

## Accurate *ab Initio* Study on the Hydrogen-Bond Pairs in Protein Secondary Structures

Zhi-Xiang Wang, Chun Wu, Hongxing Lei, and Yong Duan\*

*Genome Center and Department of Applied Science, University of California, Davis, California 95616*

Received January 17, 2007

**Abstract:** *Ab initio* calculations up to the MP2/aug-cc-pVQZ//MP2/6-311+G\*\* level have been carried out to characterize the four patterns of hydrogen-bond (H-bond) pairs in protein secondary structures. The unblocked and methyl-blocked glycine dipeptide dimers were arranged to model the H-bond pairs in  $\alpha$ -helix ( $\alpha$ HH) and antiparallel ( $A\beta\beta$ -C<sub>5</sub> and  $A\beta\beta$ -C<sub>7</sub>) and parallel  $\beta$ -sheet ( $P\beta\beta$ ) secondary structures. The study uncovers that, in addition to the primary CO $\cdots$ NH H-bonds and the crossing secondary interactions, the CH $\cdots$ OC H-bonds and the tertiary effect (as we call it) also contribute substantially. The tertiary effect is due to the interpolarization between the donor and acceptor of a H-bond. This effect, which enhances the dipole–dipole interactions between two nearby H-bonds, stabilizes the  $\beta$ -sheet-like but destabilizes the helix-like H-bond pairs. The MP2 binding energies of the complexes were further refined by extrapolating to the complete basis set limit (CBS) according to Truhlar and co-workers and by a three-basis-set-based method. The best extrapolated CBS(aD-aT-aQ) binding energies of the unblocked dimers are  $-13.1$  ( $\alpha$ HH),  $-11.3$  ( $A\beta\beta$ -C<sub>5</sub>),  $-19.2$  ( $A\beta\beta$ -C<sub>7</sub>), and  $-14.8$  kcal/mol ( $P\beta\beta$ ). For the methyl-blocked counterparts, the best extrapolated CBS(D-T-Q) binding energies are  $-14.8$ ,  $-13.4$ ,  $-20.8$ , and  $-16.7$  kcal/mol, respectively. The interactions in the parallel  $\beta$  conformations are very close to the averages of the C<sub>5</sub> and C<sub>7</sub> antiparallel  $\beta$  conformations, and both are stronger than the helical dimers. Because the additive force fields are unable to account for the tertiary effect owing to the lack of polarization, all examined additive force fields significantly overestimate the interaction energies of the helix conformations relative to the  $\beta$ -sheet conformations. Notably, the agreement between molecular mechanical and quantum mechanical binding energies is improved after turning on the polarization. The study provides reference *ab initio* structures and binding energies for characterizing the backbone H-bonds of the protein secondary structures, which can be used for the parametrization of empirical molecular mechanics force fields.

### 1. Introduction

Hydrogen bonds (H-bonds), together with other weak interactions, are some of the most important determinants of the three-dimensional structures of proteins.<sup>1,2</sup> The energy of a single H-bond, ranging from 5.0 to 10.0 kcal/mol, is comparable to the typical folding free energies of proteins. Thus, accurate characterization of these H-bonds is vital for understanding the factors stabilizing protein structures.

Accurate H-bond energies are also crucial reference data for the development of protein molecular mechanics (MM) force fields that have become powerful tools in structural biology.<sup>3</sup> Numerous studies<sup>4–16</sup> have been performed to gain insight into the underlying physical interactions of H-bonds. Among the H-bonds in proteins, the backbone C=O $\cdots$ H–N H-bonds play particularly important roles and are the major driving forces for forming the ordered secondary structures.

Modeling the backbone H-bonds has been one of the major concerns in parametrizing molecular mechanics force fields

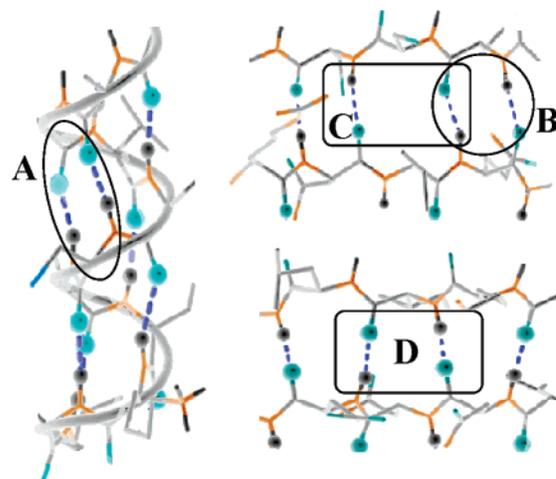
\* Corresponding author tel.: (530) 754-7632; fax: (530) 754-9658; e-mail: duan@ucdavis.edu.

for protein simulations, and various potential functions has been developed. For the physical-based force fields, because of the lack of experimental data for the backbone H-bonds, *ab initio* values were often used as reference data. As a prototype, the *trans* N-methylacetamide (NMA) dimer<sup>17–24</sup> has long been used to model such H-bonds and has been compared with the NMA–water complexes to study the relative strength of inter and intra H-bonds. With the advancement in computer hardware and software, the H-bond energies of these model complexes have been updated continuously, from Jorgensen and Swenson's<sup>23,24</sup> Hartree–Fock (HF)/minimal basis set calculations in 1985 to the most recent work of Langley and Allinger<sup>19</sup> at the MP2/6-311++G(2d,2p) level. Using the continuum solvent model, we<sup>25</sup> recently studied the solvent effect on the H-bonds. It is interesting to mention that Kelly and co-workers<sup>26</sup> have recently developed the amide-to-ester mutation approach to estimate the contribution of backbone H-bonds. But they are the free energy contributions and cannot be used to parametrize physical-based force fields directly.

A limitation of the *trans* NMA–NMA model is that it only contains one H-bond and is unable to capture the neighboring effect exerted by the nearby H-bonds on protein backbones. Recently, the aesthetic H-bond network in protein secondary structures has attracted attention from both experimentalists<sup>27,28</sup> and theorists.<sup>29–34</sup> Highlights of these efforts include the works of Wu<sup>29,30</sup> and Dannenberg<sup>31–34</sup> and their respective co-workers. In these cases, the influences on the H-bonds were assessed in the context of H-bond networks, but high-level *ab initio* calculations were difficult to perform due to the large size of the model complexes. Notwithstanding the efforts, ambiguity and controversy exist as to the contributions of the underlying physical interactions, and a detailed and reliable characterization of H-bond pairs in the context of protein secondary structures is unavailable. It is worth mentioning that Hobza and co-workers<sup>35,36</sup> have delivered highly accurate H-bond energies for nucleic acid base pairs using state-of-the-art computational chemistry methods.

An empirical approach to consider the neighboring effect in H-bond networks has been proposed by Jorgensen and Pranata,<sup>37</sup> who found that the effect in the multiple H-bonds of the nucleic acid base pairs could be accounted reasonably by the secondary interactions. This approach has been applied beyond the base pairs; because their model is consistent with additive point charge molecular mechanics force fields, widespread application of the latter implicitly renders their approach as the *de facto* model to account for the neighboring effect. However, as is well-known, the additive force fields are unable to represent the polarization effect. This approach, even for the base pairs, has been questioned by Lukin and Leszczynski<sup>38</sup> and by Dannenberg and co-workers<sup>39</sup> on the basis of the *ab initio* calculations. The fidelity of their approach in describing the H-bond pairs in protein secondary structures has not been examined despite the countless (implicit) applications.

In this study, we are interested in the typical H-bond pairs existing in protein secondary structures (Figure 1). We attempt (i) to reliably characterize the interactions between the two strands, (ii) to assess the neighboring effect between



**Figure 1.** Four patterns of H-bond pairs in protein secondary structures. (A) H-bond pair in  $\alpha$  helices; (B)  $C_5$  H-bond pair in antiparallel  $\beta$  sheets; (C)  $C_7$  H-bond pair in antiparallel  $\beta$  sheets; (D) H-bond pair in parallel  $\beta$  sheets. The notations  $C_5$  and  $C_7$  refer to the H-bond pairs in which the H of NH and the O of CO in the same strand are four ( $C_5$ ) and six ( $C_7$ ) bonds away, respectively.

two nearby H-bonds in these pairs, (iii) to obtain accurate interaction energies, which can be used to guide the parametrization of force fields, and (iv) to examine the fidelity of the modern force fields with regard to the ability to model the main-chain H-bonds of peptides.

## 2. Methods

Two sets of glycine dipeptide dimers (referred to as unblocked and methyl-blocked, respectively) were used to model the patterns of H-bond pairs shown in Figure 1. The unblocked set has the advantage to avoid the steric disturbance of the blocking methyl groups, whereas the H-bond donors and acceptors in the methyl-blocked set have the chemical-bonding environment closer to protein peptides. The four dimers in the unblocked set hereafter are referred to as  $\alpha$ HH for the H-bond pairs in the  $\alpha$  helix,  $A\beta\beta-C_5$  and  $A\beta\beta-C_7$  for the H-bond pairs in the antiparallel  $\beta$  sheet, and  $P\beta\beta$  for the H-bond pairs in the parallel  $\beta$  sheet. Their counterparts in the methyl-blocked set are labeled as  $\alpha$ HH',  $A\beta\beta-C_5'$ ,  $A\beta\beta-C_7'$ , and  $P\beta\beta'$ , respectively. Here, " $C_5$ " and " $C_7$ " denote the H-bond pairs where the hydrogen of the NH donor and the oxygen of the CO acceptor in the same strand are four ( $C_5$ ) and six bonds ( $C_7$ ) away, respectively. These dimers and their corresponding monomers are not at the energy minima. To maintain the H-bond pairs to be similar to those in protein secondary structures, we therefore fixed the backbone ( $\Phi$ ,  $\Psi$ ) torsions at the typical angles in protein secondary structures, that is, ( $-57.0^\circ$ ,  $-47.0^\circ$ ) in  $\alpha$ HH and  $\alpha$ HH'; ( $-119^\circ$ ,  $113^\circ$ ) in  $P\beta\beta$  and  $P\beta\beta'$ ; and ( $-139^\circ$ ,  $135^\circ$ ) in  $A\beta\beta-C_5$ ,  $A\beta\beta-C_7$ ,  $A\beta\beta-C_5'$ , and  $A\beta\beta-C_7'$ , respectively. It should be noted that, as these energies are applied to calibrate or parametrize empirical force fields, the backbone torsion angles in the force field calculations should be fixed to the same values.

Accurate estimation of nonbonded interactions, including H-bond interactions, has long been a challenge in compu-

**Table 1.** Energetic Results (in kcal/mol) of Unblocked and Methyl-Blocked H-bond Pairs at Various Levels, Including the BSSE-Uncorrected ( $\Delta E_{uc}$ ) and BSSE-corrected ( $\Delta E_c$ ) bonding energies, together with the BSSE corrections ( $\Delta E_{bsse}$ )<sup>a</sup>

unblocked dimers	$\alpha\text{HH}$			$A\beta\beta\text{-C}_5$			$A\beta\beta\text{-C}_7$			$P\beta\beta$			$\Delta\Delta E_{uc} = \Delta E_{uc} - \Delta E_{uc}(\alpha\text{HH})$		
	$\Delta E_{uc}$	$\Delta E_{bsse}$	$\Delta E_c$	$\Delta E_{uc}$	$\Delta E_{bsse}$	$\Delta E_c$	$\Delta E_{uc}$	$\Delta E_{bsse}$	$\Delta E_c$	$\Delta E_{uc}$	$\Delta E_{bsse}$	$\Delta E_c$	$A\beta\beta\text{-C}_5^b$	$A\beta\beta\text{-C}_7^c$	$P\beta\beta^d$
MP2/6-311+G**	-12.4	2.9	-9.5	-13.5	2.7	-9.8	-19.3	3.1	-16.2	-15.5	2.8	-12.7	1.1	-6.9	-3.1
MP2/cc-pVDZ	-14.7	6.7	-8.0	-13.7	5.1	-8.6	-22.2	7.1	-15.0	-17.6	6.1	-11.5	1.0	-7.5	-2.9
MP2/cc-pVTZ	-14.3	3.1	-11.3	-12.5	2.3	-10.3	-20.6	2.8	-17.7	-16.2	2.5	-13.6	1.8	-6.3	-1.9
MP2/cc-pVQZ	-13.9	1.4	-12.5	-12.2	1.0	-11.2	-20.2	1.2	-19.0	-15.8	1.1	-14.6	1.7	-6.3	-1.9
MP2/aug-cc-pVDZ	-15.5	3.6	-11.9	-13.8	3.0	-10.8	-21.8	3.4	-18.4	-17.2	3.1	-14.2	1.7	-6.3	-1.7
MP2/aug-cc-pVTZ	-14.6	1.9	-12.8	-13.0	1.7	-11.3	-21.1	1.9	-19.2	-16.6	1.8	-14.8	1.6	-6.5	-2.0
MP2/aug-cc-pVQZ	-14.0	1.0 <sup>e</sup>	-13.0	-12.3	1.0 <sup>e</sup>	-11.3	-20.4	1.4 <sup>e</sup>	-19.0	-15.9	1.1 <sup>e</sup>	-14.8	1.7	-6.4	-1.9
CBS(D-T) <sup>f</sup>	-15.0			-12.6			-20.7			-16.2			2.4	-5.7	-1.2
CBS(aD-aT) <sup>f</sup>	-14.2			-12.7			-20.9			-16.4			1.5	-6.7	-2.2
CBS(T-Q) <sup>f</sup>	-13.9			-12.1			-20.2			-15.7			1.8	-6.3	-1.8
CBS(D-T-Q) <sup>g</sup>	-13.3			-12.1			-20.1			-15.6			1.2	-6.8	-2.3
CBS(aT-aQ) <sup>f</sup>	-13.3			-11.4			-19.6			-15.0			1.9	-6.3	-1.7
CBS(aD-aT-aQ) <sup>g</sup>	-13.1			-11.3			-19.2			-14.8			1.8	-6.1	-1.7

blocked dimers	$\alpha\text{HH}'$			$A\beta\beta\text{-C}_5'$			$A\beta\beta\text{-C}_7'$			$P\beta\beta$			$\Delta\Delta E_{uc} = \Delta E_{uc} - \Delta E_{uc}(\alpha\text{HH}')$		
	$\Delta E_{uc}$	$\Delta E_{bsse}$	$\Delta E_c$	$\Delta E_{uc}$	$\Delta E_{bsse}$	$\Delta E_c$	$\Delta E_{uc}$	$\Delta E_{bsse}$	$\Delta E_c$	$\Delta E_{uc}$	$\Delta E_{bsse}$	$\Delta E_c$	$A\beta\beta\text{-C}_5'^b$	$A\beta\beta\text{-C}_7'^c$	$P\beta\beta^d$
MP2/6-311+G**	-16.3	4.7	-11.5	-13.8	3.1	-10.7	-20.8	3.9	-16.8	-17.2	3.7	-13.5	2.5	-4.5	-0.9
MP2/cc-pVDZ	-17.9	8.0	-9.9	-16.4	7.0	-9.4	-23.1	7.6	-15.5	-19.5	7.3	-12.2	1.5	-5.2	-1.6
MP2/cc-pVTZ	-16.3	3.4	-12.9	-14.2	2.9	-11.3	-21.5	3.2	-18.3	-17.7	3.1	-14.6	2.1	-5.2	-1.4
MP2/cc-pVQZ	-15.5	1.4	-14.1	-13.6	1.2	-12.4	-21.0	1.4	-19.6	-17.1	1.3	-15.7	1.9	-5.5	-1.6
MP2/aug-cc-pVDZ	-18.8	5.2	-13.5	-16.1	4.1	-12.0	-23.4	4.5	-18.9	-19.5	4.2	-15.2	2.7	-4.6	-0.7
MP2/aug-cc-pVTZ	-16.9			-14.8			-22.3			-18.3			2.1	-5.4	-1.4
CBS(D-T) <sup>f</sup>	-16.5			-13.9			-21.7			-17.7			2.6	-5.2	-1.2
CBS(aD-aT) <sup>f</sup>	-15.8			-14.2			-21.7			-17.7			1.6	-5.9	-1.9
CBS(T-Q) <sup>f</sup>	-15.0			-13.2			-20.8			-16.8			1.8	-5.8	-1.8
CBS(D-T-Q) <sup>g</sup>	-14.8			-13.4			-20.7			-16.7			1.4	-5.9	-1.9

<sup>a</sup> All geometries were optimized at MP2/6-311+G\*\*. <sup>b</sup>  $\Delta E_{uc}(A\beta\beta\text{-C}_5) - \Delta E_{uc}(\alpha\text{HH})$ . <sup>c</sup>  $\Delta E_{uc}(A\beta\beta\text{-C}_7) - \Delta E_{uc}(\alpha\text{HH})$ . <sup>d</sup>  $\Delta E_{uc}(P\beta\beta) - \Delta E_{uc}(\alpha\text{HH})$ . <sup>e</sup> Extrapolated using  $\text{BSSE}(n) = \text{BSSE}(0) \exp(-\alpha n)$ . <sup>f</sup> Extrapolated according to Truhlar and co-workers.<sup>48,49</sup> <sup>g</sup> Extrapolated using exponential formula based on  $\Delta E_{uc}$  and  $\Delta E_c$  calculated at cc-pVXZ or aug-cc-pVXZ (X = D, T, and Q).<sup>50</sup>

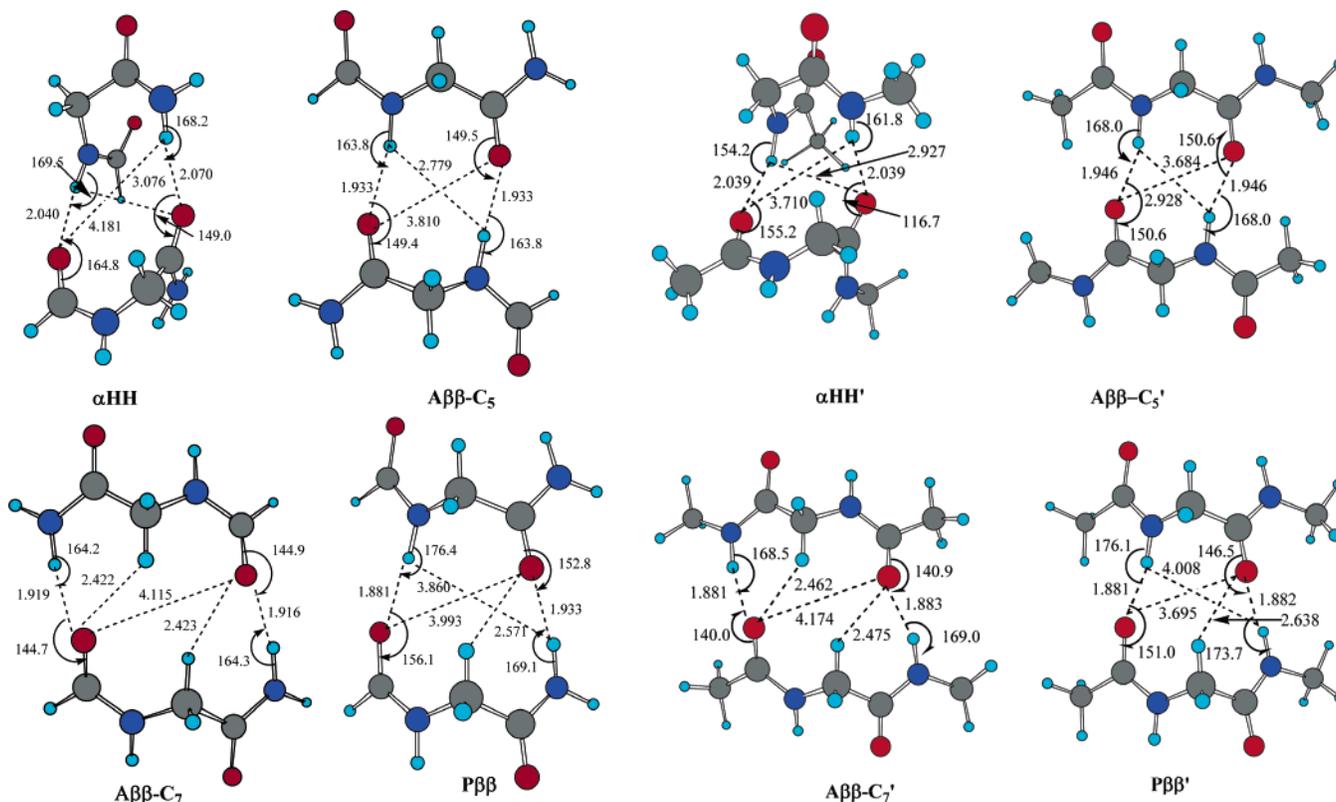
tational chemistry. On the basis of a systematic study on a set of nonbonded complexes, Rappe and Bernstein<sup>40</sup> concluded that low levels of correlation theory such as the second-order Møller–Plesset perturbation theory (MP2) can account for the full range of intermolecular interactions, and the accuracy mainly lies in the convergence with respect to the basis set expansion. In comparison, the DFT method is less reliable because of the lack of an appropriate description of the dispersion effect. According to Rappe and Bernstein<sup>40</sup> and in consideration of the size of the model complexes (18 heavy atoms and 10 hydrogen atoms for the methyl-blocked dimers) and the available computer resources, we used MP2 theory to account for the correlation energy and focused on the convergence. Because the CCSD(T) calculations even with the 6-31G\* basis set are extremely time-consuming, we decided not to account for the correlation at the CCSD(T) level. However, we note that, in the calculations of interaction energies of base pairs of nucleic acids, Hobza and co-workers<sup>35</sup> have considered the higher-order correlation at the CCSD(T)/6-31G\* level for some of their studied base pairs. They<sup>35</sup> found that the CCSD(T) corrections, ranging from 0.0 to -0.6 kcal/mol, only have a marginal effect on the relative stability of base pairs.

The geometries of the complexes were optimized at MP2/6-311+G\*\* without including basis set superposition error<sup>41,42</sup> (BSSE) correction, and the backbone ( $\Phi$ ,  $\Psi$ ) torsions

were fixed at the above angles. Interestingly, as indicated in Table 1, the MP2/6-311+G\*\* BSSE-uncorrected H-bond energies are in better agreement with the more reliable estimations than the BSSE-corrected ones. This suggests that the BSSE-uncorrected optimization at the current level might actually give better geometries than the one with the BSSE correction. We note that care should be exercised and further studies with notably higher-level optimization might help to examine this issue.

The energies were then refined by single-point calculations at the MP2/cc-pVXZ and MP2/aug-cc-pVXZ (X = D, T, and Q) levels at the MP2/6-311+G\*\* geometries. The calculations for the unblocked dimers involved up to 1672 basis functions at MP2/aug-cc-pVQZ, and those for methyl-blocked dimers involved up to 1590 basis functions at the MP2/cc-pVQZ level. The parallel Gaussian 03 package<sup>43</sup> was used to perform all ab initio calculations, which took about 2 months on the latest SGI Altix computer with 32 Itanium-2 CPUs, 128 GB of memory, and a 2.0 TB hard disk.

The interaction energies, including BSSE-uncorrected ( $\Delta E_{uc}$ ) and BSSE-corrected ( $\Delta E_c$ ), were calculated using  $\Delta E_{uc} = E_{\text{dim}} - 2E_{\text{mon}}$  and  $\Delta E_c = E_{\text{dim}} - 2E_{\text{mon}} + \Delta E_{\text{BSSE}}$ , respectively, where  $E_{\text{dim}}$  and  $E_{\text{mon}}$  are the total energies of dimers and the isolated monomers. The BSSE correction energies ( $\Delta E_{\text{BSSE}}$ ) were computed using the standard counterpoise (CP) method<sup>41,42</sup> at the dimer geometries. Because



**Figure 2.** MP2/6-311+G\*\* geometries of unblocked and methyl-blocked dimers (angles are in degrees and bond lengths are in angstroms). The C<sub>5</sub> and C<sub>7</sub> conformations refer to the two patterns of H-bond pairs in antiparallel  $\beta$  sheets (see Figure 1 caption for explanation). The atoms are colored as red (O), blue (N), gray (C), and cyan (H).

the interactions are dominated by H-bonds in these complexes, we will refer to the interaction energies as the H-bond energies.

In the present study, we focus on the interaction energies of the H-bonds in peptides with patterns similar to those in the protein secondary structures, rather than their relative stabilities of the dimers and monomers. In addition, because all of the dimers and monomers are not minima, the zero-point energies are not meaningful and are excluded in our defined binding energies. The structures and energies of alanine and glycine dipeptide monomers have been extensively studied previously.<sup>44–47</sup>

The HF and MP2 single-point energies were further extrapolated to the complete basis set (CBS) energies following Truhlar's scheme.<sup>48</sup> The HF and correlation CBS energies were estimated using  $E_{\text{CBS}}^{\text{HF}} = E^{\text{HF}}(n) - A^{\text{HF}}n^{-\alpha}$  and  $E_{\text{CBS}}^{\text{corr}} = E^{\text{corr}}(n) - A^{\text{corr}}n^{-\beta}$ , respectively, where  $n$  represents the highest angular momentum in the basis set, that is,  $n = 2, 3,$  and  $4$  corresponding to D, T, and Q basis sets, respectively, and  $\alpha = 3.39$  and  $\beta = 1.91$ , which were optimized by Truhlar and co-workers.<sup>49</sup> These energies are termed as CBS(X-Y), where X and Y can be D, T, and Q, representing the basis sets used in MP2 calculations (or aD, aT, and aQ for the augmented basis sets).

Following Dixon and co-workers,<sup>50</sup> the interaction energies were alternatively extrapolated to the complete basis set limit by the exponential relation,  $\Delta E_{\text{CBS}} = \Delta E(n) - A \exp(-Bn)$ , where  $n$  is the same as in Truhlar and co-workers' formula and  $\Delta E_{\text{CBS}}$  is the extrapolated interaction energies. The constants,  $A$  and  $B$ , were determined by the least-square fit

to six energy points including both BSSE-corrected and BSSE-uncorrected binding energies. Because these extrapolations involve energies calculated at three levels of basis sets, they are referred to as the three-basis-set exponential extrapolations and are noted as either CBS(D-T-Q) or CBS(aD-aT-aQ) for the unaugmented and augmented basis sets, respectively.

The MM energies were calculated using AMBER,<sup>51–54</sup> CHARMM,<sup>55,56</sup> and OPLS force fields.<sup>57,58</sup> The AMBER MM calculations were carried out by the AMBER 8.0 package,<sup>59</sup> and calculations with CHARMM and OPLS force fields were done with the TINKER program.<sup>60</sup> In all MM energy calculations, the backbone ( $\Phi, \Psi$ ) torsion angles were restrained to the corresponding values in the quantum mechanical (QM) calculations, and other geometrical parameters were optimized.

### 3. Results and Discussion

The MP2/6-311+G\*\* optimized structures of the complexes, together with the key geometrical parameters, are displayed in Figure 2. Table 1 compiles the energetic results, including the BSSE-corrected ( $\Delta E_c$ ) and BSSE-uncorrected binding energies ( $\Delta E_{\text{uc}}$ ) and the BSSE corrections ( $\Delta E_{\text{BSSE}}$ ). The extrapolated binding energies using Truhlar's<sup>48</sup> method are termed as CBS(X-Y), and those using the three-basis-set exponential extrapolations<sup>50</sup> are noted as CBS(X-Y-Z) (see notes in "Methods" for explanation). Also listed in Table 1 are the binding energy differences at various levels relative to the  $\alpha$ -helical dimers.

As shown in Table 1, the extrapolated CBS H-bond energies for the unblocked dimers generally decrease with the increasing basis set functions, with CBS(D-T) being the largest and CBS(aD-aT-aQ) the smallest. The two best extrapolated energy sets, CBS(aT-aQ) and CBS(aD-aT-aQ), are very close to each other, and the average differences are less than 0.3 kcal/mol. This holds true for the CBS(T-Q) and CBS(D-T-Q) energies of the methyl-blocked dimers. Although, with our available computer power, we were unable to perform MP2/aug-cc-pVQZ calculations for the methyl-blocked dimers, the trends shown in Table 1 indicate that the binding energies reached satisfactory convergence at the current level and the uncertainties are below the room-temperature thermal energies. On the basis of convergence and consistency, we consider the three-basis-set extrapolated interaction energies to be the most reliable values.

The binding energy differences relative to the helical dimers (last three columns in Table 1) are reasonably consistent at various basis sets, although the individual binding energies differ notably from the more reliable calculations. A comparison of the binding energies at the various levels with the best estimates suggests that a reasonable assessment of the relative strengths of the H-bonds could be achieved at the MP2/cc-pVTZ level. However, reliable absolute binding energies require large basis sets and extrapolation, as indicated by the trend shown in Table 1.

At the CBS(aD-aT-aQ) level, relative to  $\alpha\text{HH}$ , the binding in  $A\beta\beta\text{-C}_5$  is 1.8 kcal/mol weaker, in  $A\beta\beta\text{-C}_7$  is 6.1 kcal/mol stronger, and in  $P\beta\beta$  is 1.7 kcal/mol stronger. The average binding energy (15.3 kcal/mol) of the two antiparallel  $\beta$ -sheet conformations,  $A\beta\beta\text{-C}_5$  and  $A\beta\beta\text{-C}_7$ , which coexist in the antiparallel  $\beta$  sheet, is 2.2 kcal/mol larger than that of  $\alpha\text{HH}$ . Consistently, for the blocked dimers, relative to  $\alpha\text{HH}'$ , the binding in  $A\beta\beta\text{-C}_5'$  is 1.4 kcal/mol weaker at the CBS(D-T-Q) level, in  $A\beta\beta\text{-C}_7'$  is 5.9 kcal/mol stronger, and in  $P\beta\beta$  is 1.9 kcal/mol stronger. The average (17.1 kcal/mol) of  $A\beta\beta\text{-C}_5'$  and  $A\beta\beta\text{-C}_7'$  is 2.3 kcal/mol larger. The energetic results of both sets of model complexes indicate that the interactions in the helices are weaker than those in parallel and antiparallel  $\beta$  sheets and the interactions in the antiparallel  $\beta$  sheet is comparable to that in the parallel  $\beta$  sheets. In comparison to the unlocked dimers, the presence of the blocking groups in the methyl-blocked dipeptide dimers strengthens the H-bonds by 1.5–2.1 kcal/mol.

Other consistent trends include the systematic decrease of BSSE-uncorrected MP2 binding energies ( $\Delta E_{\text{uc}}$ ) as the basis sets expanded from DZ to TZ to QZ and, conversely, the systematic increase of the BSSE-corrected binding energies ( $\Delta E_{\text{c}}$ ). The opposite convergence trends of  $\Delta E_{\text{c}}$  and  $\Delta E_{\text{uc}}$  are indicative of the overestimation of BSSE by the CP method,<sup>41,42</sup> particularly with the small basis sets. For the unblocked dimers, the MP2/aug-cc-pVQZ binding energies with the extrapolated BSSEs are very close to the final convergent CBS(aD-aT-aQ) values, but the uncorrected ones are overestimated.

**3.1. Interaction Energies in H-Bond Pairs.** The binding energies of  $\alpha\text{HH}$  at CBS(aD-aT-aQ) and CBS(D-T-Q) are  $-13.1$  and  $-13.3$  kcal/mol, respectively, which are slightly less than the CBS(aT-aQ) and CBS(T-Q) values,  $-13.3$  and

$-13.9$  kcal/mol, respectively. However, the CBS(D-T) and CBS(aD-aT) values,  $-15.0$  and  $-14.2$  kcal/mol, respectively, are less reliable, which may be due to the less superior DZ basis set. Relative to the most sophisticated CBS(aD-aT-aQ) energy, all BSSE-uncorrected MP2 binding energies are overestimated, but the energies with large basis sets,  $-14.0$  (aug-cc-pVQZ) and  $-13.9$  kcal/mol (cc-pVQZ), are in reasonable agreement with the best extrapolated value ( $-13.1$  kcal/mol) at CBS(aD-aT-aQ). On the other hand, after BSSE corrections, the binding energies are all underestimated owing to the overestimation of BSSE error by the CP method.<sup>41,42</sup> It should be noted that the BSSE-corrected  $\Delta E_{\text{c}}$  with very large basis sets,  $-13.0$  (aug-cc-pVQZ),  $-12.8$  (aug-cc-pVTZ), and  $-12.5$  (cc-pVQZ), are quite close to the CBS(aD-aT-aQ) value, but the values with basis sets smaller than cc-pVTZ ( $-11.3$  kcal/mol) appear to be too small. In general, for basis sets larger than cc-pVTZ, the BSSE-corrected binding energies are in better agreement with the reliable extrapolated values than the uncorrected ones. For the methyl-blocked helical dimer ( $\alpha\text{HH}'$ ), the CBS(T-Q) and CBS(D-T-Q) energies are very close, being  $-15.0$  and  $-14.8$  kcal/mol, respectively. These values are between the BSSE-corrected and -uncorrected MP2/cc-pVQZ values,  $-14.1$  and  $-15.5$  kcal/mol, respectively.

The average H $\cdots$ O H-bond distances in  $\alpha\text{HH}$  and  $\alpha\text{HH}'$  are close (2.050 and 2.039 Å, respectively). However, the steric effect between the nearby methyl groups in  $\alpha\text{HH}'$  prevents the alignment of the two H-bonds from being “parallel”, which is indicated by the four H-bond angles ( $\angle\text{NHO} = 154.2^\circ$  and  $\angle\text{COH} = 155.2^\circ$  for the left-hand side and  $\angle\text{NHO} = 161.8^\circ$  and  $\angle\text{COH} = 116.7^\circ$  for the right-hand side H-bond). Although the left-hand  $\angle\text{CON}$  angle (not shown),  $148.0^\circ$ , is close to the average  $155.0^\circ$  obtained from the survey of X-ray structures in the Protein Data Bank (PDB),<sup>27</sup> the right-hand  $\angle\text{CON}$  angle (not shown),  $114.3^\circ$ , deviates significantly. In contrast, the H-bond alignment in  $\alpha\text{HH}$  is closer to that in the protein  $\alpha$  helices, and both  $\angle\text{CON}$  angles,  $162.8^\circ$  and  $152.4^\circ$ , respectively, are close to the PDB survey value ( $155.0^\circ$ ).<sup>27</sup> We note that an  $\alpha$  hydrogen in  $\alpha\text{HH}'$  approaches a nitrogen in another strand, and the distance between the two atoms is 2.58 Å. This does not occur in the unblocked helical complex  $\alpha\text{HH}$ .

The  $A\beta\beta\text{-C}_5$  ( $A\beta\beta\text{-C}_5'$ ) and  $A\beta\beta\text{-C}_7$  ( $A\beta\beta\text{-C}_7'$ ) conformations represent the two types of H-bond pairs in the antiparallel  $\beta$  sheets. All levels of calculations show that the  $C_7$  forms have much stronger binding energies than the  $C_5$  forms. The extrapolated binding energies of the two unblocked dimers at the CBS(aD-aT-aQ) level are  $-11.3$  and  $-19.2$  kcal/mol, respectively, in comparison with  $-11.4$  and  $-19.6$  kcal/mol, respectively, at the CBS(aT-aQ) level. For  $A\beta\beta\text{-C}_5'$  and  $A\beta\beta\text{-C}_7'$ , the CBS(D-T-Q) binding energies are  $-13.4$  and  $-20.7$  kcal/mol, respectively, which are almost identical to the CBS(T-Q) values of  $-13.2$  and  $-20.8$  kcal/mol.

The dimer geometries are consistent with the binding energies; the dimers with larger binding energies (i.e.,  $A\beta\beta\text{-C}_7$  and  $A\beta\beta\text{-C}_7'$ ) have shorter H $\cdots$ N H-bonds than  $A\beta\beta\text{-C}_5$  and  $A\beta\beta\text{-C}_5'$  (see Figure 2), respectively. Compared to the helical  $\alpha\text{HH}'$  dimer, the methyl groups in  $A\beta\beta\text{-C}_5'$  and  $A\beta\beta\text{-C}_7'$

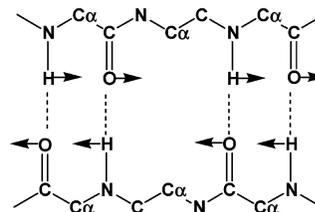
$C_7'$  have limited influence on the alignment of two H-bonds, and the H-bond angles in the two antiparallel  $\beta$  dimers are very close (see Figure 2).

The complexes similar to  $A\beta\beta-C_5'$  and  $A\beta\beta-C_7'$  have been studied previously. For the convenience of direct comparison, the following discussion about the antiparallel conformations is based on the two methyl-blocked dimers, and they can be applied to their unblocked counterparts.

The substantial binding energy difference between  $a\beta\beta-C_7'$  and  $a\beta\beta-C_5'$ ,  $-7.3$  kcal/mol at CBS(D-T-Q), has also been observed by others in spite of substantial differences in magnitude. However, the explanations have been somewhat controversial. Zhao and Wu<sup>30</sup> attributed the difference primarily to the two weak  $C-H\cdots O=C$  H-bonds in  $a\beta\beta-C_7'$  and the destabilization in  $a\beta\beta-C_5'$  due to the repulsive O/O (representing the two O atoms of CO groups in the two paired H-bonds) and H/H (representing the two H atoms of NH groups in the two paired H-bonds) secondary interactions in the H-bond pairs across the two monomers. In contrast, Dannenberg and co-workers<sup>34</sup> attributed the difference mainly to the intrastrand  $C_5 O\cdots H$  interaction and considered that the reported  $H\cdots O$  distance of the  $C-H\cdots O=C$  H-bond (2.855 Å in Zhao and Wu's work<sup>30</sup>) seems too long for such a H-bond. However, in the present MP2/6-311+G\*\* structure of  $A\beta\beta-C_7'$ , the two  $C-H\cdots O=C$  H-bond distances, 2.462 and 2.475 Å, respectively, are substantially shorter than the 2.855 Å at the HF/6-31G\* level reported previously,<sup>30</sup> strongly indicating the existence of the  $C-H\cdots O=C$  H-bonds.

Vargas et al.<sup>21,61</sup> computationally estimated the  $C-H\cdots O=C$  H-bond energy to be about 2.1 kcal/mol. Thus, the two  $C-H\cdots O=C$  H-bonds in  $A\beta\beta-C_7'$  may strengthen the binding by about 4.2 kcal/mol, which could be one of the sources of the  $\sim 7.3$  kcal/mol difference at the CBS(D-T-Q) level observed in this study. On the other hand, the intrastrand  $C_5 O\cdots H$  distance in  $A\beta\beta-C_7'$  (2.391 Å) is substantially shorter than the 2.701 Å found in  $A\beta\beta-C_5$ , suggesting the non-negligible role of the  $C_5 O\cdots H$ , although we are unable to quantify the "pure"  $C_5 O\cdots H$  interaction energy from the QM calculations. The energy difference between the two monomers in  $A\beta\beta-C_5'$  and  $A\beta\beta-C_7'$  is 0.3 kcal/mol, but the small difference does not imply that the  $C_5 O\cdots H$  interaction plays a minor role because the favorable  $C_5 O\cdots H$  interaction could be canceled by other unfavorable factors. The deformation energies (the energy difference between the monomers in the complex and the isolated one) of  $A\beta\beta-C_5'$  and  $A\beta\beta-C_7'$  are 1.2 and 0.6 kcal/mol, respectively, which contribute 0.6 kcal/mol to the difference. Dannenberg and co-workers<sup>34</sup> also studied the complexes similar to  $a\beta\beta-C_5'$  and  $a\beta\beta-C_7'$  but with a restrained  $C_s$  symmetry at the level of B3LYP/D95(D,P) + BSSE correction. The reported H-bond energies,  $-4.9$  and  $-14.0$  kcal/mol, respectively, are significantly different from our CBS(D-T-Q) values of  $-13.4$  and  $-20.7$  kcal/mol.

The difference between our results and those of others<sup>30,34</sup> could originate from two main sources. The first could be the different levels of theory used in the geometry optimizations and energy calculations. The second could arise from the symmetry restraint used by Zhao and Wu<sup>30</sup> and by



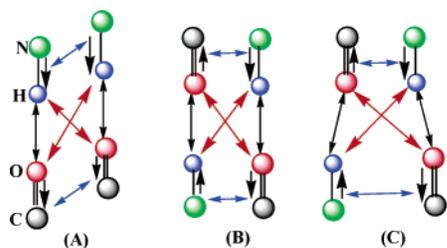
**Figure 3.** Moving tendency of atoms as the antiparallel  $\beta$  sheets are formed.

Dannenberg and co-workers.<sup>34</sup> For computational efficiency, they constrained the ( $\Phi$ ,  $\Psi$ ) main-chain torsion angles to 180.0° to maintain the  $C_s$  symmetry, which might have contributed to their too-long  $C-H\cdots O=C$  distance (2.855 Å). In comparison, the ( $\Phi$ ,  $\Psi$ ) torsion angles in our model complexes were constrained to the typical values adopted in the antiparallel  $\beta$  sheets in proteins. The overestimation of BSSE corrections and the lack of accuracy of the B3LYP/D95(D,P) method in accounting for the nonbonded interactions further contribute to the much smaller binding energies reported by Dannenberg and co-workers.<sup>34</sup>

In the methyl-blocked  $A\beta\beta-C_5'$  dimer, because the crossing O/O secondary repulsion is larger than the crossing H/H repulsion, the two carbonyl O's inevitably moved away and two amidic H's moved closer. As a consequence, the O/O distance, 3.684 Å, is larger than the 2.928 Å of the H/H distance. In contrast, because the H-bonds in  $A\beta\beta-C_7'$  are separated by four bonds (the crossing O/O distance is 4.174 Å), the secondary repulsions are weaker, allowing the carbonyl groups to move closer to one of the  $\alpha$ -hydrogen atoms in the crossing strand to form a  $C-H\cdots O=C$  H-bond. It is interesting to note that the movements in  $a\beta\beta-C_5'$  and  $a\beta\beta-C_7'$  are concerted (Figure 3), and therefore, the two patterns of H-bond pairs can be combined without introducing excess strain as they coexist in a long antiparallel  $\beta$  sheet.

The unblocked  $P\beta\beta$  complex represents the unique H-bond pattern in the parallel  $\beta$  sheets. Its extrapolated binding energies are  $-15.7$  kcal/mol [CBS(T-Q)],  $-15.6$  kcal/mol [CBS(D-T-Q)], and  $-15.0$  kcal/mol [CBS(aT-aQ)], converging to  $-14.8$  kcal/mol at the CBS(aD-aT-aQ) level. The average H-bond length in  $P\beta\beta$ , 1.907 Å, is compared with 1.918 Å in  $A\beta\beta-C_7$  and 1.933 Å in  $A\beta\beta-C_5$ . The MP2/6-311+G\*\* optimized structures apparently indicate the existence of  $C-H\cdots O=C$  H-bonds in the parallel  $\beta$  sheet. But the longer  $C-H\cdots O=C$  distances, 2.571 Å in  $P\beta\beta$  and 2.638 Å in  $P\beta\beta'$ , than the 2.423 Å and 2.422 Å in  $A\beta\beta-C_7$  and 2.462 Å and the 2.475 Å in  $A\beta\beta-C_7'$ , respectively, suggest that the  $C-H\cdots O=C$  H-bond in the parallel  $\beta$  may be weaker than those in the antiparallel  $\beta$  sheets. The ordering of the binding energies of the three  $\beta$ -sheet dimers at the CBS-(aD-aT-aQ) level,  $-11.3$  ( $A\beta\beta-C_5$ ),  $-14.8$  kcal/mol ( $A\beta\beta-C_7$ ), and  $-19.2$  kcal/mol ( $P\beta\beta$ ), is consistent with the fact that there are zero, one, and two  $C-H\cdots O=C$  H-bonds in  $A\beta\beta-C_5$ ,  $P\beta\beta$ , and  $A\beta\beta-C_7$ , respectively. The methyl-blocked  $\beta$ -sheet dimers follow the same ordering,  $-13.4$  ( $A\beta\beta-C_5'$ ),  $-16.7$  ( $P\beta\beta'$ ), and  $-20.7$  kcal/mol ( $A\beta\beta-C_7'$ ), respectively, at the CBS(D-T-Q) level.

In the study on the cooperativity of H-bonds in the  $\beta$  sheets, Zhao and Wu<sup>30</sup> have optimized the  $C_s$ -restrained



**Figure 4.** Schematic illustration of primary, secondary, and tertiary interactions in helix (A), antiparallel (B), and parallel (C)  $\beta$ -sheet dimers. The black  $\rightarrow$  lines represent the induced dipoles. The black, red, and blue  $\leftrightarrow$  lines represent the primary, secondary, and tertiary interactions, respectively.

ACE-(GLY)<sub>2</sub>-NH<sub>2</sub> dimers in which  $\Phi$  and  $\Psi$  were fixed at 180.0° to model the H-bonds in the parallel and antiparallel  $\beta$  sheets. Their HF/6-31G\* optimization led to significantly different H-bond lengths in the antiparallel and parallel  $\beta$ -sheet conformations; the two types of H-bonds in the parallel- $\beta$ -like dimer, 2.618 Å and 2.653 Å, respectively, were substantially longer than 2.150 Å in the antiparallel- $\beta$ -like dimer, leading the authors to conclude that the H-bonds in the parallel  $\beta$  sheet may be quite weaker.<sup>30</sup> In contrast, the geometries optimized at the current MP2/6-311+G\*\* level reveal that the H-bond lengths in the parallel  $\beta$  conformations are comparable to those in the C<sub>5</sub> and C<sub>7</sub> conformations of the antiparallel  $\beta$  sheets (the differences are less than 0.05 Å). It should be noted that the survey of protein crystal structures in PDB showed that the average O/N distance in the parallel  $\beta$  sheet H-bonds, 2.905 Å, is actually slightly shorter than 2.925 Å in the antiparallel  $\beta$  sheets. This is consistent with our results at the correlation level that the average O/N distance in P $\beta\beta'$ , 2.896 Å, is also slightly shorter than the average (2.918 Å) in A $\beta\beta$ -C<sub>5'</sub> and A $\beta\beta$ -C<sub>7'</sub>. The calculated binding energies show no obvious preference for the H-bond in antiparallel  $\beta$  sheets; the parallel  $\beta$  H-bonds are weaker than those in the antiparallel  $\beta$  C<sub>7</sub> forms but are stronger than those in the antiparallel C<sub>5</sub> form and are comparable to the average of the two antiparallel forms.

On the basis of the survey of the X-ray structures of proteins, Derewenda et al.<sup>62</sup> also suggested the existence of the C–H $\cdots$ O=C H-bonds in parallel  $\beta$  sheets. This is supported by the present study at the correlation level. In contrast, the earlier HF/6-31G\* study was not able to uncover such interactions in the C<sub>5</sub>-restrained ACE-(GLY)<sub>2</sub>-NH<sub>2</sub> dimers. In spite of the differences between ours and previous model complexes, the different pictures that emerged from the studies with regard to the structures and interaction energies underscore the need of high levels of theory for reliably characterizing the H-bonds.

The carbonyl and amide groups in both strands of A $\beta\beta$ -C<sub>5</sub> (or A $\beta\beta$ -C<sub>5'</sub>) are separated by two bonds. Intuitively, one might anticipate that this characteristic enhances the H-bond directionality and strength. In comparison, the carbonyl and amide groups in P $\beta\beta$  (or P $\beta\beta'$ ) are separated by four bonds in one strand and by two in the other, making them possibly misaligned. Therefore, the A $\beta\beta$ -C<sub>5</sub> (or A $\beta\beta$ -C<sub>5'</sub>) H-bonds could be conceivably stronger than the P $\beta\beta$  (or P $\beta\beta'$ )

H-bonds. This view has led to the popular belief that the perceived distortion to the H-bond geometry in the parallel  $\beta$  sheet weakens its H-bonds.<sup>63</sup> As discussed above, the current study reveals a more complex picture; the N–H $\cdots$ O=C H-bonds, C–H $\cdots$ O=C H-bonds, crossing O/O and H/H repulsions (i.e., the secondary interactions), and the aforementioned tertiary effect all contribute. Because antiparallel  $\beta$  C<sub>5</sub> and C<sub>7</sub> coexist in antiparallel  $\beta$  sheets, as an approximation, we compare the average binding energies of the two antiparallel  $\beta$  conformations with those of parallel  $\beta$  ones. At the CBS(aD-aT-aQ) level, the average energy, –15.3 kcal/mol, is only slightly larger than the –14.9 kcal/mol of P $\beta\beta$ . Similarly, the average value of –17.1 kcal/mol of A $\beta\beta$ -C<sub>5'</sub> and A $\beta\beta$ -C<sub>7'</sub> at CBS(D-T-Q) is also comparable to the –16.7 kcal/mol of P $\beta\beta$ . Although the H-bonds in C<sub>5</sub> forms have the potential to form “ideal” H-bonds, the secondary O/O and H/H repulsions prevent them from doing so. In contrast, the inherent long crossing O/O and H/H distances in parallel  $\beta$  forms let them suffer less unfavorable secondary repulsions, and the H-bonds are able to improve the linearity with minor adjustment without introducing repulsive secondary interactions; the two groups that are separated by two bonds move outward while the groups connected by four bonds move inward. As a consequence, the H-bond linearity in the parallel  $\beta$  sheet is actually better than that in the antiparallel  $\beta$  and C<sub>5</sub> forms, as indicated by the N–H $\cdots$ O H-bond angles shown in Figure 2.

**3.2. Quality of Additive Force Fields in Modeling H-Bond Pairs.** The large size of biological molecules (e.g., proteins) and the complexity of biological processes (e.g., protein folding) impose a tremendous limitation to the application of quantum-mechanics-based methods in the studies of the biological systems and processes. Molecular-mechanics-based modeling is an affordable alternative. In molecular mechanics calculations, the force fields underlie all modeling approaches and their quality is essential. In the development of protein force fields, appropriate representation of the backbone H-bonds is one of the most important concerns. While the individual backbone H-bond often exists as components of an ordered H-bond network, the typical approach in the parameter development and calibration is to compare with the single H-bond in the NMA–NMA dimer which, as mentioned above, is incapable of representing the neighboring effect between nearby H-bonds. Although the additive force fields do include the crossing O/O and H/H secondary electrostatic interactions, the suitability of this approach in describing the H-bond network in protein secondary structures has not been clarified. The results reported in this study can serve as the benchmarks to examine the existing additive force fields and reference data for future force field development.

Table 2 compares the ab initio energies of the two sets of dimers with the various empirical force fields. The partial charges for the unblocked dimers in AMBER ff94 force fields<sup>18</sup> were refitted using the same strategy as in ff94; they were fitted to the electrostatic potentials of the monomer calculated at the HF/6-31G\*\*/HF/6-31G\* level by using the restricted electrostatic potential (RESP) approach.<sup>64</sup> The other parameters, including those for bonds, angles, torsions, and

**Table 2.** H-Bond Energies (in kcal/mol) of H-Bond Pair Complexes, Their Relative Values to That of  $\alpha$  Helical Conformations, and the Contributions ( $E_{\text{pol}}$ ) Due to Polarization in the Polarizable Force Fields<sup>a</sup>

	$\Delta E(\alpha\text{HH})$	$\Delta E(A\beta\beta\text{-}C_5)$	$\Delta E(A\beta\beta\text{-}C_7)$	$\Delta E(P\beta\beta)$	$\frac{\Delta E(A\beta\beta\text{-}C_5) - \Delta E(\alpha\text{HH})}{\Delta E(\alpha\text{HH})}$	$\frac{\Delta E(A\beta\beta\text{-}C_7) - \Delta E(\alpha\text{HH})}{\Delta E(\alpha\text{HH})}$	$\frac{\Delta E(P\beta\beta) - \Delta E(\alpha\text{HH})}{\Delta E(\alpha\text{HH})}$	$\frac{\Delta E(A\beta\beta\text{-}C_7) - \Delta E(A\beta\beta\text{-}C_5)}{\Delta E(A\beta\beta\text{-}C_5)}$
QM	-13.1	-11.3	-19.2	-14.8	1.8	-6.1	-1.7	-7.9
AMBER-FF94	-16.5 (-3.4)	-12.8 (-1.5)	-17.9 (1.3)	-15.0 (-1.0)	3.8	-0.6	1.7	-5.1
OPLS_AA	-17.4 (-4.3)	-10.9 (0.4)	-18.8 (0.4)	-14.8 (0.0)	6.5	-1.4	2.6	-7.9
CHARM19-UA	-17.5 (-4.4)	-13.6 (-2.3)	-14.9 (4.3)	-14.1 (0.7)	3.9	-2.5	3.4	-1.3
FF94+Pol	-14.6 (-1.5)	-13.9 (-2.6)	-17.6 (1.6)	-15.9 (-1.1)	0.7	-3.0	-1.3	-3.7
$E_{\text{Pol}}$	1.7	-1.1	-0.2	-1.1				
QM	-14.8	-13.4	-20.7	-16.7	1.4	-5.9	-1.9	-7.3
FF94	-18.6 (-3.8)	-14.0 (-0.6)	-17.8 (2.9)	-15.7 (1.0)	4.6	-0.8	2.9	-3.8
FF03	-17.2 (-2.4)	-12.1 (1.3)	-17.2 (3.5)	-14.2 (2.5)	5.1	-0.0	3.0	-5.1
CHARM19	-19.0 (-4.2)	-15.2 (-1.8)	-16.2 (4.5)	-16.6 (0.1)	3.8	2.8	2.4	-1.0
CHARM27	-18.6 (-3.8)	-13.5 (-0.1)	-17.2 (3.5)	-15.4 (1.3)	5.1	1.4	3.2	-3.7
OPLS-AA	-18.2 (-3.4)	-12.9 (0.5)	-17.5 (3.2)	-15.2 (1.5)	5.3	0.7	3.0	-4.6
OPLS-UA	-20.9 (-6.1)	-14.7 (-1.3)	-18.7 (2.0)	-17.3 (-0.6)	6.2	2.2	3.6	-4.0
FF02 (Pol)	-17.6 (2.8)	-14.8 (-1.4)	-19.5 (1.2)	-17.5 (-0.8)	2.8	-1.9	0.1	-4.7
$E_{\text{Pol}}$	1.7	-1.0	-0.6	-1.0				
FF94+Pol	-17.3 (2.5)	-15.2 (-1.8)	-17.0 (3.7)	-16.7 (0.0)	2.1	0.3	0.6	-1.8
$E_{\text{Pol}}$	1.4	-1.2	-0.7	-1.0				
FF03+Pol	-15.7 (0.9)	-13.8 (-0.4)	-16.8 (4.1)	-15.2 (1.5)	1.9	-1.1	0.5	-3.0
$E_{\text{Pol}}$	1.0	-0.8	-0.7	-0.8				

<sup>a</sup> The differences relative to the QM energies are given in the parentheses.

Lennard-Jones, were taken from AMBER ff94, whereas the charges for the unblocked dimers in AMBER ff02 and ff03 force fields were refitted using the strategies consistent with those of ff02<sup>53</sup> and ff03.<sup>52</sup> Despite some striking agreements between the ab initio and the MM binding energies in Table 2, attention should be paid to the balance among different conformations. It is often the case that, for a given force field, good agreement with the ab initio values can be found for some dimers but not for the rest. However, in spite of the different behavior of these force fields, they all share one common feature: significantly overestimating the binding energy in the helical dimers. For example, the ab initio data shows that the binding energy of  $A\beta\beta\text{-}C_5$  is 1.8 kcal/mol stronger than that of  $\alpha\text{HH}$ . Yet, all force fields favor the helical conformation by 3.8–6.5 kcal/mol. The same holds for the methyl-blocked complexes. Given the fact that these force fields were developed by different groups on the basis of different strategies, we attribute the common feature to the inherent deficiency of the additive (point charge) molecular mechanics models.

Figure 4 schematically illustrates the major contributions to the binding energies accounted for by an additive force field. In addition to the contributions due to van der Waals interactions and deformations, the electrostatic interactions are the dominant components of the binding energies. The helical form has two primary (represented by the black double-headed lines in Figure 4) and two favorable secondary (represented the red double-headed lines in Figure 4) interactions, but the two secondary interactions in the  $\beta$  forms are unfavorable. Because both helical and  $\beta$ -sheet dimers have similar amidic H-bond donors and acceptors, the primary interactions are approximately the same. Therefore, the interactions in helical dimers tends to be stronger than that in the corresponding  $\beta$  forms if no other interactions are involved (e.g., the  $\text{C-H}\cdots\text{O}=\text{C}$  H-bond). This is in

qualitative agreement with the ab initio results; the binding energy of  $\alpha\text{HH}$  is 1.8 kcal/mol larger than that of  $A\beta\beta\text{-}C_5$ . But the substantial overestimations by force fields clearly indicate the limitation of the additive force field. We next consider the contribution due to the deformation. At the MP2/aug-cc-pVQZ level, the deformation energies of  $\alpha\text{HH}$  and  $A\beta\beta\text{-}C_5$  are 0.8 and 1.4 kcal/mol, respectively. Therefore, if the deformation contribution was excluded, the binding energy difference between the two conformations would become even smaller, which indicates server overestimation by the various force fields.

In the following, we suggest possible explanations for the significant disagreement. Because the antiparallel- $\beta$   $C_7$  and parallel- $\beta$  conformations involve  $\text{C-H}\cdots\text{O}=\text{C}$  H-bonds, which complicates the analysis, they are excluded in the following discussion.

A well-known defect in the additive force fields is the omission of the instantaneous polarization. As illustrated in Figure 4, when a H-bond forms, the donor and acceptor of the H-bond polarize their partners, making the polar  $\text{C}=\text{O}$  and  $\text{N-H}$  groups more polar in comparison to those in the monomer. In helical conformation, the enhanced polarization increases the energetically unfavorable intrastrand repulsions between the CO and CO groups in one strand and between NH and NH groups in the other (indicated by the blue double-headed lines in Figure 4). In contrast, this effect strengthens the favorable intrastrand attractions between the NH and CO in both strands of the  $A\beta\beta\text{-}C_5$ . Obviously, the neglect of the effect results in the overestimation of binding energy in  $\alpha\text{HH}$  and the underestimation of binding energy in  $A\beta\beta\text{-}C_5$ , which could be one of the sources for the large disagreement between MM and ab initio data. Following Jorgensen and Pranata,<sup>37</sup> we call this effect tertiary interaction. Due to the overestimation, the point-charge force fields also give wrong relative binding energies of  $\alpha\text{HH}$  to  $P\beta\beta$ .

Because of a favored C–H···O=C H-bond in the parallel  $\beta$  conformation, as predicted by ab initio results, the binding of the parallel  $\beta$  conformation is 1.7 kcal/mol stronger than the helical conformation  $\alpha$ HH at the CBS(aD-aT-aQ) level. In contrast, all force fields predict the former to be 1.7–3.4 kcal/mol weaker than the latter. The relative binding energies of  $A\beta\beta$ - $C_7$  to the helix form are also substantially underestimated (see the column 7).

Because the torsions were fixed in both monomers and complexes, the binding energies do not reflect the contribution of ( $\Phi$ ,  $\Psi$ ) torsion energies. The similar behavior of various additive force fields only applies to the electrostatic (H-bond) interactions and does not reflect the overall behavior of these force fields. Because ( $\Phi$ ,  $\Psi$ ) torsion energies are also important for the conformational preference, the overestimation of the binding energies of helical over  $\beta$  conformations does not necessarily imply that all additive force fields are biased to the helical conformations. Nevertheless, appropriate modeling of the H-bond pairs is necessary for accurate and balanced protein force fields.

Various nonadditive polarizable force fields have been pursued in several groups to account for the instantaneous polarization.<sup>53,54,65–73</sup> Because most of them are not publicly available, Table 2 only gives the binding energies calculated by the AMBER force fields including one specifically designed as a polarizable force field (ff02)<sup>53</sup> and those developed as fixed-charge models with ad hoc addition of the polarizability (ff94+pol, ff03+pol); the latter two are included solely for the purposes of comparison. Table 2 also gives the contributions of polarization to the binding energies.

After turning on the polarization in the additive AMBER force fields, the polarization effect systematically weakens the interactions in  $\alpha$ HH and  $\alpha$ HH' dimers and strengthens the binding in the  $\beta$ -sheet dimers. This is consistent with the positive polarization contributions in the former and negative ones in the latter. The unanimous positive contributions of polarization (ca. 1.7 kcal/mol) in the helical conformations indicate that the polarization is energetically unfavorable to the H-bond, which is consistent with the earlier reasoning that the tertiary effect is energetically unfavorable in helical conformations. In contrast, the polarization effect in all  $\beta$  forms is energetically favorable (–0.2~–1.2 kcal/mol), which is consistent with the notion that the tertiary effect enhances the binding in  $\beta$  dimers. As a consequence, the balance between the helix and the  $\beta$ -sheet conformers is notably improved in comparison with the ab initio results. For the unblocked dimers, the binding energy differences (relative to  $\alpha$ HH) of  $A\beta\beta$ - $C_5$ ,  $A\beta\beta$ - $C_7$ , and  $P\beta\beta$ , being 0.7, –3.0, and –1.3 kcal/mol, calculated by polarizable force field (ff94+pol), are compared to the ab initio values, 1.8, –6.1, and –1.7 kcal/mol, respectively, and are notably better than 3.8, –0.6, and 1.7 kcal/mol, calculated by additive force field ff94. Similar improvements can be observed among the blocked dimers after turning on the polarization in ff94 and ff03. It is noteworthy that both ff94 and ff03 were designed as the fixed-charge models (i.e., without the polarizability). Thus, the improvement is indicative of the positive roles that the polarizability plays in describing the main-chain H-bonds. Furthermore, because the two H-bonds

in antiparallel  $C_7$  forms are farther apart than those in antiparallel  $C_5$  and parallel  $\beta$  forms, the polarization contributes more in the latter than in the former. The polarization energies of the former  $C_7$  is less than that of the former  $C_5$ 's (see Table 2). The detailed comparisons between the polarizable force field (ff02) and the ab initio data also indicate that further improvements are necessary to obtain the correct absolute and relative binding energies.

Because of the favorable C–H···O=C H-bonds in the antiparallel  $\beta$   $C_7$  forms and the unfavorable crossing secondary interactions in the  $C_5$  forms, the binding energy of the  $C_7$  forms are much larger (more than 7.0 kcal/mol) than those of the  $C_5$  forms. However, except for the OPLS-AA, all other force fields, including both additive and polarizable force fields, underestimate the relative binding energies of the unblocked  $C_7$  to  $C_5$  dimers. Among many possible factors, a lack of consideration of the C–H···O=C H-bonds in parametrizations may take the main responsibility.

#### 4. Conclusions

The unblocked and methyl-blocked glycine dipeptide dimers, which were arranged to model the four patterns of backbone H-bond pairs in the protein secondary structures, have been investigated by ab initio calculations. The study provides reference structures and energetics for characterizing the protein backbone H-bonds. On the basis of the structures optimized at the MP2/6-311+G\*\* level and the energies at various high levels, the following conclusions can be drawn.

In addition to the conventionally concerned primary N–H···O=C H-bonds and the crossing secondary interactions, the C–H···O=C H-bonds and the other neighboring effect (e.g., tertiary effect) also contribute substantially. Unlike previous HF/6-31G\* optimization in which the C–H···O=C H-bond can only be observed in the antiparallel  $\beta$ -sheet-like complex, the current MP2/6-311+G\*\* optimization demonstrates that the C–H···O=C H-bonds exist in both parallel and antiparallel  $\beta$ -sheet-like conformers, which is in agreement with the PDB survey study.

The best extrapolated binding energies [CBS(aD-aT-aQ)] of unblocked dimers are –13.1, –11.3, –19.2, and –14.8 kcal/mol, and the best values [CBS(D-T-Q)] for the methyl-blocked dimers are –14.8, –13.4, –20.7, and –16.7 kcal/mol, respectively. Because the binding energies of parallel  $\beta$ -sheet conformations are only marginally weaker than the average of the two antiparallel  $\beta$ -sheet conformations, we conclude that the H-bond energies in the parallel and antiparallel  $\beta$  sheets are comparable. Consistently, the H-bond lengths in the two types of conformations are very close. This conclusion is different from the previous views, which concluded that the H-bond interaction in the parallel  $\beta$  sheet could be weaker than that in the antiparallel  $\beta$  sheets on the basis of the HF/6-31G\*\* optimization.

The secondary interactions, which are included in the additive force fields, are not able to account for the neighboring effects completely. Because other neighboring effects such as tertiary effect are also important, all additive force fields significantly overestimate the interactions in the helical conformations with respect to the  $\beta$ -sheet conformations. For instance, the energy difference between  $\alpha$ HH and

$\text{A}\beta\text{-C}_5$ , ranging from 3.8 to 6.5 kcal/mol, estimated by various force fields, is much larger than the ab initio value 1.8 kcal/mol. However, after inclusion of the polarization in the AMBER conventional force fields, the agreement with ab initio results is notably improved, which shows the promise of polarizable force fields to account for such interactions.

**Acknowledgment.** This work was supported by research grants from NIH (GM64458, GM67168, GM079383 to Y.D.). Dr. P. Cieplak provided the ff02 partial charges of unblocked glycine dipeptide.

### References

- (1) Carl, B.; Toozee, J. *Introduction to Progein Structure*, 2nd ed.; Garland Publishing, Inc: New York, 1999.
- (2) Nelson, D. L.; Cox, M. M. *Principles of Biochemistry*, 4th ed.; W. H. Freeman and Company: New York, 2005.
- (3) Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 321.
- (4) Dessent, C. E. H.; Muller-Dethlefs, K. *Chem. Rev.* **2000**, *100*, 3999.
- (5) Dykstra, C. E. *Chem. Rev.* **1993**, *93*, 2339.
- (6) Hobza, P.; Havlas, Z. *Chem. Rev.* **2000**, *100*, 4253.
- (7) Kollman, P.; Allen, L. C. *Chem. Rev.* **1972**, *72*, 283.
- (8) Muller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143.
- (9) Neusser, H. J.; Siglow, K. *Chem. Rev.* **2000**, *100*, 3921.
- (10) Schermann, J. P.; Carles, S.; Desfrancois, C. *Chem. Rev.* **2000**, *100*, 3943.
- (11) Neusser, H. J.; Siglow, K. *Chem. Rev.* **2000**, *100*, 3921.
- (12) Sherrington, D. C.; Taskinen, K. A. *Chem. Soc. Rev.* **2001**, *30*, 83.
- (13) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997.
- (14) Scheiner, S. *Hydrogen Bonding; A Theoretical Perspective*; Oxford University Press: New York, 1997.
- (15) Perczel, A.; Imre, J.; Csizmadia, I. G. *Chem.—Eur. J.* **2003**, *9*, 5332.
- (16) Perczel, A.; Gaspari, Z.; Csizmadia, I. G. *J. Comput. Chem.* **2005**, *26*, 1155–1168.
- (17) Mackerell, A. D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 10559.
- (18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (19) Langley, C. H.; Allinger, N. L. *J. Phys. Chem. A* **2003**, *107*, 5208.
- (20) Dixon, D. A.; Dobbs, K. D.; Valentini, J. J. *J. Phys. Chem.* **1994**, *98*, 13435.
- (21) Vargas, R.; Garza, J.; Friesner, R. A.; Stern, H.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105*, 4963.
- (22) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104.
- (23) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 1489.
- (24) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569.
- (25) Wang, Z.-X.; Duan, Y. *J. Theor. Comput. Chem.* **2005**, *4*, 689.
- (26) Deechongkit, S.; Dawson, P. E.; Kelly, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 16762.
- (27) Koch, O.; Bocola, M.; Klebe, G. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 310.
- (28) Lario, P. I.; Vrieling, A. *J. Am. Chem. Soc.* **2003**, *125*, 12787.
- (29) Wu, Y. D.; Zhao, Y. L. *J. Am. Chem. Soc.* **2001**, *123*, 5313.
- (30) Zhao, Y. L.; Wu, Y. D. *J. Am. Chem. Soc.* **2002**, *124*, 1570.
- (31) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124.
- (32) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 14065.
- (33) Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 10389.
- (34) Viswanathan, R.; Asensio, A.; Dannenberg, J. J. *J. Phys. Chem. A* **2004**, *108*, 9205.
- (35) Spomer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142.
- (36) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
- (37) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008.
- (38) Lukin, O.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 6775.
- (39) Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441.
- (40) Rappe, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.
- (41) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- (42) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.
- (43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (44) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L.; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989.
- (45) Iwaoka, M.; Okada, M.; Tomoda, S. *THEOCHEM* **2002**, *586*, 111.
- (46) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400.
- (47) Wang, Z. X.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699.

- (48) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, 294, 45.
- (49) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *J. Chem. Phys.* **1999**, 111, 2921.
- (50) Feyereisen, M. W.; Feller, D.; Dixon, D. A. *J. Phys. Chem.* **1996**, 100, 2993.
- (51) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, 115, 9620.
- (52) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, 24, 1999.
- (53) Cieplak, P.; Caldwell, J.; Kollman, P. A. *J. Comput. Chem.* **2001**, 22, 1048.
- (54) Wang, Z.-X.; Zhang, W.; Wu, C. H. L.; Cieplak, P.; Duan, Y. *J. Comput. Chem.* **2006**, 27, 781.
- (55) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, 4, 187.
- (56) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, 102, 3586.
- (57) Jorgensen, W. L.; Tiradorives, J. *J. Am. Chem. Soc.* **1988**, 110, 1666.
- (58) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, 118, 11225.
- (59) Case, D. A.; Darden, T. A. T. E.; Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, 2004.
- (60) Ponder, J. W. *Software Tools for Molecular Design*, 4.2 ed.; Washington University School of Medicine: St. Louis, MO, 2004.
- (61) Vargas, R.; Garza, J.; Dixon, D. A.; Hay, B. P. *J. Am. Chem. Soc.* **2000**, 122, 4750.
- (62) Derewenda, Z. S.; Lee, L.; Derewenda, U. *J. Mol. Biol.* **1995**, 252, 248.
- (63) Voet, D.; Voet, J. G. *Biochemistry*, 2nd ed.; John Wiley & Sons: New York, 1995; p 150.
- (64) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, 97, 10269.
- (65) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, 101, 6141.
- (66) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R. H.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, 103, 4730.
- (67) Stuart, S. J.; Berne, B. J. *J. Phys. Chem.* **1996**, 100, 11934.
- (68) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. *J. Comput. Chem.* **2002**, 23, 1515.
- (69) Patel, S.; Brooks, C. L. *J. Comput. Chem.* **2004**, 25, 1.
- (70) Patel, S.; Brooks, C. L. *J. Chem. Phys.* **2005**, 122, PAGE.
- (71) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. *J. Chem. Theory Comput.* **2005**.
- (72) Ren, P. Y.; Ponder, J. W. *J. Comput. Chem.* **2002**, 23, 1497.
- (73) Ren, P. Y.; Ponder, J. W. *J. Phys. Chem. B* **2003**, 107, 5933.

CT700021F