Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. IV. Pairs of different hydrophobic side chains

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · OCTOBER 2008

Impact Factor: 3.3 · DOI: 10.1021/jp803896b · Source: PubMed

Published in final edited form as:
J Phys Chem B. 2008 September 11; 112(36): 11385–11395. doi:10.1021/jp803896b.

Simple Physics-Based Analytical Formulas for the Potentials of Mean Force for the Interaction of Amino Acid Side Chains in Water. IV. Pairs of Different Hydrophobic Side Chains

Mariusz Makowski[†,‡], **Emil Sobolewski**[†,§], **Cezary Czaplewski**[‡], **Stanisław Ołdziej**[‡,§], **Adam Liwo**[‡], and **Harold A. Scheraga**[*,‡]

[†] Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland [‡] Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301 [§] Intercollegiate Faculty of Biotechnology, University of Gdańsk, Medical University of Gdańsk, ul. Kładki 24, 80-822 Gdańsk, Poland

Abstract

The potentials of mean force of 21 heterodimers of the molecules modeling hydrophobic amino acid side chains: ethane (for alanine), propane (for proline), isobutane (for valine), isopentane (for leucine and isoleucine), ethylbenzene (for phenylalanine), methyl propyl sulfide (for methionine), and indole (for tryptophane) were determined by umbrella-sampling molecular dynamics simulations in explicit water as functions of distance and orientation. Analytical expressions consisting of the Gay–Berne term to represent effective van der Waals interactions and the cavity term proposed in our earlier work were fitted to the potentials of mean force. The positions and depths of the contact minima and the positions and heights of the desolvation maxima, including their dependence on the orientation of the molecules, are well represented by the analytical expressions for all systems; large deviations between the MD-determined PMF and the analytical approximations are observed for pairs involving the least spheroidal solutes: ethylbenzene, indole, and methyl propyl sulfide at short distances at which the PMF is high and, consequently, these regions are rarely visited. When data from the PMF within only 10 kcal/mol above the global minimum are considered, the standard deviation between the MD-determined and the fitted PMF is from 0.25 to 0.55 kcal/mol (the relative standard deviation being from 4% to 8%); it is larger for pairs involving nonspherical solute molecules. The free energies of contact computed from the PMF surfaces are well correlated with those determined from protein–crystal data with a slope close to that relating the free energies of transfer of amino acids (from water to *n*-octanol) to the average contact free energies determined from protein–crystal data. These observations justify future use of the determined potentials in coarse-grained protein-folding simulations.

Introduction

In our previous papers[1–3] of this series, we derived and tested a simple analytical expression for the cavity terms of the potentials of mean force of hydrophobic solutes in water, based on an analysis of the water content in different parts of the solvation sphere,[1,2] and later generalized it to spheroidal solute particles.[3] In one of these papers,[2] we showed that the derived expression for the cavity creation term reproduces the features of the cavity part of the potentials of mean forces of homo- and heterodimers of simple spherical solutes in water very well. In

* Corresponding author. has5@cornell.edu. Phone: (607) 255-4034. Fax: (607) 254-4700.

the third paper[3] of the series, we considered models of pairs of identical nonpolar amino acid side-chain models. It has been demonstrated[3] that a new side-chain–side-chain interaction potential expressed as a sum of the van der Waals term and the cavity creation term proposed previously[1] is a good candidate to replace the present knowledge-based side-chain–side-chain interaction potentials, computed from PDB data and used in our united-residue (UNRES) force field[4–11] whose other components are physics-based potentials.

This paper is a continuation of our work on physics-based potentials to express side-chain–side-chain interactions for pairs of different hydrophobic amino acid side chains. In this work, we determined the dependence of the potentials of mean force on intermolecular distance and their relative orientation by means of molecular umbrella-sampling molecular dynamics (MD) simulations for the following molecules, simulating nonpolar side chains: ethane (Et, to model the alanine side chain); propane (Prp, to model the proline side chain); isobutane (iBut, to model the valine side chain); isopentane (iPen, to model the leucine or isoleucine side chains), ethylbenzene (PhEt, to model the phenylalanine side chain), methyl propyl sulfide (MePrpS, to model the methionine side chain), and indole (Ind, to model the tryptophan side chain). Subsequently, analytical expressions were fitted to the PMFs calculated for all possible combinations between the molecules mentioned above. For proline, the C^{α} atom is considered to be part of a side chain, as in the UNRES model.[4–11]

Theory

The general expressions for the potential of mean force (PMF) of a pair of nonpolar solutes (W_{nn}) in water were derived in our previous papers.[1–3] The PMF of a pair of nonpolar solutes (W_{nn}) in water is expressed by

$$W_{nn}=E_{vdw}+\Delta F_{cav} \tag{1}$$

where E_{vdW} is the van der Waals term and ΔF_{cav} is the cavity-creation term. As in our earlier work on the UNRES model,[5] we assume that united nonpolar side chains can be represented by ellipsoids of revolution (see Figure 1); consequently, the potentials of interactions have spheroidal symmetry. The defining equations for E_{vdW} and ΔF_{cav}, which express the PMF for pairs of nonpolar solutes, were reported in [ref 3] and are given here by eqs 2–13. For the E_{vdW} energy term, we use the Gay–Berne-type potential[12] expressed by eq 2.

$$E_{vdW}=4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}^{0}}{r_{ij}-\sigma_{ij}+\sigma_{ij}^{0}}\right)^{12}-\left(\frac{\sigma_{ij}^{0}}{r_{ij}-\sigma_{ij}+\sigma_{ij}^{0}}\right)^{6}\right] \tag{2}$$

where r_{ij} is the distance between the centers of the particles, σ_{ij} is the distance corresponding to the zero value of E_{vdW} for arbitrary orientation of the particles (σ_{ij}^{0} is the distance corresponding to the zero value of E_{vdW} for the side-to-side approach of the particles), and ε_{ij} (depending on the relative orientation of the particles) is the van der Waals well depth. The dependence of ε_{ij} and σ_{ij} on the orientation of the particles is given by eqs 3–5 and eq 6, respectively.[5]

$$\varepsilon_{ij}=\varepsilon(\omega_{ij}^{(1)},\omega_{ij}^{(2)},\omega_{ij}^{(12)})=\varepsilon_{ij}^{0}\varepsilon_{ij}^{(1)}\varepsilon_{ij}^{(2)} \tag{3}$$

$$\varepsilon_{ij}^{(1)}=[1-\chi_{ij}^{(1)}\chi_{ij}^{(2)}\omega_{ij}^{(12)2}]^{-1/2} \tag{4}$$

$$\varepsilon_{ij}^{(2)}=\left[1-\frac{\chi'^{(1)}_{ij}\omega_{ij}^{(1)2}+\chi'^{(2)}_{ij}\omega_{ij}^{(2)2}-2\chi'^{(1)}_{ij}\chi'^{(2)}_{ij}\omega_{ij}^{(1)}\omega_{ij}^{(2)}\omega_{ij}^{(12)}}{1-\chi'^{(1)}_{ij}\chi'^{(2)}_{ij}\omega_{ij}^{(12)2}}\right]^2 \tag{5}$$

$$\sigma_{ij}=\sigma_{ij}^0\left[1-\frac{\chi_{ij}^{(1)}\omega_{ij}^{(1)2}+\chi_{ij}^{(2)}\omega_{ij}^{(2)2}-2\chi_{ij}^{(1)}\chi_{ij}^{(2)}\omega_{ij}^{(1)}\omega_{ij}^{(2)}\omega_{ij}^{(12)2}}{1-\chi_{ij}^{(1)}\chi_{ij}^{(2)}\omega_{ij}^{(12)2}}\right] \tag{6}$$

with

$$\omega_{ij}^{(1)}=\widehat{\mathrm{u}}_{ij}^{(1)}\cdot\widehat{\mathrm{r}}_{ij}=\cos\theta_{ij}^{(1)} \tag{8}$$

$$\omega_{ij}^{(2)}=\widehat{\mathrm{u}}_{ij}^{(2)}\cdot\widehat{\mathrm{r}}_{ij}=\cos\theta_{ij}^{(2)} \tag{9}$$

$$\omega_{ij}^{(12)}=\widehat{\mathrm{u}}_{ij}^{(1)}\cdot\widehat{\mathrm{u}}_{ij}^{(2)}=\cos\theta_{ij}^{(1)}\cos\theta_{ij}^{(2)}+\sin\theta_{ij}^{(1)}\sin\theta_{ij}^{(2)}\cos\phi_{ij} \tag{10}$$

where $\widehat{\mathrm{u}}_{ij}^{(1)}$ and $\widehat{\mathrm{u}}_{ij}^{(2)}$ are unit vectors along the principal axes of the interacting sites (in this work identified with the C^{α}–SC axes), $\hat{\mathbf{r}}_{ij}$ is the vector linking the centers of the sites, r_{ij} is the distance between the side-chain centers (Figure 1), the parameters $\chi_{ij}^{(1)}$ and $\chi_{ij}^{(2)}$ are the anisotropies of the van der Waals distance, the parameters $\chi'^{(1)}_{ij}$ and $\chi'^{(2)}_{ij}$ are the anisotropies of the van der Waals well depth, and the parameter ε_{ij}^0 is the well depth corresponding to the side-to-side orientation of the interacting particles. In this work, the parameters mentioned above were determined by least-squares fitting.

The expression for ΔF_{cav} of spheroidal particles was derived and published in paper I[1] and is given by eq 11:

$$\Delta F_{\mathrm{cav}}=\frac{\alpha_{ij}^{(1)}[(x\cdot\lambda)^{1/2}+\alpha_{ij}^{(2)}x\cdot\lambda-\alpha_{ij}^{(3)}]}{1+\alpha_{ij}^{(4)}(x\cdot\lambda)^{12}} \tag{11}$$

with

$$x=\frac{r_{ij}}{\sqrt{\sigma_i^2+\sigma_j^2}} \tag{12}$$

$$\lambda=\left[1-\frac{\chi''^{(1)}_{ij}\omega^{(1)2}_{ij}+\chi''^{(2)}_{ij}\omega^{(2)2}_{ij}-2\chi''^{(1)}_{ij}\chi''^{(2)}_{ij}\omega^{(1)}_{ij}\omega^{(2)}_{ij}\omega^{(12)}_{ij}}{1-\chi''^{(1)}_{ij}\chi''^{(2)}_{ij}\omega^{(12)2}_{ij}}\right]^2 \tag{13}$$

where the symbols $\omega^{(1)}_{ij}$, $\omega^{(2)}_{ij}$, and $\omega^{(12)}_{ij}$ are defined by eqs 8–10, r_{ij} is the distance between the centers of the particles, $\chi''^{(1)}_{ij}$ and $\chi''^{(2)}_{ij}$ are anisotropies pertaining to ΔF_{cav}, and σ_i and σ_j can be identified with the minimum distance between the center of particle *i* or *j*, respectively. The parameters $\alpha^{(1)}_{ij}$, $\alpha^{(2)}_{ij}$, and $\alpha^{(3)}_{ij}$, σ_i, σ_j, and the anisotropies are determined by least-squares fitting.

Methods

MD simulations were carried out with the AMBER[13] suite of programs, using the AMBER 7.0 force field.[14] Each pair of nonpolar molecules was placed in a periodic box containing explicit water molecules in an amount corresponding to the experimental water density at 298 K. The TIP3P water model[15] was used in the MD simulations. The box dimensions and the number of water molecules for the systems studied are given in Table 1.

The charges on the atoms of the solute molecules, needed for the AMBER 7.0 force field, were determined for each side-chain model by using a standard procedure,[16] i.e., by fitting the point-charge electrostatic potential to the molecular electrostatic potential computed by using the electronic wave function calculated at the restricted Hartree–Fock (RHF) level with the 6-31 G* basis set. The program GAMESS[17] was used to carry out the quantum-mechanical calculations, while the program RESP[15] of the AMBER 7.0 package was used to compute the fitted charges. The charges and the AMBER atom types are shown in Figure 2.

MD simulations were carried out in two steps. In the first step, each system was equilibrated in the NPT ensemble (constant number of particles, pressure, and temperature) at 298 K for 100 ps (picoseconds), and the integration step was 2 fs (femtoseconds). After equilibration, MD simulations were run in the NVT ensemble (constant number of particles, volume, and temperature) for 10 ns, and the integration step was 2 fs. A 9 Å cutoff for all nonbonded interactions, including electrostatic interactions, was imposed. For each system, a series of 20 windows of 10 ns simulations was run with different harmonic-restraint potentials imposed on the distance between the atoms closest to the center of the mass of each of the molecules, as given by

$$V=\frac{1}{2}k(r-\mathrm{r}_i^{\circ})^2 \tag{14}$$

where *k* is a force constant [assumed to be 2 kcal/(mol Å^2)], *r* is the interparticle distance, and $r_i°$ is the center of the restraint in the *i*th simulation (window); the values of $r°$ were 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, and 13 Å. A total number of 50 000 configurations were collected for each window.

To determine the potentials of mean force (PMFs) of the systems studied, we processed the results of all restrained MD simulations for each system by using the weighted histogram analysis method (WHAM).[18,19] For a given system, four-dimensional histograms in r_{ij}, $\theta_{ij}^{(1)}$, $\theta_{ij}^{(2)}$, and φ_{ij} (Figure 1) were constructed. The ranges and bin sizes were 3.5 Å $\leq r_{ij} \leq$ 13 Å, or 5 Å $\leq r_{ij} \leq$ 13 Å, depending on the system, with one bin side distance 0.2 Å, with angles 0° $\leq \theta_{ij}^{(1)} \leq$ 180° (with bin side 60°), with angles 0° $\leq \theta_{ij}^{(2)} \leq$ 180° (with bin side 60°), and with angles −180° $\leq \varphi_{ij} <$ 180° (with bin side 60°). Consequently, each distance corresponded to 54 bins, each containing counts from different orientations of the molecules. For illustration, the PMF's are plotted vs distance for the “side-to-side” (parallel), “edge-to-edge” (linear), “side-to-edge” (perpendicular), and “edge-to-side” (perpendicular) orientations; these orientations are depicted schematically in Figure 3.

Fitting of analytical formulas to the PMFs was accomplished by minimizing (Φ) of eq 15, the sum of the squares of the differences between the PMF values computed from analytical formulas (eqs 2–6 and 11–13) and from those determined from the MD simulations, by using the Marquardt method.[20]

$$\min_{\mathrm{y},C}\Phi(\mathrm{y})=\sum_i w_i \left\{\left[\exp[-\beta W^{\mathrm{MD}}(r_i,\theta_{ij}^{(1)},\theta_{ij}^{(2)},\varphi_{ij})] - \exp[-\beta W^{\mathrm{anal}}(r_i,\theta_{ij}^{(1)},\theta_{ij}^{(2)},\varphi_{ij};\mathrm{y})]\right] / \left[\exp[-\beta W^{\mathrm{MD}}\left(r_i,\theta_{ij}^{(1)},\theta_{ij}^{(2)},\varphi_{ij}\right)]\right]^2 \quad (15)$$

where $\beta = 1/RT$ with T being the absolute temperature and R the universal gas constant, respectively, $W^{\mathrm{MD}}(r_i,\theta_{ij}^{(1)},\theta_{ij}^{(2)},\varphi_{ij})$ is the PMF value determined by simulations for the distance r_{ij} and orientation $(\theta_{ij}^{(1)},\theta_{ij}^{(2)},\varphi_{ij})$; $W^{\mathrm{anal}}(r_i,\theta_{ij}^{(1)},\theta_{ij}^{(2)},\varphi_{ij};\mathrm{y})=W_{nn}(r_i,\theta_{ij}^{(1)},\theta_{ij}^{(2)},\varphi_{ij};\mathrm{y})+C$; W_{nn}, defined by eq 1 and C, being an additive constant, is the analytical approximation to the PMF at distance r_{ij} calculated with parameters given by the vector **y**, whose components plus the additive constant C are the adjustable parameters of eqs 2–6 and 11–13; $W_i=1/W^{\mathrm{MD}}(r_i,\theta_{ij}^{(1)},\theta_{ij}^{(2)},\varphi_{ij})$ is the weight of the ith point. The constant C can be interpreted as the average value of the MD-determined PMF at large distances. The weighting factor favors points with low PMF values. It should be noted that the PMF values determined from MD simulations were always positive, because they were shifted by C.

Results and Discussion

The PMFs of the pairs considered, together with curves computed with the fitted analytical expressions, are plotted as functions of the distance between the centers of the molecules in Figure 4a–u. Dashed (black), dotted (red), dot-dashed (blue), and dash-double-dotted (green) lines in Figure 4a–u refer to the “side-to-side”, “edge-to-edge”, “side-to-edge”, and “edge-to-side” orientations, respectively, and the thinner lines of a given style in Figure 4 refer to the PMFs expressed by their analytical function (eq. 1). We compared the plots corresponding to parallel and antiparallel orientations, which are equivalent in the spheroidal model. These pairs of plots are nearly overlapping even for molecules that are apparently not symmetric with respect to the axis or plane perpendicular to the C^{α}…SC axis (MePrS, Ile, Trp, etc.; data not shown). Consequently, there is no need to differentiate orientations which are equivalent in the framework of the spheroidal model.

For most systems, the PMF curves have characteristic shapes with one deep minimum. This minimum is referred to as the contact minimum. The positions of these minima depend on the

size of the system studied and the orientation of monomers within a dimer. The minima occur at the shortest distances for the side-to-side orientation, and the longest distances for the edge-to-edge orientations, respectively. The positions of the contact minima are between 4 and 8 Å depending on orientation and pair type, and their depths range from about −0.5 kcal/mol, for pairs composed of small solute molecules at the edge-to-edge orientation, to −2.25 kcal/mol for the iBut–MePrpS pair at the side-to-side orientation. It should be noted that, because the PMF's are defined in a 4-dimensional space (consisting of the distance between the centers of the solute molecules and the three angles defining their orientation), their accuracy is lower than the accuracy of those defined only as a function of the distance. The orientation dependence of the positions of the contact minima and the desolvation maxima arises because the distance from the center of a molecule to its surface varies with its orientation with respect to the other molecule.

For heterodimers with ethylbenzene and indole, wider contact-minimum regions are observed than for pairs of aliphatic solute molecules; this occurs because the two cyclic solute molecules have significantly different axes perpendicular to the C^{α}–SC axis and, consequently, cannot be represented accurately as ellipsoids of revolution about this axis. Because of the disklike shape of the cyclic solute molecules considered in this study, their contact distances corresponding to parallel orientation of the long axis of the approximating spheroid toward the other molecule will range from a short distance, at which the plane of the ring is oriented toward the other molecule, to a long distance, at which a side of the ring is oriented toward the other molecule. Because such configurations involve contacts between the atoms of the interacting molecules, they both make the most significant contribution to the PMF and result in significantly negative PMF values; this explains the increased width of the region of the contact minimum of the PMF.

The increase of the depth of the contact minimum with the size of the contact surface of the hydrocarbon molecules is consistent with the results of an earlier study by Némethy and Scheraga[21] who calculated the free energies of hydrophobic interactions between nonpolar amino acid side chains based on the free energies of transfer of hydrocarbon molecules from the organic to the aqueous phase. These authors found that the decrease of the free energy is caused by the fact that more water molecules are eliminated from the solvation sphere of larger hydrocarbon molecules upon dimer formation.

Most plots possess a desolvation maximum. This maximum is higher for pairs with smaller nonpolar particles (Figure 4a–p), and disappears for pairs of larger particles, i.e., for PhEt or Ind (Figure 4q–u). Also, in Figure 4q–u the solvent-separated minimum (the second minimum in the PMF curve) is not observed. The loss of the desolvation maximum and the solvent-separated minimum is caused by the elimination of water from the solvation sphere of the two interacting molecules.[21,22]

The results of fitting the PMFs, using the analytical expressions given by eq 1 with components defined by eqs 2 and 11, respectively, are also shown in Figure 4 (the thinner dashed (black), dotted (red), dot-dashed (blue), and dash-double-dotted (green) lines in Figure 4a–u refer to the “side-to-side”, “edge-to-edge”, “side-to-edge”, and “edge-to-side” orientations, respectively), while the best fitted parameters of the expressions for the PMF components E_{vdW} and ΔF_{cav} (eqs 2 and 11, respectively) are presented in Table 2. The absolute (σ_W and σ_{W10}) and relative (σ_{Wrel} and σ_{W10rel}) deviations of the fitted PMF from the MD-determined PMF, both for all points (σ_W and σ_{Wrel}) and for the points at which the MD-determined PMF is within 10 kcal/mol above the minimum PMF value (σ_{W10} and σ_{W10rel}) are summarized in Table 3.

As can be seen from Figure 4a–u and Table 3, the analytical expressions (eq 1 with components defined by eqs 2–12) fit the PMF curves very well for pairs of small aliphatic solutes, such as Et-iPrp (Figure 4b) and iBut-Et (Figure 4i). For these systems, the absolute errors of the PMF computed over all points are less than 1 kcal/mol and the relative errors are 5% and 6%, respectively. However, for the following pairs involving the cyclic solute molecules (PhEt and Ind) as well as MePrpS: Prp-PhEt, Prp-MePrpS, MePrpS-iPen, MePrpS-Phet, PhEt-iPen, Ind-Et, Ind-iPen, Ind-Prp, Ind-iBut, Ind-PhEt, and Ind-MePrpS, the values of the fitted PMF curves corresponding to the edge-to-edge and edge-to-side orientations increase with decreasing distances for distances of about 1–1.5 Å smaller compared to the MD-determined PMF curves (see Figure 4g,h,m–u). This observation is confirmed by outstandingly high absolute standard deviations between the fitted and MD-determined PMF's for the pairs composed of MePrpS-PhEt, Ind-iPen, Ind-Prp, Ind-iBut, Ind-PhEt, and Ind-MePrpS (Table 3). This feature, in turn, arises from the apparent inadequacy of the spheroidal model to represent the shape of these three solute molecules. The fitting procedure is directed at reproducing low-free-energy regions, which are broader than could be reproduced by the spheroidal model; see the discussion of the shapes of the PMF's earlier in this section, particularly at the edge-to-edge and edge-to-side orientation for pairs involving these three (PhEt, Ind, MePrpS) molecules, because of their nonspheroidal shape than can be represented by the model. Consequently, the positions and depths of the contact minima are represented well but these minima appear to be too narrow in the fitted potentials.

One feature of the UNRES model is that the side chains are represented as ellipsoids of revolution with their main axes being the C^{α}–SC axes, where SC denotes a side-chain center.[5] Therefore, while it is possible to extend both the Gay–Berne potential (eq 2) and the Gaussian-overlap hydrophobic-association potential (eq 11) from spheroidal to a general-ellipsoid pattern to improve the fit of the PMF's of pairs involving PhEt, Ind, and MePrpS, it would require introducing an additional degree of freedom corresponding to an ellispoid rotation, which would complicate the present equations of motion and, consequently, the molecular dynamics algorithm for the present UNRES[23] model significantly. However, as mentioned earlier, significant differences between the fitted and MD-determined PMF's for these three model compounds occur only in high-energy regions which are rarely visited. When only low-energy regions (less than 10 kcal/mol above the global minimum) are considered, only some of the absolute standard deviations slightly exceed 0.5 kcal/mol and the highest relative standard deviation is 8% for the Ind-MePrpS pair (Table 3). We can, therefore, conclude that the simulated PMF's are reproduced by the analytical expressions developed in our work (eq 1, with components defined by eqs 2 and 11) sufficiently well.

The solvent-separated minimum does not appear in any of the fitted curves; however, it is not included in our Gaussian-overlap-based solvation model.[3] Moreover, the solvent-separated minimum is not accurately determined from the simulations because of noise in the data and is so shallow that its reproduction in the side-chain–side-chain interaction potential does not seem to be crucial for constructing a mesoscopic force field.

Based on the parameters collected in Table 2, two sample plots of the E_{vdW} and ΔF_{cav} components of the fitted PMFs of Et-iPen, and Ind-PhEt are shown in Figure 5. It can be seen that ΔF_{cav} has a desolvation maximum for both systems (and also for all other systems not shown in Figure 5), although it is less pronounced in systems in which the desolvation maximum is weak in the PMF (Figure 4 q–u). The desolvation maximum shifts to greater distances when the orientation is changed from side-to-side through side-to-edge to edge-to-edge.

To compare the MD-derived PMF's with the related quantities calculated from protein crystal data which were used in our earlier work[5] to derive the side chain-side chain potentials

($U_{SC_iSC_j}$) used in the current version of UNRES, we calculated the free energies of contacts, using the PMF's determined in this work and in our preceding paper of this series,[3] and compared them with the contact free energies determined from the protein-crystal data.[5] The contact free energies were calculated using eq 16.

$$e_{ij} = -RT \ln\left\{\left(\frac{4}{3}\pi r_c^3\right)^{-1} \frac{1}{2}\int_0^{r_c}\int_0^{\pi}\int_0^{\pi}\int_0^{2\pi} \exp[-\beta(W^{MD}(r,\theta^{(1)},\theta^{(2)},\varphi) - C)] \times r^2, \mathrm{Sin}\theta^{(1)}, \mathrm{Sin}\theta^{(2)}, \mathrm{d}\theta^{(1)}, \mathrm{d}\theta^{(2)}, \mathrm{d}\varphi\right\} \quad (16)$$

where C is the additive constant obtained from fitting the analytical PMF to the MD-determined PMF (see the text below eq 15); C can be interpreted as the MD-determined PMF value at infinity; $r_c = 8$ Å is the cutoff distance and its value is the same as that used in our earlier work;[5] the factor of (1/2) appears before the integral because of the symmetry of the relative geometries of spheroids, and the quantity $(4/3)\pi r_c^3$ is the value of the integral in eq 16 (the partition function) in the absence of any direct or indirect interactions between the solute molecules and is introduced because the contact energies determined from the PDB data are relative to side chains attached to a random-flight polypeptide chain.[5] The integral was approximated by summation over bins. The calculated contact free energies and those determined from the PDB[5] data (denoted as e_{ij}^{PDB}) are summarized in Table 4 and plotted against each other in Figure 6.

It can be seen from Figure 6 that there is a good correlation between the values of e_{ij} and e_{ij}^{PDB}, indicated quantitatively by eq 17,

$$e_{ij}^{PDB} = 1.39(0.20)e_{ij} + 0.21(0.40) \qquad R = 0.812 \quad (17)$$

where R denotes the correlation coefficient, and the numbers in parentheses are the standard deviation of the slope and intercept, respectively. However, the pair iBut-MePrpS is a clear outlier, because its e_{ij} value is significantly lower than the other contact energies (Table 4). After eliminating this pair, the correlation coefficient increases remarkably (eq 18).

$$e_{ij}^{PDB} = 1.84(0.17)e_{ij} + 1.05(0.33) \qquad R = 0.908 \quad (18)$$

It should be noted that the slopes of eqs 17 and 18 (1.39 and 1.84, respectively) are comparable to the one that relates the free energies of transfer of amino acids, from water to *n*-octanol, to residue hydrophobicities calculated from the contact energies determined from the PDB in our earlier work;[5] the latter equals 1.6 (eq 23 in [ref 5]). This finding demonstrates that the PMF's and, consequently, the $U_{SC_iSC_j}$ potentials determined in this work are related to the thermodynamics of hydrophobic interactions.

Conclusions

We determined the potentials of mean force of 21 pairs of different molecules modeling hydrophobic side chains in water as functions of distance and orientation of the molecules, and fitted an approximate expression to the PMFs composed of the anisotropic Gay–Berne potential to represent van der Waals interactions and a cavity term. Both components of the approximate

PMFs depend on distance and orientation of the spheroidal particles. We demonstrated that the analytical expression introduced to approximate the free energy of the interaction of nonpolar side chains fits the PMFs reasonably well in low-PMF regions and that the contact free energies calculated from the MD-determined PMF's correspond well to experimental free energies of transfer of amino acid residues from *n*-octanol to water, suggesting that they are good candidates for the physics-based mean-field potentials of side-chain–side-chain interactions in our mesoscopic UNRES force field[4–10] for the simulation of the structure and dynamics of proteins.

Acknowledgments

This work was supported by grants from the U.S. National Institutes of Health (GM-14312), the U.S. National Science Foundation (MCB05-41633), the National Institutes of Health Fogarty International Center Grant TW007193, and the Polish Ministry of Science and Informatization (1 T09A 099 30). M.M. was supported by a grant from the "Homing" program of the Foundation for Polish Science (FNP) and MF EOG resources. This research was conducted by using the resources of (a) our 818-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (b) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, (c) our 45-processor Beowulf cluster at the Faculty of Chemistry, University of Gdansk, (d) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdansk, and (e) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) at the University of Warsaw.

References and Notes

1. Makowski M, Liwo A, Scheraga HA. J Phys Chem B 2007;111:2910. [PubMed: 17388416]
2. Makowski M, Liwo A, Maksimiak K, Makowska J, Scheraga HA. J Phys Chem B 2007;111:2917. [PubMed: 17388417]
3. Makowski M, Sobolewski E, Czaplewski C, Liwo A, Odziej S, Scheraga HA. J Phys Chem B 2007;111:2925. [PubMed: 17388418]
4. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. Protein Sci 1993;2:1697. [PubMed: 7504550]
5. Liwo A, Odziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. J Comput Chem 1997;18:849.
6. Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA. Proc Natl Acad Sci, USA 1999;96:5482. [PubMed: 10318909]
7. Lee J, Liwo A, Ripoll DR, Pillardy J, Scheraga HA. Proteins: Struct, Funct Genet 1999:204. [PubMed: 10526370]
8. Lee J, Liwo A, Ripoll DR, Pillardy J, Saunders JA, Gibson KD, Scheraga HA. Int J Quantum Chem 2000;71:90.
9. Odziej S, Kozowska U, Liwo A, Scheraga HA. J Phys Chem A 2003;107:8035.
10. Liwo A, Odziej S, Czaplewski C, Kozowska U, Scheraga HA. J Phys Chem B 2004;108:9421.
11. Czaplewski C, Liwo A, Odziej S, Scheraga HA. Polymer 2004;45:677.
12. Gay JG, Berne BJ. J Chem Phys 1981;74:3316.
13. Case DA, Pearlman DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel G, Kollman PA. Comput Phys Commun 1995;91:1.
14. Case, DA.; Pearlman, DA.; Caldwell, JW.; Cheatham, TE., III; Wang, J.; Ross, WS.; Simmerling, CL.; Darden, TA.; Merz, KM.; Stanton, RV.; Cheng, AL.; Vincent, JJ.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, RJ.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, GL.; Singh, UC.; Weiner, PK.; Kollman, PA. AMBER 7. University of California; San Francisco: 2002.
15. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J Chem Phys 1983;79:926.
16. Jorgensen WL, Briggs JM. Mol Phys 1988;63:547.
17. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JA, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA. J Comput Chem 1993;14:1347.
18. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. J Comput Chem 1992;13:1011.
19. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. J Comput Chem 1995;16:1339.
20. Marquardt DW. J Soc Ind Appl Math 1963;11:431.

21. Némethy G, Scheraga HA. J Phys Chem 1962;66:1773.
22. Huang X, Margulis CJ, Berne BJ. J Phys Chem B 2003;107:11742.
23. Khalili M, Liwo A, Rakowski F, Grochowski P, Scheraga HA. J Phys Chem B 2005;109:13785. [PubMed: 16852727]

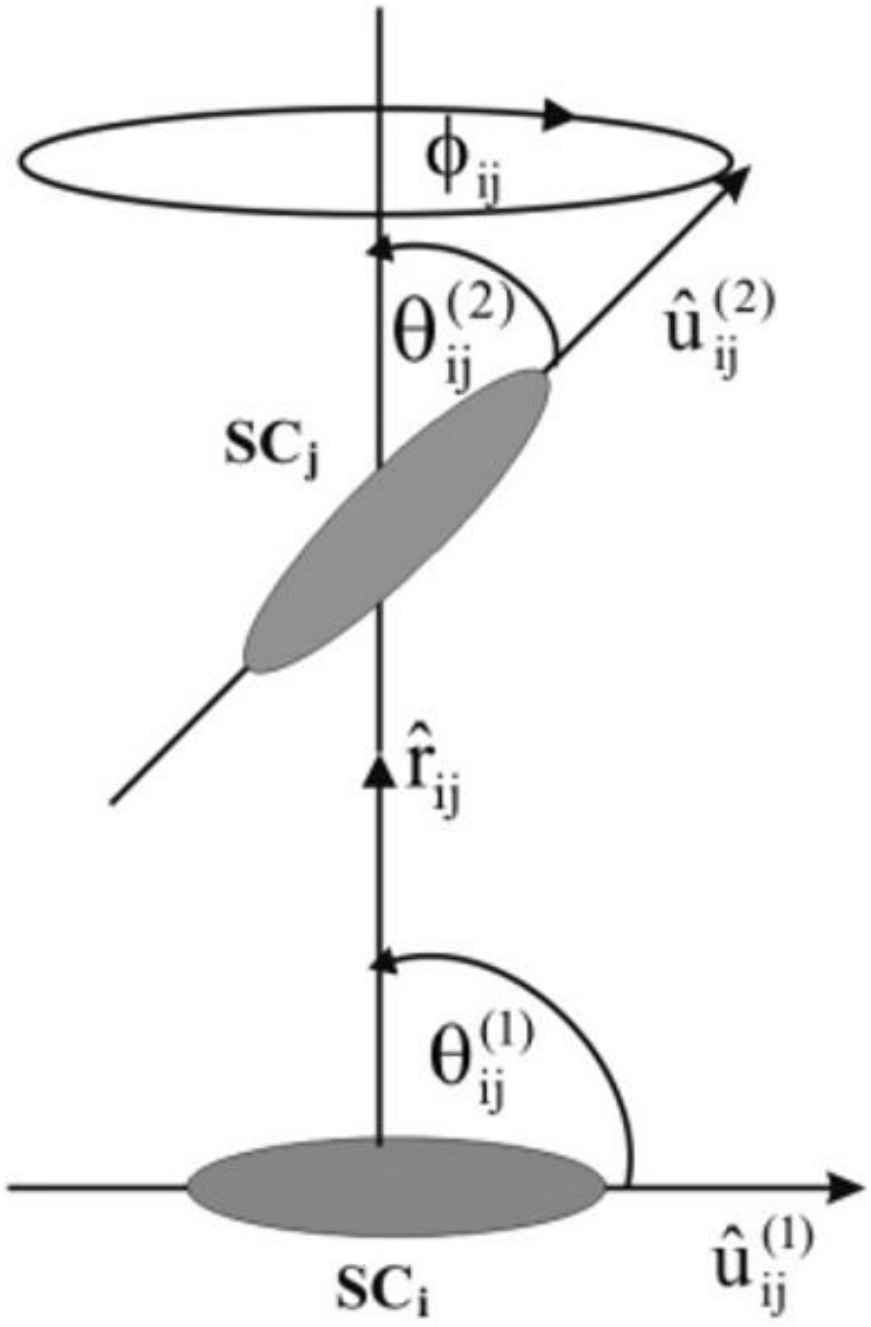


Figure 1.

Definition of variables describing the location of two spheroidal particles (*i* and *j*) with respect to each other. The vector $\widehat{u}_{ij}^{(1)}$ is the unit vector of the long axis of particle *i*, $\widehat{u}_{ij}^{(2)}$ is the unit vector of the long axis of particle *j*, $\hat{\mathbf{r}}_{ij}$ is the unit vector pointing from particle *i* to particle *j*, $\theta_{ij}^{(1)}$ and $\theta_{ij}^{(2)}$ are the angles between the vector $\hat{\mathbf{r}}_{ij}$ and vectors $\widehat{u}_{ij}^{(1)}$ and $\widehat{u}_{ij}^{(2)}$, respectively, and φ_{ij} is the angle of counterclockwise rotation of the vector $\widehat{u}_{ij}^{(2)}$ about the vector $\hat{\mathbf{r}}_{ij}$ from the plane

defined by the vector $\widehat{u}_{ij}^{(1)}$ and the vector $\hat{\mathbf{r}}_{ij}$ when looking from the center of particle *j* toward the center of particle *i*.

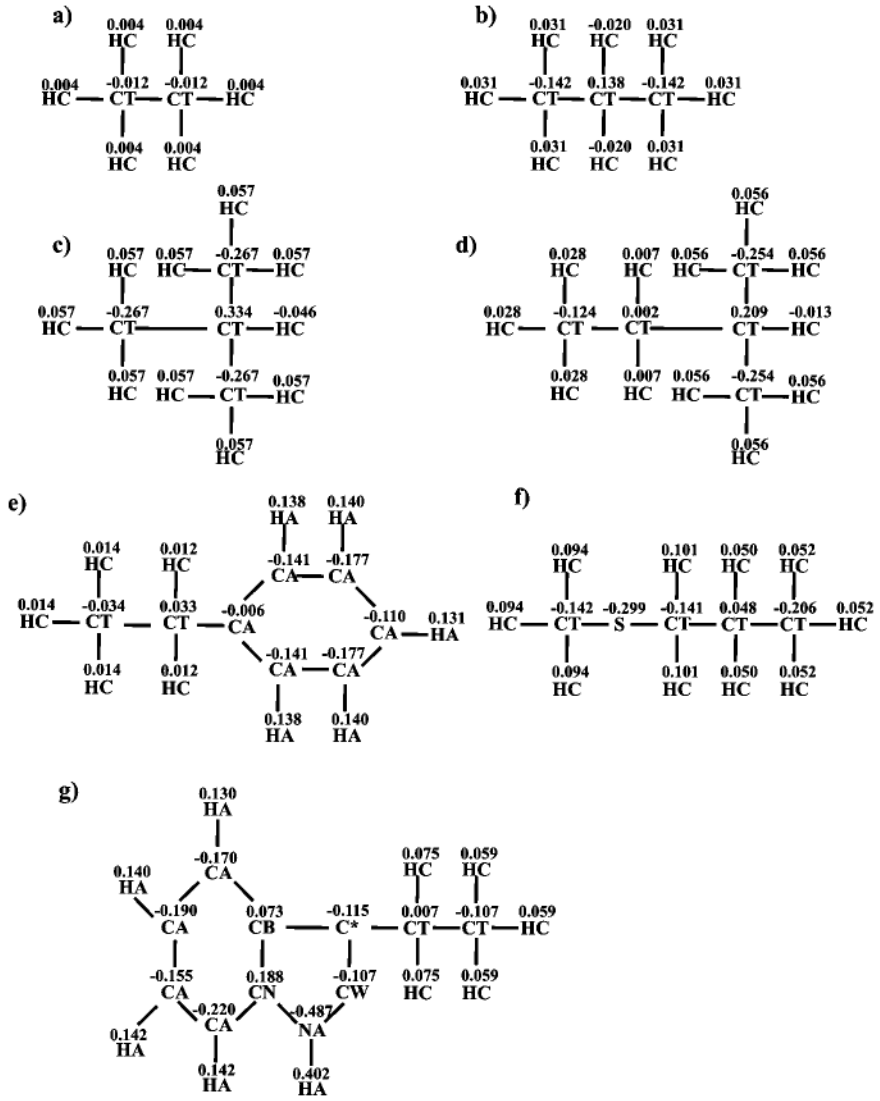


Figure 2.

Partial atomic charges (in electron charge units) of the ethane (a), propane (b), isobutene (c), isopentane (d), ethylbenzene (e), methyl propyl sulfide (f), and indole (g) molecules calculated by using the RESP method[16] based on HF/6-31G* calculations carried out with GAMESS[17] used in the calculations with the AMBER force field. The atoms are labeled with standard AMBER atom-type symbols, i.e., CT for an sp^3 carbon atom, HC for a hydrogen atom connected to a carbon atom, and CA for an aromatic or an olefinic carbon atom.

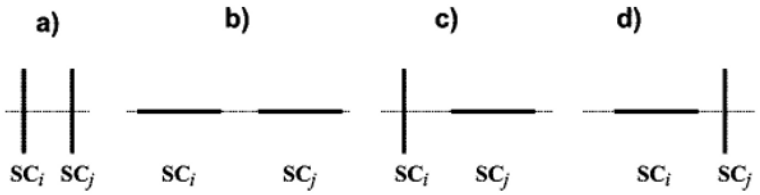
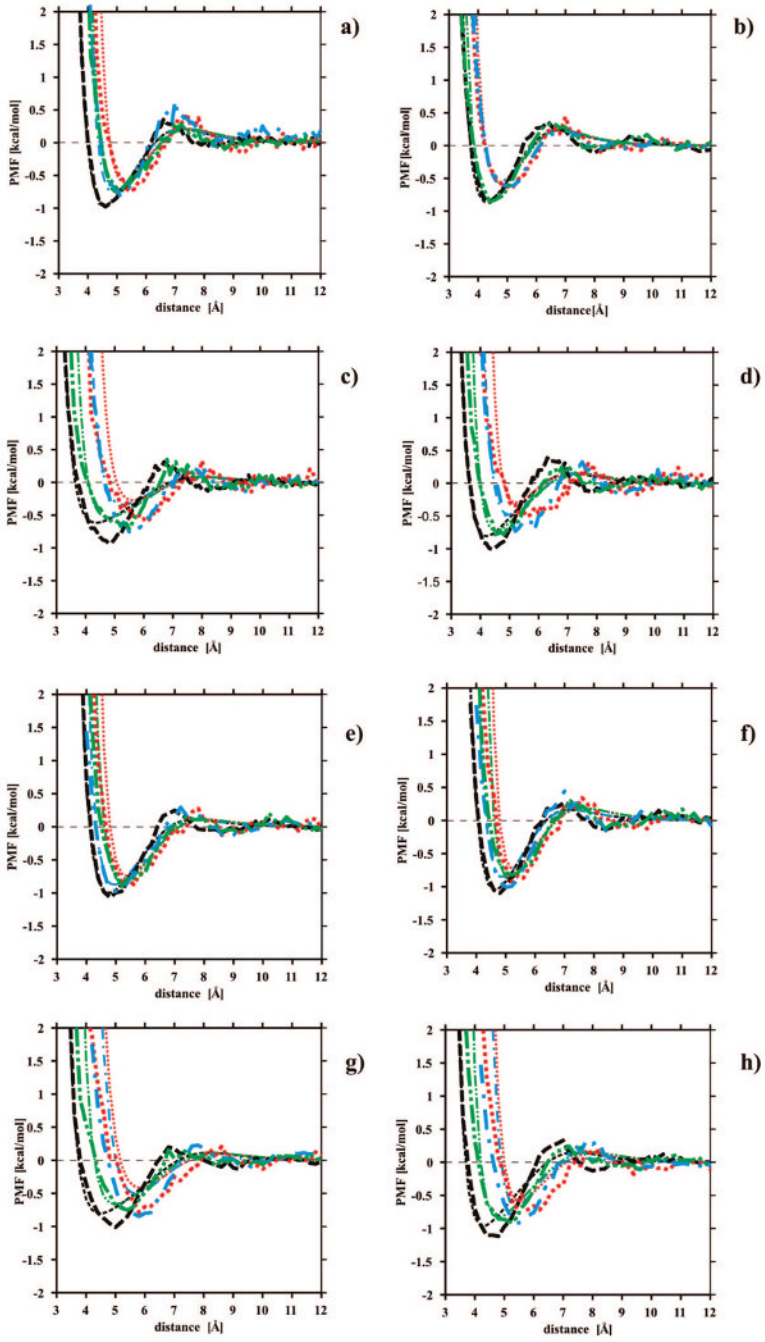
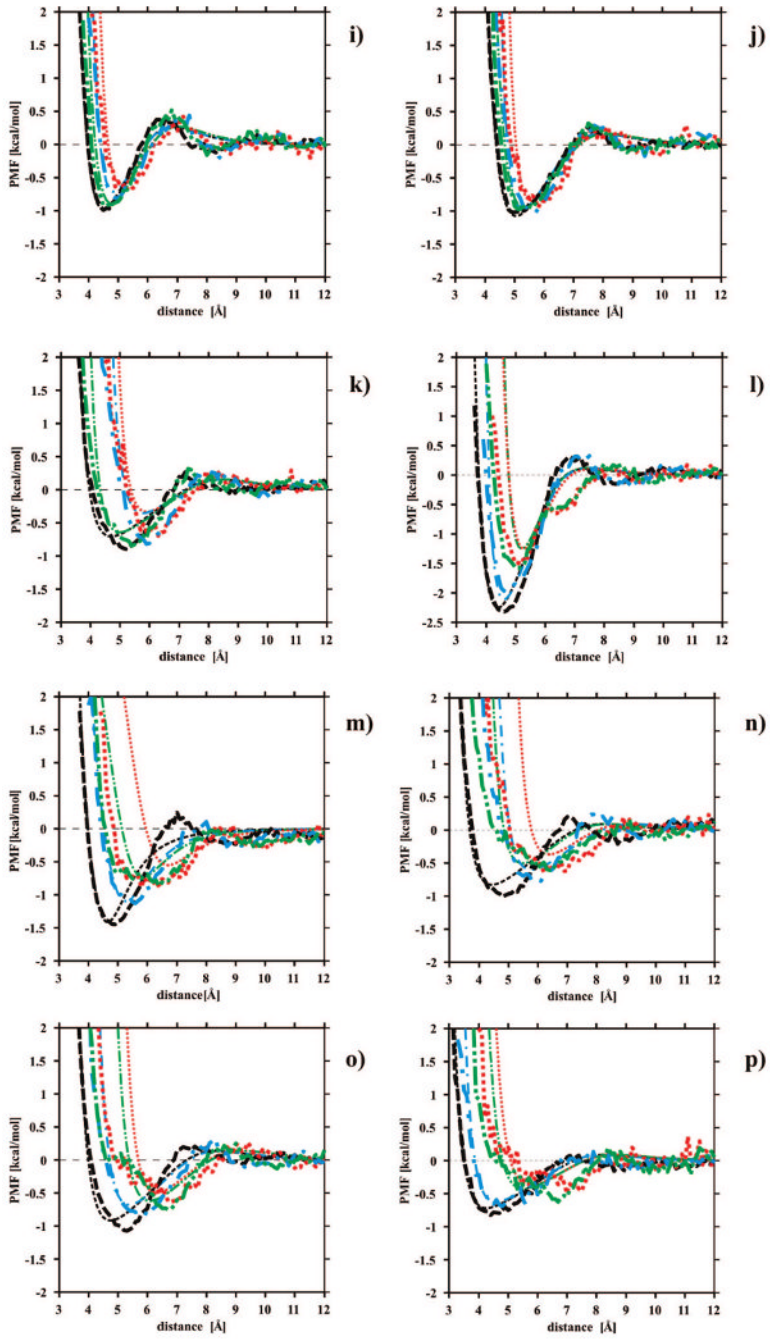


Figure 3.

Illustration of the (a) side-to-side, (b) edge-to-edge, (c) side-to-edge, and (d) edge-to-side orientation of two different spheroidal particles (*i* and *j*). The lines represent the long axes of the spheroids *i* and *j*, respectively. The orientation variables (see Figure 1 for definition) are $\theta_{ij}^{(1)}=90°$, $\theta_{ij}^{(2)}=90°$, and φ_{ij} = 0°, or 180° (a); $\theta_{ij}^{(1)}=0°$, $\theta_{ij}^{(2)}=0°$, or $\theta_{ij}^{(1)}=0°$, $\theta_{ij}^{(2)}=180°$, or $\theta_{ij}^{(1)}=180°$, $\theta_{ij}^{(2)}=0°$, or $\theta_{ij}^{(1)}=180°$, $\theta_{ij}^{(2)}=180°$, and φ_{ij} undefined (b); $\theta_{ij}^{(1)}=90°$, $\theta_{ij}^{(2)}=180°$, or $\theta_{ij}^{(1)}=90°$, $\theta_{ij}^{(2)}=0°$, or $\theta_{ij}^{(1)}=0°$, $\theta_{ij}^{(2)}=90°$, and φ_{ij} undefined (c); $\theta_{ij}^{(1)}=180°$, $\theta_{ij}^{(2)}=90°$, or $\theta_{ij}^{(1)}=0°$, $\theta_{ij}^{(2)}=90°$, and φ_{ij} undefined (d).





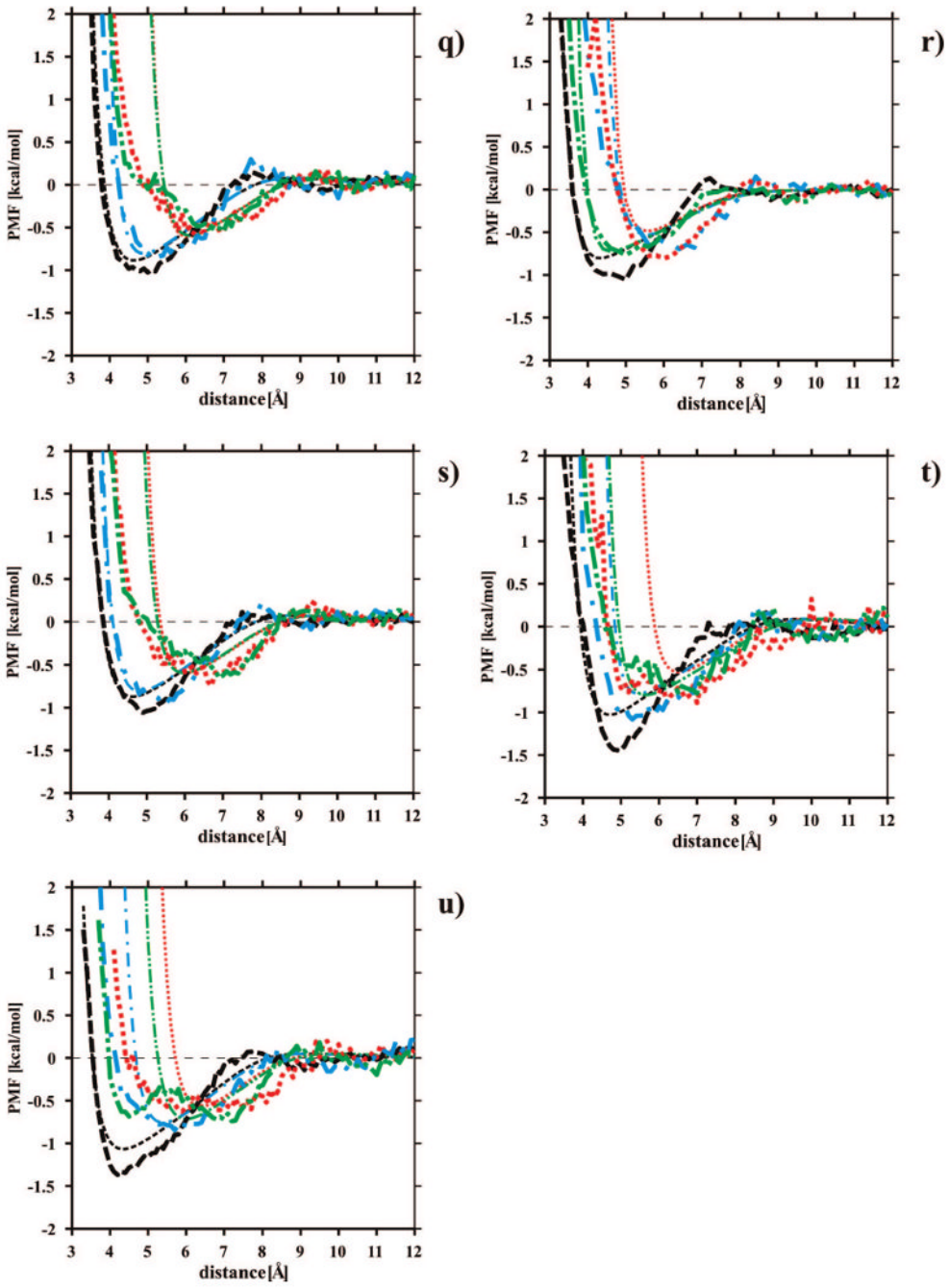


Figure 4.

PMF curves for the Et-iPen (a); Et-Prp (b); Et-PhEt (c); Et-MePrpS (d); Prp-iPen (e); Prp-iBut (f); Prp-PhEt (g); Prp-MePrpS (h); iBut-Et (i); iBut-iPen (j); iBut-PhEt (k); iBut-MePrpS (l); MePrpS-iPen (m); MePrpS-PhEt (n); PhEt-iPen (o); Ind-Et (p); Ind-iPen (q); Ind-Prp (r); Ind-iBut (s); Ind-PhEt (t); Ind-MePrpS (u); heterodimers. The dashed (black), dotted (red), dot-dashed (blue), and dash-double-dotted (green) lines correspond to PMFs determined for the side-to-side (Figure 3a), edge-to-edge (Figure 3b), side-to-edge (Figure 3c), and edge-to-side (Figure 3d) orientation, respectively, obtained by MD simulations. For each orientation, a thinner line of a given color and style corresponds to the analytical approximation to the PMFs, while a stronger line of the same style and color corresponds to the PMF determined from MD

simulations. The coefficients of the analytical expression to the PMF, with the Gay–Berne potential (eq 2) to represent the van der Waals interactions between the solute molecules, and eq 11 to represent the cavity potential were determined by least-squares fitting (eq 15).

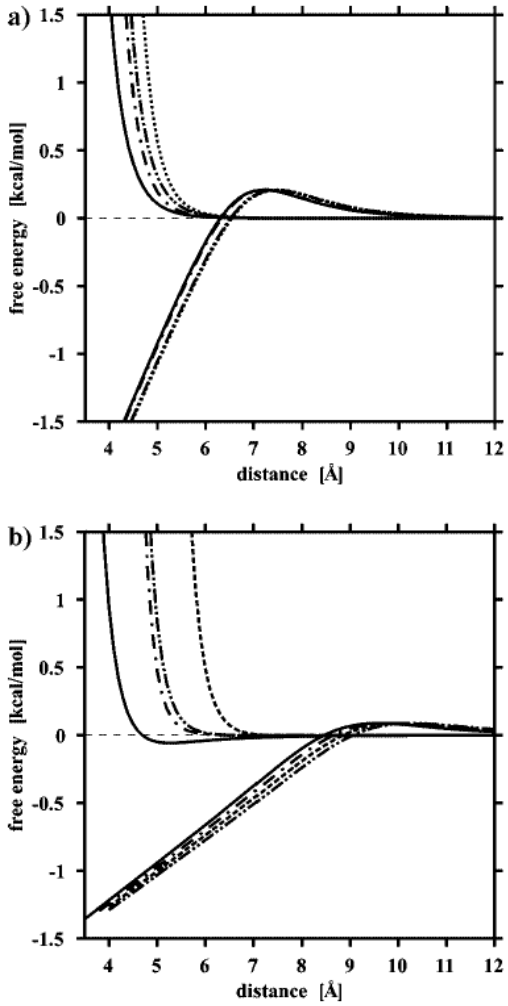


Figure 5.

Plots of the energy components of the analytical approximations to the PMFs for: Et-iPen (a) and Ind-PhEt (b) heterodimers. Lower curves: ΔF_{cav} (eq 11). Upper curves: E_{vdW} (eq 2). Solid lines: side-to-side orientation (Figure 3a), dashed lines: edge-to-edge orientation (Figure 3b), dot-dashed lines: side-to-edge orientation (Figure 3c), dash-double-dotted lines: edge-to-side orientation (Figure 3d).

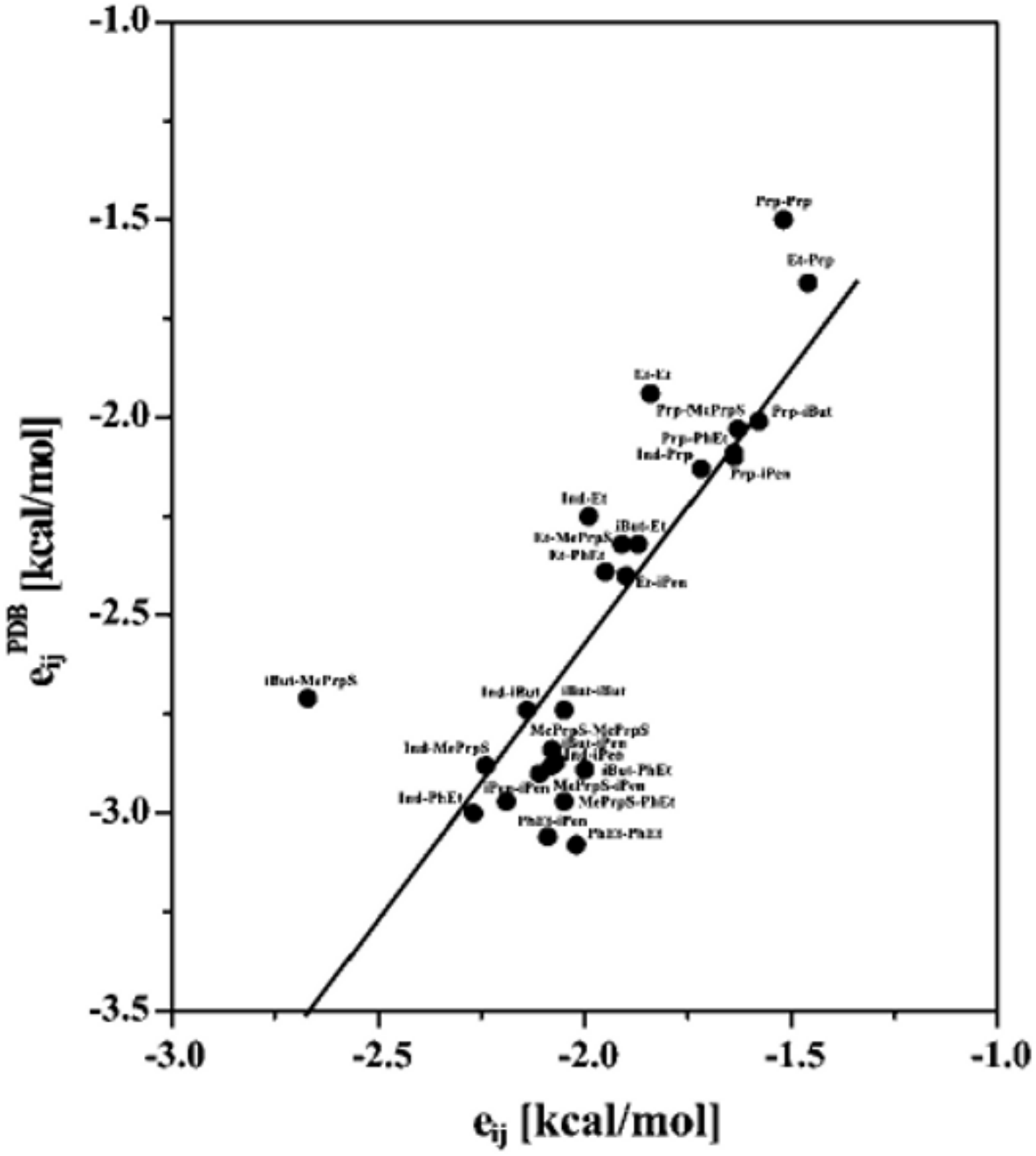


Figure 6.

Correlation of the calculated contact free energies (e_{ij}) for pairs of identical[3] and different hydrophobic amino acid side-chain models with those (e_{ij}^{PDB}) calculated by Liwo et al.[5]

TABLE 1

Box Dimensions and Number of Water Molecules for the Systems Studied during the Molecular Dynamics Simulations

system[a]	box dimensions (Å^3)	no. of water molecules
Et-iPen	37 × 37 × 37	1750
Et-Prp	34 × 34 × 34	1240
Et-PhEt	43 × 43 × 43	2637
Et-MePrpS	42 × 42 × 42	2550
Prp-iPen	33 × 33 × 33	1153
Prp-iBut	34 × 34 × 34	1233
Prp-PhEt	36 × 36 × 36	1526
Prp-MePrpS	33 × 33 × 33	1179
iBut-Et	31 × 31 × 31	1010
iBut-iPen	34 × 34 × 34	1231
iBut-PhEt	37 × 37 × 37	1597
iBut-MePrpS	37 × 37 × 37	1728
MePrpS-iPen	36 × 36 × 36	1576
MePrpS-PhEt	39 × 39 × 39	1997
PhEt-iPen	36 × 36 × 36	1534
Ind-Et	40 × 40 × 40	2048
Ind-iPen	36 × 36 × 36	1553
Ind-Prp	38 × 38 × 38	1783
Ind-iBut	37 × 37 × 37	1588
Ind-PhEt	40 × 40 × 40	2130
Ind-MePrpS	35 × 35 × 35	1464

[a] Abbreviations: E, ethane; Prp, propane; iBut, isobutane; iPen, isopentane; PhEt, ethylbenzene; MePrpS, methyl propyl sulfide; Ind, indole.

TABLE 2

Best-Fitted Parameters of E_{vdW} and ΔF_{cav} (Eqs 2 and 11, Respectively) Determined[a] by Minimization of the Function Φ Defined by Eq 15

system[b]	ε_{ij}^0 (kcal/mol)	σ_{ij}^0 (Å)	$\chi_{ij}^{(1)}$[c]	$\chi_{ij}^{(2)}$[c]	$\chi'^{(1)}_{ij}$ [c]	$\chi'^{(2)}_{ij}$ [c]	$\chi''^{(1)}_{ij}$ [c]	$\chi''^{(2)}_{ij}$ [c]	σ_i (Å)	σ_j (Å)	$\alpha_{ij}^{(1)}$ (kcal/mol)	$\alpha_{ij}^{(2)}$[c]	$\alpha_{ij}^{(3)}$[c]	$\alpha_{ij}^{(4)}$[c]
Et-iPen	0.0004	7.20	0.154	0.249	0.088	0.704	0.096	0.126	5.34	7.41	37.01	−0.003	0.25	31.25
Et-Prp	0.0052	5.36	0.127	0.127	0.318	−0.390	0.044	0.385	5.08	7.06	17.01	0.140	0.37	31.26
Et-PhEt	0.0001	7.17	0.142	0.206	−0.006	−0.750	0.074	0.092	3.90	7.33	0.52	4.39	3.96	4.32
Et-MePrpS	0.0074	5.00	0.377	0.383	0.821	0.386	0.099	0.831	5.38	7.46	36.96	−0.01	0.21	31.26
Prp-iPen	0.0056	5.93	0.219	0.215	0.181	0.556	0.119	0.001	5.46	7.58	37.02	−0.06	0.24	31.25
Prp-iBut	0.0934	4.73	0.407	0.079	0.663	−0.564	0.206	0.029	5.16	7.16	5.59	0.99	0.97	33.09
Prp-PhEt	0.0140	4.81	0.476	0.462	0.955	0.408	0.088	0.044	6.36	8.01	9.81	0.07	0.31	32.92
Prp-MePrpS	0.1760	4.08	0.343	0.456	0.464	−0.390	0.204	0.138	5.46	7.58	4.86	0.59	0.67	31.61
iBut-Et	0.0005	7.01	0.111	0.317	0.121	0.869	0.081	0.066	5.08	7.05	27.60	0.08	0.34	32.07
iBut-iPen	0.0236	5.69	0.425	0.245	0.990	0.250	0.067	−0.198	5.69	7.89	30.05	0.05	0.30	31.88
iBut-PhEt	0.0018	6.02	0.308	0.414	0.734	0.314	0.042	0.201	4.28	8.04	1.00	2.77	2.49	3.82
iBut-MePrpS	0.0168	5.21	0.586	0.253	0.969	0.531	0.064	−0.067	4.63	6.42	12.76	0.50	0.75	10.95
MePrpS-iPen	0.0108	5.45	0.444	0.464	0.419	0.960	0.267	0.213	6.90	7.04	42.35	−0.07	0.22	32.74
MePrpS-PhEt	0.0008	6.01	0.374	0.370	−0.487	0.078	0.279	0.172	7.78	3.78	0.63	4.39	4.00	4.54
PhEt-iPen	0.0006	6.79	0.346	0.279	−0.372	0.371	0.254	0.128	8.00	3.89	0.80	4.44	3.96	4.50
Ind-Et	0.0001	7.03	0.324	0.103	−0.290	−0.096	0.084	−0.118	8.01	3.90	0.49	4.41	3.94	4.50
Ind-iPen	0.0062	5.33	0.547	0.445	0.428	0.978	0.232	0.017	8.23	4.00	0.62	4.37	4.03	4.46
Ind-Prp	0.0157	4.58	0.500	0.534	0.998	0.507	0.019	0.092	4.61	7.84	22.31	−0.12	0.21	9.64
Ind-iBut	0.0024	5.72	0.505	0.310	0.424	0.779	0.186	−0.075	8.23	4.00	0.60	4.35	4.05	4.46
Ind-PhEt	0.0460	4.64	0.591	0.681	0.896	0.998	0.060	0.532	10.34	5.03	0.57	3.04	5.19	4.10
Ind-MePrPs	0.0018	5.46	0.481	0.411	−0.505	0.191	0.217	0.155	8.16	3.97	0.62	4.24	4.12	4.48

[a] The calculated van der Waals energy was based on the Gay–Berne-type potential (eq 2), and the calculated cavity creation energy term was obtained from eq 11.

[b] Abbreviations: Et, ethane; Prp, propane; iBut, isobutane; iPen, isopentane; PhEt, ethylbenzene; MePrpS, methyl propyl sulfide; Ind, indole.

[c] Dimensionless.

TABLE 3

Absolute and Relative Standard Deviations of the Fitted PMF from the PMF Obtained by MD Simulations

system[a]	σ_W[b] (kcal/mol)	σ_{W10}[c] (kcal/mol)	σ_{Wrel}[d]	σ_{W10rel}[e]
Et-iPen	1.23	0.25	0.06	0.03
Et-Prp	0.42	0.26	0.05	0.04
Et-PhEt	3.07	0.34	0.11	0.05
Et-MePrpS	2.59	0.34	0.10	0.04
Prp-iPen	0.98	0.32	0.07	0.04
Prp-iBut	1.79	0.40	0.10	0.05
Prp-PhEt	29.1	0.41	0.17	0.05
Prp-MePrpS	10.3	0.38	0.16	0.05
iBut-Et	0.79	0.25	0.06	0.03
iBut-iPen	2.35	0.37	0.10	0.05
iBut-PhEt	29.1	0.41	0.17	0.05
iBut-MePrpS	18.8	0.55	0.18	0.07
MePrpS-iPen	21.9	0.41	0.18	0.05
MePrpS-PhEt	194.	0.53	0.25	0.07
PhEt-iPen	27.1	0.40	0.20	0.05
Ind-Et	5.10	0.37	0.13	0.05
Ind-iPen	9732	0.40	0.24	0.05
Ind-Prp	264	0.35	0.19	0.05
Ind-iBut	98.8	0.40	0.20	0.05
Ind-PhEt	7.2e+12	0.50	0.30	0.06
Ind-MePrpS	1.5e+23	0.55	0.30	0.08

[a] Abbreviations: Et, ethane; Prp, propane; iBut, isobutane; iPen, isopentane; PhEt, ethylbenzene; MePrpS, methyl propyl sulfide; Ind, indole.

[b] $\sigma_W=[(1/n)\sum_{i=1}^{n}(W_i^{\mathrm{MD}}-W_i^{\mathrm{anal}})^2]^{1/2}$, where n is the number of points, W_i^{MD} is the PMF determined from MD simulations, and W_i^{anal} is the PMF calculated from the analytical expression (defined in the text under eq 15) at the ith point.

[c] $\sigma_{W10}=[1/n_{10}\sum_i(W_i^{\mathrm{MD}}-W_i^{\mathrm{anal}})^2]^{1/2}, W_i^{\mathrm{MD}}-W_{\mathrm{min}}^{\mathrm{MD}}<10\mathrm{kcal/mol}$, where $n10$ is the number of point with PMF within 10 kcal/mol cut-off over the minimum PMF, $W_{\mathrm{min}}^{\mathrm{MD}}$.

[d] $\sigma_{W\mathrm{rel}}=[(1/n)\sum_{i=1}^{n}[(W_i^{\mathrm{MD}}-W_i^{\mathrm{anal}})/(\max\{W_i^{\mathrm{MD}},1\mathrm{kcal}\cdot\mathrm{mol}\})]^2]^{1/2}$.

[e] $\sigma_{W10rel}=[1/n_{10}\sum_i(W_i^{\mathrm{MD}}-W_i^{\mathrm{anal}}/\max\{W_i^{\mathrm{MD}},1\mathrm{kcal/mol}\})^2]^{1/2}, W_i^{\mathrm{MD}}-W_{\mathrm{min}}^{\mathrm{MD}}<10\mathrm{kcal/mol}$.

TABLE 4

Calculated Contact Free Energies e_{ij} for Pairs of Identical[3] and Different Hydrophobic Amino Acid Side-Chain Models[a]

system[b]	e_{ij} (kcal/mol)[c]	e_{ij}^{PDB} (kcal/mol)[d]
Et-Et	−1.84	−1.94
iBut-iBut	−2.05	−2.74
iPen-iPen	−2.19	−2.97
Prp-Prp	−1.52	−1.50
PhEt-PhEt	−2.02	−3.08
MePrpS- MePrpS	−2.08	−2.84
Et-iPen	−1.90	−2.40
Et-Prp	−1.46	−1.66
Et-PhEt	−1.95	−2.39
Et-MePrpS	−1.91	−2.32
Prp-iPen	−1.64	−2.10
Prp-iBut	−1.58	−2.01
Prp-PhEt	−1.64	−2.09
Prp-MePrpS	−1.63	−2.03
iBut-Et	−1.87	−2.32
iBut-iPen	−2.07	−2.87
iBut-PhEt	−2.00	−2.89
iBut-MePrpS	−2.67	−2.71
MePrpS-iPen	−2.11	−2.90
MePrpS-PhEt	−2.05	−2.97
PhEt-iPen	−2.09	−3.06
Ind-Et	−1.99	−2.25
Ind-iPen	−2.08	−2.88
Ind-Prp	−1.72	−2.13
Ind-iBut	−2.14	−2.74
Ind-PhEt	−2.27	−3.00
Ind-MePrps	−2.24	−2.88

[a] For comparison with the contact free energies (e_{ij}^{PDB}) calculated by Liwo et al.[5]

[b] Abbreviations: Et, ethane; Prp, propane; iBut, isobutane; iPen, isopentane; PhEt, ethylbenzene; MePrpS, methyl propyl sulfide; Ind, indole.

[c] Values obtained by integration of the PMF curves calculated in ref 3 and this work.

[d] Values taken from ref 5.