Supporting Text

Gō Model Simulations

The selected homodimers were simulated with the Gō model (1), which takes into account only interactions that exist in the native structure, and therefore does not include energetic frustration (or, alternatively said, only includes topological frustration). We use here an off-lattice Gō model, where each residue is represented by a single bead centered on its α-carbon (C_α) position (2). Adjacent beads are strung together into a polymer chain by means of a potential encoding bond length and angle constraints. The secondary structure is encoded in the dihedral angle potential and the nonbonded (native contact) potential. The interaction energy U at a given protein conformation Γ is given by

$$U(\Gamma,\Gamma_0) = \sum_{bonds}^{N-1} K_b (b_i - b_{0i})^2 + \sum_{angles}^{N-2} K_\theta (\theta_i - \theta_{0i})^2$$
$$+ \sum_{dihedrals}^{N-3} \{K_\phi^{(1)}[1-\cos(n\times(\phi_i-\phi_{0i}))] + K_\phi^{(3)}[1-\cos(n\times(\phi_i-\phi_{0i}))]$$
$$+ \sum_{\substack{native-contacts \\ |i-j|>3}} \left\{ \varepsilon \left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10} \right] \right\} + \sum_{\substack{non-native \\ contacts,|i-j|>3}} \varepsilon \left(\frac{C}{r_{ij}}\right)^{12} \qquad [1]$$

In the equation, b_i, θ_i, and ϕ_i stand for the ith virtual bond length between ith and (i + 1)th residue, the virtual bond angle between (i - 1)th and ith bonds, and the virtual dihedral angle around the ith bond, respectively. The parameters b_{0i}, θ_{0i}, and ϕ_{0i} stand for the corresponding variables at the native structure. In the framework of the model, all native contacts (as defined by the CSU software) are represented by the 10-12 Lennard Jones form without any discrimination between the various chemical types of interaction. Moreover, both the intra- and intermonomeric

contacts (interfacial contacts) are treated in the same way without any bias toward separate folding or toward binding. The r_{ij} and r_{0ij} are the C_α-C_α distances between the contacting residues *i* and *j* in conformation Γ and Γ_0 (the PDB structure), respectively. In the summation over nonnative contacts, *C* (=4.0 Å) parameterizes the excluded volume repulsion between residues pairs that do not belong to the given native contact set. In the paper, all temperatures and energies are reported in units of ε. For other parameters, we use similar values that have been used in several folding studies (2-4), namely, $K_b = 100.0$, $K_\theta = 20.0$, $K_\phi^{(1)} = 1.0$, $K_\phi^{(3)} = 0.5$, $\varepsilon = 1.0$.

To enhance the sampling of binding events, the two identical subunits of each homodimer are linked by a polyglycine chain. This linker acts to hold the two unbound subunits (folded or unfolded) in close proximity during their motions; essentially the local concentrations are enhanced. The linker's length was determined by the distance between the C terminus of subunit A and the N terminus of subunit B. This length is sufficient to ensure the linker will not interfere with any intra- or intersubunit contacts that stabilize the folded dimer. To optimize its conformation with respect to the dimer, a minimization was performed on the linker including the two residues to which the linker is directly connected. Covalently linked Arc repressor (5) has been experimentally found to be fully functional with an enhanced folding rate and stability, suggesting that the linker plays a passive, largely entropic role of keeping the unbound monomers at high local concentrations during folding. To further ensure the linker's role is only entropic, it has no nonbonded interaction (native contacts) with both subunits. All the parameters for the bonded terms of the linker residues were chosen to be smaller by one order of magnitude to enhance its flexibility and to reduce its energetic contributions. To test the effect of the linker on the association kinetics, λ Cro repressor was studied with a linker of 12, 20, and 30 glycine residues and without a linker, where the distance between the center of mass of the two subunits was constrain. For each of the studied homodimers several constant temperature molecular dynamics simulations were performed (using the simulation package AMBER6 as an integrator; ref. 6) starting from either the dimeric conformation or the unfolded monomers. In the free energy calculations, the energy terms associated with the linker residues were not taken into account to enable a comparison between a dimer and an isolated monomer folding.

Studied Homodimers

In addition to the 11 homodimers that studied by Gō model simulations (Table 1), the phase diagram (Fig. 1) includes other homodimers that are experimentally classified as either two- or three-state dimers. The two-state homodimers include 2zta (GCN4 leucine zipper) (7) and 1jun (c-Jun). The three-state homodimers include: 1dor (dihydroorotate dehydrogenase) (8), 1qm4 (methionine adenosyltransferase) (9), 1a43 (HIV-1 capsid protein) (10), 1tim (triose phosphate isomerase) (11), 1lyn (sperm lysin) (12), 1tar (aspartate aminotransferase) (13), 1glq (glutathione *S* transferase) (14), and 1gsd (Glutathione transferase) (15).

1. Ueda, Y., Taketomi, H. & Go, N. (1978) *Biopolymers* **17**, 1531-1548.
2. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000) *J. Mol. Biol.* **298,** 937-953.
3. Koga, N. & Takada, S. (2001) *J. Mol. Biol.* **313,** 171-180.
4. Kaya, H. & Chan, H. S. (2003) *J. Mol. Biol.* **326,** 911-931.
5. Robinson, C. R. & Sauer, R. T. (1996) *Biochemistry* **35,** 13878-13884.
6. Case, D. A., Pearlman, D. A., Caldwell, J. W., III, T. E. C., Ross, S. W., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V. & Cheng, A. L. (1999) AMBER6 (Univ. of California, San Francisco).
7. Neet, K.E. & Timm, D.E. (1994) *Protein Sci.* **3**, 2167-2174.
8. Ottosen, M.B. Bjornberg, O., Norager, S., Larsen, S., Palfey, B. A. & Jensen, K. F. (2002) *Protein Sci.* **11,** 2575-2583.
9. Gasset, M., Alfonso, C., Neira, J.L., Rivas, G. & Pajares, M.A. (2002) *Biochem. J.* **361**, 307-315.
10. Mateu, M.G. (2002) *J. Mol. Biol.* **318**, 519-531.
11. Borchert, T., Zeelen, J. P., Schliebs, W., Callens, M., Minke, W., Jaenicke, R. & Wierenga, R. K. (1995) *FEBS Lett.* **367**, 315-318.
12. Shaw, A., Fortes, P.A.G., Stout, C.D. & Vacquier, V.D. (1995) *J. Cell Biol.* **130**, 1117-1125.
13. Herold, M. & Kirschner, K. (1990) *Biochemistry* **29**, 1907-1913.
14. Aceto, A., Caccuri, A. M., Sacchetta, P., Bucciarelli, T., Dragani, B., Rosato, N., Federici, G. & Di Ilio, C. (1992) *Biochem. J.* **285**, 241-245.

15. Luo, J. K., Stevens, J. M., Wallace, L. A., Kaplan, W., Armstrong, R. N. & Dirr, H. W. (2000) *Biochemistry* **39**, 12336-12344.