/MM Study. *J. Am. Chem. Soc.* **2006**, *128*, 3924–3925. (c) Kumar, D.; Hirao, H.; Shaik, S.; Kozlowski, P. M. Proton-Shuffle Mechanism of O-O Activation for Formation of a High-Valent Oxo-Iron Species of Bleomycin. *J. Am. Chem. Soc.* **2006**, *128*, 16148–16158.

(30) (a) Lide, D. R. *Handbook of Chemistry and Physics,* 84th ed.; CRC Press: Boca Raton, FL, 2003. (b) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pK_a Values of The Ionizable Groups of Proteins. *Protein Sci.* **2006**, *15*, 1214–1218. (c) Dawson, R. M. C.; Elliot, D. C.; Elliot, W. H.; Jones, K. M. *Data for biochemical research,* 3rd ed.;Clarendon Press: Oxford, 1995. (d) Bombarda, E.; Morellet, N.; Cherradi, H.; Spiess, B.; Bouaziz, S.; Grell, E.; Roques, B. P.; Mély, Y. Determination of the pK_a of the four Zn^{2+}-Coordinating Residues Of The Distal Finger Motif of the HIV-1 Nucleocapsid Protein: Consequences on the Binding of Zn^{2+}. *J. Mol. Biol.* **2001**, *310*, 659–672. (e) Dudev, T.; Lim, C. Factors Governing The Protonation State of Cysteine In Proteins: An Ab initio/CDM study. *J. Am. Chem. Soc.* **2002**, *124*, 6759–66.

(31) Noar, M. M.; Jenson, J. H. Determinants of Cysteine pKa Values in Creatine Kinase and α-Antitrypsin. *Proteins: Struct., Funct., Bioinfor.* **2004**, *57*, 799–803.

(32) (a) Marino, T.; Russo, N.; Tocci, E.; Toscano, M. Density Functional Computations of Proton Affinity and Gas-Phase Basicity of Proline. *J. Mass. Spectrom.* **2001**, *36*, 301–305. (b) Sapse, A.-M. *Molecular orbital calculations for amino acids and peptides*; Birkhaüser, Boston, MA, 2000.

(33) Joseph, J.; Jemmis, E. D. Red-, Blue-, or No-Shift in Hydrogen Bonds: A Unified Explanation. *J. Am. Chem. Soc.* **2007**, *129*, 4620–4632.

(34) Krier, M.; de Araújo-Júnior, J. X.; Schmitt, M.; Duranton, J.; Justiano-Basaran, H.; Lugnier, C.; Bourguignon, J.-J.; Rognan, D. Design of Small-Sized Libraries by Combinatorial Assembly of Linkers and Functional Groups to a Given Scaffold: Application to the Structure-Based Optimization of a Phosphodiesterase 4 Inhibitor. *J. Med. Chem.* **2005**, *48*, 3816–3822.

CT700261B

JCTC Journal of Chemical Theory and Computation

Proton Transfer at Metal Sites in Proteins Studied by Quantum Mechanical Free-Energy Perturbations

Markus Kaukonen, Pär Söderhjelm, Jimmy Heimdal, and Ulf Ryde*

Department of Theoretical Chemistry, Lund University, Chemical Centre, P.O. Box 124, SE-221 00 Lund, Sweden

Received December 22, 2007

Abstract: Catalytic metal sites in enzymes frequently have second-sphere carboxylate groups that neutralize the charge of the site and share protons with first-sphere ligands. This gives rise to an ambiguity concerning the position of this proton, which has turned out to be hard to settle with experimental, as well as theoretical, methods. We study three such proton-transfer reactions in two proteins and show that, in [Ni,Fe] hydrogenase, the bridging Cys-546 ligand is deprotonated by His-79, whereas in oxidized copper nitrite reductase, the His-100 ligand is neutral and the copper-bound water molecule is deprotonated by Asp-98. We show that these reactions strongly depend on the electrostatic interactions with the surrounding protein and solvent, because there is a large change in the dipole moment of the active site (2−6 D). Neither vacuum quantum mechanical (QM) calculations with large models, a continuum solvent, or a Poisson−Boltzmann treatment of the surroundings, nor combined QM and molecular mechanics (QM/MM) optimizations give reliable estimates of the proton-transfer energies (mean absolute deviations of over 20 kJ/mol). Instead, QM/MM free-energy perturbations are needed to obtain reliable estimates of the reaction energies. These calculations also indicate what interactions and residues are important for the energy, showing how the quantum system may be systematically enlarged. With such a procedure, results with an uncertainty of ∼10 kJ/mol can be obtained, provided that a proper QM method is used.

Introduction

The understanding of reaction mechanisms of enzymes has been a major goal in biochemistry for a long time. Most enzymes contain ionizable groups in their active site, and they often have perturbed properties (e.g., acid constants) compared to the same group in a water solution. This is especially pronounced in the neighborhood of metal sites, owing to the strong electrostatic field from the metal. Therefore, it is often crucial for the understanding of enzyme mechanisms to predict the acid constants of active-site residues, or at least the location of protons in the active site. This is a formidable task both for experimental and theoretical methods, and therefore many different methods have been developed with this aim.

A common way to study the reaction mechanism in protein is to perform a quantum mechanical (QM) calculation of only a few residues from the active site.[1] Such a procedure has the advantage of allowing for calculations with big basis sets and the inclusion of zero-point energies. On the other hand, the surrounding protein is only modeled by a few explicit residues, whereas the rest of the protein is either ignored or modeled as a featureless continuum with a dielectric constant of ∼4. Likewise, the dynamics of the protein are ignored, and entropic and thermal effects are estimated from a harmonic analysis of the normal-mode frequencies of the modeled system, a convenient but approximate approach.

At the next level of approximation, the whole enzyme is included in the calculations. This can be done in different ways. First, the whole protein may be described by molecular mechanics (MM) or the active site by QM methods and the surrounding protein by a point-charge model or by MM (QM/MM methods).[2–4] Second, a single conformation (a mini-

* Corresponding author tel.: +46 - 46 2224502; fax: +46 - 46 2224543; e-mail: Ulf.Ryde@teokem.lu.se.

10.1021/ct700347h CCC: $40.75
Published on Web 05/21/2008

mized or experimental structure) or a thermodynamic ensemble of structures can be studied. Finally, the investigation may be restricted to enthalpies or free energies may be estimated by either strict statistical mechanics methods (e.g., free-energy perturbations or thermodynamic integration) or more approximate methods (e.g., from the change in solvent-exposed surface area).

One popular way to estimate acid constants in proteins is to consider only electrostatic effects, estimated by the solution of the Poisson−Boltzmann (PB) equation.[5–9] The charges of the active site are typically obtained from a QM calculation, and the solvation energies are combined with the corresponding QM energies. Normally, only a single structure is considered, and this is compensated for by the use of an effective dielectric constant (ϵ) of 4 or higher inside the protein.

The most accurate estimates of free energies in proteins are obtained by free-energy perturbations (and similar methods) based on a QM description of at least part of the protein. Several approaches along these lines have been suggested, differing in the level of QM calculations, the interaction between the QM and MM systems, and the methods to perform the sampling and estimate the free energies.[10–19]

A problem with all methods that include explicitly the surrounding protein is the magnitude of the electrostatic interactions: Electrostatic atom−atom interactions are often ~50 kJ/mol in a point-charge approximation, but the great majority of these interactions cancel. This makes the calculations sensitive to the treatment of solvation and the value of the dielectric constant in continuum approaches. This was succinctly illustrated in a recent study of the active copper site in nitrite reductase.[8] In this, the position of a proton in a hydrogen bond between the imidazole group of a histidine (His) copper ligand and the carboxylate group of a glutamate (Glu) residue was studied. Although the two structures differ only in the movement of a proton by less than 0.7 Å, estimates of the energy difference of the two states with different methods differed by up to 93 kJ/mol. In fact, a scaling of the point-charge model of the surrounding protein with a factor (effective dielectric constant) of 4 changed the relative energy by 88 kJ/mol. Likewise, theoretical estimates of pK_a values of metal ligands may easily be off by over 10 pK_a units.[20]

In this paper, we study the transfer of a proton within hydrogen bonds close to active-site metal ions with various theoretical methods, ranging from pure QM methods to a QM/MM free-energy perturbation approach (QTCP).[18,19] We study two test cases, copper nitrite reductase, mentioned above, and the transfer of a proton between a bridging cysteine ligand and a nearby His residue in [Ni,Fe] hydrogenase. By comparing the QTCP results with those obtained with cheaper methods, we may discuss the usefulness of various approximations. We also suggest a method to decide the size of the QM system and discuss whether large proteins may be truncated in the calculations.

Methods

QM/MM Calculations. The QM/MM calculations were carried out with the ComQum program.[21,22] In this approach, the protein and solvent are split into three subsystems: The QM system (system 1) contains the most interesting atoms and is relaxed by QM methods. System 2 consists of all residues within 15 Å of any atom in system 1 and is relaxed by a full MM minimization in each step of the QM/MM geometry optimization (in the QM/MM_free calculations; in the other calculations, only system 1 was optimized). Finally, system 3 contains the remaining part of the protein and surrounding solvent molecules and is kept fixed at the original (crystallographic) coordinates.

In the QM calculations, system 1 is represented by a wave function, whereas all of the other atoms are represented by an array of partial point charges, one for each atom, taken from MM libraries. Thereby, the polarization of the QM system by the surroundings is included in a self-consistent manner. When there is a bond between systems 1 and 2 (a junction), the hydrogen link-atom approach is employed: The QM system is capped with hydrogen atoms (H junction atoms), the positions of which are linearly related to the corresponding carbon atoms (C junction atoms) in the full system.[21,23] Charges on atoms bound to junction atoms are zeroed, and the charges are evenly distributed on the other MM atoms in that residue.[21,22]

The total QM/MM energy in ComQum is calculated as:

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM1+ptch}} - E_{\mathrm{MM1}} + E_{\mathrm{MM123}} \qquad (1)$$

where $E_{\mathrm{QM1+ptch}}$ is the QM energy of the QM system truncated by hydrogen atoms and embedded in the set of point charges (but excluding the self-energy of the point charges). E_{MM1} is the MM energy of the QM system, still truncated by hydrogen atoms, but without any electrostatic interactions. Finally, E_{MM123} is the classical energy of all atoms in the system with C junction atoms and with the charges of the QM system set to zero (to avoid double-counting of the electrostatic interactions). By this approach, which is similar to the one used in the Oniom method,[24] errors caused by the truncation of the QM system should cancel.

The geometry optimizations were continued until the energy change between two iterations was less than 2.6 J/mol (10^{-6} au) and the maximum norm of the Cartesian gradients was below 10^{-3} au. Restrained structures were optimized with a harmonic restraint with a force constant of 9375 kJ/(mol·Å^2) = 1 au, giving a deviation of less than 0.001 Å in the final structure.

QM Calculations. To be consistent with our earlier calculations,[8,25] the QM calculations on the two proteins were slightly different. For [Ni,Fe] hydrogenase (H2ase), the calculations were performed with the Becke 1988−Perdew 1986 (BP86) density functional[26,27] together with the 6-31G* basis set[28] for H, C, N, O, and S and the DZP basis set for Fe and Ni.[29,30] The calculations were sped up by expanding the Coulomb interactions in auxiliary basis sets, the resolu-

tion-of-identity approximation.[31,32] For nitrite reductase (NIR), we instead used the three-parameter hybrid B3LYP method, as implemented in the Turbomole package.[33–35] These calculations employed the 6-31G* basis set for all atoms,[9] except for copper, for which we used the DZP basis sets of Schäfer et al.,[29,30] enhanced with *p*-, *d*-, and *f*-type functions with exponents of 0.174, 0.132, and 0.39 (DZpdf). All QM calculations were performed with the Turbomole 5.8 and 5.9 software.[35]

To obtain more accurate energies, a single-point B3LYP calculation was performed for the optimized structures. In these calculations, the 6-311+G(2d,2p) basis set was used for the light atoms,[9] whereas the DZP basis set for the Fe and Ni was enhanced with *s*-, *p*-, *d*-, and two *f*-type functions with the following exponents: 0.013 772, 0.041 843, 0.1244, 2.5, and 0.8 for Fe and 0.145 763, 0.146 588, 0.044 447, 0.1458, 6.74, and 1.04 for Ni, and the basis set for Cu was enhanced with *s*, *p*, and *f* functions with exponents of 0.0155, 0.046 199, and 3.55.

Vibrational frequencies were obtained with the same method and basis sets as for the geometry optimizations for H2ase (BP86/6-31G*). The QM systems were first reoptimized in a vacuum, and then vibrational frequencies were calculated for those structures. From these frequencies, zero-point energies and thermal corrections to the Gibbs free energy (at 298 K and 1 atm pressure) were obtained, using an ideal-gas approximation.[36] In nitrite reductase, a water molecule hydrogen-bonded to the glutamate model in the QM system had to be added to prevent the QM system from relaxing to the Both state.

In some cases, solvation effects were estimated by single-point calculations using the continuum conductor-like screening model (COSMO).[37,38] These calculations were performed at the same level of theory as the geometry optimization and with default values for all parameters (implying a water-like probe molecule) and a dielectric constant (ϵ) of 4. For the generation of the cavity, a set of atomic radii has to be defined. We used the optimized COSMO radii in Turbomole (1.30, 2.00, 1.83, and 1.72 Å for H, C, N, and O, respectively, and 2.223 Å for the metals).[39] These energies are called $E_{QM1+\epsilon=4}$ in the following.

MM Calculations. All MM calculations were run with the sander module in the AMBER 8 software,[40] using the Amber 1999 force field.[41,42] The QM system was represented by charges fitted to the electrostatic potential, calculated in 115 000−260 000 points selected at random around the QM system up to a distance of 8 Å and outside the van der Waals envelope of the QM system. The fit used singular-value decomposition methods to ensure that all of the fitted charges are significant. The charges were constrained to exactly reproduce the QM dipole moment and quadrupole moments, as well as the Boltzmann-weighted electrostatic potential, the CHELP-BOW method.[43] The charge on each C junction atom was adapted so that the total charge of the amino acid (including both QM and MM atoms) was the same as the sum of QM charges of the corresponding QM fragment.[22] Thereby, we ensure that the total charge of the simulated system is an integer, but we still allow charge transfer within the QM system (the amino acids containing junctions have noninteger total charges). Moreover, the charges on the C junction atoms are changed from what is typical for a hydrogen atom to what is more typical for carbon atoms.

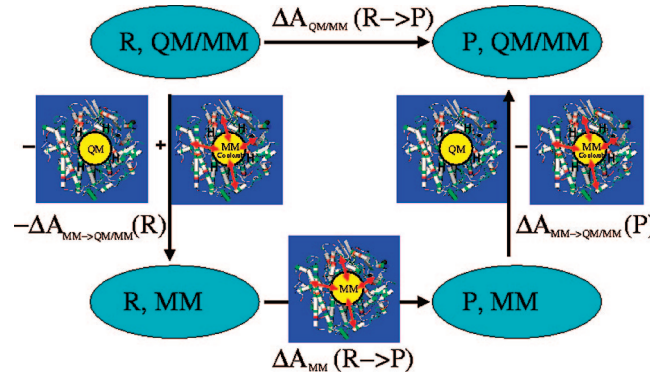


Figure 1. The QTCP cycle and energies.[18,19] We want to obtain the QM/MM free-energy difference between the reactant (R) and the product (P, upper horizontal line, $\Delta A_{QM/MM}(R \rightarrow P)$). It is obtained by calculating the computationally much cheaper MM free-energy difference ($\Delta A_{MM}(R \rightarrow P)$, lower horizontal line), as well as the two QM → QM/MM correction terms ($\Delta A_{MM \rightarrow QM/MM}$, the two vertical lines). All energies can be obtained by MD at the MM level only.

All bond lengths involving hydrogen atoms were constrained by the SHAKE algorithm.[44] The water solvent was described explicitly using the TIP3P model.[45] The electrostatics were treated with the particle-mesh Ewald method[46,47] with a grid size of 80^3 Å, a fourth-order B-spline interpolation, a tolerance of 10^{-5}, and a real-space cutoff of 8 Å. The temperature was kept constant at 300 K using the Berendsen weak-coupling algorithm[48] with a time constant of 1 ps. The molecular dynamics (MD) time step was 2 fs, and the nonbonded pair list was updated every 50 fs.

QTCP and QM/MM-FE. QTCP (QM/MM thermodynamic cycle perturbation) is a method to estimate free-energy differences at the QM/MM level.[18,19] It is based on the thermodynamic cycle in Figure 1. Thus, we estimate the QM/MM free-energy difference between the reactant (R) and the product (P), $\Delta A_{QM/MM}(R \rightarrow P)$, by performing three free-energy perturbations. The first one (the horizontal line in Figure 1) is the corresponding free-energy difference at the MM level, $\Delta A_{MM}(R \rightarrow P)$ (if the difference in free energy between the reactant and product is too large, this perturbation may be done in several smaller steps). The second and third perturbations change the description of the R and P states from the MM level to the QM/MM level (the vertical lines in Figure 1). Summing up these three energies, we obtain the free-energy change from the reactant to product at the QM/MM level:

$$\Delta A_{QTCP}(R \rightarrow P) = \Delta A_{MM}(R \rightarrow P) + \Delta A_{MM \rightarrow QM/MM}(P) - \Delta A_{MM \rightarrow QM/MM}(R) \quad (2)$$

The advantage with this approach is that the phase space needs to be sampled only at the MM level, thereby avoiding the extremely time-consuming sampling at the QM/MM level. A similar approach has been used by Warshel and co-workers,[12,13] but it led to severe convergence problems. We solved these by keeping the QM system fixed during

the simulations.[18,19] Test calculations at the semiempirical level have shown that this does not affect the final free energies significantly.[49] By this procedure, we also avoid the need of bonded MM parameters for the QM system.

In practice, the QTCP calculations were performed in the following way (see http://www.teokem.lu.se/~ulf/Methods/qtcp.html for a more detailed description).[18,19] We start from the structures obtained by the QM/MM (for H2ase,[25] keeping system 2 fixed) or quantum-refinement (nitrite reductase[8]) optimizations. Then, the proteins were better solvated by using an octahedral water box, extending at least 9 Å from the protein. This system was then minimized by 100 steps of MM minimization, keeping the QM atoms fixed and restraining the heavy atoms in the protein by a harmonic force constant of 100 kcal/(mol·Å^2). This was followed by a MD equilibration in the NPT ensemble (1 atm pressure and 300 K temperature) for 70 ps, restraining the QM atoms by a harmonic force constant of 500 kcal/(mol·Å^2) to the QM/MM structure. For the truncated H2ase, heavy atoms near the cut were restrained by harmonic forces of 10 kcal/(mol·Å^2). After this equilibration of the volume of the simulation cell, the QM system was moved back to the exact position in the QM/MM calculations (the QM system moves slightly during the constant-pressure simulation) and was fixed at this position for the remaining simulations. Finally, the system was equilibrated in the NVT ensemble for 200 ps, and snapshots were saved every 10 ps for an additional time of at least 200 ps.

Finally, free energies were calculated from these snapshots (strictly speaking, Helmholtz free energies are calculated, rather than Gibbs free energies, but the difference between these two quantities is small for proteins).[50] The free energy at the MM level ($\Delta A_{MM}(R \rightarrow P)$ in eq 2) was calculated by a free-energy perturbation (FEP), in which the coordinates and the charges of the QM atoms (with C junction atoms) were changed from the R state to the P state (or vice versa) and the energy difference was calculated:

$$\Delta A_{MM}(R \rightarrow P) = -k_B T \ln\left\langle e^{-\frac{E(P)-E(R)}{k_B T}} \right\rangle_R \qquad (3)$$

where k_B is the Boltzmann constant, T is the temperature, and the angular brackets indicate an average over a MD ensemble, sampled for the R state (Figure 1). An estimate of the accuracy of this free energy is obtained by reversing the R and P states, using snapshots from the simulation of the P state instead (the hysteresis).

Likewise, the vertical MM → QM/MM free energies are obtained by a FEP, in this case for the same state described at two different levels of theory (Figure 1). As has been described before,[19] this is accomplished by the following equation:

$$\Delta A_{MM \rightarrow QM/MM}(X) = -k_B T \ln\left\langle e^{-\frac{E'_{QM1+ptch}(X)-E_{el12}(X)}{k_B T}} \right\rangle_X \qquad (4)$$

where X is either the R or P state and $E'_{QM1+ptch}$ is essentially the same energy as $E_{QM1+ptch}$ in eq 1 (i.e., the QM energy of the QM system embedded in a point-charge model of the surrounding protein, but excluding the self-energy of the point charges), with the small exception that, in the QM/MM calculations, a few point charges near the junction atoms were deleted, which is not the case for $E'_{QM1+ptch}$, in which all point charges of the MM system were included. E_{el12} is the electrostatic interaction energy between the QM (with H junction atoms) and MM systems (excluding interactions within the two systems), calculated at the MM level (i.e., by Coulomb's law), using the same point charges for the MM system as in $E'_{QM1+ptch}$ and the QM fitted point charges for the QM system (i.e., the same point charges used in the MD simulation), with the exception for the junction atoms, for which the original CHELP-BOW point charges were used (in the MD simulations, the point charges on the junction atoms were adapted so that the total charge of the protein was an integer; in the calculation of E_{el12}, the sum of the point charges of the QM system is an integer, whereas that of the MM system is not an integer). Thereby, we obtain an energy that is as similar as possible to that in the QM calculation ($E'_{QM1+ptch}$).

The vertical MM → QM/MM free energies ($\Delta A_{MM \rightarrow QM/MM}(P)$ and $\Delta A_{MM \rightarrow QM/MM}(R)$) are computationally expensive to obtain. Therefore, approaches have been suggested that employ only a single QM calculation for each state (i.e., $\Delta A_{MM \rightarrow QM/MM}(P) - \Delta A_{MM \rightarrow QM/MM}(R)$ is approximated by $E_{QM}(P) - E_{QM}(R)$). This corresponds to the QM/MM-FE approach[14,15] (called QM-FE if the QM system is optimized in a vacuum, rather than by QM/MM[10,11]). We have shown that this was a rather good approximation to QTCP for our test case, the methyl transfer by catechol *O*-methyltransferase, and we have also discussed how the $E_{QM}(P) - E_{QM}(R)$ energy difference is best calculated.[18,19] In this paper, we also test the QM/MM-FE approach. The $E_{QM}(P) - E_{QM}(R)$ energy difference is calculated by optimizing the QM wave functions with a point-charge model of the surroundings. Then, the point charges are removed and the energy is estimated by a single self-consistent field iteration (i.e., without reoptimizing the wave function and using a final grid size of 3 in Turbomole). Thereby, we obtain the energy of the QM system polarized by the surroundings, but excluding the electrostatic interactions between the QM system and the point charges, E_{QM1pol}. However, essentially the same results are obtained by using vacuum QM calculations (i.e., without any point-charge model, E_{QM1}) directly on the QM/MM structures.

In some QTCP calculations, solvent-exposed charged amino acids were neutralized (not in the MD simulations, but only in the final FEPs). Several methods to neutralize the residues were tried, for example, changing them to the corresponding neutral residue in the Amber libraries or scaling down all charges by a factor (dielectric constant) of 80, but the results were very similar. Therefore, we use in the presented results the simplest approach, namely, zeroing all charges in the neutralized residues. The solvent exposure of the residues was determined by counting the number of (non-hydrogen and nonwater) atoms within 15.5 Å of the center of the charged group (the side-chain nitrogen atom of Lys and the amino terminal; C^ζ of Arg; the average position of the two carboxylate oxygen atoms in Asp, Glu,

and the carboxy terminal; and the center of the imidazole ring of doubly protonated His). Groups with less than 575 protein atoms within this radius were considered solvent-exposed.[51]

Poisson−Boltzmann Solvation Energies. For comparison, the solvation energy of the QM system in the surrounding protein and water was estimated by solving the Poisson−Boltzmann equation using the solvinprot module of the software Mead 2.2,[52] following the same procedure used before for both H2ase and NIR.[8,25] Similar procedures have been used for several other proteins.[5−7,9,53−55] The QM system was modeled with the same QM point charges as in the QTCP calculations. The dielectric constants of the QM system, the protein, and the water solvent were set to 1, 4, and 80, respectively. No explicit water molecules were included in the calculations (except the three H_2O ligands of Mg in the H2ase). The positions of the atoms were taken from the QM/MM structures (H2ase) or from the ComQum-X structures (NIR). The Mead calculations were performed with 351^3 grid points and a spacing of 0.25 Å, centring the grid on the QM system. The reported energies are the average of seven calculations in which the grid origin was moved 0.1 Å in the positive and negative directions along each Cartesian axis. The maximum difference among the seven calculations was 2 kJ/mol. Parse radii[56] were used for all atoms. These values have previously been shown to provide reasonable agreement with experimental results.[9,57,58] The MEAD solvation energies (ΔG_{PB}) were added to the polarized QM energy to get a full free-energy estimate:

$$\Delta E_{PBtot} = \Delta G_{PB} + \Delta E_{QM1pol} \quad (5)$$

[Ni,Fe] Hydrogenase. The calculations on H2ase were based on the $P2_12_12_1$ crystal structure of the Ser499Ala mutant from *D. fructosovorans*,[59] which was the crystal structure with the best resolution (1.81 Å) at the start of our investigation. The hydroxide group of Ser-499 forms a hydrogen bond to one of the CN^- ligands of Fe, but in the mutant, a water molecule replaces this group, leading to unchanged enzymatic properties, vibrational frequencies, and structure, but much better diffracting crystals.[59]

Two different sizes of the simulated system were tested. One set included the full protein: 818 protein residues and either 5797 (QM/MM) or 14 790 (QTCP) water molecules, giving a total of 29 689 or 56 668 atoms. In the second set of calculations, all residues more than 27 Å from the Ni ion were deleted, and solvation water molecules were added to the protein, forming a sphere with a radius of 33 Å (602 protein residues and 1042 water molecules, giving a total of 12 178 atoms; this is the system used in our previous QM/MM study of this enzyme[25]). In the QTCP calculations, this system was embedded in an octahedral box with 6701 water molecules, yielding a total of 29 149 atoms in the system.

For metal sites outside the active site (the two [4Fe−4S] clusters, the [3Fe−4S] cluster, and a six-coordinated Mg ion), we used Merz−Kollman electrostatic potential (ESP) charges, taken from QM calculations of truncated models of each site. Force constants for the bond, angle, and dihedral terms were estimated from the Hessian matrix, obtained from a QM frequency calculation of the optimized models, using the approach by Seminario.[60,61]

The QM system, depicted in Figure 2, consisted of the Ni and Fe ions and their first-sphere ligands (two CN^- ions, CO, Cys-72, Cys-75, Cys-543, and Cys-546). Two second-sphere residues, Glu-25 and His-79, were also included because they share hydrogen atoms with Cys-543 and Cys-546, respectively. The Cys residues were modeled by CH_3S^-, whereas His was modeled by imidazole and Glu-25 by acetic acid. The aim of this investigation was to study the relative stability of the two states in which the proton shared by Cys-546 and His-79 resides on either His-79 (called the HIP state) or on Cys-546 (called the HID state).

In some calculations, the QM system was enlarged with some surrounding amino acids, as is illustrated in Figure 2. The seven different QM systems, N, NR, NACG, NACG′, NACGR, NCHACG, and NCHACGR, contained 46, 59, 66, 86, 79, 91, and 104 atoms, respectively.

Hydrogen atoms were added to the protein using the leap module in AMBER.[40] The protonation status of histidine residues was determined by an inspection of the local surroundings and hydrogen-bond structure. This gave protonation of the N^1 atom for residues S5 (an initial S refers to the small subunit, whereas residue numbers without S refer to the large subunit), S92, S160, S243, 367, 481, and 549; protonation of the $N^{\epsilon 2}$ atom for S13, S184, 27, 66, 113, 118, 121, 123, 188, 210, 228, 349, and 419; and protonation of both of these atoms for S61, S192, 79, 115, 204, 305, and 538. All Lys, Arg, Asp, and Glu residues were assumed to be charged, except Glu-25, which shares a hydrogen atom with the Ni ligand Cys-543, and Glu-S16, which is involved in the proton-transfer path from Cys-543 to the protein surface.

The positions of the hydrogen atoms and solvation water molecules were optimized by a simulated-annealing molecular dynamics calculation followed by a MM minimization, as has been described before,[62] with the QM system in the HIP state. Test calculations in which this optimization was performed for the HID state instead changed the final QTCP free energies by up to 8 kJ/mol (Table 1, the last row).

The oxidation states of metals were Fe^{II} [63,64] and Ni^{III}. This gives a neutral QM system when Arg-476 is included and a -1 *e* charge otherwise. The QM system was considered in the closed-shell singlet state.

Nitrite Reductase. The calculations on copper NIR were based on the crystal structure of oxidized NIR at pH 6.0 (Protein Data Bank file 1NIC at 1.9 Å resolution).[65] The protein is composed of three identical subunits (although only one subunit is present in the crystallographic unit cell), each with 333 residues and two copper ions (one type 1 blue copper ion, used for electron transfer, and the catalytic type 2 copper ion).

The QM structures for NIR were obtained in our previous study[8] using the ComQum-X quantum refinement approach. It is essentially a QM/MM minimization, in which the structure is restrained to remain close to crystallographic raw data (the structure factors).[8,66]

Hydrogen atoms and water molecules were added to the protein and were optimized in the same way as for H2ase. His-100, -135, and -306 were protonated on the $N^{\delta 1}$ atom and His-60, -95, and -145 on the $N^{\epsilon 2}$ atom, and the other

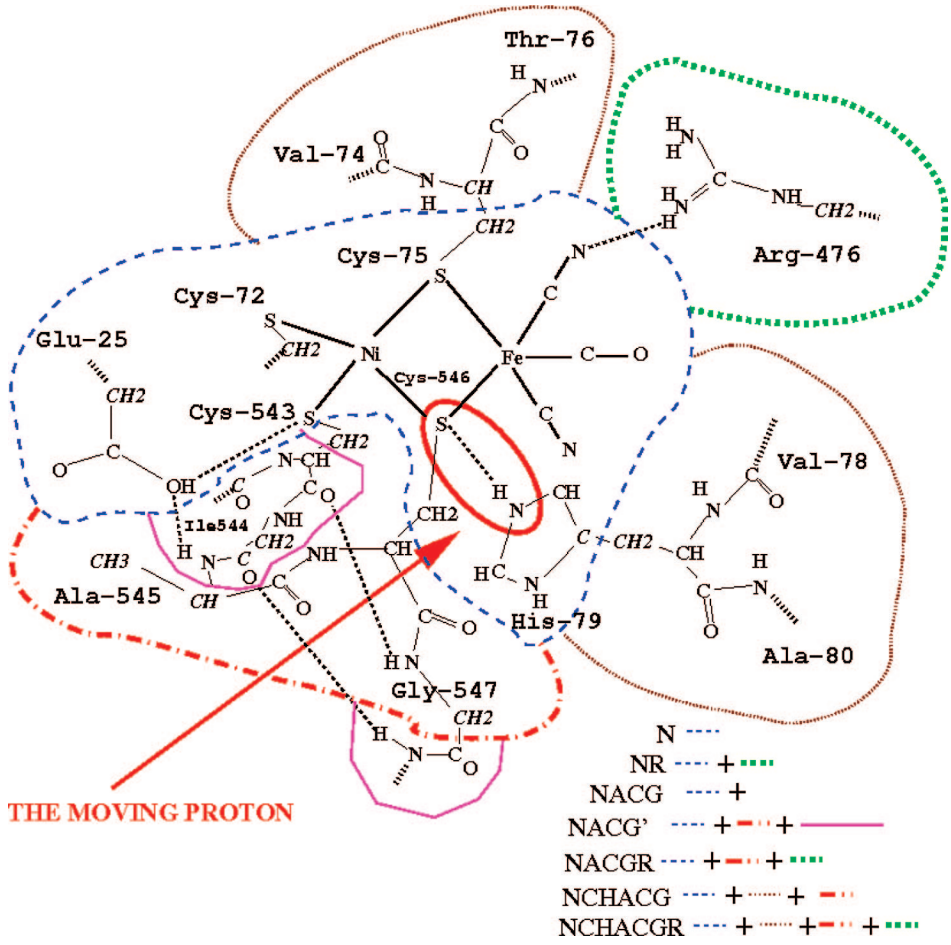


Figure 2. The seven QM systems used for the [Ni,Fe] hydrogenase (N, NR, NACG, NACG′, NACGR, NCHACG, and NCHACGR). In reality, Arg-476 is located above and His-79 below the plane of the paper. Cut bonds that are truncated by hydrogen atoms in the calculations are indicated by dotted lines and hydrogen bonds with dashed lines. The moving proton is indicated by the red circle.

His residues were protonated on both nitrogen atoms. The only Cys residue is a copper ligand and was therefore assumed to be deprotonated.

The QM system consisted of the Cu ion, three imidazole groups, as models of the His ligands, two acetate ions, as models of Asp-98 and Glu-279, and a water molecule. The QM system is shown schematically in Figure 3. The copper ion was studied in the oxidized Cu^{II} doublet state.[67] Therefore, the QM system was neutral. In some calculations, the QM system was enlarged by His-255 and Lys-262, resulting in a charge of the QM system of +2 *e*. The other copper site was treated by pure MM methods, using our previously developed and calibrated parameters (from QM calculations of the oxidized state).[68]

In the QTCP calculations, the whole trimeric protein (1005 amino acids) was considered. This was accomplished by copying the structure of the QM system from the ComQum-X calculation[8] to the other two subunits of the protein. For the QM energies in QTCP, only subunit A was considered, to save computer time, whereas the MM energies were obtained by perturbing all three copper sites simultaneously (but the presented free energies correspond to a single Cu site).

The QM/MM calculations for NIR were carried out on the quantum-refined structure, to which hydrogen atoms and a solvation layer of ~4000 water molecules were added and optimized by a 90-ps simulated-annealing calculation, followed by a MM minimization. The latter calculations were performed in the Both state to obtain the same environment for all four states. QM/MM_free structures were then obtained by optimizing system 2 for each of the four states separately by a MM minimization until the root-mean-square force was below 10^{-3} au.

Results and Discussion

[Ni,Fe] Hydrogenase: QTCP. Hydrogenases are enzymes that catalyze the seemingly simple reversible reaction[69]

$$H_2 \rightleftarrows 2H^+ + 2e^- \qquad (6)$$

There are several types of these enzymes. The H2ases have an active site with one Ni ion coordinated to four Cys residues and an Fe ion, which coordinates to two of the same Cys ligands and also two CN^- ions and one CO molecule (Figure 2).[70] Many crystal structures of H2ases have been presented,[59,70] but several details of the reaction are still controversial because the reactants and products of eq 6 are invisible in normal crystal structures. For example, the reaction involves protons, but it is not fully clear where these bind: In principle, any of the four Cys ligands of Ni may be protonated, and there are also other possible proton acceptors in the active site.

Table 1. QTCP Results for [Ni,Fe] Hydrogenase (kJ/mol)[a]

Protein	QM system	ΔA_{MM}	hysteresis	$\Delta A_{MM \rightarrow QM/MM}$	ΔA_{QTCP}
truncated	N[b]	−38.0	1.2	−1.2	−39.1
	N	−36.6	1.8	−0.7	−37.3
	N[c]	−33.9	2.8	−0.8	−34.7
	NR	0.9	0.4	−31.5	−30.6
	NACG	−18.1	1.5	−22.4	−40.5
	NACG′	−20.3	0.9	−23.3	−43.5
	NACGR	6.8	2.7	−47.9	−41.1
	NCHACG	−8.0	1.9	−40.7	−48.7
	NCHACGR	20.3	0.9	−64.2	−43.9
	NCHACGR[d]	51.0	0.6	−63.6	−12.6
full	N	−48.4	0.8	−2.3	−50.6
	NR	−14.4	4.0	−37.8	−52.2
	NACG	−46.0	1.5	−20.0	−66.0
	NACGR	−9.8	0.5	−53.1	−62.9
	NCHACG	−38.2	0.9	−39.7	−77.9
	NCHACGR	−6.7	2.5	−75.0	−81.8
	NCHACGR[d]	13.2	2.5	−59.2	−46.0
	NCHACGR[e]	−1.6	3.8	−72.8	−74.3

[a] A negative energy means that the HIP state is most stable. The calculations were performed both on the full and on the truncated protein and with seven different sizes of the QM system (cf. Figure 2). ΔA_{MM} and $\Delta A_{MM \rightarrow QM/MM}$ are the MM and MM → QM/MM parts of the QTCP free-energy difference, respectively (cf. Figure 1). The second column gives the hysteresis in the ΔA_{MM} term. The sum of the ΔA_{MM} and $\Delta\Delta A_{MM \rightarrow QM/MM}$ terms yields the QTCP free energy, ΔA_{QTCP}. [b] Six-step perturbation. [c] Nine Na^+ ions were added to the MM system to neutralize the protein. [d] The QTCP calculations were carried out without the water solvent. [e] The hydrogen atoms were equilibrated for the HID state, rather than for the HIP state.

Many theoretical investigations of the H2ases have also been presented.[25,71–76] In principle, it should be possible to deduce the energetically most favorable protonation sites by theoretical methods, but it is hard to obtain converged energies in such calculations.[25] In this investigation, we use a somewhat simpler test case, namely, the transfer of a single proton between the Ni ligand Cys-546 and His-79 along a hydrogen bond (Figure 2). We try to estimate the relative free energies of the state with the proton on Cys-546 (called the HID state) and the state with the proton on His-79 (called the HIP state).

We started to study this proton transfer in the truncated protein with the normal (N) QM system (cf. Figure 2) by dividing it into six separate steps, with H−N distances of 1.08 (HIP state), 1.14, 1.26, 1.44, 1.62, 1.80, and 1.98 Å (HID state). The corresponding QTCP free-energy profile is shown in Figure 4 (upper, black curve). It can be seen that the QTCP results indicate that the HIP state is 39 kJ/mol more stable than the HID state and that there is no barrier (activation energy) for the proton transfer. The hysteresis in each step is less than 1 kJ/mol, indicating that the step size is small enough. In fact, the proton transfer can be performed in a single step, giving a similar result (37 kJ/mol; cf. Table 1, first two rows), with a hysteresis of only 2 kJ/mol. Therefore, we have used single-step perturbations in the following investigations, where we study the effect of increasing the quantum system, with both the truncated and full proteins.

The Size of the QM System. An advantage with the MM free energies in the QTCP calculations is that it is straightforward to calculate approximate contributions to the free-energy difference from each QM atom or from each MM residue. Moreover, the interactions can be divided into contributions from the various types of MM energy terms, that is, bonds, angles, dihedrals, electrostatics, and van der Waals interactions.

The results of such an analysis are shown in Table 2. It can be seen that the energies are completely dominated by the electrostatic interactions, with only minimal contributions from the bonded and van der Waals interactions. It can also be noted that there are many residues with quite large contributions, but they to a large extent cancel. The largest contribution comes from Arg-476 (−28 kJ/mol), which is hydrogen-bonded to the CN ligand of the iron ion (cf. Figure 2). Two aspartate residues, Asp-114 and Asp-541, which are hydrogen-bonded to Arg-476, more than cancel the effect of Arg-476 (+15 and +19 kJ/mol, respectively). The MM parts of three QM residues also have significant contributions: Cys-75 (−9 kJ/mol), His-79 (+9 kJ/mol), and Cys-546 (−10 kJ/mol). In addition, the neighboring Gly-547 residue contributes by −12 kJ/mol, and His-481 (which forms a hydrogen bond to the proton acceptor His-79) favors the HIP state by 11 kJ/mol. Residues Arg-70 and Ala-71, which are close to the QM system, favor the HID state by ∼10 kJ/mol each.

These results indicate that we may systematically improve the results by enlarging the QM system by the residues with the largest MM contributions. To this end, the QM system was extended with the rest of residues, Cys-75 and His-79 (including the CO and NH groups of the preceding and succeeding residues, called CH), Ala-545, Cys-546, and Gly-547 (ACG), as well as the guanidine group of Arg-476 (R) in various combinations, as illustrated in Figure 2. In addition, we also tested a larger variant of the NACG system, in which the backbone atoms of Ile-544 and the amide linkage to Val-548 were also included (called NACG′).

The results of the QTCP calculations with the extended QM system calculations are listed in Table 1 for both the truncated and full-protein simulations. It can be seen that including Arg-476 in the QM system (R) has a small and varying effect on the QTCP energy (−4 to +9 kJ/mol, with an average of +1 kJ/mol), which probably mostly reflects the accuracy of the method (note that the hysteresis is 1−4 kJ/mol). On the other hand, there seems to be a more consistent effect of adding both Cys-546 (ACG) and Cys-75 plus His-79 (CH) to the QM system (average effects of −10 kJ/mol for both residues). These two contributions are reasonably additive. Thus, the results suggest that, if residues covalently connected to the QM system have large MM components, it is advisable to include them in the QM system, whereas other interactions (e.g., hydrogen bonds) are sufficiently well described by the MM force field.

The Size of the Simulated System. As mentioned above, we have performed QTCP calculations both with a spherically truncated system and for the full H2ase protein. The latter calculations are ∼5 times more expensive in terms of computer time, and similar truncations are frequently used in theoretical simulations of large proteins. As discussed above, calculations with the two systems give similar trends for the enlarged QM systems, and the HIP state is clearly favored in both calculations. However, the results in Table

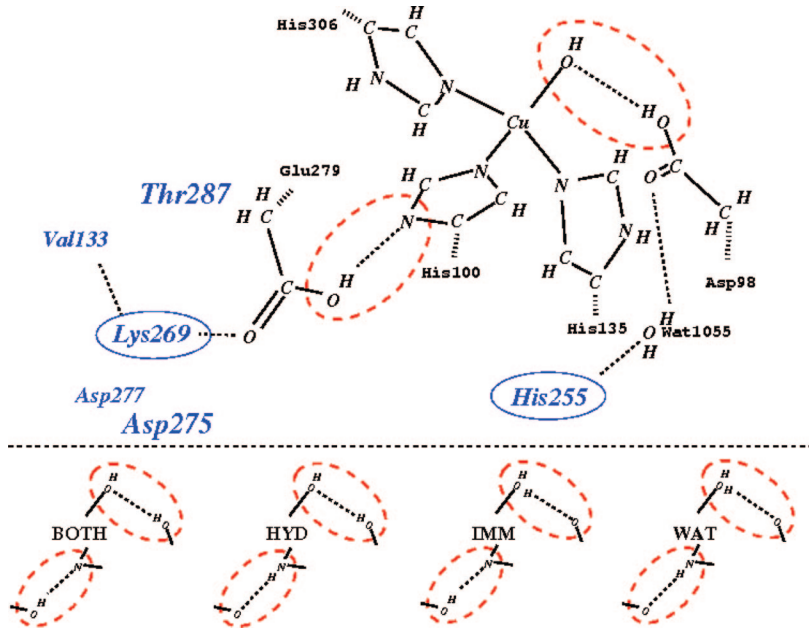


Figure 3. The normal (N) QM system for copper nitrite reductase. The moving protons are indicated as dashed ellipses. The resulting four geometries of the protonation of Glu-279, His-100, and Asp-98 and copper-bound water are sketched and labelled in the lower part of the figure. In some calculations, this QM system is extended with Lys-269 and His-255 (these residues are in ellipses in the figure; NHK). The location of the energetically important residues Val-133, Thr-287, Asp-275, and Asp-277 are also indicated.

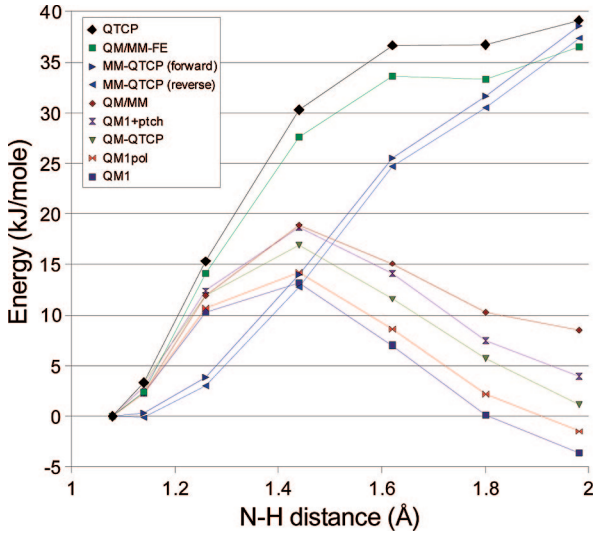


Figure 4. Various energy terms (defined in the Methods section) for the six-step proton transfer between His-79 (the HIP state at 1.08 Å) and Cys-546 (the HID state at 1.98 Å) in [Ni,Fe] hydrogenase. The data are from the truncated protein and the normal (N) QM system.

1 show that there is a shift in the free-energy difference when going from the truncated to the full protein. This shift is 13−38 kJ/mol (average 26 kJ/mol), favoring the HIP state in the full protein. The shift cannot be attributed to any single residue left out in the truncated protein; instead, it is spread over a large number of residues. The individual contributions of residues common to the full and truncated protein are similar (within 2 kJ/mol), as can be seen in Table 3. Thus, the truncation of a protein, even as far as 27 Å from the active site, can have a significant influence on the energetics of the protein.

Table 2. The Largest (Approximate) Contributions to the MM Free-Energy Difference (ΔA_{MM}, kJ/mol) between the HIP and HID States from the Surrounding Protein in the QTCP Simulation of Truncated [Ni,Fe] Hydrogenase with the Normal (N) QM System[a]

residue	bonded terms	van der Waals	electrostatics	total
Arg-70	0	0	10	10
Ala-71	0	−1	12	11
Cys-75	−1	0	−8	−9
His-79	0	0	9	9
Asp-114	0	0	15	15
Arg-476	0	−2	−26	−28
His-481	0	−3	−8	−11
Asp-541	0	0	19	19
Cys-546	0	1	−11	−10
Gly-547	0	0	−12	−12
Mg ion	0	0	9	9
water (sum)	0	0	−11	−11

[a] A negative value indicates that the HIP state is favored. The total QTCP free-energy difference for this QM system is −37 kJ/mol.

To verify that the shift really is caused by the truncated parts of the protein, and not by the water solvation, the calculation with the largest QM system (NCHACGR) was repeated without any water molecules in the system (except the three water ligands of the Mg^{2+} ion). This was achieved by stripping off the water molecules after the MD simulation. This led to a difference of 33 kJ/mol between the truncated and full simulations (cf. Table 1), clearly showing that the shift comes from the truncated part of the protein. Furthermore, it can be seen from the table that the water solvation favors the HIP state by ∼30 kJ/mol in both the full and truncated proteins.

Naturally, the full protein plus the solvent is larger than the truncated protein plus the solvent, and it is conceivable

Table 3. The Largest (Approximate) Contributions to the MM Free-Energy Difference (ΔA_{MM}, kJ/mol) between the HIP and HID State from the Surrounding Protein in the QTCP Simulation of the Full or Truncated [Ni,Fe] Hydrogenase with Various QM Systems[a]

residue	N	N	NR	NACG	NACGR	NACGR	NCHACG	NCHACGR
protein	full	truncated	full	full	full	truncated	full	full
Arg-70	10	10	10	12	11	12	11	11
Ala-71	10	11	12	10	13	14	13	11
Cys-75	−10	−9	−15	4	−4	−3	QM	QM
His-79	11	9	10	12	11	11	QM	QM
Asp-114	16	15	19	15	18	17	16	18
Hip-115	−7	−7	−8	−7	−7	−8	−7	−7
Arg-428	−9	−9	−9	−8	−9	−9	−9	−9
Arg-476	−29	−28	−3	−27	−4	−4	−27	−4
His-481	−10	−11	−11	−12	−11	−13	−9	−10
Asp-541	20	19	23	19	22	21	19	22
Cys-543	7	5	16	5	8	7	12	12
Ala-545	8	7	6	−11	−10	−9	−10	−11
Cys-546	−9	−10	−5	QM	QM	QM	QM	QM
Gly-547	−14	−12	−12	9	4	4	6	7
Mg	8	9	10	8	8	8	7	7
water (sum)	−10	−11	−20	−12	−15	−13	−12	−15

[a] A negative value indicates that the HIP state is energetically favored by the residue. QM indicates that the whole residue was included in the QM system.

that the shift is caused by the extra size in the full system. Therefore, we estimated this effect by calculating the bulk solvation energy from the change in dipole moment (the HID state of the isolated QM system in a vacuum has a larger dipole than the HIP state, 19 and 15 D for NCHACGR and 16 and 14 D in NACGR, respectively), using the Onsager formula.[77] However, this effect is only ∼5 kJ/mol, and it actually increases the shift.

Next, nine Na^+ ions were added to the simulated system to compensate for the different total charge between the truncated and whole protein. The QTCP free energies (Table 1) show that the Na^+ ions favor the HIP state, but their effect is only marginal (∼5 kJ/mol) and cannot explain the energy difference between the full and truncated proteins.

Finally, we studied the MM part of the free energies in more detail. In Figure 5, the cumulative sum of ΔA_{MM} from each residue is plotted as a function of its closest distance to the QM system. It can be seen that the curves from the simulation of the full and truncated proteins differ by 3 kJ/mol already at zero distance (i.e., for the residues that are partly in the QM system). The curves then run roughly in parallel, with differences of less than 6 kJ/mol up to 16.5 Å, at which point the first missing residue in the truncated system is found (these distances are measured from the whole QM system, whereas the protein was truncated on the basis of the distance from the Ni atom alone; moreover, the surrounding protein is flexible). From there on, the difference increases and ends up at 40 kJ/mol (note that neither $\Delta A_{MM \rightarrow QM/MM}$ nor ΔA_{MM} from the water molecules are included in Figure 5).

These differences are caused by at least three effects. First, the charges on the QM atoms differ in the two simulations, because they are estimated from the QM ESP, on the basis of the QM/MM structures, which are slightly different for the full and truncated proteins (both the QM and MM coordinates affect the QM charges, because we used wave functions polarized by the point charges of the protein in the charge calculations). This effect can be estimated by using

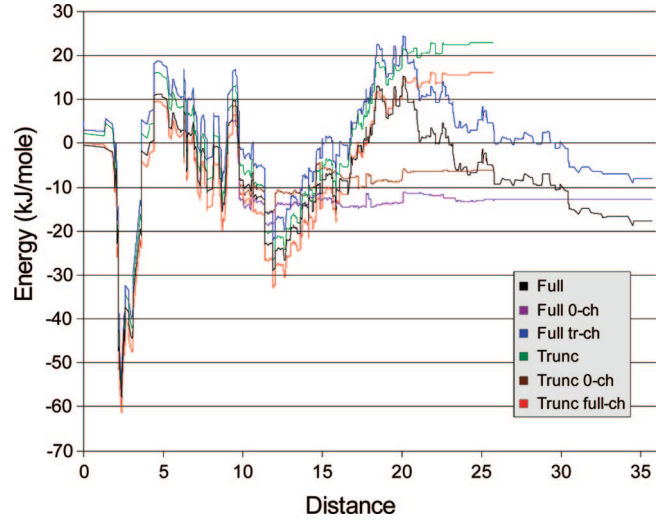


Figure 5. The cumulative residue components of ΔA_{MM} for the NCHACG QM system of [Ni,Fe] hydrogenase as a function of the distance of the residue from the QM system. Six different simulations are shown, three with the full protein (Full) and three with the truncated protein (Trunc). Two used the normal charges, whereas in two, all of the charged residues with a distance greater than 20 Å from the QM system were neutralized (0-ch; note that those curves are identical to the standard-charge curve up to approximately this distance). In the last two simulations, the charges of the QM system were taken from the other simulation (Full trunc-ch and Trunc full-ch).

the QM charges from the truncated protein for the full-protein QTCP calculation, and vice versa (the charges were switched only in the FEP, not in the MD simulations). Those results are also included in Figure 5, and it can be seen that the charges change the free energies by 7−10 kJ/mol, most of which arises from interactions with residues within 5 Å of the active site. This changes the difference between the simulations with the full and truncated protein somewhat, but the difference for the total QTCP free energy is still −27 and −26 kJ/mol, using the full-protein or truncated-protein

Table 4. The Protein Part of the MM Free-Energy Difference (ΔA_{MM}, kJ/mol) between the HIP and HID State in [Ni,Fe] Hydrogenase (NCHACG QM System) As a Function of How Far from the Active Site Solvent-Exposed Charges Are Scaled down by an Effective Dielectric Constant of 80[a]

distance	full	truncated	difference
∞	−17.7	22.8	40.5
25	−4.2	22.9	27.1
20	12.5	19.8	7.3
15	−6.1	−1.4	4.8
10	−7.2	−3.5	3.7
5	−11.6	−8.0	3.6

[a] At 5 Å, all solvent-exposed charges are scaled down (the closest one is at 9.4 Å). The results for simulations with the full and the truncated protein are given, as well as their difference.

charges in both simulations, respectively). Thus, the differences in the QM charges cannot explain the difference between the full and truncated protein.

The second effect is the difference in the coordinates of the two systems, both for the QM and MM atoms (which also indirectly affect the charges, as we saw above). This effect explains the difference between the simulations using the same charges, at distances up to 16.5 Å in Figure 5. It can be seen that the effect is modest, up to 5 kJ/mol.

Consequently, the difference between the full- and truncated-protein simulations actually comes directly from the truncation of the protein, that is, from the long-range electrostatics of deleted residues in the truncated simulations. This is an alarming and problematic observation for theoretical simulations of proteins.

However, the effect comes almost entirely from charged residues: From Figure 5, it can be seen that, if we neutralize all solvent-exposed charged residues, the difference between the full and truncated system is reduced to 6 kJ/mol (10 kJ/mol for the total QTCP free energy). This shows that the difference between the full and truncated protein is almost entirely caused by charged groups at large distances from the active site. Such effects have been much discussed. It is know experimentally that charged solvent-exposed groups have little influence on acid constants, redox potentials, and ligand-binding affinities.[78,79] It has also been observed in theoretical studies of such properties that more accurate and stable results are obtained if interactions with charged solvent-exposed groups are ignored or scaled down with a large effective dielectric constant.[79–82] Therefore, it is likely that this large effect of distant charged groups is an artifact.

The only remaining question is how to treat this problem in practical simulations. Apparently, interactions with solvent-exposed charges should be scaled down, but it remains to be settled how these residues should be selected (e.g., if all solvent-exposed charged residues should be neutralized or only those outside a certain distance from the quantum system). From Table 4, it can be seen that the free-energy difference between the full and truncated protein calculations varies by less than 4 kJ/mol if the distance between the QM system and the down-scaled residues (by $\epsilon = 80$) is varied between 0 and 20 Å. However, the absolute free-energy difference changes drastically. Therefore, we suggest that the only reasonable choice is to scale down all solvent-exposed charged residues, as was done in Figure 5.

Finally, we need also to decide how strongly the charges should be scaled down. Values between ∼20 and infinity have been discussed.[79] However, from Table 4, it is clear that the total effect is only up to 40 kJ/mol, meaning that the various choices differ by less than 2 kJ/mol (from 40/20 to 40/∞ kJ/mol). Therefore, we have decided to simply ignore these interactions, to avoid introducing another, quite arbitrary constant. The results of such QTCP calculations are given in the first column of Table 5 ($\Delta A_{QTCP\_0\text{-}ch}$). The energies are 4−14 kJ/mol less negative than the QTCP energies of the full protein with no charges removed.

Extrapolation. All QM calculations described up to now were performed with the BP86 functional and the DZP/6-31G* basis sets. This basis set is only medium-sized and not fully converged. For accurate results, the QTCP free energies should therefore be extrapolated to calculations employing larger basis sets and perhaps also a more-accurate DFT functional.[18,19] Consequently, we have extrapolated the QTCP results using calculations with the B3LYP functional and the 6-311+G(2d,2p) basis set. In addition, we have performed frequency calculations for the HID and HIP states, optimized in a vacuum to correct the results for the zero-point energies and the entropy and enthalpy of the QM system (the QM system is fixed in the QTCP calculations).[18,19,49]

The results of these calculations are collected in Table 5. It can be seen that the results depend only slightly on the type of the functional and the size of the basis set. The effect of the basis set is 3−8 kJ/mol and that of the functional is 1−6 kJ/mol, both favoring the HIP state (together by 5−14 kJ/mol). On the other hand, the zero-point energy favors the HID state by +5 kJ/mol, whereas the entropy correction is +4 kJ/mol, thus roughly canceling the functional and basis set dependency. If these corrections are combined with our best QTCP estimates ($\Delta A_{QTCP\_0\text{-}ch}$), we arrive at our final estimate that the HIP state is ∼75 kJ/mol more stable than the HID state (with the two largest QM systems).

Comparison with Other Methods. In this section, we will compare the QTCP results with results obtained by other methods, both to get a deeper understanding of the energies and to see if results of similar quality can be obtained with cheaper methods. Let us first go back to the six-step perturbation of the normal QM system in truncated H2ase, which gave a QTCP free-energy difference of 39 kJ/mol in favor of the HIP state. It can be seen from Figure 4 that the original QM/MM energies are quite different, indicating that the HIP state is only 9 kJ/mol more stable than the HID state. The QM/MM curve follows rather closely the QM with point charges (QM+ptch) curve, with a largest difference of 5 kJ/mol. This shows that the difference in the MM correction ($E_{MM123} - E_{MM1}$; cf. eq 1) is less than 5 kJ/mol.

In the figure, it can also be seen that the effect of the point charges ($E_{QM1+ptch} - E_{QM1}$) is less than 8 kJ/mol. Finally, the effect of the polarization of the wave function by the point charges ($E_{QM1pol} - E_{QM1}$) is 2 kJ/mol in all of the intermediates. Interestingly, all three QM energies (E_{QM1}, E_{QM1pol}, and $\Delta A_{MM \rightarrow QM/MM}$) are close to zero, and the E_{QM1} energy actually indicates that the HID state is 4 kJ/mol more

Table 5. Extrapolated Results for the Relative Energy (kJ/mol) of the HID and HIP States in the full [Ni,Fe] Hydrogenase (A Negative Sign Indicates That the HIP State Is More Stable)[a]

QM system	$\Delta A_{QTCP\_0\text{-}ch}$	BP86/B1	BP86/B2	B3LYP/B2	ΔA_{corr}	extrapolated
N	−37.4	3.6	−0.1	−1.1	9.4	−32.7
NR	−47.9	−28.7	−31.6	−33.6		−43.4
NACG	−52.2	−13.2	−20.4	−25.0		−54.5
NACGR	−54.0	−45.2	−50.5	−55.1		−54.5
NCHACG	−71.8	−31.4	−39.4	−45.5		−76.5
NCHACGR	−71.8	−62.2	−68.0	−73.5		−73.7

[a] The table shows vacuum QM energies of the isolated QM system (ΔE_{QM}), calculated with the BP86 and B3LYP methods, and with either the 6-31G* or 6-311+G(2d,2p) basis sets (called B1 and B2). In addition, the zero-point energy difference between the HID and HIP states, as well as the entropy and thermal corrections to ΔA is included, obtained from a harmonic analysis of the vibrational frequencies of the QM system, optimized in a vacuum (ΔA_{corr}). Only the smallest QM system (N) was used in those calculations, because for larger systems, major rearrangements of the second-sphere ligands can be expected. Finally, we have added all corrections to the QTCP free-energy estimate for the full protein with solvent-exposed charges neutralized ($\Delta A_{QTCP\_0\text{-}ch}$), to obtain the extrapolated results in the last column.

Table 6. Relative QM Energies (kJ/mol) of the HIP and HID States (A Negative Energy Indicates That HIP State Is Most Stable) in the Truncated and the Full [Ni,Fe] Hydrogenase with Different Sizes of the QM System[a]

protein	QM system	ΔE_{QM1}	ΔE_{QM1pol}	$\Delta E_{QM1+ptch}$	$\Delta E_{QM1+\epsilon=4}$	ΔE_{PBtot}	$\Delta E_{QM/MM}$	$\Delta E_{QM/MM\_free}$	$\Delta A_{QM/MM\text{-}FE}$	ΔA_{QTCP}
truncated	N	3.6	1.5	−3.8	−54.2	−68.2	−8.5	35.6	−28.4	−37.3
	NR	−28.7	−28.5	−3.4	−70.3	−76.4	−8.7	−14.5	−27.6	−30.6
	NACG	−13.2	−16.3	−17.6	−70.1	−95.4	−4.0	−58.0	−34.4	−40.5
	NACG′	−18.9	−21.0	−16.9	−72.8	−59.1	−19.3	−29.9	−41.3	−43.5
	NACGR	−45.2	−45.7	−18.2	−85.5	−71.9	−1.8	−21.7	−38.8	−41.1
	NCHACG	−31.4	−35.5	−37.7	−78.2	−66.9	−21.3	−41.6	−43.5	−48.7
	NCHACGR	−62.2	−63.7	−34.0	−91.5	−80.4	−34.6	2.4	−43.3	−43.9
full	N	2.3	2.4	−11.8	−56.5	−82.9	−18.2	−28.9	−46.0	−50.6
	NR	−29.9	−30.0	−12.7	−72.0	−85.1	−19.2	−23.5	−44.4	−52.2
	NACG	−15.5	−17.6	−22.7	−72.2	−102.4	−24.5	−18.6	−63.6	−66.0
	NACGR	−46.2	−46.9	−23.1	−85.1	−101.4	−25.6	−37.8	−56.7	−62.9
	NCHACG	−30.7	−33.1	−38.5	−78.1	−101.9	−38.9	−106.3	−71.3	−77.9
	NCHACGR	−63.1	−64.6	−42.3	−93.9	−105.9	−42.6	−54.4	−71.4	−81.8
both	MAD	26.4	25.1	30.3	23.3	32.4	31.5	28.6	5.1	0.0

[a] All energy terms are defined in the Methods section (they are, from the left, the QM energy of the isolated QM system in a vacuum, the QM system polarized by the point-charge model, the QM system with a point-charge model or dissolved in a continuum solvent with $\epsilon = 4$, the QM + Poisson-Boltzmann solvation energy, the QM/MM energy with system 2 fixed at the starting structure or relaxed, as well as the QM/MM-FE and QTCP free energies). The mean absolute difference (MAD) between the various energies and ΔA_{QTCP} is given on the last line.

stable than the HIP state in a vacuum. Thus, the stabilization of the HIP state comes entirely from the surrounding protein (as the ΔA_{MM} energies show), and neither of these energies gives any good approximation to the full QTCP free-energy difference.

However, the QM/MM-FE results are much better. This method uses the same MD simulations and ΔA_{MM} values as QTCP; the only difference between these two methods is in the MM → QM/MM correction ($\Delta A_{MM \rightarrow QM/MM}$). In the more-accurate QTCP approach, this term is estimated by FEP, using ∼200 QM calculations of the same QM system, but with different coordinates of the surrounding point charges. In the QM/MM-FE approach, this term is approximated by a single QM calculation for the reactant and the product (we have used the ΔE_{QM1pol} energy, but the results in Figure 4 show that similar results would have been obtained also with the vacuum ΔE_{QM1} energy). From Figure 4, it can be seen that the E_{QM1pol} energy is within 4 kJ/mol of the $\Delta A_{MM \rightarrow QM/MM}$ term throughout the reaction. Therefore, the QM/MM-FE results are very similar to the QTCP results, as has been seen also before for three other proteins.[19,43,83]

Similar results are obtained in all of the other simulations, as shown in Table 6. The ΔE_{QM1pol} results are always close to the vacuum values (ΔE_{QM1}; maximum deviation 4 kJ/mol). Thus, the polarization of the QM system has a small effect on the relative energies of the HIP and HID states.

All QM-only energies are poor approximations of the QTCP results: The mean absolute deviation (MAD) between ΔE_{QM1} or ΔE_{QM1pol} and QTCP is 26 or 25 kJ/mol, respectively. This is interesting, because most theoretical groups try to estimate reaction energies in proteins using small models of the active site.[1] In fact, ΔE_{QM1} does not converge to QTCP when the system is enlarged (for the largest QM system with 104 atoms, the difference is still ∼20 kJ/mol), and there is no way to decide when convergence is reached. Moreover, we have seen that this is a quite inefficient technique, because hydrogen-bond interactions are well-described by MM.

If the QM system is embedded in a continuum solvent with a dielectric constant of 4 (as is done in many QM-only studies of proteins;[1] $\Delta E_{QM1+\epsilon=4}$), the result is somewhat improved (MAD = 23 kJ/mol, or 20 kJ/mol if compared to ΔA_{QTCP_0ch}). On the other hand, if the protein is modeled by a seemingly more accurate point-charge model ($\Delta E_{QM1+ptch}$), the results get worse (MAD = 30 kJ/mol).

Somewhat unexpectedly, including the solvation of both the surrounding protein and solvent (by solving the Poisson−Boltzmann equation, ΔE_{PBtot} from eq 5), as in

several earlier studies,[5–7,9,53–55] does not improve the results; the MAD from the QTCP results is still 32 kJ/mol. Moreover, this method consistently favors the HIP state.

The original QM/MM results are of similar quality: They underestimate the free-energy difference by 9−42 kJ/mol (average 32 kJ/mol; $\Delta E_{QM/MM}$ in Table 6). This is quite unexpected, because QM/MM typically gives exaggerated electrostatic interactions.[83] The results were not improved if all protein atoms within 15 Å of the QM system and all water molecules were relaxed in the QM/MM optimization ($\Delta E_{QM/MM\_free}$): The energies change much, but they fluctuate more, and the results differ by −28 to +73 kJ/mol from the QTCP free energies, with a similar MAD (29 kJ/mol). This shows that QM/MM energies are sensitive to the conformation of the surrounding protein and solvent, and that a proper sampling of conformations is essential in obtaining stable and reliable energies.

Thus, the only method that gives reasonable results for all models is QM/MM-FE: It closely parallels the QTCP results, with a MAD of only 5 kJ/mol, although the QM/MM-FE results are always more positive than QTCP. Thus, we can conclude that, for H2ase, QM/MM-FE is a reasonable approximation to QTCP that saves much computer time.

Nitrite Reductase. Copper NIR is a bacterial enzyme that catalyzes the one-electron reduction of nitrite to gaseous NO:[65]

$$NO_2^- + e^- + 2H^+ \rightarrow NO + H_2O \quad (7)$$

It contains two copper ions, one electron-transfer blue-copper site, and the catalytic site. In the latter, the copper ion is bound to three His ligands and a solvent molecule. One of the His ligands, His-100, forms a hydrogen bond to the carboxylate group of Glu-279, and the solvent molecule forms a hydrogen bond to the carboxylate group of Asp-98. The status of the solvent molecule (water or OH^-) is important for the catalysis.[8] Therefore, we have in a previous article studied the protonation of this residue and the His-100−Glu-279 pair with vacuum QM and QM/MM calculations, as well as quantum refinements of crystal structures.[8] This gave detailed information about the structures of the various protonation states, but it was not possible to obtain reliable estimates of the relative energies of the four possible protonation states, which we called Both (protons on Asp-98 and Glu-279), Hyd (protons on Asp-98 and His-100), Imm (protons on water and Glu-279), and Wat (protons on water and His-100), cf. Figure 3. On the contrary, different treatments of the electrostatics and solvation gave estimates that differed by up to 93 kJ/mol.[8] Thus, this is an ideal test case for methods that aim at estimating accurate energies in proteins.

The results of the QTCP calculations for NIR are listed in the Table 7. Two sets of calculations were performed, because the protonation state of the His-255 residue is not clear. His-255 is proposed to be part of the proton conducting chain from the active site to the solvent, and its protonation state may vary during the catalytic cycle of NIR[65,84] (cf. Figure 3). Therefore, we run calculations with His-255 protonated either on $N^{\epsilon 2}$ (Hie) or on both nitrogen atoms (Hip). With the Hie-255 state, the calculations predict that the Hyd state is 37, 42, and 73 kJ/mol more stable than the Both, Wat, and Imm states, respectively. If His-255 is protonated on both nitrogen atoms (Hip state), these numbers change by up to 18 kJ/mol to 26, 47, and 55 kJ/mol. The results are rather well-converged for most of the reactions—the hysteresis is less than 10 kJ/mol for ΔA_{MM} in all steps, except for the Both → Imm perturbation in the Hie-255 state (14 kJ/mol). The hysteresis in this step could be reduced to 4 kJ/mol by using a time step of 1 fs and a cutoff of 10 Å for the nonbonded interactions in the MD simulation (instead of the usual values of 2 fs and 8 Å), but the actual QTCP result remained the same within 1 kJ/mol. This hysteresis is caused mainly by the water molecules.

Table 7. QTCP Results for the Four Protonation States in Nitrite Reductase (kJ/mol)[a]

reaction	ΔA_{MM}	hysteresis	$\Delta A_{MM \rightarrow QM/MM}$	ΔA_{QTCP}
		N, Hie-255		
Hyd → Both	87.3	6.1	−50.3	37.0
Imm → Wat	−78.8	9.9	47.3	−31.5
Hyd → Wat	2.5	3.5	39.7	42.2
Both → Imm	−7.5	13.8	42.7	35.3
		N, Hip-255		
Hyd → Both	71.8	1.8	−45.7	26.2
Imm → Wat	−64.1	9.7	58.8	−9.7
Hyd → Wat	12.5	2.8	34.2	46.8
Both → Imm	6.5	5.1	20.1	26.6
		NHK, Hip-255		
Hyd → Both	27.1	2.1	7.6	34.7
Imm → Wat	−16.3	5.7	−2.6	−19.0
Hyd → Wat	2.6	1.7	63.2	65.8
Both → Imm	−8.9	5.5	58.2	49.3

[a] ΔA_{MM} and $\Delta A_{MM \rightarrow QM/MM}$ are the MM and MM → QM parts of the QTCP free-energy difference, respectively (cf. Figure 1). The second column gives the hysteresis in the ΔA_{MM} term. The sum of the ΔA_{MM} and $\Delta A_{MM \rightarrow QM/MM}$ terms yields the final QTCP free energy, ΔA_{QTCP}.

Table 8. The Largest (Approximate) Contributions from Various Residues to the MM Free-Energy Difference (ΔA_{MM}) between the Four States in Nitrite Reductase in the Hip-255 State (kJ/mol) and the Normal QM System[a]

residue	Hyd → Both	Imm → Wat	Hyd → Wat	Both → Imm
Val-133	−6	−10	0	0
His-255	−33	26	22	29
Lys-269	68	−72	6	10
Asp-275	−9	13	0	−4
Asp-277	−13	13	0	−3
Thr-287	16	−18	0	0
Wat-1055	0	5	−7	−15

[a] A negative energy indicates that the product is stabilized.

The residues with the largest QTCP ΔA_{MM} free-energy difference contributions are shown in Table 8. In general, the reactions involving proton transfer between Glu-279 and His-100 (Hyd → Both and Imm → Wat) are much more affected by the surrounding protein than the other two reactions. This can be attributed to the fact that there are only two charged residues within 7 Å of the water ligand, Lys-269 and His-255. The results in Table 8 show that these two residues give the largest contributions to all four reaction energies. The distance from the $N^{\epsilon 2}$ of His-255 to the active-site Cu ion is only ∼3.5 Å, and it forms a hydrogen bond to the water molecule (Wat-1055 in the Figure 3) hydrogen-

bonded to Asp-98. By electrostatics, the doubly protonated His-255 favors states in which the moving protons are as far as possible from His-255, that is, Both and Hyd. Lys-269 forms an ion pair with Glu-279, the acceptor of the proton from His-100, involved in the Hyd → Both and Imm → Wat reactions (Figure 3). As expected, Lys-262 has a strong effect on the protonation state of Glu-279: the deprotonated forms of Glu-279 (Hyd and Wat) are favored by ∼70 kJ/mol over the protonated ones by this residue.

Interestingly, two neutral residues also have large contributions: Val-133 and Thr-287, both favoring a protonated His-100 (the Hyd and Wat states). For Val-133, this is caused by the hydrogen bond from its backbone NH to the other carboxylate oxygen atom of Glu-279. For Thr-287, the Coulomb interaction between the CB atom and the carboxylate oxygen atoms of Glu is the largest factor favoring the protonation of His-100. The negatively charged residues Asp-275 and Asp-277 both favor a proton on Glu-279 (the Both and Imm states).

If His-255 (protonated) and Lys-262 are included in the QM system, the QTCP results change by 9−23 kJ/mol (cf. Table 7). The energetic ordering of the states remains the same as that for the smaller QM system: the Hyd state is lowest, followed by Both (35 kJ/mol), Wat (66 kJ/mol), and Imm (84 kJ/mol). The hysteresis with the larger QM system is somewhat smaller than in the previous calculations, 2−6 kJ/mol.

Finally, we have also extrapolated the QTCP free energies with the 6-311+G(2d,2p) basis set and calculated zero-point and thermal corrections for the QM system from frequency calculations. The results in Table 9 show that the larger basis set changes the energies by up to 6 kJ/mol, whereas the zero-point and thermal effects are less than 3 kJ/mol for the relative energies. We also examined the effect of the solvent-exposed charges by repeating the QTCP calculations with those charges neutralized. These results are also included in Table 9, and they show that the charges change the energy difference by up to 20 kJ/mol. Thus, surface charges have a pronounced effect on the energies also in NIR.

Comparison with Other Methods. Finally, we tried to estimate the relative energies of the four protonation states of NIR with other methods. The results in Table 10 are quite similar to those for H2ase: The vacuum (ΔE_{QM1}) and vacuum-polarized (ΔE_{QM1pol}) energies are poor approximations of the QTCP free energies with MADs of 28 and 36 kJ/mol, respectively (and therefore also the original ComQum-X energies, which differ from ΔE_{QM1} by less than 5 kJ/mol[8]). The same applies to the continuum ($\Delta E_{QM1+\epsilon=4}$) or Poisson−Boltzmann solvated energies (ΔE_{PBtot}), with MADs of 23 and 39 kJ/mol, respectively. However, the results for the former improve considerably if they are compared to ΔA_{QTCP_0ch} instead (MAD = 12 kJ/mol). For all except ΔE_{PBtot}, the results are significantly better for the larger QM system with MADs of 15−20 kJ/mol, and for $\Delta E_{QM1+\epsilon=4}$, the results coincide with ΔA_{QTCP_0ch} within 4 kJ/mol. However, this good performance is probably mainly coincidental, considering that the corresponding MAD for H2ase was 19 kJ/mol and no improvement for the larger QM systems was noticed.

Table 9. Extrapolated Results for the Relative Energies (kJ/mol) of the Four Protonation States in Nitrite Reductase[a]

reaction	ΔA_{QTCP_0ch}	B3LYP/ B1	B3LYP/ B2	ΔA_{corr}	extrapolated
		N, Hie-255			
Hyd → Both	23.7	−30.9	−34.8	−1.0	18.7
Imm → Wat	−25.0	21.8	26.2	0.1	−20.6
Hyd → Wat	27.8	18.7	13.0	1.8	23.9
Both → Imm	26.0	27.7	21.6	2.7	22.6
		N, Hip-255			
Hyd → Both	8.0	−30.9	−34.8	−1.0	3.0
Imm → Wat	−9.4	21.8	26.2	0.1	−5.0
Hyd → Wat	35.6	18.7	13.0	1.8	31.7
Both → Imm	18.9	27.7	21.6	2.7	15.5
		NHK, Hip-255			
Hyd → Both	15.7	5.6	7.2	−1.0	16.3
Imm → Wat	−4.9	−2.6	−2.3	0.1	−4.5
Hyd → Wat	48.2	68.7	67.8	1.8	49.0
Both → Imm	33.6	62.9	62.9	2.7	33.5

[a] The table shows vacuum QM energies of the isolated QM system (ΔE_{QM}), calculated with the B3LYP method and with either the 6-31G* or 6-311+G(2d,2p) basis sets (called B1 and B2). In addition, the zero-point energy difference between the various states, as well as the entropy and thermal corrections to ΔA, is included, obtained from a harmonic analysis of the vibrational frequencies of the QM system, optimized in a vacuum (ΔA_{corr}). Only the smallest QM system (N) was used in those calculations, because for larger systems, major rearrangements of the second-sphere ligands can be expected. Finally, we have added all corrections to the QTCP free-energy estimate for the full protein with solvent-exposed charges neutralized ($\Delta A_{QTCP\_0\text{-}ch}$), to obtain the extrapolated results in the last column.

On the other hand, the simple point-charge model ($\Delta E_{QM1+ptch}$) gives unexpectedly good results with a MAD of only 12 kJ/mol for all 12 reactions and as low as 7 kJ/mol for the large QM system. Thus, in contrast to H2ase, a simple point-charge model seems to work very well for NIR.

The QM/MM energies ($E_{QM/MM}$) give a MAD of 18 kJ/mol, which is somewhat better than that for $E_{QM/MM}$ in H2ase but inferior to that for the point-charge model ($E_{QM1+ptch}$). Relaxing system 2 for each of the four states (E_{QM/MM_free}) gives much worse results. This is because the environment ends up in different local minima. Thus, it is clear that a proper sampling of the surroundings is required.

As usual, the QM/MM-FE results are similar to the QTCP ones, but the deviations are larger than for H2ase and other systems,[18,19,49,83] up to 20 kJ/mol (MAD = 8 kJ/mol). These larger deviations probably come from the large effect of polarization in NIR: In fact, ΔE_{QM1} and ΔE_{QM1pol} show larger differences than we have seen for any system before, up to 45 kJ/mol, but only for the small QM system. This indicates that the QM system is strongly polarized by Lys-269 (there is little difference between the Hip-255 and Hie-255 calculations). In fact, $\Delta A_{MM \rightarrow QM/MM}$ lies in between ΔE_{QM1} and ΔE_{QM1pol}, but QM/MM-FE based on ΔE_{QM1} is significantly worse (MAD = 14 kJ/mol), with a maximum difference of 33 kJ/mol. This shows the advantage of using ΔE_{QM1pol} in the QM/MM-FE approximation.

Conclusions

In this paper, we have studied two similar proton-transfer reactions between a metal ligand and a second-sphere group

Table 10. Relative Energies (kJ/mol) of the Four Protonation States in Nitrite Reductase, Using the Normal (N) or the Enlarged (NHK) QM System[a]

state	ΔE_{QM1}	ΔE_{QM1pol}	$\Delta E_{QM1+ptch}$	$\Delta E_{QM1+\epsilon=4}$	ΔE_{PBtot}	$\Delta E_{QM/MM}$	$\Delta E_{QM/MM\_free}$	$\Delta A_{QM/MM-FE}$	ΔA_{QTCP}
				N, Hie-255					
Hyd → Both	−30.8	−54.9	39.7	−10.4	17.1	39.1	128.6	32.4	37.0
Imm → Wat	21.8	67.0	−31.4	19.0	−5.4	8.4	0.2	−11.8	−31.5
Hyd → Wat	18.5	41.4	46.6	27.9	35.8	49.4	28.9	43.9	42.2
Both → Imm	27.5	29.2	38.3	19.2	24.1	1.9	−100.0	21.7	35.3
				N, Hip-255					
Hyd → Both	−30.8	−47.4	2.2	−10.4	−42.7	32.4	−115.4	24.5	26.2
Imm → Wat	21.8	66.9	1.8	19.0	75.9	−1.6	−7.0	2.9	−9.7
Hyd → Wat	18.5	45.6	82.0	27.9	90.0	66.2	−83.9	58.1	46.8
Both → Imm	27.5	26.0	78.0	19.2	56.9	35.3	38.5	32.5	26.6
				NHK, Hip-255					
Hyd → Both	5.5	5.6	44.9	19.3	−40.4	39.1	−73.9	32.6	34.7
Imm → Wat	−2.5	1.4	−25.7	−7.1	−31.1	8.4	−28.0	−14.9	−19.0
Hyd → Wat	68.7	76.9	60.9	48.1	40.4	49.4	−70.0	79.5	65.8
Both → Imm	65.7	69.9	41.7	35.8	111.9	1.9	31.8	61.0	49.3
MAD	27.9	35.9	13.5	23.2	38.9	18.4	69.1	8.6	0.0

[a] All energy terms are defined in the Methods section (they are, from the left, the QM energy of the isolated QM system in a vacuum, the QM system polarized by the point-charge model, the QM system with the point-charge model or dissolved in a continuum solvent with $\epsilon = 4$, the QM + Poisson-Boltzmann solvation energy, the QM/MM energy with the same or different MM environments for each state, as well as the QM/MM-FE and QTCP free energies). The mean absolute difference (MAD) between the various energies and the QTCP free energy is given on the last line.

with QM/MM free-energy perturbations. Previous studies have shown that these reactions strongly depend on the surrounding protein and that it is very hard to obtain converged and reliable energies for them.[8,25] This is of course a major problem in the theoretical modeling of enzyme mechanisms, because similar second-sphere interactions are found in most metalloenzymes and the reactivity of the metal site strongly depends on the exact position of this proton.[85–88] Therefore, it is of great interest to develop methods that can accurately predict the position of shared protons.

Our results show that it is mandatory to model the surrounding protein and the solvent in order to obtain reliable results. The proton-transfer energies strongly depend on the electrostatic interactions from the surroundings, and the effect is very long-range. In fact, Figure 5 shows that the cumulative energy does not start to converge until residues up to 10−15 Å from the active site have been included. This shows that QM calculations on isolated models of the active site will never give the correct results[1] (for the right reason), not even if the surroundings are modeled as a continuum solvent with a low dielectric constant (Tables 6 and 10 show that the MAD is 23−28 kJ/mol for ΔE_{QM1} and $\Delta E_{QM1+\epsilon=4}$).

A point-charge model seems to give reasonable results for NIR (Table 10), but not for H2ase. The same applies to the original QM/MM results for H2ase, which are quite similar to the $\Delta E_{QM1+ptch}$ results. This shows that averaging over many protein and solvent geometries also is important. Poisson−Boltzmann solvation models with an effective dielectric of ∼4 have previously been applied to many systems with rather good results,[5–7,9,53,54,56] but the results in Tables 6 and 10 show that such a procedure works unexpectedly poorly both for H2ase and NIR—in fact, it does not give any significant improvement compared to the vacuum calculations. Thus, this popular method seems to be useless for reactions of the type studied in this paper.

The only methods that give reliable results are full QM/MM free-energy perturbation (QTCP) and QM/MM-FE, which both give a full detailed account of the surroundings and sample their dynamic effects at the MM level. For H2ase, QM/MM-FE gives similar good results to those obtained previously for three other proteins,[19,49,83] with differences of 10 kJ/mol or less. However, for NIR, the QM/MM-FE results are slightly worse (errors of up to 20 kJ/mol). This problem can be attributed to an unusually strong polarization of the QM system, which can be identified by an unusually large difference between the ΔE_{QM1} and ΔE_{QM1pol} energies, and it can be solved by enlarging the QM system.

In conclusion, FEP is needed to obtain accurate and reliable energies in proteins, even at a QM/MM level. An advantage with FEP at the MM level is that it allows us to determine approximate contributions to the free energy, both from the various components in the MM energy function (bonded, van der Waals, and electrostatics) and from the various residues in the protein. The latter can be used to identify in an unbiased way important groups in the protein, which may be included in the QM system. Our results indicate that, as a rule of thumb, electrostatic interactions and hydrogen bonds are rather well-described by MM (and therefore need not be included in the QM system), whereas if MM atoms covalently bound to the QM system give large energy contributions, they should be included in the QM system.

A natural question is how accurate the QTCP results are. We have run many calculations with variations in the QM system, the MM system, the simulation protocol, and various parameters. These allow us to obtain an estimate of the final energy. First, the hysteresis in the calculation of ΔA_{MM} gives an estimate of the convergence of this FEP. We obtained hystereses of up to 4 kJ/mol for H2ase and up to 14 kJ/mol for NIR for single-step perturbations. Of course, the hysteresis can be reduced by dividing the reaction into more steps, as was done for H2ase (Figure 4).

Moreover, Tables 1 and 7 show that the size of the QM system strongly affects the results. However, if the QM

system is enlarged according to our rules above, the variations in the final free energies are 5 kJ/mol for H2ase and up to 23 kJ/mol for NIR, that is, similar to the hysteresis. Likewise, reasonable variations in the simulation protocol have a similar influence on the final energies: The QM charges affect the energies by 5 kJ/mol (Figure 5), and the state, for which the MM system was equilibrated, affects the energies by 8 kJ/mol.

However, the largest effect was observed for the surface charges. Table 4 shows that distant charges may change the energy by up to 37 kJ/mol in H2ase. For NIR, the effect is somewhat smaller, up to 20 kJ/mol. We attribute this effect to an incomplete solvation of these groups, in line with previous experimental and theoretical observations and the common use of neutralizing solvent-exposed charges.[78–82] If these charges are neutralized or scaled down by an effective dielectric constant of 40−80, the uncertainty from surface charges is reduced to ∼5 kJ/mol, as was seen in Table 4. Thus, the inherent uncertainty of QTCP seems to be ∼10 kJ/mol, to which should be added the hysteresis of the ΔA_{MM} term, which in our case gives an uncertainty of the final QTCP results of 10 kJ/mol for H2ase and 15 kJ/mol for NIR.

With this estimate, we can finally conclude that the shared proton in H2ase certainly prefers to bind to His-79, rather than to Cys-546 (by ∼75 kJ/mol), although the two states are almost degenerate in vacuum calculations with our normal QM model N (cf. Table 6). In most previous theoretical investigations of H2ase, His-79 was not considered, because the most common QM system includes only the first-sphere metal ligands (i.e., our N model without Glu-25 and His-79).[76,89] However, Stadler et al.[90] considered all possible protonation states of His-79 and concluded that a complex with a deprotonated Cys-546 and His-79 singly protonated on the $N^{\epsilon 2}$ atom (i.e., with one proton less than the states considered in this paper) reproduces experimental EPR parameters best for the oxidized unready A state of the protein. Siegbahn has studied even larger QM models (N + Arg-476 and an Asp formyl group forming hydrogen bonds with Arg-476) and considered the full reaction mechanism for both a neutral and a positively charged His-79.[71,76] He found quite large changes in the energetics, especially for the proton- and electron-transfer steps, with a doubly protonated His-79 giving the best agreement with experiments. However, he assumed that the proton stays on His-79 throughout the reaction and does not present any results for the HID state.

In our previous investigation,[25] we also considered all possible protonation states of His-79 and concluded that the HIP and HID states are close in energy and that protonation of the Cys-546−His-79 system is competitive compared to the protonation of the other three Cys ligands of the catalytic nickel ion, although the energies were sensitive to the surroundings and to the theoretical method. In this paper, we have developed a reliable method that may allow us to continue this investigation of how protons are transported to and from the active site in H2ase.

Likewise, we can conclude that the catalytic copper ion in NIR is in the Hyd state, that is, that the proton shared by His-100 and Glu-279 resides on His and that the proton shared by Asp-98 and the copper-bound water resides on Asp-98. This state is ∼16 kJ/mol more stable than the Both state (with Glu-279 protonated), although the latter state is 31 kJ/mol more stable in a vacuum with the normal size of the QM system (Table 10) and also more stable with several other methods.[8] However, the stability of the Hyd state is confirmed by the absence of any Cu-to-imidazolate charge-transfer transitions in the electronic spectra of the catalytic copper ion.[67] This conclusion is independent of the protonation state of His-255 (although the exact energy estimate will vary).

Again, this result has strong bearings on the study of the reaction mechanism of this enzyme. Most importantly, it shows that the copper ligand is deprotonated in the oxidized resting state, but it also shows that Asp-98 may serve as a reservoir and relay for the protons involved in the reaction (cf. eq 7). All previous theoretical studies of NIR have omitted Glu-279 and Lys-269 in the QM system,[91] although some have included Asp-98 and sometimes also His-255[92] and even Ile-257.[93,94] Our results show that a neutral imidazole group is a reasonable model for His-100, but it is likely that the negatively charged Glu-279 may tune the properties of the copper site, partly neutralizing the charge of the copper ion.

Finally, it seems appropriate to ask why proton-transfer reactions of this type are so sensitive to the surrounding protein, although the geometric difference between the two studied states is minimal (the proton moves 0.9 Å, some atoms in the His-79 ring move up to 0.3 Å, whereas only the Cys-546 S^{γ} atom and the oxygen atom of CO move more than 0.1 in the QM system in H2ase; for NIR, the difference is even smaller—the proton moves by 0.5 Å, whereas only the acceptor oxygen and $H^{\epsilon 1}$ in the His-100 moves more than 0.1 Å in the Hyd → Both transition). The answer is clearly the large change in dipole moment during the proton transfer: We have seen that, for H2ase, the HID state has a 2−4 D higher dipole moment, whereas in NIR, the Hyd state has a 6 D higher dipole moment than the Both state. Thus, it is clear that, if such large changes in the dipole moment are encountered, the results will strongly depend on the surroundings, and free-energy perturbation techniques combined with QM/MM are needed for accurate energies.

Acknowledgment. This investigation has been supported by grants from the Swedish research council, the Wenner-Gren foundation, the Crafoord foundation, and by computer resources of Lunarc at Lund University.

References

(1) Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* **2000**, *100*, 421–437.

(2) Mulholland, A. J. In *Theoretical biochemistry - processes and properties of biological systems*; Eriksson, L. A., Ed.; Elsevier Science: Amsterdam, The Netherlands, 2001; Vol. 9, pp 597−653.

(3) Ryde, U. *Curr. Opin. Chem. Biol.* **2003**, *7*, 136–142.

(4) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173–290.

(5) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochem.* **1993**, *32*, 8045–8056.

(6) Noodleman, L.; Lowell, T.; Han, W.-G.; Li, J.; Himo, F. *Rev. Chem.* **2004**, *104*, 459–508.

(7) Ullmann, G. M.; Noodleman, L.; Case, D. A. J. *Biol. Inorg. Chem.* **2002**, *7*, 632–639.

(8) Källrot, N.; Nilsson, K.; Rasmussen, T.; Ryde, U. *Int. J. Quantum Chem.* **2005**, *102*, 520–541.

(9) Olsen, L.; Rasmussen, T.; Hemmingsen, L.; Ryde, U. *J. Phys. Chem. B* **2004**, *108*, 17639–17648.

(10) Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974–2975.

(11) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72–79.

(12) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516–17524.

(13) Olsson, M. H. M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.

(14) Zhang, Y.; Liu, H.; Yang, W. J. *Chem. Phys.* **2000**, *112*, 3483–3492.

(15) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2003**, *125*, 12035–12048.

(16) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715–17733.

(17) Li, G.; Zhang, X.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643–8653.

(18) Rod, T. H.; Ryde, U. *Phys. Rev. Lett.* **2005**, *94*, 138302.

(19) Rod, T. H.; Ryde, U. *J. Chem. Theory Comput.* **2005**, *1*, 1240–1251.

(20) Fisher, C. L.; Chen, J.-L.; Li, J.; Bashford, D.; Noodleman, L. *J. Phys. Chem.* **1996**, *100*, 13498–13505.

(21) Ryde, U. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153–164.

(22) Ryde, U.; Olsson, M. H. M. *Int. J. Quantum Chem.* **2001**, *81*, 335–347.

(23) Reuter, N. I.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem.* **2000**, *104*, 1720–1735.

(24) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

(25) Söderhjelm, P.; Ryde, U. *J. Mol. Struct. (Theochem)* **2006**, *770*, 199–219.

(26) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098–3100.

(27) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822–8824.

(28) Hehre, W. J.; Radom L.; Schleyer P. v. R.; Pople, J. A. In *Ab initio molecular orbital theory*; Wiley-Interscience: New York, 1986; pp 65−88.

(29) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(30) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

(31) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283–290.

(32) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119–126.

(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(34) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345–351.

(35) Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346–354.

(36) Jensen, F. In *Introduction to Computational Chemistry*; John Wiley & Sons: New York, 1993; pp. 298−308.

(37) Klamt, A.; Schüürmann, J. *J. Chem. Soc., Perkin Trans. 2* **1993**, *5*, 799–805.

(38) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193.

(39) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem.* **1998**, *102*, 5074–5085.

(40) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, 2004.

(41) Cornell, W. D.; Cieplak, P. I.; 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.

(42) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(43) Sigfridsson, E.; Ryde, U. *J. Comput. Chem.* **1998**, *19*, 377–395.

(44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(46) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(47) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8592.

(48) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(49) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452–461.

(50) Chan, H. S. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 543–571.

(51) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704–721.

(52) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473–486.

(53) Li, J.; Nelson, M. R.; Peng, C. Y.; Bashford, D.; Noodleman, L. J. *Phys. Chem. A* **1998**, *102*, 6311–6324.

(54) Rabenstein, B.; Ullman, G. M.; Knapp, E.-W. *Biochemistry* **2000**, *39*, 10487–10496.

(55) Popovic, D. M.; Quenneville, J.; Stuchebrukhov, A. A. *J. Phys. Chem. B* **2005**, *109*, 3616–3626.

(56) Sitkoff, T.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(57) Ullmann, G. M.; Noodleman, L.; Case, D. A. *J. Biol. Inorg. Chem.* **2002**, *7*, 632–639.

(58) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochemistry* **1993**, *32*, 8045–8056.

(59) Volbeda, A.; Montet, Y.; Vernède, X.; Hatchikian, E. C.; Fontecilla-Camps, J. C. *Int. J. Hydrogen Energy* **2002**, *27*, 1449–1461.

(60) Seminario, J. M. *Int. J. Quantum Chem., Quant. Chem. Sympos.* **1996**, *30*, 59–65.

(61) Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. *Acta Crystallogr., Sect. D* **2003**, *59*, 274–289.

(62) Pierloot, K.; De Kerpel, J. O. A.; Ryde, U.; Olsson, M. H. M.; Roos, B. O. *J. Am. Chem. Soc.* **1998**, *120*, 13156–13166.

(63) Frey, M.; Fontecilla-Camps, J. C. A.; Volbeda, A. In *Handbook of metalloproteins*; Messerschmidt, A., Huber, R., Poulos, T., Wieghardt, K., Eds.; Wiley: New York, 2001; pp 880−896.

(64) Huyett, J. E.; Carepo, M.; Pamplona, A.; Franco, R.; Moura, I.; Moura, J. J. G.; Hoffman, B. M. *J. Am. Chem. Soc.* **1997**, *119*, 9291–9292.

(65) Adman, E. T.; Godden, J. W.; Turley, S. *J. Biol. Chem.* **1995**, *270*, 27458–27474.

(66) Ryde, U.; Olsen, L.; Nilsson, K. *J. Comput. Chem.* **2002**, *23*, 1058–1070.

(67) Adman, E. T.; Murphy, M. E. P. In *Handbook of Metalloproteins*; Messerschmidt, A., Huber, R., Poulos, T., Wieghart, K., Eds.; J. Wiley & Sons: Chichester, U.K., 2001; pp 1381−1390.

(68) De Kerpel, J. O. A.; Ryde, U. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 157–174.

(69) Vignais, P. M.; Billoud, B.; Meyer, J. *FEMS Microbiol. Rev.* **2001**, *25*, 455–501.

(70) Volbeda, A.; Fontecilla-Camps, J. C. *Coord. Chem. Rev.* **2005**, *249*, 1609–1619.

(71) Siegbahn, P. E. M. *Adv. Inorg. Chem.* **2004**, *56*, 101–125.

(72) Stein, M.; Lubitz, W. *Curr. Opin. Chem. Biol.* **2002**, *6*, 243–249.

(73) Fan, H.-J.; Hall, M. B. *J. Biol. Inorg. Chem.* **2001**, *6*, 467–473.

(74) Bruschi, M.; Zampella, G.; Fantucci, P.; De Gioia, L. *Coord. Chem. Rev.* **2005**, *249*, 1620–1640.

(75) Amara, P.; Volbeda, A.; Fontecilla-Camps, J. C.; Field, M. J. *J. Am. Chem. Soc.* **1999**, *121*, 4468–4477.

(76) Siegbahn, P. E. M.; Tye, J. W.; Hall, M. B. *Chem. Rev.* **2007**, *107*, 4414–4435.

(77) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

(78) André, I.; Kesvatera, T.; Jönsson, B.; Åkerfeldt, K. S.; Linse, S. *Biophys. J.* **2004**, *87*, 1929–1938.

(79) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400–417.

(80) Mehler, E. L.; Eichele, G. *Biochemistry* **1984**, *23*, 3887–3891.

(81) Penford, R.; Warwicker, J.; Jönsson, B. *J. Phys. Chem. B* **1998**, *108*, 8599–8610.

(82) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786–3791.

(83) Heimdal, J.; Rydberg, P.; Ryde, U. *J. Phys. Chem. B* **2008**, *112*, 2501–2510.

(84) Boulanger, M. J.; Kukimoto, M.; Nishiyama, M.; Horinouchi, S.; Murphy, M. E. P. *J. Biol. Chem.* **2000**, *275*, 23957–23964.

(85) Poulos, T. L. *J. Biol. Inorg. Chem.* **1996**, *1*, 356–359.

(86) De Santis, L.; Carloni, P. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 611–618.

(87) Jensen, K. P.; Ryde, U. *J. Mol. Struct. (Theochem)* **2002**, *585*, 239–255.

(88) Jensen, K. P.; Ryde, U. *Mol. Phys.* **2003**, *101*, 2003–2018.

(89) Pardo, A.; De Lacey, A. L.; Fernández, V. M.; Fan, H.-J.; Fan, Y.; Hall, M. B. *J. Biol. Inorg. Chem.* **2006**, *11*, 286–306.

(90) Stadler, C.; De Lacey, A. L.; Montet, Y.; Volbeda, A.; Fontecilla-Camps, J. C.; Conesa, J. C.; Fernández, V. M. *Inorg. Chem.* **2002**, *41*, 4424–4434.

(91) Silaghi-Dumitrescu, R. *J. Inorg. Biochem.* **2006**, *100*, 396–402.

(92) De Marothy, S. A.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Comput. Chem.* **2007**, *28*, 528–539.

(93) Sundarajan, M.; Surendran, R.; Hillier, I. H. *Chem. Phys. Lett.* **2006**, *418*, 96–99.

(94) Periyasamy, G.; Sundarajan, M.; Hillier, I. H.; Burton, N. A.; McDouall, J. J. W. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2498–2506.

CT700347H

JCTC Journal of Chemical Theory and Computation

Gas-Phase Interaction of Calcium (Ca^{2+}) with Seleno Derivatives of Uracil

Al Mokhtar Lamsabhi,*,† Otilia Mó,† Manuel Yáñez,† and Russell J. Boyd‡

Departamento de Química C-9, Facultad de Ciencias, Universidad Autónoma de Madrid, Cantoblanco, 28049-Madrid, Spain, and Department of Chemistry, Dalhousie University, Halifax, Nova Scotia, B3H 4J3 Canada

Received January 16, 2008

Abstract: The structures and relative stabilities of the complexes between Ca^{2+} and 2-selenouracil, 4-selenouracil, and 2,4-diselenouracil have been investigated through the use of B3LYP/6−311++G(3df,2p)//B3LYP/6−31+G(d,p) density functional theory (DFT) calculations. In those systems where both types of basic centers, a carbonyl or a selenocarbonyl group, are present, Ca^{2+} association with the oxygen is favored. For 2,4-diselenouracil the nitrogen atom at position 3 is the most basic site toward Ca^{2+} attachment followed by heteroatoms attached to positions 4 and 2. Although the enolic and selenol forms of selenouracils should not be observed in the gas phase, the corresponding Ca^{2+} complexes are the most stable ones. More importantly, all the activation barriers associated with the corresponding tautomeric processes are lower than the entrance channel, and therefore not only these complexes should be observed but also they should be the dominant species in the gas phase. Also, Ca^{2+} association has a clear catalytic effect on these tautomerization processes, whose activation barriers decrease between 10 and 15 kcal mol^{-1}.

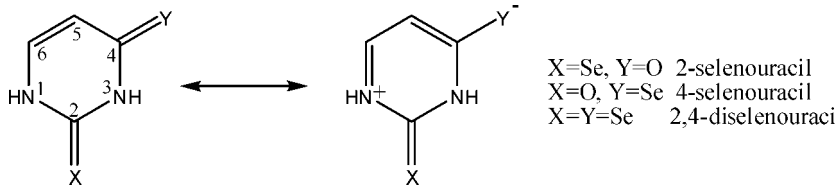
Introduction

Genetic information storage, gene expression, and catalysis are some of the important biological functions in living systems in which nucleic acids participate.[1–4] Watson and Crick base pairing and π stacking allow the formation of stable duplexes that are central to genetic information storage, transcription, and replication. The replacement of the oxygen on the nucleobases with sulfur[5,6] has provided insight into DNA duplex stability, recognition, and replication at the atomic level.[7,8] Recent studies on these sulfur modifications have revealed enhanced base-pairing selectivity[9]and replication efficiency and fidelity, especially with the 2-thiothymidine.[10] What would be the implications if oxygen is replaced by selenium? As selenium is in the same group, but is much larger than oxygen, the replacement of O by Se will provide insight into base pairing selectivity. Recently, Salon et al.[11,12] have synthesized some Se derivatives of the DNA nucleobases and studied their crystal structures, their thermostabilities, and the impact of their incorporation into oligonucleotides. In addition, specific pyrimidines in natural tRNAs have been derivatized by incorporation of Se on the nucleobases. The functionality of Se is not well understood although it was suggested that Se might be involved in the tRNA anticodon.[13–15] In view of the increased interest in nucleobase seleno derivatives, our attention has been directed to the exploration of the reactivity changes caused by replacing oxygen by selenium in small biochemical systems. In this respect we have recently found, for instance, that the reactivity of selenourea with respect to Ca^{2+} mimics that of thiourea but presents significant dissimilarities with respect to urea.[16] The aim of this paper is to investigate the Se derivatives of uracil, namely, 2-selenouracil (2SeU), 4-selenouracil (4SeU), and 2,4-diselenouarcil (24dSeU), in order to contribute to the understanding of its involvement in biological compounds when interacting with Ca^{2+}. It is well-known that Ca^{2+} takes part in a wide range of biological processes, including the regulation of muscle contraction, transduction, glycolysis, gluconeogenesis, ion transport, and the stabilization of interprotein complexes.[17,18] This moti-

* Corresponding author e-mail: mokhtar.lamsabhi@uam.es.
‡ Dalhousie University.
† Universidad Autónoma de Madrid.

10.1021/ct800017j CCC: $40.75
Published on Web 05/10/2008

Scheme 1



vated a growing interest in the study of the interaction of biochemical systems with this metal dication, both from the experimental and the theoretical viewpoints.[19–31]

Computational Details

Geometries were optimized by using density functional theory with the hybrid functional B3LYP[32,33] as implemented in the Gaussian 03 suite of programs,[34] in conjunction with the 6−31+G(d,p) basis set. Harmonic vibrational frequencies were computed (at the same level) to classify stationary points as local minima or transition structures (TS) and to estimate their zero-point vibrational energy (ZPVE) corrections (scaled by 0.986).[35] In order to obtain more reliable energies for the local minima, single point energies have been evaluated by using the same functional combined with the 6−311++G(3df,2p) basis set at B3LYP/6−31+G(d,p) geometries.

The corresponding Ca^{2+} binding energies, D_0, were evaluated by subtracting from the energy of the most stable complex the energy of the neutral and that of Ca^{2+}, after including the corresponding ZPE corrections. The basis set superposition error (BSSE) was not included in the calculation of D_0, because as it has been previously reported that for DFT and DFT/HF hybrid methods this error is usually small, when the basis set expansion is sufficiently flexible.[36]

The bonding characteristics were analyzed by means of the atoms in molecules (AIM) theory.[37,38] For this purpose we have located the relevant bond critical points (BCP) and evaluated the electron density for each of them, by means of the AIMPAC series of programs.[39] With the aim of further exploring the nature of calcium bonding in these complexes, we turn here to the usefulness of the topological analysis of the electron localization function (ELF), a direct measure of the local Pauli principle. The reader is referred to several reviews on this powerful technique of bonding analysis.[40–42] As the ELF is a scalar function, the analysis of its gradient field can be carried out in order to locate its attractors (the local maxima) and the corresponding basins. To carry out these calculations the TopMod suite of programs has been used.[43]

Results and Discussion

Structure and Stability of Selenouracils−Ca^{2+} Adducts. Since selenouracils may exist in several tautomeric forms, the first question to be addressed, in order to rationalize their intrinsic reactivity is which tautomers are predominant in the gas phase. Previous studies have shown,[44,45] that oxo-seleno and diseleno are the most stable forms and that the energy barriers connecting them with other tautomers are very high, and therefore the oxo-seleno and diseleno forms will be the only ones present in the gas phase under normal conditions.

As expected the heteroatoms bonded to carbons at positions 2 and 4 are the most favorable sites for electrophilic attack, as was previously found for uracil and its thio-derivatives.[46–51] These interactions lead to complexes **1** and **4,** as presented in Figure 1. Another possibility is π bonding, where the metal lies perpendicular to the plane of the molecule, as it has been found[47,52,53] for the thio-analogues. However, for selenouracils all attempts to locate a conventional π-complex failed, as they collapsed to structure **7** where the metal interacts with the lone pair of the nitrogen at position 3. Conversely, structure **7** has never been observed before in uracil−M^{2+} complexes (where M = Be, Mg, Ca, Cu)[48,52,53] where the topology of the electron density favors a π interaction. This is not surprising if we consider that oxygen is much more electronegative than selenium, so when oxygen is replaced by selenium there is an accumulation of electron density on the N3 lone pair. To this first effect, a second important one is added. Selenium is much more volumi-

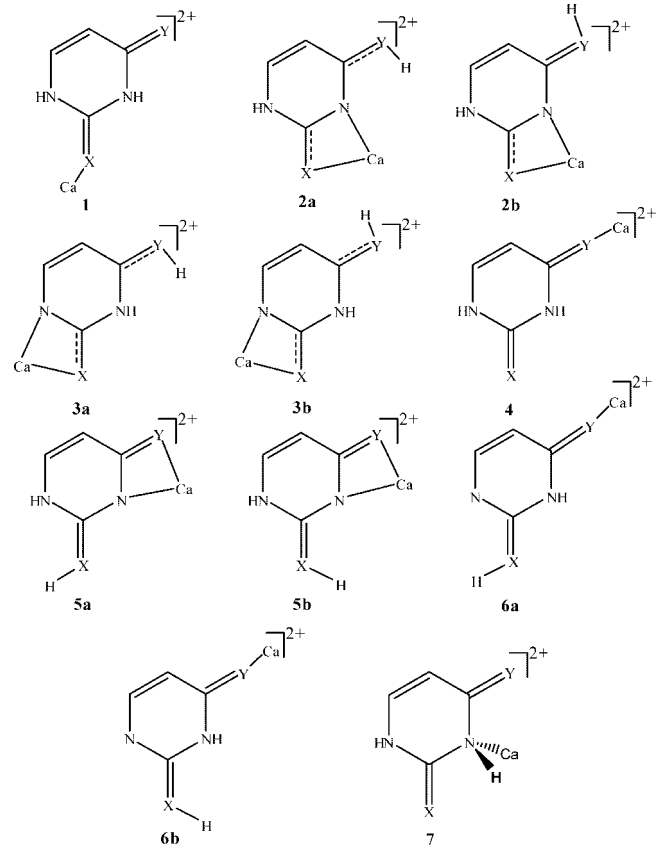


Figure 1. Schematic representation of different tautomers of selenouracil−Ca^{2+} complexes in all possible conformers.

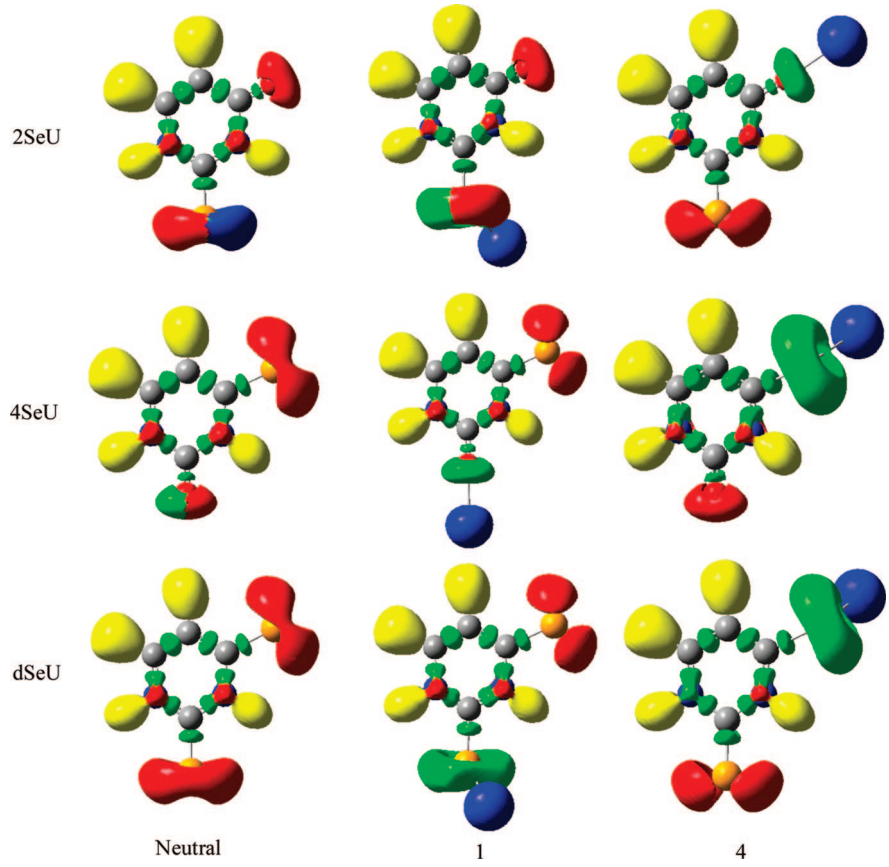


Figure 2. Schematic representation of attractors localized in neutral 2- and 4-selenouracils and their associated Ca^{2+} complexes **1** and **4** at ELF = 0.8. Yellow lobes correspond to V(C,H) or V(N,H) basins, and red lobes correspond to V(O) or V(Se) basins associated with O or Se lone pairs. Green lobes correspond to V(C,C), V(C,O), V(C,N), and V(C,Se) basins. Blue lobes correspond to the metal core, C(Ca).

nous and much more polarizable than oxygen, as clearly reflected in the ELF description of the neutral compounds (see Figure 2). Therefore, the interaction with the metal dication is much more favorable when the molecule contains only seleno groups (24dSeU). In fact, in the case of diselenouracil, complex **7** is more stable than the complex in which the metal interacts with the heteroatom at position 4 by approximately 2 kcal/mol (see Table 1). The strong polarization of both Se atoms is mirrored in the existence of BCPs between them and Ca^{2+} (see Figure 3). Conversely, the presence of a carbonyl group in the molecule significantly decreases the stability of structure **7** which, for 2-selenouracil, becomes 21.7 kcal mol^{-1} less stable than the complex in which Ca^{2+} binds the carbonyl group.

Table 1 also shows that for 2,4-diselenouracil, complex **4** is 5.9 kcal mol^{-1} more stable than complex **1**. A similar behavior was reported for the protonation[46] and the Cu^{+} association[47] of uracil and 2,4-dithiouracil, reflecting the contribution of zwitterionic resonance structures (see Scheme 1) which accumulate negative charge on the heteroatom at position 4. It is worth noting that this resonance structure also locates a positive charge at N1, explaining why π-type complexes evolve to an N3 attached species (complex **7**) but never to a N1 attached structure.

This explanation in terms of the contribution of the zwiterionic forms of Scheme 1, which may be valid for 2,4-diselenouracil, 2,4-dithiouracil, and uracil, where the heteroatoms at positions 2 and 4 are identical cannot be extended to the 2- and 4-derivatives, in which the heteroatom's nature at these positions differs. Indeed, as shown in Table 1 for 2SeU complex **4** is more stable (16.6 kcal mol^{-1}) than complex **1**; but for 4SeU both complexes are nearly degenerate, indicating that the ability of the basic center to bind Ca^{2+} depends not only on its position within the ring but also on

Table 1. Relative Energies (ΔE, kcal mol^{-1}) of the Different Stationary Points of the [Selenouracil−Ca]$^{2+}$ Complexes

	2-selenouracil ΔE	4-selenouracil ΔE	2,4-diselenouracil ΔE
1	**31.5**	25.1	33.6
2a	8.3	2.2	3.5
2b	**0.0**	0.0	0.8
3a	7.6	0.4	0.1
3b	**3.9**	0.2	0.0
4	**14.9**	25.7	27.7
5a	**1.1**	8.2	4.3
5b	3.1	12.3	6.9
6a	17.6	30.2	28.4
6b	18.6	39.2	29.2
7	**36.8**		**25.9**
TS1_4		25.7	
TS1_7	36.8		36.7
TS7_4	35.3		34.0
TS7_5a			51.6
TS7_2d			50.7
TS5a-4	37.8	63.3	51.6
TS2b-1	62.4	43.8	52.8

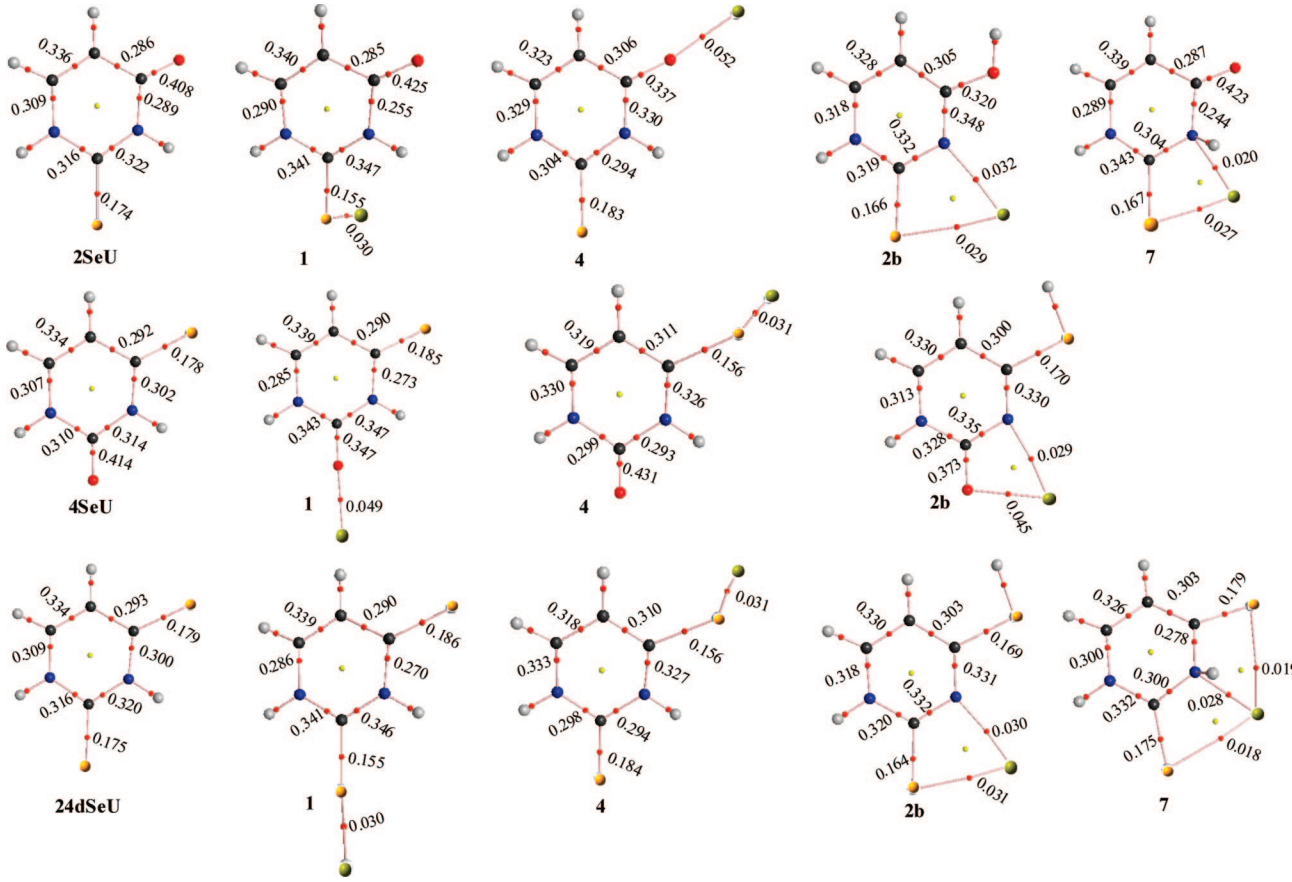


Figure 3. Molecular graphs of some selected selenouracil−Ca^{2+} complexes. Red dots represent bond critical points and yellow dots ring critical points. Electron densities are in au.

its nature. As a matter of fact the binding energy when Ca^{2+} is attached to the oxygen atom at position 4 in uracil (−104.7 kcal mol^{-1})[54] is greater than that computed for 4SeU (−94.5 kcal mol^{-1}), where this position is occupied by a Se atom. In order to separate both effects we have considered it useful to add to the set of uracil and selenouracils the subset of noncyclic molecules which contain the same basic sites in similar molecular environments, namely, acetamide, urea, and their Se-containing analogues. The corresponding binding energies are plotted in Figure 4, and the calculated values are reported in Table S3 of the Supporting Information. In this figure, points named [Ca4,O2] and [Ca4,Se2] correspond to complexes in which Ca^{2+} is attached to the heteroatom at position 4 (O or Se), and the heteroatom at position 2 is O or Se, respectively. Similarly, [Ca2,O4] and [Ca2,Se4] correspond to complexes in which Ca^{2+} is attached to the heteroatom at position 2 (O or Se), and the heteroatom at position 4 is O or Se, respectively (see Scheme 2).

It is apparent that, in agreement with our previous discussion, Ca^{2+} binding energies are always greater when the metal is attached to position 4. It is also worth noting that the binding energy to the heteroatom in position 4 (or 2) is not affected much by the nature of the heteroatom at position 2 (or 4). It is also evident that the Ca^{2+} binding energy for urea (selenourea) is larger than that of acetamide (selonoacetamide), because the presence of two amino groups bonded to the carbonyl (selenocarbonyl) group enhances the resonance stabilization of the molecular dication. However, one of the most significant findings of the plot in Figure 4 is the preference of Ca^{2+} to attach to the oxygen atoms. In fact, the Ca^{2+} binding energy of selenourea is estimated to be −97.5 kcal mol^{-1} while that of urea is about −107.3 kcal mol^{-1} which shows the preference of Ca^{2+} for oxygen. The same conclusion is reached when acetamide is compared with selenoamide. In the case of 2-selenouracil and 4-selenouracil, the preference of calcium to interact with the carbonyl group is in competition with the enhanced basicity of the heteroatom at position 4. This is evident if we analyze the relative stability of complexes **1** and **4.** Indeed, for 2SeU both effects are in the same direction, and accordingly complex **4** (in which Ca^{2+} is attached to a C═O group in position 4) is about 16.6 kcal mol^{-1} more stable than complex **1** (where Ca^{2+} is attached to C═Se group in position 2). Conversely, for 4SeU both effects counterbalance each other. Although in general oxygen attachment should be preferred to selenium attachment, the selenocarbonyl group exhibits an enhanced basicity because it occupies position 4, and the result is that complexes **1** and **4** are nearly degenerate.

Bonding and Bonding Perturbation upon Ca^{2+} Association. The small value of the electron density at the BCPs of the Ca−Se, Ca−N, and Ca−O bonds (see Figure 3) as well as the fact that the energy density is positive[55] indicates that the bonding in [Ca−selenouracils]$^{2+}$ complexes is essentially ionic. However, a comparison between the values obtained for the isolated bases and the bases within the complexes clearly shows the existence of strong polariza-

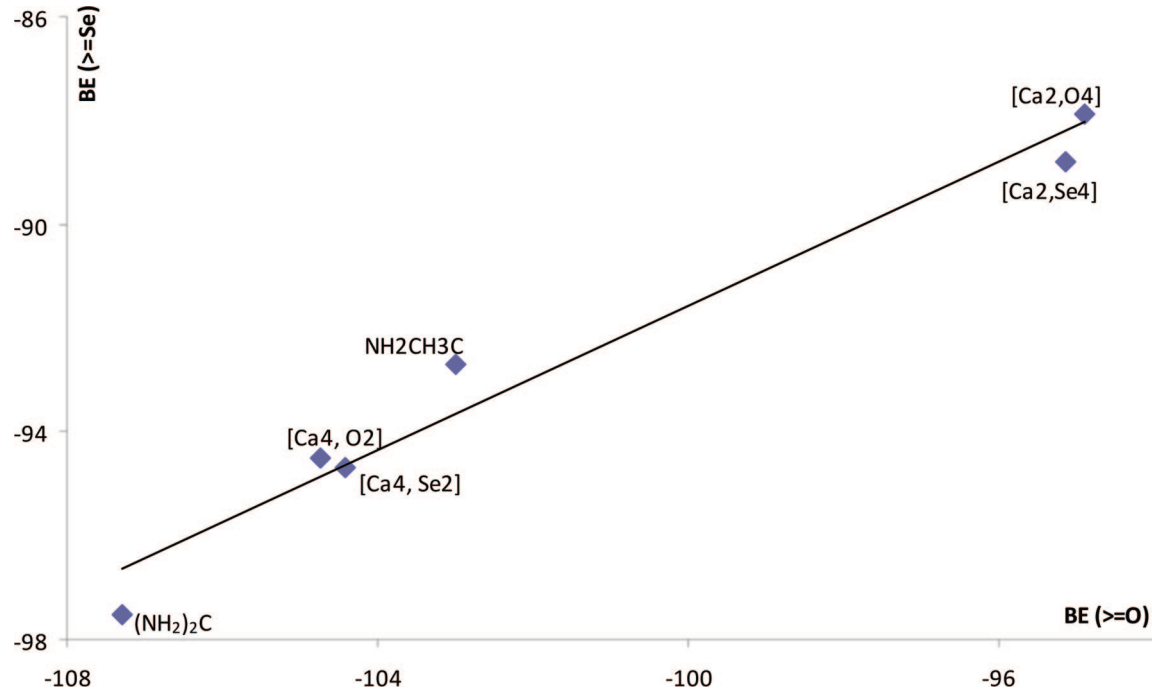
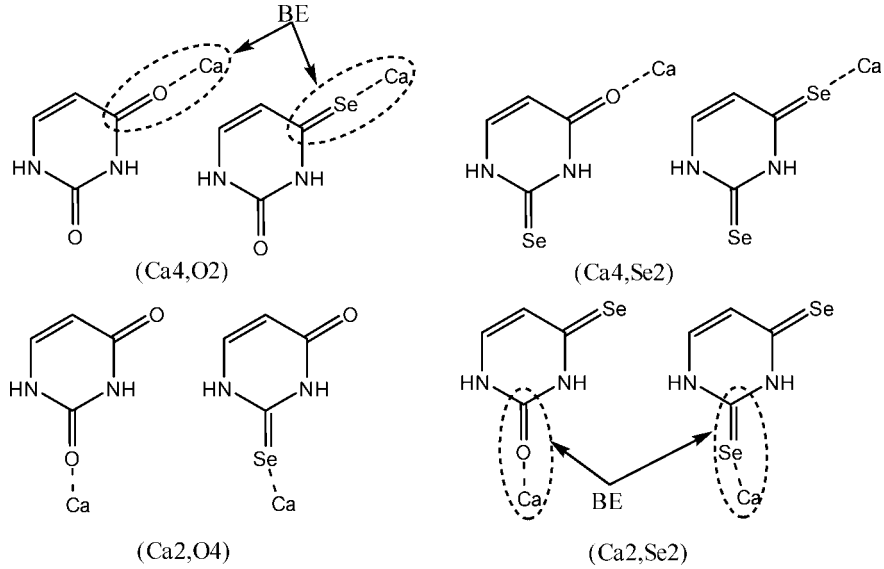


Figure 4. Carbonyl−Ca^{2+} binding energies versus selenocarbonyl−Ca^{2+} ones in the different selenouracils under consideration ($r^2 = 0.968$). $(NH_2)_2C$ and NH_2CH_3C denote urea and acetamide, respectively. [Ca4,O2], [Ca4,Se2], [Ca2,O4], and [Ca2,Se4] are defined in Scheme 2 (see also text).

Scheme 2



tion effects which cause a significant reorganization of the electron density of the base. For instance, when the alkaline-earth metal is associated directly with one of the heteroatoms at position 2 or 4, the electron density at the C═X (X = O, Se) BCP decreases significantly (by about 0.02 e.au^{-3} when X═Se and by about 0.07 e.au^{-3} when X = O). The much smaller effect observed for C═Se bonds clearly reflects the larger polarizability of Se with respect to oxygen. This different behavior of O-attached and Se-attached complexes is also reflected in the ELF of complexes **1** and **4** (see Figure 2), which shows that, upon Ca^{2+} association, an expansion of the basin associated with the basic site lone pairs takes place when this site is a Se atom, whereas a contraction of this basin is observed when the basic site is oxygen. Similar polarization effects are detected when complexes **7** are formed. In fact, the electron density at the bonds in which N3 participates also decreases significantly ($\Delta\rho_{C2-N3} = 0.020$ and $\Delta\rho_{N3-C4} = 0.022$) (see Figure 3).

The fact that electrostatic and polarization interactions are the dominant factors in these complexes explains the structural differences between urea−Ca^{2+} and selenouracil−Ca^{2+} adducts. While the angle C═O−Ca in urea−Ca^{2+} adducts[56] is 180°, the C═Se−Ca angle in selenouracil−Ca^{2+} adducts is about 110°. This behavior resembles that found for the attachment of Li^+ to formaldehyde and thioformaldehyde.[57] In the former case, a linear C−O−Li arrangement is predicted for the equilibrium conformation of the H_2CO-Li^+ complex, while a bent C−S−Li conformation

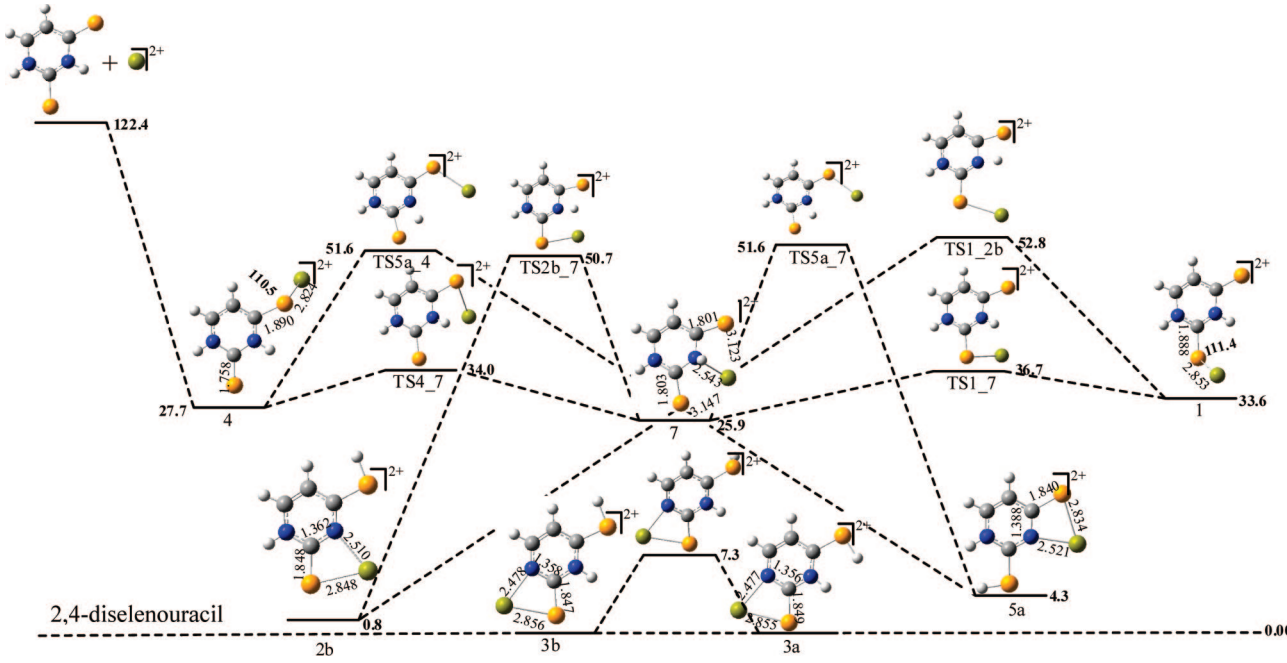


Figure 5. Energy profile for the isomerization process of dSeU−Ca^{2+} adducts. Relative energies are in kcal mol^{-1}.

is found for the equilibrium structure of the H_2CS-Li^+ system. A suitable explanation in terms of the differences in the molecular electrostatic potentials associated with the two neutral systems and polarization effects was offered by Alcamí et al.[58] The same arguments apply here. The primary interaction between the neutral molecule and the dication is electrostatic. Due to the size difference between oxygen and selenium atoms, the distance from the nuclei to the electrostatic potential minima is not only much shorter for oxygen than for selenium but also shorter than the typical dication-oxygen distance in the complex. As a consequence, the dication moves along isopotential lines connecting the two minima associated with the oxygen lone pairs. Since the potential is the same whatever the position of the dication along these lines, it nests between the two lone pairs to favor its simultaneous polarization. Conversely, when the cation approaches a selenium atom, the Ca−Se distances are similar to the distance between Se nuclei and the potential minima associated with its lone pairs, and the cation is trapped in either of these minima. This is nicely reflected in the ELF analysis of the valence basin associated with the reactive heteroatoms (see Figure 2). As far as the carbonyl group is concerned the electronic population in both monosynaptic basins associated with the oxygen lone pairs decreases. Conversely, for the selenocarbonyl only one monosynaptic basin of the selenium atom is actually affected (see Table S1).

Catalytic Effects on the Tautomerization Processes. In addition to the complexes resulting from the direct attack of Ca^{2+} on selenouracil (**1**, **4**, **7**), we have also considered the complexes that can be formed by Ca^{2+} attachment to the different tautomers formed by suitable hydrogen shifts (structures **2**, **3**, **5**, and **6** in Figure 1). The letters ***a*** and ***b*** have been added in order to distinguish between the various conformers with a different orientation of the hydrogen bonded to X or Y (see Figure 1). Their relative energies are summarized in Table 1. Total energies and ZPE corrections are given in Table S2 of the the Supporting Information. It should be mentioned that in some cases 4 conformers might be presented; in Figure 1 we have only included the two most stable ones. The optimized geometries of the global minima and transition states are given in the Supporting Information.

The data in Table 1 indicate that all complexes **2, 3,** and **5**, in which Ca^{2+} interacts simultaneously with the carbonyl (selenocarbonyl) group and the adjacent deprotonated ring nitrogen, are among the most stable selenouracils−Ca^{2+} complexes. As a matter of fact all of them are more stable than complexes **1** and **4**. The enhanced stability of complexes **2, 3**, and **5** arises from the fact that the alkaline-earth dication is able to polarize simultaneously both the heteroatom at positions 2 or 4 and the imino-type nitrogen which is very basic. This is nicely reflected in the topology of the corresponding electron densities. Although in complexes **2** the electron density at the X−Ca (X = O, Se) BCP is slightly smaller than in complexes **1** or **4**, this effect is clearly counterbalanced by the formation of a new N−Ca bond, which in complexes **1** and **4** is not possible (see Figure 3).

In view of the high stability of tautomers **2, 3,** and **5**, we investigated the possible tautomerization processes which connect them with the adducts **1**, **4**, and **7**. The corresponding energy profiles are plotted in Figures 5–7. The most significant finding is that all activation barriers lie below the entrance channel. This means that although the direct association of Ca^{2+} to selenouracils should lead exclusively to complexes **1**, **4**, and **7** since the neutrals only exist in the oxo-seleno and diseleno forms in the gas phase, the exothermicity of Ca^{2+} association is enough to trigger the tautomerization of the system, and, therefore, complexes **2b** should be the dominant ones for 2SeU−Ca^{2+} and 4SeU−Ca^{2+} species. For 2,4-SeU, a mixture of forms **3a** (34%), **3b** (41%), and **2b** (25%) should be found, because these three forms are very close in energy. The isomerization process connecting form **1** with forms **3a** and **3b** of 2,4-

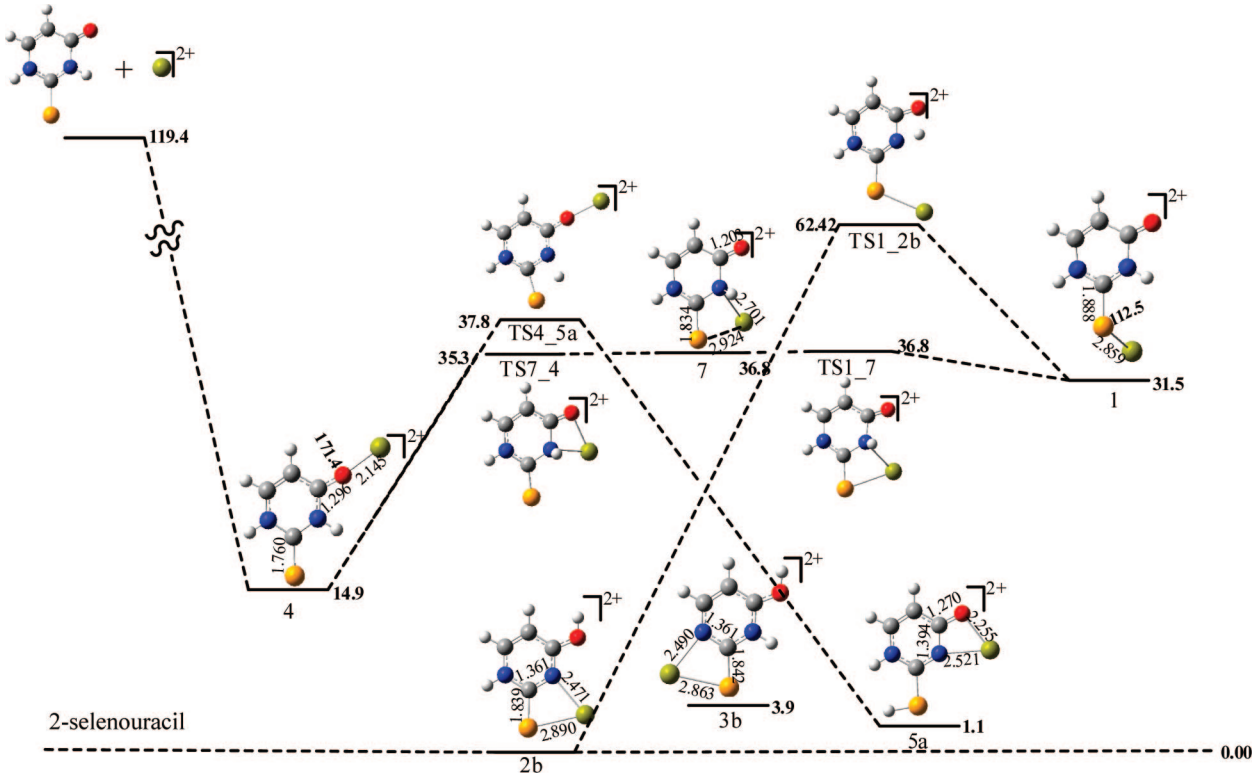


Figure 6. Energy profile for the isomerization process of 2SeU−Ca^{2+} adducts. Relative energies are in kcal mol^{-1}.

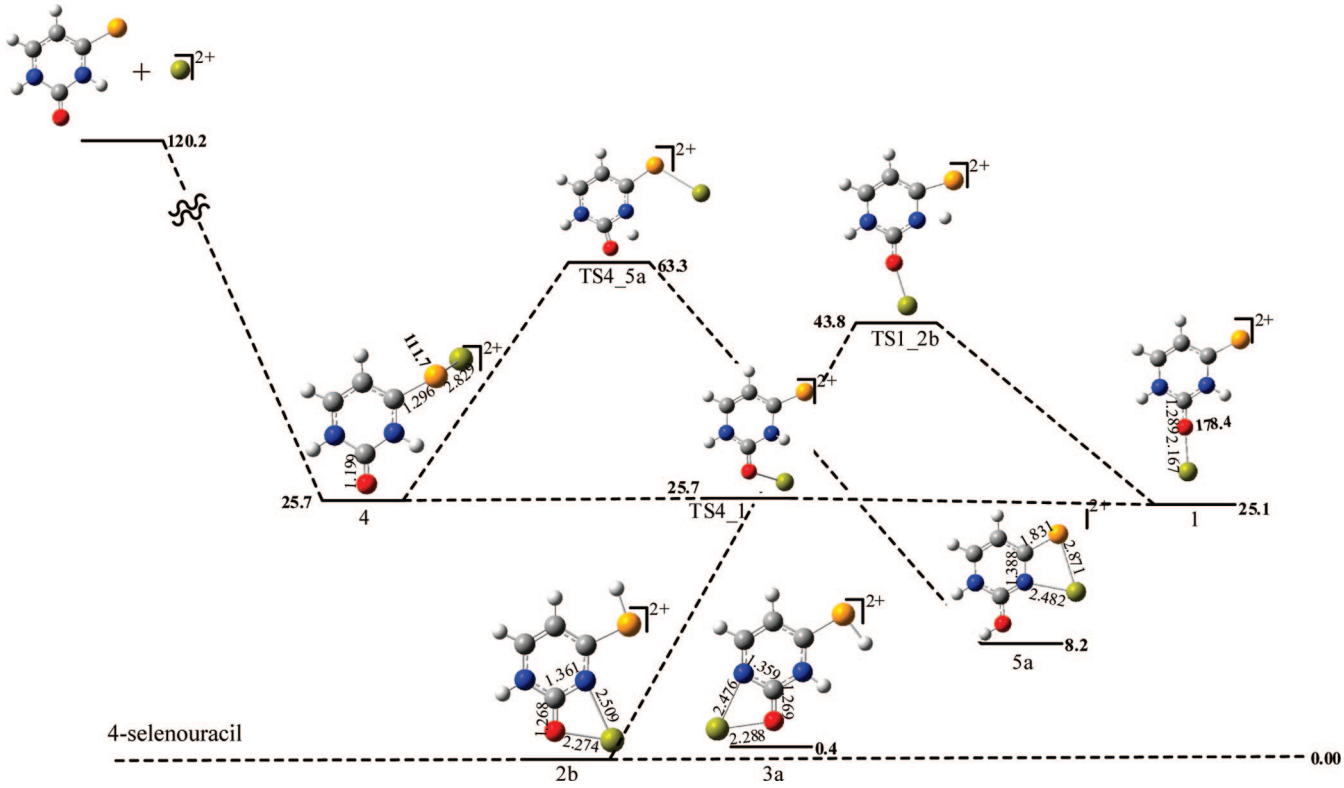


Figure 7. Energy profile for the isomerization process of 4SeU−Ca^{2+} adducts. Relative energies are in kcal mol^{-1}.

diselenouracil is presented in Figure S1 of the Supporting Information, to not overload Figure 5.

Figures 6 and 7 also show that the connections between forms **1** and **4** for 2SeU and 4SeU are essentially barrierless processes, either through form **7** as an intermediate (in the case of 2SeU) or directly (in the case of 4SeU, for which form **7** does not exist). The barriers for the same interconversion in the case of 2,4-diselenouracil are not negligible, but still rather small, so the **1**−**4** isomerization process must be also very facile.

To reach forms **2b** and **5a** a hydrogen shift is implicated, but these processes are clearly catalyzed by the presence of Ca^{2+}. If one refers, for instance, to 2-SeU, the activation barrier associated with the H-shift that connects forms **1** and **4** which for the isolated base is 40.9 kcal/mol,[45] for the Ca^{2+} complex reduces to 30.9 kcal mol^{-1}. The same applies to the barrier connecting forms **4** and **5**. For the Ca^{2+} complexes this barrier is 22.9, while for the isolated base it rises to 37.7 kcal mol^{-1}.[45] The same applies to 4-SeU and 2,4-SeU. For 4-SeU the **1**-**2b** and **4**-**5a** activation barriers change from 32.9 kcal mol^{-1} and 47.1 kcal mol^{-1} for the isolated base[45] to 18.7 and 37.6 kcal mol^{-1} in the corresponding Ca^{2+} complexes, respectively. The corresponding values for 2,4-SeU are 32.5 and 36.6 kcal mol^{-1} (isolated base) vs 19.2 and 23.9 kcal mol^{-1} (Ca^{2+} complexes).

Conclusions

The substitution of oxygen by selenium in uracil leads to significant changes in the reactivity of the system. The most important variation is the increased electron density of the nitrogen lone pair at position 3 and therefore its increased basicity. While in uracil and thiouracils the association of Cu(II)[49], Cu(I),[47] and proton[46] takes place exclusively at positions 2 and 4, in selenouracil−Ca^{2+} complexation at the nitrogen atom at position 3 leads to an alternative local minimum of the potential energy surface in the case of 2-SeU and 2,4-SeU. Furthermore, this adduct is predicted to be the most stable one in the case of 2,4-SeU, because in this position Ca^{2+} is able to interact with both Se atoms which are very polarizable.

Our results also show a clear preference of Ca^{2+} to be attached to oxygen, even though the enhanced basicity of the heteroatom at position 4 may alter this tendency, and in 4-SeU, the O-attached and the Se-attached complexes are nearly degenerate.

Perhaps the most important conclusion of this study is that although the enolic and selenol forms of selenouracils should not be observed in the gas phase,[44,45] the corresponding Ca^{2+} complexes are the most stable ones. More importantly, all the activation barriers associated with the corresponding tautomeric processes are lower than the entrance channel, and therefore not only these complexes should be observed but also they should be the dominant species in the gas phase. Also, Ca^{2+} association has a clear catalytic effect on these tautomerization processes, whose activation barriers decrease between 10 and 15 kcal mol^{-1}.

Acknowledgment. This work has been supported by the José Castillejo (ref:JC2007-00182) and Juan de la Cierva Programs from the Ministerio de Educación y Ciencia of Spain, by the DGI Project No. BQU2006-00894, and by the Project MADRISOLAR (ref.: S-0505/PPQ/0225) of the Comunidad Autónoma de Madrid. A generous allocation of computational time at the CCC of the Universidad Autónoma de Madrid is also acknowledged. The financial support of the Natural Sciences and Engineering Research Council of Canada, to R.J.B., is gratefully acknowledged.

Supporting Information Available: Tables of total energies, zero-point energies of the different selenouracil−Ca^{2+} complexes, binding energies plotted in Figure 4, and electronic populations in some relevant structures and the optimized geometries of the global minima and the transition states. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Blount, K. F.; Zhao, F.; Hermann, T.; Tor, Y. Conformational constraint as a means for understanding RNA-aminoglycoside specificity. *J. Am. Chem. Soc.* **2005**, *127*, 9818.

(2) Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2001**, *2*, 919.

(3) Latham, M. R.; Brown, D. J.; McCallum, S. A.; Pardi, A. NMR methods for studying the structure and dynamics of RNA. *ChemBioChem* **2005**, *6*, 1492.

(4) Storz, G. An expanding universe of noncoding RNAs. *Science* **2002**, *296*, 1260.

(5) Sprinzl, M.; Scheit, K. H.; Cramer, F. Preparation in-vitro of a 2-thiocytidine-containing yeast transfer-RNA PHE-A73-C74-S2C75-A76 and its interaction with para-hydroxymercuribenzoate Eur. *J. Biochem.* **1973**, *34*, 306.

(6) Lezius, A. G.; Scheit, K. H. enzymatic synthesis of DNA with 4-thio-thymidine triphosphate as substitute for dttp Eur. *J. Biochem.* **1967**, *3*, 85.

(7) Kutyavin, I. V.; Rhinehart, R. L.; Lukhtanov, E. A.; GornVv; Meyer, R. B.; Gamper, H. B. Oligonucleotides containing 2-aminoadenine and 2-thiothymine act as selectively binding complementary agents. *Biochemistry* **1996**, *35*, 11170.

(8) Coleman, R. S.; Kesicki, E. A. Synthesis and postsynthetic modification of oligodeoxynucleotides containing 4-Thio-2′-Deoxyuridine (D(S4)U). *J. Am. Chem. Soc.* **1994**, *116*, 11636.

(9) Sismour, A. M.; Benner, S. A. The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucl. Acids. Res.* **2005**, *33*, 5640.

(10) Sintim, H. O.; Kool, E. T. Remarkable sensitivity to DNA base shape in the DNA polymerase active site. *J. Am. Chem. Soc.* **2006**, *128*, 396.

(11) Salon, J.; Sheng, J.; Jiang, J. S.; Chen, G. X.; Caton-Williams, J.; Huang, Z. Oxygen replacement with selenium at the thymidine 4-position for the Se base pairing and crystal structure studies. *J. Am. Chem. Soc.* **2007**, *129*, 4862.

(12) Sheng, J.; Jiang, J. S.; Salon, J.; Huang, Z. Synthesis of a 2′-Se-thymidine phosphoramidite and its incorporation into oligonucleotides for crystal structure study. *Org. Lett.* **2007**, *9*, 749.

(13) Stadtman, T. C. Selenium biochemistry. *Science* **1974**, *183*, 915.

(14) Stadtman, T. C. Selenium-dependent enzymes. *Annu. Rev. Biochem.* **1980**, *49*, 93.

(15) Ching, W. M.; Stadtman, T. C. Selenium-containing transfer RNA-GLU from clostridium-sticklandii - correlation of aminoacylation with selenium content. *Proc. Natl. Acad. Sci. U.S. Am. Biol. Sci.* **1982**, *79*, 374.

(16) Trujillo, C.; Mó, O.; Yáñez, M.; Tortajada, J.; Salpin, J.-Y. Selenourea-Ca^{2+} Reactions in Gas Phase. Similarities and dissimilarities with urea and thiourea. J. Phys. Chem. B. **2008**, in press.

(17) Fraudo da Silva, J. J. R.; Williams, R. J. P. *The Biological Chemistry of Elements*; Oxford University Press: Oxford, 1991; p 180.

(18) Forsen, S.; Kordel, J. *Bioinorganic Chemistry*; University Science Books: Mill Valley, CA, 1994; p 107.

(19) Burda, J. V.; Sponer, J.; Leszczynski, J.; Hobza, P. Interaction of DNA base pairs with various metal cations (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Cu^{+}, Ag^{+}, Au^{+}, Zn^{2+}, Cd^{2+}, and Hg^{2+}): Nonempirical ab initio calculations on structures, energies, and nonadditivity of the interaction. *J. Phys. Chem. B* **1997**, *101*, 9670.

(20) Sponer, J.; Burda, J. V.; Sabat, M.; Leszczynski, J.; Hobza, P. Interaction between the guanine-cytosine Watson-Crick DNA base pair and hydrated group IIa (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}) and group IIb (Zn^{2+}, Cd^{2+}, Hg^{2+}) metal cations. *J. Phys. Chem. A* **1998**, *102*, 5951.

(21) Sponer, J.; Sabat, M.; Burda, J. V.; Leszczynski, J.; Hobza, P. Interaction of the adenine-thymine Watson-Crick and adenine-adenine reverse-Hoogsteen DNA base pairs with hydrated group IIa (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}) and IIb (Zn^{2+}, Cd^{2+}, Hg^{2+}) metal cations: Absence of the base pair stabilization by metal-induced polarization effects. *J. Phys. Chem. B* **1999**, *103*, 2528.

(22) Peschke, M.; Blades, A. T.; Kebarle, P. Binding energies for doubly-charged ions $M^{2+} = Mg^{2+}$, Ca^{2+} and Zn^{2+} with the ligands $L = H_2O$, acetone and N-methylacetamide in complexes ML_n^{2+} for n=1 to 7 from gas phase equilibrium determinations and theoretical calculations. *J. Am. Chem. Soc.* **2000**, *122*, 10440.

(23) Herron, S. R.; Scavetta, R. D.; Garrett, M.; Legner, M.; Jurnak, F. Characterization and implications of Ca^{2+} binding to pectate lyase C. *J. Biol. Chem.* **2003**, *278*, 12271.

(24) Russo, N.; Toscano, M.; Grand, A. Gas-phase absolute Ca^{2+} and Mg^{2+} affinity for nucleic acid bases. A theoretical determination. *J. Phys. Chem. A* **2003**, *107*, 11533.

(25) Poater, J.; Sodupe, M.; Bertran, J.; Sola, M. Hydrogen bonding and aromaticity in the guanine-cytosine base pair interacting with metal cations ($M = Cu^{+}$, Ca^{2+} and Cu^{2+}). *Mol. Phys.* **2005**, *103*, 163.

(26) Reddy, A. S.; Sastry, G. N. Cation [$M = H^{+}$, Li^{+}, Na^{+}, K^{+}, Ca^{2+}, Mg^{2+}, NH_4^{+}, and NMe_4^{+}] interactions with the aromatic motifs of naturally occurring amino acids: A theoretical study. *J. Phys. Chem. A* **2005**, *109*, 8893.

(27) Allen, R. N.; Shukla, M. K.; Burda, J. V.; Leszczynski, J. Theoretical study of interaction of urate with Li^{+}, Na^{+}, K^{+}, Be^{2+}, Mg^{2+}, and Ca^{2+} metal cations. *J. Phys. Chem. A* **2006**, *110*, 6139.

(28) Nicolas, I.; Castro, M. Theoretical study of the complexes of horminone with Mg^{2+} and Ca^{2+} ions and their relation with the bacteriostatic activity. *J. Phys. Chem. A* **2006**, *110*, 4564.

(29) Remko, M.; Rode, B. M. Effect of metal ions (Li^{+}, Na^{+}, K^{+}, Mg^{2+}, Ca^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+}) and water coordination on the structure of glycine and zwitterionic glycine. *J. Phys. Chem. A* **2006**, *110*, 1960.

(30) Liu, H. C.; Zhang, L.; Li, P.; Cukier, R. I.; Bu, Y. X. Exploration of the Ca^{2+} interaction modes of the nifedipine calcium channel antagonist. *ChemPhysChem* **2007**, *8*, 304.

(31) Andersson, J.; Hauser, K.; Karjalainen, E. L.; Barth, A. Protonation and hydrogen bonding of Ca^{2+} site residues in the E2P phosphoenzyme intermediate of sarcoplasmic reticulum Ca^{2+}-ATPase studied by a combination of infrared Spectroscopy and electrostatic calculations. *Biophys. J.* **2008**, *94*, 600.

(32) Becke, A. D. Density-Functional Thermochemistry 0.3. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648.

(33) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle and Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* **1988**, *37*, 785.

(34) 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.; 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.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, J.; Cammi, R.; Pomelli, C.; Ochterski, J.; 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. *Gaussian03; Gaussian03, Revision C.02 ed.*; Gaussian, Inc.: Wallingford, CT, 2003.

(35) Scott, A. P.; Radom, L. Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.* **1996**, *100*, 16502.

(36) Hertwig, R. H.; Koch, W.; Schroder, D.; Schwarz, H.; Hrusak, J.; Schwerdtfeger, P. A comparative computational study of cationic coinage metal-ethylene complexes $(C_2H_4)M(+)$ (M=Cu, Ag, and Au). *J. Phys. Chem.* **1996**, *100*, 12253.

(37) Bader, R. F. W. *Atoms In Molecules: A Quantum Theory*; Clarendon Press: Oxford University, Oxford, 1990; p 1.

(38) Matta, C. F.; Boyd, R. J. *The Quantum Theory of Atoms in Molecules*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2007; p 1.

(39) Biegler-König, F.; Schönbohm, J. AIM2000, 2nd ed.; 2002.

(40) Alikhani, M. E.; Fuster, F.; Silvi, B.; Silvi, B. What can tell the topological analysis of ELF on hydrogen bonding? *Struct. Chem.* **2005**, *16*, 203.

(41) Savin, A.; Nesper, R.; Wengert, S.; Fassler, T. F. ELF: The electron localization function. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1809.

(42) Silvi, B.; Savin, A. Classification of chemical-bonds based on topological analysis of electron localization functions. *Nature (London)* **1994**, *371*, 683.

(43) Noury, S.; Krokidis, X.; Fuster, F.; Silvi, B. *TopMod Package*; Université Pierre et Marie Curie: 1997.

(44) Leszczynski, J.; Sponer, J. 2,4-Diselenouracil tautomers: Structures, energies, and a comparison with uracil and 2,4-dithiouracil. *J. Mol. Struct. (Theochem)* **1996**, *388*, 237.

(45) Trujillo, C.; Mó, O.; Yáñez, M. A theoretical study of hydration effects on the prototropic tautomerism of selenouracils. *Org. Biomol. Chem.* **2007**, *5*, 3092.

(46) Lamsabhi, A. M.; Alcamí, M.; Mó, O.; Bouab, W.; Esseffar, M.; Abboud, J. L.-M.; Yáñez, M. Are the thiouracils sulfur bases in the gas-phase. *J. Phys. Chem. A* **2000**, *104*, 5122.

(47) Lamsabhi, A. M.; Alcamí, M.; Mó, O.; Yáñez, M. Gas-phase reactivity of uracil, 2-thiouracil, 4-thiouracil, and 2,4-dithiouracil towards the Cu^+ cation: A DFT study. *ChemPhysChem* **2003**, *4*, 1011.

(48) Lamsabhi, A. M.; Alcamí, M.; Mó, O.; Yáñez, M.; Tortajada, J.; Salpin, J. Y. Unimolecular reactivity of uracil-Cu^{2+} complexes in the gas phase. *ChemPhysChem* **2007**, *8*, 181.

(49) Lamsabhi, A. M.; Mó, O.; Yáñez, M.; Alcamí, M.; Tortajada, J. Association of Cu^{2+} with uracil and its thio derivatives: A theoretical study. *ChemPhysChem* **2004**, *5*, 1871.

(50) Guillaumont, S.; Tortajada, J.; Salpin, J.-Y.; Lamsabhi, A. M. Experimental and computational study of the gas-phase interactions between lead(II) ions and two pyrimidic nucleobases: Uracil and thymine. *Int. J. Mass Spectrom.* **2005**, *243*, 279.

(51) Safi, Z.; Lamsabhi, A. M. Gas-phase reactivity of 2,7-dimethyl-[1,2,4]-triazepine thio derivatives toward Cu^+ cation: A DFT study. *J. Phys. Chem A* **2007**, *111*, 2213.

(52) Lamsabhi, A. M.; Alcamí, M.; Mó, O.; Yáñez, M.; Tortajada, J. Gas-phase deprotonation of uracil-Cu^{2+} and thiouracil-Cu^{2+} complexes. *J. Phys. Chem. A* **2006**, *110*, 1943.

(53) Zhu, W. L.; Luo, X. M.; Puah, C. M.; Tan, X. J.; Shen, J. H.; Gu, J. D.; Chen, K. X.; Jiang, H. L. The multiplicity, strength, and nature of the interaction of nucleobases with alkaline and alkaline earth metal cations: A density functional theory investigation. *J. Phys. Chem. A* **2004**, *108*, 4008.

(54) Trujillo, C.; Lamsabhi, A. M.; Mó, O.; Salpin, J.-Y.; Tortajada, J.; Yáñez, M. Manuscript in preparation, 2008.

(55) Cremer, D.; Kraka, E. Chemical-bonds without bonding electron-density - does the difference electron-density analysis suffice for a description of the chemical-bond. *Angew. Chem.* **1984**, *96*, 612.

(56) Corral, I.; Mó, O.; Yáñez, M.; Salpin, J.-Y.; Tortajada, J.; Radom, L. Gas-phase reactions between urea and Ca^{2+}: The importance of Coulomb explosions. *J. Phys. Chem. A* **2004**, *108*, 10080.

(57) Alcamí, M.; Mó, O.; Yáñez, M. *In Molecular Electrostatic Potentials: Concepts and Applications*; Elsevier: Amsterdam, 1996; Vol. 3, p 407.

(58) Alcamí, M.; Mó, O.; Yáñez, M.; Abboud, J. L. M.; Elguero, J. Bond activation by protonation in the gas-phase. *Chem. Phys. Lett.* **1990**, *172*, 471.

CT800017J

JCTC Journal of Chemical Theory and Computation

How Is cis−trans Isomerization Controlled in Dronpa Mutants? A Replica Exchange Molecular Dynamics Study

Samuel L. C. Moors, Servaas Michielssens, Cristina Flors,† Peter Dedecker, Johan Hofkens, and Arnout Ceulemans*

Department of Chemistry and INPAC Institute for Nanoscale Physics and Chemistry, K. U. Leuven, Leuven, Belgium

Received February 1, 2008

Abstract: The reversibly photoactivatable green fluorescent protein analog Dronpa holds great promise as a marker for various new cellular imaging applications. Using a replica exchange method which combines both Hamiltonian and temperature exchanges, the ground-state dynamics of Dronpa and two mutants with increased switching kinetics, Val157Gly and Met159Thr, were compared. The dominant chromophore state was found to be the cis isomer in all three proteins. The simulation data suggest that both mutations strongly increase the chromophore flexibility and cis−trans isomerization rate. We identify three key amino acids, Val157, Met159, and Phe173, which are able to impede the bottom hula-twist transition path, depending on their position and rotameric state. We believe our insights will help to understand the switching process and provide useful information for the design of new variants with improved fluorescence properties.

Introduction

Dronpa, a monomeric mutant of a fluorescent protein from the coral *Pectiniidae Echinophyllia* sp. SC22, sharing 81% sequence identity with the fluorescent protein KikG from the coral *Favia favus*, emits 518 nm light with a high fluorescence quantum yield ($\Phi_{FL} = 0.85$).[1,2] At neutral or basic pH, Dronpa can be photoswitched from a bright (fluorescent) state to a dark (nonfluorescent) state by intense 488 nm light. The photoswitched dark state has a half-lifetime of 14 h.[3] Illumination with 405 nm light efficiently switches the dark state back to the bright state. This on−off switching process is highly reversible (>100 cycles),[4] opening up a host of new applications including dynamic optical labeling and tracking of proteins, organelles, or cells and detection of protein interactions using high-resolution imaging.[5,6]

Recently, through semirandom mutagenesis, Ando et al. discovered two Dronpa mutants, Dronpa Met159Thr (Dronpa-2) and Dronpa Val157Ile/Met159Ala (Dronpa-3), which display increased photoinduced and spontaneous switching kinetics.[7] Compared to Dronpa, both mutants can be turned off more efficiently with 488 nm light and returned more quickly to their emissive states in the dark but have much lower fluorescence quantum yields ($\Phi_{FL} = 0.33$ and 0.28 respectively). Ensemble and single-molecule fluorescence experiments suggested increased conformational freedom of the chromophore in Dronpa-2 and Dronpa-3 and formation of a dark intermediate state arising from cis−trans isomerization.[8] Independently, Stiel et al.[3] also discovered Dronpa-2, together with another fast-switching mutant, Dronpa Val157Gly (rsFastLime), with a slightly decreased fluorescence quantum yield ($\Phi_{FL} = 0.77$). Both variants were designed by structural comparison of Dronpa with asFP595 from the sea anemone *Anemonia sulcata*. Like Dronpa, asFP595 is a reversibly photoactivatable protein. With only four differing residues, the immediate chromophore environments of asFP595 and Dronpa are very similar. Unlike the Dronpa chromophore, which adopts a fluorescent cis conformation, the equilibrated asFP595 chromophore is in a nonfluorescent trans conformation. The residues Val157 and Ser142 in Dronpa are replaced in asFP595 at the equivalent positions by serine and alanine, respectively. As a result, the favorable H-bond between the cis chromophore and

* Corresponding author tel.: (32)16/32.73.63; fax: (32)16/32.79.92; e-mail: Arnout.Ceulemans@chem.kuleuven.be.

† Present address: School of Chemistry, University of Edinburgh, UK.

10.1021/ct8000359 CCC: $40.75
Published on Web 05/20/2008

Ser142 in Dronpa is replaced by an H-bond between the trans chromophore and Ser157 in asFP595. The spectroscopic properties of asFP595 are also different from Dronpa. Upon photoactivation with 540−560 nm light, the asFP595 chromophore isomerizes to a cis coplanar state, which is converted back to the nonfluorescent trans state, either thermally or by illumination with 450 nm light.[9,10] By modeling a trans chromophore into the cis X-ray structure of Dronpa, Stiel et al. suggested that mutations at Val157 and Met159 reduce the steric hindrance for a similar cis−trans isomerization in Dronpa.[3] Remarkably, all three mutants, Dronpa-2, Dronpa-3, and rsFastLime, arise from the replacement of only Val157 or Met159, or both.

X-ray diffraction of bright-state Dronpa crystals, which were photoswitched to the dark state and subsequently flash-frozen in liquid nitrogen to trap the chromophore in its dark state conformation, showed that cis-to-trans isomerization did indeed occur, accompanied by structural rearrangements of several nearby amino acid residues.[11] The transition mechanism however remains largely speculative. Upon photobleaching, the anionic chromophore of Dronpa is protonated at the phenolate oxygen to the neutral state.[1] The switching process is further complicated by the existence of several dark intermediate states and an excited-state proton transfer reaction.[4,12]

To gain deeper insight in the dynamics of the electronic ground state at the molecular level, we investigated the cis−trans equilibrium distributions and transition paths in solution of Dronpa, rsFastLime, and Dronpa-2 using a replica exchange method which combines exchanges between replicas with different Hamiltonians and temperatures (HT-REM). Both Dronpa mutants display increased chromophore flexibility and lower cis−trans isomerization barriers. Our simulations point out three key residues, Val157, Met159, and Phe173, which control the isomerization process. Substitution of Val157 to Gly greatly increases the conformational freedom of the chromophore by allowing a Phe173 shift away from the chromophore and a consequent reshuffling of the Met159 rotamer population. Replacement of the bulky Met159 with the smaller threonine elevates the steric hindrance with the chromophore in the cis state and along the hula-twist isomerization pathway and allows the chromophore to shift toward the protein surface, which decreases the stability of the trans state by steric hindrance with Val157.

Methods

Force Field Parameters. The chromophore force field parameters for both the anionic and neutral forms were modeled from quantum chemical calculations using Gaussian 03.[13] The atomic charges were determined with the RED program,[14] which automates the optimization of the molecular electrostatic potential and the restricted electrostatic potential charges from HF/6-31G* calculations on 4′-hydroxybenzylidene-2,3-dimethyl-imidazolinone (HBDI; Figure 1), compatible with the Amber force field. The dihedral angle parameters for φ and τ were fitted from B3LYP/6-31G* calculations on HBDI at various φ and τ angles. Using the dihedral energy term $E_{\text{dihedral}} = k_\varphi[1 + \cos(n\varphi - \varphi_0)]$,

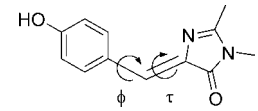


Figure 1. Structure of HBDI, used as a model for the calculation of the neutral chromophore force field parameters. For the anionic chromophore, the phenolate form of HBDI was used.

we obtained force constants $k_\varphi = 3.015$ kcal mol^{-1} and $k_\tau = 3.900$ kcal mol^{-1} and $k_\phi = 1.185$ kcal mol^{-1} and $k_\tau = 5.900$ kcal mol^{-1} for the anionic and neutral chromophores, respectively, which leads to total energy barriers $\Delta U_\varphi = 24.120$ kcal mol^{-1} and $\Delta U_\tau = 31.200$ kcal mol^{-1} and $\Delta U_\varphi = 9.480$ kcal mol^{-1} and $\Delta U_\tau = 47.200$ kcal mol^{-1} for the anionic and neutral chromophores, respectively. Similar values were obtained by Reuter et al.: $k_\varphi = 2.700$ kcal mol^{-1} and $k_\tau = 3.900$ kcal mol^{-1} and $k_\varphi = 1.400$ kcal mol^{-1} and $k_\tau = 6.840$ kcal mol^{-1} for the anionic and neutral chromophores, respectively.[15] The remaining parameters were obtained from analogues' functional groups already present in the Amber force field. A complete list of the chromophore parameters and charges can be found in the Supporting Information.

Molecular Dynamics Simulations. All molecular dynamics (MD) simulations were started from the bright-state X-ray structure of Dronpa (Protein Databank (PDB) code 2IE2), which was mutated at the relevant positions with the Swiss-PdbViewer.[16] The Amber 2003 force field[17] was used with a modified version of the Amber 8 software.[18] The smooth particle mesh Ewald method[19] was employed to accommodate long-range electrostatic forces. The nonbonded cutoff for van der Waals interactions was set to 8 Å. Covalent bonds involving hydrogen atoms were constrained using SHAKE.[20] A time step of 2 fs was used. Samples were collected every 0.5 ps. The temperature was controlled using the Berendsen weak-coupling algorithm.[21] The His193 pyrrole was protonated. The charged protein, with crystallographic waters included, was neutralized with sodium ions and solvated in a truncated octahedral TIP3P[22] water box, which extended the system with at least 8 Å at each side of the box. After rotating the chromophore to the trans isomer and mutation of the relevant residues, the system was minimized and equilibrated for 2.5 ns at 300 K and 1 bar.

Hamiltonian and Temperature Replica Exchange. To overcome the cis−trans energy barrier, normally not accessible using conventional MD, a combined Hamiltonian[23] and temperature[24,25] REM (HT-REM) simulation was performed. The exchange probability P_{acc} between two replicas i and j with reduced coordinates X_i and X_j, reciprocal temperatures β_i and β_j, and differing potential energy functions $U_i(X)$ and $U_j(X)$ is given as

$$P_{\text{acc}}(i,j) = \min\{1, \exp[\beta_i(U_i(X_i) - U_i(X_j)) - \beta_j(U_j(X_i) - U_j(X_j))]\} \quad (1)$$

Exchanges were attempted between replicas with differing temperature and force field parameters. A total of 20 replicas were distributed evenly over the 300−350.5 K temperature range, while gradually lowering the φ and τ force constants

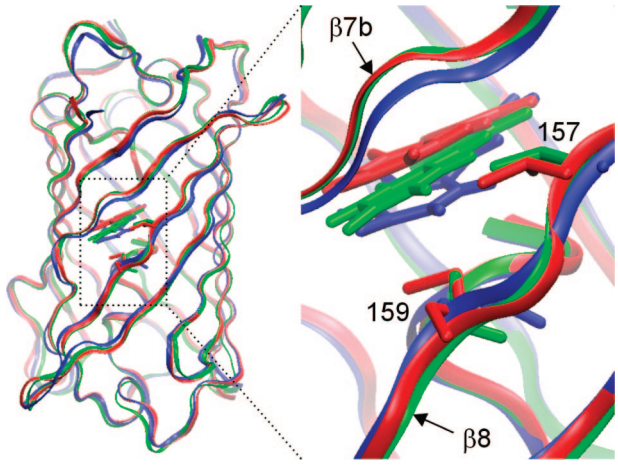


Figure 2. Ribbon diagrams of neutral Dronpa (red), rsFastLime (blue), and Dronpa-2 (green), with the chromophore and respective residues in positions 157 and 159. Each structure was generated by taking the average over 20 random snapshots. Compared to Dronpa, strand β7b is located closer to strand β8, while the Dronpa-2 chromophore is shifted towards the solvent.

to zero at 350.5 K. In the highest temperature replica, where the φ and τ dihedral terms are removed, the intrinsic potential surface of the chromophore displays minima at φ and τ angles of approximately +90°/−90°, about 2.5 kcal mol^{-1} lower than the coplanar conformation. This inverted surface enhances the sampling of the cis−trans transition state. Note that the inverted rotational barriers are much lower than the calculated barriers for HBDI. The time constant for temperature coupling was 0.5 ps. Exchanges were attempted every 0.5 ps. Acceptance probabilities between neighboring temperatures varied between 0.19 and 0.27. After each successful exchange, all velocities were randomly reassigned from a Maxwell−Boltzmann distribution. The HT-REM simulation was carried out for 37.5 ns per replica, the last 12.5 ns of which were used to calculate the reported thermodynamic properties. For Dronpa-2, 25 ns per replica were simulated, the last 12.5 ns of which were used. The following criteria for H-bond interaction were taken: the distance between the donor (D) and acceptor (A) must be smaller than 3.3 Å, and the D−H−A angle must be larger than 120°. Standard errors of the thermodynamic averages were calculated using the statistical inefficiency method.[26] Free energy plots were calculated using a multistate Bennett acceptance ratio implementation, which is suitable for the analysis of multiple simulations conducted under arbitrary conditions.[27,28]

Results and Discussion

Secondary Structure. HT-REM simulations were performed on Dronpa, rsFastLime, and Dronpa-2 with the chromophore in its neutral form as well as on the anionic form of rsFastLime. A total of 20 replicas were distributed evenly over the 300−350.5 K temperature range, with concomitant gradual lowering of the φ and τ force constants to zero at 350.5 K. As shown in Figure 2, in all simulations, the β-barrel secondary structure is very similar. The average backbone root-mean-square deviations from the Dronpa X-ray structue are around 1.1 Å. The cleft between strands β7 and β10, which was highlighted by Stiel et al.,[3] is preserved in all simulations and provides a dynamic water bridge between the chromophore and the bulk solvent. This cleft originates from a kink in strand β7 at Pro141, which divides the strand into two parts, β7a and β7b. Interestingly, the β7 kink also causes a disruption of the H-bonding pattern between strands β7b and β8, from Pro141 to Thr143, close to the chromophore phenol moiety (Figure 2). In rsFastLime, the backbone distance between strands β7b and β8 is significantly reduced; the average distance between the Thr143 O and the Asn158 N reduces from 5.3 Å in Dronpa to 4.5 Å, effectively filling up the space that arises from Val157Gly mutation. In Dronpa-2, no such adjustment of secondary structure was observed. Instead, the freed space from replacing the bulky methionine by threonine is partly filled by the chromophore, which is shifted toward the protein surface; the distance between the chromophore phenol oxygen and Glu140 O reduces by 0.3 Å. As a result, the phenol ring is more exposed to the solvent (via the cleft) and the degree of H-bonding with a nearby water molecule is increased from 4% in Dronpa to 33% in Dronpa-2.

cis−trans Equilibria. The equilibrium distributions of the chromophore and several rotational isomeric states of surrounding residues are listed in Table 1. All HT-REM simulations started from a sample in which the chromophore had been equilibrated for 2.5 ns in the trans state. In Figure 3, the equilibration of the trans chromophore population over time is shown. In neutral Dronpa-2, the trans population rapidly decreases to zero after 6 ns. Also in anionic rsFastLime, the trans population gradually fades away. In contrast, a small fraction of trans remains for the neutral forms of rsFastLime (5%) and Dronpa (2%). Irradiation with 405 nm light of thermally equilibrated rsFastLime resulted in a 2-fold increase of fluorescence.[3] Thus, there must be a nonfluorescent state with 50% population or a less fluorescent state with even higher population. Since a trans population of at most 5% is found in rsFastLime, our results indicate that trans-to-cis isomerization alone cannot explain this 2-fold fluorescence increase. As an alternative, a neutral cis state could be responsible for the residual dark state population in thermal equilibrium.

The increased stability of the trans isomer in the neutral chromophore compared to the anionic chromophore can be explained by a favorable H-bond interaction of the phenolate oxygen with the Ser142 hydroxy group (Figure 4). In rsFastLime, the anionic cis chromophore is indeed more tightly H-bonded with Ser142 than the neutral chromophore (75% H-bond population versus 20%, respectively). In Dronpa-2, the stability of the trans state is very low compared to the cis state (*vide infra*). The chromophore shift toward Glu140 O (see above) causes steric hindrance with Val157 in the trans conformation. Additionally, the cis state is stabilized by reduced steric interactions of the cis chromophore with Thr159 in Dronpa-2 in comparison with Met159 in Dronpa.

Along with cis−trans isomerization, several surrounding residues undergo significant displacement. The Phe173 side chain appears to play a central role. Contrary to all available

Table 1. Populations of Chromophore Isomer States and Several Surrounding Residue Rotamer States

		cis				
rotational isomeric state		Dronpa (X-ray)[a]	Dronpa (neutral)	Dronpa-2 (neutral)	rsFastLime (neutral)	rsFastLime (anionic)
chromophore τ	0°	1	0.98 ± 0.01	1	0.95 ± 0.02	1
Phe173 χ1	180°	0	0.007 ± 0.004	0.10 ± 0.02	0.02 ± 0.01	0
	+60°	0	0.72 ± 0.03	0.84 ± 0.03	0.94 ± 0.02	0.83 ± 0.03
	−60°	1	0.28 ± 0.03	0.05 ± 0.03	0.04 ± 0.01	0.17 ± 0.03
Met159/Thr159 χ1	180°	0.0	0	0.45 ± 0.04	0	0.01 ± 0.01
	+60°	0	0.55 ± 0.04	0.51 ± 0.04	0.86 ± 0.04	0.53 ± 0.05
	−60°	1	0.45 ± 0.04	0.03 ± 0.01	0.14 ± 0.04	0.47 ± 0.04
Val157 χ1	180°	0.2	0.28 ± 0.05	0.37 ± 0.03		
	+60°	0.8	0.72 ± 0.05	0.62 ± 0.03		
	−60°	0	0	0		
Ser142 χ1	180°	0	0.64 ± 0.03	0.42 ± 0.04	0.22 ± 0.06	0.16 ± 0.05
	+60°	1	0.02 ± 0.01	0.12 ± 0.02	0.36 ± 0.10	0.69 ± 0.07
	−60°	0	0.34 ± 0.03	0.46 ± 0.04	0.41 ± 0.07	0.16 ± 0.05

		trans				
rotational isomeric state		Dronpa (X-ray)[b]	Dronpa (neutral)	Dronpa-2 (neutral)	rsFastLime (neutral)	rsFastLime (anionic)
chromophore τ	180°	1	0.02 ± 0.01	0	0.05 ± 0.02	0
Phe173 χ1	180°	0	0		0	
	+60°	0	0		0	
	−60°	1	1		1	
Met159/Thr159 χ1	180°	0	0		0.06 ± 0.03	
	+60°	0	0.69 ± 0.17		0.53 ± 0.09	
	−60°	1	0.31 ± 0.17		0.42 ± 0.09	
Val157 χ1	180°	0	0			
	+60°	1	1			
	−60°	0	0			
Ser142 χ1	180°	0.5	0.41 ± 0.12		0.4 ± 0.2	
	+60°	0.5	0.59 ± 0.12		0.5 ± 0.2	
	−60°	0	0		0.06 ± 0.03	

[a] Average values taken from all chains of PDB X-ray structures 2IE2, 2IOV, and 2GX0. [b] Average values taken from all chains of PDB X-ray structure 2POX.

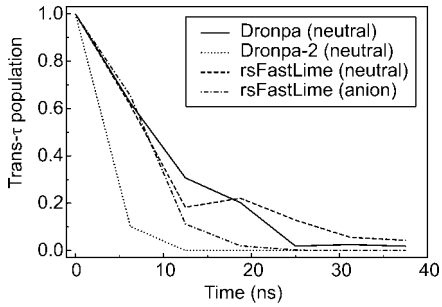


Figure 3. Evolution of the trans chromophore population at 300 K as a function of simulation time.

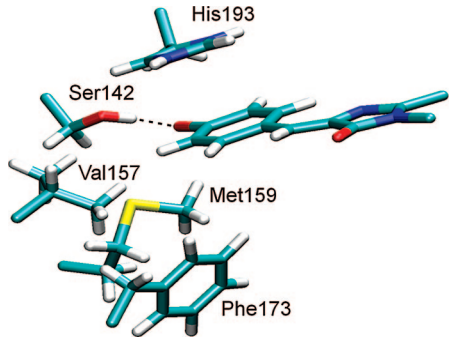


Figure 4. X-ray structure of the anionic Dronpa chromophore, H-bonded with Ser142, and surrounding residues.

Dronpa X-ray structures, in all simulated proteins, the Phe173 χ1 dihedral is mostly in its +60° rotameric state (Table 1), the phenyl ring leaning toward residue 157. This difference between the MD predicted rotamers in solution and the X-ray data is possibly due to packing effects in the crystal structure. Note that, in the Dronpa crystal structures, the Phe173 χ2 dihedral considerably deviates from its ideal 90° angle (χ2 = 19−44°). In both neutral Dronpa and rsFastLime simulations, Phe173 χ1 undergoes a rotamer shift from +60° in cis to −60° in trans, due to a steric clash between the trans chromophore and the Phe173 χ1 (=+60°) rotamer (Figure 5a). In rsFastLime, Met159 χ1 is correlated with Phe173 χ1. Steric hindrance with the Phe173 χ1 (=+60°) state restricts Met159 χ1 in the +60° rotamer, while in the Phe173 χ1 (= −60°) state, Met159 χ1 is about equally distributed over +60° and −60° (Figure 5b). In Dronpa, this correlation is not apparent. Steric interaction with the Val157 isopropyl group in *cis*-Dronpa moves the Phe173 χ1 (= +60°) phenyl ring away from Met159 CG, allowing Met159 χ1 to occupy both +60° and −60°.

In anionic rsFastLime, Ser142 is mainly positioned toward the chromophore with χ1 = +60° (Table 1), in agreement with the X-ray structure of bright-state Dronpa. In the protein with the neutral chromophore, however, Ser142 χ1 is distributed over the three rotamers, in accordance with the dark-state X-ray structure of Dronpa,[11] which exhibits a disordered Ser142 residue (χ1 = −164°/69°). In the χ1 = −60° rotamer, the Ser142 hydroxy group is H-bonded with the Glu140 carbonyl oxygen, whereas in the χ1 = 180° rotamer, a H-bond is formed with the Asn155 carbonyl

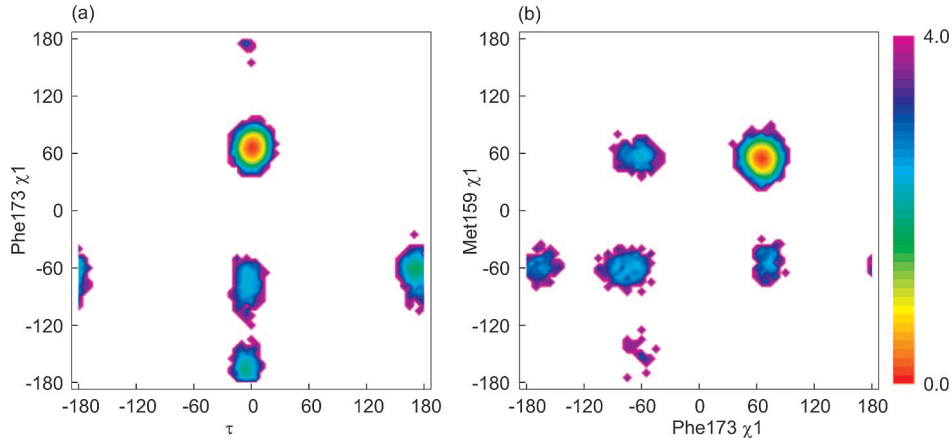


Figure 5. Free energy as a function of (a) τ and Phe173 χ1 and (b) Phe173 χ1 and Met159 χ1, from rsFastLime (neutral chromophore) samples at 300 K in kcal mol^{-1}. Values of τ equal to 0° and 180° correspond to the cis and trans isomers respectively.

Table 2. Chromophore Deviation From Planarity[a]

	cis			trans		
	τ (deg)	φ (deg)	θ (deg)	τ (deg)	φ (deg)	θ (deg)
Dronpa (X-ray)	1.5[b]/12.5[c]/1.1[d]	−6.9[b]/−17.4[c]/−0.3[d]	7.6[b]/13.4[c]/2.1[d]	160.5[e]	36.3[e]	26.7[e]
Dronpa (neutral)	0.3(6.1)	−7.1(5.9)	10.0(5.5)	181.1(6.8)	−10.0(7.4)	17.9(8.7)
Dronpa-2 (neutral)	2.1(6.0)	−3.3(10.5)	10.4(5.9)			
rsFastLime (neutral)	0.4(6.6)	−16.2(13.0)	19.8(10.5)	171.7(6.8)	−25.8(11.1)	35.8(8.5)
rsFastLime (anion)	3.0(7.0)	−4.6(9.2)	11.6(6.2)			
Chromo (neutral)[f]	−0.07(7.0)	−0.5(17.0)	16.2(9.3)			

[a] Average values of θ were calculated as the angle (in degrees) between the planes formed by the atoms C_9, C_{10}, and N_1 and C_{11}, C_{14}, and C_{15}. Standard deviations are given between parentheses. [b] PDB X-ray data are average angles over all chains taken from 2IE2. [c] PDB X-ray data are average angles over all chains taken from 2IOV. [d] PDB X-ray data are average angles over all chains taken from 2GX0. [e] PDB X-ray data are average angles over all chains taken from 2POX. [f] Data extracted from a 6 ns HT-REM simulation of the neutral chromophore in water under identical conditions as the proteins (not shown).

oxygen. This result also supports the hypothesis that the trans and cis chromophore states are neutral and anionic, respectively. Like in the dark-state X-ray structure, Val157 changes its position toward the solvent to accommodate the new conformation of the chromophore phenol ring.[11] However, the structural rearrangements of His193 and Arg66 in the dark-state X-ray structure were not observed in any of our simulations at 300 K. Note that the trans X-ray structure does not represent a fully relaxed conformation, since it was formed upon irradiation of the Dronpa molecules within the solid crystal, which strongly limits the dynamics of the protein. In contrast, the simulations that we performed allowed for full relaxation of the trans form in solution at 300 K. Specifically in the highest-temperature simulations, both His193 and Arg66 side groups undergo frequent rotamer transitions, suggesting that the absence of His193 and Arg66 rotamers at 300 K is *not* due to insufficient sampling.

In our simulations, we have only considered the anionic and neutral chromophore. The 390 and 503 nm absorption maxima of Dronpa, which correspond respectively to the neutral and anionic states of the phenolic hydroxyl of the chromophore,[1] are similar to those of other green fluorescent protein mutants. Hybrid quantum-classical calculations of the equilibria between the different protonation states of the solvated chromophore (cation, neutral, zwitterion, and anion) indicated that, at neutral pH, only the neutral and anionic states are populated in the ground state.[29] This suggests that a zwitterionic or cationic state can only exist if the protonated imidazolinone ring is sufficiently stabilized by the surrounding protein matrix. In the X-ray structure of asFP595, where a zwitterionic chromophore state has been proposed,[30] protonation of the imidazolinone nitrogen is stabilized by a H-bond with the nearby Glu215. In both cis and trans X-ray structures of Dronpa, no such stabilizing H-bond can be formed. These data suggest that the neutral and anionic chromophores are the most important ground states in Dronpa, although the existence of a zwitterionic or cationic intermediate state cannot be excluded.

Chromophore Flexibility. In both cis and trans isomers of the free chromophore in the ground state, the phenol and imidazolidone rings adopt a coplanar conformation.[31] The angle θ between the planes of the two rings is determined by φ and τ. At 300 K, θ mostly correlates with rotation about φ. Table 2 lists the average τ, φ, and θ values and standard deviations for the simulated proteins. For both rsFastLime and Dronpa, the average θ value in trans is much higher than in cis, in agreement with the X-ray structures of cis and trans Dronpa. In rsFastLime, the deviation from planarity and flexibility are much higher for the neutral than for the anionic protein, due to the higher φ force constant k_φ for anionic rsFastLime.

Compared to Dronpa and Dronpa-2, the average value and standard deviation of θ in rsFastLime are twice as high. In rsFastLime, the Thr59 methyl group collides with the chromophore phenol from below at one side, causing rotation

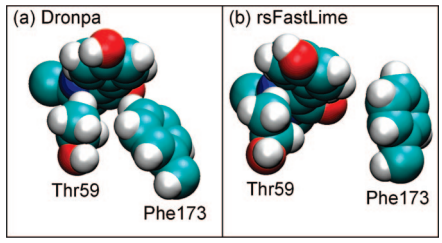


Figure 6. van der Waals representation of the chromophore in interaction with residues Thr59 and Phe173. (a) In Dronpa, φ rotation is largely offset by steric effects with Phe173, which is pushed toward the chromophore by Val157. (b) In rsFastLime, collision with Thr59 on the side of the chromophore phenol causes a disruption of coplanarity by φ rotation.

of the φ dihedral (Figure 6b). van der Waals interactions with Val157 in Dronpa cause the Phe173 phenyl ring to shift considerably toward the chromophore. As a result, the Phe173 phenyl ring is also positioned below the chromophore phenol ring, preventing large φ deviations (Figure 6a).

The cis−trans Transition Pathway. To investigate the isomerization pathways of the electronic ground state, we calculated the unbiased free energy as a function of τ and φ, on the basis of all replicas (Figure 7). For all simulated proteins, the cis-to-trans transition starts by concerted rotation of both φ and τ, the bridging methine group rotating up toward His193 and the phenol group pointing down. This so-called bottom hula-twist pathway minimizes the motion of the chromophore phenol ring and its surroundings. However, once the transition state ($\tau \sim 90°$) is crossed, the pathway becomes less defined. In addition to continuing the hula-twist path, the φ dihedral may also rotate back to its starting angle. This alternative pathway is possible thanks to the increased rotational freedom of the phenol ring around φ in the trans state. Simulations by Andresen et al.[10] on asFP595 suggested that the chromophore undergoes a trans−cis isomerization through a similar bottom hula-twist mechanism. Using force probe MD, the nonequilibrium photoisomerization process was simulated by shifting the φ and τ dihedral potentials over a short time scale. The HT-REM scheme used in this work, however, does not force the chromophore in any direction. Instead, in the highest temperature replica, the φ and τ potentials are completely removed, and the cis−trans transition path is determined solely by the surrounding protein matrix.

Despite limited sampling of the transition state, barriers of isomerization for both rsFastLime (neutral and anonic) and Dronpa-2 appear lower compared to Dronpa (Figure 7). For rsFastLime (neutral and anionic), cis−trans isomerization is hindered when Phe173 or Met159 or both are in the −60° χ1 rotameric state (Figure 8). Before trans-to-cis transition takes place, Phe173 χ1 and Met159 χ1 undergo a −60° to +60° rotamer shift. In Dronpa, sampling of the transition region is very low; isomerization is hindered by both Phe173 χ1 conformations −60° and +60°. Thus, we conclude that Val157 affects the transition barrier indirectly through both Phe173 and Met159.

In Dronpa-2, trans conformations are infrequently sampled and only in the highest temperature replicas (Figure 7c). As a result of the increased free energy of the trans state, the trans-to-cis barrier is drastically lowered, which correlates well with the strongly increased thermal recovery rate of the Dronpa-2 bright state compared to Dronpa.[3,7] Full rotation to the trans isomer is prevented by steric interaction of the phenol oxygen with the Val157 side chain. Since τ remains below 160°, Phe173 χ1 stays mainly in its +60° rotameric state, favorable for cis−trans isomerization. Also, no change in the Thr159 χ1 rotamer distribution was observed in the transition region.

Remarkably, the anionic chromophore of rsFastLime remains H-bonded to Ser142 from the cis state to well past the transition state region ($\tau \sim -30°$ to +120°). Considering the increased stability of the H-bonded chromophore, this suggests that the neutral trans chromophore deprotonates before returning to the cis state. Indeed, Fron et al.[12] compared the transient absorption properties of deuterated and nondeuterated Dronpa samples and found that exited-state proton transfer is involved in the first step of the off-to-on photoconversion, which takes place within a 4 ps time frame.

Comparing the calculated free energy barriers of Dronpa and Dronpa-2 with those obtained from fluorescence experiments (26.0 kcal mol^{-1} and 7.2 kcal mol^{-1}, respectively[8,32]), we find that the calculated barriers are much higher. Possible explanations for this difference are (i) overestimated dihedral force field parameters for φ and τ, (ii) neglect of polarization effects of the surrounding residues on the chromophore,[33] (iii) insufficient sampling of the transition state region, or (iv) involvement of a photoswitching process which does not depend on cis−trans isomerization.

Fluorescence Quantum Yield. Several sources of fluorescence quantum yield loss in fluorescent proteins have been suggested in the literature, most notably excited-state rotation about τ (twisting) with subsequent fast internal conversion, which is accelerated by increased chromophore flexibility or noncoplanarity.[34–37] Excited state quenching with molecular oxygen[38] or with water molecules through H-bonding[39] has also been proposed.

When Dronpa and its mutants are compared, the fluorescence quantum yield and switching rates appear highly correlated. As shown in Figure 7, the free energy basin of the cis state of rsFastLime and Dronpa-2 is considerably broadened toward the hula-twist transition state in comparison with Dronpa. This increased conformational freedom of τ might at least partially explain the reduced fluorescence quantum yield of rsFastLime and Dronpa-2. In rsFastLime, the high average value and standard deviation of θ in the cis state (Table 2) could also accelerate internal conversion. In Dronpa-2, the chromophore shift toward the protein surface causes increased solvent accessibility and H-bonding with nearby water molecules, which may well account for the additional reduction of the fluorescence quantum yield in Dronpa-2. Assuming that the photoswitched off state is trans, the high nonplanarity of the trans state of Dronpa and

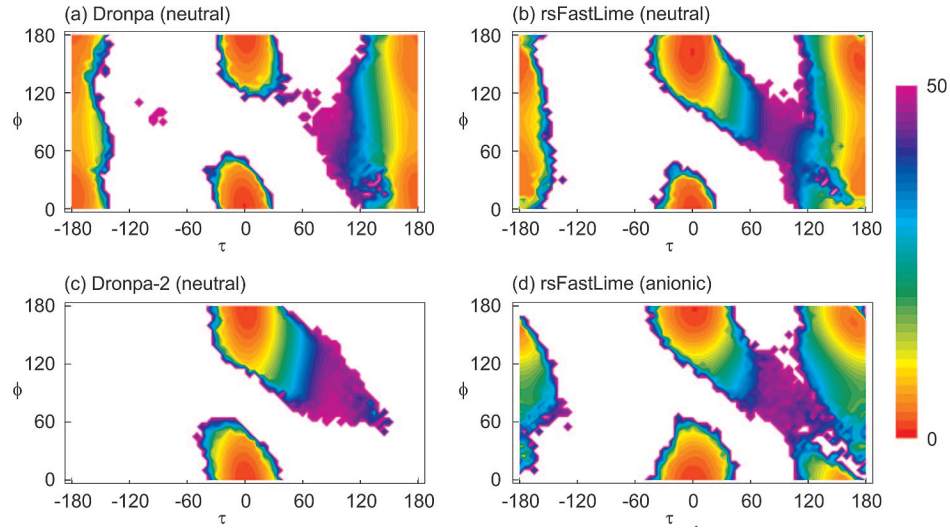


Figure 7. Unbiased free energy as a function of τ and φ in kcal mol^{-1}. (a) Dronpa with neutral chromophore (cis-to-trans free energy barrier $\Delta G^{\ddagger} = 50$ kcal mol^{-1}). (b) rsFastLime with neutral chromophore ($\Delta G^{\ddagger} = 45$ kcal mol^{-1}). (c) Dronpa-2 with neutral chromophore ($\Delta G^{\ddagger} = 46$ kcal mol^{-1}). (d) rsFastLime with anionic chromophore ($\Delta G^{\ddagger} = 45$ kcal mol^{-1}).

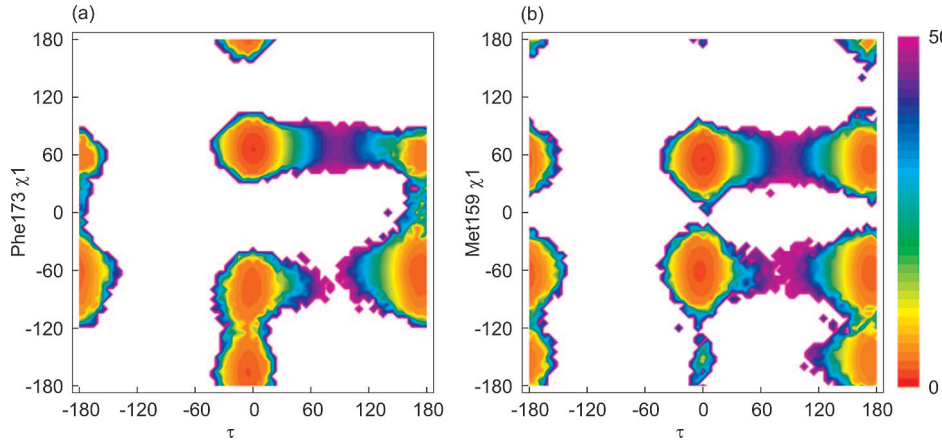


Figure 8. Unbiased free energy as a function of (a) τ and Phe173 χ1 and (b) τ and Met159 χ1, from rsFastLime (neutral chromophore) in kcal mol^{-1}.

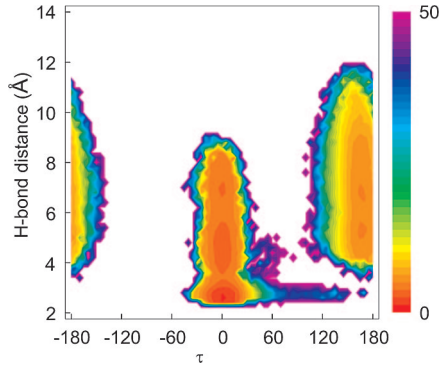


Figure 9. Unbiased free energy as a function of τ and the H-bond distance between the chromophore phenolate oxygen and Ser142 carbonyl oxygen from rsFastLime (anionic chromophore) in kcal mol^{-1}.

rsFastLime (Table 2) could be responsible for the very weak fluorescence of the photoswitched off state.

Conclusion

Canonical HT-REM simulations on the native state show that three residues, Val157, Met159, and Phe173, play a key role in the cis−trans equilibrium and bottom hula-twist transition path of Dronpa and mutants. In rsFastLime, cis−trans transition is most favorable when Phe173 and Met159 χ1 are in the +60° rotameric states. Compared to rsFastLime, the increased isomerization free energy barrier in Dronpa is attributed to steric interactions of the chromophore with the Phe173 χ1 (= +60°) rotamer in the transition state, due in turn to steric interactions of Phe173 with Val157, which cause a displacement of the Phe173 phenyl ring toward the chromophore. In comparison with Dronpa, the Dronpa-2 isomerization free energy barrier is lowered by replacement of the bulky Met159 by Thr159. The cis chromophore isomer predominated in all simulated proteins. The Dronpa-2 trans state is strongly destabilized by steric interaction of the chromophore with Val157, causing a drastic reduction of the trans-to-cis barrier. While so far mutagenic studies have only pointed to the relevance of Val157 and Met159, the present study draws attention to Phe173 as a third residue, which may strongly influence photoswitching. Subsequent study will be focused on further

increasing the photoinduced and spontaneous switching kinetics of Dronpa and variants.

Acknowledgment. We thank John Chodera (University of California, San Francisco) for providing a preview of ref 27 and kind assistance with the calculation of the unbiased free energy plots from the HT-REM simulations. This work was financially supported by the "Fonds voor Wetenschappelijk Onderzoek − Vlaanderen". P.D. is a fellow of the FWO (aspirant van het FWO). This research was conducted utilizing high-performance computational resources provided by the University of Leuven (http://ludit.kuleuven.be/hpc).

Supporting Information Available: Chromophore charges and parameters compatible with the Amber force field. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Ando, R.; Mizuno, H.; Miyawaki, A. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **2004**, *306* (5700), 1370–1373.

(2) Wilmann, P. G.; Turcic, K.; Battad, J. M.; Wilce, M. C. J.; Devenish, R. J.; Prescott, M.; Rossjohn, J. The 1.7 angstrom crystal structure of Dronpa: A photoswitchable green fluorescent protein. *J. Mol. Biol.* **2006**, *364* (2), 213–224.

(3) Stiel, A. C.; Trowitzsch, S.; Weber, G.; Andresen, M.; Eggeling, C.; Hell, S. W.; Jakobs, S.; Wahl, M. C. 1.8 angstrom bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants. *Biochem. J.* **2007**, *402*, 35–42.

(4) Habuchi, S.; Ando, R.; Dedecker, P.; Verheijen, W.; Mizuno, H.; Miyawaki, A.; Hofkens, J. Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (27), 9511–9516.

(5) Dedecker, P.; Hotta, J. I.; Flors, C.; Sliwa, M.; Uji, I. H.; Roeffaers, M. B. J.; Ando, R.; Mizuno, H.; Miyawaki, A.; Hofkens, J. Subdiffraction imaging through the selective donut-mode depletion of thermally stable photoswitchable fluorophores: Numerical analysis and application to the fluorescent protein Dronpa. *J. Am. Chem. Soc.* **2007**, *129* (51), 16132–16141.

(6) Lukyanov, K. A.; Chudakov, D. M.; Lukyanov, S.; Verkhusha, V. V. Photoactivatable fluorescent proteins. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (11), 885–891.

(7) Ando, R.; Flors, C.; Mizuno, H.; Hofkens, J.; Miyawaki, A. Highlighted generation of fluorescence signals using simultaneous two-color irradiation on Dronpa mutants. *Biophys. J.* **2007**, *92* (12), L97–L99.

(8) Flors, C.; Hotta, J. I.; Uji, I. H.; Dedecker, P.; Ando, R.; Mizuno, H.; Miyawaki, A.; Hofkens, J. A Stroboscopic Approach for Fast Photoactivation-Localization Microscopy with Dronpa Mutants. *J. Am. Chem. Soc.* **2007**, *129* (45), 13970–13977.

(9) Chudakov, D. M.; Feofanov, A. V.; Mudriku, N. N.; Lukyanov, S.; Lukyanov, K. A. Chromophore environment provides clue to "kindling fluorescent protein" riddle. *J. Biol. Chem.* **2003**, *278* (9), 7215–7219.

(10) Andresen, M.; Wahl, M. C.; Stiel, A. C.; Grater, F.; Schafer, L. V.; Trowitzsch, S.; Weber, G.; Eggeling, C.; Grubmuller, H.; Hell, S. W.; Jakobs, S. Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (37), 13070–13074.

(11) Andresen, M.; Stiel, A. C.; Trowitzsch, S.; Weber, G.; Eggeling, C.; Wahl, M. C.; Hell, S. W.; Jakobs, S. Structural basis for reversible photoswitching in Dronpa. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (32), 13005–13009.

(12) Fron, E.; Flors, C.; Schweitzer, G.; Habuchi, S.; Mizuno, H.; Ando, R.; De Schryver, F. C.; Miyawaki, A.; Hofkens, J. Ultrafast excited-state dynamics of the photoswitchable protein dronpa. *J. Am. Chem. Soc.* **2007**, *129* (16), 4870–4871.

(13) Frisch, M. J. et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford CT, 2004.

(14) Pigache, A.; Cieplak, P.; Dupradeau, F. Y. Automatic and highly reproducible RESP and ESP charge derivation: Application to the development of programs RED and X RED. *Abstr. Pap.—Am. Chem. Soc.* **2004**, *227*, U1011–U1011.

(15) Reuter, N.; Lin, R.; Thiel, W. Green fluorescent proteins: Empirical force field for the neutral and deprotonated forms of the chromophore. Molecular dynamics simulation's of the wild type and S65T mutant. *J. Phys. Chem. B* **2002**, *106* (24), 6310–6321.

(16) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18* (15), 2714–2723.

(17) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012.

(18) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, 2004.

(19) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.

(20) 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.

(21) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.

(22) 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* (2), 926–935.

(23) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116* (20), 9058–9067.

(24) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141–151.

(25) Moors, S. L. C.; Hellings, M.; De Maeyer, M.; Engelborghs, Y.; Ceulemans, A. Tryptophan rotamers as evidenced by X-ray, fluorescence lifetimes, and molecular dynamics modeling. *Biophys. J.* **2006**, *91* (3), 816–823.

(26) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.

(27) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. arXiv:0801.1426v2 [physics.comp-ph] 2008.

(28) Maragakis, P.; Spichty, M.; Karplus, M. Optimal estimates of free energies from multistate nonequilibrium work data. *Phys. Rev. Lett.* **2006**, *96* (10), 100602.

(29) Scharnagl, C.; Raupp-Kossmann, R. A. Solution pK(a) values of the green fluorescent protein chromophore from hybrid quantum-classical calculations. *J. Phys. Chem. B* **2004**, *108* (1), 477–489.

(30) Schafer, L. V.; Groenhof, G.; Klingen, A. R.; Ullmann, G. M.; Boggio-Pasqua, M.; Robb, M. A.; Grubmuller, H. Photoswitching of the fluorescent protein asFP595: Mechanism, proton pathways, and absorption spectra. *Angew. Chem., Int. Ed.* **2007**, *46* (4), 530–536.

(31) Weber, W.; Helms, V.; McCammon, J. A.; Langhoff, P. W. Shedding light on the dark and weakly fluorescent states of green fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (11), 6177–6182.

(32) Habuchi, S.; Dedecker, P.; Hotta, J. I.; Flors, C.; Ando, R.; Mizuno, H.; Miyawaki, A.; Hofkens, J. Photo-induced protonation/deprotonation in the GFP-like fluorescent protein Dronpa: mechanism responsible for the reversible photoswitching. *Photochem. Photobiol. Sci.* **2006**, *5* (6), 567–576.

(33) He, X.; Bell, A. F.; Tonge, P. J. Ground state isomerization of a model green fluorescent protein chromophore. *FEBS Lett.* **2003**, *549* (1−3), 35–38.

(34) Altoe, P.; Bernardi, F.; Garavelli, M.; Orlandi, G.; Negri, F. Solvent effects on the vibrational activity and photodynamics of the green fluorescent protein chromophore: A quantum-chemical study. *J. Am. Chem. Soc.* **2005**, *127* (11), 3952–3963.

(35) Mandal, D.; Tahara, T.; Meech, S. R. Excited-state dynamics in the green fluorescent protein chromophore. *J. Phys. Chem. B* **2004**, *108* (3), 1102–1108.

(36) Litvinenko, K. L.; Webber, N. M.; Meech, S. R. Internal conversion in the chromophore of the green fluorescent protein: Temperature dependence and isoviscosity analysis. *J. Phys. Chem. A* **2003**, *107* (15), 2616–2623.

(37) Wilmann, P. G.; Petersen, J.; Pettikiriarachchi, A.; Buckle, A. M.; Smith, S. C.; Olsen, S.; Perugini, M. A.; Devenish, R. J.; Prescott, M.; Rossjohn, J. The 2.1 angstrom crystal structure of the far-red fluorescent protein HcRed: Inherent conformational flexibility of the chromophore. *J. Mol. Biol.* **2005**, *349* (1), 223–237.

(38) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormo, M.; Remington, S. J. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (6), 2306–2311.

(39) Follenius-Wund, A.; Bourotte, M.; Schmitt, M.; Iyice, F.; Lami, H.; Bourguignon, J. J.; Haiech, J.; Pigault, C. Fluorescent derivatives of the GFP chromophore give a new insight into the GFP fluorescence process. *Biophys. J.* **2003**, *85* (3), 1839–1850.

CT8000359

JCTC Journal of Chemical Theory and Computation

Homology Models and Molecular Modeling of Human Retinoic Acid Metabolizing Enzymes Cytochrome P450 26A1 (CYP26A1) and P450 26B1 (CYP26B1)

Magnus Karlsson,[†,‡] Åke Strid,[†] Allan Sirsjö,[§] and Leif A. Eriksson*[,†]

Department of Natural Sciences and Örebro Life Science Center, Modeling and Simulation Research Center, and Department of Clinical Medicine, Örebro University, 70182 Örebro, Sweden

Received January 30, 2008

Abstract: Homology models of cytochrome P450 26A1 and cytochrome P450 26B1 were constructed using the crystal structures of human, CYP2C8, CYP2C9, and CYP3A4 as templates for the model building. The homology models generated were investigated for their docking capacities against the natural substrate all-*trans*-retinoic acid (atRA), five different tetralone-derived retinoic acid metabolizing blocking agents (RAMBAs), and R115866. Interaction energies (IE) and linear interaction energies (LIE) were calculated for all inhibitors in both homology models after molecular dynamics (MD) simulation of the enzyme−ligand complexes. The results revealed that the homologues had the capacity to distinguish between strong and weak inhibitors. Important residues in the active site were identified from the CYP26A1/B1−atRA complexes. Residues involved in hydrophobic interactions with atRA were Pro113, Phe222, Phe299, Val370, Pro371, and Phe374 in CYP26A1 and Leu88, Pro118, Phe222, Phe295, Ile368, and Tyr272 in CYP26B1. Hydrogen bonding interactions were observed between the atRA carboxylate group and Arg 90 in CYP26A1 and with Arg76, Arg95, and Ser369 in CYP26B1.

1. Introduction

Retinoic acid (RA) is the most active metabolite of vitamin A and is present in a multitude of human tissues.[1] It plays a crucial role in growth and differentiation during embryogenesis and organogenesis but also in regulation of gene expression, cellular differentiation and proliferation of epithelial cells. The all-*trans*-retinoic acid isomer (atRA) has been studied in a number of clinical situations, especially in oncology against promyelocytic leukemia[2] and in dermatology against cystic acne and photodamaged skin.[3] However, because there are common mechanisms underlying cancer and cardiovascular diseases, the use of retinoids may also show therapeutic value in some cardiovascular diseases. Restenosis and arteriosclerosis are initiated by endothelial injury in the vessel. During injury, the regulation of vascular homeostasis is disturbed leading to induction of smooth muscle cell (SMC) dedifferentiation, growth and migration.[4,5] The processes in a vessel following balloon angioplasty have been studied in animal models.[4] These studies show that atRA has the ability to reduce neointimal growth and increase the luminal area. The fact that retinoids are involved in the regulation of a variety of cellular processes suggests that they have potential abilities as therapeutic agent against vascular disorders.[6]

A rapid decrease in plasma levels of atRA has been observed in acute promyelocytic leukemia patients, which indicates an autoregulatory mechanism and the ability of atRA to induce its own catabolism.[7] This main pathway of atRA metabolism starts with hydoxylation at the C-4 position of the cyclohexenyl ring.[8] Although many microsomal cytochrome P450 enzymes are capable to initiate atRA metabolism through 4-hydroxylation, their specificity is generally low.[9] CYP26, a new family of cytochrome P450

* Corresponding author e-mail: leif.eriksson@nat.oru.se.

[†] Department of Natural Sciences and Örebro Life Science Center, Örebro University.

[‡] Modeling and Simulation Research Center, Örebro University.

[§] Department of Clinical Medicine and Örebro Life Science Center, Örebro University.

10.1021/ct800033x CCC: $40.75
Published on Web 05/02/2008

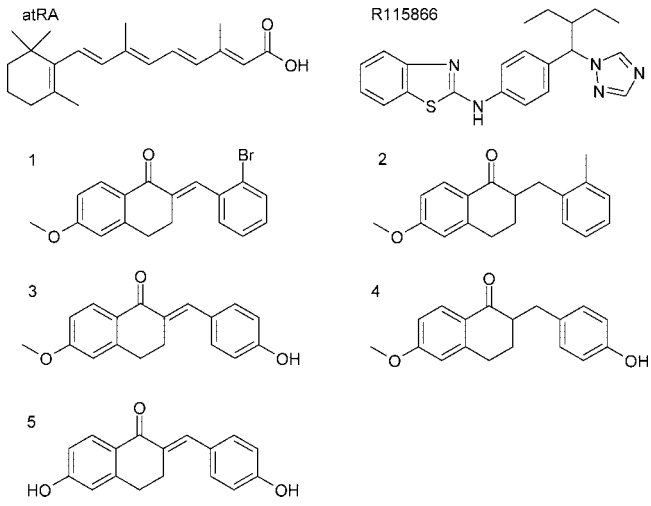


Figure 1. Ligands used for evaluation of docking capabilities in the CYP26A1 and CYP26B1 homology models. 1−5: tetralone-derived RAMBAs reported by Yee et al.[22]

enzymes, consists of three members, CYP26 A1, B1, and C1, that metabolize atRA mainly into 4-OH-RA but also to 5,6-epoxy-RA and 18-OH-RA. CYP26A1 and B1 have high specificity for atRA, whereas CYP26C1 efficiently metabolizes both atRA and 9-*cis*-RA.[10,11] CYP26B1 seems to be upregulated by lower concentrations of atRA than CYP26A1 in intimal SMCs which suggests that CYP26B1 is the dominating isoform in this celltype.[12] The discovery of retinoic acid response elements on the CYP26 promoter by Loudig et al. revealed the mechanism by which atRA induces the mRNA expressions of CYP26s.[13]

Fast metabolism and CYP26-mediated resistance are the main causes behind the search for inhibitors against CYP-mediated metabolism of atRA. The antimycotic substance ketoconazole was the first compound evaluated as a potential inhibitor against atRA metabolizing enzymes. Van Wauwe et al. demonstrated in their studies that ketoconazole prolonged the half-life of exogenously administrated atRA to animals.[14] Unfortunately, adverse side effects were observed for ketoconazole. Extensive structure−activity relationship studies on imidazole derivatives resulted in the discovery of liarozole, which is the most studied retinoic acid metabolism blocking agent (RAMBA) so far. The antitumoral effects of liarozol shown in animal studies correlate with an increase in tumor differentiation, following accumulation of atRA.[15] Liarozole has been a lead compound in the search for other RAMBAs. New inhibitors with high potency and specificity for atRA-cytochrome P450 hydroxylase based on liarzole as standard have been designed and evaluated.[16] Triazole derivatives R115866 and R116010 are highly potent (100-fold more potent than liarozole itself) and selective second-generation inhibitors of CYP26-induced atRA metabolism. R116010 was demonstrated to inhibit RA-metabolism, subsequently leading to inhibition of tumor growth at lower concentrations than atRA.[17] Stoppie et al. demonstrated the inhibitory effects of R115866 on atRA metabolism in rodents, which resulted in increased endogenous atRA and retinoidal effects including inhibition of vaginal hyperkeratinization.[18] Besides of R115866 and R116010, a large number of new RAMBAs have been reported with prominent inhibitory ability against CYP26. Recent studies of azolyl retenoids,[19] benzenacetic acid derivates[20] and 2,6-distributed naphtalenes[21] have resulted in new candidates for therapeutic use.

Lack of crystal structures of CYP26A1 and CYP26B1 makes virtual screening in the search for new and improved inhibitors difficult. In this study, bioinformatics and molecular mechanics tools necessary to generate new homologies of CYP26A1 and B1 have been utilized. The architectures of the active site of the homology models have been investigated and evaluated by their docking capabilities against known CYP26A1 inhibitors recently reported in the literature,[22] Figure 1. The homology model of CYP26A1 is also compared with the only previous homology model of CYP26A1 available, reported by Gomaa et al.[23] For CYP26B1, the current study represents the first model of that enzyme.

2. Materials and Methods

Computational Approaches. All molecular modeling was performed using the Molecular Operating Environment (MOE) 2005.06 and 2006.08 programs.[24,25] All structures of ligands, CYP26A1 and CYP26B1 models, and water−ligand and enzyme−ligand complexes were geometry minimized using the AMBER99 molecular mechanics force field[26] to within an rms gradient of 0.5 kcal mol^{-1} $Å^{-1}$. All water−ligand complexes were formed by adding a sphere of water molecules containing a layer width of 10 Å. Throughout, all systems were also surrounded by a distance dependent dielectric model.[27]

To obtain relaxed geometries, short molecular dynamics (MD) simulations were performed, followed by a final energy minimization step. The MD simulations on the CYP26A1 and CYP26B1 models, water−ligand, and enzyme−ligand complexes were performed using a canonical ensemble, NVT, with initial temperature 150 °K; heating for 50 ps; simulation temperature 300 °K; duration 500 ps; time step 1 fs (2 fs for water-ligand-complexes); temperature response 1 ps; pressure response 0.5 ps and constraint tolerance 1×10^{-9} ps. The haem group and the Fe−N or Fe−O distances to inhibitors with triazole or alcohol groups were held fixed during the simulations.

Model Building. The protein sequences of human CYP26A1 and CYP26B1 were obtained from the NCBI server (no. 2688846 and no. 9845285). Homology models of CYP26A1 and CYP26B1 were built using crystal structures of human CYP2C8 (PDB no. 1PQ2), CYP2C9 (PDB no. 1OG2) and CYP3A4 (PDB no. 1TQN) as templates.[23] Sequence similarity between CYP26A1 and B1, and the three template structures, was between 22 and 24% for all systems. Sequence alignment was performed in the MOE-align panel using default settings with alignment constraints on the haem cysteine residue, the query sequence, and the corresponding templates. CYP3A4 was used as primary template (Figure 2) and the homology models were built taking the best of ten intermediate models minimized to within a rms gradient of 0.1. The haem group was inserted and positioned with the same coordinates as the primary template, and protonation state of titratable groups at pH 7.4 were calculated.

```
CYP26A1     -----MGLPALLASALCTFVLPLLLFLAAIKLWDLYCVSGRDRSCALPLP 45
CYP26B1     MLFEGLDLVSALATLAACLVSVTLLLAVSQQLWQLRWAATRDKSCKLPIP 50
CYP3A4      ------------MDLIPNFAMETWVLVATSLVLLYIYGTHSHKLFKKLGI 38
                              :.    :: .:  :      :  .:

CYP26A1     PGTMGFPFFGETLQMVLQRRKFLQMKRRKYGFIYKTHLFGRPTVRVMGAD 95
CYP26B1     KGSMGFPLIGETGHWLLQGSGFQSSRREKYGNVFKTHLLGRPLIRVTGAE 100
CYP3A4      PGPTPLPFLGTILFYLRGLWNFDRECNEKYGEMWGLYEGQQPMLVIMDPD 88
             *.  :*::*     :     *    ..*** ::  :   :* : : ..:

CYP26A1     NVRRILLGEHRLVSVHWPASVRTILGSGCLSNLHDSSHKQRKKVIMRAFS 145
CYP26B1     NVRKILMGEHHLVSTEWPRSTRMLLGPNTVSNSIGDIHRNKRKVFSKIFS 150
CYP3A4      MIKTVLVKECYSVFTNQMPLGPMGFLKSALSFAEDEEWKRIRTLLSPAFT 138
             :: :*: *   * ..        :  . :*   ..  :. :.::   *:

CYP26A1     REALECYVPVITEEVGSSLEQWLSCGERGLLVYPEVKRLMFRIAMRILLG 195
CYP26A1     HEALESYLPKIQLVIQDTLRAWSSHP-EAINVYQEAQKLTFRMAIRVLLG 200
CYP3A4      SVKFKEMVPIISQCGDMLVRSLRQEAENSKSINLKDFFGAYTMDVITGTL 188
               ::  :* *       :.   .   ..  :  :     : : :

CYP26A1     CEPQLAGDGDSEQQLVEAFEEMTRNLFSLPIDVPFSGLYRGMKARNLIHA 245
CYP26B1     FSIPEEDLG----HLFEVYQQFVDNVFSLPVDLPFSGYRRGIQARQILQK 246
CYP3A4      FGVNLDSLNNPQDPFLKNMKKLLKLDFLDPFLLLISLFPFLTPVFEALNI 238
                  . .     :.:  :::    *  *. : :*       . : ::

CYP26A1     R--------IEQNIRAKICGLRASEAGQGCKDALQLLIEH----SWERGE 283
CYP26A1     G--------LEKAIREKLQ----CTQGKDYLDALDLLIES----SKEHGK 280
CYP3A4      GLFPKDVTHFLKNSIERMKESRLKDKQKHRVDFFQQMIDSQNSKETKSHK 288
                      : :    ::         :   * :: :*:     . :  :

CYP26A1     RLDMQALKQSSTELLFGGHETTASAATSLITYLGLYPHVLQKVREELKSK 333
CYP26B1     EMTMQELKDGTLELIFAAYATTASASTSLIMQLLKHPTVLEKLRDELRAH 330
CYP3A4      ALSDLELVAQSIIIIFAAYDTTSTTLPFIMYELATHPDVQQKLQEEIDAV 338
             :    *   :  ::*..: **::: . ::  *  :* * :*:::*: :

CYP26A1     GLLCKS--NQDNKLDMEILEQLKYIGCVIKETLRLNPPVPGGFRVALKTF 381
CYP26B1     GILHSGGCPCEGTLRLDTLSGLRYLDCVIKEVMRLFTPISGGYRTVLQTF 380
CYP3A4      LPNKAP-------VTYDALVQMEYLDMVVNETLRLFPVVSRVTRVCKKDI 381
                         :  : *  :.*:. *::*.:** . :.   *.  : :

CYP26A1     ELNGYQIPKGWNVIYSICDTHDVAEIFTNKEEFNPDRFMLPHPEDAS-RF 430
CYP26B1     ELDGFQIPKGWSVMYSIRDTHDTAPVFKDVNVFDPDRFSQARSEDKDGRF 430
CYP3A4      EINGVFIPKGLAVMVPIYALHHDPKYWTEPEKFCPERFSKKN-KDSIDLY 430
            *::*  ****  *: .*   *. .  :.: : * *:**   . :*    :

CYP26A1     SFIPFGGGLRSCVGKEFAKILLKIFTVELAR--HCDWQLLN-GPPTMKTS 477
CYP26B1     HYLPFGGGVRTCLGKHLAKLFLKVLAVELAS--TSRFELATRTFPRITLV 478
CYP3A4      RYIPFGAGPRNCIGMRFALTNIKLAVIRALQNFSFKPCKETQIPLKLDNL 480
             ::***.* *.*:* .:*   :*: .:.            .     :

CYP26A1     PTVYPVDNLPARFTHFHGEI---------------497
CYP26B1     PVLHPVDGLSVKFFGLDSNQNEILPETEAMLSATV 512
CYP3A4      PILQPEKPIVLKVHLRDGITSGP------------503
            * : * . :  :.   ..
```

Figure 2. Sequence alignment of CYP26A1, CYP26B1, and CYP3A4 with ClustalW (1.81). "*", identical residues; ":", conserved constitution; and ".", semiconserved.

Docking. Possible active sites in the receptor homology models were identified by using Alpha Site Finder.[28] Once defined, ligands were docked into the enzyme with a retain of 500 poses using the alpha triangle placement methodology with affinity ΔG as scoring function.[25] In the docking studies, flexible ligand structures were generated suing a Monte Carlo algorithm, whereas the receptors were held fixed according to the minimized geometries.

Energy Calculations of the Enzyme−Ligand Complex. The potential energies of the whole system after MD-simulation and minimization for the enzyme-ligand-complex (EL) as well as for the unbound ligand and enzyme (E+L) were calculated. The interaction energy (IE) was obtained by taking the energy difference between the two systems.

$$I_E = E_{TOT}(EL) - E_{TOT}(E+L) \quad \text{(kcal/mol)} \qquad (1)$$

The binding free energies of the ligands were calculated through the linear interaction energy (LIE). LIE basically depicts the ligand as being solvated in two different media: water and the macromolecule.[29]

$$\Delta G = \alpha\Delta(U_{e-w}^{vdw}) + \beta\Delta(U_{e-w}^{el}) \quad \text{(kcal/mol)} \qquad (2)$$

In eq 2, the ensemble averages of the interaction energies, $\Delta(U_{e-w}^{vdw})$ and $\Delta(U_{e-w}^{el})$ were obtained by calculating the van

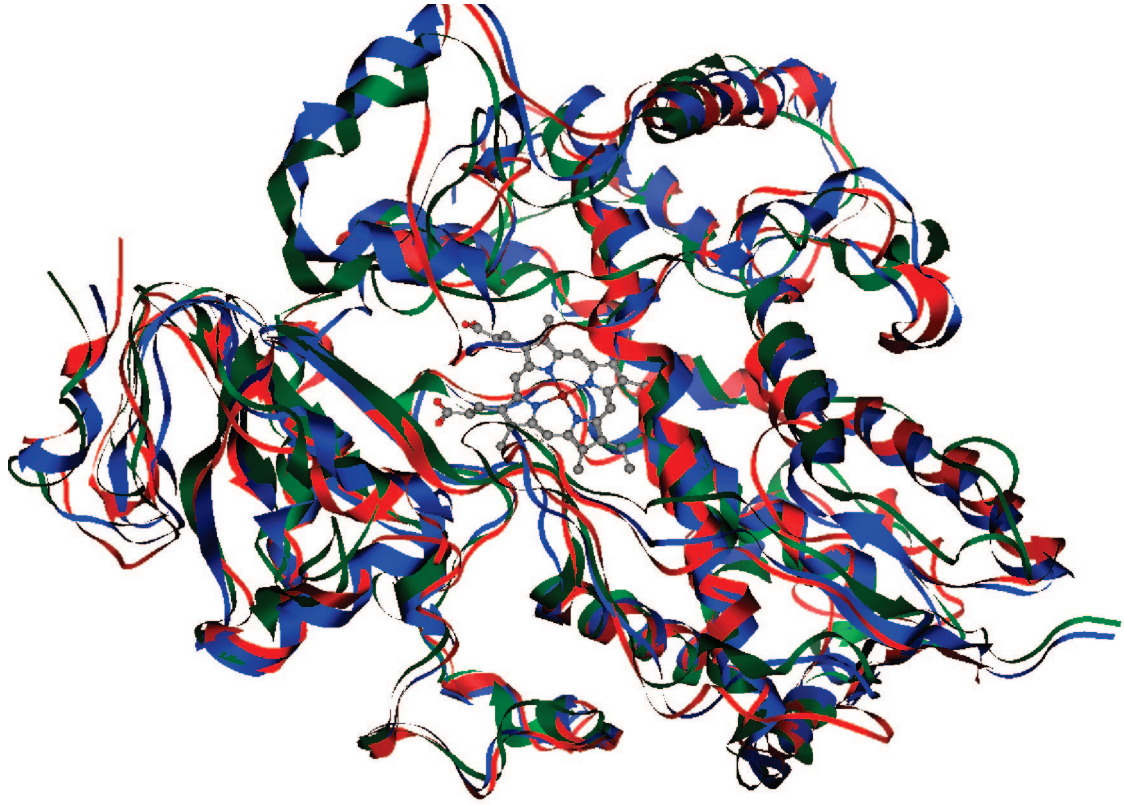


Figure 3. Superposed backbones of homologes; CYP26A1 in red, CYP26B1 in green, and CYP3A4 (1TQN-A) in blue.

der Waals and electrostatic interaction energies from ten MD trajectories for ligands free in water (w) and ligand bound to the enzyme (e). α was set to 0.9 for all compounds and β was set to 0.37 for alcohols and 0.43 for all other compounds.

3. Results and Discussion

The homology models of CYP26A1 and B1 superpose well with each other and with the major template CYP3A4, Figure 2 and 3. From the RAMPAGE server,[30] the percentage of conformations in favored regions obtained from Ramachandran plots of the superposed structures in Figure 3 was calculated to 82.9%. The active sites however display some interesting differences in architecture between the created homologes. In the CYP26 models, atRA was docked into active site orientated for oxidation with the C4 atom positioned 4.71 Å from the haem iron in CYP26A1 and 5.08 Å in CYP26B1, Figure 4. AtRA formed multiple hydrophobic interactions with the amino acid residues Pro113, Phe222, Phe299, Val370, Pro371, and Phe374 in CYP26A1 and with Leu88, Pro118, Phe222, Phe295, Ile368, and Tyr272 in CYP26B1.

Of these, equivalent residues located in the active site between the A1−B1 models were Pro113-Pro118, Phe222-Phe222, and Phe299-Phe295. Even if there are similarities, the architecture between these two homology models has sufficient differences in their active sites to render possibilities to create a selective inhibitor for each of them. In particular, the significantly large number of hydrogen bonding groups in the active site of B1 could serve as an important discriminatory factor.

Evaluation of the homology models of CYP26A1 and B1 was performed by docking known inhibitors reported in literature and calculating the properties, Figure 1. All experimental data was obtained from a MCF-7 (CYP26A1) assay for metabolic inhibition of atRA reported by Yee et al.[22]. No experimental data for inhibition of atRA metabolism in CYP26B1 is available for the inhibitors studied in this paper. The results from the CYP26B1 calculations are thus comparable only to experimental and theoretical data from CYP26A1. It should be emphasized that, as atRA is a charged acid, whereas the now studied inhibitors are not, comparison of interaction energies between inhibitors and atRA become less valid.

The theoretically calculated values for the tetralone-derived RAMBAs and R115866 in CYP26A1 agree well with the trends seen in the experimental IC_{50} data reported by Yee et al.; Table 1. Tetralone 5 was the only inhibitor that stabilized the complex (-43.44 kcal/mol) more than it should according to the experimental findings. Interestingly, tetralone 5 also represented a slightly lower free energy value (−2.63 kcal/mol) from the LIE calculation compared to the other tetralones, Table 1. One possible explanation for this behavior could be that tetralone 5 has two hydroxyl-groups in its structure, which results in increased interaction between the ligand and the enzyme when comparing with the free ligand, whereas in the LIE calculations, the "free ligand" is surrounded by hydrogen bonding water molecules. This gives an enhanced IE in the enzyme−ligand complex, but a reduced binding free energy (ΔG) from the LIE analysis.

As seen from the computed data in Tables 1 and 2, the homology models developed in this work were capable of

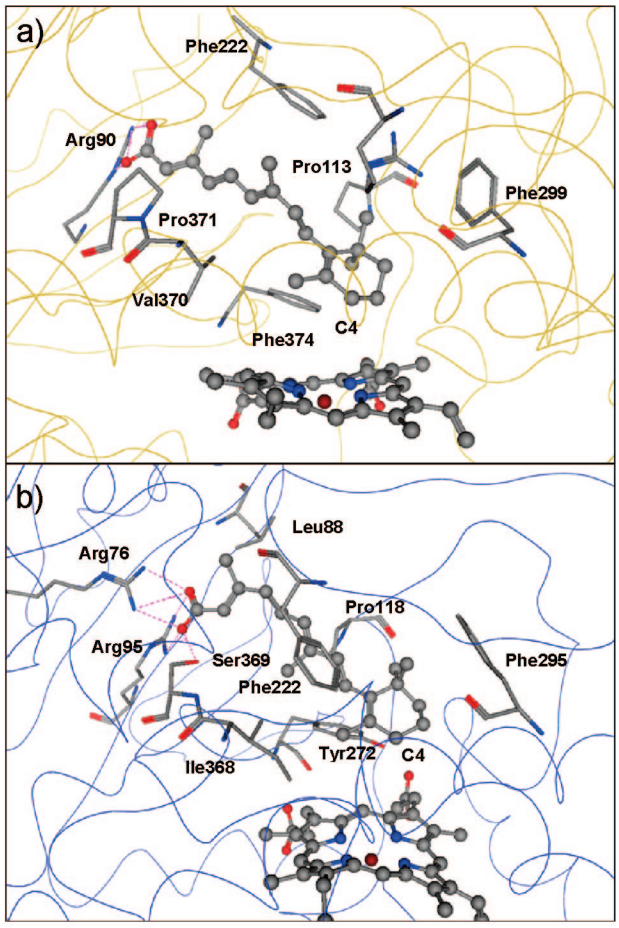


Figure 4. Active site after MD simulation of (a) CYP26A1 and (b) CYP26B1 with atRA bound. Key amino acids in stick-drawing, atRA, and haem group in ball-and-stick. Hydrogen bonds displayed as pink lines.

differentiating between "weak inhibitors" and the potent inhibitor R115866 in both CYP26A1 and CYP26B1. However, the theoretically calculated values differ between the homology models of CYP26A1 and CYP26B1. The interaction energies for the tetralones in CYP26B1, Table 2, are generally lower compared to CYP26A1, Table 1. The calculated values for the tetralones containing hydroxy-groups (tetralones 3, 4, and 5) in the CYP26B1 model differed most compared to the experimental and calculated data for CYP26A1. In Figure 5 we display the calculated interaction energies vs IC_{50} values (5a), and the calculated free energy of interaction based on LIE analysis (5b), respectively. The different LIE data for tetralone 3−5 indicate the possibility to find a selective inhibitor specific toward either A1 or B1.

Further, the tetralones in the active site seems to reach inhibitory effects in different ways. Tetralones 1 and 2 could not form any explicit interactions with the haem iron because of the lack of hydroxyl groups in their structures. Instead, they obtained their inhibitory effect mainly through hydrophobic interactions by acting in an atRA mimetic way. Tetralones 3 − 5 binds to the haem iron through their 4-hydroxyphenyl substituent, resulting in additional inhibitory activity in most cases. Tetralone 5, with an additional hydroxyl substituent, manages to obtain hydrogen bonding with Arg 90 in the CYP26A1 model and with Arg 95 in the CYP26B1 model. One essential residue for hydrophobic interactions between the tetralones and the active sites in both models is Phe222.

Comparing the CYP26A1 homologue reported by Gomaa et al. with the CYP26A1 homology model created in this study displays some differences in the active site architecture and position of atRA. In this study, the distance between the C4 atom on the cyclohexyl ring in atRA and the haem iron was 4.71Å, which is closer than the distance reported by Gomaa et al. (5.3Å). The residues in the active site involved in hydrophobic interaction are almost identical. Residues Phe222, Phe299, Pro371, and Phe374 are synonymous in both studies while Pro113 and Val370 in the current model replaced Trp112 and Phe84. The biggest difference

Table 1. Interaction Energy (eq *1*) and Linear Interaction Energy (eq *2*) Calculations from MD-Simulated CYP26A1 Complexes (kcal/mol)

ligand	IC50 (μM)[a]	E_{TOT}	E_M	I_E	$\alpha\Delta(U_{e-w}^{vdw})$	$\beta\Delta(U_{e-w}^{el})$	ΔG
atRA	inducer						
R115866	0.005	−6945.28	−6891.03	−54.25	5.943	−18.874	−12.93
tetralon 1	9	−7016.87	−6975.56	−41.31	10.451	−13.451	−3.00
tetralon 2	30	−7077.84	−7043.90	−33.94	5.046	−5.793	−0.75
tetralon 3	7	−6948.66	−6905.82	−42.86	9.600	−13.700	−3.10
tetralon 4	5	−6876.57	−6830.80	−45.57	5.646	−11.906	−6.26
tetralon 5	9	−6986.61	−6943.17	−43.44	9.418	−12.047	−2.63

[a] Results obtained from MCF-7 CYP26A1 cell assay reported by Yee et al.[22]

Table 2. Interaction Energy (eq *1*) and Linear Interaction Energy (eq *2*) Calculations from MD-Simulated CYP26B1 complexes (kcal/mol)

ligand	IC50 (μM)[a]	E_{TOT}	E_M	I_E	$\alpha\Delta(U_{e-w}^{vdw})$	$\beta\Delta(U_{e-w}^{el})$	ΔG
atRA	inducer						
R115866	0.005	−6464.79	−6410.79	−54.00	5.233	−16.609	−11.38
tetralon 1	9	−6563.46	−6520.57	−42.89	10.176	−13.861	−3.69
tetralon 2	30	−6564.84	−6528.83	−36.01	4.494	−5.054	−0.56
tetralon 3	7	−6618.86	−6572.32	−46.54	8.408	−15.224	−6.82
tetralon 4	5	−6508.45	−6461.30	−47.15	2.767	−6.800	−3.35
tetralon 5	9	−6611.97	−6571.48	−40.49	10.744	−15.458	−4.71

[a] Results obtained from MCF-7 CYP26A1 cell assay reported by Yee et al.[22]

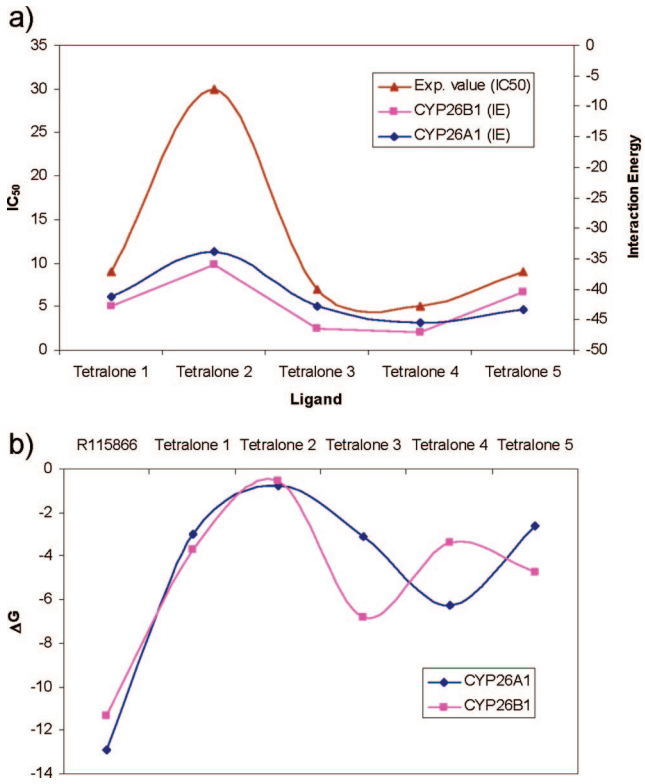


Figure 5. Comparison of (a) interaction energies of tetralone derived RAMBAs in CYP26A1 and B1 with experimental data in MCF-7 CYP26A1 cell assay[22] and (b) linear interaction energies of R115866 and tetralone derived RAMBAs in CYP26A1 and B1.

between the two models is however the residue hydrogen bonding to the carboxyl group in atRA. In our model, it binds to Arg90, whereas in the model by Gomaa et al. it binds to Arg86. We furthermore note that although the Ramachandran plots for our superposed structure in Figure 3 gave 82.9% of the conformations in favored regions, Gomaa et al.[23] obtained 81.4% for their CYP26A1-CYP3A4 model.

The differences in the active site can be explained by how the homologues were built, and which forcefield and algorithm that was used. Gomaa et al. used MOE 2004.03 with the CHARMM22 molecular mechanics forcefield and the Cartesian average model to generate the homologue. Ligands were docked using the FlexX docking of SYBYL, and the MD simulations were performed with Gromacs 3.2 in the Gromacs forcefield with a time step of 2 fs during 800 ps.

Conclusions

We have in the current work created homology models of the retinoic acid-metabolizing cytochromes, CYP26A1 and CYP26B1, and evaluated the models by docking, MD-simulation and calculation of the IE and LIE of known inhibitors bound in the active site. The IE calculations are in very good agreement with experimental data, Table 1 and 2, showing that the created models are reliable and suitable for use of further analysis in finding new RAMBAs. The results also illustrate the homologues ability to distinguish between novel and highly potent RAMBAs. The study also indicates different interactions with the active site for the inhibitors. Tetralones 1 and 2 obtained their inhibitory effects mainly through hydrophobic interactions but do not form any explicit interaction with the haem iron. Tetralones 3−5 and R115866; on the other hand, bind to the haem iron through their hydroxyl or triazole groups resulting in additional inhibitory activity in most cases. The differences in residue composition between the active sites in CYP26A1 and CYP26B1 and LIE data for tetralones 3−5 indicate the possibility of finding a selective inhibitor specific toward either A1 or B1.

Acknowledgment. We acknowledge the Modeling and Simulation Research Center (MoS) at Örebro University for a PhD scholarship grant (M.K.), and the support from the faculties of Science and Technology (L.A.E., Å.S.) and Medicine (A.S.) for additional funding. The work was also supported by the Swedish Science Research Council (VR).

References

(1) Russell, R. M.; Ross, A. C.; Trumbo, P. R.; West, K. P. Retinol equivalency ratio of beta-carotene. *J. Nutr.* **2003**, *133*, 2915–2916.

(2) Estey, E. H.; Giles, F. J.; Kantarjian, H.; O'Brien, S.; Cortes, J.; Freireich, E. J.; Lopez-Berestein, G.; Keating, M. Molecular remissions induced by liposomal-encapsulated all-trans retinoic acid in newly diagnosed acute promyelocytic leukemia. *Blood* **1999**, *94*, 2230–2235.

(3) Ahmad, N.; Mukhtar, H. Cytochrome P450: A target for drug development for skin diseases. *J. Invest. Dermatol.* **2004**, *123*, 417–425.

(4) Miano, J. M.; Kelly, L. A.; Artacho, C. A.; Nuckolls, T. A.; Piantedosi, R.; Blaner, W. S. all-trans-retinoic acid reduces neointimal formation and promotes favorable geometric remodeling of the rat carotid artery after balloon withdrawal injury. *Circulation* **1998**, *98*, 1219–1227.

(5) Gidlof, A. C.; Ocaya, P.; Krivospitskaya, O.; Sirsjo, A. Vitamin A: a drug for prevention of restenosis/reocclusion after percutaneous coronary intervention. *Clin. Sci.* **2008**, *114*, 19–25.

(6) Gidlof, A. C.; Romert, A.; Olsson, A.; Torma, H.; Eriksson, U.; Sirsjo, A. Increased retinoid signaling in vascular smooth muscle cells by proinflammatory cytokines. *Biochem. Biophys. Res. Commun.* **2001**, *286*, 336–342.

(7) Muindi, J.; Frankel, S. R.; Miller, W. H.; Jakubowski, A.; Scheinberg, D. A.; Young, C. W.; Dmitrovsky, E.; Warrell, R. P. Continuous Treatment with All-Trans Retinoic Acid Causes a Progressive Reduction in Plasma Drug Concentrations - Implications for Relapse and Retinoid Resistance in Patients with Acute Promyelocytic Leukemia. *Blood* **1992**, *79*, 299–303.

(8) Napoli, J. L. Retinoic acid biosynthesis and metabolism. *Faseb J.* **1996**, *10*, 993–1001.

(9) McSorley, L. C.; Daly, A. K. Identification of human cytochrome P450 isoforms that contribute to all-trans-Retinoic Acid 4-hydroxylation. *Biochem. Pharmacol.* **2000**, *60*, 517–526.

(10) White, J. A.; Becket Jones, B.; Guo, Y. D.; Dilworth, F. J.; Bonasoro, J.; Jones, G.; Petkovich, M. cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RAI) identifies a novel family of cytochromes P450 (CYP26). *J. Biol.*

Chem. **1997**, *272*, 18538–18541.

(11) Taimi, M.; Helvig, C.; Wisniewski, J.; Ramshaw, H.; White, J.; Amad, M.; Korczak, B.; Petkovich, M. A novel human cytochrome P450, CYP26C1, involved in metabolism of 9-cis and all-trans isomers of retinoic acid. *J. Biol. Chem.* **2004**, *279*, 77–85.

(12) Ocaya, P.; Gidlof, A. C.; Olofsson, P. S.; Torma, H.; Sirsjo, A. CYP26 inhibitor R115866 increases retinoid signaling in intimal smooth muscle cells. *Arterioscler., Thromb., Vasc. Biol.* **2007**, *27*, 1542–1548.

(13) Loudig, O.; MacLean, G. A.; Dore, N. L.; Luu, L.; Petkovich, M. Transcriptional co-operativity between distant retinoic acid response elements in regulation of Cyp26A1 inducibility. *Biochem. J.* **2005**, *392*, 241–248.

(14) Vanwauwe, J. P.; Coene, M. C.; Goossens, J.; Vannijen, G.; Cools, W.; Lauwers, W. Ketoconazole Inhibits the Invitro and Invivo Metabolism of All-Trans-Retinoic Acid. *J. Pharmacol. Exp. Ther.* **1988**, *245*, 718–722.

(15) Acevedo, P.; Bertram, J. S. Liarozole Potentiates the Cancer Chemopreventive Activity of and the up-Regulation of Gap Junctional Communication and Connexin43 Expression by Retinoic Acid and Beta-Carotene in 10t1/2 Cells. *Carcinogenesis* **1995**, *16*, 2215–2222.

(16) Njar, V. C. O.; Gediya, L.; Purushottamachar, P.; Chopra, P.; Vasaitis, T. S.; Khandelwal, A.; Mehta, J.; Huynh, C.; Belosay, A.; Patel, J. Retinoic acid metabolism blocking agents (RAMBAs) for treatment of cancer and dermatological diseases. *Bioorg. Med. Chem.* **2006**, *14*, 4323–4340.

(17) Van Heusden, J.; Van Ginckel, R.; Bruwiere, H.; Moelans, P.; Janssen, B.; Floren, W.; van der Leede, B. J.; van Dun, J.; Sanz, G.; Venet, M.; Dillen, L.; Van Hove, C.; Willemsens, G.; Janicot, M.; Wouters, W. Inhibition of all-TRANS-retinoic acid metabolism by R116010 induces antitumour activity. *Br. J. Cancer* **2002**, *86*, 605–611.

(18) Stoppie, P.; Borgers, M.; Borghgraef, P.; Dillen, L.; Goossens, J.; Sanz, G.; Szel, H.; Van Hove, C.; Van Nyen, G.; Nobels, G.; Vanden Bossche, H.; Venet, M.; Willemsens, G.; Van Wauwe, J. R115866 inhibits all-trans-retinoic acid metabolism and exerts retinoidal effects in rodents. *J. Pharmacol. Exp. Ther.* **2000**, *293*, 304–312.

(19) Patel, J. B.; Huynh, C. K.; Handratta, V. D.; Gediya, L. K.; Brodie, A. M. H.; Goloubeva, O. G.; Clement, O. O.; Nanne, N. P.; Soprano, D. R.; Njar, V. C. O. Novel retinoic acid metabolism blocking agents endowed with multiple biological activities are efficient growth inhibitors of human breast and prostate cancer cells in vitro and a human breast tumor xenograft in nude mice. *J. Med. Chem.* **2004**, *47*, 6716–6729.

(20) Vasudevan, J.; Johnson, A. T.; Huang, D.; Chandrartna, R. A. U.S. Patent 6.252.090, 2005.

(21) Mulvihill, M. J.; Kan, J. L. C.; Cooke, A.; Bhagwat, S.; Beck, P.; Bittner, M.; Cesario, C.; Keane, D.; Lazarescu, V.; Nigro, A.; Nillson, C.; Panicker, B.; Smith, V.; Srebernak, M.; Sun, F. L.; O'Connor, M.; Russo, S.; Fischetti, G.; Vrkljan, M.; Winski, S.; Castelhano, A. L.; Ernerson, D.; Gibson, N. W. 3-[6-(2-dimethylamino-1-imidazol-1-yl-butyl)-naphthalen-2-yloxy]-2,2-dim ethyl-propionic acid as a highly potent and selective retinoic acid metabolic blocking agent. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2729–2733.

(22) Yee, S. W.; Jarno, L.; Gomaa, M. S.; Elford, C.; Ooi, L. L.; Coogan, M. P.; McClelland, R.; Nicholson, R. I.; Evans, B. A. J.; Brancale, A.; Simons, C. Novel tetralone-derived retinoic acid metabolism blocking agents: Synthesis and in vitro evaluation with liver microsomal and MCF-7 CYP26A1 cell assays. *J. Med. Chem.* **2005**, *48*, 7123–7131.

(23) Gomaa, M. S.; Yee, S. W.; Milbourne, C. E.; Barbera, M. C.; Simons, C.; Brancale, A. Homology model of human retinoic acid metabolising enzyme cytochrome P450 26A1 (CYP26A1): Active site architecture and ligand binding. *J. Enzyme Inhib. Med. Chem.* **2006**, *21*, 361–369.

(24) *Molecular Operating Environment*, version 2005.06; Chemical Computing Group: Montreal, QC, 2005.

(25) Molecular Operating Environment,version 2006.08; Chemical Computing Group: Montreal, QC, 2006.

(26) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(27) Distance-dependent dielectric: $E_{SOL} = 0$, $E_{ELE} = q_i q_j / 4\pi\varepsilon_0 \varepsilon r_{ij}^2$

(28) Edelsbrunner, H. *Weighted Alpha Shapes*; technical paper of the Department of Computer Science; University of Illinois: Champaign, IL

(29) Aqvist, J.; Medina, C.; Samuelsson, J. E. New Method for Predicting Binding-Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385–391.

(30) RAMPAGE server: http://mordred.bioc.cam.ac.uk/~rapper/rampage.php (accessed Dec 12, 2007).

CT800033X