Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology
ELSEVIER

Sequences and topology: intrinsic disorder in the evolving universe of protein structure

Editorial overview

A Keith Dunker and Julian Gough

Current Opinion in Structural Biology 2011, **21**:1–3

0959-440X/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.sbi.2011.04.002

A Keith Dunker

Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA

A Keith Dunker studied virus and phage structure and biology, making the observation in 1990 that a capsid protein of fd phage evidently lost structure in order to gain function. This discovery along with subsequent work led to his total emersion into the study of intrinsically disordered proteins using a combination of bioinformatics and experiment.

Julian Gough

Department of Computer Science, Merchant Venturer's Building, Woodland Road, University of Bristol, Bristol BS8 1UB, UK
e-mail: gough@cs.bris.ac.uk

Julian Gough studied protein structure analysis; then combined it with sequence comparison methods to produce a bioinformatics resource mapping all known structural domains to all completely sequenced genomes. This resource has provided the data for him, his colleagues and others in the field to understand the evolution of the protein repertoire.

To compare and contrast structured and disordered proteins, we collected some reviews on each topic, including reviews that consider evolutionary aspects of these proteins. On the structured side, the sequence → structure → function view dominates thinking about these proteins. The first two reviews are riveted deeply into this view of the protein world.

The review by Wass, Alessia, and Sternberg starts with the assumption of a structured protein and uses this model to construct algorithms that predict the binding of small molecules to their binding sites. This lock-and-key-based molecular recognition is the most iconic of all protein functions. The authors further discuss the conservation of binding sites on which many algorithms rely and new developments in the docking algorithms themselves.

While proteins are typically structured they are not immutable and can change their binding and catalytic functions over evolutionary time. In their review, Meng and Babbitt investigate the general principles that underlie changes among structure, function and evolution for families of enzymes. They pay particular attention to how new functions evolve from older ones via evolutionary sequence variations that couple changes in structure with changes in function.

Our understanding of protein structure and disorder is being continually improved by the ever expanding body of sequence data. Completely sequenced genomes have been growing in their coverage of the tree of life for over a decade now, but Godzik's review describes the more recent metagenomics projects, which are sampling sequences in new ways and are thereby helping us to see a more complete picture of the protein universe. Godzik points out that most of the protein diversity in the new environments likely arose from functional divergence of known families, probably by mechanisms like those described by Meng and Babbitt, rather than from the emergence of new protein families.

Not only the prediction of binding and conservation, but also the comprehensive analysis of structure and disorder depend on the ability of comparison methods to detect relationships between sequences. Söding and Remmert review recent advances in the methods used for sequence comparison. These methods include some that are applicable to, but are not always in common use by, various researchers in the areas reviewed by others in this section.

The standard view of structured proteins is that they are comprised of helices, strands and loops. The review by Rost and collaborators provides a

Please cite this article in press as: Dunker AK, Gough J. Sequences and topology: intrinsic disorder in the evolving universe of protein structure, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.04.002

broader view of protein structure by discussing loops not having an ordered structure, by discussing what types of such loops there might be, and by discussing what these loops conserve if not structure. Comparisons across the major kingdoms of life indicate the generality of such disordered loops.

Given the existence of disorder in nature, Tompa takes us into the mechanics of unstructured proteins, describes progress in the last decade — both with regard to disorder itself and also with regard to several component research areas used to investigate these proteins — and looks toward a future where we can actually model protein disorder. Like the preceding review and others, Tompa calls for the further development of bioinformatics tools.

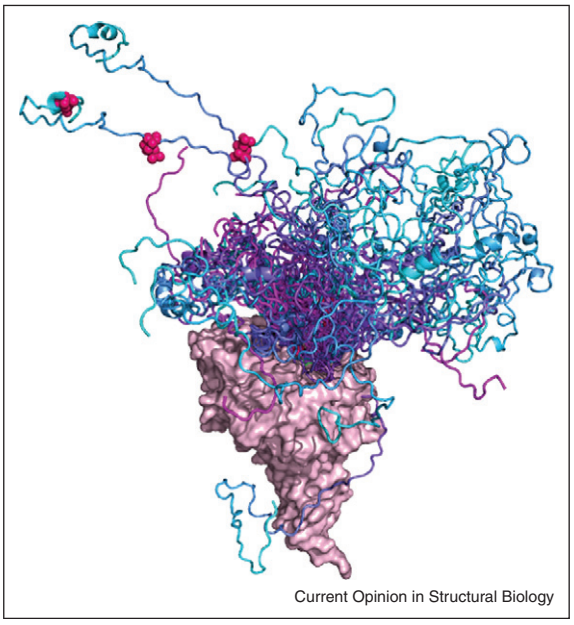
Disordered proteins exist not as a single structure but as conformational ensembles whose members stochastically interconvert into each other over a range of time scales. Fisher and Stultz compare the various methods used to build disordered protein ensembles that span an appropriate range of accessible states and that also provide a reasonable fit to the experimental data. While much has been accomplished, the authors indicate that even more remains to be done, and they suggest that it would be useful to put all of the ensembles into a common database to facilitate future development of this field.

The review by Babu and coworkers tells us that disordered proteins lend themselves well to signaling and regulatory functions. This review examines how disordered proteins themselves undergo tighter regulation as compared to structured proteins, and, like Tompa, makes the connection between disorder and disease.

This issue on sequence and topology is concluded with a review by Brown and collaborators that brings together the themes of evolution and disorder with the mechanical and functional aspects of these proteins. In brief, the authors review several lines of evidence showing that the evolution of disordered proteins is clearly distinct from the evolution of structured proteins. These evolutionary distinctions between structured and disordered proteins are consistent with, and therefore corroborate, the *in vitro* laboratory experiments and indicate that structured and disordered proteins are distinct *in vivo* as well.

The observations and concepts presented in this collection of reviews can be envisaged more easily when projected onto an example (Figure 1), which is a depiction of work carried out by Mittag *et al.* [1]. The topics covered here help us to understand this stunning new picture involving molecular interactions between a structured protein, Cdc4, and its disordered partner, Sic1. When a disordered protein binds to a structured partner, the interaction very often induces a disorder-to-order transition thereby forming stable structures in both partners. However, as mentioned in the review by Tompa, sometimes a disordered protein can retain significant disorder even after binding, as in the example given here.

Figure 1



The Sic1 ensemble associated with Cdc4. Sic1 remains as an ensemble upon associating with Cdc4 [1]. The previously determined structure of Cdc4 [2] is shown underneath in pink. On top, the ensemble of multiple Sic1 structures [3] are shown with colour gradients from cyan to magenta shaded from their N-termini to C-termini; these were calculated based on NMR and SAXS data using the program ENSEMBLE [4]. Individual Sic1 structures from the ensemble bind to Cd4 via different phosphoserines and phosphothreonines; three of which are depicted in red on one of the Sic1 structures. Tanja Mittag and Julie Forman-Kay developed this illustration and provided it to us for use in this Editorial Overview.

As shown in Figure 1, Sic1 remains a largely disordered ensemble even after binding to its structured partner, the receptor Cdc4 [1]. How can this possibly be? The Cdc4 receptor has a single binding site [2] that can associate with either a phosphoserine or a phosphothreonine when located within a certain sequence motif. Sic1 has, not one, but many suitable motif-embedded phosphoserine/phosphothreonine moieties, each of which can bind, but weakly — in part because the motifs are suboptimal. NMR spectra show that all of these suitable Sic1 moieties bind transiently to the single Cdc4 binding site [1]. Such transient binding to a single site in one protein by multiple phosphoserines or phosphothreonines at different positions cannot lead to a single bound structure but must instead lead to an ensemble.

Please cite this article in press as: . Sequences and topology: intrinsic disorder in the evolving universe of protein structure, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.04.002

This bound Sic1 ensemble has been modeled [3] using the methods developed by the same research group [4]. While the binding of each phosphoserine or phosphothreonine is weak, their transient binding in aggregate leads to a high enough overall affinity for productive binding. Furthermore, the strength of this binding can be regulated by changing the total number of phosphoserines plus phosphothreonines [5].

The interested reader can readily determine that most of the reviews in this section on sequence and topology provide information about a variety of methods that could be usefully applied to the Sic1–Cdc4 interaction. In fact, these reviews show that intrinsic disorder has a crucial role in many structural interactions. Disorder holds a significant place in the protein universe, and has played an important part in its evolution.

If you have enjoyed this collection of reviews, then we would like to draw your attention to a recent commentary by Chouard [6] that might also be of interest you.

References

1. Mittag T, Orlicky S, Choy WY, Tang X, Lin H, Sicheri F, Kay LE, Tyers M, Forman-Kay JD: **Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor**. *Proc Natl Acad Sci U S A* 2008, **105**:17772-17777.

2. Orlicky S, Tang X, Willems A, Tyers M, Sicheri F: **Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase**. *Cell* 2003, **112**:243-256.

3. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD: **Structure–function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase**. *Structure* 2010, **18**:494-506.

4. Marsh JA, Forman-Kay JD: **Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints**. *J Mol Biol* 2009, **391**:359-374.

5. Nash P, Tanga X, Orlicky S, Chen Q, Gertler FB, Mendenhall MD, Sicheri F, Pawson T, Tyers M: **Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication**. *Nature* 2001, **414**:514-521.

6. Chouard T: **Breaking the protein rules**. *Nature* 2011, **471**:151-153.

Please cite this article in press as: . Sequences and topology: intrinsic disorder in the evolving universe of protein structure, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.04.002