Intrinsic protein disorder and protein function

Jintao Liu[1], James R. Faeder[2,3], and Carlos J. Camacho[2,4]

***Short Abstract* — We studied the role of intrinsic protein disorder in protein function with a standard thermodynamic model. By relating disorder with the free energy of folding, we showed that: For *Protein Binding*, to maintain high binding efficiency, weak binding prefers ordered structure, while strong binding can tolerate disorder. For *Catalysis*, ordered structure is preferred to achieve high catalytic activity. In addition, disorder in strong *Binding* proteins can increase the specificity of molecular recognition. Further genomic analysis supports our predictions, and also raises interesting questions about the role of disorder in eukaryotic transcriptional proteins and prokaryotic proteins.**

***Keywords* — thermodynamic analysis, folding free energy, optimize function, genome wide survey.**

MANY proteins have been found to be without stable structure in their native states [1]. They are called intrinsically disordered proteins. Their ubiquitous presence undercuts the principle that a protein's structure determines its function [1]. It has been suggested that disorder itself plays a functional role by, *e.g.*, allowing for multiple interaction partners [2] and enabling functional diversity [3-5], which are particularly important in cell signaling and cancer [6]. However, the origins of disorder and its role in protein's function are still not well understood. This motivated us to look for the general principles that might link protein function and disorder. We constructed a thermodynamic model of the two broadest functional categories in the Gene Ontology (GO) [7] classification: *Protein Binding* and *Catalytic Activity*, and showed that evolution may act differentially upon the level of disorder for these two categories to optimize protein function. A comparative genomic analysis of disorder further supports this idea.

It has been found that folding is involved in the functioning of disordered proteins [1]. In our analysis, we use the standard model of folding in which a positive folding free energy (ΔG_f) favors the disordered state [8]. We assume that only the folded state is functionally active. Without losing generality, we further assume that folding is independent of substrate binding, since our conclusion only relies on (quasi)equilibrium properties.

We model *Binding* as the formation of stable complexes, and *Catalysis* as the rate of substrate conversion. For each case, we relate optimal performance to ΔG_f over the range of parameters found in nature. We find that for *Binding* proteins optimal binding requires weakly-interacting proteins to have ordered structure, while strongly-interacting proteins may be disordered but still maintain high binding efficiency. Optimal *Catalysis*, on the other hand, requires ordered structure ($\Delta G_f < 0$). Moreover, disorder in strong *Binding* proteins can increase the specificity of molecular recognition.

To support our model, we did genome-wide analysis of the disorder broken down by functional categories according to the GO classification. For each protein, the fraction of disordered amino acid residues was estimated using disorder prediction tools (VSL2B [9], DisEMBL [10] and FoldIndex [11]). In eukaryotes, *Catalysis* strongly favors ordered structures, while *Binding* proteins exhibit a broad distribution of disorder. These findings are consistent with the predictions from our thermodynamic analysis. Two additional findings from our survey deserve further study: the high fraction of disorder in eukaryotic proteins with *Transcription Regulator Activity*, which implies the importance of disorder in transcription regulation; and the low levels of disorder in prokaryotic proteins, which reveals intriguing difference between prokaryotes and eukaryotes on the molecular level. This last finding may be explained if protein-protein interactions in prokaryotes are generally weaker than in eukaryotes, which preliminary analysis suggests may be the case.

Acknowledgements: This work was partially supported by the National Science Foundation MCB-0444291/0744077.

[1]Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA. E-mail: jil41@pitt.edu

[2]Department of Computational Biology, University of Pittsburgh, Pittsburgh, PA

[3]E-mail: faeder@pitt.edu

[4]E-mail: ccamacho@pitt.edu

REFERENCES

[1] Wright PE and Dyson HJ, *J. Mol. Biol.* **293**, 321 (1999).
[2] Haynes C *et al.*, *PLoS Comput. Biol.* **2**, e100 (2006).
[3] Romero PR *et al.*, *Proc. Natl. Acad. Sci. USA* **103**, 8390 (2006).
[4] Ward JJ *et al.*, *J. Mol. Biol.* **337**, 635 (2004).
[5] Xie H *et al.*, *J. of Proteome Res.* **6**, 1882 (2007).
[6] Iakoucheva LM *et al.*, *J. Mol. Biol.* **323**, 573 (2002).
[7] Ashburner M *et al.*, *Nat. Genet.* **25**, 25 (2000).
[8] Schulz GE, *Molecular Mechanism of Biological Recognition*, Elsevier/North-Holland Biomedical Press, pp. 79-94 (1979).
[9] Peng K *et al.*, *BMC Bioinformatics* **7**, 208 (2006).
[10] Linding R *et al.*, *Structure* **11**, 1453 (2003).
[11] Prilusky J *et al.*, *Bioinformatics* **21**, 3435 (2005).

Nothing should be here on page 2! Please limit your abstract to a single page, and submit it as a one-page editable file (e.g., Word .doc format).