Characterization of Aromatic−Amide(Side-Chain) Interactions in Proteins through Systematic ab Initio Calculations and Data Mining Analyses

Guilin Duan,[†] Vedene H. Smith, Jr.,*[,†] and Donald F. Weaver[†,‡]

Departments of Chemistry and Medicine, Queen's University, Kingston, Ontario, Canada K7L 3N6

Received: September 21, 1999; In Final Form: December 15, 1999

In this study, noncovalent interactions between aromatic groups and side-chain amides in proteins were characterized. To elucidate the nature and structure−strength relationship of the interaction, the geometries and interaction potential energy surfaces for the benzene−formamide model complex were exhaustively and systematically studied at the MP2 level of theory. The effects of basis set size and basis set superposition error were investigated for 15 selected complex structures. The results indicate that the aromatic−amide (side-chain) interaction can achieve a significant binding energy of up to 4.0 kcal/mol over a wide conformational space. The interaction involves the entire side-chain amide group rather than only its amine portion. Both dispersion and electrostatic interactions are the major contributors for the binding energy, and the π electron charge distributions in both groups and the dipole moment of the side-chain amide group are crucial to the interaction. The importance of such an interaction in proteins was verified through data mining analyses of 1029 X-ray protein structures. The interaction naturally occurs in proteins with a frequency of more than one per two proteins on a statistical average and is of significance for some protein structure. The interaction was also found to play a role in determining the biological activity of some proteins. Our study not only emphasizes the significance of aromatic−amide(side-chain) interactions in proteins but also deepens our understanding of noncovalent interactions involving benzene or other aromatic groups.

I. Introduction

Noncovalent interactions involving benzene molecules or aromatic groups have emerged as important subjects in recent years, experimentally and theoretically, due to their chemical and biological significance. Accordingly, various complexes such as C_6H_6-X ($X = N_2$, He, Ne, Ar),[1] C_6H_6-HCl,[2] $C_6H_6-CO_2$,[3] C_6H_6-CO,[4] C_6H_6-HCO,[4] $C_6H_6-H_2O$,[5−7] $C_6H_6-NH_3$,[8,9] and $C_6H_6-C_6H_6$ [10−27] have been employed as model systems. The benzene−benzene interaction has been extensively studied because it models the aromatic−aromatic interaction in proteins. Experimental results indicate that there are at least two stable low-energy configurations associated with the benzene−benzene complex.[23,24] At the MP2/6-311(2d,2p) level, Jaffe and Smith[25] found that the displaced face-to-face configuration corresponds to the largest binding energy (3.33 kcal/mol); Hunter and Sanders[26] observed that the main contributions to the binding energy come from both dispersion and energetically favorable quadrupole−quadrupole electrostatic interactions. Protein structure database searches[22,27] confirm the general occurrence and significance of such an interaction in protein tertiary structures. Kollman and co-workers[22] found that, in X-ray protein structures, the perpendicular and the parallel-displaced phenyl−phenyl configurations are more common than the sandwich ones. They explained this observation as follows: the former two arrangements, especially the perpendicular one which exposes three phenyl faces to the outside, offer greater possibility for "secondary" interactions with other secondary groups. Such an observation raises a question—in addition to the aromatic ring itself, what other possible groups significantly interact with aromatic groups in a protein?

On the basis of the "surprising" strength of the benzene−cation interaction,[28,29] Dougherty suggested that a cationic residue should be a chemical species for such a function.[32,33] In the early 1990s, several studies on $C_6H_6-H_2O$ and $C_6H_6-NH_3$ interactions demonstrated the benzene molecule's capability of accepting an electron-deficient hydrogen atom from an H−O or H−N bond to form a π-type hydrogen bond.[6−9] This finding led to investigations of the aromatic−amine interaction in proteins. Statistical analyses of this interaction in experimental protein structures showed that only the minority of the close aromatic−amine contacts have the nitrogen atoms located above the aromatic ring. In these cases, the amine prefers to lie parallel to the aromatic ring. The hydrogen bonding structure, in which the amine is approximately perpendicular to the aromatic ring with one of its hydrogen atoms pointing toward the center of the aromatic ring, is unexpectedly rare between an aromatic group and isolated amine groups (−NH−).[30,31] Furthermore, the aromatic−amine interaction is weak—the largest binding energy is achieved by the optimum hydrogen bonding structure and is approximately 1.5 kcal/mol.[9,32,34] These facts favor the conclusion that the interaction between an aromatic ring and an isolated amine is not of general significance for protein structure.

Mitchell et al.'s evaluations[31] of two benzene−formamide interaction patterns using single excitation IMPT/3-21G energies plus dispersion energies from the model of Huiszoon and Mulder[36] indicate a significant binding interaction between aromatics and amides. Similarly, we proposed the amide group (in both protein side chains and backbones) to be an important structural entity binding aromatic groups in proteins.[35] From a qualitative point of view, a significant aromatic−amide interac-

* To whom correspondence should be addressed. Telephone: (613) 533-2650. Fax: (613) 533-6669. E-mail: vhsmith@chem.queensu.ca.
[†] Department of Chemistry.
[‡] Department of Medicine.

10.1021/jp993381f CCC: $19.00
Published on Web 04/12/2000

tion is understandable. The amide group (−NHC(O)−) has π electron charge distribution and possesses not only a quadrupole moment but also a dipole moment that the benzene molecule does not have. Consequently, an aromatic−amide interaction may have a larger electrostatic interaction energy and thus a greater binding energy than an aromatic−aromatic one. Quantitatively, our theoretical calculations with the MP2/6-311G-(2d,2p) method on 10 selected benzene−formamide structures show that such interactions can achieve a binding energy of up to 4.0 kcal/mol.[35] More importantly, the amide group is so abundant in proteins that the energetically significant aromatic−amide interaction may have significant frequency. This study presents further insight into the interaction between aromatic groups and side-chain amides (amide(S)). In particular, the benzene−formamide model complex is exhaustively and systematically calculated to provide a clarification concerning the nature of this interaction and to establish its structure−strength relationship. The significance of such interactions in proteins is further demonstrated with examples in protein X-ray structures.

This study is presented in two parts. In part A, we investigate the benzene−formamide complex using ab initio molecular orbital calculations to evaluate both the fundamental characteristics and the strength of the aromatic−amide(S) interaction and to identify a reliable theoretical method to describe such nonconvalent interaction systems. We optimized 15 benzene−formamide configurations (Figure 1). These 15 configurations were selected to reflect the configurational diversity as well as the nature and magnitude of the interaction. The interaction potential energy surfaces (IPESs) with respect to intermolecular geometries were systematically calculated. On the basis of the results of these calculations and the molecular electrostatic potentials (MEP) of both benzene and formamide, the nature and characteristics of the interaction were analyzed. In part B, the biological significance of such interactions in protein structures was explored through data mining analyses of a protein structural database. Due to the complexity of the interaction, the data mining analysis was strategically applied to the interaction between the phenyl group (phenyl(F)) of the phenylalanine residue and the amide(S) group.

II. Methods and Calculations

(A) Ab Initio Molecular Orbital Calculations. *(1) Geometry.* For each of 15 benzene−formamide configurations, the intermolecular geometries defined in Figure 1 were obtained through two optimization steps. First, a scan of the rigid potential energy surface was performed at the HF/6-31G** level; then the configuration corresponding to the lowest HF/6-31G** molecular energy was optimized at the MP2/6-31G** level of theory. All dihedral angles were kept at 0, 90, or 180°, as indicated for a given configuration in Figure 1. Throughout the calculations, the geometries of both benzene and formamide were kept rigid and planar, and their experimental bond lengths and bond angles[37,38] were used.

(2) Interaction Energy. Interaction energies were calculated using the ab initio supermolecular method at the MP2 level of theory. In such a computational scheme, the total interaction energy ΔE of a complex formed by two monomers A and B is obtained by

$$\Delta E = E(\mathrm{AB}) - E(\mathrm{A}) - E(\mathrm{B}) = \Delta E^{(\mathrm{HF})} + \Delta E^{(\mathrm{MP2})}$$

where E(AB), E(A), and E(B) are the molecular energies of the complex A−B, monomer A, and monomer B respectively; $\Delta E^{(\mathrm{HF})}$ and $\Delta E^{(\mathrm{MP2})}$ are respectively the Hartree−Fock and the MP2−correlation interaction energy. Note that the "MP2−correlation interaction energy" represents the interaction energy contributed by the inclusion of the electronic correlation at the MP2 level. The partitioning scheme due to Chalasinski and Szczesniak[39−41] shows that the Hartree−Fock term includes electrostatic, exchange, and induction interactions, and the MP2−correlation part includes dispersion interactions as well as correlation corrections to the electrostatic and exchange interactions at the corresponding electronic correlation level. The drawback for such a computational scheme is the existence of the basis set superposition error (BSSE), which has its origin in the incompleteness of the basis set applied. To obtain a physically meaningful interaction energy, the BSSE must be eliminated.[42−45] In this study, the counterpoise method (CP) of Boys and Bernardi[46] was used for this purpose. The calculation of the interaction energies of these 15 configurations was performed at the MP2 level with five standard polarization or diffused polarization basis sets with double- or triple-ζ valence orbitals, i.e., 6-311G(d,p), 6-31+G(2d,p), 6-311G(2d,p), 6-311G-(2d,2p), and 6-311+G(2d,p). The interaction potential energy surfaces were calculated with the MP2/6-311G(2d,2p) method.

All calculations were performed using the Gaussian 94 package[47] running on an IBM eight-node SP2 processor.

(B) Database Search. To identify the importance of the aromatic−amide(S) interaction in proteins, data mining analyses of a protein structural database were performed. Due to the complexity of such an interaction system, we elucidated its characteristics and significance in proteins by investigating the interaction between the phenyl(F) group of the phenylalanine residue and the amide(S) group of various side chains. Six geometrical parameters P, L, A1, A2, DA1, and DA2, illustrated in Figure 2, define a phenyl(F)−amide(S) interaction. Based on results for the benzene−formamide complex (see below), the following cutoffs are the suitable criteria for a significant phenyl(F)−amide(S) interaction: $P \leq 6.0$ Å; $L \leq 3.0$ Å; and $R \leq 6.0$ Å (R is the distance between X1 and X2 points in Figure 2). In total, 1029 crystallographic protein structures in the Brookhaven Protein Data Bank (PDB)[48] with a resolution less than 1.8 Å were analyzed. The data mining analyses were performed with the program BEAM[49a] written in Perl[49b] by one of the authors (G.D.).

III. Results and Discussion

(A) Ab Initio Molecular Orbital Calculations. *(1) Optimized Intermolecular Geometries of 15 Benzene−Formamide Configurations.* The HF- and MP2-optimized intermolecular geometries are summarized in Table 1. Compared with HF-optimized results, the MP2 optimization tends to bring the two monomers closer. This is because the dispersion interaction cannot be fully reproduced at the HF level of theory. However, for each configuration, the MP2 optimization maintains the configurational features obtained by the HF optimization, which is reflected by the small difference between the two optimization results. Therefore, our results are in agreement with the known conclusion that the optimal intermolecular geometries of a weakly bonded system are not sensitive to the theoretical level applied and that the mutual orientation between two monomers is determined primarily by the electrostatic interaction even for the case in which the electrostatic interaction is not a dominant contributor to the binding energy.[50] Hobza et al. demonstrated that the MP2-optimized intermolecular geometry of the benzene−benzene complex was almost the same as that obtained with CCSD(T) optimizations.[50]

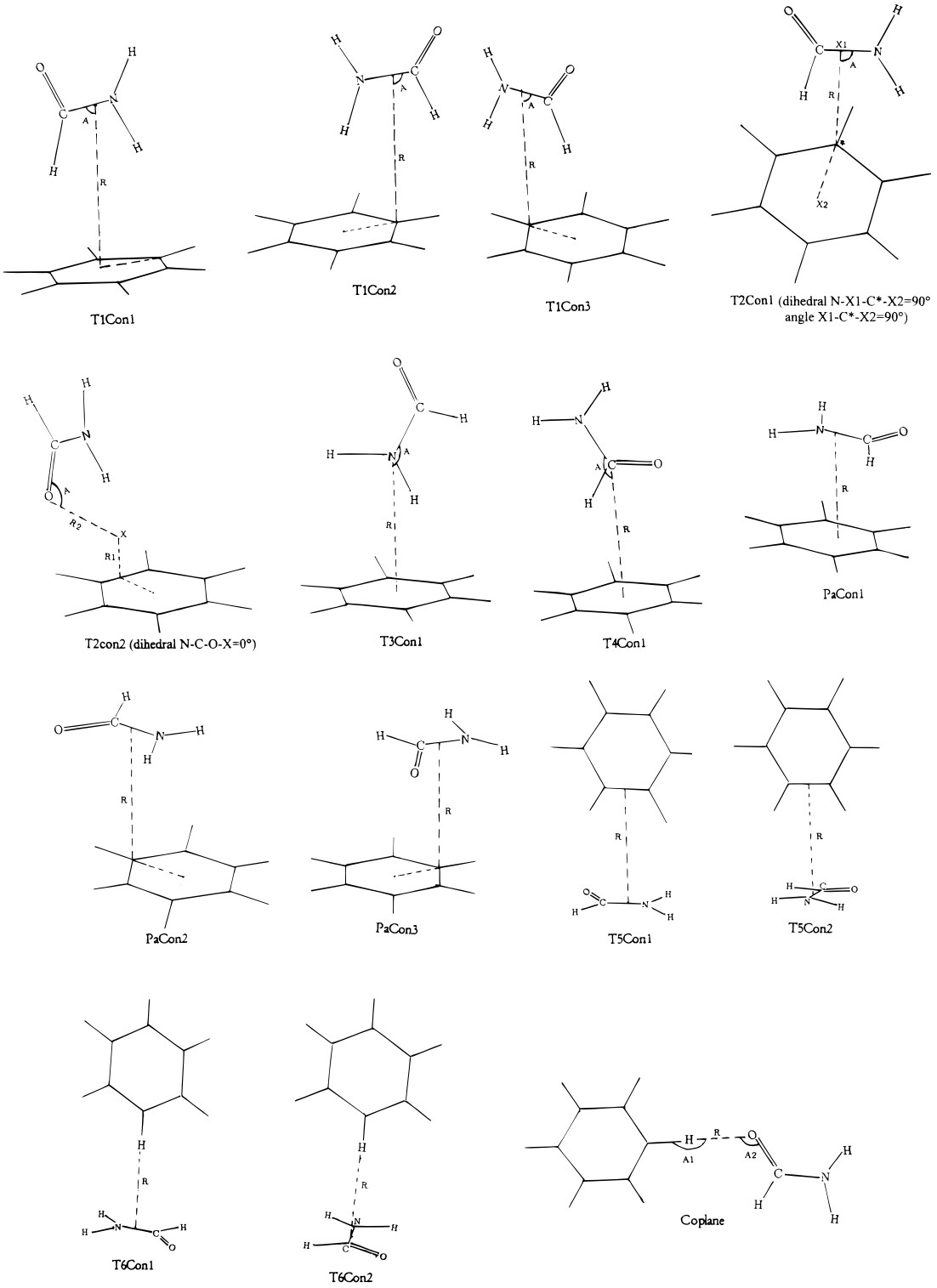

Figure 1. 15 calculated benzene−formamide interaction configurations and their notations.

(2) Interaction Energies of the Benzene−Formamide Interaction. The CP-corrected total interaction energies $\Delta E^{(cp)}$ and corresponding BSSEs for 15 calculated configurations at their MP2-optimized geometries with five basis sets are given in Table 2. A negative energy value means a binding energy or a stabilization. On the basis of these results, each of these interactions corresponds to stabilization.

Reliability of the Calculated Interaction Energies. For these 15 configurations, the calculated total interaction energies at the 6-31+G(2d,p), 6-311G(2d,p), 6-311G(2d,2p), and 6-311+G-(2d,p) basis sets are quite close. This suggests that the calculated results may converge at 6-311+G(2d,p) and that the application of a larger basis set will not significantly increase the binding energy at the MP2 level. Compared to the four basis sets, 6-311G(d,p) tends to produce the smallest binding energies for all 15 configurations. For the face-to-face configurations, PaCon1, PaCon2, and PaCon3, the calculated binding energies at this basis set are about 1.0 kcal/mol smaller than those at the other four basis sets. For the other 12 configurations, the energies are about 0.5 kcal/mol lower. Therefore, at the MP2 level, basis sets smaller than 6-311G(d,p) underestimate the binding energy of the benzene−formamide interaction.

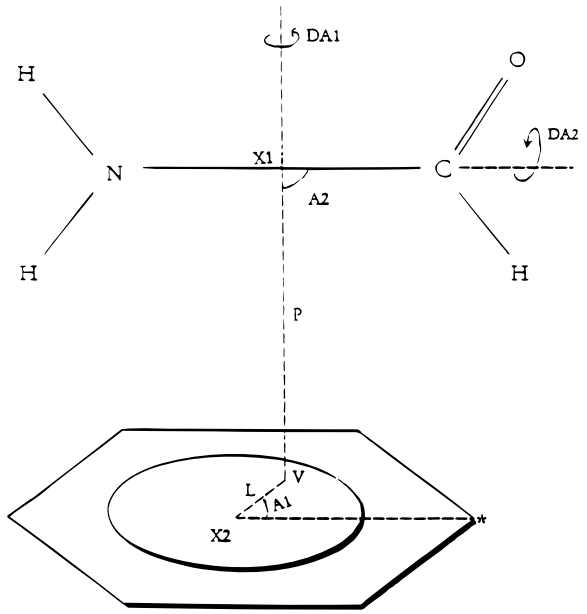

Figure 2. Definition of the benzene−formamide interaction. X1, X2, and *V* are the midpoint of formamide's C−N bond, the centroid of the benzene ring, and the projection of X1 on the benzene molecular plane, respectively. *P*, distance X1−*V*; L, distance X2−*V*; A1, angle *V*−X2−C*; A2, angle C−X1−*V*; DA1, dihedral angle C−X1−*V*−X2; and DA2, dihedral angle O−C−X1−*V*. A benzene−formamide interaction can be defined by these six parameters. An aromatic−amide(S) interaction can also be specified with these six parameters.

TABLE 1: HF- and MP2-Optimized Intermolecular Geometrical Parameters for 15 Calculated Benzene−Formamide Interaction Configurations (Distance (*R*), Å; Angle (*A*), deg)

configuration	HF/6-31G**	MP2/6-31G**
T1Con1	$R = 3.95$, $A = 110.1$	$R = 3.64$, $A = 113.8$
T1Con2	$R = 3.88$, $A = 101.8$	$R = 3.49$, $A = 101.9$
T1Con3	$R = 4.01$, $A = 124.4$	$R = 3.77$, $A = 125.6$
T2Con1	$R = 3.98$, $A = 80.2$	$R = 3.31$, $A = 81.0$
T2Con2	R1 = 1.31, R2 = 3.24, $A = 83.0$	R1 = 1.08, R2 = 3.10, $A = 81.9$
T3Con1	$R = 3.52$, $A = 151.2$	$R = 3.25$, $A = 150.3$
T4Con1	$R = 4.00$, $A = 150.0$	$R = 3.49$, $A = 144.9$
PaCon1	$R = 3.97$	$R = 3.40$
PaCon2	$R = 3.94$	$R = 3.31$
PaCon3	$R = 3.99$	$R = 3.36$
T5Con1	$R = 3.89$	$R = 3.78$
T5Con2	$R = 3.69$	$R = 3.57$
T6Con1	$R = 4.03$	$R = 3.85$
T6Con2	$R = 4.02$	$R = 3.90$
Coplane	$R = 2.70$, A1 = 179.0, A2 = 125.0	$R = 2.38$, A1 = 178.2, A2 = 118.8

The magnitude of the BSSE ranges from 0.5 to 2.0 kcal/mol, which is substantial relative to the total binding energy. The BSSEs are somewhat larger for the face-to-face configurations than for the other configurations. Two diffuse basis sets, 6-31+G(2d,p) and 6-311+G(2d,p), correspond to smaller BSSEs than the other four basis sets without diffuse functions. The CP method is an approximate and at times controversial method for the correction of BSSE.[51−54] Nevertheless, it is a prevailing method. Calculations on weakly bonded systems using the supermolecular scheme employ the CP method and have confirmed its effectiveness and reliability. Persuasive proofs in favor of this method include the following: (i) the CP-corrected interaction energy is generally agreeable to experimental results,[55−57] and (ii) only the CP-corrected interaction energy agrees quantitatively with the value calculated using intermolecular Møller−Plesset perturbation theory (IMPT) which is free of the BSSE.[58,59] Accordingly, for the system under study, we believe that the CP-corrected total interaction energies are reliable.

Relative to the results calculated at the MP4(SDTQ) or CCSD(T) level, MP2 tends to overestimate the binding energy of the benzene−benzene interaction, especially the binding energies of its face-to-face configurations in which the dispersion energy is both large and dominant.[25,50,60] Compared with MP2, both CCSD(T) and MP4(SDTQ) are more reliable methods for the description of weakly bonded systems.[61] However, one cannot negate the applicability of the MP2 method to weakly bonded systems for the following reasons: (i) with quality basis sets, the MP2 method can produce interaction energies in good agreement both with experimental values and with the results from the CCSD(T) or MP4(SDTQ) methods for a wide variety of complexes (e.g. the strongly bonded systems,[58,62,63] H_2O-H_2O, HF−HF, NH_3-NH_3, and $C_2H_2-H_2O$; the charge-transfer system, $C_2H_2-Cl_2$;[64] and the weak dispersion systems, $C_6H_6-N_2$ [1,65] and CH_4-H_2O [62,63]); (ii) the phenomenon of overestimation of the binding energy by MP2 occurs only in systems which assume the face-to-face configuration such that the dispersion energy is both large and dominantly important to the total interaction energy; (iii) for the face-to-face configuration, this phenomenon does not always occur (e.g. MP2 and CCSD(T) produce very similar interaction energies for the formamide−formamide complex with the face-to-face configuration[61]); (iv) even for the benzene−benzene interaction with a face-to-face configuration, at medium size basis sets, the MP2 method can produce interaction energies similar to those calculated with MP4(STDQ) or CCSD(T) at larger basis sets—the overestimation due to MP2 is compensated for by the incompleteness of the basis set applied. Jaffe and Smith[25] conclude that, with the 6-311G(2d,2p) basis set, MP2 produces reliable interaction energies for the benzene−benzene interaction.

Our calculations show that the dispersion interaction in the benzene−formamide complex is not so important as that in the benzene−benzene one. Table 3 lists the CP-corrected HF and MP2−correlation interaction energies of 15 calculated interaction configurations at five basis sets and at their MP2-optimized geometries. The HF interaction energies are favorable energies between −0.55 and −0.95 kcal/mol for the configurations T1Con1, T1Con2, T1Con3, T2Con1, T2Con2, T3Con1, and Coplane and are unfavorable between 0.54 and 3.02 kcal/mol for the other eight configurations. The MP2−correlation interaction energies are always favorable (negative) with magnitudes ranging from 0.5 to 4.33 kcal/mol. For the benzene−benzene interaction, the corresponding HF interaction energies for most of the calculated configurations are unfavorable, while the favorable MP2−correlation interaction energies are much larger in magnitude than those in the benzene−formamide interaction.[17,25] For example, at the 6-31+G(2d,p) basis set, for the benzene−formamide interaction, the configuration PaCon1 has the most attractive CP-corrected MP2−correlation interaction energy (−4.07 kcal/mol) among the 15 calculated configurations. For the benzene−benzene interaction, at the same basis set, a similar face-to-face configuration corresponds to a non-CP-corrected MP2−correltion interaction energy as large as −10.40 kcal/mol.[25] After considering the BSSE correction (usually about 2.50 kcal/mol in magnitude[17,25]), its CP-corrected MP2−correlation interaction energy should be approximately −7.9 kcal/mol, which is much larger in magnitude than that associated with the benzene−formamide interaction. Table 4 contains the CP-corrected HF and MP2−correlation interaction energies of several benzene−formamide interaction configura-

TABLE 2: CP-Corrected Total Interaction Energies ($\Delta E^{(cp)}$, kcal/mol) and Corresponding BSSEs for 15 Calculated Benzene−Formamide Interaction Configurations at the MP2 Level with Various Basis Sets and MP2-Optimized Geometries

	6-31+G(2d,p)		6-311G(d,p)		6-311G(2d,p)		6-311G(2d,2p)		6-311+G(2d,p)	
configuration	$\Delta E^{(cp)}$	BSSE	$\Delta E^{(cp)}$	BSSE	$\Delta E^{(cp)}$	BSSE	$\Delta E^{(cp)}$	BSSE	$\Delta E^{(cp)}$	BSSE
T1Con1	−3.87	1.30	−3.10	2.00	−3.88	1.76	−4.03	1.46	−4.01	1.39
T1Con2	−3.59	1.39	−2.97	2.06	−3.75	1.78	−3.92	1.41	−3.89	1.37
T1Con3	−2.79	1.05	−2.36	1.35	−2.82	1.24	−2.97	0.92	−2.98	1.00
T2Con1	−3.25	1.17	−2.75	1.57	−3.32	1.52	−3.44	1.18	−3.46	1.17
T2Con2	−3.25	1.21	−2.56	2.76	−3.30	2.58	−3.47	2.17	−3.58	1.26
T3Con1	−3.58	1.37	−2.95	1.86	−3.73	1.61	−3.86	1.23	−3.82	1.35
T4Con1	−1.59	1.28	−0.77	2.03	−1.42	1.84	−1.56	1.58	−1.77	1.32
PaCon1	−1.80	1.46	−0.74	2.27	−1.56	2.34	−1.64	2.15	−2.03	1.52
PaCon2	−2.28	1.40	−1.21	2.01	−1.86	2.07	−1.99	1.85	−2.42	1.45
PaCon3	−1.11	1.41	−0.14	2.26	−0.81	2.27	−0.97	2.02	−1.31	1.63
T5Con1	−0.81	0.79	−0.56	1.38	−0.83	1.49	−0.95	1.22	−0.88	0.76
T5Con2	−0.78	0.97	−0.43	1.51	−0.80	1.45	−0.93	1.09	−0.89	0.95
T6Con1	−0.65	0.73	−0.46	0.87	−0.68	0.93	−0.73	0.63	−0.69	0.70
T6Con2	−0.57	0.72	−0.39	0.80	−0.59	0.83	−0.54	0.52	−0.61	0.71
Coplane	−1.46	0.68	−1.22	1.54	−1.41	1.53	−1.44	1.23	−1.47	0.72

TABLE 3: CP-Corrected HF and MP2−Correlation Interaction Energies ($\Delta E^{(HF)}$ and $\Delta E^{(MP2)}$ kcal/mol) for 15 Calculated Benzene−Formamide Interaction Configurations at Various Basis Sets and at MP2-Optimized Geometries

	6-31+G(2d,p)		6-311G(d,p)		6-311G(2d,p)		6-311G(2d,2p)		6-311+G(2d,p)	
configuration	$\Delta E^{(HF)}$	$\Delta E^{(MP2)}$	$\Delta E^{(HF)}$	$\Delta E^{(MP2)}$	$\Delta E^{(HF)}$	$\Delta E^{(MP2)}$	$\Delta E^{(HF)}$	$\Delta E^{(MP2)}$	$\Delta E^{(HF)}$	$\Delta E^{(MP2)}$
T1Con1	−0.70	−3.17	−0.72	−2.38	−0.95	−2.93	−0.84	−3.19	−0.80	−3.21
T1Con2	−0.61	−2.96	−0.56	−2.41	−0.81	−2.94	−0.73	−3.19	−0.65	−3.24
T1Con3	−0.67	−2.32	−0.69	−1.67	−0.75	−2.07	−0.70	−2.27	−0.65	−2.33
T2Con1	−0.78	−2.47	−0.78	−1.97	−0.88	−2.44	−0.81	−2.63	−0.77	−2.69
T2Con2	−0.90	−2.35	−0.73	−1.83	−0.78	−2.52	−0.77	−2.70	−0.84	−2.74
T3Con1	−0.61	−2.97	−0.55	−2.40	−0.83	−2.90	−0.73	−2.13	−0.61	−3.20
T4Con1	1.25	−2.84	1.64	−2.41	1.27	−2.69	1.11	−2.67	1.23	−3.00
PaCon1	2.27	−4.07	2.35	−3.08	2.37	−3.93	2.36	−4.00	2.20	−4.23
PaCon2	1.19	−3.47	1.88	−3.09	1.33	−3.19	1.36	−3.35	2.85	−4.15
PaCon3	2.90	−3.01	2.97	−3.14	2.96	−3.79	2.96	−3.93	3.02	−4.33
T5Con1	0.77	−1.58	0.76	−1.32	0.77	−1.57	0.76	−1.71	0.78	−1.66
T5Con2	1.25	−2.03	1.24	−1.67	1.18	−1.98	1.18	−2.11	1.24	−2.13
T6Con1	0.59	−1.24	0.60	−1.06	0.54	−1.22	0.57	−1.30	0.61	−1.30
T6Con2	0.62	−1.19	0.60	−0.99	0.56	−1.15	0.58	−1.12	0.63	−1.24
Coplane	−0.62	−0.84	−0.72	−0.50	−0.61	−0.80	−0.56	−0.88	−0.63	−0.84

TABLE 4: CP-Corrected Total Interaction Energies ($\Delta E^{(cp)}$, kcal/mol), HF and MP2−Correlation Interaction Energies ($\Delta E^{(HF)}$ and $\Delta E^{(MP2)}$, kcal/mol) of Several Benzene−Formamide Interaction Configurations at the 6-311G(2d,2p) Basis Set and at HF-Optimization Geometries

configuration	$\Delta E^{(HF)}$	$\Delta E^{(MP2)}$	$\Delta E^{(cp)}$
T1Con1	−2.05	−1.67	−3.72
T1Con2	−1.99	−1.63	−3.62
T1Con3	−1.40	−1.35	−2.75
T2Con1	−1.37	−1.52	−2.89
T2Con2	−1.72	−1.47	−3.19
PaCon1	0.26	−1.80	−1.54
PaCon2	−0.28	−1.48	−1.76

tions with their HF-optimized geometries at the 6-311G(2d,2p) basis set level. The results indicate that, for each of these configurations, the contribution of the favorable HF interaction energy to the total binding energy at the HF-optimized geometry is larger than at the MP2-optimized one, and even outweighs the MP2−correlation interaction energy for the T1Con1, T1Con2, T1Con3, and T2Con2 configurations. On the basis of these results and analyses, we can conclude that dispersion interaction in the benzene−formamide interaction is not as important for the total binding energy as in the benzene−benzene interaction, and the contrary is true for their electrostatic interactions. From the charge distribution point of view, such a conclusion is quite understandable and obvious, because the possible energetically favorable electrostatic interactions in a benzene−formamide interaction contain not only the quadrupole−quadrupole contribution but the stronger dipole−quadrupole interaction, while in the benzene−benzene interaction, the quadrupole−quadrupole interaction is the only electrostatic term.

Therefore, we believe that the calculated interaction energies at the MP2/6-311G(2d,2p) level are reliable for the benzene−formamide interaction; it does not overestimate the binding energy of the benzene−formamide interaction. MP2 with 6-31+G(2d,p), 6-311G(2d,p), or 6-311+G(2d,2p) basis set can also produce reasonable results. However, with basis sets smaller than 6-311G(d,p), the MP2 method underestimates the binding energy.

Strength of the Benzene−Formamide Interaction. Based on the calculated results, the benzene−formamide interaction can achieve a binding energy between 0.5 and 4.0 kcal/mol, which depends heavily on the mutual orientation between benzene and formamide. According to the MP2/6-311G(2d,2p) results, configurations T1Con1, T1Con2, T1Con3, T2Con1, T2Con2, and T3Con1 have binding energies of 3.0−4.0 kcal/mol, which are equivalent to a hydrogen bonding interaction (for instance, at the MP2 level with reliable basis sets, the binding energies for H_2O-H_2O, HF−HF, and NH_3-NH_3 are 4.57, 4.05, and 2.76 kcal/mol, respectively[62,63]); T4Con1, PaCon1, PaCon2, and Coplane correspond to binding energies between 1.4 and 2.0 kcal/mol; and PaCon3, T5Con1, T5Con2, T6Con1, and T6Con2 achieve small binding energies of 0.5−1.0 kcal/mol. At the MP2/6-311G(2d,2p) level, the largest binding energy for the benzene−formamide interaction is achieved in the T1Con1 configuration and is 4.03 kcal/mol. At the same theoretical level,

TABLE 5: Various Interaction Potential Energy Surfaces (IPESs) for the Benzene−Formamide Interaction (Distance, Å; Angle, deg)

IPESs	ΔE(P)	ΔE(L)	ΔE(A2)	ΔE(DA1)	ΔE(DA2)
fixed geometrical parameters	$L = 0.02$, A1 = 0 A2 = 90 DA1 = 180	$P = 4$, A1 = 0 A2 = 90 DA1 = 180	$P = 4$, $L = 0.02$ A1 = 0 DA1 = 180	$P = 4$, $L = 1.4$ A1 = 0 A2 = 90	$P = 4$, $L = 0.02$ A1 = 0 A2 = 90 DA1 = 180
cases	Con1: DA2 = 180 Con2: DA2 = 90 Con3: DA2 = 0	Con1: DA2 = 180 Con2: DA2 = 90 Con3: DA2 = 0	Con1: DA2 = 180 Con2: DA2 = 90 Con3: DA2 = 0	Con1:DA2 = 180 Con2: DA2 = 90 Con3: DA2 = 0	

the largest binding energy of the benzene−benzene interaction is achieved with a displaced face-to-face interaction configuration and is 3.33 kcal/mol.[25] We did not find a configuration which has a larger binding energy than T1Con1. Therefore, we believe that the benzene−formamide interaction is, in general, slightly stronger than the benzene−benzene interaction and that the T1Con1 configuration is the most energetically favorable structure.

(3) Interaction Potential Energy Surfaces (IPES) for the Benzene−Formamide Interaction. The benzene−formamide interaction configuration may be defined by the parameters *P*, *L*, A1, A2, DA1, and DA2, as illustrated in Figure 2. A complete IPES of such an interaction can thus be expressed as a function of these six geometrical parameters, or $\Delta E(P, L,$ A1, A2, DA1, DA2). ΔE is not sensitive to changes in A1; for example, for T1Con1 and PaCon1, when this angle varies from 0 to 30°, the binding energy changes in magnitude by only 0.001 and 0.003 kcal/mol, respectively, at the MP2/6-311G(2d,2p) level. Therefore, A1 was fixed at 0° for calculations of IPESs. The dependence of the interaction energy on the other five parameters was investigated by varying each one individually. The cases investigated are listed in Table 5, and the CP-corrected IPESs at MP2/6-311G(2d,2p) for these cases are shown in Figure 3a−e.

$\Delta E(P)$. In Figure 3a, ΔE is plotted as a function of *P* for cases Con1, Con2, and Con3. It can be seen that the most energetically favorable *P* value between benzene and formamide is between 3.5 and 4.0 Å for Con1, less than 3.5 Å for Con2, and about 4.5 Å for Con3. The interactions for Con2 and Con3 vanish at about 5.5 and 6.0 Å, respectively; for Con1, it vanishes at a further distance of about 8.0 Å. When parameter *P* is approximately 4.0 Å, a benzene−amino π-type hydrogen bond may be achieved for both Con1 and Con3. However, the Con3 case corresponds to either a large repulsive energy or a very small binding energy, while Con1 always corresponds to a significant binding energy. This is due to differences in the location of the carbonyl's oxygen atom relative to the benzene ring in these two cases. In Con3, the oxygen atom is less than 4.0 Å above the benzene ring and there exists a benzene−oxygen repulsion (see below for further analysis).

$\Delta E(L)$. The dependence of ΔE on *L* is reflected in Figure 3b. For all three cases, the interaction energy may be neglected when *L* is larger than 5.0 Å. When *L* is less than 1.5 Å, Con3 has a smaller binding energy than Con1. This occurs because there is a large benzene−oxygen repulsion in Con3. The benzene−oxygen interaction may be negligible in Con1 because the oxygen is located sufficiently distant from the benzene ring. When *L* is larger than 2.0 Å, the benzene−oxygen repulsion in Con3 can also be neglected, such that Con3 and Con1 have similar binding energies.

$\Delta E(A2)$. Dependence of interaction energies on A2 is shown in Figure 3c. The interaction energy of the case Con3 varies sharply with A2. When A2 is less than 85°, it is highly energetically unfavorable due to the benzene−oxygen repulsion. When A2 goes from 120 to 165°, it has as large a binding energy (3.5 kcal/mol) as Con1. This indicates that when the carbonyl oxygen atom is located more than about 4.0 Å above the benzene ring, the effect imparted by the location and directionality of the oxygen atom (or the benzene−oxygen repulsion) almost disappears.

$\Delta E(DA1)$. Figure 3d reflects the change of the interaction energy with respect to DA1. The bindng energies for both Con1 and Con2 are not sensitive to changes in DA1, which suggests that, for these two classes of configurations, the energy barrier is very small for the rotation of formamide with respect to the (X1V) axis (see Figure 2). Due to the influence of the benzene−oxygen interaction, such an energy barrier associated with Con3 is quite large.

$\Delta E(DA2)$. In Figure 3e, interaction energy is expressed as a function DA2. Binding energy increases as DA2 goes from 0 to 180°. The increase in the binding energy is more dramatic from 105 to 180° than from 0 to 105°. This again demonstrates that the position of the carbonyl oxygen atom is an important factor when it is located less than about 4.0 Å above the benzene molecular plane.

Overall, these interaction potential energy surfaces indicate that the benzene−formamide interaction can achieve a significant binding energy over a wide configurational space. The characteristic (attractive or repulsive) and the magnitude of the interaction is sensitive to the relative orientation between benzene and formamide, especially in term of location of the carbonyl oxygen atom relative to the benzene ring.

(4) Nature of the Benzene−Formamide Interaction. As a noncovalent interaction system, the benzene−formamide interaction includes electrostatic, exchange, induction, and dispersion components. The exchange interaction is always repulsive; the induction and the dispersion interactions are always favorable; the electrostatic term depends on the relative orientation between the two monomers, both in its characteristics (attractive or repulsive) and in its magnitude. Relative to the dispersion and electrostatic interactions, induction is of secondary importance.[26a] Moreover, the induction term is generally much less orientation-dependent than the electrostatic term.[26b] To understand the nature of the interaction, the relationship between the dispersion and electrostatic interactions and interaction configurations must be examined.

Dispersion Interaction. The magnitude of this interaction is proportional to the area of overlap between the atoms of two subsystems. In general, the atomic overlap is more efficient in the face-to-face configuration; thus, such structural patterns should correspond to a larger dispersion energy. In addition, the dispersion energy is a principal contribution to the correlation interaction energy.[16] Among the 15 calculated configurations in Figure 1, the favorable CP-corrected MP2−correlation interaction energies are larger for the three face-to-face configurations, PaCon1, PaCon2, and PaCon3, than for the others (refer to Table 3). This is additional evidence for such a conclusion.

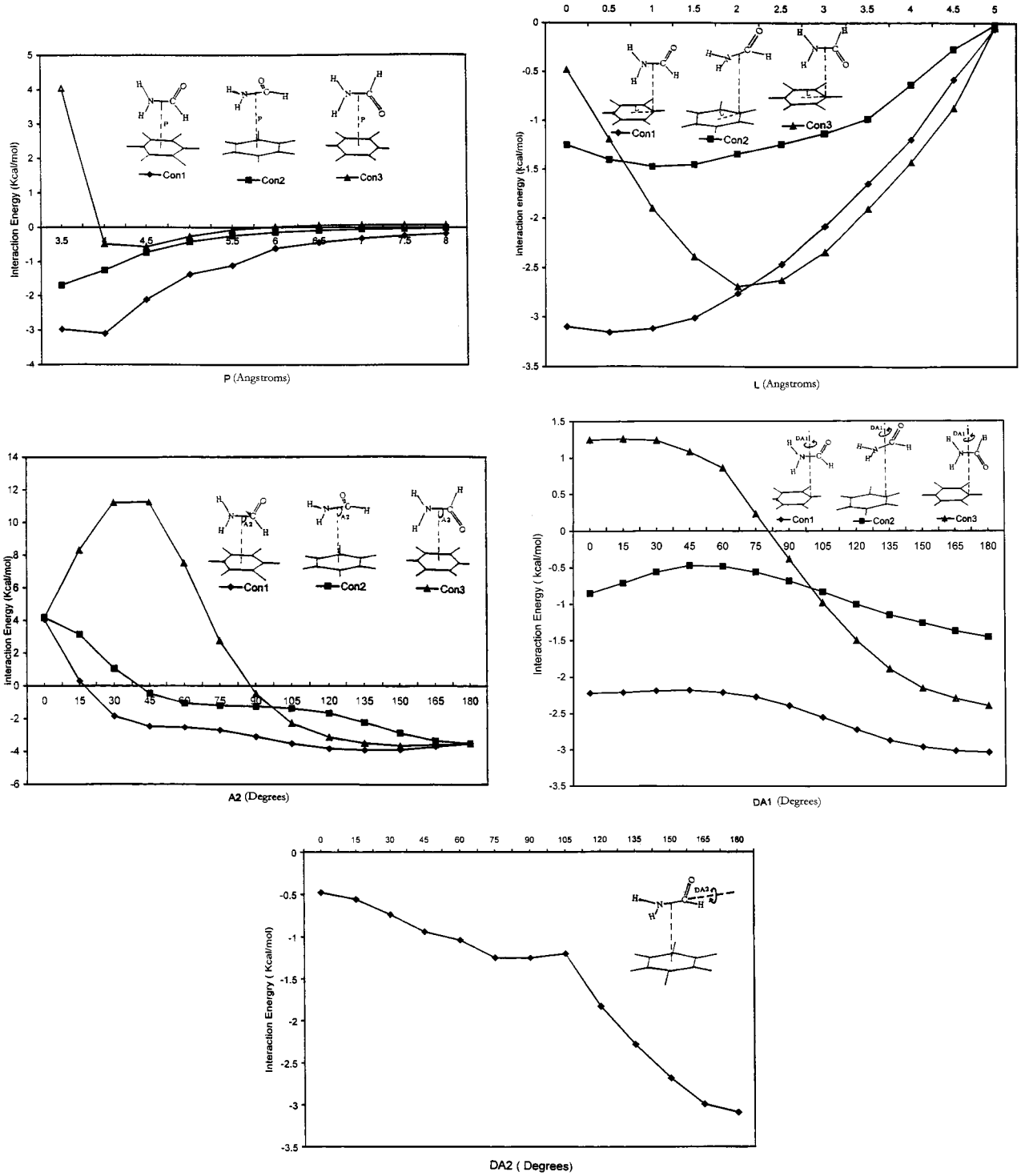

Figure 3. CP-corrected MP2/6-311G(2d,2p) interaction potential energy surfaces (IPESs) of the benzene−formamide interaction (refer to Table 5 for the geometries of cases Con1, Con2, and Con3, and Figure 2 for the parameters *P*, *L*, A2, DA1, and DA2): (a, top left) IPES on *P*; (b, top right) IPES on *L*; (c, middle left) IPES on A2; (d, middle right) IPES on DA1; (e, bottom) IPES on DA2.

Electrostatic Interactions. This interaction is much more complex. Considering that the molecular electrostatic potential (MEP) is a good tool for a qualitative understanding of the features of the electrostatic interaction between two chemical species,[66] we calculated the MEPs for both benzene and formamide on different planes at the MP2/6-311G(2d,2p) level. Parts a−d of Figure 4 illustrate the MEPs of a benzene molecule at its molecular plane and at 1.0, 1.3, and 1.7 Å above the molecular plane. The MEPs on the plane 2.0 Å above the molecular plane are similar to those in the 1.7 Å plane, except being less negative. In the plane 2.5 Å above the molecular plane, the MEPs are slightly negative around the benzene ring. These MEPs reflect the distribution of benzene's π electrons and indicate that the π electrons begin to appear in the plane about 1.3 Å above the molecular plane, concentrating mainly between 1.7 and 2.0 Å above the molecular plane. Parts a and b of Figure 5 are the MEPs of formamide at its molecular plane and 1.7 Å above the molecular plane. The negative MEPs around the oxygen atom on the molecular plane are due to its two sp^2 lone pairs, each oriented 120° from the C═O axis. In the plane 1.7 Å above the molecular plane, the MEP is more extensively negative because of the buildup of the π electrons in this plane.

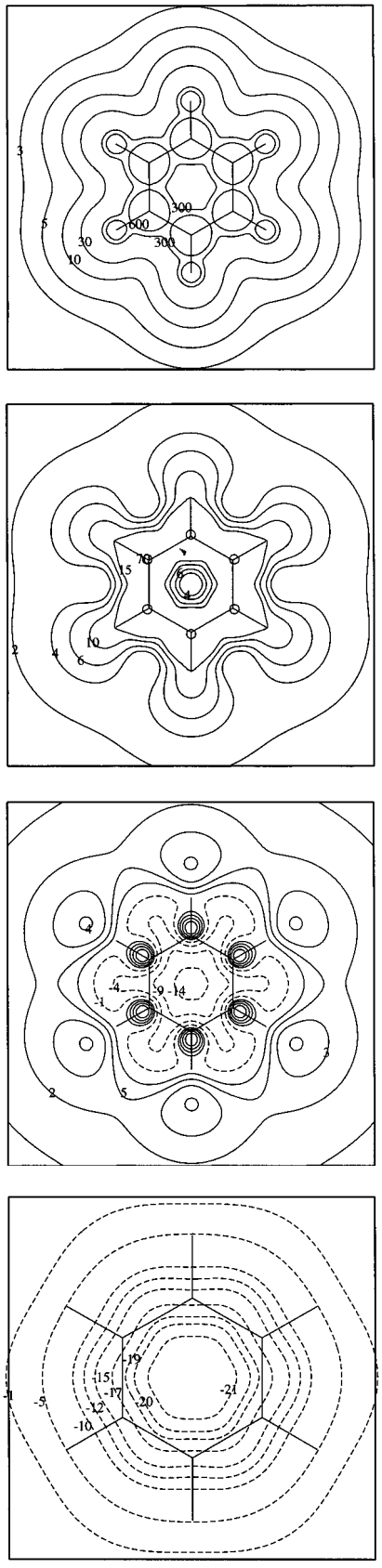

Figure 4. MP2/6-311G(2d,2p) molecular electrostatic potentials (kcal/mol) of benzene: (a, top) in the molecular plane; (b, second from top) in the plane 1.0 Å above the molecular plane; (c, third from top) in the plane 1.3 Å above the molecular plane; (d, bottom) in the plane 1.7 Å above the molecular plane.

However, unlike the benzene π electrons, the formamide's π electrons undergo an attraction from the oxygen atom, such that,

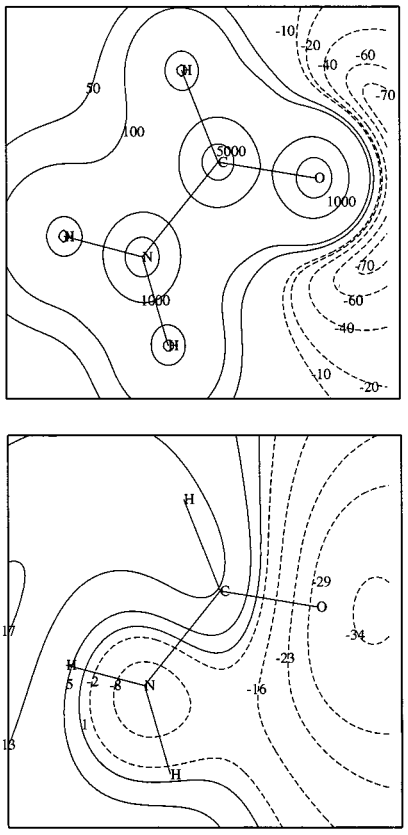

Figure 5. MP2/6-311G(2d,2p) molecular electrostatic potentials (kcal/mol) of formamide: (a, top) in the molecular plane; (b, bottom) in the plane 1.7 Å above the molecular plane.

in the plane in which the π electrons build up, only the MEPs around the O and N are negative, while the MEPs in the regions around both H and C are positive.

In the T1Con1, T1Con2, T1Con3, T2Con1, and T3Con1 configurations, the main electrostatic interactions arise from the attraction between formamide's positive MEPs (both in its molecular plane and in the plane above its molecular plane) and benzene's negative MEPs (above the molecular plane). In nature, they correspond to the attractive dipole(formamide)−quadrupole(benzene) and quadrupole−quadrupole electrostatic interactions, respectively. The dominant electrostatic interaction involved in T2Con2 is the attraction between the negative MEPs around formamide's oxygen atom and the positive MEPs in the benzene molecular plane which can be identified as a dipole-(formamide)−quadrupole(benzene) attraction, or the so-called "benzene−oxygen attraction". It is due to the favorable electrostatic interactions in these configurations that they have attractive HF interaction energies and thus rather large binding energies (refer to Tables 2 and 3). In T4Con1, the main electrostatic interaction is the repulsion between the negative MEPs around formamide's oxygen atom due to its lone pairs and the negative MEPs of benzene due to its π electrons, which can be identified as a dipole(formamide)−quadrupole(benzene) repulsion, or "benzene−oxygen repulsion". Because of this electrostatic repulsion, such a configuration corresponds to a larger unfavorable HF interaction energy and thus a small binding energy. In PaCon1, PaCon2, and PaCon3, the electrostatic interactions are dominated by the repulsion between negative MEPs around both oxygen and nitrogen atoms of formamide in its molecular plane and in the planes above its molecular plane and negative MEPs above benzene's molecular plane, which are quadrupole−quadrupole repulsions.

For a more accurate discussion on the structure−electrostatic interaction relationship, the influence of induction on the molecular electrostatic potentials should be considered. However, since benzene is a weakly polar molecule (quadrupole moment), its influences on the formamide electrostatic potential should be limited; benzene is not easily polarized and its out-of-plane polarizability is smaller than its in-plane polarizability.[32] With an intermediate intermolecular separation, the induction effect would not dramatically modify the overall profile of the electrostatic potentials near two molecules. Therefore, the above analyses provide a qualitatively reasonable insight into the orientation−electrostatic interaction relationship for the benzene−formamide complex.

Locations of Carbonyl Oxygen Atoms Relative to the Benzene Ring. Due to the benzene−oxygen repulsion or attraction, the benzene−formamide interaction is sensitive to the location of the carbonyl oxygen atom. As discussed above, IPESs reflect the importance of the location of the carbonyl oxygen atom for the characteristics (attractive or repulsive) and strength of the benzene−formamide interaction. The results for the 15 calculated interaction configurations indicate the same conclusion. In T1Con1, T1Con2, T1Con3, T2Con1, and T3Con1, the oxygen atoms of formamide are located more than 4.0 Å above benzene's molecular plane. In these interactions, the lone electron pairs on the carbonyl oxygen are far enough from the π electron of benzene and thus the benzene−oxygen repulsion can be neglected. In T4Con1, PaCon1, PaCon2, and PaCon3, the formamide oxygen atom is located approximately 3.5−4.0 Å above the benzene molecular plane such that the benzene−oxygen interaction is repulsive. In T2Con2, the formamide oxygen atom is located about 1.3 Å above the benzene molecular plane and there exists a favorable benzene−oxygen attraction. In the structure Coplane, a favorable electrostatic results from the C−H···O-type hydrogen bond (the MP2-optimized geometry for this hydrogen bond shows that a benzene H atom points directly toward one of the two sp^2 electron pairs of on foramide oxygen atom—the most favorable geometry for a hydrogen bond interaction).

Comprehensively, we can reach the following conclusions regarding the impact of oxygen's location on the benzene−formamide interaction: (i) With the oxygen atom located in the regions greater than 4.0 Å above the benzene plane, the benzene−oxygen repulsion may be neglected. (ii) With the oxygen located in the region 1.7−4.0 Å above the benzene plane, there exists a significant benzene−oxygen repulsion. (iii) With the oxygen located in regions 0−1.7 Å above the benzene plane, there is a benzene−oxygen attraction. These results are in agreement with those of the protein database searches performed by Singh and Thornton[30] and Thomas et al.,[67] namely, that the oxygen atoms prefer to pack near the benzene molecular plane with the angle OX2C* ranging from 0 to 20° (refer to Figure 2 for the points O, X2 ,and C*).

How Important Is the Benzene−Amine Component? On the basis of our results, the benzene−amine interaction is by no means dominant in the benzene−formamide interaction, although it does contribute partially to the total binding energy. The benzene−amine interaction is not the only contributor to the binding energy for the benzene−formamide interaction. For example, in configurations T2Con2 and T2Con1, the amino hydrogen atoms point far away from the benzene ring and it is impossible to achieve a strong benzene−amino interaction (a π-type hydrogen bond cannot be achieved); nevertheless these configurations have favorable HF interaction energies and quite large total binding energies. Face-to-face configurations were thought to be the structural pattern in which benzene−amino hydrogen bonds do not exist.[30,31] However, both PaCon1 and PaCon2 have binding energies larger than 2.0 kcal/mol, which is equivalent to a benzen−water π-type hydrogen bond (up to 2.0 kcal/mol [6]), and PaCon3 has a binding energy of 1.5 kcal/mol, which is comparable to the magnitude of a π-type benzene−amine-type hydrogen bond (up to 1.5 kcal/mol[9,32,34]). The T1Con1, T1Con2, and T3Con1 configurations can achieve a benzene−amine π-type hydrogen bond, but they have much larger binding energies than would be predicted from an isolated π-type hydrogen bonding interaction. Furthermore, the characteristics and strength of the benzene−formamide interaction are more sensitive to the benzene−oxygen interaction than to the benzene−amine interaction. The most important consideration for achieving a significant benzene−formamide binding interaction is to avoid a benzene−oxygen repulsion rather than to obtain a favorable benzene−amine attraction. These features distinguish the aromatic−amide(S) interaction from the aromatic−amine one and other noncovalent binding forces.

(B) Aromatic−Amide(S) Interactions in Proteins. In total, 498 phenyl(F)−amide(S) interactions were identified in the 1029 protein structures analyzed. Among them, 216 such interactions are located within the face-to-face configurational pattern with $3.0 \le P \le 5.0$ Å, $0 \le L \le 3.0$ Å, $0° \le$ A1 $\le 180°$, $60° \le$ A2 $\le 120°$, $0° \le$ DA1 $\le 180°$, and $60° \le$ DA2 $\le 120°$(see Figure 2 for the geometrical parameters). Based on the evaluation of the benzene−formamide interaction, each of these face-to-face interactions should have a binding energy between 1 and 2.5 kcal/mol. The face-to-face phenyl(F)−amide(S) configuration is a significantly energetically favorable structure for a protein. In addition to the attractive interaction between phenyl(F) and amide(S) groups themselves, both the hydrogen and oxygen atoms of the amide(S) are free to establish the normal hydrogen bonding interactions. Furthermore, the face-to-face structure is likely to be efficient from a packing point of view and thus benefits the formation of a overall compact protein structure. These face-to-face interactions constitute 43.4% of 498 phenyl(F)−amide(S) interactions and are the preferred interaction patterns. There are 103 interactions located within the configurational pattern with $4.0 \le P \le 6.0$ Å, $0 \le L \le 3.0$ Å, $0° \le$ A1 $\le 180°$, $60° \le$ A2 $\le 180°$, $0° \le$ DA1 $\le 180°$, $120° \le$ DA2 $\le 180°$. These 103 interactions are similar to T1Con1, T1Con2, T1Con3, T2Con1, and T3Con1 and can be designated as T-shaped structures. In these interactions, the amide oxygen points away from the phenyl ring and is located more than 4.0 Å above the phenyl molecular plane. Based on the evaluation of the benzene−formamide interaction, each of these should achieve a binding energy of 3.0−4.0 kcal/mol. The remaining 179 phenyl(F)−amide(S) contacts are scattered in other configurational patterns. These close contacts should also correspond to energetically favorable interactions. Taking into consideration the aromatic groups of both the tyrosine and tryptophan residues, the number of such attractive aromatic−amide(S) interactions would increase. As a result, there should, on a statistical average, be more than one significantly attractive aromatic−amide(S) interaction per two protein structures (498 phenyl(F)−amide(S) in the 1029 protein structures). These 1029 protein structures contain 9463 phenyl(F) groups. This means that most of the aromatic groups do not have an amide(S) moiety as their nearest neighbor. This occurs because the phenyl(F) is more likely to be found in the interior of a protein,[32] while the more strongly polar amide(S) group should predominantly occur on the surface of a protein. However, this interaction can achieve a binding energy up to 4.0 kcal/mol. It is generally stronger

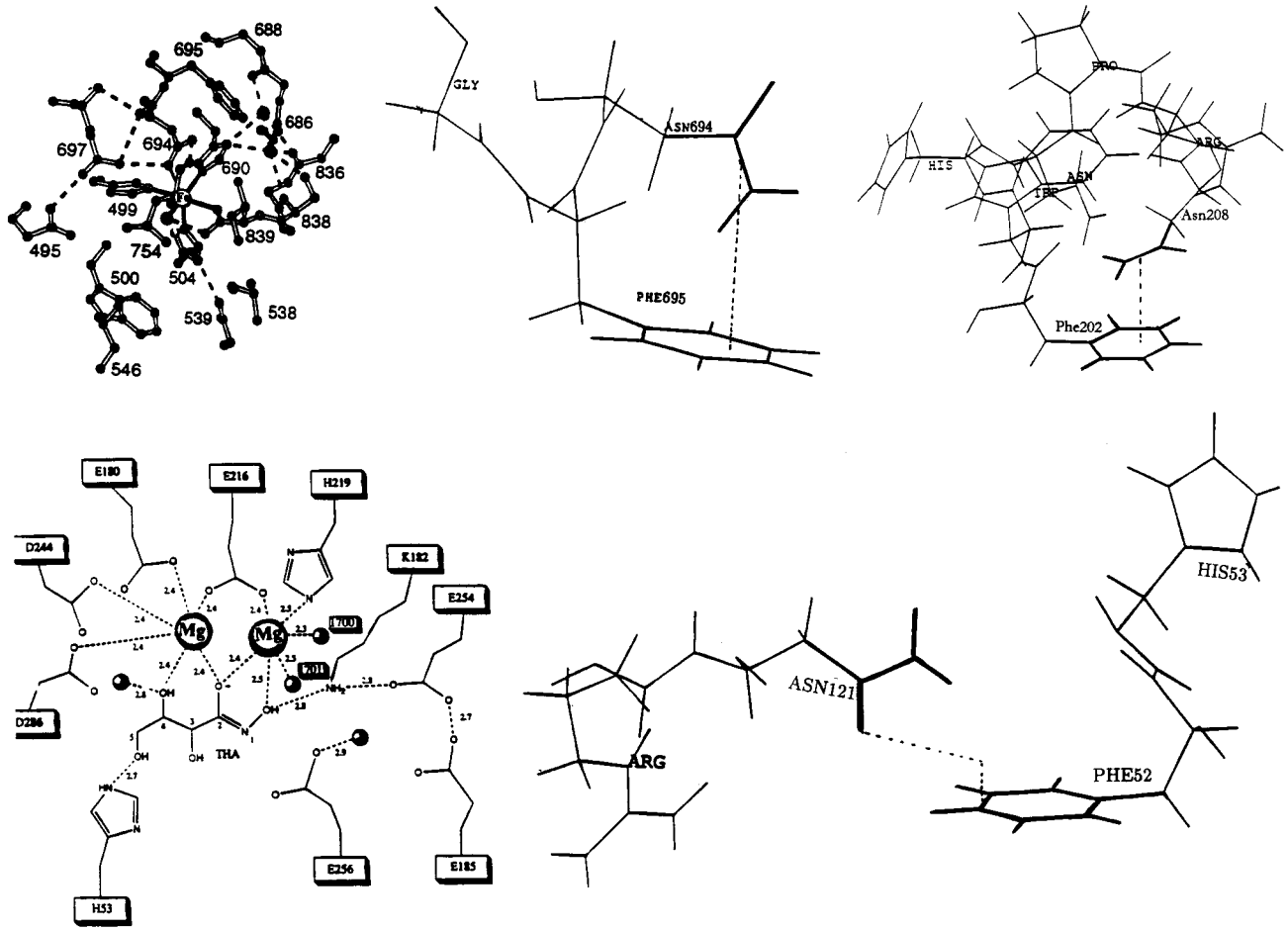

Figure 6. Representative aromatic−amide(S) interactions in three proteins. The interacting aromatic and groups(S) are illustrated with heavy lines. (a, top left and middle) Soybean lipoxygenase L-1: active site (left, adapted from Minor et al.[66]) and phenyl(Phe695)−amide(Asn694) interaction (right, $P = 3.6$ Å, $L = 0.3$ Å, A2 = 146.7°, DA1 = 106.0°, DA2 = 163.8°) (refer to Figure 2 for the geometric parameters P, L, A2, DA1, and DA2); (b, top right) 14.3.d T-cell antigen receptor: phenyl(Phe202)−amide(Asn208) interaction ($P = 3.2$ Å, $L = 0.67$ Å, A2 = 105.2°, DA1 = −167.3°, DA2 = −86.3°); (c, bottom) xylose isomerase−THA: active site (left, adapted from Allen et al.[68]) and phenyl(Phe52)−amide(Asn121) interaction (right, $P = 2.4$ Å, $L = 4.2$ Å, A2 = 62.6°, DA1 = −159.1°, DA2 = −1.9°).

than the well-documented aromatic-containing interactions in proteins such as aromatic−aromatic (up to 3.33 kcal/mol [25]), aromatic−amino (up to 1.5 kcal/mol [9,32,34]), and aromatic−water(hydroxyl) (up to 2.0 kcal/mol [6]). We believe that this interaction is significant for some protein structures due to its strength and natural occurrence. However, before determining the contribution of these effects to overall protein stability, environmental effects, such as hydration and protein packing, must be explicitly considered.

Parts a−c of Figure 6 demonstrate aromatic−amide(S) interactions in three representative proteins. Minor et al.[68] determined the crystal structure of soybean lipoxygenase L-1 at 1.4 Å resolution in 1996. The active site region is illustrated in the left part of Figure 6. The active site Fe has five ligands at distances less than 2.6 Å: the side chains of His499, His504, and His690, one oxygen atom from the carboxylate group of the terminal residues(Ile839), and one water molecule. The side-chain amide oxygen atom of Asn694, which is 3.05 Å from the Fe atom, is positioned in a proper direction to serve as the sixth ligand. Based on its position relative to the Fe, the side-chain amide oxygen atom of Asn694 is only weakly bonded to the Fe atom but enables an approximately octahedral coordination geometry. On the basis of the structure determined, we find that there is a significant interaction between the phenyl ring of the Phe695 residue and the side-chain amide of the Asn694 (refer to the right part of Figure 6a). This interaction has a configuration similar to T3Con1 (refer to Figure 1). According to the geometry, it should have an attractive interaction of 3.5−4.0 kcal/mol in the gas phase, which is significantly large. Obviously, this interaction plays an important role in fixing the orientation of the side-chain amide of Asn694 and thus its oxygen's location relative to the Fe atom in the active site.

The β-chain of the T-cell antigen receptor plays a key role in regulating early differentiation events in the thymus and in the binding complexes of MHC class II molecules. Figure 6b demonstrates a face-to-face phenyl−amide(S) interaction between Phe202 (in helix200−203) and Asn208 in the β-chain of the 14.3.d T-cell antigen receptor as determined by Boulot et al.[69] at a resolution of 1.7 Å. This face-to-face interaction is of a configuration similar to PaCon2 and thus should have an attractive interaction of 2.0−2.5 kcal/mol in the gas phase.

Allen et al.[70] determined the structure of the xylose isomerase−THA complex to a resolution of 1.6 Å. The xylose isomerase−THA interactions are illustrated in the left part of Figure 6c. The THA is bound to the xylose isomerase mainly by the metal−carboxylate ligands. The only non-covalent interactions between the active-site residues and THA are two hydrogen bonds, one between the C5 hydroxyl of the THA and His53 and the other between the N1 hydroxyl of the THA and Lys182. We find that there exists a phenyl−amide(S) interaction between Phe52 and Asn121. Phe52 is neighboring to His53. This interaction is illustrated in the right of Figure 6c, which assumes

a configuration similar to T2Con2 (refer to Figure 1). On the basis of the geometry, we estimate that this interaction corresponds to a binding energy of 3.0−3.5 kcal/mol in the gas phase. This interaction influences the conformation of Phe52 and thus position and orientation of the His53.

IV. Conclusions

In this study, the benzene−formamide complex which models the aromatic−amide(S) interaction has been exhaustively and systematically investigated at the MP2 level of theory. The applicability of different basis sets was first tested using specific interaction configurations. The results show that 6-31+G(2d,p), 6-311G(2d,p), 6-311G(2d,2p), and 6-311+G(2d,p) are reasonable basis sets for describing such interactions. Larger basis sets are not necessary, and basis sets smaller than 6-311G(d,p) would underestimate binding energies. The IPESs of the interaction were calculated at the MP/6-311G(2d,2p) level. These calculations of the benzene−formamide model complex demonstrate the unique characteristics of the aromatic−amide(S) interaction. The aromatic−amide(S) interaction involves the entirety of the amide(S) group rather than only its amine portion, the aromatic−amine interaction is not the only source for the binding energy, and the carbonyl portion has greater influence on both its characteristics (attractive or repulsive) and its strength. The binding energy arises mainly from dispersion and electrostatic interactions. The latter contributor originates from the attractive interaction between the quadrupole moment of the aromatic and quadrupole and dipole moments of the amide(S). The interaction can achieve a binding energy up to 4.0 kcal/mol in the gas phase, which indicates that it is, in general, slightly stronger than the aromatic−aromatic interaction (up to 3.33 kcal/mol in the gas phase at the same level of theory[25]) and much stronger than the aromatic−amine interaction (up to 1.5 kcal/mol in the gas phase[9,32,34]). The interaction can achieve a significant binding energy, such as larger than 1.5 kcal/mol, over a wide configurational space. In contrast, the aromatic−amine interaction corresponds to such a binding energy within a very narrow configurational space—only the optimal π-type structure can achieve a binding energy of up to 1.5 kcal/mol.[9,32,34] These features distinguish the aromatic−amide(S) interaction from other non-covalent interactions, such as the aromatic−amine one, and are enough to make it a unique type of biological binding force.

The significance of the aromatic−amide(S) for the protein structure was analyzed through data mining of the1029 X-ray protein structures. A set of 498 phenyl(F)−amide(S) interactions was identified in these protein structures. Among them, 103 interactions have a binding energy of 3.0−4.0 kcal/mol and 216 interactions have a binding energy between 1.0 and 2.5 kcal/mol. When taking into account the aromatic rings from tyrosine and tryptophan, there should be more such energetically significant aromatic−amide(S) interactions in proteins. The frequency of its occurrence should, on statistical average, be more than one per two proteins. This study shows that such a relatively high frequency of occurrence is partly due to the characteristic of the interaction that the significant binding energy can be achieved over a wide configurational space. We believe that this interaction is of significance to some protein structures due to both its strength and its frequent occurrence. It is also demonstrated that such interactions can play a role in influencing the biological activity of some proteins.

Acknowledgment. V.H.S. and D.F.W. acknowledge grants-in-aid from the Natural Science and Engineering Research Council (NSERC) of Canada.

References and Notes

(1) Hobza, P.; Selzle, H. L.; Schlag, E. W. *Chem. Rev.* **1994**, *94*, 1767.
(2) Walter, E. A.; Grover, J. R.; White, M. G.; Hui, E. T. *J. Phys. Chem.* **1985**, *89*, 3814.
(3) Nowak, R.; Menapace, J. A.; Bernstein, E. R. *J. Chem. Phys.* **1988**, *89*, 1309.
(4) Nagy, P. I.; Ulmer, C. W.; Smith, D. A. *J. Chem. Phys.* **1995**, *102*, 6813.
(5) Cheng, B. M.; Grover, J. R.; Waters, E. A. *Chem. Phys. Lett.* **1995**, *232*, 364.
(6) Suzuki, S.; Green, P. G.; Bumgarner, R. E.; Dasgupta, S.; Goddard, W. A.; Blake, G. A. *Science* **1992**, *257*, 942.
(7) Atwood, J. L.; Hamada, F.; Robinson, K. D.; Orr, G. W.; Vincent, R. L. *Nature* **1991**, *349*, 683.
(8) Wasksman, G.; Kominos, D.; Robertson, S. C.; Pant, N.; Baltimore, D.; Birge, R. B.; Cowburn, H.; Hanafusa, H.; Mayer, B. J.; Overduin, M.; Resh, M. D.; Rios, C. B.; Silverman. L.; Kuriyan, J. *Nature* **1992**, *358*, 646.
(9) Rodham, D. A.; Susuki, S.; Suenram, R. D.; Lovas, F. J.; Dasgupta, S.; Goddard, W. A.; Blake, G. A. *Nature* **1993**, *362*, 735.
(10) Henson, B. F.; Hartland, C. V.; Venturo, V. A.; Felker, P. M. J. *J. Chem. Phys.* **1992**, *97*, 2189.
(11) Arunan, E.; Gutowsky, H. S. J. *J. Chem. Phys.* **1993**, *98*, 4294.
(12) Ebata, T.; Hamakado, M.; Moriyama, S.; Morioka, Y.; Ito, M. *Chem. Phys. Lett.* **1992**, *199*, 33.
(13) Karlstrom, G.; Linse, P.; Wallqvist, A.; Jonssen, B. *J. Am. Chem. Soc.* **1983**, *105*, 377.
(14) Carsky, P.; Selzle, H. L.; Schlag, E. W. *Chem. Phys.* **1988**, *125*, 165.
(15) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1990**, *93*, 5893.
(16) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1993**, *97*, 3937.
(17) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500.
(18) Sun, S.; Berstein, B. R. *J. Phys. Chem.* **1996**, *100*, 13384.
(19) Tsuzuki, S.; Uchimaru, T.; Tanabe, K. *J. Mol. Struct. (THEOCHEM)* **1994**, *307*, 107.
(20) Garrett, A. W.; Zwier, T. S. *J. Chem. Phys.* **1992**, *96*, 3402.
(21) Venturo, V. A.; Felker, P. M. *J. Chem. Phys.* **1993**, *99*, 748.
(22) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 11217.
(23) Felker, P. M.; Maxton, P. M.; Schaeffer, M. W. *Chem. Rev.* **1994**, *94*, 1787.
(24) Law, K. S.; Schauer. M.; Bernstein, E. R. *J. Chem. Phys.* **1984**, *81*, 4871.
(25) Jaffe, R.; Smith, G. D. *J. Chem. Phys.* **1996**, *105*, 1780.
(26) (a) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525. (b) Stone, A. J. *Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 1996; Chapter 7.
(27) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23.
(28) Deakyne, C. A.; Meot-Ner, M. *J. Am. Chem. Soc.* **1985**, *107*, 474.
(29) Sunner, J.; Nishizawa, K.; Kebarle, P. *J. Phys. Chem.* **1981**, *85*, 1814.
(30) Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1990**, *211*, 595.
(31) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *239*, 315.
(32) Dougherty, D. A. *Science* **1996**, *271*, 163.
(33) Kumpfand, R. A.; Dougherty, D. A. *Science* **1993**, *261*, 1708.
(34) Admas, H.; Harris, K. D. M.; Hembury, G. A.; Hunler, C. A.; Livingston, D.; McCabe, J. F. *Chem. Commun.* **1996**, 2531.
(35) Duan, G.; Smith, V. H., Jr.; Weaver, D. F. *Chem. Phys. Lett.* **1999**, *310*, 323.
(36) Huiszoon, C.; Mulder, F. *Mol. Phys.* **1979**, *38*, 1497.
(37) Hirota, E.; Sugisaki, R.; Vielsen, C. J.; Sorensen, O. *J. Mol. Spectrosc.* **1974**, *49*, 251.
(38) Brown, R. D.; Godfrey, P. D.; Kleibomer, B. K. *J. Mol. Spectrosc.* **1987**, *124*, 34.
(39) Chalasinski, G.; Szczesniak, M. M. *Mol. Phys.* **1988**, *63*, 205.
(40) Chalasinski, G.; Szczesniak, M. M.; Cybulski, S. M. *J. Chem. Phys.* **1990**, *92*, 2481.
(41) Sadlej, J.; Buch, V. *J. Chem. Phys.* **1994**, *100*, 4272.
(42) Gutowski, M.; Chalasinski, G. J. *J. Chem. Phys.* **1993**, *98*, 5540.
(43) Liu, B.; Mclean, A. D. *J. Chem. Phys.* **1989**, *91*, 2348.
(44) Yang, J.; Kestner, N. R. *J. Phys. Chem.* **1991**, *95*, 9241.
(45) Eggenberger, R.; Gerber, R.; Huber. S.; Searls, D. *Chem. Phys. Lett.* **1991**, *183*, 223.
(46) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(47) Frisch, M. J.; Trucks, G. W.; et al. *Gaussian 94*, Gaussian Inc.: Pittsburgh, PA, 1994.

(48) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Mayer, G. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasami, M. *J. Mol. Biol.* **1977**, *112*, 535.
(49) (a) Duan, G. Ph.D. Thesis, Department of Chemistry, Queen's University, 2000. (b) Wall, L.; Christiansen, T.; Schwartz, R. L. *Programming Perl* 2nd ed.; O'Reilly & Associates, Inc.: Cambridge, U.K., 1996.
(50) Hobza, P.; Heinrich, L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790.
(51) Mayer, I.; Vibok, A. *Int. J. Quantum Chem.* **1991**, *40*, 139.
(52) Sadlej, A. *Chem. Phys.* **1991**, *95*, 6705.
(53) Cook, D. B.; Sordo, T. L.; Sordo, J. A. *J. Chem. Soc., Chem. Commun.* **1990**, 2185.
(54) Karlstrom. G.; Sadlej, A. J. *Theor. Chim. Acta* **1982**, *61*, 1.
(55) Novoa, J. J.; Planas, M.; Rovira, M. C. *Chem. Phys. Lett.* **1996**, *251*, 33.
(56) Feyereisen, M. W.; Dixon, D. A. *J. Phys. Chem.* **1996**, *100*, 2993.
(57) Kang, H. C. *Chem. Phys. Lett.* **1996**, *254*, 135.
(58) Cybulski, S. M.; Chalsinski, G. *Chem. Phys. Lett.* **1992**, *197*, 591.
(59) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723.
(60) Sponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 163.
(61) Hobza, P.; Sponer, J. *J. Mol. Struct. (THEOCHEM)*, **1996**, *388*, 115.
(62) Novoa, J. J.; Sosa, C. *J. Phys. Chem.* **1995**, *99*, 15837.
(63) Latajka, Z.; Bouteiller, Y. *J. Chem. Phys.* **1994**, *101*, 9793.
(64) Ruiz, E.; Salahub, D. R.; Vela, A. *J. Phys. Chem.* **1996**, *100*, 12265.
(65) Hobza, P.; Bludsky, O.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1993**, *98*, 6223.
(66) Politzer, P.; Daiker, V. C. Molecular Electrostatic Potential and Chemical Reactivity. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: Cambridge, U.K., 1995.
(67) Thomas, K. A.; Smith, G. M.; Thomas, T. M.; Feldman, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4843.
(68) Minor, W.; Steczko, J.; Stec, B.; Otwinowski, Z.; Bolin, J. T.; Walter, R.; Axelrod, B. *Biochemistry* **1996**, *35*, 10687.
(69) Boulot, G.; Bentley, G. A.; Karjalainen, K.; Mariuzza, R. A. *J. Mol. Biol.* **1994**, *235*, 795.
(70) Allen, K. N.; Lavie, A.; Petsko, G. A.; Ringe, D. *Biochemistry*, **1995**, *34*, 3742.