The protein trinity—linking function and disorder

A. Keith Dunker and Zoran Obradovic

Interpreting function in terms of specific three-dimensional structure has dominated thinking about proteins for more than 100 years, starting with the lock-and-key proposal of Fischer[1] and continuing with the equating of denaturation with loss of specific structure by Wu[2] and independently at a slightly later date by Mirsky and Pauling[3]. This dependence of function on structure is even embedded in our language: unfolded protein and denatured protein are used interchangeably. Furthermore, the avalanche of protein three-dimensional structures determined by X-ray diffraction and by nuclear magnetic resonance (NMR)[4] has diverted attention away from alternative views.

Numerous counterexample proteins have surfaced over the years—proteins for which lack of three-dimensional structure is required for function. One clear example is calcineurin, a serine/threonine phosphatase that becomes activated by the binding of the Ca^{2+}–calmodulin complex to a region that exists as a disordered ensemble[5,6]. The disorder spans the calmodulin binding site and is essential for calcineurin function. That is, when calmodulin binds to its target helix, the helix becomes completely surrounded[7]. Thus, the open, flexible disordered region of calcineurin provides the space needed by calmodulin so it can completely surround its target helix.

Even though hundreds of other examples of proteins with intrinsic disorder have surfaced over the past 50 years, review articles on this topic are only just now beginning to appear[8–10]. Wright and Dyson[8] suggested that the existence of proteins with intrinsic protein disorder calls for a reassessment of the protein–structure–function paradigm.

Since amino acid sequence determines three-dimensional structure, amino acid sequence should also determine lack of three-dimensional structure. Furthermore, if intrinsic disorder provides the basis for some biological functions, then the operation of natural selection should conserve the lack of folding and thereby preserve those functions that depend on this property.

If disorder is indeed encoded by the amino acid sequence, then predictors of disorder should exceed the accuracies expected by chance. Work in our group has used literature and database searches to collect a set of proteins structurally characterized to have regions of disorder, some of which were indicated by NMR to be wholly disordered under physiological conditions. Using this set of proteins with intrinsic disorder, we have set out to construct the predictors needed to test the hypothesis.

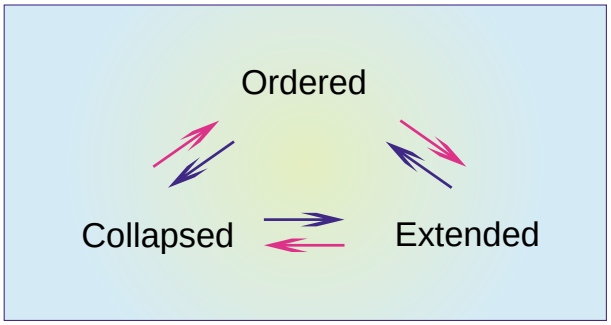

Figure 1. The protein trinity. Native proteins may exist in one of three states—ordered, collapsed-disordered, or extended-disordered.

This large jump in putatively disordered proteins in multicelled, rather than single-celled, organisms is both remarkable and unexpected

For datasets containing equal numbers of ordered and disordered residues, our predictors of natural disordered regions consistently identify out-of-training examples of order and disorder, with accuracies that were initially about 70% (ref. 11) and that now are above 80% (ref. 12). The latter study contained 16,785 putatively disordered residues from 145 nonhomologous proteins, balanced by an equal number of ordered residues. As these accuracies are far above the 50% expected by chance, the hypothesis that intrinsic disorder is encoded by the sequence is strongly supported. To test whether intrinsic disorder tends to be conserved, we have also used our predictor algorithm on aligned members of a protein family. The estimated tendencies for disorder were conserved despite substantial variations in amino acid sequence within the putative regions of disorder[13].

When we used our algorithm to search the amino acid sequences from more than 30 prokaryota, archaebacteria, and eukaryota and summarized our findings as percentage of proteins in each proteome predicted to contain disordered regions of 40 consecutive residues or longer, we observed an interesting difference. In 22 bacteria and 7 archea studied, the percentages of proteins with predicted regions of disorder ranged from 7% to 33% and from 9% to 37%, respectively[14]. In contrast, in the five eukaryota tested, disorder ranged from 36% to 63% (ref. 14). This large jump in putatively disordered proteins in multicelled, rather than single-celled, organisms is both remarkable and unexpected.

Why the large jump in intrinsic disorder for the eukaryota? Our results are too new and too unexplored to answer this question with certainty, but there are some hints to be tested by further experiments. In our training sets of disordered proteins, we notice that, like the calcineurin example given above, many of the disordered regions and most if not all of the completely disordered proteins are involved in cell signaling or regulation. Furthermore, from our small set of examples, the association between regulatory function or signaling and intrinsic disorder appears to be conserved across all three kingdoms. Qualitatively, it seems reasonable that highly flexible proteins would provide a better basis for responding to changes in the environment than rigid ones.

More specifically, disordered regions can bind partners with both high specificity and low affinity[15], so the regulatory interactions

A. Keith Dunker is professor of biophysics at the School of Molecular Biosciences, Washington State University, Pullman, WA 99124 (dunker@disorder.chem.wsu.edu). Zoran Obradovic is professor and director at the Center for Information, Science and Technology, Temple University, Philadelphia, PA 19122 (zoran@ist.temple.edu).

can be specific and also can be easily dispersed: turning a signal off is as important as turning it on. Also, as pointed out by Wright and colleagues[16], conformational disorder mediates binding diversity, so a significant advantage of intrinsic disorder is to allow one regulatory region or one regulatory protein to bind to many different partners. The ability to partner with many other proteins and other ligands, such as nucleic acids, might be of central importance. Indeed, database comparisons show that proteins making multiple interactions are more likely to lead to lethality if deleted[17]. An interesting example is HMGI(Y): this founding member of a new protein class called architectural transcription factors binds to 18 known protein partners as well as to several specific DNA structures[18]. In keeping with the present theme, in the absence of its partners, HMGI(Y) is disordered from one end to the other[19].

These observations challenge traditional thinking concerning the specific three-dimensional structure of proteins and their function. Clearly we can no longer ignore the role of disorder in determining protein activity in (especially higher) organisms

These observations challenge traditional thinking concerning the specific three-dimensional structure of proteins and their function. Clearly we can no longer ignore the role of disorder in determining protein activity in (especially higher) organisms. Native proteins should be conceptualized as part of the "protein trinity" (see Fig. 1). They exist in one of three states—ordered, collapsed-disordered, or extended-disordered. And protein function derives from any one of these three states or from transitions between them.

1. Fischer, E. *Ber. Dt. Chem. Ges.* **27**, 2985–2993 (1894).
2. Wu, H. *Chinese J. Physiol.* **1**, 219–234 (1931).
3. Mirsky, A.E. & Pauling, L. *Proc. Natl. Acad. Sci. USA* **22**, 439–447 (1936).
4. Berman, H.M. *et al. Nucleic Acids Res.* **28**, 235–242 (2000).
5. Manalan, A.S. & Klee, C.B. *Proc. Natl. Acad. Sci. USA* **80**, 4291–4295 (1983).
6. Kissinger, C.R. *et al. Nature* **378**, 641–644 (1995).
7. Meador, W.E., Means, A.R. & Quiocho, F.A. *Science* **257**, 1251–1255 (1992).
8. Wright, P.E. & Dyson, H.J. *J. Mol. Biol.* **293**, 321–331 (1999).
9. Dunker, A.K. *et al. J. Mol. Graphics Modeling* **19**, 26–59 (2001).
10. Namba, K. *Genes Cells* **6**, 1–12 (2001).
11. Romero, P., Obradovic, Z., Kissinger, C.R., Villafranca, J.E. & Dunker, A.K. *Proc. I.E.E.E. International Conference on Neural Networks* **1**, 90–95 (1997).
12. Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K. & Obradovic, Z. Methods for improving protein disorder prediction. Proc. International Joint INNS-IEEE Conference on Neural Networks, Washington, DC **4**, 2718–2723 (2001).
13. Iakoucheva, L.M. *et al. Protein Sci.* **10**, 560–571 (2001).
14. Dunker, A.K, Obradovic, Z., Romero, P., Garner, E.C. & Brown, C.J. *Genome Informat.* **11**, 161–171 (2000).
15. Schulz,G.E. In *Molecular mechanism of biological recognition*. (ed. Balaban, M.) 79–94 (Elsevier/North-Holland Biomedical Press, New York, NY; 1979).
16. Kriwacki, R.W., Hengst, L., Tennant, L., Reed, S.I. & Wright, P.E. *Proc. Natl. Acad. Sci. USA* **93**, 11504–11509 (1996).
17. Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, Z.N. *Nature* **411**, 41–42 (2001).
18. Reeves, R. *Genes,* in press (2001).
19. Huth, J.R. *et al. Nat. Struct. Biol.* **4**, 657–665 (1997).