Hydrogen Bonding between Amino Acid Backbone and Side Chain Analogues: A High-Level ab Initio Study

Kyungsun Kim and Richard A. Friesner*

Contribution from the Department of Chemistry, Columbia University, New York, New York 10027

Received June 4, 1997. Revised Manuscript Received October 9, 1997[⊗]

Abstract: We have carried out calculations of hydrogen-bonding structures and energies, using the pseudospectral local MP2 methodology and a high-quality triple ζ basis set for a large set of amino acid side chain analogues. Both neutral and charged amino acid analogues are examined, and interactions between donors and the π electron acceptor moiety in a benzene ring are considered. A total of >140 structures have been studied, representing all possible hydrogen-bonding interactions between a set of 11 amino acid side chain analogues. The effects of electron correlation, basis set size, and basis set superposition error are analyzed in detail for this data set. A particular focus of the paper in terms of chemically interesting effects is the influence of resonance interactions upon hydrogen-bonding strength, which is elucidated quantitatively for a significant number of donor−acceptor pairs. Finally, it is observed that donors and acceptors fall into "strong" and "weak" categories, with the weak species severely damping the variation in hydrogen-bonding strength as the partner donor or acceptor is varied.

I. Introduction

Hydrogen bonding plays a crucial role in the determination of protein structure and is equally central in many aspects of biological function. For this reason, a crucial objective in computational modeling of biological systems is an accurate description of hydrogen bonding. An enormous variety of hydrogen bonds, both between various side chain functional groups and involving the backbone peptide group, are possible. Verifying that the theoretical representation of all these interactions is reliable, and achieving a physical understanding of what controls hydrogen-bonding strengths, is a formidable task.

There are two principal approaches to modeling hydrogen bonding in proteins. The first is to employ molecular mechanics force fields, such as CHARMm,[1−3] AMBER,[4,5] or OPLS.[6] In general, hydrogen-bonding interactions in force fields are primarily represented by electrostatic interactions, based upon the use of partial atomic charges obtained either via low level quantum chemistry (e.g. 6-31G* Hartree−Fock calculations, as is used in the parameterization of AMBER) or adjusted to fit liquid state data (as in OPLS). Modification of this interaction is made via van der Waals terms which are empirically scaled to reproduce a subset of experimental and/or quantum chemical data.

Recently, we have carried out an extensive investigation of the accuracy of solvation free energy calculations.[7] While this work principally examined dielectric continuum models, we discovered serious problems with the representation of hydrogen-bonding energies via electrostatic models which affect explicit solvent simulations as well. For example, for a series of methylated amines hydrogen bonding to a water molecule, there is a poor correlation between classical electrostatic interaction energies and quantum chemical hydrogen-bonding calculations. This poor correlation is directly reflected in large errors in the prediction of the solvation free energy of the amines in water; discrepancies as large as 2−3 kcal/mol are observed between theory and experiment for several members of the series. Similar errors (although smaller in magnitude) were identified for a substantial number of additional functional groups, including ethers, formates, carboxylic acids, phenols, and thiols. The conclusion of this work is that the representation of hydrogen bonding in current molecular modeling force fields through electrostatic models, while often qualitatively reasonable, has not yet achieved quantitative accuracy. In several force fields, the quantum chemical hydrogen-bonding energies are "scaled" to reproduce "condensed phase effects". We have argued elsewhere[8] that this procedure, while having some justification, is likely to be unreliable, and that a rigorous treatment must involve the use of a polarizable force field.

An alternative approach is to calculate hydrogen-bonding energies directly from ab initio quantum chemistry for molecular pairs. Such calculations have been used as an adjunct to parameterization of existing force fields (see, for example, refs 3, 6, 9, and 28). However, the majority of such studies to date have been limited in several important ways: (1) Typically, a low level of quantum chemistry (e.g. Hartree−Fock theory with a small basis set) is employed. While claims have been made in the literature concerning cancellation of error for such approaches, there is no rigorous validation of these claims on a large database of donor−acceptor pairs. (2) The great majority of calculations have involved interaction of a water molecule with various functional groups (see, for example, ref 26). While this is undoubtedly an important interaction, obtaining good results for a water hydrogen bond does not guarantee that other hydrogen bonds will be accurately represented as well. This is particularly worrisome given the results described in ref 7, which indicates that different donors and acceptors can behave quite

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1997.

(1) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, J.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(2) Momany, F. M.; Rone, R. J. *J. Comput. Chem.* **1992**, *13*, 888.
(3) Mackerell, A. D., Jr.; Wiórkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.
(4) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A.; *J. Comput. Chem.* **1986**, *7*, 230.
(5) McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*, 7743.
(6) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(7) 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.
(8) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.
(9) Pavelites, J. J.; Gao, J.; Bash, P. A.; Mackerell, A. D., Jr. *J. Comput. Chem.* **1997**, *18*, 221.

S0002-7863(97)01836-2 CCC: $14.00

differently with regard to the correlation of electrostatics with hydrogen-bonding energetics. (3) The evaluation of a "hydrogen-bonding strength" for multifunctional chemical groups such as the peptide moiety is very complicated because in many cases it is impossible to isolate a single donor−acceptor interaction. There are often several different hydrogen-bonding geometries even for small molecules containing such functional groups, and the interpretation of the binding energy must take into account multiple functional group interactions.

All of these considerations suggest that it would be very useful to assemble a large database of high-quality ab initio calculation of hydrogen-bounded structures and energies of protein backbone and side chain functional groups. As little as 5 years ago, this would have presented a formidable computational challenge. With the development of new quantum chemical methods and enhancements in the performance of computational hardware, however, such an undertaking is quite feasible. By utilizing the local MP2 (LMP2)[17,29] approach and pseudospectral numerical methods,[10−15] correlated results in which basis set superposition error (BSSE) is for the most part eliminated[17] can readily be obtained for a relevant series of donor−acceptor pairs while employing large basis sets. The ultimate goal is to develop a complete database covering functional groups for all 20 naturally occurring amino acids and considering all possible hydrogen-bonding structures.

In the present paper, we take an initial step in this direction. We have chosen to study small molecules containing the carbonyl, amine, carboxylic acid, alcohol, thiol, hydrocarbon, peptide, and benzyl functional groups. In addition, we examine one positively charged group, the ammonium cation, and one negatively charged group, an ionized carboxylic acid moiety. All possible hydrogen-bonded structures that can be formed between these groups are examined below. This represents a complete coverage of the following 12 amino acids: methionine, serine, threonine, asparagine, glutamine, cysteine, aspartic acid, glutamic acid, phenylalanine, lysine, alanine, and glycine. The remaining amino acids have either larger hydrocarbon side chains or larger ring structures. These are perfectly feasible to treat by the methods used here, requiring some additional computation time in the latter case.

Our hope is that the data provided in this paper will serve two purposes. First, it gives experimental chemists working on protein structure or protein−ligand interactions a convenient way to assess the strength of any given hydrogen-bonding interaction in their particular problem; the rapid estimation of relative hydrogen-bonding strengths of different functional groups may also be important for ligand design. Similarly, the structural data will allow an assessment of the strain imposed in hydrogen-bonded structures in the protein as opposed to the gas phase dimers studied here. Secondly, we provide to force field developers a database of structures and energies which can be used to evaluate the accuracy of their intermolecular potential functions.

This paper is organized as follows. In section II, we briefly discuss the quantum chemical methodology and summarize the list of donors and acceptors to be studied. Section III presents the structural and energetic results in a convenient tabular form, and discusses general trends that emerge from an analysis of the data. Section IV, the conclusion, presents future directions of this reserach.

II. Methods

A. Ab Initio Quantum Chemical Calculations. We use the PS-GVB suite of ab initio electronic structure programs[15] for all calculations reported here. The level of theory that we have chosen to employ for our initial studies is (a) geometry optimization at the HF/6-31G** [19,20] level; (b) single-point binding energy calculations at the LMP2/cc-pVTZ(-f/d)[16] level (the notation -f/d indicates that we have deleted the f functions on heavy atoms and d functions on hydrogen). For the latter, we correct the binding energies for basis set superposition error (BSSE) in the Hartree−Fock energy only using the counterpoise correction,[18] the use of local MP2 having to a great extent eliminated BSSE in the correlation energy.

Extensive testing of our PS-LMP2 code, and of the protocol described above, has been carried out over the past several years. For example, Murphy et al.[29] studied a suite of 36 small-molecule conformational energy differences for which experimental data is available, assembled by Halgren and co-workers.[23] Geometry optimizations were carried out at the 6-31G**/MP2 and 6-31G**/HF levels, followed in each case by a single point cc-pVTZ(-f)/LMP2 evaluation of the energy. Excellent agreement with experiment was obtained (average deviation 0.35 kcal/mol), superior to DFT results and to conventional MP2 results with a smaller basis set. The substitution of HF for MP2 geometries produced errors no larger than a few tenths of a kcal/mol in all cases.

This level of theory is not a "benchmark" level: there are possible errors in the extent of electron correlation, as compared to, for example, QCISD(T) or MRPT methods, basis set (several workers have found that diffuse functions and f functions can make measurable contributions to the binding energy[21,27]), and method of geometry optimization. However, our objective here is not to produce results accurate to 0.1 kcal/mol for a few structures but rather to survey a wide variety of structures at a level of theory that represents a significant improvement as compared to the overwhelming majority of calculations currently in the literature. For the water dimer, the present method obtains a binding energy of 4.8 kcal/mol[17,29] as opposed to a best estimate (with a very large basis set and coupled cluster type methods) of 5.0 kcal/mol.[27] In general, our expectation is that our results will be within ±0.5 kcal/mol of the converged gas phase answer and that the majority of the results have a precision that is better than this. Future work will involve detailed, systematic comparisons with higher levels of theory. While the level of theory we employ here is far from perfect, as pointed out above, the use of a high-quality TZ2P level basis set at the MP2 level for systems of this size is highly nontrivial.

In what follows, we shall define "donor" to refer to a hydrogen bond donor (that is, a hydrogen atom bonded to some chemical functional group) and "acceptor" to refer to hydrogen bond acceptor, typically the lone pair of an electronegative atom such as O, N, or S embedded in a specific functional group but including as well the benzene π

(10) Friesner, R. A. *Chem. Phys. Lett.* **1985**, *116*, 39.
(11) Friesner, R. A. *J. Chem. Phys.* **1986**, *85*, 1462.
(12) Friesner, R. A. *J. Chem. Phys.* **1987**, *86*, 3522.
(13) Friesner, R. A. *J. Phys. Chem.* **1988**, *92*, 3091.
(14) Friesner, R. A. *Annu. Rev. Phys. Chem.* **1991**, *42*, 341.
(15) Ringnalda, M. N.; Langlois, J.; Greeley, B. H.; Murphy, R. B.; Russo, T. V.; Cortis, C.; Muller, R. P.; Murlen, B.; Donnelly, R. E., Jr.; Mainz, D. T.; Wright, J. R.; Pollard, W. T.; Cao, Y.; Won, Y.; Miller, G. H.; Goddard, W. A. III; Friesner, R. A. PS-GVB V2.1, Schrödinger, Inc., 1995.
(16) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
(17) Saebø, S.; Tong, W.; Pulay, P. *J. Chem. Phys.* **1993**, *98*, 2170.
(18) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(19) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
(20) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(21) Jordan, K. D., personal communication.
(22) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.
(23) St.-Amant, A.; Cornell, W. D.; Kollman, P. A.; Halgren, T. A. *J. Comput. Chem.* **1995**, *16*, 1483.
(24) Murphy, R. B.; Pollard, W. T.; Friesner, R. A. *J. Chem. Phys.* **1997**, *106*, 5073.
(25) Pudzianowski, A. *J. Phys. Chem.* **1996**, *100*, 4781.
(26) *Modeling the Hydrogen Bond*; Smith, D., Ed.; ACS Symposium Series 589; American Chemical Society: Washington, DC, 1994.
(27) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104.
(28) Halgren, T. A. *Comput. Chem.* **1996**, *17*, 520.
(29) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *J. Chem. Phys.* **1995**, *103*, 1481.

Table 1. Side Chain Analogues in Amino Acids Studied for Donor−Acceptor Hydrogen-Bonding Interaction

side chain analogue	common amino acid	donor	acceptor
CH_4	alanine	C-H	
NH_2CH_3	lysine	N-H	N
NH_2COCH_3	asparagine, glutamine	N-H	N, O
$(NH_3CH_3)^+$	N-terminal glycine	N-H	
CH_3OH	serine	O-H	O
CH_3CHO	threonine	C-H	O
CH_3COOH	aspartic acid, glutamic acid	O-H	O
CH_3COO^-	C-terminal glycine		O
C_6H_6	phenylalanine	C-H	π
CH_3SCH_3	methionine		S
CH_3SH	cysteine	S-H	S

system. All of the complexes we study here involve polar donor and acceptor groups (even the benzene π system qualifies as a polar acceptor, to some extent) except the methane complexes, which are dominated by dispersion interactions. In these cases, we believe that HF geometry optimization does not lead to errors beyond the range suggested above. If the dispersion interaction is dominant, geometry optimization with a method like local MP2 (in which the BSSE is removed from the effective dispersive interaction) is likely to yield effective results. For this reason, we will consider interactions of aliphatic hydrogens with acceptor groups further in a separate study.

Initial geometries of all possible donor−acceptor complexes are designed using MacroModel.[22] Although geometry optimization uses C_1 symmetry to get lowest energy structures, many of complexes have a symmetry plane and these structures are reoptimized using C_s symmetry via use of **Z**-matrix to check if this way gives more stable structures. The use of C_s symmetry reproduces very close structural parameters and energies in C_1 symmetry.

A final cautionary note concerns the single case that we include containing a negative charge, CH_3COO^-. The values obtained for this case probably have larger errors than in other cases because one should use a diffuse basis set in treating negative ions. Nevertheless, the results reported here are likely to be reasonable because the basis set we used, cc-pVTZ(-f/d), does contain some small exponent functions. This issue will be examined further in future publications.

B. Donor and Acceptor Groups for the Side Chain Analogues. Table 1 presents a list of the side chain analogues that we study in this paper. The amino acid associated with the analogue is listed in the table. The remainder of the table enumerates the donor and acceptor groups in the molecule. For compounds containing oxygen, there are two acceptor modes for each O atom, corresponding to the two lone pair groups. In some cases, these will be equivalent (by symmetry), but in others there are two binding modes and the energies can be significantly different. Similarly, in an NH_2 group the two protons do not have to be equivalent, and in this case we treat each one as a separate donor proton. Because a major purpose of the paper is to examine different hydrogen-bonding conformations in detail, we calculate the interaction of every donor with every acceptor in the table.

III. Results

Tables 2−5 present binding energy data for all donor−acceptor pairs (a total of >140 in all) by four different methods: HF/6-31G** without a counterpoise correction, HF/6-31G** with a counterpoise correction, HF/cc-pVTZ(-f/d) with a couterpoise correction, and LMP2/cc-pVTZ(-f/d) results with a counterpoise correction for the HF part of the energy. The HF/6-31G** uncorrected results are included because it has repeatedly been claimed in the literature that accurate results are obtainable with this protocol (due to cancellation of correlation and basis set error); however, this assertion has not been explored with a large data set and high-quality comparison data. The HF/6-31G** results with counterpoise allow the magnitude of this effect to be manifested. The counterpoise-corrected cc-pVTZ (-f/d) results are presented at both the HF and LMP2 levels so that the magnitude of the correlation contribution to the binding can be assessed. For purposes of the subsequent discussion, we shall be referring to the data in Table 5 (i.e. the LMP2/cc-pVTZ(-f/d) results) unless otherwise specified.

It is not feasible to display all of the >140 structures in this article, so we instead present a representative sample of structures. These are shown in Figure 1. Each donor and acceptor is represented in at least one complex shown. Various sites of donor and acceptor abilities within a molecule leads to the variation in hydrogen bonded structure that are possible even in a simple molecule, e.g., sets of complexes II, V, VI, VII and VI, VII, X in Figure 1. Structure V shows two (multiple) hydrogen bonds from the two interaction sites of both donor and acceptor. Donor−acceptor distances for the complexes are presented in Table 6. These distances are then plotted against binding energies in Figures 2 and 3 for individual donor (columns in Table 5) or acceptor (rows in Table 5) and in Figure 4 for all (results in Table 2 are also plotted for comparison).

Several interesting observations can be made with regard to the plots of binding energy versus donor−acceptor distance presented in Figures 2−4. First, the data for fixed acceptor, varying the donor, appears to be considerably more regular than that for fixed donor with variable acceptor. In the former case, the data for neutral acceptors generally falls on a straight line (excluding the single charged donor species) with the exception of acetic acid, where it is likely that multiple functional group effects induces dispersion. Even the data for the charged acceptor COO^- group is relatively linear, albeit with a significantly larger slope than that of the neutral acceptors. In contrast, the donor plots, while predominantly linear in a number of cases, exhibit significantly more dispersion. This suggests that the acceptor exerts the greatest control over the hydrogen-bonding geometry. An examination of the entire data set in Figure 4 reveals the more general presentation of a biphasic pattern, representing two different slopes which change dramatically at ∼3 Å. The charged species are clearly in one of the biphasic regions, while the "weak" donor−acceptor pairs are in the other. The transitional region contains many of the remaining groups and exhibits significant dispersion.

A complete analysis of the implications of the data in Table 5 is beyond the scope of this paper. However, quite a lot can be learned from a relatively straightforward comparison among the molecules in the test suite. Before beginning the detailed analysis, a number of preliminary observations are important. First, there are some particular structures in which multiple functional group interactions between donor and acceptor are clearly significant. It is useful to include such structures in our database because they will be valuable in testing force fields for transferability (in particular, many body polarization effects will impact on how additive such interactions are). However, in trying to draw conclusions about the intrinsic hydrogen-bonding strengths of individual functional groups, this data is obviously not the most incisive. Consequently, we have indicated in the tables which structures we believe have a large multiple functional group contribution; in the discussion below, we ignore these structures. Fortunately, it is often the case that there is an alternative structure for a given donor−acceptor pair where the interactions are dominated by a single donor−acceptor interaction.

A second general observation is the prevalence of resonance effects in modifying donor and acceptor strengths due to the charge redistribution induced by the resonance forms. Our database allows quantitative insight into the impact of resonance on binding energies to be obtained. This idea is certainly not new, but it is useful to see its emergence from a large high-quality data set such as the present one. Further studies are obviously suggested by the results (in terms of variation of

Table 2. Binding Energies, ΔE (kcal/mol), of Binary Hydrogen-Bonding Complexes between the Amino Acid Side Chain Analogues at the PS-HF/6-31G** Level in the PS-HF/6-31G** Optimized Geometries[a]

complex	C[H]H_3	CH_3C[H]O	C_6[H]H_5	N[H]-HCH_3	N[H_α]-$HCOCH_3$	NH[H_β]-$COCH_3$	N[H]-$H_2CH_3^+$	CH_3O[H]	CH_3COO[H]	CH_3S[H]
[N]H_2CH_3	−0.66, −0.76	−2.17, −2.23	−1.52	−2.86	−6.94	−6.10	−23.21	−6.50, −6.58	−10.13[b]	−2.29
[N]H_2COCH_3	−0.19	−1.26		−0.17	−5.82[b]	−1.98		−2.62	−3.95	−0.17
CH_3[O]H	−0.76, −0.77	−2.23	−1.30	−2.66	−3.77, −6.27	−5.53	−19.16	−5.39	−6.54	−1.83
CH_3CH[$\backslash$O]	−0.65, −0.66	−2.95	−1.09	−3.67	−3.14	−4.92	−20.72	−5.74	−5.74	−2.06
CH_3CH[O$'$]	−0.61, −0.63	−2.34	−1.11	−2.82	−2.85	−4.95	−21.40	−5.24	−5.67	−2.10
NH_2C[$\backslash$O]CH_3	−0.84, −0.87	−3.49		−4.07	−4.25	−6.88	−28.30	−7.05	−7.74	−2.88
NH_2C[O$'$]CH_3	−0.88, −0.93	−7.46[b]	−1.29	−7.76[b]	−4.41	−6.71	−27.31	−7.44	−8.30	−2.85
CH_3C[$\backslash$O]OH	−0.64	−3.40	−0.92	−3.71	−3.29	−4.97	−21.24	−5.78	−6.03	−2.03
CH_3C[O$'$]OH	−0.70	−9.35[b]	−0.71	−10.44[b]	−2.51	−4.53	−19.85	−9.98[b]	−6.58	−1.96
CH_3CO[O]H	−0.24	−0.37		−1.10	−4.10[b]	−2.51	−7.07	−3.49	−3.74	−0.85
CH_3C[O][O]$^-$	−3.32 to −4.10	−10.73 to −15.47[b]		−9.03 to −10.16	−14.31 to −22.20	−22.93 −25.68		−17.19 to −18.26	−22.22 to −24.07	−10.46 to −12.13
C_6H_6[π]	−0.03, −0.05	−1.26	−0.55	−1.14	−1.98	−2.30	−13.16	−2.24		
CH_3[S]CH_3	−0.12	−1.03		−1.16	−1.27	−2.45	−13.74	−2.85	−2.93	−0.90
CH_3[S]H	−0.08	−0.86		−0.71	−1.40	−2.18	−11.73	−2.43	−2.59	−0.82

[a] A molecule in row times a molecule in column represents each optimized donor−acceptor complex. Atoms in brackets represent hydrogen donors and acceptors. The [H_α] and [H_β] in NH_2COCH_3 represent and hydrogen cis and trans to the carbonyl group, respectively. The [$\backslash$O] and [O$'$] represent lone-pair directions on the oxygen atom, and in CH_3C[O][O]$^-$ case, there are six lone-pair directions, three from each [O], giving ranges from several data points. Some complexes have two data points due to the different orientation of donors. Two of $(NH_3CH_3)^+$ complexes are missing due to severe deformations from initial geometries. [b] These binding energies include a large multiple functional group contribution.

Table 3. BSSE-Corrected Binding Energies, ΔE_{BSSE} (kcal/mol), of Binary Hydrogen-Bonding Complexes between the Amino Acid Side Chain Analogues at the PS-HF/6-31G** Level in the PS-HF/6-31G**-Optimized Geometries[a]

complex	C[H]H_3	CH_3C[H]O	C_6[H]H_5	N[H]-HCH_3	N[H_α]-$HCOCH_3$	NH[H_β]-$COCH_3$	N[H]-$H_2CH_3^+$	CH_3O[H]	CH_3COO[H]	CH_3S[H]
[N]H_2CH_3	−0.05, −0.17	−1.58	−1.00	−2.21	−5.35	−5.30	−22.19	−5.70, −5.74	−90.13[b]	−1.44
[N]H_2COCH_3	+0.09	−0.89		+0.30	−4.45[b]	−1.50		−2.45	−3.46	+0.24
CH_3[O]H	−0.06	−1.33	−0.63	−1.74	−3.21, −4.71	−4.58	−18.14	−4.24	−5.72	−1.00
CH_3CH[$\backslash$O]	−0.03, −0.06	−1.89	−0.61	−2.33	−2.41	−4.11	−19.78	−4.45	−4.84	−1.37
CH_3CH[O$'$]	+0.02, −0.08	−1.69	−0.61	−2.07	−2.20	−4.21	−20.53	−4.37	−5.01	−1.40
NH_2C[$\backslash$O]CH_3	−0.13, −0.18	−2.54		−2.90	−3.37	−5.76	−27.17	−5.96	−6.80	−1.90
NH_2C[O$'$]CH_3	−0.16, −0.22	−5.97[b]	−0.44	−6.19[b]	−3.38	−5.59	−26.23	−6.18	−7.09	−1.98
CH_3C[$\backslash$O]OH	−0.14	−2.27	−0.32	−2.50	−2.67	−4.20	−20.41	−4.72	−5.33	−1.44
CH_3C[O$'$]OH	−0.09	−8.01[b]	+0.03	−9.25[b]	−1.79	−3.72	−19.06	−8.65[b]	−5.57	−1.31
CH_3CO[O]H	+0.20	+0.09		−0.56	−2.83[b]	−1.54	−6.36	−2.31	−3.02	−0.29
CH_3C[O][O]$^-$	−1.65 to −2.13	−8.48 to −11.80[b]		−6.95 to −7.76	−12.04 to −18.68	−19.29 to −22.55		−14.30 to −15.43	−19.30 to −21.33	−7.64 to −9.22
C_6H_6[π]	+0.28, +0.14	−0.78	−0.04	−0.60	−1.10	−1.39	−12.13	−1.55		
CH_3[S]CH_3	−0.01	−0.88		−0.96	−1.02	−2.16	−13.48	−2.46	−2.57	−0.54
CH_3[S]H	+0.01, −0.01	−0.75		−0.80	−1.21	−2.15	−11.10	−2.16	−2.44	−0.57

[a] A molecule in row times a molecule in column represents each optimized donor−acceptor complex. Atoms in brackets represent hydrogen donors and acceptors. The [H_α] and [H_β] in NH_2COCH_3 represent and hydrogen cis and trans to the carbonyl group, respectively. The [$\backslash$O] and [O$'$] represent lone-pair directions on the oxygen atom, and in CH_3C[O][O]$^-$ case, there are six lone-pair directions, three from each [O], giving ranges from several data points. Some complexes have two data points due to the different orientation of donors. Two of $(NH_3CH_3)^+$ complexes are missing due to severe deformations from initial geometries. [b] These binding energies include a large multiple functional group contribution.

neighboring functional groups, for example, to further characterize resonance effects) and will be pursued in another publication.

A. Analysis of Results for Neutral Donors and Acceptors. We discuss first the uncharged donor/acceptor species, reserving an examination of the results for charged species for a separate section below. We analyze each molecule in the donor−acceptor pair database in terms of its behavior as a donor and acceptor, focusing when relevant on resonance structures. The simplest cases are NH_2CH_3, CH_3CHO, CH_3SCH_3, CH_3SH, and CH_3OH. These molecules all contain a single heteroatom with either hydrocarbon or hydrogen substituents, and resonance effects are irrelevant except in the case of CH_3CHO. As acceptors, the order of strength is N > O > S, with effects at the 0.5−1 kcal/mol level differentiating the alcohol from the carbonyl oxygen, or the SCH_3 moiety from the SH. These trends are readily understood from the acid/base properties of the various atoms, which in turn are related to the size of the nuclear charge and efficacy of atomic screening. Nitrogen, for example, is an excellent base due to its small atomic charge (as compared to oxygen, e.g.) and hence is also an excellent hydrogen-bonding acceptor. In previous work, we have shown that the hydrogen-bonding strength of the acceptor is, in contrast, not particularly well correlated with formal charge: the nitrogen in trimethylamine, for example, actually has a stronger hydrogen bond to water than ammonia, despite the fact that the ESP-fit charge on the latter is −1.1 eu and that on the former is only −0.2 eu. As donors, behavior depends as expected upon the acidity of the proton (again, this must be contrasted with models based upon classical electrostatic charges, which would yield very different conclusions). This leads to the unsurprising result that OH ≫ NH ∼ SH > CHO. The weakness of NH as a donor has been noted by us before and has significant consequences for solvation free energies of amines in water, for example. The aldehyde proton is a better donor than an aliphatic group like CH_4 would be, due to a resonance structure associated with putting a negative charge on the carbonyl oxygen and a positive charge on the proton. This is still not enough of an effect to move the proton up beyond NH or SH as a donor.

We next consider the carboxylic acid, CH_3COOH. As a donor, the proton here is the most acidic of all uncharged donating groups considered, as would be expected from the well-known important resonance structure. The resonance effect is 0.5−1 kcal/mol in magnitude as compared to the hydrogen bonds formed by methanol. The same resonance structure

Table 4. BSSE-Corrected Binding Energies, ΔE_{BSSE} (kcal/mol), of Binary Hydrogen-Bonding Complexes between the Amino Acid Side Chain Analogues at the PS-HF/cc-pVTZ(-f/d) Level in the PS-HF/6-31G**-Optimized Geometries[a]

complex	C[H]H_3	CH_3C[H]O	C_6[H]H_5	N[H]-HCH_3	N[H_α]-HCOCH_3	NH[H_β]-COCH_3	N[H]-$H_2CH_3^+$	CH_3O[H]	CH_3COO[H]	CH_3S[H]
[N]H_2CH_3	+0.09, −0.07	−1.40, −1.42	−0.81	−1.70	−4.61	−4.25	−20.31	−4.57, −4.66	−7.99[b]	−1.02
[N]H_2COCH_3	+0.09	−0.82		+0.47	−4.28[b]	−1.34		−1.84	−2.72	+0.22
CH_3[O]H	−0.03, −0.02	−1.13	−0.55	−1.25	−2.68, −4.07	−3.85	−16.72	−3.42	−4.80	−1.13
CH_3CH[$\backslash$O]	−0.03, −0.07	−1.85	−0.61	−1.99	−2.26	−3.71	−19.20	−3.99	−4.36	−1.28
CH_3CH[O$'$]	+0.04, −0.08	−1.68	−0.59	−1.81	−1.89	−3.83	−20.05	−3.83	−4.50	−1.30
NH_2C[$\backslash$O]CH_3	−0.09, −0.15	−2.57		−2.50	−3.32	−5.67	−26.81	−5.44	−6.38	−1.88
NH_2C[O$'$]CH_3	−0.10, −0.18	−5.59[b]	−0.50	−5.32[b]	−3.29	−5.54	−25.85	−5.65	−6.61	−1.84
CH_3C[$\backslash$O]OH	−0.12	−2.28	−0.31	−2.22	−2.59	−4.01	−19.81	−4.29	−4.84	−1.32
CH_3C[O$'$]OH	−0.11	−7.56[b]	+0.11	−8.15[b]	−1.76	−3.58	−18.58	−7.51[b]	−4.92	−1.22
CH_3CO[O]H	+0.22	+0.05		−0.43	−2.51[b]	−1.25	−5.50	−1.74	−2.39	−0.25
CH_3C[O][O]$^-$	−1.49 to −1.96	−8.12 to 11.70[b]		−6.29 to −7.12	−11.83 to −18.08	−18.90 to −21.75		−13.48 to −14.37	−18.03 to −19.70	−7.83 to 8.85
C_6H_6[π]	+0.16, +0.11	−0.83	−0.37	−0.48	−1.26	−1.44	−12.92	−1.27		
CH_3[S]CH_3	−0.02	−0.90		−0.93	−1.19	−2.28	−14.03	−2.47	−2.08	−0.48
CH_3[S]H	0.00, −0.02	−0.74		−0.82	−1.38	−2.05	−11.89	−2.05	−2.33	−0.46

[a] A molecule in row times a molecule in column represents each optimized donor−acceptor complex. Atoms in brackets represent hydrogen donors and acceptors. The [H_α] and [H_β] in NH_2COCH_3 represent and hydrogen cis and trans to the carbonyl group, respectively. The [$\backslash$O] and [O$'$] represent lone-pair directions on the oxygen atom, and in $CH_3C[O][O]^-$ case, there are six lone-pair directions, three from each [O], giving ranges from several data points. Some complexes have two data points due to the different orientation of donors. Two of $(NH_3CH_3)^+$ complexes are missing due to severe deformations from initial geometries. [b] These binding energies include a large multiple functional group contribution.

Table 5. BSSE-Corrected Binding Energies, ΔE_{BSSE} (kcal/mol), of Binary Hydrogen-Bonding Complexes between the Amino Acid Side Chain Analogues at the PS-LMP2/cc-pVTZ(-f/d) Level in the PS-HF/6-31G**-Optimized Geometries[a]

complex	C[H]H_3	CH_3C[H]O	C_6[H]H_5	N[H]-HCH_3	N[H_α]-HCOCH_3	NH[H_β]-COCH_3	N[H]-$H_2CH_3^+$	CH_3O[H]	CH_3COO[H]	CH_3S[H]
[N]H_2CH_3	−0.81, −0.74	−1.91, −1.90	−1.97	−2.80	−6.65	−6.20	−24.62	−6.35, −6.39	−10.29[b]	−2.89
[N]H_2COCH_3	−0.38	−0.13		+0.48	−4.45[b]	−1.59		−2.07	−3.62	+0.09
CH_3[O]H	−0.71, −0.65	−1.75	−1.72	−2.34	−3.79, −5.98	−5.32	−19.22	−4.83	−6.07	−2.37
CH_3CH[$\backslash$O]	−0.52, −0.60	−2.30	−1.59	−3.30	−2.63	−4.45	−19.62	−5.11	−5.07	−2.17
CH_3CH[O$'$]	−0.44, −0.47	−1.76	−1.52	−2.40	−2.68	−4.46	−20.39	−4.42	−5.01	−2.19
NH_2C[$\backslash$O]CH_3	−0.67, −0.69	−2.83		−3.70	−3.81	−6.59	−28.34	−6.39	−7.20	−3.12
NH_2C[O$'$]CH_3	−0.76, −0.58	−6.79[b]	−1.37	−7.68[b]	−3.68	−6.11	−26.75	−6.78	−7.69	−2.81
CH_3C[$\backslash$O]OH	−0.68	−2.90	−1.18	−3.56	−2.89	−4.47	−20.09	−5.19	−5.46	−2.11
CH_3C[O$'$]OH	−0.66	−8.31[b]	−0.74	−10.75[b]	−2.10	−3.94	−18.52	−8.90[b]	−5.70	−1.98
CH_3CO[O]H	−0.24	−0.56		−1.27	−3.95[b]	−2.91	−8.69	−3.42	−3.69	−1.49
CH_3C[O][O]$^-$	−2.84 to −3.30	−8.95 to −12.64[b]		−8.27 to −8.99	−15.01 to −21.02	−21.97 to −24.30		−16.04 to −16.60	−21.63 to −22.54	−11.06 to −12.18
C_6H_6[π]	−0.80, −0.91	−2.36	−1.66	−2.21	−2.89	−3.32	−15.24	−3.11		
CH_3[S]CH_3	−0.38	−1.41		−1.66	−2.15	−3.42	−15.68	−3.68	−3.52	−2.12
CH_3[S]H	−0.24, −0.31	−1.11		−1.37	−2.06	−2.86	−13.33	−2.85	−3.15	−1.56

[a] A molecule in row times a molecule in column represents each optimized donor−acceptor complex. Atoms in brackets represent hydrogen donors and acceptors. The [H_α] and [H_β] in NH_2COCH_3 represent and hydrogen cis and trans to the carbonyl group, respectively. The [$\backslash$O] and [O$'$] represent lone-pair directions on the oxygen atom, and in $CH_3C[O][O]^-$ case, there are six lone-pair directions, three from each [O], giving ranges from several data points. Some complexes have two data points due to the different orientation of donors. Two of $(NH_3CH_3)^+$ complexes are missing due to severe deformations from initial geometries. [b] These binding energies include a large multiple functional group contribution.

putting a negative charge on the carbonyl oxygen stabilizes the carbonyl as an acceptor, although this effect is not as easy to extract from Table 5 due to multiple functional group interference in quite a few cases. In contrast, the alcoholic OH in the carboxylic acid is destabilized as an acceptor, due to the positive charge accrued in the resonance structure.

Finally, we consider acetamide, NH_2COCH_3. The resonance structure here leads to the well-known planarity of the peptide bond. This is quite a strong resonance, and the effects on donor and acceptor properties are very large in comparison with the results we have been discussing above. The nitrogen as expected becomes a significantly worse acceptor, due to the formal positive charge in the resonance structure. As compared to the nitrogen in methylamine, differences on the order of 3 kcal/mol in hydrogen-bonding strength are observed. The resonance structures similarly effects all other donor and acceptor species: the N-H protons become significantly more acidic than those in methylamine, and the carbonyl is a better acceptor than in acetaldehyde. Multiple functional group effects can again be substantial here, which is unsurprising in view of the molecular structure.

Having examined each molecule in the test suite individually, we now attempt to assess general trends that can be extracted from the data in Table 5 and Figures 2 and 3. One striking result is that there is a very wide variation among donors and acceptors in the spread of binding energies as the partner changes, leaving aside the charged groups. Consider first the sulfur acceptors and the benzene π system. The variation in binding energy is a surprisingly small 3.3 kcal/mol at the maximum, with most variations much smaller. We designate these acceptors as dominant acceptors, in the sense that they control the size of the binding energy. Dominant acceptors are weak; there is something about the accepting group that simply does not permit large binding energies to be achieved for an uncharged donor, no matter how acidic the proton in question. Similarly, there are a group of weak, dominant donors; these included CH_3CHO, NH_2CH_3, benzene, CH_3SH, and CH_4 with much the same quantitative characteristics. Cases of multiple functional group interactions have been discarded in this analysis.

If an acceptor is strong, as in the case of the nitrogen in methylamine or the various oxygen acceptors, the strength of

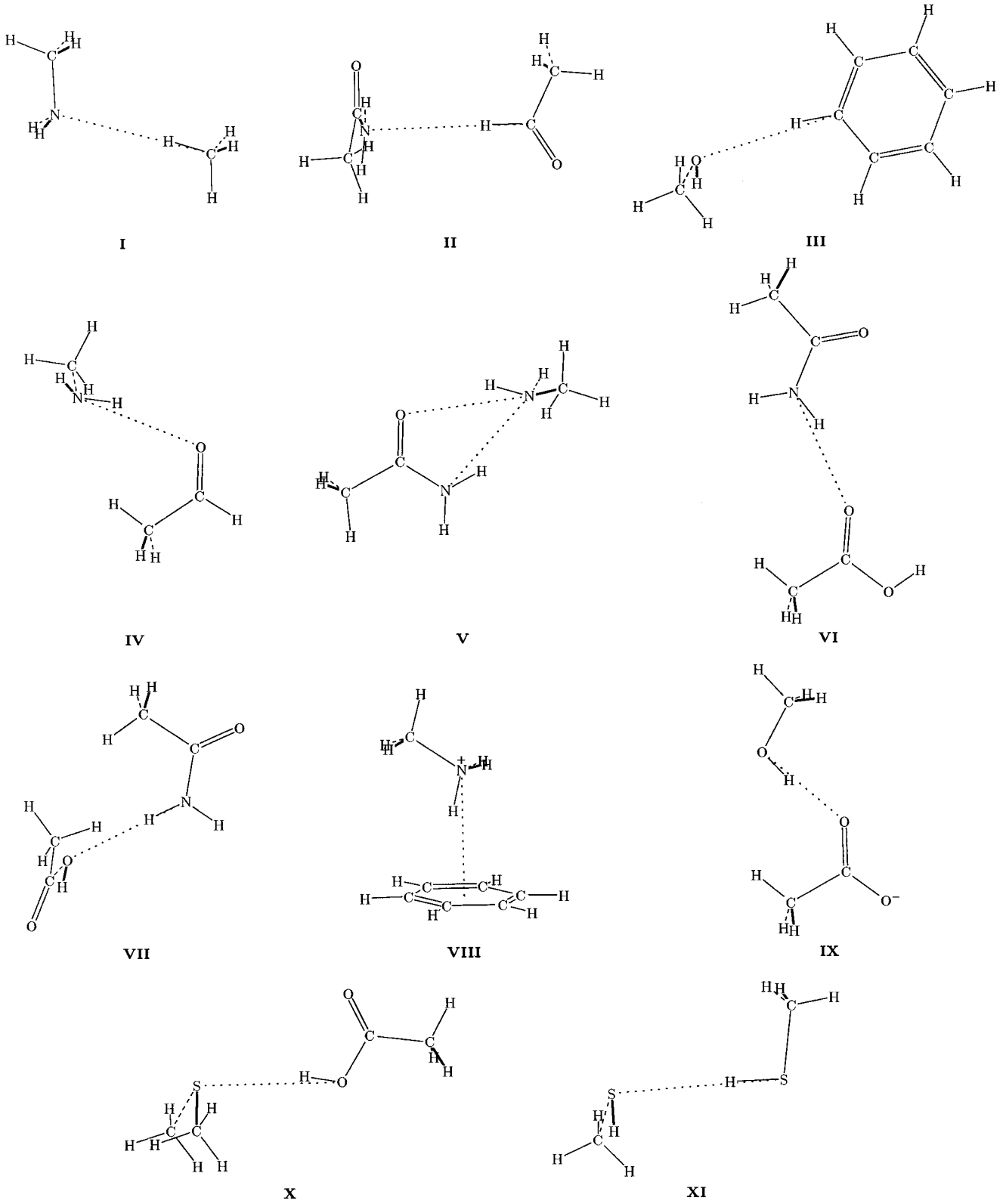

Figure 1. A representative sample of donor−acceptor structures between the amino acid side chain analogues optimized at the PS-HF/6-31G** level.

the interaction depends upon the donor, spanning a relatively wide range of values. Similarly, a strong donor, such as the alcohol proton in methanol or the carboxylic acid group proton in acetic acid, exhibits a wide range of interaction strengths as well. Intermediate behavior can also be observed, as for the carbonyl oxygen in acetic acid. These observations suggest that it should be possible to predict with reasonable reliability the binding energy of a new donor−acceptor pair as long as there is some data on each of their behaviors with a small set of test molecules. This sort of model, which will be the subject of a future publication, could prove to be extremely valuable in predicting the interaction energies for a very large list of donors and acceptors. A brute force approach to enumerating the interaction energies would scale as the product of the number of donors with the number of acceptors. In contrast, if one could carry out a small number of calculations for each molecule in the list and then extrapolate the results for the remaining pairs, linear rather than quadratic scaling could be achieved.

Table 6. Equilibrium Bond Distances, R_e (Å), of Binary Hydrogen-Bonding Complexes between the Amino Acid Side Chain Analogues Optimized at the PS-HF/6-31G** Level, Where Distances Are Hydrogen Bond Lengths between Heavy Atom Elements of Hydrogen Donor and Hydrogen Acceptor in Complexes[a]

complexes	C[H]H_3	CH_3C[H]O	C_6[H]H_5	N[H]-HCH_3	N[H_α]-$HCOCH_3$	NH[H_β]-$COCH_3$	N[H]-$H_2CH_3^+$	CH_3O[H]	CH_3COO[H]	CH_3S[H]
[N]H_2CH_3	3.79, 3.96	3.86, 3.91	3.96	3.56	3.09	3.18	2.83	3.03, 3.01	2.86, 3.24[b]	3.71
[N]H_2COCH_3	4.27	4.06		3.62	3.29, 3.25[b]	3.54		3.31	3.03	4.16
CH_3[O]H	3.89, 3.87	3.58	3.66	3.31	3.15, 3.08	3.09	2.77	2.96	2.83	3.68
CH_3CH[$^\backslash$O]	3.85, 3.90	3.56	3.75	3.31	3.19	3.12	2.76	2.95	2.90	3.76
CH_3CH[O$^/$]	3.73, 3.92	3.71	3.65	3.40	3.19	3.16	2.75	2.97	2.89	3.71
NH_2C[$^\backslash$O]CH_3	3.82, 3.88	3.64		3.24	3.09	3.06	2.67	2.92	2.81	3.63
NH_2C[O$^/$]CH_3	3.80, 3.88	3.27, 3.14[b]	3.48	3.16, 3.10[b]	3.08	3.06	2.69	2.89	2.80	3.64
CH_3C[$^\backslash$O]OH	3.93	3.53	3.50	3.27	3.22	3.12	2.74	2.95	2.85	3.74
CH_3C[O$^/$]OH	3.95	3.23, 2.83[b]	3.48	3.11, 2.85[b]	3.27	3.17	2.77	2.83, 2.79[b]	2.84	3.77
CH_3CO[O]H	3.84	3.95		3.59	3.25, 3.48[b]	3.21	2.92	3.05	2.97	3.83
CH_3C[O][O]$^-$	3.41−3.47	3.02−3.37[b]		2.97−3.01	2.77−3.01	2.79−2.86		2.67−2.72	2.46−2.51	3.18−3.23
C_6H_6[π]	4.18, 4.22	4.02	4.22	3.87	3.71	3.79	3.17	3.57		
CH_3[S]CH_3	4.72	4.33		4.09	3.87	3.82	3.44	3.64	3.61	4.18
CH_3[S]H	4.70, 4.80	4.49		4.16	3.96	3.90	3.48	3.72	3.65	4.27

[a] A molecule in row times a molecule in column represents each optimized donor−acceptor complex. Atoms in brackets represent hydrogen donors and acceptors. The [H_α] and [H_β] in NH_2COCH_3 represent and hydrogen cis and trans to the carbonyl group, respectively. The [$^\backslash$O] and [O$^/$] represent lone-pair directions on the oxygen atom, and in CH_3C[O][O]$^-$ case, there are six lone-pair directions, three from each [O], giving ranges from several data points. Some complexes have two data points due to the different orientation of donors. Two of $(NH_3CH_3)^+$ complexes are missing due to severe deformations from initial geometries. [b] These two bond lengths are due to a multiple functional group interaction.

We next examine briefly variation in the binding energy for different structures between the same donor and acceptor pair. Variations as large as 8 kcal/mol are observed, which of course reflects multiple functional group interactions. The interpretation of these results without a molecular mechanics model does not strike us as productive, so we defer such an analysis to another publication. Nevertheless, we believe that presenting the data here is quite useful as it serves to illustrate just how important such effects can be and also because it can readily be used to test the performance of a molecular mechanics force field.

B. Analysis of Results for Charged Donor or Acceptor Species. We have examined one positively charged donor, $(NH_3CH_3)^+$, and one negatively charged acceptor, CH_3COO^-, each of which is an important functional group in actual protein structures. Unsurprisingly, the hydrogen-bonding energies are significantly larger when a charged species is involved, due to the increased magnitude of the electrostatic interaction. Charged species are also clearly "strong" donors and acceptors, with a wide range of binding energies evidenced from Table 5.

Indeed, the variation in hydrogen-bonding strength as the partner of the charged species is varied is as much as an order of magnitude larger than the corresponding difference for neutral pairs. For example, the binding energy of $(NH_3CH_3)^+$ with the carbonyl oxygen in *N*-methylacetamide is 28 kcal/mol, but is only 19 kcal/mol with the oxygen in methanol, a difference of 9 kcal/mol. In contrast, the binding energies of these two molecules with a typical neutral donor, for example the N−H hydrogen in methylamine, differ by only 1 kcal/mol. This difference in behavior reflects the much greater variation in electrostatic energies which occurs when one species has a net charge.

C. Effects of Basis Set, Superposition Error, and Electron Correlation on Hydrogen-Bonding Energetics. We examine two interesting questions in this section with regard to the influence of the level of quantum chemical theory on the accuracy of hydrogen bonding energies. First, how large are the specific effects of basis set, electron correlation, and superposition error in the Hartree−Fock component of the energy? Secondly, is the commonly used approximation of computing these energies at the 6-31G** level without counterpoise correction (ostensibly leading to a cancellation of basis set and correlation errors with superposition error) a reliable one, and how large are the errors from a quantitative and qualitative point of view?

One can get both quantitative and qualitative feeling for the answers to the first question by determining the maximum and minimum difference between the results of the various levels of theory presented in Tables 2−5. First, consider the LMP2 and HF results with the cc-pVTZ(-f/d) basis set, both corrected for superposition error at the HF level (Tables 4 and 5). In this case, the maximum energy difference is ∼2.6 kcal/mol (carbonyl oxygen acceptor CH_3COOH with proton donor NH_2-CH_3) to ∼0.1 kcal/mol (CH_3CHO as oxygen acceptor with CH_3-CHO as proton donor), with the LMP2 results always displaying stronger binding. The differences vary across this entire range, and at first glance it is not obvious how they could be predicted in advance. The magnitude of basis set effects at the HF level (comparing Tables 3 and 4) appears to range from ∼0 to 1 kcal/mol, with the larger basis set displaying the smaller binding energy. The BSSE effects, obtained by examining the 6-31G** results with and without counterpoise corrections (Tables 2 and 3) are on the order of ∼0−1.6 kcal/mol.

We next perform a similar comparison of the BSSE-corrected LMP2/cc-pVTZ(-f/d) results (Table 5) with the *uncorrected* HF/6-31G** results (Table 2). Some cancellation of error is in fact manifested (shown clearly in Figure 3): the absolute maximum deviation for neutral donor−acceptor pairs between the two tables is ∼1 kcal/mol. However, this deviation can go in either direction, so the spread in errors is still 2 kcal/mol. For this case, the sign of the error appears to be more specifically correlated with donor and acceptor species: for example, sulfur acceptors appear to be lower in energy at the LMP2 level (reflecting the fact that the sulfur atom has a relatively large binding contribution of electron correlation), as does the benzene π system, for the same reason. On the basis of a very preliminary analysis, it looks as though it should be possible to employ such a lower level approximation in conjuction with some sort of empirical correction based upon the donor and acceptor groups and obtain reasonably accurate results. The raw numbers, however, can be considered reliable only to ±1 kcal/mol. Whether this is an adequate level of precision depends upon the use to which the calculations are to be put, a subject beyond the scope of the present discussions.

Finally, we consider the effects of the level of theory upon the results for charged groups. The deviations are larger in terms

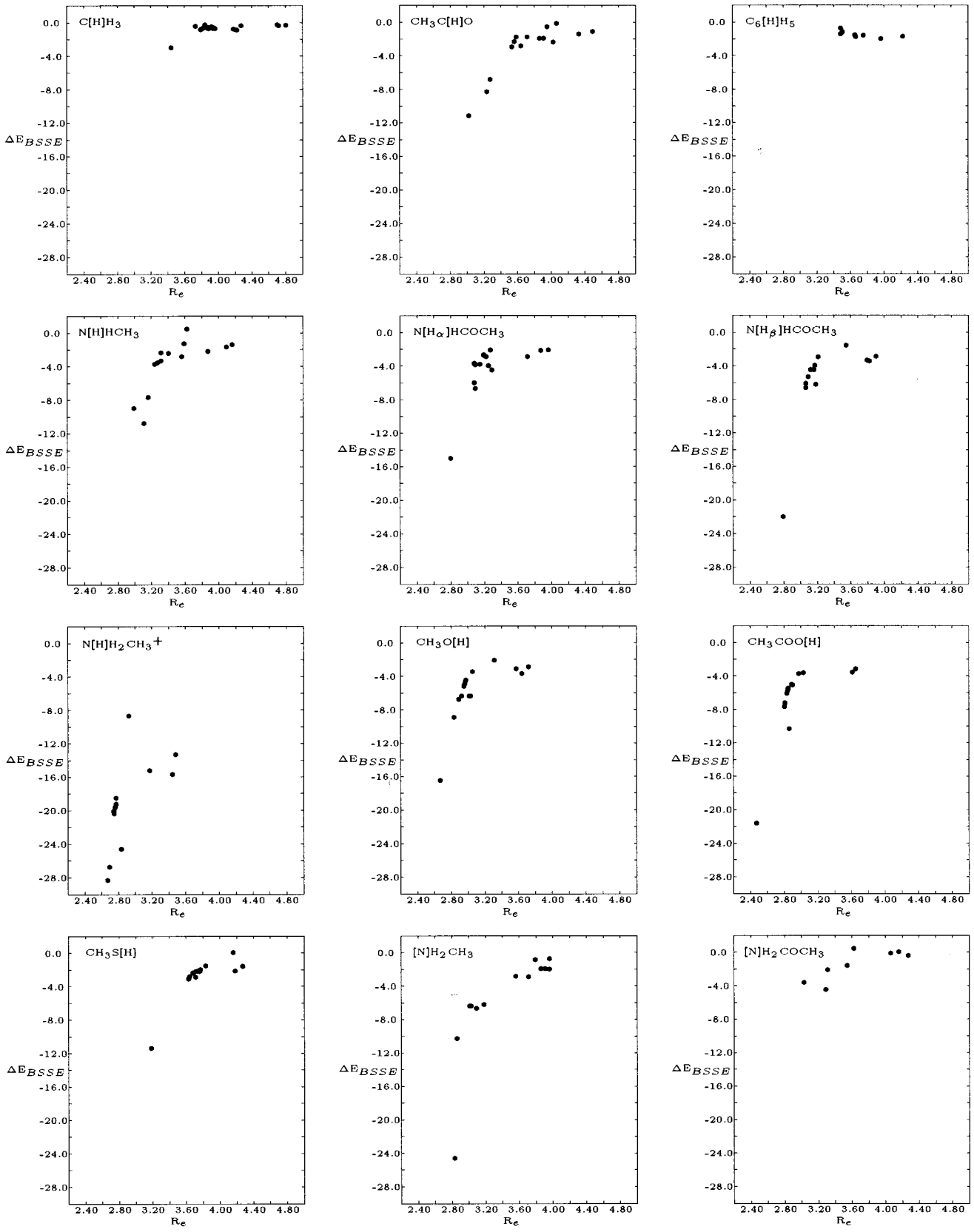

Figure 2. BSSE-corrected binding energies, ΔE_{BSSE} (kcal/mol), vs donor−acceptor distances, R_{e} (Å), at the PS-LMP2/cc-pVTZ(-f/d) level in the PS-HF/6-31G** optimized geometries for some individual donor and acceptor complexes.

of absolute energy (maximum ∼4 kcal/mol) but comparable if one considers percentage errors for the charged and neutral cases. Again, the question of what is an acceptable level of theory cannot be answered without the context of a particular application.

IV. Conclusion

The calculations described above are an initial attempt to systematically study hydrogen-bonding interactions at a high level of quantum chemical theory for a wide range of donors

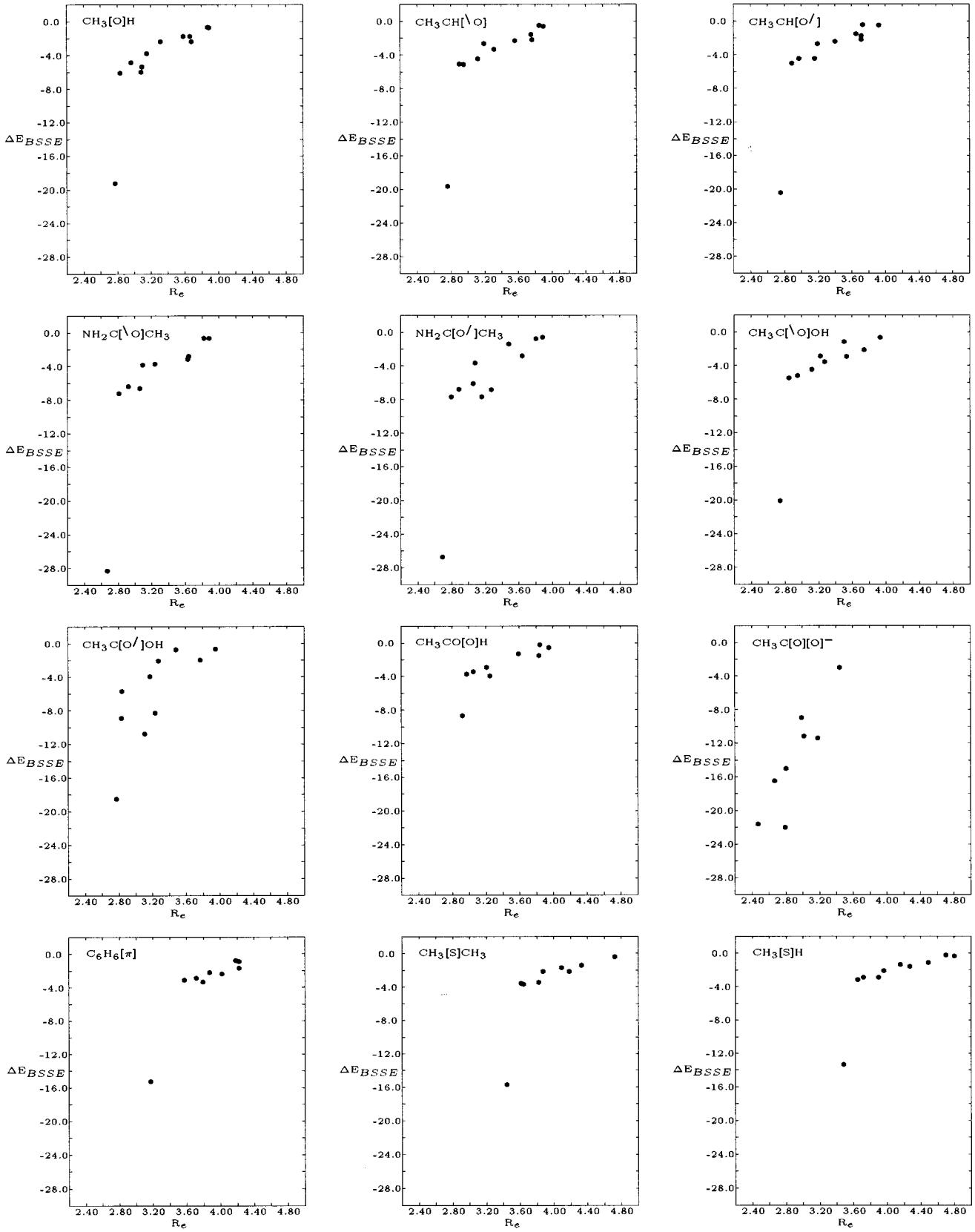

Figure 3. BSSE-corrected binding energies, ΔE_{BSSE} (kcal/mol), vs donor−acceptor distances, R_e (Å), at the PS-LMP2/cc-pVTZ(-f/d) level in the PS-HF/6-31G** optimized geometries for some individual donor and acceptor complexes.

and acceptors, considering variation in structure as well as species. While far from definitive, the results clearly surpass in breadth any that have been assembled in the literature to date and are superior in quality of the calculation to all but a few small molecules studies.

Nevertheless, there is much more that needs to be done before a comprehensive, reliable data base has been generated and a deep understanding has been achieved. Our own research directions for the future includes the following:

(1) Improvement of the level of quantum chemical theory, with regard to both the basis set, electron correlation method, and level at which geometry optimization is carried out. It will

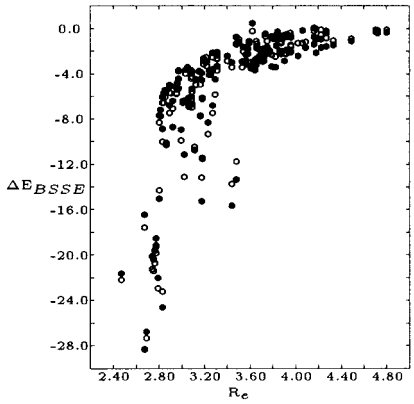

Figure 4. BSSE-corrected binding energies, ΔE_{BSSE} (kcal/mol) vs donor−acceptor distances, R_e (Å), at the PS-LMP2/cc-pVTZ(-f/d) level in the PS-HF/6-31G**-optimized geometries for all donor−acceptor complexes. Binding energies, ΔE (kcal/mol), at the PS-HF/6-31G** level marked by white circles (○) are drawn for comparison.

be particularly valuable to compare our anticipated multireference MP2 results with those obtained from other high-level methods such as QCISD(T); convergence of more than one high-level method to the same answer will provide confidence that a key component of the energy is not missing. This is a very tricky issue as is shown in our recent work on conformational energetics, where the QCISD(T) results appear to be significantly in error for at least one apparently innocuous case.[24]

(2) Extension of the present study to a much larger number of donors and acceptors. The investigation here of resonance effects and intrinsic hydrogen-bonding strengths has barely scratched the surface, at least from the standpoint of having a clear understanding of the size of the effect for an arbitrary functional group. While the results obtained make intuitive sense, that intuition may not transfer over to the wide range of functionality existing in the chemical literature.

(3) Development of a heuristic model for estimation of hydrogen-bonding strengths. The patterns observed above have numerous regularities, and it should be possible to build an approximate model which captures these to a reasonable level of precision. While such models have been investigated for many years in an empirical sense, there is an opportunity to improve the precision substantially, by making use of the ability to rapidly generate benchmark quantum chemical data. Achieving a reliability in the 0.25−0.5 kcal/mol range for neutral species does not seem out of the question.

(4) A much more extensive investigation of charged species needs to be carried out. There has been some recent work along these lines in the literature,[25] but it did not attempt the kind of systematic comparisons across a large number of functional groups that we present here. As mentioned above, diffuse functions should definitely be used to improve the precision of negative ion studies. Similarly, the studies of the benzene π system interacting with various donors is a very limited investigation of aromatic systems serving as hydrogen-bonding acceptors (although a highly relevant one for understanding some important features of protein structure), and additional work on this kind of system would be desirable.

(5) A deeper understanding of what causes intrinsic variations in donor−acceptor behavior would be desirable. While a great deal of work in the literature has been published in this area (e.g. various decompositions of the binding energy in hydrogen bonded systems), consideration of a large data set such as is presented here in this context has not yet been attempted.

Acknowledgment. This work was supported in part by a grant from the NIH (GM-40526) and the NIH Division of Research Resources (P41-RR06892). Grants of supercomputer time via the MetaCenter program of the NSF High Performance Computing Centers is acknowledged.

JA971836D