

Coarse-Grained Lattice Model Simulations of Sequence-Structure Fitness of a Ribosome-Inactivating Protein

Mark A. Olson,¹ In-Chul Yeh,² Michael S. Lee^{1,2,3}

¹ Department of Cell Biology and Biochemistry, U.S. Army Medical Research Institute of Infectious Diseases, Frederick, MD 21702

² Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Frederick, MD 21702

³ Computational and Information Sciences Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD 21005

Received 1 August 2007; revised 30 October 2007; accepted 30 October 2007

Published online 5 November 2007 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.20880

ABSTRACT:

Many realistic protein-engineering design problems extend beyond the computational limits of what is considered practical when applying all-atom molecular-dynamics simulation methods. Lattice models provide computationally robust alternatives, yet most are regarded as too simplistic to accurately capture the details of complex designs. We revisit a coarse-grained lattice simulation model and demonstrate that a multiresolution modeling approach of reconstructing all-atom structures from lattice chains is of sufficient accuracy to resolve the comparability of sequence-structure modifications of the ricin A-chain (RTA) protein fold. For a modeled structure, the unfolding–folding transition temperature was calculated from the heat capacity using either the potential energy from the lattice model or the all-atom CHARMM19 force-field plus a generalized Born solvent

approximation. We found, that despite the low-resolution modeling of conformational states, the potential energy functions were capable of detecting the relative change in the thermodynamic transition temperature that distinguishes between a protein design and the native RTA fold in excellent accord with reported experimental studies of thermal denaturation. A discussion is provided of different sequences fitted to the RTA fold and a possible unfolding model. © 2007 Wiley Periodicals, Inc.[†]

Biopolymers 89: 153–159, 2008.

Keywords: lattice model; multiscale modeling; protein engineering; simulated calorimetry

This article was originally published online as an accepted preprint. The “Published Online” date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

A challenge in the application of in silico protein design is the problem of detecting the optimal sequence fitness of a protein fold. In a simplified approach, a sequence is mounted onto a three-dimensional structure and a sequence-structure compatibility profile is calculated as amino acid residues are positioned or deleted along the polypeptide chain. Certain structural regions are thermodynamically “stable,” and hence have high compatibility scores, whereas other regions are less

Correspondence to: Mark A. Olson; e-mail: molson@ncifcrf.gov
Contract grant sponsor: United States Defense Threat Reduction Agency (DTRA)
Contract grant number: 3.10010_06_RD_B
Contract grant sponsors: United States Department of Defense High Performance Computing Modernization Program Office



© 2007 Wiley Periodicals, Inc. [†]This article is a US Government work and, as such, is in the public domain in the United States of America.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE OCT 2007		2. REPORT TYPE		3. DATES COVERED 00-00-2007 to 00-00-2007	
4. TITLE AND SUBTITLE Coarse-Grained Lattice Model Simulations of Sequence-Structure Fitness of a Ribosome-Inactivating Protein				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command,Biotechnology High Performance Computing Software Applications Institute,Telemedicine and Advanced Technology Research Center,Fort Detrick,MD,21702				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Many realistic protein-engineering design problems extend beyond the computational limits of what is considered practical when applying all-atom molecular dynamics simulation methods. Lattice models provide computationally robust alternatives, yet most are regarded as too simplistic to accurately capture the details of complex designs. We revisit a coarse-grained lattice simulation model and demonstrate that a multiresolution modeling approach of reconstructing all-atom structures from lattice chains is of sufficient accuracy to resolve the comparability of sequence-structure modifications of the ricin A-chain (RTA) protein fold. For a modeled structure, the unfolding?folding transition temperature was calculated from the heat capacity using either the potential energy from the lattice model or the all-atom CHARMM19 force-field plus a generalized Born solvent approximation. We found, that despite the low-resolution modeling of conformational states, the potential energy functions were capable of detecting the relative change in the thermodynamic transition temperature that distinguishes between a protein design and the native RTA fold in excellent accord with reported experimental studies of thermal denaturation. A discussion is provided of different sequences fitted to the RTA fold and a possible unfolding model.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 7	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

favorable. While there are many computational methods that apply force fields or statistical potentials to assess comparability of sequences fitted to single protein chain structures, they suffer from the lack of ability to treat unfolding–folding transitions that truly govern the thermodynamics of protein stability.

The most direct approach to evaluate sequence fitness is to calculate calorimetric observables. Generalized-ensemble simulations^{1–5} offer efficient methods that naturally invoke the concept of energy landscape. The connection to thermodynamics from landscape theory is calculation of expectation values from the density of states. All-atom molecular-dynamics simulations provide the most rigorous sampling method to generate conformational states over a wide range of temperatures. While significant progress has been made on massively parallel all-atom, explicit solvent molecular-dynamics simulations of protein folding,^{6–9} reported studies of computing a heat capacity or other calorimetric observables have been limited to small protein systems.^{10,11} Alternatively, lattice and off-lattice simulation models^{12–14} are computationally tractable approaches, yet their general applicability remains uncertain beyond “toy problems” or well-defined folding models.

In this brief article, we examine a multiresolution modeling approach based on a coarse-grained lattice model applied to a realistic structure-based protein design. Ricin A-chain (RTA) is an *N*-glycosidase that acts as a “ribosome-inactivating protein” (RIP) to remove a specific adenine base of the 28S ribosomal RNA (reviewed in Ref. 15). An immunogenic polypeptide was recently developed for RTA as a modified RIP fold by dedifferentiation of the molecule into a nontoxic, compact single-domain scaffold for presentation of protective B-cell epitopes.¹⁶ Experimental melting temperatures were reported and indicated that the altered protein fold showed greater resistance to thermal denaturation than the native RTA structure.^{16,17} Here, we estimate the unfolding–folding transition temperature (T_f) from the heat capacity and compare our results with experiments. We show that, regardless of the low resolution of conformational states generated from the lattice model, the potential energy function of the reduced protein representation and the scoring of reconstructed all-atom structures from lattice chains using the CHARMM19 forcefield with a generalized Born (GB) implicit solvent model¹⁸ are sufficiently accurate to yield the correct relative change in T_f . We discuss these results and provide an assessment of lattice simulations to model sequence-structure fitness and unfolding of the RTA protein fold.

METHODS

Generation of chain conformations were based on Monte Carlo sampling of a cubic lattice using the MONSSTER program devel-

oped by Skonick et al. and Kolinski and Skolnick.^{19,20} MONSSTER implements the SICHO (Side Chain Only) model where each amino acid in a protein chain is represented by a single virtual particle located at the side-chain center of mass and projected onto a cubic lattice with 1.45-Å grid spacing. The SICHO model forcefield consists of short- and long-range potential energy interactions, hydrogen-bonding cooperativity, and a mean force potential that describes hydrophobic interactions, and is given by

$$U = U_{\text{short}} + U_{\text{long}} + U_{\text{H-bond}} + U_{\text{hydro}}, \quad (1)$$

where each term contains sequence-independent, sequence-dependent, and restraint components. The sequence-specific potential was derived through geometric statistics of known protein structures and accounts for short-range interactions between nearest neighbors along the polypeptide chain, as well as long-range, pairwise, soft-core repulsive interactions. Sequence-independent potential terms model protein-like conformational biases of the chain.

To enhance the exploration of the potential energy landscape, we applied replica-exchange simulations^{1–5} implemented in the Multiscale Modeling Tools for Structural Biology (MMTSB).²¹ The wild-type ricin A-chain structure (denoted as wt-RTA) was taken from the PDB using ID:1rtc and crystallographic waters were deleted. The starting structure for the RTA immunogen was built by removing the residues in the loop region 34–43 and the C-terminal region of residues 199–267. We designate this modeled structure as RTA1-33/44-198.¹⁶ The remaining two modeled structures are RTA with residues located in the C-terminal region 199–267 replaced with the hydrophobic residue Ile (the protein chain is denoted as RTA-Ile), and the second are the same residues replaced with the sequence corresponding to pokeweed antiviral protein (PAP), using a structure–structure alignment with the PDB structure 1pag. We denote our hybrid as RTA/PAP. The three built structures (RTA1-33/44-198, RTA-Ile, and RTA/PAP) were each subjected to structural refinement by applying steepest descent minimization and adopted-basis Newton-Raphson minimization for a total of 100 steps. The forcefield was set with CHARMM19 and solvent effects during minimization were treated by a protein distance-dependent dielectric screening model of $\epsilon = r$. Nonbonded interaction cutoff parameters for electrostatics and van der Waals (vdW) terms were set at a radius of 22 Å with a 2-Å potential switching function.

The grid size for the cubic lattice was set at a value of 500 lattice units in each direction. Selection of the lattice size was to ensure adequate sampling of unfolded conformations without restricting the protein radius of gyration. The number of lattice simulation cycles at each temperature was set to 200 and the number of Monte Carlo moves per cycle was 200. Culled conformations from the lattice simulations consisted per replica of 4000 structures extracted from a sampled population of 1.6×10^9 conformations. The initial 20 million conformations per replica were applied for equilibration and not used in the analysis. A total of 32 replicas were used and exponentially spaced from a reduced temperature T of 0.8 to 2.0, where T is normalized by a reference temperature such that $\beta^{-1} = k_B T$ represents the energy unit (where k_B is the Boltzmann constant). We set $T = 1$ to represent the distribution of conformations modeled by the SICHO forcefield at 298 K. All-atom structures were reconstructed from the lattice simulations by using the

MMTSB procedure developed by Feig et al.²² The lattice-based structures were refined by a combined application of steepest descent minimization and adopted-basis Newton-Raphson minimization for a total of 100 steps. To change the representation from lattice to all-atom model, we selected the CHARMM19 force-field as a stepwise increase in resolution and modeled solvent effects during minimization by setting $\epsilon = r$. Nonbonded interaction cutoff parameters were set at values stated above. The final scoring of the reconstructed structures was based on a CHARMM19/GB model.¹⁸

Thermodynamic properties were calculated by using the temperature weighted histogram analysis method (WHAM).^{23–25} Given two parameters, for example, the potential energy, U , and the radius of gyration, R_g , the density of states is given by

$$\Omega(U, R_g) = \frac{\sum_{i=1}^L N_i(U, R_g)}{\sum_{j=1}^L n_j \exp(f_j - \beta_j U)}, \quad (2)$$

where n_j is the number of data points in the j th simulation, L is the number of replica-exchange temperatures, and β_j is the inverse temperature. The function $N_i(U, R_g)$ is the histogram of U and R_g calculated from the i th simulation, and f_j is the scaled free energy obtained by solving the following equations self-consistently,

$$P_\beta(U, R_g) = \frac{\sum_{i=1}^L N_i(U, R_g) \exp(-\beta U)}{\sum_{j=1}^L n_j \exp(f_j - \beta_j U)} \quad (3)$$

and

$$\exp(-f_j) = \sum_{U, R_g} \Omega(U, R_g) \exp(-\beta U), \quad (4)$$

where $P_\beta(U, R_g)$ is the probability density at the inverse temperature β . The heat capacity at constant volume, C_v , can be determined from the probability density by the following expression²⁶

$$C_v(T) = \frac{\sigma_U^2}{k_B T^2} = \frac{\langle U^2 \rangle - \langle U \rangle^2}{k_B T^2}, \quad (5)$$

where

$$\langle U^n \rangle = \frac{\sum_{U, R_g} U^n P_\beta(U, R_g)}{\sum_{U, R_g} P_\beta(U, R_g)} \quad (6)$$

with $n = 1$ or 2 .

RESULTS AND DISCUSSION

Figure 1 shows the calculated temperature-dependent profiles of the heat capacity function for each of the RTA structural models (wt-RTA, RTA1-33/44-198, RTA/PAP, and RTA-Ile). The temperature corresponding to the maximum in the C_v function is modeled as an apparent unfolding–folding

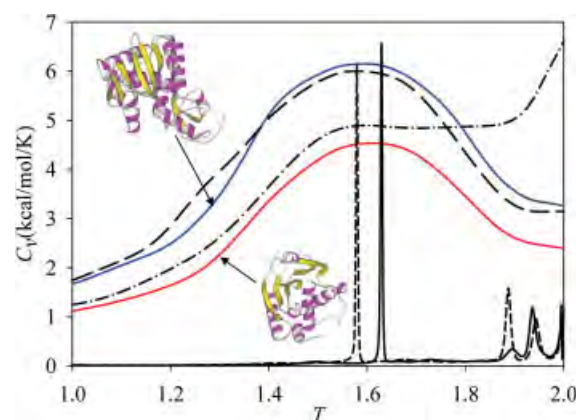


FIGURE 1 Temperature-dependent profiles of the heat capacity C_v calculated from the lattice simulations of proteins wt-RTA (blue colored line), RTA1-33/44-198 (red line), the hybrid RTA/PAP (long dashed line), and RTA-Ile (dashed-dot-dashed line). Heat capacities were computed from a WHAM analysis of an equilibrated set of 128,000 lattice chains extracted from a total sampling of over 5×10^{10} chain configurations. A comparison is provided of C_v computed by scoring the all-atom structures reconstructed from the lattice chains for wt-RTA (black solid line) and RTA1-33/44-198 (black dashed line). Superimposed on the graph are the starting all-atom structures for the native wt-RTA fold (corresponding to the blue-colored profile) and a model of RTA1-33/44-198 (corresponding to the red-colored profile), where the loop region 34–43 and the C-terminal region 199–267 were removed by protein engineering.¹⁶ The bumps in the all-atom C_v are an artifact of the high T conversion.

temperature of the protein chain. We first consider the wt-RTA structure and the polypeptide immunogen (RTA1-33/44-198). For each model, the simulation yields a broad heat capacity of unfolding–folding transitions with a fitted maximum value at T_f of 1.592 or roughly 474 K for the wild-type structure and a fitted T_f of 1.616 or 482 K for the immunogen. As described further below, the broad nature of C_v over the wide range of temperatures indicates that the model for unfolding–folding is nearly continuous and lacks significant cooperativity, leading to what can best be described as a polyamorphic transition between a highly-frustrated “glassy” state and a configurationally labile protein.^{27,28} While the SICHO model may not have realistically reproduced the high-resolution calorimetric features of the RTA molecules, the difference given by ΔT_f between the two structures is approximately 8 K and is in remarkable concurrence with the experimental determination of roughly 7 K.^{16,17} This correct prediction of a small change in protein stability needs, however, to be cautiously weighted by the ambiguity in determining T_f from a broad C_v function. Although each heat capacity has converged from Monte Carlo sampling over 5×10^{10} lattice configurations and the error in the WHAM numerical

solution is negligible, the statistical uncertainty approaches the scale of ΔT_f .

To help locate the maximum transition in the C_v function, we applied a multiresolution modeling approach and reconstructed all-atom structures from the lattice chains. After nominal energy refinement, the all-atom structures were scored using the CHARMM19/GB model. Each heat capacity was recomputed using the CHARMM19/GB energies according to Eq. (5) and the values were linearly scaled to allow for a direct comparison with the calculations using lattice energies. We note that, while the distributions were analyzed by WHAM, the Boltzmann factors containing CHARMM19/GB energies were not derived from a detailed balance. Without resorting to a complete all-atom simulation, the most straightforward approach to achieve the transformation of statistical weights from one distribution to another is to apply an umbrella reweighting technique to the probability density distribution generated by the lattice and use as corrective terms the all-atom energies of the structures that are allowed to readjust on their minimalist energy landscape. Given the molecular size of the proteins in our study and the 128,000 conformations extracted from the replicas, applying this approach and achieving numerical convergence of the population distribution profile would be a challenging task.

Figure 1 compares the temperature-dependent profiles for the all-atom models and reveals an unambiguous determination of ΔT_f , yielding approximately 15 K. As before, the calculated trend is in excellent accord with experiments. The effect of changing the statistical distributions to all-atom energies mimics a highly-ordered system (crystalline vs. glassy) with a percolation to random chains. Interestingly, the profiles now resemble those of off-lattice models with very sharp transition peaks (see, e.g., Ref. 14). In either case, the lattice-generated conformations demonstrate that the protein design of truncating primarily the C-terminal region yields a protein fold with greater resistance to thermal denaturation.

To test further the sensitivity of the SICHO potential energy function to model sequence compatibility of the C-terminal region on unfolding rather than merely a random generation of lattice chains, two additional lattice simulations were carried out. The first is a structural model of a hybrid protein RTA/PAP, where the region 199–267 is replaced with the sequence from pokeweed antiviral protein, which contains greater percentage of hydrophilic residues than RTA (30% vs. 15%). PAP is a single-chain RIP that lacks a B-chain for cell-binding induced translocation and it catalytically cleaves the same specific adenine base from 28 rRNA as does RTA. A very similar hybrid protein has been experi-

mentally tested and showed no significant differences in ribosome specificity and catalytic activity from that observed for RTA.²⁹ Although there is a predicted small decrease in the T_f of the hybrid protein compared to wt-RTA, our simulations show that the two models behaved very similarly and reflects the sequence fitness in their heating profiles rather than arbitrary chain excursions. This similarity would indicate that the sequence divergence between RTA and PAP in the C-terminal region does not markedly alter the mechanism of thermal unfolding.

As a final demonstration that the lattice simulation model can yield insight into unfolding of RTA, the final model is construction of an artificial protein by replacing the C-terminal region with all isoleucine residues. The temperature profile of the heat capacity for the RTA-Ile model displays comparable heating up to the transition point of the immunogen (truncated RTA1-33/44-198) and then becomes flat until reaching the upper bound temperature where significant unfolding occurs. By comparing the four calculated heat capacities, the simulations suggest a theoretical model where RTA unfolds initially through a “molten-globule” state consisting of a partially unfolded N-terminal domain and the C-terminal domain is unfolded.^{17,30} Because of the hydrophobic collapse of the C-terminal region against the N-terminal domain of RTA-Ile, the protein fails to dislodge the C-terminal as an unfolded state until a high temperature is reached, thus the initial heat capacity appears similar to the truncated RTA. The rate-limiting step to unfolding is the N-terminal domain and the early onset of the unfolded C-terminal domain lowers the T_f of the wt-RTA compared to the truncated molecule. This qualitative model of unfolding RTA is consistent with reported experimental denaturation studies.^{16,17,30}

Figure 2 illustrates two-dimensional (2D) contour maps of U versus R_g near the transition temperature for the lattice simulations and the reconstruction to all-atom models for the wt-RTA and RTA1-33/44-198 structures. The scale of probability density distributions is tabulated as $-\log P_\beta(U, R_g)$, where P_β is given by Eq. (8) computed near or at the T_f . We define U for the lattice simulation by Eq. (1), and for all-atom structures, U is the CHARMM19/GB total energy. For application of lattice energies, the contour maps at the T_f show no apparent separation among the structures into well-defined conformational basins that contribute to thermal unfolding of either protein molecule. While these results could have been anticipated from the broad heat capacities, the SICHO model is suitably parameterized to correctly produce a funnel-like shape in the conformational energy as the protein moves from low R_g values of folded structures to larger values representing the unfolded state. Nonetheless,

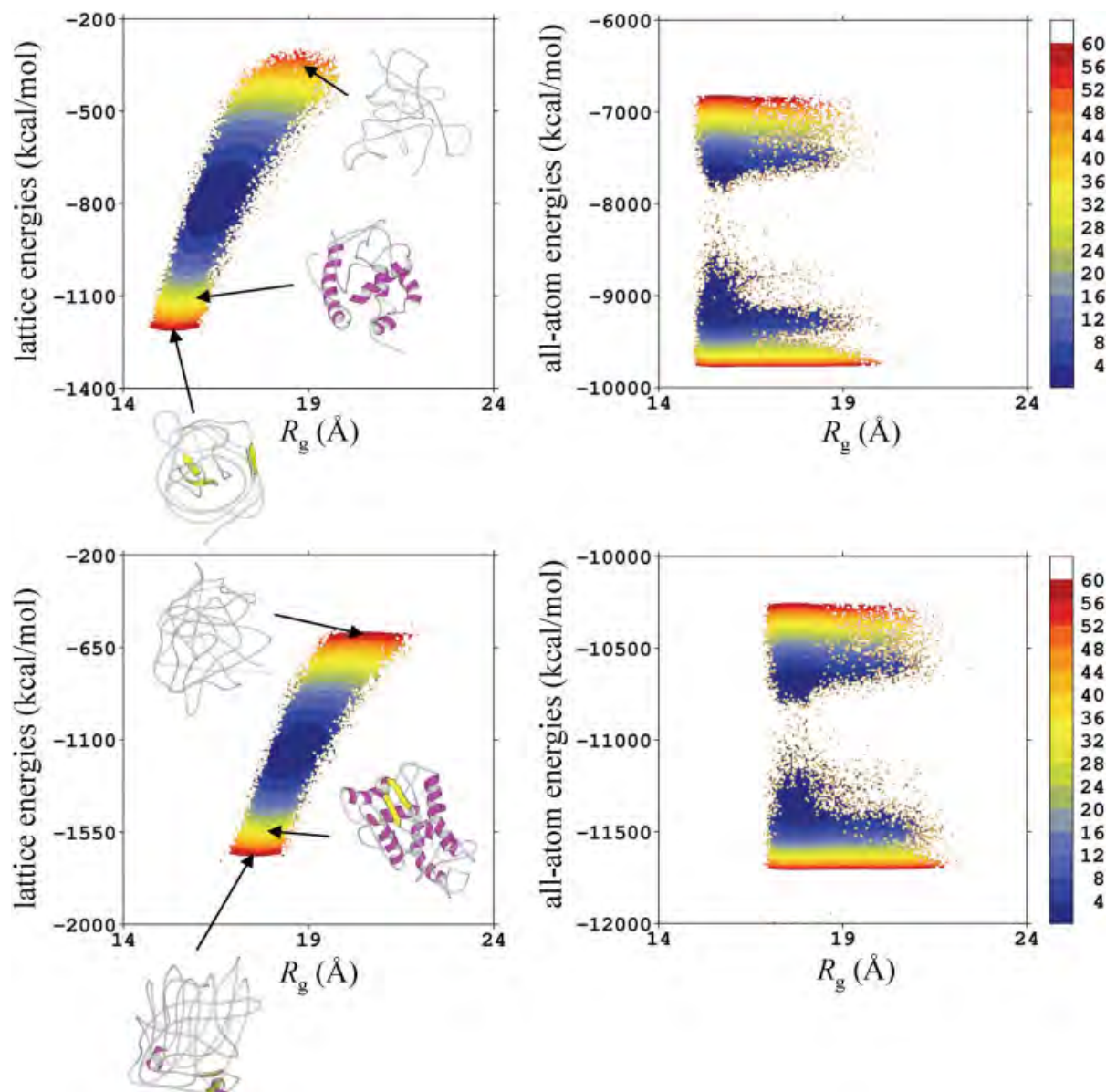


FIGURE 2 Two-dimensional probability density distributions as function of the potential energy U and the radius of gyration R_g . Profiles were computed at the observed T_f values from Figure 1 and their populations are clustered as $-\log P_\beta(U, R_g)$, where P_β is the probability density evaluated at the inverse temperature β . Shown is a comparison of lattice energies versus CHARMM19/GB energies for RTA1-33/44-198 (top two panels) and the native wt-RTA fold (bottom two panels). Superimposed structures on the plots include the best scoring lattice-energy conformations, conformations that show the lowest root-mean-square deviation from the starting structures (compare with Figure 1) and the unfolded conformations. A native-like conformation for wt-RTA displays a lattice potential energy of -1545 kcal/mol and for RTA1-33/44-198, a potential energy of -1164 kcal/mol.

because of the lack of a rigorous treatment of protein interactions and solvent effects in Eq. (1), the lowest-energy states are more compacted than native-like conformations. This outcome is a characteristic observation of statistical poten-

tials applied to structural refinement of de novo generated protein models using conformational energy sampling techniques.³¹ Typical applications of the SICH0 model is to include secondary structure as a bias for some

interactions,^{19,20,32} while our simulations lack any additional restraints entered into the force-field. Introduction of secondary structure biases would have decreased the compacting of structures, although an incorrect strength of the force constants and their lack of temperature dependence can drastically affect the modeling of T_f .

The lattice simulations with the defined grid size provided sufficient sampling of conformational space, yielding a wide range of global root-mean-square deviations (RMSD) of the backbone coordinates from the starting structures. For the unfolded conformations, the generated ensembles comprise RMSD values of approximately 23–26 Å. The lowest RMSD from the starting model (comparison of structures shown in Figures 1 and 2) for RTA1-33/44-198 is 5.33 Å and for the native RTA structure, the value is 5.05 Å. Although there is some loss of β -strands for both near-native RMSD structures, the lattice replica-exchange simulations exhibited conformational states that maintained the starting folds fairly well and preserved the structural integrity of the human B-cell epitopes.³³ However, the population density of native-like tertiary structures is low and consequently their energies make only marginal contributions to the broad transition. The transition is dominated by the coexistence of multiple states with weak structural order and is located in the highly-dense distribution of mid-range energies (dark-blue colored clusters in Figure 2).

In contrast with lattice energies, the scoring of reconstructed all-atom structures with CHARMM19/GB yields bifurcation into clusters and shows an unmistakable free-energy barrier to unfolding at the T_f . Because of the structure regularization of each lattice chain, there appears to be a less pronounced funnel-like organization among the most populated conformations (see Figure 2). The general failure to discriminate structures of similar energies with different R_g values illustrates the challenge of applying all-atom forcefields to low-resolution structures and is the outcome of a mismatch of forcing atomic resolution on structures containing steric clashes, unfavorable packing, irregular chain configurations, etc. Other than the greater disorder of the wt-RTA clusters and larger R_g values, it is difficult to extract directly from $P_\beta(U, R_g)$ the effect of the C-terminal region without having highly-refined conformational states and their structures for input into a principal component analysis. Nevertheless, the CHARMM19/GB model detects a transition in short- and long-range order plus interactions of the continuum solvent environment of the protein conformations that populate the highly-dense clusters at the T_f .

There are several alternative approaches that could conceivably yield improvement in the resolution of the folding–

unfolding free-energy profiles. One possible development is to replace the simple pairwise GB solvent model and the united-atom CHARMM19 forcefield with a modified analytical GB molecular-volume model³⁴ combined with the all-atom CHARMM22 forcefield. The latter method is known to reproduce more accurately Poisson solvation free energies of single protein chain conformations.³⁵ A more rigorous approach is a hybrid explicit/implicit solvent model to treat at the all-atom level the local hydrogen bond interactions between protein surface atoms and water that are missing in GB models, while modeling the bulk hydration by an analytical Poisson-based reaction field.³⁶ This approach, although more computationally intensive, eliminates the mean-field modeling of the protein dielectric solvent boundary based on continuum electrostatics and applies a thermodynamic integration method for determining the solvation free energy of each protein chain.³⁷ It is also possible to gain improvement in the resolution of the generated protein chains from recent reported developments in lattice-based reduced models.^{38,39} In either case, the SICHO potential energy function and other reduced model functions are generic in their representations and are similar to all-atom functions of modeling protein interactions. Generalization of the computational methodology to other model systems is straightforward; however, the applicability is restricted to modeling small to medium sized proteins and the effects of large-scale modifications, rather than the more detailed problem of detecting free-energy changes from single amino acid substitutions. Despite the limitations of our reduced model, the end result from the calculations is an apparent two-state model of unfolding–folding with the protein design exhibiting a larger free-energy barrier to thermal denaturation than the native structure.

CONCLUSIONS

In this study, we determined that a multiresolution modeling approach of pairing a coarse-grained lattice simulation model and all-atom reconstruction of structures from lattice chains is capable of modeling a realistic protein design problem. Given the size of the modeled proteins and their structural complexity, the potential energy functions are of sufficient resolution to resolve favorable changes in the sequence-structure compatibility of the RTA fold which leads to an improved thermodynamic stability. Within the framework of coarse-grained simulations, the methodology of multiscale modeling should be suitable to other examples where computing relative changes in the protein heat capacity is simply outside the practicality of applying all-atom simulations.

We thank Drs. C. B. Millard and J. H. Carra for helpful discussions. Computational time was provided by the U.S. Army Research Laboratory Major Shared Resource Center and the Advanced Biomedical Computing Center at the National Cancer Institute. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution.

REFERENCES

1. Sugita, Y.; Okamoto, Y. *Chem Phys Lett* 1999, 314, 141–151.
2. García, A. E.; Sanbonmatsu, K. Y. *Proteins* 2001, 42, 345–354.
3. Sanbonmatsu, K. Y.; García, A. E. *Proteins* 2002, 46, 225–234.
4. García, A. E.; Sanbonmatsu, K. Y. *Proc Natl Acad Sci USA* 2002, 99, 2782–2787.
5. Earl, D. J.; Deem, M. W. *Phys Chem Chem Phys* 2005, 7, 3910–3916.
6. Zagrovic, B.; Snow, C. D.; Khaliq, S.; Shirts, M. R.; Pande, V. S. *J Mol Biol* 2002, 323, 153–164.
7. Zagrovic, B.; Snow, C. D.; Shirts, M. R.; Pande, V. S. *J Mol Biol* 2002, 323, 927–937.
8. Jayachandran, G.; Vishal, V.; Pande, V. S. *J Chem Phys* 2006, 124, 164902.
9. Jayachandran, G.; Vishal, V.; García, A. E.; Pande, V. S. *J Struct Biol* 2007, 157, 491–499.
10. Pitera, J. W.; Swope, W. *Proc Natl Acad Sci USA* 2003, 100, 7587–7592.
11. Lei, H.; Wu, Ch.; Liu, H.; Duan, Y. *Proc Natl Acad Sci USA* 2007, 104, 4925–4930.
12. Skolnick, J.; Kolinski, A. *Science* 1990, 250, 1121–1125.
13. Miyazawa, S.; Jernigan, R. L. *J Mol Bio* 1996, 256, 623–644.
14. Borreguero, J. M.; Nikolay, V.; Dokholyan, N. V.; Buldyrev, S. V.; Shakhnovich, E. I.; Stanley, H. E. *J Mol Biol* 2002, 318, 863–876.
15. Marsden, C. J.; Smith, D. C.; Roberts, L. M.; Lord, J. M. *Expert Rev Vaccines* 2005, 4, 229–237.
16. Olson, M. A.; Carra, J. H.; Roxas-Duncan, V.; Wannemacher, R. W.; Smith, L. A.; Millard, C. B. *Protein Eng Des Sel* 2004, 17, 391–397.
17. McHugh, C. A.; Tammariello, R. F.; Millard, C. B.; Carra, J. H. *Protein Sci* 2004, 13, 2736–2743.
18. Dominy, B. N.; Brooks, C. L., III. *J Phys Chem B* 1999, 103, 3765–3773.
19. Skonick, J.; Kolinski, A.; Ortiz, A. R. *J Mol Biol* 1997, 265, 217–241.
20. Kolinski, A.; Skonick, J. *Proteins* 1998, 32, 475–494.
21. Feig, M.; Karanicolas, J.; Brooks, C. L., III. *J Mol Graph Model* 2004, 22, 377–395.
22. Feig, M.; Rothiewicz, P.; Kolinski, A.; Skolnick, J.; Brooks, C. L., III. *Proteins* 2000, 41, 86–97.
23. Ferrenberg, A. M.; Swendsen, R. H. *Phys Rev Lett* 1989, 63, 1195–1198.
24. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J Comput Chem* 1992, 13, 1011–1021.
25. Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J Phys Chem B* 2005, 109, 6722–6731.
26. McQuarrie, D. A. *Statistical Mechanics*. Harper and Row: New York, 1976.
27. Angell, C. A. *Proc Natl Acad Sci USA* 1995, 92, 6675–6682.
28. Nymeyer, H.; García, A.; Onuchic, J. *Proc Natl Acad Sci USA* 1998, 95, 5921–5928.
29. Chaddock, J. A.; Monzingo, A. F.; Robertus, J. D.; Lord, J. M.; Roberts, L. M. *Eur J Biochem* 1996, 235, 159–166.
30. Argent, R. H.; Parrott, A. M.; Day, P. J.; Roberts, L. M.; Stockley, P. G.; Lord, J. M.; Radford, S. E. *J Biol Chem* 2000, 275, 9263–9269.
31. Lee, M. S.; Olson, M. A. *J Chem Theory Comput* 2007, 3, 312–324.
32. Kolinski, A.; Klein, P.; Romiszowski, P.; Skolnick, J. *Biophys J* 2003, 85, 3271–3278.
33. Castelletti, D.; Fracasso, G.; Righetti, S.; Tridente, G.; Schnell, R.; Engert, A.; Colombatti, M. *Clin Exp Immunol* 2004, 136, 365–372.
34. Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L., III. *J Comput Chem* 2003, 24, 1348–1356.
35. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III. *J Comput Chem* 2003, 25, 265–284.
36. Lee, M. S.; Salsbury, F. R., Jr.; Olson, M. A. *J Comput Chem* 2004, 25, 1967–1978.
37. Lee, M. S.; Olson, M. A. *J Phys Chem B* 2005, 109, 5223–5236.
38. Kmiecik, S.; Kurcinski, M.; Rutkowska, A.; Gront, D.; Kolinski, A. *Acta Biochim Pol* 2006, 53, 131–144.
39. Kmiecik, S.; Kolinski, A. *Proc Natl Acad Sci USA* 2007, 104, 12330–12335.

Reviewing Editor: Nils Walter