

Predicting the conformational preferences of proteins using a physics-based free energy method

Arijit Roy, Alberto Perez, Ken A. Dill, and Justin L. MacCallum

Laufer Center for Physical and Quantitative Biology
and Departments of Physics and Chemistry
Stony Brook University
Stony Brook, NY 11794-5252.

May 3, 2013

1 Preparation of input model for chameleon sequences

We prepared computer generated models in both α and β conformations for all five chameleon sequences mentioned in the main text. Crystallographic structure of GA95 and GB95 sequence have α and β fold respectively. When, we attempt to prepare the α and β fold with GA95 sequence we start with the crystallographic structure of GA95 and GB95 with GA95 sequence. However, we remove the sidechains at three mutated positions in both the fold and call the program SCWRL4¹ to generate the sidechain conformation of the mutated residues and its neighbors. Followed by this step, we do long molecular dynamics minimization to remove any possible bad contact. This procedure was followed to generate all α and β conformations of five sequences that are used for free energy calculation.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

2 What can we say about low resolution models from CASP experiments?

In the main text, we have seen that the method is good at predicting preferences when the structures are not very far from the native. But the question remains how far from native can we go and still see that the method produces correct result. In this section we explored this question with models from extracellular domain of the jumping translocation breakpoint protein (pdb id: 2KJX). Most of the group could only generate low resolution models for this CASP9 target (id: T0531) . In our comparison, we choose five models by the group MUFOOLD-MD, which was the best performing group for this target with their best model had a GDT_TS value of 44. The result presented in Figure S6 shows: 1.) native is correctly identified as expected and 2.) Surprisingly there is a high level of correlation between the GDT ranking and the free energy ranking for model 1 and model 3, the rest three structures with GDT_TS score less than 35 are ordered incorrectly. It is worth to note that models 2 and 3 have the same GDT and very different free energies, meaning that the actual ordering could change a lot². It is encouraging that at least the method can pick out the best model even though it is got a low GDT score: 44.

[Figure 6 about here.]

References

- [1] Krivov, G.G.; Shapovalov, M.V.; Dunbrack, RL Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins.* (2009) 77, 778-795.
- [2] Perez, A.; Yang, Z.; Bahar, I.; Dill, K.A.; MacCallum, J.L.; FlexE: Using Elastic Network Models to Compare Models of Protein Structure. *J. Chem. Theory Comput.*, 2012, 8, 3985-3991.

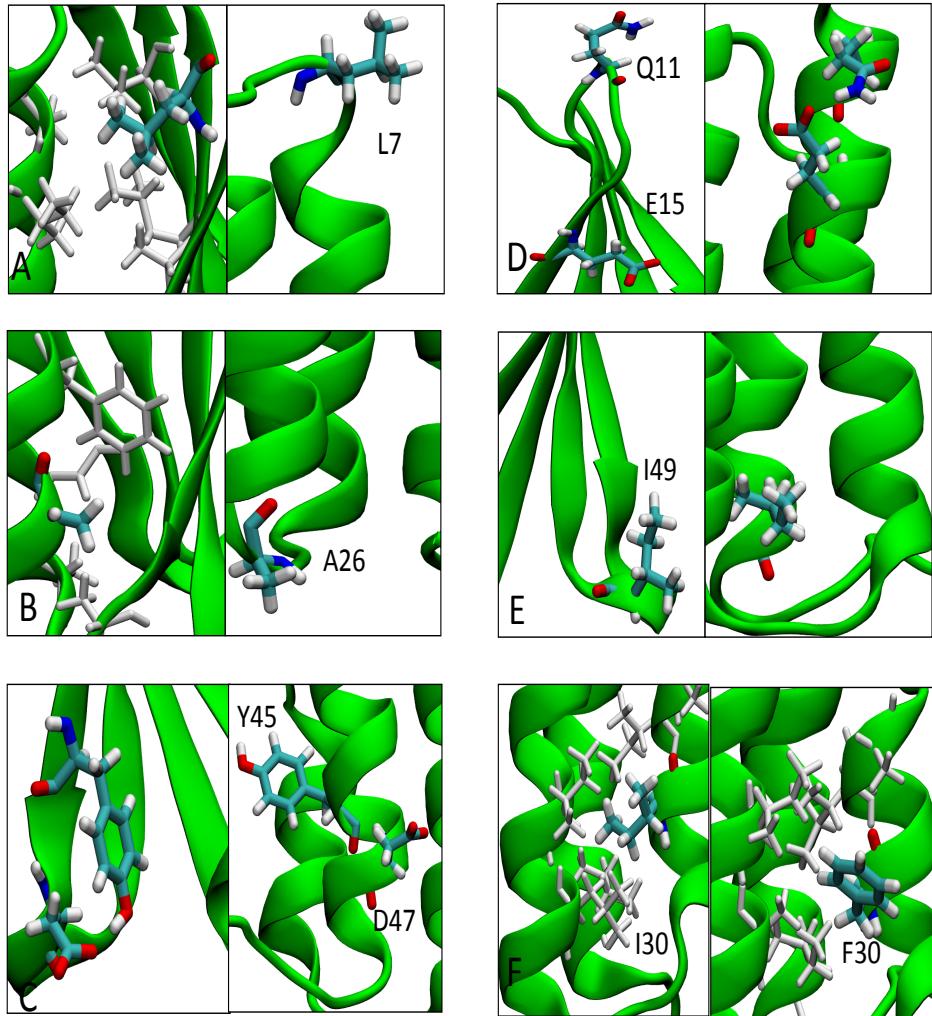


Figure S1: Per residue free energy calculation indicate significant differences of sidechain orientation of key amino acid residues. These amino acid residues control the free energy equilibrium to either β or α conformation. All figures in the left are in β conformation, whereas the right ones are in α conformation except in Figure (F). Sidechains of hydrophobic residues L7 in (A) and A26 in (B) are oriented towards the protein interior in β conformation. In such conformation they have hydrophobic interactions with other hydrophobic residues colored in white. In α conformation these residues are exposed to the protein surface. (C) Y45 forms H-bond with D47 with GB95 sequence and β conformation. (D) Q11 and E15 forms a salt bridge in β conformation which is absent in the α conformation. (E) Hydrophobic I49 is exposed to the protein surface in β conformation, whereas in α conformation it is inside protein hydrophobic core. (F) Sidechain of I30 is inside the protein hydrophobic core as it has a smaller sidechain with GA95 sequence and α conformation. In contrast, Phe-30 with larger sidechain is exposed to the protein surface with GB95 sequence and β conformation.

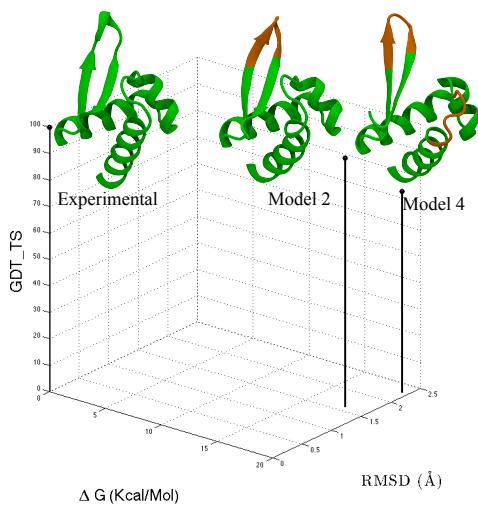


Figure S2: Native structure and two models (from group “Splicer”) for target T0560.

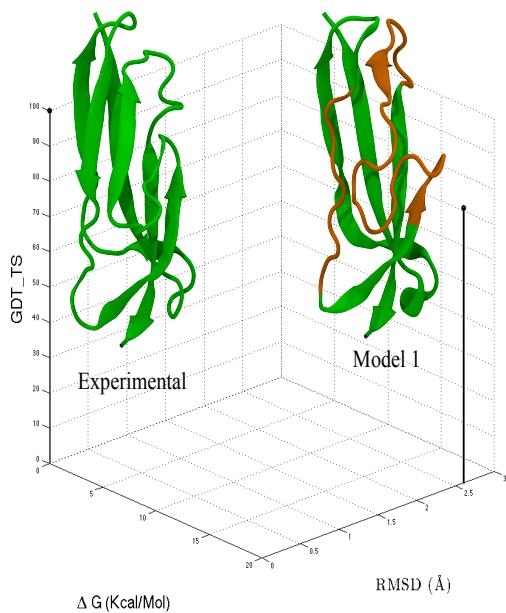


Figure S3: Native and best computer generated model structure of CASP Target T0569.

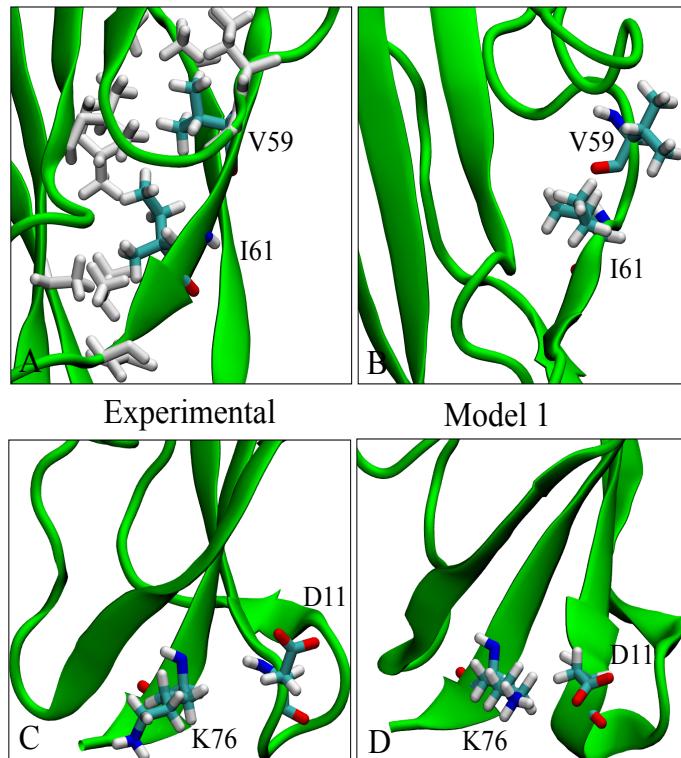


Figure S4: Key differences between experimental structure and the computer generated model as indicated by per residue free energy calculation. The sidechain of hydrophobic residues, V59 and I61 are oriented towards the protein hydrophobic core in (A) structure, (B) exposed to the surface in model. The beta sheet containing these residues is disordered in the generated model. A salt bridge between K76 and D11 in (D) computer generated structure, which is absent in the (C) native structure.

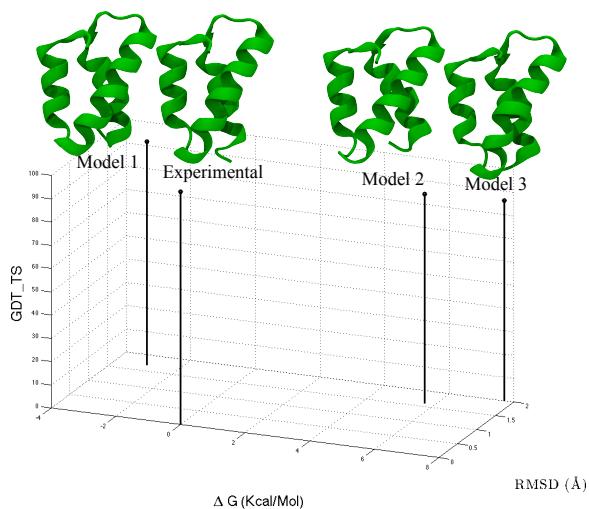


Figure S5: The native and three model structure of engineered protein from Asr4154 protein (PDB ID: 2L09 and CASP code:T0538). The model 1,2 and 3 are from the group PconsR, Shell and FOLDIT respectively.

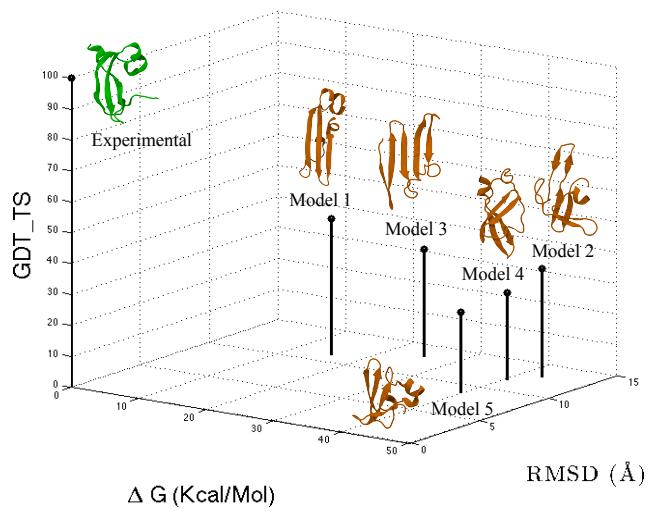


Figure S6: The native structure and 5 models of extracellular domain of the jumping translocation breakpoint protein (pdb id: 2KJX and the CASP code: T0531).