Hydrophobic, Hydrophilic, and Charged Amino Acid Networks within Protein

Md. Aftabuddin and S. Kundu
Department of Biophysics, Molecular Biology & Genetics, University of Calcutta, Kolkata 700009, West Bengal, India

ABSTRACT The native three-dimensional structure of a single protein is determined by the physicochemical nature of its constituent amino acids. The 20 different types of amino acids, depending on their physicochemical properties, can be grouped into three major classes: hydrophobic, hydrophilic, and charged. The anatomy of the weighted and unweighted networks of hydrophobic, hydrophilic, and charged residues separately for a large number of proteins were studied. Results showed that the average degree of the hydrophobic networks has a significantly larger value than that of hydrophilic and charged networks. The average degree of the hydrophilic networks is slightly higher than that of the charged networks. The average strength of the nodes of hydrophobic networks is nearly equal to that of the charged network, whereas that of hydrophilic networks has a smaller value than that of hydrophobic and charged networks. The average strength for each of the three types of networks varies with its degree. The average strength of a node in a charged network increases more sharply than that of the hydrophobic and hydrophilic networks. Each of the three types of networks exhibits the "small-world" property. Our results further indicate that the all-amino-acids networks and hydrophobic networks are of assortative type. Although most of the hydrophilic and charged networks are of the assortative type, few others have the characteristics of disassortative mixing of the nodes. We have further observed that all-amino-acids networks and hydrophobic networks bear the signature of hierarchy, whereas the hydrophilic and charged networks do not have any hierarchical signature.

INTRODUCTION

Network analysis is increasingly being recognized as a powerful tool to study complex systems. It helps us to understand the interaction among individual components and hence to characterize the whole system. Several researchers have worked to shed light on the topology, growth, and dynamics of different kinds of networks including the world-wide web, food webs, gene coexpression networks, metabolic networks, and protein-protein interaction networks, etc. (1–9).

Efforts have also been made to transform a protein structure into a network where amino acids are nodes and their interactions are edges (10–17). However, these protein structure networks have been constructed with varying definition of nodes and edges. This network approach has been used in a number of studies, such as protein structural flexibility, prediction of key residues in protein folding, identification of functional residues, and residue contribution to the protein-protein binding free energy in given complexes (10–14). Several groups have also studied the protein network to understand its topology, small world properties, and behaviors of long-range and short-range interactions of the amino acid nodes, etc. (15–17).

In almost all of the previous studies on the protein structure networks, the protein has been considered as an unweighted network of amino acids. Very recently, we have considered it as a weighted network (18). This investigation has focused on degree and strength distribution, signature of hierarchy, and assortative-type mixing behavior of the amino acid nodes.

A protein molecule is a polymer of different amino acids joined by peptide bonds. These 20 different amino acids have different side chains and hence different physicochemical properties. When a protein folds in its native conformation, its native three-dimensional structure is determined by the physicochemical nature of its constituent amino acids. Depending on the physicochemical properties, the different amino acids fall into three major classes: hydrophobic, hydrophilic, and charged residues. In this context, it would be interesting to study the network structures of hydrophobic, hydrophilic, and charged residues separately. We have also recently studied the hydrophobic and hydrophilic networks (19). Our analysis has mainly focused on the degree, the degree distribution, and small world properties. We have found that the average degree of a hydrophobic node is larger than that of a hydrophilic node. We have also observed the existence of small world properties in both cases. The hydrophobic and hydrophilic networks we have studied previously (19) are unweighted networks, but the study presented here considers both the weighted and unweighted networks of hydrophobic, hydrophilic, and charged residues' networks. We have analyzed these networks to focus on their topology including degree, strength, strength-degree relationships, clustering coefficients, shortest path length, existence of small world property and hierarchical signature, if any, and mixing behavior of the nodes. In summary, in our investigations, we have studied the anatomy of hydrophobic, hydrophilic, and charged residues' networks and have also

Submitted September 25, 2006, and accepted for publication December 1, 2006.

Address reprint requests to Sudip Kundu, Dept. of Biophysics, Molecular Biology & Genetics, University of Calcutta, 92 APC Road, Kolkata 700009, West Bengal, India. E-mail: skbmbg@caluniv.ac.in.

Editor: Eugene Shakhnovich.

0006-3495/07/07/225/07 $2.00

doi: 10.1529/biophysj.106.098004

performed a comparative study among them as well as with all-amino-acids networks.

METHODS

Construction of hydrophobic, hydrophilic, charged, and all-amino-acids networks

Primary structure of a protein is a linear arrangement of different types of amino acids in one-dimensional space where any amino acid is connected with its nearest neighbors through peptide bonds. But when a protein folds in its native conformation, distant amino acids in the one-dimensional chain may also come close to each other in three-dimensional space, and hence, different noncovalent interactions are possible among them depending on their orientations in three-dimensional space. Moreover, each of the 20 amino acids has a different side chain and different physicochemical properties. These different 20 amino acid residues have been grouped into three major classes: hydrophobic (F, M, W, I, V, L, P, A), hydrophilic (N, C, Q, G, S, T, Y), and charged (R, D, E, H, K). Here we are interested in studying the hydrophobic, hydrophilic, and charged networks within proteins.

Any network has two basic components: nodes and edges. Only the hydrophobic residues are considered as nodes of a hydrophobic network, whereas hydrophilic and charged residues are considered as the nodes of hydrophilic and charged networks, respectively. If any two atoms from two different amino acids (nodes) are within a cutoff distance (5 Å), the amino acids are considered to be connected or linked. The cutoff distance is within the higher cutoff distance of London-van der Waals forces (20). Further, in our calculations we have not considered the interaction of any of the backbone atoms; we have included only the interactions of the side-chain atoms.

Thus, in a hydrophobic network, hydrophobic residues are nodes, and the possible links among them are edges. The same logic is followed to construct the other networks.

Because we have also compared the network parameters of hydrophobic, hydrophilic, and charged networks with those of an all-amino-acids network within protein, we have also constructed the networks taking into account all amino acids without any classifications. Thus, we have obtained the unweighted networks of all-amino-acids (AN), hydrophobic (BN), hydrophilic (IN), and charged (CN) amino acid network types.

Next, we discuss the basis of transforming the protein structure into a weighted network. When we consider a protein's three-dimensional structure, several atoms of any amino acid in a protein may be within the cutoff distance of several atoms of another amino acid. This results in possible multiple links between any two amino acids. These multiple links are the basis of the weight of the connectivity, which may vary for different combinations of amino acids as well as for different orientations of them in three-dimensional conformational space. The intensity w_{ij} of the interaction between two amino acids i and j is defined as the number of possible links between the ith and the jth amino acids. Considering the intensity of interaction between any two amino acids, we have constructed the weighted BN, IN, CN, and AN.

We have collected a total of 161 protein structures from a protein crystal structure data bank (21) with the following criteria:

1. Maximum percentage identity: 25.
2. Resolution: 0.0–2.0.
3. Maximum R-value: 0.2.
4. Sequence length: 500–10,000.
5. Non-x-ray entries: excluded.
6. CA-only entries: excluded.
7. CULLPDB by chain.

In some of the crystal structures, the atomic coordinates of some of the residues are missing. We have not considered those structures because they may give erroneous values of different network parameters (degree, clustering coefficient, etc.). A final set of 85 crystal structures was taken for the calculation and analysis of network properties. We have generated the BN, IN, CN, AN of each of the 85 proteins using the three-dimensional atomic coordinates of the protein structures. Although ANs for each of the proteins form a single cluster, the BN, IN, and CN, in general, have more than one subnetwork. The number of nodes of these subnetworks varies over a wider range. The subnetworks having at least 30 nodes have been collected and analyzed.

Network parameters

Each of the networks has been represented as an adjacency matrix (A). Any element of adjacency matrix (A), a_{ij}, is given as

$$a_{ij} = 1, \text{ if } i \neq j \text{ and } i \text{ and } j \text{ nodes are connected by an edge.}$$
$$0 \text{ if } i \neq j \text{ and } i \text{ and } j \text{ nodes are not connected.}$$
$$0 \text{ if } i = j.$$

The degree of any node i is represented by $k_i = \Sigma_j a_{ij}$.

The number of possible interactions between any two amino acids may vary depending on their three-dimensional orientations and the number of atoms in their side chains. If w_{ij} is the number of possible interactions between any ith and jth amino acids, then the strength (s_i) of a node i is given by $s_i = \Sigma_j a_{ij} w_{ij}$.

This parameter represents the number of connectivities of any two amino acids and is thus a characteristic of a weighted network. It should be clearly mentioned that the weighted network analysis depends on 1), the number of possible interactions between amino acid residues and also on 2), the energy of interactions between them. Because the total energy of interactions again depends on the total number of interactions between residues, we, for the sake of simplicity of analysis, have considered only the number of interactions between residues.

We have determined the characteristic path length (L) and the clustering coefficient (C) of each network. The characteristic path length L of a network is the path length between two nodes averaged over all pairs of nodes. The clustering coefficient C_i is a measure of local cohesiveness. Traditionally the clustering coefficient C_i of a node i is the ratio between the total number (e_i) of the edges actually connecting its nearest neighbors to the ith node and the total number of all possible edges between all these nearest neighbors ($k_i(k_i - 1)/2$ if the ith vertex has k_i neighbors) and is given by $C_i = 2e_i/k_i(k_i - 1)$. The clustering coefficient of a network is the average of all of its individual C_is. For a random network having N nodes with average degree $\langle k \rangle$, the characteristic path length (L_r), and the clustering coefficient (C_r) have been calculated using the expressions $L_r \approx lnN/ln\langle k \rangle$ and $C_r \approx \langle k \rangle/N$ given by Watts and Strogatz (3). To ascertain if there is any small world property in a network, we have followed Watts and Strogatz's method (3). According to them, a network has the small world property if $C \gg C_r$ and $L \geq L_r$. Combining the topological information with the weight distribution of the network, Barrat et al. (22) have introduced an analogous parameter to C and that is known as weighted clustering coefficient, C_i^w. It takes into account the importance of the clustered structure on the basis of amount of interaction intensity (number of possible interactions between amino acids) actually found on the local triplets and is given by $C_i^w = [1/s_i(k_i - 1)]\sum_{j,h}(w_{ij} + w_{ih})a_{ij}a_{ih}a_{jh}/2$.

To study the tendency for nodes in networks to be connected to other nodes that are like (or unlike) them, we have calculated the Pearson correlation coefficient of the degrees at either ends of an edge. For our undirected unweighted protein network, its value has been calculated using the expression suggested by Newman (23) and is given as:

$$r = (M^{-1}\textstyle\sum_i j_i k_i - [M^{-1}\sum_i 0.5(j_i + k_i)]^2) \div (M^{-1}\sum_i 0.5(j_i^2 + k_i^2) - [M^{-1}\sum_i 0.5(j_i + k_i)]^2).$$

Here j_i and k_i are the degrees of the vertices at the ends of the ith edge, with $i = 1, ..M$. The networks having positive r values are assortative in nature.

RESULTS AND DISCUSSION

We have constructed the hydrophobic, hydrophilic, and charged residues' networks for each of the 85 proteins. It has been observed that all the hydrophobic residues of the BN for each of all the proteins do not form a single cluster. In general, they form one (in some cases more than one) giant cluster associated with small subclusters and isolated nodes. The same feature has also been observed for both IN and CN. Thus, all of the above three types of networks are sparse networks. On the other hand, when we consider an AN within a protein, the nodes (amino acids) do form a single cluster. We have also observed that in each of the 85 proteins, the total number of subclusters and isolated nodes of the BN is smaller than that of the CN and IN. In only one protein the number of subclusters and isolated nodes of the IN is higher than that of the BN. However, CNs of 56 proteins (of the 85 proteins) show higher numbers of subclusters and isolated nodes than the respective INs. Thus, we may say that INs and CNs within a protein are more sparse in nature than BNs.

To calculate and analyze different network properties, we have selected those subclusters that have at least 30 nodes. Thus, we have finally obtained 92 hydrophobic, 99 hydrophilic, and 69 charged subclusters with the criteria of having at least 30 nodes. We have further observed that the average number of nodes (amino acids) of hydrophobic subclusters is nearly double and quadruple, respectively, those of hydrophilic and charged subclusters, as is evident from Table 1.

It should be clearly mentioned that all the network parameters we have further calculated and analyzed are the result of our finally selected different subclusters or ANs. In the remainder of this article, we refer to these subclusters as networks.

Average degree of the networks

For each of the four types of networks (BN, IN, CN, and AN) we have calculated the average degree $\langle k \rangle$. The values are listed in Table 1. We find that the average degree of BNs ($\langle k^{b} \rangle$), INs ($\langle k^{i} \rangle$), CNs ($\langle k^{c} \rangle$), and ANs ($\langle k^{a} \rangle$) varies from 2.97 to 5.47, from 2.22 to 3.81, from 2.06 to 4.18, and from 6.75 to 10.09, respectively. The average of the $\langle k^{b} \rangle$ values for all of the BNs, $\langle k^{b}_{av} \rangle$, was found to be 4.84 with a standard deviation 0.35. The average of the $\langle k^{i} \rangle$ values for all of the INs, $\langle k^{i}_{av} \rangle$, was found to be 2.97 with a standard deviation 0.29. For the CNs, the average ($\langle k^{c}_{av} \rangle$) was found to be 2.72 with a standard deviation 0.33.

It has been observed that the average of the $\langle k^{a} \rangle$ of all of the ANs, $\langle k^{a}_{av} \rangle$ shows expected higher values than that of BN, IN, and CN. Our results also clearly show that $\langle k^{b}_{av} \rangle > \langle k^{i}_{av} \rangle \approx \langle k^{c}_{av} \rangle$. The Mann-Whitney *U*-test shows that these three populations are significantly different (level of significance is 0.001). To verify whether the observed trend is a result of the network size or is purely the characteristic of the nature of the nodes of the network, we have compared the $\langle k \rangle$ values of different networks with similar sizes (i.e., nearly the same number of nodes). The result confirms the trend previously described. Hence, our observation ($\langle k^{b}_{av} \rangle > \langle k^{i}_{av} \rangle \approx \langle k^{c}_{av} \rangle$) is clearly an inherent nature of the network. We have also observed that within the same populations the value of the average degree does not depend on the network size (i.e., on the number of amino acids of the protein).

Average strength of the networks

Next we have studied the strength of the nodes within different types of weighted networks. The average strength of the BNs ($\langle s^{b} \rangle$) varies from 17.28 to 35.21, whereas that of the INs ($\langle s^{i} \rangle$) and CNs ($\langle s^{c} \rangle$) varies from 6.76 to 27.74 and from 14.71 to 50.63, respectively. On the other hand, the average strength of AN ($\langle s^{a} \rangle$) varies from 34.85 to 83.86. The average of $\langle s^{a} \rangle$ for all of the ANs was found to be 41.94 with a standard deviation 5.61. The average of the $\langle s^{b} \rangle$ values for all of the BNs, $\langle s^{b}_{av} \rangle$, is nearly equal to that ($\langle s^{c}_{av} \rangle$) of the CNs, whereas that of INs has smaller value than those of BNs and CNs.

TABLE 1 Different network properties

Network type	$\langle N_r \rangle$	$\langle k \rangle$	$\langle s \rangle$	$\langle L \rangle$	$\langle C \rangle$	$\langle C^{w} \rangle$	$\langle r \rangle$*	$\langle p \rangle$	$\langle q \rangle$	$\langle \beta \rangle^{\dagger}$	$\langle \beta^{w} \rangle^{\dagger}$
BN	221.22 ± 73.29	4.84 ± 0.35	23.72 ± 2.74	7.45 ± 1.59	0.46 ± 0.02	0.23 ± 0.01	0.30 ± 0.07	20.76 ± 6.71	2.18 ± 0.34	0.254 ± 0.125	0.231 ± 0.124
IN	92.78 ± 56.08	2.97 ± 0.29	14.59 ± 2.95	7.96 ± 2.38	0.49 ± 0.05	0.25 ± 0.02	0.20 ± 0.10	14.98 ± 8.22	1.94 ± 0.40		
CN	45.97 ± 18.60	2.72 ± 0.33	22.46 ± 5.59	6.73 ± 1.74	0.52 ± 0.06	0.27 ± 0.03	0.19 ± 0.12	8.67 ± 3.12	1.74 ± 0.32		
AN	612.15 ± 134.82	7.58 ± 0.38	41.94 ± 5.61	6.61 ± 0.88	0.37 ± 0.07	0.19 ± 0.01	0.30 ± 0.04	29.71 ± 6.46	2.09 ± 0.22	0.208 ± 0.110	0.166 ± 0.106

Average number of nodes ($\langle N_r \rangle$), average degree ($\langle k \rangle$), average strength ($\langle s \rangle$), average characteristic path length ($\langle L \rangle$), average clustering coefficients of unweighted ($\langle C \rangle$) and weighted ($\langle C^{w} \rangle$) networks, Pearson correlation coefficient ($\langle r \rangle$), the average ratios ($\langle p \rangle$ and $\langle q \rangle$), average scaling coefficients of unweighted ($\langle \beta \rangle$) and weighted ($\langle \beta^{w} \rangle$) networks of hydrophobic (BN), hydrophilic (IN), charged (CN), and all-amino-acids (AN) networks.

*Data shown only for positive $\langle r \rangle$.

†Because there is no clear functional relation in the case of IN and CN, the values of scaling coefficients are not listed.

It should be mentioned that the sizes of most of the hydrophobic and charged residues are larger than those of the hydrophilic ones. We have also observed that there is a relation of the volume of an amino acid with its strength. This may be one of the causes of the higher values of the strengths of the BNs and CNs.

Strength-degree relations

To understand the relation between the strength of a node and its degree, k, we have further studied the average strength $\langle s^{b}\rangle(k)$, $\langle s^{i}\rangle(k)$, and $\langle s^{c}\rangle(k)$ as a function of k. The result is shown in Fig. 1. We have observed that the strength of a vertex changes with its degree, k. The average strength for all of the hydrophobic networks varies linearly with its degree, k. On the other hand, the average strength of CNs and INs increases linearly with k for smaller values of k but sharply for higher values. It has been further noted that the slope of the best-fit line is different for different types of networks. The average strength of a node in CNs increases more sharply than that of the BN and IN, as is evident from Fig. 1.

Small world property

To examine whether the networks have the ''small world'' property or not, we have calculated the average clustering coefficient $\langle C\rangle$ and the characteristic path length $\langle L\rangle$ for each of the networks and their respective values ($\langle C_r\rangle$ and $\langle L_r\rangle$) for the random network having the same N (number of nodes) and $\langle k\rangle$. The averages of the $\langle C\rangle$ and $\langle L\rangle$ values for all of the hydrophobic networks are given in Table 1. Those of IN and CN are also presented in Table 1. The ratios $[p = \langle C\rangle/\langle C_r\rangle]$ of average clustering coefficients of BN to that of a classical random graph vary from 3.55 to 40.37. The ratios for IN and CN vary from 5.14 to 42.55 and from 3.69 to 24.32, respectively. On the other hand, it has been observed that the characteristic path length is of the same order as that of a corresponding random graph, as is evident from $q = \langle L\rangle/\langle L_r\rangle$ values listed in Table 1. Although the ratios (p) for networks under study are not of the order of 10^2–10^4 as observed in the cases of scientific collaboration networks and networks of film actors, there are several other networks where p may have smaller values (2,6,19,24,25). For example, the ratio (p) for metabolic network, protein-protein interaction network, food webs, and network of *C. elegans* have values 5.0, 4.4, 12.0, and 5.6, respectively. Even a recent study on amino acid networks within proteins reported that the ratio (p) varies from 4.61 to 25.20 depending on the size of the network (18). Thus, we may conclude that each of the three different types of networks (BN, IN, and CN) has small world property. We have also examined the ANs of the same proteins. We find that the ANs also have the small world property as is evident from the p and q values listed in Table 1.

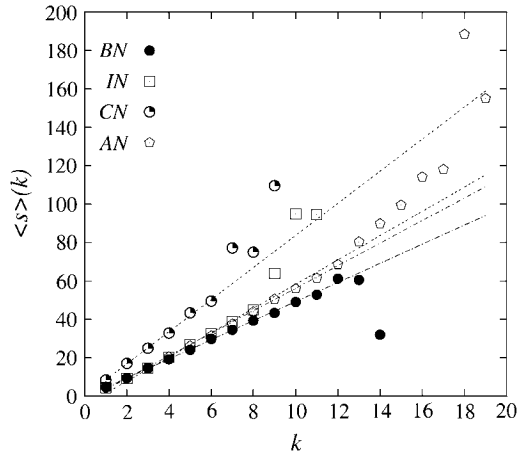

FIGURE 1 Average strength $\langle s\rangle(k)$ as a function of degree k of hydrophobic (BN), hydrophilic (IN), charged (CN), and all-amino-acids (AN) networks.

We have further studied the dependencies of p and q on N, number of nodes. The results are shown in Fig. 2. We find that both the ratios p and q vary with N, but with different relations. The ratio (p) of clustering coefficients varies linearly with N, whereas the ratio (q) of characteristic path lengths varies logarithmically with N. It should be mentioned that the p values of ANs vary from 23.10 to 60.66. The higher p values of ANs obtained in our study than those reported by Aftabuddin and Kundu (18) may be because of the larger size of networks.

Mixing behavior of the nodes

We have also calculated Pearson correlation coefficients (r) for each of the networks. Almost all the BNs (except one) have positive r^{b} values, which vary from 0.02 to 0.43 with an average 0.30. Although most of the hydrophobic networks have positive r values, both the INs and the CNs have both positive and negative r values. The positive r value of a network suggests that the mixing behavior of the nodes of that network is assortative type, whereas the negative r value implies that the network is of disassortative type. The percentage of INs having negative r values is significantly higher and lower than that of BNs and CNs, respectively. Among the networks having nonnegative r values, the r values of INs (r^{i}) vary from 0.00 to 0.52, and those of CNs (r^{c}) vary from 0.00 to 0.51. The average of the r^{i} values was found to be 0.20 with a standard deviation 0.10, whereas that of r^{c} values was found to be 0.19 with a standard deviation 0.12. In the case of ANs, the r^{a} values varied from 0.22 to 0.43. The average of the r^{a} values was found to be 0.30 with a standard deviation 0.04.

The r values of different networks suggest that the ANs are of the assortative type. The hydrophobic networks (except one) are also of assortative type. Although most of

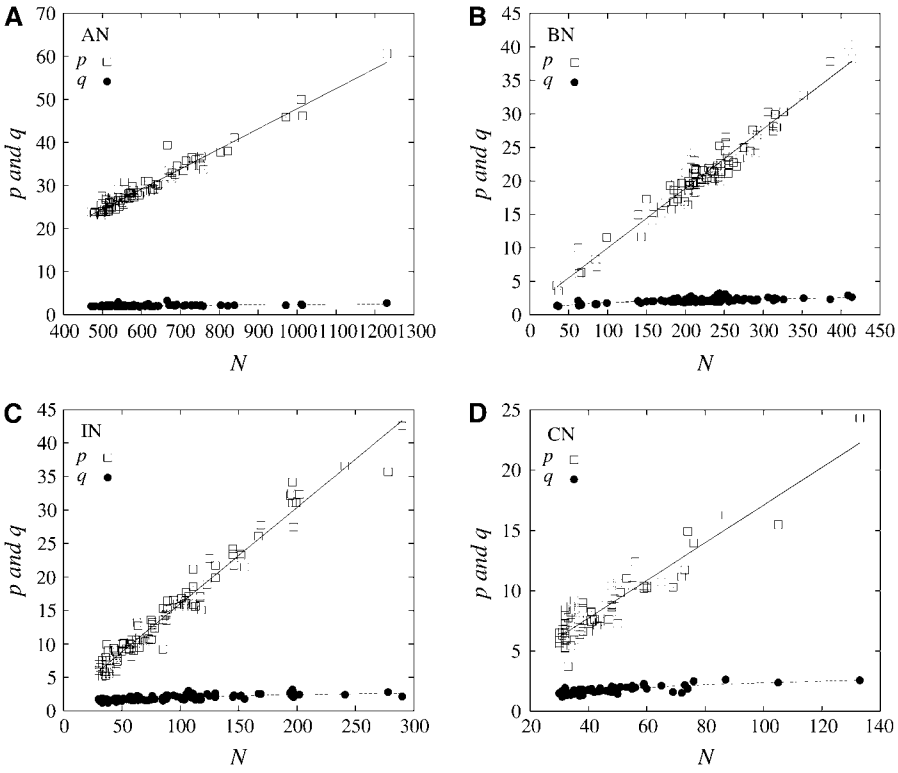

FIGURE 2 Ratios p (= $\langle C\rangle/\langle C_r\rangle$) and q (= $\langle L\rangle/\langle L_r\rangle$) as a function of network size N. The ratio p varies linearly with N, whereas the ratio q varies logarithmically with N. The best-fit curves are shown by lines for the two ratios. (*A*) All-amino-acids networks (AN). (*B*) Hydrophobic networks (BN). (*C*) Hydrophilic networks (IN). (*D*) Charged networks (CN).

the INs and CNs are of assortative type, a few others have the characteristics of disassortative mixing of the nodes, as is evident from the r values (data are not shown for negative r values). Thus, we may say that, in almost all of the BNs, the hydrophobic residues (nodes) with high degree have tendencies to be attached to the hydrophobic residues having high k values. Most of the hydrophilic and charged residues within their respective networks do follow the same behavior as followed by the hydrophobic residues. In a very few networks having negative r values, the mixing pattern of amino acid residues is different. Here the amino acids (nodes) having high k values have a tendency to be attached to amino acids with smaller degree. A protein, in general, has hydrophobic, hydrophilic, and charged residues. Thus, an AN is basically a composite network of these three types (BN, IN, and CN) of networks. When we consider ANs, we obtain the r values, which represent a cumulative effect of either all positive r values or a mixture of positive and negative r values. Thus, we find that the ANs always have positive r values.

Weighted and unweighted clustering coefficients of networks

We have calculated the weighted and unweighted clustering coefficients of each of the BNs, INs, and CNs. The average clustering coefficients of BNs, INs, and CNs are assembled separately to make the ensemble of each type. The average of each of the ensembles has been calculated and is listed in Table 1.

In the study presented here, the unweighted clustering coefficients of BNs vary from 0.41 to 0.55, whereas those of IN and CNs vary from 0.38 to 0.63 and from 0.38 to 0.67, respectively. It is evident from Table 1 that $\langle C_{\rm av}^{\rm b}\rangle < \langle C_{\rm av}^{\rm i}\rangle < \langle C_{\rm av}^{\rm c}\rangle$. We also find that the average weighted clustering coefficients of BNs, INs, and CNs vary from 0.21 to 0.28, from 0.19 to 0.33, and from 0.19 to 0.34, respectively. We have also observed that $\langle C_{\rm av}^{\rm w,b}\rangle < \langle C_{\rm av}^{\rm w,i}\rangle < \langle C_{\rm av}^{\rm w,c}\rangle$. The average weighted clustering coefficient is always nearly half that of unweighted networks. In summary, the two major observations are 1), both the unweighted and weighted clustering coefficient values of INs are higher than those of BNs but are smaller than those of CNs, and 2), the average unweighted clustering coefficients are double those of weighted clustering coefficients. The second observation indicates that the topological clustering is generated by edges with low weights. It further implies that the largest part of interactions (i.e., interactions between two amino acids) is occurring on edges (amino acids) not belonging to interconnected triplets. Therefore, the clustering has only a minor effect in the organization of each of the three different (BN, IN, and CN) types of networks. On the other hand, the unweighted clustering coefficient is a measure of local cohesiveness, and the weighted clustering coefficient takes into account the strength of the local cohesiveness. Thus, the

first observation implies that IN have higher and lower local cohesiveness than BN and CN, respectively.

Is there any hierarchical signature within the networks?

We have also studied the relation of the clustering coefficients for both weighted and unweighted networks with their degree k. We find that for most of the hydrophobic networks having $k > 8$, both the unweighted ($\langle C^{b}\rangle(k)$) and weighted ($\langle C^{b,w}\rangle(k)$) clustering coefficients change with their degree k. The results are plotted in Fig. 3. It has been observed that the nodes with smaller k values have higher clustering coefficients than the nodes with higher k values. It is known that the hierarchical signature of a network lies in the scaling coefficient of $C(k) \approx k^{-\beta}$. The network is hierarchical if β has a value of 1, whereas for a nonhierarchical network the value of β is 0 (6,26). The low-degree nodes in a hierarchical network generally belong to well-interconnected communities (high clustering coefficients) with hubs connecting many nodes that are not directly connected (small clustering coefficient). Because in most of the hydrophobic networks, $C(k)$ significantly changes with k, we intend to study the possibility of hierarchy in the hydrophobic network. Here, both the $\langle C^{b}\rangle(k)$ and $\langle C^{w,b}\rangle(k)$ exhibit a power-law decay as a function of k, as is evident from Fig. 3. It should be noted that we are aware of the problem in drawing conclusions about the power-law scaling and deriving exponents as well with such a limited range of values. But this small range of k values is actually a limitation of this real physical network. At the same time we have observed that both the $\langle C^{b}\rangle(k)$ and $\langle C^{w,b}\rangle(k)$ decrease significantly with k. So, it may be worthwhile to get an idea about the scaling coefficient values and, hence, also about the nature of networks. However, the scaling coefficient (β) for the $\langle C^{b}\rangle(k)$ varies from 0.005 to 0.750 with an average of 0.254, whereas the corresponding coefficient (β^{w}) for $\langle C^{w,b}\rangle(k)$ varies from 0.025 to 0.755 with an average of 0.231. We observe a power-law decay for both $\langle C^{b}\rangle(k)$ and $\langle C^{w,b}\rangle(k)$, but the average values (β and β^{w}) of the scaling coefficients lie very close to neither 0 nor 1 but take intermediate values. The values of the scaling coefficients imply that the networks have a tendency to hierarchical nature.

The unweighted and weighted clustering coefficients of both the hydrophilic and charged residues do not show any clear functional relation with their degree k, as is evident from Fig. 3. We have already mentioned that the small range of k values imposes a problem in drawing conclusions about the power-law scaling and deriving its exponents. Despite the limitations, we may say that the BNs bear the signature of hierarchy, whereas the INs and CNs do not have any hierarchical signature. We have further observed that ANs exhibit a signature of hierarchy as is evident from Fig. 3 and from the values of scaling coefficients listed in Table 1. The same observation has also been mentioned by Aftabuddin and Kundu (18). Thus, we may say that the hierarchical

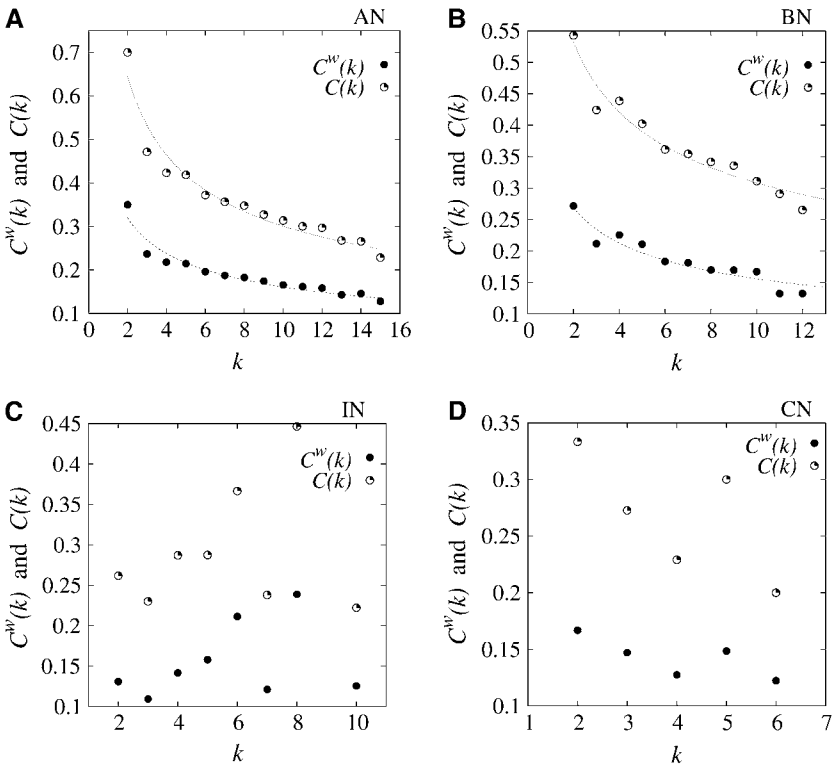

FIGURE 3 Topological clustering coefficient $C(k)$ and weighted clustering coefficient $C^{w}(k)$ as a function of degree k for different types of networks: (*A*) all-amino-acids networks (AN), (*B*) hydrophobic networks (BN), (*C*) hydrophilic networks (IN), and (*D*) charged networks (CN) for a representative protein (PDB Id:8ACN). The best-fit curves are shown by lines.

signature of ANs mainly originated from the hierarchical behavior of hydrophobic residues network.

Degree and strength distribution

We have also studied the probability degree and strength distributions of AN, BN, IN, and CN. We have observed that the probability degree distribution of network connectivities of all four types of networks (AN, BN, IN, and CN) has a peak followed by a decay whose exact nature is difficult to determine because of the small number of k values (data not shown). On the other hand, the probability strength distributions exhibit a large number of fluctuations (data not shown), which makes difficult to find the exact nature of the distributions.

CONCLUSION

In summary, all three types of networks (BN, IN, and CN) as well as ANs have the small world property. Although BNs, INs, and CNs are sparse in nature, ANs do not have any subclusters or isolated nodes. The total number of subclusters and isolated nodes in BNs of each of the proteins we have studied is significantly smaller than that of INs and CNs. The average degree of the BNs has a significantly higher value than those of the INs and CNs. On the other hand, the average strength of the INs has a smaller value than those of the BNs and CNs. We have also observed that the average strength of the charged networks is nearly equal to that of BNs. Whereas the average strength of the nodes (residues) for each of the three types of networks (BN, IN, and CN) varies with its degree, k, the average strength of a node in CNs increases more sharply than those of BNs and INs. We have further observed that ANs and BNs are of the assortative type. Although most of the INs and CNs are of the assortative type, few others have the characteristics of disassortative mixing of the nodes. We have also observed that ANs and BNs bear the signature of hierarchy, whereas the INs and CNs do not have any hierarchical signature.

The authors acknowledge the computational support provided by Distributed Information Center of Calcutta University.

REFERENCES

1. Barabasi, A. L., and R. Albert. 1999. Emergence of scaling in random networks. *Science*. 286:509–512.
2. Montoya, J. M., and R. V. Sole. 2002. Small world patterns in food webs. *J. Theor. Biol.* 214:405–412.
3. Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature*. 393:440–442.
4. Williams, R. J., E. L. Berlow, J. A. Dunne, A. L. Barabasi, and N. D. Martinez. 2002. Two degrees of separation in complex food webs. *Proc. Natl. Acad. Sci. USA*. 99:12913–12916.
5. van Noort, V., B. Snel, and M. A. Huynen. 2004. The yeast co-expression network has a small-world scale-free architecture and can be explained by a simple model. *EMBO Rep.* 5:280–284.
6. Ravasz, E., A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. 2002. Hierarchical organization of modularity in metabolic networks. *Science*. 297:1551–1555.
7. Maslov, S., and K. Sneppen. 2002. Specificity and stability in topology of protein networks. *Science*. 296:910–913.
8. Uetz, P., and M. J. Pankratz. 2004. Protein interaction maps on the fly. *Nat. Biotechnol.* 22:43–44.
9. Fell, D. A., and A. Wagner. 2000. The small world of metabolism. *Nat. Biotechnol.* 18:1121–1122.
10. Vendruscolo, M., N. V. Dokholyan, E. Paci, and M. Karplus. 2002. Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E.* 65: 061910.
11. Dokholyan, N. V., L. Li, F. Ding, and E. I. Shakhnovich. 2002. Topological determinants of protein folding. *Proc. Natl. Acad. Sci. USA*. 99:8637–8641.
12. Amitai, G., A. Shemesh, E. Sitbon, M. Shklar, D. Netanely, I. Venger, and S. Pietrokovski. 2004. Network analysis of protein structures identifies functional residues. *J. Mol. Biol.* 344:1135–1146.
13. Atilgan, A. R., P. Akan, and C. Baysal. 2004. Smallworld communication of residues and significance for protein dynamics. *Biophys. J.* 86:85–91.
14. del Sol, A., H. Fujihashi, D. Amoros, and R. Nussinov. 2006. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* 2006;2:2006.0019. Epub 2006.
15. Brinda, K. V., and S. Vishveshwara. 2005. A network representation of protein structures: implications for protein stability. *Biophys. J.* 89:4159–4170.
16. Greene, L. H., and V. A. Higman. 2003. Uncovering network systems within protein structures. *J. Mol. Biol.* 334:781–791.
17. Kannan, N., and S. Vishveshwara. 1999. Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* 292:441–464.
18. Aftabuddin, M., and S. Kundu. 2006. Weighted and unweighted network of amino acids within protein. *Physica A*. 396:895–904.
19. Kundu, S. 2005. Amino acids network within protein. *Physica A*. 346:104–109.
20. Tinoco, I., Jr., K. Sauer, and J. C. Wang. Physical Chemistry: Principles and Application in Biological Sciences. Prentice-Hall, Englewood Cliffs, NJ. 456–544.
21. PDB. Protein Data Bank, http://www.rcsb.org/
22. Barrat, A., M. Barthelemy, R. Pastor-Satorras, and A. Vespigani. 2004. The architecture of complex weighted network. *Proc. Natl. Acad. Sci. USA*. 101:3747–3752.
23. Newman, M. E. J. 2002. Assortative mixing in networks. *Phys. Rev. Lett.* 89:208701–208704.
24. Newman, M. E. J. 2001. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*. 98:404–409.
25. Barabasi, A. L., H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. 2002. Evolution of the social network of scientific collaborations. *Physica A*. 311:590–614.
26. Barabasi, A. L., and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5:101–113.