PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Mimicking the action of folding chaperones by Hamiltonian replica-exchange molecular dynamics simulations: Application in the refinement of de novo models

Hao Fan,[1] Xavier Periole,[2] and Alan E. Mark[2,3]*

[1] Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94158-2330

[2] Department of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute (GBB) and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

[3] School of Chemistry and Molecular Bioscience, and the Division of Chemistry and Structural Biology Institute for Molecular Biosciences, University of Queensland, St Lucia, Queensland 4072, Australia

ABSTRACT

The efficiency of using a variant of Hamiltonian replica-exchange molecular dynamics (Chaperone H-replica-exchange molecular dynamics [CH-REMD]) for the refinement of protein structural models generated de novo is investigated. In CH-REMD, the interaction between the protein and its environment, specifically, the electrostatic interaction between the protein and the solvating water, is varied leading to cycles of partial unfolding and refolding mimicking some aspects of folding chaperones. In 10 of the 15 cases examined, the CH-REMD approach sampled structures in which the root-mean-square deviation (RMSD) of secondary structure elements (SSE-RMSD) with respect to the experimental structure was more than 1.0 Å lower than the initial de novo model. In 14 of the 15 cases, the improvement was more than 0.5 Å. The ability of three different statistical potentials to identify near-native conformations was also examined. Little correlation between the SSE-RMSD of the sampled structures with respect to the experimental structure and any of the scoring functions tested was found. The most effective scoring function tested was the DFIRE potential. Using the DFIRE potential, the SSE-RMSD of the best scoring structures was on average 0.3 Å lower than the initial model. Overall the work demonstrates that targeted enhanced-sampling techniques such as CH-REMD can lead to the systematic refinement of protein structural models generated de novo but that improved potentials for the identification of near-native structures are still needed.

Proteins 2012; 80:1744–1754.

Key words: protein structure prediction; protein structure refinement; replica-exchange molecular dynamics; chaperone; statistical potential.

INTRODUCTION

Knowing the structure of a protein at atomic resolution is often the key to understanding its biological function. However, while during the last 7 years, the number of experimentally determined protein structures deposited in the Protein Data Bank (PDB) increased by a factor of 3 from 23,096 to 67,421 (November 2010),[1] and the number of sequences in the Universal Protein Resource (UniProt) increased by an order of magnitude from 1.2 to 12.8 million, over the same period.[2] The gap between the number of sequences that are identified and the number of high resolution protein structures solved experimentally continues to grow and is a problem that can only be bridged by improving the reliability of protein structure prediction methods,[3] including both comparative modeling and de novo approaches.

Template-based comparative modeling (homology modeling and fold recognition) is the most reliable and widely used method to obtain high quality structural models for a given target protein.[4] The utility of compar-

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: Australian Research Council; Grant numbers: DP0878608, DP110100327.
*Correspondence to: Alan E. Mark, School of Chemistry and Molecular Bioscience, and the Institute for Molecular Biosciences, University of Queensland, St Lucia, Queensland 4072, Australia. E-mail: a.e.mark@uq.edu.au
Received 7 December 2011; Revised 11 February 2012; Accepted 3 March 2012
Published online 13 March 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24068

ative modeling is limited, however, by the availability of suitable template structures. In cases where no structure of an analogous protein has been determined, the only option is to attempt to predict the structure of the protein de novo.[5,6] Over the last decade, significant progress in the de novo prediction of protein structures has been made. A major milestone was the work by Baker *et al.*,[7] in which they were able to predict de novo 5 of 16 small single-domain proteins (<85 residues) to a resolution <1.5 Å using the ROSETTA all-atom scoring function and a Monte Carlo (MC) global search method. However, a recent Critical Assessment of Techniques for Protein Structure Prediction exercise suggests that the root-mean-square deviation (RMSD) of Cα atoms between de novo models and the native structure is more typically greater than 4 Å.[8–13] The primary problem in these cases is not the identification of the overall fold but the ability of current approaches to refine the final model.

The refinement of the structure of a protein when starting from a proposed model generated either by homology or de novo depends on two factors: (a) the ability to search conformational space around the native structure efficiently and (b) a function that can be used to identify the native state. In general, different refinement procedures are used depending on how close the model is expected to lie to the native structure. At lower resolution, considerable success has been achieved using fragment-based approaches, combined with MC simulation techniques based on a statistical potential.[7,14] The challenge is in attempting to go from a low-resolution model to a high-resolution model. In principle, molecular dynamics (MD) simulation techniques would appear well suited for the refinement of low-resolution protein models.[15] However, in general, MD simulation has proved to be ineffective. The potential energy surface of a protein at an atomic level is extremely rugged leading to an inability to sample a sufficient region of conformational space on the time scales accessible. Nevertheless, MD simulations on times scales of tens to hundreds of nanoseconds in explicit water have been shown to lead to significant improvements in the structures of small proteins.[16] One way refinement protocols can be improved is by the use of enhanced sampling techniques such as replica exchange molecular dynamics (REMD).[17–21] For example, in one study, REMD, combined with a generalized Born solvation model and structural restraints, resulted in significant improvement in four of five template-based models.[22] REMD in explicit water has also proved to be effective in the global refinement of small proteins, improving the average backbone RMSD of secondary structure elements (SSE-RMSD) with respect to the corresponding experimental structure by 0.82 Å for 21 template-based models.[23]

Low-resolution template-based or de novo models have many similarities to misfolded proteins. Although they have not completely adopted their native fold, they normally have extensive regions of near-native secondary structure and are metastable on extended timescales. This makes it (a) difficult to recognize if the protein is not correctly folded and (b) difficult to sample alternative conformations using standard simulations techniques. In vivo, the folding of a protein to its native state is often promoted by molecular chaperones. Chaperones act by driving the refolding of misfolded proteins in an energy-dependent manner. This can be achieved by repeated cycles of binding and release and/or sequestering the protein in a confined space.[24–26] The mechanism by which chaperones promote folding has been the subject of multiple simulation studies.[27–33] For example, some of the effects of chaperones can be mimicked by periodically increasing and decreasing the partial charges of water molecules driving repeated cycles of folding and unfolding. When this was incorporated into standard MD simulations, significant improvements in the overall fold of three proteins predicted de novo was achieved.[34,35] This chaperone approach has also been implemented as a part of a Hamiltonian REMD (H-REMD) protocol. In this case, the protein–water interaction was attenuated as a function of a coupling parameter.[36] This method was used for successfully folding a small alpha helical protein, 3K(I), within 4 ns.

Here, the efficiency of varying the nature of the interaction between the protein and the environment (the chaperone effect) as implemented as part of a H-REMD (Chaperone H-REMD [CH-REMD]) scheme to enhance sampling during structural refinement is investigated. The methodology has been evaluated in terms of its ability to sample near-native conformations for 15 small proteins starting from models that had been generated de novo. The models used in this work were generated originally by Baker *et al.*[35] using ROSETTA and were chosen as they had been used previously to evaluate the use of MD in protein structure refienment.[16] The conformations sampled during the simulations were assessed by determining the backbone SSE-RMSD with respect to the corresponding experimental structures. They were also ranked using three alternative scoring functions: the DFIRE potential,[37] a modified RAPDF potential,[23,37,38] and the ROSETTA scoring function.[39] Configurations taken from a series of independent MD simulations starting from the same initial conformation were used as a control. Overall, CH-REMD was found to sample near-native regions of conformational space much more efficiently than multiple independent MD simulations. The lowest SSE-RMSD sampled by CH-REMD was on average 1.14 Å lower than the SSE-RMSD of the initial model. The DFIRE potential[37] and the modified RAPDF potentials[23,37,38] selected near-native conformations more reliably than the ROSETTA scoring function.[39] The models selected by DFIRE and RAPDF showed an average improvement of 0.46 and 0.42 Å in the SSE-RMSD, respectively.

MATERIALS AND METHODS

Test set

Previously,[16] the efficiency of using classical MD simulation techniques to refine protein structures predicted

Table I
Properties of the 15 Proteins Used to Test the Refinement Protocol

PDB code	Description	Exp.	N_{res}	SCOP	SSE%	N_{charge}
1a1z	FADD death-effector domain	NMR	83	All α	75	−3
1afi	Mercuric ion binding protein MerP	NMR	72	α + β	61	3
1ail	RNA-binding domain of the influenza virus NS1 protein	X-ray	70	All α	86	2
1aoy	Arginine repressor (ArgR), N-*terminal DNA-binding domain*	NMR	78	All α	54	3
1coo	C-*terminal domain of RNA polymerase alpha subunit*	NMR	81	All α	44	−3
1lea	LexA repressor, N-*terminal DNA-binding domain*	NMR	72	All α	58	2
1sap	DNA-binding protein	NMR	66	All β	70	6
1sro	S1 RNA-binding domain of polyribonucleotide phosphorylase, PNPase	NMR	76	All β	36	1
1stu	Maternal effect protein staufen	NMR	68	α + β	68	5
1tuc	alpha-Spectrin, SH3 domain	X-ray	61	All β	46	0
1vcc	Vaccinia DNA topoisomerase I, 9 kDa N-terminal fragment	X-ray	77	α + β	47	−1
1vif	R67 dihydrofolate reductase	X-ray	60	All β	48	0
2bby	DNA-binding domain from rap30	NMR	69	All α	64	3
2ezh	I γ subdomain of the Mu end DNA-binding domain of phage Mu transposase	NMR	65	All α	69	−2
2fmr	KH1 domain of FMR1	NMR	65	α + β	54	−4

Exp., experimental method used to determine the protein structure; N_{res}, total number of residues; SCOP, secondary structure class; SSE%, the proportion of residues in secondary structures among the protein; N_{charge}, the net charge of the protein at pH 7.0.

de novo was evaluated using a test set of 15 proteins (Table I). These proteins range in size from 60 to 83 amino acid residues. Four have been determined using X-ray crystallography, and 11 were determined using NMR spectroscopy. According to the SCOP classification system,[40] of the 15 proteins, 7 are all-α, 4 are all-β, and

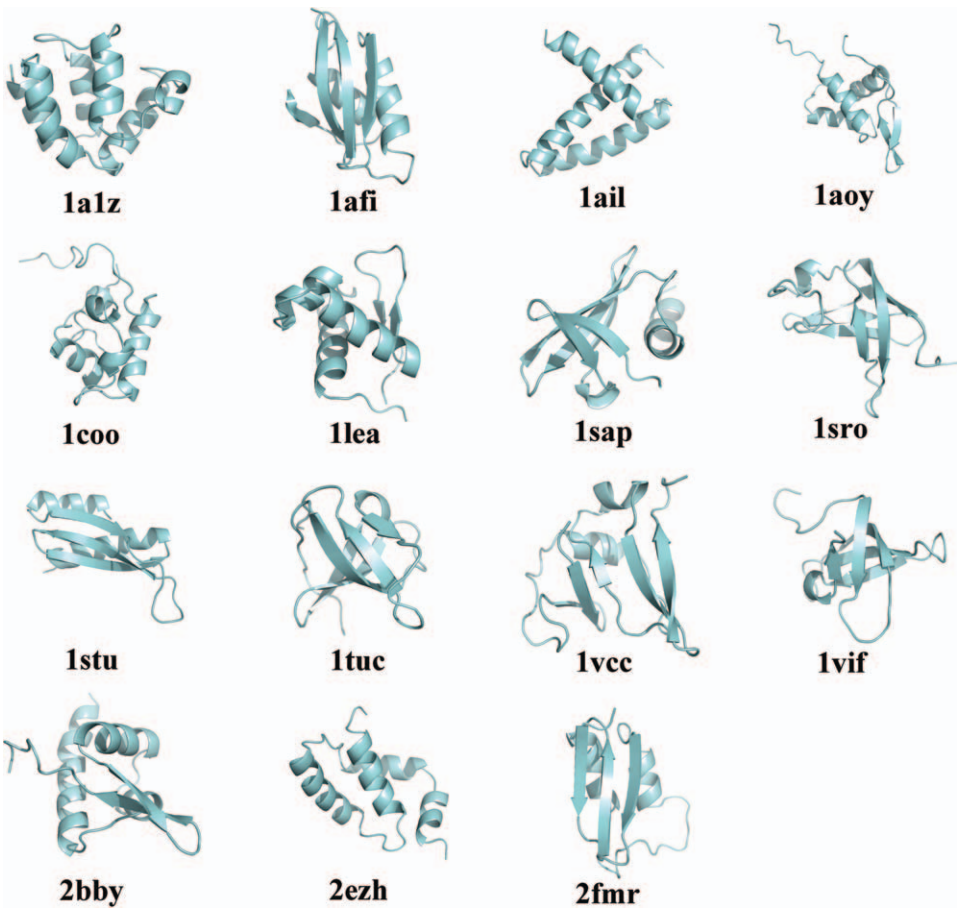

Figure 1
A cartoon representation of the experimental structures of the 15 proteins used to test the refinement protocol. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table II
Analysis of Conformations Sampled During the Replica Exchange Molecular Dynamics and Multiple Independent Molecular Dynamics Simulations*

		20 ns REMD (12 replicas)		12 × 20 ns MD	
Model	$RMSD_{init}$	$RMSD_{min}$	Perc. (%)	$RMSD_{min}$	Perc. (%)
1a1z	5.31	**2.80** (0.992)	73.6 (37.5)	**3.18**	30.2
1afi	2.46	**1.02** (1.008)	60.8 (45.8)	**1.27**	81.8
1ail	5.22	**4.22** (1.008)	8.1 (7.5)	**4.04**	14.9
1aoy	4.38	**3.30** (1.016)	29.1 (26.5)	*3.82*	7.5
1coo	3.68	**2.49** (0.918)	38.0 (22.5)	*3.08*	5.3
1lea	2.97	*2.21* (1.008)	22.7 (20.9)	*2.46*	17.7
1sap	3.26	*2.60* (0.931)	1.5 (2.1)	3.06	0.2
1sro	2.86	*2.25* (1.008)	15.2 (28.6)	*2.22*	19.6
1stu	5.22	**3.79** (0.983)	20.8 (6.7)	*4.72*	3.5
1tuc	2.67	2.25 (0.983)	5.6 (1.1)	2.41	0.6
1vcc	5.48	**3.86** (1.016)	53.9 (71.8)	*4.75*	35.6
1vif	5.06	**3.76** (0.931)	24.9 (32.7)	**4.00**	11.5
2bby	5.21	**4.20** (1.024)	63.0 (41.8)	*4.55*	18.6
2ezh	2.87	*2.24* (0.964)	0.5 (0.0)	2.57	0.0
2fmr	3.95	**2.50** (1.008)	47.8 (63.6)	**2.23**	57.6
Average	4.04	2.90	31.0 (27.3)	3.22	21.0

*RMSD (in Å), the root mean square positional deviation of the backbone of residues involved in elements of secondary structure with respect to the experimentally determined structure; **Model**, the model structure of the protein target; **$RMSD_{init}$**, the RMSD of the initial structure used in the REMD or multiple-MD simulations; **$RMSD_{min}$**, the minimum RMSD sampled during the REMD or multiple independent MD simulations (the number in parenthesis corresponds to the value of the scaling factor λ of the replica in which the configuration corresponding to $RMSD_{min}$ was found); **Perc.**, the percentage of conformations that have lower RMSD than **$RMSD_{init}$** (the number in parenthesis corresponds to the percentage of considering only the replica in which λ was equal to 1). $RMSD_{min}$ values that are at least 1.0 Å and 0.5 Å lower than the corresponding $RMSD_{init}$ values are in bold and italic, respectively.

4 are α+β (Fig. 1). In the previous study, 60 models, 4 for each protein were used to test the refinement protocol. In the current work, only a single model for each of the test proteins was chosen. As the focus of this study is on the final stages of refinement, the model with the lowest SSE-RMSD with respect to the corresponding experimental structure was used. The SSE-RMSD with respect to the corresponding experimental structure for these 15 models range from 2.5 to 5.5 Å (Table II).

Refinement protocol

The refinement protocol consisted of two stages:

1. The sampling of conformational space around the initial protein model.
2. The selection of native-like models using a variety of scoring functions.

Conformational sampling

The sampling of conformational space around the original model was performed in two ways: (a) using standard MD simulation techniques and (b) using a variant of Hamiltonian replica-exchange molecular dynamics (CH-REMD) in which the interaction between the protein and its environment varied as described below. All simulations were performed in explicit solvent using the GROMACS (Groningen Machine for Chemical Simulation) package[41,42] in conjunction with the GROMOS96 43a1 force field[43] for condensed phases. The protonation state of each ionizable amino acid residue was set appropriate for pH 7.0. No counter ions were added to neutralize the system. Each protein model was solvated in a rhombic dodecahedron box using the SPC water model.[44] The minimum distance between any atom of the protein and the wall of the unit cell was 8 Å. Nonbonded interactions were evaluated using a twin-range method. Coulomb and van der Waals interactions within a short-range cutoff of 0.9 nm were evaluated every time step. Longer range Coulomb and van der Waals interactions between 0.9 and 1.4 nm were updated every five time steps together with the pairlist. To minimize the effects of truncating the electrostatic interactions beyond the 1.4 nm long-range cutoff, a reaction field correction[45] was applied using a relative dielectric constant of 78. A time step of 2 fs was used to integrate the equations of motion. Covalent bonds in the proteins were constrained using the LINCS algorithm.[46] The SETTLE algorithm[47] was used to constrain the geometry of the water molecules. To generate the starting configuration for each system, the following protocol was used. After energy minimization using a steepest descent algorithm a 100 ps MD simulation with position restraints on the protein (PRMD) was performed to relax the system. During PRMD, a constant temperature of 300 K and a constant pressure of 1 bar were maintained by coupling to an external heat and an isotropic pressure bath.[48] The structures obtained after PRMD were used as the initial structures for the unrestrained MD simulations and the CH-REMD simulations.

Molecular dynamics

For each of the 15 systems generated as described above, a series of 12 independent 20 ns simulations with different starting velocities were performed. During these simulations, the configurations were saved every 2.5 ps for analysis.

CH-REMD

In Hamiltonian REMD,[49–51] the exchange probability between two replicas i and j, simulated using the potential energy functions E_m and E_n, respectively is given by:

$$\mathrm{acc}(i \rightarrow j) = \min[0, \exp(-\Delta)] \qquad (1)$$

with

$$\Delta = \beta[E_m(X_j) + E_n(X_i) - E_m(X_i) - E_n(X_j)] \qquad (2)$$

where X_i *and* X_j represent the configurations of the replicas i and j, respectively. β is the reciprocal temperature ($\beta = 1/k_B T$), which is held constant. In the CH-REMD scheme used in this work, only the electrostatic interaction between the protein and water was varied. All other inter-

actions were unchanged. Thus, Δ in Eq. 2 can be expressed simply as:

$$\Delta = \beta(\lambda_n - \lambda_m)[U_{elec}^{prot-water}(X_i) - U_{elec}^{prot-water}(X_j)] \quad (3)$$

where λ is a scaling factor that determines the electrostatic interaction between the atoms of the protein and the surrounding water molecules. When $\lambda = 1$, the protein experiences a normal water environment. When $\lambda > 1$, the polarity of the solvent is increased promoting hydrogen bonding interactions between the protein and solvent, thereby destabilizing elements of secondary structure. When $\lambda < 1$, the polarity of the solvent is decreased. This increases the propensity of the protein to form intramolecular hydrogen bonds and thereby the formation of secondary structure. By changing the polarity of the environment, the protein is forced to undergo a series of partial unfolding and refolding events mimicking different stages in the refolding cycle driven by chaperones such as GroEL.

Each system was equilibrated for 1 ns at five different λ values, 0.90, 0.94, 0.98, 1.02, and 1.06. The average protein-solvent electrostatic energies from these five trail simulations were fitted by a polynomial of λ. Equation 1 was solved iteratively in the range $\lambda = 0.90–1.06$, such that an exchange ratio of approximately 0.15 was achieved. For each protein model, 12 replicas (λ = 0.918, 0.931, 0.943, 0.954, 0.964, 0.974, 0.983, 0.992, 1.000, 1.008, 1.016, 1.024) were generated. More replicas were simulated with a λ value lower than 1 to limit the extent of unfolding and promote refolding. Starting from this set of initial structures, 20 ns CH-REMD simulation was performed at 300 K and constant volume (NVT) with exchanges attempted every 5 ps. Snapshots were recorded every 2.5 ps for analysis resulting in a total of 8000 conformations per replica.

Model selection

Three scoring functions that are widely used for protein structure prediction were used for the model selection. The scoring functions were DFIRE,[37] RAPDF,[38] and ROSETTA.[39,52] Both DFIRE and RAPDF are atom-based, distance-dependent pairwise statistical potentials. The DFIRE potential uses a distance-scaled, finite-gas reference state calculated by using a set of uniformly distributed noninteracting points in finite spheres. RAPDF potential uses the conditional probability reference state. In this study, the RAPDF potential was modified,[23] with the distance binning procedure of DFIRE. The DFIRE and RAPDF potentials were compared with the default ROSETTA score[39] The default ROSETTA score contains seven terms, including rama (Ramachandran torsion preferences), LJ (Lennard-Jones interactions), hb (hydrogen bonding), solv (solvation), pair (residue pair interactions such as electrostatics and disulfides), dun (rotamer self-energy), and ref (unfolded state reference energy). In this study, the standard weighting factors were used for all terms, and the softened Lennard-Jones (LJ) potential option was used to compensate for the differences between the atom radii used in the MD simulations (GROMOS96 43a1) and those used in ROSETTA (CHARMM22[53]).

To select the best model for each protein based on the scoring function, the following procedure was used. First, all the models generated during the CH-REMD simulations were scored and ranked. The top five ranked models were then analyzed in terms of their similarity to the experimental structure based on their SSE-RMSD. The backbone RMSD of just the elements of secondary structure was used as a measure of similarity as opposed to the backbone RMSD of entire protein, as the SSE-RMSD is less sensitive to changes in the conformation of loops and the termini of the chains, which can be affected by crystal packing effects. Thus, the SSE-RMSD reflects better the accuracy of the overall fold. An analysis of the model selection based on the backbone RMSD of entire protein is given as Supporting Information Table S2. The best model was defined as the one with the lowest SSE-RMSD among the top five ranked models (Table III). The SSE-RMSD of the best-ranked model according to the scoring function itself is also given in Table III.

RESULTS

Conformational sampling

CH-REMD

The positional RMSD of secondary structure elements (SSE-RMSD) with respect to the corresponding experimental structure was used as the main criterion to evaluate the sampling efficiency of CH-REMD. For each protein, the conformation with the lowest SSE-RMSD value ($RMSD_{min}$) generated in CH-REMD is given in Table II and compared with the SSE-RMSD of the initial model ($RMSD_{init}$). For all of the 15 proteins, conformations that were closer to the experimental structures than the initial model were sampled during the CH-REMD simulations. For 14 proteins, the $RMSD_{min}$ was more than 0.5 Å lower than the $RMSD_{init}$. For 10 proteins, the $RMSD_{min}$ was more than 1.0 Å lower than the $RMSD_{init}$. On average, the $RMSD_{min}$ was 1.14 Å lower than the $RMSD_{init}$. Although the configuration with $RMSD_{min}$ was not observed in the replica corresponding to the native interaction $\lambda = 1.0$ for any of the proteins, the configuration with $RMSD_{min}$ appeared to be evenly distributed in the replicas and in general close to the replica in which $\lambda = 1.0$. This suggests that the effect of varying the interaction is comparatively mild and that the native state is likely to be most stable when $\lambda = 1.0$. Furthermore, the proportion of conformations that lay closer to the native

Table III
Analysis of the Conformations Selected by Three Alternative Scoring Functions*

			ROSETTA		RAPDF		DFIRE	
Protein	$RMSD_{init}$	$RMSD_{min}$	Lowest RMSD	Best score	Lowest RMSD	Best score	Lowest RMSD	Best score
1a1z	5.31	**2.80**	**3.86** (3)	*4.47*	**4.27** (2)	*4.38*	**3.95** (3)	*4.41*
1afi	2.46	**1.02**	**1.29** (2)	**1.39**	**1.28** (4)	**1.43**	**1.27** (1)	**1.27**
1ail	5.22	**4.22**	5.89 (4)	7.10	5.15 (1)	5.15	5.03 (1)	5.03
1aoy	4.38	**3.30**	4.05 (1)	4.05	*3.49* (3)	4.47	*3.57* (4)	4.50
1coo	3.68	**2.49**	3.34 (1)	3.34	*2.73* (3)	3.06	3.59 (3)	3.75
1lea	2.97	*2.21*	2.78 (2)	3.38	2.49 (3)	2.81	*2.43* (2)	2.82
1sap	3.26	*2.60*	4.38 (5)	4.69	3.40 (1)	3.40	3.58 (2)	3.64
1sro	2.86	*2.25*	2.90 (5)	2.94	3.01 (1)	3.01	2.84 (1)	2.84
1stu	5.22	**3.79**	4.89 (5)	6.64	5.41 (5)	5.54	**4.08** (2)	**4.21**
1tuc	2.67	2.25	3.59 (5)	3.60	2.97 (2)	4.04	3.31 (4)	3.55
1vcc	5.48	**3.86**	5.17 (2)	6.22	5.22 (1)	5.22	**4.31** (4)	**4.37**
1vif	5.06	**3.76**	5.65 (5)	5.87	4.67 (2)	4.97	5.58 (5)	5.66
2bby	5.21	**4.20**	*4.51* (5)	4.92	*4.52* (5)	5.04	*4.26* (5)	*4.59*
2ezh	2.87	*2.24*	3.28 (5)	4.40	2.85 (5)	3.06	2.96 (2)	3.06
2fmr	3.95	**2.50**	**2.90** (1)	**2.90**	**2.85** (1)	**2.85**	**2.92** (4)	*3.17*
Average	4.04	2.90	3.90	4.39	3.62	3.90	3.58	3.79

*RMSD, the root mean square positional deviation of the backbone of residues involved in elements of secondary structure with respect to the experimentally determined structure. **Protein**, the identification code of the target protein; **$RMSD_{init}$**, the RMSD of the initial structure; **$RMSD_{min}$**, the minimum RMSD found during the REMD simulations; **Lowest RMSD**, the RMSD and the corresponding rank (in parenthesis) of the best model amongst the 5 top scoring models; **Best Score**, the RMSD of the model with the best score. RMSD values that are more than 1.0 Å and 0.5 Å lower than the corresponding $RMSD_{init}$ values are indicated bold and italic, respectively.

structure than the initial model (refined conformations) varied remarkably between the different proteins. In four cases, the majority of the conformations sampled were closer to the experimental structure than the initial model. Of these, two (PDBID 1a1z and 2bby) are all-α and the other two (PDBID 1afi and 1vcc) are α+β. In four cases, the refined conformations represented less than 10% of the conformations sampled. Of these two (PDBID 1ail and 2ezh) are again all-α class and two (PDBID 1sap and 1tuc) all-β class. There was no obvious correlation between the $RMSD_{init}$ and the percentage of closer configurations sampled. On average, 31.0% of the conformations sampled during the CH-REMD simulations were closer to the experimental structures than the original model. This percentage was slightly lower (27.3%) if only those structures sampled when λ=1 were considered (Table II).

MD simulation

The sampling efficiency of CH-REMD was compared to performing multiple independent MD simulations. In this case, each initial model was simulated 12 times for 20 ns, so that the total time simulated was the same as that using CH-REMD. Again conformations closer to the corresponding experimental structure than the initial model were sampled in all cases. However, in this case, the $RMSD_{min}$ was more than 0.5 Å lower than the $RMSD_{init}$ in 12 cases, and the $RMSD_{min}$ was more than 1.0 Å lower than the $RMSD_{init}$ in just five cases (Table II). On average, the $RMSD_{min}$ was only 0.82 Å lower than the $RMSD_{init}$. In only two cases (PDBID 1afi and 2fmr) were the majority of the conformations sampled were closer to the experimental structure than the initial model. However, in six cases, less than 10% of the conformations sampled were closer to the experimental structure than the initial model. On average, 21.0% of the conformations sampled during the MD simulations were closer to the experimental structures than the original model. In 12 of the 15 cases, the percentage of improved structures found in the multiple MD simulations is less than that found in the $\lambda = 1$ replica of CH-REMD (Table II). Based on a two-way analysis of variance (ANOVA), the difference between results obtained using CH-REMD and the multiple independent MD simulations is statistically significant, both in terms of the $RMSD_{min}$ ($P = 0.002$) and the enrichment of refined conformations ($P = 0.05$).

Selection of the best model

The conformations generated in CH-REMD were scored and ranked based on their ROSETTA, RAPDF, and DFIRE scores. For each protein, the SSE-RMSD of the best model was given in Table III ("Lowest RMSD" column in Table III). Using ROSETTA, for 9, 4, and 3 of the 15 proteins, the SSE-RMSD of the best model was decreased from that of the initial model by more than 0.0, 0.5, and 1.0 Å, respectively. For RAPDF and DFIRE, the corresponding numbers were 11, 6 and 3, and 11, 8 and 5, respectively. The average improvement in SSE-RMSD with respect to the initial model was 0.14, 0.42, and 0.46 Å for ROSETTA, RAPDF, and DFIRE, respectively. Based on a two-way ANOVA, the differences between the three scoring functions were statistically significant ($P = 0.03$) and suggest that for these cases DFIRE and RAPDF may be slightly more effective than ROSETTA for the selection of near-native conformations generated using MD-based methods. The average SSE-RMSD of the best-ranked model (col-

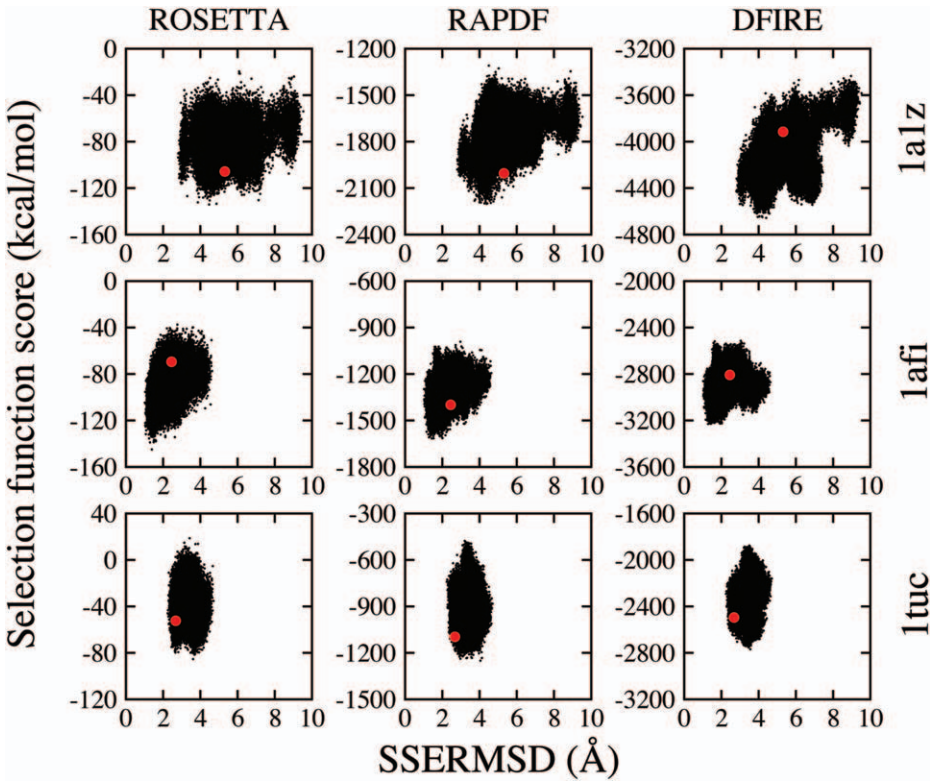

Figure 2

The correlation between three different scoring functions and the secondary structure root mean squared deviation (SSE-RMSD) from the experimental structure for three models: 1a1z, 1afi, and 1tuc. For each model, the ROSETTA, RAPDF, and DFIRE scores (*y* axis) are plotted against SSE-RMSD (*x* axis) for all model conformations generated during the REMD simulations. The red circle indicates the initial model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

umns 5, 7, and 9 in Table III) is 4.39, 3.90, and 3.79 Å for ROSETTA, RAPDF, and DFIRE, respectively, supporting this hypothesis. However, none of scoring functions tested reliably identified the model with the lowest SSE-RMSD as the best model. Furthermore, the best model was often not ranked first among the top five ranked models. Of the top five ranked models, the average rank of the best model was 3.4, 2.6, and 2.9 using ROSETTA, RAPDF, and DFIRE, respectively.

Detailed results for three representative systems

The analysis of three targets (PDBID 1a1z, 1afi, and 1tuc) is presented in more detail (Figs. 2 and 3). These three targets were chosen because they represent (a) different structural classes (e.g., the SCOP class), (b) a range of initial model quality (e.g., $RMSD_{init}$), and (c) highlight differences in the performance of the refinement protocol.

1a1z

1a1z is an NMR-derived structure of the Fas-associated death domain protein (FADD).[54] The protein contains 83 amino acids (Table I). The structure corresponds to an average over the ensemble of NMR structures generated, which was then minimized. The structure is globular in shape consisting of six antiparallel, amphipathic α-helices connected by short loops. In the initial model, the orientation of these six helices differed from the proposed experimental structure and did not form a densely packed structure. The SSE-RMSD value between the NMR structure and the initial model is 5.31 Å (Table II). With respect to the NMR structure, 73.6% of the conformations populated in the CH-REMD simulation showed a smaller SSE-RMSD value than the initial model. The lowest SSE-RMSD found in the simulation was 2.80 Å. In this conformation, the six helices form a densely packed core (Fig. 3). The relative orientation of these helices was also similar to that proposed in the NMR structure. However, this conformation was not ranked highly by any of the scoring functions used in this study. The ROSETTA, RAPDF, and DFIRE scores were plotted against the SSE-RMSD values for all the conformations sampled in the simulation (Fig. 2). As can be seen, there is little correlation between the SSE-RMSD and the ROSETTA score ($R = 0.12$) or the RAPDF score ($R = 0.19$). The correlation with the DFIRE score ($R = 0.42$) is weak. Using these three scoring functions, a number of

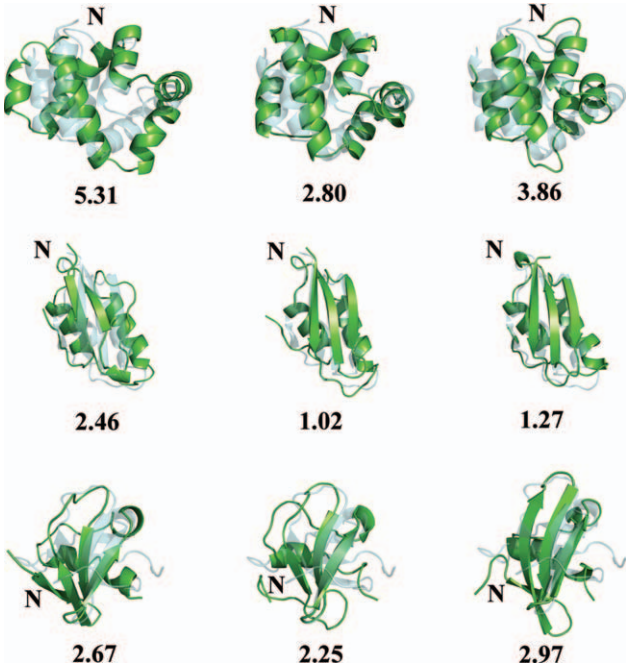

Figure 3

The superposition of model conformations (green) and the corresponding experimental structure (cyan, transparent) is shown for the proteins 1a1z, 1afi, and 1tuc, from top to bottom, respectively. For each protein, the original model conformation, the conformation with the lowest SSE-RMSD found in the REMD simulation and the best model conformation selected by three scoring functions are shown from left to right, respectively. The number below each model conformation corresponds to the SSE-RMSD of that conformation with respect to the experimental structure.

conformations with SSE-RMSD values of approximately 4.0 Å were considered optimal. Of the five best scoring structures selected using ROSETTA, RAPDF, and DFIRE, the best models had SSE-RMSD values of 3.86, 4.27, and 3.95 Å, respectively. In the structure with a SSE-RMSD value 3.86 Å, four of the six helices (α1, α4, α5, and α6) were organized in a similar manner to in the NMR structure (Fig. 3). However, the first turn of α2 was misfolded, and the orientations of both α2 and α3 are different to those in the NMR structure.

1afi

1afi is the first structure in an ensemble of 20 structures generated based on NMR-derived restraints for the reduced form of merP protein.[55] The merP protein binds Hg(II) and transfers it to the membrane transport protein merT. The protein consists of 72 amino acid residues and has a βαββαβ fold with the two α helices overlaying a four-strand antiparallel β-sheet. Overall, the initial model has the same fold as the proposed NMR structure, with a SSE-RMSD of 2.46 Å (Fig. 3). However, in the initial model, the alignment of the two helices differs from the NMR structure while the strands β3 and β4 are modeled as loops. Again the majority of conformations (60.8%) sampled during CH-REMD are closer to the proposed experimental structure than the initial model (Table II). The lowest SSE-RMSD sampled during the simulation was 1.02 Å. In this conformation, the two helices are packed closely together and lie below a region of β-sheet formed by stands β1, β2, and β3. Most notably this region of β-sheet expanded during the simulations. Only the *C*-terminal β4 strand did not adopt the proposed experimental conformation and remained as a loop. As can be seen in Figure 2, the ROSETTA and RAPDF scores are weakly correlated with the SSE-RMSD (R = 0.42 and 0.41, respectively). There is effectively no correlation with the DFIRE score (R = 0.26). Nevertheless, the SSE-RMSD values of the best model selected by ROSETTA, RAPDF, and DFIRE were 1.29, 1.28, and 1.27 Å, respectively. In all cases, the β4 strand was not incorporated into the β-sheet (Fig. 3).

1tuc

The structure 1tuc corresponds to α-Spectrin SH3[56] and was determined by X-ray crystallography at 2.02 Å resolution. The protein contains 61 residues. In the crystal structure, five antiparallel β-strands form two orthogonal β-sheets in a β-sheet sandwich. Despite a small SSE-RMSD value of 2.67 Å, the fold of the initial model differs markedly from that of the X-ray structure: (1) in the crystal structure, the first ten residues at the *N*-terminus are in a loop, whereas in the initial model, these residues form a β-sheet together with strands β1, β2, and β3 strands; (2) in the crystal structure, the β1 strand lies in the middle of a β-sheet sandwich, bridging the two orthogonal β-sheets, whereas in the proposed model, there is only one β-sheet formed by strands β1, β2, and β3 and the misfolded *N*-terminal β strand; (3) strands β4 and β5 in the crystal structure are assigned to be an α-helix and a loop respectively in the proposed model. In this case, only 5.6% of the conformations sampled during REMD lie closer to the crystal structure than the initial model. The lowest SSE-RMSD found in the simulation was 2.25 Å. In the lowest SSE-RMSD model, the length, position, and relative orientation of the strands β1, β2, and β3, is similar to that found in the crystal structure (Fig. 3). The *N*-terminal β strand and helix unfolded, but did not fold into a more appropriate conformation. Clearly, in this case, the sampling was limited, and there was no correlation between the SSE-RMSD and any of the three scoring functions (R = −0.07, 0.06, and 0.15 for ROSETTA, RAPDF, and DFIRE, respectively; Fig. 2). The SSE-RMSD values of the best of the five best scoring models using ROSETTA, RAPDF, and DFIRE were 3.59, 2.97, and 3.31 Å, respectively. All were higher than the initial model. The best of these had the same incorrect arrangement of the secondary structure as

found in the initial model. The β1, β2, and β3 strands were extended.

DISCUSSION

For most proteins of the size considered in this study, the amino acid sequence alone determines its native, three-dimensional structure. In such cases, the ideal approach to predict the structure would be to numerically simulate the process of folding under reversible conditions. However, this goal has only been truly achieved for systems containing small numbers of amino acid residues. An 11-residue β-peptide could be folded reversibly.[57] The 36-residue villin headpiece was folded to conformations that were on average 1.7 Å distance-based RMSD from the native state.[58] The folding of the 20-residue Trp-cage peptide has also been simulated in implicit solvent with conformations that are <1 Å Cα RMSD from structures in the corresponding NMR ensemble,[59] being sampled using temperature-based replica-exchange molecular dynamics (T-REMD). Recently, reversible folding of 12 fast-folding proteins that range in size from 10 to 80 amino acid residues was reported, over simulation periods ranging between 100 μs and 1 ms.[60] However, folding is not always possible even in a simulation time 10 times that of the experimental folding time, as shown for the FPB28-WW domain.[61] For this reason, approaches based on bioinformatics are still the only means that can be used to predict what is the native structure given a series of models generated by comparative modeling or de novo methods.

The two fundamental challenges in the high-resolution structure refinement are therefore to (a) efficiently sample near-native conformations and (b) identify the native conformation from this ensemble of states. In a previous study, we showed that the sampling of near-native conformations during MD simulations could be enhanced by oscillating environment to mimic some of the effects of molecular chaperones.[34] However, the utility of this approach was limited by the difficulty of obtaining optimized parameters as different proteins showed different sensitivity to the variation in the environment. To circumvent this problem instead of oscillating the environment, the approach was implemented as part of a Hamiltonian replica-exchange molecular dynamics (CH-REMD) scheme, in which the electrostatic interactions between the protein and the water molecules was varied across the different replicas. In this way, a protein could experience cycles of unfolding, refolding, and release by exchange between the replicas. One advantage of CH-REMD over the previous chaperone scheme is that it avoids the arbitrarily imposition of the length of time the molecule is exposed to different environments. This advantage is evident in the case of the FADD death-effector domain (PDB 1a1z): a 20 ns of CH-REMD simulation was much more effective than a single 50 ns MD simulation consisting of five chaperone cycles.[34] The multiple replicas also allow the protein to be exposed to different magnitudes of the perturbation.

Comparing the sampling efficiency of CH-REMD to that of multiple classical MD simulations, it is clear that the performance of CH-REMD is significantly better with respect to both the range of configurations sampled as indicated by the SSE-RMSD values and the percentage of improved conformations (Table II). On average the best conformations sampled were 1.14 Å closer to the experimental structure than the initial model. For 9 of the 15 targets, CH-REMD led to near-native conformations ($RMSD_{min}$ < 3.0 Å) being sampled. Note, the initial models for the remaining six targets were all relatively far from their native states (five targets have >5 Å $RMSD_{init}$ and one target has >4 Å $RMSD_{init}$). Nevertheless, for these models, CH-REMD led to conformations being sampled that were closer to the native structure by >1.0 Å (Table II). This suggests that CH-REMD could be used to refine de novo models of small proteins, which are often within 2–6 Å of the corresponding experimental structure.[7,62,63] It is also expected that the sampling efficiency of the current CH-REMD approach could be improved further, for instance by incorporating accelerated MD methods into individual replicas.[64–66] We also note that the latest version of the GROMOS force field GROMOS 54A7[67] shows improved structural properties for proteins. Alternatively, force fields such as AMBER99SB-ILDN[68] might be used.

The most challenging aspect in structure refinement, however, remains the identification of the "native" configuration. In this study, three scoring functions, ROSETTA, RAPDF, and DFIRE, were used for model selection. As noted by one of the referees, alternative versions of the DFIRE scoring function have been published which have been claimed to perform better than the original DFIRE potential under certain conditions.[69,70] The most successful of these, the dipolar DFIRE (dDFIRE), was therefore also tested. It was found that for the purpose of model selection, the performance of the original DFIRE scoring function was essentially identical to that of the alternative dDFIRE scoring function. A comparison of the results obtained using DFIRE and dDFIRE are provided as Supplementary Information Table S1. Overall, DFIRE and RAPDF were more effective than ROSETTA at identifying near-native conformations as measured by the SSE-RMSD with respect to the experimental structure (Table III). Furthermore, in the case of the nine targets for which a high proportion of near-native conformations were sampled, the average improvement in the SSE-RMSD with respect to the experimental structure was 0.19, 0.46, and 0.35 Å for ROSETTA, RAPDF, and DFIRE, respectively. For the remaining six targets, the average improvement in SSE-RMSD was 0.07, 0.35, and 0.62 Å for ROSETTA, RAPDF, and DFIRE, respectively. Note, when conformations very close to the

native structure (SSE-RMSD $\sim$ 1 Å) were sampled (1afi), good scores were obtained with all three scoring functions. This might indicate that if the conformational sampling could be improved further, the current scoring functions might perform better.

In summary, we have shown that CH-REMD, which attempts to mimic some of the effects of folding chaperones by allowing proteins to move between environments that favor the formation or disruption of secondary structure, is a simple but effective method to improve conformational sampling as part of structure refinement protocols. Furthermore, we have shown that although it is possible to sample near native conformations using techniques such as CH-REMD, scoring functions remain the major limiting factor in protein structure refinement.

ACKNOWLEDGMENTS

The authors thank Dr. Jiang Zhu and Dr. Min-Yi Shen for valuable suggestions in regard to the use of the statistical potentials.

REFERENCES

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
2. Bairoch A, Bougueleret L, Altairac S, Amendolia V, Auchincloss A, Puy GA, Axelsen K, Baratin D, Blatter MC, Boeckmann B, Bollondi L, Boutet E, Quintaje SB, Breuza L, Bridge A, Saux VBL, deCastro E, Ciampina L, Coral D, Coudert E, Cusin I, David F, Delbard G, Dornevil D, Duek-Roggli P, Duvaud S, Estreicher A, Famiglietti L, Farriol-Mathis N, Ferro S, Feuermann M, Gasteiger E, Gateau A, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hulo N, Innocenti A, James J, Jain E, Jimenez S, Jungo F, Junker V, Keller G, Lachaize C, Lane-Guermonprez L, Langendijk-Genevaux P, Lara V, Le Mercier P, Lieberherr D, Lima TD, Mangold V, Martin X, Michoud K, Moinat M, Morgat A, Nicolas M, Paesano S, Pedruzzi I, Perret D, Phan I, Pilbout S, Pillet V, Poux S, Pozzato M, Redaschi N, Reynaud S, Rivoire C, Roechert B, Sapsezian C, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey AL, Vitorello C, Yip L, Zuletta LF, Apweiler R, Alam-Faruque Y, Barrell D, Bower L, Browne P, Chan WM, Daugherty L, Donate ES, Eberhardt R, Fedotov A, Foulger R, Frigerio G, Garavelli J, Golin R, Horne A, Jacobsen J, Kleen M, Kersey P, Laiho K, Legge D, Magrane M, Martin MJ, Monteiro P, O'Donovan C, Orchard S, O'Rourke J, Patient S, Pruess M, Sitnov A, Whitefield E, Wieser D, Lin Q, Rynbeek M, di Martino G, Donnelly M, van Rensburg P, Wu C, Arighi C, Arminski L, Barker W, Chen YX, Crooks D, Hu ZZ, Hua HK, Huang HZ, Kahsay R, Mazumder R, McGarvey P, Natale D, Nikolskaya AN, Petrova N, Suzek B, Vasudevan S, Vinayaka CR, Yeh LS, Zhang J, Consortium U. The universal protein resource (UniProt). Nucleic Acids Res 2008;36:D190–D195.
3. Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001;294:93–96.
4. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Ann Rev Biophy Biomol Struct 2000;29:291–325.
5. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
6. Bonneau R, Baker D. Ab initio protein structure prediction: Progress and prospects. Ann Rev Biophys Biomol Struct 2001;30:173–189.
7. Baker D, Bradley P, Misura KMS. Toward high-resolution de novo structure prediction for small proteins. Science 2005;309:1868–1871.
8. Moult J, Fidelis K, Kryshtafovych A, Rost B, Tramontano A. Critical assessment of methods of protein structure prediction-Round VIII. Proteins 2009;77:1–4.
9. Lee J, Lee J, Lee J, Sasaki TN, Sasai M, Seok C. De novo protein structure prediction by dynamic fragment assembly and conformational space annealing. Proteins 2011;79:2403–2417.
10. Zhang Y. I-TASSER: fully automated protein structure prediction in CASP8. Proteins 2009;77:100–113.
11. Baker D, Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei JM, Kim D, Kellogg E, DiMaio F, Lange O, Kinch L, Sheffler W, Kim BH, Das R, Grishin NV. Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins 2009;77:89–99.
12. Levy Y, Ben-David M, Noivirt-Brik O, Paz A, Prilusky J, Sussman JL. Assessment of CASP8 structure predictions for template free targets. Proteins 2009;77:50–65.
13. Xu D, Zhang JF, Wang QG, Barz BD, He ZQ, Kosztin I, Shang Y. MUFOLD: a new solution for protein 3D structure prediction. Proteins 2010;78:1137–1152.
14. Skolnick J, Zhang Y. Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci USA 2004;101:7594–7599.
15. Moult J, Venclovas C, Zemla A, Fidelis K. Comparison of performance in successive CASP experiments. Proteins 2001:163–170.
16. Fan H, Mark AE. Refinement of homology-based protein structures by molecular dynamics simulation techniques. Protein Sci 2004;13:211–220.
17. Swendsen RH, Wang JS. Replica monte-carlo simulation of spinglasses. Phys Rev Lett 1986;57:2607–2609.
18. Okamoto Y, Sugita Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999;314:141–151.
19. Mitsutake A, Sugita Y, Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. Biopolymers 2001;60:96–123.
20. Geyer CJ, Thompson EA. Annealing Markov-chain Monte-Carlo with applications to ancestral inference. J Am Stat Assoc 1995;90:909–920.
21. Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc Jpn 1996;65:1604–1608.
22. Brooks CL, Chen JH. Can molecular dynamics simulations provide high-resolution refinement of protein structure? Proteins 2007;67:922–930.
23. Zhu J, Fan H, Periole X, Honig B, Mark AE. Refining homology models by combining replica-exchange molecular dynamics and statistical potentials. Proteins 2008;72:1171–1188.
24. Hartl FU. Molecular chaperones in cellular protein folding. Nature 1996;381:571–580.
25. Feldman DE, Frydman J. Protein folding in vivo: the importance of molecular chaperones. Curr Opin Struc Biol 2000;10:26–33.
26. Walter S, Buchner J. Molecular chaperones—cellular machines for protein folding. Angew Chem Int Edit 2002;41:1098–1113.
27. Chan HS, Dill KA. A simple model of chaperonin-mediated protein folding. Proteins 1996;24:345–351.
28. Sfatos CD, Gutin AM, Abkevich VI, Shakhnovich EI. Simulations of chaperone-assisted folding. Biochemistry 1996;35:334–339.
29. Thirumalai D, Betancourt MR. Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. J Mol Biol 1999;287:627–644.
30. Klimov DK, Newfield D, Thirumalai D. Simulations of beta-hairpin folding confined to spherical pores using distributed computing. Proc Natl Acad Sci USA 2002;99:8019–8024.
31. Baumketner A, Jewett A, Shea JE. Effects of confinement in chaperonin-assisted protein folding: rate enhancement through smoothing of the folding energy landscape. Abstr Pap Am Chem S 2003;226:U424–U424.
32. Shea JE, Jewett AI, Baumketner A. Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: creation of an

alternate fast folding pathway. Proc Natl Acad Sci USA 2004;101:13192–13197.
33. Takagi F, Koga N, Takada S. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. Proc Natl Acad Sci USA 2003;100:11367–11372.
34. Fan H, Mark AE. Mimicking the action of folding chaperones in molecular dynamics simulations: application to the refinement of homology-based protein structures. Protein Sci 2004;13:992–999.
35. Baker D, Simons KT, Bonneau R, Ruczinski I. Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins 1999:171–176.
36. Berne BJ, Liu P, Huang XH, Zhou RH. Hydrophobic aided replica exchange: an efficient algorithm for protein folding in explicit solvent. J Phys Chem B 2006;110:19018–19022.
37. Zhou HY, Zhou YQ. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci 2002;11:2714–2726.
38. Moult J, Samudrala R. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol 1998;275:895–916.
39. Baker D, Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL. Design of a novel globular protein fold with atomic-level accuracy. Science 2003;302:1364–1368.
40. Murzin AG, Brenner SE, Hubbard T, Chothia C. Scop—a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
41. Berendsen HJC, Vanderspoel D, Vandrunen R. Gromacs—a message-passing parallel molecular-dynamics implementation. Comput Phys Commun 1995;91:43–56.
42. van der Spoel D, Lindahl E, Hess B. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 2001;7:306–317.
43. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: the GROMOS96 manual and user guide. Zürich, Switzerland: Vdf Hochschulverlag AG an der ETH Zürich; 1996. pp 1–1042.
44. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular Forces. Dordrecht: Reidel; 1981. pp 331–342.
45. Tironi IG, Sperb R, Smith PE, Vangunsteren WF. A generalized reaction field method for molecular-dynamics simulations. J Chem Phys 1995;102:5451–5459.
46. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–1472.
47. Miyamoto S, Kollman PA. Settle—an analytical version of the shake and rattle algorithm for rigid water models. J Comput Chem 1992;13:952–962.
48. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
49. Takada S, Fukunishi H, Watanabe O. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. J Chem Phys 2002;116:9058–9067.
50. Hansmann UHE, Aleksenko V, Kwak W. Generalized-ensemble simulations of all-atom protein models. Physica A 2005;350:28–37.
51. Kwak W, Hansmann UHE. Efficient sampling of protein structures by model hopping. Phys Rev Lett 2005;95:138102.
52. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using rosetta. Method Enzymol 2004;383:66–93.
53. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
54. Fesik SW, Eberstadt M, Huang BH, Chen ZH, Meadows RP, Ng SC, Zheng LX, Lenardo MJ. NMR structure and mutagenesis of the FADD (Mort1) death-effector domain. Nature 1998;392:941–945.
55. Steele RA, Opella SJ. Structures of the reduced and mercury-bound forms of MerP, the periplasmic protein from the bacterial mercury detoxification system. Biochemistry 1997;36:6885–6895.
56. Viguera AR, Blanco FJ, Serrano L. The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. J Mol Biol 1995;247:670–681.
57. Periole X, Mark AE. Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. J Chem Phys 2007;126:014903.
58. Zagrovic B, Snow CD, Shirts MR, Pande VS. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. J Mol Biol 2002;324:1051–1051.
59. Pitera JW, Swope W. Understanding folding and design: replica-exchange simulations of "Trp-cage" fly miniproteins. Proc Natl Acad Sci USA 2003;100:7587–7592.
60. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science 2011;334:517–520.
61. Periole X, Allen LR, Tamiola K, Mark AE, Paci E. Probing the free energy landscape of the FBP28 WW domain using multiple techniques. J Comput Chem 2009;30:1059–1068.
62. Skolnick JR, Zhang Y, Arakaki AK. TASSER: an automated method for the prediction of protein tertiary structures in CASP6. Proteins 2005;61:91–98.
63. Dill KA, Ozkan SB, Weikl TR, Chodera JD, Voelz VA. The protein folding problem: when will it be solved? Curr Opin Struc Biol 2007;17:342–346.
64. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys 2004;120:11919–11929.
65. Grubmuller H, Lange OE, Schafer LV. Flooding in GROMACS: accelerated barrier crossings in molecular dynamics. J Comput Chem 2006;27:1693–1702.
66. Yang W, Zheng LQ. Essential energy space random walks to accelerate molecular dynamics simulations: convergence improvements via an adaptive-length self-healing strategy. J Chem Phys 2008;129:014105.
67. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J Biophy 2011;40:843–856.
68. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010;78:1950–1958.
69. Yang YD, Zhou YQ. Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. Protein Sci 2008;17:1212–1219.
70. Yang YD, Zhou YQ. Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins 2008;72:793–803.