Amino Acids (2014) 46:1419–1439
DOI 10.1007/s00726-014-1710-6

REVIEW ARTICLE

The construction of an amino acid network for understanding protein structure and function

Wenying Yan · Jianhong Zhou · Maomin Sun · Jiajia Chen · Guang Hu · Bairong Shen

Received: 5 April 2013 / Accepted: 21 February 2014 / Published online: 13 March 2014

Abstract Amino acid networks (AANs) are undirected networks consisting of amino acid residues and their interactions in three-dimensional protein structures. The analysis of AANs provides novel insight into protein science, and several common amino acid network properties have revealed diverse classes of proteins. In this review, we first summarize methods for the construction and characterization of AANs. We then compare software tools for the construction and analysis of AANs. Finally, we review the application of AANs for understanding protein structure and function, including the identification of functional residues, the prediction of protein folding, analyzing protein stability and protein–protein interactions, and for understanding communication within and between proteins.

Keywords Amino acid network · Network properties · Software tools · Protein structure and function

Introduction

Proteins are biological macromolecules made up of a linear chain of amino acids that fold into three-dimensional structures comprised of different secondary structure elements. They play essential roles in biological systems and act as structural materials, catalysts, adapters, transporters, and regulators through their unique three-dimensional (3D) structures (Anfinsen 1973). Many experimental and theoretical methods have been developed toward the understanding of protein structures and functions (Go 1983; Alm and Baker 1999; Matsumura et al. 1989; Shoichet et al. 1995). One computational technique recently explored is network characterization, which transforms the protein 3D structure into an amino acid network. The network-based analysis of amino acid interactions in protein structures provides a new perspective in the study of protein systems.

In the last two decades, network analysis has become a powerful tool in many real-world complex networks, such as social networks (Wasserman and Faust 1994), internet networks (Faloutsos et al. 1999), road networks (Kalapala et al. 2006), gene co-expression networks (Stuart et al. 2003; Yan et al. 2010; Tang et al. 2010; Chis et al. 2011; Chen et al. 2013), protein–protein interaction networks (Ito et al. 2001), and metabolic networks (Ma and Zeng 2003). Analysis of the properties of such networks can shed light on the interactions among individual components and the whole system. Many real-world networks have shown some common network structure properties such as small worldness (Watts and Strogatz 1998) and scale freeness (Barabási and Albert 1999). In network models, nodes denote the system elements and edges denote their interactions. A weight may be assigned to edges to characterize edge strength (e.g., affinity, intensity, or probability).

Protein structures can be modeled as undirected networks composed of amino acids and their interactions (Zhou et al. 2014; Hu et al. 2013, 2014) (Fig. 1a). These networks are termed residue interaction networks, protein structure networks, or amino acid networks. In this review, we employ the term amino acid network (AAN) to distinguish it from the protein–protein interaction network. Compared with traditional structure-based methods, studying a protein from a network perspective can provide

W. Yan and J. Zhou: contributed equally to the work.

W. Yan · J. Zhou · M. Sun · J. Chen · G. Hu · B. Shen (✉)
Center for Systems Biology, Soochow University, Suzhou 215006, Jiangsu, China
e-mail: bairong.shen@suda.edu.cn

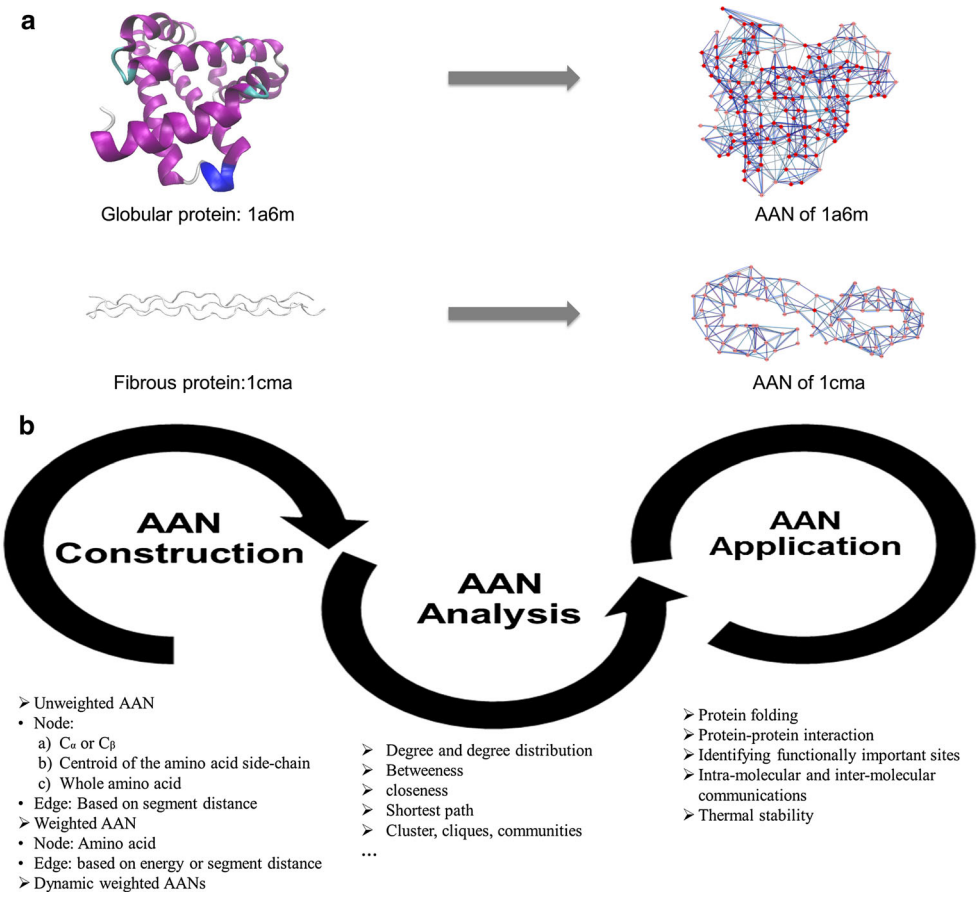

Fig. 1 Methods and tasks for AAN construction and applications. **a** 3D protein structure of the globular protein myoglobin (PDB identifier: 1a6m) and the fibrous protein met repressor (PDB identifier: 1cma) and their corresponding AANs. 3D structures are constructed by VMD with a new cartoon representation. Network structures were built with the RINerator software. **b** Schematic diagram of AAN construction, characteristics, and applications

topological information and capture the global connectivity in a protein molecule. It also permits investigations into the role of each individual amino acid within the complex interacting network (Vendruscolo et al. 2002; del Sol and O'Meara 2005). However, the AAN simplification of protein 3D structure may cause the loss of some spatial information and hinder investigations into the steric effects of protein under specific biological conditions.

In the last years, some reviews on amino acid networks have already been published. Susan primarily reviewed the topological characteristics of amino acid networks and emphasized specific aspects of short-range and long-range links (Susan 2010). Greene reviewed the construction of AANs based on the C_α/C_β backbone and its applications in studies on protein folding (Greene 2012). Di Paola et al. (2013). summarized methods for the generation of amino acid networks and discussed the application of network properties to protein signaling, allosterism, and folding.

Other reviews have focused on the application of AANs. For example, Csermely reviewed the use of network analysis to predict active centers in proteins (Csermely 2008). Csermely et al. (2012) also contributed an interesting and comprehensive review on the relationship between protein disorder and network disorder, and the effect of network disorder on protein structure, dynamics, and function. Most recently, Zhang and colleagues discussed the evolution of protein sequence, structure, and interaction from the view of amino acid networks (Zhang et al. 2013). Hegedus et al. (2013) described network-related methods for understanding the allosteric coupling and importance of intramolecular interactions in ABC (ATP-binding cassette) proteins. Although the amino acid network is becoming a popular model, a comprehensive overview of the construction, characteristics, software, and applications of AANs is yet to be summarized.

In this paper, we aim to review recent work on the construction, analysis, and application of AANs (Fig. 1b). We first review amino acid network construction methods and topological characterization of the networks. Next, recently emerging tools for AAN construction and analysis are summarized. Finally, we concentrate on the biological insights obtained from AANs, including the identification of functional residues, the prediction of protein folding, protein stability, and protein–protein interaction, and the understanding of communication within and between proteins.

Table 1 Methods for amino acid network construction

Nodes	Links	Network type
C_α	Node distance less than a threshold 7 Å (Bartoli et al. 2007; Susan 2011; Gaci and Balev 2009; Bagler and Sinha 2005; Vishveshwara et al. 2009), 8 Å (Bagler and Sinha 2007) and 8.5 Å (Dokholyan et al. 2002)	Unweighted
C_β (C_α for Gly)	Node distance less than a threshold 7 Å and 8.5 Å (Atilgan et al. 2004)	Unweighted
Centroids of side chain	Node distance less than $R_c = 8.5$ Å (Alves and Martinez 2007)	Unweighted
Amino acid	if distance between any atoms from the whole amino acids (Greene and Higman 2003; Bartoli et al. 2007) or only from the side chain (Aftabuddin and Kundu 2007) is less than $R_c = 5$ Å	Unweighted
	Strength of non-covalent interactions based on atom–atom contact and only consider atom from side chain. Atom–atom distance cutoff: $R_c = 5$ Å (Vishveshwara et al. 2009; Brinda and Vishveshwara 2005)	
Amino acid	Links: if distance between any atoms from the whole amino acid residue (Aftabuddin and Kundu 2006) or only from the side chain (Aftabuddin and Kundu 2007) is less than $R_c = 5$ Å	Weighted
	Weight: the number of possible atom–atom links	
Atom	The summation of the electrostatic interaction energy (Coulomb potential) and the van der Waals interaction energy (Lennard-Jones potential) between two atoms (Veloso et al. 2007)	Weighted
Amino acid	The summation of the electrostatic interaction energy (Coulomb potential) and the van der Waals interaction energy (Lennard-Jones potential) between two amino acid (Vijayabaskar and Vishveshwara 2010)	Weighted
Geometrical center of the side chain	Links: Node distance less than a threshold	Weighted
	Weight: Miyazawa and Jernigan contact energy between two residues (Jiao et al. 2007)	
Amino acid	Links: within a cutoff distance or for at least 75 % of an MD	Weighted
	Weight: based on cross-correlation between the monomer over the course of the MD simulation (Sethi et al. 2009; Vanwart et al. 2012)	

Amino acid network construction

Amino acid networks (AANs) are usually constructed from the Cartesian coordinates of amino acid residues of protein molecules stored in the protein data bank (Berman et al. 2000). The nodes represent amino acid residues in the protein while the edges represent interactions between the amino acid residues. The different methods for the construction of AANs are summarized in Table 1. The amino acids (nodes) can be represented by atoms like C_α (Huang et al. 2007; Deb and Vishveshwara 2009; Morita and Takano 2009; Susan 2011; Bagler and Sinha 2005; Bartoli et al. 2007) and C_β (Estrada 2010; Atilgan et al. 2007) in the center of the amino acid, or by the centroid of the side chain (Alves and Martinez 2007) obtained from the spatial coordinates of heavy atoms in the amino acid residue (Greene and Higman 2003). To date, unweighted amino acid networks are more widely used than weighted ones. In many unweighted AANs, the edges are established if the physical distances between the nodes are less than a cutoff distance, R_c, which ranges from 5 to 9 Å. Especially for the AANs whose nodes are denoted as whole amino acids, edges between two nodes represent the interaction between all atoms in the amino acids. Hence, there are multiple edges between two nodes and much larger node degrees (Greene and Higman 2003). Based on the all-atom interaction model, Aftabuddin and Kundu (2006) assigned the weight of the edge between residues as the number of atom–atom contacts and constructed a "weighted" amino acid network. AANs based on distance between the single atoms or centroids of the side chain are regarded as coarse-grain level models since they only indicate the general protein structure shape. In comparison, AANs based on the distances between all atoms between residues can be more accurate (although computational intensive) as they reflect more information regarding physical contacts (Bartoli et al. 2007).

The edges of another type of unweighted amino acid network are defined based on distance-based functions. One example is the side-chain level model (Brinda and

Vishveshwara 2005; Vishveshwara et al. 2009), where edges are denoted based on the normalized strength of non-covalent interactions between the side chains of two amino acid residues in the protein. The strength I_{ij} between residue i and residue j is evaluated by the percentage of atom–atom contacts from the side chains of the two residues i and j. Then a cutoff value $I_{\min}$ is chosen and the residues i and j are connected if $I_{ij} > I_{\min}$. In this model, the connections between covalently linked backbone neighbors are ignored and the non-covalent interactions are focused on instead. This method not only describes the interactions between side chains but also quantifies the strength of interactions.

In the construction of AANs based on the distance between amino acid residues, regardless of the method used to define edges (by direct distance or a distance-based function), the cutoff distance, R_c, is the key factor which affects the performance of the network model. For the all-atom distance model, a favorable cutoff is in the range of 4.5–5 Å, the largest distance that does not accommodate water molecules between two neighbor residues (Hinds and Levitt 1992). Based on this cutoff, to get the minimal average Hamming distance between the networks based on all-atom links and C_α links, the threshold of C_α is determined as 7 Å (Bartoli et al. 2007). da Silveira et al. (2009) carried out more comprehensive work to investigate how the distance cutoff influences amino acid contact by a comparative analysis between cutoff-dependent and cutoff-free methods. They found that 7 Å emerged as an ideal lower bond cutoff to represent contact in proteins for both the C_α and side-chain geometric center representation model. This value is independent of these two representations because all contacts are complete and legitimate (not occluded) in the model at this distance. Moreover, they also reported that biases could be introduced with a cutoff set below 6.8 Å. In this case, C_α representation models may introduce a bias toward alpha proteins and side-chain geometric center representation models introduce a bias toward beta proteins.

The amino acid network based on distance between atoms is an abstract description of protein structure, solely taking into account amino acid contacts at a geometric level and not the chemical properties of the protein. Therefore, another approach for the simulation of protein residue interactions is proposed, employing the energy between amino acid residues and assigning the energy as weight of the edges. In one scenario, energy is composed of two separate energy terms: electrostatic interaction energy (Coulomb potential) and van der Waals interaction energy (Lennard-Jones potential). Veloso et al. (2007) constructed the weighted networks of 12 myoglobins, with atoms as the nodes and the energy between the atoms as the weighted edges, and took into account the aspects of occluded interactions. Vijayabaskar and Vishveshwara (2010) also modeled protein structures as weighted AANs by summing the two energy compositions at the residue level averaged over the equilibrium ensemble. Another scenario is to estimate the interaction between amino acid residues as weighted links according to the magnitude of the contact energy between two residues, as suggested by Miyazawa and Jernigan (Jiao et al. 2007).

The dynamic-weighted amino acid network is a comprehensive method that combines network theory with dynamical information from molecular dynamics simulations. The methods could provide equilibrium global properties and have good applications in the analysis of protein allostery (Vanwart et al. 2012) and signaling pathways in protein–RNA complexes (Sethi et al. 2009). The weights of edges between nodes are obtained from the residuewise cross-correlation values in the simulation (Sethi et al. 2009; Vanwart et al. 2012). The dynamics-weighted AANs could then be constructed by the software PSN-Ensemble (Bhattacharyya et al. 2013), Wordom (Seeber et al. 2011) and NetworkView (Eargle and Luthey-Schulten 2012). The detailed description of these software tools will be shown in a later section on software tools for construction and analysis of AANs.

Characterization of amino acid network

Degree and degree distribution

In networks, the degree (k) of a node is the number of connections it has to other nodes. The mean degree ($\langle k \rangle$) of a network is the average degree of all nodes and it reflects the connectivity of the network. Amino acid networks have an average degree with relatively weak variation when protein size increases (Alves and Martinez 2007; Gaci and Balev 2009; Susan 2010). The mean degree $\langle k \rangle$ increases relatively quickly for small proteins and then stabilizes in large proteins for all structure classes (SCOP classification: α, β, α + β, and α/β) (Bagler and Sinha 2005). There is a predominantly lower $\langle k \rangle$ value for all α proteins compared with other classes as alpha proteins have a lower percentage of more-connected residues than other proteins (Alves and Martinez 2007).

In Gaci and Balev's model (2009), where protein structure is represented as a secondary structure elements (SSE) interaction network (SSE–IN) (a subgraph induced by the set of amino acids participating in SSE that ignores the amino acids not in SSE), regardless of the size of the network, the average degree is always between 5 and 8, even for networks with a size ratio of 100. The bound of the average degree is also related to the percentage of edges connected to SSEs. Moreover, the mean degree of

each SSE subgraph also ranges from 5 to 8 independent of SSE size and type.

Degree distribution, p_k, is the probability that a node has degree k and the cumulative degree distribution, p_k, is the probability that a node has a degree of at least k. Amino acid networks have a bell-shaped, Gaussian-like degree distribution (Greene and Higman 2003; Atilgan et al. 2004; Bagler and Sinha 2005; Susan 2011), rather than a scale-free power-law degree distribution, although the latter is often seen in other biological networks. The degree distribution of amino acid networks is independent of the amino acid spatial location; that is, the Gaussian distribution is valid for both the hydrophobic core and the molten surface of the globular proteins (Atilgan et al. 2004). One of the major reasons for deviations from a scale-free degree distribution is the limited binding capacity of a given amino acid (excluding volume effects) (Bode et al. 2007).

However, there are two kinds of subnetworks whose degrees do not follow the Gaussian distribution as whole AANs. Greene and Higman (2003), and Bagler and Sinha (2007) have constructed, respectively, a type of subnetwork that only contains long-range interactions, showing an exponential type of degree distribution as a single-scale function with a fast decaying tail. Another type of subnetwork is induced by the set of amino acids participating in secondary structure elements (SSE) (Gaci and Balev 2009). This is a truncated scale-free network whose cumulative degree distribution has a power-law regime followed by a sharp exponential cutoff (Amaral et al. 2000). Average degree values constitute a threshold for protein SSE–IN cumulative degree.The cumulative degree slowly decreases when it is lower than the mean degree and then quickly decreases when it is higher than this threshold.

Assortativity

Assortativity refers to a preference of a network's nodes to attach to other nodes and is often examined in terms of node degree (Newman 2003). Assortativity (or assortative mixing) is a measure of degree correlation: if the correlation is positive, there is a preference to high (low) degree nodes to other high (low) degree nodes (otherwise there is disassortativity). There are two prominent parameters to capture such a correlation: neighbor connectivity (Pastor-Satorras et al. 2001) and the assortativity coefficient (Newman 2002). Based on these two measurements, both weighted and unweighted amino acid networks exhibit an assortative mixing pattern (Alves and Martinez 2007; Bagler and Sinha 2007; Susan 2011; Aftabuddin and Kundu 2006). According to Bagler and Sinha's work (2007), both amino acid networks and LINs (only contain long-range interactions) show assortativity, with the correlation coefficient up to 0.58. At the same time, Alves and Martinez (2007) confirmed this property of AANs and additionally found that the average degree of the neighboring nodes ($\langle k_{nn}(k)\rangle$) on degree k can be described by a linear relationship. Moreover, the average of $\langle k_{nn}(k)\rangle$ over all nodes in one network attained a stationary value for a large network size. Similarly, the assortativity coefficients had a stationary pattern and the assortative mixing properties of AANs were independent of protein size (Susan 2011).

Shortest path length

The shortest path length, L_{ij}, is the size of the shortest edge that connects two nodes (i and j) (Watts and Strogatz 1998). The average path length, L, of a network is the mean length of the set of shortest paths between all node pairs. Atilgan et al. (2004) showed that the average of the shortest path lengths of a residue to all other residues in a protein consistently decreases for residues at greater depths, implying that residues in the core of the protein are connected to other residues in a fewer number of steps, and that the shortest path lengths are also highly correlated with residue fluctuations. The average path length, L, of AANs was around 4–7 and of the same order as the corresponding random networks (Bartoli et al. 2007; Vendruscolo et al. 2002; Bagler and Sinha 2007; Susan 2011; Bagler and Sinha 2005). Bagler and Sinha (2005) showed that L logarithmically increases with network size, regardless of the structural classification of proteins. Furthermore, they found that fibrous proteins had a larger diameter ($D_{\text{fibrous}} = 15$), the largest of all shortest path lengths in the network, compared to globular proteins ($D_{\text{globular}} = 8.57$). The larger diameter of fibrous proteins was expected because of its elongated structure. Bartoli et al. (2007) found that the mean values of L for 1,753 protein chains were exactly the same for different network node representations by C_α or all atomic contacts in amino acids. In addition, according to Susan's work (2010), the shortest path length of AANs in 64 proteins demonstrated Gaussian-like distributions. Finally, the shortest path from energy-based AANs was suitable for investigating finer details such as communication paths between distal residues (Vijayabaskar and Vishveshwara 2010).

Cluster and motif

The clustering coefficient, Ci, of a residue i in an AAN reflects the probability that the neighbors of a node are also neighbors of each other. The clustering coefficient of a network is the average over the clustering coefficients of all nodes and the clustering coefficient of a node is the fraction of neighboring node pairs that are connected (Watts and Strogatz 1998). Atilgan et al. (2004) investigated the C_i of residue i as a function of residue depth in an amino acid

network and found that C_i of the residues at greater depth approached a fixed value of $\sim$0.35 beyond a depth of about 4 Å irrespective of protein size. The core residues were found surrounded by other residues and not exposed to the solvent, and the local packing arrangements of the protein were always the same. The clustering coefficient, C, of the AAN was significantly larger than its corresponding random network and the mean of C in different datasets ranged from 0.5 to 0.6 (Susan 2010, 2011; Bagler and Sinha 2005; Vendruscolo et al. 2002; Alves and Martinez 2007). The C values of AANs did not significantly change with an increase of protein size (Bagler and Sinha 2005; Susan 2010; Alves and Martinez 2007). Bagler and Sinha (2005) studied the C values of proteins in different structural classes and found a small but definite difference between the C values of α proteins and β proteins where the α proteins had a slightly larger C value than β proteins ($C_\alpha = 0.588$, $C_\beta = 0.538$). Aftabuddin and Kundu (2006) reported that C values of weighted networks are less than unweighted networks, implying that edges with low weight mainly contribute to topological clustering (Aftabuddin and Kundu 2006).

A high clustering coefficient is suggestive of potential modularity in amino acid networks. Larger units, made up of nodes more densely connected to each other than the rest of the network, often indicate clusters or communities and are usually the essential structural units of real networks (Girvan and Newman 2002). Kannan and Vishveshwara (1999) presented a method to detect side-chain clusters in protein structure in an amino acid network using a network based on the self-defined interaction strength between residues. They detected a variety of side-chain clusters, such as hydrophobic clusters, which showed a good correlation with the folding intermediates observed in experiments. Moreover, according to their method, the determined clusters of AANs corresponded to protein domains. Vishveshwara et al. proposed a method to identify sub-clusters as domains through amino acid network analysis (Sistla et al. 2005). Residue cluster identification and function will be discussed in the following section.

The largest cluster size is often used to assess whether there is a phase transition from the percolation point of view. A giant connected cluster percolates the network, whereas only smaller clusters are present below a threshold (critical point) (Deb and Vishveshwara 2009). Brinda and Vishveshwara (2005) monitored the normalized largest cluster size (in terms of number of amino acids) in the AANs at different cutoffs. A remarkable similarity was found in proteins of various folds and sizes, whereas a transition was observed within a very narrow range of the cutoff. By investigating the role of non-covalent connections in an amino acid network based on C_α contacts and side-chain interactions, only the latter type of AAN at a high level of connections (low cutoff I_{min}) was capable of clique percolation and none of the random models (including one very similar to that of proteins) exhibited clique percolation (Deb and Vishveshwara 2009). In terms of complex network theory, if a clique had k nodes, then a community could be defined as the collection of adjacent k-cliques where each of these cliques share k-1 nodes with the adjacent clique (Derenyi et al. 2005).

The largest community is a percolated clique. Deb et al., therefore, used the terminology of "largest community" for clique percolation. A similar percolation behavior pattern was also found in the energy-based unweighted amino acid network (Vijayabaskar and Vishveshwara 2010), where the largest cluster size could be calculated as the function of energy cutoff. They concluded that most of the residues in proteins are connected with energies between 0 and −10 kJ/mol and that these connections disappear at high interaction energies ($\leq$−20 kJ/mol). This confirms the concept that hydrophobic interactions hold together the local clusters of highly interacting residues, keeping the protein topology intact.

Small worldness

The small-world network is a model often used to describe biological complex systems such as protein interaction networks (Bork et al. 2004), gene co-expression networks (van Noort et al. 2004), and complex brain networks (Sporns et al. 2004). Small-world topology is characterized by a small average shortest path length, L, and larger clustering coefficient, C (L should logarithmically increase with network size) (Watts and Strogatz 1998; Watts 1999). Amino acid networks represent small-world properties in physical distance-based and energy-based models (Jiao et al. 2007; Veloso et al. 2007). Vendruscole et al. (2002) first reported the small-world behavior of amino acid networks. Greene and Higman (2003) confirmed that amino acid networks have small-world properties while long-range amino acid interaction networks do not have small-world topology.

Amino acid networks exhibit small-world topology regardless of their structure classes (e.g., SCOP classification (α, β, α + β, and α/) or classification as fibrous or globular proteins) (Bagler and Sinha 2005), distance cutoff values for construction of the amino acid network (Atilgan et al. 2004; Bartoli et al. 2007), selected protein representations (C_α or all-atom), and residue composition in the studied protein (Bartoli et al. 2007). Furthermore, both hydrophobic and hydrophilic amino acid networks of protein complexes exhibit small-world properties (Chang et al. 2008). In addition, work by Yu et al. (2009) indicated that AANs constructed by peptide and hydrogen bonds, London-van der Waals forces, and hydrophobic effects, show

Table 2 Software tools for amino acid networks

Software	Description/google scholar time cited	Platform	AAN type	AAN analysis	AAN visualization	URL
Wordom module: PSN (Seeber et al. 2011)	One of Wordom module to construct amino acid network (Times Cited: 8)	Wordom	Strength of non-covalent interactions from side chain based	Averaged interaction strength Hub information Cluster compositions and largest cluster size	False	http://wordom.sf.net/
RING (Martin et al. 2011)	A web server that can be used to derive amino acid networks at atomic level (Times Cited: 20)	Web Server	Van der Waals contacts-based Closest atom distance based C_α backbone based	Node degree Subnetworks Energy score function Protein sequence conservation	False	http://protein.bio.unipd.it/ring/
RINerator (Doncheva et al. 2011)	Construct a weighted network with multiple edges (Times Cited: 24)	Python	Edge types Generic residue interaction Interatomic contact Hydrogen bond Overlapping Weights: proportional to the strength of the interaction	False	True	http://rinalyzer.de/rinerator.php
xPyder (Pasi et al. 2012)	Identify and visualize common network parameters (Times Cited: 8)	PyMOL	Dynamic-weighted AAN	Hubs Isolated components Intra- or intermolecular interaction	True	http://linux.btbs.unimib.it/xpyder/
Jamming (Cusack et al. 2007)	Capture the crucial residues for protein function from AANs (Times Cited: 7)	Java	Atom distance based Centers of the residue mass based	Crucial residues	True	http://bis.ifc.unam.mx/jamming/
GraProStr (Vijayabaskar et al. 2011)	Web tool for analyzing protein structures as networks (Times Cited: 3)	Web Server	Side chain based C_α/C_β backbone-based protein ligand networks	Clusters Hubs Cliques Communities	True	http://vishgraph.mbu.iisc.ernet.in/GraProStr/
JGromacs (Munz and Biggin 2012)	a Java package for analyzing protein simulations (Times Cited: 2)	Java	Dynamic-weighted AAN	Atoms or nodes distance Structure similarity	False	http://sbcb.bioch.ox.ac.uk/jgromacs/

Table 2 continued

Software	Description/google scholar time cited	Platform	AAN type	AAN analysis	AAN visualization	URL
NetworkView (Eargle and Luthey-Schulten 2012)	VMD plugin for displaying and analyzing of protein–RNA interaction networks (Times Cited: 7)	VMD	Dynamic AAN Node: one node to each amino acid and two nodes to each nucleotide Egde: within a cutoff distance or for at least 75 % of an MD Dynamic-weighted AAN	Communities Optimal and Suboptimal Paths	True	http://www.ks.uiuc.edu/Research/vmd/plugins/networkview/
PSN-Ensemble (Bhattacharyya et al. 2013)	An automated approach to network features of protein structure ensembles (Times Cited:3)	Perl Matlab	Side chain based C_α backbone based Dynamic-weighted AAN	Hubs, Cluster, K-cliques Community Paths and cost of communication 'Junction residues' between domains	True	http://vishgraph.mbu.iisc.ernet.in/PSN-Ensemble/

strong, very strong, and relatively weak small-world properties, respectively. According to the results of Bartoli et al. (2007), backbone constraints play a fundamental role in determining small-world properties. They also concluded that the average path length, *L*, and clustering coefficient, *C*, are not useful quantities for "protein fingerprinting". However, Morita and Takano (2009) concluded that the amino acid network of native protein structure fractal networks is not a small-world network.

Software tools for construction and analysis of AANs

Several software tools have been developed for the construction of AANs in the past years (see Table 2). AANs are usually constructed from PDB files. Seeber et al. (2011) developed a model for constructing AANs based on the strength of non-covalent interactions from side chains, proposed by Vishveshwara et al., using the Wordom program (Kannan and Vishveshwara 1999; Brinda and Vishveshwara 2005). RING is a web server that derives amino acid networks using three different definitions of the edge: Van der Waals contacts, closest atomic distance, and alpha carbon distance (Martin et al. 2011). The resulting networks can be visualized with Cytoscape (Saito et al. 2012; Smoot et al. 2011; Cline et al. 2007; Shannon et al. 2003). RINerator distinguishes different residue interaction types and quantifies the strength of individual interactions, resulting in an undirected weighted network with multiple interaction edges (Doncheva et al. 2011). Weights of edges between residues are proportional to the strengths of the interactions.

Tools have recently been developed to generate the dynamics of AANs from Molecular Dynamic (MD) simulations or a set of protein structures. Using MD trajectory data, JGromacs constructed the dynamic weights of AANs, with weights defined from correlations between monomers calculated over the course of the MD simulation (Munz and Biggin 2012). NetworkView focuses on the generation of AAN dynamics in protein–RNA complexes (Eargle and Luthey-Schulten 2012). More recently, Bhattacharyya et al. (2013) developed a standalone and exhaustive tool named PSN-Ensemble to generate three types of AANs: side chain-based AANs, C_α-based and weighted AANs from MD simulations, as well as NMR and multiple X-ray structures.

The construction tools mentioned above have provided functions from amino acid network analysis. For example, the Wordom module PSN writes an output file that contains detailed information about average interaction strength, hub information, stable cluster composition, and largest cluster size. RING can produce subnetworks with a user-defined conservation or relative solvent accessibility

Table 3 Application of amino acid network in protein science

Network type	Network parameter	Application	References
Node: residue Edge: use different distance cutoff	Degree: maximum degree and average degree	Discriminate the non-native and native structures	(Muppirala and Li 2006; Pabuwal and Li 2008)
Node: C_α Edge: the residues are in contact if the distance between the nodes less than $R_c = 8$ Å	Average degree Normalized complexity Network flow Weighted flow Connectivity Weighted connectivity	Discriminate the non-native and native structures by Z score and enrichment score	(Vassura et al. 2009)
Node: C_α Edge: the residues are in contact if the distance between the nodes less than $R_c = 6.8$ Å	Degree Clustering coefficient Second connectivity	Discriminate the non-native and native structures by classification methods	(Küçükural et al. 2008)
Node: residue Edge: strength of non-covalent interactions based on atom–atom contact from side chain	The number of non-covalent interactions (ncov) Largest cluster size (slclu) The clustering coefficient of the largest cluster (ccoe) The size of the top large communities (ccoms)	Discriminate the non-native and native structures by Z score	(Chatterjee et al. 2012)
Node: C_α Edge: the residues are in contact if the distance between the nodes less than $R_c = 8.5$ Å	Betweenness	Key residues for protein folding	(Dokholyan et al. 2002)
Node: C_α Edge the residues are in contact if the distance between the nodes less than $R_c = 8$ Å	Impact of edge removal per residue based on average path length Clustering coefficient	Protein unfolding rate	(Jung et al. 2005)
Node: C_α Edge: the residues are in contact if the distance between the nodes less than $R_c = 8$ Å	Coefficient assortativity Clustering coefficient	Protein-folding rate	(Bagler and Sinha 2007)
Node: residue Edge: strength of non-covalent interactions based on atom–atom contact from side chain	Largest cluster size	Unfolding process kinetics	(Ghosh et al. 2007)
Node: C_α the residues are in contact if the distance between the nodes less than $R_c = 8.5$ Å	Average shortest path length	Protein-folding kinetics: Pretransition state Posttransition state	(Dokholyan et al. 2002)
Node: residue Edge: strength of non-covalent interactions based on atom–atom contact from side chain	Hub distribution	Compare the mechanical unfolding of wild-type rhodopsin with mutants	(Fanelli and Seeber 2010)
Node: residue Edge: strength of non-covalent interactions based on atom–atom contact from side chain	Laplacian matrix Cluster	Identify the side-chain cluster and crucial residues at the dimer interface of α-subunit in Escherichia coli RNA polymerase	(Kannan et al. 2001)
Node: residue Edge: strength of non-covalent interactions based on atom–atom contact from side chain	Laplacian matrix Cluster	Analysis of homodimeric proteins interface	(Brinda et al. 2002)

Table 3 continued

Network type	Network parameter	Application	References
Node: residue Edge: different cutoff for different type of interactions Weight: One for atomic backbone–side chain interaction Two for side-chain–side-chain interaction	Clustering	Characterize the modular architecture of protein–protein interfaces	(Reichmann et al. 2005)
Node: residue Edge: strength of non-covalent interactions based on atom–atom contact from side chain	Laplacian matrix Cluster Hubs	Characterize the oligomeric interface of lectins	(Brinda et al. 2005)
Node: residue Edge: two residue are in contact if at least one pair atoms of the residues at a distance $\leq$5 Å	Small world Betweenness	Identify key residues in protein–protein interaction and find their correlation to the hotspots	(del Sol and O'Meara 2005; del Sol et al. 2005)
Node: residue Edge: Two residues (one from each chain) are in contact if the distance between at least one pair atoms of the residues smaller than the sum of their van der Waals radii plus a 0.5 Å tolerance Two residues (within one chain) are in contact if the distance between the C_α of the residues is less than 6 Å Edges weight: knowledge-based potentials	Minimum cut tree	Identify key residues in protein–protein interaction, especially the hotspots and hot regions	(Tuncbag et al. 2010)
Node: residue Edge: two residues are in contact if at least one pair atoms of the residues at a distance smaller than specific distance cutoff Weight: affinity values based on the number of heavy atoms of two residues and total number of atom–atom contacts between the residues	EVT values based on random walk	Model allosteric signal propagation by ETV profile and found correlation between hotspots and residues with high ETV values	(Park and Kim 2011)
Node: residue Edge: two residue are in contact if at least one pair atoms of the residues at a distance $\leq$5 Å	Degree Clustering coefficient Average shortest path length	Define protein–protein docking scoring function and find the difference between the AANs of correct docked complex and incorrect ones.	(Chang et al. 2008)
Node: geometrical center of amino acid side chain Edge: two nodes are in contact If the distance between the node is less than 6.5 Å Weights: knowledge-based potential (contact energy)	Weighted average nearest neighbors degree	Define protein–protein docking scoring function	(Jiao and Chang 2011)
Node: residue Edge: based on the atomic contacts from each residue	Closeness centrality	Identify functional important residues of protein	(Amitai et al. 2004)

Table 3 continued

Network type	Network parameter	Application	References
Node: residue Edge: two residue are in contact if at least one pair atoms of the residue side chain at a distance ≤5 Å	Closeness centrality	Identify functional important residues in protein families	(del Sol et al. 2006)
Node: residue Edge: two residue are in contact if At least one pair atoms of the residues within the cutoff Center of mass were within the cutoff	Shortest path Betweenness Dynamic connectivity and its distribution	Identify functional important residues of protein	(Cusack et al. 2007)
Node: residue Edge: two residue are in contact if at least one pair atoms of the residues at a distance ≤5 Å	Betweenness Shortest path length	Study the relationship of the substrate entrance in the active site and protein stability of mammalian histidine decarboxylase	(Pino-Angeles et al. 2010)
Node: residue Edge: two residues are in contact if the distance between at least one pair atoms of the residues smaller than the sum of their van der Waals radii plus 2 Å	Degree Clustering coefficient Hubscore Cocitation Constraint Betweenness Closeness centrality Shortest path length	Predict catalytic residues in proteins	(Li et al. 2011a)
Node: residue Edge: two residue are in contact if at least one pair atoms of the residues at a distance ≤5 Å Weight: co-evolution score	Degree Clustering coefficient	Identify functional residues and functional modules clusters of rhodopsin	(Park and Kim 2012)
Node: geometrical center of amino acid side chain Edge: two nodes are in contact If the distance between the node is less than 6.5 Å	Degree Clustering coefficient Closeness Betweenness	Predict disease-associated substitution of a single amino acid	(Li et al. 2011b)
Node: C_β and C_α for glycine residues Edges: if the distance between the node is less than 6.7 Å Weights: knowledge-based potential (contact energy)	Connectivity Shortest path length	Characterize the shortest path length in reduced network that under screen edges that have higher than a cutoff potential	(Atilgan et al. 2007)
Node: residue Weights: the summation of the electrostatic interaction energy (Coulomb potential) and the van der Waals interaction energy (Lennard-Jones potential) between two amino acid	Clusters Hubs Shortest path Closeness index	Identify the residues crucial for stabilizing the folding units and the paths can communication between distal residues in the protein.	(Vijayabaskar and Vishveshwara 2010)
Node: residue Edges: strength of non-covalent interactions based on atom–atom contact from side chain	Shortest path Hubs	Identify key amino acids that are responsible for communication between the extracellular and intracellular poles of the LHR	(Angelova et al. 2011)

Table 3 continued

Network type	Network parameter	Application	References
Node: C_α Edges: distance cutoff between nodes $R_c = 6.5$ Å	Degree Betweenness Closeness Clustering coefficient	Predict protein surface residues	(Huang et al. 2007)
Model one: Node: C_α Edges: distance cutoff between nodes $R_c = 6.5$ Å Model two: Node: residues Edges: strength of non-covalent interactions based on atom–atom contact from side chain	Laplacian matrix Cluster	Identify domains and domain interface residues in proteins	(Sistla et al. 2005)
Node: residues Edges: strength of non-covalent interactions based on atom–atom contact from side chain	Shortest path	Communication pathways between the anticodon region and aminoacylation site in methionyl-tRNAsynthetase	(Ghosh and Vishveshwara 2007)
Node: residues Edges: strength of non-covalent interactions based on atom–atom contact from side chain	Cluster Hub Degree distribution Size of largest cluster Shortest path	Intra and intermolecular communications in proteins	(Vishveshwara et al. 2009)
Node: residues Edges: strength of non-covalent interactions based on atom–atom contact from side chain	Laplacian matrix Cluster	Identify functionally important amino acid residue clusters in the vicinity of the protein–DNA interaction site	(Sathyapriya and Vishveshwara 2004)
Node: residues and Nucleotides Edges: strength of non-covalent interactions based on atom–atom contact from residue side chain and nucleotide	Clusters Hubs	Characterize the protein–DNA interactions by an undirected bipartite graph	(Sathyapriya et al. 2008)
Node: residues Edges: strength of non-covalent interactions based on atom–atom contact from two residue side chain	Cliques Communities Hubs Shortest path	Clique and community pattern in protein structures and inter-domain communications	(Ghosh and Vishveshwara 2008)
Node: residues Edges: strength of non-covalent interactions based on atom–atom contact from two residue side chain	Shortest path Clusters	Intra and inter subunit communication in human tryptophanyl-tRNA	(Hansia et al. 2009)
Node: residues Edges: strength of non-covalent interactions based on atom–atom contact from side chain	Laplacian matrix Cluster	Investigate the factors contribute to the thermal stability of thermophilic proteins	(Kannan and Vishveshwara 2000)
Node: Residues Edges: strength of non-covalent interactions based on atom–atom contact from side chain	Largest cluster size Hub Degree distribution Edge number	Protein stability	(Brinda and Vishveshwara 2005)

threshold. PSN-Ensemble can also calculate network parameters such as hubs, clusters, K-cliques, communities, paths, and cost of communication.

Beyond these construction tools, several tools based on different platforms have recently been developed for the analysis and visualization of AANs. Cusack et al. (2007) developed a Java-based program called JAMMING to capture the crucial residues for protein function from AANs. RINalyzer is a Cytoscape plug-in that especially visualizes an amino acid network and analyzes weighted AANs. It is the only tool that supports the simultaneous view of an amino acid network in 2D and the corresponding protein structure in 3D by connecting Cytoscape to the UCSF Chimera molecular structure viewer (Doncheva et al. 2011). In particular, Doncheva et al. (2012) even provided a workflow and protocol that demonstrates how to combine Network Analyzer and RINalyzer for the comparison of multiple AANs. Moreover, some tools, such as the xPyder (Pasi et al. 2012) and NetworkView (Eargle and Luthey-Schulten 2012), not only deal with amino acid networks from the PDB file but also analyze and visualize AANs from MD trajectory data, providing more insight into the dynamic features of the residues. xPyder, a PyMOL plugin, is a useful tool to identify and visualize common network parameters, including hubs and isolated components, as well as intra- or intermolecular interactions (Pasi et al. 2012). In comparison, NetworkView pays more attention to the display and analysis of protein–RNA interaction networks, and identifies major communication pathways by edge betweenness (Eargle and Luthey-Schulten 2012).

We compared the number of Google Scholar citations of these software tools before 2013 (see Table 2) and found that Wordom, RINerator, and RINalyzer were the most cited tools (cited more than 20 times), as well as general network analysis software for amino acid network analysis, including R packages, [e.g., igraph (Csardi and Nepusz 2006), sna (Handcock et al. 2008), tnet (Opsahl et al. 2010)], the Java framework JUNG (Madadhain et al. 2005), the Python package NetworkX (Hagberg et al. 2008), CFinder (Adamcsek et al. 2006), and Cytoscape plug-ins such as NetworkAnalyzer (Assenov et al. 2008) and ClusterMaker (Morris et al. 2010).

Application of amino acid networks

The amino acid network view of a protein has brought new insight into protein science. In this section, we summarize the application of AANs to hot topics in protein science, including protein-folding, protein–protein interactions, identification of functionally important residues, communication within and between proteins, and the thermal stability of proteins. The type of networks and their corresponding applications are summarized in Table 3.

Protein folding

Since most proteins attain their functionality through their unique and native three-dimensional states, accurately predicting protein folding is a central objective in structural biology. Rationalizing the formation of native states from a topological point of view is an important strategy for this objective. Alm and Baker (1999) even suggested that ''protein folding mechanisms and landscapes are largely determined by the topology of the native state and are relatively insensitive to details of interatomic interactions.'' Betweenness with larger values based on AANs have been used to determine the role of three key residues in the process of protein folding (Vendruscolo et al. 2001, 2002).

Discrimination of native structures from non-native structures

An important aspect of understanding the protein-folding mechanism is to discriminate native structures from non-native ones. Different types of AANs and many topological-based score functions that have been used to distinguish native and non-native structures are completely independent of the primary protein structure. Using the maximum degree and the number of residues with that maximum degree, in non-native state and native state AANs, to identify non-native structures exhibited a better performance in most cases compared to hydrophobic fitness (HF) score and Evolutionary Trace (ET)-based scores (Muppirala and Li 2006). Non-native structures had either a smaller value of the maximum degree or fewer numbers of residues with the same maximum degree than native ones. Similar results were also found when discriminating non-native states from the native-state of helical membrane proteins (Pabuwal and Li 2008), using an AAN model with a backbone level network that takes the C_α as a node. Vassura et al. (2009) also employed C_α representation AANs and seven network properties (average degree, contact order, normalized complexity, network flow, weighted flow, connectivity, and weighted connectivity) to define the Z score and enrichment score to check whether the native structure ranked high. The results indicated backbone geometry as one of the most relevant pieces of information.

Using several classification methods, Küçükural et al. (2008) defined a function to combine the degree, clustering coefficient, second connectivity, and Jernigan potential scores to distinguish native structures from decoys and found that network properties have more discriminating power. In particular, the degree and clustering coefficient

accuracy was as high as 98.72 %. Recently, Chatterjee et al. (2012) used AANs based on the strength of non-covalent interactions from atom–atom side-chain contacts to identify native structures from decoys. First, they characterized the network properties of AANs of native structures that could provide a reference scale to assess the quality of decoy structures. They investigated four-network parameters: the number of non-covalent interactions (NCov), the largest cluster size (SLClu), the clustering coefficient of the largest cluster (CCoe), and the size of the top large communities (CComS). After providing a statistical range of the native protein from the parameters, rank and Z scores of the native structures compared to decoys for the parameters showed a good performance but were dataset specific. Moreover, higher-order network parameters, like CCoe and CComS, identified native structures more effectively than NCov and SLClu. Most recently, we successfully developed an integrated score function (SVR_CAF) which incorporates the contact energy-based score (CE_score), amino acid network-based score (AAN_score), and the fast Fourier transform-based score (FFT_score) to discriminate native structures from decoys (Zhou et al. 2013).

Protein-folding kinetics

The problem of protein-folding kinetics is an important issue and the network concept can be used to investigate the different aspects. First, key residues critical for protein folding can be predicted by the network method. Dokholyan et al. (2002) showed that the key residues, acting as nucleation for protein folding, have large betweenness values in the protein transition states. Second, some network properties show some degree of a relationship with protein unfolding and folding rates. Jung et al. (2005) found that the impact of edge removal per residue, the ratio of the change of the average path length divided by the protein's size to edge removal probability, had a high correlation with protein unfolding rate. Bagler and Sinha (2007) found that the network parameter coefficient of assortativity has a positive correlation with protein-folding rates at both short and long contact scales. In contrast, the clustering coefficient of AANs only constructed by long interactions exhibited a negative correlation with protein-folding rate. From these two works, neither the folding rate nor unfolding rate showed any significant relationship with the clustering coefficient of the AANs.

Finally, by investigating the topological properties of AANs and network evolution of molecular dynamic trajectory data, the whole process of folding and unfolding can be monitored. For example, the transition states during folding and unfolding have revealed important features about the process. Using two well-characterized proteins, CI2 and C-Src SH3, Dokholyan et al. (2002) determined the topological difference of AANs between two protein transition states, pretransition states and posttransition states. They found that average shortest path lengths for posttransition conformation AANs were smaller than those for pretransition conformation AANs. The difference in average shortest path length, $L(i)$ (defined as the mean shortest path length between the node i and the rest of the network nodes), between the pre- and posttransition state had the largest values for the experimentally identified folding nucleus. During the protein-folding process, most connected residues decreased the number of edges and the small worldness of the AANs increased as the protein structure becomes more and more compact (Dokholyan et al. 2002). Ghosh et al. (2007) monitored the unfolding simulations of T4 lysozyme by following the largest cluster size, and found that the cluster size was large in the equilibrium simulation and considerably decreased as the simulation temperature increased. Jiao et al. (2007) constructed contact energy-based weighted AANs on the MD trajectory of protein CI2 and found that protein unfolding is mainly caused by the derogation of the hydrophobic core and that the shortest path lengths increased in the unfolding process. Moreover, the unfolding behavior is also associated with the loss of many hubs in the native state and the native hubs showed a correlation in experimental mutation results. Some non-mutated hub residues are predicted to be the targets of mutation for protein destabilization. Fanelli and Seeber (2010) combined atomistic steered molecular dynamics, single-molecule force spectroscopy, and AAN analysis on bovine rhodopsin to compare the mechanical unfolding of wild-type rhodopsin with that of mutants. From the analysis of the distribution of highly connected nodes they found a diffuse intramolecular communication inside and between two poles of the helix bundle and that selected mutations shared a more or less marked ability to impair hubs in the AAN.

Protein–protein interactions

Identifying and understanding the underlying principles of protein–protein interactions are other essential applications of AANs, since many proteins carry out their cellular functions by interacting with other proteins. Different methods are available to investigate protein–protein interactions based on the characterization of local geometrical, chemical, and energetic features of the interfaces. Analysis of AANs of protein–protein complexes provides a global perspective on the interaction across the interface, which is difficult to obtain from pairwise interaction analysis or from the loss of accessible surface area calculations.

Characteristics of protein interfaces

By applying amino acid network methods, several attempts have been made to capture the characteristics of protein interfaces. Kannan et al. (2001) applied AANs to identify the side-chain cluster at the dimeric interface of the α-subunit in *Escherichia coli* RNA polymerase (RNAP). A nine-residue cluster was identified in the subunit interface adjacent to the hydrophobic core and most of the cluster residues were topologically and sequentially conserved in *T. aquaticus* RNAP crystal structures. Two residues, F35 and I46, were predicted as crucial for the stability of the α-dimer interface. Brinda et al. (2002) also used the same AAN representation to identify side-chain clusters at the dimer interface and the center of these clusters in a set of homodimeric proteins. Their results indicated that both charged and hydrophobic residues are involved in stabilizing the dimer interface, and that arginine, histidine, phenylalanine, tyrosine, and glutamic acid seem to be the most preferred residues. Furthermore, from a weighted analysis of AANs, the interface between proteins was found to have a modular architecture, with clusters of residues containing both strong and weak intra-cluster interactions (Reichmann et al. 2005). Mutation in one module did not affect residues located in a neighboring module but caused complex energetic and structural consequences within the module (Reichmann et al. 2005). In addition, the network-based method performed with the sequence alignment method has provided signature sequence motifs characterizing the protein–protein interface in multimeric proteins like lectins (Brinda et al. 2005) and the AAN approach was found ideally suited for the detection such motifs through clusters at the interface.

AAN analysis identified the important binding hot spots that significantly contribute to binding free energy at the protein interface. Brinda et al. (2002) found that the side-chain clusters detected at the interface of a set of homodimeric proteins correlated well with the location of hot spots at the protein–protein interfaces. From the small-world network perspective, (del Sol and O'Meara 2005; del Sol et al. 2005) used betweenness to identify the highly central residues that frequently occur at protein interfaces. These identified residues were either experimentally determined hot spots (clustered into tightly packed regions called hot regions) or in direct contact with experimental hot spots. Based on weighted amino acid networks, Tuncbag et al. (2010) proposed a minimum cut tree method to extract hot spots and their organization. The most connected residues in the mincut tree generally corresponded to hotspots and the cluster residues corresponded to hot regions. Recently, by constructing an amino acid network and modeling signal transmission with a Markov random walk, Park and Kim (2011) defined the expected visiting time (EVT), which describes the signal-induced effects caused by signal transmission through all possible routes, to estimate the global perturbation effects from particular signal initiation sites. As a result, they found that hot spots have significantly high ETV values.

Protein–protein docking

Amino acid network analysis is used to assess protein–protein docking to forecast complex structure based on the structure of two monomers. Chang et al. (2008) constructed a hydrophobic and hydrophilic amino acid network of a protein–protein complex and defined the simple scoring terms that composed the network parameters, the degree and clustering coefficient. Scoring functions including these two network-based scoring terms from the hydrophobic and hydrophilic amino acid network showed a better performance than the energy function score, and it also permitted improvements to the scoring function of RosettaDock when combined with other energy terms. Moreover, they also found that correctly docked complex conformations had higher degree values and lower clustering coefficients than incorrect ones. Jiao and Chang (2011) constructed a weighted amino acid network by assigning the contact energy to the network links on 42 dimer complexes. The strength (the sum of all link weights of a network) and the weighted average nearest neighbor degree were employed to define a scoring function. Compared with the pair potential scoring functions, the network-based scoring function had a similar performance to the pair potentials on some items, but a better success rate due to its simplicity and clarity in calculation.

Identifying functionally important sites

Identifying functionally important sites in proteins is a difficult task and further complicated when there are few or no recognizable homologs. Fortunately, viewing protein structure as an amino acid network provides a novel aspect into the identification of functional sites, as it does not rely on sequence conservation or any prior knowledge. The closeness value, a network parameter used to measure residue centrality, was found to associate with active sites (Amitai et al. 2004; del Sol et al. 2006). Combining closeness centrality and surface accessibility, Amitai et al. (2004) predicted 70 % of active-site residues in enzyme families. The correlation between centrally conserved positions and active-site residues was particularly strong for enzyme families compared with non-enzyme families. Many centrally conserved amino acids were clustered with active-site residues in cavities or clefts (del Sol et al. 2006). Cusack et al. (2007), identified crucial residues for protein function through determining the most traversed residues in

networks based on shortest path and betweenness. Recently, Li et al. (2011a) provided a more comprehensive method to predict catalytic residues, taking several network parameters into account, especially the novel descriptor description of network signal communication and the layered description of the structure environment they developed based on shortest path length. Furthermore, Huang et al. (2007) defined four-network parameters by normalizing common network parameters (e.g., degree, clustering coefficient, closeness, betweenness) to predict protein surface residues and found that the best new parameter was based on degree.

Network parameters are also used to investigate functional modules in proteins. The aromatic clusters are located close to the active site of the thermophilic enzyme (Kannan and Vishveshwara 2000). Pino-Angeles et al. (2010) found that the betweenness centrality set conforms one patch in the active site in each monomer of the enzyme histidine decarboxylase (HDC) and that the active site particularly influences the loop region of HDC. Park and Kim (2012) encoded typical co-evolutionary information into the amino acid network using structure-based correlation mutation analysis (SCMA) to identify functional residues and functional module clusters with distinct functional roles in rhodopsin.

Network topological features can be also used to predict disease-associated single amino acid polymorphisms. Since disease-associated mutations usually occur at residues with a high centrality value and/or high degree values, neighboring residues around a mutation site can help determine whether a mutation is disease-related (Li et al. 2011b).

Intramolecular and intermolecular communications

Intramolecular and intermolecular protein communication is crucial for their biological functions. Studies based on intramolecular communications in protein amino acid networks have mainly focused on communications between protein segments, such as residues and domains. The communication paths in protein structures could be evaluated as the shortest path between protein segments. From this perspective, Atilgan et al. (2007) constructed unweighted and weighted amino acid networks based on knowledge-based potential, screened edges higher than a cutoff potential, and calculated the shortest path lengths in these reduced networks, while keeping chain connectivity. The shortest path lengths from the reduced unweighted networks, with only the strongest few non-bonded pairs, closely reproduced the strong shortest path lengths from the weighted networks. In addition, a diverse optimal path emerged for robust residue communication under different perturbations.

Recently, combined with in vitro mutational analysis, Angelova et al. (2011) used an amino acid network to analyze wild-type luteinizing hormone receptor (LHR) and two active mutants, and found that the conserved amino acids, network hubs in both inactive and active states, were responsible for communication between the extracellular and intracellular poles of the receptor. Moreover, AAN analysis successfully separated domains and identified domain interface residues in multidomain proteins (Sistla et al. 2005). Pino-Angeles et al. (2010) even found that the central patches defined by betweenness communicate significantly better with the loop region of the opposite monomer than with the rest of the histidine decarboxylase (HDC).

Vijayabaskar and Vishveshwara (2010) constructed a weighted amino acid network based on energy. They identified and elucidated the features of energetically favorable paths between functionally important residues. In addition, Vijayabaskar and Vishveshwara (2010) used their energy-based AAN to examine allosteric communications in PDZ domains. Changes in AANs upon ligand binding altered the shortest paths and closeness index of a small portion of residues, indicating that allosteric communication is anisotropic in PDZ domains.

Aminoacyl-tRNA synthetases are another excellent example of proteins used for analyzing the influence of the intra-communication of a ligand bound to a protein as there is long distance communication between the aminoacylation site and the anticodon binding site. Based on AAN analysis, several case-specific studies have been performed to investigate the intra-communication in these proteins. Combined with molecular dynamic simulations, amino acid network analysis can be used to obtain the communication paths between the anticodon region and the aminoacylation region of different liganded states of methionyl-tRNAsynthetase (MetRS) (Ghosh and Vishveshwara 2007). The communication path is strongly correlated and unique to the enzyme complex. The path lengths between the anticodon sites and aminoacylation sites are significantly short and strongly correlated when MetRS binds to both tRNA and methionyl-AMP (Vishveshwara et al. 2009; Ghosh and Vishveshwara 2007). In addition, four communication paths between the active-site region and the anticodon region of MetRS were identified (Ghosh and Vishveshwara 2007). The paths between the residues, which take part in cliques, communities, and hubs in the MetRS complex were also investigated to understand detailed inter-domain communications (Ghosh and Vishveshwara 2008).

A similar analysis was also performed on human tryptophanyl-tRNA synthetase (TrpRS), which requires two molecules of tRNA for aminoacylation (Hansia et al. 2009). The communication path between the anticodon binding domain and the catalytic domain of the TrpRS dimer is far more complex compared to monomeric

MetRS, as the communication path between these two domains is short and more frequent than intra-subunit communications and the frequency is different. In addition, the strength of the dimer interface interaction depends on the type of ligand that binds to the protein. Most recently, Szalay and Csermely (2013) defined a measure of dynamic network centrality, which they termed perturbation centrality, to correctly identify the amino acids participating in allosteric signaling in AANs and binding sites.

Besides inter-communication in protein–protein complexes, intermolecular interaction in protein–DNA complexes is another important issue to understand. To gain insight into the mechanisms of protein–DNA binding and recognition, investigations were carried out on local interactions at the interface whereas the amino acid network could provide a whole view of the interacting surface. Protein–DNA complexes have been classified on the basis of structural descriptors (Prabakaran et al. 2006). Side-chain clusters comprised of Arg, Lys, Asn, Gln, and aromatic residues from proteins were found to interact with DNA by the work of Sathyapriya and Vishveshwara (2004). They found that half of the proteins in their dataset interacted with DNA through the amino acid clusters, and the rest interacted through individual residues. The clusters contain predominantly more Arg (than Lys, Asn, and Gln) and more residues (Phe, Tyr, and Arg) have been identified as conserved and interactive with DNA. However, the interacting DNA nucleotides were not considered part of the networks; hence, Vishveshwara and the collaborators improved the model by representing the protein–DNA complex as a bipartite graph where the amino acids comprised one node set and the nucleotides constituted the other node set (Sathyapriya et al. 2008). In this study, clusters of interacting residues and hubs were identified along the protein–DNA interface and have brought out the specific features of interactions in β-sheet proteins. Moreover, they also proposed a potential classification scheme for protein–DNA complexes based on the interface clusters.

Thermal stability

Various factors have been found to contribute to the thermal stability of thermophilic proteins, such as the number of salt bridges, internal packing, and hydrogen bonding (Jaenicke and Bohm 1998; Ladenstein and Antranikian 1998). Based on AAN network analysis, Kannan and Vishveshwara (2000) found that aromatic side-chain clusters in thermophilic proteins were involved in imparting thermal stability to proteins and that the clusters were mostly located on the protein surface, usually in small sizes. Applying the same strategy, Brinda and Vishveshwara (2005) identified amino acid hubs that contribute to the stability of thermophilic proteins and predicted a few residues in thermophilic and mesophilic proteins that can be mutated to alter their thermal stability. Since many network parameters (hubs, number of edges, edge/note ratio, and largest cluster size) were higher in thermophilic proteins than their mesophilic counterparts, they suggested that network parameters were able to account for the additional stability of thermophilic proteins, although there was no single parameter that can be used to predict the stability.

Conclusions and perspectives

The amino acid network provides a holistic view of proteins and protein complexes. Residues in proteins are in contact with each other, but their positions and conformations are restricted to ensure the maintenance of protein structures. Therefore, AAN analysis allows one to focus on individual amino acid residues and their interactions, and it also provides additional insight into protein function, structure, stability, and folding.

An amino acid network may be represented through different aspects (geometry, chemistry, and energy) and at different levels, such as the single-atom (C_α, C_β, and centroid of side chain) level or all-atom level, by appropriately defining the nodes and edges. In this review, we summarized various AAN construction methods with different definitions for nodes and edges and investigated their applications in protein science. AANs, depending on their construction, can be applied to achieve different biological objectives.

AANs, where C_α, C_β, or the geometric center of the side chain represents nodes and edges, exist when the distance between the nodes are within a cutoff. They can reveal the gross information of proteins and make calculations computationally inexpensive. However, such AANs are unlikely to change when the protein experiences minor perturbations, such as the binding of a ligand; thus, AAN simplification may not be suitable for studies under this context.

AAN models, including the all atoms (Greene and Higman 2003) and only side-chain atoms (Kannan and Vishveshwara 1999; Brinda and Vishveshwara 2005) of the residue model, can capture details of residue interactions. The wealth of information about intra- and intermolecular interactions allows a better understanding of protein dynamics, such as ligand-induced conformational changes. Weighted AANs, based on interaction energy or contact energy among amino acids, consider not only the geometry but also the chemistry of the interaction, and thus have the potential to provide more insights into protein structure, stability, and function (Vijayabaskar and Vishveshwara 2010; Jiao and Chang 2011).

The advantage of AAN representation is that it takes into account the global topology, chemistry, and even energy information, rather than only the local information of proteins. In addition, the analysis of AANs is computationally much less expensive compared to molecular dynamics simulations and many other computational tools, and are also extremely efficient. Moreover, complex network parameters (e.g., betweenness, shortest path, clustering coefficient, clique, and small world) are likely to provide new insight into the structure, function, stability, folding and even the evolution of proteins. Finally, the perspective of network evolution as the rewiring of amino acid networks may also play a role in protein dynamics in the future.

Acknowledgments We gratefully acknowledge financial support from the National Natural Science Foundation of China grants (31170795, 91230117, 91029703), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20113201110015), International S&T Cooperation Program of Suzhou (SH201120) and the National High Technology Research and Development Program of China (863 program, Grant No. 2012AA02A601).

Conflict of interest The authors declare no conflicts of interest.

References

Adamcsek B, Palla G, Farkas IJ, Derenyi I, Vicsek T (2006) CFinder: locating cliques and overlapping modules in biological networks. Bioinformatics 22(8):1021–1023. doi:10.1093/bioinformatics/btl039

Aftabuddin M, Kundu S (2006) Weighted and unweighted network of amino acids within protein. Phys A 369(2):895–904. doi:10.1016/j.physa.2006.03.056

Aftabuddin M, Kundu S (2007) Hydrophobic, hydrophilic, and charged amino acid networks within protein. Biophys J 93(1):225–231. doi:10.1529/biophysj.106.098004

Alm E, Baker D (1999) Matching theory and experiment in protein folding. Curr Opin Struct Biol 9(2):189–196. doi:10.1016/S0959-440X(99)80027-X

Alves NA, Martinez AS (2007) Inferring topological features of proteins from amino acid residue networks. Phys A 375(1):336–344. doi:10.1016/j.physa.2006.09.014

Amaral LA, Scala A, Barthelemy M, Stanley HE (2000) Classes of small-world networks. Proc Natl Acad Sci USA 97(21):11149–11152. doi:10.1073/pnas.200327197

Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, Pietrokovski S (2004) Network analysis of protein structures identifies functional residues. J Mol Biol 344(4):1135–1146. doi:10.1016/j.jmb.2004.10.055

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181(4096):223–230. doi:10.1126/science.181.4096.223

Angelova K, Felline A, Lee M, Patel M, Puett D, Fanelli F (2011) Conserved amino acids participate in the structure networks deputed to intramolecular communication in the lutropin receptor. Cell Mol Life Sci 68(7):1227–1239. doi:10.1007/s00018-010-0519-z

Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. Bioinformatics 24(2):282–284. doi:10.1093/bioinformatics/btm554

Atilgan AR, Akan P, Baysal C (2004) Small-world communication of residues and significance for protein dynamics. Biophys J 86(1 Pt 1):85–91. doi:10.1016/S0006-3495(04)74086-2

Atilgan AR, Turgut D, Atilgan C (2007) Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication. Biophys J 92(9):3052–3062. doi:10.1529/biophysj.106.099440

Bagler G, Sinha S (2005) Network properties of protein structures. Phys A Stat Mech Appl 346(1–2):27–33. doi:10.1016/j.physa.2004.08.046

Bagler G, Sinha S (2007) Assortative mixing in protein contact networks and protein folding kinetics. Bioinformatics 23(14):1760–1767. doi:10.1093/bioinformatics/btm257

Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512. doi:10.1126/science.286.5439.509

Bartoli L, Fariselli P, Casadio R (2007) The effect of backbone on the small-world properties of protein contact maps. Phys Biol 4(4):L1–L5. doi:10.1088/1478-3975/4/4/L01

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242 (gkd090)

Bhattacharyya M, Bhat CR, Vishveshwara S (2013) An automated approach to network features of protein structure ensembles. Protein Sci 22(10):1399–1416. doi:10.1002/pro.2333

Bode C, Kovacs IA, Szalay MS, Palotai R, Korcsmaros T, Csermely P (2007) Network analysis of protein dynamics. FEBS Lett 581(15):2776–2782. doi:10.1016/j.febslet.2007.05.021

Bork P, Jensen LJ, von Mering C, Ramani AK, Lee I, Marcotte EM (2004) Protein interaction networks from yeast to human. Curr Opin Struct Biol 14(3):292–299. doi:10.1016/j.sbi.2004.05.003

Brinda KV, Vishveshwara S (2005) A network representation of protein structures: implications for protein stability. Biophys J 89(6):4159–4170. doi:10.1529/biophysj.105.064485

Brinda KV, Kannan N, Vishveshwara S (2002) Analysis of homodimeric protein interfaces by graph-spectral methods. Protein Eng 15(4):265–277. doi:10.1093/protein/15.4.265

Brinda KV, Surolia A, Vishveshwara S (2005) Insights into the quaternary association of proteins through structure graphs: a case study of lectins. Biochem J 391(Pt 1):1–15. doi:10.1042/BJ20050434

Chang S, Jiao X, Li CH, Gong XQ, Chen WZ, Wang CX (2008) Amino acid network and its scoring application in protein–protein docking. Biophys Chem 134(3):111–118. doi:10.1016/j.bpc.2007.12.005

Chatterjee S, Bhattacharyya M, Vishveshwara S (2012) Network properties of protein-decoy structures. J Biomol Struct Dyn 29(6):606–622. doi:10.1080/07391102.2011.672625

Chen J, Zhang D, Zhang W, Tang Y, Yan W, Guo L, Shen B (2013) Clear cell renal cell carcinoma associated microRNA expression signatures identified by an integrated bioinformatics analysis. J Transl Med 11:169. doi:10.1186/1479-5876-11-169

Chis O, Dumitru O, Concu R, Shen B (2011) Yeast network and report of new stochastic-credibility cell cycle models. Curr Bioinform 6(1):35–43

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2(10):2366–2382. doi:10.1038/nprot.2007.324

Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJ Complex Syst 1695

Csermely P (2008) Creative elements: network-based predictions of active centres in proteins and cellular and social networks. Trends Biochem Sci 33(12):569–576. doi:10.1016/j.tibs.2008.09.006

Csermely P, Sandhu KS, Hazai E, Hoksza Z, Kiss HJ, Miozzo F, Veres DV, Piazza F, Nussinov R (2012) Disordered proteins and network disorder in network descriptions of protein structure, dynamics and function: hypotheses and a comprehensive review. Curr Protein Pept Sci 13(1):19–33

Cusack MP, Thibert B, Bredesen DE, Del Rio G (2007) Efficient identification of critical residues based only on protein structure by network analysis. PLoS ONE 2(5):e421. doi:10.1371/journal.pone.0000421

da Silveira CH, Pires DE, Minardi RC, Ribeiro C, Veloso CJ, Lopes JC, Meira W Jr, Neshich G, Ramos CH, Habesch R, Santoro MM (2009) Protein cutoff scanning: a comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. Proteins 74(3):727–743. doi:10.1002/prot.22187

Deb D, Vishveshwara S (2009) Understanding protein structure from a percolation perspective. Biophys J 97(6):1787–1794. doi:10.1016/j.bpj.2009.07.016

del Sol A, O'Meara P (2005) Small-world network approach to identify key residues in protein–protein interaction. Proteins 58(3):672–682. doi:10.1002/prot.20348

del Sol A, Fujihashi H, O'Meara P (2005) Topology of small-world networks of protein–protein complex structures. Bioinformatics 21(8):1311–1315. doi:10.1093/bioinformatics/bti167

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. Protein Sci 15(9):2120–2128. doi:10.1110/ps.062249106

Derenyi I, Palla G, Vicsek T (2005) Clique percolation in random networks. Phys Rev Lett 94(16):160202

Di Paola L, De Ruvo M, Paci P, Santoni D, Giuliani A (2013) Protein contact networks: an emerging paradigm in chemistry. Chem Rev 113(3):1598–1613. doi:10.1021/cr3002356

Dokholyan NV, Li L, Ding F, Shakhnovich EI (2002) Topological determinants of protein folding. Proc Natl Acad Sci USA 99(13):8637–8641. doi:10.1073/pnas.122076099

Doncheva NT, Klein K, Domingues FS, Albrecht M (2011) Analyzing and visualizing residue networks of protein structures. Trends Biochem Sci 36(4):179–182. doi:10.1016/j.tibs.2011.01.002

Doncheva NT, Assenov Y, Domingues FS, Albrecht M (2012) Topological analysis and interactive visualization of biological networks and protein structures. Nat Protoc 7(4):670–685. doi:10.1038/nprot.2012.004

Eargle J, LutheySchulten Z (2012) NetworkView: 3D display and analysis of protein.RNA interaction networks. Bioinformatics 28(22):3000–3001. doi:10.1093/bioinformatics/bts546

Estrada E (2010) Universality in protein residue networks. Biophys J 98(5):890–900. doi:10.1016/j.bpj.2009.11.017

Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the Internet topology. SIGCOMM Comput Commun Rev 29(4):251–262. doi:10.1145/316194.316229

Fanelli F, Seeber M (2010) Structural insights into retinitis pigmentosa from unfolding simulations of rhodopsin mutants. FASEB J 24(9):3196–3209. doi:10.1096/fj.09-151084

Gaci O, Balev S (2009) Node degree distribution in amino acid interaction networks. In: Balev S (ed) pp 107–112. doi:10.1109/BIBMW.2009.5332113

Ghosh A, Vishveshwara S (2007) A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. Proc Natl Acad Sci USA 104(40):15711–15716. doi:10.1073/pnas.0704459104

Ghosh A, Vishveshwara S (2008) Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes. Biochemistry 47(44):11398–11407. doi:10.1021/bi8007559

Ghosh A, Brinda KV, Vishveshwara S (2007) Dynamics of lysozyme structure network: probing the process of unfolding. Biophys J 92(7):2523–2535. doi:10.1529/biophysj.106.099903

Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99(12):7821–7826. doi:10.1073/pnas.122653799

Go N (1983) Theoretical studies of protein folding. Annu Rev Biophys Bioeng 12:183–210. doi:10.1146/annurev.bb.12.060183.001151

Greene LH (2012) Protein structure networks. Brief Funct Genomics 11(6):469–478. doi:10.1093/bfgp/els039

Greene LH, Higman VA (2003) Uncovering network systems within protein structures. J Mol Biol 334(4):781–791. doi:10.1016/j.jmb.2003.08.061

Hagberg A, Swart P, S Chult D (2008) Exploring network structure, dynamics, and function using networkx. In: Varoquaux G, Vaught T, Millman J (eds) Proceedings of the 7th python in science conference (SciPy2008), pp 11–15

Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2008) statnet: software tools for the representation, visualization, analysis and simulation of network data. J Stat Softw 24(1):1548–7660

Hansia P, Ghosh A, Vishveshwara S (2009) Ligand dependent intra and inter subunit communication in human tryptophanyl tRNA synthetase as deduced from the dynamics of structure networks. Mol BioSyst 5(12):1860–1872. doi:10.1039/b903807h

Hegedus T, Gyimesi G, Gaspar ME, Szalay KZ, Gangal R, Csermely P (2013) Potential application of network descriptions for understanding conformational changes and protonation states of ABC transporters. Curr Pharm Des 19(23):4155–4172

Hinds DA, Levitt M (1992) A lattice model for protein structure prediction at low resolution. Proc Natl Acad Sci USA 89(7):2536–2540

Hu G, Zhou J, Yan W, Chen J, Shen B (2013) The topology and dynamics of protein complexes: insights from intra-molecular network theory. Curr Protein Pept Sci 14(2):121–132

Hu G, Yan W, Zhou J, Shen B (2014) Residue interaction network analysis of Dronpa and a DNA clamp. J Theor Biol. doi:10.1016/j.jtbi.2014.01.023

Huang J, Kawashima S, Kanehisa M (2007) New amino acid indices based on residue network topology. Genome Inform 18:152–161 (9781860949920_0015 [pii])

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98(8):4569–4574. doi:10.1073/pnas.061034498

Jaenicke R, Bohm G (1998) The stability of proteins in extreme environments. Curr Opin Struct Biol 8(6):738–748. doi:10.1016/S0959-440X(98)80094-810.1016/S0959-440X(98)80094-8

Jiao X, Chang S (2011) Scoring function based on weighted residue network. Int J Mol Sci 12(12):8773–8786. doi:10.3390/ijms12128773

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 Stat Nonlin Soft Matter Phys 75(5 Pt 1):051903. doi:10.1103/PhysRevE.75.051903

Jung J, Lee J, Moon HT (2005) Topological determinants of protein unfolding rates. Proteins 58(2):389–395. doi:10.1002/prot.20324

Kalapala V, Sanwalani V, Clauset A, Moore C (2006) Scale invariance in road networks. Phys Rev E 73(2):026130

Kannan N, Vishveshwara S (1999) Identification of side-chain clusters in protein structures by a graph spectral method. J Mol Biol 292(2):441–464. doi:10.1006/jmbi.1999.3058

Kannan N, Vishveshwara S (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins. Protein Eng 13(11):753–761. doi:10.1093/protein/13.11.753

Kannan N, Chander P, Ghosh P, Vishveshwara S, Chatterji D (2001) Stabilizing interactions in the dimer interface of alpha-subunit in *Escherichia coli* RNA polymerase: a graph spectral and point mutation study. Protein Sci 10(1):46–54. doi:10.1110/ps.26201

Küçükural A, Sezerman OU, Ercil A (2008) Discrimination of native folds using network properties of protein structures. In: Paper presented at the APBC

Ladenstein R, Antranikian G (1998) Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. Adv Biochem Eng Biotechnol 61:37–85. doi:10.1007/BFb0102289

Li Y, Li G, Wen Z, Yin H, Hu M, Xiao J, Li M (2011a) Novel feature for catalytic protein residues reflecting interactions with other residues. PLoS ONE 6(3):e16932. doi:10.1371/journal.pone.0016932

Li Y, Wen Z, Xiao J, Yin H, Yu L, Yang L, Li M (2011b) Predicting disease-associated substitution of a single amino acid by analyzing residue interactions. BMC Bioinform 12:14. doi:10.1186/1471-2105-12-14

Ma H, Zeng AP (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19(2):270–277. doi:10.1093/bioinformatics/19.2.270

Madadhain J, Fisher D, Smyth P, White S, Boey YB (2005) Analysis and visualization of network data using JUNG. J Stat Softw 10:1–35

Martin AJ, Vidotto M, Boscariol F, Di Domenico T, Walsh I, Tosatto SC (2011) RING: networking interacting residues, evolutionary information and energetics in protein structures. Bioinformatics 27(14):2003–2005. doi:10.1093/bioinformatics/btr191

Matsumura M, Signor G, Matthews BW (1989) Substantial increase of protein stability by multiple disulphide bonds. Nature 342(6247):291–293. doi:10.1038/342291a0

Morita H, Takano M (2009) Residue network in protein native structure belongs to the universality class of a three-dimensional critical percolation cluster. Phys Rev E Stat Nonlin Soft Matter Phys 79(2 Pt 1):020901. doi:10.1103/PhysRevE.79.020901

Morris JH, Meng EC, Ferrin TE (2010) Computational tools for the interactive exploration of proteomic and structural data. Mol Cell Proteomics 9(8):1703–1715. doi:10.1074/mcp.R000007-MCP201

Munz M, Biggin PC (2012) JGromacs: a Java package for analyzing protein simulations. J Chem Inf Model 52(1):255–259. doi:10.1021/ci200289s

Muppirala UK, Li Z (2006) A simple approach for protein structure discrimination based on the network pattern of conserved hydrophobic residues. Protein Eng Des Sel 19(6):265–275. doi:10.1093/protein/gzl009

Newman ME (2002) Assortative mixing in networks. Phys Rev Lett 89(20):208701. doi:10.1103/PhysRevLett.89.208701

Newman ME (2003) Mixing patterns in networks. Phys Rev E Stat Nonlin Soft Matter Phys 67(2 Pt 2):026126

Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. Soc Netw 32(3):245–251. doi:10.1016/j.socnet.2010.03.006

Pabuwal V, Li Z (2008) Network pattern of residue packing in helical membrane proteins and its application in membrane protein structure prediction. Protein Eng Des Sel 21(1):55–64. doi:10.1093/protein/gzm059

Park K, Kim D (2011) Modeling allosteric signal propagation using protein structure networks. BMC Bioinform 12(Suppl 1):S23. doi:10.1186/1471-2105-12-S1-S23

Park K, Kim D (2012) Structure-based rebuilding of coevolutionary information reveals functional modules in rhodopsin structure. Biochim Biophys Acta. doi:10.1016/j.bbapap.2012.05.015

Pasi M, Tiberti M, Arrigoni A, Papaleo E (2012) xPyder: a PyMOL plugin to analyze coupled residues and their networks in protein structures. J Chem Inf Model 52(7):1865–1874. doi:10.1021/ci300213c

Pastor-Satorras R, Vazquez A, Vespignani A (2001) Dynamical and correlation properties of the internet. Phys Rev Lett 87(25):258701. doi:10.1103/PhysRevLett.87.258701

Pino-Angeles A, Morreale A, Negri A, Sanchez-Jimenez F, Moya-Garcia AA (2010) Substrate uptake and protein stability relationship in mammalian histidine decarboxylase. Proteins 78(1):154–161. doi:10.1002/prot.22587

Prabakaran P, Siebers JG, Ahmad S, Gromiha MM, Singarayan MG, Sarai A (2006) Classification of protein–DNA complexes based on structural descriptors. Structure 14(9):1355–1367. doi:10.1016/j.str.2006.06.018

Reichmann D, Rahat O, Albeck S, Meged R, Dym O, Schreiber G (2005) The modular architecture of protein–protein binding interfaces. Proc Natl Acad Sci USA 102(1):57–62. doi:10.1073/pnas.0407280102

Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T (2012) A travel guide to Cytoscape plugins. Nat Methods 9(11):1069–1076. doi:10.1038/nmeth.2212

Sathyapriya R, Vishveshwara S (2004) Interaction of DNA with clusters of amino acids in proteins. Nucleic Acids Res 32(14):4109–4118. doi:10.1093/nar/gkh733

Sathyapriya R, Vijayabaskar MS, Vishveshwara S (2008) Insights into protein–DNA interactions through structure network analysis. PLoS Comput Biol 4(9):e1000170. doi:10.1371/journal.pcbi.1000170

Seeber M, Felline A, Raimondi F, Muff S, Friedman R, Rao F, Caflisch A, Fanelli F (2011) Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. J Comput Chem 32(6):1183–1194. doi:10.1002/jcc.21688

Sethi A, Eargle J, Black AA, Luthey-Schulten Z (2009) Dynamical networks in tRNA: protein complexes. Proc Natl Acad Sci USA 106(16):6620–6625. doi:10.1073/pnas.0810961106

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504. doi:10.1101/gr.1239303

Shoichet BK, Baase WA, Kuroki R, Matthews BW (1995) A relationship between protein stability and protein function. Proc Natl Acad Sci USA 92(2):452–456. doi:10.1073/pnas.92.2.452

Sistla RK, Brinda KV, Vishveshwara S (2005) Identification of domains and domain interface residues in multidomain proteins from graph spectral method. Proteins 59(3):616–626. doi:10.1002/prot.20444

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27(3):431–432. doi:10.1093/bioinformatics/btq675

Sporns O, Chialvo DR, Kaiser M, Hilgetag CC (2004) Organization, development and function of complex brain networks. Trends Cogn Sci 8(9):418–425. doi:10.1016/j.tics.2004.07.008

Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302(5643):249–255. doi:10.1126/science.1087447

Susan K (2010) Static and dynamic characteristics of protein contact networks. Proteins 40. doi:10.1007/s12064-011-0135-y

Susan K (2011) Towards an integrated understanding of the structural characteristics of protein residue networks. Theory Biosci. doi:10.1007/s12064-011-0135-y

Szalay KZ, Csermely P (2013) Perturbation centrality and turbine: a novel centrality measure obtained using a versatile network dynamics tool. PLoS ONE 8(10):e78059. doi:10.1371/journal.pone.0078059

Tang B, Wu X, Tan G, Chen SS, Jing Q, Shen B (2010) Computational inference and analysis of genetic regulatory networks via a supervised combinatorial-optimization pattern. BMC Syst Biol 4(Suppl 2):S3. doi:10.1186/1752-0509-4-s2-s3

Tuncbag N, Salman FS, Keskin O, Gursoy A (2010) Analysis and network representation of hotspots in protein interfaces using minimum cut trees. Proteins 78(10):2283–2294. doi:10.1002/prot.22741

van Noort V, Snel B, Huynen MA (2004) The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. EMBO Rep 5(3):280–284. doi:10.1038/sj.embor.7400090

Vanwart AT, Eargle J, Luthey-Schulten Z, Amaro RE (2012) Exploring residue component contributions to dynamical network models of allostery. J Chem Theory Comput 8(8):2949–2961. doi:10.1021/ct300377a

Vassura M, Margara L, Fariselli P, Casadio R (2009) A graph theoretic approach to protein structure selection. Artif Intell Med 45(2–3):229–237. doi:10.1016/j.artmed.2008.07.016

Veloso CJ, Silveira CH, Melo RC, Ribeiro C, Lopes JC, Santoro MM, Meira W Jr (2007) On the characterization of energy networks of proteins. Genet Mol Res 6(4):799–820

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(6820):641–645. doi:10.1038/35054591

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 Stat Nonlin Soft Matter Phys 65(6 Pt 1):061910. doi:10.1103/PhysRevE.65.061910

Vijayabaskar MS, Vishveshwara S (2010) Interaction energy based protein structure networks. Biophys J 99(11):3704–3715. doi:10.1016/j.bpj.2010.08.079

Vijayabaskar MS, Niranjan V, Vishveshwara S (2011) GraProStr—graphs of protein structures: a tool for constructing the graphs and generating graph parameters for protein structures. Open Bioinform J 5(2011):6. doi:10.2174/1875036201105010053

Vishveshwara S, Ghosh A, Hansia P (2009) Intra and inter-molecular communications through protein structure network. Curr Protein Pept Sci 10(2):146–160. doi:10.2174/138920309787847590

Wasserman S, Faust K (1994) Social network analysis: methods and applications (structural analysis in the social sciences). Cambridge University Press, Cambridge

Watts DJ (1999) Small worlds: the dynamics of networks between order and randomness. Princeton University Press, Princeton

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442. doi:10.1038/30918

Yan W, Zhu H, Yang Y, Chen J, Zhang Y, Shen B (2010) Effects of time point measurement on the reconstruction of gene regulatory networks. Molecules (Basel, Switzerland) 15(8):5354–5368. doi:10.3390/molecules15085354

Yu T, Zou X, Huang SY, Zou XW (2009) Cutoff variation induces different topological properties: a new discovery of amino acid network within protein. J Theor Biol 256(3):408–413. doi:10.1016/j.jtbi.2008.09.042

Zhang X, Perica T, Teichmann SA (2013) Evolution of protein structures and interactions from the perspective of residue contact networks. Curr Opin Struct Biol 23(6):954–963. doi:10.1016/j.sbi.2013.07.004

Zhou J, Yan W, Hu G, Shen B (2013) SVR_CAF: an integrated score function for detecting native protein structures among decoys. Proteins. doi:10.1002/prot.24421

Zhou J, Yan W, Hu G, Shen B (2014) Amino acid network for the discrimination of native protein structures from decoys. Curr Protein Pept Sci 15 (in press)