A STUDY OF NODE DEGREE DISTRIBUTIONS IN AMINO ACID INTERACTION NETWORKS

Omar Gaci*

Abstract

A protein interaction network 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 study the properties of these networks. In a first time, we lead a topological description of structural families to observe how proteins from the same family have homogeneous topological properties. Second, we compare the studied graphs to the general model of scale-free networks. In particular, we are interested in the degree distribution and the mean degree of vertices. The results show a correlation between these two measures.

Key Words

Interaction network, protein structure, scale-free network

1. Introduction

In recent years, graph-theoretic descriptions have been applied to describe and analyse a number of complex biological systems. Such an approach is now applied to study protein structures and in particular the network of interactions between amino acids. The main interest is to investigate the usefulness of graph-theory measures to describe the relationship between protein sequence, structure and function by constructing graphs of protein structures.

In this study, we treat proteins as networks of interacting amino acid pairs [1]. In particular, we consider the subgraph induced by the set of amino acids participating in the secondary structure also called secondary structure elements (SSEs). We term this graph SSE interaction network (SSE-IN). We carry out a study to describe the SSE-INs according to a general model of interaction networks. We begin by describing the SSE-INs relying on common graph theory metrics. Because proteins belonging to the same family have homogeneous topological properties, we are interested in comparing this type of graphs to the scale-free networks. Thus, we study the SSE-IN node degree distributions to deduce that those distributions are specific and confirm relative works.

* LITIS Laboratory, EA 4108, Le Havre University, 25 rue Phillipe Lebon, 76600 Le Havre, France; e-mail: omar.gaci@univ-lehavre.fr

Recommended by Dr. L. Elnitski
(10.2316/J.2010.210-1026)

1.1 Amino Acid Interaction Networks

The 3D structure of a protein is represented by the coordinates of its atoms. We consider the residues of proteins to represent them. From files recorded in Protein Data Bank (PDB) [2], we compute the distances between pair of amino acids by considering that the C_α atom is their centre.

We consider a contact map matrix which 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. A contact is defined according to the distance between two residues, when this distance is inferior to 7 Å [3], a contact exists between these residues.

We construct 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.

In this paper, we consider the subgraph induced by the set of amino acids participating in secondary structures. We call this graph secondary structure interaction network (SSE-IN). 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. Figure 1 gives an example of a protein and its SSE-IN.

To generate a SSE-IN graph, we start from a PDB file from which we extract specific data to build the graph. For this purpose, we have developed a parser that is able to build the set of nodes representing the protein amino acids and the set of edges considered as the node interactions. Once the graph is generated, it may be displayed in two-or three-dimensional space. Towards this goal, we exploit the GraphStream library [4] which allows the manipulation of graphs.

2. A Topological Description

Through the results presented in this section, we want to offer a graph theory interpretation of the protein structural

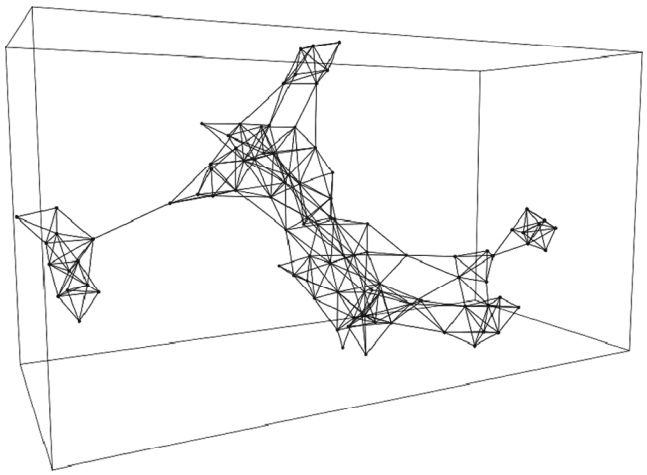

Figure 1. A representation of the isolated catalytic domain of diphtheria toxin (1DTP) SSE-IN. The nodes represent the amino acids composing the secondary structure, they participate in a-helices and b-sheets.

properties. Thus, when a protein belongs to a family according to its structural properties, by analogy, the protein SSE-IN also belongs to a specific topological family. Therefore, the SSE-IN topological properties depend on the protein structural family.

Here, we select some measures form graph theory to characterize the SSE-INs as it is presented in [5].

2.1 Topological Metrics

The distance in a graph $G=(V,E)$ between two vertices $u,v \in V$, denoted by $d(u,v)$, is the length of the shortest path connecting u and v. If there is no path between u and v, we suppose that $d(u,v)$ is undefined.

A graph diameter, D, is the longest shortest path between any two vertices of a graph:

$$D = \max\{d(u,v) : u,v \in V\}$$

The density, δ, is defined as the ratio between the number of edges in a graph and the maximum number of edges which it could have:

$$\delta = \frac{2m}{n(n-1)} \sim \frac{2m}{n^2}$$

The density of a graph is a number between 0 and 1. When the density is close to one, the graph is called dense, when it is close to zero, the graph is called sparse [6].

The clustering measures the local density of vertices [7]. For each node v, the local clustering around its neighbourhood is defined by the following way:

$$C_v = \frac{1}{2}k_v(k_v-1)$$

The clustering coefficient is a ratio between the number of edges and the maximum number of possible edges in the vertex neighbourhood. If we extend the previous definition to the entire graph, the clustering is given by the expression:

$$C_L = \frac{1}{n}\sum_{v\in V}\frac{\text{Number of connected neighbour pairs}}{C_v}$$

Nevertheless, this last definition is mainly local because for each node, the clustering involves only the node neighbourhood.

The global clustering was studied by Newman *et al.* [8] and can be measured by the following formula:

$$C_G = \frac{3 \times \text{Number of triangles in the graph}}{\text{Number of connected triplets of vertices}}$$

A triangle is formed by three vertices which are all connected and a triplet is constituted by three nodes and two edges. The global clustering coefficient C_G is the mean probability that two vertices that are neighbours of the same other vertex will themselves be neighbours.

2.2 Experimental Results

To study the SSE-INs, we need to select them according to their SSE arrangements, we consider only proteins which have one domain. We work with proteins which belong to the CATH v3.1.0 topology level or the SCOP v1.73 fold level. For each of the two classifications, we study three families as shown in Table 1. Those six families have been chosen because of their huge protein number. Thus, each family provides a broad sample guarantying more general results and avoiding fluctuations. Moreover, these six families contain proteins of very different sizes, varying from several dozens to several thousands amino acids in SSE.

Table 1
Families Studied to Put in Evidence the Parallel between Structural and Topological Properties

Name	Type	Class	Proteins
Rossmann fold	CATH	$\alpha\ \beta$	2,576
TIM Barrel	CATH	$\alpha\ \beta$	1,051
Lysozyme	CATH	Mainly α	871
Globin-like	SCOP	All α	733
TIM β/α-barrel	SCOP	α/β	896
Lysozyme-like	SCOP	$\alpha+\beta$	819

The results, see (Table 2, column D), show very close diameters between the families *TIM Barrel* and *TIM β/α-barrel* and also between the families *Lysozyme* and *Lysozyme-like*. It is correlated to the family compositions. Indeed, among these pair of families, one can observe that they contain almost the same proteins. In other words, *Lysozyme* topology in CATH is the equivalent of *Lysozyme-like* fold level in SCOP.

Table 2
Families Studied to Describe the SSE-INs according to Their Structural Families

Name	D	δ	C_L	C_G
Rossmann fold	18.84	0.033	0.63	0.56
TIM Barrel	19.83	0.030	0.64	0.57
Lysozyme	12.81	0.038	0.65	0.58
Globin-like	15.65	0.034	0.63	0.57
TIM β/α-barrel	20.09	0.029	0.64	0.57
Lysozyme-like	12.85	0.042	0.66	0.58

We observe clearly that the topological properties depend on the structural family. Indeed, the topological description is different from a family to another.

As the density is concerned, we observe (column δ) that the families *TIM Barrel* and *TIM β/α-barrel* have the minimum density. It has an impact on the SSE-IN topology. Indeed, as the density is low, the network is less connected and, therefore, the diameter is higher.

The local clustering C_L measures the fraction of pairs of a vertex's neighbours and the global clustering C_G gives the probability that among three vertices at least two are connected. The results presented (column C_L and C_G) show that the clustering coefficients are close for different families and cannot be correlated to density values. Consequently, the neighbour density remains independent of the previously studied properties.

2.3 A Topological Space

Until now, we have proposed an alternative means to describe a protein structural family when proteins are represented as SSE-INs. In particular, we have enumerated some topological properties, like diameter and density, which allow discriminating two distinct families, whereas others, like clustering coefficients, are general properties of all SSE-IN. Consequently, we except that proteins which have similar structural properties or biological functions will also have similar SSE-IN properties. In this way, our model allows us to draw a parallel between biology and graph theory.

In [5], we illustrate this parallel by building a topological space where proteins are described by their SSE-IN topologies. We except that proteins from a same structural family have SSE-INs which are grouped in this topological space. Thus, we build a 3D topological space whose dimensions are the topological metrics enough discriminant between SSE-INs belonging to different families.

From a dataset composed of 10 families in SCOP v1.73 classification, we consider a SSE-IN as a dot in our 3D topological space, see Fig. 2. The x axis represents the SSE-IN size, denoted N, the y axis represents the densities, denoted G, and the z axis represents the diameters, denoted D. The plots present only a sample of our results. It appears clearly that proteins belonging to the same family are grouped.

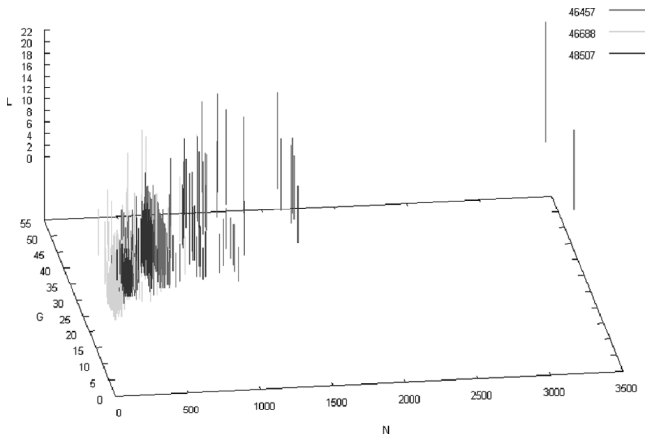

Figure 2. A 3D topological space. The x axis represents the SSE-IN size, y the density and z the diameters. The families are identified by their SCOP id in the v1.73 classification.

Thus, our topological space provides a means to discriminate structural families when proteins are represented by SSE-INs. Therefore, the parallel between structural and topological properties can be illustrated through the topological space we propose.

3. Comparison with the Scale-Free Network

In the previous section, we have shown that the structural properties of proteins can be translated into a topological space described by discriminant metrics. Then, we continue to study the parallel between structural and topological properties by comparing the SSE-INs to a general model of interaction networks.

3.1 The Scale-Free Model

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, as these are the only solutions to $f(an) = b\ f(n)$, where n is the number of vertices [9]. 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 [10].

Definition 1. *The degree of a vertex u, k_u, is the number of edges incident to u. The mean degree of a graph, denoted z, is defined as follows:*

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

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

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

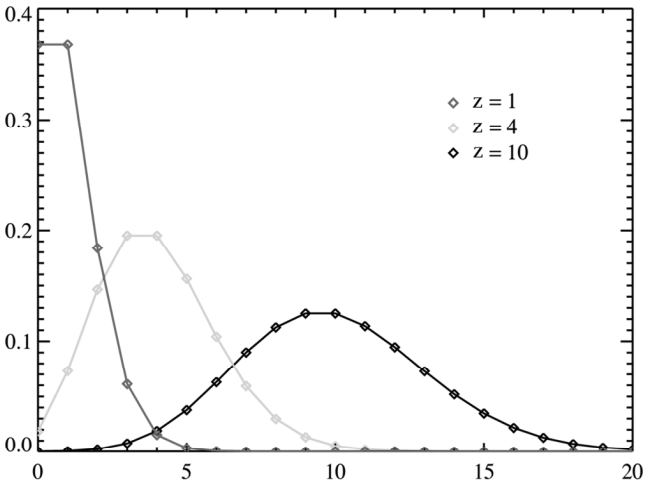

Figure 3. Examples of Poisson distributions. The curves are centered around the value of z when $p_k = \frac{z^k e^{-z}}{k!}$ with $z = 1, 4$, and 10.

The degree distribution is an important characteristic of interaction networks because it affects their properties and behaviour [11]. The probability p_k that a randomly chosen vertex is connected to exactly k others is [12]:

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

When n tends to infinity, this becomes:

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

which is a Poisson distribution. As we see in Fig. 3, Poisson distributions have different behaviour for different mean degree 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 any exponential.

The degree distribution can also be expressed *via* the cumulative degree function [13]:

$$P_k = \sum_{k'=k}^{\infty} p_{k'} \tag{1}$$

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 following [9]:

$$P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)} \tag{2}$$

and the exponential distribution is defined by the next formula:

$$P_k \sim \sum_{k'=k}^{\infty} e^{-k'/\alpha} \sim e^{-k/\alpha} \tag{3}$$

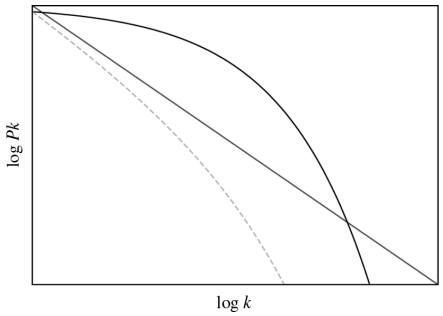

Figure 4. Degree distribution for each of the three models described by Amaral [15]. The red line follows a power law, a function with a relatively "fat tail" as for scale-free networks. The green line corresponds to truncated scale-free networks because it describes a power law regime followed by a sharp cut-off. The black curve has a fast decaying tail, typically exponential, and corresponds to single-scale networks.

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

$$P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} e^{-k'/\alpha} \sim k^{-(\alpha-1)} e^{-k/\alpha} \tag{4}$$

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 cumulative degree distribution is a power law distribution $P_k \sim k^{-\alpha}$ over a part of its range, see (2).
- The distribution exponent satisfies: $2 < \alpha \leq 3$ [14].

Amaral *et al.* [15] have studied networks whose cumulative degree distribution shape lets appear three kinds of networks, see Fig. 4. First, scale-free networks whose distribution decays as a power law with an exponent α satisfying bounds seen above (2). Second, (3), single-scale networks whose degree distribution decays fast like an exponential. Third, (4), broad-scale or truncated scale-free networks whose degree distribution has a power law regime followed by a sharp cut-off.

3.2 Experimental Results

As we explained previously, to study the SSE-INs, we build a dataset composed of proteins grouped according to their structural families. Here, to compare SSE-INs to the scale-free model, we rely on the SCOP v1.73 classification and notably the fold family level to select a total of 18,296 proteins, see Table 3. Also, a protein belongs to a SCOP fold level if all its domains are the same.

Table 3
Structural Families Studied

Class	Number of Families	Number of Proteins
All α	12	2,970
All β	17	6,372
α/β	18	5,197
$\alpha+\beta$	16	3,757

For each class, we select template families according to their protein numbers. By this way, a family is described by enough proteins to provide a general topological description. The total number of studied proteins is 18,296. We have worked with the SCOP 1.73 classification.

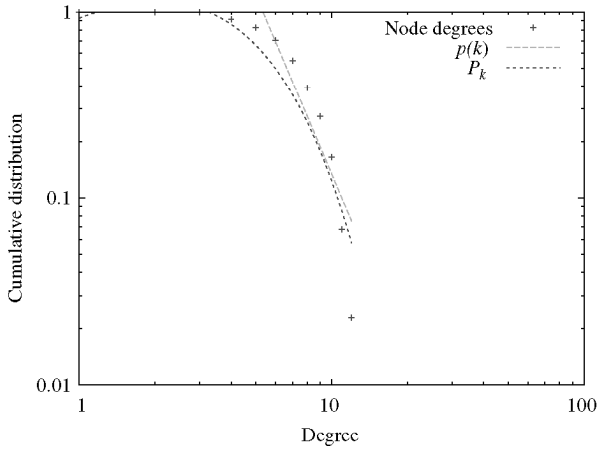

Figure 5. Cumulative degree distribution for 1COY SSE-IN. This distribution is approximated by the function P_k (6), and in the same time it follows a power law regime (5).

3.3 General Behaviour

To compute the cumulative degree distribution of protein SSE-INs (denoted $P_k = a\ k^{-b} \exp^{-k/c}$, see (4), we have divided our dataset into two parts whose first one is composed of 20% of the total studied proteins. Then, we fit our specific sub dataset relying on a numerical approximation using the method of least squares. Once, we have obtained the coefficients from our sub dataset, we apply them for the others studied proteins. A sample of our results is presented in Fig. 5. We can remark that the cumulative distributions describe a power law regime followed by a sharp cut-off. The power law function is expressed as following:

$$p(k) = 213.413\, k^{-\alpha},\ \text{where } \alpha = 3.2 \pm 0.6 \qquad (5)$$

while the distribution is approximated by the next function:

$$P_k = 1.48347\, k^{0.962515}\ \exp^{-k/2.12615} \qquad (6)$$

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 such a behaviour.

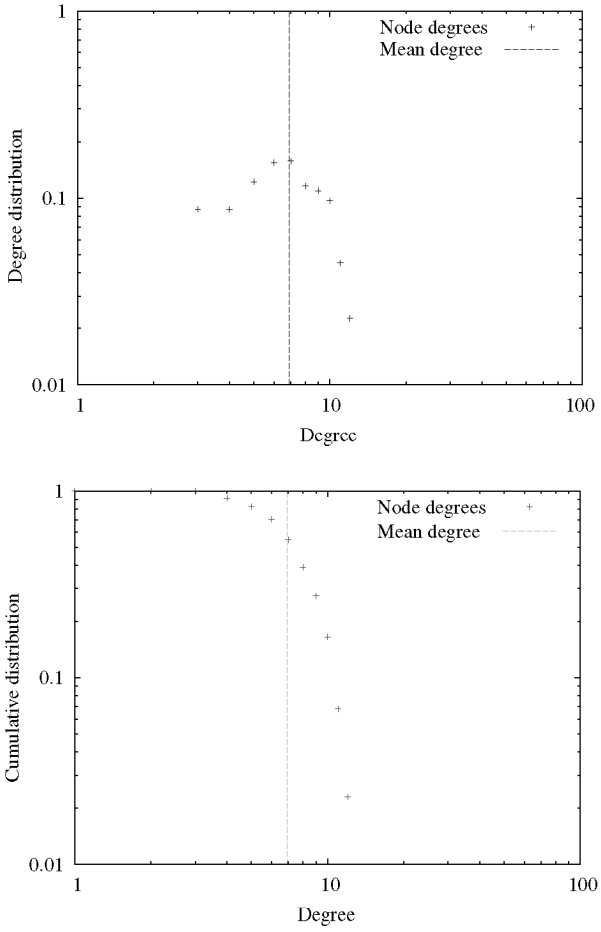

Figure 6. Degree and cumulative distribution for 1COY SSE-IN. Top, the curve follows a Poisson distribution whose peak equals to the mean degree z. Bottom, the curve decreases quickly for degrees superior to z.

First, we are interested in the degree distribution and mainly its shape, see Fig. 6. We can see that degree distribution follows a Poisson distribution whose peak is reached for a degree near z. 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 has a connectivity enough 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 as 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 4,998. 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 de-

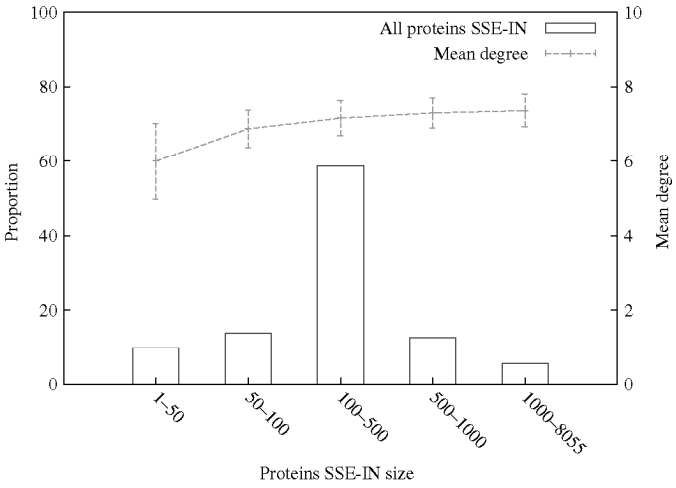

Figure 7. Mean degree distribution according to protein SSE-IN size. It evolves with values enough close, between 5 and 8.

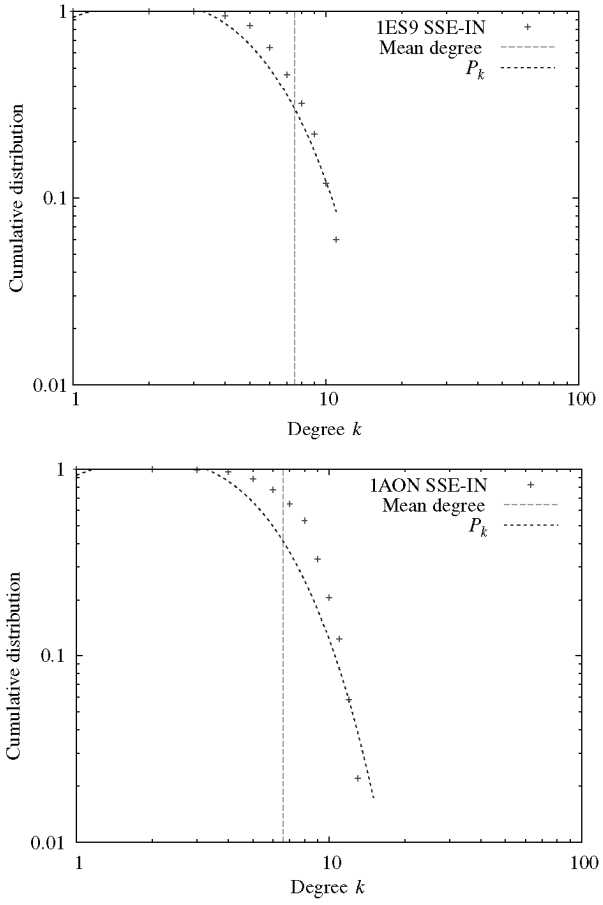

Figure 8. Cumulative degree distribution of 1SE9 and 1AON SSE-IN whose size equals 50 and 4,988. Despite their important size difference, their mean degree stay close and worth, respectively, 6.6 and 7.5.

gree distribution. For degrees lower than the mean degree, it decreases slowly and after this threshold its decrease is fast compared to an exponential one, as shown Figs 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 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. Thus, we find a way to approximate all protein SSE-INs cumulative degree distribution by the function P_k which can be adjusted. This result is also confirmed by previous studies [16–18].

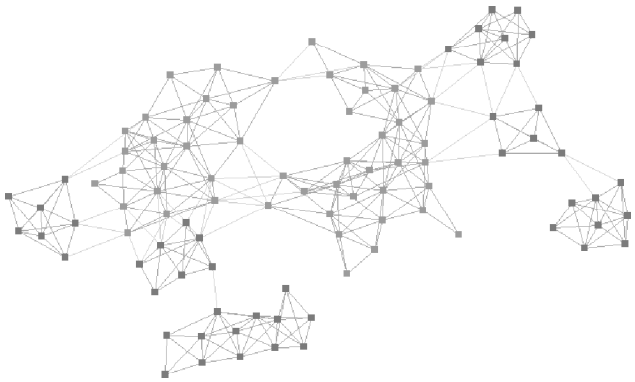

Figure 9. SSE-IN of 1DTP protein. The edges connecting different SSE are green.

3.4 Mean Degree Evolution

As the mean degree plays the role of a threshold beyond which the cumulative degree distribution decreases exponentially, it is interesting to study its evolution with the size of the network. Figure 7 shows that the mean degree increases very slightly with the size of the network. Even for networks with size ratio of 100, the mean degree ratio is only 1.15 (as an example, see Fig. 8).

Whatever the size of the network is, we observe that the mean degree is always between 5 and 8. This mean degree interval is a common property characterizing all SSE-IN. To explain this property, let us consider the structure of our networks. They are composed of densely connected subgraphs corresponding to SSEs (see Fig. 9). The number of edges connecting different subgraphs is relatively small, but these edges are the most important, because they correspond to interactions determining the tertiary structure.

We start by computing the mean degree in each SSE subgraph. The results are shown in Fig. 10. We can see that the mean degree evolution at microscopic level is almost the same as at macroscopic level (compare to Fig. 7). Independently of the SSE size and type, the mean degree of each SSE subgraph, z_{SSE} is always bounded:

$$z_{\min} < z_{\mathrm{SSE}} < z_{\max} \tag{7}$$

when the size of the network is more than 10. In the general case $z_{\min} = 5$ and $z_{\max} = 8$, but when we consider a specific SSE size and type, finer bounds can be found (see Fig. 10).

Now let us consider a whole protein. Suppose that it contains s SSEs and let the element i has n_i vertices and m_i edges, $i = 1, \ldots, s$. Then, the total number of

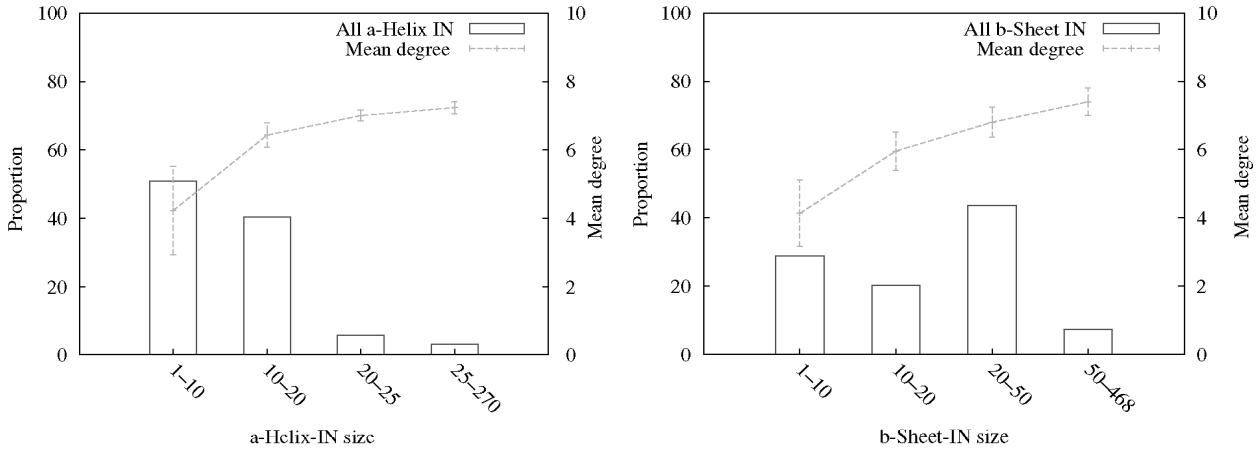

Figure 10. SSE subgraphs size distribution and mean degree as a function of the size. The mean degree evolves between 3 and 5.

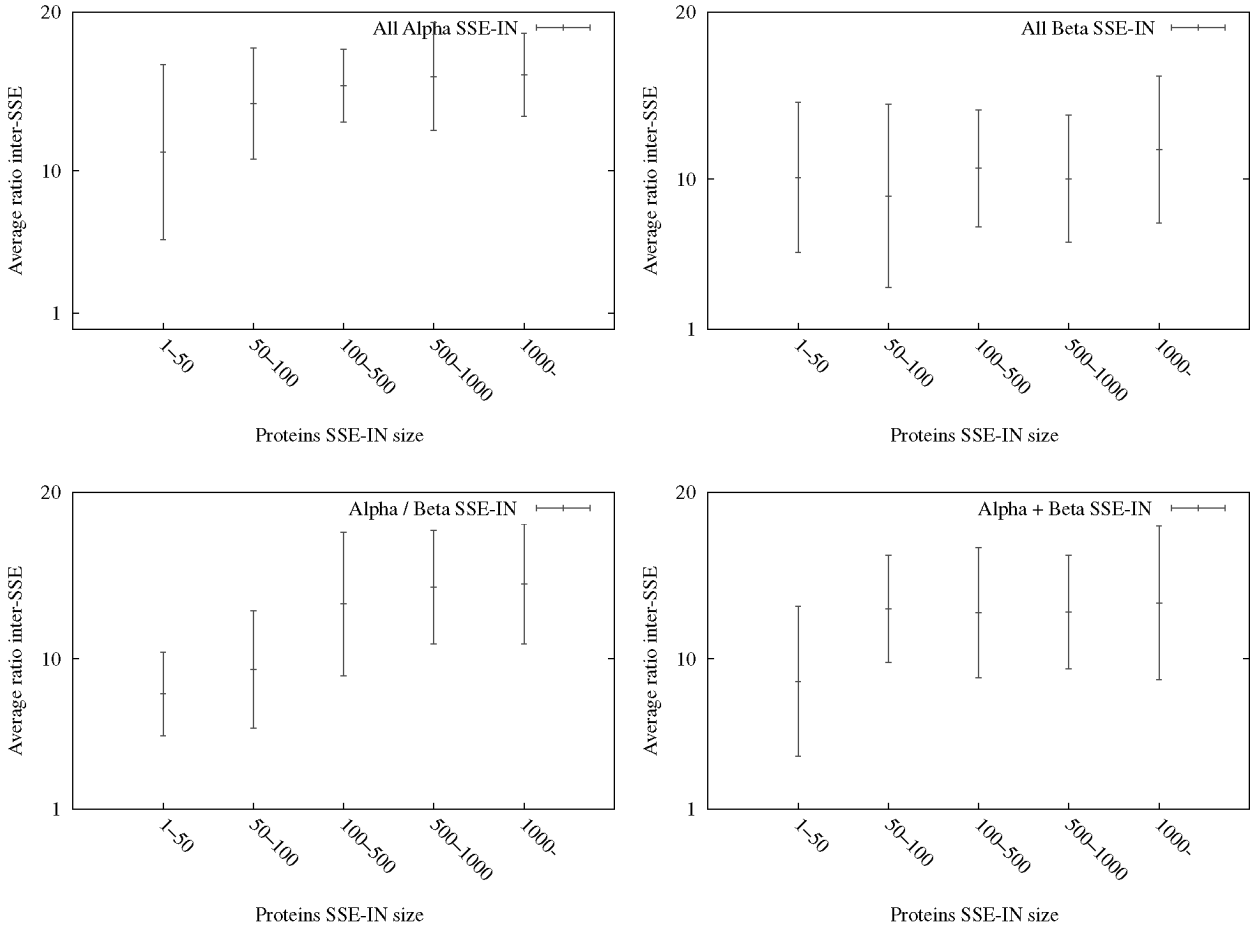

Figure 11. Ratio of inter-SSE edges (r) as a function of the network size for the four classes studied. It is bounded and never exceed 20% of the total edge numbers.

vertices is $n = \sum_{i=1}^{s} n_i$ and the total number of edges is $m = \sum_{i=1}^{s} m_i + m_{\text{inter}}$, where m_{inter} is the number of edges connecting vertices from different SSEs. Let $r = m_{\text{inter}}/m$ be the ratio of inter-SSE edges. Then:

$$\frac{m}{n} = \frac{\sum\limits_{i=1}^{s} m_i + m_{\text{inter}}}{\sum\limits_{i=1}^{s} n_i} = \frac{\sum\limits_{i=1}^{s} m_i}{\sum\limits_{i=1}^{s} n_i} + r\frac{m}{n} \tag{8}$$

and hence for the mean degree z we have:

$$z = \frac{2m}{n} = \frac{2}{1-r}\frac{\sum\limits_{i=1}^{s} m_i}{\sum_{i=1}^{s} n_i} \tag{9}$$

On the other hand, from (7) it follows that:

$$\frac{z_{\min}}{2} n_i < m_i < \frac{z_{\max}}{2} n_i,\ i = 1, \ldots, s \tag{10}$$

By summing up the last equation we obtain:

$$\frac{z_{\min}}{2} < \frac{\sum_{i=1}^{s} m_i}{\sum_{i=1}^{s} n_i} < \frac{z_{\max}}{2} \tag{11}$$

which together with (9) gives:

$$\frac{z_{\min}}{1-r} < z < \frac{z_{\max}}{1-r} \tag{12}$$

The last equation gives finer bounds on the mean degree. It shows that the bounds on z depend not only on the bounds on z_{SSE}, but also on the ratio of inter-SSE edges. A higher proportion of inter-SSE edges shifts up the bounds. Proteins with bigger size have more SSEs and hence more links between different SSEs. This explains the increase of the mean degree with the size of the networks. Figure 11 shows that the number of inter-SSE edges is quite variable, but it never exceeds 20%. It is the consequence of the excluded volume effect, because the number of residues that can physically reside within a given radius is limited. This last property explains why the mean degree is homogeneous.

4. Conclusion

In this paper, we consider proteins as interaction network of amino acids. We study some of the properties of these networks. It appears that specific properties, like diameter and density, allow discriminating two distinct families, whereas others are common to all SSE-INs. Thus, proteins whose structural properties are similar will also have similar SSE-IN properties. In this way, our model allows us to draw a parallel between biology and graph theory.

We compare the amino acid interaction network to the model of scale-free networks. We show that we can approximate all protein SSE-IN cumulative degree distributions 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. Protein SSE-INs are so truncated scale-free networks. This node distribution implies that there exist amino acids whose degree is marginal (greater than the mean degree).

The characterization we propose constitutes a first step of a new approach to the protein folding problem. The properties identified here, but also other properties we studied [19], can give us an insight on the folding process. They can be used to guide a folding simulation in the topological pathway from unfolded to folded state.

References

[1] A.R. Atilgan, P. Akan, & C. Baysal, Small-world communication of residues and significance for protein dynamics, *Biophysical Journal*, *86*(1), 2004, 85–91.

[2] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, & P.E. Bourne, The protein data bank, *Nucleic Acids Research*, *28*(1), 2000, 235–242.

[3] N.V. Dokholyan, L. Li, F. Ding, & E.I. Shakhnovich, Topological determinants of protein folding, *Proceedings of National Academy of Sciences USA*, *99*(13), 2002, 8637–8641.

[4] A. Dutot, F. Guinand, D. Olivier, & Y. Pigné, GraphStream: A tool for bridging the gap between complex systems and dynamic graphs, *Proc of EPNACS: Emergent Properties in Natural and Artificial Complex Systems*, Dresden, Germany, 2007, 137–143.

[5] O. Gaci & S. Balev, *Advances in computational biology* (Princeton, NJ: Springer, 2010).

[6] T.F. Coleman & J.J. Moré, Estimation of sparse Jacobian matrices and graph coloring problems, *SIAM Journal on Numerical Analysis*, *20*(1), 1983, 187–209.

[7] D.J. Watts, *Small worlds* (Princeton, NJ: Princeton University Press, 1999).

[8] M.E.J. Newman, The structure of scientific collaboration networks, *Proceedings of National Academy of Sciences USA*, *98*(2), 2001, 404–409.

[9] M.E.J. Newman, The structure and function of networks, *Computer Physics Communications*, *147*(1–2), 2002, 40–45.

[10] A. Réka, H. Jeong, & A.-L. Barabási, Error and attack tolerance of complex networks, *Nature*, *406*(1), 2000, 378–382.

[11] A. Réka & A.-L. Barabási, Topology of evolving networks: Local events and universality, *Physical Review Letters*, *85*(24), 2000, 5234–5237.

[12] M.E.J. Newman, S.H. Strogatz, & D.J. Watts, Random graphs with arbitrary degree distributions and their applications, *Physical Review E*, *64*(2), 2001, 026118.

[13] P. Erdõs & A. Rényi, On random graphs I, *Publicationes Mathematicae*, *6*(1), 1959, 290–297.

[14] K.-I. Goh, B. Kahng, & D. Kim, Universal behavior of load distribution in scale-free networks, *Physical Review Letters*, *87*(27), 2001, 278701.

[15] L.A.N. Amaral, A. Scala, M. Barthélémy, & H.E. Stanley, Classes of small-world networks, *Proceedings of National Academy of Sciences USA*, *97*(21), 2000, 11149–11152.

[16] L.H. Greene & V.A. Higman, Uncovering network systems within protein structures, *Journal of Molecular Biology*, *334*(4), 2003, 781–791.

[17] C. Bode, I.A. Kovács, M.S. Szalay, R. Palotai, T. Korcsmáros, & P. Csermely, Network analysis of protein dynamics, *FEBS Letters*, *581*(15), 2007, 2776–2782.

[18] J. Huang, S. Kawashima, & M. Kanehisa, New amino acid indices based on residue network topology, *Genome Informatics*, *18*(1), 2007, 152–161.

[19] O. Gaci, A topological description of hubs in amino acid interaction networks, *Advances in Bioinformatics*, *2010*(1), 2010, doi: 10.1155/2010/257512.

Biography

Omar Gaci received, in 2005, his Professional Master of Science in Computer Science with a specialization in distributed object systems. He obtained his Research Master of Science in Computer Science, specialization in mathematic and computer science applied to complex systems, in 2006.

From October 2006, he is a Ph.D. student in the LITIS laboratory located in Le Havre University, France. His researches deal with interaction networks obtained from folded proteins. The study of these interaction networks helps to describe the folding process relying on the graph theory and by exploiting the behaviour of interaction networks.