Biophysical Journal, Volume *111*

Supplemental Information

Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape

Paulo Ricardo Mouro, Vinícius de Godoi Contessoto, Jorge Chahine, Ronaldo Junio de Oliveira, and Vitor Barbanti Pereira Leite

Supporting Material

Structure-based C_α model

The Cα model is a structure-based model which considers the amino acid sequence as simple spheres centralized at the position of the alpha carbon, in accordance with the structure data deposited in the Protein Data Bank (PDB) (1). During the dynamic to which the model is submitted, the potential that defines the energy of the conformations is of Gō-type, in which the main idea is to give importance to the interactions between amino acids residing in native contacts, and then choose the energy of contacts that minimize the total energy of the native state (1, 2).

The expression defining the energy of a configuration Γ based on the native conformation Γ^0 for the model Cα-model is given by,

$$\begin{aligned} V(\Gamma,\Gamma_o) = & \sum_{bonds} \epsilon_r (r-r_o)^2 + \sum_{angles} \epsilon_\theta (\theta-\theta_o)^2 \\ & + \sum_{dihedrals} \epsilon_\phi \left\{ [1-cos(\phi-\phi_o)] + \frac{1}{2}[1-cos(3(\phi-\phi_o))] \right\} \\ & + \sum_{contacts} \epsilon_C \left[5\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{d_{ij}}{r_{ij}}\right)^{10} \right] \\ & + \sum_{non-contacts} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} \end{aligned} \tag{S1}$$

In Equation S1, the first term refers to a harmonic potential representing the bond between two adjacent α-carbons in which r_0 is the distance between the two carbons connected in the native structure. The second sum also forms a harmonic potential, but this time, an angular harmonic potential formed by three α-carbons in sequence in the polypeptide chain. θ_0 is the angle formed by the three residues in the native conformation. The third term of the expression takes into consideration the torsion carried by the chain in which, ϕ_0 is the dihedral angle formed by four α-carbons in sequence. The fourth term accounts for the interaction between non-bonded α-carbons i and j , but which make contact in the native structure. For this is used a potential 10-12, in which d_{ij} is the value of the distance between the carbons which make a native contact. A native contact is defined by a map created by CSU software (3). Finally, the last term refers to all α-carbons that do not form a native contact. This term is used to keep the maximum approach distance between α-carbons. $\sigma_{NC} = 4\mathring{A}$ (volume of the carbon atom in the model). The constants ($\epsilon_r = 100$, $\epsilon_\theta = 20$, $\epsilon_\phi = 1$ and $\epsilon_{NC} = 1$) are given in ϵ_c units (2, 4).

The potential used to add frustration to the Cα-model is shown in the equation S2.

$$V_f(r) = -\epsilon_{NN} \exp\left[-\frac{(r_{ij} - \bar{d})^2}{{\sigma_f}^2}\right] \tag{S2}$$

$\bar{d}$ is the average of the distances of the native contacts, $\sigma_f = 1\ \mathring{A}$ and ϵ_{NN} is the frustration parameter in ϵ_c units. In this study $\epsilon_{NN} = 0.00$; $\epsilon_{NN} = 0.05$; $\epsilon_{NN} = 0.1$ and $\epsilon_{NN} = 0.2$ are used.

Simulation details

The topology and initial coordinates files were generated by Structure-based Models in Gromacs (SMOG) available online (5). The dynamics was performed using molecular dynamics package *GROMACS version 4.5-5* (6) using a stochastic integrator and the Berendsen thermal coupling. The simulations without frustration were made with 10^9 steps (having the information stored every 5000 steps with 0.5 fs of time step) and equilibrated after 10^7 steps. The simulations using $\epsilon_{NN} \neq 0$ were made with 5×10^8 steps (having the information stored every 5000 steps with 0.5 fs of time step) and equilibrated after 10^6 steps. The native contact fraction (Q) was the order parameter used to accompany the folding process, and a native contact was accepted in a Γ conformation if the distance between the amino acid in this conformation was less than $1.2d_{ij}$, in which d_{ij} is the distance between the residues in the native structure. The dynamic was realized with the computational resources of GridUnesp.

The thermodynamic profiles, such as thermal energy, free energy, entropy and specific heat were calculated by the method of multiple histograms (WHAM - *Weighted Histograms Analysis Method (7, 8)* in which the folding temperature was defined as the temperature at which F(0) $\approx$ F(1).

Figure

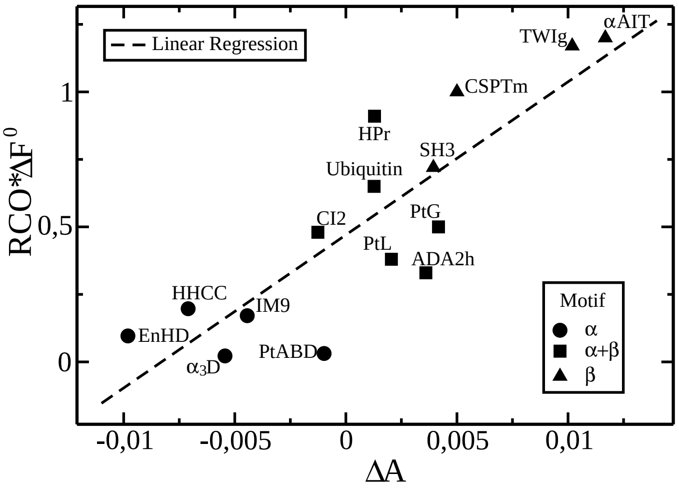

Figure S1

Figure S1. Free energy barrier and relative contact order ($\Delta F^0 \times RCO$) as a function of nonnative contact fraction variation (ΔA) for all proteins studied. Proteins are represented by their fold motif according to the SCOP database criterion (9): α (circles), β (triangles) and $\alpha + \beta$ (squares). The linear fit correlation to the data is 0.84. The data were extracted from Table I by using (ΔF) without energetic frustration.

Supporting References

1. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The protein data bank. Nucleic. Acids. Res. 28:235–242.

2. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: What determines the structural details of the transition state ensemble and ”en-route“ intermediates for protein folding? An investigation for small globular proteins. J. Mol. Biol. 298:937–953.

3. Sobolev, V., R. Wade, G. Vried, and M. Edelman. 1996. Molecular docking using surface complementarity. Proteins: Struct. Funct. Genet. 25:120–129.

4. Contessoto, V. G., D. T. Lima, R. J. Oliveira, A. T. Bruni, J. Chahine, and V. B. P. Leite. 2013. Analyzing the effect of homogeneous frustration in protein folding. Proteins: Struct. Funct. Bioinf. 81:1727–1737.

5. Noel, J. K., P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic. 2010. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. Nucleic Acids Research. .

6. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. J. Comp. Chem. 26:1701–1718.

7. Ferrenberg, A. M., and R. H. Swendsen. 1988. New monte carlo technique for studying phase transitions. Phys. Rev. Lett. 61:2635–2638.

8. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized monte-carlo data analysis. Phys. Rev. Lett. 63:1195–1198.

9. Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247:536–540.