

## SMALL WORLD NETWORK STRATEGIES FOR STUDYING PROTEIN STRUCTURES AND BINDING

Neil R. Taylor<sup>a,\*</sup>

**Abstract:** Small world network concepts provide many new opportunities to investigate the complex three dimensional structures of protein molecules. This mini-review explores the published literature on using small-world network approaches to study protein structure, with emphasis on the different combinations of descriptors that have been tested, on studies involving ligand binding in protein-ligand complexes, and on protein-protein complexes. The benefits and success of small world network approaches, which change the focus from specific interactions to the local environment, even to non-local phenomenon, are described. The purpose is to show the different ways that small world network concepts have been used for building new computational models for studying protein structure and function, and for extending and improving existing modelling approaches.

### MINI REVIEW ARTICLE

#### Introduction

A fundamental tenet in the study of chemistry is the direct relationship between structure, and chemical and physical properties. In the field of molecular biology, this relationship becomes structure and function [1]. For proteins, structure plays a key role in catalysis, messaging, activation, and disease states [1-2]. In protein structures, what is particularly interesting is that both rigidity and flexibility are critical properties, often both being needed to fulfil functional requirements [2]. Proteins are synthesised as linear polypeptides that fold into compact three dimensional conformations consisting of a large number of inter-residue close-contacts. They bind with other folded proteins and/or small molecules, often adapting to the shape and electrostatic requirements of these complexes. It is the intrinsic complexity of the three dimensional conformations of proteins, and the important role that proteins play in living organisms, that lead many scientists to study them.

In textbooks, protein structures are described at their different levels of complexity: primary, secondary, tertiary, and quaternary structure [1-2]. In this article, the focus is on details of three dimensional (3D) structure. In practice, the way that scientists understand and work with 3D structures is closely determined by how they are represented in a computer. Figure 1 illustrates some examples of different representations of 3D protein structure. For protein crystallographers, for example, a protein structure is the atomic model that best represents observed electron density intensities. Atom level details are typically viewed and explored, overlaid with electron density maps. The quality of the electron density measurements and the fit to the atomic model, particularly for key atoms and residues of interest, provide the critical information needed to correctly interpret the data [3].

For computational chemists working in pharmaceutical research, the typical approach is to use a more detailed all atom representation

of just the binding site region; usually including hydrogen atoms directed toward their hydrogen bonding partners. An in-depth analysis of the fine-details of specific non-covalent protein-ligand interactions, of local molecular surface shape and electrostatic profile, the role of specific water molecules, etc, provides a vast array of new information for understanding ligand binding [4].

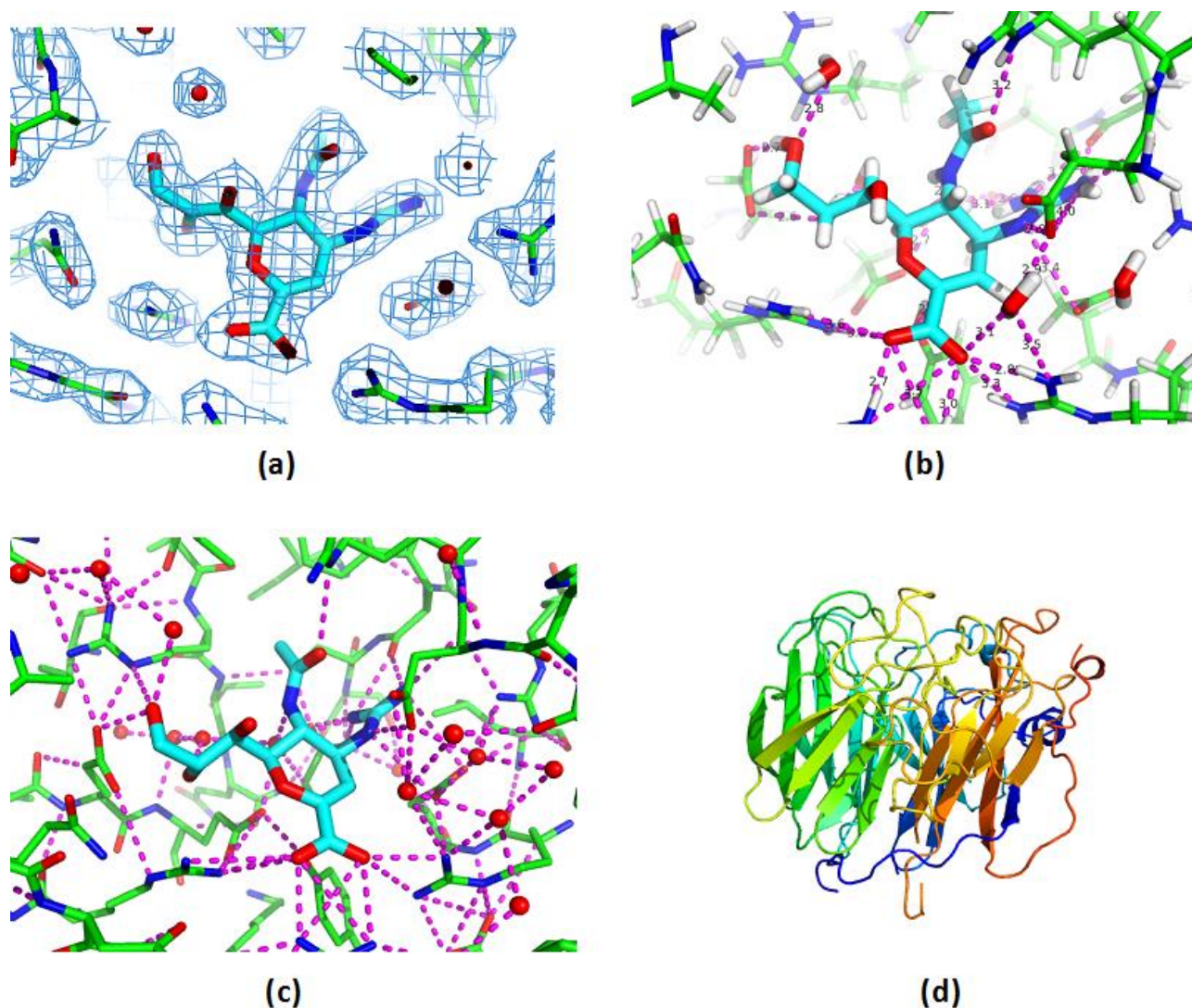
For scientists working in structural bioinformatics, the typical view is often a less detailed schematic representation of 3D structure, highlighting the alpha helices, beta sheets, and loops. Despite the lower detail, this level of analysis provides the best insight into overall topology and stability. Furthermore, comparing and contrasting schematic representations helps with the elucidation of function and functional relationships; leads to powerful classification schemes; and provides insights into evolutionary relationships [5].

This article describes another approach for representing and investigating 3D protein structures, one that has grown rapidly over the last ten years, both in use and scope: the network approach. More specifically, it explores methods that model protein structure as a network of nodes (typically aminoacid residues) and edges (close-contacts between residues) and use a small world network analysis to account for key aspects of structure and function. This type of analysis changes the focus from individual residues and individual interactions to local connectedness and the connectedness of the whole protein. The change allows non-additivity effects and non-local effects to be incorporated into modelling. An example network view for a protein is shown in Figure 1c.

The aim of this paper is to review the different ways that small world network concepts have been used as the foundation for building new computational models to study protein structure and function, and to extend and improve existing modelling approaches. The focus is on cheminformatics and bioinformatics, describing when and how network descriptors have been combined with standard properties to account for the contributions of non-local and global effects. A brief introduction to networks and different network types, and the main descriptors used to explore them, is provided. A short historical perspective on small world networks is also included. Since the focus is on molecular structure, in the main part of this review the literature is divided according to whether the research is on single proteins, protein-ligand binding, or protein-protein binding.

<sup>a</sup>Desert Scientific Software Pty Ltd, Level 5 Nexus Building, Norwest Business Park, 4 Columbia Court, Norwest, NSW, 2153, Australia

\* Corresponding author. Tel.: +61 288606466; Fax: +61 296807520  
E-mail address: [neil.taylor@desertsci.com](mailto:neil.taylor@desertsci.com)



**Figure 1.** Different representations of 3D protein structure provide different types of understanding. (a) The electron density map for ligand and surrounding binding site residues shows a good quality, rigid model. (b) Typical computational chemistry view, showing that protein-ligand binding involves many short hydrogen bonds. (c) A network view of all favourable polar interactions in the binding site region shows how the protein-ligand hydrogen bonds are parts of highly connected local environments, which also involve numerous bound water molecules. (d) Secondary structure view, suited for structural bioinformatics analysis. Protein is Neuraminidase in complex with Zanamivir. Images generated using PyMol ([www.pymol.org](http://www.pymol.org)), PDB Ids 2cml and 1nnc.

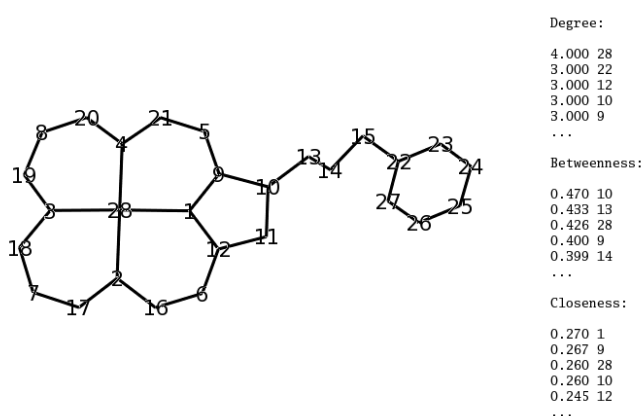
### Introduction to Networks

When a system is defined as a network, one is describing it at a highly abstract level, reducing it to a collection of simple nodes joined to one another by simple edges. It should be mentioned that the terms network and graph can be used inter-changeably, as can node/vertex, and also edge/connection/wiring. Network studies are less interested in the attributes of individual nodes and edges, and more interested in the properties arising from local and global topologies that are the result of the connectedness of the nodes, see for example Bollobas [6]. Typically, edges do not have a directional component, nor do they have different strengths (networks of these types are described as unweighted, undirectional graphs). Two important properties when describing any network are node degree, the count of the number of edges a node has, and shortest path length - the count of the minimum number of edges needed to be traversed to get from one node to another. Key information can be obtained from both the averages and distribution functions of these values over an entire network [6].

The main network types encountered in the literature are regular, random, small-world, and scale-free. A regular network is a regular array of nodes and edges, with high local symmetry - all nodes having identical connectivity - in other words, a regular lattice or mesh. An example of this type of network, well known to all scientists, is the diamond lattice, the arrangement of atoms and covalent bonds in diamond. A random network is one in which edges occur randomly, that is, the rules governing attachment are entirely random (not based on node proximity, for instance). In a random network, most nodes have approximately the same number of edges, and node degree has a Poisson distribution [6]. A small world network typically arises when one very simple rule governs attachment, that rule being "popularity is attractive" [7-8]. That is, a new node in the network is more likely to form an edge with a node that already has a higher than average number of attachments, or with another node that has a short path length to such a node. A scale-free network can be thought of as an extreme case of small world behaviour, having a degree distribution that follows a power-law distribution. In a scale free network it can be possible to find nodes with degree five to ten log units higher than

the average degree [8] - such nodes can never possibly arise in random networks. Hub is the name given to the few nodes with the highest number of links.

Global network descriptors enable different types of networks to be distinguished from one another. Within a graph, local descriptors can be used to distinguish the different roles played by different nodes, based just on the connections. The most commonly used local network descriptors for nodes are explained in Figure 2. The two most commonly used global network descriptors are: characteristic path length,  $L$ , the average of the shortest path lengths between all pairs of nodes (small for random graphs); and clustering coefficient,  $C$ , the average over all nodes of the fraction of the number of connected pairs of neighbours for each node (large for regular networks). Watts and Strogatz [9] showed that the different network types have different magnitudes and ratios of  $L$  and  $C$ : regular networks have large  $L$  and large  $C$ ; random networks have small  $L$  and small  $C$ ; and small world networks have small  $L$  and large  $C$ .



**Figure 2.** Simple chemical graph showing the three most widely used network descriptors for nodes, with the top five values for each. Node 28 has the highest degree, that is, the most connections (and is therefore a hub). Node 10 has the highest betweenness, which is a function of the fraction of shortest paths through a node (removing this node creates the two largest disconnected fragments). Node 1 has the highest closeness, that is, the inverse of the average of the shortest paths to all other nodes. This particular graph has a low clustering coefficient - no pair of connected nodes is connected to the same third node (no triangles). The characteristic path length is low (near 5) though this is not particularly meaningful as it is only a very small network.

A number of introductory books have been written on the subject of small world networks [10-11]. For readers interested in a more detailed coverage of the underlying physics and mathematical aspects, Dorogovtsev et al. [8] is recommended. A comprehensive survey of the scientific literature is available in Boccaletti et al. [12], with nearly 900 references; within the field of medicines research, including protein structure networks as discussed herein, see Csermely et al. [13] which has more than 1100 cited references. Surveys, introducing more of the available literature on protein structure networks, include Boede et al. [14] and also Krishnan et al. [15].

A wide range of powerful and versatile software tools are available for researchers interested in using network approaches in their work, such as Cytoscape [www.cytoscape.org], NetMiner [www.netminer.com], Networkx [networkx.lanl.gov], JUNG [jung.sourceforge.net]. See Csermely et al. [13] for many more examples. In the field of protein structure networks, on-line research tools are also available, including RING [protein.cribi.unipd.it/ring], RINalyzer [www.rinalyzer.de], and Scorpion [www.desertsci.com].

## Historical perspective

For many years, the study of networks focused on the properties of random graphs with normal or Poisson distributions of connections between nodes [16-17]. The focus changed in the late 1990s in parallel with the incredibly rapid growth and development of the internet and the world wide web. Subsequently, the world wide web was observed to have: (1) an unexpectedly low average shortest path length between any pair of nodes, and (2) a fat-tailed degree distribution, that is, the number of connections for some nodes is many orders of magnitude higher than the average number [7]. They termed these small world networks, and numerous studies have found them throughout nature [7,18-22]. Not only have they been identified in many different realms in the natural world, but also in man-made systems including communications systems, financial systems, scientific citations, and throughout numerous human social organisations [8]. The reason why small world networks are so frequently observed is believed to be due to their stability - stability here meaning maintaining integrity and minimising the possibility of failure. This stability is believed to arise from optimised communication pathways within these networks, or in other words, from short path lengths between all parts of the network. Since 2002, small world network concepts have been incorporated more and more into the fields of chemistry and structural biology [13]. The thinking behind using the network approach in protein structure analysis is that it allows for contributions from non-local effects to be included into a model.

To clarify, there are two completely different fields of research involving proteins and networks. The subject of this paper is the field of protein conformation. The other field, which has been more widely investigated and reported, involves communication pathways between whole proteins, and commonly referred to as protein networks, protein contact networks, or protein-protein interaction networks (see Csermely et al. [13] for an overview). That field lies more within biology, specifically systems biology, than within chemistry.

## Individual protein domains

Many articles have been published describing protein domains using small world network approaches, based on the network of proximal aminoacid residues [23-45]. Network analyses of aminoacid graphs have been used to tackle many different aspects of topology, stability, folding, and communication pathways. Analysis typically involves a comparison of the calculated characteristic path length and clustering coefficient against a [comparable] random network. Articles consistently report that protein domains have small world network characteristics, or more precisely, have topologies that lie somewhere in the middle between the two extremes of random networks and scale free networks. The network of aminoacid sidechains cannot be fully scale free due to two important boundary conditions. The first is the limit of packing multiple residues in the coordination sphere of a central residue, also known as the excluded volume affect; the second involves the numerous constraints imposed by the protein backbone, for example, that residue  $i$  is covalently constrained by residues  $i-1$  and  $i+1$ . Despite these constraints, evidence that small world network effects contribute to the structural integrity of protein structures is highly persuasive. Mention is sometimes made of the role of evolution: that proteins are the product of evolutionary optimisation processes, and that topologies with small world network characteristics have been selected due to advantages provided for overall stability, robustness, and protection against

failure of function due to mutations [27]. The point is often made that modelling a protein structure as network means modelling as a communication system, which provides many opportunities for new insights into structure and function [34,42,44].

Table I lists a subset of articles, demonstrating the broad scope of the problems studied, and methods used. In the earlier articles on residue networks for protein domains (and complexes), graphs were built using C alphas as the nodes (or C betas) and an edge defined when two nodes were within a threshold cutoff, typically in the range 7.0 to 9.0 Angs. It is now more common to define edges using a higher level of atomic detail - when a pair of heavy atoms is within a threshold cutoff, typically in the range 4.0 to 5.0 Angs. Although this method is more consistent with the energetics of atom-atom close-contacts, it has the drawback of not allowing for through-water interactions between residues, or through-metal interactions. There are many different descriptors that can be computed for any network, however, on their own, these descriptors do not provide such interesting predictions for proteins. It is when they are combined with other structural and sequence properties, or added into existing models, that they do become really useful. The purpose of listing the different methods in Table I is to provide a comprehensive overview of the different cheminformatics and bioinformatics methods that have been explored to date, so that these may provide ideas for further research.

### Active site analysis and protein-ligand binding

The small world network approach has been particularly successful in the prediction of binding site residues [27-29,46-50], for example, using closeness centrality with solvent accessibility scores [27]; combining closeness centrality and phylogenetics [29]. Betweenness centrality, when averaged over a patch of residues, was found to improve the predictive power of a model for identifying residues involved in binding RNA [48]. Heme-binding residue prediction [49] was shown to work well using standardised network descriptors in combination with a sequence profile descriptor and several additional structural descriptors, including solvent accessibility, measures of local concavity and convexity of the protein surface. Interestingly, the clustering coefficient was found to be lower for residues that bind heme, that is, they are less packed, suggested to be important for allowing for some flexibility in the binding site. Network descriptors have also been incorporated into a method for predicting DNA-binding residues [50], with a weighted average betweenness centrality measure combined with features from evolutionary profiles, interface propensity, and side-chain solvent accessibility. In a network study of DHFR [47], the network model included the nitrogen and oxygen atoms of the cofactor. Changes in closeness values showed that cofactor binding, specifically at the binding site, has a significant affect on the network, and that in the complex, most network interactions involved the cofactor. Ligand binding sites are unique regions on the surfaces of proteins, and the ability of network approaches to successfully identify these regions strongly hints at the importance of including network effects into modelling.

A small world network approach has also been used to build a quantitative model for predicting ligand binding affinity [51]. In this work, the small-world network architecture of protein structure was assumed, and a novel set of localised network descriptors developed. A reduced graph definition of protein structure was used, with separate nodes for each sidechain and mainchain group (including C alpha), and included HET groups, metal atoms, and tightly bound water molecules. Network edges were created from all covalent and

favourable non-covalent interactions between atoms, identified using a comprehensive classification scheme. Network descriptors were constructed in such a way as to maximise potential small-world network influences, and incorporated all short paths, rather than just shortest paths, as in most previous studies. The network model served to increase the affinity contributions of non-covalent interactions when the local environment was more highly connected. In this work, global descriptors were found not to be generally applicable as they were overly sensitive to individual close-contacts. In some cases ligand binding affinity was found to track closely with ligand molecular weight, and additional descriptors did not significantly improve predictions. In other cases, tight binding clearly involved additional factors, and the incorporation of non-local affects [via the network model] helped to account for these.

### Protein-protein binding

Models based on residue interaction networks have been used in a number of studies of protein-protein complexes, see for example [52-54]. In an exploration of protein-protein interaction sites, conserved residues with high betweenness scores, buried upon dimerisation, correlated well with experimentally determined hotspot residues [52]. At the same time, network descriptors for uncomplexed monomers were found not to be good predictors for the protein-protein interface - the majority of high-betweenness values at the interface are created with dimerisation.

Several small world network approaches have been developed to improve the performance of in silico protein-protein docking [54,55]. Chang et al. [54] computed network descriptors from two networks, one with hydrophobic residues as nodes (Ile, Leu, Val, Phe, Met, Trp, Cys, Tyr, Pro, Ala) and one with hydrophilic residues as nodes (Gly, Lys, Thr, Ser, Gln, Asn, Glu, Asp, Arg, His). Edges were created for every atom pair within 5.0 Angs of one another (for atoms from two different residues). The two networks themselves were shown to be small-world networks. Clustering coefficient and characteristic path length were computed, then combined with other parameters, which included terms for vdW interactions (attractive and repulsive), solvation, hydrogen bonding, long-range electrostatics (attractive and repulsive), constraining side chains rotamers, and a residue-residue pair probability metric. To account for differences in protein size, network parameters were standardised by calculating their standard deviation from the mean for each protein. Parameterisation was done by maximally separating true docked solutions from decoys. They showed that the addition of the new network terms to the scoring function improved the discrimination capabilities of the method. Interestingly, they observed that correct docking solutions exhibit lower characteristic path lengths than incorrect solutions, suggesting that correct solutions better preserve the characteristic path length of native protein structures than incorrect solutions.

Pons et al. [55] achieved improved results using network descriptors to help score rigid body protein-protein docking solutions. Apo proteins were modelled as networks, with the C alpha atoms of all aminoacids as nodes, and edges when two C alpha atoms were within 8.5 Angs (results mostly unchanged if C beta atoms used instead). Four different network parameters: degree, closeness, betweenness, and clustering coefficient, were investigated to determine their utility in protein-protein docking studies. Best results were obtained when combining calculated closeness descriptors with the default scoring function [56]. The article is also of interest in that a number of diverse influences were explored to determine under which conditions the network descriptors work best, including the type of complexes, flexibility, complex size, and protein shape.

**Table 1.** Overview of publications using network methods to model the three dimensional conformations of protein domains. The summaries focus on the different cheminformatics and bioinformatics methods employed, and the properties investigated.

Publication	Topic of Inquiry	Methods Used
Dokholyan et al. 2002 [24]	Studied folding by creating networks based on conformations generated from Monte Carlo simulations	Showed that post-transition state protein conformations are more small-world like, that is, globally tighter, than pre-transition state conformations. Characteristic path length was found to be the best performing global descriptor examined
Vendruscolo et al. 2002 [25]	Studied folding by creating a weighted graph from transition state ensembles of protein structures generated by Monte Carlo sampling	Edge weights obtained by dividing the number of structures with a specific edge by the total number of structures in the ensemble. Betweenness values were highest for key residues involved in folding
Greene et al. 2003 [26]	Studied protein conformation	Used a network with weighting factors derived from the presence of multiple links between residues
Brinda et al. 2005 [32]	Studied protein structure and stability	Used edge weighting based on the number of close contacts between a pair of residues divided by normalisation factors (which take into account residue size and the propensity to make a large number of contacts) for each residue, and allowing for varying interaction strengths
Paszkiwicz et al. 2006 [33]	Used network methods to identify a predictor of suitable regions of circular permutation	Closeness was shown to useful descriptor; exploration of relative side chain area and sequence conservation measures also undertaken
del Sol et al. 2006 [34]	Study of communication aspects of the aminoacid residue network	Used a characteristic path length metric, corresponding to the change in path length resulting from removal of a node
Muppurala et al. 2006 [35]	Aim was to distinguish correctly folded from incorrectly folded proteins	Used more stringent distance constraints for hydrogen bonding (and also angle cutoffs), and similarly for hydrophobic and ionic contacts, and disulphide bonds, and conserved hydrophobic aminoacids were mapped onto the network to examine their roles
Aftabuddin et al. 2007 [36]	Protein structure analysis	Separated amino acid sidechains into different types, hydrophobic (F, M, W, I, V, L, P, A), hydrophilic (N, C, Q, G, S, T, Y), and charged (R, D, E, H, K), and the weighted (according to the number of close-contacts between a pair of residues) and unweighted networks separately analysed
Gaci et al. 2008 [41]	Examined subgraphs involving residues participating in secondary structure elements, alpha helices and beta sheets	Explored distribution of characteristic path length and clustering coefficient as a function protein size
Li et al. 2008 [42]	Examined protein folding and communication between residues	Found that key residues in the folding process - found to be global centrals rather than local centrals - could be identified using solvent accessibility and network terms together
Milenkovic et al. 2009 [43]	Sought to determine the best null model for discriminating network motifs, including secondary structure motifs, in aminoacid graphs	Examined different choices for nodes - just using side chain atoms, just using backbone atoms, using all residue atoms - and different wiring models explored, including 3D geometric, random, scale free, and stickiness-index based networks
Vijayabaskar et al. 2010 [44]	Investigated protein structure integrity and communication pathways using energy weighted networks	Protein energy networks (PENs) were created with edges weighted according to calculated interaction energies, obtained from sampling structures from a molecular dynamics simulation, and fluctuations from the mean
Petersen et al. 2012 [45]	Protein structure analysis, seeking patterns in the packing of amino acid pairs	Created an eight dimensional descriptor space encompassing residue proximity, solvent accessibility, sequence distance, secondary structure, and sequence length, and uncovered a scale free organization in aminoacid pair interactions

## Summary and Outlook

The published literature on using network models to study protein structure consistently shows that the structures exhibit small world characteristics. In addition to the experimental support provided by analysis of characteristic path length and clustering coefficient, further verification is provided by the success of incorporating network descriptors into existing models and observing better quality predictions. The small world network approach is attractive from the conceptual perspective, providing better understanding of the stability of protein topology, both in terms of overall structural integrity and in terms of robustness against failure of function due to mutations. The network approach is also attractive from the modelling perspective, enabling non-local and global affects to be incorporated into computational methods.

In this review it has been shown how small world network approaches have successfully contributed to a range of studies involving protein structure and function. The network approach has been shown to help investigation of global phenomena, such as protein folding, and local phenomena, such as ligand binding and protein-protein docking. The network approach itself is highly abstract and so really useful results cannot be achieved by network descriptors alone, that is, they need to be combined with other factors: for example, with physico-chemical properties and/or sequence conservation measures; or added onto existing models. A list of different methods using network descriptors has been included to provide new researchers in the field with a better understanding of what has been done and possible directions for future research.

Following recent reported success using descriptors derived from network approaches, particularly in protein-ligand binding and protein-protein docking, one could expect to see more widespread use of network based methods, particularly to augment existing models. As algorithms improve and computer hardware becomes more powerful, one would expect to see more investigations using weighted graphs. Furthermore, as experience grows using network descriptors in work with proteins, one would hope to see a wider range of descriptors explored, not limited to the descriptor types used in other fields. Sidechain inter-connectivity in proteins has been fine-tuned by constraints imposed by nature, of packing residues that are covalently linked to one another, and so the network descriptors that best fit this field are probably unlike the descriptors suited to other areas of study. The small world network concept seems well suited for investigating the role of water molecules in protein stability and binding events (which is a current topic of high interest), and for studying protein flexibility, which also has important functional implications. Future research will provide a better understanding of the conditions under which network approaches work best, and when they should be included into research programs.

### Citation

Taylor NR (2013) Small world network strategies for studying protein structures and binding. *Computational and Structural Biotechnology Journal*. 5 (6): e201302006. doi: <http://dx.doi.org/10.5936/csbj.201302006>

## References

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K et al. (2002) *Molecular Biology of the Cell*. Garland.
2. Petsko GA, Ringe D (2004) *Protein Structure and Function*. New Science Press.
3. Löll PJ, Lattman EE (2008) *Protein Crystallography: A Concise Guide*. Hopkins University Press.
4. Timmerman H, Gubernator K, Bohm HJ, Mannhold R, Kubinyi H (1998) *Structure Based Ligand Design (Methods and Principles in Medicinal Chemistry)*. Wiley-VCH.
5. Schwede T, Peitsch MC (2008) *Computational Structural Biology*. World Scientific.
6. Bollobas B (1998) *Modern Graph Theory*. Springer.
7. Barabasi AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509-512.
8. Dorogovtsev SN, Mendes JFF (2006) *Evolution of Networks*. Oxford: Oxford University Press
9. Watts DJ, Strogatz SH (1998) Collective dynamics of small world networks. *Nature* 393: 440-442.
10. Barabasi AL (2002) *Linked*. Cambridge, Massachusetts: Perseus Publishing.
11. Buchanan M (2002) *Nexus*. New York, London: WW Norton & Company.
12. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D -U (2006) Complex networks: Structure and dynamics. *Phys Rev* 424: 175-308.
13. Csermely P, Korcsmaros T, Kiss HJM, London G, Nussinov R (2012) Structure and dynamics of molecular networks: A novel paradigm of drug discovery. A comprehensive review. *Pharmac Therap* arXiv:1210.0330 <http://arxiv.org/abs/1210.0330>.
14. Boede C, Kovacs IA, Szalay MS, Palotai R, Korcsmaros T, et al. (2007) Network analysis of protein dynamics. *FEBS Letters* 581: 2776-2782.
15. Krishnan A, Zbilut JP, Tomita M, Giuliani A (2008) Proteins as networks: Usefulness of graph theory in protein science. *Curr Protein Pept Sci* 9: 28-38.
16. Erdos P, Renyi A (1959) On random graphs I. *Publ Math (Debrecen)* 6: 561-568.
17. Erdos P, Renyi A (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* 5: 17-61.
18. Fell DA, Wagner A (2000) The small world of metabolism. *Nat Biotechnol* 18: 1121-1122.
19. Maslov S, Sneppenm K (2002) Specificity and stability in topology of protein networks. *Science* 296: 910-913.
20. Montoya JM, Sole RV (2002) Small world patterns in food webs. *J Theor Biol* 214: 405-412.
21. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551-1555.
22. Williams RJ, Berlow EL, Dunne JA, Barabasi AL, Martinez ND (2002) Two degrees of separation in complex food webs. *Proc Natl Acad Sci U S A* 99: 12913-12916.
23. Vendruscolo M, Paci E, Dobson CM, Karplus M (2001) Three key residues form a critical contact network in a protein folding transition state. *Nature* 409: 641-645.
24. Dokholyan NV, Li L, Ding F, Shakhnovich EI (2002) Topological determinants of protein folding. *Proc Natl Acad Sci U S A* 99: 8637-8641.
25. Vendruscolo M, Dokholyan NV, Paci E, Karplus M (2002) Small-world view of the amino acids that play a key role in protein folding. *Phys Rev E* 65: 061910.

26. Greene LH, Higman VA (2003) Uncovering network systems within protein structures. *J Mol Biol* 334: 781-791.
27. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanel D, et al. (2004) Network analysis of protein structures identifies functional residues. *J Mol Biol* 344: 1135-1146.
28. Atilgan AR, Akan P, Baysal C (2004) Small-world communication of residues and significance for protein dynamics. *Biophysical Journal* 86: 85-91.
29. Thibert B, Bredesen DE, del Rio G (2005) Improved prediction of critical residues for protein function based on network and phylogenetic analysis. *BMC Bioinformatics* 6: 213-227.
30. Bagler G, Sinha S (2005) Network properties of protein structures. *Physica A* 346: 27-33.
31. Kundu S (2005) Amino acid network within protein. *Physica A* 346: 104-109.
32. Brinda KV, Vishveshwara S (2005) A network representation of protein structures: implications for protein stability. *Biophys J* 89:4159-4170.
33. Paszkiewicz KH, Sternberg MJE, Lappe M (2006) Structural bioinformatics prediction of viable circular permutants using a graph theoretic approach. *Bioinf* 22: 1353-1358.
34. del Sol A, Fujihashi H, Amoros D, Nussinov R (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol Sys Biol* 2: 2006.0019.
35. Muppurala UK, Li Z (2006) A simple approach for protein structure discrimination based on the network pattern of conserved hydrophobic residues. *Prot Eng Des Sel* 19: 265-275.
36. Aftabuddin M, Kundu S (2006) Weighted and unweighted network of amino acids within protein. *Physica A* 369: 895-904.
37. Aftabuddin M, Kundu S (2007) Hydrophobic, hydrophilic, and charged amino acids networks within protein. *Biophysical J* 93: 225-231.
38. Jiao X, Chang S, Li CH, Chen WZ, Wang CX (2007) Construction and application of the weighted amino acid network based on energy. *Phys Rev E* 75: 051903.
39. Chea E, Livesay DR (2007) How accurate and statistically robust are catalytic site predictions based on closeness centrality? *BMC Bioinformatics* 8: 153-156.
40. Huang J, Kawashima S, Kanehisa M (2007) New amino acid indices based on residue network topology. *Genome Informatics* 18: 152-161.
41. Gaci O, Balev S (2008) A general model for amino acid interaction networks. *World Acad Sci Eng Tech* 20: 401-405.
42. Li J, Wang J, Wang W (2008) Identifying folding nucleus based on residue contact networks of proteins. *Proteins: Struc Func Bioinf* 71: 1899-1907.
43. Milenkovic T, Filippis I, Lappe M, Przulj N (2009) Optimized null model for protein structure networks. *PLoS ONE* 4: e5967 doi:10.1371/journal.pone.0005967 <http://arxiv.org/abs/1210.0330>.
44. Vijayabaskar MS, Vishveshwara S (2010) Interaction energy based protein structure networks. *Biophys J* 99: 3704-3715.
45. Petersen SB, Neves-Petersen MT, Henriksen SB, Mortensen RJ, Geertz-Hansen HM (2012) Scale-free behaviour of amino acid pair interactions in folded proteins. *PLoS ONE* 7: e41322 doi:10.1371/journal.pone.0041322 <http://dx.doi.org/10.1371/journal.pone.0041322>.
46. del Sol A, Fujihashi H, Amoros D, Nussinov R (2006) Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families. *Prot Sci* 15: 2120-2128.
47. Hu Z, Bowen D, Southerland WM, del Sol A, Pan Y, et al. (2007) Ligand binding and circular permutation modify residue interaction network in DHFR. *PLoS Comput Biol* 3: e117 doi:10.1371/journal.pcbi.0030117 <http://dx.doi.org/10.1371/journal.pcbi.0030117>.
48. Maetschke SR, Yuan Z (2009) Exploiting structural and topological information to improve prediction of RNA-protein binding sites. *BMC Bioinformatics* 10: 341-354.
49. Liu R, Hu J (2011) Computational prediction of heme-binding residues by exploiting residue interaction network. *PLoS ONE* 6: e25560 doi:10.1371/journal.pone.0025560 <http://dx.doi.org/10.1371/journal.pone.0025560>.
50. Xiong Y, Xia J, Zhang W, Liu J (2011) Exploiting a reduced set of weighted average features to improve prediction of DNA-binding residues from 3D structures. *PLoS ONE* 6: e28440 doi:10.1371/journal.pone.0028440 <http://dx.doi.org/10.1371/journal.pone.0028440>.
51. Kuhn B, Fuchs J, Reutlinger M, Stahl M, and Taylor NR (2011) Rationalizing tight ligand binding through cooperative interaction networks. *J Chem Inf Model* 51: 3180-3198.
52. del Sol A, Fujihashi H, O'Meara P (2005) Topology of small-world networks of protein-protein complex structures. *Bioinformatics* 21: 1311-1315.
53. del Sol A, O'Meara P (2005) Small-world network approach to identify key residues in protein-protein interaction. *Proteins: Struc Func Bioinf* 58: 672-682.
54. Chang S, Gong XQ, Jiao X, Li CH, Chen WZ, et al. (2010) Network analysis of protein protein interaction. *Chinese Sci Bull* 55: 814-822.
55. Pons C, Glaser F, Fernandez-Recio J (2011) Prediction of protein-binding areas by small-world residue networks and application to docking. *BMC Bioinformatics* 12: 378-387.
56. Cheng TM-K, Blundell TL, Fernandez-Recio J (2007) pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins: Struc Func Bioinf* 68: 503-515.

**Keywords:**

Small world network, protein structure, protein-ligand binding, protein-protein binding

**Competing Interests:**

The authors have declared that no competing interests exist.



© 2013 Taylor.

Licensee: Computational and Structural Biotechnology Journal.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are properly cited.

**What is the advantage to you of publishing in *Computational and Structural Biotechnology Journal (CSBJ)* ?**

- ✚ Easy 5 step online submission system & online manuscript tracking
- ✚ Fastest turnaround time with thorough peer review
- ✚ Inclusion in scholarly databases
- ✚ Low Article Processing Charges
- ✚ Author Copyright
- ✚ Open access, available to anyone in the world to download for free

[WWW.CSBJ.ORG](http://WWW.CSBJ.ORG)