Published in final edited form as:

J Phys Chem B. 2009 June 25; 113(25): 8738–8744. doi:10.1021/jp901788q.

An improved functional form for the temperature scaling factors of the components of the mesoscopic UNRES force field for simulations of protein structure and dynamics

Hujun Shen, Adam Liwo, and Harold A. Scheraga *

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca NY, USA 14853

Abstract

Coarse-grained or mesoscopic models of proteins and the corresponding force fields are of great importance because they enable us to reduce the folding simulation time by several orders of magnitude compared to the all-atom approach and, consequently, reach the millisecond time scale of simulations. In the coarse-grained UNRES model for simulations of protein structure and dynamics, developed by our group, each amino-acid residue is represented by a united side chain and a united peptide group located in the middle between the two neighboring α -carbon atoms, which assist only in the definition of the geometry.. The prototype of the UNRES force field has been defined as a potential of mean force or restricted free energy function corresponding to averaging out the degrees of freedom not present in the coarse-grained representation, which has further been approximated by a truncated Kubo cumulant series to enable us to derive analytical expressions for the corresponding terms. This force field should depend on temperature and, in its simplest form, a term corresponding to the cumulant of order n should be multiplied by $f_n = 1/T^{n-1}$. The temperature dependence has been introduced in recent work (*J. Phys. Chem. B*, 2007, 111, 260–285) and, in order to prevent too steep a variation with temperature, the factors at the n -th order cumulant terms were assumed to have a form $f_n = \ln[\exp(1) + \exp(-1)] / \ln\{\exp[(T/T_o)^{n-1}] + \exp[-(T/T_o)^{n-1}]\}$, where $T_o = 300$ K is the reference temperature. In this work, we have introduced a modified scaling factor $f_n = \ln[\exp(c) + \exp(-c)] / \ln\{\exp[c(T/T_o)^{n-1}] + \exp[-c(T/T_o)^{n-1}]\}$ where c is an adjustable parameter, and determined c by fitting the analytical approximation of the temperature dependence of the virtual-bond torsional term corresponding to rotation about the $C^\alpha \dots C^\alpha$ virtual bond in terminally-blocked di-alanine to the respective potential of mean force calculated from the MP2/6-31G(d,p) ab initio energy surfaces of terminally-blocked alanine (Ac-Ala-NHMe) and, independently, by optimizing it to obtain a sharp heat-capacity curve and the lowest ensemble-averaged root-mean-square deviation over the C^α atoms of 1GAB used as a training protein. Both approaches gave consistent results, and $c = 1.4$ has been selected as the optimal value of this parameter. The force field with the new temperature-scaling factors has been optimized using 1GAB as the training protein. The new force field has been tested on a series of medium size α -helical proteins and found to perform better than that with the original temperature-scaling factors.

Keywords

protein folding; united-residue force fields; potentials of mean force; temperature dependence; replica exchange molecular dynamics

*Author to whom correspondence should be addressed. Phone: (607) 255-4034, fax: (607) 254-4700, has5@cornell.edu.

1. Introduction

Coarse-grained or mesoscopic models of proteins have received much attention because they enable us to reduce the folding simulation time by several orders of magnitude compared to all-atom models.^{1–4} In the past decade, we have been developing a physics-based united-residue force field, which is referred to as UNRES.^{5–20} With mesoscopic dynamics implemented,^{14–16} UNRES was shown to offer a 4000-fold speed-up compared to all-atom molecular dynamics and, consequently, to be capable of simulating ab initio folding of medium-sized proteins; this feature is due primarily to averaging out the secondary degrees of freedom, which results in smoothing the energy landscape compared to an all-atom representation.^{15,16} In contrast to most other coarse grained models, which are largely knowledge-based potentials,^{21–23} or, in the case of force fields parameterized on a physical basis, are composed of terms taken directly from all-atom force fields,^{24,25} UNRES has been carefully derived, based on the physics of interactions, as a potential of mean force (PMF) or restricted free energy (RFE) function of the united-residue chain^{7,8} in which the degrees of freedom that do not belong to the coarse-grained representation are averaged out; a similar approach to treat protein systems, based on force averaging, has recently been proposed by Voth and coworkers.²⁶ Factorization of the total RFE into components corresponding to parts of the polypeptide chain and expansion of each of the factors in a series of Kubo generalized cumulants²⁷ enabled us^{7,8} to derive approximate analytical formulas for the respective terms of the force field, including the multibody or correlation terms. The UNRES force field was ultimately parameterized^{9,11,12} based on the concept of a hierarchical protein energy landscape.²⁸ Subsequently,^{19,20} in order to accelerate the calculation of ensemble-averaged quantities, we have implemented the replica-exchange^{29–34} (REMD) and multiplexing replica exchange^{35,36} (MREMD) extensions of molecular dynamics. For a detailed description of UNRES, the reader is referred to our earlier work^{5–20} and, for a discussion of different coarse-grained approaches to treat biomolecular and soft-material systems, the reader is referred to the recent book edited by Voth⁴ and to recent review articles.^{1–3}

The definition of the prototype of the UNRES energy function as a potential of mean force implies its dependence on temperature. In our recent work,¹⁷ we proposed a form of temperature dependence by multiplying the respective components of the generalized-cumulant expansion by appropriate temperature-dependent factors, which are given by eq 1 of the Methods section. In this paper, we introduce a more general parametric functional form of the scaling factors, and determine the parameter by fitting the analytical virtual-bond-dihedral angle potential for rotation about the Ala-Ala C^α...C^α virtual bond to the set of the respective PMF's calculated by numerical integration from the MP2/6-31G(d,p) ab initio energy surfaces of the Ac-Ala-NHMe system and, independently, optimized the parameters using 1GAB as a training protein.

The paper is organized as follows. In section 2.1, we recall the UNRES model and the force field, together with the temperature dependence of the force field and introduce the improved functional form of the temperature factors. In section 2.2, we describe the calculation of the temperature-dependent virtual-bond torsional energy surfaces. In section 2.3, we provide the details of MREMD simulations and the evaluation of thermodynamic quantities based on the results of MREMD simulations. In sections 3.1 and 3.2, we describe the parameterization of the temperature-scaling factors based on virtual-bond-torsional potentials and by calibration using the 1GAB protein, respectively. Finally, in section 3.3, we describe the optimization of the force field with the new scaling factors and tests of the optimized force field with five selected α -helical proteins.

2. Methods

2.1. UNRES model of polypeptide chains and force field

In the UNRES model,^{5–20} a polypeptide chain is represented by a sequence of α -carbon atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p). Each united peptide group is located in the middle between two consecutive α -carbons. Only these united peptide groups and the united side chains serve as interaction sites; the α -carbons serve only to define the chain geometry, as shown in Figure 1. As mentioned in the Introduction, the UNRES force field has been derived as a restricted free energy (RFE) function of an all-atom, polypeptide chain plus the surrounding solvent, where the all-atom energy function is averaged over the degrees of freedom that are lost when passing from the all-atom to the simplified system (viz., the degrees of freedom of the solvent, the dihedral angles, χ , for rotation about the bonds in the side chains, and the torsional angles, λ , for rotation of the peptide groups about the $C^\alpha \dots C^\alpha$ virtual bonds).^{7,8} The RFE is further decomposed into factors coming from interactions within and between a given number of united interaction sites.⁸ Expansion of the factors into generalized Kubo cumulants²⁷ enabled us to derive approximate analytical expressions for the respective terms, including the multibody or correlation terms, which are derived in other force fields from structural databases.^{21–23} The theoretical basis of the force field is described in detail in our earlier work.^{5–8} The UNRES energy function is expressed by eq 1.

$$\begin{aligned}
 U = & w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{p_i p_j} + w_{pp}^{el} f_2(T) \sum_{i < j-1} U_{p_i p_j} \\
 & + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) + w_{tord} f_3(T) \sum_i U_{tord}(\gamma_i, \gamma_{i+1}) + w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) \\
 & + \sum_{m=3}^6 w_{corr}^{(m)} f_{m-1}(T) U_{corr}^{(m)} + w_{turn}^{(3)} f_2(T) U_{turn}^{(3)} + w_{turn}^{(4)} f_3(T) U_{turn}^{(4)} + w_{turn}^{(6)} f_5(T) U_{turn}^{(6)} \\
 & + w_{bond} \sum_i U_{bond}(d_i) + w_{SS} \sum_i U_{SS,i}
 \end{aligned} \tag{1}$$

where $U_{SC_i SC_j}$, $U_{SC_i p_j}$, and $U_{p_i p_j}$ denote the effective free energies of interaction between the side chains in water, between the side chains and the peptide groups, and between the peptide groups, respectively, U_{bond} , U_b , and U_{rot} are the effective potentials for stretching of virtual bonds, for bending the virtual-bond angles, and for the energetics of the rotameric states of virtual side chains, respectively, U_{tor} and U_{tord} are torsional and double-torsional potentials for rotation about the $C^\alpha \dots C^\alpha$ virtual bonds, $U_{corr}^{(m)}$, $m=3,4,5,6$,⁸ are correlation or multibody terms pertaining to coupling between backbone-local and backbone-electrostatic interactions, $U_{turn}^{(m)}$, $m=3,4,6$,⁸ are the correlation terms pertaining to pairs of interacting peptide groups separated by 1, 2, and 4 peptide groups, respectively (there is no $U_{turn}^{(5)}$ term⁸), (where m denotes the order of correlation), and U_{SS} is a term accounting for the energetics of disulfide bonds.¹⁸ In the recent versions of UNRES,¹² we use correlation terms of the order up to 4, which saves computer time and provides as good results as using the correlation terms of higher order. Each of these terms is multiplied by an appropriate weight, w (the initial weights used in this work were optimized in ref ¹⁴ and are given in the second column of Table 1) and, wherever appropriate, by the temperature factors $f_n(T)$ defined by eq 2, introduced in ref ¹⁷.

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

where T_0 is the reference temperature and we set $T_0 = 300$ K, and where n is the lowest order of energy component in terms of the Kubo cumulant expansion.²⁷ With these scaling factors, the temperature-dependent force field was reparameterized based on decoys corresponding to canonical ensembles at finite temperatures with our revised hierarchical optimization procedure. The force field optimized using 1GAB as a training protein proved successful to predict structures of some α proteins with simple topologies and reasonably well to predict more complex α proteins.¹⁷

The functional form of the temperature-scaling factor defined by eq 2 originates from the cluster-cumulant expansion of the respective factors of the RFE. In this expansion, a cumulant of order n is multiplied by $T^{-(n-1)}$. Consequently, a plausible functional form of the temperature-scaling factor could be $f_n(T) = (T_0/T)^{n-1}$, where T_0 is an arbitrary reference temperature. However, such a functional form implies too steep a dependence on temperature for $T < T_0$ and, hence, we introduced the form given by eq 2.¹⁷ In this work, we propose an extended parametric form of $f_n(c; T)$, as given by eq 3,

$$f_n(c; T) = \frac{\ln[\exp(c) + \exp(-c)]}{\ln[\exp[c(T/T_0)^{n-1}] + \exp(-c(T/T_0)^{n-1})]} \quad (3)$$

where c is an adjustable parameter. It can be seen that eq 2 can be obtained from eq 3 by setting $c = 1$. In this work, we varied c from 0.8 to 5.0 to select its optimal value.

2.2 Calculation of the virtual-bond-torsional potential-surfaces

As mentioned in the Introduction, in one approach to the determination of the factor c in eq 3, we used the virtual-bond torsional potential, U_{tor} , for rotation about the $C^\alpha \dots C^\alpha$ virtual bond of the Ala-Ala sequence. We have chosen alanine because it represents all natural amino acids except proline and glycine as far as backbone-local interactions are concerned.^{8,10} This potential can be calculated numerically from the energy surfaces of the terminally-blocked L-alanine residue, as given by eq 7 of ref ¹⁰, which we quote here as eq 4 for the reader's convenience.

$$\begin{aligned} F_{tor,XY}(\gamma) = & -RT \ln \left\{ \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \exp \left[-\frac{1}{RT} [e_X(\lambda_1, \gamma - \pi - \lambda_2) + e_Y(\lambda_2, \lambda_3)] \right] d\lambda_1 d\lambda_2 d\lambda_3 \right\} \\ & + RT \ln \left\{ \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \exp \left[-\frac{1}{RT} e_X(\lambda_1, \lambda_2) \right] d\lambda_1 d\lambda_2 \right\} \\ & + RT \ln \left\{ \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \exp \left[-\frac{1}{RT} e_X(\lambda_2, \lambda_3) \right] d\lambda_2 d\lambda_3 \right\} \end{aligned} \quad (4)$$

where e_X and e_Y are the Ramachandran surfaces for the terminally-blocked residues of type X and Y, respectively (in this work $X = Y = \text{Ala}$, and we implemented the MP2/6-31G(d,p) ab initio energy surfaces $e_{Ala}(\lambda_1, \lambda_2)$ calculated in ref ¹⁰), R is the universal gas constant, and T is the absolute temperature. The angles λ_1 , λ_2 , and λ_3 are angles for rotation of the peptide groups about the respective $C^\alpha \dots C^\alpha$ virtual bonds defined by Nishikawa et al.³⁷ and γ is the virtual-bond dihedral angle. The dipeptide system with which to define U_{XY} is illustrated in Figure 2 (which is Figure 3a in ref ¹⁰).

Having calculated $F_{tor, AlaAla}$ from eq 4, we can now determine the constant c in eq 3 by minimization of the following target function, $\Phi(c)$:

$$\Phi(c) = \sum_T \sum_r \left[F_{\text{tor,AlaAla}}(\gamma, T) - U_{\text{tor,AlaAla}}(\gamma) \times f_2(T; c) \right]^2 \quad (5)$$

where $U_{\text{tor,AlaAla}}$ is the analytical torsional potential at the reference temperature $T_0 = 300$ K, expressed as a truncated Fourier series in γ [eq 9 in ref ¹⁰ and eq 6 below] and determined in ref ⁷.

$$U_{\text{tor,XY}}(\gamma) = a_0 + \sum_{m=1}^N a_m \cos(m\gamma) + b_m \sin(m\gamma) \quad (6)$$

where N is the order of the Fourier series; for $U_{\text{tor,AlaAla}}$, $N = 9$.

2.3 Simulation Details

We used the MREMD method^{35,36} which we have recently implemented in UNRES^{19,20} to carry out simulations with the UNRES force field. This approach results in much faster convergence of ensemble-based quantities compared to canonical MD simulations.^{19,20} All simulations were carried out with the Berendsen thermostat,³⁸ the implementation of which in UNRES/MD has been described in our earlier papers.^{15,16} Consequently, the random and stochastic forces were not included. In our earlier work,^{15,16} we have shown that Newtonian dynamics with the Berendsen thermostat leads to faster simulated folding compared to that of Langevin dynamics, which justifies its use in the present work, where the purpose of the simulations was to obtain converged thermodynamic functions and ensemble averages and not to determine kinetic functions and ensemble averages or kinetics of folding. The coupling parameter for the Berendsen thermostat was assumed to be $\tau = 1$ mtu (where 1mtu = 48.9 fs), as in our earlier work.¹⁵ The velocity Verlet (VV) algorithm,³⁹ with the variable time step extension developed in our earlier work,¹⁴ was used, and the basic time step in integrating the equations of motion was 5 fs. Each MREMD simulation was carried out at 20 temperatures over the range of temperatures from 200 to 440 K, 6 trajectories per temperature (a total of 120 replicas), and 20,000,000 steps per replica. Replica exchanges were attempted every 20,000 steps during the simulation.

To compute thermodynamic quantities (such as the partition function, total energy and heat capacity) from the results of the simulations carried out at different temperatures, the weighted histogram analysis method (WHAM)⁴⁰ was used, as described in ref ¹⁷. The expressions for thermodynamic quantities are given by eqs.17–21 of ref ¹⁷.

The drawings of the structures of the proteins considered in this work were prepared with the MOLMOL program.⁴¹

3 Result and Discussion

3.1 Determination of the factor c in eq 3 by analysis of temperature dependence of the $F_{\text{tor,AlaAla}}$ term

First, we computed the $F_{\text{tor,AlaAla}}$ PMF surfaces defined by eq 4 for the temperature range from $T = 200$ K to $T = 600$ K. The PMFs at four representative temperatures ($T = 200, 250, 300$, and 350 K) are plotted in Figure 3. It can be seen that the PMF profiles calculated from the MP2/6-31G(d,p) ab-initio-method Ramachandran surface are almost symmetric, with the torsional potential disfavoring conformations with too small virtual-torsional angles (e.g. $\gamma = 0^\circ$) and favoring extended conformations (e.g. $\gamma = 180^\circ$). Specifically, the Ala-Ala torsional

potential has three maxima at $\gamma = 0^\circ$ and $\gamma = \pm 135^\circ$, and three minima at $\gamma = \pm 105^\circ$ and $\gamma = 180^\circ$. It can also be observed that the regions between the two maxima at $\gamma = \pm 135^\circ$ and the two minima at $\gamma = \pm 105^\circ$ become flat with increasing temperature, which justifies their scaling down by a temperature-dependent factor. To determine the constant c of the modified temperature-dependence expression (eq 3) of the $U_{\text{tor,AlaAla}}$ term of the UNRES force field, we have plotted the target function $\Phi(c)$ (defined by eq 5) against c in Figure 4. It can be seen that $\Phi(c)$ has a minimum at $c = 1.6$; however, the region of the minimum is very flat and any value of c between $c = 1.4$ and $c = 1.8$ results in virtually the same fit of the analytical to the numerical PMF and, consequently, considered as a plausible value of this parameter.

3.2 Selection of the optimal parameter c in eq 3 by force-field calibration with 1GAB

In the second approach to determine the factor c in eq 3, we carried out MREMD simulations of the 1GAB protein⁴² (a 53-residue three- α -helix bundle; Figure 5) with $c=0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 3.0$, and 5.0 , respectively, using the weights determined in ref¹⁷, which are summarized in the second column of Table 1.

First, we examined the heat capacity curves obtained in simulations with different values of c . The respective plots are shown in Figure 6. It is known that the process of protein folding is a first-order phase transition, which means that the plot of heat capacity versus temperature should have one clear peak at the folding transition temperature, and the peak should coincide with the jump in native-likeness.^{43,44} It can be seen from Figure 6 that the peak in heat capacity becomes sharper with increasing c but converges to a maximum height of about 12 kcal/(mol $\times$ K) for $c \geq 1.4$. The transition temperature does not change much and remains near $T_f = 320$ K, as shown in Table 2. Therefore, with regard to obtaining a sharp peak in heat capacity, $c \geq 1.4$.

Next, we examined the quality of the native-like structure obtained in the simulations. The ensemble-averaged root-mean-square deviation over the C^α atoms between the experimental structures and the computed structures (RMSD) was calculated at $T = 280$ K which is close to and below the folding transition temperature and at which the curve of the dependence of RMSD on temperature becomes flat (Figure 7); the results are shown in Table 2 for different values of c . The best result for the ensemble averages is found when $c = 1.2$, and the second best is found when $c = 1.4$. We have also examined the average of the RMSD of the 10 best structures encountered in the whole simulation, and the values are also presented in Table 2. The best result is found when $c = 1.4$, and the second best result is found when $c = 1.6$.

Taking into account the improvement of the sharpness of the folding transition and the quality of the native-like structures, a value $c = 1.4$ is selected as the optimal parameter of the temperature factors in eq 3. This result is in good agreement with the range of c determined by fitting the UNRES virtual-bond torsional energy term $U_{\text{rot,AlaAla}}$ to the respective numerically calculated PMF presented in section 3.1.

3.3 Optimization and tests of the modified UNRES force field

Because the temperature-dependent UNRES force field introduced recently¹⁷ was optimized on 1GAB with the scaling factors defined in eq 2 (which is equivalent to that defined by eq 3 with $c = 1$), we had to reparameterize the UNRES force field using the new scaling factors. We used 1GAB as the training protein, and applied exactly the same procedure as described in ref¹⁷, starting from the parameters obtained in our earlier work (Table 3 in ref¹⁷, also the second column of Table 1). We optimized only the energy-term weights, whose optimized values are given in column 3 of Table 1. It can be seen that the final force field does not differ much from the starting values. The heat capacity and RMSD curves corresponding to the force field of ref¹⁷ (with the temperature factors defined by eq 2) are compared with those of the

optimized force field (with the temperature factors defined by eq 3, in which $c = 1.4$) in Figure 7. It can be seen that the heat capacity curve became sharper and the RMSD curve became steeper with smaller RMSD values before the transition temperature.

Finally, we tested the optimized force field on the same five α -helical proteins: 1BDD (Figure 8a),⁴⁵ 1CLB (Figure 8b),⁴⁶ 1POU (Figure 8c),⁴⁷ 1PRU (Figure 8d),⁴⁸ and 1KOY (Figure 8e),⁴⁹ which were used in ref¹⁷ to test the original force field. For each protein, we carried out MREMD simulations with settings specified in the Methods section. The first of these proteins (1BDD) has been studied in the recent years using the ECEPP/3^{50,51} and AMBER⁵² force fields and was found to produce some native-like conformations in electrostatically-driven Monte Carlo,⁵⁰ replica-exchange Monte Carlo,⁵¹ and molecular dynamics⁵² simulations. In order to determine the dominant conformational ensembles after the folding transition, we used the procedure described in ref¹⁷ in which, after determining the heat-capacity curve from the results of MREMD simulations and with the use of the WHAM method, a temperature corresponding to the base of the ascending branch of the heat-capacity curve is selected as the working temperature, and the probabilities of all conformations are calculated at that temperature (eq 26 of ref¹⁷). The conformations are then sorted according to decreasing probability, and a cluster analysis including those conformations whose probabilities sum up to a given threshold (0.99 in this work) is carried out with the single-link method.^{53,54} The RMSD cutoff in the cluster analysis was selected as a compromise between the need to obtain a minimum number of families and compact clusters of similar conformations; the RMSD cutoff values are collected in Table 3. The probabilities of the clusters are calculated as the sums of probabilities of the constituent conformations, and the clusters are ranked according to decreasing probability. The results of the cluster analysis are summarized in Table 3. In particular, the rank of the clusters of the native-like conformations (which determines the ability of the force field to locate native-like structures as candidate models in blind prediction) and the average RMSD of the native-like cluster from the experimental structure are listed for each protein.

As can be seen from Table 3, the new force field provides major improvement compared to the old force field, for 1CLB and 1POU, both in terms of the rank of the native-like cluster and the ensemble-averaged RMSD. For 1BDD both force fields predict the native-like structure as the most probable cluster; however, the ensemble-averaged RMSD is slightly lower with the new force field. Both force fields provide low probabilities of the native-like clusters and high ensemble-averaged RMSD for 1PRU and 1KOY; nevertheless, the probabilities of the native-like clusters are slightly higher with the new force field. Consequently, these results suggest that the temperature-scaling factor defined by eq 2 should be replaced with that defined eq 3 with $c = 1.4$.

4. Conclusions

We have proposed a parametric functional form (with a common parameter) for the temperature-scaling factors (eq 3) of the energy terms of the UNRES force field and determined its parameter by fitting the virtual-bond-torsional energy term (corresponding to the rotation about the $C^\alpha \dots C^\alpha$ virtual bond of terminally-blocked dialanine) to the respective potential of mean force and, independently, by calibration with the 1GAB protein to obtain a sharp folding transition and low ensemble-averaged RMSD for the region of thermal stability of the folded structure. Both approaches gave consistent results and enabled us to select the optimal value of the parameter. Replacing the old scaling factors (eq 2) with the new ones (eq 3) results in sharper heat-capacity peaks and lower ensemble-average RMSD for the 1GAB protein. With the new scaling factors, we have re-optimized the UNRES force field with 1GAB as the training protein and tested the predictive power of the optimized force field with five selected α -helical

proteins with different helix topology. We found that the new force field performs better than the old one, which suggests that the new temperature-scaling factor should replace the old one.

Acknowledgement

This work was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB05-41633), and the Polish Ministry of Science and Education (0490/B/H03/2008/35). This research was conducted by using the resources of (a) our 800-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (b) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, (c) the John von Neumann Institute for Computing at the Central Institute for Applied Mathematics, Forschungszentrum Jülich, Germany, (d) our 45-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk, (e) the Informatics Center of the Metropolitan Academic Network (ICMAN) in Gdańsk, and (f) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) at the University of Warsaw.

References

1. Kolinski A, Skolnick J. *Polymer* 2004;45:511–524.
2. Tozzini V. *Curr. Opinion Struct. Biol* 2005;15:144–150.
3. Clementi C. *Curr. Opinion Struct. Biol* 2008;18:10–15.
4. Voth, G. *Coarse-Graining of Condensed Phase and Biomolecular Systems*. Boca Raton, FL USA: Taylor& Francis Group; 2008. p. 1
5. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *J. Comp. Chem* 1997;18:849–873.
6. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *J. Comp. Chem* 1997;18:874–887.
7. Liwo A, Kazmierkiewicz R, Czaplewski C, Groth M, Oldziej S, Wawak RJ, Rackovsky S, Pincus MR, Scheraga HA. *J. Comp. Chem* 1998;19:259–276.
8. Liwo A, Czaplewski C, Pillardy J, Scheraga HA. *J. Chem. Phys* 2001;115:2323–2347.
9. Liwo A, Arlukowicz P, Czaplewski C, Oldziej S, Pillardy J, Scheraga HA. *Proc. Natl. Acad. Sci. U.S.A* 2002;99:1937–1942. [PubMed: 11854494]
10. Oldziej S, Kozłowska U, Liwo A, Scheraga HA. *J. Phys. Chem. A* 2003;107:8035–8046.
11. Oldziej S, Liwo A, Czaplewski C, Pillardy J, Scheraga HA. *J. Phys. Chem. B* 2004;108:16934–16949.
12. Oldziej S, Łagiewka J, Liwo A, Czaplewski C, Chinchio M, Nancias M, Scheraga HA. *J. Phys. Chem. B* 2004;108:16950–16959.
13. Oldziej S, Czaplewski C, Liwo A, Chinchio M, Nancias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kazmierkiewicz R, Ripoll DR, Pillardy J, Scheraga HA. *Proc. Natl. Acad. Sci. U.S.A* 2005;102:7547–7552. [PubMed: 15894609]
14. Khalili M, Liwo A, Rakowski F, Grochowski P, Scheraga HA. *J. Phys. Chem. B* 2005;109:13785–13797. [PubMed: 16852727]
15. Khalili M, Liwo A, Jagielska A, Scheraga HA. *J. Phys. Chem. B* 2005;109:13798–13810. [PubMed: 16852728]
16. Liwo A, Khalili M, Scheraga HA. *Proc. Natl. Acad. Sci. U.S.A* 2005;102:2362–2367. [PubMed: 15677316]
17. Liwo A, Khalili M, Czaplewski C, Kalinowski S, Oldziej S, Wachucik K, Scheraga HA. *J. Phys. Chem. B* 2007;111:260–285. [PubMed: 17201450]
18. Chinchio M, Czaplewski C, Liwo A, Oldziej S, Scheraga HA. *J. Chem. Theory and Comput* 2007;3:1236–1248.
19. Nancias M, Czaplewski C, Scheraga HA. *J. Chem. Theory Comput* 2006;2:513–528. [PubMed: 18797518]
20. Czaplewski C, Kalinowski S, Scheraga HA. *J. Chem. Theory Comput* 2009;5:627–640.
21. Kolinski A, Skolnick J. *J. Chem. Phys* 1992;97:9412–9426.
22. Kolinski A, Godzik A, Skolnick J. *J. Chem. Phys* 1993;98:7420–7433.
23. Kolinski A. *Acta Biochim. Pol* 2004;45:511–524.

24. Favrin G, Irbäck A, Wallin S. *Proteins: Struct. Funct. Genet* 2002;47:99–105. [PubMed: 11933057]
25. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman GP, Marrink S-J. *J. Chem. Theor. Comput* 2008;4:819–834.
26. Zhou J, Thorpe I, Izvekov S, Voth GA. *Biophys. J* 2007;92:4289–4303. [PubMed: 17400700]
27. Kubo R. *J. Phys. Soc. Jpn* 1962;17:1100–1120.
28. Hardin C, Eastwood MP, Prentiss M, Luthey-Schulten Z, Wolynes PG. *J. Comput. Chem* 2002;23:138–146. [PubMed: 11913379]
29. Geyer CJ, Thompson EA. *J. Am. Stat. Ass* 1995;90:909–920.
30. Hukushima K, Nemoto K. *J. Phys. Soc. Jpn* 1996;65:1604–1608.
31. Tesi MC, van Rensburg EJJ, Orlandini E, Whittington SG. *J. Stat. Phys* 1996;82:155–181.
32. Marinari, E.; Parisi, G.; Ruiz-Lorenzo, JJ. *Spin Glasses and Random Fields*. Young, AP., editor. Singapore: World Scientific; 1998. p. 58-98.
33. Hansmann UHE. *Chem. Phys. Lett* 1997;281:140–150.
34. Sugita Y, Okamoto Y. *Chem. Phys. Lett* 1999;314:141–151.
35. Swendsen RH, Wang J-S. *Phys. Rev. Lett* 1986;57:2607–2609. [PubMed: 10033814]
36. Rhee YM, Pandé VS. *Biophys. J* 2003;84:775–786. [PubMed: 12547762]
37. Nishikawa K, Momany FA, Scheraga HA. *Macromolecules* 1974;7:797–806. [PubMed: 4437206]
38. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. *J. Chem. Phys* 1984;81:3684–3690.
39. Swope WC, Andersen HC, Berens PH, Wilson KR. *J. Chem. Phys* 1982;76:637–649.
40. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. *J. Comput. Chem* 1992;13:1011–1021.
41. Koradi R, Billeter M, Wuethrich K. *J. Mol. Graph* 1996;14:51–55. [PubMed: 8744573]
42. Johansson MU, de Chateau M, Wikstrom M, Forsen S, Drakenberg T, Bjorck L. *J. Mol. Biol* 1997;266:859–865. [PubMed: 9086265]
43. Camacho CJ, Thirumalai D. *Europhys. Lett* 1996;35:627–632.
44. Klimov DK, Thirumalai D. *Phys. ReV. Lett* 1996;76:4070–4073. [PubMed: 10061184]
45. Gouda H, Torigoe H, Saito A, Sato M, Arata Y, Shimada I. *Biochemistry* 1992;31:9665–9672. [PubMed: 1390743]
46. Skelton NJ, Kordel J, Chazin WJ. *J. Mol. Biol* 1995;249:441–462. [PubMed: 7783203]
47. Assa-Munt N, Mortishire-Smith RJ, Aurora R, Herr W, Wright PE. *Cell* 1993;73:193–205. [PubMed: 8462099]
48. Nagadoi A, Morikawa S, Nakamura H, Enari M, Kobayashi K, Yamamoto H, Sampei G, Mizobuchi K, Schumacher MA, Brennan RG, et al. *Structure* 1995;3:1217–1224. [PubMed: 8591032]
49. Fukushima K, Kikuchi J, Koshiha S, Kigawa T, Kuroda Y, Yokoyama S. *J. Mol. Biol* 2002;321:317–327. [PubMed: 12144788]
50. Vila JA, Ripoll DR, Scheraga HA. *Proc. Natl. Acad. Sci. USA* 2003;100:14812–14816. [PubMed: 14638943]
51. Trebst S, Hansmann UHE. *Eur. Phys. J. E* 2007;24:311–316. [PubMed: 18071628]
52. Jang S, Kim E, Shin S, Pak YJ. *Am. Chem. Soc* 2003;125:14841–14846.
53. Murtagh F. *Comput. J* 1983;26:354–359.
54. Murtagh, F.; Heck, A. *MultiVariate data analysis*. Dordrecht, Holland: Kluwer Academic; 1987. p. 63-65.

Figure 1.

The UNRES model of the polypeptide chain. Dark circles represent united peptide groups (p) and open circles represent the C^{α} atoms. Ellipsoids represent side chains, with their centers of mass at the SC's. The p 's are located halfway between two consecutive C^{α} atoms. The virtual-bond angles, θ , the virtual-bond dihedral angles, γ , and the angles, α_{sc} and β_{sc} , that define the location of a side chain with respect to the backbone, are also indicated.

Figure 2.

Illustration of the model terminally-blocked dipeptide constructed to compute the integrals of eq 4. X and Y denote Ala, Gly or Pro. The angles λ_1 , λ_2 , and λ_3 for rotation of the consecutive peptide groups about the $C^\alpha \dots C^\alpha$ virtual bonds and the virtual-bond-dihedral angle γ are indicated in the Figure.

Figure 3. Plot of the calculated torisional potentials of mean force (PMF's) F_{tor} of the UNRES force field for the Ac-Ala-Ala-NHMe system at four representative temperatures 200 K (black-filled diamonds), 250 K (red-filled squares), 300 K (green-filled triangles), and 350 K (purple-filled circles), obtained with the MP2/6-31G(d,p) ab initio method.

Figure 4.
Plot of the function $\Phi(c)$ of eq 5 vs. c .

Figure 5.
Ribbon diagram of the experimental structure of 1GAB.

Figure 6.

Converged heat capacities of 1GAB obtained from UNRES/MREMD simulations for different values of the parameter c in the temperature-scaling factor (eq 3) and the weights in the second column of Table 1. Curves corresponding to different values of c are plotted with different colors.

Figure 7. Converged heat capacities (solid lines) and RMSD (dashed lines) of the 1GAB protein obtained from the new force field (the final force field with the new scaling factors given in eq 3; red) and the old force field (the starting force field with the old scaling factors given in eq 1; black).

Figure 8.

C^α -traces of the average structures (gray) of the clusters of native-like conformations superposed on those of the experimental structures (black) of (a) 1BDD,³⁷ (b) 1CLB,³⁸ (c) 1POU,³⁹ (d) 1PRU,⁴⁰ and (e) 1KOY,⁴¹ calculated with the new force field. The numerical characteristics of the computed structures are collected in Table 3.

Table 1

Initial and final energy-term weights resulting from optimization of the UNRES force field using the new scaling factors defined by eq 3 and 1GAB as a training protein

weight	Starting force field	Final force field
w_{SC}	1.35279	1.37474
w_{SCp}	1.59304	1.64019
w_{pp}^{el}	0.71534	0.72886
w_{pp}^{VDW}	0.11371	0.11371
w_h	1.13873	1.14620
w_{rot}	0.16258	0.15985
w_{tor}	1.98599	1.96848
w_{tord}	1.57069	1.58564
$w_{corr}^{(3)}$	0.16036	0.13589
$w_{corr}^{(4)}$	0.42887	0.42940
$w_{turn}^{(3)}$	1.68722	1.65648
$w_{turn}^{(4)}$	0.66230	0.65863

Table 2

Results of MREMD simulations of 1GAB at 280K, using force-field parameters of ref ¹⁴ and the scaling factors of eq 3 with various values of c .

c^a	Folding temperature (K)	RMSD ^b (ensemble average at 280 K) (Å)	RMSD ^b (average of ten best structures encountered in the whole simulation) (Å)
0.8	321	4.26	2.18
1.0	321	4.15	2.15
1.2	319	4.06	2.15
1.4	319	4.11	2.03
1.6	318	4.52	2.09
2.0	318	5.02	2.12
3.0	319	5.03	2.29
5.0	319	5.67	2.39

^aThe value of c in the function defined in eq 3.

^bRMSD from experimental structure.

Table 3
Results of tests of the force fields parameterized with 1GAB

PDB id	N ^d	The force field of ref ¹⁴ optimized with 1GAB							The new force field with the scaling factors given by eq 3 with <i>c</i> = 1.4					
		ρ_{cut}^b [Å]	T^c [K]	ρ_{min}^d [Å]	ρ_{min}^e [Å]	N ^f	P^g	ρ_{cut}^b [Å]	T^c [K]	ρ_{min}^d [Å]	ρ_{min}^e [Å]	N ^f	P^g	
1BDD	46	3.0	300	4.6	2.2	1	0.90	3.0	300	4.3	2.2	1	0.93	
1CLB	75	4.0	300	6.4	3.9	4	0.08	4.0	300	4.9	4.0	2	0.38	
1POU	71	4.0	300	7.0	5.3	7	0.03	4.0	300	5.5	4.4	5	0.12	
1PRU	56	4.0	300	7.0	5.3	9	0.01	4.0	300	6.9	5.5	6	0.014	
1KOY	62	4.0	280	5.9	4.4	7	0.002	4.0	280	5.8	4.6	4	0.01	

^aNumber of residues.

^bThe RMSD cutoff in clustering.

^cTemperature at which the probabilities of the clusters were calculated.

^dBoltzmann averaged RMSD value over the conformations of the native cluster from the experimental structure.

^eLowest RMSD from the experimental structure encountered in the MREMD simulation.

^fThe rank(s) (according to probability defined in ref. 14) of the native cluster(s).

^gProbability of the native cluster, defined in ref. 14.