Supplementary Information: *In silico* design of misfolding resistant proteins: the role of structural aspect of competing conformational ensemble for optimization of frustration

Bondeepa Saikia and Anupaul Baruah[a)]
Department of Chemistry, Dibrugarh University, Dibrugarh-786004, India

(Dated: 27 February 2024)

The expressions for the stability gaps; Δ_1, Δ_2, Δ_3 and Δ_4 can be given as

$$\Delta_1 = E_f - < En_1 > \quad (S1)$$
$$\Delta_2 = E_f - < En_2 > \quad (S2)$$
$$\Delta_3 = E_f - < En_3 > \quad (S3)$$
$$\Delta_4 = E_f - < En_4 > \quad (S4)$$

Here, all terms have the same meaning as described in the manuscript.

For Eq. 19, there is a set of 2732 transcendental equations which determines the Lagrange undetermined multipliers. The variational functionals, V_1, is maximized with respect to 2600 (= 130×20) site identity monomer probabilities, $\omega_i(\alpha)$. This gives a set of 2600 nonlinear equations. Thus, equation 1 to 2600 are differential equations and the general form of this is expressed below,

$$\frac{\partial V_1}{\partial \omega_i(\alpha)} = \frac{\partial}{\partial \omega_i(\alpha)}(S - \lambda_{norm} f_{norm} - \lambda_1 \Delta_1 - \lambda_2 \Delta_2)$$

There are 130 normalization equations (equation 2601 to 2730) and can be obtained from Eq. 17. The general form of this is expressed below,

$$-1 + \sum_{\alpha=1}^{m} \omega_i(\alpha) = 0$$

Equations 2731 and 2732 are for foldability criteria Δ_1 and Δ_2 respectively. Thus, there are a set of 2732 (=2600+130+1+1) nonlinear transcendental equations for Eq. 19. Similarly, for Eq. 20 to Eq. 24 there is a set of 2732 transcendental equations each.

DEGREE OF FRUSTRATION IN DESIGNED SEQUENCES

Presence of frustration is an important aspect of protein folding and stability. The sections 3.3.3 and 3.3.4 provide insightful information about frustration in the designed sequences. Here, the Monte Carlo designed sequences corresponding to each method are analyzed for their degree of frustration at the target conformation based on the developed potential. It is already stated in the manuscript that the developed potential has specific potential values for each of the 20 amino acids at each of the six structural contexts, k=1, 2, 3, 4, 5, and 6. For the calculation of degree of frustration, first for an amino acid, α at i^{th} site in a designed sequence, the structural context of that site in the target structure is determined. Next, the potential values of an amino acid, α are ranked from least to highest across all structural contexts. Based on this ranking, the degree of frustration are assigned to the amino acid, α at i^{th} site of the target structure (refer to Table S1). Then, structural context of amino acid, α at i^{th} site of the target is matched with its structural context in the potential and the corresponding degree of frustration is attributed. This is repeated for each amino acid in the designed sequence. Finally the total degree of frustration for each designed sequence is calculated by adding all these frustration values. The following equations are used to calculate degree of frustration, $F_s^i(\alpha)$ of an amino acid, α of a s^{th} sequence at i^{th} site of the target structure; total degree of frustration, $F_s^{tot}(\alpha)$ in the whole sequence and its average, $F_s^{avg}(\alpha)$ where N=130.

TABLE S1. Attribution of degree of frustration to the amino acids at the target structure based on developed potential.

Rank	Degree of frustration
1	0
2	+1
3	+2
4	+3
5	+4
6	+5

$$F_s^i(\alpha) = \begin{cases} 0, & \text{if k is at rank 1} \\ +1, & \text{if k is at rank 2} \\ +2, & \text{if k is at rank 3} \\ +3, & \text{if k is at rank 4} \\ +4, & \text{if k is at rank 5} \\ +5, & \text{if k is at rank 6} \end{cases}$$

$$F_s^{tot}(\alpha) = \sum_{i=1}^{N} F_s^i(\alpha)$$
$$F_s^{avg}(\alpha) = \frac{F_s^{tot}(\alpha)}{N}$$

In section 3.3.3, it is shown that among the design methods, method M_4 designs more frustrated sequences. Moreover, it

a)Electronic mail: anupaulbaruah@dibru.ac.in

TABLE S2. Determination of degree of frustration of the designed sequences at the target structure based on developed potential.

Method	Sl. No. of Sequence in TSR	Total Frustration	Average of Total Frustration
M_2	1	247	252.8
	2	265	
	3	264	
	4	251	
	5	237	
M_3	1	214	213.8
	2	217	
	3	216	
	4	211	
	5	211	
M_4	1	246	**244.8**
	2	244	
	3	246	
	4	246	
	5	242	
M_5	1	231	225.4
	2	218	
	3	224	
	4	229	
	5	225	
M_6	1	192	209.2
	2	217	
	3	226	
	4	195	
	5	216	
M_7	1	239	258.2
	2	262	
	3	274	
	4	252	
	5	264	

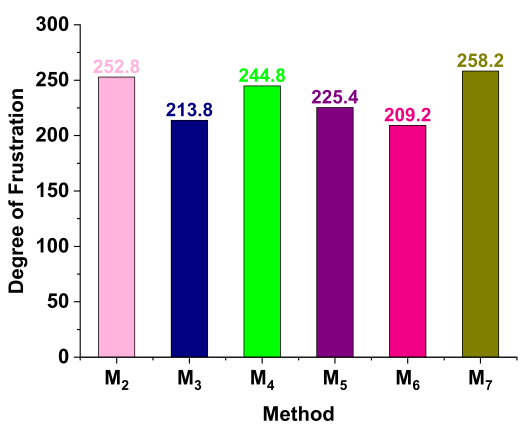
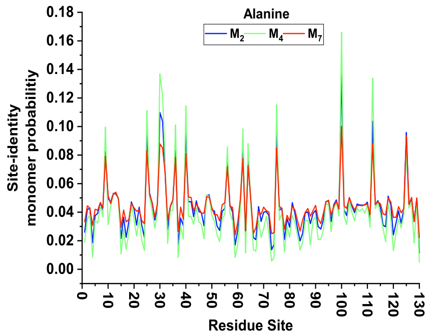


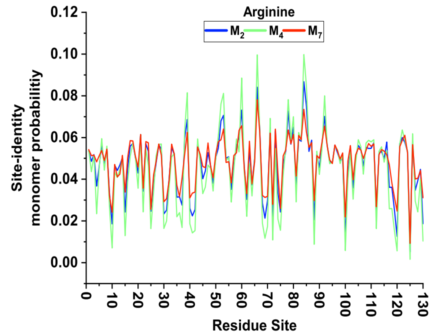
Fig. S1. Plot of degree of frustration for the sequences designed using method M_2, M_3, M_4, M_5, M_6 and M_7.

is also shown that the method M_4 is successful in designing sequences which are highly frustrated and specifically destabilize the non-native conformations. From Table S2 and Fig. S1, it can be seen that all the methods designs frustrated sequences. However, the average degree of frustration for the sequences designed by method M_4 is 244.8. Although this average degree of frustration is not highest, however, this is in between the two extremes 209.2 (method M_6) and 258.2 (method M_7). This result indicates that method M_4 designs frustrated sequence. It is also shown in the section 3.2.1 and 3.3.1 that method M_4 designs more stable sequences in terms of percentage of Monte Carlo steps populated in the target state ensemble. This observation has bolstered our previous findings that method M_4 designs frustrated, however, stable sequences at the target structure.

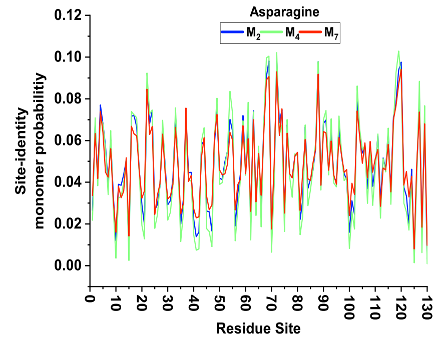
Self consistent mean field theory provides sequences in the form site identity monomer probabilities. Therefore, distribution graphs (Fig. S2(a)- S2(t)) are plotted for the comparison of site-identity monomer probabilities of some sequences designed by the methods M_2, M_4 and M_7. The comparison is done between two DSG methods M_2, M_4 and method M_7. The two DSG methods are chosen based on the percentage of Monte Carlo steps population in the target state ensemble, method M_2 having the lowest percentage of Monte Caro steps population and method M_4 with its highest value (see Table 5). While method M_7 is having lowest percentage of Monte Carlo steps population among all methods. A distribution plot, say Fig. S2(a), shows the monomer probabilities of alanine at all 130 residue sites of the sequences designed by the three selected methods M_2, M_4 and M_7. Similar distribution graphs are plotted for the remaining amino acid residues showing their monomer probabilities at each of the 130 sites. Since only generalized conclusion can be drawn from the method applied here for designing sequences, it is difficult to comment whether a specific residue will be conserved at particular site of the target structure.



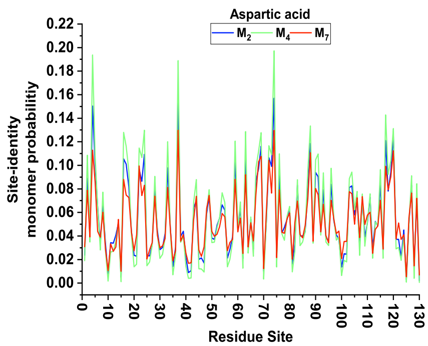
(a)



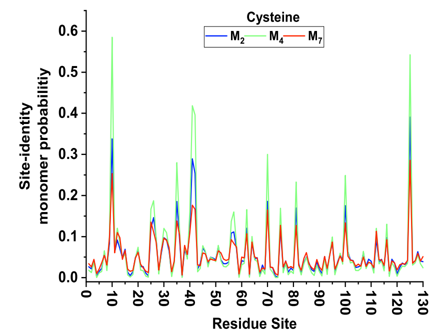
(b)



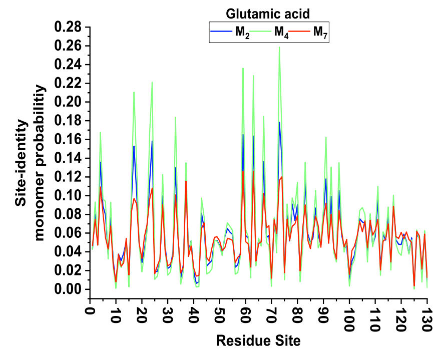
(c)



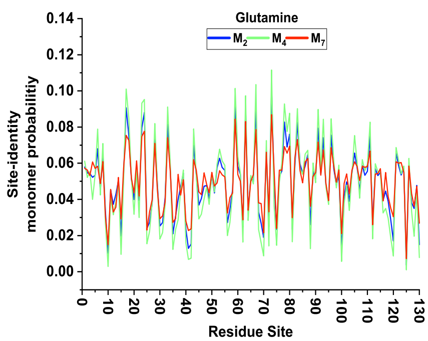
(d)



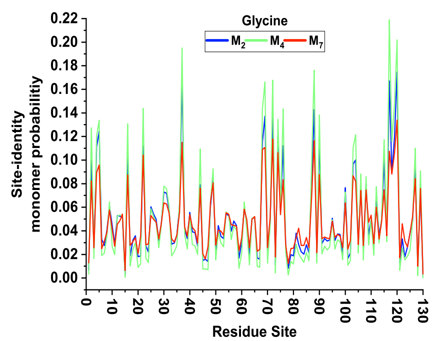
(e)



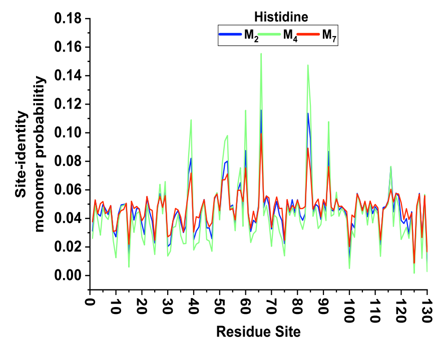
(f)



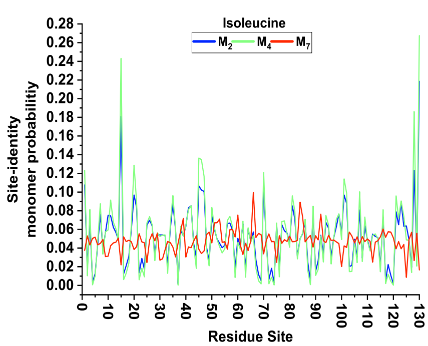
(g)



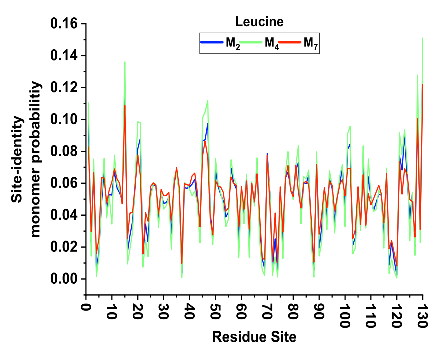
(h)



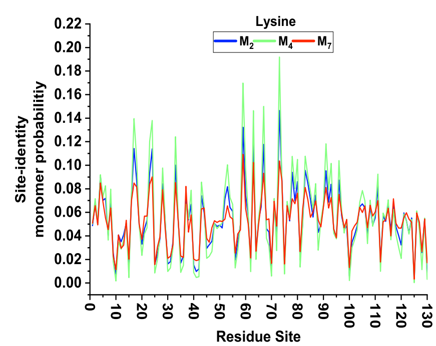
(i)



(j)



(k)



(l)

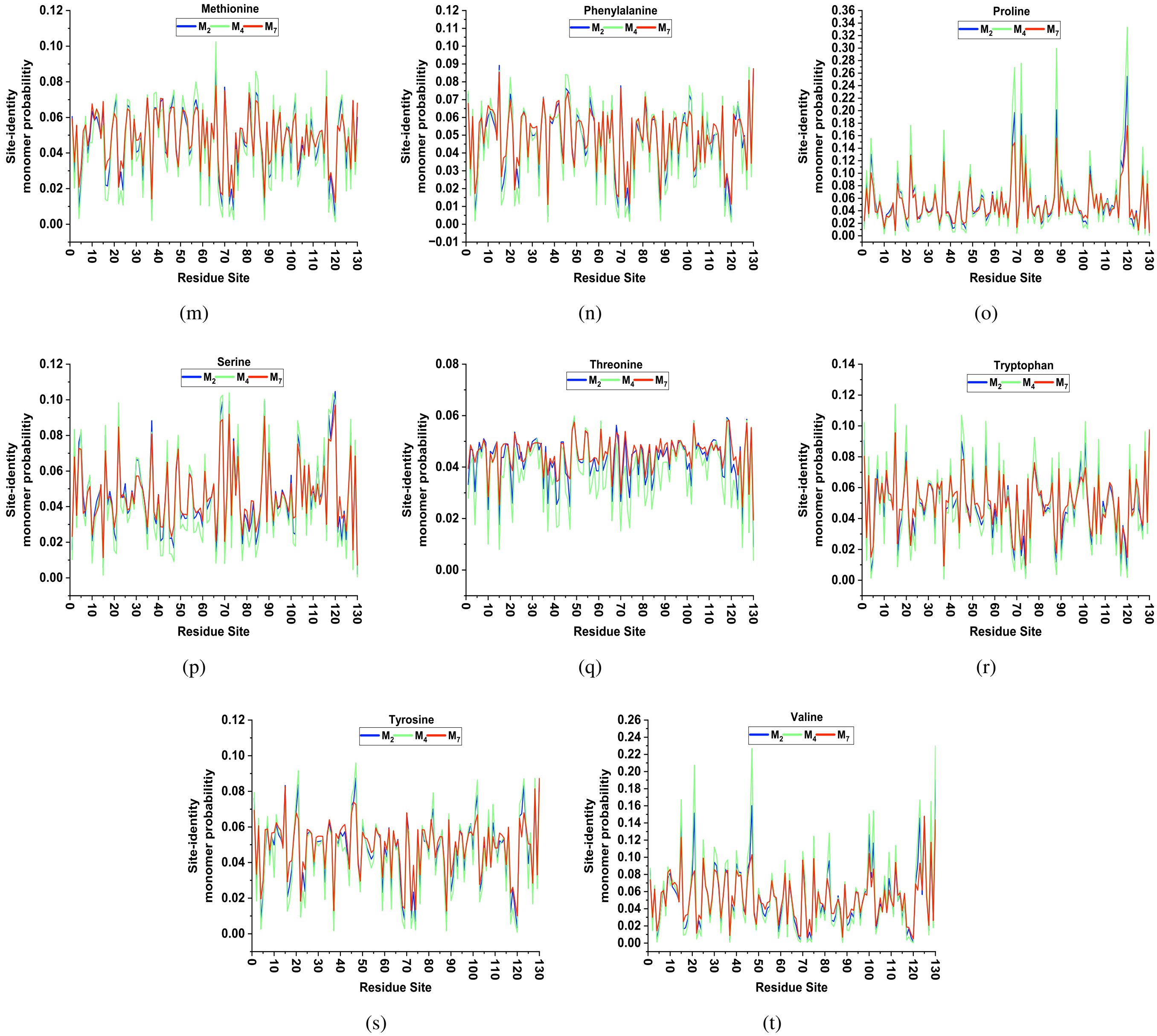


Fig. S2. Distribution plot of site-identity monomer probability vs. residue sites for the sequences corresponding to lowest foldability values designed by methods M_2, M_4, and M_7.

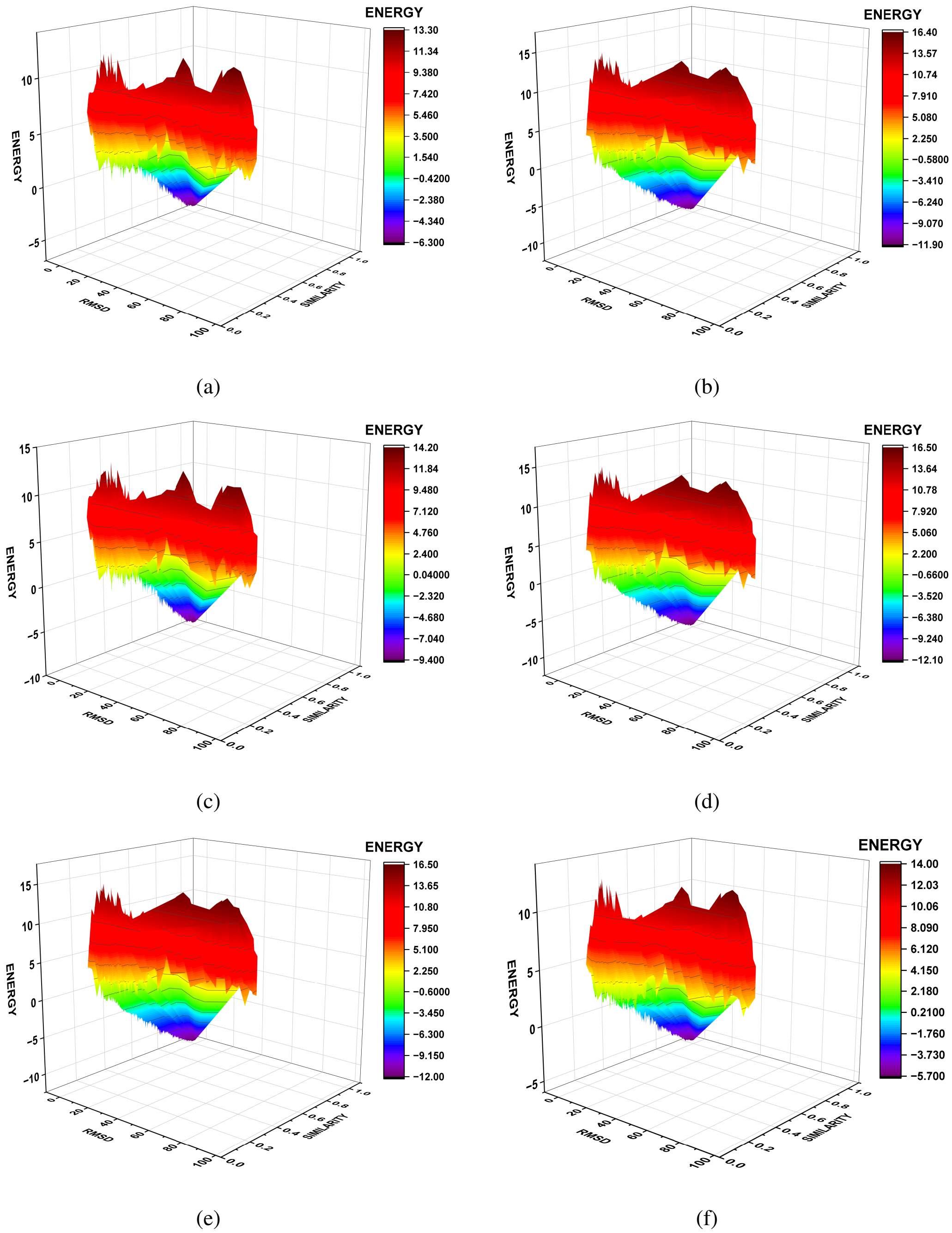


Fig. S3. Energy landscape contours of the sequences having lowest Δ value designed using (a)M_2, (b)M_3, (c)M_4, (d)M_5, (e)M_6, (f)M_7

TABLE S3. Results of Monte Carlo simulations on the sequences designed by Monte Carlo simulation method.

Method	Sequence in TSR	Number of refolding	Steps in target(%)	Average of steps in target (%)
M_2	Δ_1=−1.2805 Δ_3=−6.8465	7674	63.57	54.73
	Δ_1=−1.2806 Δ_3=−6.8283	7654	46.75	
	Δ_1=−1.2806 Δ_3=−6.8307	7644	47.25	
	Δ_1=−1.2806 Δ_3=−6.8313	7674	63.57	
	Δ_1=−1.2806 Δ_3=−6.8376	7666	52.51	
M_3	Δ_1=−1.2105 Δ_4=−13.2776	7643	56.07	55.91
	Δ_1=−1.2105 Δ_4=−13.2782	7649	56.16	
	Δ_1=−1.2106 Δ_4=−13.2768	7644	56.34	
	Δ_1=−1.2106 Δ_4=−13.2775	7642	56.16	
	Δ_1=−1.2103 Δ_4=−13.2768	7626	54.82	
M_4	Δ_2=−4.1528 Δ_3=−8.7561	7655	74.78	**73.96**
	Δ_2=−4.1527 Δ_3=−8.7395	7655	73.48	
	Δ_2=−4.1523 Δ_3=−8.7372	7654	73.49	
	Δ_2=−4.1508 Δ_3=−8.7470	7655	73.48	
	Δ_2=−4.1555 Δ_3=−8.7147	7654	74.57	
M_5	Δ_2=−3.1738 Δ_4=−13.4279	7660	56.01	62.40
	Δ_2=−3.1734 Δ_4=−13.4176	7684	68.08	
	Δ_2=−3.1736 Δ_4=−13.4195	7678	68.45	
	Δ_2=−3.1737 Δ_4=−13.4120	7675	55.26	
	Δ_2=−3.1737 Δ_4=−13.4462	7684	64.22	
M_6	Δ_3=−7.0678 Δ_4=−13.2386	7668	61.41	66.33
	Δ_3=−7.0645 Δ_4=−13.2442	7655	66.42	
	Δ_3=−7.0623 Δ_4=−13.2155	7688	70.38	
	Δ_3=−7.0647 Δ_4=−13.2221	7662	63.49	
	Δ_3=−7.0638 Δ_4=−13.2105	7697	69.97	
M_7	Δ=−6.2101	7526	68.05	64.65
	Δ=−6.2102	7108	60.41	
	Δ=−6.2103	7609	64.69	
	Δ=−6.2104	7664	59.81	
	Δ=−6.2105	7378	70.27	

TABLE S4. Comparison of the sequences designed by Monte Carlo simulation method with the sequences designed by mean field method. (**Bold figures in black color indicates class**; number of sites in parentheses are in blue color; percentages in brackets are in magenta color.)

Method	Sequence in TSR	**Class** Number of sites Percentage (%)			
		1–3	**4–6**	**7–10**	**11–20**
M_2	Δ_1=−1.2805 Δ_3=−6.8465	(34) [26.15]	(30) [23.07]	(31) [23.84]	(35) [26.92]
	Δ_1=−1.2806 Δ_3=−6.8283	(31) [23.84]	(38) [29.23]	(30) [23.07]	(31) [23.84]
	Δ_1=−1.2806 Δ_3=−6.8307	(33) [25.38]	(34) [26.15]	(30) [23.07]	(33) [25.38]
	Δ_1=−1.2806 Δ_3=−6.8313	(34) [26.15]	(30) [23.07]	(30) [23.07]	(36) [27.69]
	Δ_1=−1.2806 Δ_3=−6.8376	(33) [25.38]	(31) [23.84]	(29) [22.30]	(37) [28.46]
M_3	Δ_1=−1.2105 Δ_4=−13.2776	(40) [30.76]	(28) [21.53]	(27) [20.76]	(35) [26.92]
	Δ_1=−1.2105 Δ_4=−13.2782	(40) [30.76]	(28) [21.53]	(27) [20.76]	(35) [26.92]
	Δ_1=−1.2106 Δ_4=−13.2768	(40) [30.76]	(29) [22.30]	(26) [20.00]	(35) [26.92]
	Δ_1=−1.2106 Δ_4=−13.2775	(40) [30.76]	(28) [21.53]	(27) [20.76]	(35) [26.92]
	Δ_1=−1.2103 Δ_4=−13.2768	(41) [31.53]	(29) [22.30]	(24) [18.46]	(36) [27.69]
M_4	Δ_2=−4.1528 Δ_3=−8.7561	(48) [36.92]	(28) [21.53]	(25) [19.23]	(29) [22.30]
	Δ_2=−4.1527 Δ_3=−8.7395	(48) [36.92]	(28) [21.53]	(24) [18.46]	(30) [23.07]
	Δ_2=−4.1523 Δ_3=−8.7372	(47) [36.15]	(29) [22.30]	(25) [19.23]	(29) [22.30]
	Δ_2=−4.1508 Δ_3=−8.7470	(48) [36.92]	(28) [21.53]	(25) [19.23]	(29) [22.30]
	Δ_2=−4.1555 Δ_3=−8.7147	(47) [36.15]	(28) [21.53]	(25) [19.23]	(30) [23.07]
M_5	Δ_2=−3.1738 Δ_4=−13.4279	(34) [26.15]	(25) [19.23]	(33) [25.38]	(38) [29.23]
	Δ_2=−3.1734 Δ_4=−13.4176	(41) [31.53]	(26) [20.00]	(31) [23.84]	(32) [24.61]
	Δ_2=−3.1736 Δ_4=−13.4195	(37) [28.46]	(31) [23.84]	(34) [26.15]	(28) [21.53]
	Δ_2=−3.1737 Δ_4=−13.4120	(38) [29.23]	(27) [20.76]	(27) [20.76]	(38) [29.23
	Δ_2=−3.1737 Δ_4=−13.4462	(42) [32.30]	(25) [19.23]	(29) [22.30]	(34) [26.15]
M_6	Δ_3=−7.0678 Δ_4=−13.2386	(35) [26.92]	(38) [29.23]	(25) [19.23]	(32) [24.61]
	Δ_3=−7.0645 Δ_4=−13.2442	(41) [31.53]	(27) [20.76]	(23) [17.69]	(39) [30.00]
	Δ_3=−7.0623 Δ_4=−13.2155	(38) [29.23]	(28) [21.53]	(27) [20.76]	(37) [28.46]
	Δ_3=−7.0647 Δ_4=−13.2221	(42) [32.30]	(22) [16.92]	(32) [24.61]	(34) [26.15]
	Δ_3=−7.0638 Δ_4=−13.2105	(35) [26.92]	(36) [27.69]	(28) [21.53]	(31) [23.84]
M_7	Δ=−6.2101	(24) [18.46]	(28) [21.53]	(35) [26.92]	(43) [33.07]
	Δ=−6.2102	(30) [23.07]	(23) [17.69]	(40) [30.76]	(37) [28.46]
	Δ=−6.2103	(22) [16.92]	(34) [26.15]	(33) [25.38]	(41) [31.53]
	Δ=−6.2104	(30) [23.07]	(16) [12.30]	(40) [30.76]	(44) [33.84]
	Δ=−6.2105	(29) [22.30]	(29) [22.30]	(33) [25.38]	(39) [30.00]

TABLE S5. The averages over the percentage of number of sites belonging to class 1–3 for each method.(**Bold figures indicate class;** number of sites in parentheses are in blue color; percentages in brackets are in magenta color.)

Method	Sequence in TSR	**Class** / Number of sites / Percentage (%) / **1–3**	Average over the percentage(%)
M_2	Δ_1=−1.2805, Δ_3=−6.8465	(34) [26.15]	25.38
	Δ_1=−1.2806, Δ_3=−6.8283	(31) [23.84]	
	Δ_1=−1.2806, Δ_3=−6.8307	(33) [25.38]	
	Δ_1=−1.2806, Δ_3=−6.8313	(34) [26.15]	
	Δ_1=−1.2806, Δ_3=−6.8376	(33) [25.38]	
M_3	Δ_1=−1.2105, Δ_4=−13.2776	(40) [30.76]	30.91
	Δ_1=−1.2105, Δ_4=−13.2782	(40) [30.76]	
	Δ_1=−1.2106, Δ_4=−13.2768	(40) [30.76]	
	Δ_1=−1.2106, Δ_4=−13.2775	(40) [30.76]	
	Δ_1=−1.2103, Δ_4=−13.2768	(41) [31.53]	
M_4	Δ_2=−4.1528, Δ_3=−8.7561	(48) [36.92]	**36.61**
	Δ_2=−4.1527, Δ_3=−8.7395	(48) [36.92]	
	Δ_2=−4.1523, Δ_3=−8.7372	(47) [36.15]	
	Δ_2=−4.1508, Δ_3=−8.7470	(48) [36.92]	
	Δ_2=−4.1555, Δ_3=−8.7147	(47) [36.15]	
M_5	Δ_2=−3.1738, Δ_4=−13.4279	(34) [26.15]	29.53
	Δ_2=−3.1734, Δ_4=−13.4176	(41) [31.53]	
	Δ_2=−3.1736, Δ_4=−13.4195	(37) [28.46]	
	Δ_2=−3.1737, Δ_4=−13.4120	(38) [29.23]	
	Δ_2=−3.1737, Δ_4=−13.4462	(42) [32.30]	
M_6	Δ_3=−7.0678, Δ_4=−13.2386	(35) [26.92]	29.38
	Δ_3=−7.0645, Δ_4=−13.2442	(41) [31.53]	
	Δ_3=−7.0623, Δ_4=−13.2155	(38) [29.23]	
	Δ_3=−7.0647, Δ_4=−13.2221	(42) [32.30]	
	Δ_3=−7.0638, Δ_4=−13.2105	(35) [26.92]	
M_7	Δ=−6.2101	(24) [18.46]	20.76
	Δ=−6.2102	(30) [23.07]	
	Δ=−6.2103	(22) [16.92]	
	Δ=−6.2104	(30) [23.07]	
	Δ=−6.2105	(29) [22.30]	

TABLE S6. Comparison of Monte Carlo designed sequences with the target sequence based on structural context similarity.

Method	Sequence in TSR	Number of sites with same k as in target	Percentage(%)	Average over the percentage(%)
M_2	Δ_1=−1.2805 Δ_3=−6.8465	28	21.53	21.99
	Δ_1=−1.2806 Δ_3=−6.8283	28	21.53	
	Δ_1=−1.2806 Δ_3=−6.8307	29	22.30	
	Δ_1=−1.2806 Δ_3=−6.8313	28	21.53	
	Δ_1=−1.2806 Δ_3=−6.8376	30	23.07	
M_3	Δ_1=−1.2105 Δ_4=−13.2776	36	27.69	28.31
	Δ_1=−1.2105 Δ_4=−13.2782	36	27.69	
	Δ_1=−1.2106 Δ_4=−13.2768	37	28.46	
	Δ_1=−1.2106 Δ_4=−13.2775	37	28.46	
	Δ_1=−1.2103 Δ_4=−13.2768	38	29.23	
M_4	Δ_2=−4.1528 Δ_3=−8.7561	26	20.00	20.30
	Δ_2=−4.1527 Δ_3=−8.7395	26	20.00	
	Δ_2=−4.1523 Δ_3=−8.7372	27	20.76	
	Δ_2=−4.1508 Δ_3=−8.7470	26	20.00	
	Δ_2=−4.1555 Δ_3=−8.7147	27	20.76	
M_5	Δ_2=−3.1738 Δ_4=−13.4279	33	25.38	25.84
	Δ_2=−3.1734 Δ_4=−13.4176	34	26.15	
	Δ_2=−3.1736 Δ_4=−13.4195	33	25.38	
	Δ_2=−3.1737 Δ_4=−13.4120	34	26.15	
	Δ_2=−3.1737 Δ_4=−13.4462	34	26.15	
M_6	Δ_3=−7.0678 Δ_4=−13.2386	36	27.69	27.38
	Δ_3=−7.0645 Δ_4=−13.2442	36	27.69	
	Δ_3=−7.0623 Δ_4=−13.2155	35	26.92	
	Δ_3=−7.0647 Δ_4=−13.2221	36	27.69	
	Δ_3=−7.0638 Δ_4=−13.2105	35	26.92	
M_7	Δ=−6.2101	30	23.07	21.53
	Δ=−6.2102	30	23.07	
	Δ=−6.2103	23	17.69	
	Δ=−6.2104	27	20.76	
	Δ=−6.2105	30	23.07	

TABLE S7. Determination of structural context (k) similarity between the k of Monte Carlo designed sequences and the k of unfolded conformations.

Method	Sequence in TSR	Average Percentage of structural context similarity over unfolded conformations(%)	Average over the percentage(%)
M_2	Δ_1=−1.2805 Δ_3=−6.8465	16.80	16.84
	Δ_1=−1.2806 Δ_3=−6.8283	16.41	
	Δ_1=−1.2806 Δ_3=−6.8307	16.61	
	Δ_1=−1.2806 Δ_3=−6.8313	16.80	
	Δ_1=−1.2806 Δ_3=−6.8376	17.59	
M_3	Δ_1=−1.2105 Δ_4=−13.2776	20.46	20.94
	Δ_1=−1.2105 Δ_4=−13.2782	20.60	
	Δ_1=−1.2106 Δ_4=−13.2768	21.12	
	Δ_1=−1.2106 Δ_4=−13.2775	21.12	
	Δ_1=−1.2103 Δ_4=−13.2768	21.41	
M_4	Δ_2=−4.1528 Δ_3=−8.7561	14.21	**14.51**
	Δ_2=−4.1527 Δ_3=−8.7395	14.28	
	Δ_2=−4.1523 Δ_3=−8.7372	15.03	
	Δ_2=−4.1508 Δ_3=−8.7470	14.19	
	Δ_2=−4.1555 Δ_3=−8.7147	14.84	
M_5	Δ_2=−3.1738 Δ_4=−13.4279	20.40	19.46
	Δ_2=−3.1734 Δ_4=−13.4176	18.75	
	Δ_2=−3.1736 Δ_4=−13.4195	19.41	
	Δ_2=−3.1737 Δ_4=−13.4120	20.28	
	Δ_2=−3.1737 Δ_4=−13.4462	18.49	
M_6	Δ_3=−7.0678 Δ_4=−13.2386	21.96	21.79
	Δ_3=−7.0645 Δ_4=−13.2442	20.29	
	Δ_3=−7.0623 Δ_4=−13.2155	22.86	
	Δ_3=−7.0647 Δ_4=−13.2221	22.27	
	Δ_3=−7.0638 Δ_4=−13.2105	21.60	
M_7	Δ=−6.2101	19.53	17.75
	Δ=−6.2102	17.81	
	Δ=−6.2103	14.73	
	Δ=−6.2104	17.93	
	Δ=−6.2105	17.75	

TABLE S8. Binning of number of non-native conformations according to their percentage of structural context similarity. (Bins of structural context similarity percentage are in blue color; number of non-native conformations are in black color)

Method	Sequence in TSR	Bins of structural context similarity percentage / Number of non-native conformations						
		0–5	5–10	10–15	15–20	20–25	25–30	30–100
M_2	Δ_1=−1.2805 Δ_3=−6.8465	6	2047	68914	154633	46660	302	14
	Δ_1=−1.2806 Δ_3=−6.8283	2	1925	77713	160209	32456	263	8
	Δ_1=−1.2806 Δ_3=−6.8307	2	2121	75956	151677	42624	186	10
	Δ_1=−1.2806 Δ_3=−6.8313	6	2039	68993	154637	46574	312	15
	Δ_1=−1.2806 Δ_3=−6.8376	2	842	51760	149938	69019	988	27
M_3	Δ_1=−1.2105 Δ_4=−13.2776	0	331	18353	90368	134034	29376	114
	Δ_1=−1.2105 Δ_4=−13.2782	0	283	16956	87435	136981	30785	136
	Δ_1=−1.2106 Δ_4=−13.2768	0	283	13912	77717	139199	41041	424
	Δ_1=−1.2106 Δ_4=−13.2775	0	283	13912	77717	139199	41041	424
	Δ_1=−1.2103 Δ_4=−13.2768	0	256	13474	74442	132387	50497	1520
M_4	Δ_2=−4.1528 Δ_3=−8.7561	18	11343	157152	97387	6468	177	31
	Δ_2=−4.1527 Δ_3=−8.7395	14	10599	155573	99552	6642	172	24
	Δ_2=−4.1523 Δ_3=−8.7372	3	5766	130149	123732	12728	173	25
	Δ_2=−4.1508 Δ_3=−8.7470	18	11659	157956	96403	6334	175	31
	Δ_2=−4.1555 Δ_3=−8.7147	17	7183	137299	115782	12080	187	28
M_5	Δ_2=−3.1738 Δ_4=−13.4279	0	363	17392	88901	143527	22028	365
	Δ_2=−3.1734 Δ_4=−13.4176	16	1077	36665	123095	103157	8545	21
	Δ_2=−3.1736 Δ_4=−13.4195	6	807	26509	110915	123019	11302	18
	Δ_2=−3.1737 Δ_4=−13.4120	2	348	18157	91455	141936	20582	96
	Δ_2=−3.1737 Δ_4=−13.4462	6	1126	41871	126218	96388	6940	27
M_6	Δ_3=−7.0678 Δ_4=−13.2386	2	303	8877	58851	147721	55035	1787
	Δ_3=−7.0645 Δ_4=−13.2442	6	648	19565	92771	133238	26000	348
	Δ_3=−7.0623 Δ_4=−13.2155	1	205	5425	37508	154336	71622	3479
	Δ_3=−7.0647 Δ_4=−13.2221	7	455	9407	52643	143138	63903	3023
	Δ_3=−7.0638 Δ_4=−13.2105	4	341	9919	63939	152687	44608	1078
M_7	Δ=−6.2101	7	547	18619	112418	135190	5755	40
	Δ=−6.2102	4	1105	44245	150045	76238	919	20
	Δ=−6.2103	17	6497	138236	120908	6761	135	22
	Δ=−6.2104	7	764	37980	157644	74680	1460	41
	Δ=−6.2105	2	1135	44407	153360	73116	537	19

TABLE S9. Similarity of misfolded structures to the native structure.

Method	Average similarity of misfolded structures to the native structure
M_2	93.19%
M_3	93.10%
M_4	93.53%
M_5	93.11%
M_6	93.08%
M_7	92.79%