Journal of Biology

BioMed Central

Minireview

Liaison amid disorder: non-native interactions may underpin long-range coupling in proteins

Hue Sun Chan and Zhuqing Zhang

Address: Department of Biochemistry, Department of Molecular Genetics, and Department of Physics, University of Toronto, 1 King's College Circle, MSB 5207, Toronto, Ontario, M5S 1A8 Canada.

Correspondence: Hue Sun Chan. Email: chan@arrhenius.med.toronto.edu

Published: 13 March 2009

Journal of Biology 2009, **8**:27 (doi:10.1186/jbiol126)

The electronic version of this article is the complete one and can be found online at http://jbiol.com/content/8/3/27

Abstract

A lattice-model study of double-mutant cycles published in *BMC Structural Biology* underscores how interactions in non-native conformations can lead to thermodynamic coupling between distant residues in globular proteins, adding to recent advances in delineating the often crucial roles played by disordered conformational ensembles in protein behavior.

How do the conformational structures, dynamics and biological function of a protein emerge from the interactions among its amino acid residues? A significant part of current ideas about protein behaviors is based on structures in the Protein Data Bank (PDB) and notions of contact-like interactions between amino acid residues in spatial proximity. While useful, this picture is limited. In particular, studies of allostery and mutational analyses have demonstrated that energetic coupling can exist between residues at positions far apart in a protein's native structure. An intriguing possibility is that such apparently long-range coupling may arise from the residues' transient association in the unfolded state. This scenario was elucidated by an extensive computational study using two-dimensional lattice protein models published recently in *BMC Structural Biology* by the groups of Ron Unger and Amnon Horovitz (Noivirt-Brik *et al.* [1]). Their study provides a theoretical framework that will be useful for guiding future experiments. It also highlights the power and versatility of simple lattice modeling. Despite the highly coarse-grained representations of polypeptide chains used, this decades-old practice offers conceptual clarity and has been proved effective time and again in discovering and elucidating fundamental biophysical principles.

Characterizing energetic coupling by double-mutant cycle

Energetic coupling between amino acid residues is difficult to discern from the static folded structure of a protein alone. Double-mutant cycle (DMC) is a direct perturbative technique to assess the degree to which the consequences of mutations at two different sites are correlated. DMC compares the sum of effects of two single mutations on two sites (one at a time) and the effect of double mutations on both of the sites. Often, as in Noivirt-Brik *et al.* [1], the effect of interest is the free energy of folding, ΔG (native state more stable for more negative ΔG). If $\Delta\Delta G(m_1)$, $\Delta\Delta G(m_2)$, and $\Delta\Delta G(m_1,m_2)$ are, respectively, the changes in ΔG resulting from two single mutations and from the double mutations ($\Delta\Delta G$ equals ΔG of the mutant minus that of the wild type), coupling is quantified by an 'interaction free energy' $\Delta\Delta G_{\text{int}} = \Delta\Delta G(m_1,m_2) - [\Delta\Delta G(m_1) + \Delta\Delta G(m_2)]$. The two sites are energetically independent if the mutational

effects are additive ($\Delta\Delta G_{int} = 0$). Otherwise they are coupled, wherein the native state is either stabilized ($\Delta\Delta G_{int} < 0$) or destabilized ($\Delta\Delta G_{int} > 0$) by coupling. Energetic coupling may also be estimated using a bioinformatics approach based on evolutionary assumptions. This indirect method has also identified likely long-range interactions, for example in PDZ domains [2].

Long-range coupling in proteins can have multiple physical origins

The existence of long-range coupling should not be surprising. After all, the folded state of a protein may be viewed as an elastic solid [3]. As such, the vibrational dynamics of distant sites can be coupled and a 'pathway of energetic connectivity' [2] inside the folded protein is physically plausible. Without discounting such folded-state mechanisms, Noivirt-Brik *et al*. [1] tackled another possibility, focusing mainly on the unfolded (denatured) state. Because native stability is determined by the balance between the folded and unfolded states, interactions in the unfolded states can have an impact on coupling. This possibility was overlooked when unfolded states were envisaged to be devoid of significant contact interactions (Figure 1a), a picture rooted in a simplistic view of cooperative, two-state-like folding. However, it is physically reasonable to expect, for instance, that two hydrophobic residues can associate in the unfolded state even if they are not in contact in the folded structure. This idea is embodied in the well-studied hydrophobic-polar (HP) model, which aims to capture essential protein physics by using only two residue types (Figure 1b). The HP model [4] illustrates the same principle as that deduced from the model with four residue types used by Noivirt-Brik *et al*. [1]. Figure 1b shows two residues (red and blue) exposed in the folded structure, but they can contact other residues as well as each other in the unfolded state.

Reverse hydrophobic effect and other manifestations of non-native interactions

Does the model in Noivirt-Brik *et al*. [1] and that shown in Figure 1b reasonably mimic reality? Ample evidence supports the existence of non-native interactions in protein unfolded states [5]. As early as 1990, the hydrophobicity of an exposed residue in the Cro repressor from bacteriophage λ was found to correlate negatively with the stability of the protein. Dubbed the 'reverse hydrophobic effect' to contrast it with the usual role of hydrophobicity in stabilizing the folded state, the phenomenon was rationalized by the proposal that the residue is partially buried; that is, it has non-native contact(s) in the unfolded state [6]. The variation in the denaturant dependence of native stability (equilibrium *m*-value, defined as the rate of decrease in native stability with respect to increase in denaturant concentration) of staphylococcal nuclease observed in earlier site-directed mutagenesis experiments also indicated variable hydrophobic burial in the unfolded state. Recent experiments suggested that non-native ionic interactions are present as well in the unfolded states of the amino-terminal domain of ribosomal protein L9 (see [5] and references therein).

Simple lattice protein models are an effective conceptual tool

Lattice models have been successful in accounting for some of these phenomena. An early HP square-lattice model study elucidated how mutations can lead to substantial changes in *m*-value, as found for staphylococcal nuclease experimentally [4]. Figure 1b shows three HP model mutants that exhibit reverse hydrophobic effect ($\Delta\Delta G < 0$). From their $\Delta\Delta G$ values, $\Delta\Delta G_{int}$ for the model DMC was determined to be significantly negative (green curve in the left plot of Figure 1b). This result indicates a long-range coupling (between the red and blue residues) underpinned by non-native interactions in the unfolded state of the HP model.

As illustrated by these examples and similar analyses by Noivirt-Brik *et al.* [1], lattice models are a powerful investigative tool. Common notions about protein energetics are sometimes fuzzy. Their precise ramifications are often obscure owing to a lack of discipline from an explicit consideration of chain connectivity and conformational entropy [7]. Lattice models account for these key ingredients, albeit in a simplified fashion. By virtue of their computational tractability, lattice models can clarify the logic between assumptions and testable consequences, generate new hypotheses, and ask 'what if' questions to advance conceptual understanding.

It goes without saying that lattice models are limited. Learning from both their strengths and limitations, concrete progress often requires comparative evaluation of models embodying different physical ideas. Notably, extensive analyses over the past decade have shown that traditional lattice protein models - the HP model included - fold much less cooperatively than real, two-state proteins [7]. In the light of this knowledge, it is instructive to explore whether the predictions about long-range coupling obtained by Noivirt-Brik *et al*. [1] and from the HP model are robust.

Folding cooperativity may dampen but cannot eliminate non-native interactions

Contact interactions such as that in the model used by Noivirt-Brik *et al*. [1] and the HP model do not fully capture protein energetics. More subtle physical chemistry has

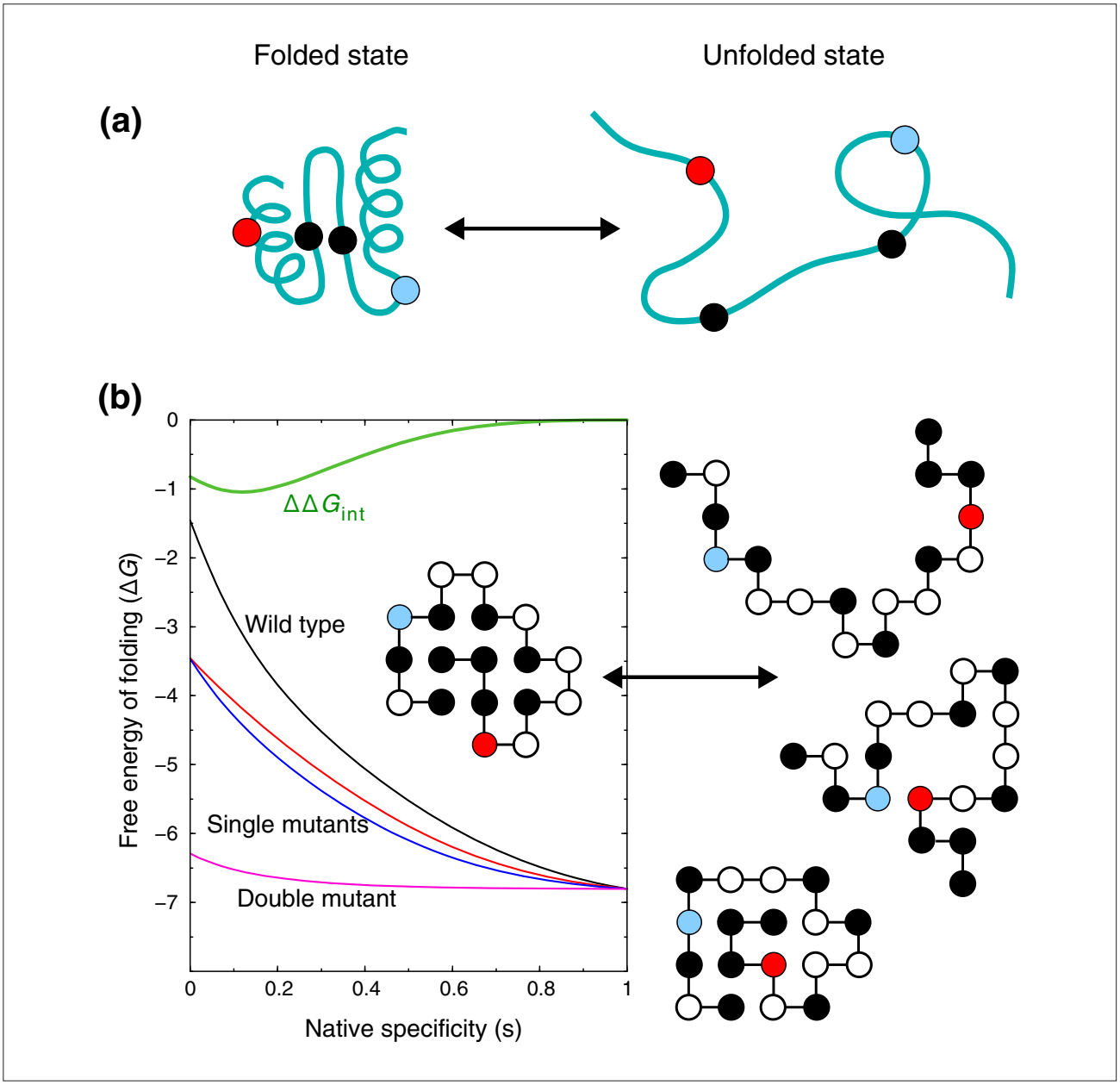

Figure 1

Non-native interactions in the unfolded state affect native protein stability. **(a)** Schematic diagram of the equilibrium between the natively folded and the unfolded (non-native, or denatured) states. Selected exposed and buried residues are marked by circles. A simplistic view of cooperative folding envisages all conformations in the unfolded ensemble to be open, with negligible residue-residue contact, as exemplified by the chain on the right. **(b)** Double-mutant cycles (DMC) in square-lattice models are simulated using different hypothetical interaction schemes to explore a range of native specificity - from the HP model (s = 0), which allows for non-native interactions [4], to the Gō model (s = 1), which precludes them (the Gō model was formulated originally by Nobuhiro Gō and co-workers in 1975 and favors only native interactions). Native specificity is the ability of a set of interactions to discriminate against non-native attractions and is indicated here by the parameter s. Hydrophobic (H) and polar (P) residues are drawn, respectively, as black and white circles. The wild-type sequence has H at both mutation sites (red and blue). Two single mutants and one double mutant that preserve the wild-type native structure (which is shown on the left) are created by changing either one or both of these sites to P. Depicted on the right are three example unfolded conformations (in an ensemble of around 6 million) that have (from top to bottom) no, one, and two contacts involving the mutation sites. The plot on the left shows how the free energy of folding (ΔG) of the wild type (black curve) and the mutants (red, blue, and magenta curves) as well as the coupling energy $\Delta\Delta G_{int}$ (green curve) depend on the native specificity parameter s. Results are presented for model contact energy $\varepsilon = -5k_BT$, where k_B is the Boltzmann constant and T is absolute temperature. Free energies are in units of k_BT.

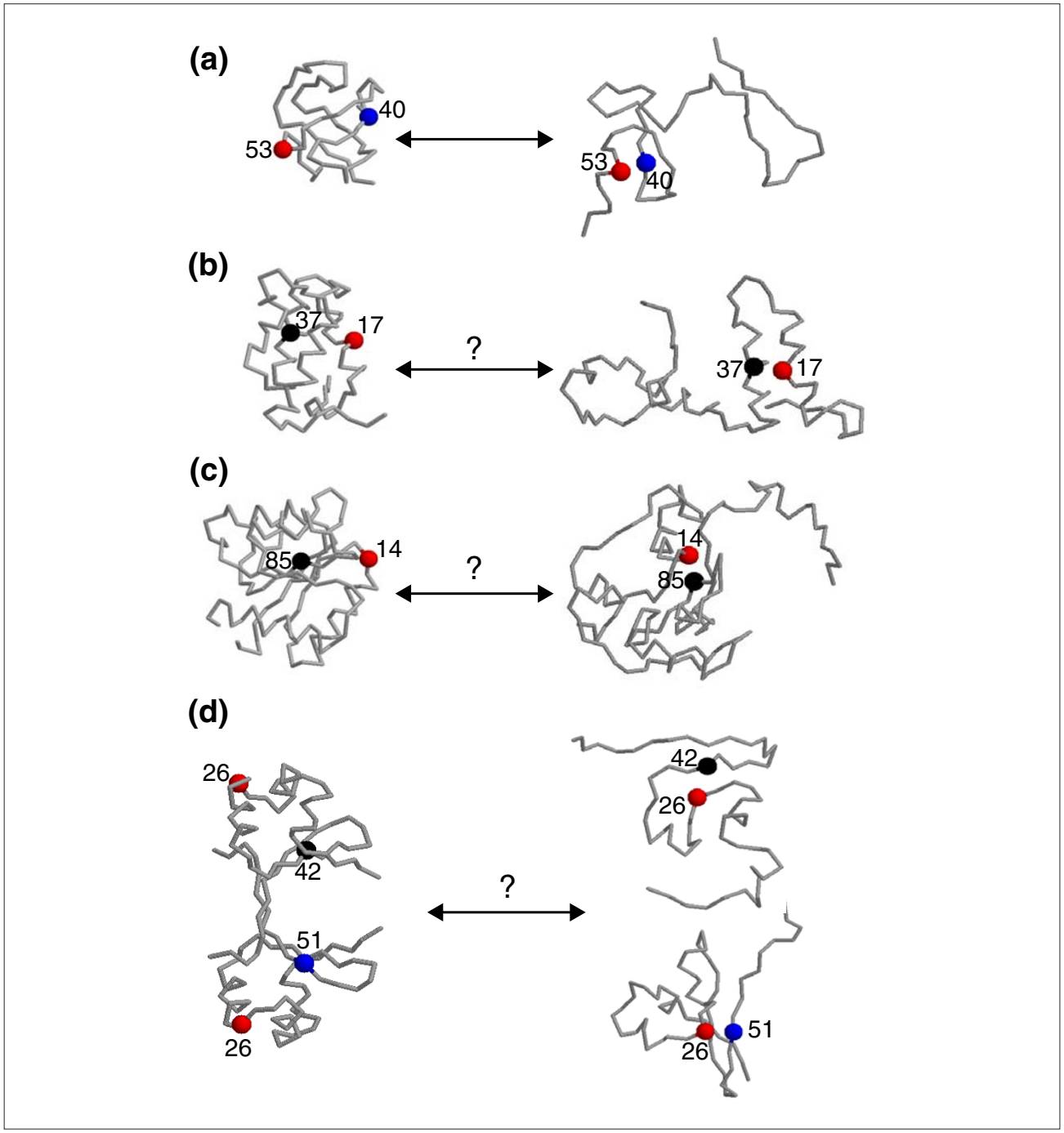

Figure 2
Non-native interactions underpin the reverse hydrophobic effect. Representative unfolded conformations (right) based on PDB structures (left) were simulated using a coarse-grained continuum chain model that allows sequence-dependent non-native hydrophobic interactions [10]. (**a**) An unfolded conformation (right) of a double mutant of the Fyn SH3 domain (PDB 1shf) containing a non-native contact between positions 40 and 53 as implicated by DMC [10]. (**b-d**) Residue positions in red are known experimentally to contribute to the reverse hydrophobic effect [6,8,9]. Those in black or blue are their most likely unfolded-state non-native interacting partners in our simulations. (b) The H1P variant of bacterial immunity protein Im9 (PDB 1imq) [9], non-native contact Ile17-Val37. (c) Chemotactic protein CheY (PDB 3chy) [8], Phe14-Met85. (d) λ Cro repressor (PDB 5cro), which unfolds from a dimer to two monomer chains [6], Tyr26-Leu42 and Tyr26-Tyr51. Question marks in (b-d) emphasize that the predicted non-native interactions are yet to be tested by experiment.

enabled higher native specificity and more cooperative folding to be achieved in natural proteins. Hence, the probabilities of non-native interactions and the long-range coupling they engender in real proteins are likely to be lower than those stipulated by these models. This point is illustrated in Figure 1b using a class of energy functions $E = (1 - s)\ E_{HP} + sE_{G\bar{o}}$ interpolating between the HP model (E_{HP}) and a Gō model ($E_{G\bar{o}}$) that favors only native interactions (formulated originally by Nobuhiro Gō and co-workers in 1975; see reference to Gō in [7]). Here, s is the weight of Gō energy and thus a parameter for native specificity. As the strength of favorable non-native interactions decreases with increasing s, the associated long-range coupling diminishes. Could all non-native interactions be 'designed out' by evolution?

Experiments on reverse hydrophobic and other effects of non-native interactions suggested otherwise. There are physical limits to evolutionary and artificial protein design. Unlike Gō models, non-native interactions are present in some real proteins that fold cooperatively [5,10]. (Conversely, Gō models are often insufficiently cooperative [7].) From a modeling standpoint, a mixture of HP- and Gō-like components (with s somewhere between 0 and 1) may best capture the balance between physical constraints and the drive toward native specificity. A continuum version of such a modeling construct has successfully predicted non-native interactions in the unfolded and folding transition states of the SH3 domain of the protein kinase Fyn [10]. As an illustration of the method, Figure 2 applies the same model to obtain putative non-native interactions in several other proteins [6,8,9].

From biophysics to biological functions of non-native protein conformations

The main point of the study of Noivirt-Brik *et al.* [1] - that non-native interactions are the origin of some long-range coupling - is thus on a firm physical and molecular biological footing. A deeper question is whether non-native interactions are mere annoying necessities imposed by physics, a feature that should be designed out if possible by evolution, or whether they can serve biological purposes? With our increasing appreciation of the regulatory functions of intrinsically disordered proteins [11], there is no reason to believe that biology would not exploit every opportunity presented by physics. A case in point is that non-native conformations can have 'promiscuous' biological functions different from the dominant function of a protein, and that selection for promiscuous functions can speed up evolution considerably [12]. In this case as well, simple lattice modeling has afforded the pertinent biophysical principles (see accompanying article of [12]). More discoveries lie ahead as protein scientists broaden our sight beyond well-ordered folded native structures.

Acknowledgements

We acknowledge support from the Canada Research Chairs Program and funding from the Canadian Institutes of Health Research to HSC (grant MOP-84281). We thank Matt Cordes of University of Arizona for a helpful discussion on the λ Cro repressor. Because journal policy placed a limit on the number of references, we apologize to colleagues whose important contributions we were not able to cite.

References

1. Noivirt-Brik O, Unger R, Horovitz A: **Analysing the origin of long-range interactions in proteins using lattice models.** *BMC Struct Biol* 2009, **9:**4
2. Lockless SW, Ranganathan R: **Evolutionarily conserved pathways of energetic connectivity in protein families.** *Science* 1999, **286:**295-299.
3. Haliloglu T, Bahar I, Erman B: **Gaussian dynamics of folded proteins.** *Phys Rev Lett* 1997, **79:**3090-3093.
4. Shortle D, Chan HS, Dill KA: **Modeling the effects of mutations on the denatured states of proteins.** *Protein Sci* 1992, **1:**201-215.
5. Bowler BE: **Thermodynamics of protein denatured states.** *Mol BioSyst* 2007, **3:**88-99.
6. Pakula AA, Sauer RT: **Reverse hydrophobic effects relieved by amino-acid substitutions at a protein surface.** *Nature* 1990, **344:**363-364.
7. Chan HS, Shimizu S, Kaya H: **Cooperativity principles in protein folding.** *Methods Enzymol* 2004, **380:**350-379.
8. Muñoz V, Lopez EM, Jager M, Serrano L: **Kinetic characterization of the chemotactic protein from *Escherichia coli*, CheY. Kinetic analysis of the inverse hydrophobic effect.** *Biochemistry* 1994, **33:**5858-5866.
9. Cranz-Mileva S, Friel CT, Radford SE: **Helix stability and hydrophobicity in the folding mechanism of the bacterial immunity protein Im9.** *Protein Eng Des Sel* 2005, **18:**41-50.
10. Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, Davidson AR, Chan HS: **Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding.** *Proc Natl Acad Sci USA* 2008, **105:**9999-10004.
11. Mittag T, Forman-Kay JD: **Atomic-level characterization of disordered protein ensembles.** *Curr Opin Struct Biol* 2007, **17:**3-14.
12. Amitai G, Gupta RD, Tawfik DS**: Latent evolutionary potentials under the neutral mutational drift of an enzyme.** *HFSP J* 2007, **1:**67-78.