Comparison of Quantum Mechanics and Molecular Mechanics Dimerization Energy Landscapes for Pairs of Ring-Containing Amino Acids in Proteins

Alexandre V. Morozov,[†,‡] Kira M. S. Misura,[†] Kiril Tsemekhman,[§] and David Baker*[,†]

Department of Biochemistry, University of Washington, Box 357350, Seattle Washington 98195-7350, and Department of Chemistry, University of Washington, Box 351700, Seattle, Washington 98195-1700

Received: December 4, 2003; In Final Form: April 6, 2004

A promising approach to developing improved potential functions for modeling macromolecular interactions consists of combining protein structural analysis, quantum mechanical calculations on small molecule models, and molecular mechanics potential decomposition. Here we apply this approach to the interactions of pairs of ring-containing amino acids in proteins. We find reasonable qualitative agreement between molecular mechanics and quantum chemistry calculations, both over one-dimensional projections of the binding free energy landscape for amino acid homodimers and over a set of homodimers and heterodimers from experimentally observed protein crystal structures. The molecular mechanics landscapes are a sum of charge−charge and Lennard−Jones contributions; short-range quantum mechanical effects such as charge transfer appear not to be significant in ring side chain interactions. We also find a reasonable degree of correlation between the molecular mechanics energy landscapes and the distributions of dimer geometries observed in protein structures, suggesting that the intrinsic dimer interaction energies do contribute to packing of side chains in proteins rather than being overwhelmed by the numerous interactions with other protein atoms and solvent. These results demonstrate that interactions involving aromatic residues and proline can be fairly well modeled using current molecular mechanics force fields, but there is still room for improvement, particularly for interactions involving proline and tyrosine.

1. Introduction

Interactions between cyclic side chains (including phenylalanine (PHE), tyrosine (TYR), tryptophan (TRP), histidine (HIS), and proline (PRO)) play an important role in protein energetics. They contribute to the specificity of molecular recognition[1] and to secondary structure and hydrophobic core formation in protein folding.[2,3] The importance of $\pi-\pi$ interactions in biological macromolecules has attracted considerable attention to the problem of predicting dimerization energies and low energy conformations in these systems.

A benzene dimer is a simple model system for aromatic interactions that has been extensively studied in the literature by ab initio quantum mechanical (QM) methods.[4−9] Three conformations are usually distinguished in the dimer, corresponding to parallel stacked (parallel ring planes), T-shaped (perpendicular ring planes, ring plane of one monomer bisects the other), and parallel displaced (parallel but horizontally displaced ring planes) conformations. The latter two were shown to be almost isoenergetic using high level quantum chemistry methods,[7−9] whereas the parallel stacked conformation is higher in energy. It was also noted in these studies that benzene dimerization is primarily due to electrostatic and dispersion forces, and hence Hartree−Fock (HF) theory is insufficient since it describes each electron in an average field of the other electrons and thus is incapable of correctly estimating dispersion energetics which depend on explicit electron−electron correlations. Moreover, Density Functional Theory (DFT) implementations based on the local density approximation are only partially successful in accounting for dispersion interactions[10] and are therefore expected to be less accurate than the wave function correlation methods, such as the Moller−Plesset (MP) perturbation theory and the coupled cluster method with singles, doubles, and perturbative triples (CCSD(T)). However, MP and CCSD(T) methods were found to be basis set dependent; large basis sets are necessary for accurate binding energy estimates. In particular, second order perturbation theory (MP2) overestimates the effect of electron correlation when large basis sets are employed, while producing fairly accurate dimer geometries.[7,9−11]

Quantum mechanical calculations have been previously carried out for a toluene dimer which is an alternative model of the PHE side chain. In these studies, the potential energy surface (PES) was shown to be affected by the extra methyl groups, with the parallel stacked and parallel displaced conformations becoming more favorable than the T-shaped one at the MP2 level.[12,13] These papers employed molecular mechanics (MM) in addition to ab initio QM methods in order to sample the PES more effectively and to carry out binding energy calculations for aromatic dimers in solvent. A similar approach was later extended to more general aromatic dimers involving PHE, TYR, and TRP, as well as to the TRP−HIS complex.[14−17] The PES of these dimers is considerably more complex, due to the formation of classical and nonclassical hydrogen bonds,[18,19] with aromatic rings acting as hydrogen bond acceptors in the latter case. Numerous local minima have been identified for these complexes by first sampling conformational space with molecular dynamics simulations and then carrying out ab initio calculations on minimized MM structures,[14,17] and good agree-

* Corresponding author. E-mail: dabaker@u.washington.edu; phone: (206) 543-1295; fax: (206) 685-1792.

[†] Department of Biochemistry.

[‡] Present address: Center for Studies in Physics and Biology, The Rockefeller University, Box 25, 1230 York Ave., New York, NY 10021-6399.

[§] Department of Chemistry.

10.1021/jp037711e CCC: $27.50
Published on Web 05/21/2004

ment was found between the QM and MM binding energies of the minimum energy complexes. Alagona et al.[15,16] studied the energetics of TRP−HIS pairs from the Protein Data Bank (PDB) and produced one-dimensional projections of the TRP−HIS ab initio PES by varying the distance between monomer centroids and keeping all other degrees of freedom fixed. In addition, MM calculations have been carried out for several aromatic dimers in a vacuum and in various solvents, and the resulting potentials of mean force have been compared with experimental data.[20−22] The geometries of aromatic dimers in protein structures were investigated in refs 4, 23−26. Systematic deviations from random distributions of geometric degrees of freedom characterizing mutual orientation of aromatic side chains were observed, and in refs 23 and 26, comparisons were made for the benzene dimer with calculations based on empirical MM potentials. Finally, introduction of off-atom partial charges for a better description of π orbitals in aromatic molecules was discussed in refs 23 and 27.

The overall focus of our research is to develop improved energy functions which can be used for accurate description of biological macromolecules. For this purpose, we use QM calculations on small molecule model systems (representative of different aspects of protein energetics) in order to check the accuracy of MM force fields. We also test the ability of QM and MM methods to reproduce experimentally observed features of proteins, such as relative orientations of ring side chains in protein crystal structures, or hydrogen bonding geometries. In this combined QM, MM, and protein structure analysis approach, full energy landscapes are more informative than identification of minimum energy structures, since not all individual side chain-side chain interactions are necessarily optimized in proteins. We recently found that in the hydrogen bonding case, MM force fields do not produce accurate binding energy landscapes.[28] On the other hand, QM calculations successfully reproduced experimentally observed hydrogen bonding geometries. The discrepancy between QM and MM results is attributed to the inability of force fields based on fixed atom-centered partial charges to reproduce effects related to the partially covalent character of hydrogen bonds, such as electronic polarization.

In this paper, we apply this program to the analysis of energetics of ring-containing side chain dimers (including PHE, TYR, TRP, HIS, and PRO) in proteins. We carry out high level QM calculations and compare the binding energies to those predicted by MM force fields. In the first part of the paper, we sample PESs of five homodimers by creating one-dimensional projections of their dimerization energy landscapes. The projections are made by changing one geometric parameter characterizing relative orientation of the monomers in a dimer at a time and keeping all other degrees of freedom fixed. We use both QM and MM methods on this dimer set, which allows us to check the force field method accuracy, and to evaluate the separate contributions of Lennard−Jones and electrostatic interactions to dimerization energy landscapes of aromatic and PRO complexes. To our knowledge, this is the first study of dimerization energy landscapes for all five homodimers. In the second part of the paper, we compute QM and MM dimerization energies on a set of a few hundred homo- and heterodimers obtained from high-resolution crystal structures. This alternative approach to sampling the PES focuses on the energetics of experimentally observed dimer geometries. It is similar to that employed in ref 29 for cation−π interactions; however, here we employ quantum chemistry methods that take explicit electron−electron correlations into account. Finally, we investigate the extent to which the spatial distributions of dimers obtained from a large collection of protein crystal structures are consistent with the MM dimerization energy landscapes.

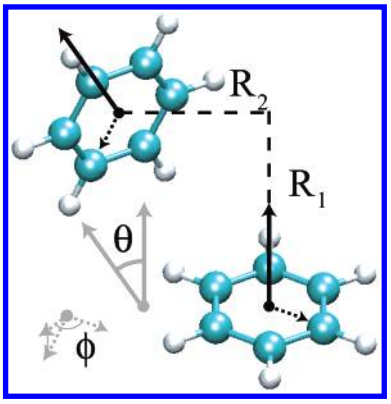

Figure 1. Schematic representation of the degrees of freedom used to describe ring dimer orientation. R_1: vertical component of the centroid−centroid separation (vertical offset); R_2: horizontal component of the centroid−centroid separation (horizontal offset); θ: angle between normals to ring planes of the monomers (interplanar angle); ϕ: angle between the projection of the in-plane vector of one monomer onto the plane of the other monomer and the in-plane vector of the other monomer. The vertical offset is determined by projecting the centroid−centroid separation onto the ring normal of one of the monomers. Ring normals are shown as solid arrows, and in-plane vectors (defined as connecting ring centroids with one of the ring atoms) are shown as dashed arrows.

2. Methods

2.1. Small Molecule Models. Small molecule models were generated for each of the ring-containing amino acids by truncating the side chain at an aliphatic carbon, which we then replace with an aliphatic hydrogen using standard bond lengths and bond angles. The carbonyl group was retained in the PRO analogue to preserve the chemical environment of the ring nitrogen. PHE is modeled as benzene, TYR is modeled as phenol, HIS is modeled as imidazole, TRP is modeled as indole, and PRO is modeled as 1-pyrrolidinecarboxaldehyde. The imidazole ring is neutral, with the hydrogen atom on the $\delta_1(\pi)$ nitrogen. These molecules are chosen over their methylated counterparts because of the smaller number of atoms and the absence of the methyl group torsional angle. Adding the methyl group would require extra optimization with respect to its rotation which would be prohibitive, and its presence makes the PES considerably more complex.[12,13]

2.2. Dimer Set Description and Geometric Degrees of Freedom. We use three different dimer sets in this work. The first set consists of five homodimers (PHE−PHE, TYR−TYR, TRP−TRP, HIS−HIS, and PRO−PRO) created for sampling one-dimensional projections of the dimerization energy landscape. Convenient degrees of freedom for describing the relative ring orientation in a dimer are the vertical (R_1) and the horizontal (R_2) offsets (displacements) of the centroids of the two rings defined by the projection onto the ring plane normal of one of the monomers, and the interplanar angle (θ) (see Figure 1). For the first dimer set, we also consider the in-plane rotation angle (ϕ) of one ring relative to the other (Figure 1); while this degree of freedom may not be informative in the case of the highly symmetric benzene ring, it provides important information for less symmetric dimers and for dimers with polar atoms which may form a hydrogen bond. For each dimer series, one of the parameters (R_1, R_2, θ, ϕ) was changed at a time while the others were fixed at the values shown in Table 1. All degrees of

TABLE 1: Ranges of Geometric Parameters for Homodimers of Ring-Containing Amino Acid Side Chains[a]

series	R_1, Å	R_2, Å	θ, deg	ϕ, deg
		PHE−PHE, TYR−TYR, TRP−TRP		
a	3.0−8.0	0.0	0.0	0.0
b	3.7	−6.0−6.0	0.0	0.0
c	5.0	0.0	0.0−180.0	0.0
d	3.7	0.0	0.0	0.0−360.0
		HIS−HIS		
a	3.0−8.0	0.0	0.0	180.0
b	3.4	−6.0−6.0	0.0	180.0
c	5.0	0.0	0.0−180.0	180.0
d	3.4	0.0	0.0	0.0−360.0
		PRO−PRO		
a	4.0−8.0	0.0	0.0	0.0
b	4.4	−6.0−6.0	0.0	0.0
c	6.0	0.0	0.0−180.0	0.0
d	5.0	0.0	0.0	0.0−360.0

[a] R_1 = vertical offset, R_2 = horizontal offset, θ = interplanar angle, ϕ = in-plane rotation angle (Figure 1).

freedom depend on the ring centroid position, which was defined as an average of all non-hydrogen atom positions in each side chain ring. Prior to creating the homodimer set, the energy of each monomer was minimized separately as described below.

The second set consists of 269 homo- and heterodimers of 15 types, which were obtained from experimentally observed nonhomologous protein structures resolved to 1.3 Å or better by X-ray crystallography. For each dimer type, the energies of 20 different conformations were computed, except for the TRP−TRP and TRP−PRO pairs with 12 conformations each due to the size of the molecules, and for the PRO−PRO pair where only five suitable conformations were observed. Purely geometric criteria were used to designate a given pair of amino acids as a dimer in this set, with a 6.0 Å cutoff in the vertical offset (R_1) and a 7.0 Å cutoff in the horizontal offset (R_2). In addition, PRO pairs adjacent in sequence were excluded since they are likely to be influenced by the mainchain geometry (the average sequence separation in the PRO−PRO dimers was about 32 residues). Five out of 269 dimers were excluded due to high MM dimerization energies (>10 kcal/mol) which indicate poor geometries.

The third set contains 46 708 homo- and heterodimers of eight types, extracted from a set of approximately 3500 protein crystal structures resolved to 2.5 Å or better and with less than 40% sequence identity to other proteins in the set. This set was used to compare the distributions of dimer geometries observed in protein structures with the MM dimerization energy landscapes. Only R_1, R_2, and θ degrees of freedom were considered for this dimer set; θ and $\pi - \theta$ interplanar angles were treated as equivalent, thereby reducing the range of θ from 0 to 180° to 0−90°. In this set, the centroid position and the ring plane were defined using four side chain atom coordinates (for TRP, only atoms from the indole ring were used; for PRO, N, CA, CB, and CD were used). In addition to the same R_1, R_2 cutoffs as in the second dimer set, a more stringent energetic cutoff was applied:[29] All dimers with the force field binding energy weaker than −1.0 kcal/mol were excluded from the set. Out of 15 dimer types, HIS−HIS, PRO−HIS, TRP−HIS, PHE−HIS, and TYR−HIS pairs were excluded because the HIS protonation state was impossible to deduce from the structural data. Furthermore, PRO−PRO and TRP−TRP pairs were excluded because fewer than 2000 dimers were found in the database, making it difficult to sample dimer geometry distributions.

In the second and third dimer sets, only carbon, nitrogen, and oxygen atomic coordinates were obtained from the PDB; hydrogen atoms were added using standard bond lengths and angles from the AMBER94 force field.[30] This creates ambiguities in two cases: HIS, for which a singly protonated state with the hydrogen atom on the $\delta_1(\pi)$ nitrogen atom was chosen, and TYR, for which the rotatable hydroxyl hydrogen bond was fixed to have the hydroxyl hydrogen−oxygen dipole in the ring plane. All dimer set geometries and energies are available from the authors upon request.

2.3. Computational Details of Energy Calculations. All electronic structure calculations were carried out using NWChem 4.5 quantum chemistry software.[31] We used MP2 perturbation theory applied to HF self-consistent field with the *aug-cc-pVDZ* basis set for all dimerization energy calculations. The counterpoise (CP) correction[32] was applied to account for the basis set superposition error. Geometries of the first (homodimer) set were constructed from (MP2/*aug-cc-pVDZ*) fully optimized monomers; geometries of the second and third dimer sets were extracted from the PDB and protonated as described above. On these dimer sets, single point CP-corrected dimerization energy calculations were carried out.

All molecular mechanics calculations were done with the TINKER 4.0 molecular modeling package[33] (http://dasher.wustl.edu/tinker/). We used CHARMM27[34] and OPLS-AA[35] force fields for MM calculations in the first and second dimer sets. In the third dimer set, CHARMM27 was used for all energy calculations. R_1, R_2, and θ were computed for each dimer from this set (ϕ had to be excluded because of the limited number of counts in the data), and the force field energies were binned as follows: R_1 into 12 bins of 0.5 Å, R_2 into 14 bins of 0.5 Å, and θ into 3 bins of 30° (corresponding to the parallel, oblique, and perpendicular dimer classes). MM dimerization energies were averaged in each bin:

$$E_{\mathrm{MM}}(i,j,k) = \frac{1}{N(i,j,k)} \sum_{m=1}^{N(i,j,k)} E^m_{\mathrm{MM}}(i,j,k) \qquad (1)$$

where (i, j, k) label a particular (R_1, R_2, θ) bin and $N(i, j, k)$ is the number of dimers in the (i, j, k) bin. $E_{\mathrm{MM}}(i, j, k)$ is compared to the bin energies inferred from experimentally observed dimer distributions, defined as follows:

$$E_{\mathrm{PDB}}(i,j,k) \sim -\log\frac{p_{\mathrm{PDB}}(i,j,k)}{\Omega(i,j,k)} \qquad (2)$$

where

$$p_{\mathrm{PDB}}(i,j,k) = \frac{N(i,j,k)}{\sum_{\mathrm{bins}} N(i,j,k)}$$

is the observed probability of being in the (i, j, k) bin, and

$$\Omega(i,j,k) = \frac{1}{2}[R^2_{2,\max}(i) - R^2_{2,\min}(i)][\cos\theta_{\min}(k) - \cos\theta_{\max}(k)]$$

corresponds to the phase volume element in the (R_1,R_2,θ) coordinates: $\int_{bin} R_2 \sin\theta \mathrm{d}R_1 \mathrm{d}R_2 \mathrm{d}\theta$. $\theta_{\min}(k)$, $\theta_{\max}(k)$ are the minimum and maximum angles in the angular bin k, and $R_{2,\min}$, $R_{2,\max}$ are the minimum and maximum values of R_2 in the horizontal displacement bin i.

3. Results and Discussion

Our main objectives in this paper are to evaluate the accuracy of MM energy landscapes for ring-containing amino acids by comparison with MP2 QM landscapes and to assess the extent

to which geometric distributions observed in protein structures reflect intrinsic (vacuum) dimerization energies.

3.1. Homodimer Dimerization Energy Landscapes. In this section, we compare MP2 and HF dimerization energy landscapes with the landscapes given by two empirical force fields widely used in structural biology: CHARMM27[34] and OPLS-AA.[35] Since dimerization energy landscapes are multidimensional, it is impossible to adequately sample them using high level ab initio methods; therefore, one-dimensional PES projections become necessary. Starting from the superimposed homodimer configuration, we create series of dimers by changing one degree of freedom (R_1, R_2, θ, or ϕ; Figure 1) at a time. While not necessarily intersecting the PES minima in a given homodimer, these series can be used to consistently compare different methods for computing binding energies, and to identify the contributions essential in forming dimerization energy landscapes. This test is more comprehensive than comparing binding energies for a few minimized dimer conformations.

MP2/*aug-cc-pVDZ* level ab initio calculations were used as a compromise between computational efficiency and accuracy.[7,9−11] Indeed, benzene dimer binding energies computed at the MP2/*aug-cc-pVDZ* level of theory were shown in ref 9 to be just 1.09, 0.42, and 1.5 kcal/mol lower than the estimated complete basis set limit binding energies (corrected for electron correlation using the CCSD(T) method) for the parallel stacked, T-shaped, and parallel displaced conformations, respectively. Moreover, we expect the difference in dimerization energies of two alternative molecule conformations computed using the *aug-cc-pVDZ* basis set to be sufficiently accurate when CP-corrected, even if the absolute dimerization energy is slightly off because of the basis set limitations.

Figure 2 shows that in general, force field landscapes (shown in red/solid and green/short dashes) reproduce MP2 landscapes (shown in blue/long dashes) reasonably well. The general features of the landscapes, and in particular the locations of minima and maxima in a given one-dimensional projection, are similar; however, the absolute values of force field energies are not quantitatively accurate, with energy differences in excess of 2 kcal/mol in some cases, notably in the R_2 dependence of the PHE homodimer and the ϕ and θ dependence of the PRO homodimer (Figure 2). The MP2 calculations do somewhat overestimate electron correlations (as manifested by the difference between MP and CCSD(T) results discussed above), but the pronounced underestimation of dimerization energies for the aromatic dimers by the MM potentials is greater than the expected error in the MP2 calculations. For example, the energy difference between the parallel stacked and parallel displaced conformations of the benzene dimer is 0.97 kcal/mol in the basis set limit CCSD(T) binding energies, 1.38 kcal/mol in the MP2/*aug-cc-pVDZ* calculations,[9] and 0.33 kcal/mol in the MM landscape shown in Figure 2.

We observe discrepancies between MP2 and HF ab initio landscapes which are greater on average than the differences between force field and MP2 ab initio landscapes: HF captures electrostatic interactions and atomic repulsion but is unable to take dispersion interactions into account and thus often grossly underpredicts binding energies. For example, Figure 2 shows that binding is unfavorable at the HF level for the PHE−PHE pair (cyan/long dashes and dots). To investigate this question further, we plotted charge−charge (magenta/dots) and Lennard−Jones (black/short dashes and dots) contributions to the PES using CHARMM27 force field (OPLS-AA results are very similar; data not shown). The sum of these two contributions forms the MM energy landscape; short-range force field components such as bond stretching and angle bending do not contribute to the dimerization PES. Consistent with the failure of HF theory to account for dispersion interactions, the HF results are quite similar to the force field charge−charge (Coulomb) interaction energies. Clearly, charge−charge interactions alone are not sufficient to predict binding energies in ring dimers, and attractive van der Waals interactions are crucial in creating accurate local minima positions.[27] Short-range QM effects not modeled by force fields, such as charge transfer and polarization, are probably not important in protein interactions involving aromatic residues and PRO. The difference between HF and charge−charge MM energies is attributable to the hard-core atomic repulsion which is captured only by the former. For example, the peak in the HF energy that occurs around $\theta = 90°$ in PHE is due to the atomic repulsion, which is clearly seen in the Lennard−Jones interaction energy in the PHE−PHE θ panel of Figure 2.

It is instructive to consider how the electrostatic and Lennard−Jones components of the force fields (based on atom-centered partial charges and van der Waals parameters) combine to qualitatively reproduce the MP2/*aug-cc-pVDZ* landscape. In more sophisticated models of $\pi-\pi$ and cation$-\pi$ interactions, negative partial charges are placed in the center of π-electron clouds (π charges), and positive partial charges are placed at the nuclei (σ charges).[23,27,36] Van der Waals interactions tend to favor a maximally superimposed (sandwich) configuration, where favorable dispersion interactions are maximized. However, the sandwich conformation is not optimal for charge−charge interactions, which instead favor the T-shaped and parallel displaced structures, with the aromatic protons located close to the π electrons.[27] For example, for the PHE homodimer, the charge−charge R_2 curves exhibit shallow minima at about $R_2 = \pm 4.5$ Å. The location of these minima corresponds to the major offset of the π-stacked structure which is necessary for the π-electron−proton attraction to overcome $\pi-\pi$ repulsion. Adding electrostatic interactions to dispersion interactions, which favor fully stacked monomers (i.e. with $R_2 = 0$), shifts the energy minima to about 2.4 Å, much closer to the MP2 dimerization energy minima (Figure 2). The parallel stacked PHE conformations would not dimerize at all without attractive van der Waals interactions, as can be seen from the charge−charge R_1 curve in Figure 2. The T-shaped minimum of the electrostatic energy corresponds to the other conformation in which π-electron−proton electrostatic interactions dominate, as can be seen from the plot of charge−charge interaction energy vs θ for PHE homodimer (Figure 2).

The MP2 and MM landscapes of the other aromatic ring dimers also exhibit parallel stacked, T-shaped, and parallel displaced energy minima. The θ and ϕ curves, however, become considerably more complex than in a benzene dimer. This is due to the lower degree of symmetry around the axes of rotation which define these angles, and the presence of polar atoms in the rings. These deviations are especially striking in the HIS dimer where one of the parallel displaced conformations is significantly more favorable, and the T-shaped conformation is slightly skewed, with $\theta < 90°$. The PES complexity of aromatic dimers with polar atoms has been observed before in energy calculations on fully optimized dimer geometries, and likely reflects hydrogen bonding.[14−17,21,22] PRO is the only protein ring side chain which is not aromatic and thus has a completely different landscape structure.

The agreement between CHARMM27 and OPLS-AA force fields is very good for the PHE−PHE, TYR−TYR, and TRP−

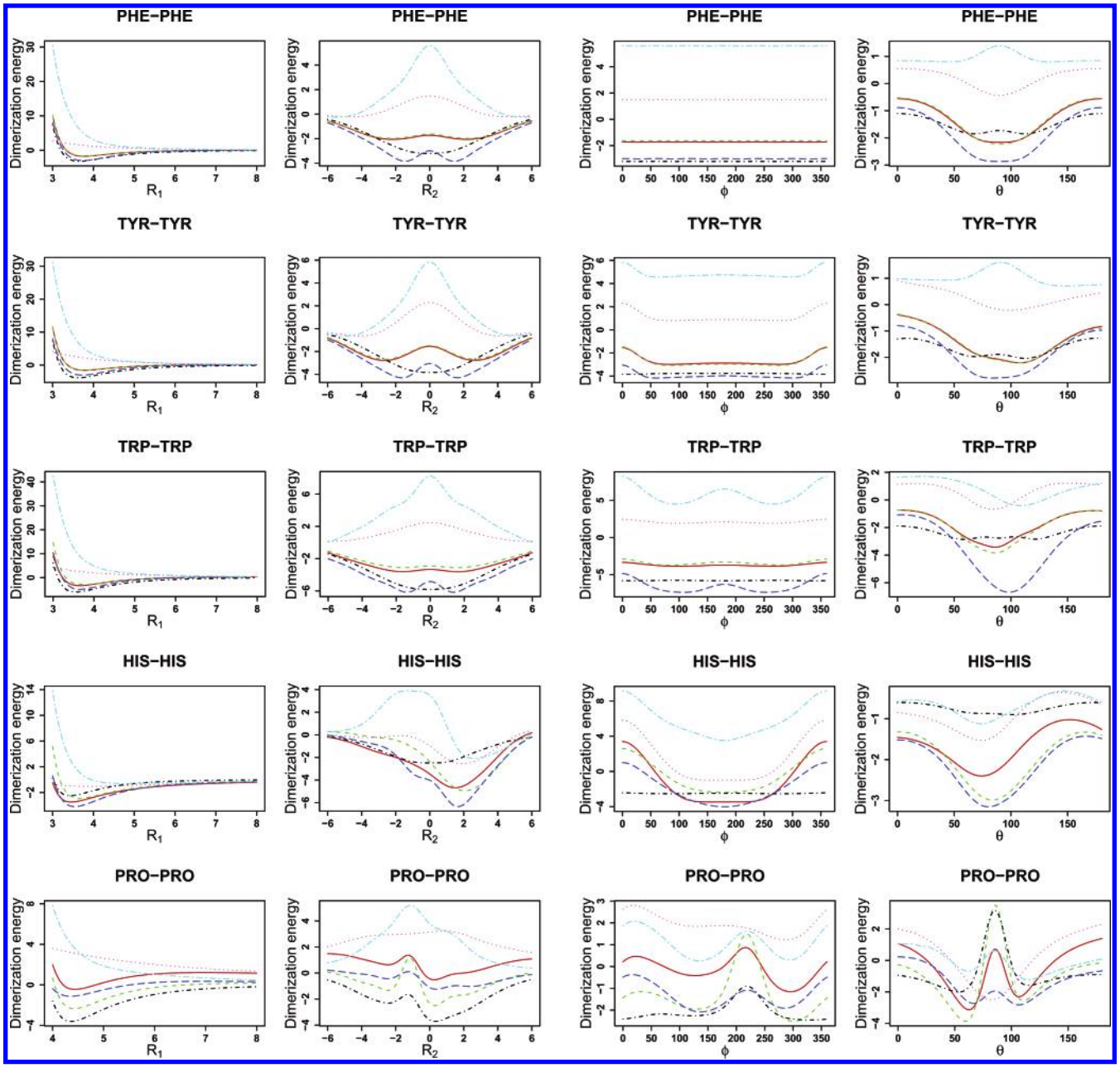

Figure 2. One-dimensional PES projections as functions of R_1 (Å), R_2 (Å), ϕ (deg), and θ (deg) (defined in Figure 1) for PHE, TYR, TRP, HIS, and PRO homodimers. All energies are in kcal/mol. Red (solid lines): CHARMM27 total dimerization energy; green (short dashes): OPLS-AA total dimerization energy; magenta (dots): charge−charge component of CHARMM27 total dimerization energy; black (short dashes and dots): Lennard−Jones component of CHARMM27 total dimerization energy; blue (long dashes): MP2 total dimerization energy; cyan (long dashes and dots): HF total dimerization energy.

TRP pairs, but there are some deviations for the HIS−HIS and PRO−PRO pairs (Figure 2; compare red/solid and green/short dashes curves). Modeling these side chains is a challenge for force fields, requiring introduction of specialized atom types. Analysis of electrostatic and Lennard−Jones contributions showed that the difference in peak heights for the two force fields in the R_2, ϕ and θ panels for the PRO−PRO dimer in Figure 2 (red/solid and green/short dashes) is mostly attributable to the difference in van der Waals parameters. Repulsive interactions are more prominent in PRO because ring pucker allows some atoms to get closer to one another than in planar aromatic rings when PES projections are made. In contrast, the discrepancies between the force fields for the HIS−HIS dimer reflect the difference in electrostatic energies (partial charges). Overall, CHARMM27 and OPLS-AA are about equally consistent with QM calculations and hence the two MM force fields appear to be equally good models for ring side chain dimers.

3.2. MM and QM Comparison Using PDB Heterodimers. An alternative approach to PES sampling is to consider dimers found in experimentally observed protein structures. Unlike the landscapes described in the previous section, these dimers occupy experimentally observed conformations found in proteins, and therefore the ability to accurately model their energetics is of special interest. The dimer set we use here was collected from high-resolution protein crystal structures as described in Methods; both negative and positive dimerization energies were considered in order to see whether the ab initio energetics of both stable and unstable dimers is well reproduced by the force fields. The upper plot of Figure 3 shows the comparison of MP2 and CHARMM27 dimerization energies.

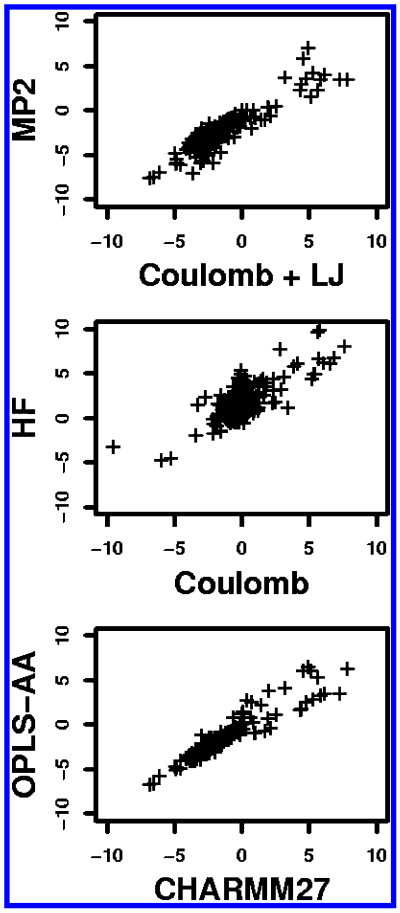

Figure 3. Dimerization energies (in kcal/mol) of a set of aromatic and PRO heterodimers (of 15 types) extracted from high-resolution protein crystal structures. Upper plot: MP2 energy vs CHARMM27 total energy (sum of charge−charge and Lennard−Jones components; Coulomb + LJ). Middle plot: HF energy vs CHARMM27 charge−charge energy (Coulomb). Lower plot: OPLS-AA total energy vs CHARMM27 total energy. Correlation coefficients are (from top to bottom): $r = 0.91$, $r = 0.77$, $r = 0.94$.

Consistent with our previous observations on the homodimer set, we see reasonable agreement between QM and MM total energies, with correlation coefficient $r = 0.91$ and the rmsd between QM and MM dimerization energies of 0.86 kcal/mol. This observation holds for all dimer types and for both positive and negative dimerization energies. When MP2 dimerization energies are plotted vs CHARMM27 charge−charge and Lennard−Jones components, the agreement is considerably worse as expected, with correlation coefficients $r = 0.77$ and $r = 0.53$, respectively (data not shown). The correlation of Lennard−Jones energies with MP2 results is particularly poor because the $1/r^{12}$ distance dependence in the former leads to sensitivity to errors in geometry. The middle plot of Figure 3 shows the extent to which HF dimerization energies are reproduced by the charge−charge component of the CHARMM27 force field. The agreement is reasonable (with $r = 0.77$), but there is more scatter than in the MP2/CHARMM27 plot. HF calculations capture both hard-core repulsion and Coulomb interactions; however, the former is missing in the charge−charge component of the force field.

There are no significant differences from the situation described above when the OPLS-AA rather than CHARMM27 force field is compared to the ab initio calculations. The correlation coefficients are 0.90 for MP2 vs OPLS-AA total, 0.77 for MP2 vs OPLS-AA charge−charge, 0.51 for MP2 vs OPLS-AA Lennard−Jones, and 0.79 for HF vs OPLS-AA charge−charge interaction energies. The results of the two force fields are closely correlated, with $r = 0.94$ for the total energies (lower plot of Figure 3), $r = 0.95$ for the charge−charge component, and $r = 0.92$ for the Lennard−Jones component (data not shown). The agreement is slightly less pronounced for repulsive dimer energies, where empirical force fields are expected to be less accurate because of the approximations made in modeling repulsive interactions and because force fields in general are parametrized to reproduce dimerization energy minima.

The tests of MM force fields carried out above lead us to believe that simple empirical models with atom-centered partial charges and standard van der Waals parameters reproduce computationally expensive ab initio QM calculations with a reasonable degree of accuracy. Therefore, force fields can be used on extensive sets of dimers, where application of QM methods is not feasible. A similar approach was employed for cation−π systems where the charge−charge component was found to be correlated with HF calculations;[29] in contrast, here we find that the total force field dimerization energy is preferable to any of its components taken separately, when compared with MP2 ab initio calculations. CHARMM27 and OPLS-AA appear to be equally good models of interactions involving aromatic side chains and PRO.

3.3. MM Dimerization Energies and Experimental Ring Side Chain Orientation Distributions. We carried out CHARMM27 calculations on an extensive set of dimers comprising about 47 000 side chain pairs of eight dimer types (see Methods). We excluded seven dimer types because of insufficient data and the uncertainty of the HIS imidazole ring protonation state, as discussed in Methods. In addition, we excluded all dimers that are unbound or only weakly bound (i.e. have energies > -1.0 kcal/mol) according to the force field total energy, since such pairs are likely to be strongly influenced by protein environment, rather than by intrinsic mutual interaction. We compare MM force field energies defined by eq 1 with the energies inferred from experimentally observed dimer distributions. In a set of systems frozen in low energy states, where the total energy is the sum of many independent contributions which are functions of some geometric parameter p, the negative logarithm of the probability of occurrence of a particular value of p is proportional to the interaction energy for that value of p.[37] A set of protein crystal structures constitutes such an ensemble to a very good approximation, and hence experimental probabilities $p_{PDB}(i, j, k)$ for a given dimer to be found in a particular (R_1, R_2, θ) bin can be related to the effective interaction energies according to the Boltzmann-like expression (eq 2).

Effective interaction energies E_{PDB} are compared with force field energies E_{MM} in Figure 4. High correlation between the two sets of energies is expected if the potentials are accurate and if the dimer interaction energies are not overwhelmed by forces involving interactions of all protein side chains, such as hydrophobic burial and steric packing in protein folding. The relatively low correlation (with average correlation coefficient of 0.44) could reflect either the protein environment influence or inaccuracies in the MM force field. It is possible in some cases to distinguish which of the two effects plays a major role in the discrepancies between MM force field predictions and experimentally observed distributions. For example, in the case of TYR, experimental side chain orientations are very similar for PHE−PHE and TYR−TYR dimers (see ref 38), which suggests that $\pi-\pi$ interactions are more important than the interactions related to the TYR hydroxyl group dipole. However,

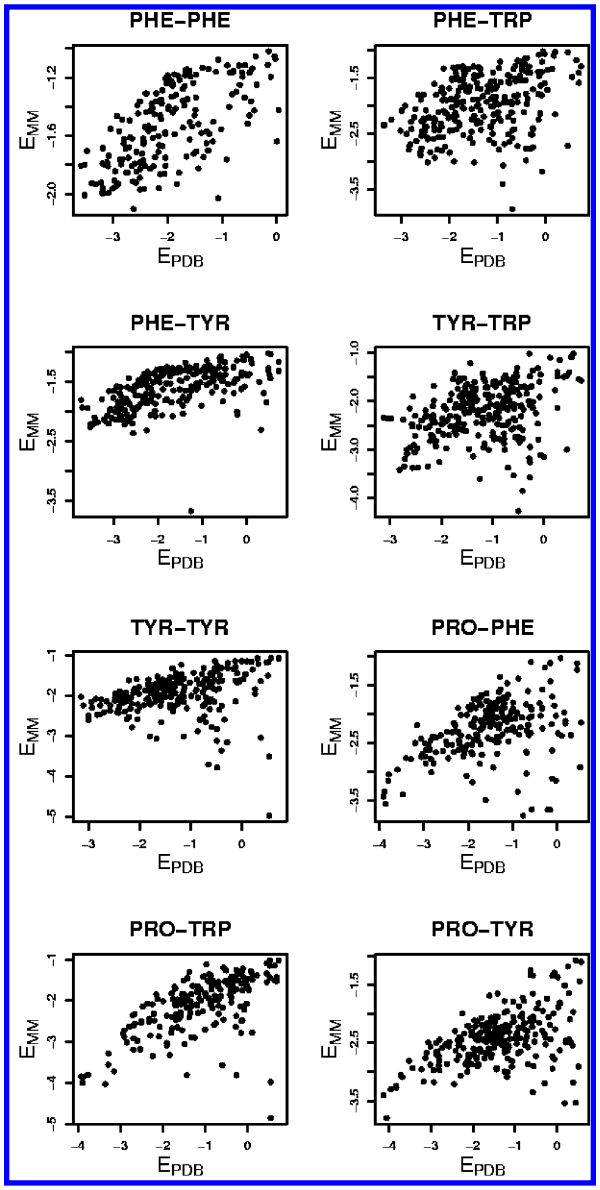

Figure 4. Correlation between energies inferred from PDB geometries using the assumption of a Boltzmann-like dimer distribution (eq 2) vs MM (charge−charge + Lennard−Jones) energies computed using CHARMM27 force field (eq 1). MM energies are in kcal/mol. Correlation coefficients are 0.63 for PHE−PHE, 0.36 for PHE−TRP, 0.56 for PHE−TYR, 0.32 for TYR−TRP, 0.21 for TYR−TYR, 0.40 for PRO−PHE, 0.51 for PRO−TRP, and 0.49 for PRO−TYR dimers. The average correlation coefficient is 0.44.

MM electrostatic interactions in the TYR homodimer are dominated by the dipole−dipole interactions of the hydroxyl group, making the MM landscape quite different from the experimentally observed one (with the correlation coefficient of 0.21). Consistent with this observation, optimizing the hydroxyl proton positions in isolated dimers, which makes the electrostatic interactions of the hydroxyl group still more dominant, makes the correlation with the observed distributions of TYR-containing dimers even worse (data not shown). The relative orientation of the hydroxyl group dipole is likely to be influenced by electrostatic and hydrogen bonding interactions with other protein atoms and cannot be obtained by MM energy minimization of the dimer alone. Nevertheless, the correlations between the geometric distributions and the MM energy landscapes are significant for many ring side chain pairs (Figure 4), suggesting that the intrinsic dimer interactions both are reasonably well modeled by the MM potentials and contribute to the arrangement of residues in proteins.

4. Conclusions

Our main conclusion in this paper is that QM dimerization energy landscapes of aromatic and PRO side chains in proteins are fairly well captured by empirical force fields, and that interactions between cyclic side chains contribute to the geometric distributions observed in protein structures. We justify these conclusions by showing first that one-dimensional projections of homodimer dimerization energy landscapes computed using MM force fields and high level ab initio QM (MP2/*aug-cc-pVDZ*) methods share the same general trends and features, although the absolute values of MM and MP2 energies are not always in quantitative agreement. Both charge−charge and Lennard−Jones interactions are important in creating qualitatively accurate landscapes; charge−charge interactions alone tend to follow HF calculations which are unable to predict well-known parallel stacked, parallel displaced, and T-shaped minima of aromatic homodimers (Figure 2). Furthermore, we carried out QM and MM calculations on a set of homo- and heterodimers found in high resolution protein crystal structures. This dimer set is complementary to the first one and allows us to check energetics of side chain pairs that typically occur in proteins. A reasonable correlation of MP2 and MM energies observed on this set suggests that the force field representation of aromatic and PRO energetics is qualitatively accurate when both charge−charge and Lennard−Jones components are taken into account, but the average error of 0.86 kcal/mol also indicates considerable room for improvement. Finally, our tests showed that both CHARMM27 and OPLS-AA force fields reproduce QM calculations with an equal degree of accuracy; however, there are systematic deviations between force field results for dimers involving HIS and PRO.

The qualitative correspondence of QM and MM energies of ring-containing protein side chains is in contrast with hydrogen bonded systems, where we have shown that MM force fields are unable to capture general trends in dimerization energy landscapes.[28] In the hydrogen bonding case, this failure was attributed to the absence of off atom charges, higher order multipoles, and electronic polarizability in current force fields. Aromatic and PRO ring interactions are less covalent in character than hydrogen bonds and thus are expected to be dominated by Lennard−Jones and Coulomb interactions. However, it is encouraging that atom-centered rather than separate π and σ charges provide a reasonable description of electrostatics in aromatic rings.

Finally, we applied the computationally efficient MM methods to an extensive dataset of ring-containing side chain dimers obtained from about 3500 experimentally observed protein structures. On this set, even HF calculations with a small basis set would not be feasible. We find that dimerization energies predicted using CHARMM27 and inferred from experimentally observed geometry distributions reproduce one another to a limited extent. The overall agreement is limited by the protein environment influence and the impossibility to reliably deduce TYR hydroxyl proton position and HIS protonation state from the structural data. However, MM force field inaccuracies are apparent in TYR-containing dimers: while the experimental geometric distributions are fairly close for PHE−PHE and TYR−TYR side chain pairs,[38] the correlation with the force fields is much worse for the TYR−TYR dimer because of the dominant contribution of the oxygen−hydrogen dipole.

Force field treatments capture the qualitative features of the interactions between cyclic amino acid side chains in proteins

and are thus superior to a purely Lennard−Jones packing-based treatment in the protein structure prediction and protein design methods being developed in our group and others. However, it should be possible to create an improved but still computationally efficient model based on the comparison between QM, MM, and experimentally observed landscapes carried out in this paper. An improved model would have a larger energy difference between parallel stacked and parallel displaced arrangements of PHE and TYR dimers consistent with the QM data, a repulsive Lennard−Jones treatment for PRO that provides a better match with the QM landscapes, and a treatment of the TYR hydroxyl group that does not distort the dimerization energy landscape significantly away from that observed in experimental protein crystal structures.

Acknowledgment. We thank Jim Havranek and Tanja Kortemme for their valuable advice throughout this project. K.T. was funded by the Division of Materials Science and Engineering, Office of Basic Energy Sciences, US Department of Energy. K.M. was supported by a fellowship from Helen Hay Whitney Foundation. A.M. and D.B. were supported by the Howard Hughes Medical Institute. The computational part of this research was performed using a grant from the William R.Wiley Environment Molecular Sciences Laboratory, located at Pacific Northwest National Laboratory.

References and Notes

(1) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210−1250.
(2) Bhattacharyya, R.; Chakrabarti, P. *J. Mol. Biol.* **2003**, *331*, 925−940.
(3) Dill, K. A. *Biochem.* **1990**, *29*, 7133−7155.
(4) Burley, S. K.; Petsko, G. A. *J. Am. Chem. Soc.* **1986**, *108*, 7995−8001.
(5) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1993**, *97*, 3937−3938.
(6) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500−3506.
(7) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790−18794.
(8) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104−112.
(9) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887−10893.
(10) Tsuzuki, S.; Lüthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949−3957.
(11) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *Chem. Phys. Lett.* **2000**, *319*, 547−554.
(12) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 11217−11224.
(13) Gervasio, F. L.; Chelli, R.; Procacci, P.; Schettino, V. *J. Phys. Chem. A* **2002**, *106*, 2945−2948.
(14) Gervasio, F. L.; Chelli, R.; Procacci, P.; Schettino, V. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 117−125.
(15) Alagona, G.; Ghio, C.; Monti, S. *J. Phys. Chem. A* **1998**, *102*, 6152−6160.
(16) Alagona, G.; Ghio, C.; Monti, S. *Int. J. Quantum Chem.* **1999**, *73*, 175−186.
(17) Gervasio, F. L.; Procacci, P.; Cardini, G.; Guarna, A.; Giolitti, A.; Schettino, V. *J. Phys. Chem. B* **2000**, *104*, 1108−1114.
(18) Burley, S. K.; Petsko, G. A. *FEBS Lett.* **1986**, *203*, 139−143.
(19) Levitt, M.; Perutz, M. F. *J. Mol. Biol.* **1988**, *201*, 751−754.
(20) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768−4774.
(21) Gervasio, F. L.; Chelli, R.; Marchi, M.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2001**, *105*, 7835−7846.
(22) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *J. Am. Chem. Soc.* **2002**, *124*, 6133−6143.
(23) Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837−846.
(24) Brocchieri, L.; Karlin, S. *Proc. Natl. Acad. Sci.* **1994**, *91*, 9297−9301.
(25) Mitchell, J. B. O.; Laskowski, R. A.; Thornton, J. M. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 370−380.
(26) McGaughey, G. B.; Gagné, M.; Rappé, A. K. *J. Biol. Chem.* **1998**, *273*, 15458−15463.
(27) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525−5534.
(28) Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. *Proc. Natl. Acad. Sci.* **2004**, *101*, 6946−6951.
(29) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci.* **1999**, *96*, 9459−9464.
(30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(31) Apra, E.; Bylaska, E. J.; de Jong, W.; Hackler, M. T.; Hirata, S.; Pollack, L.; Smith, D. et al. *NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.5*; Pacific Northwest National Laboratory: Richland, WA, 2003.
(32) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.
(33) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016−1024.
(34) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S. et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(35) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(36) Dougherty, D. A. *Science* **1996**, *271*, 163−168.
(37) Grzybowski, B. A.; Ishchenko, A. V.; DeWitte, R. S.; Whitesides, G. M.; Shakhnovich, E. I. *J. Phys. Chem. B* **2000**, *104*, 7293−7298.
(38) Misura, K. M. S.; Morozov, A. V.; Baker, D. *J. Mol. Biol.*, submitted.