Intrinsically Disordered Proteins 2:1, e970902; October 1, 2014; © 2014 Published with license by Taylor & Francis Group, LLC

Toward a consensus in protein structure nomenclature

Linder C DaSilva[1,4], Thomas Gurry[1,2], and Collin M Stultz[1,2,3,*]

[1]Research Laboratory of Electronics; Massachusetts Institute of Technology; Cambridge, MA USA; [2]Computational and Systems Biology Initiative; Massachusetts Institute of Technology; Cambridge, MA USA; [3]Dept. of Electrical Engineering and Computer Science & Institute for Medical Engineering and Sciences; Massachusetts Institute of Technology; Cambridge, MA USA; [4]ICET-CUAl Federal University of Mato Grosso; Barra do Garças, Brazil

In a recent article, published in *Intrinsically Disordered Proteins*, a valuable consensus view regarding the nomenclature for disordered proteins was presented.[1] In this work the authors present a thoughtful and systemic review of terms that have been used in the literature to describe proteins that sample a heterogeneous set of structures during their biological lifetime. We agree that the term "intrinsically disordered proteins" (IDPs) is an appropriate single descriptor to refer to this particular class of proteins, although it does not fully capture much of the nuanced complexities that are inherent to this class. In what follows we suggest a refinement to this nomenclature based on an analysis of the underlying ensemble that describes the thermally accessible states of a given IDP.

We propose that an approach to IDP taxonomy would benefit from following some of the standards that are used to classify ordered (or folded) proteins. A naming convention consistent with the existing language used to describe folded proteins would facilitate discourse between structural biochemists who study folded proteins and those who study intrinsically disordered biomolecules. For example, it is commonplace to classify folded proteins by their overall secondary structure content, i.e., folded proteins are categorized into larger classes such as all-α, all-β, α/β, etc.[2] Such classifications can, in principle, be extended to disordered proteins by focusing on quantifiable metrics that describe their underlying propensity to form secondary structure. In this vein the "residual" secondary structure propensity can form the basis for additional descriptors intended to refine any IDP taxonomy.

As we and others have argued, structural disorder is not a binary concept in the sense that proteins fall on a spectrum of disorder, e.g., folded proteins have the least disorder and IDPs have the most. In this disorder-order continuum, the amount of protein disorder can be quantified from an analysis of its thermally accessible states.[3] If the ensemble corresponding to the thermally accessible states of the protein contains only one structure (i.e., the protein samples one state), then the associated order parameter is 1. By contrast, if the ensemble contains an infinite number of structures that are all very different from one another then the associated order parameter is close to 0. We recognize that these situations are never realized in practice; however, such examples serve as limiting cases that help to frame subsequent discussion.

The preference for secondary structure content can be similarly quantified, at least in principle. By analyzing the thermally accessible states of the protein, one can obtain an order parameter that describes the average amount of secondary structure content in the underlying ensemble of the protein. A value of 0 denotes that the average secondary structure content in the ensemble is 0% and a value of 1 corresponds to an average secondary structure content of 100%. Just as every protein can be classified along a disorder-order continuum, proteins can similarly be classified along an unstructured-structured continuum. In this sense we use the word "structure" solely to refer to a given protein's propensity to form secondary structure.

Keywords: intrinsically disordered proteins, nomenclature, secondary structure, order parameter

*Correspondence to: Collin M Stultz; Email: cmstultz@mit.edu

Submitted: 06/23/2014

Accepted: 06/23/2014

http://dx.doi.org/10.4161/idp.29700

We can use these two order parameters to classify all proteins. The first order parameter, *O*, describes the amount of disorder in the protein's structural ensemble, and the second order parameter, *S*, quantifies the secondary structure content in the ensemble. However, a strict nomenclature based on this scheme would entail referring to any given protein by its propensity for disorder and its propensity for secondary structure formation—a naming convention that would be too awkward in our view. More importantly, since it is difficult to a priori quantify the extent of heterogeneity that any given protein ensemble has, the notion of using quantitative continuous order parameters has limited utility in practice. Qualitative descriptions, albeit imperfect, based on this formalism are desirable because they enable scientists to engage in meaningful discourse about proteins that fall on ends of the order-disorder continuum. In this sense, we can distinguish two classes that form the basis for a qualitative but useful naming convention for IDPs:

(1) IDP with high secondary structure propensity: These proteins have typically been referred to as premolten or molten globules in the literature,[4,5] i.e., compact states that have no stable tertiary structure but that have considerable secondary structure. Examples in this class include the von Hippel-Lindau tumor suppressor protein[6] and the nuclear coactivator binding domain of CREB binding protein.[7]
(2) IDP with low secondary structure propensity: These proteins have no stable tertiary structure and no appreciable secondary structure as measured via standard techniques such as circular dichroism spectroscopy. Proteins in this class include a number of highly studied IDPs, many of which play a role in neurodegenerative disorders; e.g., tau,[8] α-synuclein,[9] amyloid-β,[10] and huntington[11] proteins. Although these proteins can adopt conformations with well-defined secondary structural elements when bound to particular binding partners, we feel that an initial approach to their taxonomy beings with an assessment of their structural propensities in isolation.

It is important to note that while these classes above apply to entire proteins, they just easily can be applied to regions of proteins. As the authors of the aforementioned review stated, "A majority of eukaryotic protein sequences are chimeras of ordered and disordered regions. . .." Further sub-classifications can be constructed using data that provide clues as to the preferred types of secondary structural elements sampled by the IDP of interest. Such information can be garnered from experimental observables such as secondary chemical shifts and/or molecular models constructed with the aid of experimental data.[12-14]

Our goal in writing this comment is to further the dialog regarding the taxonomy of disordered proteins. By no means do we intend to proffer a definitive naming convention. We only propose that further naming conventions for disordered proteins should in some way follow the existing paradigm for ordered (or folded) proteins.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work was partially supported by CAPES and Science Without Border-Brazil, process nº 2514-13-3.

References

1. Dunker AK, Babu MM, Barbar E, et al. What's in a name? Why these proteins are intrinsically disordered. Intrinsically Disord Proteins 2013; 1:1-4; http://dx.doi.org/10.4161/idp.24157
2. Brindén C-I, Tooze J. *Introduction to protein structure*. 2nd ed. New York: Garland Pub.; 1999.
3. Fisher CK, Stultz CM. Protein structure along the order-disorder continuum. J Am Chem Soc 2011; 133:10022-5; PMID:21650183; http://dx.doi.org/10.1021/ja203075p
4. Ohgushi M, Wada A. 'Molten-globule state': a compact form of globular proteins with mobile side-chains. FEBS Lett 1983; 164:21-4; PMID:6317443; http://dx.doi.org/10.1016/0014-5793(83)80010-6
5. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002; 11:739-56; PMID:11910019; http://dx.doi.org/10.1110/ps.4210102
6. Sutovsky H, Gazit E. The von Hippel-Lindau tumor suppressor protein is a molten globule under native conditions: implications for its physiological activities. J Biol Chem 2004; 279:17190-6; PMID:14963040; http://dx.doi.org/10.1074/jbc.M311225200
7. Kjaergaard M, Teilum K, Poulsen FM. Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proceedings of the National Academy of Sciences*. July 13, 2010 2010;107 (28):12535-12540.
8. Schweers O, Schönbrunn-Hanebeck E, Marx A, Mandelkow E. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. J Biol Chem 1994; 269:24290-7; PMID:7929085
9. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry 1996; 35:13709-15; PMID:8901511; http://dx.doi.org/10.1021/bi961799n
10. Zhang S, Iwata K, Lachenmann MJ, Peng JW, Li S, Stimson ER, Lu Y, Felix AM, Maggio JE, Lee JP. The Alzheimer's peptide a β adopts a collapsed coil structure in water. J Struct Biol 2000; 130:130-41; PMID:10940221; http://dx.doi.org/10.1006/jsbi.2000.4288
11. MacDonald ME, Ambrose CM, Duyao MP, et al.; The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 1993; 72:971-83; PMID:8458085; http://dx.doi.org/10.1016/0092-8674(93)90585-E
12. Wood SP, Tickle IJ, Treharne AM, Pitts JE, Mascarenhas Y, Li JY, Husain J, Cooper S, Blundell TL, Hruby VJ, et al. Crystal structure analysis of deamino-oxytocin: conformational flexibility and receptor binding. Science 1986; 232:633-6; PMID:3008332; http://dx.doi.org/10.1126/science.3008332
13. Savarin P, Guenneugues M, Gilquin B, Lamthanh H, Gasparini S, Zinn-Justin S, Ménez A. Three-dimensional structure of kappa-conotoxin PVIIA, a novel potassium channel-blocking toxin from cone snails. Biochemistry 1998; 37:5407-16; PMID:9548922; http://dx.doi.org/10.1021/bi9730341
14. Skelton NJ, Garcia KC, Goeddel DV, Quan C, Burnier JP. Determination of the solution structure of the peptide hormone guanylin: observation of a novel form of topological stereoisomerism. Biochemistry 1994; 33:13581-92; PMID:7947768; http://dx.doi.org/10.1021/bi00250a010