



HAL
open science

Hubs identification in Amino Acids Interaction Networks

Omar Gaci, Stefan Balev

► **To cite this version:**

Omar Gaci, Stefan Balev. Hubs identification in Amino Acids Interaction Networks. The 7th ACS/IEEE International Conference on Computer Systems and Applications, May 2009, Rabat, Morocco. pp.96-102. hal-00431270

HAL Id: hal-00431270

<https://hal.science/hal-00431270>

Submitted on 11 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hubs Identification in Amino Acids Interaction Networks

Omar Gaci and Stefan Balev
Université du Havre, LITIS EA 4108
BP 540, 76058 Le Havre - France
Email: omar.gaci@univ-lehavre.fr

Abstract—In this paper we introduce the notion of protein interaction network. This is a graph whose vertices are the protein's amino acids and whose edges are the interactions between them. Using a graph theory approach, we identify a number of properties of these networks. We compare them to the general scale-free network model and we analyze the existence of nodes with a high interaction level.

I. INTRODUCTION

Proteins are biological macromolecules participating in the large majority of processes which govern organisms. The roles played by proteins are varied and complex. Certain proteins, called enzymes, act as catalysts and increase several orders of magnitude, with a remarkable specificity, the speed of multiple chemical reactions essential to the organism survival. Proteins are also used for storage and transport of small molecules or ions, control the passage of molecules through the cell membranes, etc. Hormones, which transmit information and allow the regulation of complex cellular processes, are also proteins.

Genome sequencing projects generate an ever increasing number of protein sequences. For example, the Human Genome Project has identified over 30,000 genes which may encode about 100,000 proteins. One of the first tasks when annotating a new genome, is to assign functions to the proteins produced by the genes. To fully understand the biological functions of proteins, the knowledge of their structure is essential.

In their natural environment, proteins adopt a native compact form. This process is called folding and is not fully understood. The process is a result of interactions between the protein's amino acids which form chemical bonds. In this paper we identify some of the properties of the network of interacting amino acids. We believe that understanding these networks can help to better understand the folding process.

The rest of the paper is organized as follows. In section II we briefly present the main types of amino acid interactions which determine the protein structure. In section III we introduce our model of amino acid interaction networks. Section IV presents two general network models, random graphs and scale-free networks. In section V we compare protein interaction networks to the general models and empirically characterize them based on all protein structures available in PDB. We show how the properties of these networks are related to the structure of

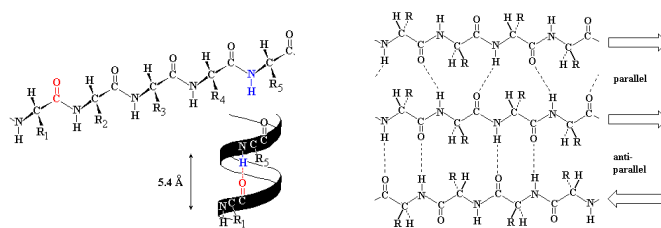


Fig. 1. Left: an α -helix illustrated as ribbon diagram, there are 3.6 residues per turn corresponding to 5.4 Å. Right: A β -sheet composed by three strands.

the corresponding proteins. Finally, in section VI we conclude and give some future research directions.

II. PROTEIN STRUCTURE

Unlike other biological macromolecules (e.g., DNA), proteins have complex, irregular structures. They are built up by amino acids that are linked by peptide bonds to form a polypeptide chain. We distinguish four levels of protein structure:

- The amino acid sequence of a protein's polypeptide chain is called its primary or one-dimensional (1D) structure. It can be considered as a word over the 20-letter amino acid alphabet.
- Different elements of the sequence form local regular secondary (2D) structures, such as α -helices or β -strands.
- The tertiary (3D) structure is formed by packing such structural elements into one or several compact globular units called domains.
- The final protein may contain several polypeptide chains arranged in a quaternary structure.

By formation of such tertiary and quaternary structure, amino acids far apart in the sequence are brought close together to form functional regions (active sites). The reader can find more on protein structure in [1].

One of the general principles of protein structure is that hydrophobic residues prefer to be inside the protein contributing to form a hydrophobic core and a hydrophilic surface. To maintain a high residue density in the hydrophobic core, proteins adopt regular secondary structures that allow non-covalent hydrogen bonds and hold a rigid and stable framework. There are two main classes of secondary structure elements (SSE), α -helices and β -sheets (see Fig 1).

An α -helix adopts a right-handed helical conformation with 3.6 residues per turn with hydrogen bonds between C'=O group of residue n and NH group of residue $n + 4$.

A β -sheet is build up from a combination of several regions of the polypeptide chain where hydrogen bonds can form between C'=O groups of one β strand and another NH group parallel to the first strand. There are two kinds of β -sheet formations, anti-parallel β -sheets (in which the two strands run in opposite directions) and parallel sheets (in which the two strands run in the same direction).

III. AMINO ACID INTERACTION NETWORKS

The 3D structure of a protein is determined by the coordinates of its atoms. This information is available in Protein Data Bank (PDB) [2], which regroups all experimentally solved protein structures. Using the coordinates of two atoms, one can compute the distance between them. We define the distance between two amino acids as the distance between their C_α atoms. Considering the C_α atom as a "center" of the amino acid is an approximation, but it works well enough for our purposes. Let us denote by N the number of amino acids in the protein. A contact map matrix is a $N \times N$ 0-1 matrix, whose element (i, j) is one if there is a contact between amino acids i and j and zero otherwise. It provides useful information about the protein. For example, the secondary structure elements can be identified using this matrix. Indeed, α -helices spread along the main diagonal, while β -sheets appear as bands parallel or perpendicular to the main diagonal [3]. There are different ways to define the contact between two amino acids. Our notion is based on spatial proximity, so that the contact map can consider non-covalent interactions. We say that two amino acids are in contact iff the distance between them is below a given threshold. A commonly used threshold is 7 Å and this is the value we use.

Consider a graph with N vertices (each vertex corresponds to an amino acid) and the contact map matrix as incidence matrix. It is called contact map graph. The contact map graph is an abstract description of the protein structure taking into account only the interactions between the amino acids. Now let us consider the subgraph induced by the set of amino acids participating in SSE. We call this graph SSE interaction network (SSE-IN) and this is the object we study in the present paper. The reason of ignoring the amino acids not participating in SSE is simple. Evolution tends to preserve the structural core of proteins composed from SSE. On the other hand, the loops (regions between SSE) are not so important to the structure and hence, are subject to more mutations. That is why homologous proteins tend to have relatively preserved structural cores and variable loop regions. Thus, the structure determining interactions are those between amino acids belonging to the same SSE on local level and between different SSEs on global level. Fig. 2 gives an example of a protein and its SSE-IN.

In [4], [5], [6], [7] the authors rely on similar models of amino acid interaction networks to study some of their properties, in particular concerning the role played by certain

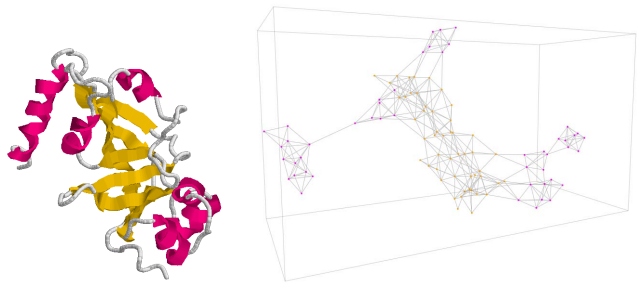


Fig. 2. Protein 1DTP (left) and its SSE-IN (right).

nodes or comparing the graph to general interaction networks models. Thanks to this point of view the protein folding problem can be tackled by graph theory approaches.

As we will see in the next section, there are three main models of interaction networks, extensively studied and whose properties are identified. The purpose of our work is to identify specific properties which associate the proteins SSE-IN with a general network model. Based on such a pattern description of SSE-IN, one can plan the study of their formation, dynamics and evolution.

IV. GENERAL MODELS OF INTERACTION NETWORKS

Many systems, both natural and artificial, can be represented by networks, that is by sites or vertices bound by links [8]. The study of these networks is interdisciplinary because they exist in scientific fields like physics, biology, computer science or information technology. The purpose of these studies is to explain how elements interact inside the network and what are the general laws which govern the observed network properties.

From physics and computer science to biology and the social sciences, researchers have found that a broad variety of systems can be represented as networks, and that there is much to be learned by studying these networks [9]. Indeed, the study of the Web [10], of social networks [11] or of metabolic networks [12] contributed to identify common non-trivial properties of these networks which have *a priori* nothing in common. The ambition is to understand how the large networks are structured, how they evolve and what are the phenomena acting on their constitution and formation [13].

In this section we present the three main models of interaction networks by describing their specific properties. We also define several measures that we use in the next section in order to study SSE-IN empirically.

A. The Random Graph Model

The random graph models are one of the oldest network models, introduced by Solomonoff and Rapoport [14] and studied by Erdős and Rényi [15], [16]. In their successive works they have identified two different random graph models, called $G_{n,m}$ and $G_{n,p}$. Each of them is defined by connection rules which are the following:

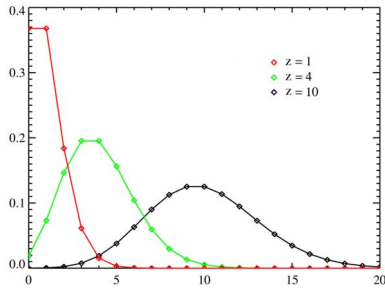


Fig. 3. Poisson distribution $p_k = \frac{z^k e^{-z}}{k!}$ with $z = 1, 4$ and 10 .

- $G_{n,m}$ regroups all graphs with n vertices and m edges. To generate a graph sampled uniformly at random from the set $G_{n,m}$, one has to put m edges between vertex pairs chosen randomly from n initially unconnected vertices.
- $G_{n,p}$ is the set of all graphs with n vertices, where each pair of vertices is connected with probability p , independently from the other pairs. To generate a graph sampled randomly, one has to begin with n initially unconnected vertices and connect each pair with probability p , or not with probability $1 - p$.

In the $G_{n,m}$ the number of edges is fixed whereas in $G_{n,p}$ the number of edges can fluctuate but its mean is fixed. When n tends to be large the two models are equivalent.

The degree of a vertex u , k_u , is the number of edges incident to u . The mean degree of a graph, denoted z for a random graph and \bar{k} more generally, is defined as follows:

$$z = \frac{1}{n} \sum_{u \in V} k_u = \frac{2m}{n}$$

The degree distribution is one of the important characteristics of this kind of networks because it affects their properties and behavior [17]. The random graph $G_{n,p}$ has a binomial degree distribution. The probability p_k that a randomly chosen vertex is connected to exactly k others is [18]:

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

When n tends to infinity, this becomes:

$$p_k = \lim_{n \rightarrow \infty} \frac{n^k}{k!} \left(\frac{p}{1-p}\right)^k (1-p)^n \simeq \frac{z^k e^{-z}}{k!}$$

which is a Poisson distribution. As we see in Fig. 3, Poisson distributions have different behavior for different mean degrees z . Each distribution has a clear peak close to $k = z$, followed by a tail that decays as $1/k!$ which is considerably quicker than exponential.

B. Scale-Free Networks

The most important property of scale-free systems is their invariance to changes in scale. The term “scale-free” refers to a system defined by a functional form $f(x)$ that remains unchanged within a multiplicative factor under rescaling of the independent variable x . Indeed, this means power-law forms,

since these are the only solutions to $f(an) = bf(n)$, where n is the number of vertices [19]. The scale-invariance property means that any part of the scale-free network is stochastically similar to the whole network and parameters are assumed to be independent of the system size [20].

If n_k is the number of vertices having degree k , we define p_k as the fraction of vertices that have degree k in the network:

$$p_k = \frac{n_k}{n}$$

The degree distribution can be expressed *via* the cumulative degree function [19], [15]:

$$P_k = \sum_{k'=k}^{\infty} p_{k'}$$

which is the probability for a node to have a degree greater or equal to k .

By plotting the cumulative degree function one can observe how its tail evolves, following a power law or an exponential distribution.

The power law distribution is defined as follows [19]:

$$P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}$$

and the exponential distribution is defined by the next formula:

$$P_k \sim \sum_{k'=k}^{\infty} e^{-k'/\alpha} \sim e^{-k/\alpha}$$

Between these two distributions, there is a mixture of them where the distribution has a power law regime followed by a sharp cut-off, with an exponential decay of the tail, expressed by the next formula:

$$P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} e^{-k'/\alpha} \sim k^{-(\alpha-1)} e^{-k/\alpha}$$

Like a power law distribution, it decreases polynomially, so that the number of vertices with weak degree is important while a reduced proportion of vertices having high degree exists. The last are called “hubs”, that is sites with large connectivity through the network, see Fig 4.

The scale-free model depends mainly on the kind of degree distribution, thus a network is defined as a scale-free if:

- The degree distribution is a power law distribution $P(k) \sim k^{-\alpha}$ over a part of its range.
- The distribution exponent satisfies $2 < \alpha \leq 3$ [21].

Amaral et al. [22] have studied networks whose cumulative degree distribution shape lets appear three kinds of networks. First, scale-free networks whose distribution decays as a power law with an exponent α satisfying bounds seen above. Second, see Fig 4, broad-scale or truncated scale-free networks whose the degree distribution has a power law regime followed by a sharp cutoff. Third, single-scale networks whose degree distribution decays fast like an exponential.

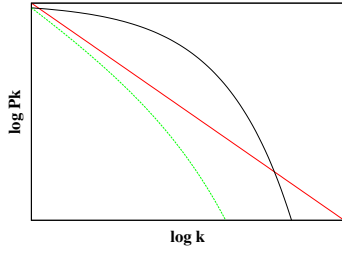


Fig. 4. Degree distribution for each of the three models described by Amaral[22]. The red line follows a power law, a function with a relatively “fat tail” as for scale-free networks. The green line corresponds to truncated scale-free networks because it describes a power law regime followed by a sharp cut-off. The black curve has a fast decaying tail, typically exponential, and corresponds to single-scale networks.

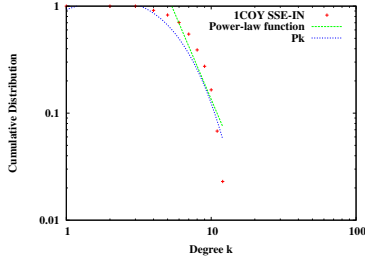


Fig. 5. Cumulative degree distribution for protein 1COY SSE-IN.

V. EXPERIMENTAL RESULTS

We have worked on the entire PDB database available in October 2007, that is 46679 files. We made our simulations by considering all proteins SSE-IN without limiting ourselves to a particular protein family. Indeed, our goal is to identify a general model for proteins SSE-IN and so that we don’t discriminate proteins due to their biological function. Thus, despite the heterogeneity we can expect among folded proteins, we assume that a general model of interaction networks emerge and can be identified from protein SSE-IN which come from combination of the 20 amino acids.

We compute the cumulative degree distribution for all proteins SSE-IN, a sample of our results is presented on Fig. 5. We can remark that the curves describe a power law regime followed by the sharp cut-off. The power law function is expressed as following:

$$p(k) = 213.413k^{-\alpha}, \text{ where } \alpha = 3.2 \pm 0.6$$

while the distribution is approximated by the next function:

$$P_k = 1.48347k^{0.962515}exp^{-k/2.12615}$$

We observe the same result for all studied proteins, that is a cumulative degree distribution approximated by the function P_k . Here, we discuss about characteristics or conditions which involve a such a behavior.

First, we are interested on the degree distribution and mainly its shape, see Fig. 6. We see that degree distribution follows a Poisson distribution whose peak is reached for a degree near

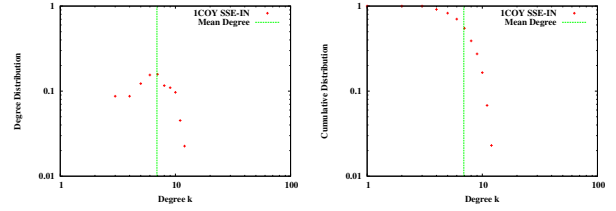


Fig. 6. Degree and cumulative degree distribution for 1COY SSE-IN. They decrease for degree values greater than the mean degree.

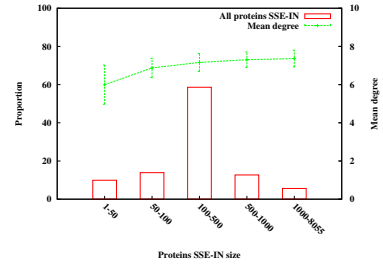


Fig. 7. Mean degree distribution according to protein SSE-IN size. The values are close and stay between 6 and 8.

\bar{k} . This result provides precision about how the vertices are connected within SSE-IN. It implies that the degree of the vertices is homogeneous. In other words a major part of them have a connectivity close to the mean degree. Consequently, the cumulative distribution depends on the mean degree value which acts as a threshold beyond which it decreases as an exponential since it is approximated *via* P_k .

Second, we study how the mean degree evolves through all SSE-IN. Its distribution, see Fig 7, indicates a relative weak variation according to the size. Even if two protein SSE-IN have size ratio around 10 or 100, their mean degree ratio is estimated to 1.05 or 1.15 and remains in the same scale order.

To illustrate the mean degree homogeneity we choose two proteins, namely 1SE9 and 1AON with sizes respectively 50 and 4998. Their size ratio is approximately 100. Even if the mean degrees are slightly different, the distributions are very similar (see Fig. 8).

To recapitulate, we show that the mean degree values constitute a threshold for protein SSE-IN cumulative degree distribution. For degrees lower than the mean degree it de-

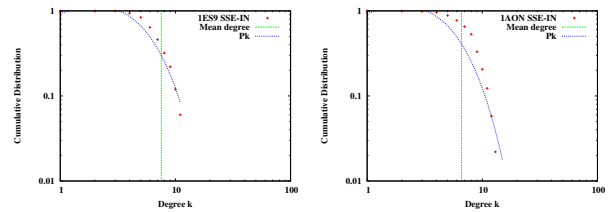


Fig. 8. Cumulative degree distribution of 1ES9 and 1AON SSE-IN whose sizes are 50 and 4988. Despite their important size difference, their mean degrees are close, respectively 6.6 and 7.5.

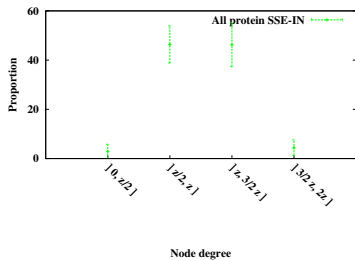


Fig. 9. Degree of nodes in protein SSE-IN as function of z . Less than 5 % of nodes are hubs.

TABLE I

STRUCTURAL FAMILIES STUDIED FOR THE SCALE-FREE PROPERTIES. WE CHOOSE ONLY FAMILIES WHICH COUNT MORE THAN 100 PROTEINS, FOR A TOTAL OF 18294 PROTEINS. WE HAVE WORKED WITH THE SCOP 1.7.3 CLASSIFICATION, A PROTEIN BELONGS TO A SCOP FOLD LEVEL IF ALL ITS DOMAINS ARE THE SAME.

Class	Number of Families	Number of Proteins
All α	12	2968
All β	17	6372
α/β	18	5197
$\alpha + \beta$	16	3757

creases slowly and after this threshold its decrease is fast compared to an exponential one, as show Fig. 5, 6 and 8.

Consequently, we find a way to approximate all proteins SSE-IN cumulative degree distribution by the function P_k which can be adjusted. This function describe a power law regime followed by a sharp cut-off which arises for degree values exceeding the mean degree. Proteins SSE-IN are so truncated scale-free networks.

Since the degree distribution depends on the mean degree value, we compare for each node its degree as function of z , see Fig 9, to illustrate how nodes interact and in particular to highlight the weak fraction of highly connected nodes, also called hubs, less than 5 % of the total node number.

An interesting study is to put in evidence the biological properties of nodes whose connectivity is marginal. Thus, we search to identify nodes which highly interact with their neighbors in the folded proteins. To do that, we have proceeded by grouping the proteins according to their secondary structure. We have already shown [23] that the protein SSE-IN topologies from structural classifications are homogeneous and established a parallel between structural and topological properties. Based on the SCOP classification and more precisely on the fold families, we have selected a total of 18294 proteins, see Table 1, and studied their SSE-IN to describe the specificity of the hubs.

For each protein SSE-IN from a SCOP fold level we identify the nodes whose degree is high, see Fig. 9. By grouping nodes as a function of amino acids they represent, we finally obtain the amino acid connectivity score by fold families. By repeating this process at the SCOP class level, we calculate and normalize the cumulative connectivity level of amino acids playing the role of hubs, see Fig 10.

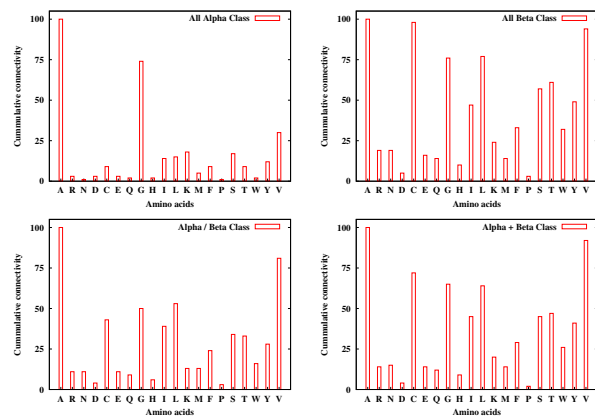


Fig. 10. Normalization of the amino acid cumulative connectivity when they act as hubs for each SCOP class level. The amino acid Ala interacts higher than the others independently of proteins classification.

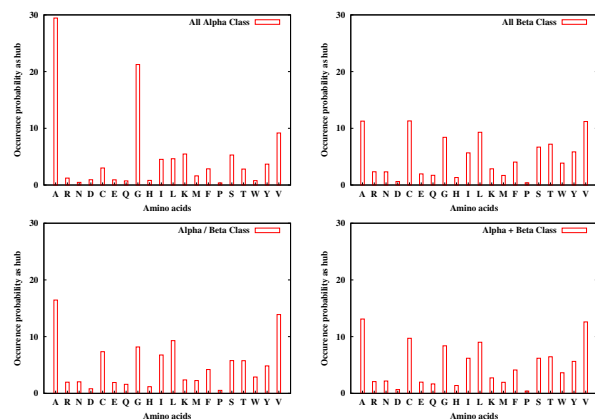


Fig. 11. Occurrence probability of hub amino acids for each SCOP Class level. The probability that an amino acid Ala has a high interaction level is greater than the others independently of proteins classification.

If we consider that the role played by an amino acid inside a folded protein is equivalent to its interaction degree, then these plots show that despite a functional diversity between the 4 SCOP Classes, there are globally the same amino acids which interact most, namely the Ala, Cys, Gly, Leu, Val. Therefore, the amino acids having a high connectivity interact independently of the protein biological function, we will explain this behavior further. Thus, the figure shows clearly that the amino acid Ala plays a central role in the studied proteins independently of the classification of their secondary structure.

We also compute the occurrence level of hubs, that is their number of appearances on each protein SSE-IN. We cumulated this score at the SCOP class level to obtain the probability for each amino acid to be a hub having a high connected occurrence in a protein SSE-IN, see Fig 11.

We can remark the existence of peaks which show clearly a high tendency of amino acids Ala, Cys, Gly, Leu and Val to have a high interactivity within the folded protein. Thus,

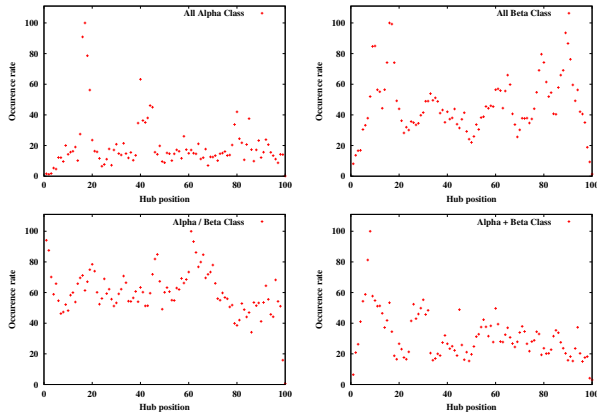


Fig. 12. Occurrence rate of hubs according to their position for each SCOP class level.

there are the same amino acids which play the role of hubs independently of the structural family, we provide below the cause of this result.

Now, we want to describe the way in which the hubs appear in the folded protein. Indeed, we study the distribution of the hubs position as a function of the structural class of the SSE-IN to identify variations dependent or not on the biological function of proteins. To make this study, we attribute an incremental position to each amino acid, beginning by 1 for the C extremity. We normalize by the total number of amino acids so that the H extremity has a position 100. Then, each time a hub exists in a SSE-IN, we increment its position occurrence number and finally normalize by the maximum to obtain the occurrence rate of hubs according to their positions in the SSE-IN, see Fig 12.

The results show the existence of favorable region in which the hub appearance is higher than somewhere else. This favorable localization is strongly visible for the All Alpha class where the hubs have a tendency to interact around the positions 20, 40 or 80. The distribution of the hub positions is the most homogeneous for the α/β class which involve a dependence to the SSE-IN topology since it not possible to find more than one strong favorable area which appear around the position 60.

To illustrate the existence of hubs favorable localization in the SSE-IN, we rely on the rich-club phenomenon [24] according to which highly connected nodes have tendency to be connected to one another. We compute the rich-club connectivity of a hub as the ratio of the number of links to the maximum number of links between nodes belonging to the rich-club.

It appears, see Fig 13, that certain hubs are isolated, mainly when the rich-club connectivity is low (position 0, 30, 60 for All Alpha SSE-IN) where as the favorable hub localizations correspond to a high coefficient.

The main observations about hubs behavior are the following. First, there are four amino acids which have higher probability to have a high connectivity. Second, hubs have

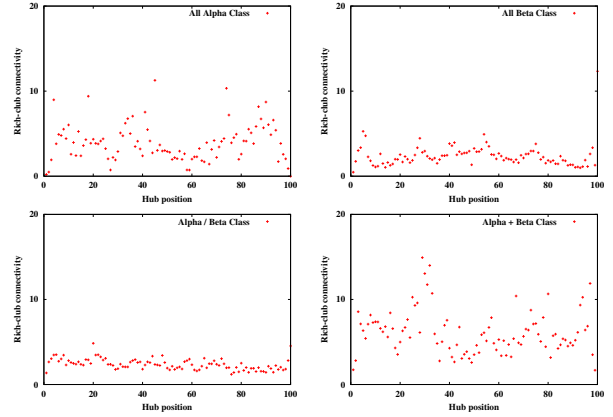


Fig. 13. Hub rich-club connectivity as a function of hub position for each SCOP Class level

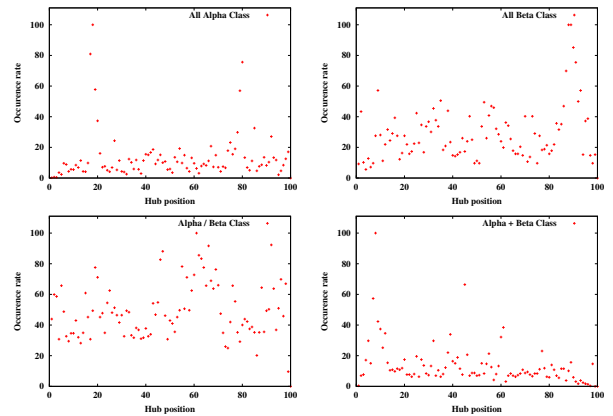


Fig. 14. Occurrence rate of the hub amino acid Ala for each SCOP Class level.

tendency to act in particular region in the SSE-IN. These two observations lead us to compute only the occurrence rate of the most frequently encountered hub Ala according to its position, see Fig 14.

By comparing figures 12 and 14, it appears clearly that the highest occurrence rate for the 4 classes corresponds to the position of the amino acid Ala. Therefore, we can establish a relation between amino acid position in the primary structure and hub apparition. Then, the amino acids Ala, Cys, Gly, Leu and Val act as hubs because they appear in the protein sequence in favorable region which involves a high interaction within the folded structure.

VI. CONCLUSION

In this paper we show that we can approximate all proteins SSE-IN cumulative degree distribution by a unique function. This function describes a power law regime followed by a sharp cut-off which arises for degree values exceeding the mean degree. Proteins SSE-IN are so truncated scale-free networks. The study of nodes whose degree is marginal shows that there exist amino acids which play a central role

independently of the protein biological functions. The nature of the hubs depends on their position in the protein sequence.

REFERENCES

- [1] C. Branden and J. Tooze, *Introduction to protein structure*. Garland Publishing, 1999.
- [2] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, 2000.
- [3] A. Ghosh, K. V. Brinda, and S. Vishveshwara, "Dynamics of lysozyme structure network: probing the process of unfolding," *Biophys J*, vol. 92, no. 7, pp. 2523–2535, April 2007.
- [4] U. K. Muppurala and Z. Li, "A simple approach for protein structure discrimination based on the network pattern of conserved hydrophobic residues," *Protein Eng Des Sel*, vol. 19, no. 6, pp. 265–275, June 2006.
- [5] K. V. Brinda and S. Vishveshwara, "A network representation of protein structures: implications for protein stability," *Biophys J*, vol. 89, no. 6, pp. 4159–4170, December 2005.
- [6] A. R. Atilgan, P. Akan, and C. Baysal, "Small-world communication of residues and significance for protein dynamics," *Biophys J*, vol. 86, no. 1 Pt 1, pp. 85–91, January 2004.
- [7] N. V. Dokholyan, L. Li, F. Ding, and E. I. Shakhnovich, "Topological determinants of protein folding," *Proc Natl Acad Sci U S A*, vol. 99, no. 13, pp. 8637–8641, June 2002.
- [8] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, pp. 268–276, 2001.
- [9] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the Web," *Computer Networks*, vol. 33, no. 1–6, pp. 309–320, 2000.
- [10] R. Albert, H. Jeong, and A.-L. Barabási, "The diameter of the world wide web," *Nature*, vol. 401, pp. 130–131, 1999.
- [11] S. Wasserman and K. Faust, *Social network analysis : methods and applications*, ser. Structural analysis in the social sciences. Cambridge: Cambridge University Press, 1994, vol. 8.
- [12] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, no. 6804, pp. 651–654, October 2000.
- [13] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.
- [14] R. Solomonoff and A. Rapoport, "Connectivity of random nets," *Bull. Math. Biophys.*, vol. 13, p. 107111, 1951.
- [15] P. Erdős and A. Rényi, "On random graphs I," *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.
- [16] —, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 7, p. 17, 1960.
- [17] A. Réka and A.-L. Barabási, "Topology of evolving networks: local events and universality," *Physical Review Letters*, vol. 85, p. 5234, 2000.
- [18] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, p. 026118, 2001.
- [19] M. E. J. Newman, "The structure and function of networks," *Computer Physics Communications*, vol. 147, pp. 40–45, 2002.
- [20] A. Réka, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, July 2000.
- [21] K.-I. Goh, K. B., and D. Kim, "Universal behavior of load distribution in scale-free networks," *Phys. Rev.*, vol. 87, 2001.
- [22] L. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proc. Natl. Acad. Sci USA*, vol. 97, no. 21, 2000.
- [23] O. Gaci and S. Balev, "Proteins: From structural classification to amino acid interaction networks," in *Proceedings of BIOCOMP'08*, vol. II. CSREA Press, 2008, pp. 728–734.
- [24] V. Colizza, A. Flammini, M. Serrano, and A. Vespignani, "Detecting rich-club ordering in complex networks," *Nature Physics*, vol. 2, p. 110, 2006.