Comparative protein analysis of the chitin metabolic pathway in extant organisms: A complex network approach

Aristóteles Góes-Neto [a,*], Marcelo V.C. Diniz [a], Leonardo B.L. Santos [b], Suani T.R. Pinho [b], José G.V. Miranda [b], Thierry Petit Lobao [c], Ernesto P. Borges [d], Charbel Niño El-Hani [e], Roberto F.S. Andrade [b]

[a] Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, Feira de Santana, Bahia 44036-900, Brazil
[b] Instituto de Física, Universidade Federal da Bahia, Campus Universitário de Ondina, Salvador, Bahia 40170-110, Brazil
[c] Instituto de Matemática, Universidade Federal da Bahia Campus Universitário de Ondina, Salvador, Bahia 40170-110, Brazil
[d] Escola Politécnica, Universidade Federal da Bahia, Rua Prof. Aristides Novis, 02 - Federação, Salvador, Bahia 40210-630, Brazil
[e] Instituto de Biologia, Universidade Federal da Bahia, Campus Universitário de Ondina, Salvador, Bahia 40170-110, Brazil

ARTICLE INFO

ABSTRACT

Chitin is a structural endogenous carbohydrate, which is a major component of fungal cell walls and arthropod exoskeletons. A renewable resource and the second most abundant polysaccharide in nature after cellulose, chitin is currently used for waste water clearing, cosmetics, medical, and veterinary applications. This work comprises data mining of protein sequences related to the chitin metabolic pathway of completely sequenced genomes of extant organisms pertaining to the three life domains, followed by meta-analysis using traditional sequence similarity comparison and complex network approaches. Complex networks involving proteins of the chitin metabolic pathway in extant organisms were constructed based on protein sequence similarity. Several usual network indices were estimated in order to obtain information on the topology of these networks, including those related to higher order neighborhood properties. Due to the assumed evolutionary character of the system, we also discuss issues related to modularity properties, with the concept of edge betweenness playing a particularly important role in our analysis. Complex network approach correctly identifies clusters of organisms that belong to phylogenetic groups without any *a priori* knowledge about the biological features of the investigated protein sequences. We envisage the prospect of using such a complex network approach as a high-throughput phylogenetic method.

© 2010 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The rapidly developing theory of complex networks, based on both graph theory and statistical mechanics, has been successfully applied to uncover the organizing principles that govern the formation and evolution of various complex biological, technological, and social systems (Barabasi and Oltvai, 2004).

A key challenge of contemporary biology is to carry out an integrated theoretical and experimental program to map out, understand, and model, in quantifiable terms, the topological and dynamic properties of diverse biological networks (Barabasi and Oltvai, 2004). Systems biology is concerned with phenomena that arise when a number of interaction partners operate in complex networks, and not only function individually (Stoll and Naef, 2004). There is, therefore, a great heuristic potential in applying the theory of complex networks in the context of systems biology, as we intend to do in our research project.

Recent studies using complex network approach in the fields of both genomics and proteomics (Gavin et al., 2004; Boone et al., 2007) have contributed to a better knowledge of the structure and dynamics of the complex webs of interactions of a living cell. Although molecular biological networks are intricately interconnected and interwoven inside an organism, at least three distinct molecular networks can be distinguished: protein interaction, transcriptional (or gene regulatory), and metabolic networks. Furthermore, proteins are also evolutionarily related through their phylogeny (Silva and Stumpf, 2005).

Several features of complex systems can be identified and understood through the recently developed complex network framework (Albert and Barabási, 2002; Boccaletti et al., 2006; Costa

* Corresponding author at: Universidade Estadual de Feira de Santana, Departamento de Ciências Biológicas, Av. Transnordestina, s/n, Bairro Novo Horizonte, Feira de Santana, Bahia 44036-900, Brazil. Tel.: +55 7532248296; fax: +55 7532248132.
E-mail addresses: arigoesneto@pq.cnpq.br, arigoesneto@gmail.com, arigoesneto@gmail.com, agoesnt@uefs.br (A. Góes-Neto).
URL: http://www.uefs.br/ppgbiotec (A. Góes-Neto).

et al., 2007). As such systems contain a large number of variables, the use of functional relationships among their constituents allows us to construct an interaction network, which can offer a first indication of how the system is structured. Of course, the identification of the proper functional relationships, which are responsible for assigning the edges between nodes, is the main and crucial step in this methodology. To use the simplest framework, we consider a non-weighted, undirected complex network R with N nodes and E edges. This means that no edge carries more information than the others, and that if node i is connected to node j, then, node j is also connected to node i.

In most of the studies that use network concepts, the properties of the systems are expressed by several parameters that describe some topological properties of the network, such us the average degree $\langle k \rangle$, the clustering coefficient C, the mean minimal distance among the nodes $\langle d \rangle$, and the network diameter D. The degree k_i of a node i counts the number of edges connected to it, while $\langle k \rangle$ is the average number of edges per node over the network.

The clustering coefficient C_i of node i is defined as the ratio between the number of edges among the immediate neighbors of i and $k_i(k_i - 1)/2$, which is the maximum number of edges between the set of neighbors of i. Again, the average of C_i over i leads to the network clustering coefficient C.

A finer description of the relationship among the nodes can be expressed by the distribution function of node degree $p(k)$, which counts the relative number of nodes with a degree k.

If two nodes, i and j, are directly connected, the minimal distance between them is 1. If this is not the case, they can be connected by a path of edges in the network. More than one distinct path can be used to go from i to j, but the shortest path $d_{i,j}$ corresponds to the path with the smallest number of edges connecting those nodes. If two nodes i and j are not connected, the distance between them is not well defined and, for the purpose of going running the process, we define $d_{i,j} = 0$. The average shortest path $\langle d_i \rangle$ of node i is the mean value, over j, of $d_{i,j}$, while the network average shortest path $\langle d \rangle$ is obtained by averaging $\langle d_i \rangle$ over the whole set of network nodes. Finally, the network diameter D is the largest value of $d_{i,j}$.

If any pair of nodes, i and j, can be connected through a path over network edges, the network is formed by a single cluster. If this is not the case, the network is split into several sub-graphs or sub-networks, each one of them constituted by one cluster. If this is the case, the further analysis of the network depends on the way the sub-graphs are constituted. In many situations, as in the network we consider in this work, it turns out that there is only one giant cluster, which has a much more number of nodes than all the other sub-graphs The identification and characterization of the largest cluster in the network, also called the largest connected component, become quite relevant, as it displays the topological features shared by the largest number of nodes in the network.

A measure of the importance of a given edge between two nodes, i and j, to the network structure is provided by the edge betweenness degree. It counts the number of shortest paths between the $N(N - 1)/2$ pairs of nodes that go through that edge. If we successively eliminate the edges with highest values of betweenness degree, as proposed by Newman and Girvan (Newman and Girvan, 2004), it is possible to identify community structures in networks that are characterized by large modularity. This concept indicates that a network is composed by several communities (or modules). The nodes within each community are densely connected with each other, but they are sparsely connected with nodes of other communities.

Here, we used a complex network approach to investigate the chitin metabolic pathway in a phylogenetic framework as it is shown in Fig. 1. Chitin, the β-1,4-linked linear homopolymer of N-acetylglucosamine, is a structural endogenous carbohydrate, which is a major component of fungal cell walls (Bowmann and Free, 2006), cephalopod beaks (Hanlon and Messenger, 1996), integuments of larvae and young nematodes (Ax, 1996), and arthropod exoskeletons (Merzendorfer, 2006). Chitin occurs only in extant eukaryotic organisms of the Metazoa-Fungal clade. This suggests that chitin may have evolved before the crown eukaryotic radiation. The same can be said of cellulose, which is highly similar to chitin (Merzendorfer, 2006) and is found in eukaryotic cell walls of Viridiplantae (plants and green algae) (Raven et al.,
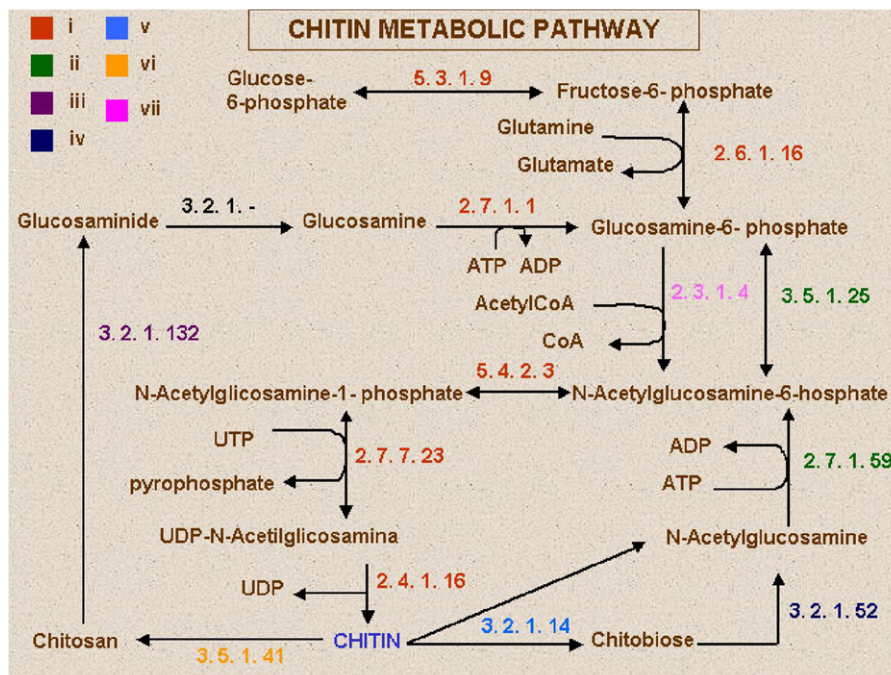

Fig. 1. Reference pathway for chitin metabolism.

Table 1
Enzymes of the chitin metabolic pathway (E = Eukarya, B = Bacteria, A = Archaea; E.C. = Enzymatic classification. Numbers in parentheses after letters are the total of organismal individual sequences per domain of each protein).

Protein	E.C. number	Metabolic role	Domain (#)
Phosphoglucoisomerase	5.3.1.9	Synthesis	E(16), B(472), A(12)
glucosaminephosphate isomerase	2.6.1.16	Synthesis	E(23), B(285), A(5)
phosphoglucosamine acetylase	2.3.1.4	Synthesis	E(3)
acetylglucosamine phosphate deacetylase	3.5.1.25	Synthesis	B(170), A(6)
acetylglucosamine phosphomutase	5.4.2.3	Synthesis	E(5)
UDP-acetylglucosamine pyrophosphorylase	2.7.7.23	Synthesis	E(2), B(324), A(2)
Chitin synthase	2.4.1.16	Synthesis	E(22)
Chitinase	3.2.1.14	Degradation	E(7), B(57)
chitin deacetylase	3.5.1.41	Degradation	E(2), B(25)
Hexosaminidase	3.2.1.52	Degradation	E(3), B(235)
acetylglucosamine kinase	2.7.1.59	Activation	E(1)
Chitosanase	3.2.1.132	Degradation	B(1)
hexokinase type IV glucokinase	2.7.1.1	Activation	E(2), B(15)

2004), and some protists, such as Oomycota (Alexopoulos et al., 1996).

Chitin is synthesized by a sequence of six successive reactions: (i) conversion of Glu-6P into Fru-6-P by phosphoglucoisomerases (E.C. 5.3.1.9); (ii) conversion of Fru-6-P into GlcN-6-P by glucosaminephosphate isomerases (E.C. 2.6.1.16); (iii) acetylation of GlcN-6-P generating GlcNAc-6-P by phosphoglucosamine acetylases (E.C. 2.3.1.4), (iv) interconversion of GlcNAc- 6-P into GlcNAc-1-P by acetylglucosamine phosphomutases (E.C. 5.4.2.3) or, alternatively, by acetylglucosamine phosphate deacetylases (E.C. 3.5.1.25); (v) uridilation of GlcNAc-1-P by UDP-acetylglucosamine pyrophosphorylases (E.C. 2.7.7.23); and (vi) conversion of UDP-GlcNAc into chitin by chitin synthases (E.C. 2.1.4.16) (Mio et al., 1998; Lagorce et al., 2002).

Chitin degradation is achieved by chitinases (E.C. 3.2.1.14), either by exochitinases, which convert chitin into N-acetylglucosamine residues, or by endochitinases, which convert chitin into chitobiose, which, in turn, may be converted into N-acetylglucosamine residues by hexoaminidases (E.C. 3.2.1.52). N-acetylglucosamine residues may be activated by acetylglucosamine kinases - acetylglucosamine-6-P, restoring the precursor of the short feedback cycle of chitin metabolism. Chitin may also be deacetylated by chitin deacetylases (E.C. 3.5.1.41), converted into chitosan, which is degraded by chitosanases (E.C. 3.2.1.132) into glucosaminide, which, when converted into glucosamine, may be activated by hexokinase type IV glucokinases (E.C. 2.7.7.1), which restore the precursor of N-acetylglucosamine-6-P, Glucosamine-6-P, configuring a longer feedback cycle (Pirovani et al., 2005).

In this paper, we use the complex network approach as a theoretical and methodological tool to perform a comparative study of the enzymes related to the chitin metabolic pathway in extant organisms of the three life domains, Archaea, Bacteria, and Eukarya, and to explore how the information derived from network structure and statistics can be used to uncover and explain biological patterns.

2. Material and Methods

Our database was composed by all the protein sequences corresponding to the enzymes of the chitin metabolic pathway of completely sequenced genomes of extant organisms pertaining to the domains Archaea, Bacteria, and Eukarya from Genebank, NCBI (Benson et al., 1999), at May 19th, 2007.

Each individual protein sequence was stored in a single file containing the protein sequence itself and all the relevant associated information, such as indexers, molecular source, structural and functional information, and complete taxonomic classification of the organism from which the sequence was derived. Protein sequences were initially categorized based on two criteria: (i) E.C. number (enzymatic classification) and (ii) presence or absence in seven distinct groups, according to all possible combinations [in only one of the three domains (3), in two domains (3), or in all the three domains (1)].

Similarity comparison of all protein sequences with each other was performed by using BLAST 2.2.15 (Altschul et al., 1997), and three indexes were extracted, a similarity index (%), score (bits), and e-value (probability). Then, according to the similarity level between protein sequences, a similarity matrix was constructed and submitted to symmetrization.

The adjacency matrix, constructed based on the information of the symmetrical similarity matrix was used to generate complex networks for enzymatic classes as well as one network containing all the sequences. In each of these complex networks, a vertex represents one protein sequence. As will be discussed in the next Section, two vertices are connected by an edge when the similarity degree between the two corresponding protein sequences, measured by the BLAST software, is larger than a similarity threshold. Statistical indexes of the networks, such as degree distribution, clustering coefficient, average path-length, and edge betweenness, were evaluated from the analysis of neighborhood matrices, which present explicitly the neighborhood order associated with each node of the network (Andrade et al., 2006).

All the programs were executed on LINUX-and WINDOWS-running computers, MySQL was used as database, and scripts and auxiliary programs were written in PERL, C, and FORTRAN 77. PAJEK (Batagelj et al., 2003) was used to generate network images.

3. Results and Discussions

A total of 1695 protein sequences corresponding to the 13 enzymes of the chitin metabolic pathway were retrieved from Genebank (Table 1). Only proteins from completely sequenced organisms were used in order to guarantee the retrieval of all possible isoforms of the 13 proteins in each sampled organism. The remarkably higher numbers of bacterial records in some protein types reflected the fact that there are much more completely sequenced organisms of the domain Bacteria in comparison with the other two domains, Archaea and Eukarya.

Although sequences from eukaryotic representatives were not found in two of the protein groups (E.C. 3.5.1.25 and E.C. 3.2.1.132), this does not mean that they are absent from eukaryotic organisms, but simply reflects that they were not found in the completely sequenced eukaryotic organisms until the date in which we downloaded the database used in the present work. The whole biochemical repertoire to carry out chitin synthesis and degradation, even in the absence of these two proteins, is only found in eukaryotic organisms.

As previously pointed out, the protein network was set up with the help of the concept of protein similarity, which can be quantified by the BLAST software. This software provides three distinct measures of the proximity between the proteins: similarity, e-value, and score. The results in this work are primarily based on the similarity value, which varies in the range [0,100], in an ascending way, according to how close two proteins are.

Each pair i,j of proteins in the selected set was probed by the BLAST software, in a directed way. This results in an evaluation of the protein similarity for each of the pairs (i,j) and (j,i). These results were stored in a similarity matrix S.

(a)

(b)
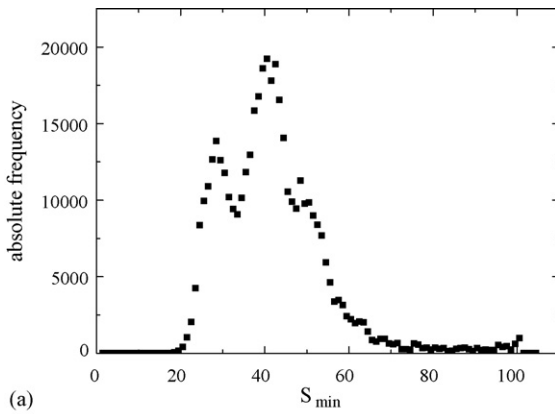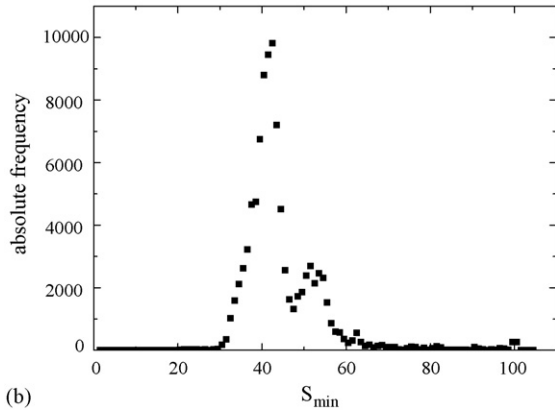
Fig. 2.
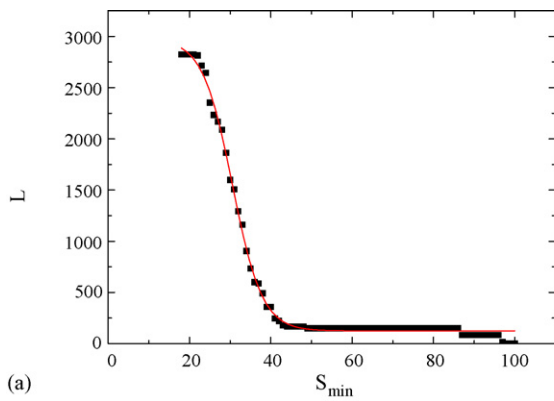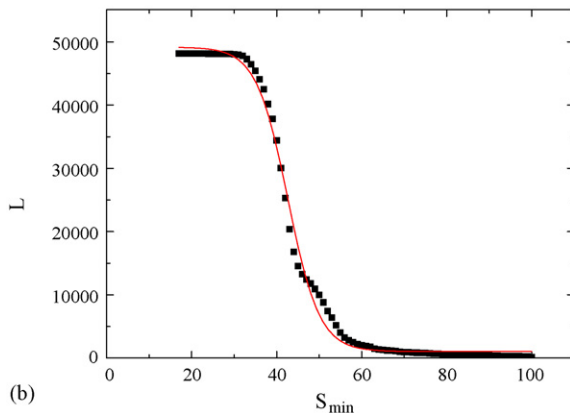
(a)

(b)
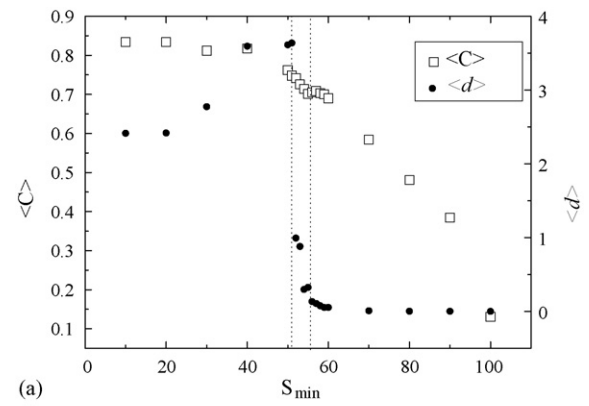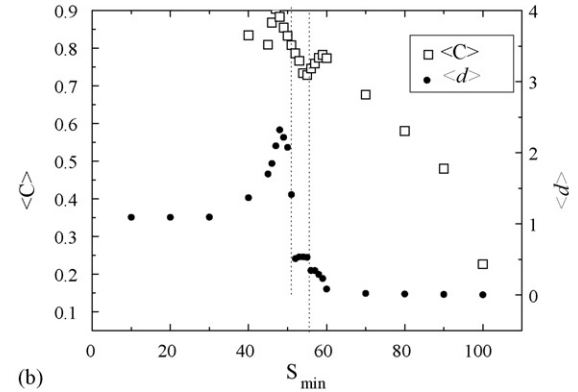
Fig. 4.

Several networks were set up and analyzed based on the matrix S. We selected different threshold values S_{min}, and constructed a network $R(S_{min})$ by imposing the following condition: set an edge between nodes i and j if, and only if, $S_{ij} \geq S_{min}$.

The data obtained indicates a slight degree of asymmetry between components S_{ij} and S_{ji}. The probability to find an asymmetry of the components of an adjacency matrix with $40 \leq S_{min} \leq 60$ (the more interesting region, according to our study) was smaller than 1%. Then, the use of an undirected network was justified. Consequently, the matrix was made symmetrical based on the highest value of a pair S_{ij} and S_{ji}.

In Fig. 2, we show the frequency distribution of similarity degrees for the whole set of proteins and for the subset of UDP-acetylglucosamine pyrophosphorylase (E.C. 2.7.7.23), which was

(a)

(b)

Fig. 3.

Fig. 5.

Fig. 6.

thoroughly explored by us in a case study also reported here. Both of them show double peaks and the mean value is in the interval (40–50).

In Fig. 3, we show the results for the number of edges L of networks constructed by the method explained above as function of S_{min}, again for the whole set of proteins and for the subset of UDP-acetylglucosamine pyrophosphorylase (E.C. 2.7.7.23). We adopted a sigmoidal function since it properly fitted the data, and subsequently, evaluated its inflexion point.

In Table 2, we show, for some subsets of proteins of the chitin metabolic pathway, the similarity mean value $\langle S \rangle$ and the inflexion point of the sigmoid curve. The correlation between these two values is very impressive, and they are related to the fact that perturbations in the value of S_{min} produce a major effect, through the removal or inclusion of a large number of edges precisely in the vicinity of the mean value of similarity.

We will now turn to the discussion of our results for some network measures, more specifically, for the values of $\langle C \rangle$ and $\langle d \rangle$, for the whole protein set and for the subset of UDP-acetylglucosamine pyrophosphorylase (E.C. 2.7.7.23) (Fig. 4a, b).

Table 2
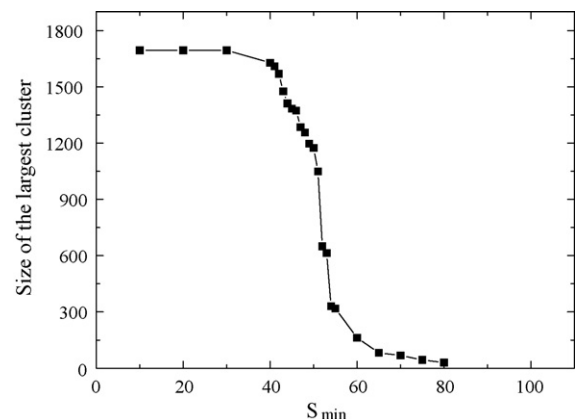Relationship between similarity mean value $\langle S \rangle$ and the inflexion point of the sigmoid curve for some subsets of enzymes of the chitin metabolic pathway.

Protein	$\langle S \rangle_s$	Inflexion point
UDP-acetylglucosamine pyrophosphorylase	41.27	42.66
Hexosaminidase	32.48	34.27
Hexokinase_type_IV_glucokinase	33.38	34.26
Acetylglucosamine_phosphate_deacetylase	32.44	32.64
Chitinase	31.66	31.79
Glucosaminephosphate_isomerase	38.55	38.44

These results show first, for $S_{min} < 40\%$, a continuous change in $\langle C \rangle$ and $\langle d \rangle$. In this region, the networks consist of single clusters, and the removal of some edges due to the increase of S_{min} does not change much of their topologies. $\langle C \rangle$ remains almost constant, while $\langle d \rangle$ increases smoothly as, on average, a larger number of steps is required to connect a pair of nodes. For $S_{min} > 40\%$, some isolated nodes and small clusters begin to appear, but the largest cluster still dominates the scene. For S_{min} ranging from 51 to 54, a sudden transition in network properties occurs, what suggests to call an interval of critical values of S_{min}. This transition can be revealed by the sharp decrease of $\langle d \rangle$, which is related to a division of the network into two large clusters, each with roughly half the number of nodes of the previously largest cluster. At the same time, we note that $\langle C \rangle$ remains large, indicating that, inside each of the clusters, the nodes continue to be highly interconnected. Nevertheless, the effect of network splitting is also revealed by the change in the derivatives of both $\langle C \rangle$ and $\langle d \rangle$ as function of S_{min}, as it may be seen in Fig. 4.

These results are corroborated by the findings presented in Fig. 5, where we show the size of the largest cluster as a function of S_{min}, for the networks constructed with the whole set of proteins.

For the purpose of obtaining the desired phylogenetic classification of the organisms in the considered data basis, it is important to consider those networks that are close to the region of critical values of S_{min}, i.e., at those values where the network topology changes abruptly. The disruption of the whole network into clusters that characterizes the critical region allows us to identify the distinct communities that are entailed in the set of considered organisms. As our results will show, such communities are related to the individual or composite classes of organisms to which the proteins belong. The fact that the critical region extends over a finite interval of S_{min} value (instead of a single point, as one might expect) is related to the fact that the different communities do not split from the largest cluster at once.
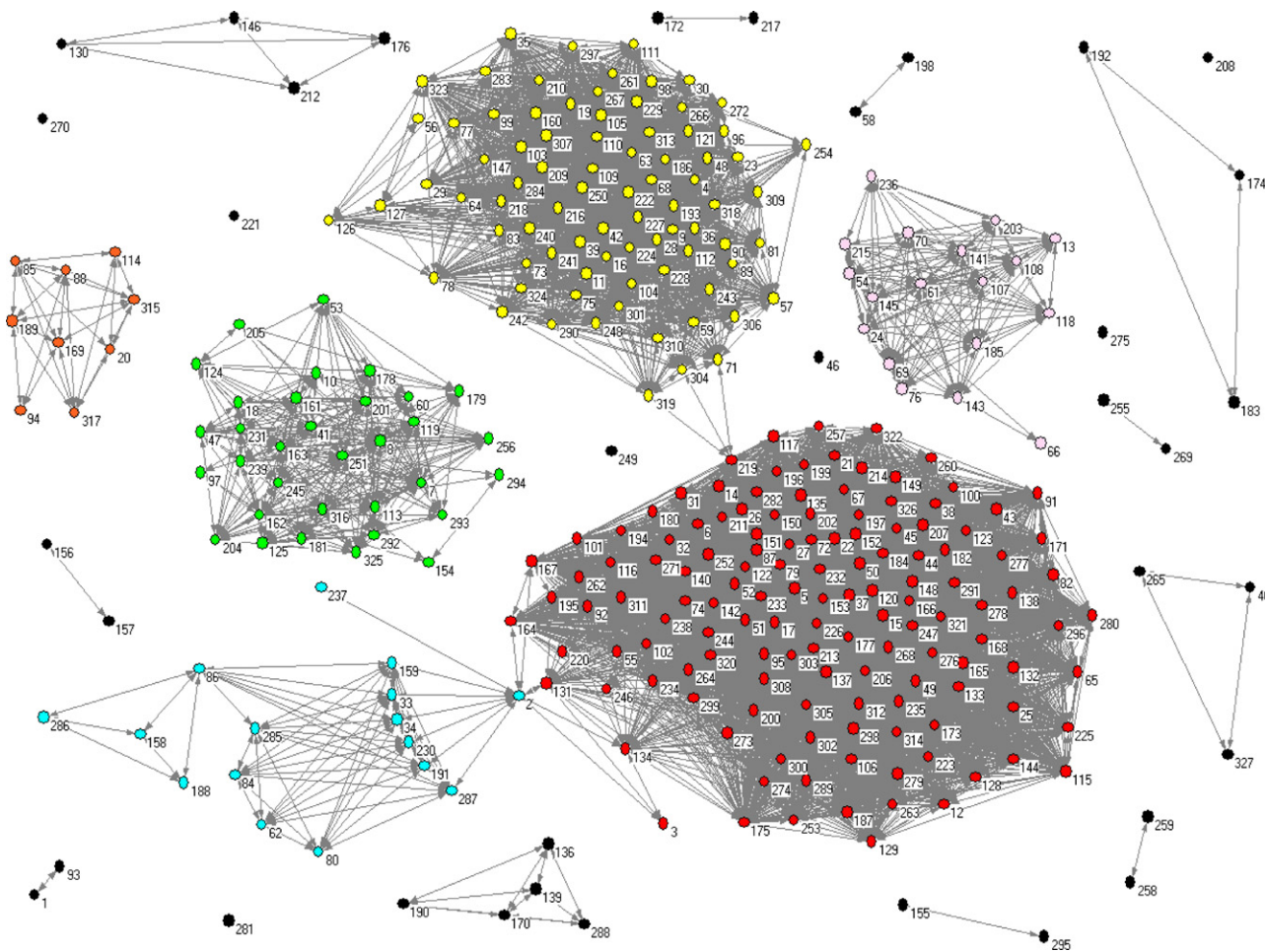
Fig. 7.

This fact can be further illustrated by Figs. 6 and 7, which display finer details of the network built for the subset of UDP-acetylglucosamine pyrophosphorylase (E.C. 2.7.7.23). Comparing Fig. 6 with Fig. 7, it is clear that, while communities (modules) can be clearly revealed when S_{min} = 51%, there are no communities (modules) when S_{min} = 40%. This confirms that the modular structure is not ubiquitously revealed by the networks built for all S_{min} values, but it is rather restricted to a critical range of such values.

In Fig. 8, we show a dendrogram for the subset of UDP-acetylglucosamine pyrophosphorylase (E.C. 2.7.7.23), at S_{min} = 51%, following the framework proposed by Newman and Girvan (2004). It shows how the elimination of the edge of largest betweenness degree also allows the identification of community structure, in the same network at S_{min} = 51%, using the protein sequences of UDP-acetylglucosamine pyrophosphorylase (E.C. 2.7.7.23).

The neighborhood matrix (Fig. 9) of the network for the subset of UDP-acetylglucosamine pyrophosphorylase (E.C. 2.7.7.23) at S_{min} = 51% not only shows once again the modular structure of the network, but also clearly depicts how far the retrieved communities (modules) are to each other.

When we cross these findings derived from the complex network approach with taxonomic and phylogenetic data, sound biological information can be promptly retrieved, even in the absence of any previous knowledge about the biological issues at stake. The modules that can be discerned at S_{min} = 51% correspond in a clear and rather precise manner to bacterial phyla and/or classes. The reason why this analysis could be readily carried out in the case of the domain Bacteria lies in the fact that most of the protein sequences in the database are derived from this domain.

Notice that, in the subset of UDP-acetylglucosamine pyrophosphorylase (E.C. 2.7.7.23), we obtained 324 bacterial sequences, and only two sequences each of Eukarya and Archaea (Table 1).

Community C1 is composed by 16 nodes, 14 of which are protein sequences from representatives of the phylum Cyanobacteria, the only bacterial group that comprises organisms capable of carrying out oxygenic photosynthesis. One of the nodes corresponds to a
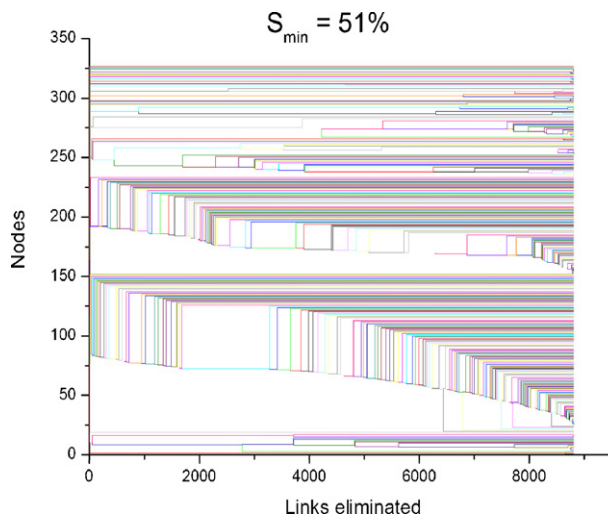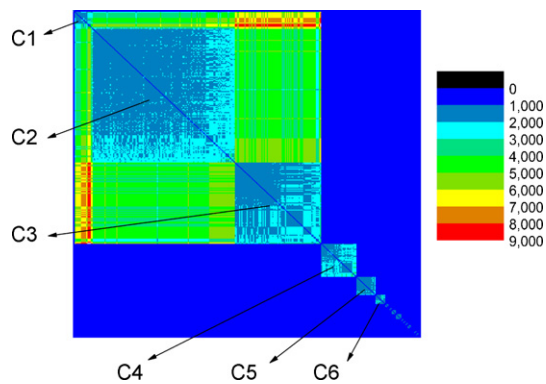

Fig. 8.

Fig. 9.

sequence from a species of Deinococcus-Thermus, a Gram-negative diderm bacterial group of extremophiles that is closely related to Cyanobacteria (Gupta, 2001).

Community C2 contains 134 nodes and, among them, 132 are sequences from species of both β-and γ-Proteobacteria, which is considered to be more closely related to each other than to any other proteobacterial class (Gupta and Sneath, 2007). Community C3 is entirely constituted by 76 sequences from Firmicutes species, low G + C Gram-positive monoderm bacteria. Community C4 contains 26 vertices, of which 24 are sequences from the presumed monophyletic group of α-Proteobacteria. Community C6 comprises only nine sequences from the putative monophyletic group of ε-Proteobacteria (Gupta and Sneath, 2007). Finally, community C5 is entirely formed by sequences from Actinobacteria, high G + C Gram-positive monoderm bacteria. Minor incongruences in the grouping of nodes in the aforementioned communities (modules) are related to the underrepresentation of eukaryal and archeal sequences in the studied data set.

To close this Section, it is important to recall once again that the results shown in Figs. 7–9 would be completely different if values of S_{min} were not chosen in the critical region. The corresponding graphs below the critical region of S_{min} do not present modular structure. Therefore, the power of the analysis reported here to resolve the relationships between taxonomic groups from a network approach to the comparison of protein sequences crucially depends on choosing the appropriate values of similarity, in which the modularity of the network occurs. The critical region was established based on the behavior of $\langle d \rangle$, $\langle C \rangle$, and the size of the largest cluster shown in Figs. 4 and 5. These measures are related to the behavior of the Euclidian distance between the neighborhood matrices of successive values of similarity that determines very precisely the critical region, as was introduced in a recent paper by some of us (Andrade et al., 2009). However, like other phylogenetic analyses, this method does not set up a completely deterministic criterion, in the sense that the number of communities is not determined by an automated manner. As already discussed, the fact that the distinct modules do not detach from the largest cluster at a single value of S_{min} prevents us to find a uniquely defined critical value where disruption occurs.

The modular structure of the network, revealed by the present method, agrees with the phylogenetic relationships of the organisms with high reliability, since 99.6% of the nodes had a neighborhood from the same taxonomic group.

4. Conclusions

The results obtained through the application of a complex network approach to the comparative analysis of protein sequences related to the chitin metabolic pathway in extant organisms suggest that this method can indeed retrieve sound biological information even in the absence of previous knowledge about the systems under analysis. Therefore, it can be used as a powerful tool to reveal relationship patterns among both organisms we have knowledge about and organisms about which we do not have much information available. The fact that the algorithm used to identify the communities (modules) and, consequently, the modularity of the network does not use any *a priori* biological information suggests, in sum, that this method can be applied as a new way of inferring molecular phylogenies, in a high-throughput and automatic manner.

The next steps in our research program will be the application of the method presented here to new sets of protein sequences, and the comparison of the results obtained with the outcome of other methods used to analyze phylogenetic relationships based on molecular data in order to reveal their advantages and limitations. Although this is a huge task, the results of which deserve to be discussed in another work, it is possible to advance that preliminary results for a much smaller data set than that used herein are promising. For such case, we have found that the phylogenetic classification between the proposed method agree with those based on the Bayesian, distance, maximum likelihood, and maximum parsimony criteria to an extent of, respectively, 46%, 51%, 51% and 33% of the total number of organisms in the set. In half of the comparisons, the agreement between the classification based on complex network approach and phylogenetic methods are above 50%, as well as between each of the four phylogenetic methods. Interestingly, the highest percentage of agreement always involves the maximum likelihood method. Since the agreement between the classification among the quoted methods using the same data set are about the same order of magnitude, varying from 17% to 66%, we may conclude that the proposed methodology already shows to be reliable.

References

Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97.
Alexopoulos, C., Mims, C., Blackwell, M., 1996. Introductory Mycology. Wiley & Sons, New York.
Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Andrade, R.F.S., Miranda, J.G.V., Petit Lobao, T., 2006. Neighborhood properties of complex networks. Phys. Rev. E 73, 046101.
Andrade, R.F.S., Pinho, S.T.R., Petit Lobao, T.C., 2009. Identification of community structure in networks using higher order neighborhood concepts. Int. J. Bifurc. Chaos 19, 2677–2685.
Ax, P., 1996. Multicellular Animals: A New Approach to the Phylogenetic Order in Nature. Springer, Berlin.
Barabasi, A.L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113.
Batagelj, V., Mrvar, A., 2003. Pajek—analysis and visualization of large networks. In: Jünger, M., Mutzel, P. (Eds.), Graph Drawing Software. Springer, Berlin, pp. 77–103.
Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F., Rapp, B.A., Wheeler, D.L., 1999. Nucleic Acids Res. 27, 12–17.
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U., 2006. Complex networks: structure and dynamics. Phys. Rep. 424, 175–308.
Boone, C., Bussey, H., Andrews, B.J., 2007. Exploring genetic interactions and networks with yeast. Nat. Rev. Genet. 8, 437–449.
Bowmann, S.M., Free, S.J., 2006. The structure and synthesis of the fungal cell wall. Bioessays 28, 799–808.
Costa, L., da, F., Rodrigues, F.A., Travieso, G., Boas, P.V., 2007. Characterization of complex networks: a survey of measurements. Adv. Phys. 56, 167–242.
Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B., Superti-Furga, G., 2004.

Proteome survey reveals modularity of the yeast cell machinery. Nature 440, 631–636.

Gupta, R.S., 2001. The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. Int. Microbiol. 4, 187–202.

Gupta, R.S., Sneath, P.H.A., 2007. The phylogeny of Proteobacteria: relationships to other eubacterial phyla and eukaryotes. J. Mol. Evol. 64, 90–100.

Hanlon, R.T., Messenger, J.B., 1996. Cephalopod Behaviour. Cambridge University Press, Cambridge.

Lagorce, A., Berre-Anton, V., Aguilar-Uscanga, B., Martin-Yken, H., Dagkessamanskaia, A., François, J., 2002. Involvement of GFA1, which encodes glutamine–fructose-6-phosphate amidotransferase, in the activation of the chitin synthesis pathway in response to cell-wall defects in *Saccharomyces cerevisiae*. Eur. J. Biochem. 269, 1697–1707.

Merzendorfer, H., 2006. Insect chitin synthases: a review. J. Comp. Physiol. B 176, 1–15.

Mio, T., Yabe, T., Arisawa, M., Yamada-Okabe, H., 1998. The eukaryotic UDP N-acetylglucosamine pyrophosphorylases: gene cloning, protein expression, and catalytic mechanism. J. Biol. Chem. 273, 14392–14397.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113.

Pirovani, C.P., Lopes, M.A., Oliveira, B.M., Dias, C.V., Souza, C.S., Galante, R.S., Santos-Junior, M.C., Silva, B.G.M., Uetanabaro, A.P.T., Taranto, A.G., Cruz, S.H., Roque, M.R.A., Micheli, F.F.L., Gesteira, A.S., Schriefer, A., Cascardo, J.C.M., Pereira, G.A.G., Góes-Neto, A., 2005. Knowledge discovery in genome database: the chitin metabolic pathway in *Crinipellis perniciosa*. In: Proceedings of IV Brazilian Symposium on Mathematical and Computational Biology/I International Symposium on Mathematical and Computational Biology, vol. 1, E-Papers Serviços Editoriais LTDA, Rio de Janeiro, pp. 122–139.

Raven, P.H., Evert, R.F., Eichorn, S.E., 2004. Biology of Plants, 7th ed. W H Freeman & Co., New York.

Silva, E. de, Stumpf, P.H., 2005. Complex networks and simple models in biology. J. R. Soc. Interface 2, 419–430.

Stoll, G., Naef, F., 2004. Effects of perturbations on dynamical properties of biological networks. In: Guimarães, K.S., Sagot, M-F. (Eds.), Texts in Algorithms (Algorithms and Computational Methods for Biochemical and Evolutionary Networks), vol. 3, pp. 115–116.