JCTC Journal of Chemical Theory and Computation

Coarse-Grained Protein Model Coupled with a Coarse-Grained Water Model: Molecular Dynamics Study of Polyalanine-Based Peptides

Wei Han*,† and Yun-Dong Wu*,†,‡

Department of Chemistry, The Hong Kong University of Science & Technology, Kowloon, Hong Kong, China, and State Key Lab of Molecular Dynamics and Stable Structures, College of Chemistry, Peking University, Beijing, China

Received June 24, 2007

Abstract: The coupling of a coarse-grained (CG) protein model with the CG water model developed by Marrink et al. (*J. Phys. Chem. B* **2004**, *108*, 750) is presented. The model was used in the molecular dynamics studies of Ac-$(Ala)_6$-Xaa-$(Ala)_7$-NHMe, Xaa = Ala, Leu, Val, and Gly. A Gly mutation in the middle of polyalanine is found to destabilize the helix and stabilize the hairpin by favoring a type-II′ turn and probably to speed up hairpin folding. The simulations allow us to derive thermodynamic parameters of, in particular, the helical propensities (*s*) of amino acids in these polyalanine-based peptides. The calculated *s* values are 1.18 (Ala), 0.84 (Leu), 0.30 (Val), and <0.02 (Gly) at 291 K, in excellent agreement with experimental values (R^2=0.970). Analyses using a structural approach method show that the helical propensity difference of these amino acids mainly comes from solvation effect. Leu and Val have lower helical propensities than Ala mainly because the larger side chains shield the solvation of helical structures, while Gly has a much poorer helical propensity mainly due to the much better solvation for the coil structures than for the helical structures. Overall, the model is at least about 10^2 times faster than current all-atom MD methods with explicit solvent.

Introduction

The problem of protein folding is an active area of experimental and theoretical research.[1,2] Computer simulations, as indispensable tools in this area, provide microscopic insights to complement the interpretation of experimental observations.[3−5] All-atom simulations, which explicitly represent every atom of proteins and solvent molecules, can reveal the maximum details but are computationally demanding. The shortest time scale of the folding of a small protein is about tens of microseconds, but most all-atom simulations can only be carried out up to microseconds.[6−8] Thus, all-atom simulations are currently impractical in describing the protein folding completely.

An alternative to all-atom simulations is coarse-grained (CG) simulation, in which a group of atoms is reduced to a single particle.[9] In addition, solvent molecules are usually implicitly represented. The gain in simulation speed comes from a large reduction in the number of particles and smoother interparticle potentials. Thus, coarse-grained simulations have been useful in the study of protein folding. There have been many successful applications of coarse-grained simulations that characterize proteins at different levels of details.[10−19] The minimalist model characterizes each amino acid with one single particle or one particle for the backbone and one for the side chain.[9−13] This model has provided insights into general folding mechanisms, but it has less predictive values since it relies on preknowledge of native structures to impose Gō-type biases onto interparticle potentials.[10,11] Models between the minimalist level and the all-atom level have been devised to reproduce reasonable Ramachandran maps and anisotropic hydrogen bond (HB)

* Corresponding author e-mail: hwer@ust.hk (W.H.) and chydwu@ust.hk (Y.-D.W.).
† The Hong Kong University of Science & Technology.
‡ Peking University.

10.1021/ct700151x CCC: $37.00
Published on Web 10/02/2007

potentials.[14−18] These models allow coarse-grained simulations to qualitatively study the protein folding or aggregation without any biased potential.

HB and hydrophobic interactions play essential roles in protein stability and dynamics. They strongly depend on the local environment, such as the solvation level, and have a many-body character.[20] These interactions in some CG models remain pair wise additive.[15,16] Interestingly, Takada et al.[14] introduced HB and hydrophobic potentials that depend on local densities of protein particles. These densities indicate the extent of exposure of particles involved in HB or hydrophobic interaction.

Another way to model the solvent effect is to use a coarse-grained (CG) solvent model. Shelley et al. proposed a CG model to study lipid aggregation in solutions.[21] Their model basically has two types of CG particles (waterlike and oil-like). A group of water molecules can be represented by one CG waterlike particle. Water−oil interactions are unfavorable. The interactions are favored with like particles. The solvation effect can be explicitly taken into account. This model has been further developed and calibrated by Marrink et al.[22] to enable semiquantitative or quantitative comparisons with experiments. The simulation speed is increased by at least 10^3-fold. This model has been successfully applied to the study of membrane fusion.[23,24] Bond et al.[25] and Shih et al.[26] have shown that some behaviors of membrane proteins can be studied by using minimalist models of CG proteins with this CG solvent/lipid model. The current limit of the CG solvent/lipid model is that it cannot deal with atomistic details of proteins.[22,26] However, recent studies by Voth's group[27,28] indicated that all-atom force fields and CG models can work together well when each water molecule is represented by one particle.

In this paper, we present a CG protein model at an intermediate resolution level. This model is developed in tandem with the CG solvent model by Marrink et al.[22] and has been applied to the molecular dynamics studies of Ac-$(Ala)_6$-Xaa-$(Ala)_7$-NHMe (Xaa = Ala, Leu, Val, and Gly). We show that (1) parameters can be optimized in a systematic way that is compatible to the procedures by Marrink et al.; with such procedures, it is easy to incorporate new models and parameters into our CG model in a consistent manner; (2) structures in atomistic details, such as α-helix, β-hairpin, and β-turn, can fold and/or refold properly in polypeptides without biased potentials; (3) the simulations are fast enough to obtain the equilibrium thermodynamic and kinetic properties of polypeptides; and (4) the properties from optimized parameters are comparable to those from experiments.

Models and Methods

The CG Protein Model. As shown in Figure 1a, in the backbone of our CG protein, each heavy atom with its attached hydrogen(s) is explicitly represented by one CG particle, resembling the model by Ding et al.[16] "R" is the side-chain group that determines the identity of the amino acid (aa). Four kinds of amino acids are studied in this paper, including Ala, Val, Leu, and Gly. Their representations are pictured in Figure 1b. All heavy atoms of Ala and Val are explicitly considered, while the isopropyl group of Leu,

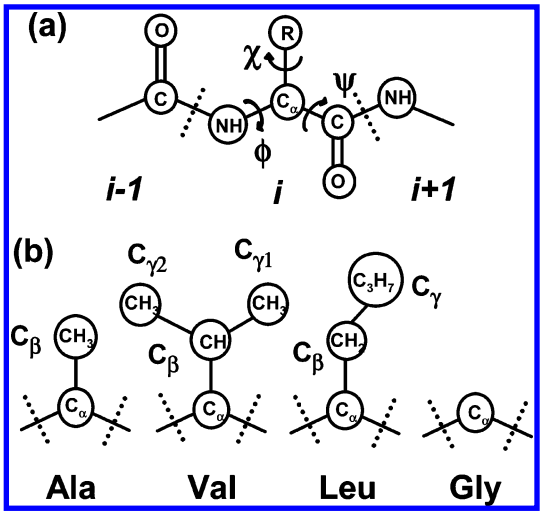

Figure 1. The scheme of the CG protein model.

which connects to C_β carbon, is replaced by one CG particle at its centroid position. The potential energy of this CG model is described by eq 1

$$V_{\text{Total}} = V_{\text{Angle}} + V_{\text{Improper}} + V_{\text{Torsion}} + V_{\text{loc-cdW}} + V_{\text{vdW}} + V_{\text{HB}} \quad (1)$$

where V_{Total} can be partitioned into the bonded terms, $V_{\text{Angle}} + V_{\text{Improper}} + V_{\text{Torsion}} + V_{\text{loc-vdW}}$, and the nonbonded terms, which are the remaining parts.

Bonded Interactions. The bonded interactions are defined as the interactions between the particles connected by the direct bonds or separated by less than four chemical bonds. The direct bonds are constrained by the LINCS algorithm[29] with the bond length of r_0. The interactions between particles i and j, which both connect to particle k, are described by eq 2 with $K_{\text{Angle}} = 72$ kcal mol^{-1} rad^{-2}. The equilibrium value of $\angle(i\text{-}k\text{-}j)$ is θ_0.

$$V_{\text{Angle}} = K_{\text{Angle}}(\theta - \theta_0)^2/2 \quad (2)$$

In order to maintain the planarity of a carbonyl group or the chirality of an sp^3 carbon, eq 3 is used

$$V_{\text{Improper}} = K_{\text{Improper}}(\xi - \xi_0)^2/2 \quad (3)$$

where ξ is the dihedral of the four particles involving in planar or the chiral groups. K_{Improper} is 72 kcal mol^{-1} rad^{-2}.

The torsional angle is defined as the dihedral angle, $\angle(i\text{-}j\text{-}k\text{-}l)$, of four particles, i, j, k, and l, which are connected by three successive chemical bonds. They are known to be critical to determining the local conformational features of amino acids. Following Takada et al.,[14] we use the combination of two kinds of potentials (eqs 4 and 5) to describe the torsion.

$$V_{\text{Torsion}} = K_{\text{Torsion}}(1 + \cos(n\phi - \phi_0)) \quad (4)$$

$$V_{\text{loc-vdW}} = \sum_{1-4\ \text{relationship}} 4\epsilon_{\text{loc}}\left(\frac{\delta_{\text{loc},ij}^{12}}{r^{12}} - \frac{\delta_{\text{loc},ij}^{6}}{r^{6}}\right) \quad (5)$$

In eq 4 ϕ is $\angle(i\text{-}j\text{-}k\text{-}l)$, and n is the multiplicity of the periodic potential. Equation 5 describes the van der Waals (vdW)

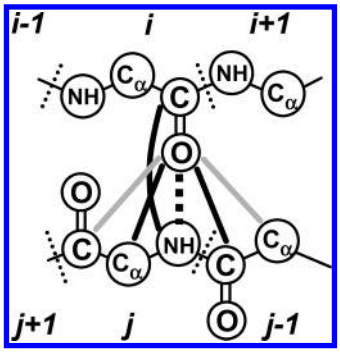

Figure 2. The scheme of the HB interaction between amide units. The *i* and *j* are residue numbers.

overlap between particles *i* and *l*. Such an interaction is atomic in nature and is weaker than nonlocal vdW interactions, where atoms are separated by more than three bonds.[14] In our model, ϵ_{loc} is set to be 0.22 kcal/mol, smaller than that for nonlocal vdW interactions, which will be introduced later. The vdW radii, $r_{vdW-loc}$, are also somewhat smaller than the nonlocal vdW radii.

Nonbonded Interactions. Nonbonded interactions in our model comprise nonlocal vdW interactions (separated by >3 bonds) and HB interactions.

Lennard-Jones (LJ) potentials (eq 6) are used to describe nonlocal vdW interactions. Following the treatment by Marrink et al.,[22] electrostatic interactions between groups with partial charges are implicitly incorporated into nonlocal vdW interactions.

$$V_{vdW} = \sum_{i<j} 4\epsilon_{ij}\left(\frac{\delta_{ij}^{12}}{r^{12}} - \frac{\delta_{ij}^{6}}{r^{6}}\right) \quad (6)$$

Another important nonbonded interaction, the HB interaction between amide units (Figure 2), is known to be crucial for α-helix, *β*-sheet, and *β*-turn conformations of proteins. The HB interaction is normally described with both dipole−dipole interactions and vdW interactions in current all-atom force fields. Takada et al. devised a simplified way to model this anisotropic interaction in their CG model.[14] The basic idea is to allow not only the attraction between the carbonyl (carbonyl is represented as one particle in their model) of one peptide unit and the nitrogen of another but also the auxiliary repulsions between particles adjacent to the carbonyl/nitrogen of one peptide unit to the nitrogen/carbonyl of the other. Our HB model is similar but slightly different since a carbonyl group is represented by C and O particles. This representation has been argued to be better for HB geometry.[16] Indeed Takada et al. reported that polyalanine folded into helices and also fold an experimentally designed peptide[30] folded into a helix bundle using the model.[14] In addition, Ding et al. also used the HB potential to successfully fold a miniprotein of 20 aas, namely Trp-cage,[31] into its NMR conformation.[16] These examples reflect the applicability of the HB potential.

Figure 2 illustrates the HB potential described by eq 7, where the black bold dotted lines indicate the attraction, and the black bold solid lines indicate the auxiliary repulsion. Nevertheless, because our HB model adopts the repulsive interaction between O particles and C/C_α particles, an extra repulsion exists (gray lines in Figure 2) although this repulsion is irrelevant to the HB interaction. This can weaken the expected HB interaction, especially in *β*-sheets, where the distances between particles associated with the extra repulsion are quite short. To counteract this extra repulsion, we increased the C−C, C−C_α, and C_α−C_α attraction (ϵ_{vdw}) between residues *i* and *j* if $|i-j| > 2$.

$$V_{HB} = \sum_{|i-j|>2}\left[4\epsilon_{attr}\left(\frac{\delta_{Oi-NHj}^{12}}{r_{Oi-NHj}^{12}} - \frac{\delta_{Oi-NHj}^{6}}{r_{Oi-NHj}^{6}}\right) + 4\epsilon_{rep}\frac{\delta_{Oi-C\alpha j}^{12}}{r_{Oi-C\alpha j}^{12}} + 4\epsilon_{rep}\frac{\delta_{Oi-Cj-1}^{12}}{r_{Oi-Cj-1}^{12}} + 4\epsilon_{rep}\frac{\delta_{Ci-NHj}^{12}}{r_{Cj-NHj}^{12}}\right] \quad (7)$$

Parameter Optimization and Optimized Parameters. The optimization procedures for all parameters are shown in Figure 3. Our strategy is to optimize the parameters for the backbone (Ala, Gly) first by molecular dynamics (MD) simulations of polyalanine. The parameters for Ala have enough details to describe Gly. Once the backbone parameters are obtained, they remain unchanged. For other amino acids, such as Leu and Val, that are presented in this paper, we need to optimize the parameters for the part of the side chains that are attached on C_β of Ala.

Parameters for Bonded Interactions. The values for all bonded interactions are listed in Table 1. Specifically, they are obtained as follows:

The parameters of bond length (r_0) and bond angle (θ_0) related to backbone atoms, such as C, C_α, NH, O, and C_β, and the side-chain atoms of Val are available in the average results of X-ray crystal structures,[32] which are basically the same with the parameters used in previous studies.[14−18] The improper (ξ_0) parameters are set here to ensure that the C_i, $C_{\alpha i}$, N_{i+1}, and O_i are placed in the same plane and that the amino acid has an L-configuration.

The torsional potentials ($K_{Torsion}$, ϕ_0, and *n*) and the local vdW parameters ($\delta_{loc,ij}$) for backbone atoms were optimized as follows: (1) An alanine dipeptide was simulated with varying torsional and local vdW parameters so that its Ramanchadran (ϕ, ψ) map achieved separated regions, such as α, *β*, and PPII regions, and the population deviation in each region was less than 10−15% compared to previous studies (see the next paragraph). (2) Ac-$(Ala)_{14}$-NHMe in the CG water was simulated repeatedly (about 150 times) with varying local vdW parameters. Each trial simulation lasted for about 1 *μ*s. Parameters were optimized so that (a) the full helical structure could refold for more than five times and the full hairpin structure (defined in Data Analyses) can refold twice from fully extended conformation and that (b) the two secondary structures, once formed, could last for nanoseconds before they were broken again.

With these parameters, the (ϕ, ψ) plots of Ac-Xaa-NHMe in CG water at 300 K are shown in Figure 4. The Ala dipeptide has about 38% *β* conformation ((ϕ, ψ) at (−135°±45°,135°±45°)), about 16% PPII conformation ((ϕ, ψ) at (−45°±45°,135°±45°)), and about 27% total helical conformation ((ϕ, ψ) at (−90°±90°,−45°±45°)). This is in good

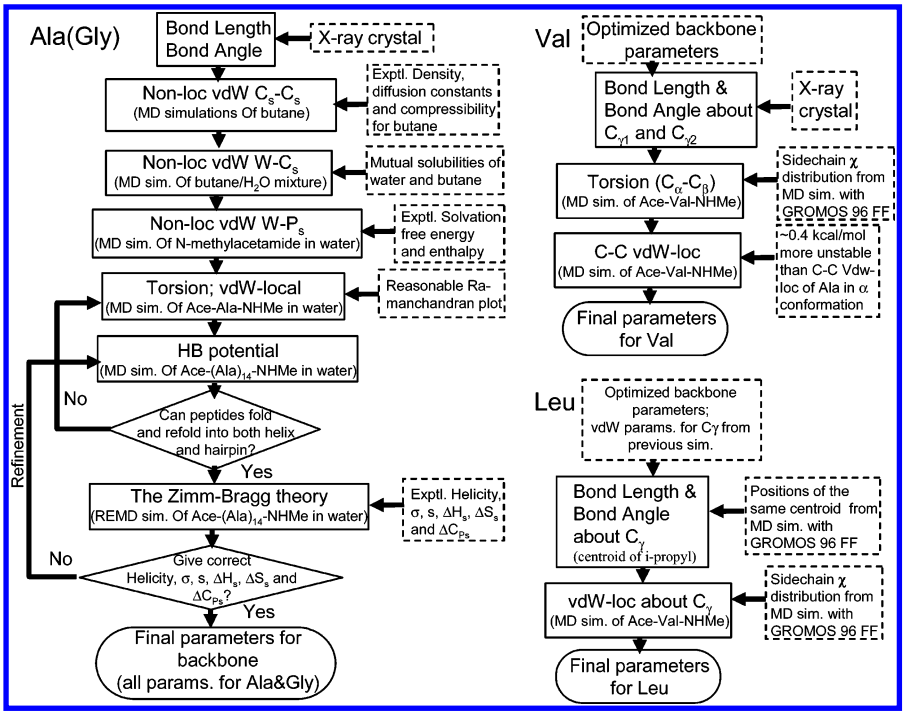

Figure 3. Flowcharts of parameter optimization. Each rectangle with a solid outline contains one optimization step. Each rectangle with a dotted outline contains the data used for optimization and/or their sources. Each optimization step is performed based on the parameters optimized from previous steps. W indicates CGW particles; C indicates CG nonpolar particles; C_s indicates a small $CH_{x\ (x=0-3)}$ group or carbonyl carbon; P_s indicates an O atom or NH group.

Table 1. Parameters of r_0, θ_0, and ξ_0

bond	r_0 (nm)	bond	r_0 (nm)
$C_\alpha-C$	0.152	$C_\beta-C_{\gamma1/\gamma2}$	0.153
C−N	0.133	$C_\beta-C_\gamma$ (Leu)	0.194
$C_\alpha-N$	0.145	C−O	0.123
$C_\alpha-C_\beta$	0.153		
angle	**θ_0 (deg)**	**angle**	**θ_0 (deg)**
$N-C_\alpha-C$	111.6	$C_\alpha-C-O$	121.0
$C_\alpha-C-N$	117.5	N−C−O	124.0
$C-N-C_\alpha$	120.0	$C_\alpha-C_\beta-C_{\gamma/\gamma/2}$	111.0
$N-C_\alpha-C_\beta$	110.0	$C_{\gamma/1}-C_\beta-C_{\gamma/2}$	111.0
$C-C_\alpha-C_\beta$	110.0	$C_\alpha-C_\beta-C_\gamma$ (Leu)	124.0
improper	**ξ_0 (deg)**	**improper**	**ξ_0 (deg)**
$C_{\alpha i}-N_i-C_i-C_{\beta i}$	35.3	$C_i-C_{\alpha i}-N_{i+1}-O_i$	0.0

agreement with the results from the OPLS/AA/L force fields,[33] which produce slightly more β (31%) conformation than PPII (25%) conformation.[34] It is in moderate agreement with a recent experiment based on PR/FTIR, which found that $(Ala)_3$ can dominantly adopt β and PPII conformations at a 1:1 ratio.[35]

The torsional and pair interactions involving the side chains of Val were optimized by matching the distribution of $N-C_\alpha-C_\beta-C_{\gamma1}$ of the CG model with that from the GROMOS96 force field[36] through the simulations of Val dipeptide in water. The matching results are shown in Figure S1 of the Supporting Information (SI).

For Leu, a similar way was used to optimize the parameters about bond $C_\beta-C_\gamma$, angle $C_\alpha-C_\beta-C_\gamma$, and the dihedral angles involved with the side chain, where C_γ represents the centroid of the isopropyl group. The r_0 of $C_\beta-C_\gamma$ and the θ_0 of $C_\alpha-C_\beta-C_\gamma$ (Table 1) are just the positions of the single narrow peaks in the distributions of $C_\beta-C_\gamma$ and $C_\alpha-C_\beta-C_\gamma$ of the Leu dipeptide from the GROMOS96[36] simulations (Figure S2a, Supporting Information), respectively. The detailed matching results about the distribution of the $N-C_\alpha-C_\beta-C_\gamma$ dihedral angle are given in Figure S2b, Supporting Information.

In addition, a local C_i-C_{i+1} vdW interaction of β-branched aa, such as Val, is modified (Table 2). A model building reveals that when ϕ is about −60° (Figure 5a), hydrogen atoms of C_γs are too close to hydrogen atoms of backbone amides in the three side-chain rotamers. The distance is about 0.21 nm, shorter than previously reported repulsive vdW diameters (0.24−0.26 nm) for hydrogen.[37] Such repulsion is absent in extended conformations ($\phi < -120°$). This effect cannot be explicitly considered if amide hydrogen is ignored. We therefore handled the repulsion by enlarging δ_{loc} of the local vdW C_i-C_{i+1} interaction. With molecules A and B (Figure 5b) used to model the ϕs of Val and Ala, respectively, quantum mechanics calculations at the B3LYP/6-311++G** level show that the energy difference between A with ϕ = −80° to −60° and A in its global minimum is about 0.4−0.6 kcal/mol higher than that for B. This gives an estimation of the repulsive effect for Val compared to Ala. According to dipeptide simulations, our modified $\delta_{loc,Ci-Ci+1}$ on average leads to about 0.4−0.5 kcal/mol more C_i-C_{i+1} repulsion for Val than it does for Ala when ϕ is −80° to −60°.

Parameters for Nonlocal vdW Interactions. One of the purposes of this work is to make our CG model compatible with the CG solvent model by Marrink et al.[22] In the original CG model, ϵ_{ij} (eq 6) of the vdW interactions has discrete

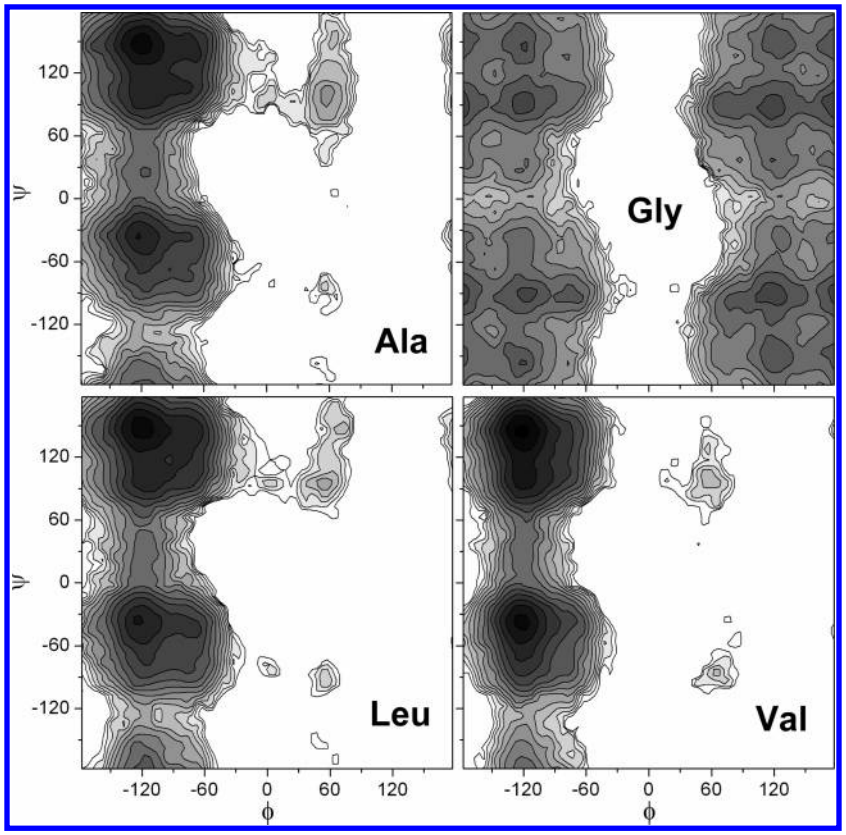

Figure 4. The Ramachandran plots of Ala, Gly, Leu, and Val. The free energy interval for the contours is 0.25 kcal/mol. The darker region has lower free energy.

Table 2. $K_{Torsion}$, ϕ_0, n, and $r_{vdW-loc}$

torsional angle	$K_{Torsion}$ (kcal/mol)	ϕ_0 (deg)	N
$C_{\alpha i-1}-C_{i-1}-N_i-C_{\alpha i}$	10.00	180.0	2
$C_{i-1}-N_i-C_{\alpha i}-C_i$	0.20	180.0	6
$N_i-C_{\alpha i}-C_i-N_{i+1}$	0.20	0.0	6
$N_i-C_{\alpha i}-C_{\beta i}-C_{\gamma 1,il\gamma,i}$	1.20	0.0	3

atom type	$r_{vdW-loc}$[a] (nm)
O	0.130
NH	0.145
C	0.165/0.155/0.168[b]
$CH/CH_2/CH_3$	0.165
C_3H_7	0.185

[a] $\delta_{loc,ij}$ is equal to $r_{vdw-loc,i}+r_{vdw-loc,j}$, where i and j are atom types. [b] 0.165 is for the interaction between C and other particles with different kinds of atom types, 0.155 is for the C−C interaction in the non-β-branched aas, and 0.168 is $^1/_2\delta_{loc,Ci-1-Ci}$ for β-branched aa i.

levels, which are 1.20, 1.01, 0.82, 0.63, and 0.44 kcal/mol in the metric of 0.19 kcal/mol. δ_{ij} is uniformly 0.47 nm. These parameters were optimized for CG particles that represent four atoms (butane or equivalent).[25,26] However, the values of ϵ_{ij} and δ_{ij} of the original model seem to be too large for our CG particles, some of which represent only one atom. Therefore, the vdW radii of our CG particles are taken from the statistical survey of crystal structures.[38] The isopropyl particle in Leu is composed of three carbon atoms, and its r_{vdw} is taken from Shelley et al.[21] In addition, we used a method similar to that of Marrink et al.[22] to reoptimize the ϵ_{ij} of interactions between the CG solvent and our CG protein. As shown in Figure 3, nonlocal vdW parameters are optimized so that simulations can reproduce important

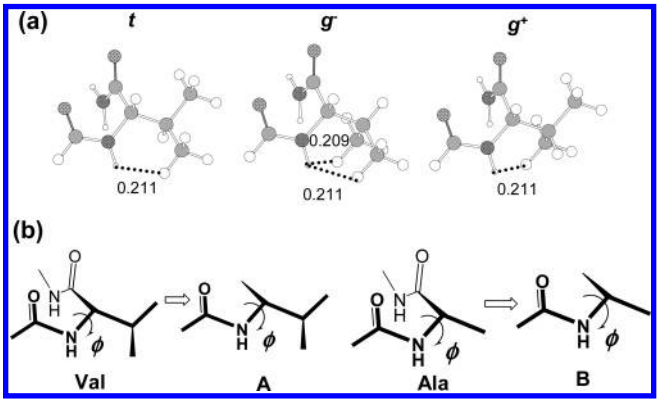

Figure 5. (a) Val in α conformation with its side chain in three rotamers and (b) model molecules for Val (A) and Ala (B).

physical properties of pure liquid and liquid mixtures. All nonlocal vdW parameters are listed in Table 3.

The ϵ_{ij} for interactions between small hydrophobic particles ($CH_{x(x=0-3)}$) (called C_s particles) is obtained by simulating a system of 400 CG butane molecules, each of which is composed of four C_s particles. ϵ_{ij} between CG water (CGW) and C_s particles is obtained by simulating mixtures of 400 CGW molecules and 400 CG butane molecules (900 ns for each simulation). The simulations are conducted at $T = 300$ K and $P = 1$ atm with a Nose-Hoover thermostat[39,40] and a Parrinello-Rahman pressure bath.[41] The resulting values for ϵ_{ij} of both C_s-C_s and $CGW-C_s$ interactions are 0.25 kcal/mol.

With these parameters, the density, compressibility, and self-diffusion constant of butane as well as free energies to

Table 3. ϵ_{ij} and r_{vdW}

ϵ_{ij} (kcal/mol)	W[a]	CH_x(x=0−3)	O	NH	C_3H_7
W(P)	1.20				
$CH_{x(x=0-3)}$	0.25	0.25			
O	1.20	0.25	0.25		
NH	1.20	0.25	0.25[a]	0.25	
C_3H_7	0.44	0.44	0.25	0.25	0.82

type	r_{vdW}[b] (nm)	type	r_{vdW}[b] (nm)
$CH_{x(x=1-3)}$	0.185	W	0.235
C	0.165	NH	0.165
C_3H_7	0.220	O	0.140

[a] The ϵ_{ij} of O_i and NH_j is valid only if $|i-j| < 3$, where i and j are residue numbers. [b] δ_{ij} in eq 6 is equal to $r_{vdW,i} + r_{vdW,j}$.

Table 4. Experimental and Calculated Physical Properties of Butane and Mutual Transfer Free Energies between Butane and CG Water at 300 K

	exptl	I[e]	II
density/g·cm^{-3}	0.58[a]	0.68	0.74
compressibility/10^{-5} bar^{-1}	>17[a]	28	8
diffusion[b]/10^{-5} cm^2·s^{-1}	>5[b]	1.9	1.8
$\Delta G_{But(But\rightarrow W)}$/kcal·mol^{-1}	5.5[c]	5.4	7.2
$\Delta G_{W(W\rightarrow But)}$/kcal·mol^{-1}	6.0[d]		6.5

[a] Measured at 293 K.[43] [b] Obtained from the slope of the mean squared displacement (MSD) curve in the long time limit. Experimental values are from ref 44. [c] Measured at 298 K.[44] [d] Measured at 294 K.[46] [e] Obtained by using one CG particle to represent one butane molecule (ref 22). [f] No detectable mutual solubility.

transfer water into butane, $\Delta G_{W(W\rightarrow But)}$, or butane into water $\Delta G_{But(But\rightarrow W)}$, which is derived from mutual solubilities of water and butane,[42] were calculated from simulations. The simulated properties are reasonably comparable to those from experiments and from the work by Marrink et al. (Table 4[43−46]), except that compressibility is about half of the experimental value and $\Delta G_{But(But\rightarrow W)}$ is 1.7 kcal/mol higher than the experimental value, indicating that $\epsilon_{ij,Cs-Cs}$ may need to be further reduced. The vdW interactions for other small particles are here supposed to be similar to those of C_s particles. Therefore, for simplicity, we add one more energy level, $\epsilon_{ij} = 0.25$ kcal/mol, which satisfies the metric of 0.19 kcal/mol in the original energetic system,[22] for interactions between small particles except for the interacting particles participating in HB interactions.

In order to parametrize ϵ_{ij} of the interactions between O and NH of amide and CGW, we investigated the solvation of *N*-methylacetamide, which is a model compound widely used in the study of peptide hydration.[47−49] An *N*-methylacetamide molecule was placed in a box with 350 CG water particles. Its solvation free energy, ΔG_{sov}, enthalpy ΔH_{sov}, and entropy $T\Delta S_{sov}$ were calculated with the thermal integration (TI) method[50] as described in the Appendices. The TI can normally compute solvation free energy accurately if that the sampling is adequate.[51] ΔG_{sov}, ΔH_{sov}, and $T\Delta S_{sov}$ with different ϵ_{ij} from the TI are shown in Table 5. Compared to experimental values (−10.1 kcal/mol), ΔG_{sov} is the best (−9.2 kcal/mol)) when $\epsilon_{ij} = 1.20$ kcal/mol.

Although TI is good for accurate calculation of the solvation free energy, it can only be applied to small or rigid molecules.[52] TI becomes impractical for polypeptides that

Table 5. Experimental (298 K) and Calculated (300 K) ΔG_{sov}, ΔH_{sov}, $T\Delta S_{sov}$, ΔG_{cav}, and $<U_{int}>$ of *N*-Methylacetamide (kcal/mol)

	ΔG_{sov} −10.1[a]		ΔH_{sov} −17.1[c]		$-T\Delta S_{sov}$ 7.0		ΔG_{cav}[d]		
exptl	TI[b]	SPT[b]	TI	SPT	TI	SPT	TI	SPT	$<U_{int}>$[e]
$\epsilon = 1.39$	−11.6	−16.5	−15.6	−18.7	4.0	2.1	10.5	5.6	−22.1
$\epsilon = 1.20$	**−9.2**	**−13.8**	**−14.2**	**−15.9**	**5.0**	**2.1**	**10.2**	**5.6**	**−19.4**
$\epsilon = 1.01$	−6.7	−11.0	−11.3	−13.2	4.5	2.1	9.8	5.6	−16.6
$\epsilon = 0.82$	−4.4	−8.4	−8.4	−10.5	4.0	2.1	9.6	5.6	−14.0

[a] From ref 47. [b] "TI" means thermal integration; "SPT" means the scaled particle theory with SAS assumption. [c] From refs 48 and 49. [d] Free energy to create a cavity in solvent, which is approximately equal to $\Delta G_{sov} - <U_{int}>$. [e] Average interaction energy between solvent and solute.

are highly flexible. For this reason, an approximate method based on the scaled particle theory (SPT) together with the assumption of a solvent accessible surface (SAS) for polyatomic molecules was used to estimate the solvation of polypeptide in our study (Appendices).[53−57] To examine the difference between the TI and SPT approaches for polypeptides, the solvation properties of *N*-methylacetamide were also computed by the SPT method and are listed in Table 5. The results reveal that the SPT based approach overestimates ΔG_{sov} by about 4−5 kcal/mol compared with the TI method. The difference is from the calculation of the free energy, ΔG_{cav}, to make a solute-sized cavity. It is calculated to be about 10 kcal/mol by the TI method, which is close to the value of 9.2 kcal/mol derived from a theoretic treatment of experimental data.[58] But the ΔG_{cav} values by the SPT approach is only 5.6 kcal/mol, about half of the TI value. Indeed, a previous study found that the SPT method could underestimate the work to create cavity.[59] Similarly, the calculated $T\Delta S_{cav}$ (−2.1 kcal/mol) and ΔH_{cav} (3.5 kcal/mol) by the SPT method are also about twice the $T\Delta S_{cav}$ (−4 to −5 kcal/mol) and ΔH_{cav} (5−6 kcal/mol) by the TI method. Consequently, when the solvation of polypeptides was calculated by the SPT with SAS assumption, $T\Delta S_{cav}$ and ΔH_{cav} were increased by 1-fold as a rough correction for the SPT approach.

It should be noted that in our present model, we used the discrete energy levels of nonbonded parameters that were used by Marrink et al. for simplicity. In our further work, we will remove this restriction so that solvation free energy may be better calculated by fine-tuning parameters such as $\epsilon_{ij,Cs-Cs}$ and $\epsilon_{ij,W-O/NH}$.

Parameters for HB Interactions. The optimized HB parameters are as follows: $\epsilon_{attr} = 3.35$ kcal/mol; $\epsilon_{rep} = 1.08$ kcal/mol; $\delta_{Oi-NHj} = 0.24$ nm; $\delta_{Oi-C\alpha j} = \delta_{Oi-Cj-1} = 0.29$ nm; $\delta_{Ci-NHj} = 0.338$ nm. In addition, the enhanced $\epsilon_{vdW,ij}$ for C and C_α particles is 0.63 kcal/mol.

Finally, during the optimization, we found that π-helices ($i\rightarrow i+5$ HB) were significantly sampled in the simulations, and sometimes their population even overwhelms that of α-helices ($i\rightarrow i+4$ HB). In real polypeptides, however, they should be rare. This may be because that the current HB model cannot differentiate between the two helical structures very well. To differentiate these two helices, their structures were inspected in detail. We found that the distance between

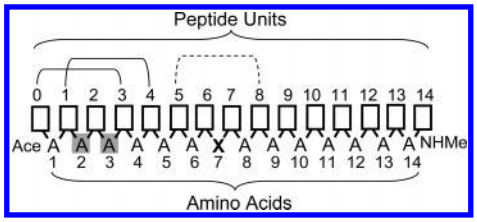

Figure 6. The scheme of Ac-$(Ala)_6$-Xaa-$(Ala)_7$-NHMe. The squares represent peptide units, and the arches indicate the helical hydrogen bonds between the peptide units.

$C_{\alpha i}$ and $C_{\beta i+4}$ in the π-helices in our model is quite short (~0.37 nm), which is not the case for α-helices. We therefore can increase δ_{vdW} for $C_{\alpha i}$ and $C_{\beta j}$ ($|i-j| \geq 2$) to selectively destabilize the π-helices. It turns out that when σ_{vdW} is 0.435 nm, the π-helices are significantly weakened and the α-helices are strengthened, while the β-hairpins are not affected.

Models. The polypeptide models used in the simulations were Ac-$(Ala)_6$-Xaa-$(Ala)_7$-NHMe, where Xaa is Ala if the peptide is a polyalanine (polyA) and Leu, Val, and Gly if the peptide is a Xaa mutant of polyA. The peptide chain includes two kinds of groups (Figure 6), a peptide unit (−CO−NH−) (PU, represented by squares) and an amino acid unit (C_α and side chain) (AU, represented by letters). Each PU/AU has two neighboring AUs/PUs. The numbering scheme for PUs and AUs is illustrated in Figure 6.

Simulation Setup. The simulations were performed with the GROMACS 3.3.1 package.[60] A peptide with a helical conformation was placed into a dodecahedron box with ~1100 CG water particles. The shortest distance between the peptide and the edges of the box was 1.5 nm. The vdw interaction had a cutoff of 1.2 nm, and it was smoothed to 0 from 0.9 to 1.2 nm. The temperature and pressure were controlled by a thermostat and a pressure bath, with coupling constants of 0.1 and 0.5 ps, respectively.[61] The time interval to integrate the Newton equations was 6 fs, and the neighboring list was updated every 10 steps. The whole system was subjected to 5000 steps of steep descent optimization and then to a 200 ps of pre-equilibrium at 300 K and 1 atm with the peptide constrained. The system was then heated at 340 K with the peptide relaxed for 100 ns. The generated conformations with no apparent helical or hairpin structures were used as starting points for the long simulations.

Replica exchange molecular dynamics (REMD) simulations provide an efficient platform to perform equilibrium simulations,[62,63] which is beneficial to parameter optimization. Our REMD simulations contained 14 replicas with temperatures ranging from 291 to 436 K at 1 atm. Each replica started with a different conformation generated from the heating simulation at 340 K. Exchanges were attempted every 1 ps. Each REMD simulation lasted for 200 ns, while the results of the last 150 ns of the simulations were analyzed.

Mass Scaling in the REMD Simulations. In the original CG solvent model,[22] the time interval to integrate the Newton equations is 25−45 fs. Since some particles in our model represent only a single atom and are connected directly by strong covalent bonds, they are so light that the time interval can only be up to 6−10 fs in order to keep the simulations from crashing. This crashing problem is especially serious for REMD since high-temperature (up to 436 K) simulations are involved. To avoid the problem we quadrupled the masses of peptide particles. This allows the interval for integration to be kept at 10 fs. Although this action changes the dynamics of those motions heavily dependent on mass, it may be feasible for the REMD simulations, whose dynamics have lost their physical meaning. The effect of mass scaling on the thermodynamics of the simulations at physiological temperatures is reported in the Results and Discussion.

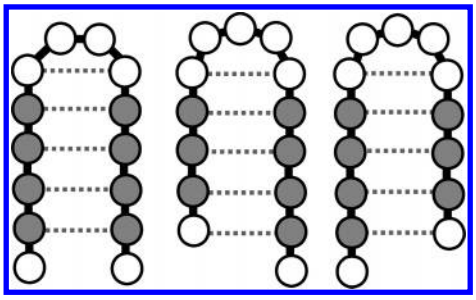

Figure 7. Three possible hairpin topologies with turns in the middle of the peptide. Dotted lines denote the conditions (<0.65 nm) between C_α atoms of different aas used to identify hairpin topologies.

Data Analyses

Definition of Helical Structures. Helical structures are formed by at least three successive aas in the α conformation, which is defined as an aa conformation with its (ϕ, ψ) at (−60°±30°,−47°±30°), as suggested by García et al.[64] The middle residue of the three aas is considered to be in the helical state in our study. As shown in Figure 6, if aas 1, 2, and 3 form a helical structure, aa 2 is in a helical state and the CO group of PU 0 will form a HB with the NH group of PU 3 (the solid arch), which is called as a helical HB. For a given peptide conformation, its helical content, h_{HLX}, is the ratio of the number of helical HBs of this conformation to the maximum number of helical HBs allowed for this peptide, which is 12. It represents the extent of the formation of helical structures.

Definition of Hairpin Structures. An aa conformation with (ϕ, ψ) at (−135°±45°,135°±45°) is defined as a β conformation. If an aa and its two neighbors all have β conformations, this aa is considered to be in a β-strand. h_{HP} is a score (0−1) to measure the extent of hairpin formation, which is defined as the ratio of the number of aas in β-strands to the maximum possible number (eight for hairpins, gray circles in Figure 7) if the peptide can have any hairpin topology shown in Figure 7.

Besides, the reverse turn of hairpins is defined by the backbone dihedrals of two aas *i* and *i+1*, with $(\phi, \psi)_i$ and $(\phi, \psi)_{i+1}$ at (−60°±45°,−30°±45°) and (−90°±45°,0°±45°) for Type I, (60°±45°,30°±45°) and (90°±45°,0°±45°) for Type I′, (−60°±45°,120°±45°) and (80°±45°,0°±45°) for Type II, and (60°±45°,−120°±45°) and (−80°±45°,0°±45°) for Type II′.[65] If aas *i* or *i+1* is in a helical structure, it will not be considered for turns.

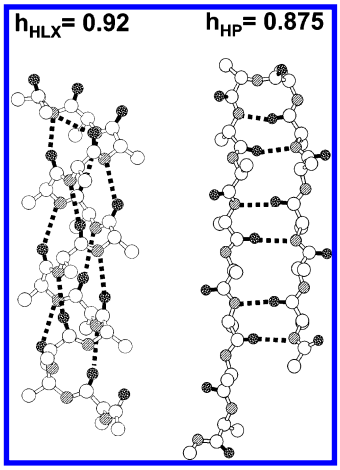

Figure 8. Typical α-helix and β-hairpin in the polyA simulation.

Table 6. Structural Properties of Peptides in Long Simulations at 310 K for PolyA and Its G Mutant

	polyA	G mutant	polyA′	G′ mutant
simulation time[a] (ns)	5000	4000	4900	5400
$<h_{HLX}>$	0.238[b]	0.072[b]	0.229[b]	0.087[b]
$<h_{HP}>$	0.007	0.024	0.007	0.032
refold (HLX)[c]	40	48	40	50
refold (HP)[c]	7	12	8	16

[a] Due to the coarse-graining, a nanosecond here does not mean the real time. [b] 0.242 for polyA and 0.053 for G mutant from AGADIR.[68] [c] Occurrence of the refolding of helices or hairpins.

Other Analyses. The extraction of the helical parameters s and σ of the Zimm-Bragg theory[66,67] is given in Appendix A. The method to calculate solvation effect is in Appendix B. The description about structural approaches is in Appendix C.

Results and Discussion

The α-Helix and β-Hairpin in Long Simulations. To examine the quality of the current CG model in reproducing structures and thermodynamics of peptides, long simulations were performed with Ac-$(Ala)_{14}$-NHMe (polyA) and Ac-$(Ala)_6$-Gly-$(Ala)_7$-NHMe (G mutant). This allowed us to examine the ability of our model to discern sequence-dependent properties of peptides.

The α-helix and β-hairpin are our major targeted peptide structures. The h_{HLX} and h_{HP} (see Methods and Models) are taken as the indicators of these structures. The typical conformations with high h_{HLX} or h_{HP} are indeed α-helix and β-hairpin (Figure 8), suggesting that these indicators are good for our purpose.

All relevant results are listed in Table 6. At 310 K, the polyA and its G mutant have 23.8% and 7.2% helices (columns "polyA" and "G mutant"). Predictions by AGADIR,[68] an accurate algorithm to predict the helical content of a special sequence based on the statistical analyses of a great number of sequences with known helical content, gave 24.2% and 5.3% for these two peptides, respectively. This agrees with the notion that Gly is a strong helix breaker.[69] As expected, β-hairpin is scarce in both peptides. Interestingly, the chance of β-hairpin formation in a G mutant is about three times that of the polyA. This implies that our model can also recognize the sequence-dependent stability of hairpins.

To examine if helices or hairpins can refold in our simulations, we monitored the change of h_{HLX} or h_{HP} with simulation time. We roughly define a refolding event of helices or hairpins as the recovery of h_{HLX} or h_{HP} to 0.5 or above after h_{HLX} or h_{HP} has been 0 for over 10 ns. This corresponds to the reformation of these structures after peptides have fully unfolded. The results in Table 6 clearly show that in these long time simulations, our CG model is capable of refolding peptides dozens of times. This guarantees statistical meaning for the calculation results of thermodynamic properties. Helices fold and refold much more frequently than do hairpins (40/48 vs 7/12), as can be expected generally. Furthermore, hairpins of the G mutant fold and refold more frequently than those of the PolyA (12 vs 7). This indicates that changing the turn sequence of a hairpin may alter the folding speed of the hairpin in our model, which coincides with the proposed zipper mechanism for hairpin folding.[70]

As demonstrated above, the turns in the middle of peptides are crucial. We therefore computed the probabilities of double aa units AG, GA, and AA in the middle of peptides to form four types of β-turns (see Methods and Models) in the long time simulations. The relative stabilities of the turns are in the following descending order: AA_I (0.0/0.0), AG_{II} (0.6/0.9), $GA_{II'}$ (0.7/1.0), AG_I (1.0/1.1), GA_I (1.0/1.1), $AA_{II'}$ (2.2/2.7), AA_{II} (2.8/1.7), $AG_{II'}$ (3.3/3.8), $AG_{I'}$ (3.8/2.1), GA_{II} (4.1/2.8), $GA_{I'}$ (4.7/2.9), and $AA_{I'}$ (6.0/3.0). The values in the parentheses are destabilization energies (kcal/mol) relative to AA_I for the formation of a given type of turn. The values before the slashes are from relative probabilities of formation of turns in simulations. The ones after the slashes are from the free energy perturbation calculations with an all-atom force field and an explicit solvent.[71] The order and magnitudes of the relative turn stabilities by the two calculation methods compare quite favorably. It is interesting that although our CG model is optimized for helices and hairpins, it can even capture many other structural features. During the simulations, 20.1% of the $GA_{II'}$ turn can occur in hairpins, which is the highest for all turn sequences. Although the AA_I turn has the highest probability to occur, only about 0.9% of the AA_I turn is found in hairpin. For polyA, the $AA_{II'}$ turn has the highest chance (6.1%) for the hairpin. This is consistent with the discovery that the type II′ turn is primarily found in hairpins.[72] Therefore, the G mutant favors a hairpin more because it can have a II′ turn with considerable stability.

To examine the effect of mass scaling on thermodynamic properties of peptides, we also carried out long simulations of the two peptides with quadrupled masses for peptide sites and with increased step size of 18 fs. These are defined as polyA′ and G′ mutant simulations, and their results are given in Table 6. They give very similar $<h_{HLX}>$ and $<h_{HP}>$ values to those of poly A and G mutant simulations. The calculated turn stabilities without (before slash) and with (after slash) mass scaling are also very similar: AA_I (0.0/0.0), AG_{II} (0.6/0.6), $GA_{II'}$ (0.7/0.5), AG_I (1.0/0.8), GA_I (1.0/

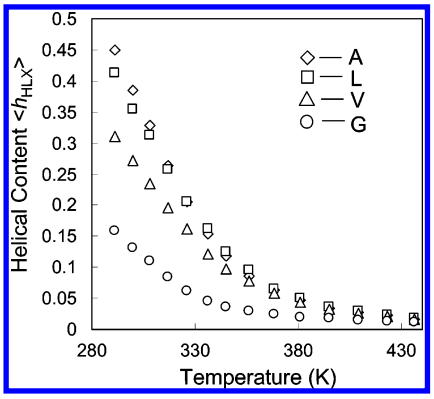

Figure 9. The helical contents of polyA and its mutants in the REMD simulations.

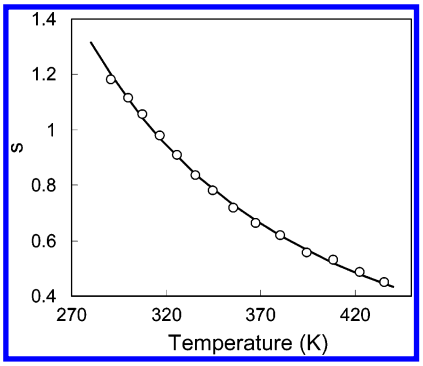

Figure 10. The fitted results for polyA. Empty circles indicate the *s* values calculated at each temperature, and the solid line is the trend line of these *s* values by eq A5.

0.9), $AA_{II'}$ (2.2/1.9), AA_{II} (2.8/3.1), $AG_{II'}$ (3.3/3.0), $AG_{I'}$ (3.8/3.8), GA_{II} (4.1/4.2), $GA_{I'}$ (4.7/4.8), and $AA_{I'}$ (6.0/5.3). These suggest that mass scaling has a negligible effect on thermodynamics of peptides. It is also interesting that the helical and hairpin structures also refold many times in these mass-scaled simulations.

Helical Propensities of Our CG Amino Acids. In the long time simulations, the polyA and its G mutant possess different $<h_{HLX}>$. This inspired us to perform a series of equilibrium REMD simulations on polyA and its Xaa mutants (X = L, V, and G). These simulations allow us to obtain helical contents of these peptides at different temperatures. As shown in Figure 9, the CG model clearly gives a sequence-dependent formation of helices for the four peptides.

In standard analyses, it is necessary to extract probabilities in helices (helical propensity, *s*) and probabilities in initiating helices (σ) (Appendix A) for different aas from these simulations in order to compare our results with the available experimental and theoretical results. The Zimm-Bragg (ZB) theory provides a way to obtain s and σ values from the average properties of a system.[66] For homopolymers like polyA, as applied by Garcia et al.[64] and Sorin et al.,[73] we used the average helical content, $<h_{HLX}>$, and the mean number of helical fragments, $<n_s>$ (eqs A2 and A4 in Appendix A), to obtain *s* and σ for Ala at each temperature. The ΔG, ΔH, and ΔS associated with *s* during the coil-helix transition of Ala were computed by fitting *s* values at different temperatures to eq A5 in Appendix A. The fitted results are shown in Figure 10. There is no experimental study of the polyA peptide. For comparison, the average helical contents of our peptides at 273−395 K were estimated by AGADIR. These data were fitted by eqs A2 and A5, assuming that $\sigma = 0.004$ which is measured by Yang et al.[74]

The fitted ΔH and $T\Delta S$ of coil-helix transition are, respectively, −1.44 and −1.34 kcal/mol at 291 K for our Ala model. These are very close to the fitted ΔH and $T\Delta S$ of Ala from AGADIR, which are −1.43 and −1.24 kcal/mol, respectively, at 291 K. Our ΔH *is also close to the values of −1.3 kcal/mol per residue for polyalanine helices measured by Scholtz et al.* with calorimetric methods.[75] Besides, the ΔC_P (-0.004 kcal·mol^{-1}·K^{-1}) of our model and the ΔC_P (0.007 kcal·mol^{-1}·K^{-1}) from AGADIR are both within the acceptable range of ± 0.008 kcal·mol^{-1}·K^{-1} for helix formation as measured by Lopez et al.[76] The helix initiation parameter, σ, of Ala in our model is 0.033 at 291 K. Such σ is larger than the *σ (0.004) obtained by Yang et al.*[74] in their CD measurements, the σ (0.004) derived from simulations with a modified AMBER-94 force field,[64] and the σ (0.007) obtained with the OPLS/AA/L force field.[34] However, our value is closer to the *σ value of 0.01−0.025 from the T-jump experiments by Thompson et al.* who intended to measure helix initiation kinetics more accurately[77] and the σ (0.027) obtained with a modified AMBER-99 force field, which showed an improved agreement of helix thermodynamics and kinetics of Fs peptide (Ac-$A_5(A_3R^+A)_3$A-NHMe) with experimental measurements.[73] Finally, we can also derive *s* and σ values for Ala from our long-time simulation of the polyA with the same procedure. They are 1.00 for *s* and 0.019 for σ at 310 K. The *s* and σ values from REMD at the same temperature are 1.02 and 0.029, respectively. Thus, the long time simulation and REMD give similar results.

Luo et al. obtained the s values of various aas (X) from the helix contents of polypeptides Ac-KA_4XA_4KGY-NH_2 at 273 K by fitting with the ZB theory.[78] Since their peptide models are similar to ours, it is desirable to compare our s_x values of aas with their experimental values. To obtain s_x values of aas other than Ala, we adopted a similar fitting procedure that was used by Luo et al.[78] and Myers et al.[79] who extracted the *s* values of single mutants in host peptides. We assumed that each Ala residue in the Xaa mutants takes the same *s* and σ values that were in polyA. The σ_x of Xaa was also assumed to be that of Ala, while its s_x was fitted by eq A3 in Appendix A.[80] The same procedures were also used to obtain the s_x values of Xaa based on the helix contents at 291 K predicted by AGADIR. The fitted s_x values (s_{sim}) are listed in Table 7. The results reveal that our *s* values for Ala, Leu, Val, and Gly at low temperature (291 K) are in good agreement with those from the AGADIR prediction (R^2=0.98), the single mutation experiments by Luo et al.[78] (R^2=0.99), and the measurements based on vast peptides with various sequences (R^2=0.97) (Table 7).[69]

Theoretically, it should be possible to derive ΔS and therefore ΔH at 291 K for aas other than Ala through the temperature dependence of s_x around 291 K. However, in practice, it is difficult because of large uncertainties of fitted s_x for Xaa when the temperature increases. This problem has

Table 7. *s* Values of Ala, Leu, Val, and Gly from Simulations, AGADIR Prediction and Experiments as Well as Linear Correlation of s_{sim} Values against Other *s* Values with $s_{other} = a \cdot s_{sim}$.

	A	L	V	G	a/R^2
s_{sim}[a]	1.18	0.84	0.30	0.02	
s_{AGADIR}[a]	1.38	0.81	0.25	0.03	1.09/0.98
$s_{exptl(A)}$[b]	1.5	1.0	0.3	0.05	1.23/0.99
$s_{exptl(B)}$[c]	1.54	0.92	0.22	0.05	1.21/0.97

[a] Derived at 291 K. [b] Measured at 273 K.[78] [c] Measured at 278 K.[69]

already been addressed by Luo et al.[78] For this reason, we used another method, as described in the following section, to estimate ΔS and ΔH of mutants at 291 K.

Structural Approaches to Helical Propensities. Although the ZB theory does not allow us to derive the ΔH and ΔS values for the coil-to-helix transition other than Ala using the current peptide models, it would still be desirable and informative to approximately calculate these components for different amino acids. This would allow us to examine the performance of the CG model in greater detail by comparing the available experimental and theoretical studies and to qualitatively analyze the factors that cause the differences of ΔG. For this purpose, we applied an approximate method, the so-called structural approach, which has been applied to decompose ΔG in other studies,[81−84] to estimate the ΔH and ΔS at 291 K.

Basically, the ΔG of various aas in our case is roughly considered as part of the free energy difference that is contributed from the aa 7 (Figure 6), between the ensemble of conformations with the aa 7 in helix (HE) and the ensemble with the aa 7 in coil (CE). The HE in our analyses is defined as the conformations with aas 5−9 in helical states, and the CE is the conformations with none of the aas 5−9 in helical states. It would be convenient to separate the overall free energy change due to the coil-to-helix transition into two parts: the free energy change of the peptide conformation, $\Delta G_{V,CE-HE,}$ and the free energy change of hydration, $\Delta G_{W,CE-HE}$:

$$\Delta G_{CE-HE} = \Delta G_{V,CE-HE} + \Delta G_{W,CE-HE} \quad (8)$$

$$\Delta G_{V,CE-HE} = \Delta\langle U_{HB}\rangle_{CE-HE} + \Delta\langle U_{vdW}\rangle_{CE-HE} + \Delta\langle U_{loc}\rangle_{CE-HE} - T\Delta S_{loc,CE-HE} \quad (9)$$

$$\Delta G_{W,CE-HE} = \Delta\langle H_W\rangle_{CE-HE} - T\Delta S_{W,CE-HE} \quad (10)$$

The enthalpy change of the peptide conformation during the coil-helix transition, $\Delta H_{V,CE-HE}$, includes three components: the local torsional energy change, $\Delta<U_{loc}>_{CE-HE}$, of the aa 7, the vdW interaction energy change $\Delta<U_{vdW}>_{CE-HE}$ between AU 7, which belongs to aa 7, and the rest of the peptide (Figure 6), and the HB interaction energy change, $\Delta<U_{HB}>_{CE-HE}$, involving both PUs 6 and 7, which surround AU 7 (Figure 6). The entropy change, $\Delta S_{loc,CE-HE}$, includes both the backbone and/or the side-chain torsional entropies of the aa 7. The details on how to calculate these quantities are given in Appendix C.

Table 8. Fitted ΔG and All Energy Components (kcal/mol) from the Structural Approaches for Simulations at 291 K

	A	L	V	G
$\Delta G_{fitting}$	−0.1	0.1	0.7	2.3
$\Delta<U_{HB}>_{CE-HE}$	−1.7	−2.0	−2.2	−2.2
$\Delta<U_{vdW}>_{CE-HE}$	−2.0	−2.5	−2.9	−2.3
$\Delta<U_{loc}>_{CE-HE}$	0.0	0.1	0.3	0.4
$-T\Delta S_{loc,CE-HE}$	1.5	1.7	1.4	1.9
$\Delta G_{V,CE-HE}$[a]	−2.2	−2.7	−3.4	−2.2
$\Delta<H_W>_{CE-HE}$	2.4	3.4	4.6	5.5
$-T\Delta S_{W,CE-HE}$	−0.3	−0.5	−0.5	−0.8
$\Delta G_{W,CE-HE}$[b]	2.1	2.9	4.1	4.7
ΔH_{CE-HE}[c]	−1.3(−1.44) [f]	−1.0	−0.2	1.4
$-T\Delta S_{CE-HE}$[d]	1.2(1.34) [f]	1.2	0.9	1.1
ΔG_{CE-HE}[e]	−0.1	0.2	0.7	2.5

[a] $\Delta G_{V,CE-HE} = \Delta<U_{HB}>_{CE-HE} + \Delta<U_{vdW}>_{CE-HE} + \Delta<U_{loc}>_{CE-HE} - T\Delta S_{loc,CE-HE}$. [b] $\Delta G_{W,CE-HE} = \Delta<H_W>_{CE-HE} - T\Delta S_{W,CE-HE}$. [c] $\Delta H_{CE-HE} = \Delta<U_{HB}>_{CE-HE} + \Delta<U_{vdW}>_{CE-HE} + \Delta<U_{loc}>_{CE-HE} + \Delta<H_W>_{CE-HE}$. [d] $\Delta S_{CE-HE} = \Delta S_{loc,CE-HE} + \Delta S_{W,CE-HE}$. [e] $\Delta G_{CE-HE} = \Delta H_{CE-HE} - T\Delta S_{CE-HE}$. [f] Values in parentheses are from fitting with the ZB theory.

To evaluate the applicability of the structural approach to the current peptide systems, we first compare the thermodynamic properties of the polyA, which have been derived from fitting with the ZB theory and by other experimental and theoretical methods. As shown in the first column of Table 8, the enthalpy difference, ΔH_{CE-HE} (−1.3 kcal/mol), and the entropy difference, $-T\Delta S_{CE-HE}$ (1.2 kcal/mol), between HE and CE are comparable to ΔH (−1.44 kcal/mol) and $-T\Delta S$ (1.34 kcal/mol) derived by the fitting with the ZB theory. The resulting ΔG_{CE-HE} for Ala agrees well with the fitted ΔG and the ΔG derived by other methods.[69,78,64,34,73] The structural approach gives a hydration enthalpy of about 2.4 kcal/mol. This is in good agreement with a recent calculation by Avbelj with the finite difference Poisson−Boltzmann (PB) method, which gives a range of 1.6−3.2 kcal/mol for solvation change to transfer an aa from β-strands to the middle of a helix.[85] In addition, the unfavorable $-T\Delta S_{CE-HE}$ in our model is mainly due to the loss of local conformational entropy $-T\Delta S_{loc,CE-HE}$ (1.5 kcal/mol), consistent with the estimated values of 1.5 kcal/mol by Wang et al.[81] and 1.37 kcal/mol by D'Aquino et al.[86]

For other aas, the structural approach derives a free energy change, ΔG_{CE-HE} of 0.2, 0.7, and 2.5 kcal/mol for the coil-to-helix transition of Leu, Val, and Gly, respectively, of our model. These are in good agreement with the values deduced from the fitting with the ZB theory ($\Delta G_{fitting}$ in Table 8), which has already been shown to agree with experimental results well (Table 7). Thus, the structural approach is also able to estimate the helix propensities of these aas well: Leu and Val have lower helical propensities than Ala, while Gly has a much lower helical propensity. The structural approach also indicates that the entropic contributions ($-T\Delta S_{CE-HE}$) of the four amino acids are quite similar, ranging in 1.2−0.9 kcal/mol, with Val slightly more favorable, and the relative helix propensities are mainly determined by the enthalpy change (ΔH_{CE-HE}) of the coil-helix transition, with Leu, Val, and Gly being less favorable by about 0.3, 1.1, and 2.7 kcal/mol with respect to Ala, respectively. This suggests that in our model, the enthalpy factor $\Delta\Delta H_{CE-HE}$

is the major determinant for the helical propensities ($\Delta\Delta G_{CE-HE}$) of different aas, while the entropy factor plays a minor role.

It has been suggested that the entropy loss from the local conformation may determine the difference in helical propensities of different aas.[87] The previous calculations showed that the relative local entropy losses ($-T\Delta\Delta S_{loc,CE-HE}$) with respect to Ala are about −0.06, −0.18, and 0.72 kcal/mol for Leu, Val, and Gly, respectively.[86,87] These are close to our calculation results of 0.2, −0.1, and 0.4 kcal/mol for Leu, Val, and Gly, respectively. However, as pointed out by Luo et al.,[78] the magnitude of this relative local entropy loss only accounts for less than one-third of $\Delta\Delta G$ obtained from their experiments on polyalanine-based peptides. They suggested that enthalpy should play a major role. The structural approach indicates that our model agrees better with the experiments by Luo et al.[78]

The structural approach allows the analysis of various factors that contribute to the free energy change in the coil-to-helix transition for our polyalanine-based model. As shown in Table 8, the coil-to-helix transition is favored by the formation of hydrogen bonds in the HE ($\Delta\langle U_{HB}\rangle_{CE-HE}$ = −1.7−2.2 kcal/mol) and the vdW interaction ($\Delta\langle U_{vdW}\rangle_{CE-HE}$ = −2.0−2.5 kcal/mol) between the aa 7 and the rest of of the peptide. It is disfavored by a solvation enthalpy change ($\Delta\langle H_w\rangle_{CE-HE}$ = 2.4−5.5 kcal/mol) due to a poorer solvation of the helix, and the loss of local conformational entropy ($-T\Delta S_{loc,CE-HE}$ = 1.4−1.9 kcal/mol). The change in local conformational enthalpy ($\Delta\langle U_{loc}\rangle_{CE-HE}$ = 0.0−0.4 kcal/mol) and hydration entropy ($-T\Delta S_{W,CE-HE}$ = −0.3−0.8 kcal/mol) is less significant. Overall, the coil-to-helix transition is favored by the free energy change of the peptide conformation ($\Delta G_{V,CE-HE}$ = −2.2 to −3.4 kcal/mol), with Leu and Val more favorable, but it is disfavored by the free energy change of hydration or solvation ($\Delta G_{W,CE-HE}$ = 2.1−4.7 kcal/mol). The lower helical propensities of Leu, Val, and Gly with respect to Ala are mainly caused by relative poorer hydration of the helical structures than the coil structures. In particular, the relative difference of hydration enthalpy, $\Delta\Delta\langle H_W\rangle_{CE-HE}$, in reference to Ala, which is the sum[88] of the relative difference of hydration enthalpy for PUs 4−9 and AUs 5−9 (Figure 6) between the mutant peptides and polyA (Appendix C), contributes about 1.0, 2.2, and 3.1 kcal/mol for Leu, Val, and Gly, respectively. This result is in agreement with the experimental observations by Luo et al.,[78] who suggest that hydration enthalpy plays a major role in determining helical propensities of different aas in polyalanine based-host/guest systems.

To further reveal the origin of the hydration effect on helical propensity, we analyzed the detailed solvation contributions from each of PUs 4−9 and AUs 5−9 to $\Delta\Delta\langle H_W\rangle_{CE-HE}$. Since the solute−solvent interaction $\langle U_{int}\rangle$ accounts for the major part of solvation enthalpy of our peptide model (Table 5), we calculated the average interaction energy $\langle U_{int}\rangle$ between each of these PUs or AUs and the solvent for HE or CE as well as the corresponding $\Delta\langle U_{int}\rangle_{CE-HE}$ during the coil-to-helix transition (Figure 11).

As shown in Figure 11a, $\Delta\langle U_{int}\rangle_{CE-HE}$ of nonpolar AUs is much smaller than that of polar PUs for all aas. The

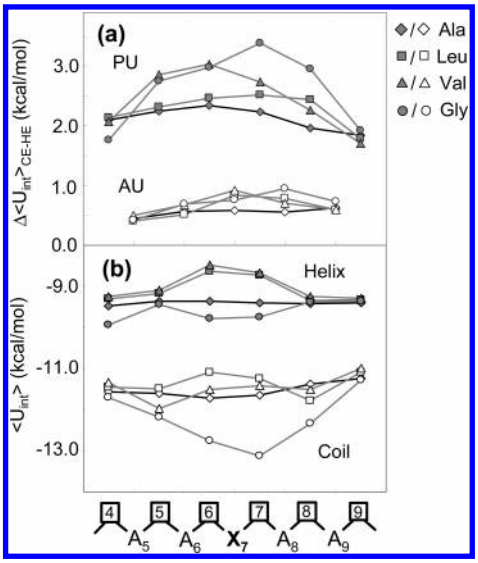

Figure 11. (a) The solvation energy change $\Delta\langle U_{int}\rangle_{CE-HE}$ of coil-helix transition for PUs 4−9 (top) and AUs 5−9 (bottom). (b) The solvation energy of PUs 4−9 for the helix ensemble (top) and coil ensemble (bottom). Ala (diamond); Leu (square); Val (triangle); Gly (circle).

difference in $\Delta\langle U_{int}\rangle_{CE-HE}$ between Ala and its mutant is also smaller for AUs than for PUs. Therefore, the hydration of polar PUs is important for different helical propensities of aas in our model. In addition, a significant change of $\Delta\langle U_{int}\rangle_{CE-HE}$ not only occurs at the site of mutation (PUs 6 and 7) but also at the neighboring aas (PUs 5 and 8, Figure 11a), in accord with previous calculation results.[88]

Figure 11b shows the solute−solvent interaction energy for each of the PUs in the HE ($\langle U_{int}\rangle_{HE}$) and CE ($\langle U_{int}\rangle_{CE}$), from which $\Delta\langle U_{int}\rangle_{CE-HE}$ can be derived. It shows that in the HE, the side chains of Val and Leu shield the solvent-PU interaction that is available to Ala. The side chain of Val shields slightly more than does the side chain of Leu. The PU of the Gly mutant is somewhat better hydrated than the PUs of Ala. The difference in $\langle U_{int}\rangle_{HE}$ among our aas is expected since the more particles attached to the C_β, the less accessible the polar backbone is, as suggested by Makhatadze[89] and by Avbelj et al.,[88] and Gly has no C_β carbon at all, allowing a greater exposure of PUs. In the CE, while its solvation is apparently better than the HE for all four aas, Leu and Val have similar solvation as Ala. On the other hand, Gly is found to have a much better solvation for its PUs. Thus, the structural approach reveals that in our model, Leu and Val have lower helical propensities than Ala mainly because the larger side chains shield the solvation of helical structures, while Gly has a much poorer helical propensity mainly due to the much better solvation for the coil structures than for the helical structures.

Simulation Speed-Up. Computational speed-up of the current CG model compared to the all-atom model comes from two reasons.

As was pointed out by Marrink et al.[22] and Shih et al.,[26] the most important reason is that the four-water-one-particle mapping greatly reduces the number of interaction sites, and the number of interaction pairs between the interaction sites therefore decreases further. To demonstrate this, we per-

formed a CG simulation of a polyA peptide in about 1000 CG solvent particles. We also carried out an all-atom simulation of a polyA peptide in about 4000 all-atom water molecules, which corresponds to the CG simulation. Both the simulations have a time interval of 2 fs and a cutoff of 0.12 nm. With the same computational power and in the same amount of time, the CG simulation samples 65-time more steps than the all-atom simulation. In addition, our CG model can use a step size of 18 fs, which gives a further speed-up by several folds. As a result, the CG model has a 200−300-fold speed-up compared to the all-atom model in explicit solvent.

The second reason is that the time scale in the CG simulation could be increased by the coarse-graining, which has been pointed out in other studies.[19,90,91] Marrink et al. found that the time scale of the CG water model, which is used in our CG model, is increased by four folds based on the comparison of the self-diffusion coefficients of water from their CG simulations and from all-atom simulations and experiments.[22] Interestingly, this time scale factor is found in all the dynamics in their CG water and CG lipid system.[22] Since most heavy atoms of our peptide model are explicitly represented, this time scale factor may only be applied to the diffusive motions in peptides.

For the above two reasons, our CG model may be about 10^3 faster than the all-atom model in the best situation where peptide motions are controlled by diffusion and may be 10^2 faster in the worst situation where local motions are dominant in peptides.

Summaries and Conclusions

We have constructed a CG protein model (for Gly, Aly, Leu, and Val) at an intermediate level coupled with the CG solvent model developed by Marrink et al.[22] A systematic method has been used to optimize parameters for protein potentials and protein−solvent interactions. The optimized CG model can fold polyalanine and its mutants into both helix and hairpin conformations without biased potentials. The calculated stabilities and dynamics of the peptides are sequence-dependent and compare very favorably with available experimental data. In particular, the helical propensities of Ala, Leu, Val, and Gly calculated by the CG model are very close to experimental values. Structural analysis indicates that the helical-forming propensities of different amino acids are mainly determined by solvation effects. Although some fine-tuning is still needed, we expect that a full development of this coarse-grained protein model for the remaining residue side chains will provide a promising tool for the study of fast folding of small proteins in aqueous solutions and in membrane environments.

Acknowledgment. We are grateful to RGCHK (HKUST6083/02M, N-HKUST 623/04) and NSFC (20225312) for financial support of the research.

Appendix A

The seminal work by Zimm and Bragg[66] (ZB) developed two parameters, σ and s, to describe helices. In their work, σ is defined as the probability of two peptide units to initiate a helical turn, and s is the probability of a peptide unit in the helical HB. In this study, we define s as the probability of an aa in the helical structures, where this aa and its two neighboring aas are all in helical conformations, and define σ as the probability of two aas at two ends of a helical sequence to initiate helices. These probabilities are relative to the coil structures. Such definition of s and σ should be equivalent to the original one as suggested by Schellman et al.[67]

As shown in Figure 6, if aas 1−4 are in helical conformations, aas 2 and 3 are in helical structures and aas 1 and 4 are at two ends. The probability of such structure is σs^2. If Xaa is mutated from Ala to another aas, it will contribute s_x and σ_x. If Xaa is at either end of a helical sequence, it contributes the weight of σ_x. The weight of a specific peptide structure can be $\sigma^i s^j$ for the polyalanine and $\sigma^i s^j \sigma_x^k s_x^n$ for its mutants. The weighted sum of all possible structures for the polyalanine (Q) and its mutants (Q') reads

$$Q = \sum_{i=0}^{3}\sum_{j=0}^{12} C_{ij}\sigma^i s^j$$

$$Q' = \sum_{i=0}^{3}\sum_{j=0}^{11}\sum_{k=0}^{1}\sum_{n=0}^{1} C_{\mathrm{i,j,k,n}}\sigma^i s^j \sigma_x{}^k s_x{}^n \qquad \text{(A1)}$$

where C is the number of structures with the same weight. It can be obtained by computer enumeration. "3" and "12" indicate the maximum numbers of helical sequences and helical HBs in a given structure according to the ZB theory.[66] Two important properties, the average helical content $<h_{\mathrm{HLX}}>$ or $<h_{\mathrm{HLX}}'>$ and the average helical fragment $<n_s>$, can read

$$<h_{\mathrm{HLX}}> = \frac{1}{12Q}\sum_{i=0}^{3}\sum_{j=0}^{12} jC_{\mathrm{i,j}}\sigma^i s^j \qquad \text{(A2)}$$

$$<h_{\mathrm{HLX}'}> = \frac{1}{12Q'}\sum_{i=0}^{3}\sum_{j=0}^{11}\sum_{k=0}^{1}\sum_{n=0}^{1} (j+n)C_{\mathrm{i,j,k,n}}\sigma^i s^j \sigma_x{}^k s_x{}^n \qquad \text{(A3)}$$

$$<n_s> = \frac{1}{Q}\sum_{i=0}^{3}\sum_{j=0}^{11} iC_{i,j}\sigma^i s^j \qquad \text{(A4)}$$

These equations are used to obtain σ, s, σ_x, and s_x by fitting the results from simulations. The relevant ΔH, ΔS, and ΔC_p can be obtained from s at different temperatures by eq A5.

$$\begin{aligned}\Delta G(T) &= -RT\,1\mathrm{n}\,s \\ &= \Delta H(T) - T\Delta S(T) \\ &= \Delta H_0 + \Delta C_p(T - T_0) - T\Delta S_0 - \Delta C_p T\,1\mathrm{n}\left(\frac{T}{T_0}\right)\end{aligned} \qquad \text{(A5)}$$

In eq A5, ΔH_0 and ΔS_0 are the changes of enthalpy and entropy at the reference temperature ($T_0 = 291$ K).

Appendix B

The Thermal Integration (TI) Approach. The solvation free energy ΔG_{sov} is defined as the free energy difference

between the state where the solute is immersed in solvent and the state where the solute is isolated from solvent. The TI method can calculate the free energy difference between two states, for instance, A and B by introducing a coupling parameter (λ). As λ gradually varies from zero to unity, state A transforms to state B. During this process, $\Delta G_{A\rightarrow B}$ can be calculated as[50]

$$\Delta G_{A-B} = G_B - G_A = \int_{\lambda=0}^{\lambda=1} d\lambda \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \qquad \text{(A6)}$$

where U(λ) is the total energy when the system is in the intermediate state λ. In our calculation of ΔG_{sov}, the interaction between solute and solvent is gradually switched off during the state transformation. In addition, a soft-core Lennard-Jones potential is applied to avoid the singularity problem when λ is close to unity or zero.[92]

To calculate ΔS, and therefore ΔH, we apply the finite difference as Smith et al. [51]

$$-\Delta S_{A-B} \approx \frac{\Delta G_{A-B}(T+\Delta T) - \Delta G_{A-B}(T-\Delta T)}{2\Delta T} \qquad \text{(A7)}$$

where $\Delta T = 5$ K in our calculation. Each simulation for the TI totally lasts for 40 ns, which is long enough for accurate calculation of both ΔG_{sov} and ΔS_{sov}.[52]

The Approach Based on the Scaled Particle Theory (SPT). The solvation free energy of a sphere in CGW can be divided into two parts:[53] (1) the work ΔG_{cav} to create a cavity and (2) the free energy change ΔG_{int} by turning on solvent−solute interactions. Since the solvent molecule is essentially Lennard-Jones sphere, ΔG_{cav} can be computed from the scaled particle theory (SPT)[54,55]

$$\Delta G_{cav} = K_0 + K_1\alpha_{12} + K_2{\alpha_{12}}^2 + K_2{\alpha_{12}}^3 \qquad \text{(A8)}$$

with

$$K_0 = RT\left\{-1n(1-y) + \frac{9}{2}[y/(1-y)]^2\right\} - (\pi P {a_1}^3)/6$$

$$K_1 = -(RT/a_1)\{6y/(1-y) + 18[y/(1-y)]^2\} + \pi P {a_1}^2$$

$$K_2 = (RT/{a_1}^2)\{12y/(1-y) + 18[y/(1-y)]^2\} - 2\pi P a_1$$

$$K_3 = \left(\frac{4}{3}\right)\pi P \qquad \text{(A9)}$$

where R is the gas constant, T is the temperature, P is the pressure, and $y = (\pi {a_1}^3 n_s)/6$ with n_s the number density of solvent molecules. The a_1 and a_2 are diameters of solvent and solute molecules, and the $a_{12} = (a_1+a_2)/2$ is the diameter of cavity. As suggested by Pierotii et al.,[56] the a_1 and a_2 can effectively be vdW diameters. ΔH_{cav} can read[56]

$$\Delta H_{cav} = \alpha_P RT^2[y/(1-y)]\left\{[6/(1-y)][2(a_{12}/a_1)^2 - (a_{12}/a_1)] - [36y/(1-y)^2]\left[(a_{12}/a_1)^2 - (a_{12}/a_1) + \frac{1}{4}\right] + 1\right\} \qquad \text{(A10)}$$

where α_P is thermal expansion coefficient of the CG water. α_P(291 K) is calculated as 0.00089 K^{-1} by finite difference between two simulations at 290 and 292 K with $P = 1$ atm. ΔG_{int} can be approximately $<U_{int}>$, the average interaction energy between solvent and solute, if $<U_{int}>$ comes from the vdW interaction.[53] Since we have

$$\begin{aligned}\Delta G_{sov} &= \Delta G_{cav} + \Delta G_{int} \\ &\approx \Delta H_{cav} - T\Delta S_{cav} + \langle U_{int}\rangle \\ &= \Delta H_{sov} - T\Delta S_{sov}\end{aligned} \qquad \text{(A11)}$$

therefore, $\Delta H_{sol} \approx \Delta H_{cav} + <U_{int}>$ and $\Delta S_{sov} \approx \Delta S_{cav}$. According to Claverie et al.,[57] ΔG_{cav} or ΔH_{cav} of the solute with complex shape can be roughly estimated by

$$\Delta G_{cav} = \sum_i \frac{A_i}{4\pi {a_{1i}}^2}\Delta G_{cav}(a_{1i})$$

$$\Delta H_{cav} = \sum_i \frac{A_i}{4\pi {a_{1i}}^2}\Delta H_{cav}(a_{1i}) \qquad \text{(A12)}$$

where a_{1i} is the cavity diameter of the ith particle composing the solute, and A_i is its accessible surface area.

Appendix C

In order to estimate the free energy difference ΔG between the conformations with Xaa (the aa 7 in our peptide model) in helical states and the conformations with Xaa in coil states, we define helical ensemble (HE) and coil ensemble (CE) for Xaa as peptide conformations where aas 4−10 are in helical conformations ((ϕ, ψ) in ($-60°\pm30°$,$-47°\pm30°$)), and where no three aas can be in helical conformations for aas 4−10, respectively. By such definition, the aa 7 is embedded in the middle of a helical stretch, and each of the PUs 6 and 7 (Figure 6) that connect to the aa 7 involves in two HB interactions. The ΔG is therefore roughly estimated by $\Delta G_{CE-HE} = G_{HE} - G_{CE}$, which is the contribution from the aa 7 to the free energy difference between the HE and the CE. The enthalpy part of $\Delta G_{CE-HE} = \Delta H_{CE-HE} - T\Delta S_{CE-HE}$ can be calculated as

$$\begin{aligned}\Delta H_{CE-HE} &= H_{HE} - H_{CE} \\ &\approx U_{HE} - U_{CE} \\ &= \langle U_V(r) + H_W(r)\rangle_{HE} - \langle U_V(r) + H_W(r)\rangle_{CE}\end{aligned} \qquad \text{(A13)}$$

where the last equality is from Wang et al.[81] $U_v(\boldsymbol{r})$ is the conformational energy of the Xaa mutant in conformation $\boldsymbol{r}$, and $H_w(\boldsymbol{r})$ is the solvation enthalpy of Xaa of this conformation. "$\langle\rangle_{HE/CE}$" indicates the average over HE or CE. For the conformational energy in the HE or CE

$$\langle U_v(r)\rangle = \langle U_{loc}(r)\rangle + \langle U_{vdW}(r)\rangle + \langle U_{HB}\rangle \qquad \text{(A14)}$$

$U_{loc}(\boldsymbol{r})$ is the local torsional energy of Xaa. $U_{vdW}(\boldsymbol{r})$ is half of the vdW energy between Xaa and other parts of the protein. Because AU 7 belongs to aa 7 and both PUs 6 and 7 can be considered as parts of aa 7 (Figure 6 and Methods and Models), for Ala in polyA, $U_{HB}(\boldsymbol{r})$ reads

$$\langle U_{HB}(r)\rangle = \langle {}^1/_2[\phi(6) + \phi(7)]\rangle \quad \text{(A15)}$$

where $\phi(i)$ is half of the HB energy of PU i with the other part of the protein. For the Xaa mutant, we assume that changes of $\phi(6)$ and $\phi(7)$ are all induced by the mutation of Ala into Xaa. The difference of $U_{HB}(\boldsymbol{r})$ between Xaa and Ala is calculated as

$$\Delta\langle U_{HB}(r)\rangle^{Ala-Xaa} = \langle U_{HB}(r)\rangle^{Xaa} - \langle U_{HB}(r)\rangle^{Ala}$$
$$= \langle \phi(6) + \phi(7)\rangle^{Xaa} - \langle \phi(6) + \phi(7)\rangle^{Ala} \quad \text{(A16)}$$

The $U_{HB}(\boldsymbol{r})$ for Xaa can therefore be derived from eq A16.

The hydration enthalpy $H_w(\boldsymbol{r})$ was calculated with the SPT based method with the SAS assumption (Appendix B). For Ala in polyA, $H_w(\boldsymbol{r})$ is computed by

$$\langle H_W(r)\rangle = \langle {}^1/_2[H_W^{PU}(6) + H_W^{PU}(7)] + H_W^{AU}(7)\rangle \quad \text{(A17)}$$

where $H_W{}^{PU}(i)$ and $H_W{}^{AU}(j)$ are the solvation enthalpy of PU i and AU j, respectively. In the case that Ala is mutated into Xaa, we follow the suggestion by Avbelj et al.[88] that the mutation induces not only the solvation change of aa 7 but also the solvation change of its several neighboring aas. Therefore, PUs 4−9 and AUs 5−9 are considered in the calculation of solvation enthalpy. The difference of $H_W(\boldsymbol{r})$ between Xaa and Ala reads

$$\Delta\langle H_W\rangle^{Ala-Xaa} = \langle H_W(r)\rangle^{Xaa} - \langle H_W(r)\rangle^{Ala}$$
$$= \langle \sum_{i=4}^{9} H_W^{PU}(i) + \sum_{j=5}^{9} H_W^{AU}(j)\rangle^{Xaa}$$
$$- \langle \sum_{i=4}^{9} H_W^{PU}(i) + \sum_{j=5}^{9} H_W^{AU}(j)\rangle^{Ala} \quad \text{(A18)}$$

The entropy part ΔS_{CE-HE} of ΔG_{CE-HE} is estimated by

$$\Delta S_{CE-HE} \approx \Delta S_{W,CE-HE} + \Delta S_{loc,CE-HE}$$
$$\approx \Delta S_{W,CE-HE} + \Delta S_{bb,CE-HE} + \Delta S_{sc,CE-HE} \quad \text{(A19)}$$

$\Delta S_{W,CE-HE}$ is solvation entropy difference of S_W between the HE and the CE. It can be obtained in the same way as described in eqs A17 and A18 as well as in Appendix B. $\Delta S_{loc,CE-HE}$ is the local conformational entropy change including backbone $\Delta S_{bb,CE-HE}$ and side-chain $\Delta S_{sc,CE-HE}$ entropy loss. The backbone entropy in the HE or the CE is calculated by dividing (ϕ, ψ) of Xaa into 36 × 36 states and summing $-R\Sigma[p_i \ln(p_i)]$, where p_i is the probability of state i. The side-chain entropy is similarly obtained by dividing χ of Xaa (Figure 1a) into 36 states.

Finally, in the calculation of solvation of peptides, we have increased both ΔH_{cav} and ΔS_{cav} from the SPT methods by 1-fold as a rough correction, which has been demonstrated in Models and Methods. Without the correction, ΔG_{CE-HE} (kcal/mol) for Ala is 0.3, and the relative $\Delta\Delta G_{CE-HE}$ of the mutants in reference to Ala is 0.5 for Leu, 1.0 for Val, and 3.1 for Gly. Compared to the corresponding values in Table 8, the removal of the correction should not make a qualitative difference from the results with correction and the results from fitting.

Supporting Information Available: Simulated structural information of valine and leucine dipeptides. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dobson, C. M. *Nature* **2003**, *426*, 884-890

(2) Hartl, F. U.; Hayer-Hartl, M. *Science* **2002**, *295*, 1851−1858.

(3) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430.

(4) Schueler-Furman, O.; Wang, C.; Bradley, P.; Misura, K.; Baker, D. *Science* **2005**, *310*, 638−642.

(5) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.

(6) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.

(7) Burton, R. E.; Huang, M. A.; Daugherty, M. A.; Fullbright, P. W.; Oas, T. G. *J. Mol. Biol.* **1996**, *263*, 311.

(8) Ding, F.; Dokholyan, N. V. *TRENDS Biotechnol*. **2005**, *23*, 450−455.

(9) Nymeyer, H.; García, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5921.

(10) Gō, N. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183.

(11) Abe, H.; Gō, N. *Biopolymers* **1980**, *20*, 1013.

(12) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860.

(13) Thirumalai, D.; Guo, Z. *Biopolymers* **1995**, *35*, 137.

(14) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **1999**, *110*, 11616−11629.

(15) Ding, F.; Borreguero, J. M.; Buldyrey, S. V.; Stanley, H. E.; Dokholyan, N. V. *Proteins: Struct., Funct., Genet.* **2003**, *53*, 220−228.

(16) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J.* **2005**, *88*, 147−155.

(17) Smith, A. V.; Hall, K. C. *J. Mol. Biol. Proteins: Struct., Funct., Genet.* **2001**, *44*, 344−360.

(18) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shaknovich, E. I. *Fold Des.* **1998**, *3*, 577−587.

(19) Sharma, S.; Ding, F.; Dokholyan, N. V. *Biophys. J.* **2007**, *92*, 1457−1470.

(20) Honig, B.; Yang, A. *Adv. Protein Chem.* **1995**, *46*, 27.

(21) Shelley, J. C.; Shelley, M.; Reeder, R.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464.

(22) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750−760.

(23) de Vries, A. H.; Mark, A. E.; Marirnk, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 4488−4489.

(24) Kasson, M. P.; Kelly, N. W.; Singhal, N.; Vrljic, M.; Brunger, A. T.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11916−11921.

(25) Bond, P. J.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2006**, *128*, 2697.

(26) Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K. *J. Phys. Chem. B* **2006**, *110*, 3674−3684.

(27) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.

(28) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

(29) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(30) Regan, L.; DeGrado, W. F. *Science* **1988**, *241*, 976.

(31) Neidigh, J. W.; Fesinmeyer, R. M.; Anderson, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425−430.

(32) Marsh, R. E.; Donohue, J. *Adv. Protein Chem.* **1967**, *22*, 249.

(33) Kaminski, G.; Friesner, R.; Tirado-Rives, J.; Jorgensen, W. *J. Phy. Chem. B* **2001**, *105*, 6474−6487.

(34) Gnanakaran, S.; Garcia, A. E. *Proteins* **2005**, *59*, 773−782.

(35) Eker, F.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2002**, *124*, 14330−14341.

(36) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P. *Biomolecular simulation: the GROMOS96 manual and user guide*; Hchschulverlag AG an der ETH: Zurich, 1996.

(37) Williams, D. E.; Craycroft, D. J. *J. Phys. Chem.* **1987**, *91*, 6365−6373.

(38) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. *J. Mol. Biol.* **1999**, *290*, 253−266.

(39) Nosé, S. *Mol. Phys.* **1984**, *52*, 255−268.

(40) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.

(41) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(42) Ben-Naim, A. *Solvation Thermodynamics*; Plenum Press: New York, 1987.

(43) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 72nd ed.; CRC Press: Boca Raton, FL, 1992.

(44) Douglass, D. C.; Mccall, D. W. *J. Phys. Chem.* **1958**, *62*, 1102.

(45) Kresheck, G. C.; Schneider, H.; Scheraga, H. A. *J. Phys. Chem.* **1965**, *69*, 3132−3144.

(46) Black, C.; Joris, G. G.; Taylor, H. S. *J. Chem. Phys.* **1948**, *16*, 537.

(47) Wolfenden, R. *Biochemistry* **1978**, *17*, 201−204.

(48) Wolfenden, R.; Anderson, L.; Cullis, P. M.; Southgate, C. C. *Biochemistry* **1981**, *20*, 849−855.

(49) Della Gatta, G. D.; Barone, G.; Elia, V. *J. Solution Chem.* **1986**, *15*, 157−167.

(50) Mezei, M.; Beveridge, D. L. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 1.

(51) Smith, D. E.; Haymet, A. D. J. *J. Chem. Phys.* **1993**, *98*, 6445−6454.

(52) Wan, S.-Z.; Stote, R. H.; Karplus, M. *J. Chem. Phys.* **2004**, *121*, 9539−9548.

(53) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(54) Reiss, H.; Frisch, H. L.; Lebowitz, J. L. *J. Chem. Phys.* **1959**, *31*, 369.

(55) Reiss, H.; Frisch, H. L.; Helfand, E.; Lebowitz, J. L. *J. Chem. Phys.* **1960**, *32*, 119.

(56) Pierotti, R. A. *J. Phys. Chem.* **1965**, *69*, 281−288.

(57) Claverie, P. In *Intermolecular Interactions: from Diatomics to Biomolecules*; Pullman, B., Ed.; J. Wiley: Chichester, 1978.

(58) Graziano, G. *J. Phys. Soc. Jpn.* **2000**, *69*, 3720−3725.

(59) Pohorille, A.; Pratt, L. R. *J. Am. Chem. Soc.* **1990**, *112*, 5066.

(60) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91* 43−56.

(61) Berendsen, H. J. C.; Pstma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(62) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(63) Okabe, T.; Kawata, M.; Okamoto, Y.; Mikami, M. *Chem. Phys. Lett.* **2001**, *335*, 435−439.

(64) García, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782−2787.

(65) Hutchinson, G.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 2207−2216.

(66) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526−535.

(67) Qian, H.; Schellman, J. A. *J. Phys. Chem.* **1992**, *96*, 3987−3994.

(68) Munoz, V.; Serrano, L. *Nature Struct. Biol.* **1994**, *1*, 399−409.

(69) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 843−852.

(70) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.

(71) Yan, Y. B.; Erickson, B. W.; Tropsha, A. *J. Am. Chem. Soc.* **1995**, *117*, 7592−7599.

(72) Sibanda, B. L.; Thornton, J. M. *Nature* **1985**, *316*, 170−174.

(73) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472−2493.

(74) Yang, J.; Zhao, K.; Gong, Y.; Vologodskii, A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **1998**, *120*, 10646−10647.

(75) Scholtz, J. M.; Marqusee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Santoro, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2854−2858.

(76) Lopez, M. M.; Chin, D. H.; Baldwin, R. L.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1298−1302.

(77) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200−9210.

(78) Luo, P. Z.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4930−4935.

(79) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2833−2837.

(80) In the fitting for Gly, a s_x value of −0.07 was obtained assuming $\sigma_x = 0.033$. The negative s_x value is mathematically possible but physically meaningless. It indicates that the σ_x of 0.033 is too large for Gly. We found that $s_x < 0.02$ is required for $\sigma_x > 0$ and $\sigma_x < 0.0045$ is required for $s_x > 0$. However, for Val and Leu, variation of σ_x from 0.01 to 0.1 has little influence on the fitted s_x. Thus, the assumption of $\sigma_x = 0.033$ should be valid in the cases of Val and Leu.

(81) Wang, J.; Purisima, E. O. *J. Am. Chem. Soc.* **1996**, *118*, 995−1001.

(82) Hermans, J.; Anderson, A. G.; Yun, R. H. *Biochemistry* **1992**, *31*, 5646−5653.

(83) Luque, I.; Mayorga, O. L.; Freire, E. *Biochemistry* **1996**, *35*, 13681−13688.

(84) Yang, A. S.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 351−365.

(85) Avdelj, F. *J. Mol. Biol.* **2000**, *300*, 1335−1359.

(86) D'Aquino, J. A.; Gómez, J.; Hilser, V. J.; Lee, K. H.; Amzel, L. M.; Freire, E. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 143−156.

(87) Creamer, T. P.; Rose, G. D. *Proteins: Struct., Funct., Genet.* **1994**, *19*, 85−97.

(88) Avbelj, F.; Luo, P. Z.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10786−10791.

(89) Makhatadze, G. I. *Adv. Protein Chem.* **2006**, *72*, 199−226.

(90) Nielson, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *J. Phys. Condens. Matter*. **2004**, *16*, R481−R512.

(91) Zhou, Y.; Karplus, M. *J. Mol. Biol.* **1999**, *293*, 917−951.

(92) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Greber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529−539.

CT700151X