

Table S1

Table 1: $Q \geq 55\%$

Φ	Δ	Γ^2_{unf}	Γ^2_{nat}
-1.5	-0.4687	2.2827	4.3452
-1.4	-0.3711	2.2862	4.3439
-1.3	-0.2729	2.2888	4.3429
-1.2	-0.1742	2.2906	4.3422
-1.1	-0.0748	2.2916	4.3418
-1.0	+0.0249	2.2918	4.3418
-0.9	+0.1253	2.2912	4.3420
-0.8	+0.2262	2.2899	4.3425
-0.7	+0.3277	2.2878	4.3432
-0.6	+0.4296	2.2849	4.3442
-0.5	+0.5320	2.2813	4.3455
-0.4	+0.6349	2.2770	4.3470
-0.3	+0.7383	2.2720	4.3487
-0.2	+0.8420	2.2664	4.3505
-0.1	+0.9462	2.2601	4.3526
+0.0	+1.0508	2.2531	4.3548
+0.1	+1.1557	2.2456	4.3571
+0.2	+1.2610	2.2374	4.3595
+0.3	+1.3666	2.2287	4.3620
+0.4	+1.4725	2.2195	4.3646
+0.5	+1.5787	2.2096	4.3672
+0.6	+1.6852	2.1993	4.3698
+0.7	+1.7919	2.1884	4.3724
+0.8	+1.8989	2.1770	4.3749
+0.9	+2.0061	2.1651	4.3774
+1.0	+2.1136	2.1526	4.3798

Table S1. Values of Φ , Δ , Γ^2_{unf} , Γ^2_{nat} in the range of the unfolding transition for $Q \geq 55\%$.

Table S2

Table 2: $Q \geq 75\%$

Φ	Δ	Γ^2_{unf}	Γ^2_{nat}
-1.5	-1.9583	4.1051	3.1884
-1.4	-1.8425	4.1283	3.2433
-1.3	-1.7275	4.1496	3.2946
-1.2	-1.6133	4.1692	3.3425
-1.1	-1.5000	4.1871	3.3870
-1.0	-1.3877	4.2034	3.4280
-0.9	-1.2762	4.2182	3.4658
-0.8	-1.1656	4.2315	3.5002
-0.7	-1.0559	4.2434	3.5314
-0.6	-0.9472	4.2539	3.5594
-0.5	-0.8394	4.2632	3.5843
-0.4	-0.7325	4.2712	3.6061
-0.3	-0.6265	4.2780	3.6248
-0.2	-0.5215	4.2836	3.6406
-0.1	-0.4173	4.2882	3.6534
+0.0	-0.3141	4.2917	3.6634
+0.1	-0.2118	4.2941	3.6705
+0.2	-0.1104	4.2956	3.6748
+0.3	-0.0098	4.2962	3.6764
+0.4	+0.0897	4.2958	3.6753
+0.5	+0.1885	4.2946	3.6717
+0.6	+0.2864	4.2925	3.6654
+0.7	+0.3834	4.2897	3.6567
+0.8	+0.4796	4.2861	3.6455
+0.9	+0.5750	4.2819	3.6319
+1.0	+0.6695	4.2769	3.6159

Table S2. Values of Φ , Δ , Γ^2_{unf} , Γ^2_{nat} in the range of the unfolding transition for $Q \geq 75\%$.

Table S3

Table 3: $Q \geq 85\%$

Φ	Δ	Γ^2_{unf}	Γ^2_{nat}
-1.5	-2.4126	4.3531	2.5297
-1.4	-2.2999	4.3536	2.5556
-1.3	-2.1877	4.3535	2.5800
-1.2	-2.0759	4.3530	2.6029
-1.1	-1.9647	4.3521	2.6245
-1.0	-1.8539	4.3509	2.6450
-0.9	-1.7435	4.3496	2.6643
-0.8	-1.6337	4.3482	2.6826
-0.7	-1.5243	4.3467	2.6999
-0.6	-1.4154	4.3452	2.7161
-0.5	-1.3071	4.3437	2.7313
-0.4	-1.1993	4.3423	2.7455
-0.3	-1.0920	4.3409	2.7586
-0.2	-0.9853	4.3396	2.7707
-0.1	-0.8792	4.3383	2.7817
+0.0	-0.7737	4.3372	2.7915
+0.1	-0.6689	4.3362	2.8002
+0.2	-0.5647	4.3353	2.8077
+0.3	-0.4611	4.3345	2.8141
+0.4	-0.3583	4.3339	2.8192
+0.5	-0.2561	4.3334	2.8230
+0.6	-0.1546	4.3331	2.8257
+0.7	-0.0539	4.3329	2.8270
+0.8	+0.0461	4.3329	2.8270
+0.9	+0.1454	4.3331	2.8258
+1.0	+0.2439	4.3334	2.8232

Table S3. Values of Φ , Δ , Γ^2_{unf} , Γ^2_{nat} in the range of the unfolding transition for $Q \geq 85\%$.

Figure S1

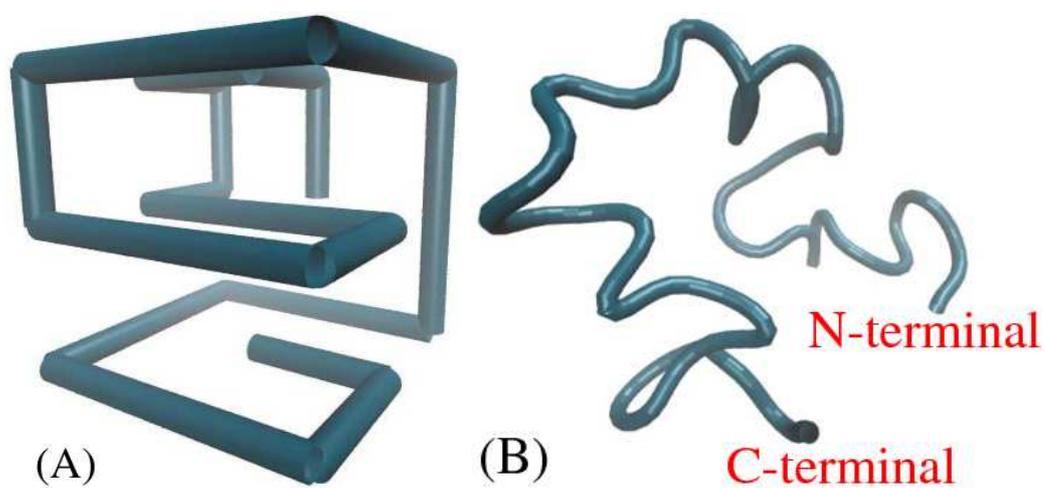


Figure S1. Backbone representation of misfolded (A) lattice and (B) real protein structures.

C_α model of Real Proteins

Figure S2

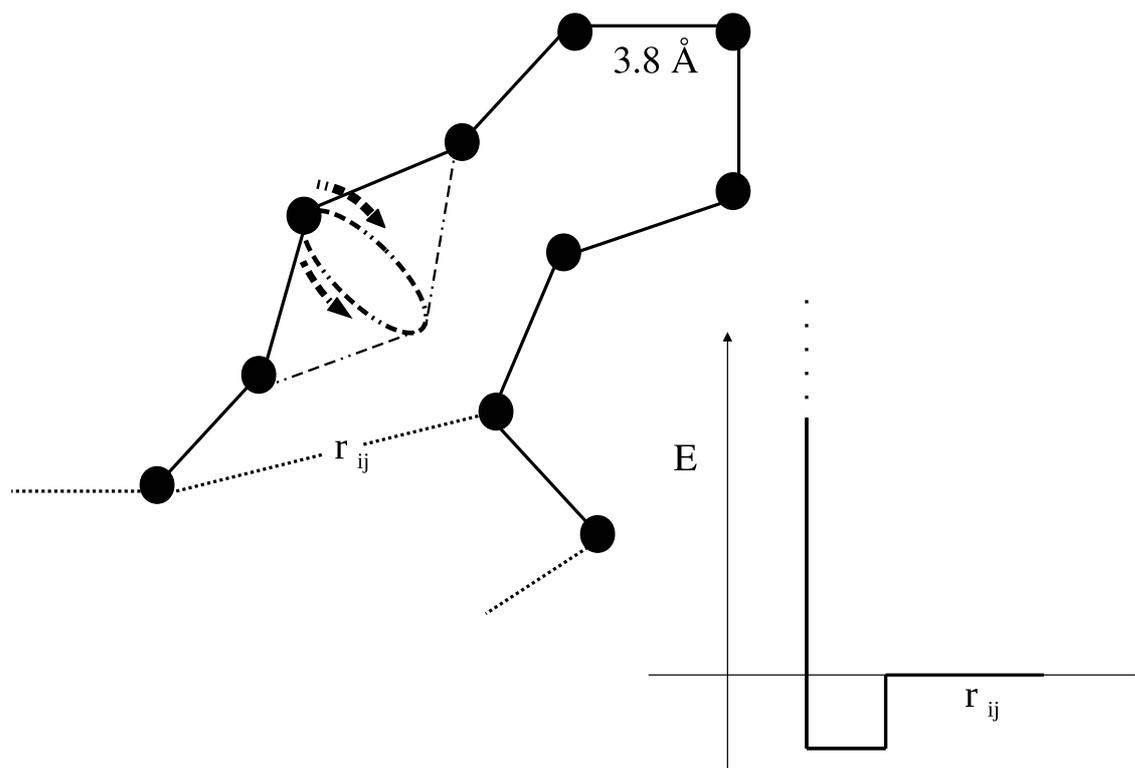


Figure S2. Sketch of the real protein backbone conformation represented by C_α coordinates (black circles) kept at distance 3.8 ± 0.15 Å apart. The arrows represent the MC updates as rotation above or below the plane of the paper.

Each MC update to build the self-avoiding protein conformations is associated with the conformational stability measured by a simple square-well potential,

$$E_{ij}^{conf} = \begin{cases} \infty & r_{ij} \leq 3.8 \text{ \AA}, \forall i \neq j \pm 1, \\ -\epsilon & 3.8 < r_{ij} \leq 7.5 \text{ \AA}, \forall i \neq j \pm 1. \end{cases} \quad (1)$$

where r_{ij} is the distance between i -th and j -th residue. At local level the repulsive term prevents unphysical sharp bents in the protein backbone while at the non-local level it mimics self-avoidance by constraining no two nonadjacent C_α residues are allowed to be at a distance closer than 3.8 Å. The attractive term stabilizes a conformation locally and has little effect to form a globally folded structure.

Figure S3

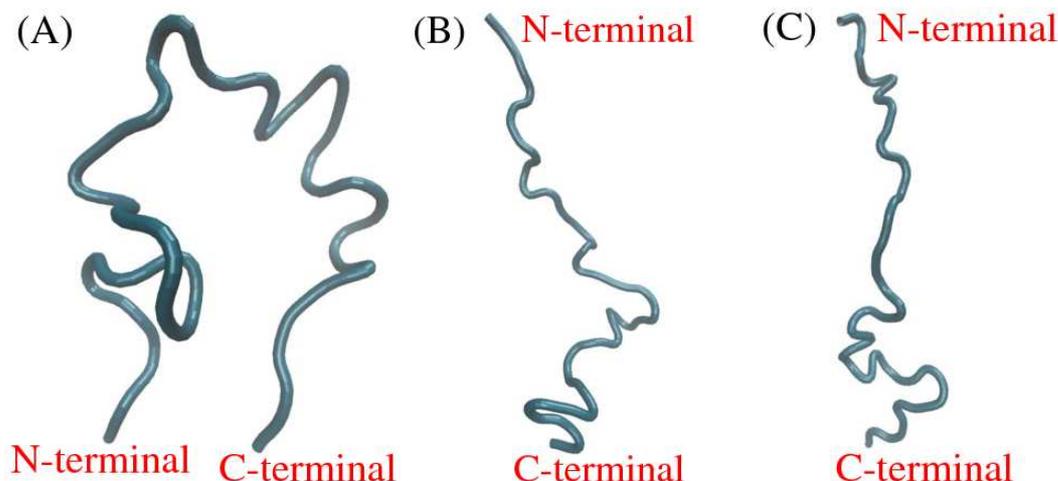


Figure S3. Some snapshots of the generated unfolded conformations with varying R_g .

To validate the generated unfolded ensemble two statistical quantities are observed: (a) average radius of gyration and (b) end-to-end distance. The geometric radius of gyration R_g for a chain is given by the following equation:

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{r}_i - \vec{r}_c)^2} \quad (2)$$

where N is the number of atoms in the protein structure, \vec{r}_i is the position of residue i in three-dimensional space given by C_α coordinates, and \vec{r}_c is the geometric center of the molecule. The ensemble-averaged radius of gyration was computed simply by averaging R_g over all chains in the ensemble and found to be 12.5 Å.

The mean squared end-to-end distance, $\langle L^2 \rangle$, is given by,

$$\langle L^2 \rangle = \frac{1}{n} \sum_{j=1}^n L_j^2 \quad (3)$$

where n is the number of conformations in the ensemble, and L_j is the end-to-end distance of conformation j , taken from the N-terminal C_α to the C-terminal C_α . For the generated ensemble, $\langle L^2 \rangle$ is found to be 836.4 Å².

The calculated values for present ensemble are then compared to the results predicted¹ from Flory's scaling law,

$$R_g = R_0 N^\nu \quad (4)$$

where R_0 is a constant related to persistence length and ν is the scaling factor depends on solvent quality. A study based on most reliable small-angle x-ray scattering (SAXS)² on a series of 28 unfolded proteins³ suggests $R_0 = 2.08 \pm 0.19$ Å and $\nu = 0.598 \pm 0.029$.

Figure S4

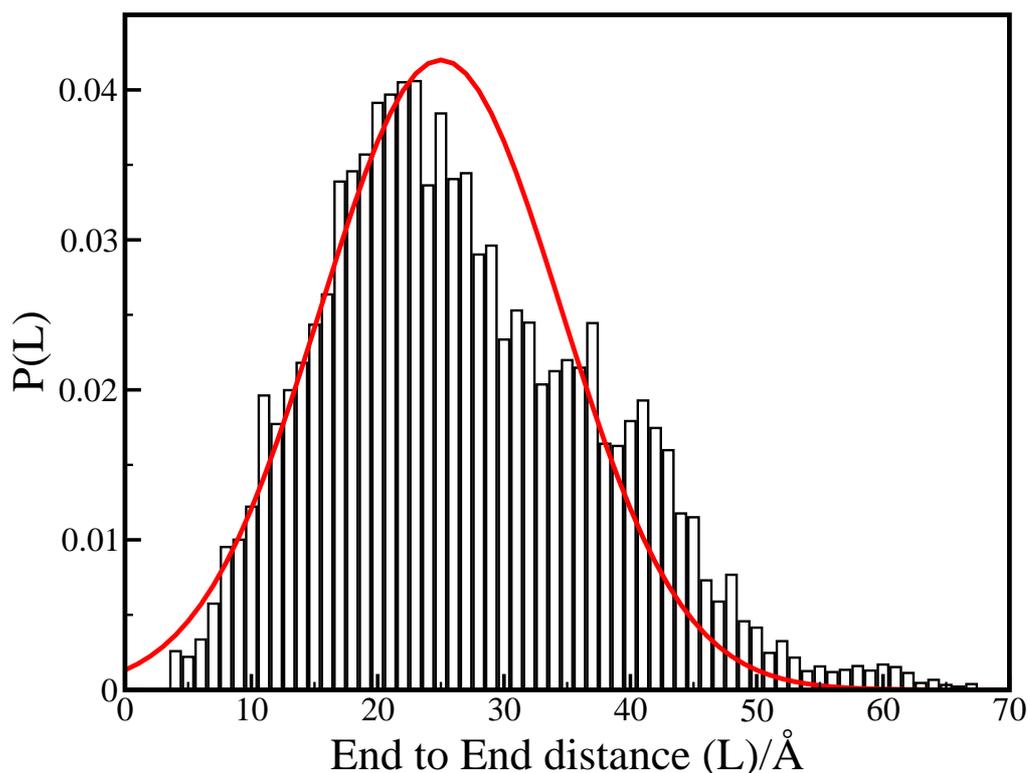


Figure S4. End-to-end distance histogram for generated unfolded ensemble (6182 conformations). Chains were grouped into 1 Å bins based on the distance from the N-terminal C_{α} to the C-terminal C_{α} . $P(L)$ is the number of chains normalized with respect to total number of chains. The solid line shows a Gaussian curve for comparison with same mean and Standard Deviation as the actual distribution has.

It may be noted that our values are bit low compared to the predicted values as the generated conformations are relatively more semi-compact (see Eq(1)) instead of completely extended chains. This is advantageous to the present situation as unlike extended chains, semi-compact structures are more significant to study misfolding. Importantly, the generated unfolded ensemble shows a Gaussian distribution(see Fig S4)) for end-to-end distances which is consistent with the behavior for random coils⁴, and some recent simulations⁵.

Figure S5

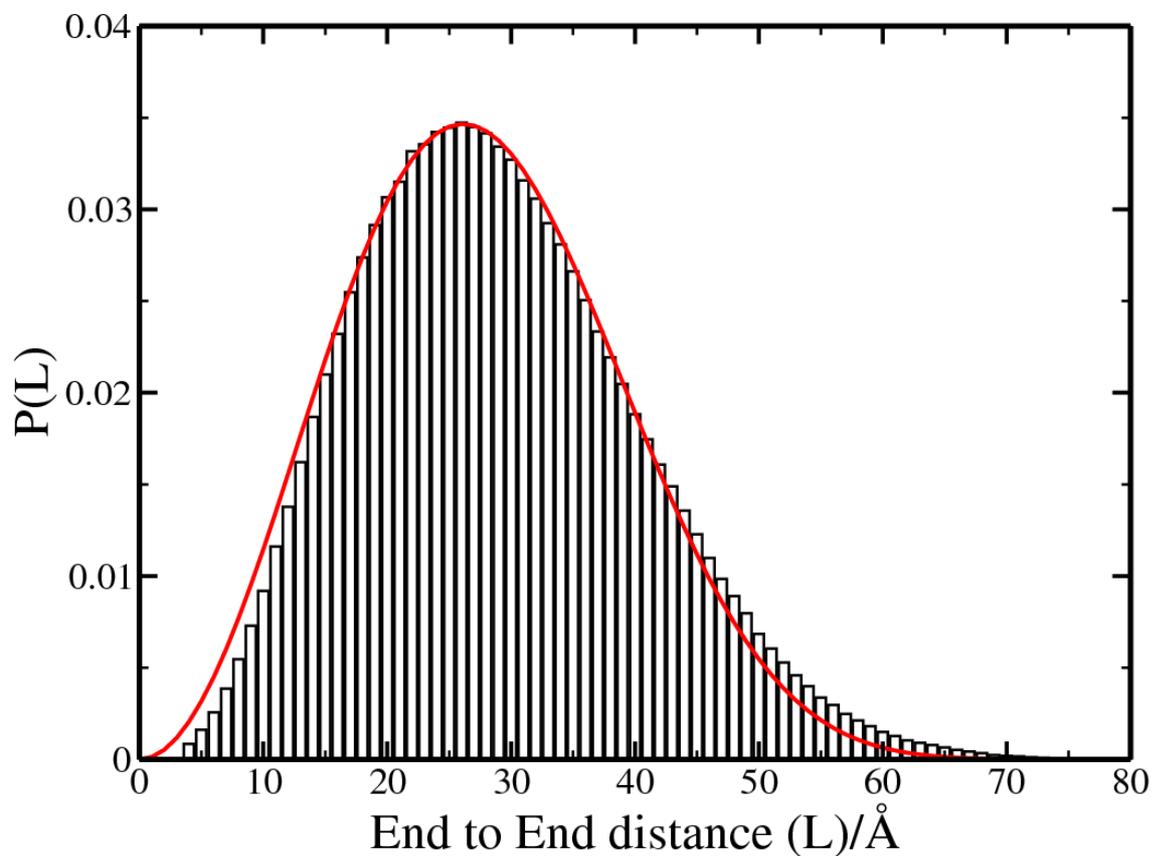


Figure S5. End-to-end distance histogram for generated unfolded ensemble (151977 conformations). Chains were grouped into 1 Å bins based on the distance from the N-terminal C_α to the C-terminal C_α . $P(L)$ is the number of chains normalized with respect to total number of chains. The solid line shows a Gaussian curve for comparison with same mean and Standard Deviation as the actual distribution has.

Figure S6

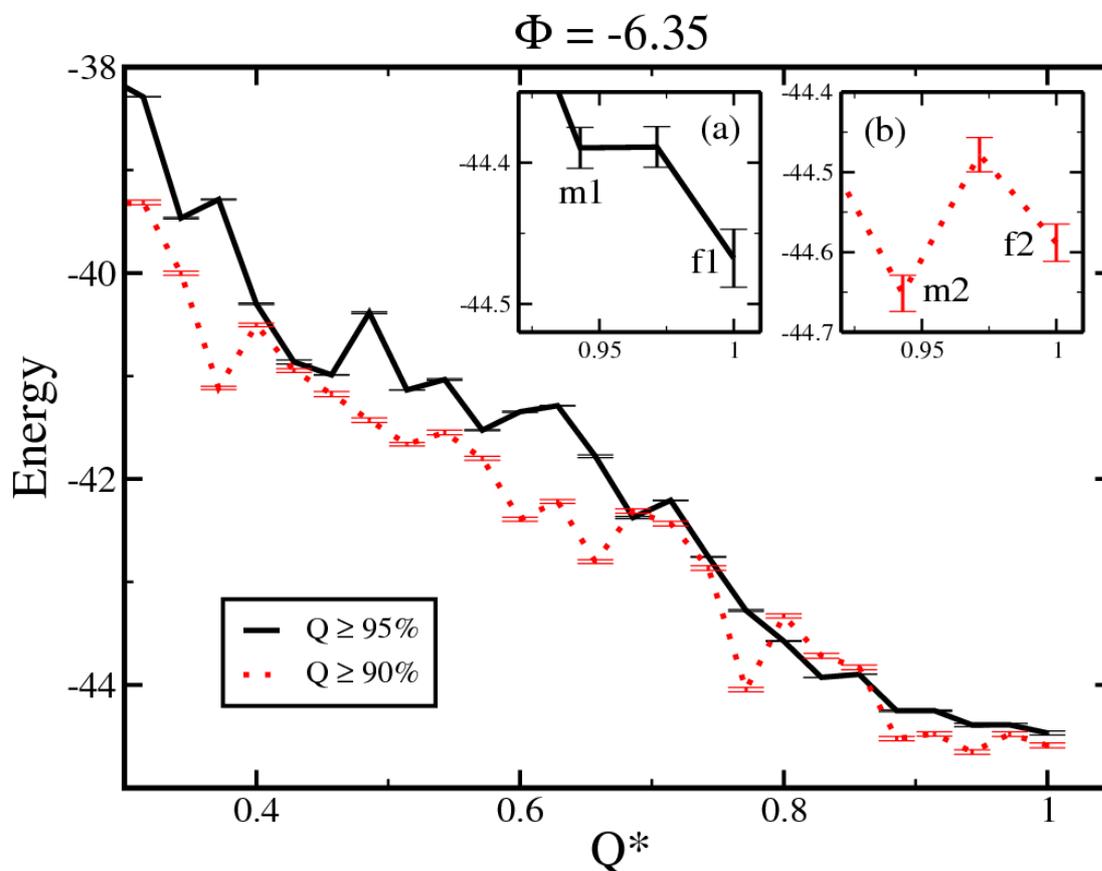


Figure S6. Plot of energies of real protein conformations (for the larger sample with 151977 conformations) vs the fraction of native structural context, Q^* for $\phi = -6.35$. In inset (a) and (b) the energy gap between the native structure and the most stable misfolded structure is magnified for the folding and the misfolding sequences respectively.

1. P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, NY), 1979, 43.
2. I. S. Millett, S. Doniach and K. W. Plaxco, *Adv. Prot. Chem.*, 2002, **62**, 241-262.
3. J. E. Kohn et al., *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 12491-12496.
4. H. S. Chan, and K. A. Dill, *Annu. Rev. Biophys. Biophys. Chem.*, 1991, **20**, 447-490.
5. D. P. Goldenberg, *J. Mol. Biol.*, 2003, **326**, 1615-1633.