Molecular dynamics of protein A and a WW domain with a united-residue model including hydrodynamic interaction

Agnieszka G. Lipska,[1,2,a)] Steven R. Seidman,[2,3,4] Adam K. Sieradzan,[1] Artur Giełdoń,[1] Adam Liwo,[1] and Harold A. Scheraga[2]
[1]*Laboratory of Molecular Modeling, Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland*
[2]*Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, USA*
[3]*Department of Mathematical Sciences, DePaul University, 2320 N. Kenmore Ave., Chicago, Illinois 60614-3210, USA*
[4]*Academic Learning Center, Argosy University, 225 N. Michigan Ave. #1300, Chicago, Illinois 60601, USA*

(Received 26 February 2016; accepted 25 April 2016; published online 13 May 2016)

The folding of the N-terminal part of the B-domain of staphylococcal protein A (PDB ID: 1BDD, a 46-residue three-α-helix bundle) and the formin-binding protein 28 WW domain (PDB ID: 1E0L, a 37-residue three-stranded anti-parallel β protein) was studied by means of Langevin dynamics with the coarse-grained UNRES force field to assess the influence of hydrodynamic interactions on protein-folding pathways and kinetics. The unfolded, intermediate, and native-like structures were identified by cluster analysis, and multi-exponential functions were fitted to the time dependence of the fractions of native and intermediate structures, respectively, to determine bulk kinetics. It was found that introducing hydrodynamic interactions slows down both the formation of an intermediate state and the transition from the collapsed structures to the final native-like structures by creating multiple kinetic traps. Therefore, introducing hydrodynamic interactions considerably slows the folding, as opposed to the results obtained from earlier studies with the use of Gō-like models. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4948710]

I. INTRODUCTION

The development of physics-based methods to study protein folding and to predict protein structures continues to be an important field of computational biology.[1–9] Coarse-grained models are of great interest in the field because they enable us to extend the time- and size-scale of simulations by orders of magnitude.[10–13] However, coarse-grained representation usually makes use of implicit solvent. The potential energy contribution from interactions of the protein molecules with the solvent is usually included in the effective coarse-grained potentials but, apart from this, the solvent influences folding dynamics through the friction and stochastic forces. These are usually accounted for by a Langevin-dynamics treatment, in which each site experiences the non-conservative forces from the solvent individually. However, these forces also lead to a solvent-mediated apparent drag of two objects moving through a liquid, which is known as hydrodynamic interactions (HIs).[14–16] These interactions can be accounted for by introducing a non-diagonal matrix of friction coefficients, referred to as, e.g., the Rotne-Prager (RP) tensor[17] and its modification, the Rotne-Prager-Yamakawa (RPY) tensor,[17,18] whose off-diagonal elements depend on the distances between the objects involved.

The influence of HIs on polymer collapse[19–21] was studied by modeling a polymer chain as beads connected by a finitely extendable nonlinear elastic potential, with an effective interaction potential represented by the Lennard-Jones function. This model showed that the introduction of HIs sped up polymer collapse.[20,21] This effect was also observed by Chang and Yethiraj[19] who found that, without HIs, the compact polymer globule could not be formed, and the system got trapped in a non-ergodic metastable state.

The role of HIs in protein folding has been discussed by Tanaka,[14] who suggested that HIs could have a major effect on protein-folding pathways, as well as on the final state. HIs in proteins, modeled by using the RPY tensor, were investigated with Gō-like models and Brownian Dynamics (BD) with the use of the Ermak-McCammon algorithm.[22] In that work, HIs were found to speed up folding by a factor of 2 to 3 for proteins composed of 56 to 149 residues. On the other hand, the decrease of the folding rate of the individual secondary-structure elements, the α-helix and the β-hairpin (16-residues-long), was observed.[16] Similar results were obtained by Cieplak and Niewieczerzał.[15] They suggested that the acceleration is due to faster initial collapse. However, those investigations concerned Gō-like protein models, in which the interaction pattern is highly simplified; the interaction potentials have minima for native long-range interactions and are repulsive-only for non-native interactions while, in reality, the character of long-range residue-residue interactions depends only on the kinds of residues involved.

HIs were also conceptualized as interactions that occur between molecules. With the use of BD simulations and the RPY tensor, Ando and Skolnick found[23] that, in

a)lipska.a@sun1.chem.univ.gda.pl

the crowded cell interior, the decrease of the diffusion coefficients for macromolecules, compared to infinitely dilute water solution, is most likely caused by HIs and excluded-volume effects. The influence of intermolecular HIs on other biological phenomena such as formation of lipid membranes,[24] protein-protein association,[25] the sliding motion of protein along DNA,[26] and the diffusion processes within an idealized biological cell[27] was also studied.

For many years, we have been developing a physics-based coarse-grained force field for the simulation of protein structure and dynamics.[28–34] We implemented Langevin dynamics in UNRES[35,36] to study protein folding. Despite the medium resolution of the force field resulting from the coarse-grained nature of the UNRES model, it proved successful in protein-structure prediction,[37–40] in simulations of the pathways, kinetics, and free-energy landscapes of protein folding,[41–44] effect of mutation on protein folding and stability,[45,46] as well as in studying biologically important processes.[7,47–50] In the study reported here, as an attempt to find a realistic mean-field description of the dynamic effect of the solvent, we introduced HIs in UNRES Langevin dynamics and assessed the effect of these interactions on the pathways and kinetics of folding of two small proteins, the N-terminal domain of staphylococcal protein A, which has the structure of a three-α-helix bundle (PDB ID: 1BDD; 46 residues)[51] and a formin-binding protein WW domain, which is a three-stranded anti-parallel β-sheet (PDB ID: 1E0L; 37 residues).[52] The experimental structures of these two proteins are shown in Figure 1. These proteins will hereafter be referred to by their PDB IDs.

II. METHODS

A. The UNRES force field

In the UNRES model,[28–34] Figure 2, a polypeptide backbone is represented as a sequence of α-carbon (C^α) atoms linked by virtual bonds. A peptide group (p) is placed in the middle of each backbone virtual bond and a side chain (SC) is attached to the respective C^α atom. Only the peptide groups and the side chains are interaction sites and the C^αs serve to define the geometry.

The UNRES energy function is expressed by

$$\begin{aligned} U = {} & w_{SC}\sum_{i<j} U_{SC_iSC_j} + w_{SCp}\sum_{i\neq j} U_{SC_ip_j} + w_{pp}^{VDW}\sum_{i<j-1} U_{p_ip_j}^{VDW} + w_{pp}^{el} f_2(T)\sum_{i<j-1} U_{p_ip_j}^{el} + w_{tor} f_2(T)\sum_i U_{tor}(\gamma_i) \\ & + w_{tord} f_3(T)\sum_i U_{tord}(\gamma_i,\gamma_{i+1}) + w_b\sum_i U_b(\theta_i) + w_{rot}\sum_i U_{rot}(\alpha_{SC_i},\beta_{SC_i}) + w_{bond}\sum_i U_{bond}(d_i) \\ & + w_{corr}^{(3)} f_3(T) U_{corr}^{(3)} + w_{corr}^{(4)} f_4(T) U_{corr}^{(4)} + w_{turn}^{(3)} f_3(T) U_{turn}^{(3)} + w_{turn}^{(4)} f_4(T) U_{turn}^{(4)}, \end{aligned} \tag{1}$$

where θ_i is the ith backbone virtual-bond angle, γ_i is the ith backbone virtual-bond-dihedral angle, α_i and β_i are the polar angles defining the location of the united side-chain center of residue i, and d_i is the length of the ith virtual bond, which is either a $C^\alpha\cdots C^\alpha$ virtual bond or a $C^\alpha\cdots$SC virtual bond, Figure 2. Each energy term, U, is multiplied by an appropriate weight, w_x, and the terms corresponding to factors of order higher than 1 are additionally multiplied by the respective temperature factors, $f_n(T)$,[30,53] defined by

$$f_n(T) = \frac{\ln\left[\exp(1)+\exp(-1)\right]}{\ln\left\{\exp\left[(T/T_\circ)^{n-1}\right]+\exp\left[-(T/T_\circ)^{n-1}\right]\right\}}, \tag{2}$$

where $T_\circ = 300$ K.

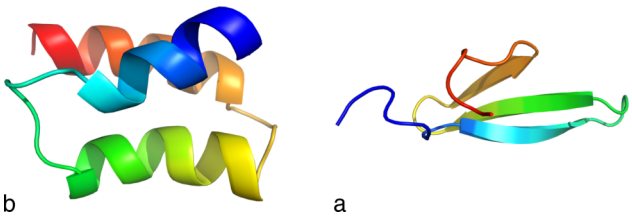

FIG. 1. Cartoon representation of model 1 of NMR experimental structures of the (a) 1BDD and (b) 1E0L.

In this study, two versions of the UNRES force field that differ in terms of parameterization were used; one was calibrated with the albumin-binding GA module (PDB ID: 1GAB[54]),[30] and the second one was calibrated with the engrailed homeodomain (PDB ID: 1ENH[55]) and the Fbp28 WW domain (PDB ID: 1E0L[52])[33] for simulations of 1BDD and 1E0L protein, respectively. Although recently[56] we parameterized a version of UNRES with better transferability, these two versions were already used with success in protein-folding studies.[42–46]

B. Langevin dynamics

The implementation of Langevin dynamics to the UNRES model has been described in earlier work.[36] The Langevin equations of motion in virtual bond coordinates are given by Eq. (3). Briefly, the variables are the $C^\alpha\cdots C^\alpha$ and $C^\alpha\cdots$SC virtual-bond vectors, denoted by $\mathbf{dC_\circ}, \mathbf{dC_1}, \mathbf{dC_2}, \ldots, \mathbf{dC_{n-1}}$ (where $\mathbf{dC_i} = \mathbf{r_{C_{i+1}^\alpha}} - \mathbf{r_{C_i^\alpha}}$, $\mathbf{r_{C_i^\alpha}}$ denotes the Cartesian coordinates of the ith C^α atom and $\mathbf{dC_\circ} = r_{C_1^\alpha}$ places the first C^α atom in the coordinate system) and $\mathbf{dX_2}, \mathbf{dX_2}, \ldots, \mathbf{dX_{n-1}}$ (where $\mathbf{dX_i} = \mathbf{r_{SC_i}} - \mathbf{r_{C_i^\alpha}}$). These vectors are combined into the vectors of generalized coordinates,

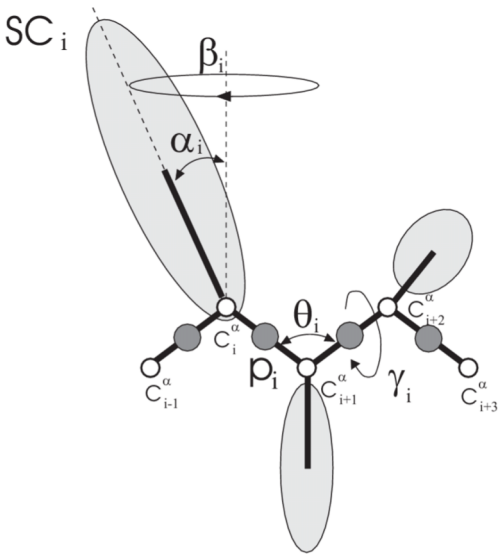

FIG. 2. The UNRES model of the polypeptide chain. Dark circles represent united peptide groups (p), white circles represent the C^{α} atoms, which serve as geometric points, and ellipsoids represent side chains (SCs). The p's are located half-way between two consecutive C^{α} atoms. The virtual-bond 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.

$$\mathbf{q} = [\mathbf{dC_o}, \mathbf{dC_1}, \mathbf{dC_2}, \ldots, \mathbf{dC_{n-1}}, \mathbf{dX_2}, \mathbf{dX_2}, \ldots, \mathbf{dX_{n-1}}]^{\mathbf{T}}$$

$$\mathbf{G\ddot{q}} = -\nabla U(\mathbf{q}) + \mathbf{f^{fric}} + \mathbf{f^{rand}}, \tag{3}$$

where the inertia-matrix $\mathbf{G}$ is defined by

$$\mathbf{G} = \mathbf{A^T M A} + \mathbf{H}, \tag{4}$$

where $\mathbf{M}$ is a diagonal matrix of the masses of all the interacting sites, and $\mathbf{H}$ is a diagonal matrix (part of the inertia matrix corresponding to the internal stretching motions of the virtual bonds). $\mathbf{A}$ is the matrix that transforms the virtual-bond coordinates to Cartesian coordinates. The frictional force in virtual bond coordinates is given by

$$\mathbf{f^{fric}} = -\mathbf{A^T \Gamma A \dot{q}}, \tag{5}$$

where, in the absence of HIs, the friction super-matrix is given by

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1 \mathbf{I} & 0 & \ldots & 0 \\ 0 & \gamma_2 \mathbf{I} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & \gamma_N \mathbf{I} \end{bmatrix}, \tag{6}$$

where $\mathbf{I}$ is a 3×3 identity matrix and the friction coefficient for the ith site and γ_i is calculated from the Stokes equation as follows:

$$\gamma_i = 6\pi(R_i + R_w)\eta_w, \tag{7}$$

where R_i and R_w are the radii of the site and water, respectively, and η_w is the water viscosity. Thus, for the purpose of the calculation of the friction and stochastic forces, each site is modeled as a sphere.[36]

In Langevin stochastic dynamics, the temperature is kept constant (on average) by applying frictional (damping) forces along with the compensating stochastic forces to the force centers. The average stochastic forces are related to the frictional forces by the fluctuation-dissipation theorem.[57] This model closely mirrors the physical dynamics of the system, i.e., the stochastic forces model the solvent molecules impinging against the protein, and the frictional forces model the Stokes drag (also due to the solvent) on each interaction site. Therefore, to obtain the information about protein dynamical motions, kinetics, and pathways, this model is an appropriate one to choose.

C. Introduction of hydrodynamic interactions into the UNRES force field

The HIs between unequal virtual-spheres representing the SCs and ps are modeled by the use of the RP tensor, and empirical averages are used for the radii of the spheres.[58,59] To compute the hydrodynamic drag, both the SCs and ps will be considered as virtual spheres that experience a Stokes drag from the solvent. This approximation had been used in previous investigations of the effect of HIs on protein folding.[16]

As each spherical site (SC or p) moves through the solvent, it creates a wake in the solvent that will perturb all surrounding force centers. This perturbing force is referred to as the HI force. It does not alter the form of the UNRES potential since the UNRES potential contains all the conservative forces acting on the coarse-grained model of the protein while the HIs modify the non-conservative damping term in the Langevin equation. A molecular dynamics (MD) study of the motion of a sphere in a viscous Lennard-Jones fluid has shown that the results of continuum hydrodynamics work even if the sphere is comparable in size to that of the fluid molecules.[60] This justifies the use of continuum fluid equations to calculate the HIs.

The effective radii for virtual spheres representing the amino-acid side chains have been taken from Levitt.[61] The effective radius for the peptide group (p) was set to 5 Å, based on the van der Waals radius of the peptide group. This radius was estimated from calculations of potential of mean force surfaces for interacting peptide groups.[62]

The force $\mathbf{f_i^h}$ that the solvent exerts on the ith particle is determined, in general, by the instantaneous momentum $\mathbf{p_i}$ and position $\mathbf{r_i}$ of all the N particles in the solvent,

$$\mathbf{f_i^h} = f_i^h(\mathbf{p_i}(t), \ldots, \mathbf{p_N}(t), \mathbf{r_i}(t), \ldots, \mathbf{r_N}(t)). \tag{8}$$

The superscript h indicates a hydrodynamic force. In our system, the N particles are virtual spheres representing ps and the SCs. For the fluid dynamics of the protein-solvent HIs, the Navier-Stokes equation, for typical diameters and velocities of the virtual spheres, and typical solvent viscosity, can be linearized allowing one to write the hydrodynamic forces as linear functions of the velocities,[63]

$$\mathbf{f_i^h} = -\sum_{j=1}^{N} \mathbf{\Gamma_{ij}(r_1, r_2, \ldots, r_N) v_j}, \tag{9}$$

where $\mathbf{\Gamma}$ is the 3×3 microscopic friction tensor. In a very dilute solution, there is approximately no HI between any of

the virtual spheres. Therefore, Eq. (5) reduces to

$$\mathbf{f_i^h} = -\gamma \mathbf{v_i}, \tag{10}$$

where γ is the friction coefficient for isolated (i.e., hydrodynamically uncoupled) virtual spheres. This is the form for the friction in the unmodified UNRES-MD algorithm.[36]

To first order HIs, for large separations between the force centers, it can be taken into account by modification of the off diagonal elements as follows:

$$\begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \\ \vdots \\ \mathbf{v_N} \end{bmatrix} = -\beta \begin{bmatrix} \mathbf{D_{11}} & \mathbf{D_{12}} & \dots & \mathbf{D_{1N}} \\ \mathbf{D_{21}} & \mathbf{D_{22}} & & \\ \vdots & & \ddots & \\ \mathbf{D_{N1}} & & & \mathbf{D_{NN}} \end{bmatrix} \begin{bmatrix} \mathbf{f_1^h} \\ \mathbf{f_2^h} \\ \vdots \\ \mathbf{f_N^h} \end{bmatrix}, \tag{11}$$

where the $\mathbf{D_{ij}}$s are the 3×3 diffusion tensors, put into a $3N \times 3N$ diffusion super-matrix to relate velocity, $\mathbf{v_i}$, of each force center (i.e., virtual sphere), to the hydrodynamic force, $\mathbf{f_i^h}$, on each virtual sphere, and $\beta = 1/k_B T$, with Boltzmann constant k_B and temperature T.

The friction super-matrix is related to the diffusion super-matrix by a generalization of the Stokes-Einstein relation

$$\mathbf{\Gamma^{-1}} = \beta \mathbf{D}. \tag{12}$$

From these equations the friction super-matrix is determined.

The random forces are computed by taking the square root of the friction super-matrix[36] according to the fluctuation-dissipation theorem[57]

$$\mathbf{f^{rand}} = \sqrt{\frac{2RT}{\delta t}} \mathbf{A^T} \Gamma^{1/2} \mathbf{N}(0,1), \tag{13}$$

where δt is the size of the time step, and $\mathbf{N}(0,1)$ is a multi-dimensional normal distribution with zero mean and unit variance.

Two specific forms for the interaction tensor will be considered. The first form assumes that the perturbing force at point j can be considered to be a point force. This assumption, along with the two approximations (Stokes' equation and incompressibility), enables us to solve for the HI tensor analytically. The result is the Oseen-Burgers tensor[64,65]

$$\mathbf{D} = \frac{kT}{8\pi\mu r}(\delta + \delta_\mathbf{r} \otimes \delta_\mathbf{r}), \tag{14}$$

where δ is a unit tensor of dimension 3×3, $\delta_\mathbf{r}$ is a unit vector pointing from the point where a force makes a disturbance in the solvent to the field point where the resulting disturbance is determined, $\delta_\mathbf{r} \otimes \delta_\mathbf{r}$ is a dyad of $\delta_\mathbf{r}$; r is the distance between these two points, and μ is the solvent viscosity. Mathematically, since the perturbing force can be represented as a Dirac-delta function, the Oseen-Burgers tensor is the Green function for the Stokes equation for incompressible flow and fluid dynamics similar to that found in protein dynamics.[66]

A second form of the interaction tensor extends Eq. (14) to consider the perturbing sphere to have a finite radius, in which case the interaction tensor is the generalized RPY tensor.[17] The interaction tensor for spheres of unequal radii a_i, a_j that are non-overlapping[17,18,58] is

$$\mathbf{D} = \frac{1}{8\pi\mu r}[(\mathbf{I} + \frac{\delta_\mathbf{r} \otimes \delta_\mathbf{r}}{r^2}) + (\frac{a_i^2 + a_j^2}{r^2})(\frac{1}{3}\mathbf{I} - \frac{\delta_\mathbf{r} \otimes \delta_\mathbf{r}}{r^2})] \quad if \quad r \geq a_i + a_j. \tag{15}$$

In the case of overlapping unequal virtual spheres, one modifies the RP tensor[66] to

$$\mathbf{D} = \frac{1}{6\pi\mu a_{ij}}[(1 - \frac{9}{32}\frac{r}{a_{ij}})\mathbf{I} + \frac{3}{32}\frac{\delta_\mathbf{r} \otimes \delta_\mathbf{r}}{r a_{ij}}] \quad if \quad r < a_i + a_j \tag{16}$$

with an average radius a_{ij} proposed on an empirical basis.[67] In this paper, a cubic average radius, $a_{ij} = (\frac{a_i^3 + a_j^3}{2})^{1/3}$, is used when the virtual-spheres overlap.

The purpose of using the RP tensor instead of the simpler Oseen-Burgers tensor is to ensure that the friction matrix remains positive-definite. This is important so that we can take the square root of this matrix to find the stochastic forces according to the fluctuation-dissipation theorem as indicated by Eq. (13). It can be proven that the RP tensor is positive-definite.[66]

The diagonal term of the friction super-matrix (or diffusion super-matrix) is also altered by the inclusion of HIs since the perturbing fluid disturbance propagates away from each virtual-sphere and is reflected off all the other virtual-spheres and propagates back to the original virtual-sphere that caused the disturbance. In effect, there is a self-interaction term. On the other hand, the HIs (solvent interaction) propagate infinitely fast only for incompressible fluids.[68] In a real fluid, the HIs will propagate at the speed of sound. For typical distances between the virtual spheres in a folding protein, the HI disturbance will have insufficient time to reflect off of the adjacent virtual spheres and cause a self-interaction; therefore, the diagonal terms in Equation (4) or (9) were left unaltered.

D. Simulation and analysis procedure

To assess the effect of HIs on the folding of 1BDD and 1E0L, multi-trajectory canonical MD simulations with UNRES were carried out for each of them in the classical Langevin mode and with HIs included, at temperatures 300 K for 1BDD and 335 K for 1E0L, respectively, which are below the simulated transition mid-point temperatures equal to 320 K[69] and 339 K,[45] respectively. In summary, four series of simulations were carried out: 1BDD without HIs, 1BDD with HIs, 1E0L without HIs, and 1E0L with HIs, respectively. Each series of simulation consisted of 200 trajectories run with the 4.89 fs time step. Each trajectory was run for 1 μs UNRES time for 1BDD (both with and without HIs), 1.2 μs UNRES time for 1E0L without HIs, and 5.0 μs UNRES time for 1E0L with HIs, which corresponds to about 1, 1.2, and 5.0 ms of real time, respectively, because of averaging out secondary degrees of freedom in UNRES.[35,36] All trajectories were started from an extended structure. To speed up the simulations, the viscosity of water was scaled down by a factor of 100, as in our earlier work.[35,36]

For each of the four series of simulations, hierarchical cluster analysis using Ward's minimum variance method[70,71] was performed for every 10th trajectory in order to find possible intermediate structures or kinetic traps that might have occurred during simulation. Then the obtained families were classified based on the C^{α}-RMSD of the mean structure of the cluster from the experimental structure as native (C^{α}-RMSD below 6 Å), intermediate (C^{α}-RMSD between 6 and 10 Å) or unfolded (C^{α}-RMSD higher than 10 Å). For each cluster, the representative conformation was defined as the conformation of the cluster which had the lowest C^{α}-RMSD from the mean conformation of the cluster. For each of the two proteins, the representatives of the clusters of the intermediate structures corresponding to simulations without (5 clusters for 1E0L and 2 clusters for 1BDD, respectively) and with HIs (8 clusters for 1E0L and 5 clusters for 1BDD, respectively) were then compared to find possible differences between the intermediates obtained with each of the two Langevin-dynamics regimes considered. The cross-cluster C^{α}-RMSD values are collected in Tables I and II of the supplementary material for 1E0L and 1BDD, respectively, and the structures are superposed in Figure SM9 of the supplementary material. As can be seen from Tables I and II of the supplementary material,[72] the RMSD values between the clusters corresponding to one (without HI or with HI) Langevin-dynamics regime do not differ from those corresponding to two different regimes; according to Student's test,[73] the differences between the same- and different-regime RMSDs are insignificant. The structures of the intermediates are also similar (Figure SM9 of the supplementary material).[72] It can, therefore, be concluded that the structure of the intermediate does not change remarkably after including HIs. Consequently, for each protein, the representative conformation of the intermediate structure was selected as the conformation closest to the mean conformation over all clusters of intermediate structures; this was the conformation with the lowest sum of cross-cluster C^{α}-RMSDs (shown in the bottom rows of Tables I and II of the supplementary material).[72]

To determine the C^{α}-RMSD thresholds for the native and intermediate structures, the probability distributions of C^{α}-RMSD from the respective experimental or intermediate structures were calculated for each quarter of the simulation. For 1BDD (both with and without HIs), the time windows from 0.00 to 0.25 μs, from 0.25 to 0.50 μs, from 0.50 to 0.75 μs, and from 0.75 to 1.00 μs were set. For 1E0L without including the HIs, time windows were set from 0.0 to 0.3 μs, from 0.3 to 0.6 μs, from 0.6 to 0.9 μs, and from

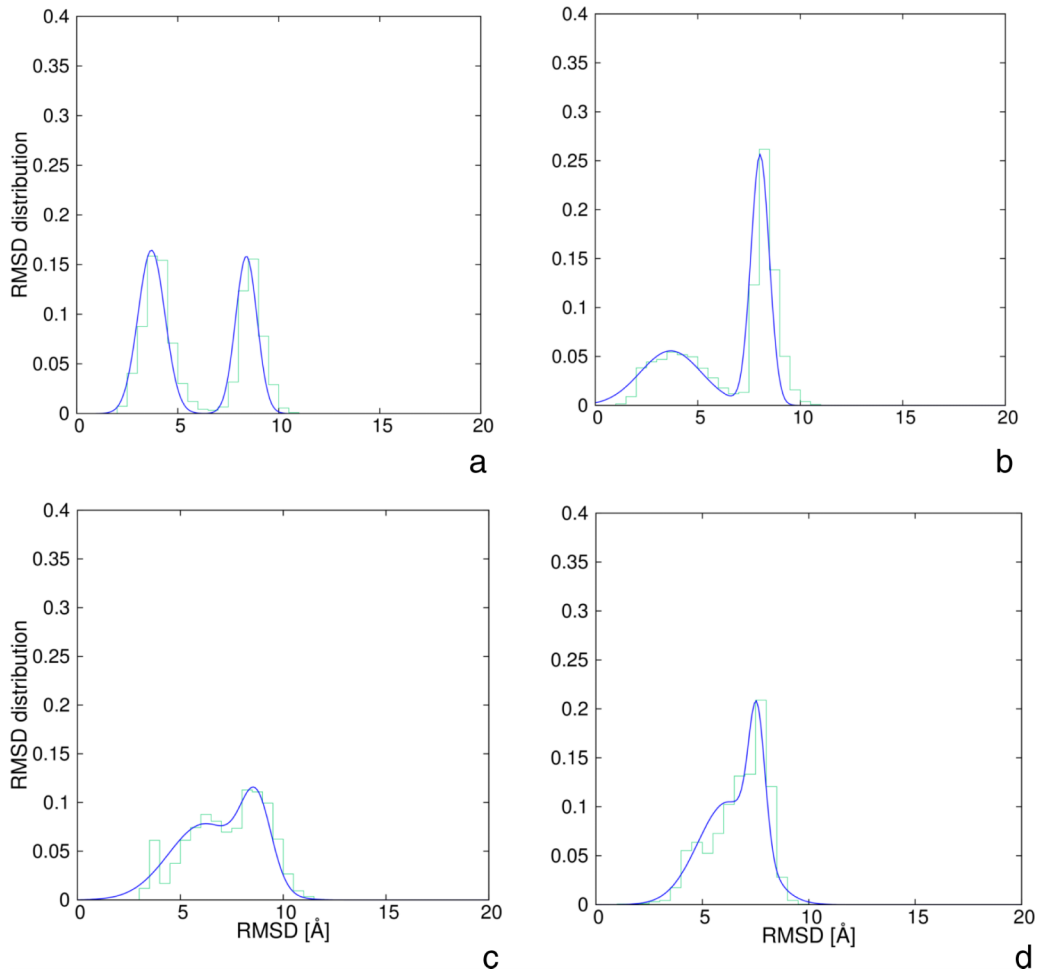

FIG. 3. Plots of probability distribution of C^{α}-RMSD values from the native and intermediate structures (green rugged lines) corresponding to the last time windows of simulations with inclusion of HIs and double-Gaussians fits to these plots (blue smooth lines) (Eq. (17)): (a) for the native structure of 1BDD, (b) for intermediate structure of 1BDD, (c) for the native structure of 1E0L, (d) for intermediate structure of 1E0L.

0.9 to 1.2 μs, respectively, while for 1E0L with inclusion of HIs the windows were set from 0.00 to 1.25 μs, from 1.25 to 2.50 μs, from 2.50 to 3.75 μs, and from 3.75 to 5.00 μs, respectively. The C^{α}-RMSD bin size was set to 0.5 Å. The C^{α}-RMSD range of values was from 1 Å (the native or intermediate structure) to 42 Å (the extended structure). For 1BDD, the plots of the C^{α}-RMSD distributions from the native and from the intermediate structure, respectively, corresponding to the last time windows of simulations with inclusion of HIs (for which the intermediates last longer) are shown in Figures 3(a) and 3(b), respectively. For 1E0L, the plots are shown in Figures 3(c) and 3(d), respectively. Detailed plots of the time evolutions of the C^{α}-RMSD distributions are shown in the supplementary material.[72] It can clearly be seen that two peaks are present. The first peak occurs at lower C^{α}-RMSD and corresponds to structures that are similar to the reference structure (native or intermediate), while the second peak corresponds to structures not similar to the respective reference structure. To facilitate the determination of the cutoffs of the reference-like structures, the distributions were fitted to a sum of two Gaussians (Eq. (17)), with the use of the nonlinear least-squares (NLLS) Marquardt-Levenberg[74] algorithm implemented in the GNUPLOT program (http://www.gnuplot.info)

TABLE I. The percentage of trajectories that folded to the native structure.

Protein	Percentage of folded trajectories (%)
1BDD without HIs	85
1BDD with HIs	59
1E0L without HIs	31
1E0L with HIs	15

$$f_g(x) = \sum_{i=1}^{N} \frac{C_i}{(2\pi)^{\frac{1}{2}}\sigma_i} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right] \tag{17}$$

where f_g is the fitted distribution at each time window obtained with the optimal parameters, x is the C^{α}-RMSD, $N = 2$ is the number of Gaussian terms, C_i is the multiplier for each Gaussian and represents a measure of the population of each Gaussian, and μ_i and σ_i are the mean and the standard deviation of each Gaussian, respectively.

Based on the C^{α}-RMSD distributions, the cutoffs on the C^{α}-RMSDs from the respective experimental structures to assign a conformation to the native-structure family were set at 5.5 Å for 1BDD and 4.5 Å for 1E0L, respectively, and the cutoffs from the respective intermediate structures to

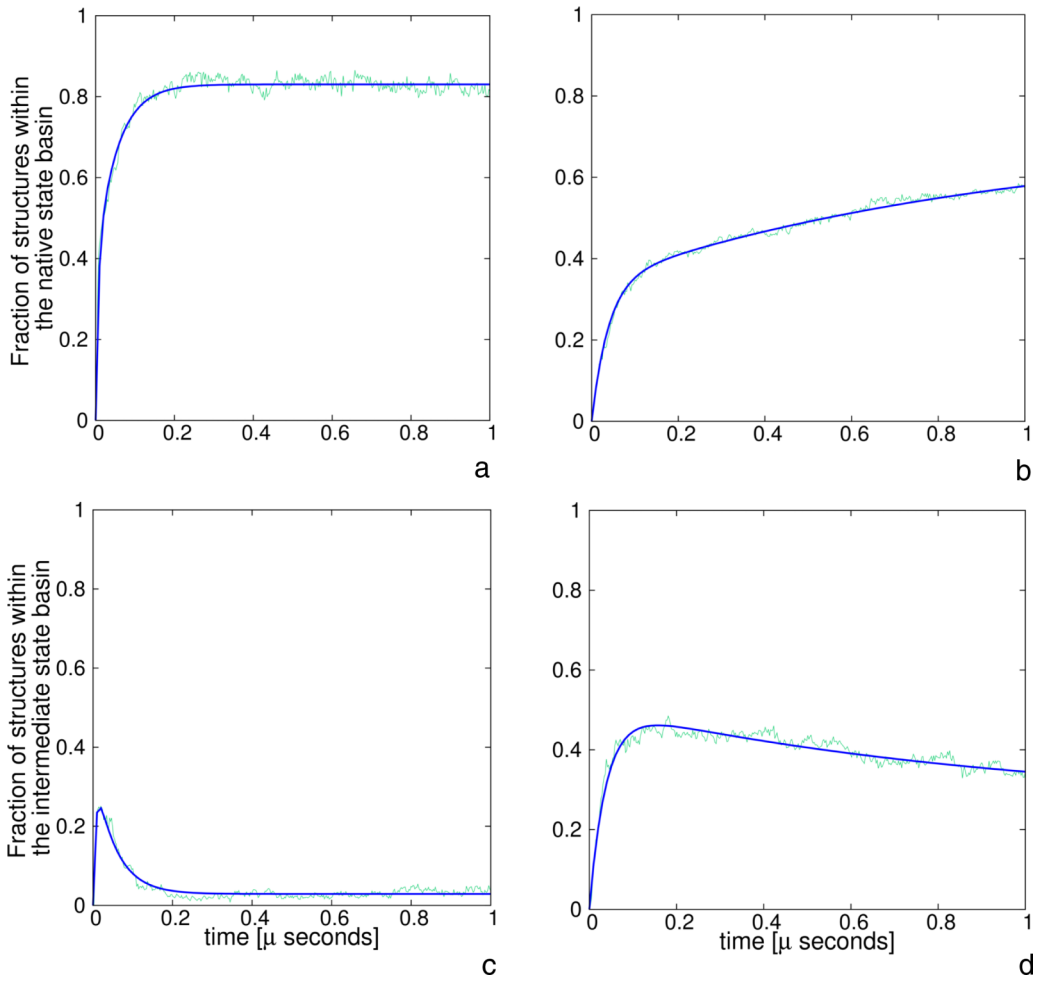

FIG. 4. Plots of the fractions of native ((a) and (b)) and intermediate ((c) and (d)) structures as a function of time for 1BDD protein without HIs ((a) and (c)) and with HIs ((b) and (d)). Green rugged lines present the simulated data and blue lines correspond to the fits of multi-exponential functions (Eqs. (20) and (21)).

assign a conformation to an intermediate-structure family were set at 5.5 Å and 5.0 Å for 1BDD and 1E0L, respectively. The cutoffs used ensured that no structure belongs to both native-like and intermediate-structure families simultaneously.

The fractions of native (f_N) or intermediate (f_I) structures as a functions of time were calculated for each of the proteins with and without HIs as averages over all 200 trajectories, as given by the following equations, respectively:

$$f_N(t) = \frac{n_N(t)}{n_{traj}}, \tag{18}$$

$$f_I(t) = \frac{n_I(t)}{n_{traj}}, \tag{19}$$

where $n_N(t)$ is the number of trajectories for which the native conformation appears at time t, $n_I(t)$ is the number of trajectories for which the intermediate conformation appears at time t and $n_{traj} = 200$ is the total number of trajectories.

Subsequently, biexponential functions (Eqs. (20) and (21)) were fitted, with the use of the nonlinear least-squares (NLLS) Marquardt-Levenberg[74] algorithm implemented in the GNUPLOT program (http://www.gnuplot.info), to the simulated $f_N(t)$ and $f_I(t)$ curves,

$$f_N(t) = C_1\left[1 - m_1 \exp\left(-\frac{t}{\tau_1}\right) - (1 - m_1)\exp\left(-\frac{t}{\tau_2}\right)\right], \tag{20}$$

$$f_I(t) = C_2\left[1 - m_2 \exp\left(-\frac{t}{\tau_1}\right) - (1 - m_2)\exp\left(-\frac{t}{\tau_2}\right)\right], \tag{21}$$

where C_1 and C_2 can be considered as the equilibrium fractions of the native and intermediate states, respectively, and τ_1 and τ_2 are the respective decay times.

III. RESULTS AND DISCUSSION

A. Effect of HIs on simulated folding

The percentages of trajectories that resulted in folded structures for all four series of simulations are collected in Table I. As can be seen, introducing HIs depletes the fraction of folded trajectories considerably. Thus, introducing HIs slows down the folding. The decrease of folding rate upon introducing HIs is apparent from Figures 4(a), 4(b), 5(a), and 5(b) in which the time evolution of the fraction of the native structure is displayed. Different results were obtained in an earlier study by Frembgen-Kesner and Elcock[16] with the use of Gō-like models, in which 11 proteins with sizes comparable to those of 1BDD and 1E0L were investigated, and HIs were

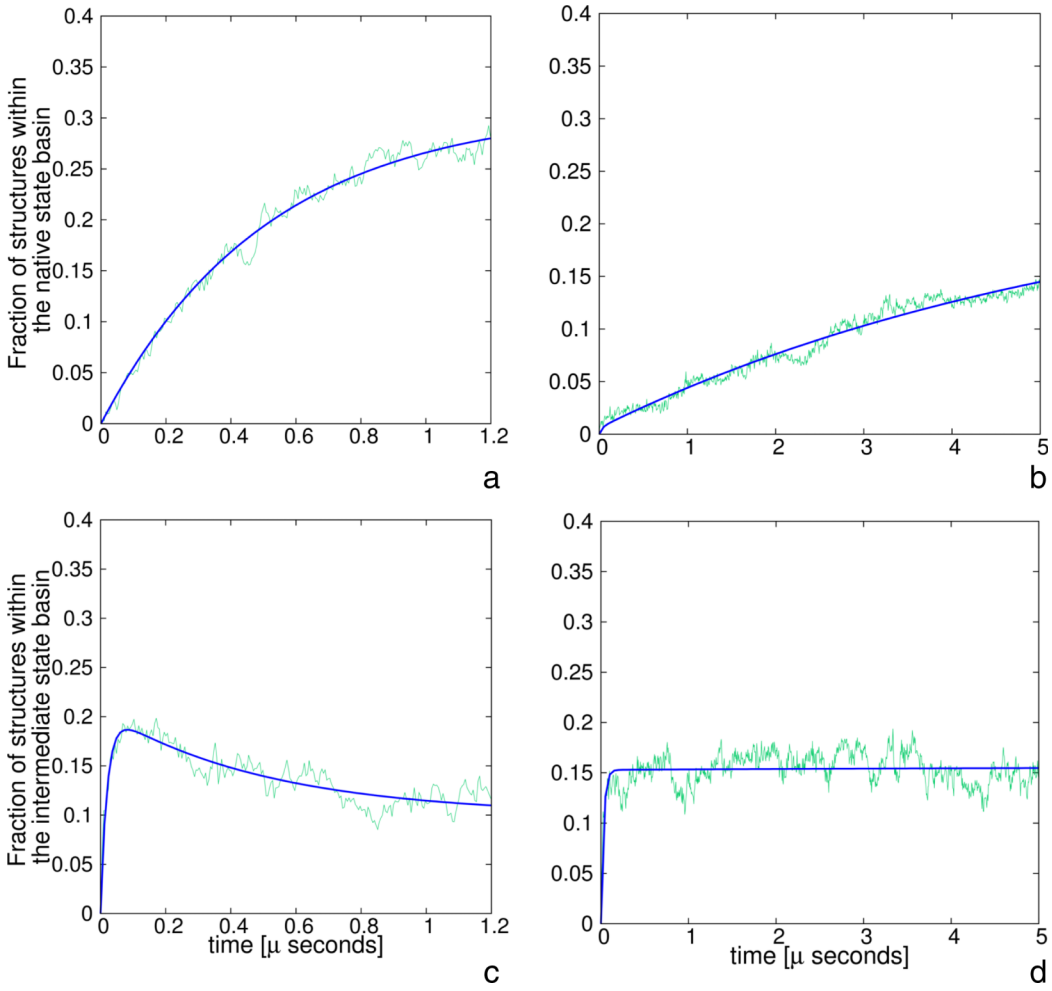

FIG. 5. Plots of the fractions of native ((a) and (b)) and intermediate ((c) and (d)) structures as a function of time for 1E0L without HIs ((a) and (c)) and with HIs ((b) and (d)). Green rugged lines present the simulated data and blue lines correspond to the fits of multi-exponential functions (Eqs. (20) and (21)).

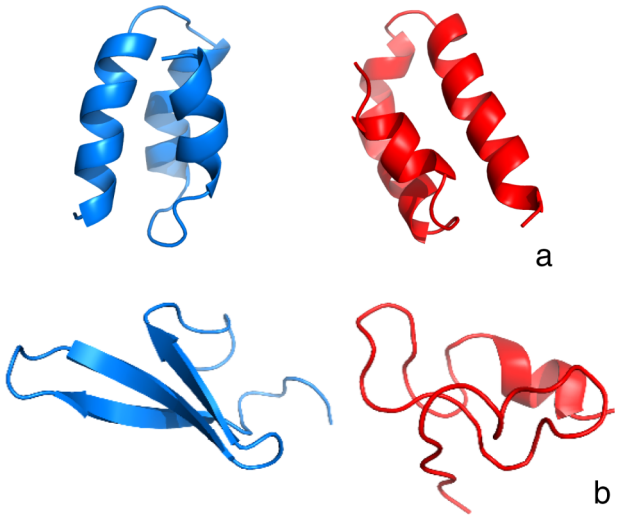

FIG. 6. Comparisons of the native (blue) and intermediate (red) structures of (a) 1BDD and (b) 1E0L proteins.

accounted for by using the RPY tensor. These researchers found that the overall folding rate increased 2 to 3 times. On the other hand, they also found that the rate of the formation of secondary-structure elements (α-helices and β-hairpins) decreased.

The reason for the slowing-down of the folding when HIs are included is the appearance of intermediates. These structures are shown in Figures 6(a) and 6(b) for 1BDD and 1E0L, respectively. The intermediate in the simulated folding of 1BDD is the mirror image of the native structure (Figure 6(a)); this intermediate was observed in our earlier studies of the folding of protein A with the UNRES force field[41–43] and in an all-atom study of this protein with the ECEPP/3 force field.[75] For 1E0L, the intermediate is a compact β-sheet-like structure but without regular peptide-group contacts and not well formed secondary structure except for a small α-helical section. This result is consistent with those of the NMR studies of the C-terminal β-hairpin of the IGG protein[76–78] and the N-terminal β-hairpin of the formin-binding WW domain[79] in which only hairpin-like structures without hydrogen bonds or even short α-helical sections were observed depending on temperature. Also, the C-terminus is bent in the opposite direction compared to the native structure (Figure 6(b)). Such structures were observed in our earlier studies of the free-energy landscapes of the folding of 1E0L and its mutants.[44–46]

Figures 4(c) and 4(d) show the time evolution of the fractions of the intermediate for 1BDD in the plain-Langevin and Langevin-with-HI simulations, while Figures 5(c) and 5(d) show the fractions of the intermediate for 1E0L. It can be seen that, for 1BDD simulations without HIs (Figure 4(c)), the fraction of the intermediate increases sharply during the first 200 ps of simulations and then decreases quickly. Conversely, inclusion of HIs results in very slow decay of the intermediate (Figure 4(d)). For 1E0L, the intermediate decays very slowly even in plain-Langevin simulations (Figure 5(c)) and effectively reaches a steady state when the HIs are included (Figure 5(d)). This result conforms well with the predictions by Tanaka[14] that HIs could have a major effect on the selection of the kinetic pathway of protein folding.

The fast formation and slow decay of intermediates after including HIs occur because of the hydrodynamic drag and create kinetic traps. This finding enables us to reconcile our results with the results obtained by Frembgen-Kesner and Elcock[16] with the use of Gō-like models. With Gō-like models, attractive interactions occur only between residues that are in contact with the native structure. Therefore, hydrodynamic drag helps to form only the native contacts; non-native interactions are always repulsive and, therefore, the respective contacts will never form. Conversely, in UNRES, the interactions depend only on the kind of residues involved and, therefore, intermediates with non-native interactions are formed quite frequently.

B. Kinetics of the simulated folding

To compare the simulated folding of the two proteins studied without and with the presence of the HIs, first-order kinetic equations (Eqs. (20) and (21)) were fitted to the fractions of native and intermediate structures. All obtained coefficients are summarized in Table II. The faster growing exponential term, characterized by the smaller value of τ, can be identified with the fast, two-state folding pathway while the slow growing exponential term, characterized by the greater value of τ, with the slow, three-state folding pathway, that includes an intermediate state. Thus, the coefficient C_1 identifies the fraction of the trajectories that represent the fast rate, while the coefficient C_2 identifies the fraction of the slow rate.

As can be seen from Table II, including HIs results in the increase of the decay times τ corresponding to both the fast and slow rates for both proteins; however the increase of τ_2 (corresponding to the slow rates that involve intermediates) is more pronounced than that of τ_1. This observation strongly suggests that HIs slow down the simulated folding, especially when intermediates are involved. This conclusion is also consistent with the facts that the C_2 coefficients of Eq. (21), which can be regarded as the "equilibrium fractions" of intermediates, are increased upon inclusion of HIs, indicating that the intermediates become persistent when HIs are included (Table II).

TABLE II. Fitted coefficients in Eqs. (20) and (21) for 1BDD and 1E0L without and with HIs.

Protein	τ_1 (μs)	τ_2 (μs)	C_1	m_1	C_2	m_2
1BDD without HIs	6.33×10^{-3}	5.32×10^{-2}	0.830	0.460	0.029	12.694
1BDD with HIs	3.85×10^{-2}	9.53×10^{-1}	0.760	0.485	0.257	1.975
1E0L without HIs	2.08×10^{-2}	5.08×10^{-1}	0.309	0.039	0.100	2.075
1E0L with HIs	2.89×10^{-2}	5.81×10^{0}	0.246	0.024	0.156	0.980

IV. CONCLUSIONS

In this work, HIs were implemented in the Langevin-dynamics simulation procedure with the UNRES force field. The influence of HIs on protein folding was investigated with the examples of a small α-helical (1BDD) and a small β-sheet (1E0L) protein. The introduction of HIs was found to slow down the folding compared to plain-Langevin simulations. We found that the decrease of the folding rate resulted from fast formation of persistent intermediates that contain non-native residue-residue contacts, when the HIs were turned on, because of the hydrodynamic drag of interacting sites toward each other. This result seems to contradict the results of an earlier study by Frembgen-Kesner and Elcock[16] carried out with the use of Gō-like models, in which it was found that HIs sped up the folding. However, in Gō-like models, the interactions between only the residues that form contacts in the native state are assigned potentials with a minimum, while all other interactions are all-repulsive. Therefore, although HIs facilitate the approach of any objects towards each other, only the native contacts will survive after they are formed. Conversely, in UNRES, the character of interactions depends on the kinds of residues involved and not on the native- or non-native character of the contact between them and, therefore, non-native contacts can survive for a longer time, creating kinetic traps. In reality, the formation of residue-residue contacts is governed by their physicochemical properties and not by the position in the sequence and, therefore, the results obtained in our study seem to be more realistic than those obtained with the use of structure-biased Gō-like models.

On the other hand, HIs are a feature of a macroscopic description of motion through a solvent, which assumes a continuous solvent governed by the Navier-Stokes equation, even though the UNRES sites (peptide groups and side chains) are not much larger in size than a water molecule. For this reason, HIs modeled in a mean-field fashion can result in too fast of an uncorrelated approach of residues, thus generating kinetic traps. The HIs can be more adequate for studies of the behavior of larger systems such as the interior of cells or lipid membrane formation investigated by Ando and Skolnick.[23,24]

ACKNOWLEDGMENTS

We would like to thank Professor Jeffrey Skolnick for suggesting this topic and Dr. Magdalena Mozolewska for many fruitful discussions.

This research was supported by Grant No. BMN 538-8370-B351-14 from the University of Gdask (to A.G.L.), Grant No. DEC-2012/06/A/ST4/00376 (to A.L., A.G.L., and A.K.S.) from the National Science Center of Poland, Grant No. Mistrz 7./2013 from the Foundation for Polish Science (to A.L.), Grant No. GM-14312 from the U.S. National Institutes of Health (to H.A.S.), and by Grant No. MCB10-19767 from the U.S. National Science Foundation (to H.A.S.). This research was supported by an allocation of advanced computing resources provided by the U.S. National Science Foundation (http://www.nics.tennessee.edu/) and by the U.S. National Science Foundation through TeraGrid resources provided by the Pittsburgh Supercomputing Center. Computational resources were also provided by (a) our 588-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (b) the U.S. National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, (c) the Beowulf cluster at the Department of Computer Science, Cornell University, (d) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdask, and (e) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) at the University of Warsaw.

[1]H. A. Scheraga, A. Liwo, S. Ołdziej, C. Czaplewski, J. Pillardy, D. R. Ripoll, J. A. Vila, R. Kaźmierkiewicz, J. A. Saunders, Y. A. Arnautova, A. Jagielska, M. Chinchio, and M. Nanias, "The protein folding problem: Global optimization of force fields," Front. Biosci. **9**, 3296–3323 (2004).
[2]A. Kryshtafovych, C. Venclovas, K. Fidelis, and J. Moult, "Progress over the first decade of CASP experiments," Proteins: Struct., Funct., Bioinf. **61**(S7), 225–236 (2005).
[3]D. Petrey and B. Honig, "Protein structure prediction: Inroads to biology," Mol. Cell **20**, 811–819 (2005).
[4]D. Baker, "Prediction and design of macromolecular structures and interactions," Philos. Trans. R. Soc., B **361**, 459–463 (2006).
[5]J. M. Bujnicki, "Protein-structure prediction by recombination of fragments," ChemBioChem **7**, 19–27 (2006).
[6]M. C. Prentiss, C. Hardin, M. P. Eastwood, C. H. Zong, and P. G. Wolynes, "Protein structure prediction: The next generation," J. Chem. Theory Comput. **2**, 705–716 (2006).
[7]A. Rojas, A. Liwo, D. Browne, and H. A. Scheraga, "Mechanism of fiber assembly; treatment of Aβ-peptide peptide aggregation with a coarse-grained united-residue force field," J. Mol. Biol. **404**, 537–552 (2010).
[8]K. A. Dill, "The protein-folding problem, 50 years on," Science **338**, 1042–1046 (2012).
[9]S. Piana, J. L. Klepeis, and D. E. Shaw, "Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations," Curr. Opin. Struct. Biol. **24**, 98–105 (2014).
[10]A. Kolinski, D. Gront, P. Pokarowski, and J. Skolnick, "A simple lattice model that exhibits a protein-like cooperative all-or-none folding transition," Biopolymers **69**, 399–405 (2003).
[11]C. Czaplewski, A. Liwo, M. Makowski, S. Ołdziej, and H. A. Scheraga, "Coarse-grained models of proteins: Theory and applications," in *Multiscale Approaches to Protein Modeling*, edited by A. Koliński (Springer, 2010), Chap. 3.
[12]A. Wagenmann and T. Geyer, "Coarse-grained simulations of protein backbone dynamics. I. Local sterics define the dihedral angles," J. Chem. Theory Comput. **8**, 4732–4745 (2012).
[13]M. Jamroz, M. Orozco, A. Kolinski, and S. Kmiecik, "Consistent view of protein fluctuations from all-atom molecular dynamics and coarse-grained dynamics with knowledge-based force-field," J. Chem. Theory Comput. **9**, 119–125 (2013).
[14]H. Tanaka, "Roles of hydrodynamic interactions in structure formation of soft matter: Protein folding as an example," J. Phys.: Condens. Matter **17**, S2795–S2803 (2005).
[15]M. Cieplak and S. Niewieczerzał, "Hydrodynamic interactions in protein folding," J. Chem. Phys. **130**, 124906 (2009).
[16]T. Frembgen-Kesner and A. H. Elcock, "Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins," J. Chem. Theory Comput. **5**, 242–256 (2009).
[17]J. Rotne and S. Prager, "Variational treatment of hydrodynamic interaction in polymers," J. Chem. Phys. **50**, 4831–4837 (1969).
[18]H. Yamakawa, "Transport properties of polymer chains in dilute solution: Hydrodynamic interaction," J. Chem. Phys. **53**, 436–443 (1970).
[19]R. Chang and A. Yethiraj, "Solvent effects on the collapse dynamics of polymers," J. Chem. Phys. **114**, 7688–7699 (2001).
[20]N. Kikuchi, A. Gent, and J. M. Yeomans, "Polymer collapse in the presence of hydrodynamic interactions," Eur. Phys. J. E **9**, 63–66 (2002).
[21]N. Kikuchi, J. F. Ryder, C. M. Pooley, and J. M. Yeomans, "Kinetics of the polymer collapse transition: The role of hydrodynamics," Phys. Rev. E **71**, 061804 (2005).
[22]D. L. Ermak and J. A. McCammon, "Brownian dynamics with hydrodynamic interactions," J. Chem. Phys. **69**, 1352–1360 (1978).

[23]T. Ando and J. Skolnick, "Crowding and hydrodynamic interactions likely dominate *in vivo* macromolecular motion," Proc. Natl. Acad. Sci. U. S. A. **107**, 18457–18462 (2010).
[24]T. Ando and J. Skolnick, "On the importance of hydrodynamic interactions in lipid membrane formation," Biophys. J. **104**, 96–105 (2013).
[25]T. Frembgen-Kesner and A. H. Elcock, "Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: The role of intermolecular hydrodynamic interactions in barnase-barstar association," Biophys. J. **99**, L75–L77 (2010).
[26]T. Ando and J. Skolnick, "Sliding of proteins non-specifically bound to DNA: Brownian dynamics studies with coarse-grained protein and dna models," PLoS Comput. Biol. **10**, e1003990 (2014).
[27]E. Chow and J. Skolnick, "Effects of confinement on models of intracellular macromolecular dynamics," Proc. Natl. Acad. Sci. U. S. A. **112**, 14846–14851 (2015).
[28]A. Liwo, S. Ołdziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, "A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data," J. Comput. Chem. **18**, 849–873 (1997).
[29]A. Liwo, C. Czaplewski, J. Pillardy, and H. A. Scheraga, "Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field," J. Chem. Phys. **115**, 2323–2347 (2001).
[30]A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik, and H. A. Scheraga, "Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins," J. Phys. Chem. B **111**, 260–285 (2007).
[31]A. Liwo, C. Czaplewski, S. Ołdziej, A. V. Rojas, R. Kaźmierkiewicz, M. Makowski, R. K. Murarka, and H. A. Scheraga, "Simulation of protein structure and dynamics with the coarse-grained UNRES force field," in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, edited by G. Voth (CRC Press, 2008), Chap. 8, pp. 107–122.
[32]Y. He, Y. Xiao, A. Liwo, and H. A. Scheraga, "Exploring the parameter space of the coarse-grained unres force field by random search: Selecting a transferable medium-resolution force field," J. Comput. Chem. **30**, 2127–2135 (2009).
[33]U. Kozłowska, G. G. Maisuradze, A. Liwo, and H. A. Scheraga, "Determination of side-chain-rotamer and side-chain and backbone virtual-bond-stretching potentials of mean force from AM1 energy surfaces of terminally blocked amino-acid residues, for coarse-grained simulations of protein structure and folding. II. Results, comparison with statistical potentials, and implementation in the UNRES force field," J. Comput. Chem. **31**, 1154–1167 (2010).
[34]A. Liwo, M. Baranowski, C. Czaplewski, E. Gołaś, Y. He, D. Jagieła, P. Krupa, M. Maciejczyk, M. Makowski, M. A. Mozolewska, A. Niadzvedtski, S. Ołdziej, H. A. Scheraga, A. K. Sieradzan, R. Ślusarz, T. Wirecki, Y. Yin, and B. Zaborowski, "A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions," J. Mol. Model. **20**, 2306 (2014).
[35]A. Liwo, M. Khalili, and H. A. Scheraga, "*Ab initio* simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains," Proc. Natl. Acad. Sci. U. S. A. **102**, 2362–2367 (2005).
[36]M. Khalili, A. Liwo, A. Jagielska, and H. A. Scheraga, "Molecular dynamics with the united-residue model of polypeptide chains. II. Langevin and berendsen-bath dynamics and tests on model α-helical systems," J. Phys. Chem. B **109**, 13798–13810 (2005).
[37]J. Lee, A. Liwo, and H. A. Scheraga, "Energy-based *de novo* protein folding by conformational space annealing and an off-lattice united-residue force field: Application to the 10-55 fragment of staphylococcal protein a and to *apo*-calbindin D9K," Proc. Natl. Acad. Sci. U. S. A. **96**, 2025–2030 (1999).
[38]A. Liwo, J. Lee, D. R. Ripoll, J. Pillardy, and H. A. Scheraga, "Protein structure prediction by global optimization of a potential energy function," Proc. Natl. Acad. Sci. U. S. A. **96**, 5482–5485 (1999).
[39]S. Ołdziej, C. Czaplewski, A. Liwo, M. Chinchio, M. Nanias, J. Vila, M. Khalili, Y. A. Arnautova, A. Jagielska, M. Makowski, H. D. Schafroth, R. Kaźmierkiewicz, D. R. Ripoll, J. Pillardy, J. Saunders, Y. Kang, K. Gibson, and H. A. Scheraga, "Physics-based protein-structure prediction using a hierarchical protocol based on the unres force field: Assessment in two blind tests," Proc. Natl. Acad. Sci. U. S. A. **102**, 7547–7552 (2005).
[40]Y. He, M. A. Mozolewska, P. Krupa, A. K. Sieradzan, T. K. Wirecki, A. Liwo, K. Kachlishvili, S. Rackovsky, D. Jagieła, R. Ślusarz, C. R. Czaplewski, S. Ołdziej, and H. A. Scheraga, "Lessons from application of the UNRES force field to predictions of structures of CASP10 targets," Proc. Natl. Acad. Sci. U. S. A. **110**, 14936–14941 (2013).
[41]M. Khalili, A. Liwo, and H. A. Scheraga, "Kinetic studies of folding of the B-domain of staphylococcal protein a with molecular dynamics and a united-residue (UNRES) model of polypeptide chains," J. Mol. Biol. **355**, 536–547 (2006).
[42]Y. Yin, G. G. Maisuradze, A. Liwo, and H. A. Scheraga, "Hidden protein folding pathways in free-energy landscapes uncovered by network analysis," J. Chem. Theory Comput. **8**, 1176–1189 (2012).
[43]K. Kachlishvili, G. G. Maisuradze, O. A. Martin, A. Liwo, J. A. Vila, and H. A. Scheraga, "Accounting for a mirror-image conformation as a subtle effect in protein folding," Proc. Natl. Acad. Sci. U. S. A. **111**, 8458–8463 (2014).
[44]R. Zhou, G. G. Maisuradze, D. Sunol, T. Todorovski, M. J. Macias, Y. Xiao, H. A. Scheraga, C. Czaplewski, and A. Liwo, "Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements," Proc. Natl. Acad. Sci. U. S. A. **111**, 18243–18248 (2014).
[45]G. G. Maisuradze, R. Zhou, A. Liwo, Y. Xiao, and H. A. Scheraga, "Effects of mutation, truncation, and temperature on the folding kinetics of a WW domain," J. Mol. Biol. **420**, 350–365 (2012).
[46]G. G. Maisuradze, J. Medina, K. Kachlishvili, P. Krupa, M. A. Mozolewska, P. Martin-Malpartida, L. Maisuradze, M. J. Macias, and H. A. Scheraga, "Preventing fibril formation of a protein by selective mutation," Proc. Natl. Acad. Sci. U. S. A. **12**, 13549–13554 (2015).
[47]Y. He, A. Liwo, H. Weinstein, and H. A. Scheraga, "PDZ binding to the BAR domain of PICK1 is elucidated by coarse-grained molecular dynamics," J. Mol. Biol. **405**, 298–314 (2011).
[48]E. I. Gołaś, G. G. Maisuradze, P. Senet, S. Ołdziej, C. Czaplewski, H. A. Scheraga, and A. Liwo, "Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics," J. Chem. Theory Comput. **8**, 1334–1343 (2012).
[49]A. G. Lipska, A. K. Sieradzan, P. Krupa, M. A. Mozolewska, S. D'Auria, and A. Liwo, "Studies of conformational changes of an arginine-binding protein from *Thermatoga maritima* in the presence and absence of ligand via molecular dynamics simulations with the coarse-grained UNRES force field," J. Mol. Model. **3**, 64 (2015).
[50]M. A. Mozolewska, P. Krupa, H. A. Scheraga, and A. Liwo, "Molecular modeling of the binding modes of the iron-sulfur protein to the Jac1 co-chaperone from *Saccharomyces cerevisiae* by all-atom and coarse-grained approaches," Proteins: Struct., Funct., Bioinf. **83**, 1414–1426 (2015).
[51]H. Gouda, H. Torigoe, A. Saito, M. Sato, Y. Arata, and I. Shimada, "Three-dimensional solution structure of the B domain of staphylococcal protein A: Comparisons of the solution and crystal structures," Biochemistry **31**, 9665–9672 (1992).
[52]M. J. Macias, V. Gervais, C. Civera, and H. Oschkinat, "Structural analysis of WW domains and design of a WW prototype," Nat. Struct. Biol. **7**, 375–379 (2000).
[53]H. Shen, A. Liwo, and H. A. Scheraga, "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," J. Phys. Chem. B **113**, 8738–8744 (2009).
[54]M. U. Johansson, M. de Chateau, M. Wikstrom, S. Forsen, T. Drakenberg, and L. Bjorck, "Solution structure of the albumin-binding GA module: A versatile bacterial protein domain," J. Mol. Biol. **266**, 859–865 (1997).
[55]N. D. Clarke, C. R. Kissinger, J. Desjarlais, G. L. Gilliland, and C. O. Pabo, "Structural studies of the engrailed homeodomain," Protein Sci. **3**, 1779–1787 (1994).
[56]A. K. Sieradzan, P. Krupa, H. A. Scheraga, A. Liwo, and C. Czaplewski, "Physics-based potentials for the coupling between backbone- and side-chain-local conformational states in the united residue (UNRES) force field for protein simulations," J. Chem. Theory Comput. **11**, 817–831 (2015).
[57]R. Kubo, "The fluctuation-dissipation theorem," Rep. Prog. Phys. **29**, 255–284 (1966).
[58]J. G. de le Torre and V. A. Bloomfield, "Hydrodynamic properties of macromolecular complexes. I. Translation," Biopolymers **16**, 1747–1763 (1977).
[59]R. G. Larson, "The rheology of dilute solutions of flexible polymers: Progress and problems," J. Rheol. **49**, 1–70 (2005).
[60]M. Vergeles, P. Keblinski, J. Koplik, and J. Banavar, "Stokes drag at the molecular-level," Phys. Rev. Lett. **75**, 232–235 (1995).
[61]M. Levitt, "A simplified representation of protein conformations for rapid simulation of protein folding," J. Mol. Biol. **104**, 59–107 (1976).
[62]A. Liwo, S. Ołdziej, C. Czaplewski, U. Kozłowska, and H. A. Scheraga, "Parameterization of backbone-electrostatic and multibody contributions

to the UNRES force field for protein-structure prediction from *ab initio* energy surfaces of model systems," J. Phys. Chem. B **108**, 9421–9438 (2004).
[63]J. K. G. Dhont, *An Introduction to Dynamics of Colloids* (Elsevier, Amsterdam, 1996).
[64]C. Oseen, *Hydrodynamik* (Akademisches Verlag, Leipzig, 1927), Vol. 1.
[65]J. Burgers, in *Second Report on Viscosity and Plasticity* (Academy of Sciences at Amsterdam, Nordemann Publishing Company, 1938), Chap. 3.
[66]B. Carrasco and J. G. de le Torre, "Hydrodynamic properties of rigid particles: Comparison of different modeling and computational procedures," Biophys. J. **76**, 3044–3057 (1999).
[67]P. Zipper and H. Durchschlag, "Calculation of hydrodynamic parameters of proteins crystallographic data using multibody approaches," in *Analytical Ultracentrifugation IV*, edited by R. Jaenicke and H. Durchschlag, Progress in Colloid and Polymer Science (Steinkopff Darmstadt, 1997), Vol. 107, pp. 58–71.
[68]M. Kröger, A. Alba-Pérez, M. Laso, and H. Öttinger, "Variance reduced Brownian simulation of a bead-spring chain under steady shear flow considering hydrodynamic interaction effects," J. Chem. Phys. **113**, 4767–4773 (2000).
[69]P. Krupa, A. K. Sieradzan, S. Rackovsky, M. Baranowski, S. Ołdziej, H. A. Scheraga, A. Liwo, and C. Czaplewski, "Improvement of the treatment of loop structures in the UNRES force field by inclusion of coupling between backbone- and side-chain-local conformational states," J. Chem. Theory Comput. **9**, 4620–4632 (2013).
[70]F. Murtagh, *Multidimensional Clustering Algorithms* (Physica-Verlag, Vienna, 1985).
[71]F. Murtagh and A. Heck, *Multivariate Data Analysis* (Kluwer Academic Publishers, 1987).
[72]See supplementary material at http://dx.doi.org/10.1063/1.4948710 for the structures of intermediates and for the probality distribution of the C^{α}-RMSD values for the native and intermediate structures of 1BDD and 1E0L proteins with and without the inclusion of HI for each quater of the simulation.
[73]M. S. Nikulin, "Student test," in *Encyclopedia of Mathematics*, edited by U. Rehmann (Springer, 2012).
[74]D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," J. Soc. Ind. Appl. Math. **11**, 431–441 (1963).
[75]J. A. Vila, D. R. Ripoll, and H. A. Scheraga, "Atomically detailed folding simulation of the B domain of staphylococcal protein A from random structures," Proc. Natl. Acad. Sci. U. S. A. **100**, 14812–14816 (2003).
[76]A. Skwierawska, W. Żmudzińska, S. Ołdziej, A. Liwo, and H. A. Scheraga, "Mechanism of formation of the C-terminal β-hairpin of the B3 domain of the immunoglobulin binding protein G from *Streptococcus*. Part II. Interplay of local backbone conformational dynamics and long-range hydrophobic interactions in hairpin formation," Proteins: Struct., Funct., Bioinf. **76**, 637–654 (2009).
[77]A. Lewandowska, S. Ołdziej, A. Liwo, and H. A. Scheraga, "Mechanism of formation of the C-terminal β-hairpin of the B3 domain of the immunoglobulin binding protein G from *Streptococcus*. Part IV. Implication for the mechanism of folding of the parent protein," Biopolymers **93**, 469–480 (2010).
[78]A. Lewandowska, S. Ołdziej, A. Liwo, and H. A. Scheraga, "β-hairpin-forming peptides; models of early stages of protein folding," Biophys. Chem. **151**, 1–9 (2010).
[79]J. Makowska, W. Żmudzińska, D. Uber, and L. Chmurzyński, "A study of the influence of charged residues on β-hairpin formation by nuclear magnetic resonance and molecular dynamics," Protein J. **33**, 525–535 (2014).