Protein structure networks

Lesley H. Greene

Advance Access publication date 4 October 2012

Abstract

The application of the field of network science to the scientific disciplines of structural biology and biochemistry, have yielded important new insights into the nature and determinants of protein structures, function, dynamics and the folding process. Advancements in further understanding protein relationships through network science have also reshaped the way we view the connectivity of proteins in the protein universe. The canonical hierarchical classification can now be visualized for example, as a protein fold continuum. This review will survey several key advances in the expanding area of research being conducted to study protein structures and folding using network approaches.

Keywords: protein structure; protein folding; long-range interactions; allostery; networks; graph theory

INTRODUCTION

Proteins are the building blocks of almost all biological processes that constitute life. Soluble, globular proteins have many function including transporting molecules such as oxygen, lipids and odorants, fighting foreign invaders such as bacteria and viruses to protect the organism, catalyzing reactions in metabolic pathways that are central to catabolism and anabolism, transcribing and translating DNA and RNA respectively, and generating signaling cascades, to name just a few critical roles. Proteins are also found in less-soluble fibrous forms that compose key structural elements. The protein α-keratin for example is found in hair and fingernails and collagen is found in connective tissue which composes cartilage, bones and blood vessels [1]. They are also integral constituents of membranes and as peripheral or integral membrane proteins play a pivotal role in controlling the transport of metabolites into and out of the cell, serve as receptors, adhesion molecules as well as participate in cell–cell communication [1]. There are also a small group of proteins which are intrinsically unstructured and these disordered proteins are gaining considerable attention in biochemistry for their unique structural and functional properties [2–4].

Protein structures have been traditionally viewed as constructs of secondary structure which pack into a 3D arrangement. This view was originally and most obviously conceived when the first 3D structures were being determined by John Kendrew, Max Perutz, Dorothy Hodgkin and other early crystallographers in the 1950's, 60's and 70's [5–7]. By the start of the 1980's the complexity of protein structures had been simplified through the systematic development and use of schematic ribbon drawings to illustrate the then approximately 75 known protein structures [8–11] and in its essence remains the most popular way to visualize these highly complex forms today. Through the development of protein structure visualization programs such as RASMOL [12], Chimera [13], Mage [14], VMD [15] and Jmol (Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/) for example we can not only see the all–atom 3D structure with atomic resolution but a generalized backbone pattern of the protein which reinforces the structure as a construct of secondary structure. Biology and biochemistry textbooks also preferentially use these ribbon-like images to convey protein structures worldwide to educate new generations.

The way in which we view proteins can affect the way we think about the biology and thus, can have serious ramifications in the development of hypotheses, experimental design and interpretation of data. During the past decade an alternative view of

Corresponding author. Lesley H. Greene, Department of Chemistry and Biochemistry, Old Dominion University, 4541 Hampton Boulevard, Norfolk, VA 23529, USA, Tel: +1-757-683-6596; Fax: +1-757-683-4628; E-mail: lgreene@odu.edu

Dr Lesley Greene is an Associate Professor at Old Dominion University. Her research interests include investigating the determinants of protein structure, folding and evolution using both experimental and computational approaches.

proteins structures has emerged in a concerted effort and begun to take hold. This view proposes that proteins structures can be modeled as an abstract network system [16–21]. In order to achieve this, we first have to translate terminology which is borrowed from graph theory, known more commonly today as network science. In the case of a graph, also defined as a network, you have nodes which can be people, cities or terminals for example which interact through links that can be friendships, highways or routers, respectively. To model a protein structure into a network you begin by adapting proteins to network terminology. In this way an amino acid can be considered a node or vertex and an interaction between residues can be considered a link or edge. In this review, we will use the term node and link.

Proteins are composed of amino acids linked by covalent peptide bonds to form polypeptide chains, also known as the primary structure. There are 20 common naturally occurring amino acids whose side chains have different physical characteristics that can be classified for example as polar, non-polar, acidic or basic in chemical nature. The packing of amino acid side chains through different types of non-covalent interactions which include hydrogen bonds, ionic interactions, van der Waals interactions and hydrophobic interactions as well as covalent interactions such as disulfide bonds ultimately confers the native 3D structure also known as the tertiary structure. This review will be focused on soluble, globular tertiary structures and not quaternary structures which involve the higher level interaction between individual protein structures to form a multimeric complex.

The abstraction of a 3D protein structure into a network considers only amino acids and their interactions through space without consideration of the polypeptide backbone, secondary structure composition or fold type. In general, there are two fundamental types of interactions: those that dictate the formation and stabilization of secondary structures (α-helices, β-sheets and turns)—known as short-range interactions; and long-range interactions that dictate the tertiary structure, which as mentioned earlier is the global organization of secondary structures. Both short-range and long-range interactions can also be interchangeably referred to as local- and non-local interactions, respectively; though we will use the more common former terms for consistency.

PROTEIN STRUCTURE NETWORKS
Approach to model proteins as networks

The traditional view of protein structures is shown in Figure 1. Here secondary structures are organized into 3D arrangements which can be further divided into fold-types, for example: an up–down α-helical bundle, a β-sandwich and a mixed α/β-barrel (Figure 1). The translation of these structures involves calculating the interactions between amino acids in 3D space and representing it in one instance as a 2D graph or in another instance as a 3D network. The key is in the calculation of the interactions between amino acids through space. This can be distilled down to two basic parameters. The first considers the number of residues apart in the primary structure for the interacting pair and the second the distance between the interacting pair in the 3D structure. Both rely on their designation as being either a short-range or a long-range interaction. The definition of a short-range interaction is straight-forward, consisting of contacts between residues within secondary structures. Hydrogen bonds within a β-sheet or within an α-helix, for example, are the most obvious descriptors. The definition of a long-range interaction is found to be more variable in the literature and is based on the selection of distances used as parameters. Fundamentally, however a long-range interaction is a contact between pairs of residues located on different secondary structures.

One early analysis which modeled proteins as network systems defined long-range interactions as the contacts between amino acids that are ≥10 residues apart in the primary structure but within 5 Å in the tertiary structure [17]. Other cutoffs can range from 4 to 18 residues apart in the primary structure [17]. Interaction distances can vary for example between 4 and 10 Å in the tertiary structure in the construction of these networks and may or may not take into account the location of interacting residues in the primary structure [16,17,20–24]. The interactions between residues or residue side chains can be calculated using all heavy atoms or select atoms such as the C_α or C_β carbons for example. Translating proteins into network systems using pair-wise amino acid interactions converts the structures into a 2D graph as represented in Figure 2 or a 3D network as shown in Figure 3. Algorithms written to conduct these distance calculations results in the generation of pairs of interacting residues that equates to linked
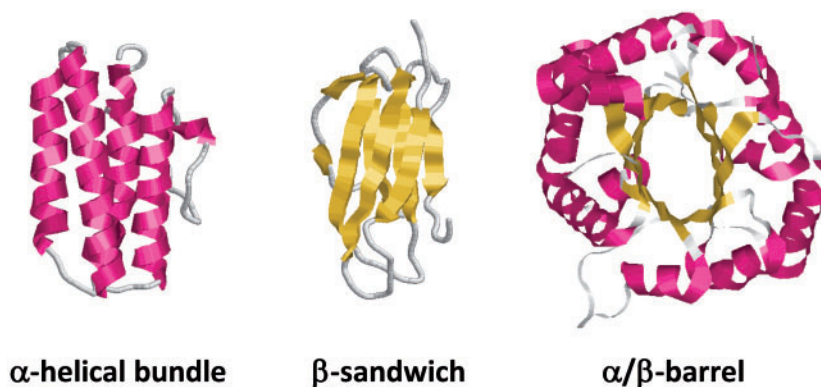
Figure 1: Examples of proteins from the all-α, all-β and mixed α/β classes in the traditional view as ribbon drawings. The all α-helical protein is myohemerythrin (pdb code: 1a7d). The all β-sheet protein is β2microglobulin (pdb code: 1bmg). The mixed α/β protein is triose phosphate isomerase (pdb code: 1tim). The α-helices are dark gray and the β-strands are light gray. The online color version displays the helices in pink and the β-stands in yellow. The structures were drawn using RasMol (www.rasmol.org).

nodes. There are also many programs such as Pajek (http://vlado.fmf.uni–lj.si/pub/networks/pajek/), Cytoscape [25] and RINalyzer [26] available to facilitate the construction, visualization and analysis of networks based on the distance calculations used to identify the nodes and links. This imagery of translating protein structures into network systems is also beautifully illustrated for example in the work of Estrada using hemoglobin and in the work of Atilgan *et al*. using interleukin 1-β (refer to Figure 1 in both references) [20,21]. The present discussion is applicable both to single domain and multidomain proteins. With respect to the later, the analysis [20] and prediction of discrete structural domains is another novel application of network science to the field of structural biology [27–29].

For general reference, there are several comprehensive analyses conducted to define meaningful contacts between amino acids. One for example, rigorously defines amino acid contacts in the tertiary structure based on residue distance separation in the primary structure for the different protein classes [30]. Another, details the effect of using different angstrom cutoff distances between contacting pairs in the 3D structure, which is highly informative [31]. These are valuable resources to facilitate deciding upon distance parameters to construct protein structure networks.

Protein structure networks include construction from wide variety of information in addition to the long– and short–range contact information outlined above. Protein structure networks may also be developed by using for example interaction energies

[32], covariance data [33], evolutionary conservation [34] and parameters that lead to formation of elastic networks such as amino acid fluctuations [35] and energy [36]. The wealth of diverse information that can be used is ultimately providing a richly layered and novel multidimensional view of these key biological molecules.

Application of network science to analyze protein structure networks

The establishment of protein structures as network systems opens up the application of the arsenal of mathematical principles in the field of network science to the analysis of these forms which until this conversion was not possible. The rigorous analysis of globular protein structure networks involving for example the following calculations: small–world, scale–free, betweeness–centrality and degrees of separation have been conducted by research groups across the world. The results revealed that protein structures have small–world properties [16–18,21] a concept originally pioneered by Watts and Strogatz in the analysis of a worm neural network, film actor collaborations and the Western United States power grid [37]. This means that protein structure networks involving both long– and short–range interactions have a high clustering coefficient C and a relatively short characteristic path length L [16–18,21]. Network models of transmembrane proteins are also now being constructed and analyzed [38]. Like the globular proteins discussed, transmembrane proteins also exhibit small–world character. The seminal work of Barabàsi and coworkers highlighted the
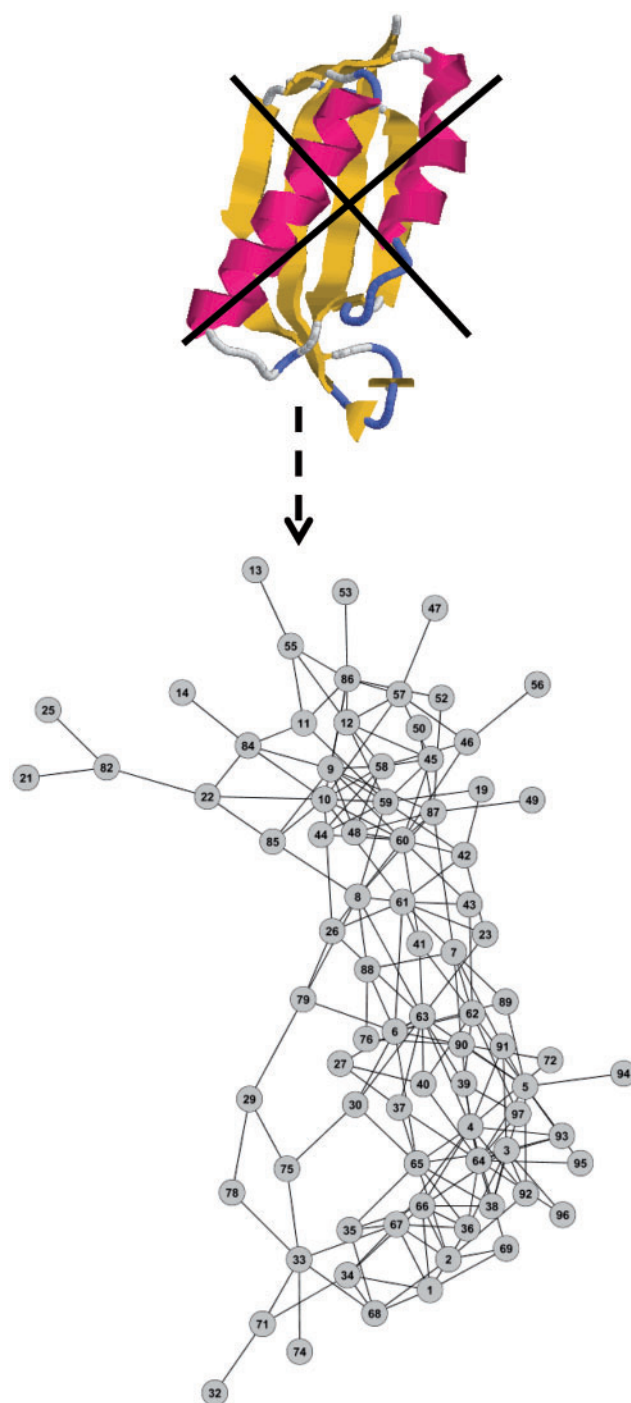
Figure 2: An alternative view of proteins as a 2D network representation. Shown is the traditional view of ribosomal S6 (pdb code: Iris) as a ribbon drawing using RasMol and a 2D representation of the long-range interaction network for this protein. The circles are the amino acids (nodes) and the lines are the long-range interactions (links). The network was constructed using long-range interactions between residues that are ten or more residues apart in the primary structure and within 5 Å in the tertiary structure. The method for generating this network is outlined in reference [l7]. The 2D network using the force directed layout is visualized used Cytoscape (www.cytoscape.org).

ubiquitous nature of scale-free distributions in network systems where the number of links per node, scale as a power law [39]. Long-range interaction networks in proteins have also been shown to share this quality for simulated partially folded structures [40]. In the native-state, when considering a network composed of both short- and long-range interactions a bell-shaped Poisson distribution was evident in the pattern of links [17,18].

Integral to the study of protein structures is gaining an understanding of the determinants of structural stability and dynamics. Excellent examples of studies analyzing stability and residue fluctuation using network parameters can be found in references [19,21,32]. Protein structure networks are now playing a valuable role in biochemistry by facilitating the identification of functionally relevant residues using for instance 'closeness centrality' measures [41].

In a traditional sense, select regions of a protein are most often correlated to functional sites, binding regions, hydrophobic cores and folding nuclei. This modularity is a feature well-suited for network modeling and analysis. The subnetworks identified in the analysis of protein allostery, is one aspect of a proteins modular nature. It is fascinating to consider amino acids communicating with one another to induce structural change or effect behavior when proteins interact or bind ligands. Allostery is an example of an inducible change in the structure of an enzyme when binding a cofactor or inhibitor. The interplay between select amino acid interactions within a protein can be modeled as a network and provides fascinating insights into structure–function relationships. Communication within the protein and hence the modeling of these networks of amino acid interactions has been studied by numerous groups and makes essential contributions by guiding the design of experiments as well as facilitating or extending the interpretation of experimental data: for example, Selvaratnam *et al.* used covariance analysis of NMR chemical shift data to map allosteric networks [33]. Other examples of the analysis of protein structure networks with respect to function include the work of del Sol *et al.* [42] which led to the determination that many residues key to maintaining the shortest path across a connected network are both conserved and essential for the biological role such as active sites among other interesting findings. This research also relates the role of network analysis in the evaluation of robustness of protein structures to mutation, where the general concept of robustness is an
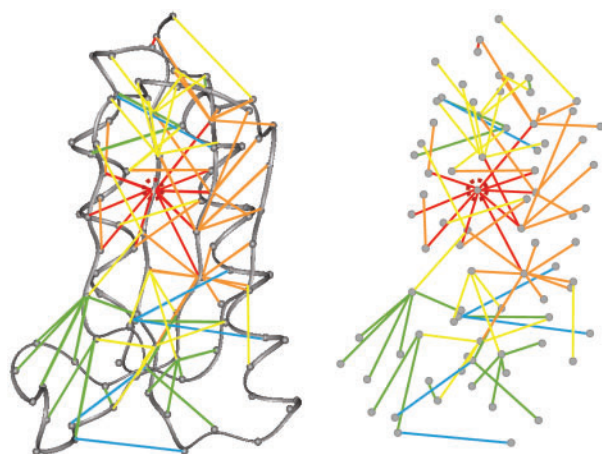
Figure 3: A view of a long-range interaction network in protein structures using ribosomal S6 as a model system. The long-range interaction network for ribosomal S6 (pdb code: Iris) is shown stemming from valine 6 in degrees of separation. The location of valine 6 is shown in a dashed circle. The image on the left shows the network in the context of the polypeptide backbone with the C_α shown as gray circles. The image on the right is the network without the polypeptide backbone and the C_α are shown in gray circles. The color version depicting the links stemming in five degrees going from red to blue is online. This figure is adapted from reference [17].

important one in both network science and biochemistry. The communication between conserved co-evolving residue networks and functional sites in proteins by Suel *et al.* provides further insight into both allostery and the intrinsic importance of subnetworks within the whole network [43]. An early and excellent example of evolutionarily conserved energetic coupling of long-range interactions between amino acids can be found in reference [44].

A clear example of the modular division of a protein through the reduction of protein structure networks into modules can be found in reference [45]. Del Sol *et al.* [45] highlight how this modularity is key to maintaining shortest pathways in signaling transmission and suggest interestingly that changes to modular boundary residues could evolve new or enhanced functions among other interesting conclusions. In the work of Krishnan *et al.*, structural modules or 'clusters' can be identified and are based in part on connectivity density within and between regions of the structure [46]. 'Multi-scale graph partitioning' also adds another dimension to the analysis of the modular nature of protein structures [47].

The identification of hydrophobic clusters was elegantly done by Kannan and Vishveshwara and found to be conserved between evolutionarily related proteins and as well as correlated to contain sites experimentally shown to be important in protein folding [27]. Protein structures may also be reduced to distinct subdivisions termed 'protein sectors' which can be used to decompose the protein structure network [48]. The dynamic nature of proteins is a crucial feature that has been harnessed to find clusters or 'structural communities' when looking at a physical parameter such as stability versus time [49]. Algorithm development such as the plug-in 'ModuLand' for Cytoscape will significantly enhance our ability to rigorously interrogate the modular nature of protein structure networks by resolving key aspects such as 'overlapping network modules' and 'hierarchical layers' [49]. The associated paper by Szalay-Bekö *et al.* also provides a an excellent discussion and illustrative examples of the 'community centrality' measure as well as a comprehensive modular analysis of the Met–tRNA synthase protein structure network [50].

Protein classification and the structure of the protein universe

Concomitant with the traditional view of protein structures is the way in which we organize and classify these forms [51]. The Worldwide Protein Databank and partners such as the RCSB Protein Databank are the central depository of all experimentally derived protein structure coordinates for the scientific community and public (http://www.wwpdb.org) [52,53]. There are two leading databases which organize this vast collection of structures using a hierarchical approach. These are the CATH [54] and SCOP [55] databases. Both databases initially group proteins according to secondary structure content such as mainly α–helical, mainly β-sheet or a combination of α and β structure which is termed 'Class' and then move down the hierarchy. In the CATH databases the primary levels in order of hierarchy are: Class, Architecture, Topology (fold family) and Homologous superfamily. In the SCOP database the primary levels in descending order are: Class, Fold, Superfamily and Family.

Where proteins have traditionally been viewed in a hierarchical manner, relationships were seen only in a vertical orientation. Protein relationships are now being visualized with horizontal connections
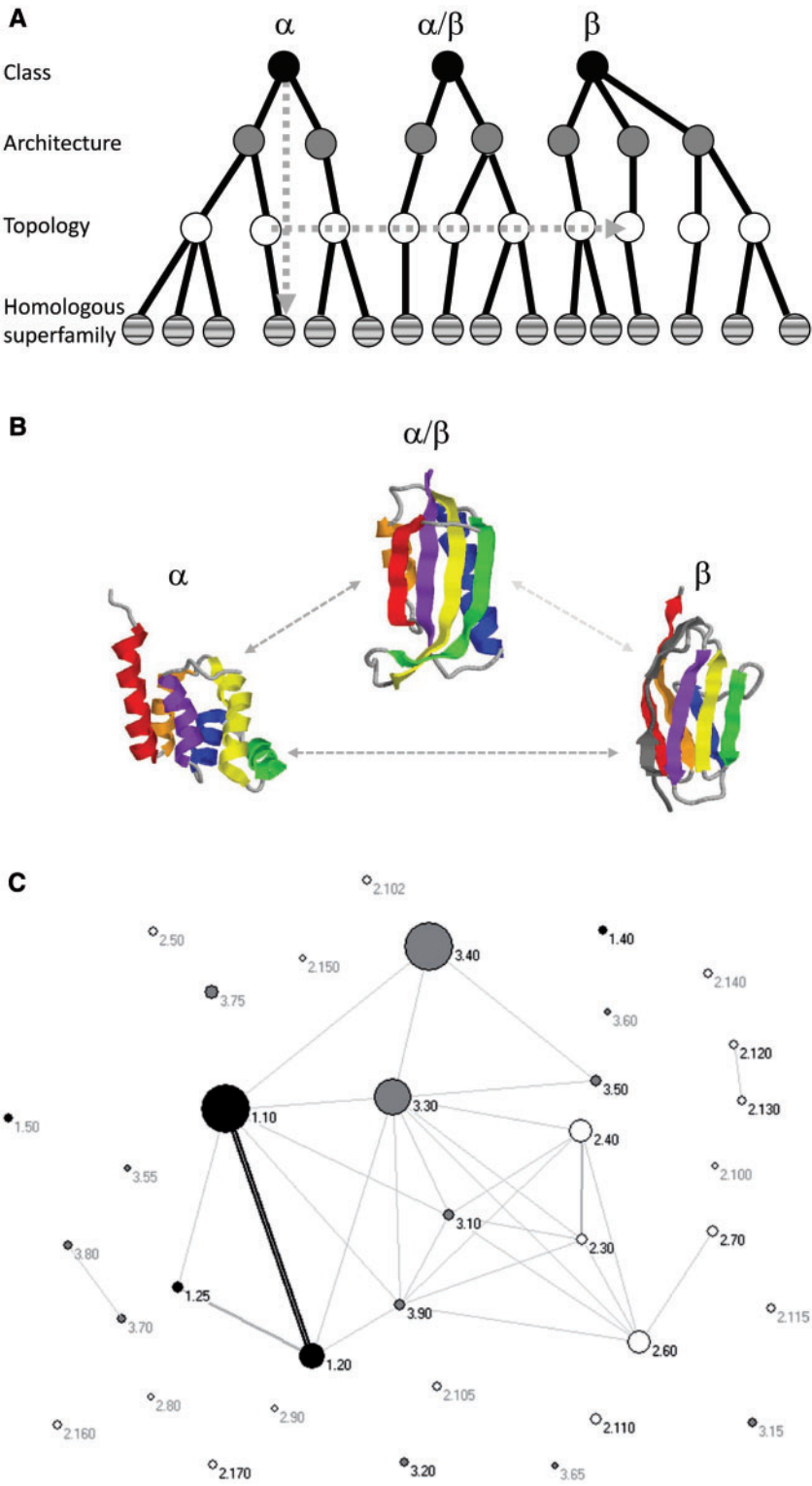
Figure 4: The interconnectivity of the protein universe. (**A**) Schematic of the CATH protein structure database organization. The four main levels in the hierarchy are class, architecture, topology (fold type) and homologous superfamily. A very simplified network view of representatives at each level is denoted by circles and their relationships by solid black lines. The solid gray arrow is drawn to emphasize the vertical nature of relationships in a hierarchical classification scheme. The dotted gray arrow provides a visual example of hypothetical horizontal relationships between proteins in traditionally different classes. (**B**) An example of three proteins which have been traditionally been assigned to different classes, architectures and topologies can be connected by a common

(continued)

meaning that proteins previously classified into different classes can be connected. For example, an all α-helical protein can share the same topology as an all β-sheet protein as in the example of the Greek-key proteins illustrated by Higman and Greene [40]. This lends itself to a network representation of fold space such as a continuum [56,57] or galaxy [58]. Both the Cuff [57] and Alva [58] papers display seminal examples of the continuous and interconnected nature of regions of fold space that can in some instances transcend hierarchical class, architecture, fold type and superfamily divisions (refer to Figures 10 and 1–3, in [57] and [58], respectively). Figure 4 in this review displays the concept of converting the hierarchical view of the protein universe into an interconnected network that embodies non-traditional links between proteins from the different classes.

PROTEIN FOLDING NETWORKS

The protein folding problem remains one of the major unsolved questions in science today. Researchers from a myriad of disciplines which include biochemistry, chemistry, computer science, mathematics and physics have sought for over 30 years to resolve the mechanism by which the primary structure dictates the tertiary structure and computationally predict this structure from the sequence. A particularly challenging facet of the protein folding problem is to elucidate the transition-state structure. Using interaction networks between residues experimentally shown to be important for folding as restraints in a Monte Carlo sampling procedure, insightful models of the transition state structure ensemble have been constructed [59]. Dokholyan *et al.* have also used a network approach to provide

important insights into the determinants of pre- and post-transition state ensembles [22]. The value of a network approach in understanding conformational space along the protein folding landscape can be highlighted for example by the work of Rao and Caflisch [60]. Here the folding of small β-sheet peptide using molecular dynamics simulations is analyzed by considering the generated conformations as nodes and the transition between the forms as links along the folding trajectory. They find among other interesting results that the network is scale-free and transition-state conformations as well as two main average folding pathways could be identified. A network approach was recently applied by Greene and Grant to propose a novel model for the formation of native protein structure networks from the transition-state in a modification of the network concept 'degrees of separation' into 'levels of separation' [61]. Further examples of the application of network parameters can be found in the work of Li *et al.* where folding nuclei were identified based on an analysis of the native state of six proteins [62], in the folding of the villin headpiece subdomain in the work of Lei *et. al.* [63] and in the unfolding of lysozyme by Ghosh *et. al.* [64].

Another advance in understanding the protein folding problem using network principles comes from the application of the network centrality measures such as, 'betweeness' [65–67]. Here nodes with high betweeness are considered to be keys to governing the network [67]. In proteins it was used to identify and characterize important residues for folding [16,40]. This was applied to several protein structures including chymotrypsin inhibitor 2, acylphosphatase, ribosomal S6 and iceberg [16,40]. It was shown in one study that the residues with high betweeness are important to forming and stabilizing

---

'Greek-key topology' thus generating horizontal relationships. Shown is the all-α helical protein, the Fas-associated death domain (pdb code:Ie3y), the mixed α/β-protein is ribosomal S6 (pdb code: Iris) and the all β-sheet protein is titin (pdb code: Itit) [40]. Each secondary element is assigned a different color in the online version. The five key secondary elements and their connectivity which share the same canonical Greek-key topology are colored in purple, blue, green, yellow and orange. For orientation the N−terminus of the protein is purple for Ie3y and Iris and gray for Itit. It is also interesting to note that they share a common network of long-range interactions between the structures [40]. (**C**) Relationships between proteins classified in different classes and architectures can be seen in this panel as an interconnected network. The numbering system denotes levels in the hierarchy with the first number specifying class and the number after the decimal point specifies the architecture classification number in the CATH database. For example, 3.30 signify the two-layer α/β-sandwich. The color of the circles corresponds to the class (black = mainly α-helical, white = mainly β-sheet and gray = mixed α/β) and the size represents the number of sequence subfamilies. The line thickness represents the number of overlapping superfamilies between the architectures. Figure 4C is reproduced from reference [57].

the transition-state structure [16] and in another study are highly conserved between proteins that share a related Greek-key topology which suggests they are topological determinants [40]. This would not have been possible were it not for the transformation of protein structures into network systems.

CONCLUSION

While great strides have been made to advance our understanding of the determinants of protein transition-state and native state structures as well as the folding landscape continued interrogation of proteins and the dynamic folding process from a network perspective will invariably bring about deeper insights. It is ultimately fascinating to know that a protein structure conceptually embodies similar characteristics to other seemingly disparate network systems such as the world-wide web, social networks, power grids and neural networks. We would also like to refer our readers to an excellent review of protein structure networks which covers a wide range of topic areas such as features of the network topology, protein function, dynamics and folding which can be found in reference [68]. The application of network science is also infiltrating the way in which we perceive protein relationships. This takes the form of alternative views of the protein structure universe and their interactions and is having profound conceptual effects.

Key points

- Protein structures can be modeled as network systems by translating amino acids into nodes and both long- and short-range interactions into links.
- By viewing proteins as network systems significant insights into the determinants of protein structure, stability and folding can be gained.
- Network models of proteins are essential for understanding a wide array of biochemistry questions such as structural relationships in protein space.

References

1. Garrett RH, Grisham CM. *Biochemistry*. Boston: Brooks/Cole, 2010.
2. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 1999;**293**:321–31.
3. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 2005;**579**:3346–54.
4. Dunker AK, Oldfield C, Meng J, *et al*. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* 2008;**9**:S1.
5. Kendrew JC, Bodo G, Dintzis HM, *et al*. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 1958;**181**:662–66.
6. Perutz MF, Rossmann MG, Cullis AF, *et al*. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature* 1960;**185**:416–22.
7. Harding MM, Hodgkin DC, Kennedy AF, *et al*. The crystal structure of insulin. II. An investigation of rhombohedral zinc insulin crystals and a report of other crystalline forms. *J Mol Biol* 1966;**16**:212–26.
8. Carter CW, Kraut J, Freer ST, *et al*. Two-angstrom crystal structure of oxidized chromatium high potential iron protein. *J Biol Chem* 1974;**249**:4212–25.
9. Richardson JS. Early ribbon drawings of proteins. *Nat Struct Mol Biol* 2000;**7**:624–25.
10. Richardson JS. The anatomy and taxonomy of protein structure. *Adv Protien Chem* 1981;**34**:168–339.
11. Richardson J. Schematic drawings of protein structures. *Methods Enzymol* 1985;**115**:359–80.
12. Bernstein HJ. Recent changes to RasMol, recombining the variants. *Trends Biochem Sci* 2000;**25**:453–55.
13. Pettersen EF, Goddard TD, Huang CC, *et al*. UCSF chimera—a visualization system for exploratory research and analysis. *J Computat Chem* 2004;**25**:1605–12.
14. Richardson DC, Richardson JS. The kinemage: a tool for scientific communication. *Protein Sci* 1992;**1**:3–9.
15. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;**14**:33–8.
16. Vendruscolo M, Dokholyan NV, Paci E, *et al*. Small-world view of the amino acids that play a key role in protein folding. *Phys Rev E* 2002;**65**:061910.
17. Greene LH, Higman VA. Uncovering network systems within protein structures. *J Mol Biol* 2003;**334**:781–91.
18. Bagler G, Sinha S. Network properties of protein structures. *Physica A* 2005;**346**:27–33.
19. Brinda KV, Vishveshwara S. A network representation of protein structures: implications for protein stability. *Biophys J* 2005;**89**:4159–70.
20. Estrada E. Universality in protein residue networks. *Biophys J* 2010;**98**:890–900.
21. Atilgan AR, Akan P, Baysal C. Small-world communication of residues and significance for protein dynamics. *Biophys J* 2004;**86**:85–91.
22. Dokholyan NV, Li L, Ding F, *et al*. Topological determinants of protein folding. *Proc Natl Acad Sci USA* 2002;**99**:8637–41.

23. Deb D, Vishveshwara S, Vishveshwara S. Understanding protein structure from a percolation perspective. *Biophys J* 2009;**97**:1787–94.

24. Gaci O. A Topological description of hubs in amino acid interaction networks. *Adv Bioinformatics* 2010; 1–9.

25. Smoot ME, Ono K, Ruscheinski J, *et al*. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011;**27**:431–32.

26. Doncheva NT, Klein K, Domingues FS, *et al*. Analyzing and visualizing residue networks of protein structures. *Trends Biochem Sci* 2011;**36**:179–82.

27. Kannan N, Vishveshwara S. Identification of side-chain clusters in protein structures by a graph spectral method. *J Mol Biol* 1999;**292**:441–64.

28. Xu Y, Xu D, Gabow HN. Protein domain decomposition using a graph-theoretic approach. *Bioinformatics* 2000;**16**: 1091–104.

29. Guo JT, Xu D, Kim D, *et al*. Improving the performance of DomainParser for structural domain partition using neural network. *Nucleic Acids Res* 2003;**31**:944–52.

30. Gromiha MM, Selvaraj S. Inter-residue interactions in protein folding and stability. *Progr Biophys Mol Biol* 2004;**86**: 235–77.

31. Da Silveira CH, Pires DEV, Minardi RC, *et al*. Protein cutoff scanning: a comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. *Proteins Struct Funct Bioinformatics* 2009;**74**:727–43.

32. Vijayabaskar M, Vishveshwara S. Comparative analysis of thermophilic and mesophilic proteins using protein energy networks. *BMC Bioinformatics* 2010;**11**:S49.

33. Selvaratnam R, Chowdhury S, Vanschouwen B, *et al*. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc Natl Acad Sci USA* 2011;**108**:6133–8.

34. Greene LH, Hamada D, Eyles SJ, *et al*. Conserved signature proposed for folding in the lipocalin superfamily. *FEBS Lett* 2003;**553**:39–44.

35. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Design* 1997;**2**:173–81.

36. Piazza F, Sanejouand Y-H. Discrete breathers in protein structures. *Phys Biol* 2008;**5**:1–14.

37. Watts DJ. *Six Degrees: The Science of a Connected Age*. London: William Heinmann, 2003.

38. Emerson IA, Gothandam KM. Network analysis of transmembrane protein structures. *Physica A Stat Mech Appl* 2012; **391**:905–16.

39. Barabasi Al AR. Emergence of scaling in random networks. *Nature* 1999;**286**:509–12.

40. Higman VA, Greene LH. Elucidation of conserved long-range interaction networks in proteins and their significance in determining protein topology. *Physica A Stat Mech Appl* 2006;**368**:595–606.

41. Amitai G, Shemesh A, Sitbon E, *et al*. Network analysis of protein structures identifies functional residues. *J Mol Biol* 2004;**344**:1135–46.

42. del Sol A, Fujihashi H, Amoros D, *et al*. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol Syst Biol* 2006;**2.**

43. Suel GM, Lockless SW, Wall MA, *et al*. Evolutionary conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 2002;**10**:59–69.

44. Lockless SW, Ranganathan R. Evolutionary conserved pathways of energetic connectivity in protein families. *Science* 1999;**286**:295–99.

45. Del Sol A, Arauzo-Bravo M, Amoros D, *et al*. Modular architecture of protein structures and allosteric communications: potential implications for signaling proteins and regulatory linkages. *Genome Biol* 2007;**8**:R92.

46. Krishnan A, Zbilut JP, Tomita M, *et al*. Proteins as networks: usefulness of graph theory in protein science. *Curr Protein Peptide Sci* 2008;**9**:28–38.

47. Delmotte A, Tate EW, Yaliraki SN, *et al*. Protein multi-scale organization through graph partitioning and robustness analysis: application to the myosin–myosin light chain interaction. *Phys Biol* 2011;**8**:055010.

48. Halabi N, Rivoire O, Leibler S, *et al*. Protein sectors: evolutionary units of three-dimensional structure. *Cell* 2009; **138**:774–86.

49. Delvenne J-C, Yaliraki SN, Barahona M. Stability of graph communities across time scales. *Proc Natl Acad Sci USA* 2010; **107**:12755–60.

50. Szalay-Bekö M, Palotai R, Szappanos B, *et al*. ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics* 2012;**28**:2202–04.

51. Levitt M, Chothia C. Structural patterns in globular proteins. *Nature* 1976;**261**:552–58.

52. Berman HM, Kleywegt GJ, Nakamura H, *et al*. The Protein Data Bank at 40: reflecting on the Past to Prepare for the Future. *Structure* 2012;**20**:391–96.

53. Rose PW, Beran B, Bi C, *et al*. The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res* 2011;**39**:D392–401.

54. Orengo CA, Michie AD, Jones S, *et al*. CATH–a hierarchic classification of protein domain structures. *Structure* 1997;**5**: 1093–108.

55. Murzin AG, Brenner SE, Hubbard T, *et al*. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;**247**:536–40.

56. Kolodny R, Petrey D, Honig B. Protein structure comparison: implications for the nature of 'fold space', and structure and function prediction. *Curr Opin Struct Biol* 2006;**16**: 393–98.

57. Cuff A, Redfern OC, Greene L, *et al*. The CATH hierarchy revisited–structural divergence in domain superfamilies and the continuity of fold space. *Structure* 2009;**17**:1051–62.

58. Alva V, Remmert M, Biegert A, *et al*. A galaxy of folds. *Protein Sci* 2010;**19**:124–30.

59. Vendruscolo M, Paci E, Dobson CM, *et al*. Three key residues form a critical contact network in a protein folding transition state. *Nature* 2001;**409**:641–45.

60. Rao F, Caflisch A. The protein folding network. *J Mol Biol* 2004;**342**:299–306.

61. Greene LH, Grant T. Protein folding by 'levels of separation': a hypothesis. *FEBS Lett* 2012;**586**:962–66.

62. Li J, Wang J, Wang W. Identifying folding nucleus based on residue contact networks of proteins. *Proteins Struct Funct Bioinformatics* 2008;**71**:1899–907.

63. Lei H, Su Y, Jin L, *et al*. Folding network of villin headpiece subdomain. *Biophys J* 2010;**99**:3374–84.

64. Ghosh A, Brinda KV, Vishveshwara S. Dynamics of lysozyme structure network: probing the process of unfolding. *Biophys J* 2007;**92**:2523–35.

65. Dorogovtsev SN, Mendes JFF. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford: Oxford University Press, 2003.

66. Newman MEJ. The structure and function of complex networks. *SIAM Rev* 2003;**45**:167–256.

67. Linton CF. A Set of measures of centrality based on betweenness. *Sociometry* 1977;**40**:35–41.

68. Böde C, Kovács IA, Szalay MS, *et al*. Network analysis of protein dynamics. *FEBS Lett* 2007;**581**:2776–82.