

Physics-Based Potentials for the Coupling between Backbone- and Side-Chain-Local Conformational States in the United Residue (UNRES) Force Field for Protein Simulations

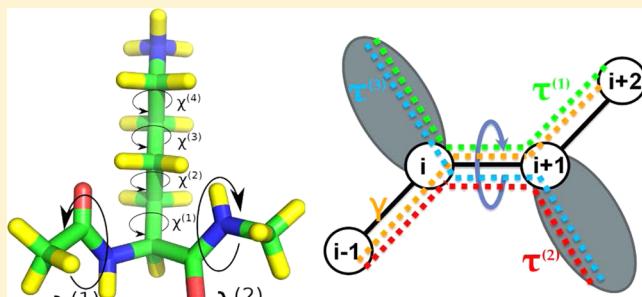
Adam K. Sieradzan,^{†,‡,§} Paweł Krupa,^{*,†,‡,§} Harold A. Scheraga,[‡] Adam Liwo,[†] and Cezary Czaplewski[†]

[†]Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-180 Gdańsk, Poland

[‡]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, United States

Supporting Information

ABSTRACT: The UNited RESidue (UNRES) model of polypeptide chains is a coarse-grained model in which each amino-acid residue is reduced to two interaction sites, namely, a united peptide group (p) located halfway between the two neighboring α -carbon atoms (C^α s), which serve only as geometrical points, and a united side chain (SC) attached to the respective C^α . Owing to this simplification, millisecond molecular dynamics simulations of large systems can be performed. While UNRES predicts overall folds well, it reproduces the details of local chain conformation with lower accuracy. Recently, we implemented new knowledge-based torsional potentials (Krupa et al. *J. Chem. Theory Comput.* 2013, 9, 4620–4632) that depend on the virtual-bond dihedral angles involving side chains: $C^\alpha\cdots C^\alpha\cdots C^\alpha\cdots SC$ ($\tau^{(1)}$), $SC\cdots C^\alpha\cdots C^\alpha\cdots C^\alpha$ ($\tau^{(2)}$), and $SC\cdots C^\alpha\cdots C^\alpha\cdots SC$ ($\tau^{(3)}$) in the UNRES force field. These potentials resulted in significant improvement of the simulated structures, especially in the loop regions. In this work, we introduce the physics-based counterparts of these potentials, which we derived from the all-atom energy surfaces of terminally blocked amino-acid residues by Boltzmann integration over the angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for rotation about the $C^\alpha\cdots C^\alpha$ virtual-bond angles and over the side-chain angles χ . The energy surfaces were, in turn, calculated by using the semiempirical AM1 method of molecular quantum mechanics. Entropy contribution was evaluated with use of the harmonic approximation from Hessian matrices. One-dimensional Fourier series in the respective virtual-bond-dihedral angles were fitted to the calculated potentials, and these expressions have been implemented in the UNRES force field. Basic calibration of the UNRES force field with the new potentials was carried out with eight training proteins, by selecting the optimal weight of the new energy terms and reducing the weight of the regular torsional terms. The force field was subsequently benchmarked with a set of 22 proteins not used in the calibration. The new potentials result in a decrease of the root-mean-square deviation of the average conformation from the respective experimental structure by 0.86 Å on average; however, improvement of up to 5 Å was observed for some proteins.



AM1 calculations of amino-acids to obtain physics-based SC-backbone correlation potentials.

1. INTRODUCTION

Simulations of molecular biosystems can give insights into molecular mechanism of folding,^{1,2} functionally important protein motions,³ protein–ligand affinity,⁴ lipid bilayer behavior,⁵ and DNA–drug interactions.⁶ Theoretical and experimental studies in this field are complementary.⁷ Molecular dynamics (MD) approaches to study ligand–receptor binding have been used in preliminary *in silico* experiments of drug development, successfully reducing the cost of designing new drugs.^{8,9}

Great progress in extending the time and size scales of all-atom simulations is constantly achieved. Many different approaches to improve calculation speed were proposed such as use of world-distributed computing (e.g., the FOLDING@HOME project),¹⁰ development of very efficient load-balanced parallel codes such as, e.g., GROMACS,¹¹ NAMD,¹² and DESMOND,¹³ implementation of all-atom molecular

dynamics programs on graphical processor units (GPUs),¹⁴ and the construction of dedicated machines such as ANTON.¹⁵ The recent advances in computation methods have facilitated the simulations of very large systems at all-atom resolution.⁷ With the ANTON machine, all-atom simulations of smaller systems (e.g., up to 100 residues) can be performed at submillisecond time scale.^{16,17} However, access to the ANTON¹⁵ supercomputer is limited, and even calculations with ANTON are restricted to either relatively small (microsecond) time scale¹⁸ or to small systems (up to 120,000 atoms with solvent).^{16,17} Owing to recent improvements of the all atom force fields and simulation techniques, ab initio folding simulations at the all-atom resolution have become feasible for small proteins.^{16,19–21}

Received: August 13, 2014

Published: January 6, 2015



The need for simulating large systems at large time scales is addressed by coarse-graining approaches, in which some of the details of a system are omitted from the model. One of such approaches to proteins is the UNited RESidue (UNRES) model of polypeptide chains, which is being developed in our laboratory.^{22–30} Owing to the use of a coarse-grained representation of polypeptide chains, simulations with UNRES are faster by 3–4 orders of magnitude with respect to all-atom molecular dynamics simulations in explicit water or by 2 orders of magnitude faster with respect to all-atom simulations in implicit solvent^{31,32} (implicit solvent is assumed in UNRES). Part of the speed-up results from the extension of the effective time scale because of averaging out fast-moving degrees of freedom such as, e.g., the solvent degrees of freedom. Thus, 1 μ s of UNRES simulations corresponds to 1 ms of all-atom simulations or 1 ms of real time.^{31,32} Another part of the speed-up is a result of the reduction of the number of interaction sites and, thereby, a lower cost of energy and force evaluation.

The UNRES force field performs well in the prediction of overall folds,^{33–37} including domain packing,³⁷ as proved in Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiments.^{33–36} In the CASP10 experiment, the predictions made with the use of UNRES for targets T0663 and T0740 were featured by the assessors as the best for these targets.³⁷ The reason for good performance of UNRES is use of anisotropic potentials for side chain–side chain interactions, represented by the Gay-Berne functional form, which has spheroidal symmetry,³⁸ and introduction of multibody terms for the potential of mean force of polypeptide chains in water,²⁴ derived in a systematic way through Kubo's cluster-cumulant expansion.³⁹ Recently, other coarse-grained models of proteins with Gay-Berne potentials for the side chain–side chain interactions have been developed.^{40,41}

On the other hand, UNRES does not reproduce local chain conformations that well. To address this problem, in our recent work⁴² we developed new torsional potentials that depend on the angles involving side-chain centers. These potentials were derived based on statistics from the Protein Data Bank (PDB).⁴³ These potentials improved the quality of UNRES-simulated structures of proteins, especially in loop regions.⁴² However, statistical potentials are dependent on a database and, moreover, cannot be used with confidence to handle D-amino-acid residues and nonstandard residues. Therefore, in this work, we focused on the improvement of the specificity of local interactions. We introduced new physics-based energy terms that account for the coupling between backbone- and side-chain conformational states.

In this study, we tuned only the weight of the new terms in the UNRES energy function to obtain the best performance of the force field with eight selected training proteins. We did not change the weights of the other energy terms except for reducing the weight of the torsional terms following the introduction of the new terms, to avoid double-counting of the same interactions. Such an approach enabled us to assess the improvement resulting from the introduction of the new terms and not from optimization of the other terms already present in the energy function. To test the force field with the new terms, we used a set of 22 proteins (Table 2), none of which was present in the training set. Both the calculations with the new terms and the reference calculations without the new terms were run on this set of test proteins.

2. METHODS

2.1. UNRES Representation of Polypeptide Chain. In the UNRES model,^{22–30} a polypeptide chain is defined by the α -carbon (C^α) trace with united side chains (SC) attached to the respective C^α 's and united peptide groups (p) positioned halfway between two consecutive C^α 's. The SC and p centers are interaction sites, while the C^α 's serve only to define backbone geometry. The effective energy function is represented by a restricted free energy (RFE) or potential of mean force (PMF) of the conformational ensemble restricted to a given coarse-grained geometry (defined by C^α 's and SCs) and is expressed by eq 1.

$$\begin{aligned} U = & w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{p_i p_j}^{VDW} \\ & + w_{pp}^{el} f_2(T) \sum_{i < j-1} U_{p_i p_j}^{el} + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) \\ & + w_{tor} f_3(T) \sum_i U_{tor}(\gamma_i, \gamma_{i+1}) + w_b \sum_i U_b(\theta_i) \\ & + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + w_{bond} \sum_i U_{bond}(d_i) \\ & + w_{corr}^{(3)} f_3(T) U_{corr}^{(3)} + w_{corr}^{(4)} f_4(T) U_{corr}^{(4)} + w_{turn}^{(3)} f_3(T) U_{turn}^{(3)} \\ & + w_{turn}^{(4)} f_4(T) U_{turn}^{(4)} + w_{ssbond} \sum_{nss} U_{ssbond}(d_{ss}) \\ & + w_{SC-corr} f_2(T) \sum_{m=1}^3 \sum_i U_{SC-corr}(\tau_i^{(m)}) \end{aligned} \quad (1)$$

where the U 's are energy terms, θ_i is the backbone virtual-bond angle between three consecutive C^α atoms, γ_i is the backbone virtual-bond-dihedral angle (defined by four consecutive C^α 's), α_i and β_i are the angles defining the location of the center of the united side chain of residue i (Figure 1) with the respect to the C_{i-1}^α , C_i^α , and C_{i+1}^α plane, d_i is the length of the i th virtual bond, which is either a $C^\alpha\cdots C^\alpha$ virtual bond or $C^\alpha\cdots SC$ virtual bond, d_{ss} is the distance between the side chains of two cysteine residues, and the angles $\tau^{(1)}\cdots \tau^{(3)}$ are the $SC\cdots C^\alpha\cdots C^\alpha\cdots C^\alpha$ ($\tau^{(1)}$), $C^\alpha\cdots C^\alpha\cdots C^\alpha\cdots SC$ ($\tau^{(2)}$), and $SC\cdots C^\alpha\cdots C^\alpha\cdots SC$ ($\tau^{(3)}$), respectively (Figure 2).

Each energy term is multiplied by an appropriate weight, w_x , and the terms corresponding to factors of order higher than 1 in the cluster-cumulant expansion of the RFE²⁴ are additionally multiplied by the respective temperature factors which were introduced in our earlier work²⁵ and which reflect the dependence of the first generalized-cumulant term in those factors on temperature, as discussed in refs 25 and 44. The factors f_n are defined by eq 2:

$$f_n(T) = \frac{\ln[\exp(1) + \exp(-1)]}{\ln\{\exp[(T/T_0)^{n-1}] + \exp[-(T/T_0)^{n-1}]\}} \quad (2)$$

where $T_0 = 300$ K.

The term $U_{SC_i SC_j}$ represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with solvent (water). The term $U_{SC_i p_j}$ denotes the excluded-volume potential of the side-chain-peptide-group interactions. The peptide-group interaction potential is split into two parts: the Lennard-Jones interaction energy between peptide-group centers (U_{pp}^{VDW}) and the average

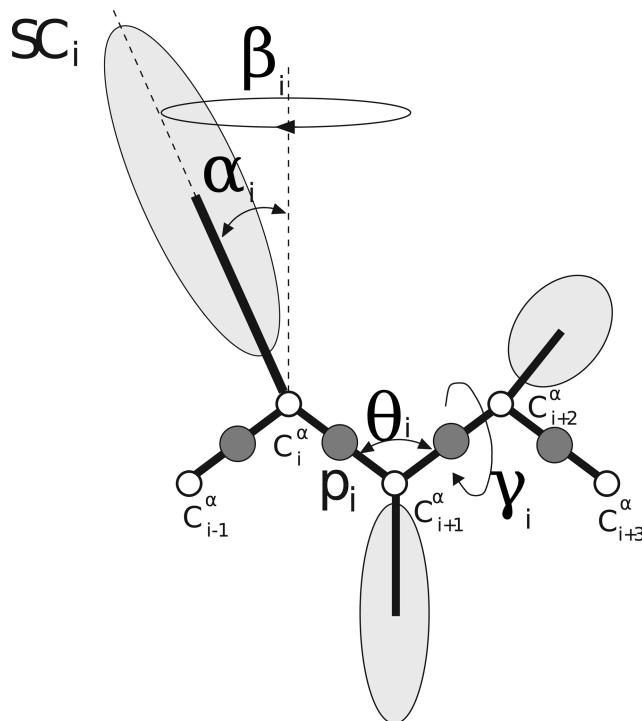


Figure 1. UNRES model of polypeptide chains. The interaction sites are peptide-group centers (p_i) and side-chain centers (SC_i) attached to the corresponding α -carbons with different $C^\alpha \cdots SC$ bond lengths, d_{SC} . The peptide groups are represented as dark gray circles and the side chains are represented as light gray ellipsoids of different size. The α -carbon atoms are represented by small open circles. The geometry of the chain can be described by the virtual-bond lengths, backbone virtual-bond angles θ_i , $i = 1, 2, \dots, n - 2$, backbone virtual-bond-dihedral angles γ_i , $i = 1, 2, \dots, n - 3$, and the angles α_i and β_i , $i = 2, 3, \dots, n - 1$, that describe the location of a side chain with respect to the coordinate frame defined by C^α_{i-1} , C^α_i , and C^α_{i+1} .

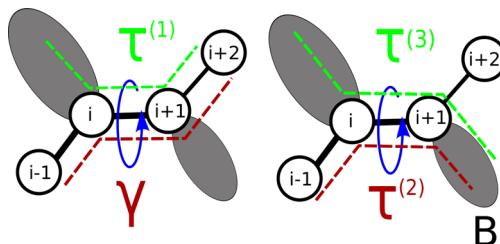


Figure 2. Illustration of backbone torsional angle γ (A, red) and side-chain backbone torsional angles $\tau^{(1)}$ (A, green), $\tau^{(2)}$ (B, red), $\tau^{(3)}$ (B, green).

electrostatic energy between peptide-group dipoles (U_{pp}^{el}); the second of these terms accounts for the tendency to form

$$\begin{aligned} U_{SC-corr}^{XY}(\tau^{(k)}) &= -\beta^{-1} \ln \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} \int_{\gamma: \tau^{(k)}(\gamma, \chi_X^{(1)}, \dots, \chi_Y^{(n)})=\tau^{(k)}} [\det H_X(\lambda_1, \tau^{(k)} - \pi - \lambda_2, \chi_X^{(1)}, \dots, \chi_X^{(m)})]^{-1/2} \\ &\quad \times [\det H_Y(\lambda_2, \lambda_3, \chi_Y^{(1)}, \dots, \chi_Y^{(n)})]^{-1/2} \exp\{\beta[e_X(\lambda_1, \gamma - \pi - \lambda_2, \chi_X^{(1)}, \dots, \chi_X^{(m)}) + e_Y(\lambda_2, \lambda_3, \chi_Y^{(1)}, \dots, \chi_Y^{(n)})]\} \\ &\quad \times d\lambda_1 d\lambda_2 d\lambda_3 d\chi_X^{(1)}, \dots, d\chi_X^{(m)} d\chi_Y^{(1)}, \dots, d\chi_Y^{(n)} d\gamma \end{aligned} \quad (3)$$

where $U_{SC-corr}^{XY}(\tau^{(k)})$ is the potential of mean force, X and Y are the first and the second residues, respectively, H is the Hessian (second derivative of energy) matrix, e is the potential energy for a given conformation, and m and n are the number of

backbone hydrogen bonds between peptide groups p_i and p_j . The terms U_{tor} , U_{tord} , U_b , U_{rot} , and U_{bond} are the virtual-bond-dihedral angle torsional terms, virtual-bond dihedral angle double-torsional terms, virtual-bond angle bending terms, side-chain rotamer, and virtual-bond-deformation terms; these terms account for the local properties of the polypeptide chain. The terms $U_{corr}^{(m)}$ represent correlation or multibody contributions from the coupling between backbone-local and backbone-electrostatic interactions, and the terms $U_{turn}^{(m)}$ are correlation contributions involving m consecutive peptide groups; they are, therefore, named turn contributions. The multibody terms are indispensable for reproduction of regular α -helical and β -sheet structures.^{24,45,46} U_{ssbond} is the energy term that describes the interactions between cysteine side chains; it has two minima, one corresponding to disulfide-bond formation and another one to nonbonded interactions,⁴⁷ and n_{SS} is the number of pairs of cystine residues. The $U_{SC-corr}$ terms are new physics-based side-chain backbone correlation potentials; in this work, those terms are based on physical models (calculated with the AM1 method) and not on the statistical analysis of the PDB, as previously.⁴² The AM1 semiempirical method was chosen as a compromise between feasibility and accuracy of computations. As shown in a previous study,⁴⁸ use of AM1 results in energy profiles qualitatively similar to those obtained by *ab initio* approaches but the energy barriers are reduced. This is not a problem, though, because the UNRES energy terms are scaled by energy-term weights, which are adjustable parameters.

The set of energy-term weights was determined by force-field calibration to reproduce the structure and folding thermodynamics of two selected training proteins:⁴⁹ the tryptophan cage (PDB code: 1L2Y)⁵⁰ and tryptophan zipper (PDB code: 1LE1). In this force field, all the energy terms (U 's) are physics-based apart from side-chain–side-chain (U_{SC-SC}) interactions, which we obtained by simulations in water to compute the potentials of mean force, and correlation terms (U_{corr}) which are knowledge-based potentials.

2.2. Determination of Potentials of Mean Force from AM1 Calculations. The potentials of mean force corresponding to the $U_{SC-corr}$ contributions to the UNRES energy function were determined from the potential energy surfaces of terminally blocked amino-acid residues, calculated in our previous work^{27,51} with the AM1 method of semiempirical quantum mechanics.⁵² The variables were the $\lambda^{(1)}$ and $\lambda^{(2)}$ dihedral angles, defined by Nishikawa et al.,⁵³ for rotation of the first and the second peptide group, respectively, about the $C^\alpha \cdots C^\alpha$ virtual-bond axes and the $\chi^{(1)}, \dots, \chi^{(m)}$ dihedral angles for rotation of heavy atoms of the side-chain, where m depends on the type of amino-acid residue (Figure 3). The grid sizes in all variables are summarized in Table 1.

The potentials of mean force were calculated from eq 3.

χ angles describing rotation of heavy atoms for residue X and Y, respectively.

The integration in backbone virtual-bond dihedral angles γ in eq 3 is carried out subject to the condition that a given value of

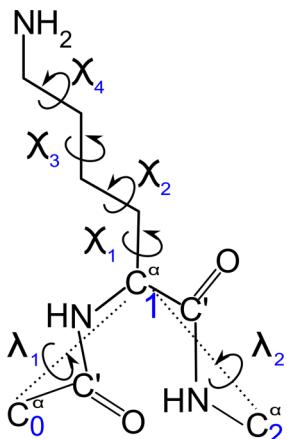


Figure 3. Illustration of variables used for calculation of potential energy surfaces with the example of the lysine residue.

Table 1. Grid Sizes in $\lambda^{(1)}$, $\lambda^{(2)}$, the Significant χ Angles (involving rotation of non-hydrogen atoms), and the Numbers of Grid Points for the 19 Natural Amino-Acid Residues with Side Chains (glycine is excluded, because it does not have a side chain)^a

n_χ^b	grid size (deg)						residue(s)	N_{grid}^c
	$\lambda^{(1)}$	$\lambda^{(2)}$	χ^1	χ^2	χ^3	χ^4		
0	—	30	—	—	—	—	Pro	12
	30	30	—	—	—	—	Ala, Gly	144
1	30	30	30	—	—	—	Cys, Ser, Thr, Val	1728
2	30	30	30	30	—	—	Asn, Asp, His, Ile, Leu, Phe, Trp, Tyr	20736
3	30	30	30	30	60	—	Glu, Gln, Met	124416
4	30	30	30	60	120	120	Arg, Lys	93312

^aThe amino-acid residues are grouped according to the number of significant χ angles. ^bNumber of significant χ angles. ^cNumber of grid points.

the virtual-bond dihedral angle $\tau^{(k)}$ has a given value. This angle depends on γ and the spherical angles α and β (Figure 1) of both side chains and, therefore, implicitly, on the side-chain angles $\chi_X^{(1)}, \dots, \chi_X^{(n)}$. To carry out this part of the integration

$$r_{XY}^{(m)} = \frac{\int_{-\pi}^{\pi} [\overline{U_{\text{SC-corr}}^{XY}}(\tau^{(m)}) - \langle U_{\text{SC-corr}}^{XY}(\tau^{(m)}) \rangle] \times [U_{\text{SC-corr}}^{XY}(\tau^{(m)}) - \langle U_{\text{SC-corr}}^{XY}(\tau^{(m)}) \rangle] d\tau^{(m)}}{\sqrt{\int_{-\pi}^{\pi} [\overline{U_{\text{SC-corr}}^{XY}}(\tau^{(m)}) - \langle U_{\text{SC-corr}}^{XY}(\tau^{(m)}) \rangle]^2 d\tau^{(m)} \int_{-\pi}^{\pi} [U_{\text{SC-corr}}^{XY}(\tau^{(m)}) - \langle U_{\text{SC-corr}}^{XY}(\tau^{(m)}) \rangle]^2 d\tau^{(m)}}} \quad (6)$$

where $\overline{U_{\text{SC-corr}}^{XY}}(\tau^{(m)})$ is the potential profile averaged over residue types (X and Y), $\langle U_{\text{SC-corr}}^{XY}(\tau^{(m)}) \rangle$ is the average value of this potential (averaged over $\tau^{(m)}$), and $\langle U_{\text{SC-corr}}^{XY}(\tau^{(m)}) \rangle$ is the average value of a given potential (averaged over $\tau^{(m)}$). These quantities are defined by eqs 7, 8, and 9, respectively; it should be noted that the overline denotes averaging over residue types and the brackets $\langle \rangle$ denote averaging over $\tau^{(m)}$.

$$\overline{U_{\text{SC-corr}}^{XY}}(\tau^{(m)}) = \frac{1}{N(m)M(m)} \sum_{\{X(m)\}} \sum_{\{Y(m)\}} U_{XY}(\tau^{(m)}) \quad (7)$$

where $N(m)$ and $M(m)$ are the numbers of residue types over which to sum for the respective type of the $\tau^{(m)}$

numerically, for each value of γ and for each orientation of both side-chain centroids, we computed the value of the respective angle $\tau^{(k)}$ and added the Boltzmann factor (to the integrand in eq 3) to the bin in $\tau^{(k)}$. The bin size was 30° . This value was a compromise between a discretization error of numerical integration, which should be as small as possible, and the feasibility of computing the potential-energy surfaces [it must be kept in mind that up to four side-chain dihedral angles χ are considered (Table 1)].

The presence of the $\gamma - \pi - \lambda_2$ terms in eq 3 arises from the fact that λ_2 is shared between residues X and Y in the dipeptide. The adiabatic energy surface, e_X , of a terminally blocked amino-acid residue X can be expressed as a function of the rotation of $\lambda^{(1)}$ and $\lambda^{(2)}$. The local angles of consecutive residues are related by eq 4:

$$\begin{aligned} \lambda_X^{(1)} &= \lambda_1 \\ \lambda_X^{(2)} &= \gamma - \pi - \lambda_2 \\ \lambda_Y^{(1)} &= \lambda_2 \\ \lambda_Y^{(2)} &= \lambda_3 \end{aligned} \quad (4)$$

The potentials for the virtual-bond-dihedral angles $\tau^{(1)}$, $\tau^{(2)}$, and $\tau^{(3)}$ that involve side-chains were determined from potential energy surfaces and were subsequently fitted to the one-dimensional Fourier series (eq 5) by the linear least-squares method:

$$U_{\text{SC-corr}}^{XY}(\tau^{(m)}) = a_{mo} + \sum_{n=1}^4 [a_{mn} \cos(n\tau^{(m)}) + b_{mn} \sin(n\tau^{(m)})] \quad (5)$$

where X and Y are types of residues and a_{mn} , $m = 1, 2, 3, n = 1, 2, \dots, 4$, and b_{mn} , $m = 1, 2, 3, n = 1, 2, \dots, 4$, are coefficients of the Fourier expansions of $U_{\text{SC-corr}}^{XY}(\tau^{(m)})$.

2.3. Analysis of the $U_{\text{SC-corr}}$ Potentials. To determine how the $U_{\text{SC-corr}}(\tau^{(m)})$ potentials for pairs of amino-acid residues are similar to each other, for each type of potential, we computed the correlation coefficients of the potentials with the potential averaged over all residue types in the first and in the second position (X and Y), respectively. The correlation coefficients are defined by eq 6.

angle: $M(1) = 20$, $N(1) = 19$; $M(2) = 19$, $N(2) = 20$; and $M(3) = N(3) = 19$, because glycine does not have a side chain and is excluded from summation when the respective $\tau^{(m)}$ angle depends on side-chain coordinates of residue X or Y, and $\{X(m)\}$ and $\{Y(m)\}$ are the sets of residue types over which to sum for a given type of τ angle.

$$\langle \overline{U_{XY}(\tau^{(m)})} \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{U_{XY}}(\tau^{(m)}) d\tau^{(m)} \quad (8)$$

$$\langle U_{\text{SC-corr}}^{XY}(\tau^{(m)}) \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} U_{\text{SC-corr}}^{XY}(\tau^{(m)}) d\tau^{(m)} \quad (9)$$

A value of $r_{XY}^{(m)} > 0.7$ indicates that $U_{\text{SC-corr}}^{XY}(\tau^{(m)})$ is correlated with (or, in other words, similar to) the average potential; $r_{XY}^{(m)}$

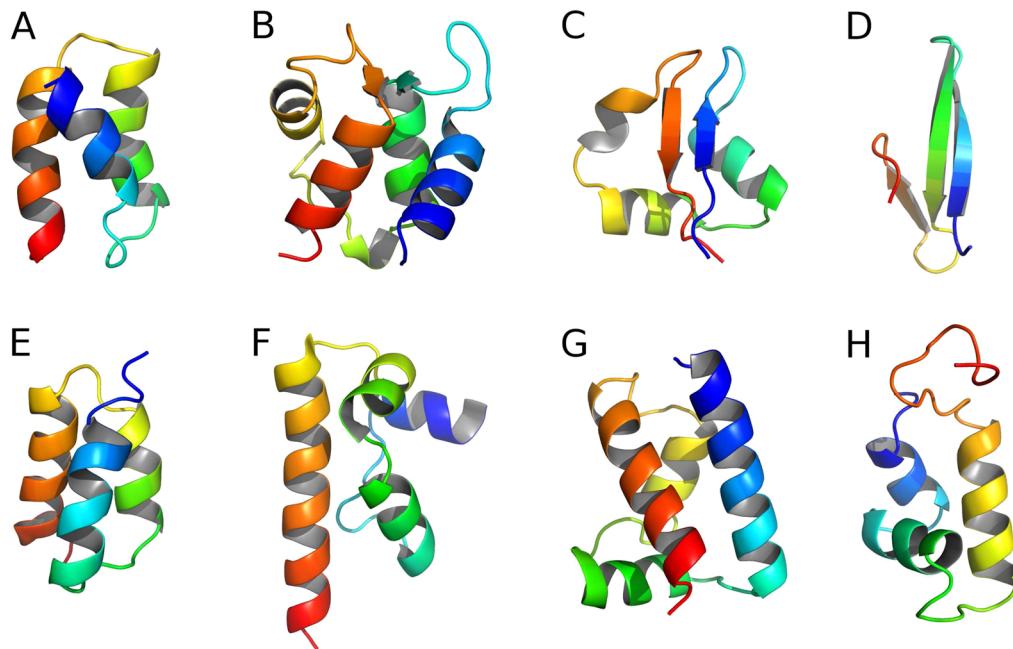


Figure 4. Cartoon representation of truncated experimental structures of training set proteins: A, 1BDD; B, 1CLB; C, 1E0G; D, 1E0L; E, 1GAB; F, 1KOY; G, 1POU; H, 1PRU. The chains are colored from blue (N-terminus) to red (C-terminus).

more negative than -0.7 indicates anticorrelation (this means 50% or more explained variance). Otherwise, the potential is not correlated or anticorrelated with the average potential.

2.4. Testing UNRES with the New Potentials. The version of the UNRES force field optimized with the 1L2Y and 1LE1 proteins was used to test the newly derived and implemented physics-based $U_{SC-corr}$ potentials. In contrast to the force-field used in our earlier work to test the statistical $U_{SC-corr}$ potentials⁴² (which was optimized with 1GAB²⁵), the force field optimized with 1LE1 and 1L2Y⁴⁹ has no knowledge-based local-interaction potentials.

A set of small proteins (37–76 residues) used by us previously^{25,42} for assessing the performance of previous versions of the UNRES force field was selected as a training set. The set consisted of the recombinant B domain (FB) of staphylococcal protein A (PDB code: 1BDD)⁵⁴ (α -helical structure), apo calbindin D9k from *Bos taurus* (PDB code: 1CLB)⁵⁵ ($\alpha + \beta$ structure), the LysM domain from *E. coli* (PDB code: 1E0G)⁵⁶ ($\alpha + \beta$ structure), the Fbp28 WW domain from *Mus musculus* (PDB code: 1E0L)⁵⁷ (β structure), the GA module (PDB code: 1GAB)⁵⁸ (α -helical structure), the DFF-C domain of DFF45/ICAD from *Homo sapiens* (PDB code: 1KOY)⁵⁹ (α -helical structure), the POU-specific domain from *Homo sapiens* (PDB code: 1POU)⁶⁰ (α -helical structure), and the purine repressor (PurR) DNA-binding domain from *E. coli* (PDB code: 1PRU)⁶¹ (α -helical structure). Short N- and C-terminal fragments of the training proteins were truncated from the structures determined by NMR if the conformations of the NMR ensemble exhibited large fluctuations in these parts. The truncated experimental structures are shown in Figure 4. The training proteins exhibit low sequence similarity to each other; the highest sequence identity between the training proteins was 24.32%, as determined in Figure 4 by the ClustalW2 program,⁶² with average identity of 9.69%.⁴²

Subsequently, 22 proteins listed in Table 2 were selected to test the force field. The sizes of the selected proteins

Table 2. Proteins Used To Test the Performance of the Force Field with the New Terms

PBD code	no. of residues	structure type
1L2Y	22	α
1LE1	12	β
1E0L	38	α
1NKL	78	α
1A6S	87	α
1BW6	56	α
1EO0	77	α
1FEX	59	α
1HYP	80	$\alpha + \beta$
1K40	126	α
1LEA	84	$\alpha + \beta$
1RES	43	α
1RJ	23	α
1TIG	94	$\alpha + \beta$
1YRF	35	α
2CRB	97	α
1ACP	77	α
1ENH	54	α
1FSD	28	$\alpha + \beta$
1LQ7	67	α
1PGA	56	α
2HEP	42	α

vary from 12 (for 1LE1) to 126 (for 1K40) amino-acid residues. Simulations were carried out for each protein of the set, using the force field with both the new terms and the best set of energy-term weights ($w_{SC-corr} = 0.25$, $w_{tor} = 1.3431$, $w_{tord} = 1.26571$) and with Table 2 the original force field.⁴⁹

All simulations were run with the use of Multiplexed Replica Exchange Molecular Dynamics (MREMD).^{26,63,64} This procedure provides much better coverage of the conformational space compared to canonical molecular dynamics. For each tested protein with each weight set, 64 trajectories were run at

Table 3. Weights of the New $U_{SC\text{-corr}}$ Potentials, the Torsional, and the Double-Torsional Potentials and Temperatures of Heat Capacity Peaks [K] for Eight Training Proteins and All Tested Sets of Energy-Term Weights

energy-term weight			temperatures [K] for proteins								
$w_{SC\text{-corr}}$	w_{tor}	w_{tord}	1BDD	1CLB	1E0G	1E0L	1GAB	1KOY	1POU	1PRU	Mean
0.00	1.84316	1.26571	320	325	330	315	320	330	325	330	324.4
0.00	1.34316	1.26571	320	335	325	325	330	325	340	330	328.8
0.25	1.84316	1.26571	325	330	335	325	335	340	325	335	331.3
0.25	1.59316	1.26571	320	325	335	315	320	335	330	330	326.3
0.25	1.34316	1.26571	325	320	325	315	325	330	335	330	325.6
0.25	1.09316	1.26571	330	335	320	320	325	330	340	330	328.8
0.25	1.34316	1.01571	305	325	320	325	325	325	335	330	323.8
0.50	1.84316	1.26571	325	335	340	315	345	345	335	345	335.6
0.50	1.34316	1.26571	315	325	335	320	325	335	320	335	326.3
0.50	0.84316	1.26571	330	335	320	320	330	330	330	325	327.5
0.57	1.84316	1.26571	320	330	335	320	325	340	325	330	328.1
1.00	1.84316	1.26571	335	340	345	330	350	340	340	355	341.9
1.00	0.00000	1.26571	335	335	315	325	335	345	325	325	330.0

32 temperatures (2 trajectories per temperature). The temperatures ranged from 200 to 500 K with a 10 K increment and, additionally, one pair of replicas was run at 295 K. Fifty million steps with length of 4.89 fs⁶⁵ (0.24 μ s total UNRES simulation time, which corresponds to about 0.24 ms because of UNRES time-scale extension resulting, in turn, from averaging out the fast degrees of freedom^{31,32}) were run for every protein considered.

The Berendsen thermostat⁶⁶ with the coupling parameter $\tau = 48.9$ fs was used to maintain constant temperature. For each protein, the simulations were started from the extended structure. The variable time step (VTS) algorithm³¹ was used to integrate the equations of motion.

The energy-term weights applied in the calculations are collected in the first three columns of Table 3. The weights of the torsional (w_{tor}) and double-torsional (w_{tord}) terms were modified from the original values determined in ref 49 by subtracting a value ranging from $w_{SC\text{-corr}}$ to $3 \times w_{SC\text{-corr}}$ because the $U_{SC\text{-corr}}$ potentials are likely to include some contributions from U_{tor} and U_{tord} . Reference runs with original values of w_{tor} and w_{tord} were also carried out.

To standardize the analysis of the resulting structures and trajectories, the sets of conformations after performing WHAM analysis^{25,67} were always divided into five clusters by using different cutoff values in Ward's minimum-variance method;⁶⁸ this number of clusters is the same as the maximum number of models of a given target that a group can submit in the CASP experiments.

For each training protein, clustering was carried out in two ways. In the first approach, the temperature to calculate the statistical weights of the conformations was chosen ≈ 10 K below the heat capacity peaks (shown in Table 3), as in ref 25. However, for some of the calculations, the heat-capacity curves possessed multiple peaks and the statistical weights of conformations related to these peaks were calculated, and clustering was carried out at 210 K, 240 K, 270 K, 290 K, and 310 K.

The C α root-mean-square deviation (C α RMSD) was used as a measure of the agreement between the calculated and the experimental structures. Because MREMD simulations produce conformational ensembles, mostly the ensemble-averaged RMSDs were used for analysis. The four analyzed RMSD values are defined by eqs 10–12, respectively.

$$\langle \rho \rangle(T_a) = \sum_i \rho_i w_i(T_a) \quad (10)$$

$$\langle \rho \rangle_{clust}^{\min}(T_a) = \min_I \sum_{i \in I} \rho_i w_i(T_a) \quad (11)$$

$$\rho^{\min} = \min_i \rho_i \quad (12)$$

where ρ_i is the C α RMSD of the i th conformation, T_a is the absolute temperature, and $w_k(T_a)$ is the weight of the k th conformation obtained in the MREMD simulations.

For each of the proteins of the test set, the lowest RMSD value from whole simulation (ρ^{\min}) and the average RMSD of the lowest-RMSD (the most native-like) cluster ($\langle \rho \rangle_{clust}^{\min}(T_a)$) at 290 K were determined.

3. RESULTS AND DISCUSSION

3.1. Analysis of the Derived $U_{SC\text{-corr}}$ Potentials of Mean Force. Plots of all 380 SC...C α ...C α ...C α [$U_{SC\text{-corr}}^{XY}(\tau^{(1)})$], all 380 C α ...C α ...C α ...SC [$U_{SC\text{-corr}}^{XY}(\tau^{(2)})$], and all 361 SC...C α ...C α ...SC [$U_{SC\text{-corr}}^{XY}(\tau^{(3)})$] potentials are shown in Figure S1 of Supporting Information. The plots of nine sample potentials (D–L) and of the three potentials (A–C) averaged over residue types ($\overline{U_{SC\text{-corr}}^{XY}}(\tau^m)$, $m = 1, 2, 3$; eq 7) are shown in Figure 5.

The average potentials (Figure 5A–C) are smooth and possess the global minimum at about $\tau^{(1)} \approx -60^\circ$ [for $\overline{U_{SC\text{-corr}}^{XY}}(\tau^{(1)})$], $\tau^{(2)} \approx 0^\circ$ [for $\overline{U_{SC\text{-corr}}^{XY}}(\tau^{(2)})$], and $\tau^{(3)} \approx 150^\circ$ [for $\overline{U_{SC\text{-corr}}^{XY}}(\tau^{(3)})$], respectively. The maxima are at $\tau^{(1)} \approx 120^\circ$, (for $\overline{U_{SC\text{-corr}}^{XY}}(\tau^{(1)})$), $\tau^{(2)} \approx -150^\circ$ (for $\overline{U_{SC\text{-corr}}^{XY}}(\tau^{(2)})$), and $\tau^{(3)} \approx -30^\circ$ (for $\overline{U_{SC\text{-corr}}^{XY}}(\tau^{(3)})$), respectively. The minima and maxima are thus separated by about 180°, which is understandable because one orientation corresponds to minimum steric repulsion of the peptide group linking the two central α -carbon atoms with the attached side chains and the other one to increased repulsion. Also, for this reason, the potentials are shifted with respect to each other; the shift angle reflects the phase-angle difference of the projection of a terminal C α ...C α and C α ...SC virtual bond onto the plane perpendicular to the central C α ...C α virtual bond. As can be seen from Figure 5A–C and SG–I, the average potential is almost identical with the potential for methionine; as can be seen from Supporting Information

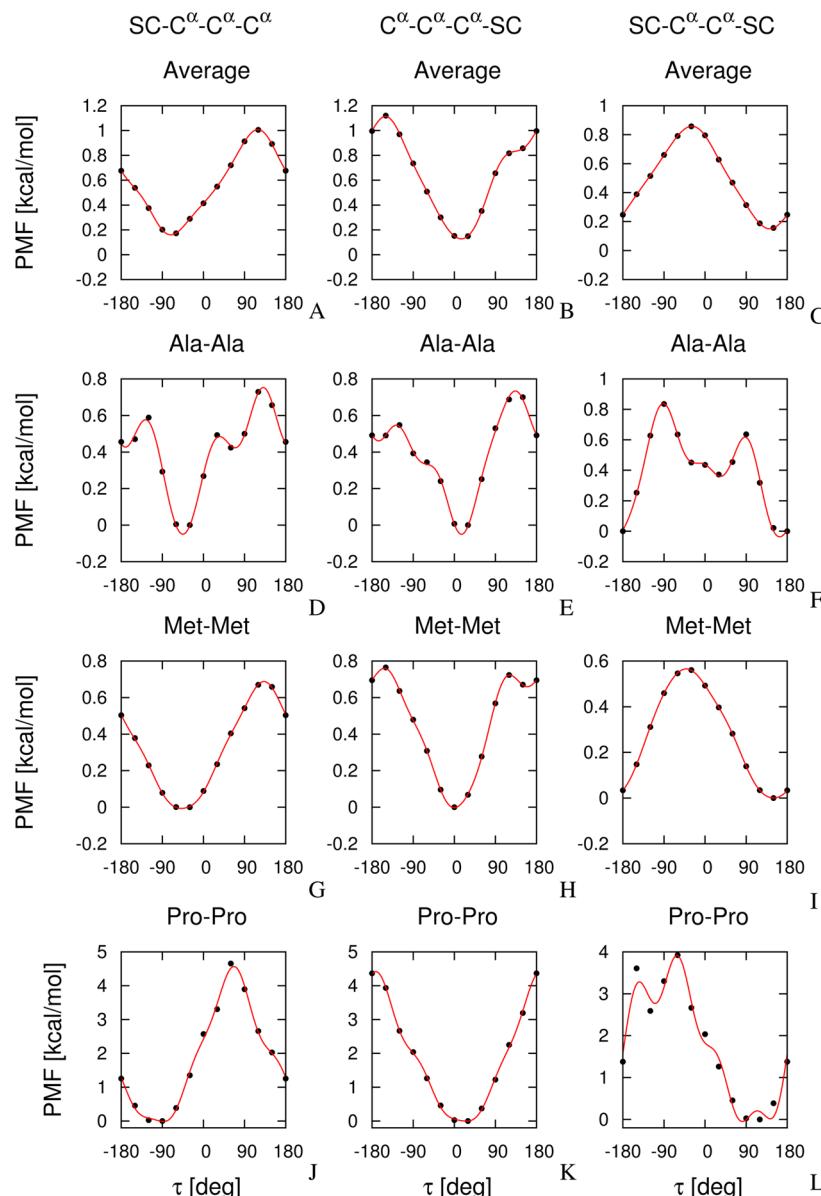


Figure 5. Plots of average side-chain-backbone potentials (A–C), $\overline{U_{SC\text{-corr}}^{XY}(\tau^{(m)})}$, $m = 1, 2, 3$; eq 7], and sample side-chain backbone correlation potentials: $U_{AlaAla}(\tau^{(m)})$, $m = 1, 2, 3$ (D–F), $U_{MetMet}(\tau^{(m)})$, $m = 1, 2, 3$, (G–I), and $U_{ProPro}(\tau^{(m)})$, $m = 1, 2, 3$ (J–L). Black circles represent the values of the dimensionless PMFs calculated from histograms (eq 3) and red lines represent one-dimensional Fourier series fits (eq 5) to the PMF values.

Figure S1, this also holds for other bulky nonbranched residues (e.g., Phe, Tyr, Trp, etc.). In contrast to this, as can be seen from Figure 5D–F and from Supporting Information Figure S1, small or branched residues such as, e.g., Ala, Val, and Leu have potentials that possess more fine structure. All these potentials, however, possess a similar free-energy span of about 1 kcal/mol. Proline (Figures 5J–L) is a special case and has a more pronounced energy span of about 5 kcal/mol. Another special case is glycine which does not have a side chain and has, therefore, potential patterns different from those of other residues (Supporting Information Figure S1). It should be noted that the $U_{SC\text{-corr}}^{XY}(\tau^{(3)})$ potentials are undefined for this residue. The different conformational preferences of proline and glycine residues were also found in other theoretical and experimental studies.^{69,70}

An inspection of the dominant conformations of the dipeptides corresponding to the global minima in the

$U_{SC\text{-corr}}^{XY}(\tau^{(m)})$ potentials ($m = 1, 2, 3$) shows that the backbone has an extended conformation and that hydrogen bonds between the carbonyl-oxygen atom of the preceding peptide group and the amide-hydrogen atom of the succeeding peptide group are formed at each residue (the C7 conformations), unless it is a proline. This is understandable because only the intraresidue energies are considered in eq 3. Thus, the specificity of the potentials lies in subtle details of the position of the minima and in differences in PMF barriers, depending on the kind of amino-acid side chains.

The correlation coefficients of the average potential with all other potentials, shown in Figure 6, confirm the above observations; for bulky residues such as cystine, methionine, phenylalanine, and tryptophan, the correlation coefficients with the average potentials are high if the type of potential includes such a side chain [or two such side chains for the $U_{SC\text{-corr}}(\tau^{(3)})$ potentials]. The correlation with the average potential is weak

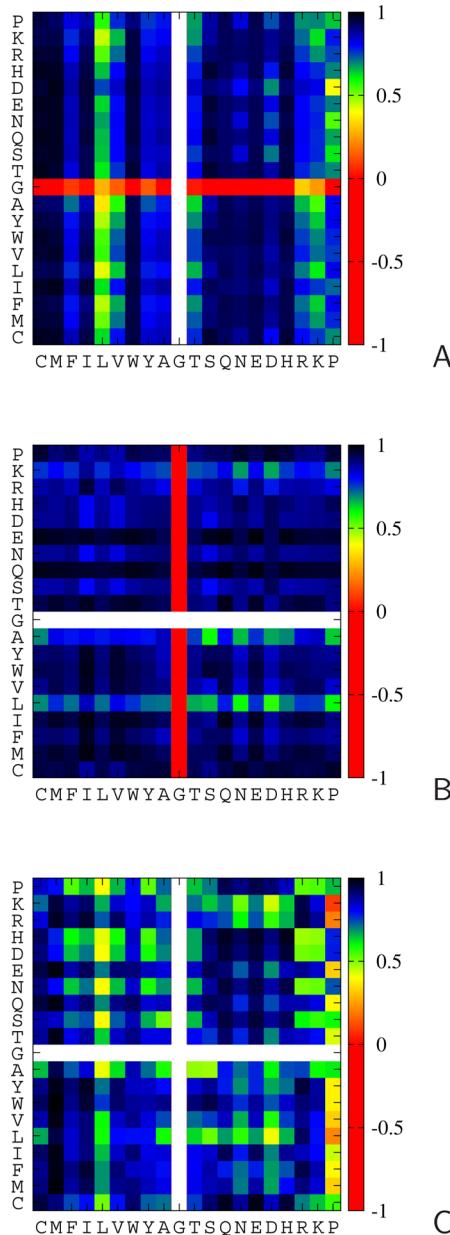


Figure 6. Color plots of the correlation coefficients of the respective $U_{SC\text{-corr}}^{XY}(\tau^{(m)})$ potentials with the respective $\overline{U_{SC\text{-corr}}^{XY}(\tau^{(m)})}$ potentials averaged over all residue types at positions X and Y (eq 6). A: $r_{XY}^{(1)}$ (for the $C^\alpha\cdots C^\alpha\cdots C^\alpha\cdots SC$ angles, $\tau^{(1)}$). B: $r_{XY}^{(2)}$ (for the $SC\cdots C^\alpha\cdots C^\alpha\cdots C^\alpha$ angles, $\tau^{(2)}$). C: $r_{XY}^{(3)}$ (for the $SC\cdots C^\alpha\cdots C^\alpha\cdots SC$ angles, $\tau^{(3)}$). The color scales are on the right bars of each panel. Types of residues X are on the abscissae and types of residues Y are on the ordinates. The color scale is shown on the right side of each panel.

for branched or small residues; additionally, it is weak for residues with oppositely charged side chains, because of hydrogen-bond formation (it should be noted that only uncharged side chains were considered when constructing energy maps).²⁷ The $U_{X-Gly}(\tau^{(1)})$ and $U_{Gly-X}(\tau^{(2)})$ potentials exhibit anticorrelation with the respective average potentials, which reflects the fact that glycine does not have a side chain. It can also be seen from Figure 6 that the side chain of the first residue has a greater influence on the potential than the side chain of the second residue. Therefore, the $U_{SC\text{-corr}}^{XY}(\tau^{(1)})$ and $U_{SC\text{-corr}}^{XY}(\tau^{(3)})$ potentials are more sequence specific than the

$U_{SC\text{-corr}}^{XY}(\tau^{(2)})$ potentials. This feature is manifested in that the specific $U_{SC\text{-corr}}^{XY}(\tau^{(2)})$ potentials (Figure 6B) are much more correlated with the average $U_{SC\text{-corr}}^{XY}(\tau^{(2)})$ than the $U_{SC\text{-corr}}^{XY}(\tau^{(1)})$ and $U_{SC\text{-corr}}^{XY}(\tau^{(3)})$ potentials (Figure 6A,C). This result is consistent with the analysis for statistical potentials,⁴² for which the $U_{SC\text{-corr}}^{XY}(\tau^{(2)})$ potentials showed greater correlation with the backbone torsional potentials than potentials for angles $\tau^{(1)}$ and $\tau^{(3)}$.

3.2. Performance of the New Potentials in ab Initio Simulations of Protein Structure. To preliminarily calibrate and test the performance of the new potentials, MREMD simulations were run on the training proteins for each set of $w_{SC\text{-corr}}$, w_{tor} , and w_{tord} values summarized in the first three columns of Table 3. For each run, heat-capacity curves were calculated during the progress of simulations and monitored for convergence, as described in our earlier work.^{25,42,64,71} Simulations were terminated when the heat-capacity curves calculated from at least two consecutive time windows of a given simulation overlapped closely. Sample plots of heat-capacity curves during the progress of the MREMD run for 1EOL are shown in Figure 7.

We noticed that the convergence of the heat capacity curves was faster and the heat-capacity profiles were narrower with the new potentials ($w_{SC\text{-corr}} > 0$) than without the new potentials ($w_{SC\text{-corr}} = 0$) (compare Figure 7A,B). When the value of w_{tor} was diminished by subtracting $w_{SC\text{-corr}}$, the convergence was even faster, but the heat-capacity profiles became wider (Figure 7C,D). Also, with increasing $w_{SC\text{-corr}}$, the average temperature of the heat capacity peaks increased by up to 17.5 K, from 324.4 for $w_{SC\text{-corr}} = 0$ to 341.9 for $w_{SC\text{-corr}} = 1$ (Table 3), which is understandable, because UNRES with the new terms has not yet been optimized to reproduce the thermodynamic properties of proteins. This optimization is currently being carried out with the maximum-likelihood method.

The values of ensemble-averaged RMSD ($\langle \rho \rangle(T_a)$) of the RMSD averaged over the conformations of the lowest-RMSD cluster ($\langle \rho \rangle_{clust}^{\min}(T_a)$) and the lowest RMSD found in a run (ρ^{\min}) averaged over all eight training proteins are shown in Figure 8 as bar diagrams. As can be seen from Figure 8, introducing the new physics-based $U_{SC\text{-corr}}$ potentials without tuning other parameters resulted in higher values of RMSD, while RMSD decreased on average when the new potentials were introduced together with reducing w_{tor} .

The lowest values of ρ^{\min} averaged over all eight training proteins (white bars in Figure 8) were observed for $w_{SC\text{-corr}} = 0.50$, $w_{tor} = 0.84316$, and $w_{tord} = 1.26571$ (by 0.55 Å) and for $w_{SC\text{-corr}} = 0.25$, $w_{tor} = 1.34316$, and $w_{tord} = 1.26571$ (by 0.62 Å), respectively. Because the lowest-RMSD structures were obtained for the force fields in which the original w_{tor} is reduced by subtracting twice the value of $w_{SC\text{-corr}}$, it can be concluded that the $U_{SC\text{-corr}}$ potentials include some information already encoded in the U_{tor} potentials.

For $\langle \rho \rangle_{clust}^{\min}(T_a)$ averaged over all eight training proteins, a significant decrease (by 0.41 Å; light-gray bars in Figure 8) at $T = 210, 240, 270, 290$, and 310 K was observed for only one set of weights: $w_{SC\text{-corr}} = 0.25$, $w_{tor} = 1.34316$, $w_{tord} = 1.26571$. However, for these weights, the $\langle \rho \rangle_{clust}^{\min}(T_a)$'s (dark-gray bars in Figure 8) computed at a temperature 10 K lower than that of the maximum of the respective heat-capacity curve are higher than those computed at the five temperatures listed above; this feature probably results from the fact that the force field with the new potentials was not yet parametrized using the

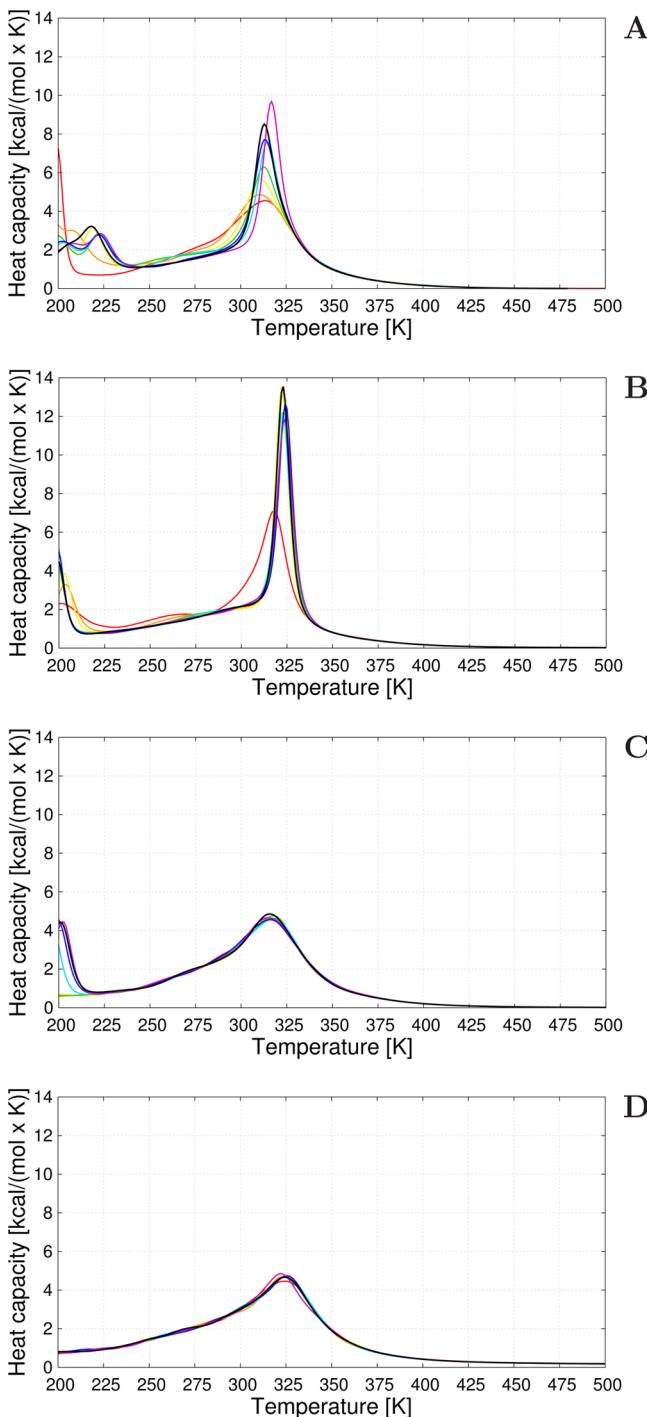


Figure 7. Sample convergence plots of the heat capacity of 1EOL for $w_{SC-corr} = 0.0$ (standard value), $w_{tor} = 1.84316$ (standard value), and $w_{tord} = 1.26571$ (standard value) (A); $w_{SC-corr} = 0.25$, $w_{tor} = 1.84316$ (standard value), and $w_{tord} = 1.26571$ (standard value) (B); $w_{SC-corr} = 0.25$, $w_{tor} = 1.34316$, and $w_{tord} = 1.26571$ (standard value); and (C) $w_{SC-corr} = 1.0$, $w_{tor} = 0.0$, and $w_{tord} = 1.26571$ (standard value) (D). Different colors denote heat-capacity curves for consecutive windows of the MREMD simulation, for the range from 10 000 to 50 000 000 MD steps divided into eight equal windows. Windows are colored in order: red, orange, yellow, green, cyan, blue, purple, black.

thermodynamic data of proteins. At temperatures 10 K lower than those of the respective heat-capacity peaks, the simulations with the following three sets of parameters resulted in reduced $\langle \rho \rangle_{clust}^{\min}$ values with respect to the force field without the new

terms: for $w_{SC-corr} = 0.25$, $w_{tor} = 1.34316$, $w_{tord} = 1.01571$ ($\langle \rho \rangle_{clust}^{\min}$ decreased by 0.46 Å), for $w_{SC-corr} = 0.25$, $w_{tor} = 1.34316$, $w_{tord} = 1.26571$ ($\langle \rho \rangle_{clust}^{\min}$ decreased by 0.28 Å), and for $w_{SC-corr} = 0.25$, $w_{tor} = 1.09316$, $w_{tord} = 1.26571$ ($\langle \rho \rangle_{clust}^{\min}$ decreased by 0.23 Å).

As can be seen from Figure 8, the ensemble-averaged RMSD values ($\langle \rho \rangle(T_a)$) did not change remarkably after implementing the new $U_{SC-corr}$ potentials (black bars in Figure 8). Together with the observation regarding $\langle \rho \rangle_{clust}^{\min}$ this observation suggests that the new potentials do not result in improving the ability of UNRES to predict overall folds but improve the quality of those UNRES-predicted structures which have correct global fold. In summary, weighting the new terms with $w_{SC-corr} = 0.25$, together with modifying the weights of the torsional and double-torsional terms ($w_{tor} = 1.34316$ and $w_{tord} = 1.26571$, respectively) consistently improves the quality of the calculated structures with respect to the force field without the torsional terms.⁴⁹

To assess the effect of reducing the torsional terms without introducing the new side-chain-torsional terms, control simulations were also run with $w_{SC-corr} = 0.0$, $w_{tor} = 1.34316$, $w_{tord} = 1.26571$. As can be seen from Figure 8, only reduction of the torsional terms results in a lower average $\langle \rho \rangle_{clust}^{\min}$; the other RMSDs increase by a small amount. Removing the new potentials with simultaneous reduction of w_{tor} and w_{tord} in particular results in deterioration of the quality of the calculated β -sheet segments, as can be seen from Figure 2S with the example of 1EOL protein (a β -structure protein).

To determine what sections of the calculated structures were most improved by introducing the new terms, plots of the deviations of the C^α atoms of the mean structures corresponding to the most native-like clusters from those of the experimental structures, as functions of residue number in the sequence, were constructed and analyzed. The deviations were calculated at optimal superposition of the computed structure on the experimental structure. As an example, the plots for 1CLB and 1KOY structures for which the most significant improvement was obtained (with $w_{SC-corr} = 0.25$, $w_{tor} = 1.34316$, $w_{tord} = 1.26571$) are shown in Figures 9 and 10, respectively. For 1CLB, the biggest improvement is observed for residues 20–25, 40–45, and 55–70, which covers the loop regions with small β -sheets, and for the C-terminal helix in the experimental structure. The calculated structure of 1KOY is improved mostly for residues 239–244, 264–268, 281–285, and 287–299, which cover the α -helical parts of the protein.

The structure of 1CLB calculated with the force field derived in ref 49 has an RMSD of 7.90 Å from the experimental structure. In the calculated structure, only the two middle α -helices are present, while the N- and the C-terminal α -helix are converted into β -sheets (Figure 11B). The structure predicted with new potentials ($w_{SC-corr} = 0.25$, $w_{tor} = 1.34316$, $w_{tord} = 1.26571$) (Figure 11B) has a RMSD of 6.04 Å and all α -helical parts are formed in that structure. The better agreement of that structure with the experimental 1CLB structure is even more evident when comparing the Global Distance Test (GDT) scores,⁷² for distance up to 4 Å which increased from 0.28 to 0.53, which means that the number of residues within 4 Å cutoff is greater by 89%. Both sets of parameters were unable to predict two small β -sheets correctly, which is the reason why α -helical parts with the new parameters are packed incorrectly.

A similar situation occurs for 1KOY. The structure calculated with the force field of ref 49 (without the new terms) forms the $\alpha + \beta$ -structure instead of α -helical structure (Figure 12B),

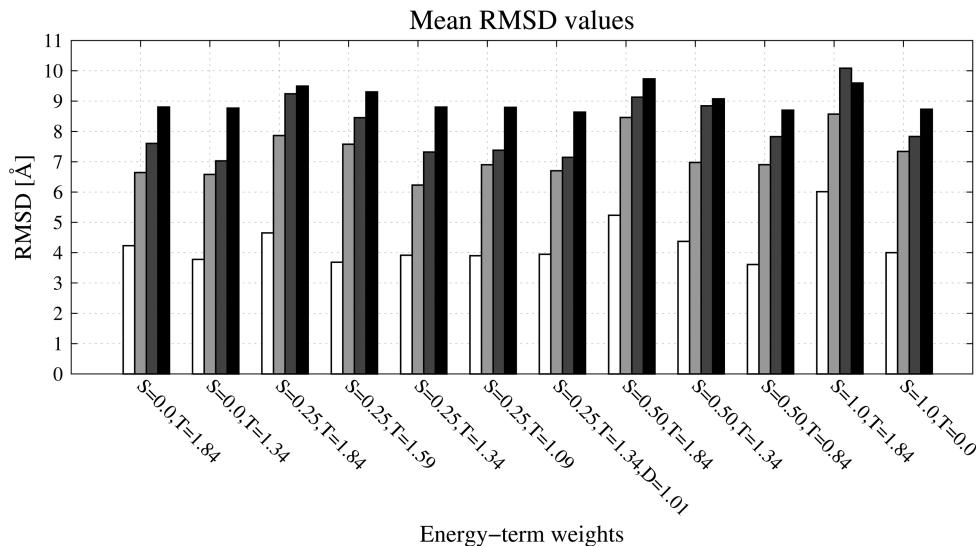


Figure 8. Bar diagrams of various RMSDs averaged over 8 proteins. White bars show the lowest RMSD obtained during the corresponding MREMD run [ρ^{\min} of eq 12], light-gray bars show the minimum of the cluster-averaged RMSD [$\langle \rho \rangle_{\text{clust}}^{\min}(T_a)$ of eq 11] from five temperatures of clustering (210, 240, 270, 290, and 310 K), dark-gray bars show the minimum of the cluster-averaged RMSD [$\langle \rho \rangle_{\text{clust}}^{\min}(T_a)$ of eq 11] from the temperature of clustering 10 K lower than the heat capacity peak, and the black bars show the RMSD averaged over the conformational ensemble generated during the MD run by WHAM in the last part of the simulation. S stands for $w_{\text{SC-corr}}$, T stands for w_{tor} , and D stands for w_{tord} .

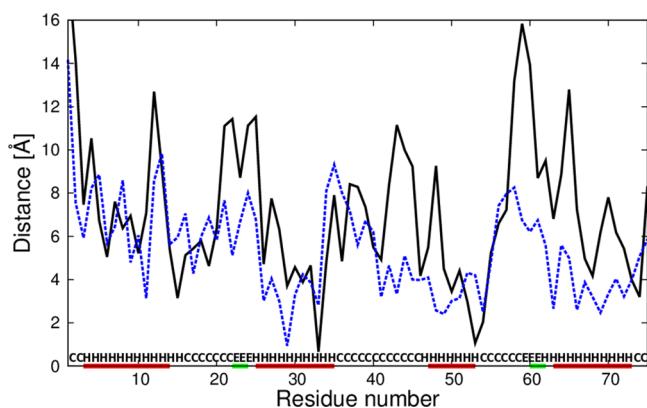


Figure 9. Plot of distances between C^α atoms of the average structures of the most native-like cluster of simulated 1CLB structures from the respective C^α atoms of the experimental 1CLB structure after optimal superposition. Solid black line: calculations with standard force field parameters ($w_{\text{SC-corr}} = 0.25$, $w_{\text{tor}} = 1.34316$, $w_{\text{tord}} = 1.26571$),⁴⁹ dashed black line: the force field that includes the $U_{\text{SC-corr}}^{\text{XY}}(\tau^{(m)})$, $m = 1, 2, 3$ potentials derived in this work ($w_{\text{SC-corr}} = 0.25$, $w_{\text{tor}} = 1.34316$, $w_{\text{tord}} = 1.26571$). Red horizontal lines on the abscissa mark α -helices in the experimental structure.

with RMSD from the native structure equal to 9.23 Å. Conversely, simulations with new parameters ($w_{\text{SC-corr}} = 0.25$, $w_{\text{tor}} = 1.34316$, $w_{\text{tord}} = 1.26571$) resulted in correct secondary structure with RMSD equal to 5.25 Å. Contrary to the previous work,⁴² in which implementation of statistical potentials results mainly in improvements of the loop regions, the physics-based version of the UNRES force field also contributes to correct recognition of regular secondary structure elements (α -helix and β -sheet).

To test the performance of the force field, calculations were carried out for another set of 22 proteins, which were not utilized in the estimation of the weight of the new terms (see Table 2 for the list of these proteins). The calculations were run with the best set of energy-term weights ($w_{\text{SC-corr}} = 0.25$, $w_{\text{tor}} = 1.34316$, $w_{\text{tord}} = 1.26571$) and, for reference, without the new

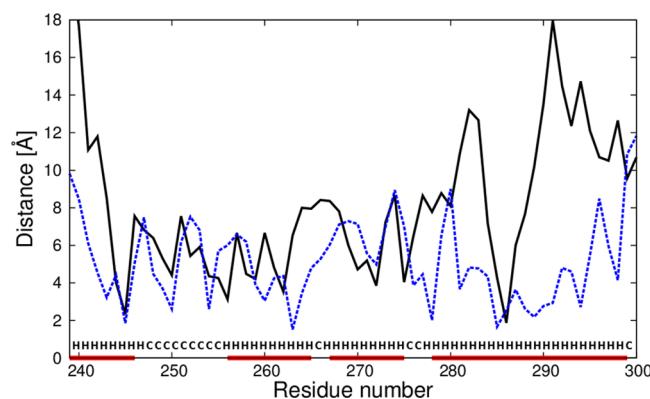


Figure 10. Plot of distances between C^α atoms of the average structure of the most native-like cluster of simulated 1KOY structure from the respective C^α atoms of the experimental 1KOY structure after optimal superposition. Solid black line: calculations with standard force field parameters ($w_{\text{SC-corr}} = 0.25$, $w_{\text{tor}} = 1.34316$, $w_{\text{tord}} = 1.26571$),⁴⁹ dashed black line: the force field that includes the $U_{\text{SC-corr}}^{\text{XY}}(\tau^{(m)})$, $m = 1, 2, 3$ potentials derived in this work ($w_{\text{SC-corr}} = 0.25$, $w_{\text{tor}} = 1.34316$, $w_{\text{tord}} = 1.26571$). Red horizontal lines on the abscissa mark α -helices in the experimental structure.

potentials ($w_{\text{SC-corr}} = 0.0$, $w_{\text{tor}} = 1.84316$, $w_{\text{tord}} = 1.26571$). The lowest RMSDs from the experimental structures are compared in Figure 13. The most significant decreases of ρ^{\min} with the new potentials (Figure 13) were observed for 1K40 (by 4.67 Å), 1TIG (by 1.63 Å), and 1LEA (by 1.09 Å), the average decrease of ρ^{\min} being 0.332 Å.

The RMSDs from the experimental structures corresponding to the mean structures of the most native-like clusters of the proteins studied are plotted, together with the respective error bars, in Figure 14. For each of the test proteins, the RMSD error was estimated by computing the standard deviation of the RMSD of the structures of the selected (most native-like) cluster from the RMSD of the mean structure of that cluster. It can be seen that noticeable improvement (over 1 Å) of $\langle \rho \rangle_{\text{clust}}^{\min}(T_a)$ (Figure 14) was observed for the following

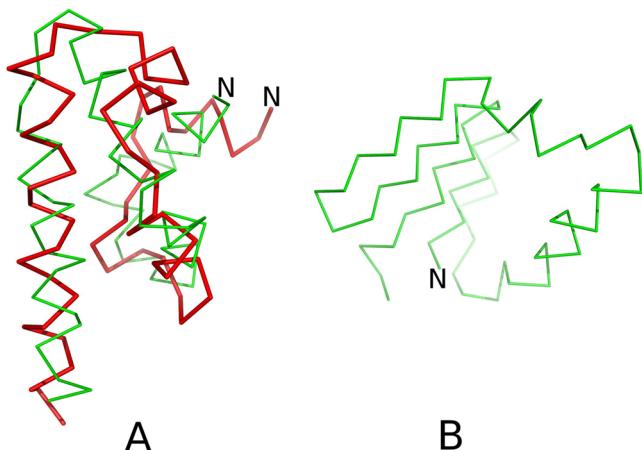


Figure 11. A: Superposition of the C^α -trace of the average structure of the most probable cluster of conformations of 1CLB obtained in MREMD simulations with inclusion of the $U_{SC-corr}^{XY}(\tau^{(m)})$ potentials derived in this work (green lines) on that of the experimental structure of 1CLB (red lines). B: Average structure of the most probable cluster of conformations of 1CLB obtained in MREMD simulations with $w_{SC-corr} = 0.0$, $w_{tor} = 1.84316$, $w_{tord} = 1.26571$ (without new potentials). The RMSDs from the experimental structures are 6.04 Å for panel A and 7.90 Å for panel B, respectively.

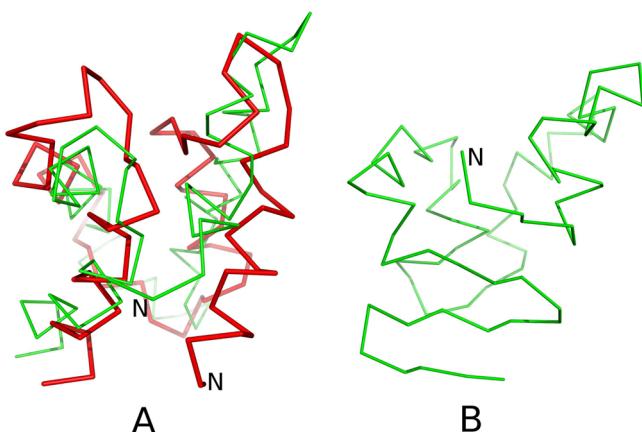


Figure 12. A: Superposition of the C^α -trace of the average structure of the most probable cluster of conformations of 1KOY obtained in MREMD simulations with inclusion of the $U_{SC-corr}^{XY}(\tau^{(m)})$ potentials derived in this work (green lines) on that of the experimental structure of 1KOY (red lines). B: Average structure of the most probable cluster of conformations of 1KOY obtained in MREMD simulations with $w_{SC-corr} = 0.0$, $w_{tor} = 1.84316$, $w_{tord} = 1.26571$ (without new potentials). The RMSDs from the experimental structure are 5.25 Å for panel A and 9.23 Å for panel B, respectively.

8 out of 22 test proteins: 1BW6, 1ENH, 1FEX, 1HYP, 1LQ7, 1NKL, 1YRF, and 2CRB with average improvement of 0.86 Å. For none of tested proteins was a noticeable RMSD increase observed (beyond 0.67 Å).

The biggest improvement in the reproduction of the experimental secondary structures was observed for 1FEX, for which the UNRES force field generated $\alpha + \beta$ -structures without the new terms instead of the native α -helical structure (Figure 15). These results, together with the results obtained with the training proteins, strongly suggests that that newly implemented physics-based potentials improve fold recognition.

To test the capacity of the UNRES force field with the new potentials to reproduce the content of secondary structure in

yet another way, as in our previous work,⁷³ we computed the change of the free-energy of α -helical structure formation upon the replacement of $x = \text{Gly}$ with a given amino-acid residue ($\Delta\Delta G_{hel}$). For this purpose, we used the KLALKLAL xx LKLALKLA host–guest peptides studied experimentally by Krause and co-workers.⁷⁴ The calculated values of $\Delta\Delta G_{hel}$ are defined by eq 14.⁷³

$$\Delta G_{hel}^{hel}(x) = -RT \ln \frac{f_{hel}(x)}{1 - f_{hel}(x)} \quad (13)$$

with

$$\Delta\Delta G^{hel}(\text{Gly} \rightarrow x) \equiv \Delta\Delta G^{hel} = \Delta G^{hel}(x) - \Delta G^{hel}(\text{Gly}) \quad (14)$$

where R is the universal gas constant, T is the absolute temperature, and $f_{hel}(x)$ is the ensemble-averaged fraction of α -helical structures in the ensemble for the host–guest peptide containing a pair of specific residues x . The temperature T was set at 298 K, as in the experiment. A residue was considered to be in the α -helical state if the peptide group preceding it formed a hydrogen-bonding contact with the third succeeding peptide group; the presence of a hydrogen-bonding contact was assessed based on the mean-field energy of interactions, which depends on the distance between the centers of the two peptide groups; the details of this method are described in ref 75.

The computed values of $\Delta\Delta G_{hel}$ are compared with the experimental data from ref 74 in Table 4. It can be seen from Table 4 that the values computed with the new potentials are about 1.9 kcal/mol closer to the experimental values⁷⁴ than those computed without the new potentials. This result suggests that the new potentials significantly improve the agreement of the thermodynamics of secondary-structure formation with the experiment.

4. SUMMARY

The new side chain–backbone correlation torsional potentials of mean force depending on the $SC \cdots C^\alpha \cdots C^\alpha \cdots C^\alpha$ ($\tau^{(1)}$), $C^\alpha \cdots C^\alpha \cdots C^\alpha \cdots SC$ ($\tau^{(2)}$), and $SC \cdots C^\alpha \cdots C^\alpha \cdots SC$ ($\tau^{(3)}$) angles (1121 potentials total) were derived from AM1 potential energy surfaces of terminally blocked amino-acid residues calculated in our earlier work.²⁷ By comparing the respective average potentials for each type of dihedral angle, the derived potentials were analyzed for similarity. Apart from obvious dissimilarity of the potentials involving the glycine and the proline residue, it was found that other residues can be grouped in three classes regarding the similarity of the potentials: bulky residues (e.g., cystine, methionine, phenylalanine, and tryptophan), branched and small residues (e.g., leucine, alanine, serine, threonine), and charged residues. This division largely overlaps with the classification of residues proposed by Solis and Rackovsky,⁷⁶ which was based on statistical analysis of physicochemical properties of amino-acid residues.

One-dimensional Fourier series were fitted to the obtained potentials, and the resulting formulas were implemented in the UNRES force field. The potentials were implemented in the UNRES force field package. It was found that introduction of the new potentials must be accompanied by reduction of the weight of the torsional terms to result in improvement of the calculated structures. On average the RMSD of the native-like clusters improved by 0.41 Å for the training set and by 0.86 Å for the test set of proteins, respectively. These results correspond to the best set of parameters: $w_{SC-corr} = 0.25$,

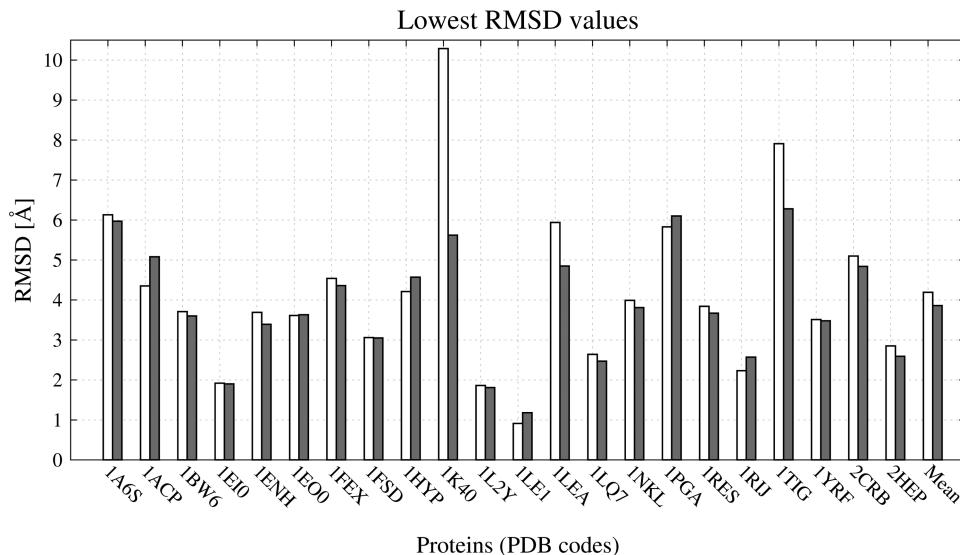


Figure 13. Bar diagrams of the lowest RMSDs from the experimental structures for 22 proteins from the testing set. White bars: the lowest RMSD obtained in MREMD simulations [ρ^{\min} of eq 12] with the UNRES force field without new terms. Light-gray bars: ρ^{\min} with the best set of new terms.

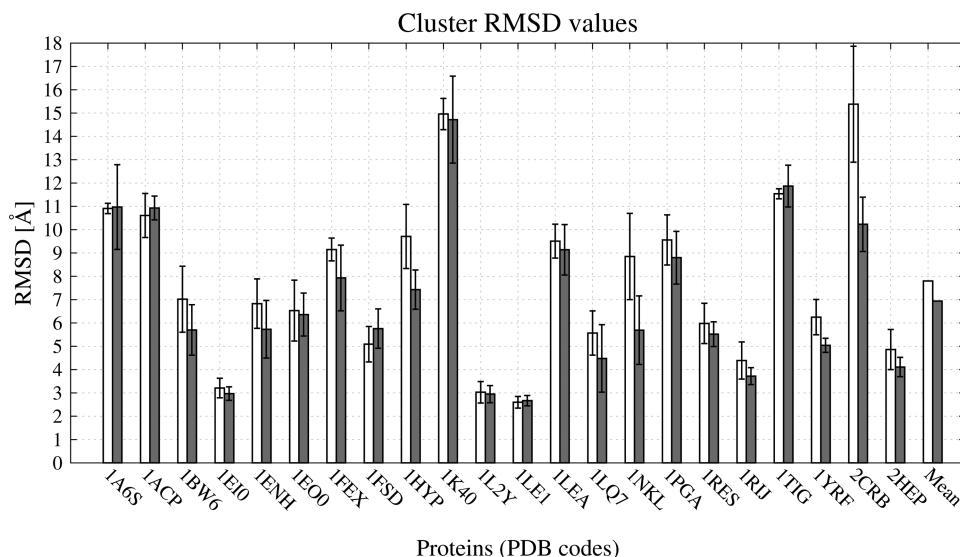


Figure 14. Bar diagrams of the RMSDs of the mean structures of the most native-like clusters [$\langle \rho \rangle_{clust}^{\min}(T_a)$ of eq 11] for the 22 test proteins corresponding to clustering at $T = 290$ K. White bars: the UNRES from force field without new terms. Light-gray bars: the UNRES force field with the new terms and optimal energy-term weights ($w_{SC-corr} = 0.25$, $w_{tor} = 1.34316$, $w_{tord} = 1.26571$). For each protein and each force field, the error bar represents the standard deviations of the RMSDs of the structures of the most native-like cluster from the RMSD of the average structure of that cluster.

$w_{tor} = 1.34316$, $w_{tord} = 1.26571$. Work on complete optimization of the new potentials to take into account the structural and thermodynamic properties of proteins is now underway in our laboratory.

From Figure 13 it can be concluded that, after introducing the new terms, the lowest RMSD from the native structures is about 4 Å on average for the set of 22 proteins with which the modified force field was tested. This value is about 2 times larger than the average value of 2 Å for the set of 12 proteins studied by Lindorff-Larsen et al. by using the all-atom simulations with the all-atom force field (the modified CHARMM force field⁷⁷) and the ANTON supercomputer.¹⁶ This difference might reflect the difference in the resolution of the coarse-grained and all-atom approaches. It should be noted, though, that the selection of the 12 proteins studied in ref 16

was made based on their foldability with the force field used, while the 22 proteins used in this study were not selected based on their foldability with UNRES. Therefore, it is also likely that the UNRES force field can still be improved by elaborating on the potentials of local interactions (as done in this work) and on force-field calibration. For example, for 1EI0 (38 residues) the lowest RMSD is below 2 Å and for 1LQ7 (67 residues) the lowest RMSD is below 3 Å.

The advantage of coarse-grained simulations over all-atom simulations is extension of the time scale and reduction of the cost of computations per MD step; both factors enable us to treat much larger systems at a much larger time scale than accessible all-atom simulations. For example, the simulations of ref 16 were carried out for over 1000 μ s to achieve folding, while the MREMD simulations of this work lasted only 0.25 μ s

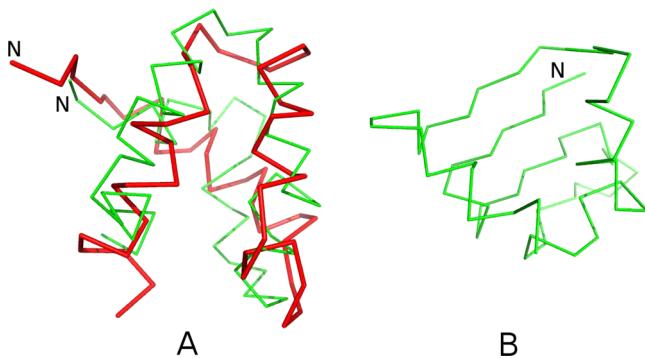


Figure 15. A: Superposition of the C^α -trace of the average structure of the most probable cluster of conformations of 1FEX obtained in MREMD simulations with inclusion of the $U_{SC\text{-corr}}^{XY}(\tau_i^{(m)})$ potentials derived in this work (green lines) on that of the experimental structure of 1FEX (red lines). B: the average structure of the most probable cluster of conformations of 1FEX obtained in MREMD simulations with $w_{SC\text{-corr}} = 0.0$, $w_{tor} = 1.84316$, $w_{tord} = 1.26571$ (without new potentials). The RMSDs from the experimental structure are 7.93 Å for panel A and 9.15 Å for panel B, respectively.

UNRES time per trajectory (which amounts to about 250 μ s real time per trajectory), and these simulations took from 4 to about 48 hours on a Beowulf cluster, depending on protein size. Within this time, all simulations converged (Figure 7). This time-scale extension and reduction of computational cost enables us to run several tens or hundreds of trajectories even on a Beowulf cluster which, in turn, enables us to use parallel-tempering and related sampling techniques such as, e.g., MREMD to estimate ensemble averages and folding thermodynamics reliably or to run multiple-trajectories canonical simulations to determine folding kinetics such as, e.g., in our earlier study of protein A⁷⁸ or in our recent study of the FBP 28 WW domain and its mutants.⁷⁹ Extension of the time

and size scale also enables us to treat biologically important processes such as, e.g., the opening of a Hsp70 chaperone recently studied by us with the used of molecular dynamics with UNRES.⁸⁰

ASSOCIATED CONTENT

S Supporting Information

Figure S1 presents plots of the PMFs for all side-chain backbone correlation potentials $U_{SC\text{-corr}}$ ($\tau_i^{(m)}$). Figure S2 presents bar diagrams for each of eight proteins from the training set. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*(P.K.) Phone: +48585235351. Fax: +48585235350. E-mail: vetinari@sun1.chem.univ.gda.pl.

Author Contributions

§(A.K.S. and P.K.) These authors contributed equally to the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by Grants MPD/2010/5 and Mistrz 7./2013 and START scholarship (100.2014) from the Foundation for Polish Science (FNP), Grant DEC-2011/01/N/ST4/01772 from the National Science Center of Poland, and Grants from the National Institutes of Health (GM-14312) and the National Science Foundation (MCB10-19767). This research was supported by an allocation of advanced computing resources provided by the National Science Foundation (<http://www.nics.tennessee.edu/>) and by the National Science Foundation through TeraGrid resources provided by the Pittsburgh Supercomputing Center. Computational resources

Table 4. Gibbs Free Energy Differences of α -Helical Structure Formation in the KLALKLALxxLKLALKLA Host–Guest Peptides with Respect to Glycine Host–Guest Peptide

residue name	$\Delta\Delta G_{hel;exp}^a$	$\Delta\Delta G_{hel,new}^b$	$\Delta\Delta G_{hel,old}^c$	$ \Delta\Delta G_{hel;exp} - \Delta\Delta G_{hel,new} $	$ \Delta\Delta G_{hel;exp} - \Delta\Delta G_{hel,old} $
Pro	6.64	0.81	0.48	5.83	0.33
Lys	-0.66	-0.51	-3.33	0.15	2.82
Arg	-1.33	-0.56	-3.48	0.77	2.92
His	-2.23	-0.35	-3.55	1.88	3.2
Asp	-2.12	-1.7	-3.49	0.42	1.79
Glu	-1.12	-0.59	-3.48	0.53	2.89
Asn	-1.44	-0.48	-3.36	0.96	2.88
Gln	-0.59	-0.58	-3.5	0.01	2.92
Ser	-1.43	-0.04	-3.31	1.39	3.27
Thr	-2.16	-0.43	-3.65	1.73	3.22
Ala	0.18	-0.06	-3.3	0.24	3.24
Tyr	1.88	-0.34	-3.75	2.22	3.41
Trp	-0.26	0.02	-3.58	0.28	3.6
Val	-2.24	0.09	-3.61	2.33	3.7
Leu	-0.23	0.44	-3.61	0.67	4.05
Ile	-0.89	0.24	-4.01	1.13	4.25
Phe	1.72	0.23	-3.84	1.49	4.07
Cys	0.58	0.23	-1.5	0.35	1.73
Met	0.15	0.21	-3.73	0.06	3.94
Average	-0.29	-0.18	-3.24	1.18	3.06

^aData from ref 74. ^bValues computed with the $U_{SC\text{-corr}}$ terms (this work) computed from eq 14. ^cValues computed from eq 14 without the $U_{SC\text{-corr}}$ terms (ref 73).

were also provided by (a) the supercomputer resources at the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, (b) the 952-processor Beowulf cluster at the Baker Laboratory of Chemistry, Cornell University, and (c) our 488-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk.

■ REFERENCES

- (1) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740–744.
- (2) Sieradzan, A. K.; Liwo, A.; Hansmann, U. H. E. *J. Chem. Theory Comput.* **2012**, *8*, 3416–3422.
- (3) Beckstein, O.; Sansom, M. S. P. *Phys. Biol.* **2006**, *3*, 147.
- (4) Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. *J. Med. Chem.* **2002**, *45*, 1412–1419 (PMID: 11906282).
- (5) Feller, S. E.; Pastor, R. W. *J. Chem. Phys.* **1999**, *111*, 1281.
- (6) Spackova, M.; Cheatham, T.; Ryjacek, F.; Lankas, F.; Van Meervelt, L.; Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2003**, *125*, 1759–1769 (PMID: 12580601).
- (7) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronen-born, A. M.; Schulten, K.; Aiken, C.; et al. *Nature* **2013**, *497*, 643–646.
- (8) Terstappen, G. C.; Reggiani, A. *Trends Pharmacol. Sci.* **2001**, *22*, 23–26.
- (9) Srinivasa Rao, V.; Srinivas, K. *J. Bioinf. Sequence Anal.* **2011**, *3*, 89–94.
- (10) Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S.; Kaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B. *Biopolymers* **2003**, *68*, 91–109.
- (11) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (12) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (13) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. *SC 2006 Conference, Proceedings of the ACM/IEEE (SC.06)*; IEEE: 2006; pp 43–43.
- (14) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. *J. Comput. Chem.* **2009**, *30*, 864–872.
- (15) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; et al. *Commun. ACM* **2008**, *51*, 91–97.
- (16) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517–520.
- (17) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134*, 3787–3791.
- (18) Sanbonmatsu, K. Y.; Joseph, S.; Tung, C.-S. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15854–15859.
- (19) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17845–17850.
- (20) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *J. Phys. Chem. B* **2013**, *117*, 12935–12942.
- (21) Piana, S.; Klepeis, J. L.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98–105.
- (22) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715–1731.
- (23) Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849–873.
- (24) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323–2347.
- (25) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. *Phys. Chem. B* **2007**, *111*, 260–285.
- (26) Liwo, A.; Czaplewski, C.; Oldziej, S.; Rojas, A. V.; Kaźmierkiewicz, R.; Makowski, M.; Murarka, R. K.; Scheraga, H. A. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G., Ed.; CRC Press: 2008; Chapter 8, pp 1391–1411.
- (27) Kozłowska, U.; Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2010**, *31*, 1154–1167.
- (28) Makowski, M.; Liwo, A.; Sobolewski, E.; Scheraga, H. A. *J. Phys. Chem. B* **2011**, *115*, 6119–6129.
- (29) Makowski, M.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2011**, *115*, 6130–6137.
- (30) Sieradzan, A. K.; Scheraga, H. A.; Liwo, A. *J. Chem. Theory Comput.* **2012**, *8*, 1334–1343.
- (31) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. *Phys. Chem. B* **2005**, *109*, 13798–13810.
- (32) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362–2367.
- (33) Liwo, A.; Lee, J.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5482–5485.
- (34) Lee, J.; Scheraga, H. A. *Int. J. Quantum Chem.* **1999**, *75*, 255–265.
- (35) Oldziej, S.; Czaplewski, C.; Liwo, A.; Chinchio, M.; Nania, M.; Vila, J. A.; Khalili, M.; Arnaudova, Y. A.; Jagielska, A.; Makowski, M.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7547–7552.
- (36) Liwo, A.; He, Y.; Scheraga, H. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16890–16901.
- (37) He, Y.; Mozolewska, M. A.; Krupa, P.; Sieradzan, A. K.; Wirecki, T. K.; Liwo, A.; Kachlishvili, K.; Rackovsky, S.; Jagiela, D.; Slusarz, R.; Czaplewski, C. R.; Oldziej, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 14936–14941.
- (38) Gay, J. G.; Berne, B. J. *J. Chem. Phys.* **1981**, *74*, 3316–3319.
- (39) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100–1120.
- (40) Wu, J.; Zhen, X.; Shen, H.; Li, G.; Ren, P. *J. Chem. Phys.* **2011**, *135*, 155104.
- (41) Shen, H.; Li, Y.; Ren, P.; Zhang, D.; Li, G. *J. Chem. Theory Comput.* **2014**, *10*, 731–750.
- (42) Krupa, P.; Sieradzan, A. K.; Rackovsky, S.; Baranowski, M.; Odziej, S.; Scheraga, H. A.; Liwo, A.; Czaplewski, C. *J. Chem. Theory Comput.* **2013**, *9*, 4620–4632.
- (43) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F. J.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
- (44) Shen, H.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2009**, *113*, 8738–8744.
- (45) Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *97*, 9412–9426.
- (46) Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Oldziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. *J. Comput. Chem.* **1998**, *19*, 259–276.
- (47) Chinchio, M.; Czaplewski, C.; Liwo, A.; Oldziej, S.; Scheraga, H. A. *J. Chem. Theory Comput.* **2007**, *3*, 1236–1248.
- (48) Oldziej, S.; Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **2003**, *107*, 8035–8046.
- (49) He, Y.; Xiao, Y.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2009**, *30*, 2127–2135.
- (50) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- (51) Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2010**, *31*, 1143–1153.
- (52) Stewart, J. J. *J. J. Comput.-Aided Mol. Design* **1990**, *4*, 1–105.
- (53) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797–806.
- (54) Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. *Biochemistry* **1992**, *31*, 9665–9672.
- (55) Skelton, N. J.; Krdel, J.; Chazin, W. *J. J. Mol. Biol.* **1995**, *249*, 441–462.
- (56) Bateman, A.; Bycroft, M. *J. Mol. Biol.* **2000**, *299*, 1113–1119.
- (57) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. *Nat. Struct. Biol.* **2000**, *7*, 375–379.
- (58) Johansson, M. U.; de Chateau, M.; Wikstrom, M.; Forsen, S.; Drakenberg, T.; Bjorck, L. *J. Mol. Biol.* **1997**, *266*, 859–865.
- (59) Fukushima, K.; Kikuchi, J.; Koshiba, S.; Kigawa, T.; Kuroda, Y.; Yokoyama, S. *J. Mol. Biol.* **2002**, *321*, 317–327.
- (60) Assa-Munt, N.; Mortishire-Smith, R. J.; Aurora, R.; Herr, W.; Wright, P. E. *Cell* **1993**, *73*, 193–205.

- (61) Nagadoi, A.; Morikawa, S.; Nakamura, H.; Enari, M.; Kobayashi, K.; Yamamoto, H.; Sampei, G.; Mizobuchi, K.; Schumacher, M. A.; Brennan, R. G. *Structure* **1995**, *3*, 1217–24.
- (62) Larkin, M.; Blackshields, G.; Brown, N.; Chenna, R.; McGettigan, P.; McWilliam, H.; Valentin, F.; Wallace, I.; Wilm, A.; Lopez, R.; Thompson, J.; Gibson, T.; Higgins, D. *Bioinformatics* **2007**, *23*, 2947–2948.
- (63) Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *84*, 775–786.
- (64) Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga, H. A. *J. Chem. Theory Comput.* **2009**, *5*, 627–640.
- (65) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. *J. Phys. Chem. B* **2005**, *109*, 13785–13797.
- (66) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (67) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (68) Späth, H. *Cluster Analysis Algorithms*; Halsted Press: New York, 1980.
- (69) Hermans, J. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3095–3096.
- (70) Grdadolnik, J.; Mohacek-Grosev, V.; Baldwin, R. L.; Avbelj, F. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1794–1798.
- (71) Nania, M.; Czaplewski, C.; Scheraga, H. A. *J. Chem. Theory Comput.* **2006**, *2*, 513–528.
- (72) Zemla, A.; Venclovas, C.; Moult, J.; Fidelis, K. *Proteins: Struct. Funct. Genet.* **2001**, *45* (S5), 13–21.
- (73) Sieradzan, A. K.; Niadzvedtski, A.; Scheraga, H. A.; Liwo, A. *J. Chem. Theory Comput.* **2014**, *10*, 2194–2203.
- (74) Krause, E.; Bienert, M.; Schmieder, P.; Wenschuh, H. *J. Am. Chem. Soc.* **2000**, *122*, 4865–4870.
- (75) Ołdziej, S.; Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 16934–16949.
- (76) Solis, A. D.; Rackovsky, S. *Proteins: Struct. Funct. Bioinf.* **2000**, *38*, 149–164.
- (77) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. *J. Comput. Chem.* **2009**, *30*, 1545–1615.
- (78) Khalili, M.; Liwo, A.; Scheraga, H. A. *J. Mol. Biol.* **2006**, *355*, 536–547.
- (79) Zhou, R.; Maisuradze, G. G.; Sunol, D.; Todorovski, T.; Macias, M. J.; Xiao, Y.; Scheraga, H. A.; Czaplewski, C.; Liwo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (51), 18243–18248.
- (80) Golas, E. I.; Maisuradze, G. G.; Senet, P.; Ołdziej, S.; Czaplewski, C.; Scheraga, H. A.; Liwo, A. *J. Chem. Theory Comput.* **2012**, *8*, 1750–1764.