

Bacillus Subtilis Protein Interaction Network Analysis

Olusola C. Idowu, Steven J. Lynden, Malcolm P. Young and Peter Andras
School of Computing Science, University of Newcastle upon Tyne, NE1 7RU, UK
o.c.idowu@newcastle.ac.uk

Abstract

Identifying functionally important proteins that are essential to the survival of a bacterial cell is of considerable interest in the development of new antimicrobial agents. Recent studies have shown that functionally important components in protein interaction networks may also be structurally important. We studied the protein interaction network of Bacillus subtilis to identify structurally essential proteins. Fifty-four percent of the structurally essential proteins identified by our methods were encoded by functionally essential genes.

1. Introduction

Bacteria remain one of the main threats to human health due to their increasing resistance to antimicrobials. Most of the existing classes of antimicrobials were discovered by systematic screening of natural products which has produced limited results in recent years. New antimicrobial drug discovery efforts now focus on developing methods for mining bacterial genome data for antimicrobial drug targets. One method is to analyse the physical interactions between proteins in order to understand their functions and predict their importance in maintaining the functional integrity of a cellular network. Protein interactions can be detected experimentally using techniques such as the high-throughput methods [1], and can also be computationally inferred from sequenced genomes. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [2] contains predicted functional associations between proteins, derived from analysing genomic associations between the genes which encode them. Protein interaction data can be represented as a network of interacting nodes and edges in which the nodes represent proteins and the edges represent interactions between proteins.

Characteristics of the network such as connectivity, clustering, and hierarchies can be inferred directly by studying the structural properties of the network (Bu *et al.* [3]). The availability of proteomic data and associated modelling techniques means studies can be conducted to identify new putative protein targets for drug discovery. One of the crucial steps in achieving this goal is to find a robust way of identifying essential protein encoding genes. Different experimental studies have been conducted to identify essential protein encoding genes. Of particular interest is the Kobayashi *et al* [4] experiment which found 271 genes among $\approx 4,100$ genes of *Bacillus subtilis* to be essential.

The objective of this work is to identify ways of selecting structurally essential proteins from protein interaction networks, and to establish if these structurally essential proteins are functionally essential to the network. For this study we used the *Bacillus subtilis* protein interaction network.

2. Methods

As a consequence of the heterogeneous connectivity distribution, scale-free connection topologies [5] are characterised by a significant minority of nodes called hubs with a very high connectivity. Hubs mediate the communication of a large proportion of nodes and are therefore essential to network integrity [6]. For a given protein P the hub value represents the proteins directly connected to P . Distinct types of hub proteins are the second-order hubs. Second-order hubs proteins are protein hubs which are connected to other hubs. They are important to the structural integrity of the network as they serve as gateways to protein hubs. The bottleneck property of a protein can be described as the degree of conflux of network traffic through a given protein. We can perceive the flow through a network along all the pathways between two clusters as movement of traffic along the entire path linking two major clusters. Proteins participating in these freeway links are bottleneck proteins.

For our analysis, we used the STRING database to construct the protein interaction network for *Bacillus subtilis*. Out of 5017 protein interactions among 801 proteins available in the STRING database as of 1, February, 2004 a total of 75 proteins (9.4%) were encoded by essential genes.

3. Results

Our study shows that there is a strong link between the network structures of the *Bacillus subtilis* protein interaction network and the functionality of proteins within the network. The top 1% hub, second-order hub, and bottleneck proteins predicted by our methods are listed in Table 1. Proteins encoded by essential genes of *Bacillus subtilis* [4] are highlighted with grey in Table 1. Only one protein (Sensor protein yycG, Swiss-Prot reference: Q45614) out of the top 1% of the hub proteins was encoded by an essential gene of *Bacillus subtilis*. However, we found that all the proteins in the top 3% of the second-order hub ranking were encoded by essential genes. The top 3% of the second-order hub proteins consist of hub proteins with 20 or more connections. Most of the proteins in the top 1% were ribosome proteins (Table 1). Ribosome proteins are involved in the complex process of mRNA translation to form a polypeptide chain. Four proteins out of the top 1% of the bottleneck proteins were encoded by essential genes (Table 1). Each of the bottleneck proteins listed in Table 1 are connected to two other proteins. Their structural importance could not be detected by looking only at their direct connections only. This highlights the importance of considering different aspects of network structure to determine structurally essential proteins. Two of the proteins in the top 1% list were hypothetical proteins. Hypothetical proteins are proteins that would appear to be encoded by genes which have been identified through the analysis of DNA sequences but have not been characterised with certainty.

4. Conclusions

In this study we have shown that structurally essential proteins are likely to be encoded by functionally essential genes. We developed methods for selecting structurally important proteins in protein interaction networks, and applied these techniques to analyse the protein interaction network of *Bacillus subtilis*. We compared our protein selection methods with random protein selections and show that our selection is statistically significant. In particular the top 3% of the second-order hub proteins (26 proteins) were encoded by essential *Bacillus subtilis* genes.

5. References

- [1] Cagney, G., Uetz, P., and Fields, S., "High-throughput screening for protein-protein interactions using the two-hybrid assay", *Meth. Enzymol*, 328, 2000, pp. 3-14
- [2] von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snel, B., "STRING: a database of predicted functional associations between proteins", *Nucleic Acids Res.* 31(1), 2003, pp. 258-261
- [3] Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., *et al.*, "Topological structure analysis of the protein-protein interaction network in budding yeast", *Nucleic Acids Research*, 31(9), 2003, pp. 2443-2450
- [4] Kobayashi, k., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., *et al.*, "Essential *Bacillus subtilis* genes", *Proc. Natl. Acad. Sci. USA*, 100, 2003, pp. 4678-4683
- [5] Barabasi, A. L. and Albert, R., "Emergence of scaling in random networks", *Science*, 286, 1999, pp. 509-512.
- [6] Jeong, H., Mason, S.P., Barabasi A.L., and Oltvai, Z.N., "Lethality and centrality in protein networks", *Nature*, 411, 2001, pp. 41-42

Table 1: This table contains the list of structurally essential proteins of *Bacillus subtilis* as determined by our method. The functionally essential proteins are highlighted in grey. Thirteen out of twenty four (54%) of the selected proteins are coded by essential genes.

	Swiss-Prot reference	Gene name	Fatality	Protein Names
Hub proteins	Q45614	YYCG	Essential	Sensor protein yycG
	P35164	RESE	Non-essential	Sensor protein resE
	P16497	KINA	Non-essential	Sporulation kinase A
	O31661	YKRQ	Non-essential	YKRQ protein
	P39764	KINC	Non-essential	Sporulation kinase C
	P23545	PHOR	Non-essential	Alkaline phosphatase synthesis sensor protein
	O07527	YHCY	Non-essential	Hypothetical protein yhcY
	O07014	RSBP	Non-essential	Phosphoserine phosphatase rsbP
Second order hub proteins	P16336	SECY	Essential	Preprotein translocase secY subunit
	P42920	RPLC	Essential	50S ribosomal protein L3
	P42921	RPLD	Essential	50S ribosomal protein L4
	P42924	RPLW	Essential	50S ribosomal protein L23
	P42919	RPLB	Essential	50S ribosomal protein L2
	P42060	RPLV	Essential	50S ribosomal protein L22
	P21465	RPSC	Essential	30S ribosomal protein S3
	P46899	RPLR	Essential	50S ribosomal protein L18
Bottleneck proteins	P05652	GYRB	Essential	DNA gyrase subunit B
	Q45066	PARC	Essential	Topoisomerase IV subunit A
	P05653	GYRA	Essential	DNA gyrase subunit A
	Q59192	PARE	Essential	Topoisomerase IV subunit B
	O07622	YHF W	Non-essential	Hypothetical protein yhfW
	P46912	QCRB	Non-essential	Menaquinol-cytochrome c reductase cytochrome b subunit
	P46913	QCRC	Non-essential	Menaquinol-cytochrome c reductase cytochrome b/c subunit
	O34655	YTHA	Non-essential	YTHA