The Topological Properties of Virus-Human Protein Interaction Networks

Jinying Tan[1, 2]   Xiufen Zou[1, *]

1 School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
2 College of Science, Huazhong Agricultural University, Wuhan 430070, China

Abstract In the research of biological networks, it is an important problem how to use the quantity to identify essential protein-protein interactions. In this paper, a new definition "representative value of networks", denoted as " RV ", is presented and applied to analyze the topological properties of three protein-protein interaction networks. From calculation we find that the relationship between human protein-protein interactions is highly clustered, for example, the percentage of value of the top 20% proteins of representative value from high to low accounts for 71.96%. The virus-human protein interactions (i.e. virus attacks human) are selective. And the attack of the virus also has a high aggregation. These results show that the proposed new definition is reasonable and can be considered as an important index for analyzing the topological properties of biological networks.

Keywords Protein-protein interaction network; Representative value (RV); Virus; degree

1 Introduction

Protein-protein interaction (PPI) plays a central role in many biological processes [1, 2], which not only is the base of normal physiological processes such as DNA replication, transcription, translation, metabolism, signal transduction and cell cycle control [3, 4], but also plays an important role in pathological processes [5, 6, 7]. In the viral infection of the host, it is also essentially expressed as interactions of viral proteins and host proteins, inhibiting activity of host proteins or denaturizing host proteins.

With the development in the high-throughput protein interaction detection technology such as the Yeast Two-Hybrid (Y2H) technology [8] and the tandem affinity purification - mass spectrometry technique (TAP-MS) [9], the protein-protein interaction networks of many species as data are uncovered, allowing the understanding of the process of life activities from the system-level of the protein-protein interaction networks [10, 11, 12, 13]. Access to the interaction networks of virus proteins and host proteins can also enable us to understand the mechanism of virus infecting host from the network-level and find new ways to solve the problems such as the toxicity for the virus which became stronger or

---

* Corresponding author, Email: xfzou@whu.edu.cn

weaker, differences among species, and the identification of targets for treatment and tumor pathogenesis.

Calderwood et al. [14] undertook a study for the interaction network of the Epstein-Barr virus (EBV) proteins and human proteins, from a broader perspective, EBV genes can be classified into two evolutionary classes, a class called the "core" genes, which contains conservative genes of the herpes viruses and their sub-classes, and the other for "non-nuclear" genes. Prior to that, the interaction networks for viral proteins have been reported for vaccinia virus (VV) [15], Varicella-Zoster virus (VZV) and Kaposi's sarcoma-associated herpes virus (KSHV) [16] by the yeast two-hybrid system. B de Chassey, et al. [17] performed a proteome-wide mapping of interactions between hepatitis C virus (HCV) and human proteins by Y2H and literature mining, to provide a comprehensive view of the cellular infection.

Recent researches showed that the attacked proteins in the virus-Human protein interaction networks have "Hub" characteristics and the higher average connectivity, corresponding to play an important role in their life activities [14, 15, 16, 17, 18] For the above different viral infection networks, it is obviously a significant problem to study whether have some common features in network structure for attacking the host proteins of viral proteins. To reflect some common features of these networks, in this study, we propose a new definition "representative value of networks" based on the degree of networks, denoted as "RV" and conduct a comparative analysis of viral infection in the networks.

2   Data sets and Methods

2.1   Data sets

In this study, we use three data sets, the human protein-protein interaction network which is composed of 44223 non-redundant PPIs between 9520 different human proteins, the HCV-human protein interaction network which contains 481 HCV-human protein interactions between 11 virus proteins and 338 human proteins and the EBV-human protein interaction network which includes 173 different EBV-human protein interactions between 40 different EBV proteins and 112 human proteins, which are adapted from the literature [17] and [14] respectively.

2.2   Methods

The function of a protein can not be completed independently in the cell. Any life course is collaboratively completed by many proteins and other molecules together. In the interaction network, each protein can be seen as a node and each interaction as an edge, which forms a linked network of protein interactions. Thus we can research the topological properties of protein interaction networks from the perspective of graphic theory [19, 20, 21, 22, 23, 24]. To better compare different virus - human protein interaction networks, we introduce a concept "representative value".

Protein-protein interaction network can be defined as a simple undirected graph $G = (V, E)$, which has n vertices and m edges. The graph vertex represents a protein

and the edge represents the interaction between two proteins.

Let $I_i = \{ j \mid (v_j, v_i) \in E \}$ $(i = 1, 2, ..., n)$ and $\mid I_i \mid = d_i$, i.e. the degree of node v_i is d_i. If $\mid I_i \mid = d_i$, then node v_i is claimed to vote to d_i nodes and the value of each vote of v_i is $1/d_i$. The representative value (RV) of node v_i is defined as $RV_i = \sum_{j \in I_i} 1/d_j$. Obviously, the mean of representative value of all nodes is 1, i.e. $\dfrac{\sum_{i=1}^{n} RV_i}{n} = 1$.
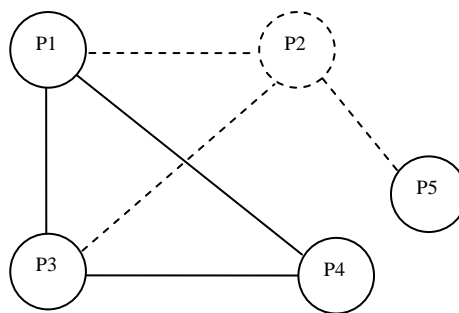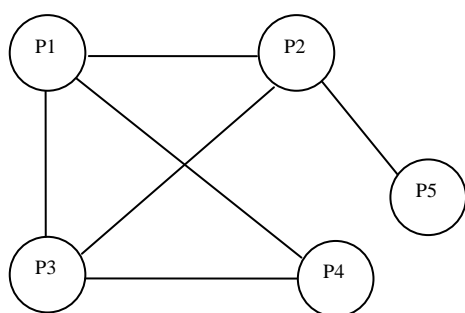

Fig.1 An example of a simple network.   Fig.2 The deletion of P2.

We give a simple network as an example shown in Fig.1. The degree of P1, P2, P3, P4 and P5 is 3, 3, 3, 2 and 1, respectively. In order to reflect the various protein importance, we can vote and score to each protein: because the degree of P1 is 3, then the value of each vote of P1 is 1 / 3, and the value of each vote of P2, P3, P4 and P5 is 1 / 3, 1 / 3, 1 / 2 and 1 / 1, respectively. If Pi votes for Pj, then Pj obtains the corresponding value of the vote. At the same time, the representative value of P1 is P2 + P3 + P4 = 1 / 3 +1 / 3 +1 / 2 = 1.1667. The rest may be deduced by analogy, the representative values of P2, P3, P4 and P5 are P1 + P3 + P5 = 1 / 3 +1 / 3 +1 / 1 = 1.6667, P1 + P2 + P4 = 1 / 3 +1 / 3 +1 / 2 =1.1667, P1 + P3 = 1 / 3 + 1 / 3 = 0.6667 and P2 = 1 / 3 =0.3333, respectively. The order of these nodes by " RV " is P2> P1 = P3> P4> P5, which is different from the order with the degree of nodes (P2 = P1 = P3> P4> P5). We can find that the node with the same degree possibly has different RV. In this simple network, it is obvious that three nodes P1, P2 and P3 have the same importance according to the degree. In fact, P2 is more important than P1 and P3 because if P2 is inhibited and lost its function, then the impact is not just P2 itself, and P5 will also be affected, whose function will be lost because at this time the only interaction with other nodes is cancel (Fig.2). Therefore, we could distinguish P2 from P1 and P3 by calculating the RV. In addition, this definition

also has an advantage, that is, the mean of the RV of all nodes in the whole network is always 1. When the virus proteins select to interact with the human proteins which have high representative value, then the mean of the RV of the sub-network of virus-human interaction will be higher than 1. Accordingly, a higher " RV " can be argued that the connectivity of the node in the network is more intense and the node is often more important.

3   Results

Here we calculate the RV of above three networks and list the names of the top 5 proteins of RV from high to low. We also calculate the percentage of total value of the top 20% proteins of RV from high to low and the results are shown in Table 1, Table 2 and Table 3, respectively.

The obvious feature of human protein-protein interaction network is highly clustered, for example, the percentage of value of the top 20% proteins of representative value from high to low accounts for 71.96%, which is similarity to the result of "Hub" characteristics [18].

From Table 1, the protein SLC2A4 has the highest RV, which is probably related with its structure and function. SLCA2A4, a kind of protein involved in the glucose transport, is widespread in Skeletal and cardiac muscles, brown and white fat. The protein localizes primarily to the perinuclear region, undergoing continued recycling to the plasma membrane where it is rapidly reinternalized. The dileucine internalization motif is critical for intracellular sequestration. [25] From Table 2 and 3, we are lucky to find that HCV and EBV don't interact with the human protein SLC2A4, defects in which may be a cause of noninsulin-dependent diabetes mellitus (NIDDM), or a cause of certain post-receptor defects in NIDDM, Otherwise, hepatitis and herpes disease would be more serous than we thought. [26, 27, 28]. The second highest RV protein ATXN1, which Locates cytoplasm or nucleus, also expresses widely throughout the body. ATXN1 may be involved in RNA metabolism, Defects in ATXN1 are the cause of spinocerebellar ataxia type 1; also known as olivopontocerebellar atrophy I (OPCA I or OPCA1). Spinocerebellar ataxia is a clinically and genetically heterogeneous group of cerebellar disorders. [29, 30, 31].

YWHAG, UBQLN4, etc., also the characteristics of the protein in the body exists in most organizations, to interact with many proteins, some biological features common to complete, and therefore high RV value [31, 32]. Of course, low RV does not mean that the protein is not important, only because of its structure or its expression in some exceptional cases to complete certain specific features with other proteins together. The protein PGCP, a kind of plasma glutamate carboxypeptidase, for example, which RV is only 0.0012, up-regulated in the majority of hepatitis C virus-associated hepatocellular carcinoma [33]. Perhaps the low RV could also be the false negative test result because the interaction with other proteins is not detected in these experiments.

The virus-human protein interactions (i.e. virus attacks human) are selective, and the virus protein interacts often more easily with the high RV of human protein (such as human protein interaction with HCV, the mean of RV for HCV-human

Table 1 Results of human-human protein interaction network

Top 5 proteins of representative value from high to low	Protein name	SLC2A4	ATXN1	YWHAG	UBQLN4	GRB2
	Value	283.2470	60.1099	52.4070	49.7931	46.2145
Top 5 proteins of representative value from low to high	Protein name	KBTBD10	PGCP	PGRMC2	PECI	CRTAP
	Value	0.0012	0.0012	0.0012	0.0012	0.0012
Sum of RV of all human proteins	9520	Representative mean	1	the percentage of total value of the top 20% proteins of representative value from high to low	71.96%

Table 2 Results of HCV-human protein interaction network

Top 5 proteins of representative value from high to low	Protein name	UBQLN4	GRB2	TGFBR1	TP53	BCL6
	Value	49.7931	46.2145	36.2796	27.0501	25.8791
Top 5 proteins of representative value from low to high	Protein name	GAA	PRRC1	FAM120B	RSHL2	TRIOBP
	Value	0.0012	0.0012	0.0055	0.0057	0.0080
Sum of RV of all infected human proteins	971.7208	representative mean	2.8749	the percentage of value of the top 20% proteins of representative value from high to low	74.40%

Table 3 Results of EBV-human protein interaction network

Top 5 proteins of representative value from high to low	Protein name	MDFI	TRAF2	PLSCR1	VIM	PPP1CA
	Value	32.3726	25.1728	15.3683	15.2561	11.0488
Top 5 proteins of representative value from low to high	Protein name	DKK3	EGFL7	CCDC14	TXNDC11	TRAF3IP3
	Value	0.0068	0.0141	0.0230	0.0341	0.0385
Sum of RV of all infected human proteins	264.2585	Representative mean	2.9039	the percentage of value of the top 20% proteins of representative value from high to low	71.75%

protein interaction sub-network is 2.8749 and the mean of RV for EBV-human protein interaction sub-network is 2.9039). TGFBR1, for example, which has higher RV, forms a receptor complex consisting of two type II and two type I transmembrane serine/threonine kinases [34]. Defects in TGFBR1 will lead to vascular disease [35, 36, 37], and will enable Hepatitis patients face yellow and thin, which is consistent with Hepatitis symptoms. Although the RV of GAA or PRRC1 is very low (0.0012), they are essential to liver function and are expressed in liver. Defects in them will cause to liver disease, one of which is hepatitis [38, 39, 40].

The attack of the virus also has a high aggregation, such as human protein interaction with HCV, of which 20% of the protein its RV accounts for 74.40%, and interaction with the EBV, of which 20% of the protein its RV accounts for 71.75%.

4 Conclusions

From these results, we can see that the purpose of the protein-protein interaction is strong. When the virus invades the body, the virus proteins often interact with the human proteins which have high RV, and inhibit the activity of these host proteins or alter their activity. Thus this causes to degrade their corresponding functions of these proteins. If the proteins which have high RV and tend to play more important role in normal physiological processes lose their functions, the body is possibly in the disease state. In conclusion, the proposed new definition is reasonable and can be considered as an important index for analyzing the topological properties of biological networks.

Acknowledges

References

[1] Eisenberg D, Marcotte E M, Xenarios I, et al. Protein function in the post-genomic era. Nature, 2000, 405(6788): 823~826

[2] Auerbach D, Thaminy S, Hottiger M O, et al. The post-genomic era of interactive proteomics: Facts and perspectives. Proteomics, 2002, 2(6): 611~623

[3] Wang J. Protein recognition by cell surface receptors: Physiological receptors versus virus interactions. Trends Biochem Sci, 2002, 27(3): 122~126

[4] Kone B C, Kuncewicz T, Zhang W, et al. Protein interactions with nitric oxide synthases: Controlling the right time, the right place, and the right amount of nitric oxide. Am J Physiol Renal Physiol, 2003, 285(2): F178~F190

[5] Cohen F E, Prusiner S B. Pathologic conformations of prion proteins. Annu Rev Biochem, 1998, 67: 793~819.

[6] Loregian A, Marsden H S, Palu G. Protein-protein interactions as targets for antiviral chemotherapy. Rev Med Virol, 2002, 12(4):239~262

[7] Selkoe D J. The cell biology of beta-amyloid precursor protein and presenilin in

Alzheimer's disease. Trends Cell Biol, 1998, 8(11): 447~453

[8] Uetz, P, Hughes, R E. Systematic and large-scale two-hybrid screens[J]. Curr Opin Microbiol, 2000, 3: 303-308.

[9] Gavin, A C, Basche M, Krause R, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes[J]. Nature, 2000, 415: 141~147.

[10] Jeong, H., Mason, S. P., Barabási, A.-L., Oltvai, Z. N., Lethality and centrality in protein networks. Nature, 2001. 411(6833): p. 41-42.

[11] Han, J.D.J., et al., Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature, 2004. 430: p. 88-93.

[12] Gandhi, T.K.B., et al., Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nature Genetics, 2006. 38: p. 285-293.

[13] Han, J.D.J., Understanding biological functions through molecular networks. Cell Research, 2008. 18: p. 224-237.

[14] Calderwood M. A. , Venkatesan K., Xing L., et al. Epstein–Barr virus and virus human protein interaction maps. PNAS, 2007, 104(18): 7606~7611.

[15] McCraith S, Holtzman T, Moss B, Fields S., Proc Natl Acad Sci USA, 2000, 97:4879–4884.

[16] Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, Rajagopala SV, Roupelieva M, Rose D, Fossum E, et al. (2006) Science 311:239~242.

[17] Chassey, B de., et al. Hepatitis C virus infection protein network. Molecular Systems Biology, 2008, 4:230.

[18] Jin, G., et al., Hubs with Network Motifs Organize Modularity Dynamically in the Protein-Protein Interaction Network of Yeast. PLoS ONE, 2007. 2(11).

[19] Drewes G., Bouwmeester T., Global approaches to protein-protein interactions [J]. Curr Opin Cell Biol, 2003, 15(2) : 199~205.

[20] Albert, R. and A.L. Barabasi, Statistical mechanics of complex networks. Reviews of Modern Physics, 2002. 74(1): p. 47-97.

[21] Newman, M.E.J., The Structure and Function of Complex Networks. SIAM Review, 2003. 45(2): p. 167-256.

[22] Kitano, H., Biological robustness. Nature Reviews Genetics, 2004. 5(11): p. 826-837.

[23] Song, C., S. Havlin, and H.A. Makse, Self-similarity of complex networks. Nature, 2005. 433(7024): p. 392-395.

[24] Costa L. Da F., et al., Characterization of complex networks: A survey of measurements. Advances In Physics, 2007. 56(1-2): p. 167-242.

[25] Verhey K.J., Birnbaum M.J. (1994). A Leu-Leu sequence is essential for COOH-terminal targeting signal of GLUT4 glucose transporter in fibroblasts. J. Biol. Chem. 269:2353-2356.

[26] Kusari J., Verma U.S., Buse J.B., Henry R.R., Olefsky J.M. (1991). Analysis of the gene sequences of the insulin receptor and the insulin-sensitive glucose transporter (GLUT-4) in patients with common-type non-insulin-dependent diabetes mellitus. J. Clin. Invest. 88:1323-1330.

[27] Choi W.H., O'Rahilly S., Buse J.B., Rees A., Morgan R., Flier J.S., Moller D.E. (1991). Molecular scanning of insulin-responsive glucose transporter (GLUT4) gene in NIDDM subjects. Diabetes 40:1712-1718.

[28] O'Rahilly S., Krook A., Morgan R., Rees A., Flier J.S., Moller D.E. (1992). Insulin receptor and insulin-responsive glucose transporter (GLUT 4) mutations and polymorphisms in a Welsh type 2 (non-insulin-dependent) diabetic population. Diabetologia 35:486-489.

[29] Banfi S., Servadio A., Chung M.-Y., Kwiatkowski T.J. Jr., McCall A.E., Duvick L.A., Shen Y., Roth E.J., Orr H.T., Zoghbi H.Y. (1994). Identification and characterization of the gene causing type 1 spinocerebellar ataxia. Nat. Genet. 7:513-519.

[30] Quan F., Janas J., Popovich B.W. (1995). A novel CAG repeat configuration in the SCA1 gene: implications for the molecular diagnostics of spinocerebellar ataxia type 1. Hum. Mol. Genet. 4:2411-2413.

[31] Horie M., Suzuki M., Takahashi E., Tanigami A. (1999). Cloning, expression, and chromosomal mapping of the human 14-3-3gamma gene (YWHAG) to 7q11.23. Genomics 60:241-243.

[32] Davidson J.D., Riley B., Burright E.N., Duvick L.A., Zoghbi H.Y., Orr H.T. (2000) Identification and characterization of an ataxin-1-interacting protein: A1Up, a ubiquitin-like nuclear protein. Hum. Mol. Genet. 9:2305-2312.

[33] Smith M.W., Yue Z.N., Geiss G.K., Sadovnikova N.Y., Carter V.S., Boix L., Lazaro C.A., Rosenberg G.B., Bumgarner R.E., Fausto N., Bruix J., Katze M.G. (2003). Identification of novel tumor markers in hepatitis C virus-associated hepatocellular carcinoma. Cancer Res. 63:859-864.

[34] Wieser R., Wrana J.L., Massague J. (1995). GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex. EMBO J. 14:2199-2208.

[35] Loeys B.L., Chen J., Neptune E.R., Judge D.P., Podowski M., Holm T., Meyers J., Leitch C.C., Katsanis N., Sharifi N., Xu F.L., Myers L.A., Spevak P.J., Cameron D.E., De Backer J.F., Hellemans J., Chen Y., Davis E.C., Webb C.L., Kress W., Coucke P.J., Rifkin D.B., De Paepe A.M., Dietz H.C. (2005). A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. Nat. Genet. 37:275-281.

[36] Ades L.C., Sullivan K., Biggin A., Haan E.A., Brett M., Holman K.J., Dixon J., Robertson S., Holmes A.D., Rogers J., Bennetts B. (2006). FBN1, TGFBR1, and the Marfan-craniosynostosis/mental retardation disorders revisited. Am. J. Med. Genet. A 140:1047-1058(2006)

[37] Matyas G., Arnold E., Carrel T., Baumgartner D., Boileau C., Berger W., Steinmann B. (2006). Identification and in silico analyses of novel TGFBR1 and TGFBR2 mutations in Marfan syndrome-related disorders. Hum. Mutat. 27:760-769.

[38] Ghaffari S.R., Sabokbar T., Tahmasebi S., Dastan J. (2006). Identification of a novel mutation in the acid alpha glucosidase gene causing juvenile form of Pompe disease in Iranian population. Submitted (AUG-2006) to the EMBL/GenBank/DDBJ databases

[39] Lin C.-Y., Shieh J.-J. (1995). Identification of a de novo point mutation resulting in infantile form of Pompe's disease. Biochem. Biophys. Res. Commun. 208:886-893.

[40] Kamakari S., Roussou A., Jefferson A., Ragoussis I., Anagnou N.P. (2005). Structural analysis and expression profile of a novel gene on chromosome 5q23 encoding a Golgi-associated protein with six splice variants, and involved within the 5q deletion of a Ph(-) CML patient. Leuk. Res. 29:17-31.