



Protein folding: coming to terms with cooperativity

Ariel Fernández

James Franck Institute & Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, USA;
Instituto de Matemática, Universidad Nacional del Sur, Consejo Nacional de Investigaciones Científicas y Técnicas,
Bahía Blanca 8000, Argentina

Most proteins fold *in vitro* under various renaturation conditions and in so doing, they create a gamut of local solvent environments. The conformation dependence of such environments arises since residues are clustered within a large-scale organization and polar and hydrophobic groups have distinctive ways of organizing solvent around themselves, shaping solvation hulls or cavities. Furthermore, the pairwise interactions of the peptide chain are inevitably sensitive to the environments the chain itself is creating [1-3]. This picture underscores what researchers have termed cooperativity, a central feature of the so-called hydrophobic collapse which suggests a concertedness of folding events requiring the participation of distant parts of the chain.

Nevertheless, in dealing with the rôle of cooperativity at the onset of large-scale organization, most approaches have so far avoided dealing with the wanton complexities involved in treating the solvent structure explicitly, let alone incorporating the conformation dependence of local solvent contexts [4-6]. On the other hand, in dealing with the protein folding problem by treating the solvent implicitly, theoreticians have based their analysis on intramolecular potentials which are pairwise additive: they seem reluctant to pay the price for circumventing the treatment of the solvent, despite the fact that by doing so they might not be able to effectively account for cooperativity [1-3].

Thus, the inconsistency of standard approaches becomes apparent since pairwise interactions (hydrophobic, coulombic, dipole-dipole, etc.) are all sensitive to the local dielectric and solvent structure the chain creates and therefore, the intramolecular potential energy must necessarily incorporate higher-order correlations [3]. In other words, we are faced with a many-body problem. But, what kind of many-body problem?

To shed light on this question, let us first focus on a specific example: an intramolecular (amide-carbonyl) hydrogen bond formed in the vicinity of a hydrophobic moiety is less prone to be targeted for water attack—which would entail a concurrent disruption of the water structure nearby—than if it were surrounded by bulk solvent [3]. This implies that solvation of the exposed chain becomes kinetically and thermodynamically inhibited by the presence of a neighbouring hydrophobic group, resulting in a net stabilization of the hydrogen bond. In this regard, researchers have recently emphasized the role of context [3, 4, 7, 8], which sometimes supersedes local propensities as a structure determinant [7, 8].

This prompts us to introduce an entirely new picture of the folding process in which the chain compactification becomes the result not only of clustering hydrophobic forces *per se*, but also of a “self-kosmotropic effect” in which

the chain organizes solvent locally with dramatic effects on its own intramolecular interactions [3, 7]. Thus, the hydrophobic collapse should not be regarded merely as a clustering event but also as a backbone-desolvation event. This effect leads to a wholesale protection of hydrogen bonds, an essential condition to warrant the development of secondary structure by scaffolding the collapse-triggering nucleus.

Furthermore, hydrophobic-polar (h-p) mismatches, energetically penalized within a zeroth-order context-insensitive pairwise approximation, might find an actual thermodynamic trade-off as soon as higher order correlations are included (cf. [3]): a hydrophobic residue k might approach a polar unit i if the latter is itself engaged in an amide-carbonyl backbone hydrogen bond with a third unit j . Thus, the protection—stabilization—of the (i,j) -interaction might easily compensate for the h-p mismatch between k and i . The reader should keep in mind that a hydrogen bond in a desolvated *quasi in-vacuo* environment is roughly ten times more stable than its “in bulk” counterpart [4].

This discussion makes it obvious that we must abandon the standard and widely used scheme: $U^{(o)} = \sum_{i<j} U^{(o)}_{ij}$, where $U^{(o)}$ is the context-insensitive in-bulk intramolecular potential energy and $U^{(o)}_{ij}$ the zeroth-order pairwise (i,j) -contribution; and replace it appropriately to incorporate the sensitivity to conformation-dependent environments.

In this regard, I propose the following *Ansatz*: the solvent environment affecting the (i,j) -contribution is articulated through third-body hydrophobic residues spatially close to the (i,j) -pair but not covalently bonded to either i or j . These third-body influences may be incorporated in the form of correlation factors which rescale the in-bulk (i,j) -term.

In a minimal model, at least two parameters would be needed: let r^* be a critical distance to either i or j beyond which no influence is exerted by hydrophobic residue k , and let L be a parameter measuring the sensitivity of the pairwise interaction to the third-body presence. Within this frame we get:

$$U_{ij} = U^{(o)}_{ij} \times [\prod_k C(i, j, k)], \quad (1)$$

where the correlation factor $C(i, j, k)$ is given by

$$C(i, j, k) = [1 + L \times h(r^* - d(i, k)) \times h(r^* - d(j, k))]^{g(i, j)} \quad (2)$$

Here $h(x)$ is the Heaviside function ($h(x) = 1$ if $x > 0$ and $h(x) = 0$ if $x < 0$) and the exponent $g(i, j)$ is +1 if the favourable (i, j) -interaction ($U^{(o)}_{ij} < 0$) is further stabilized

by desolvation, like a hydrogen bond or an attractive p-p interaction, or the h-p repulsion ($U^{(o)}_{ij} > 0$) is further destabilized as the polar unit becomes progressively more buried, while $g(i,j) = -1$ if the favorable (i,j)-interaction is destabilized by desolvation, like a hydrophobic attraction ($U^{(o)}_{ij} < 0$). In this way, the product of Heaviside functions ensures that only those third-body units within a critical distance from i and j will actually exert their influence, itself measured by the parameter L . Thus, the cumulative effect of three-body correlations quantified in eqns (1) and (2) effectively models the solvent environment shaped by nearby hydrophobic groups.

This is a primitive *Ansatz*, but it seems to be the first to deal effectively with an emerging view which suggests that novel approaches to the folding problem should incorporate the dual role of hydrophobic residues which act not only as clustering elements, but also as backbone desolvators. This second role is paramount to understand cooperativity, the *bête noire* of the theory. Thus, the protein folding process may be regarded as a struggle for the survival of intramolecular hydrogen bonds. In this scenario, the protein chain organizes solvent as it folds onto itself and in so doing becomes an endogenous kosmotrope, making surrounding water molecules less prone to attack the intramolecular hydrogen bonds.

REFERENCES

1. Vendruscolo, M. & Domany, E. Pairwise contact potentials are unsuitable for protein folding. *J. Chem. Phys.* **109** (1998) 11101–11108.
2. Vendruscolo, M. & Domany, E. Efficient dynamics in the space of contact maps. *Folding & Design* **3** (1998) 329–336.
3. Fernández, A. Conformation-dependent environments in protein folding. *J. Chem. Phys.* **114** (2001) 2489–2502.
4. Baldwin, R. L. & Rose, G. D. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* **24** (1999) 26–33.
5. Muñoz, V. & Eaton, W. A. A simple model for calculating the kinetics of protein folding from three dimensional structure. *Proc. Natl Acad. Sci. USA* **96** (1999) 11311–11316.
6. Fernández, A., Kostov, K. & Berry, R. S. From residue matching patterns to protein folding topographies: general model and bovine pancreatic trypsin inhibitor. *Proc. Natl Acad. Sci. USA* **96** (1999) 12991–12996.
7. Fernández, A., Colubri, A. & Berry, R. S. Topology to geometry in protein folding: beta-lactoglobulin. *Proc. Natl. Acad. Sci. USA* **97** (2000) 14062–14066.
8. Minor, D. & Kim, P. S. Context-dependent secondary structure formation of a designed protein sequence. *Nature (Lond.)* **380** (1996) 730–734.

* * *