Published in final edited form as:

J Phys Chem B. 2017 January 19; 121(2): 379–390. doi:10.1021/acs.jpcb.6b08541.

Simple Physics-Based Analytical Formulas for the Potentials of Mean Force of the Interaction of Amino Acid Side Chains in Water. VII. Charged–Hydrophobic/Polar and Polar–Hydrophobic/Polar Side Chains

Mariusz Makowsk^{*,†}, Adam Liwo[†], and Harold A. Scheraga[‡]

[†]Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland [‡]Baker laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, United States

Abstract

The physics-based potentials of side-chain–side-chain interactions corresponding to pairs composed of charged and polar, polar and polar, charged and hydrophobic, and hydrophobic and hydrophobic side chains have been determined. A total of 144 four-dimensional potentials of mean force (PMFs) of all possible pairs of molecules modeling these pairs were determined by umbrella-sampling molecular dynamics simulations in explicit water as functions of distance and orientation, and the analytical expressions were then fitted to the PMFs. Depending on the type of interacting sites, the analytical approximation to the PMF is a sum of terms corresponding to van der Waals interactions and cavity-creation involving the nonpolar sections of the side chains and van der Waals, cavity-creation, and electrostatic (charge–dipole or dipole–dipole) interaction energies and polarization energies involving the charged or polar sections of the side chains. The model used in this work reproduces all features of the interacting pairs. The UNited RESidue force field with the new side-chain–side-chain interaction potentials was preliminarily tested with the N-terminal part of the B-domain of staphylococcal protein A (PDBL 1BDD; a three- α -helix bundle) and UPF0291 protein YnzC from *Bacillus subtilis* (PDB: 2HEP; an α -helical hairpin).

Graphical abstract

*Corresponding Author: mariusz.makowski@ug.edu.pl. Phone: +48 58 523 50 55.

Supporting Information: The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b08541.

ORCID: Mariusz Makowski: 0000-0002-7342-722X

Notes: The authors declare no competing financial interest.

Introduction

Solvent-mediated interactions between protein side chains are one of the driving forces for the formation of tertiary structure of proteins.¹ Therefore, these interactions must be accounted for very accurately when developing models with which to simulate protein dynamics. In our previous papers of this series,^{2–7} we reported the development of our physics-based coarse-grained model of interacting side chains in water to be used in the UNited RESidue (UNRES) force field that is being developed in our laboratory.⁸ The UNRES model uses a highly reduced representation of polypeptide chains with only united side chains and united peptide groups as interaction sites; owing to this reduction, it enables us to extend the time scale of simulations by 3–4 orders of magnitude and results in a smoother effective potential energy surface compared to that from all-atom treatment.⁹ The UNRES energy function originates from the PMF of polypeptide chains in water, which is expanded into the cluster-cumulant functions that are identified with the effective interaction potential involving coarse-grained interaction sites, and most of the analytical expressions for these potentials have been obtained by using the generalized-cumulant expansion.¹⁰ In particular, this expansion enabled us to derive analytical formulas for the average electrostatic interactions between backbone peptide groups and for the coupling between backbone-local and backbone-electrostatic interactions.¹¹ Owing to the presence of these terms, the UNRES force field is able to reproduce protein secondary structure, in particular regular α -helices and β -sheets, correctly. Therefore, the UNRES force field is able to yield results consistent with the experiment (but with accuracy limited by the coarse-grained resolution of the treatment) regarding protein structure, protein dynamics, protein-folding kinetics, and thermodynamics.⁸ Other physics-based coarsegrained force fields, such as, for example, PaLaCe¹² or MARTINI,¹³ use more interaction sites per residue because the respective functional forms are largely imported from the all-atom force fields and do not include multibody terms. These force fields are not used for ab initio protein folding. The knowledge-based CABS force field¹⁴ developed by the Kolinski group that includes the multibody terms that couple adjacent hydrogen-bonding interactions can reproduce secondary structures correctly and, consequently, is used for ab initio protein folding.¹⁵ The reader is referred to a recent comprehensive review by Kmiecik et al. for more information about the coarse-grained force fields for protein.¹⁶

In the present UNRES force field, the interactions between side chains are represented by knowledge-based potentials⁹ that were determined on the basis of the statistics derived from the Protein Data Bank (PDB).^{17,18} Although these potentials have connections to the physics of interactions because they were rescaled to reproduce the free energies of transfer of amino acids from water to *n*-octanol,⁸ determining potentials from protein database does not enable us to avoid double-counting of interactions. Moreover, the statistical data are insufficient to determine the complete dependence of the potentials on side-chain orientation. In our earlier work,⁹ we assumed the Gay–Berne functional form,¹⁹ which has spheroidal symmetry. This form can be sufficient for interactions involving nonpolar side chains; however, charged and polar side chains possess nonpolar tailgroups and charged or polar headgroups and, thereby, a lower symmetry. Other statistical potentials that include the dependence on side-chain orientation are averaged over distances (i.e., they are contact potentials good for fold recognition)^{20,21} or averaged over residue types, leaving only four representative types of residue pairs.²² A way to derive the effective physics-based potential for side-chain–side-chain interactions, which is consistent with the definition of the UNRES energy function, is through the potentials of mean force (PMFs) that are usually determined by means of molecular dynamics (MD) simulations of the respective model systems in water. The PMF is one of the free-energy properties with which to calculate the physicochemical properties of the systems studied. The calculated PMFs can subsequently be used to compute measurable physicochemical characteristics of the systems studied, such as, for example, association constants,^{23–25} and are also suitable when direct experiment is inapplicable (for example to study hydrophobic association^{26,27}) or to lower the cost of experiment.^{28–30} The PMFs can also be used to determine the preferred mode of interaction of the molecules or groups under study in water. Many simulation studies of the PMF have been reported in the literature;^{31–35} however, in most of them, the distance between the centers of the systems is chosen as the sole reaction coordinate, which ignores the dependence of the PMF on the orientation of the interacting objects. However, it should be noted that the inaccuracies inherent in the all-atom force field used for this purpose translate to errors in the determined PMFs and the mean-field potentials of interactions between protein side chains and also between protein side chains and protein backbone are not pairwise additive.^{36,37} Nevertheless, these inaccuracies can be compensated for by appropriate calibration of the entire force field.⁸

In our previous work^{2–7} on the development of the new side-chain–side-chain interaction potentials (U_{SCSC}), a Gaussian-overlap model of hydrophobic association was developed and introduced² and then tested with simple spherical solute particles, and also, the model and functional forms for the interactions between charged and charged and hydrophobic³ side chains were proposed. On the basis of this initial work, we determined the distance- and orientation-dependent PMFs for pairs of charged^{6,7} and nonpolar^{4,5} side chains and then fitted the respective analytical expressions to them.

In this work, we complete the determination of the physics-based side-chain–side-chain potentials for UNRES by treating pairs of charged and nonpolar, charged and polar, polar and polar, and polar and nonpolar natural amino acid side chains. We determined the respective PMFs as functions of distances between the molecules and their relative orientation (four variables in total) by means of umbrella-sampling MD simulations. The respective analytical expressions were subsequently fitted to the PMFs. In each side-chain

model, the C^α atom is considered to be part of a side chain, as in the UNRES model.⁸ The character of the side chains was assigned as follows: nonpolar (Gly, Ala, Pro, Val, Leu, Ile, Phe, Met, Trp), polar (Ser, Thr, Cys, Asn, Gln, His, Tyr), and charged (Asp, Glu, Lys, Arg). A total of 144 systems were considered. Because the formation of disulfide bonds in UNRES was treated in our earlier work,³⁸ here, we considered only nonbonded cysteine pairs. We used the well-established AMBER 9.0 force field³⁹ to run MD simulations to determine the PMFs. Although this force field is not error-free, the fact that it and its earlier generations were used to fold a number of proteins, such as, for example, villin headpiece,⁴⁰ the N-terminal B-domain of staphylococcal protein A⁴¹ and tryptophan cage⁴² enable us to consider the obtained results as reliable.

Theory

In the model developed in our earlier work,^{2–7} nonpolar amino acid side chains are represented by ellipsoids of revolution (also termed spheroids), whereas for the charged and polar ones, an additional spherical headgroup is introduced, which bears the net charge or dipole moment for the charged and polar side chains, respectively. The variables describing the location of two spheroidal particles (i and j) with respect to each other are illustrated in Figure 1. Additionally, Figure 2 illustrates the model of amphiphilic side chains, which consists of a spheroidal nonpolar tail group and a spherical charged or polar headgroup whose center is located on the long axis of the ellipsoid, which is collinear with the C^α -SC axis.

The effective PMFs of pairs of charged–nonpolar (W_{ch}), charged–polar (W_{cp}), polar–nonpolar (W_{ph}), and polar–polar (W_{pp}) solutes in water are approximated by eqs 1–4, respectively, to give the effective side chain–side-chain interaction potentials

$$W_{\text{ch}} = E_{\text{GBerne}} + E_{\text{pol}} + \Delta F_{\text{cav}} \quad (1)$$

$$W_{\text{cp}} = E_{\text{GBerne}} + E_{\text{pol}} + \Delta F_{\text{cav}} + E_{\text{cp}} + E_{\text{LJ}} \quad (2)$$

$$W_{\text{ph}} = E_{\text{GBerne}} + \Delta F_{\text{cav}} \quad (3)$$

$$W_{\text{pp}} = E_{\text{GBerne}} + \Delta F_{\text{cav}} + E_{\text{pp}} \quad (4)$$

where E_{GBerne} is the van der Waals term corresponding to the interactions between the nonpolar sites represented by the anisotropic Gay–Berne potential;¹⁹ E_{pol} is the polarization energy coming from the interactions between the charged and nonpolar sites (but not from their interactions with the solvent); F_{cav} is the cavity term corresponding to the nonpolar

sections of the side chains, which is the difference between the cavity contribution to the free energy of hydration of the nonpolar sections of the side chains in the dimer and that in the isolated monomers; the isotropic Lennard-Jones term (E_{LJ}) expresses the van der Waals interaction energy between two amphiphilic headgroups; E_{cp} is the interaction energy between charged and polar sites; and E_{pp} is the potential of interaction between two polar headgroups.

The E_{GBerne} energy term is expressed by eq 5.

$$E_{GBerne} = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}^0}{r_{ij} - \sigma_{ij} + \sigma_{ij}^0} \right)^{12} - \left(\frac{\sigma_{ij}^0}{r_{ij} - \sigma_{ij} + \sigma_{ij}^0} \right)^6 \right] \quad (5)$$

where r_{ij} is the distance between the centers of the side chains, σ_{ij} is the distance corresponding to the zero value of E_{GBerne} for arbitrary orientation of the particles (σ_{ij}^0 is the distance corresponding to the zero value of E_{GBerne} 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 6–8 and eq 9, respectively.

$$\varepsilon_{ij} \equiv \varepsilon(\omega_{ij}^{(1)}, \omega_{ij}^{(2)}, \omega_{ij}^{(12)}) = \varepsilon_{ij}^0 \varepsilon_{ij}^{(1)} \varepsilon_{ij}^{(2)} \quad (6)$$

$$\varepsilon_{ij}^{(1)} = [1 - \chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(12)2}]^{-1/2} \quad (7)$$

$$\varepsilon_{ij}^{(2)} = \left[1 - \frac{\chi_{ij}^{(1)} \omega_{ij}^{(1)2} + \chi_{ij}^{(2)} \omega_{ij}^{(2)2} - 2\chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(1)} \omega_{ij}^{(2)} \omega_{ij}^{(12)2}}{1 - \chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(12)2}} \right]^2 \quad (8)$$

$$\sigma_{ij} = \sigma_{ij}^0 \left[1 - \frac{\chi_{ij}^{(1)} \omega_{ij}^{(1)2} + \chi_{ij}^{(2)} \omega_{ij}^{(2)2} - 2\chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(1)} \omega_{ij}^{(2)} \omega_{ij}^{(12)2}}{1 - \chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(12)2}} \right] \quad (9)$$

with

$$\omega_{ij}^{(1)} = \hat{\mathbf{u}}_{ij}^{(1)} \cdot \hat{\mathbf{r}}_{ij} = \cos \theta_{ij}^{(1)} \quad (10)$$

$$\omega_{ij}^{(2)} = \hat{\mathbf{u}}_{ij}^{(2)} \cdot \hat{\mathbf{r}}_{ij} = \cos\theta_{ij}^{(2)} \quad (11)$$

$$\omega_{ij}^{(12)} = \hat{\mathbf{u}}_{ij}^{(1)} \cdot \hat{\mathbf{u}}_{ij}^{(2)} = \cos\theta_{ij}^{(1)} \cos\theta_{ij}^{(2)} + \sin\theta_{ij}^{(1)} \sin\theta_{ij}^{(2)} \cos\phi_{ij} \quad (12)$$

where $\hat{\mathbf{u}}_{ij}^{(1)}$ and $\hat{\mathbf{u}}_{ij}^{(2)}$ are the unit vectors along the principal axes of the interacting sites (identified in this work with the C^α -SC axes), $\hat{\mathbf{r}}_{ij}$ is the unit vector pointing from the center of site i to that of site j , r_{ij} is the distance between the side-chain centers (Figures 1–2), parameters $\chi_{ij}^{(1)}$ and $\chi_{ij}^{(2)}$ are the anisotropies of the van der Waals distance, parameters $\chi_{ij}'^{(1)}$ and $\chi_{ij}'^{(2)}$ are the anisotropies of the van der Waals well depth, and 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 of the analytical expression for the free energy of two side chains interacting in water (eqs 1–4) to the PMFs determined from MD simulations, as in our earlier work.^{3,4,6,7}

The polarization component of the interactions between charged or polar and nonpolar particles is expressed by eq 13 (see Figure 2)⁶.

$$E_{\text{pol}} = \alpha_{ij}^{\text{pol}} \left(\frac{1}{f_{\text{GB}}(r_{ij}'')} \right)^4 + \alpha_{ji}^{\text{pol}} \left(\frac{1}{f_{\text{GB}}(r_{ji}'')} \right)^4 \quad (13)$$

where α_{ij}^{pol} and α_{ji}^{pol} are related to the polarizability of the nonpolar parts of side chain i and side chain j , respectively.

The expression for F_{cav} of spheroidal particles was derived in ref 3 on the basis of the Gaussian-overlap model of hydrophobic association introduced in that work and is given by eq 14. This term accounts for the free-energy contribution due to restructuring water molecules around a hydrophobic dimer and has been derived and discussed in detail in ref 2.

$$\Delta F_{\text{cav}} = \frac{\alpha_{ij}^{(1)} [(x \cdot \lambda)^{1/2} + \alpha_{ij}^{(2)} x \cdot \lambda - \alpha_{ij}^{(3)}]}{1 + \alpha_{ij}^{(4)} (x \cdot \lambda)^{12}} \quad (14)$$

with

$$x = \frac{r_{ij}}{\sqrt{\sigma_i^2 + \sigma_j^2}} \quad (15)$$

$$\lambda = \left[1 - \frac{\chi_{ij}^{(1)} \omega_{ij}^{(1)2} + \chi_{ij}^{(2)} \omega_{ij}^{(2)2} - 2\chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(1)} \omega_{ij}^{(2)} \omega_{ij}^{(12)}}{1 - \chi_{ij}^{(1)} \chi_{ij}^{(2)} \omega_{ij}^{(12)2}} \right]^2 \quad (16)$$

where symbols $\omega_{ij}^{(1)}$, $\omega_{ij}^{(2)}$, and $\omega_{ij}^{(12)}$ are defined by eqs 10–12, r_{ij} is the distance between the centers of the particles, $\chi_{ij}^{(1)}$ and $\chi_{ij}^{(2)}$ are anisotropies pertaining to F_{cav} , and σ_i and σ_j can be identified with the minimum distance between the centers of particles i and j , respectively. Parameters $\alpha_{ij}^{(1)}$, $\alpha_{ij}^{(2)}$, $\alpha_{ij}^{(3)}$, $\alpha_{ij}^{(4)}$, σ_i and σ_j and the anisotropies are determined by least-squares fitting of the analytical expression for the free energy of two side chains interacting in water (eqs 1–4) to the PMFs determined from MD simulations.

The Lennard-Jones potential (E_{LJ}) describing the van der Waals interactions between two charged or polar headgroups is expressed by eq 17

$$E_{\text{LJ}} = 4 \cdot \varepsilon_{ij} \cdot \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^6 \right] \quad (17)$$

where R_{ij} is the distance between the centers of the amphiphilic headgroups, σ_{ij} is the distance corresponding to the zero value of E_{LJ} , and ε_{ij} is the van der Waals well depth.

The E_{cp} interaction potential between charged and polar sites is given by eq 18

$$E_{\text{cp}} = w_{\text{dip}}^{(1)} \cdot \left(\frac{q \cdot \cos\theta_1}{R_{ij}^2} \right) - w_{\text{dip}}^{(2)} \cdot \left(\frac{q^2 \cdot \sin\theta_{12}}{R_{ij}^4} \right) \quad (18)$$

where $w_{\text{dip}}^{(1)}$ and $w_{\text{dip}}^{(2)}$ are the parameters determined by least-squares fitting of the analytical expressions (eqs 1–4) to the PMFs, q is the net charge of the charged headgroup, and R_{ij} is the distance between the centers of the amphiphilic head-groups.

The average energy of the interaction between two polar-group dipoles (E_{pp}) is expressed by eq 19 (ref 32)

$$E_{pp} = \left[\frac{w_{p1}}{R_{ij}^3} \cdot (\cos\omega_{ij}^{(12)} - 3 \cdot \cos\theta_{ij}^{(1)} \cdot \cos\theta_{ij}^{(2)}) - \left[\frac{w_{p2}}{R_{ij}^6} \cdot (4 + (\cos\omega_{ij}^{(12)} - 3 \cdot \cos\theta_{ij}^{(1)} \cdot \cos\theta_{ij}^{(2)})^2) - 3 \cdot (\cos^2\theta_{ij}^{(1)} + \cos^2\theta_{ij}^{(2)}) \right] \right] \quad (19)$$

where w_{p1} and w_{p2} are the parameters determined by least-squares fitting of the analytical expressions (eqs 1–4) to the PMFs and R_{ij} is the distance between the centers of the polar headgroups.

Methods

The charges on the atoms of the solute molecules needed for the AMBER 9.0 force field³⁹ were determined from the molecular electrostatic potentials calculated with the Hartree–Fock method with the 6-31G* basis set. The GAMESS⁴³ program was used to carry out quantum mechanical calculations. The partial charges were obtained by fitting to the molecular electrostatic potentials, computed as described above, by the RESP⁴⁴ program of the AMBER 9.0 package. The structures used as models of regular and amino acid side chains are shown in Figure S1 of the Supporting Information. The charges and atom force-field types used in MD simulations are presented in the same Figure S1 in electron charge units.

MD simulations were performed with the AMBER 9.0 package and force field.³⁹ Regular amino acid side chains were modeled by the molecules collected in Table S1. As mentioned in the Introduction, all pairs of charged and nonpolar, charged and polar, polar and polar, and polar and nonpolar natural amino acid side chains were treated (a total of 144 pairs). The pairs of nonpolar side chains and charged side chains were treated in our previous work.^{4–7}

Each of the pairs of molecules was placed in a periodic box containing approx. 1800 TIP3P water molecules.⁴⁵ Box sizes were adjusted to keep the density in the simulation box about 1 g/cm³. For pairs bearing a net charge, the chloride or sodium counterions were added, respectively, to keep the system neutral.

MD simulations were carried out in two steps. First, each system was equilibrated for $t = 100$ ps in the NPT ensemble (constant number of particles, pressure, and temperature) at $T = 298$ K and $p = 1$ atm. Production simulations were run in the NVT ensemble (constant number of particle, volume, and temperature) at $T = 298$ K for 10 ns. The integration time step was 2 fs. A 10 Å cut-off distance was applied for Lennard-Jones interactions. The bonds to hydrogen atoms were constrained with the SHAKE algorithm.⁴⁶ The particle-mesh Ewald summation⁴⁷ was used for estimating long-range effects of electrostatic interactions. Consequently, the PMFs must tend to zero with increasing solute–solute distance.²⁵

To cover the distance range between amino acid residues, the umbrella-sampling⁴⁸ method was used. For each system, a series of 10 windows of 10 ns simulations per window was run. The distance between the molecules was restrained with a biasing harmonic potential defined by eq 20.

$$V(\xi) = k(\xi - d_0)^2 \quad (20)$$

where k is the force constant ($k = 2 \text{ kcal/mol/\AA}^2$), ξ is the distance between the centers of the side-chain models, and d_0 is the equilibrium distance, equal to 3 Å, 4–12 Å for the respective simulation windows (10 windows in total). Snapshots from MD simulations were saved after every 0.2 ps; as a result, 50 000 trajectory frames were obtained (i.e., every 100th frame was saved) for each window at 298 K. Unrestrained sampling in angles covers the angle space well enough. Imposing the restraints on all four reaction coordinates would increase the computational effort by at least more than orders of magnitude; for example, if a 30° grid on $\theta^{(1)}$, $\theta^{(2)}$, and ϕ was used, there would be $6 \times 6 \times 12 = 432$ angle windows points for every distance window. Consequently, angles were not restrained during MD simulations.

To determine the PMFs of the systems studied, we processed the results of all restrained MD simulations for each system using the weighted histogram analysis method (WHAM).^{49,50}

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 $4.0 \text{ \AA} \leq r_{ij} \leq 13 \text{ \AA}$ with a bin size equal to $0.2 \text{ \AA}$, $0^\circ \leq \theta_{ij}^{(1)} \leq 180^\circ$ with a bin size equal to 60° , $0^\circ \leq \theta_{ij}^{(2)} \leq 180^\circ$ with a bin size equal to 60° , and $-180^\circ \leq \phi_{ij} < 180^\circ$ with a bin size equal to 60° . Consequently, each distance corresponded to 54 bins, each containing counts from different orientations of the molecules.

Fitting the analytical formulas to the PMFs was accomplished by minimizing the sum of the squares of the differences between the PMF values computed from the analytical formulas and determined from MD simulations, (Φ) defined by eq 21 using the Marquardt method.⁵¹

$$\min_x \Phi(y) = \sum_i w_i [W^{\text{MD}}(r_i, \theta_{ij}^{(1)}, \theta_{ij}^{(2)}, \phi_{ij}) - W^{\text{anal}}(r_i, \theta_{ij}^{(1)}, \theta_{ij}^{(2)}, \phi_{ij}; \mathbf{y})]^2 \quad (21)$$

$$w_i = \exp \left[- \frac{W^{\text{MD}}(r_i, \theta_{ij}^{(1)}, \theta_{ij}^{(2)}, \phi_{ij}) - W_{\min}}{RT} \right] \quad (22)$$

where $W^{\text{MD}}(r_{ij}, \theta_{ij}^{(1)}, \theta_{ij}^{(2)}, \phi_{ij})$ is the PMF value determined by simulations for distance r_{ij} and orientation $(\theta_{ij}^{(1)}, \theta_{ij}^{(2)}, \phi_{ij})$; $W^{\text{anal}}(r_{ij}, \theta_{ij}^{(1)}, \theta_{ij}^{(2)}, \phi_{ij}; \mathbf{y})$ is the analytical approximation to the

PMF at that point calculated with the parameters given by vector y , these parameters are the adjustable parameters of eqs 5–19; and w_i defined by eq 22 (in which $W_{\min}$ is the minimum PMF obtained in the simulations, R is the gas constant, and $T = 298$ K is the absolute temperature), is the weight of the i th point. Weighting the data points by the Boltzmann factor (eq 22) gives greater importance to low-energy regions of the free-energy surface.

In Figure 3, four selected orientations; “side-to-side” (parallel, black) (a), “head-to-head” (linear, red) (b), “head-to-side” (perpendicular, green) (c), and “side-to-head” (perpendicular, blue) (d), for which the distance dependence of the PMF is discussed in the next section; are illustrated.

Results and Discussion

We determined the PMFs [$W(r, \theta^{(1)}, \theta^{(2)}, \phi)$ of eq 21] for pairs of amino acid side-chain models in water, calculated with MD simulations performed in the NVT scheme at 298 K. Subsequently, the analytical formulas appropriate for the respective types of side-chain pairs (charged and nonpolar, charged and polar, polar and polar, polar and nonpolar; eqs 1–4) were fitted to the PMFs to derive the parameters. Because the side-chain–side-chain interaction PMFs depend on distance and orientation, we analyzed the PMFs for four selected orientations, namely, side-to-side (a), head-to-head (b), head-to-side (c), and side-to-head (d), illustrated in Figure 3. For glycine, which is represented by a methane molecule, only the orientation of the first side chain in a pair matter and they are denoted by the same colors as the side-to-side (when the side of the first side chain is oriented toward the methane molecule) and head-to-head (when the head of the first side chain is oriented toward the methane molecule). The PMF curves (plotted as functions of the distance between the side-chain centers for the four selected orientations), together with the fitted curves for selected pairs, are shown in Figures 4a–d (pairs of charged and nonpolar side chains of amino acids), 5a–d (pairs of charged and polar side chains of amino acids), 6a–d (pairs of polar and nonpolar side chains of amino acids), and 7a–d (pairs of polar and polar side chains of amino acids), whereas plots for the remaining pairs are shown in Figures S2–S5. The black, red, green, and blue lines correspond to the PMFs determined for the side-to-side, head-to-head, head-to-side, and side-to-head orientations, respectively. The definitions of these four orientations are indicated below each plot. Because of the large number of figures considered, we denote the interacting pairs in each plot by the three-letter symbols of the corresponding amino acid residues.

Generally, the positions and depths of the contact minima and the positions and heights of the desolvation maxima, including their mutual orientations, depend on the character of the interacting pairs. More distinct minima are observed for charged and nonpolar (Figure 4a–d) and polar pairs (Figures 6a–d, S2, and S5, respectively).

The results of fitting the PMFs, using the analytical expressions given by eqs 1–4, are also shown in Figures 4–7 and S2–S5 (solid lines). The fitted parameters of the expressions of the PMF components for E_{GBerne} (eq 5), E_{pol} (eq 13), F_{cav} (eq 14), E_{LJ} (eq 17), E_{cp} (eq 18), and E_{pp} (eq 19) and the distances of the charged or polar headgroups from the centers of the nonpolar tailgroups (Figure 1) are summarized in Tables S2–S5.

As can be seen from Figures 4–7 and S2–S5, the analytical expressions reproduce the shapes of the PMFs reasonably well. Particularly good fit is obtained for the side-to-side orientation. The order of the depths of the minima for the side-to-side, head-to-head, head-to-side, and side-to-head orientations is also reproduced.

The PMFs of the four classes of pairs of interacting side chains considered in this study (charged and nonpolar, charged and polar, polar and nonpolar, and polar and polar) are discussed in the following four subsections.

Pairs of Charged and Nonpolar Amino Acid Side Chains

The PMF curves for charged and nonpolar amino acid side-chain models are shown in Figure 4a–d and in Figure S2. As could be expected, the side-to-side approach minima occur at the shortest and the head-to-head minima at the longest distance. For pairs composed of negatively charged (Asp and Glu) and nonpolar residues, the head-to-head approach minima are deeper than the side-to-side approach minima or have nearly the same depth as that of these minima (Figures 4a,b and S2a–S2n). Conversely, for pairs containing one positively charged side chain (Arg and Lys), the side-to-side approach minima are the deepest, the side-to-head minima the second deepest, and the head-to-head minima are the shallowest. The PMFs corresponding to these three orientations are shown in Figure 4c,d. A shallow minimum is usually observed for the head-to-side approach, but for some pairs (such as, e.g., Lys–Ala and Lys–Ile/Leu; Figures S2w and S2y), it disappears producing all-repulsive PMF profiles. As for the pairs of positively charged and nonpolar residues, the head-to-side approach minima are the shallowest and, for some pairs, disappear (see, e.g., Figures S2c, S2j, and S2n). The relatively deeper minima corresponding to the side-to-side orientation for the pairs of charged and nonpolar side chains result from the fact that the positively charged Lys and Arg side chains have larger nonpolar sections compared to those of the negatively charged Asp and Glu side chains, making the hydrophobic association with nonpolar side chains stronger.

It can be observed that the PMF minima for pairs with positively charged side chains are about 1.0–2.0 kcal/mol deep for the side-to-side orientation. For pairs of negatively charged and nonpolar side chains, the depth of the minima is between 0.5 and 1.0 kcal/mol for both the head-to-head and side-to-side orientations.

The analytical expression for the PMF (eqs 1 and 5–17) reproduces the PMF profiles reasonably well, the fit being particularly good for the side-to-side orientation for which the greatest numbers of points have been collected (Figures 4a–d and S2). The positions and depths of the contact minima are also reproduced reasonably well.

As mentioned in the Introduction, there are only a few similar studies of the determination of the orientation dependence of the side-chain–side-chain interaction potentials. The closest study was carried out by Mukherjee et al.²² These authors determined statistical potentials of four pairs corresponding to the interactions of nonpolar, like-charged, oppositely charged, and charged and nonpolar residue pairs. The arginine–phenylalanine pair was selected to represent the interaction between charged and nonpolar side chains; the respective potential profiles were shown in Figure 11 of ref 22, whereas the potential profiles determined in this

work are shown in Figure S2t. It can be seen that the minima occur generally at shorter distances in the statistical potential, which is reasonable because of crystal packing. The order of the minimum distance is the same for both potential profiles, that is, side-to-side < side-to-head < head-to-side < head-to-head (assuming that the arginine side chain is the first side chain in the pair); however, the minimum corresponding to the side-to-head orientation occurs at a shorter distance in the potential determined in this work, and its distance is clearly shorter than that of the head-to-side orientation, whereas the minimum distance is comparable for these two orientations for the statistical potential. Because arginine is a larger side chain than phenylalanine and has a greater ratio of the long to the short axis, the feature of the PMF determined in this work seems to reflect the interaction pattern of these side chains better, whereas the statistical potentials can be influenced by crystal packing. The order of the depths of the minima is comparable for both potentials, with the side-to-side orientation corresponding to the lowest and the head-to-head orientation to the highest potential.

Pairs of Charged and Polar Amino Acid Side Chains

The PMF curves for selected pairs of charged and polar amino acid side-chain models are shown in Figure 5a–d, and the plots for the remaining pairs are shown in Figure S3. The curves have essentially the same shapes as those corresponding to pairs of charged and nonpolar side chains. Likewise, for pairs composed of a positively charged and a nonpolar side chain, the minima corresponding to the side-to-side orientation are remarkably deeper than those for the head-to-head orientation, whereas the head-to-head minima are deeper or of comparable depth for the pair composed of a negatively charged and a nonpolar side chain.

Pairs of Polar and Nonpolar Amino Acid Side Chains

The PMF curves for polar and nonpolar amino acid side-chain models are shown in Figure 6a–d and in Figure S4.

The shapes of the PMFs and their dependence on the orientation and side-chain size match those observed for the pairs of charged and nonpolar and charged and polar side chains discussed in the two preceding sections. Again, the contact minima are the deepest for the side-to-side orientation if at least one of the side chains is large, whereas the head-to-head minimum also becomes pronounced for smaller side chains. The head-to-side orientation (with the polar group of the first side chain in a pair facing the nonpolar side chain in the pair) has a shallower minimum compared to that corresponding to the side-to-head orientation.

Pairs of Polar Amino Acid Side Chains

The PMF curves for polar and polar amino acid side-chain models are shown in Figure 7a–d and in Figure S5.

The shapes of the PMF curves for different orientations and the order of the position and depth of the contact minima in the PMF curves are similar to those discussed in the preceding three sections except that for identical side chains the head-to-side and side-to-

head orientations are equivalent and there is less difference between these two orientations for pairs of non-identical side chains because both side chains in the pair contain polar groups. Also, for the Tyr–Tyr and His–His pairs (Figure 7c,d), the PMF curves contain no minima for the head-to-head orientation, which can be explained by the cancellation of the attractive interactions between the polar groups of the two side chains and their interactions with the solvent.

Preliminary Tests of the Force Field with New Side-Chain–Side-Chain Potentials

The complete set of side-chain–side-chain interaction potentials determined in this and in our previous work^{2–7} was introduced into the UNRES force field. We tested the modified force field with the N-terminal part of the B-domain of staphylococcal protein A (PDBL 1BDD; a three- α -helix bundle) and UPF0291 protein YnzC from *Bacillus subtilis* (PDB: 2HEP; an α -helical hairpin), carrying out replica exchange MD (REMD)⁵² simulations adapted to UNRES in our previous work⁵³ (20 trajectories at temperatures ranging from 250 to 500 K, each trajectory 40 000 000 time steps in total, 4.89 fs time-step size) and processing the results with the WHAM⁴⁹ and cluster analysis to determine conformational families, as described in our previous work.⁵⁴ The results are shown in Figure 8.

For both proteins, clusters of native-like structures were found with C $^{\alpha}$ -rmsd values of 4.0 and 4.6 Å, for protein A (Figure 8a) and 2HEP (Figure 8b), respectively. On the other hand, the structures are still of middle resolution; therefore, the force field with the new potentials needs to be calibrated. The calibration procedure is aimed at reproducing the native structures and thermodynamics of folding of selected training proteins.⁸ Recently,⁵⁵ we developed a new method on the basis of maximum-likelihood fitting of the conformational ensembles simulated at various temperatures to the respective experimental ensembles determined by NMR. This work is now being carried out in our laboratory.

Conclusions

In this work, we determined the PMFs for 144 pairs composed of charged and nonpolar, charged and polar, polar and nonpolar, and polar amino acid side-chain models, as functions of the distances between the interacting particles and their orientations. The respective approximate analytical expressions were fitted to the PMFs to give the physics-based side-chain–side-chain potentials that will be used with the coarse-grained UNRES force field or in other coarse-grained force fields.⁸ We carried out preliminary tests of these potentials with the two small α -helical proteins, obtaining clusters of native-like structures. However, for transferability and to achieve better resolution, the force field with the new potentials needs to be recalibrated with several training proteins. This work is currently being carried out in our laboratory.

Despite the different character of the side chains considered in this study, the PMF surfaces exhibit very similar shapes. If at least one side chain in the pair is large, the deepest minimum in the PMF plot corresponds to the side-to-side orientation, whereas for smaller side chains, the head-to-head minimum has comparable depth or even becomes deeper than the side-to-side minimum. This feature can be explained by the fact that for small side chains, which usually exhibit a less pronounced anisotropy, the molecular surface of the

dimer does not depend much on the orientation and, on the other hand, the screening of the polar group from the solvent is not complete even when it contacts its counterpart. Conversely, for larger side chains, the molecular surface of the dimer strongly depends on the orientation, being the smallest for the side-to-side orientation, and the screening of the polar groups is more pronounced for the head-to-head contacts. For pairs composed of a charged or polar and a nonpolar side chain, there is a strong asymmetry between the head-to-side and side-to-head orientations, the first one often resulting in a repulsive-only potential profile. This is because of the efficient screening of the charged or polar group from the solvent; this screening becomes more pronounced with increasing side-chain size. It should be noted that this asymmetry is not accounted for in the side-chain–side-chain potentials used in the present UNRES model,⁸ which have the Gay–Berne¹⁹ functional form of spheroidal symmetry.

It should also be noted that the PMFs of side-chain–side-chain interactions determined in this and our earlier work^{4–7} describe the energetics of solvent-mediated mean-field interactions between amino acid side chains in general. Therefore, they can find other applications, for example, in the analysis of the interactions in proteins for the purpose of determining the stability of mutants, and so forth.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a grant from the Polish National Science Centre (UMO-2013/10/E/ST4/00755), the U.S. National Science Foundation (MCB-10-19767), and the National Institutes of Health (GM-14312). This research was conducted using resources of (a) our 818-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (b) our cluster *Piasek* at the Faculty of Chemistry, University of Gdańsk, (c) the Academic Computer Center (CI TASK) in Gdańsk, and (d) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) at the University of Warsaw.

References

1. Dill KA. Dominant Forces in Protein Folding. *Biochemistry*. 1990; 29:7133–7155. [PubMed: 2207096]
2. Makowski M, Liwo A, Scheraga HA. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force for the Interaction of Amino Acid Side Chains in Water. 1. Approximate Expression for the Free Energy of Hydrophobic Association Based on a Gaussian-Overlap Model. *J Phys Chem B*. 2007; 111:2910–2916. *Erratum J Phys Chem B* **2010**, 114, 1226. [PubMed: 17388416]
3. Makowski M, Liwo A, Maksimiak K, Makowska J, Scheraga HA. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force for the Interaction of Amino Acid Side Chains in Water. 2. Tests with Simple Spherical Systems. *J Phys Chem B*. 2007; 111:2917–2924. [PubMed: 17388417]
4. Makowski M, Sobolewski E, Czaplewski C, Liwo A, Oldziej S, No JH, Scheraga HA. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force for the Interaction of Amino Acid Side Chains in Water. 3. Calculation and Parameterization of the Potentials of Mean Force of Pairs of Identical Hydrophobic Side Chains. *J Phys Chem B*. 2007; 111:2925–2931. [PubMed: 17388418]
5. Makowski M, Sobolewski E, Czaplewski C, Oldziej S, Liwo A, Scheraga HA. 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. *J Phys Chem B*. 2008; 112:11385–11395. [PubMed: 18700740]
6. Makowski M, Liwo A, Sobolewski E, Scheraga HA. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force of the Interaction of Amino-Acid Side Chains in Water. V. Like-Charged Side Chains. *J Phys Chem B*. 2011; 115:6119–6129. [PubMed: 21500792]
7. Makowski M, Liwo A, Scheraga HA. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force of the Interaction of Amino-Acid Side Chains in Water. VI. Oppositely-Charged Side Chains. *J Phys Chem B*. 2011; 115:6130–6137. [PubMed: 21500791]
8. Liwo A, Baranowski M, Czaplewski C, Goła E, He Y, Jagieła D, Krupa P, Maciejczyk M, Makowski M, Mozolewska MA, et al. A Unified Coarse-Grained Model of Biological Macromolecules Based on Mean-Field Multipole-Multipole Interactions. *J Mol Model*. 2014; 20(2306)
9. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. Prediction of Protein Conformation on the Basis of a Search for Compact Structures: Test on Avian Pancreatic Polypeptide. *Protein Sci*. 1993; 2:1715–1731. [PubMed: 8251944]
10. Kubo R. Generalized Cumulant Expansion Method. *J Phys Soc Jpn*. 1962; 17:1100–1120.
11. Khalili M, Liwo A, Jagielska A, Scheraga HA. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model α -Helical Systems. *J Phys Chem B*. 2005; 109:13798–13810. [PubMed: 16852728]
12. Pasi M, Lavery R, Ceres N. PaLaCe: A Coarse-Grain Protein Model for Studying Mechanical Properties. *J Chem Theory Comput*. 2013; 9:785–793. [PubMed: 26589071]
13. Monticelli L, Kandasamy SK, Periole X, Larson RL, Tieleman DP, Marrink SJ. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput*. 2008; 4:819–834. [PubMed: 26621095]
14. Kolinski A. Protein Modeling and Structure Prediction with a Reduced Representation. *Acta Biochim Pol*. 2004; 51:349–371. [PubMed: 15218533]
15. Kmiecik S, Kolinski A. Folding Pathway of the B1 Domain of Protein G Explored by Multiscale Modeling. *Biophys J*. 2008; 94:726–736. [PubMed: 17890394]
16. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-Grained Protein Models and Their Applications. *Chem Rev*. 2016; 116:7898–7936. [PubMed: 27333362]
17. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. A United-Residue Force Field for Off-Lattice Protein-Structure Simulations. I: Functional Forms and Parameters of Long-Range Side-Chain Interaction Potentials from Protein Crystal Data. *J Comput Chem*. 1997; 18:849–873.
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28:235–242. [PubMed: 10592235]
19. Gay J, Berne BJ. Modification of the Overlap Potential to Mimic a Linear Site-Site Potential. *J Chem Phys*. 1981; 74:3316–3319.
20. Buchete NV, Straub JE, Thirumalai D. Anisotropic Coarse-Grained Statistical Potentials Improve the Ability to Identify Nativelike Protein Structures. *J Chem Phys*. 2003; 118:7658.
21. Buchete NV, Straub JE, Thirumalai D. Orientational Potentials Extracted from Protein Structures Improve Native Fold Recognition. *Protein Sci*. 2004; 13:862–874. [PubMed: 15044723]
22. Mukherjee A, Bhimalapuram P, Bagchi B. Orientation-dependent potential of mean force for protein folding. *J Chem Phys*. 2005; 123(014901)
23. Makowska J, Makowski M, Gieldo A, Liwo A, Chmurzyński L. Theoretical Calculations of Heteroconjugation Equilibrium Constants in Systems Modeling Acid-Base Interactions in Side Chains of Biomolecules Using the Potential of Mean Force. *J Phys Chem B*. 2004; 108:12222–12230.
24. Makowska J, Makowski M, Chmurzyński L, Liwo A. Theoretical Calculation of Homoconjugation Equilibrium Constants in Systems Modeling Acid-Base Interactions in Side-Chains of Biomolecules Using the Potential of Mean Force. *J Comput Chem*. 2005; 26:235–242. [PubMed: 15599950]
25. Wiñiewska M, Makowski M. Theoretical Studies on Anionic Association of Phenol and Its Derivatives in Acetonitrile. *J Mol Struct*. 2014; 1076:165–173.

26. Sobolewski E, Makowski M, Oldziej S, Czaplewski C, Liwo A, Scheraga HA. Towards Temperature-Dependent Coarse-Grained Potentials of Side-Chain Interactions. I. Molecular Dynamics Study a Pair of Methane Molecules in Water at Various Temperatures. *Protein Eng, Des Sel.* 2009; 22:547–552. [PubMed: 19556395]
27. Bartosik A, Wi niewska M, Makowski M. Potentials of Mean Force for Hydrophobic Interactions Between Hydrocarbons in Water Solution: Dependence on Temperature, Solute Shape, and Solute Size. *J Phys Org Chem.* 2015; 28:10–16.
28. Makowska J, Makowski M, Chmurzy ski L. Ab Initio Studies on Acid-Base Equilibria of Substituted Phenols. *J Phys Chem A.* 2004; 108:10354–10358.
29. Sobolewski E, Oldziej S, Wi niewska M, Liwo A, Makowski M. Towards Temperature-Dependent Coarse-Grained Potentials of Side-Chain Interactions for Protein Folding Simulations. II. Molecular Dynamics Study of Pairs of Different Types of Interactions in Water at Various Temperatures. *J Phys Chem B.* 2012; 116:6844–6853. [PubMed: 22475198]
30. Wi niewska M, Sobolewski E, Oldziej S, Liwo A, Scheraga HA, Makowski M. Theoretical Studies of Interactions Between O-phosphorylated and Standard Amino-Acid Side-Chain Models in Water. *J Phys Chem B.* 2015; 119:8526–8534. [PubMed: 26100791]
31. Kovalenko A, Hirata F. Potentials of Mean Force of Simple Ions in Ambient Aqueous Solution. I. Three-Dimensional Reference Interaction Site Model Approach. *J Chem Phys.* 2000; 112:10391–10402.
32. Kovalenko A, Hirata F. Potentials of Mean Force of Simple Ions in Ambient Aqueous Solution. II. Solvation Structure from the Three-Dimensional Reference Interaction Site Model Approach, and Comparison with Simulations. *J Chem Phys.* 2000; 112:10403–10417.
33. Rozanska X, Chipot C. Modeling Ion-Ion Interaction in Proteins: A Molecular Dynamics Free Energy Calculation of the Guanidinium-Acetate Association. *J Chem Phys.* 2000; 112:9691–9694.
34. Villarreal M, Montich G. Energetic and Entropic Contributions to the Interactions Between Like-Charged Groups in Cationic Peptides: A Molecular Dynamics Simulation Study. *Protein Sci.* 2002; 11:2001–2009. [PubMed: 12142454]
35. Masunov A, Lazaridis T. Potentials of Mean Force Between Ionizable Amino Acid Side Chains in Water. *J Am Chem Soc.* 2003; 125:1722–1730. [PubMed: 12580597]
36. Rank JA, Baker D. A Desolvation Barrier to Hydrophobic Cluster Formation May Contribute to the Rate-Limiting Step in Protein Folding. *Protein Sci.* 1997; 6:347–354. [PubMed: 9041636]
37. Meral D, Toal S, Schweitzer-Stenner R, Urbanc B. Water-Centered Interpretation of Intrinsic pP_{II} Propensities of Amino Acid Residues: In Vitro-Driven Molecular Dynamics Study. *J Phys Chem B.* 2015; 119:13237–13251. [PubMed: 26418575]
38. Chinchio M, Czaplewski C, Liwo A, Oldziej S, Scheraga HA. Dynamic Formation and Breaking of Disulfide Bonds in Molecular Dynamics Simulations with the UNRES Force Field. *J Chem Theory and Comput.* 2007; 3:1236–1248. [PubMed: 26633198]
39. Case, DA., Darden, TA., Chaetham, TE., III, Simmerling, CL., Wang, J., Duke, RE., Luo, R., Merz, KM., Pearlman, DA., Crowley, M., et al. AMBER 9. University of California; San Francisco: 2006.
40. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science.* 1998; 282:740–744. [PubMed: 9784131]
41. Jang S, Kim E, Shin S, Pak Y. Ab initio folding of helix bundle proteins using molecular dynamics simulations. *J Am Chem Soc.* 2003; 125:14841–14846. [PubMed: 14640661]
42. Kannan S, Zacharias M. Folding of Trp-cage Mini Protein Using Temperature and Biasing Potential Replica—Exchange Molecular Dynamics Simulations. *Int J Mol Sci.* 2009; 10:1121–1137. [PubMed: 19399241]
43. Schmidt MW, Baldrige KK, Boatz JA, Elbert ST, Gordon MS, Jensen JA, Koseki S, Matsunaga N, Nguyen KA, Su S, et al. General Atomic and Molecular Electronic Structure System. *J Comput Chem.* 1993; 14:1347–1363.
44. Bayly CI, Cieplak P, Cornell WD, Kollman PA. Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - The Resp Model. *J Phys Chem.* 1993; 97:10269–10280.

45. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys.* 1983; 79:926–935.
46. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J Comput Phys.* 1977; 23:327–341.
47. Darden T, York D, Pedersen L. Particle Mesh Ewald: An $N\log(N)$ Method for Ewald Sums in Large Systems. *J Chem Phys.* 1993; 98:10089–10092.
48. Torrie GM, Valleau JP. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation. Umbrella Sampling. *J Comput Phys.* 1977; 23:187–199.
49. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J Comput Chem.* 1992; 13:1011–1021.
50. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J Comput Chem.* 1995; 16:1339–1350.
51. Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J Soc Ind Appl Math.* 1963; 11:431–441.
52. Sugita Y, Okamoto Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem Phys Lett.* 1997; 1-3:140–150.
53. Czaplewski C, Kalinowski S, Liwo A, Scheraga HA. Application of Multiplexed Replica Exchange Molecular Dynamics to the UNRES Force Field: Tests with α and $\alpha + \beta$ Proteins. *J Chem Theory Comput.* 2009; 5:627–540. [PubMed: 20161452]
54. Liwo A, Khalili M, Czaplewski C, Kalinowski S, Ołdziej S, Wachucik K, Scheraga HA. Modification and Optimization of the United-Residue (UNRES) Potential Energy Function for Canonical Simulations. I. Temperature Dependence of the Effective Energy Function and Tests of the Optimization Method with Single Training Proteins. *J Phys Chem B.* 2007; 111:260–285. [PubMed: 17201450]
55. Zaborowski B, Jagieła D, Czaplewski C, Hałabis A, Lewandowska A, Mudzińska W, Ołdziej S, Karczyńska A, Omieczynski C, Wirecki T, Liwo A. A Maximum-Likelihood Approach to Force-Field Calibration. *J Chem Inf Model.* 2015; 55:2050–2070. [PubMed: 26263302]

Figure 1.

Definition of variables describing the location of two spheroidal particles (i and j) with respect to each other. $\hat{\mathbf{u}}_{ij}^{(1)}$ is the unit vector of the long axis of particle i , $\hat{\mathbf{u}}_{ij}^{(2)}$ 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 vector $\hat{\mathbf{r}}_{ij}$ and vectors $\hat{\mathbf{u}}_{ij}^{(1)}$ and $\hat{\mathbf{u}}_{ij}^{(2)}$, respectively, and ϕ_{ij} is the angle of counterclockwise rotation of vector $\hat{\mathbf{u}}_{ij}^{(2)}$ about vector $\hat{\mathbf{r}}_{ij}$ from the plane defined by vector $\hat{\mathbf{u}}_{ij}^{(1)}$ and vector $\hat{\mathbf{r}}_{ij}$ when looking from the center of particle j toward the center of particle i .

Figure 2.

Illustration of the new model for the interactions of charged and polar side chains. A side chain of this type consists of a nonpolar site (represented by an ellipsoid of revolution) and a polar/charged site (represented by a shaded sphere). The center of the polar/charged site of side chain i is at the distance $d_i^{(1)}$ from the geometric center of that side chain (SC_i) (which is located between the polar/charged and nonpolar centers and represented by a small sphere in the figure) and that of side chain j is at the distance $d_j^{(1)}$ from the side-chain center (SC_j), whereas the centers of the nonpolar sites of side chains i and j are at distances $d_i^{(2)}$ and $d_j^{(2)}$, respectively, from their geometric centers (SC_i and SC_j , respectively). $\hat{u}_{ij}^{(1)}$ is the unit vector of the long axis of the nonpolar site of side chain i , $\hat{u}_{ij}^{(2)}$ is the unit vector of the long axis of the nonpolar site of side chain j , $\hat{r}_{ij}$ is the unit vector pointing from the geometric center of the nonpolar site of side chain i to that of side chain j , R_{ij} is the distance between these two centers, r_{ij} is the distance between the centers of the charged/polar site of side chains i and j , r''_{ij} is the distance between the centers of the charged site of side chain i and the nonpolar site of side chain j , and r''_{ji} is the distance between the centers of the charged site of side chain j and the nonpolar site of side chain i .

Figure 3.

Illustration of the (a) side-to-side, (b) head-to-head, (c) head-to-side, and (d) side-to-head orientations of two charged and spheroidal particles. The lines represent the long axes of the spheroids. The orientation variables (see Figure 1 for definition) are as follows: $\theta_{ij}^{(1)}=90^\circ$, $\theta_{ij}^{(2)}=90^\circ$, and $\phi_{ij}=0^\circ$ (a); $\theta_{ij}^{(1)}=0^\circ$, $\theta_{ij}^{(2)}=180^\circ$, and ϕ_{ij} undefined (b); $\theta_{ij}^{(1)}=0^\circ$, $\theta_{ij}^{(2)}=90^\circ$, and $\phi_{ij}=0^\circ$ (c); $\theta_{ij}^{(1)}=90^\circ$, $\theta_{ij}^{(2)}=180^\circ$, and ϕ_{ij} undefined (d). The filled circles at one of the ends of each particle refer to the charged “head” of the particle.

Figure 4.

PMF curves for 4 out of 32 possible charged and hydrophobic side chain models of interactions of (a) Asp–Val, (b) Glu–Val, (c) Arg–Val, and (d) Lys–Val amino acid pairs. The dashed black, red, green, and blue lines correspond to the PMFs determined for the side-to-side (Figure 3a), head-to-head (Figure 3b), head-to-side (Figure 3c), and side-to-head (Figure 3d) orientations, respectively, obtained by MD simulations at 298 K. The solid lines of the same colors correspond to the analytical approximation to the PMFs, with coefficients determined by least-squares fitting (eq 21).

Figure 5.

PMF curves for 4 out of 28 possible charged and polar side chain models of interactions of (a) Asp-Asn, (b) Glu-Asn, (c) Arg-Asn, and (d) Lys-Asn amino acid pairs. The dashed black, red, green, and blue lines correspond to the PMFs determined for the side-to-side (Figure 3a), head-to-head (Figure 3b), head-to-side (Figure 3c), and side-to-head (Figure 3d) orientations, respectively, obtained by MD simulations at 298 K. The solid lines of the same colors as those of dashed ones correspond to the analytical approximation to the PMFs, with coefficients determined by least-squares fitting (eq 21).

Figure 6.

PMF curves for 4 out of 56 possible polar and hydrophobic side chain models of interactions of (a) Asn–Val, (b) Ser–Val, (c) His–Val, and (d) Tyr–Val amino acid pairs. The dashed black, red, green, and blue lines correspond to the PMFs determined for the side-to-side (Figure 3a), head-to-head (Figure 3b), head-to-side (Figure 3c), and side-to-head (Figure 3d) orientations, respectively, obtained by MD simulations at 298 K. The solid lines of the same colors as those of dashed ones correspond to the analytical approximation to the PMFs, with coefficients determined by least-squares fitting (eq 21).

Figure 7.

PMF curves for 4 out of 28 possible polar and polar side chain models of interactions of (a) Ser-Ser, (b) Asn-Asn, (c) His-His, and (d) Tyr-Tyr amino acid pairs. The dashed black, red, green, and blue lines correspond to the PMFs determined for the side-to-side (Figure 3a), head-to-head (Figure 3b), head-to-side (Figure 3c), and side-to-head (Figure 3d) orientations, respectively, obtained by MD simulations at 298 K. The solid lines of the same colors as those of dashed ones correspond to the analytical approximation to the PMFs, with coefficients determined by least-squares fitting (eq 21).

Figure 8. Results of fitting of the C^α -trace of the experimental structures (grey sticks) and the best theoretical structures (colored ribbons) of protein A (a) and 2HEP (b). The root-mean-square deviation (rmsd) values from the native structures, averaged over the entire ensembles at 300 K were 4.0 and 4.6 Å, respectively.