

细胞黏附分子互作网络的构建与分析*

韦晓兰**

(重庆工商大学环境与生物工程学院, 药物化学与化学生物学研究中心, 重庆 400067)

摘要 借助网络分析可对基因调控、蛋白质互作和信号转导等细胞活动进行全局和局部性质分析。以细胞黏附的蛋白质相互作用为对象, 通过数据挖掘和可视化软件构建了整合蛋白介导的黏附分子互作网络, 该分子互作网络由 156 种蛋白质通过 690 种相互作用相连, 其平均节点度为 8.66、平均聚集系数为 0.24, 平均路径长度为 2.6。黏附分子互作网络中包含数个功能模块, 这些模块涉及网络内部多种分子相互作用的启动与停止, 并进一步影响细胞的黏附、迁移和骨架组织。对黏附分子网络进行模体筛选和比较, 发现一些数量相对较少、以三元复合物为主要结构的关键模体, 同时对各网络模块和模体对细胞黏附的调控作用进行了探讨。

关键词 蛋白质相互作用, 网络分析, 系统生物学, 黏附分子
学科分类号 Q50, Q51

DOI: 10.3724/SP.J.1206.2010.00532

生命活动可以表示为一个复杂的动态变化的网络。网络中的节点是各种生物分子, 如 DNA、RNA、蛋白质, 它们之间的关联表示为边^[1-2]。系统生物学认为这种网络关系正是生命复杂性的源泉, 各种复杂的生命现象都是由网络中各节点之间的相互作用、调控关系的组合和动态变化产生的。因此, 对于生命活动的研究需要从整体出发, 研究生命网络中这些复杂的相互作用和调控关系, 而不是仅仅孤立地研究一个个网络中的节点^[3-5]。

细胞黏附是细胞通过黏着斑与其他表面连接并调节细胞与胞外基质的相互作用, 在细胞活动诸如分裂、发育、迁移和侵入等中起着重要作用。黏着斑内包含由受体蛋白、细胞骨架蛋白以及多种衔接蛋白组成的复合体, 这些蛋白质复合体使细胞能感应由胞外基质几何、物理和化学性质所决定的多种细胞外信号。例如, 细胞在二维和三维基质上有不同的生长行为, 能区分不同的胞外基质成分, 可感应不同的黏附配体密度、力学扰动和表面刚性变化^[6-8]。

通过分析细胞黏附的相关蛋白或信号转导通道, 只能局部了解细胞黏附某些响应的分子机制, 无法解释黏附“分子机器”如何感知环境信号并做

出响应。只有在整体上对细胞黏附的分子机制进行网络分析, 才能阐明细胞黏附各蛋白质相互作用和调控关系。为此, 本文利用已有的大量细胞黏附实验研究结果, 从蛋白质相互作用数据出发构建细胞黏附分子互作网络, 并从中计算筛选出与功能密切相关的蛋白质复合体或模体, 以在多层次上探索细胞黏附及其信号传导的分子基础。

1 黏附分子互作网络

细胞黏附受体蛋白中整合蛋白占绝大多数, 由整合蛋白介导的细胞黏附目前研究最为广泛, 因此本文以基于整合蛋白的细胞黏附为代表, 通过整合各种细胞以及所有整合蛋白的黏附分子数据, 构建细胞黏附分子互作网络。已发现整合蛋白黏着斑内有 90 种黏附内在蛋白, 66 种与黏附内在蛋白作用

* 国家自然科学基金(31070830), 重庆市自然科学基金(CSTC2008BB5329)和重庆市教委自然科学基金(KJ080706)资助项目。

** 通讯联系人。

Tel: 023-62769785, E-mail: weishine@ctbu.edu.cn

收稿日期: 2010-10-15, 接受日期: 2011-01-24

并影响其活性的黏附关联成分;另外,还可能不存在检测到的关联蛋白和在黏附位点发挥作用后消失的瞬态内在蛋白^[9].在黏附分子互作网络中,各种黏附分子作为节点,黏附分子之间的相互作用表示为节点之间的连接.

1.1 黏附分子互作数据库与网络构建

对已知 156 种整合蛋白黏附成分(152 种蛋白质、3 种类脂和钙离子),通过大型蛋白质-蛋白质相互作用数据库——BIND^[9]和 HPRD^[10],以及大量文献检索(Pubmed),获得这些黏附分子之间已知的大部分相互作用.各种相互作用分为结合(节点无向连接)和调控(节点有向连接)两大类,后者根据调控性质定义为活化或抑制,如磷酸化、GEF 活性和 GTP 酶活化属于活化,而抑制则包括去磷酸化、GAP 作用、泛素化以及蛋白质水解.

在黏附互作分子网络中,可将 156 个黏附分子节点分为不同的功能组:25 个衔接蛋白节点、24 个骨架蛋白节点、9 个肌动蛋白结合蛋白节点、10 个丝氨酸/苏氨酸蛋白激酶节点、3 个丝氨酸/苏氨酸蛋白磷酸酶节点、9 个酪氨酸磷酸酶节点、8 个酪氨酸激酶节点、8 个 GAPs 节点、8 个 GEFs 节点、7 个跨膜受体节点、6 个黏附蛋白节点、5 个 GTP 酶节点和 34 个其他成分节点.搜索出这些节点间有连接 690 个,其中无向连接 379 个、有向连接 311 个(活化 213 个+抑制 98 个).采用 AVIS 软件^[11-12]构建出可视化的黏附分子互作网络如图 1 所示,其平均节点度为 8.66,节点度分布如图 2 所示,与代谢网络相似遵循幂律分布^[13];平均聚集系数(节点间联系的紧密程度)为 0.24,平均路径长度为 2.6.已知细胞内蛋白质互作网络如酵母和突触后蛋白质互作网络的平均节点度分别为 3.99^[9]和 5.14^[14],与之相比,黏附分子互作网络的平均节点度即分子之间相互作用程度明显偏高,这可能是由黏附分子形成紧密的复合体(黏着斑)所致.黏附分子互作网络的这一特性使其具有很高的稳定性,即去掉多个节点分子也能保持该互作网络完整;同时节点分子的节点度越高,该节点对互作网络的功能也越重要.如节点度小于 7 的张力蛋白、波形蛋白等的缺失并不产生严重后果,而节点度高于 30 的整合蛋白、酪氨酸激酶 FAK 等的缺失则可使鼠胚致死(<http://www.bioscience.org/knockout/alphabet.htm>).

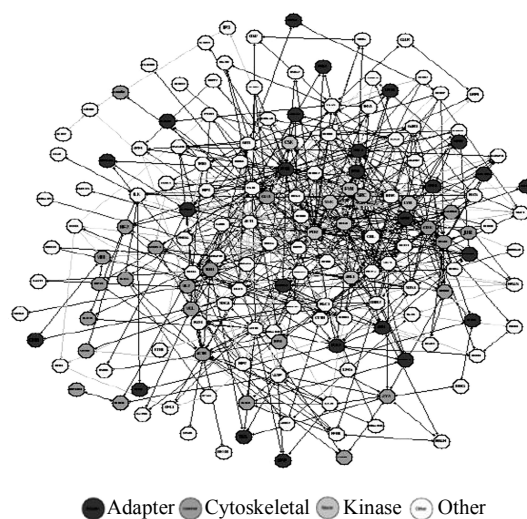


Fig. 1 Molecular interaction network of integrin adhesion

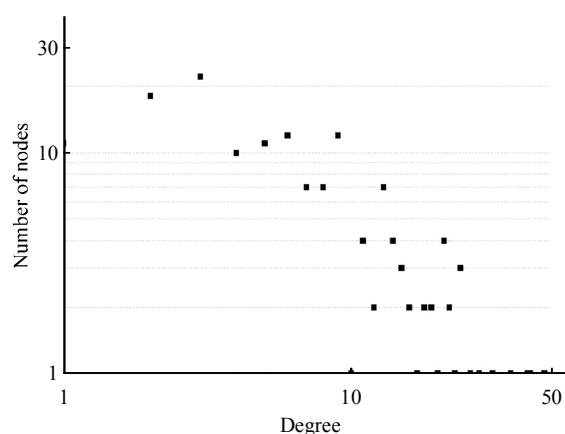


Fig. 2 Degree distributions of the molecular interaction network of integrin adhesion

1.2 黏附分子互作网络的结构与功能

采用基于效用构建子网络可将黏附分子互作网络有着相似活性的蛋白质归属为功能家族,其中衔接蛋白、黏附受体和肌动蛋白调节子三大家族构成了黏附网络的结构支架——将细胞膜与肌动蛋白骨架连接起来,其他家族主要包括调节黏附组装与解离以及细胞内外信号传导的各类酶蛋白.黏附分子互作网络所包含的主要蛋白质家族的节点数、平均节点度以及家族间的连接示于图 3.

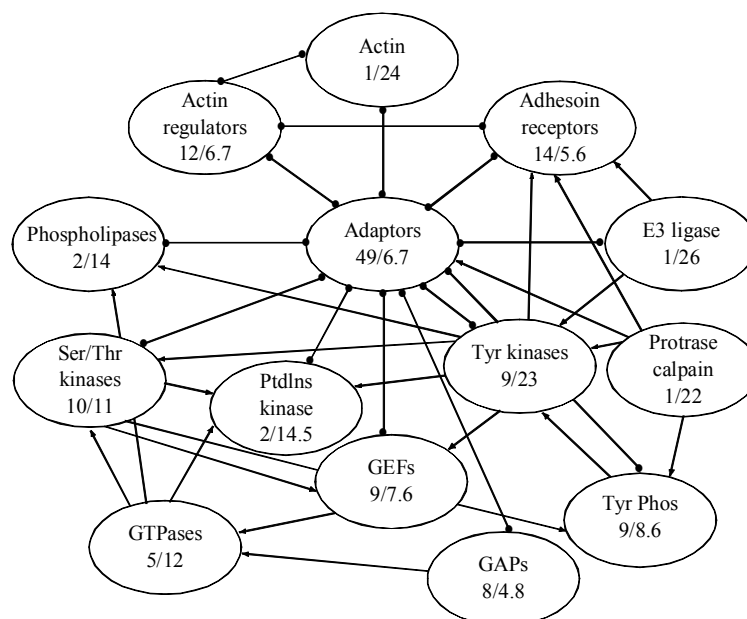


Fig. 3 Functional families in the molecular interaction network

Families: Node number/average degree; Arrows: Regulating interactions; Lines: Binding interactions.

功能家族间的连接揭示出重要的生物信息, 包括 GAP 和 GEF 蛋白受酪氨酸激酶调控, GTP 酶调控 PtdIns 激酶和丝氨酸 / 苏氨酸激酶, 而且 PtdIns 激酶也受酪氨酸激酶和丝氨酸 / 苏氨酸激酶调控, 脂质体调控衔接蛋白和肌动蛋白调节子等。同时, 无连接趋势也可提供一些有用的信息, 如黏附受体不与肌动蛋白相互作用、E3 连接酶 Cbl 不作用于肌动蛋白调控子、钙蛋白酶也不作用于 GAPs 或 PtdIns 激酶等。另外, 由黏附分子互作网络功能家族间的相互作用分析还可看到: a. 衔接蛋白家族为该互作网络的枢纽, 其他所有黏附体系的蛋白家族都与之连接; b. 酪氨酸激酶家族与除肌动蛋白和肌动蛋白调节子家族外所有与衔接蛋白家族联接的蛋白质家族之间存在调控关系; c. 丝氨酸 / 苏氨酸激酶主要调控其他激酶。

在此基础上, 可将黏附分子互作网络分解为以下 6 个功能模块。

跨膜连接结构模块: 该模块节点有 75 个, 包含骨架蛋白、黏附蛋白、受体蛋白、衔接蛋白、肌动蛋白及其调节蛋白; 连接有 144 条, 主要特征是所有分子互作均为结合作用的无向连接, 特别是受体蛋白通过衔接蛋白和肌动蛋白结合蛋白与肌动蛋白间接相连。已有结果表明有 5 种蛋白, 包括踝蛋白、张力蛋白、网蛋白、纤维蛋白和 α - 辅肌动蛋

白, 在受体蛋白和肌动蛋白之间形成桥接。

丝氨酸 / 苏氨酸磷酸化调控模块: 该模块节点有 54 个, 包含丝氨酸 / 苏氨酸激酶、磷酸酶及其底物和调节蛋白; 连接有 79 条, 主要特征是丝氨酸 / 苏氨酸激酶被其自身磷酸化、被丝氨酸 / 苏氨酸磷酸酶去磷酸化, 并受酪氨酸激酶、脂质和 Rho GTP 酶调控, 但不被钙蛋白酶(PKC 除外)或泛素 E3 连接酶蛋白分解, 也不调控丝氨酸 / 苏氨酸磷酸酶。

酪氨酸磷酸化调控模块: 该模块节点有 69 个, 包含酪氨酸激酶和磷酸酶及其底物和调节蛋白; 连接有 154 条, 主要特征是调控受体蛋白、肌动蛋白调节蛋白、丝氨酸 / 苏氨酸和 PtdIns 激酶以及多种衔接蛋白, 其显著特征是活化远远多于抑制, 可能是由于酪氨酸磷酸酶有较高的广连性。

Rho GTP 酶调控模块: 该模块节点有 44 个, 包含 GTP 酶及其效应物、GAPs 和 GEFs 及其上游调控物; 连接有 63 条, 主要特征是 Rho GTP 酶调控 PtdIns 激酶与丝氨酸 / 苏氨酸激酶, Rho GEFs 和 GAPs 受黏附内在成分调控, 而 Rac GEFs 和 GAPs 则受黏附关联成分调控。

脂质体调控模块: 该模块有节点 31 个和连接 37 条; 主要参与肌动蛋白(如踝蛋白、纽蛋白)与衔接蛋白结合、促进肌动蛋白成束和交联蛋白(如

丝蛋白)的调控,在一定条件下,与脂质体的结合会开启或关闭肌动蛋白的结合活性;另外,丝氨酸/苏氨酸激酶的活性也受脂质体调控。

蛋白质水解活性调控模块:该模块有43个节点与46条连接;所有黏附蛋白的降解只受钙蛋白酶和泛素E3连接酶蛋白(Cbl)的调控,钙蛋白酶受钙和丝氨酸/苏氨酸磷酸化调控,其分解底物主要是衔接蛋白、肌动蛋白调节蛋白以及2种酪氨酸激酶与2种酪氨酸磷酸酶;Cbl受酪氨酸磷酸化调控,而其E3连接酶活性下调酪氨酸激酶的信号传导并促进受体整合蛋白的降解。

2 黏附分子网络模体筛选

上述基于已有细胞黏附研究结果对黏附分子互作网络的模块解析存在粗浅和偏差的问题。通过鉴别互作网络中各种内部高度连接的节点集合,可以将功能模块进一步分解成不同的结构模体。在实际网络中,存在大量的模体如前馈环和反馈环等,具有重要的信息处理和调控功能,但各种模体不是同样显著的,各种网络都会有一系列独特的模体类型。这些模体是发挥特定的生物学功能的基本单位,揭示了相互作用模式的特点,表征了该网络的特征^[15-16]。

2.1 网络模体定义及其筛选

网络模体是满足下列条件的子图^[15]: a. 该子图在与真实网络对应的随机网络中出现的次数(频率)大于它在真实网络中出现的概率是很小的,通常要求这个概率小于某个阈值 P ,如 $P = 0.10$ 。 b. 该子图在真实网络中出现的次数 N_{real} 不小于某个下限 U ,如 $U = 4$ 。 c. 该子图在真实网络中出现的次数 N_{real} 明显高于它在随机网络中出现的次数 N_{rand} ,一般要求 $(N_{\text{real}} - N_{\text{rand}}) > 0.1 N_{\text{rand}}$ 。

网络模体筛选^[17-18]:给定一个真实网络和参数 k ,找出所有 k 规模的网络模体。典型的网络模体发现算法包括3个步骤: a. 根据真实网络的性质生成一组对应的随机网络; b. 在真实网络和随机网络中搜索特定规模的子图,确定哪些子图是同构的,并将同构的子图归为一类; c. 通过比较每一类子图在真实网络和随机网络中出现的次数以确定其统计意义,从而确定网络模体。

模体的统计意义一般用 $Z\text{-score}$ 来表示:

$$Z = (N_{\text{real}} - \langle N_{\text{rand}} \rangle) / \sigma_{\text{rand}}$$

式中, N_{real} 表示该子图在真实网络中的出现次数, $\langle N_{\text{rand}} \rangle$ 和 σ_{rand} 分别表示该子图在随机网络集

合中的平均出现次数和标准差。

2.2 黏附分子网络关键模体筛选

采用基于Mfinder^[19]软件进行黏附网络模体筛选。该软件列出所有 n -节点模式并将其分成几种拓扑学上不同的子图,各子图的计数与100个随机网络(节点数与真实网络相同、且各节点进出连接数相等但随机连接)获得的计数相比较,其中出现频率真实网络显著高于随机网络的子图视为网络模体。所有3节点和4节点模体列于表1,模体结构示于图4。

Table 1 Key motifs in molecular interaction network of integrin adhesion

Motif	N_{real}	$\langle N_{\text{rand}} \rangle \pm \sigma_{\text{rand}}$	$Z\text{-score}$	P
1	181	135 ± 12	3.8	0.00
2	105	82 ± 10	2.3	0.02
3	57	48 ± 7	1.3	0.10
4	138	124 ± 10	1.4	0.10
5	98	64 ± 16	2.1	0.04
6	273	144 ± 16	1.8	0.00
7	83	49 ± 14	2.4	0.02
8	171	110 ± 27	2.3	0.02
9	59	30 ± 10	2.9	0.01
10	66	22 ± 9	4.9	0.00
11	201	127 ± 27	2.7	0.00
12	39	18 ± 6	3.5	0.00
13	84	52 ± 13	2.5	0.01
14	84	50 ± 13	2.6	0.01

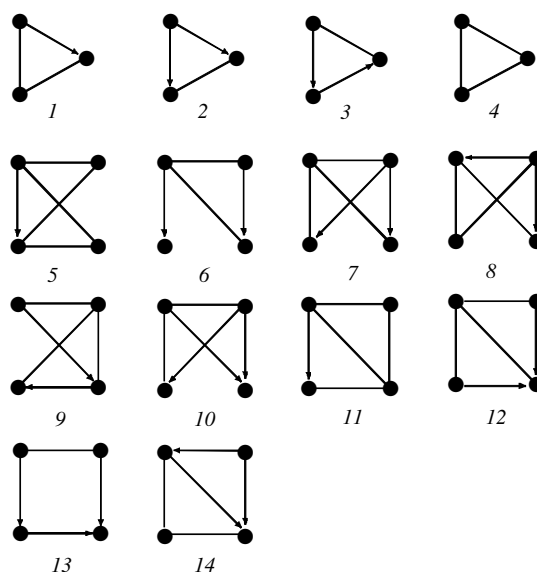


Fig. 4 Key motifs ($n = 3, 4$) in the molecular interaction network of integrin adhesion

模体 1~4: 模体 1 由一种酶及其底物构成, 两者均与另一第三蛋白有结合作用. 该模体自身出现 181 次, 同时含在大量 4 节点模体之内; 模体 2 包含一酶蛋白与两底物, 两底物间存在结合作用; 模体 3 和 4 在统计学上不如模体 1 和 2 重要, 但在黏附分子互作网络中出现相当频繁; 模体 3 为两种酶的级联, 底物与第一个酶结合; 模体 4 为一个三合体.

模体 5~8: 模体 5 为两蛋白同时支架一酶蛋白及其底物; 模体 6 由两酶蛋白及其相应底物构成, 其中一酶蛋白和底物支架另一酶蛋白; 模体 7 和 8 均包含一结合蛋白与一酶蛋白及两底物, 模体 7 结合蛋白与酶蛋白和底物均有结合, 而模体 8 结合蛋白只与酶蛋白及一底物结合.

模体 9~14: 模体 9 包含一酶蛋白、底物、下游底物及一结合蛋白(与前三者均有结合); 模体 10 由相互作用的两酶蛋白和两底物构成; 模体 11 中两蛋白支架一酶蛋白调控一底物; 模体 12 中两酶蛋白作用于同一底物; 模体 13 和 14 都包含一前馈环(FEL), 模体 9 为一非连续的 FFL, 而模体 10 是一连续 FFL; 模体 11~14 都形成闭环.

上述网络模体代表不同的信号转导通路, 这里并非所有的连接作用都同时发生(如 Abl 使 Cas 磷酸化, Cas 随后与 Crk 或 CrkL 结合). 由图 2 观察这些模体的结构, 可发现: a. 大多模体(除模体 6 和 13 外)有一重要共性, 即如蛋白 A 同时与蛋白 B 和 C 相互作用, 则蛋白 B 与 C 也可能结合; b. 几乎所有 4 节点模体的基本构件是 3 节点模体, 表明网络模体存在嵌套.

另外, 通常的生物网络模体是信息加工的静态环路, 而大多黏附分子网络模体实际上自身是动态变化的. 图 5 所示为模体 1 可由三种不同的动态模

式进行组合和分解. 所有模体都包含具有开关调控功能的结合作用, 而模体中结合节点与信号转导节点的连接则反映了支架模体和调控模体之间的相互作用, 以及转换多种相互作用的能力.

3 结 论

在细胞黏附分子和信号转导已有数据的基础上, 实现了整合蛋白黏附的分子互作网络的构建与可视化, 并表征出网络架构、内部层次、功能模块和关键模体; 对网络功能模块的解析与关键模体的筛选揭示出黏附分子互作网络的基本构架和功能单位.

由于构建黏附分子互作网络数据的高可变性, 不可能完全确定所筛选网络模体的统计显著性, 但关键模体的高丰度表明它们极有可能具有重要的生物学作用. 实验研究这些模体在黏附分子网络中的功能作用将具有重要意义, 如由此建立关于黏附介导的细胞信号转导机制, 并提出可行的扰动策略, 应用于药物筛选. 总之, 细胞黏附分子互作网络的构建与分析, 将为认识细胞分子机器如何感应环境等重要科学问题提供理论依据.

参 考 文 献

- [1] Strogatz S H. Exploring complex networks. *Nature*, 2001, **410**(6825): 268-276
- [2] Boccaletti S, Latora V, Moreno Y, *et al.* Complex networks: Structure and dynamics. *Phys Rep*, 2006, **424**(4-5): 175-308
- [3] Alon U. Biological networks: the tinkerer as an engineer. *Science*, 2003, **301**(5641): 1866-1867
- [4] Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. *Mol Sys Biol*, 2007, **3**(88): 1-13
- [5] 刘中扬, 李 栋, 朱云平, 等. 蛋白质相互作用网络进化分析研究进展. *生物化学与生物物理进展*, 2009, **36**(1): 13-24
Liu Z Y, Li D, Zhu Y P, *et al.* *Prog Biochem Biophys*, 2009, **36**(1): 13-24
- [6] Geiger B, Spatz J P, Bershadsky A D. Environmental sensing through focal adhesions. *Nat Rev Mol Cell Biol*, 2009, **10**(1): 21-33
- [7] Arnaout M A, Goodman S L, Xiong J P. Structure and mechanics of integrin-based cell adhesion. *Curr Opin Cell Biol*, 2007, **19**(5): 495-507
- [8] Harburger D S, Calderwood D A. Integrin signalling at a glance. *J Cell Sci*, 2009, **122**(2): 159-163
- [9] Alfaro C, Andrade C E, Anthony K, *et al.* The biomolecular interaction network database and related tools 2005 update. *Nucl Acids Res*, 2005, **33** (suppl 1): D418-D424
- [10] Keshava Prasad TS, Goel R, Kandasamy K, *et al.* Human protein

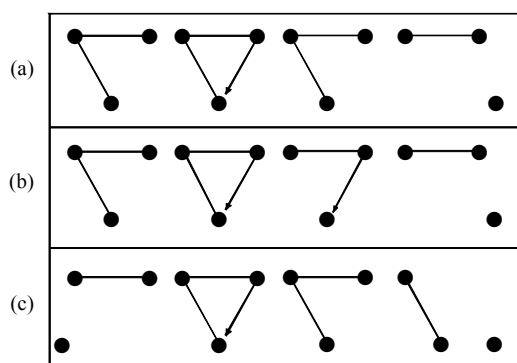


Fig. 5 Dynamic modes of motif 1

(a) Scaffolding. (b) Binding. (c) Unbinding.

- reference database-2009 update. Nucl Acids Res, 2009, **37**(Database issue): D767– D772
- [11] Hu Z, Mellor J, Wu J, *et al.* Towards zoomable multidimensional maps of the cell. Nat Biotechnol, 2007, **25**(5): 547–554
- [12] Berger S I, Iyengar R, Ma'ayan A. AVIS: Ajax viewer of interactive signaling networks. Bioinformatics, 2007, **23**(20): 2803–2805
- [13] Jeong H, Tombor B, Albert R, *et al.* The large-scale organization of metabolic networks. Nature, 2000, **407**(6804): 651–654
- [14] Collins M O, Husi H, Yu L, *et al.* Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. J Neurochem, 2005, **97**(s1): 16–23
- [15] Milo R, Shen-Orr S, Itzkovitz S, *et al.* Network motifs: simple building blocks of complex networks. Science, 2002, **298**(5594): 824–827
- [16] Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet, 2007, **8**(6): 450–461
- [17] Ciriello G, Guerra C. A review on models and algorithms for motif discovery in protein-protein interaction networks. Brief Funct Genomic Proteomic, 2008, **7**(2): 147–156
- [18] Qin G M, Gao L, Hu J L. A review on algorithms for network motif discovery in biological networks. Acta Electronica Sinica, 2009, **37**(10): 2258–2265
- [19] Kashtan N, Itzkovitz S, Milo R, *et al.* Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. Bioinformatics, 2004, **20**(11): 1746–1758

Construction and Analysis for Molecular Interaction Network of Cell Adhesion*

WEI Xiao-Lan**

(College of Environmental and Biological Engineering & Research Center for Pharmaceutical Chemistry and Chemical Biology,
Chongqing Technology and Business University, Chongqing 400067, China)

Abstract Complex network has been used to analyze the global and local properties of cell processes such as gene regulation, protein-protein interaction and signal transduction. A molecular network of integrin-mediated cell adhesions is constructed by data-mining and presented by a visualizing software. The network consists of 156 components linked by 690 interactions, with an average node degree of 8.66, an average cluster coefficient of 0.24 and an average path length of 2.6. Several functional modules are involved in switching on or off many of the molecular interactions within the network, consequently affecting cell adhesion, migration and cytoskeletal organization. Screening and examination of the adhesome network motifs reveals a relatively small number of key motifs, dominated by 3-nodes (three-component complexes). The role of the different network modules and motifs in regulating cell adhesions is also discussed.

Key words protein interaction, network analysis, systems biology, adhesion molecules

DOI: 10.3724/SP.J.1206.2010.00532

*This work was supported by grants from The National Natural Science Foundation of China(31070830), Natural Science Foundation of Chongqing (CSTC2008BB5329) and Natural Science Foundation of Chongqing Municipal Education Commission (KJ080706).

**Corresponding author.

Tel: 86-23-62769785, E-mail: weishine@ctbu.edu.cn

Received: October 15, 2010 Accepted: January 24, 2011