Supplementary Materials

Geometry of the Protein Model: To model hydrogen bonding more accurately, we add the oxygen atoms into the backbone of the original four-bead model. For the amino acids that are neither beta-branched nor bulky, the gamma beads are positioned at in the geometrical center of the group of all heavy atoms of the sidechain except C_β (see Figure 1 in the main text). The two C_γ beads of the β-branched amino acids are centered in the geometrical center of the two groups of heavy atoms forming the branches. For Lys and Arg, the effective C_γ beads are located in the position of the actual C_δ atom. The effective C_δ bead of Lys is located in position of the charged N_ζ atom. Similarly, the effective C_δ bead of Arg plays the role of the positive charge center and coincides with the actual C_ζ atom. For Trp, the C_γ bead is centered in the five-atom ring and the C_δ bead is centered in the six-atom benzene ring.

In order to model the bond lengths and bond angles, we introduce constraints between the neighboring beads. We list the parameters for bonded pairs in Table S1. We model the non-bonded interactions by assigning stepwise potentials between pairs. Each bead is modeled as an interacting soft ball with a hardcore radius and its interaction range. To assess the hardcore radius and interaction range for various sidechain beads, we make statistical evaluation of the available crystal structures from protein databank (PDB). First, we define the existence of a contact between two effective sidechain beads, if any two atoms from the two groups of actual sidechain heavy atoms which the two effective beads represent are within 4.5Å from each other. Next, we calculate the distributions of distances between two effective sidechain beads that are in contact. From this distribution, we estimate the corresponding hardcore radius, *HC*, and the interaction range, *IR*, which are also listed in Table S1.

Dihedral angles: Since the model contains up to three effective sidechain beads for the amino acids, we are able to model the sidechain dihedral angles χ_1 and χ_2. It is well known that the rotamers have limited freedom of rotation. We model the behavior of rotamers by introducing effective bonds between the C' and the effective γ_1 bead for χ_1 and between C_α and the effective δ bead for χ_2, with the following potential,

$$U_{1,4} = \begin{cases} +\infty, & d < d_{\min}; d > d_{\max} \\ \varepsilon_\chi, & d_{\min} \le d < d_0; d_1 \le d < d_2 \\ 0, & d_0 \le d < d_1; d_2 \le d < d_{\max} \end{cases}, \qquad \text{(S.1)}$$

where $d_{min} < d_0 < d_1 < d_2 < d_{max}$ (Figure S1). As it is demonstrated in the schematic diagram of Figure S1A, the values of d_0 and d_1 and d_2 determine the distribution of correct rotamer angles. We calculate the distributions of distances between the effective gamma bead and C' for different amino acids by sampling over thousands of crystal structures from PDB. For instance, we present in Figure S1C the distribution for valine. The parameters related to the constraints for different residues are listed in Table S2. In Figure S1D, we show the distribution of the χ_1 angles for an unfolded poly-valine peptide from DMD simulations. In our model, the gamma and/or delta beads are coarse-grained atoms and if the gamma and/or delta beads for a certain amino acid are very flexible the corresponding χ_1 and χ_2 angles have no well defined values in the frame of current model. Therefore, in this model we do not assign any constraints to confine the rotamer angles for the amino acids with flexible effective gamma and/or delta beads: Arg, Glu, Gln, Lys, and Met. Trp residue contains a well-defined C_δ bead and we introduce a similar constraint between the C_α and the C_δ bead to model χ_2 (see Table S2).

Proline is a special imino acid because its sidechain is linked by a covalent bond with its backbone amide. Therefore, its distribution of the χ_1 angle differs from such distributions for

other amino acids. We assign for proline a covalent bond between gamma bead and its backbone nitrogen bead with an average distance as 1.80Å and the allowed fluctuations of ±0.09Å. Covalently connected to its backbone, proline also has unusual Φ angle distributions (Figure S1E). We introduce a constraint between the prime carbon of previous residue and the beta carbon of proline residue with the distance of 3.63±0.05Å. In Figure S1F, we present the distribution of the dihedral angles of proline from a DMD simulation of poly-proline. The experimental and simulated distributions are in agreement with each other.

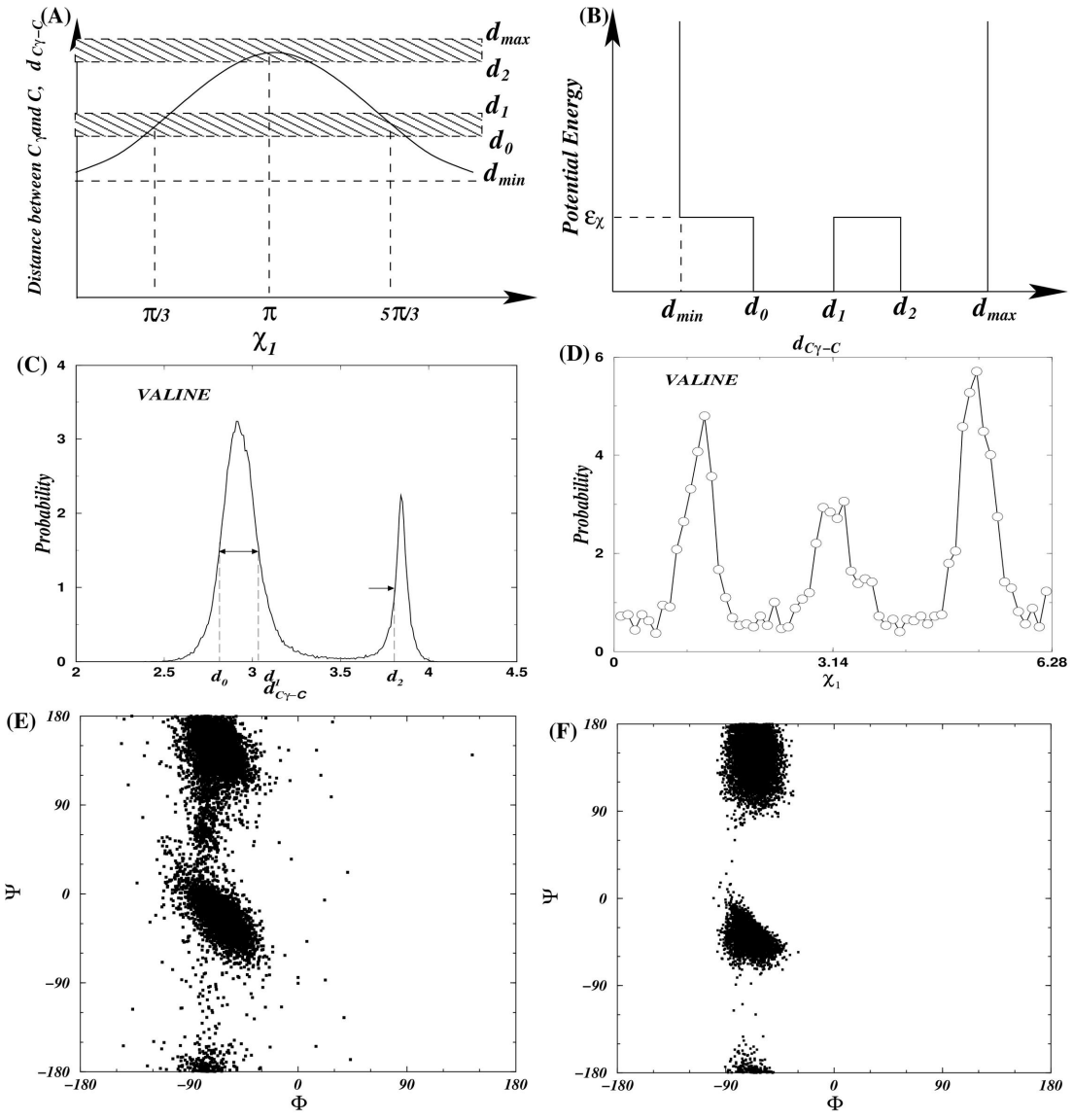

Figure S1: The schematic diagram for the χ_1 constraint. (A) The distance between C_γ and C' beads is drawn as the function of rotamer angle χ_1. The shaded regions correspond to the allowed rotamer angle regions around π/3, π, and 5π/3. (B) The introduced potential between the C_γ and C' beads. (C) The probability distribution of the distance, $d_{cc'}$ for valine, which is calculated from available PDB structures. (D) The probability of χ_1 angles from DMD simulation of unfolded poly-valine. The Ramachandran plot of proline from (E) various crystal structures from PDB, and from (F) the DMD simulations of a poly-proline peptide.

Table S1: We denote the distance of the covalent bonds between beta and gamma beads as $d_{\beta\gamma}$, and the distances of the auxiliary bonds between alpha and gamma beads as $d_{\alpha\gamma}$.We denote hardcore radii as HC. We also denote the interaction range as IR. For the second gamma beads for Thr, Val, Ile, we denote the distance between beta and gamma2 beads as $d_{\beta\gamma2}$, the distance between alpha and gamma2 beads as $d_{\alpha\gamma2}$, the distance between gamma and gamma2 beads as $d_{\gamma\gamma2}$. The hardcore radius and the interaction radius are denoted as $HC_{\gamma2}$ and IRγ2. For the bulky amino acids, Arg, Lys and Trp, we introduce a delta bead. For the distance constraints, we allow a variance of ±2% unless it is specified in the table.

Residue	Gamma Bead				Gamma2 Bead					Delta Bead			
	$d_{\beta\gamma}$,Å	$d_{\alpha\gamma}$,Å	HC_{γ},Å	IR_{γ},Å	$d_{\beta\gamma2}$,Å	$d_{\alpha\gamma2}$,Å	$d_{\gamma\gamma2}$,Å	$HC_{\gamma2}$,Å	$IR_{\gamma2}$,Å	$d_{\gamma\delta}$,Å	$d_{\beta\delta}$,Å	HC_{δ},Å	IR_{δ},Å
CYS	1.83	2.80	1.70	2.25									
MET	2.76	3.71±0.46	1.85	2.90									
PHE	2.91	3.79	2.00	3.20									
ILE	1.52	2.52	1.65	2.25	1.94	2.87	2.85	1.65	2.95				
LEU	1.94	3.04	2.00	3.00									
VAL	1.52	2.50	1.65	2.20	1.52	2.50	2.49	1.65	2.20				
TRP	2.69	3.60	1.90	2.90						2.15	4.00	2.00	3.20
TYR	3.30	4.16	2.00	3.20									
THR	1.52	2.49	1.65	2.25	1.43	2.41	2.42	1.35	2.20				
SER	1.45	2.43	1.35	2.25									
GLN	2.47	3.40±0.38	1.85	2.90									
ASN	1.94	2.88	1.75	2.70									
GLU	2.47	3.40±0.38	1.85	2.90									
ASP	1.94	2.88	1.75	2.70									
HIS	2.65	3.55	1.90	2.90									
ARG	2.51	3.12±0.38	1.65	2.75						2.47	4.30	1.85	2.90
LYS	3.40	3.12±0.38	1.65	2.75						2.51	4.55	1.50	2.75
PRO	1.83	2.28	1.65	2.60									

Table S2 The parameters of the rotamer constraints: d_0, d_1 and d_2. The parameters d_{min} and d_{max} is not sensitive (Figure 1a). Therefore, we assign 2.0Å and 6.0Å for d_{min} and d_{max}, respectively. For Trp the constraint to model χ_2 is between Cα and the delta beads.

Residue	d_0,Å	d_1,Å	d_2,Å
CYS	3.00	3.30	4.10
PHE	3.70	4.18	5.12
ILE	2.80	3.05	3.79
LEU	3.28	3.55	4.25
VAL	2.80	3.05	3.79
TRP	3.62	4.07	4.89
TYR	4.00	4.54	5.47
THR	2.80	3.05	3.79
SER	2.68	3.06	3.68
ASN	3.12	3.40	4.16
ASP	3.12	3.40	4.16
HIS	3.57	4.05	4.87
TRP(C_δ)	4.56	4.90	5.30

Table S3: The types of various sidechain beads. The available types are hydrophobic (H), amphipathic (A), aromatic (AR), neutral polar (P), positive charge (PC), and negative charged (NC).

Residue	C_β	Gamma bead	Gamma2 bead	Delta bead
CYS	A	H		
MET	A	H		
PHE	A	H,AR		
ILE	A	A	H	
LEU	A	H		
VAL	H	A	A	
TRP	A	A		H,AR
TYR	A	A,AR		
ALA	A			
GLY				
THR	P	A	P	
SER	P	P		
GLN	A	P		
ASN	P	P		
GLU	A	NC		
ASP	P	NC		
HIS	P	P		
ARG	A	A		PC
LYS	A	A		PC
PRO	P	A		

Table S4: Hydrogen bonding interaction parameters.

	Pair	d_{min}^{HB}, Å	d_{max}^{HB}, Å
Backbone	N_i, O_j	2.80	3.12
	N_i, C_j	3.80	4.23
	$C_{\alpha i}$, O_j	3.60	4.04
	C_{i-1}, O_j	3.60	4.00
Thr(HBA)	$Gamma2_i$, N_j	2.87	3.27
	$Gamma2_i$, $C_{\alpha j}$	3.64	4.08
	$Gamma2_i$, C_{j-1}	3.77	4.23
Ser(HBA)	$Gamma_i$, N_j	2.87	3.27
	$Gamma_i$, $C_{\alpha j}$	3.64	4.08
	$Gamma_i$, C_{j-1}	3.77	4.23
ASN(HBA)	$Gamma_i$, N_j	3.52	4.04
	$Gamma_i$, $C_{\alpha j}$	4.08	4.76
	$Gamma_i$, C_{j-1}	4.42	4.94
ASP(HBA)	$Gamma_i$, N_j	3.52	4.04
	$Gamma_i$, $C_{\alpha j}$	4.08	4.76
	$Gamma_i$, C_{j-1}	4.42	4.94
ASN(HBD)	$Gamma_i$, O_j	3.29	3.59
	Gammai, C_j	3.16	4.00
GLN(HBD)	$Gamma_i$, O_j	3.50	4.06
	Gammai, C_j	4.35	4.99
SER(HBD)	$Gamma_i$, O_j	2.60	3.00
	Gammai, C_j	3.53	4.13
THR(HBD)	$Gamma2_i$, O_j	2.60	3.00
	Gamma2i, C_j	3.53	4.13