

# 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.

April 24, 2013

## 1 Preparation of input model for chameleon sequences

We prepared computer generated models in both  $\alpha$  and  $\beta$  conformation for all five chameleon sequences mentioned in the main text. Crystallographic structure of GA95 and GB95 sequence have  $\alpha$  and  $\beta$  fold respectively. When, we attempt to prepare the  $\alpha$  and  $\beta$  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 SCWRL4<sup>1</sup> 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  $\alpha$  and  $\beta$  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.]

[This fig isn't very readable. We need to focus into the sidechains, keeping only the nearby pieces of chain and leaving the rest out or making it fuzzy or something. We need to see some point that you are making by looking at the figure.]

## 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.

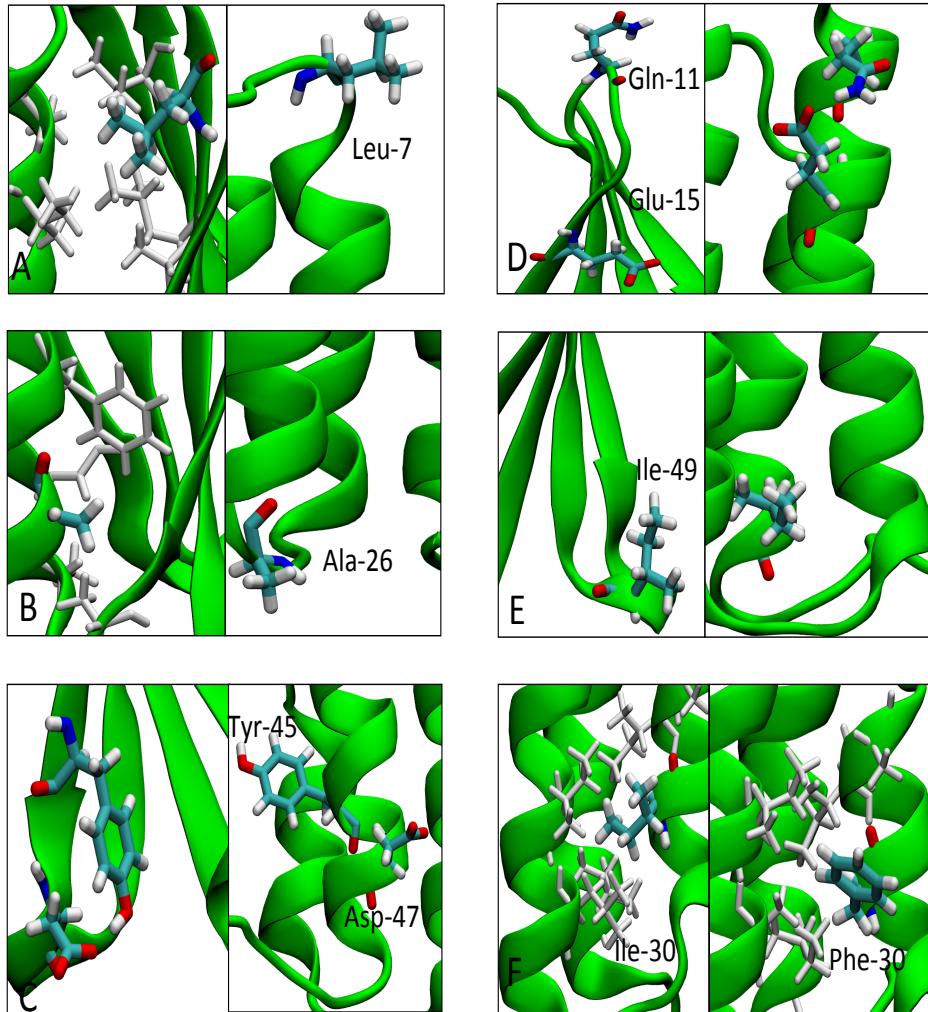


Figure 1: Per residue free energy calculation indicate significant differences of sidechain orientation of key amino acid residues. These amino acid residues control the free energy equilibrium to either  $\beta$  or  $\alpha$  conformation. All figures in the left are in  $\beta$  conformation, whereas the right ones are in  $\alpha$  conformation except in Figure (F). Sidechains of hydrophobic residues Leu-7 in (A) and Ala-26 in (B) are oriented towards the protein interior in  $\beta$  conformation. In such conformation they have hydrophobic interactions with other hydrophobic residues colored in white. In  $\alpha$  conformation these residues are exposed to the protein surface. (C) Tyr-45 forms H-bond with Asp-47 with GB95 sequence and  $\beta$  conformation. (D) Gln-11 and Glu-15 form a H-bond in  $\beta$  conformation which is absent in the  $\alpha$  conformation. (E) Hydrophobic Ile<sup>49</sup> is exposed to the protein surface in  $\beta$  conformation, whereas it is inside protein hydrophobic core in  $\alpha$  conformation. (F) Sidechain of Ile-30 is inside the protein hydrophobic core in as it has a smaller sidechain with GA95 sequence in  $\alpha$  conformation. In contrast, Phe-30 with larger sidechain is exposed to the protein surface with GB95 sequence.

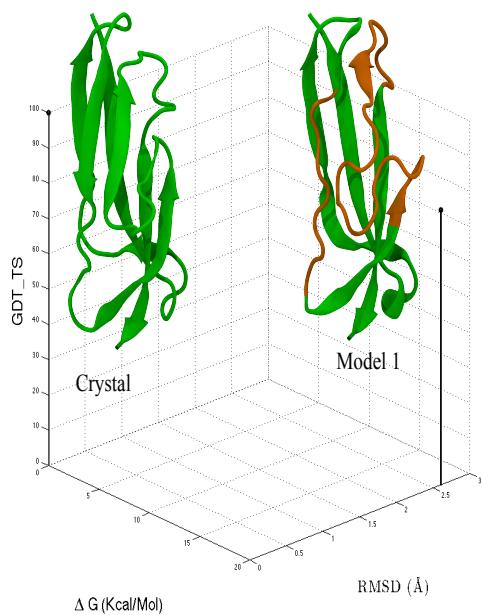


Figure 2: Native and best model structure of a domain of adhesion exoprotein from *Pediococcus pentosaceus* (pdb id: 2KWF and CASP code: T0569). The best model was from the group "Mufold".

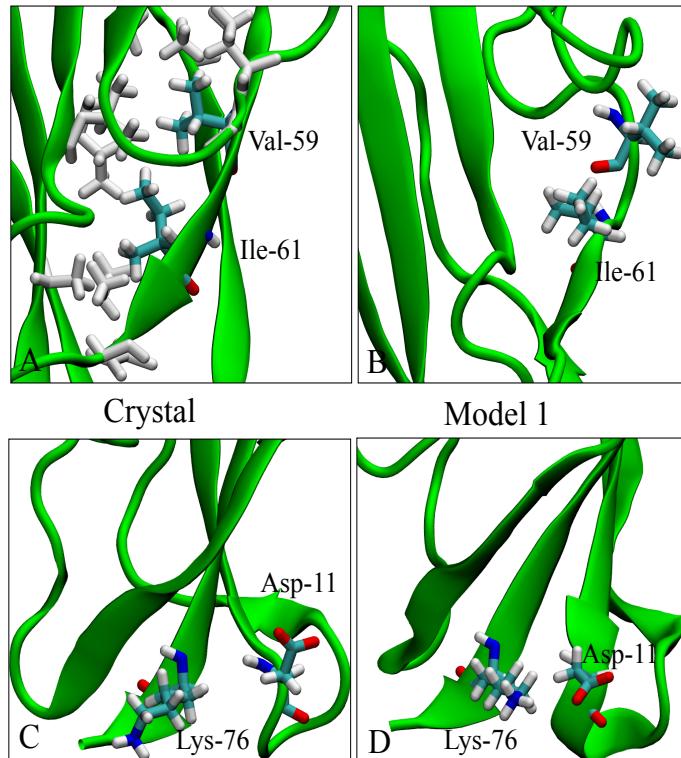


Figure 3: Key differences between native crystal structure and the model as indicated by per residue free energy calculation. The sidechain of hydrophobic residues, Val-59 and Ile-61 are oriented towards the protein hydrophobic core in (A) crystal structure, (B) exposed to the surface in model. The beta sheet containing these residues is disordered in the generated model. There is a H-bond between Lys-76 and Asp-11 in (D) model structure, which is absent in the (C) native structure.

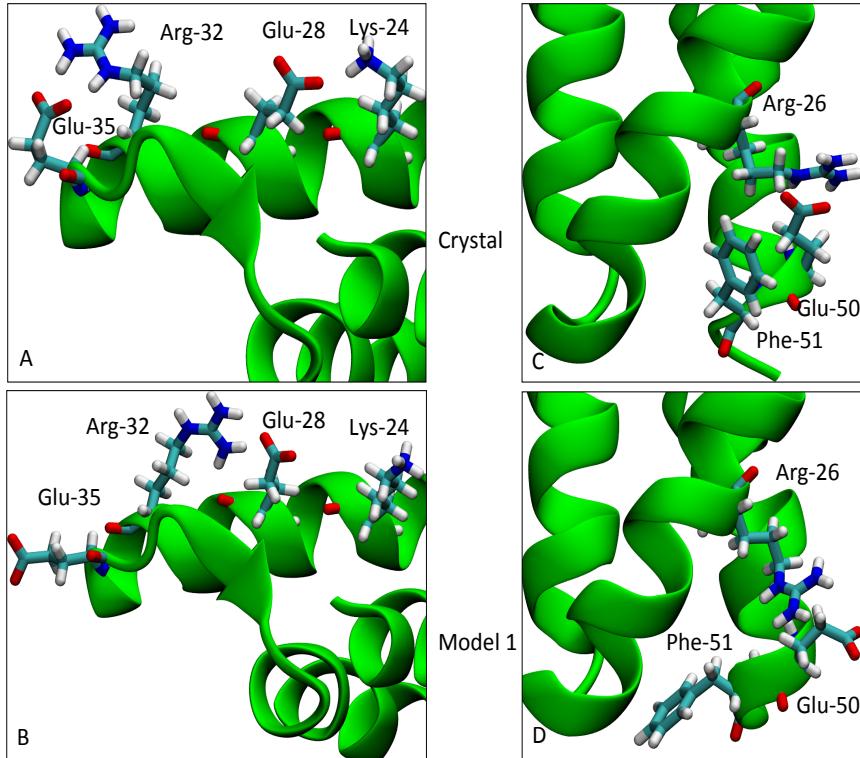


Figure 4: Differences of the sidechain orientation in model 1 and crystallographic structure of CASP target T0538 as determined by per residue free energy calculation. (A) Two salt bridges between residues Glu-35-Arg-32 and Glu-28-Lys-24 exist in the native crystallographic structure. These are absent in the model structure. Instead (B) a new salt bridge is formed between residues Arg-32-Glu-28 in model. (C) Another salt bridge between residues Arg-26-Glu-50 in the native structure, (D) This salt bridge is absent in the model structure. In comparison to the crystallographic structure, orientation of Phe-51 differs in model1.