

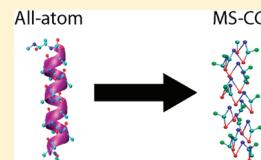
Exploration of Transferability in Multiscale Coarse-Grained Peptide Models

Ian F. Thorpe,^{†,II} David P. Goldenberg,[‡] and Gregory A. Voth^{*,†,§}

[†]Department of Chemistry and [‡]Department of Biology, University of Utah, Salt Lake City, Utah 84112, United States

[§]Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics, and Computation Institute, University of Chicago, Chicago, Illinois, 60637, United States

ABSTRACT: Coarse-grained models can facilitate the efficient simulation of complex biological systems. In earlier studies the multiscale coarse-graining (MS-CG) method was employed to examine the folding landscape for two small peptides. In those studies, MS-CG force fields specific to each peptide were employed. We extend here the scope of that work with the goal of obtaining a transferable MS-CG force field which can be used to simulate the folded conformations of peptides with disparate structural motifs. Information obtained via MS-CG modeling was used to understand the characteristics of CG interactions which govern their capacity to be transferred between different peptide systems. We find that polar CG groups are least transferable in general, with interactions between CG sites representing the CO and NH groups on the peptide backbone being particularly resistant to facile transfer. Our results additionally suggest that, while there are limitations to the approach, the MS-CG method may provide a systematic path toward obtaining rigorously defined CG interactions with at least some degree of transferability. These studies also indicate that it may be possible to enhance the transferability of the MS-CG approach by identifying novel ways to combine information from different MS-CG force fields.



1. INTRODUCTION

Coarse-grained (CG) models for computer simulation have risen in prominence in recent years.^{1–6} These methods involve generating simplified (CG) representations of fully detailed molecular systems. The promise of such methods is that they can allow the study of larger systems for longer times than conventional fully detailed simulation methods. In principle such methods can allow the length and time regimes important for mesoscopic processes to be explored via molecular simulations, illuminating the physical origin of these processes at the molecular level. CG methods have been applied to study a variety of non-biomolecular systems including ionic liquids,⁷ liquid water,^{8–10} polymers,¹¹ and electrolyte solutions.¹²

There has, however, been a particular emphasis on biomolecular systems because many important biological processes occur on length and time scales that remain inaccessible to conventional all-atom molecular dynamics (MD) simulations. Diverse approaches have been employed to carry out CG modeling of biological systems such as membranes,^{13–16} proteins and protein complexes,^{17–23} and nucleic acids,^{24–26} as well as various combinations of these^{27–34} and other^{35,36} biopolymers. The broad diversity of CG approaches makes it difficult to classify such methods. However, one class of these methods attempts to systematically construct CG models based on a molecular level description of fully detailed systems.^{1,17,32,37–45}

In generating a CG model in such a manner, one typically endeavors to ensure that the properties of the CG system are consistent with corresponding properties evaluated for the original fully detailed system.^{43–45} From this perspective, the process of coarse-graining can be framed as the search for an “energy” function which accounts for the parts of the original

system which have been coarse-grained away by incorporating these effects into CG effective interactions. Thus, the most rigorously defined (in terms of statistical mechanics) CG energy function should in principle be a many-body potential of mean force (PMF),^{43–45} in which the details of the original fully detailed system have been subsumed into the CG energy function. While simplified, such CG interactions are intended to correctly account for structural and thermodynamic properties of the original fully detailed system.

There have been numerous approaches employed to identify the effective energy functions appropriate for generating a CG model. Knowledge based potentials obtained by inverting the probability distributions found in peptide databases or generated by molecular simulations are examples of one method that is commonly used.^{46–48} As ensemble averages, such probability distributions implicitly include contributions from those parts of the system which have been removed to generate a CG representation. However, the physical principles which underlie such statistical potentials are not always transparent. Another widely employed method is to use the structural information in radial distribution functions to obtain parameters describing the desired CG energy function.^{8,10,49,50} The work of Savelyev et al. embodies another interesting approach in which concepts from renormalization group theory are used to define CG interactions for biomolecular systems.^{12,25} With regard to peptides and proteins, extensive studies have been carried out by Scheraga and colleagues using the UNRES method, which employs

Received: May 12, 2011

Revised: September 6, 2011

Published: September 11, 2011

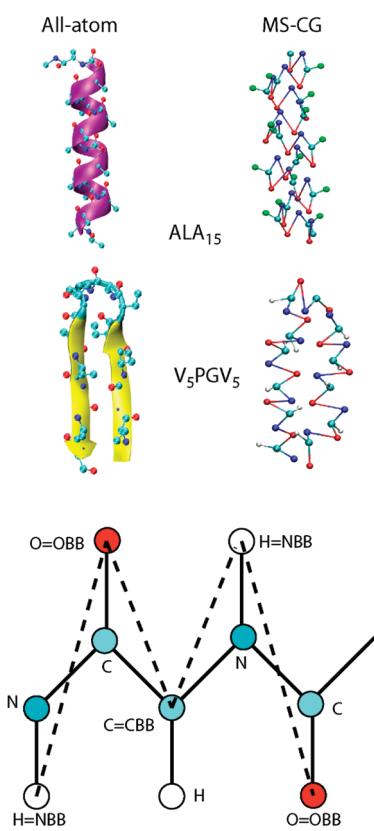


Figure 1. Depiction of peptide coarse-graining scheme. The first and second panels display all-atom (left) and MS-CG (right) representations of helical C-peptide (CPEP) and the β -hairpin GB1, respectively. The bottom panel shows a schematic view of the placement of CG sites in relation to atomic groups on the peptide backbone. Solid lines denote bonds between backbone atoms, while dashed lines denote bonds between CG sites. Side chain sites are attached to CBB and are located above and below the plane of the peptide backbone.

restricted free energy functions to describe interactions between CG groups.^{51–53}

The multiscale coarse-graining (MS-CG) approach embodies the rigorous, bottom-up approach to CG modeling.^{39,40,42–45} The method evaluates the many-body PMF underlying CG descriptions of fully detailed atomistic systems by using such simulations to systematically (i.e., variationally) parametrize CG models. In the MS-CG method the many-body PMF required to obtain a CG model is evaluated by computing the gradient of the requisite free energy terms (i.e., the coarse-grained forces) rather than by evaluating these free energy terms directly. It has been shown that the necessary CG forces are related to ensemble averaged forces which can be evaluated directly from simulations with full atomic detail.⁴⁰ This formulation of the problem makes it amenable to solution using efficient numerical techniques.^{42,45} In addition, the resulting concrete statistical mechanical link between data from fully detailed simulations and CG effective interactions allows one to have confidence that properties of the CG model possess a real physical origin. Previous studies used the MS-CG method to reproduce the equilibrium structural properties of helical Ala₁₅ and the β -hairpin V₅PGV₅³⁷ and evaluated the folding propensities of the corresponding CG models.⁴¹ In those studies, distinct MS-CG force fields were developed for each peptide. The present work seeks to understand the fundamental degree of

transferability of CG peptide models rigorously developed using the MS-CG approach.

Representative structures of the MS-CG peptide models studied in this work are shown in Figure 1, while pairwise effective interaction profiles between select CG sites (as described in section 2.3) are shown in Figure 2. As revealed by the energy functions displayed in Figure 2, the CG interactions present in each peptide are somewhat similar for distinct interaction pairs. Features common to the different interaction functions include a repulsive core and relatively gentle attractive components, such as one might expect from a collection of Lennard-Jones spheres. However, there do appear to be marked differences between certain pairs, in particular those that include polar groups. It is not clear simply by examining these curves how relevant the observed differences are to maintaining the folded structure of the peptides. To help to answer this question, it is reasonable to ask whether these interactions can be used interchangeably in CG simulations to see if the folded conformations of diverse peptides can be well described. Specifically, we sought to determine whether MS-CG models can describe the folded structures of peptides with different amino acid sequences using a single, shared set of nonbonded CG interactions.

The results of these studies revealed that it was not possible to obtain direct transferability with the MS-CG effective interactions shown in Figure 2. This result is significant because it shows that CG models developed consistently and systematically using statistical mechanics may have limited transferability. However, it was possible to obtain a CG peptide force field displaying improved transferability by employing additional assumptions and approximations. Our results suggest that MS-CG methods may thus offer a path toward obtaining reasonably transferable CG force fields for complex biological systems.

2. METHODS

2.1. Peptide Systems Studied. We investigated several short (<20 residues) peptides which are expected to demonstrate some degree of stable secondary structure. These included peptides displaying either α -helical or β -hairpin folded states.

Helical Peptides. The first peptide, Ala₁₅, was also studied in our previous work.^{37,41} Alanine is known to have one of the highest helical propensities of any amino acid,^{54,55} and alanine-based peptides have become canonical models for the structural and thermodynamic properties of the α -helix.^{48,56–60} Another helical peptide studied was Leu₁₅. While also found in helical secondary structure, Leu is thought to exhibit somewhat reduced helical propensity compared to Ala,⁵⁴ while being more hydrophobic.⁶¹ The final helix examined in this work is the 13 residue N-terminal fragment of ribonuclease A (also known as C-peptide, or CPEP). This helical fragment is one of the smallest peptides known to exhibit stable secondary structure, although it is thought to possess marginal stability.^{62,63}

Hairpin Peptides. V₅PGV₅, the β -hairpin studied in our previous work,^{37,41} is also discussed in this study. Valine is known to exhibit a propensity to form β -strands.⁶⁴ Moreover, the fused ring in the Pro residue which immediately precedes the β -turn in the peptide reduces its entropy substantially, greatly facilitating hairpin formation. This peptide has previously served as a model to examine the mechanism of β -hairpin formation.^{65,66} GB1, another β -hairpin discussed in this work, has been extensively studied using both experimental^{67–69} and computational^{70–78} approaches as a model of a canonical β -hairpin. This peptide is

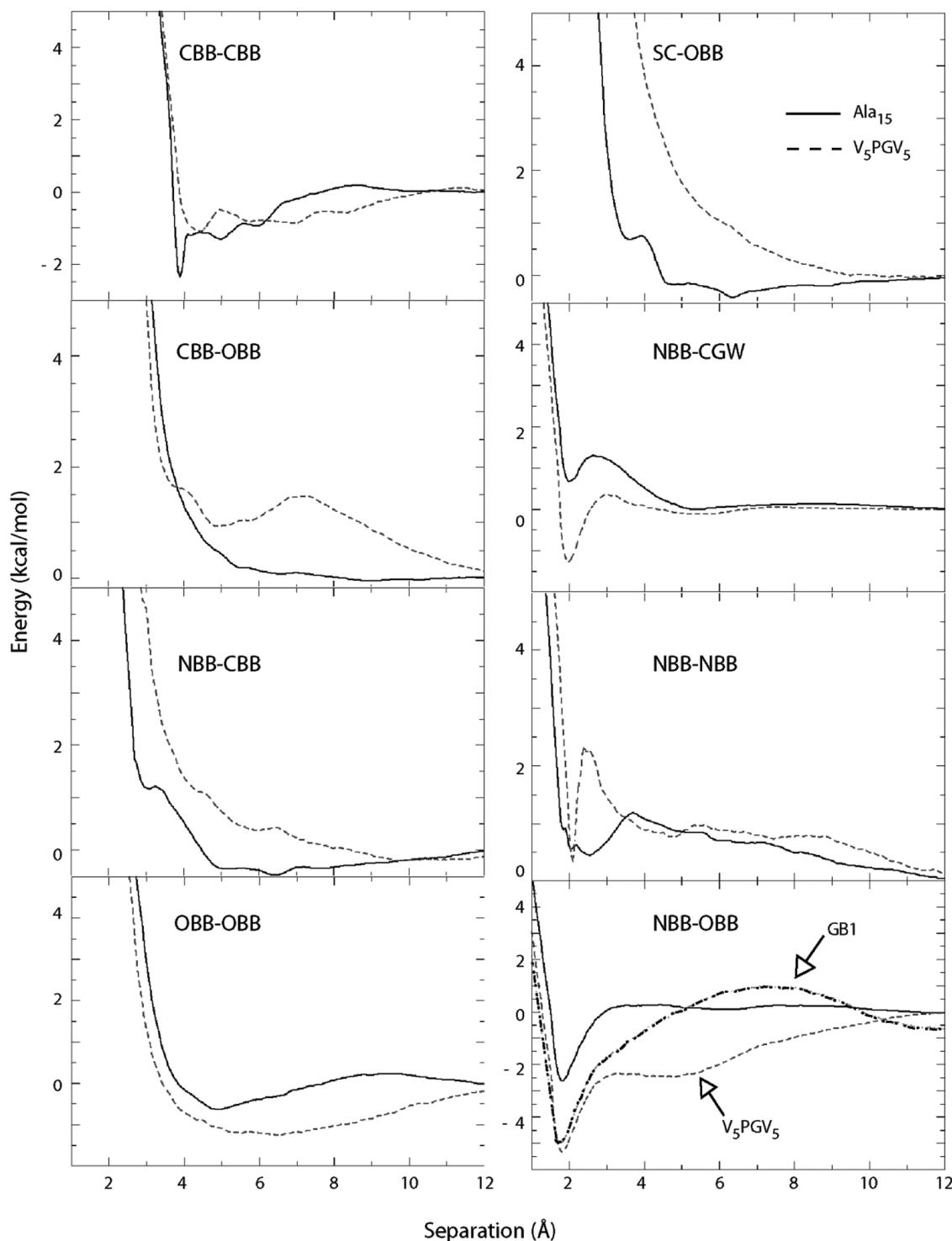


Figure 2. Nonbonded effective interaction energy between MS-CG sites in different peptide systems. The solid and dashed lines denote profiles obtained from all-atom simulations carried out for Ala_{15} and V_5PGV_5 , respectively. The dashed-dotted line in the lower right panel denotes an interaction obtained from a simulation of the GB1 peptide. The SC acronym denotes the amino acid side chain; for V_5PGV_5 this is shown for valine only. CGW denotes the coarse-grained water model. Other acronyms are as shown in Figure 1.

the C-terminal fragment of the immune protein Protein G but has also been shown to be relatively stable in isolation. GB1 has served as the basis for the design of a series of β -hairpins known as tryptophan zippers (TrpZip peptides). TrpZip peptides 1, 2, and 4 were also examined in this work and are referred to as WZ1, WZ2, and WZ4, respectively, in Table 1. These variants of GB1

were designed for increased stability relative to the wild type peptide sequence and have also been extensively studied.^{79–83} The amino acid sequences of all peptides studied in this work are provided in Table 1.

2.2. All-Atom Simulations. The MS-CG force fields used in the present study are based to a large extent on the MS-CG

Table 1. Amino Acid Sequences^a of Peptides Studied in This Work

		force field	
secondary structure	peptide designation	amino acid sequence	
			CHARMM OPLS-AA
α	Ala ₁₅	AAAAAAAAAAAAAAA	×
	Leu ₁₅	LLLLLLLLLLLLLL	×
	CPEP	KETAAAKFERQHM	×
β	V ₅ PGV ₅	V ₅ PGV ₅	×
	GB1	GEWTY <u>D</u> DATKT <u>F</u> T <u>V</u> TE	×
	WZ4	GEWTW <u>D</u> DATKT <u>W</u> TW <u>T</u> TE	×
	WZ1	SWTWE <u>G</u> GNKWTWK	×
	WZ2	SWTWE <u>N</u> NGKWTWK	×

^a Underlined positions denote residue substitutions that differentiate peptides in successive rows of the table.

peptide models of Ala₁₅ and V₅PGV₅ which were studied in our previous work.^{37,41} Only a limited amount of additional simulation data was employed to obtain new MS-CG interactions. This was done in order to replicate a situation where one intends to generate a CG model of a complex system but has limited resources with which to do so. Thus, we selected to employ information from only Ala₁₅, V₅PGV₅, and GB1 in studying the transferability of the MS-CG models.

CHARMM Force Field. As these simulations have already been presented previously, we will only discuss them briefly here. Conformations of Ala₁₅ and V₅PGV₅ were generated in full atomic detail using the CHARMM⁸⁴ molecular simulation program. For folded simulations the initial conformation of Ala₁₅ was set to be an idealized α -helix while V₅PGV₅ was generated as a β -hairpin with D-Pro and Gly forming the midpoint of the β -turn. For each system, five peptides were solvated in a cube of TIP3P⁸⁵ water with edge length of 40 Å. The energy of the entire system was reduced using 1000 steps of steepest descent minimization followed by 1000 steps of conjugate gradient minimization. The DLPOLY⁸⁶ molecular dynamics (MD) engine and CHARMM 22 all-atom force field without the CMAP correction were used to simulate the peptides at 310 K in cubic unit cells. Simulations were performed in the constant NVT ensemble using a Nosé–Hoover thermostat⁸⁷ with a relaxation time of 0.5 ps. Periodic boundary conditions were applied, and the particle mesh Ewald method was used for electrostatic interactions. A 10 Å cutoff for nonbonded interactions was employed. A 2 fs time step was used, and coordinates were written every 2 ps. After an equilibration period of 6 ps, trajectories of 4 ns were propagated for each folded peptide.

OPLS-AA Force Field. Initial conformations of Ala₁₅ and Leu₁₅ were set to be idealized α -helices. The remaining peptides were obtained from structures deposited in the Protein Data Bank (PDB; www.rcsb.org). The initial structure for CPEP was obtained by employing the 13 N-terminal residues of ribonuclease A from the PDB structure 9RAT. The initial structure for GB1 was obtained by employing the 16 C-terminal residues of Protein G from PDB structure 2GB1. Initial structures for WZ1, WZ2, and WZ4 were obtained by using the first NMR structure provided in the PDB files 1LE0, 1LE1, and 1LE3, respectively. For each system, a single peptide was solvated in cube of SPC⁸⁸ water so that there was solvent out to a distance of at least 12 Å from the nearest protein residue. The resulting system was then

subjected to 5000 steps of steepest descent minimization. The GROMACS⁸⁹ (version 3.3) MD engine and the OPLS-AA⁹⁰ force field were used to simulate each peptide at 300 K. Simulations were performed in the constant NVT ensemble using a Nosé–Hoover thermostat with a relaxation time of 0.5 ps. Periodic boundary conditions were applied, and the particle mesh Ewald method was used for electrostatic interactions. A 10 Å cutoff for nonbonded interactions was employed. A 2 fs time step was used, and coordinates were written every 0.5 ps. After an equilibration period of 50 ps, trajectories of 10 ns were propagated for each folded peptide.

2.3. Coarse-Grained Simulations. Coarse-Graining Scheme (See Figure 1). Explicit solvent is represented by a single CG site (CGW) located at the geometric center of each water molecule. Except for Pro and Gly, each amino acid residue is represented by four CG sites: three for the peptide backbone and one for each side chain. Separate CG sites are used to represent the CO, NH, and CH groups of the peptide backbone and are referred to as OBB, NBB, and CBB, respectively. CBB was placed at the α -carbon of each CH group, NBB was located at the hydrogen atom of each NH group, and OBB was located at the oxygen atom of each CO group. The backbone N atom of Pro was considered a separate group while the entire backbone of Gly was represented by a single site; these CG sites were placed at the centers of mass of the constituent atoms. CG side chains were represented by a single site located at the centers of mass of each side chain.

We note that the manner in which a given physical system is coarse-grained can have a marked effect on transferability. During the course of these studies we discovered that it was not possible to maintain folded structures for any of the peptides unless the NBB and OBB groups were situated at the hydrogen atom of the NH group and at the oxygen atom of the CO group, respectively. Thus, simply placing these CG sites at the centers of mass of the constituent groups did not preserve stable folded structures for the peptides and led to rapid unfolding. Locating these CG sites at their centers of mass destabilizes the peptide backbone, perhaps because the OBB and NBB sites are too far from the point of action of the CG force. Similar conclusions were suggested in our earliest studies where several different CG schemes were investigated.³⁷ These findings indicate it might be necessary to place CG sites at distinct locations in order to target the resulting CG force fields for specific purposes (such as improved transferability between peptide systems). Recent studies by Mullinax and Noid⁹¹ also demonstrated that the manner in which a system is coarse-grained has marked effects on the CG interactions obtained from a given collection of all atomic structures. These researchers observed that certain CG schemes lead to the generation of CG effective interactions which were more transferable and were associated with improved description of the properties of the all-atom simulations from which they were generated. Thus, this observation may be quite general in scope.

Coarse-Grained Force Field. We divide the CG effective force field interactions into bonded and nonbonded energy terms. CG bonds are defined based on chemical connectivity, with bonds linking different CG sites that contain covalently connected atoms. Bonded energy terms involve sites separated by one, two, or three CG bonds and are described by standard functional forms including harmonic terms for intersite separations or angles and cosine series for dihedral angles. The probability distribution along a given coordinate x in the all-atom simulations, $P(x)$, was employed

to obtain the energy term U_x defining fluctuations along this coordinate by fitting the following expression: $P(x) = C_x \exp[-\beta U_x]$, where C_x is an arbitrary constant. Each U_x is defined using parameters (such as the average bond length or angle) obtained directly from the all-atom trajectories. Nonbonded interactions were evaluated using the MS-CG force matching methodology. The primary feature of this approach is that it minimizes the difference between predicted CG forces and computed atomistic forces projected onto CG sites.^{39,40,43,45} The resulting effective forces are radial and pairwise decomposable, depending only on intersite separations. Detailed descriptions of the numerical methodology have been presented previously.^{42,45} For the transferable peptide potential, a single nonbonded interaction type was employed for all side chains. This was taken to be the side chain interaction computed for Ala_{15} . However, each side chain was placed at the center of mass of the constituent atomic groups. Thus, while they possess the same nonbonded interactions, the side chains in each peptide possess unique bonded interactions and positions relative to the peptide backbone. The center of mass for an individual side chain is computed from the positions of its constituent atoms during the atomistic trajectory. The force constants for side chain bonded interactions result from fluctuations of center-of-mass positions. Thus, both sequence-specific and conformation-specific effects are indirectly encoded by the bonded interactions.

At the time we carried out our previous studies of Ala_{15} and V_5PGV_5 with the CHARMM force field, we were unaware that this version of the force field induces the conversion of small α -helical peptides into π -helices, which are not often observed in proteins.⁹² This issue has now been rectified in recent versions of CHARMM which include the CMAP energy term.^{93,94} Thus, the helical peptide model of Ala_{15} generated in our last study was actually based on a π -helical version of the peptide. As a result, references to the helical structure of Ala_{15} in that work should be taken to mean the π -helix rather than the α -helix. However, all other observations reported in that work remain the same. In contrast to the previous CHARMM force field, OPLS-AA was observed to be compatible with the α -helical conformation of Ala_{15} .

A consequence of the π -helical preference of the previously employed CHARMM force field was that the MS-CG Ala_{15} model generated from those simulations was most stable in a π -helical conformation. This finding highlights one of the features of the MS-CG method: it will incorporate any errors present in the atomistic data from which effective interactions are generated. However, we discovered that it is possible to generate an Ala_{15} model that displays the correct α -helical structure from the previous π -helical MS-CG model simply by altering parameters for the energy terms which govern the backbone dihedral angles. The newly modified dihedral parameters are presented in Table 2 alongside the original parameters for comparison. All other properties of the model remain the same, including all nonbonded interactions.

Simulations of the CG peptides were initiated from CG versions of the same initial structures as the all-atom MD simulations. This allows one to determine how the free energy landscape in the vicinity of a given initial conformation differs between atomistic and CG systems. As a consequence of the coarse-graining procedure, all nonbonded (electrostatics and van der Waals) interactions from the all-atom simulations are subsumed into a short-range CG effective force field. These effective

Table 2. CG Backbone Torsional Parameters^a for Ala_{15}

dihedral	π -helical CG model			α -helical CG model		
	A	m	δ	A	m	δ
CBB–OBB–NBB–CBB	31.16	1	0.0	31.16	1	0.0
NBB–CBB–OBB–NBB	8.50	1	135.7	8.50	1	165.9
OBB–NBB–CBB–OBB	9.91	1	113.3	9.91	1	105.7

^a Dihedral energy $U_{\text{dih}}(\phi) = A[1 + \cos(m\phi - \delta)]$. Dihedral angle = ϕ , multiplicity = m , phase angle = δ (degrees), and force constant = A ($\text{kcal}\cdot\text{mol}^{-1}$).

interactions were precomputed and tabulated for use in the CG simulations. Simulations were performed under periodic boundary conditions in the constant NVT ensemble at 300 K using a Nosé–Hoover thermostat with a relaxation time of 1 ps. Simulations were carried out using the GROMACS⁹⁵ (version 4.0) MD engine. A 2 fs time step was employed for comparison with all-atom data and coordinates written out every 5 ps. After an equilibration period of 10 ns, trajectories of 100 ns were propagated for each peptide.

2.4. Obtaining a CG Force Field with Enhanced Transferability. Considering the four different peptide sites and CGW, there are $(5 \cdot 4)/2 = 10$ pairs composed of dissimilar CG sites and five pairs composed of identical CG sites, giving a total of 15 unique pairs of CG interactions present for each peptide. Coupled with the fact that interactions could be taken from either Ala_{15} or V_5PGV_5 , this leads to 2^{15} combinations of pair interactions which could, in principle, be used to construct transferable CG peptide force fields. However, attempting to investigate each of these combinations would be prohibitively time-consuming. Instead, the following course of action was employed.

Each of the 15 unique interactions between CG pairs was individually switched between Ala_{15} and V_5PGV_5 in order to perform simulations. While this was done, the remaining CG interaction pairs were maintained as they had been in the original peptide force fields. Thus, each interaction pair was placed in the context of the original collection of CG interactions obtained for the other peptide. Below, we will refer to this single CG pair interaction as the “guest” interaction while the structural and energetic environment into which it is placed will be referred to as the “host”. Considering the 2^{15} possible combinations of pair interactions, the host displays maximal dissimilarity from the environment in which the guest originated. As such, the ability of the resulting guest–host CG force field to maintain the folded structure of the peptide represents a stringent test of the ability of the guest interaction to be used in a transferable way.

The goal of performing this procedure was to quickly identify CG interactions which are incompatible with vastly dissimilar structural and energetic environments and thus induce significant structural perturbations when placed in such environments. Such interactions are less likely to be successfully used interchangeably in a variety of contexts. The utility of this approach is that it significantly reduces the number of CG interactions which must be tested for transferability by ensuring that the least transferable interactions are rapidly removed from consideration. Thus, instead of testing 2^{15} combinations, it was only necessary to test $2 \cdot 15 = 30$ combinations of interactions. While this approach does not exhaustively delineate every possible combination, it does quickly identify guest interactions which have a low likelihood of being transferable.

We note that a possible caveat to employing this approach is that it does not account for possible correlations in the CG interaction profiles. It is possible that lack of transferability of a given guest interaction could be ameliorated by the simultaneous substitution of other guest interactions which could compensate for its deficiencies. As a result, we do not claim that the set of transferable CG interactions identified in this study is the only possible combination. Nonetheless, by identifying at least one example of a transferable CG peptide force field, we hope to gain insight into the fundamental characteristics which modulate transferability.

Successful transferability in this study was defined as the ability of a CG peptide force field to maintain the structure of each peptide studied within 2 Å root-mean-square distance (RMSD) of the folded conformation for at least 20 ns. A time period of 20 ns was chosen because it is at least as long as the all-atom simulations used to validate the CG force field. Though modest, we will demonstrate that this criterion represents a significant improvement over an arbitrarily selected CG interaction set. If a matching pair of guest interactions from Ala_{15} and V_5PGV_5 was successfully deployed in both hosts, this guest interaction was classified as fully transferable. In contrast, if only one of a matching pair of guest interaction could be deployed successfully in both hosts, this behavior was classified as one-way transferability. If neither of the pair of a given guest interaction could be successfully deployed in the respective host, this was classified as a complete lack of transferability. Once a collection of transferable interactions was identified as outlined above, individual interactions were combined to produce a single CG peptide force field.

2.5. Data Analysis. An important metric in comparing MS-CG representations to all-atom peptides is the free energy landscape obtained for each. Such landscapes demonstrate the ability of the MS-CG models to reproduce complex properties of the all-atom peptides. Relative free energy plots were computed by inverting the probability distributions obtained from molecular dynamics simulations, with the free energy F given by $F = -k_B T[\ln P(x,y)]$, where k_B is the Boltzmann constant, T is temperature, and $P(x,y)$ is the normalized probability distribution occurring along variables x and y . To allow for direct comparison of the all-atom and CG data, data from all-atom simulations were coarse-grained before analysis. The landscapes were projected onto various coordinates such as RMSD from the folded conformation and radius of gyration (R_g). The former metric is particularly important in this context as it measures the degree to which peptide conformations are similar to the folded basin, which is of prime interest in this work. RMSD was computed using the peptide backbone sites only. Except as otherwise stated, the reference structure employed for the RMSD calculations was the initial minimized structure of each peptide in the folded state.

2.6. All-Atom Reconstruction. The moderate level of coarse-graining employed in this study facilitated further scrutiny of the agreement between all-atom and CG free energy landscapes by reconstructing missing atomistic detail into the MS-CG models using the available coordinates of the MS-CG sites. Based on known features of peptide geometry, several relationships between the positions of known and missing atoms can be obtained.⁴¹ These provide approximate values for the coordinates of atoms with unknown atomic positions located in the immediate vicinity of these backbone sites. Such an approach to rebuilding missing molecular details is exemplified by recent

studies by Gopal et al.³⁴ Employing such a procedure greatly simplifies the task of reconstruction by converting the process into a local optimization problem rather than one requiring global optimization of the entire peptide structure. Estimates for the locations of side chains were obtained by assigning the atom closest to the side chain center of mass to the position of the MS-CG peptide side chain site.

The entire set of known and estimated all-atom coordinates was output as a PDB file and read into the CHARMM⁸⁴ program. Positions for the missing atomic coordinates were obtained by deducing internal coordinates for the all-atom representations based on standard values present in CHARMM 22 topology files. As a result of the approximations made, the estimated sets of coordinates contain inaccuracies. In particular, the positions of side chain atoms may be unrealistic because of the simple rules used to assign values for unknown coordinates. In order to correct for errors in the estimated coordinates, the generated structure was subjected to 1000 steps of steepest descent minimization using the CHARMM 22 force field. During the initial minimization positions of the carbonyl oxygen, amino hydrogen and C_α atoms for the peptide backbone were kept fixed since the locations of these sites are known with more confidence. By keeping the positions of these atoms fixed, the conformation of the peptide backbone is made to closely correspond to that obtained in the MS-CG simulations. The constraints on the peptide backbone limit the possible positions of side chain atoms. This factor, in concert with the positions of the CG side chain site and sites in surrounding residues, provides enough information to definitively place side chain atoms. A final minimization step consisting of 500 steps of steepest descent was then employed in which all atoms are allowed to move freely to remove any remaining unfavorable contacts.

The goal of this minimization procedure is to locate the local minimum that best approximates the CG coordinates. In doing so it is not expected that the CG and all-atom structures should agree exactly. There is no reason to expect such agreement since any given CG structure represents, in principle, information from an ensemble of all-atom structures. Rather, the objective of this procedure is to obtain an all-atom conformation that is representative of the MS-CG structure so that further all-atom simulations can be carried out to characterize the free energy landscape local to the CG conformation. In this regard our approach is similar in spirit to that of Clementi and colleagues.^{2,96}

Simulations Initiated from Reconstructed Structures. Reconstructed peptides were placed in a new simulation cell of the same size as the original MS-CG system, and the cell was filled with the appropriate number of atomically detailed solvent molecules. The peptides were held fixed and the solvent was allowed to equilibrate around the structure during 50 ps of MD at 300 K using the original OPLS-AA force field. This conformation was then used to initiate additional all-atom trajectories with the same simulation conditions as the CG peptide simulations. Simulations were continued for 50 ns to provide a representation of the all-atom free energy landscape in the immediate vicinity of the original CG snapshot. With regard to data analysis, data from both reconstructed simulations and original all-atom simulations are plotted on the same axes in order to clearly show how the free energy landscape differs between all-atom simulations started from the folded conformations and those started from reconstructed all-atom peptide coordinates.

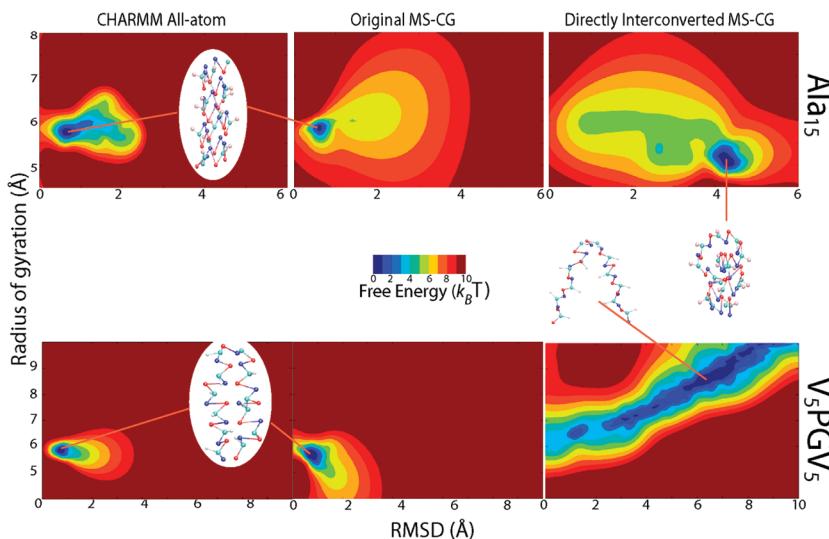


Figure 3. Comparison of original all-atom free energy landscapes (left), original MS-CG landscapes (center), and landscapes obtained when all MS-CG interactions for the peptide backbone are simply exchanged between Ala_{15} and V_5PGV_5 (right). Simulations were started from the folded conformation of each peptide and the landscapes projected onto the root-mean-square distance (RMSD; x -axis) from this conformation and the radius of gyration (R_g ; y -axis). The rightmost panel embodies a baseline for results one might expect to obtain by combining sets of disparate MS-CG interactions. The figure demonstrates that such a procedure should not be expected to be successful. Note that the data in the far left plots refer to the all-atom simulations carried out with the CHARMM force field while those in the center plots refer to the original MS-CG force field derived from this force field and used in our previous studies.^{45,46}

3. RESULTS AND DISCUSSION

3.1. Stability of the Folded Conformation Is a Measure of Transferability. As discussed in section 2, we systematically interchanged guest MS-CG nonbonded pair interactions between Ala_{15} and V_5PGV_5 host environments in order to understand the determinants of transferability in the peptide models. Simulations were then carried out starting from the folded conformation of each peptide using the host–guest CG force fields. As an initial test of the protocol, we performed an experiment in which every MS-CG effective interaction describing the peptide backbone for Ala_{15} and V_5PGV_5 was simultaneously interchanged between the two peptides. Performing this procedure within a few thousand simulation time steps disrupted the folded conformation of each peptide (see Figures 3–5). In contrast, the original MS-CG force fields maintain the folded conformations of the peptides for at least 2×10^7 time steps. The free energy minima found in the landscapes for which MS-CG interactions were interchanged are in very different locations from both the folded all-atom simulations and the original MS-CG force fields. In addition, the shapes of these landscapes differ substantially. In the case of the helix, the free energy minimum displays a compact conformation which does not appear to possess residual secondary structure. This conformation is much more compact than the original folded peptide and exhibits a lower R_g . In the case of the hairpin, the resulting free energy minimum is broad and diffuse and primarily consists of diverse conformations in which the strands of the hairpin are misaligned.

The obliteration of the folded conformations of the peptides can be observed from the representative figures inset into the free energy landscapes. It is apparent from Figures 3–5 that employing a combination of MS-CG interactions which were optimized to describe dissimilar peptide environments should not, in general, allow a CG peptide to remain folded. This result indicates that the ability to maintain a peptide's folded

conformation does represent a sensitive test of the transferability of a given CG force field.

Figures 4 and 5 reveal that modifying even a single MS-CG pair interaction can dramatically alter the peptides' free energy landscapes. The degree to which replacing these interactions alters the properties of each peptide can be assessed by comparing free energy landscapes before and after replacement. Figures 4 and 5 also provide a concrete illustration of the rationale for considering a CG force field successful if it allows the folded conformation of a peptide to be maintained within 2 Å RMSD of the native basin. It is apparent from the plots that each of the all-atom simulations satisfies this criterion. In addition, CG simulations which maintain folded conformations typically exhibit free energy minima located well within this 2 Å RMSD cutoff.

3.2. Systematic Examination of Transferability. In the assembly of a transferable CG force field from the individual MS-CG profiles for different peptides, the force fields were first tested to see whether they could maintain the folded conformation of Ala_{15} and V_5PGV_5 . Once this goal was achieved, the resulting CG force field was tested with GB1. Only when a CG force field was deployed successfully for each of these peptides was it evaluated in the context of the remaining peptides listed in Table 1. In the majority of the cases examined, the individual nonbonded interaction profiles for Ala_{15} and V_5PGV_5 could be employed interchangeably without a noticeable detrimental effect on the peptides' folded structures. This is in contrast to the observation that interchanging all of the backbone interactions between the peptides abrogated their ability to remain folded. This observation may indicate that only a few of the MS-CG interactions were responsible for mediating the unfolding demonstrated in Figures 3–5. It may also indicate that at the reduced level of detail afforded by the MS-CG models peptides are able to tolerate a range of interactions at most positions. Whenever we discovered that MS-CG interactions could be employed interchangeably, the curve from Ala_{15} was selected for

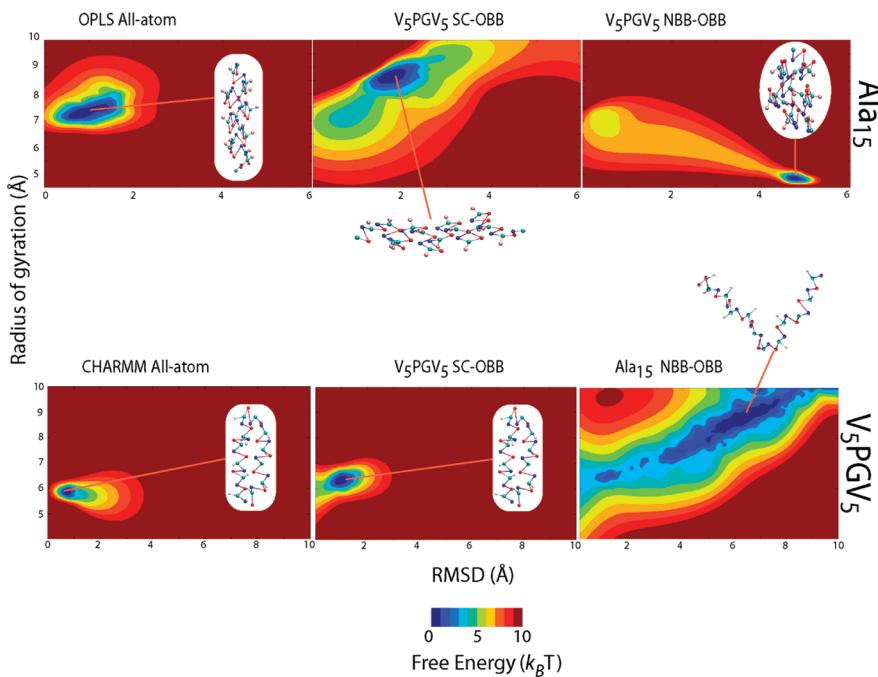


Figure 4. Free energy landscapes displaying results from various stages in the process of identifying a transferable CG force field time projected onto the RMSD (*x*-axis) and R_g (*y*-axis). From left to right, the plots contain information from (i) all-atom simulations, (ii) CG simulations in which SC–OBB from $V_5\text{PGV}_5$ was employed for both peptides, and (iii) CG simulations in which only NBB–OBB interaction was interchanged. Here (ii) is an example of the “one-way transferability” referred to in the text while (iii) illustrates the lack of transferability observed for NBB–OBB.

inclusion in the final transferable CG force field by default. Exceptions to this general observation of interchangeability are listed in Table 3 and discussed below.

Two examples of one-way transferability were identified in which the interaction profiles originally taken from $V_5\text{PGV}_5$ could be used to maintain the folded structure of Ala_{15} , but not the converse. This situation was observed for the NBB–CGW and NBB–NBB interactions. The MS-CG potential energy profiles for these two interactions are displayed in Figure 2. In the case of NBB–CGW (backbone NH group and water), the deeper minimum present in the interaction curve from $V_5\text{PGV}_5$ was found to be essential to maintain the structure of the peptide. This minimum reflects the increased presence of water near the β -hairpin backbone NH groups; these groups may assist to stabilize the backbone in its more extended state. In contrast, the NBB–CGW interaction profile generated from Ala_{15} contains a shallow minimum that induces both $V_5\text{PGV}_5$ and GB1 to collapse into compact, misfolded structures. The most important characteristic of this interaction function seems to be the depth of the minimum and thus the magnitude of the barrier one must traverse to leave this minimum: this barrier is significantly higher for the $V_5\text{PGV}_5$ interaction profile.

The NBB–NBB interaction curve (Figure 2) exhibits a minimum which is shifted slightly toward smaller separations in $V_5\text{PGV}_5$ and incorporates a significantly higher barrier to dissociation. As for NBB–CGW, using the NBB–NBB interactions generated from simulations of Ala_{15} caused both $V_5\text{PGV}_5$ and GB1 to unfold. The reason for this observation may be that the intersite separations allowed by the NBB–NBB interaction in $V_5\text{PGV}_5$ are a subset of those allowed for the Ala_{15} . Thus, Ala_{15} is able to adjust to the presence of the β -hairpin interaction by exploring a smaller fraction of the configuration space previously available to it. In contrast, the helical NBB–NBB interaction

function allowed exploration of conformations which were not typically accessible to β -hairpins and likely incompatible with other requirements of β -hairpin structure.

A final example of an interaction that is transferable in only one direction is that of SC–OBB. “SC” and “OBB” refer to the CG sites representing each amino acid side chain and backbone carbonyl groups, respectively. In this case it was possible to substitute the interaction from helical Ala_{15} into $V_5\text{PGV}_5$, but not the converse. The SC–OBB interaction curve (Figure 2) is much more repulsive for $V_5\text{PGV}_5$ than for Ala_{15} . This reflects the fact that side chains cannot approach OBB groups closely in $V_5\text{PGV}_5$ due to the secondary structure of the peptide. This deficit in density of OBB groups around SC sites is manifest in the SC–OBB interaction curve as an increased repulsion between SC and OBB for the hairpin relative to that displayed in the Ala_{15} interaction curve. In contrast, OBB groups in the helix are able to approach SC sites much more closely. Thus, if the more repulsive SC–OBB curve from $V_5\text{PGV}_5$ is employed in simulations of Ala_{15} , it induces the helix to partially unwind (see Figure 4 and 5). However, if the SC–OBB interaction from Ala_{15} is employed for either $V_5\text{PGV}_5$ or GB1, both are able to maintain stable folded conformations.

These results suggest that a prerequisite for successful interchange is that the low-energy regions of the effective energy terms in question overlap. However, this is not sufficient to guarantee transferability. The manner in which these energy terms interact with other components of the CG force field and the conformations sampled by each peptide are also relevant. By interchanging the interactions we were able to understand which features of the profiles are compatible with transferability and which may be unique to a given secondary structure. Our observations indicate that there are fundamental physical characteristics of the effective interactions that are shared by diverse

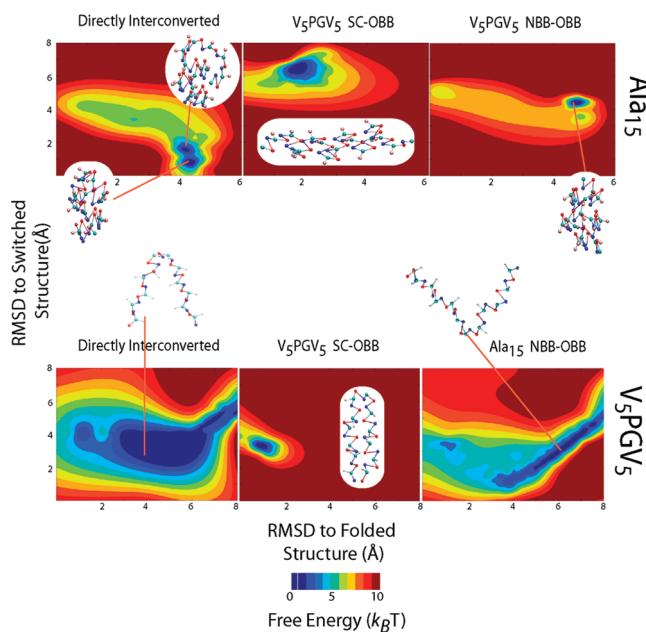


Figure 5. Free energy landscapes from various stages in the process of identifying a transferable CG force field, projected onto the RMSD from the folded conformation (*x*-axis) and the RMSD from the final structure obtained for the simulations in which only the NBB–OBB interaction was interchanged between the two peptides (*y*-axis). From left to right the panels depict data from the following: (i) CG simulations in which all backbone interactions were interchanged between the two peptide models, (ii) CG simulations in which SC–OBB from $V_5\text{PGV}_5$ was employed for both peptides, and (iii) CG simulations in which only the NBB–OBB interaction was interchanged. Here (i) is another representation of the far left panels in Figure 3 while (ii) and (iii) are alternative representations of the corresponding panels in Figure 4. These landscapes illustrate the conformational diversity which occurs when the MS-CG peptides unfold. Despite the broad distribution observed for unfolded states of $V_5\text{PGV}_5$, the representative snapshots displayed for each plot indicate that unfolded conformations for the individual peptides are structurally quite similar.

peptide systems and which can be extracted using MS-CG modeling. In general, we observe that the interaction curve in each pair that possesses the deeper minimum allows both peptides to remain folded. These deeper minima impart additional stability to the folded peptide structure and usually allow such interactions to be easily transferred between the two peptide environments. However, as we will describe below, the exception to this general observation highlights potential difficulties of generating fully transferable CG effective interactions using the MS-CG approach.

3.3. Interactions between NBB and OBB Sites on the Peptide Backbone Limit Transferability. The sole exception to the general observation of at least limited transferability is provided by the NBB–OBB interaction. The NBB and OBB sites represent the NH and CO peptide groups, respectively. NBB–OBB effective interactions obtained for the test peptides are not transferable: not even in a limited, one-directional sense. Thus, using the NBB–OBB interaction obtained from $V_5\text{PGV}_5$ simulations to describe Ala_{15} induced the resulting helix to be much too compact (see Figures 4 and 5). Similarly, employing the NBB–OBB profile obtained from GB1 simulations also leads to rapid compaction of the helix. Conversely, employing the NBB–OBB profile from Ala_{15} in simulations of either $V_5\text{PGV}_5$

Table 3. Host–Guest Simulations To Evaluate Transferability of MS-CG Interactions^a

CG site	NBB	OBB	CBB	SC
NBB	−/+			
OBB	−/−	+/+		
CBB	+/+	+/+	+/+	
SC	+/+	+/−	+/+	+/+
CGW	−/+	+/+	+/+	+/+

^a In each cell of the table the specific pair interaction is indicated by the row and column of that cell. The value preceding the slash denotes the results obtained when the interaction pair from Ala_{15} serves as the guest interaction, while the value after the slash denotes results obtained when the interaction pair from $V_5\text{PGV}_5$ serves as the guest interaction. In each case, “+” indicates that the guest interaction was able to maintain the folded structure of the peptide within 2 Å RMSD of the native basin for at least 20 ns while “−” indicates the guest interaction caused the peptide to unfold within 20 ns. Only the lower half of the symmetric matrix is shown. A column is not shown for CGW–CGW because this interaction was not altered during this procedure. SC = CG side chain.

or GB1 caused the β -hairpins to rapidly unfold. This observation suggests that the NBB–OBB curve obtained from the Ala_{15} helix is not sufficiently attractive to maintain β -hairpin folds. In support of this assertion, one can observe in Figure 2 that the energy minimum in the curve obtained from Ala_{15} is much less deep than that obtained for either $V_5\text{PGV}_5$ or GB1. As hydrogen bonding effects are subsumed into these interaction curves, these results suggest a fundamental difference between hydrogen bonding in the hairpin and helical secondary structural motifs.

However, this lack of transferability is not due merely to secondary structural differences. While the NBB–OBB curve in GB1 was found to work well for $V_5\text{PGV}_5$, using the corresponding curve from $V_5\text{PGV}_5$ to simulate GB1 disrupted that peptide’s folded conformation. This observation was the result of non-native association between several NBB and OBB sites. Similarly to the observations made for helical Ala_{15} , these non-native interactions induced GB1 to exhibit conformations which are too compact for the peptide to remain folded. This observation suggests that the NBB–OBB interaction in $V_5\text{PGV}_5$ is incompatible with the structure of other peptides in general. This would suggest that there are features of $V_5\text{PGV}_5$ which imbue unique characteristics to hydrogen bonding in this peptide.

These differences likely result from shorter, stronger hydrogen bonding present in $V_5\text{PGV}_5$ on average. This could be a consequence of the greater rigidity of $V_5\text{PGV}_5$ compared to the other peptides as well as the geometry of the peptide. The former effect would tend to reduce thermal fluctuations that decrease hydrogen bonding strength, while the latter would place hydrogen bonding partners into more optimal orientations. While we attempted to determine the primary source of the differences between the chemical environments of NBB and OBB groups in $V_5\text{PGV}_5$ and the other two peptides, our analyses were inconclusive. Nonetheless, we were ultimately able to generate a version of the NBB–OBB curve with a much improved degree of transferability.

3.4. Obtaining a Transferable NBB–OBB Interaction. While the NBB–OBB interaction was not directly transferable between peptide systems, we used the NBB–OBB curves in the different peptides as building blocks with which to construct a more transferable NBB–OBB interaction profile. The inability of the NBB–OBB curve from $V_5\text{PGV}_5$ to stabilize folded

conformations of the other two peptides strongly suggests that this peptide is not likely to be a good starting point to develop a transferable NBB–OBB interaction. Consequently, we investigated this issue using curves from GB1 and Ala₁₅. We considered the need to account for the strongly attractive nature of hydrogen bonding at short distances in β -hairpins, and yet to limit the strength of this interaction to prevent excessive compaction of the peptides or spurious association of NBB and OBB sites. Each of the curves obtained from all-atom simulations of the three peptides displays an attractive component at short distances. However, for Ala₁₅ this attraction decays to zero more quickly. This feature of the curve prevents both spurious interactions and excessive compaction of the peptide structure. Consequently, we chose to represent the NBB–OBB interaction at short distance using information from GB1 while employing the curve from Ala₁₅ to describe interactions at longer distances.

A straightforward way to accomplish this goal is to employ a linear combination of the Ala₁₅ and GB1 NBB–OBB force profiles. In this context, the final transferable CG force curve, $F'_{\text{NBB} - \text{OBB}}(r)$, may be defined as a function of the separation r between NBB and OBB sites and the individual force profiles of each peptide, $F'_{\text{NBB} - \text{OBB},i}(r)$, by

$$F'_{\text{NBB} - \text{OBB}}(r) = \frac{\sum_i^n \chi_i(r) F'_{\text{NBB} - \text{OBB},i}(r)}{\sum_i^n \chi_i(r)} \quad (1)$$

In this expression $\chi_i(r)$ denotes the relative weight of the MS-CG forces calculated from a given peptide force field in the hybrid interaction profile and the summation runs over all n CG force fields being used to generate the profile. Note that this expression was applied to the effective forces obtained for each MS-CG peptide model. Effective energy curves such as those shown in Figure 2 are obtained from the effective forces by numerical integration. In principle the approach outlined above provides a very flexible way to combine data from multiple MS-CG force fields. In the present case, $n = 2$ as only data from the Ala₁₅ and GB1 peptides was used. We employed the simple choice that

$$\begin{cases} \chi_i(r) = 1, \chi_j(r) = 0 \} & r \leq r_0 \\ \chi_i(r) = 0, \chi_j(r) = 1 \} & r > r_0 \end{cases} \quad (2)$$

where the i and j subscripts denote data from Ala₁₅ and GB1, respectively, and r_0 is a distance that was systematically varied to control at which point the hybrid curve would begin to contain force information from the respective peptides.

Sample NBB–OBB effective energy curves resulting from applying this procedure to the Ala₁₅ and GB1 force profiles are displayed in Figure 6. The performance of each of the interaction energy curves in conferring stability to folded peptides conformations was found to be quite similar, though the ultimately selected curve appeared to impart marginally better stability to folded peptides than the others. This curve was used to generate the final set of CG interactions which allowed us to represent folded states of both α - and β -peptides with a high degree of stability. Below we will examine how this more transferable CG force field affects the structure and stability of the different peptides by referring to the free energy landscapes observed for each. This final force field is referred to as the T-CG force field below.

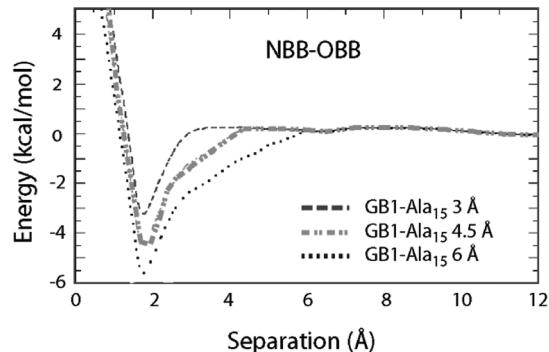


Figure 6. NBB–OBB interaction profiles obtained by combining interactions generated from Ala₁₅ and GB1 peptide simulations. In each case data from GB1 were used to generate the first part of the curve while data from Ala₁₅ were used to generate the latter part of the curve as described in section 3.4. The distance at which the curves are joined together is given in angstroms. The profile employed in the most transferable T-CG force field is denoted by the dashed and dotted line between the other two curves.

3.5. The T-CG Force Field Allows a Single Dominant Free Energy Minimum To Be Populated for Several Peptides. The results of applying the T-CG force field to describe the Ala₁₅ and V₅PGV₅ peptides are shown in Figure 7. It is apparent from Figure 7 that T-CG is able to capably represent the folded conformation of both peptides, despite their very different secondary structures. In both cases a compact basin with significant similarity to the folded state is obtained. This observation is in stark contrast to the free energy landscapes shown in Figures 4 and 5, where the CG peptide models obtained from interchanging interactions between the two peptides were largely unfolded. To determine whether T-CG is able to reproduce folded states of peptides with diverse amino acid sequences, we assessed whether the force field can allow the other peptides listed in Table 1 to remain in their respective folded conformations.

We observed that several of the remaining peptides listed in Table 1 displayed free energy landscapes with one dominant free energy minimum which remained relatively close to the folded basin for the entire length of the T-CG trajectories. These include WZ2, CPEP, and Leu₁₅. Like V₅PGV₅ and Ala₁₅, the peptide WZ2 displays a free energy landscape that is quite similar to that observed in the atomistic system. The structure of WZ2 within the T-CG free energy minimum does not differ greatly from that displayed in the all-atom folded conformation, though both RMSD and R_g are slightly larger in the T-CG model than for the all-atom simulations. The representative structures displayed for this peptide reveal that this observation is due to the fact that the two β -strands in the hairpin are not held together as closely in the T-CG model.

The T-CG representation of CPEP generates a free energy landscape that appears the least similar to the folded configuration of the three peptides shown in Figure 8, with a minimum located about 3 Å RMSD from the folded state. In addition, the T-CG peptide exhibits a much smaller R_g than the atomistic peptide. Despite these observations, hydrogen bonding analysis indicates that the T-CG force field maintains about 30% of the native hydrogen bonds in CPEP while the original all-atom peptide contains between 30 and 100% native hydrogen bonds (data not shown). Thus, the peptide still retains some residual folded structure. As noted in section 2, CPEP is thought to

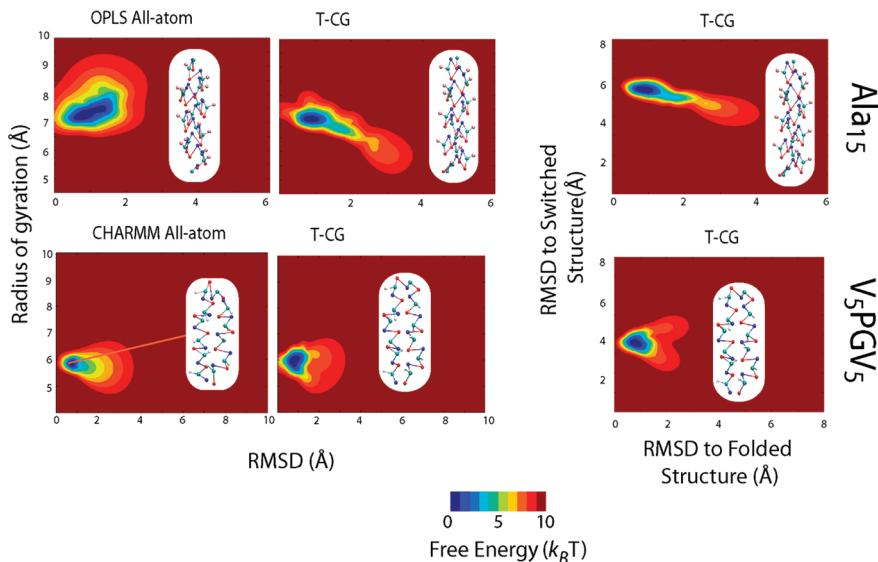


Figure 7. Free energy landscapes of Ala_{15} and V_5PGV_5 generated using the T-CG force field. The leftmost plots contain landscapes from all-atom simulations projected onto the RMSD from the folded basin (x -axis) and R_g (y -axis). The center plots contain results from T-CG simulations projected onto the same coordinates. The rightmost plots contain T-CG landscapes computed from the same simulations as the center plots but projected onto the coordinates employed in Figure 5.

exhibit marginal stability.^{62,63} This peptide was specifically chosen to test the limits of this CG modeling approach. We do not anticipate that it would be possible to reproduce secondary structures accurately for peptides smaller than CPEP using the current CG approach given that the model only preserves the minimum number of native hydrogen bonds seen in the atomistic simulation and displays markedly different RMSD and R_g values than observed for the atomistic trajectory. Peptides smaller than CPEP would possess fewer sites with which to encode their secondary structures and thus are likely to exhibit even worse agreement for these properties than is observed for CPEP.

Finally, we turn our attention to the free energy landscape of Leu_{15} in Figure 8. Both all-atom and T-CG landscapes of Leu_{15} display significant sampling around the folded basin. In addition, a significantly populated local minimum is present in both T-CG and all-atom simulations. Within this minimum, Leu_{15} develops a distinct bend in both atomistic and T-CG simulations, though this bend is more pronounced in the former. This observation was unexpected because the model was not explicitly parameterized to reproduce such behavior. Moreover, this region of the free energy landscape is quite far from the folded basin, while the T-CG representation was generated using information from the vicinity of the folded state. Many of the interactions present in Leu_{15} are the same as those employed in Ala_{15} (see section 2). Thus, we would expect the peptide to behave much like Ala_{15} and give rise to helical configurations without such a bend (see Figure 7).

The most significant difference between Leu_{15} and Ala_{15} is the size of the side chain, with the hydrophobic side chain being more distant from the peptide backbone in Leu_{15} and thus interacting more with solvent. Due to their hydrophobic nature, increased interaction of the side chain sites with the solvent exacerbates the hydrophobic effect. Thus, Leu_{15} becomes curved in an effort to minimize the solvent-exposed surface of the peptide. One might consider the presence of this bent conformation to be an emergent phenomenon arising from the combination of

MS-CG effective interactions. This result suggests that the T-CG force field may be able to reproduce certain physical properties without having been explicitly designed to do so. While further studies will be required to determine whether this holds true for other situations, it would be quite useful if this observation proves to be a general feature of T-CG models.

3.6. The T-CG Force Field Also Allows Non-native Minima To Be Explored. To varying degrees, the remaining peptides explored non-native minima in addition to the folded basin when described with the T-CG force field. The simulations typically began to leave the folded basin after about 80 ns of T-CG simulations. This result may stem from the inability of the T-CG force field to stabilize any given folded peptide conformation to the same extent as an MS-CG model specifically designed to describe this peptide. That said, the length of time the T-CG simulations remained in the folded conformation is greater than the length of the original atomically detailed simulations from which these models were generated. It may be that these non-native minima represent legitimate regions of conformational space that became newly accessible because of the ability of the T-CG models to enhance exploration of the underlying free energy landscapes.

To investigate this possibility, we reconstructed missing atomic detail into representative snapshots taken from the non-native local minima observed in the T-CG simulations (see section 2). Using all-atom reconstruction provides one measure of the degree to which the T-CG force field is consistent with the original all-atom force field. The resulting data are shown in Figure 9. Though the reconstructed simulations were carried out separately from the original all-atom simulations, data from both sets of simulations are presented in the same plot to highlight differences between the local minima obtained via atomic reconstruction and the all-atom folded basin. Data from atomistic simulations started from the folded basin are encircled by the dashed lines in the left panels of Figure 9. Thus, regions of the atomistic landscape to the right of this circle (i.e., with larger RMSD) were generated from

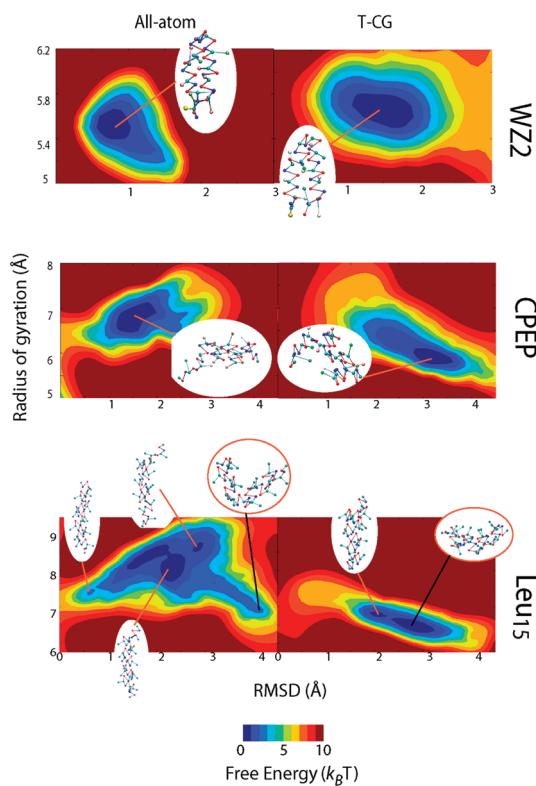


Figure 8. Free energy landscapes of peptides displaying a single dominant minimum when simulated with the T-CG force field. These landscapes have been projected onto the RMSD from the folded basin (*x*-axis) and R_g (*y*-axis). In each case the descriptions of the folded basin by the T-CG simulations are in good agreement with those obtained from the all-atom trajectories. All-atom data are shown on the left while T-CG data are displayed on the right. In the lower panel describing Leu₁₅, structures circled in red are examples of configurations outside the folded basin that are present for both T-CG and all-atom landscapes. In such structures the hydrophobic Leu side chains have curved the peptide α -helix in order to shield themselves from solvent. This feature of the peptide was not explicitly designed into the T-CG model.

simulations initiated from reconstructed all-atom structures. For comparison, the folded basin of the MS-CG representation in the leftmost panels of each plot is also encircled. The conformations employed for the reconstruction process were taken from the non-native minima present in the T-CG simulations. Positions of these minima in the landscape are denoted by an “X” in each plot.

As reflected in Figure 9, the T-CG simulation of GB1 initiated from a folded conformation spends most of its time around the folded basin centered at 3 Å RMSD and 6.5 Å R_g . The peptide displays reduced barriers between the folded basin and the rest of the landscape compared to the all-atom peptide, allowing the T-CG peptide to escape the native basin. For the T-CG simulations, one local minimum exists in a region that displays relatively high free energy in the atomistic landscape. The free energy landscape of atomistic simulations initiated from the reconstructed version of this local minimum exhibits a broad and diffuse region of relatively low free energy. The lowest point in the non-native region of the landscape is quite far from the folded basin at 7 Å RMSD and 8 Å R_g . Thus, the reconstructed atomistic system goes even further from the folded basin than the T-CG simulation.

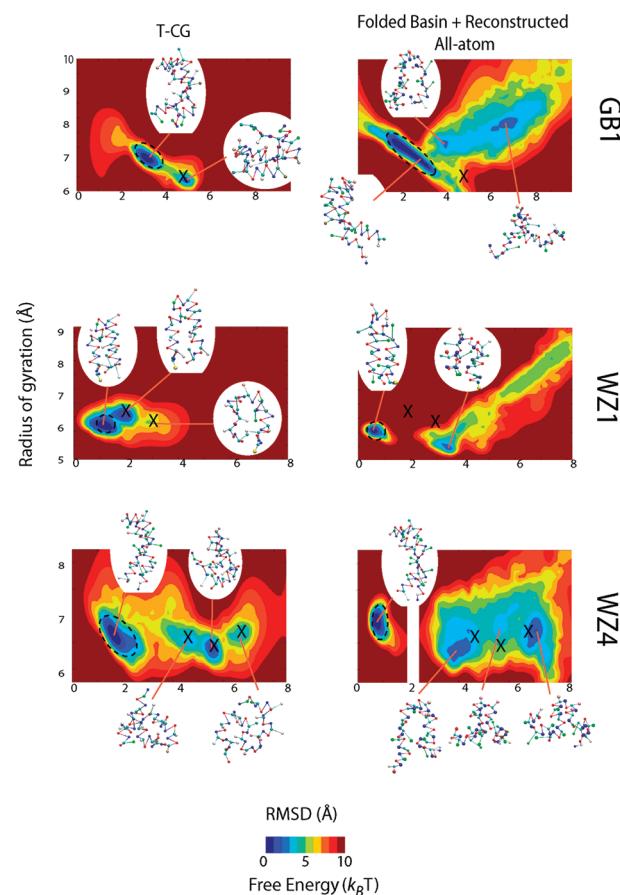


Figure 9. Free energy landscapes of peptides that display multiple minima in their T-CG representations. T-CG data are displayed on the left while all-atom data are shown on the right. The folded basin in each plot is encircled by a black dashed line. Sample conformations (inset) were collected from the T-CG non-native minima (denoted by each “X”) and used to reconstruct atomically detailed structures of the peptides. The reconstructed all-atom MD configurations were used to initiate additional simulations with the all-atom force field. Results from these simulations are shown together with data from all-atom simulations initiated in the folded basin. Thus, regions in the all-atom landscapes outside the folded basin result from simulations initiated from reconstructed configurations. Note that for WZ4 the original all-atom landscape and that from the reconstructed simulations do not overlap, as denoted by the gap of white space in the landscape. The regions of the landscape on the right side of the white space have been shifted relative to the folded basin so that the lowest free energy point is at the same level as the lowest free energy non-native minimum observed in the T-CG simulations.

While the native basin in all-atom simulations of WZ1 is fairly similar to that displayed in the T-CG representation, the remainder of the landscape differs more dramatically. T-CG simulations of WZ1 reveal two non-native minima located at 2 and 3 Å RMSD from the folded state. As is the case for GB1, the part of the landscape generated from these minima via all-atom reconstruction exhibits broad regions of relatively low free energy. Unfortunately, the most stable non-native minimum in the reconstructed atomistic simulations does not correspond to any of the minima obtained from the T-CG simulations. Also, there are no atomistic local minima which correspond to the two local minima obtained from the T-CG simulations. Moreover, there is a distinct barrier between the

native basin and the remainder of the landscape in the all-atom simulations.

The T-CG representation of WZ4 displays a folded basin that agrees well with that obtained in the atomistic simulations. However, there are three local minima in the free energy landscape which are quite far from the native basin. Three minima were also obtained in the part of the landscape originating from reconstructed atomistic conformations. However, the locations of these minima did not coincide in the two landscapes. As for GB1 and WZ1, there is a distinct barrier between the folded basin and the remainder of the free energy landscape for the all-atom data. In contrast to the previous two cases, there appears to be a barrier in the T-CG landscape of WZ4 in approximately the same location observed for the all-atom data. However, this barrier is much smaller for the T-CG representation. Because of the barrier between the folded basin and the non-native minima in the atomistic landscape, there is no sampling in this region of the landscape. Thus, the atomistic landscapes on the right (see Figure 9) are not contiguous. This situation is denoted by the white space between the two regions of the landscape. Complete lack of overlap between the left- and right-hand sides of the landscape prevents one from deducing a free energy difference between the two. For convenience, the regions of the landscape on the right side of the white space have been shifted relative to the folded basin so that the lowest free energy point is at the same level as the lowest free energy non-native minimum observed in the T-CG simulations.

Overall, the T-CG simulations of GB1, WZ1, and WZ4 reflect both commonalities and differences compared to the original all-atom landscape. They allow the folded basin to be stabilized for reasonably long times but also allow the exploration of non-native minima that often do not coincide with minima in the all-atom landscape. The observation that GB1, WZ1, and WZ4 were able to leave the folded basin during the course of the T-CG simulations indicates that barriers to unfolding were reduced in these peptides, which would tend to encourage exploration of diverse peptide conformations. This result contrasts with our previous studies, where MS-CG models developed from folded peptide conformations displayed a distinct bias toward sampling such conformations.⁴¹ This decrease in barrier height around the native basin may be a general feature of CG force fields with some degree of transferability, which by their nature should not heavily favor a particular peptide conformation. This attribute may be advantageous if one wishes to use the CG models to explore as much conformational space as possible.^{97,98} A reduced bias for any given peptide conformation is likely to be more similar to the situation present in atomistic studies of peptides.

However, many of the structures in the atomistic simulations seem to move away from the T-CG conformations. This observation might suggest that the T-CG potential exhibits decreased transferability outside of the α -helical and β -hairpin native structures. This behavior may be a consequence of the fact that the force field was explicitly developed to represent these secondary structural motifs. Thus, the free energy landscapes resulting from the T-CG force field likely do not as accurately describe regions of the landscape (such as unfolded peptide conformations) outside of these two motifs.

In summary, the T-CG force field provides a qualitative description of the all-atom peptide free energy landscapes but does not replicate them exactly. The overall shapes of T-CG and atomistic free energy landscapes are similar but the details differ; in particular, the specific locations of free energy minima located

outside the folded conformation and the heights of free energy barriers, which tend to be smaller in the T-CG simulations. This suggests that, while the T-CG representations may be quite useful in a qualitative sense, it may be challenging to directly interpret results from the models in a quantitative manner, particularly for peptide conformations which are not folded into regular secondary structures. However, the T-CG force field may still be of utility in allowing more of the free energy landscape to be explored than all-atom simulation methods by removing barriers to traversal of the landscape. If necessary, atomic detail could then be reconstructed into the CG models after sufficient sampling has been carried out in order to generate more accurate descriptions of atomistic peptide properties.^{97,98}

3.7. Combining Information from Diverse MS-CG Force Fields May Improve Transferability. Our efforts to improve transferability of CG peptide force fields suggest it may be possible to generate such force fields by manipulating MS-CG interactions in novel ways. In this study, we identified a simple yet effective approach to increase the transferability of the CG peptide force field via combining individual MS-CG interactions generated from different peptides. However, it may be necessary to use more sophisticated approaches for other systems, including employing more complex weighting functions for different components of the CG force fields. In addition to combining CG effective interactions after the fact, as done in this work, one could envision applying such an approach during the calculation of MS-CG forces which best approximate the many-body CG PMF. For example, one might wish to combine all-atom data obtained from different molecular dynamics simulations. In this case, unique weighting constants could be applied to the information present for each data set. By carefully choosing the weights given to different sets of all-atom data, it may be possible to generate CG interactions which are more transferable.

We note that Mullinax and Noid employed an approach that is similar in spirit to that described above in their recent exploration of transferability in CG models of liquid mixtures.⁹¹ These authors chose to optimize CG force fields for a large collection of data obtained from diverse molecular simulations in order to obtain transferable CG interactions. They termed this approach the use of an "extended ensemble" to solve the equations employed in the conventional MS-CG method.^{39,43} While these authors did suggest that it might be possible to adjust the weighting factors employed for different data sets in their approach, they chose to give each data set equal weight in calculating transferable CG effective interactions from the extended ensemble. Recently these authors extended their approach to systematically recover the transferable interaction potential used to generate structural ensembles in an idealized protein data bank.⁹⁹ Their work suggests that approaches based on the MS-CG method can be used to identify transferable interaction potentials if the appropriate structural ensembles are employed.

Other recent studies by Hills and Voth employed multiple atomistic ensembles, as well as a more elaborate CG representations of amino acid side chains, to help identify transferable CG interaction potentials which can be used to model realistic proteins.¹⁰⁰ The resulting CG model provided reasonable agreement with atomistic simulations for protein native state structures. While the model did have limitations, including imparting excessive stabilization to non-native states, these studies illustrate the potential value of CG force fields obtained using multiple atomistic ensembles in describing protein complexes and transitions between well-defined structural states.

As a complement to incorporating information from multiple atomistic ensembles into the MS-CG approach, it may be possible to use physical intuition to design appropriate weighting functions to combine the all-atom data employed in such a manner. Moreover, generating transferable CG interactions by differentially weighting data from fully detailed simulations allows for the possibility of incorporating experimental information into the CG models. Such information could be employed to adjust the weights between individual MS-CG energy terms or, in the context of an extended ensemble approach, between entire all-atom data sets. Analogous ideas have been suggested by Matysiak and Clementi in the context of defining CG protein models.¹⁰¹ Employing experimental data in this manner could circumvent the difficulties of performing multiple all-atom simulation at different thermodynamic state points. Along these lines, the recent work of Mullinax and Noid demonstrates in principle how the MS-CG approach could be applied to structural databases such as the PDB in order to identify transferable CG interaction potentials for proteins.⁹⁹ In this way one could gain access to much longer time and length scales than are typically accessible by even the most extensive all-atom MD simulations. Thus, incorporating experimental data into MS-CG force fields could allow the resulting CG models to include contributions from macroscopic ensembles while maintaining a link to the underlying physical properties derived from all-atom simulations. Implementing such approaches may provide an opportunity to greatly expand the scope of force fields based on the MS-CG framework.^{91,99}

4. CONCLUDING REMARKS

The present work has shed light on the strengths and limitations of physical interaction-based CG peptide modeling in general and the MS-CG approach in particular, information that can aid the future development and application of such CG models. In this context, complete transferability of physical CG models between different peptide systems may be quite difficult to achieve because the specific physical environment in which different MS-CG force fields are generated may endow them with unique features which make it difficult to apply them in other contexts. While including environmental effects is typically desired when generating a CG model, doing so can also reduce the transferability of such models. In particular, these studies provide evidence that the peptide backbone environment is distinct in helical and β -peptides and provides insight into the physical phenomena that underlie this observation.

Nonetheless, a certain degree of transferability was achieved in this study, enabling reasonably stable descriptions of the folded conformations of a variety of peptides encompassing the two dominant secondary structural motifs in proteins. The T-CG force field generated in this study seems to provide a faithful reproduction of the all-atom peptide landscape in the vicinity of the folded basin, with a less accurate description of unfolded states. Since the T-CG force field was developed by employing information from folded peptides, this observation may suggest that the MS-CG method has the potential to capture properties which are common to similar but distinct physical environments, but that this ability degrades the further one departs from the physical conditions under which the CG force field was generated. These studies provide a concrete demonstration of how information from diverse environments can easily and effectively be used to improve transferability of rigorously developed CG

models. This knowledge may be useful as part of a general framework for developing transferable CG models.

Our results also indicate that the transferability of CG force fields may be improved by incorporating information from the range of physical systems that one intends to model. For example, it may be possible to achieve a better description of non-native states by explicitly considering unfolded peptide conformations in developing the CG force field. Further studies will be necessary, however, to determine the most effective manner in which to integrate data from diverse physical environments into CG models. The possibility of applying such ideas to extend the applicability of the MS-CG approach is attractive because the method allows for transparent connections to be made between CG effective interactions and physical properties of fully detailed atomistic systems. One would expect that this characteristic will also be conveyed to transferable CG force fields having their foundation in the MS-CG approach.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gavoth@uchicago.edu.

Present Addresses

[†]Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA.

■ ACKNOWLEDGMENT

This research was supported by a National Science Foundation Collaborative Research in Chemistry grant (NSF Grant CHE-0628257/CHE-1047323). The computations were supported in part by the National Science Foundation through TeraGrid resources provided by the Texas Advanced Computing Center and Indiana University.

■ REFERENCES

- (1) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.
- (2) Heath, A. P.; Kavraki, L. E.; Clementi, C. *Proteins: Struct., Funct., Bioinf.* **2007**, *68*, 646–661.
- (3) *Coarse-Graining of Condensed Phase and Biomolecular Systems*, 1st ed.; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009; 456 pp.
- (4) Karttunen, M.; Murtola, T.; Vattulainen, I. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; 1st ed.; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009; pp 83–106.
- (5) Sherwood, P.; Brooks, B. R.; Sansom, M. S. P. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630–640.
- (6) Tozzini, V. *Acc. Chem. Res.* **2010**, *43*, 220–230.
- (7) Wang, Y. T.; Izvekov, S.; Yan, T. Y.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 3564–3575.
- (8) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. *J. Chem. Phys.* **2007**, *126*, 144509.
- (9) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2006**, *125*, 151101.
- (10) Hadley, K. R.; McCabe, C. *J. Phys. Chem. B* **2010**, *114*, 4590–4599.
- (11) Carbone, P.; Varzaneh, H. A.; Chen, X.; Muller-Plathe, F. *J. Chem. Phys.* **2008**, *128*, 064904.
- (12) Savelyev, A.; Papoian, G. A. *J. Phys. Chem. B* **2009**, *113*, 7785–7793.
- (13) Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2009**, *113*, 4413–4424.

- (14) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (15) Winter, N. D.; Schatz, G. C. *J. Phys. Chem. B* **2010**, *114*, 5053–5060.
- (16) Lu, L.; Voth, G. A. *J. Phys. Chem. B* **2009**, *113*, 1501–1510.
- (17) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Phys. Chem. B* **2007**, *111*, 9390–9399.
- (18) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (19) Kim, Y. C.; Hummer, G. J. *Mol. Biol.* **2008**, *375*, 1416–1433.
- (20) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. *J. Mol. Biol.* **2009**, *385*, 312–329.
- (21) Maisuradze, G. G.; Senet, P.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **2010**, *114*, 4471–4485.
- (22) Krishna, V.; Ayton, G. S.; Voth, G. A. *Biophys. J.* **2010**, *98*, 18–26.
- (23) Ayton, G. S.; Voth, G. A. *Biophys. J.* **2010**, *99*, 2757–2765.
- (24) Basdevant, N.; Ha-Duong, T.; Borgis, D. *J. Chem. Theory Comput.* **2006**, *2*, 1646–1656.
- (25) Saveliev, A.; Papoian, G. A. *Biophys. J.* **2009**, *96*, 4044–4052.
- (26) Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2010**, *31*, 1644–1655.
- (27) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. *J. Struct. Biol.* **2007**, *157*, 593–605.
- (28) Voltz, K.; Trylska, J.; Tozzini, V.; Kurkali-Siebert, V.; Langowski, J.; Smith, J. *J. Comput. Chem.* **2008**, *29*, 1429–1439.
- (29) Vorobyov, I.; Li, L. B.; Allen, T. W. *J. Phys. Chem. B* **2008**, *112*, 9588–9602.
- (30) Ayton, G. S.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2009**, *19*, 138–144.
- (31) Arkhipov, A.; Yin, Y.; Schulten, K. *Biophys. J.* **2009**, *97*, 2727–2735.
- (32) Ayton, G. S.; Lyman, E.; Voth, G. A. *Faraday Discuss.* **2010**, *144*, 347–357.
- (33) Wee, C. L.; Ulmschneider, M. B.; Sansom, M. S. P. *J. Chem. Theory Comput.* **2010**, *6*, 966–976.
- (34) Gopal, S. M.; Mukherjee, S.; Cheng, Y. M.; Feig, M. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1266–1281.
- (35) Shih, A. Y.; Freddolino, P. L.; Arkhipov, A.; Schulten, K. *J. Struct. Biol.* **2007**, *157*, 579–592.
- (36) Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Huenenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.
- (37) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A. *Biophys. J.* **2007**, *92*, 4289–4303.
- (38) Makowski, M.; Sobolewski, E.; Czaplewski, C.; Oldziej, S.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2008**, *112*, 11385–11395.
- (39) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
- (40) Ayton, G. S.; Noid, W. G.; Voth, G. A. *MRS Bull.* **2007**, *32*, 929–934.
- (41) Thorpe, I. F.; Zhou, J.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 13079–13090.
- (42) Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954–965.
- (43) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
- (44) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4116–4127.
- (45) Noid, W. G.; Liu, P.; Wang, Y. T.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.
- (46) Jernigan, R. L.; Bahar, I. *Curr. Opin. Struct. Biol.* **1996**, *6*, 195–209.
- (47) Betancourt, M. R.; Thirumalai, D. *Protein Sci.* **1999**, *8*, 361–369.
- (48) van Giessen, A. E.; Straub, J. E. *J. Chem. Phys.* **2005**, *122*, 024904.
- (49) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730–3737.
- (50) Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624–1636.
- (51) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323–2347.
- (52) Murarka, R. K.; Liwo, A.; Scheraga, H. A. *J. Chem. Phys.* **2007**, *127*, 155103.
- (53) Kozlowska, U.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2010**, *31*, 1143–1153.
- (54) Pace, C. N.; Scholtz, J. M. *Biophys. J.* **1998**, *75*, 422–427.
- (55) Rohl, C. A.; Fiori, W.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3682–3687.
- (56) Levy, Y.; Jortner, J.; Becker, O. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2188–2193.
- (57) Gnanakaran, S.; Garcia, A. E. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 773–782.
- (58) Couch, V. A.; Cheng, N.; Nambiar, K.; Fink, W. *J. Phys. Chem. B* **2006**, *110*, 3410–3419.
- (59) Morozov, A. N.; Lin, S. H. *J. Phys. Chem. B* **2006**, *110*, 20555–20561.
- (60) Buchete, N. V.; Hummer, G. *J. Phys. Chem. B* **2008**, *112*, 6057–6069.
- (61) Wimley, W. C.; Creamer, T. P.; White, S. H. *Biochemistry* **1996**, *35*, 5109–5124.
- (62) Fairman, R.; Shoemaker, K. R.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biophys. Chem.* **1990**, *37*, 107–119.
- (63) Shoemaker, K. R.; Fairman, R.; Schultz, D. A.; Robertson, A. D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1990**, *29*, 1–11.
- (64) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510–5517.
- (65) Sung, S. *S. Biophys. J.* **1999**, *76*, 164–175.
- (66) Wang, H.; Varady, J.; Ng, L.; Sung, S. S. *Proteins: Struct., Funct., Bioinf.* **1999**, *37*, 325–333.
- (67) Honda, S.; Kobayashi, N.; Munekata, E. *J. Mol. Biol.* **2000**, *295*, 269–278.
- (68) Kobayashi, N.; Honda, S.; Yoshii, H.; Munekata, E. *Biochemistry* **2000**, *39*, 6564–6571.
- (69) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238–7243.
- (70) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062–9067.
- (71) Roccatano, D.; Amadei, A.; Di Nola, A.; Berendsen, H. J. *Protein Sci.* **1999**, *8*, 2130–2143.
- (72) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068–9073.
- (73) Ma, B.; Nussinov, R. *J. Mol. Biol.* **2000**, *296*, 1091–1104.
- (74) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931–14936.
- (75) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345–354.
- (76) Bolhuis, P. G. *Biophys. J.* **2005**, *88*, 50–61.
- (77) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. *Proteins* **2005**, *61*, 795–808.
- (78) Bonomi, M.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2008**, *130*, 13938–13944.
- (79) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578–5583.
- (80) Snow, C. D.; Qiu, L.; Du, D.; Gai, F.; Hagen, S. J.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4077–4082.
- (81) Yang, W. Y.; Pitera, J. W.; Swope, W. C.; Gruebele, M. *J. Mol. Biol.* **2004**, *336*, 241–251.
- (82) Yang, L. J.; Shao, Q.; Gao, Y. Q. *J. Phys. Chem. B* **2009**, *113*, 803–808.
- (83) He, Y.; Xiao, Y.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2009**, *30*, 2127–2135.
- (84) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. *Comput. Chem.* **2009**, *30*, 1545–1614.
- (85) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (86) Smith, W. *Mol. Simul.* **2006**, *32*, 933–1121.
- (87) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

- (88) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, 1981; p 331.
- (89) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (90) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 16557–11666.
- (91) Mullinax, J. W.; Noid, W. G. *J. Chem. Phys.* **2009**, *131*, 104110.
- (92) Feig, M.; MacKerell, A. D.; Brooks, C. L. *J. Phys. Chem. B* **2003**, *107*, 2831–2836.
- (93) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.
- (94) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (95) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (96) Shehu, A.; Kavraki, L. E.; Clementi, C. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 837–851.
- (97) Liu, P.; Voth, G. A. *J. Chem. Phys.* **2007**, *126*, 045106.
- (98) Liu, P.; Shi, Q.; Lyman, E.; Voth, G. A. *J. Chem. Phys.* **2008**, *129*, 114103.
- (99) Mullinax, J. W.; Noid, W. G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19867–19872.
- (100) Hills, R. D., Jr.; Lu, L.; Voth, G. A. *PLoS Comput. Biol.* **2010**, *6*, e1000827.
- (101) Matysiak, S.; Clementi, C. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*, 1st ed.; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009; pp 157–170.