

Alzheimer's 疾病相关蛋白质相互作用 网络构建及其相互作用预测*

蒋雄飞^{1,2}, 杨洁^{1**}, 王炜²

(1. 南京大学生物医药技术国家重点实验室, 南京, 210093;

2. 南京大学物理系, 南京, 210093)

摘要: 依据无标度网络的相关理论, 提出一种预测蛋白质-蛋白质相互作用的算法, 并预测潜在的作用位点。收集了所有与阿尔兹海默疾病 (Alzheimer's disease) 相关疾病的人类蛋白, 并找出与之能发生相互作用的蛋白质, 绘制一个有关人类 AD 相关蛋白质相互作用网络。分析和计算了该网络的一些详细属性, 并对网络进行生物信息学分析, 找出网络中的蛋白质调控路径, 该路径可能涉及到 Apoptosis-antagonizing transcription factor (AATF) 蛋白对于阿尔兹海默疾病的抑制作用。几个潜在与 AATF 蛋白作用的位点被自己开发的程序预测出, 将它们的潜在作用位点做了比对, 最后得到一个高度收敛的序列模体, 其正则表达式为 $I-x(0,1)-E-x(2)-[AENT]-x-[EK]$ 。

关键词: 阿尔兹海默氏病, 蛋白质相互作用, 复杂网络, 无标度网络, 相互作用预测, 序列模体

中图分类号: Q 615

The Complex Network of Protein-Protein Interaction of Alzheimer's Disease Associated Proteins and an Interaction Predicting

Jiang Xiong-Fei^{1,2}, Yang Jie¹, Wang Wei²

(1. State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing, 210093, China; 2. Department of Physics, Nanjing University, Nanjing, 210093, China)

Abstract: Proteins potentially associated with the pathology of Alzheimer's disease were gathered into our database, and were then mapped into a protein interaction network. Here, we reported a novel approach to predict protein-protein interactions and their potential interaction sites, based on the characteristics of the free-scale network. Meanwhile, the detailed properties of network were analyzed, some proteins meditative pathways were also found, which may be involved in the antagonizing PAWR mediated induction of aberrant amyloid peptide production in Alzheimer's disease. Then several potential protein interactions and their sites were predicted through program compiled by ourselves. One interaction between O96018 and P05067 confirmed in experiment was re-predicted through the PPIP.

Pending proteins predicted to interact with Apoptosis-antagonizing transcription factor protein with high possibility were picked out, and their potential interaction sequences were in alignment. A well conservative pattern

* 基金项目: 国家自然科学基金 (30171094, 30271497)

收稿日期: 2006-07-01

** 通讯联系人, E-mail: yangjie@nju.edu.cn

(motif) maybe involved in molecular mechanism of interaction with Apoptosis-antagonizing transcription factor protein was generated, with a regular expression of $I-x(0,1)-E-x(2)-[AENT]-x-[EK]$

Key words: Alzheimer's disease, protein interaction, complex network, free-scale network, interaction prediction, motif

阿尔茨海默氏病是中枢神经系统退行性疾病中的一种,病理表现主要为神经元纤维缠结老年斑和颗粒空泡变性,其内在机制到目前尚不非常清楚.但很显然,它的产生是由于某些蛋白质异常行为而引起的^[1~3],很多报道指出其与淀粉样蛋白(APP)密切相关^[4,5].

到目前为止,生物学研究一直侧重在对单个蛋白的研究,即研究蛋白个体的特征及能够和其发生相互作用的其他蛋白.现面临的课题是如何将这些已研究过的基于单体的实验信息整和起来,以更为系统和全面的眼光来看待生物体内在分子水平上的相互作用及其调控路径.为此,我们选择以复杂网络作为工具来研究蛋白分子水平上的整体特征.

对于蛋白质相互作用网络的研究不仅具有理论上的重要意义,也具有潜在的实际应用前景.在新药研发时,它可使药物指向特定靶点,对网络进行相应的调控或者阻碍操作,而取代以往简单的以给定的蛋白质功能为目标靶点.对于细胞系统明显的网络特性和代谢途径中普遍存在的瓶颈步骤,当前发现药物方法都仍依赖于对确定的单个目标进行破坏,这本身就有局限性^[6].迄今为止,有 500 种人类的蛋白已被确认可作为药物作用的靶子,这些蛋白代表了基因功能和更高水平的细胞表型之间直接的一对一关系^[7].

同时,预测蛋白的高级结构及其功能一直是计算生物学中的一个核心的问题,一套好的预测方法能给实验人员以有效的理论指导,如何预测蛋白之间是否存在相互作用也是其中一项任务,很多预测两个蛋白或者亚基契合的算法被开发出来,比如 Jones 和 Thornton 曾使用包括疏水性、表面区域(accessible surface

area)、形状以及残基偏好性等来比对结构特性^[8].近来的研究集中在使用残基疏水性来预测相互作用的接触位点,如 Korn 和 Burnett 使用疏水分析来预测在二聚体蛋白相互作用中的作用表面位置^[9],Young 进一步发展他们的方法,提出基于蛋白簇残基(clusters of residues)疏水分析的自动预测算法^[10].

根据 Anfinsen 的蛋白质序列决定其高级结构理论,在蛋白质之间存在着相互作用时,可以看作是两个蛋白的结合区域的三维结构正好相契,所以当有另一蛋白的某一区域的三维结构与其中某一蛋白的结构相似,那么,它就很有可能与另一蛋白发生作用.以此思想为基础,通过比对,即可能从已有的相互作用序列片断中找出未知蛋白是否可能与靶蛋白发生作用,若有,可计算出大致发生的位置.以此方法,我们开发程序预测未知蛋白会与某些 AD (Alzheimer's disease) 相关蛋白发生作用,从而将其纳入人类 AD 相关蛋白相互作用网络.

1 蛋白质相互作用网络

1.1 构建蛋白质相互作用网络 首先,有关所有中枢神经系统退行性疾病的相关蛋白被收集起来,建立 CNSNDD 数据库(Central Nervous System Neurodegenerative Diseases Database),而 AD 则是 CNS 退行性疾病中典型的一种疾病.CNSNDD 中的蛋白相互作用数据来源于 IntAct^[11],基于网络规模,考虑三度以内的蛋白,数据均可以在 CNSNDD 方便的查询到.

为更有效的分析网络性质,在开始只选取与 AD 相关的人类蛋白构建网络,未推广至所有的模式动物蛋白.一共确定 119 个蛋白与

AD 相关,其中有 70 个人类蛋白.以该 70 个蛋白作为骨架,获取其三度以内的相互作用蛋白,最后构建了人类 AD 相关蛋白相互作用网络.利用可视化网络构建软件,将整个网络图画出来.如图 1 所示.

深色点均是 70 个 CNSDD 中的部分蛋白,都已在实验上证实与 AD 的发病机理有关,其中圆点代表引起或促进 AD 发病的蛋白,三

角形点代表抑制 AD 发病的蛋白;而浅色点表示它们暂时仅仅确定是深色点的相互作用蛋白,而没有包含于 CNSDD 数据库中,即尚未在实验上发现与 AD 的发病有关联.由于实验的限制,在该 70 个蛋白中,还有部分蛋白尚无相互作用蛋白,而无法联入网络.虚线表明为预测的相互作用,将在后面叙述.

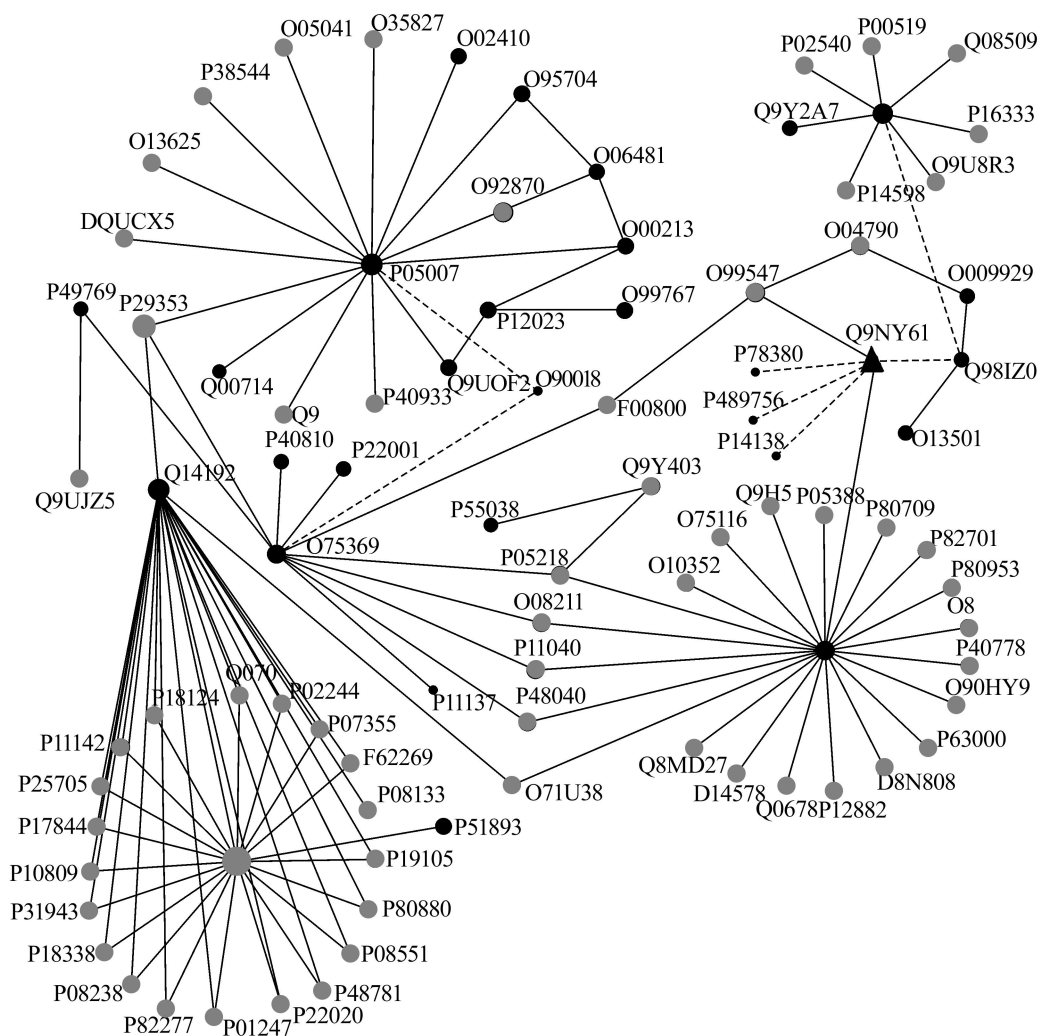


图 1 人类 AD 相关蛋白质相互作用图

Fig. 1 Illustration showing protein-protein interaction of human's Alzheimer's disease associated proteins

1.2 蛋白相互作用网络 (Scale-free Network) 的性质 这里,首先列举出一些在工作中要使用的量^[12]:

(1) 网络是由节点(Node)与它们两两之间的连线(Connection)构成.在蛋白相互作用网络中,节点代表着蛋白,而线则意味着它们之间有相互作用.

(2) 度(Degree):一个节点所联结的所有线,即某蛋白所能作用的蛋白数量,以 k 表示,度分布 $P(k)$ 则意味着拥有某个度 k 的节点在所有的可能的 k 中的可能性.

(3) 平均度(Mean Degree):网络中所有节点的 k 的平均值,以 $\langle k \rangle$ 表示.

无标度网络^[12,13]具有一些奇特的性质,这里只关注那些与本文工作密切相关的性质.有两个机制主导着无标度的性质,一、网络的生长是新的节点连接到网络原有的节点上的过程,二、也是最重要的,在大多数真实的网络中,拥有大的连接数,即度(k)大的节点拥有更大的几率连接新的节点,这就是优先连接(preferential attachment)^[12,13]原则.

在无标度网络中,一个具有大的连接度的节点($k \gg \langle k \rangle$)从统计角度说要比其他的节点重要的多,这些节点被称为中心节点(Hub),网络的属性在很大程度上由这些只占很小部分的中心节点决定.一个重要的后果是无标度网络具有抗随机扰动能力,但对加在中心节点上的作用很敏感.换言之,随机节点的故障不会破坏网络的完整性,然而,如有目的的将中心节点切除,则网络结构会迅速瓦解^[14,15].这个特点对于生物系统尤为重要,因为它反映了生物医学网络对于随机突变的鲁棒性^[16].

1.3 AD 相关蛋白相互作用网络的属性特征的详细分析 整个网络中,共有 91 个节点(不包含孤立蛋白),119 个连接,平均度, $\langle k \rangle = 1.31$. 7 个蛋白被选择为中心蛋白,它们与 AD 的发病机理密切相关,分别是 P05067、P29353、Q14192、O75369、P10636、Q9N Y61 和 Q8IZP0 (表 1),度分别为 15、3、23、11、22、6 和 8,除 P29353 只有 3,其他均远大于平均度 $\langle k \rangle =$

1.31,而 P29353 的入选是因为其重要的连接作用.尤其 P05067, Amyloid beta A4 protein,即 APP 蛋白,有 6 个深色,即实验上证实于 AD 相关的蛋白与其直接相连.如果考虑 3 度以内,即 3 步以内,很明显整个网络可以被分割为两个主要的子网络.

第一个子网络是以 APP 为中心,而在此 3 度以内的深色节点迅速上升到 14 个,占到 27 个深色节点的 55.6%.从某种程度上说,APP 在整个网络中的地位是最重要的.最近的研究普遍认为,APP 的聚集在 AD 的发病机理起着至关重要的作用^[17~19].所以,可以看出,无论是实验,还是本文的生物信息学方法,发现 APP 确实在 AD 的发病机理中十分关键.

另一个子网络的中心蛋白为 apoptosis-antagonizing transcription factor (AATF),即网络右边的三角形点(图 1),聚集在它周围的深色节点则为 4 个.从宏观角度看,AD 相关蛋白相互作用网络就是通过上面提到的两组蛋白质的相互作用.

Q9N Y61,也就是 AATF,最近被确认是 DAP-like kinase (Dlk) - DAP (death-associated protein) kinase 家族的成员之作用伴侣.AATF 会阻抗 PAWR (Q96IZ0) 诱导的细胞凋亡,这表明 AATF 直接或间接的参与 PAWR 活性的调节.AATF 的表达会阻碍由 PAWR 诱导的 APP 异常表达和分泌,同时 AATF/PAWR 的复合系于 AATF 对 APP 异常分泌的抑制作用起着最基本的作用.这些结果意味着 AATF 是 PAWR 活性和 APP 在凋亡环境下异常表达、分泌的内生性阻抗物^[20].然而,尚无实验报道 AATF 与 PAWR 能够与 APP 直接发生相互作用.所以,根据以上的推理,我们认为从 AATF 到 APP 之间应该存在着某条调控路径,通过调控路径上的蛋白级联作用而最后使得 AATF 对 APP 的异常表达产生抑制作用,并对 AD 起到一定的阻止或抑制作用.从网络中,得出几条潜在的调控路径,如下:

1. Q9N Y61 (AATF) Q99547 P60866

O75369 P29353 **P05067 (APP)**

2. **Q9NY61 (AATF)** P10636 Q71U36
Q14192 P29353 **P05067 (APP)**

3. **Q9NY61 (AATF)** P10636 P05218
(or Q08211, or P11940, or P46940) O75369
P29353 **P05067 (APP)**

(注: Q99547: M - phase phosphoprotein 6; P60866: 40S ribosomal protein S20; O75369: Filamin - B; P29353: SHC transforming protein 1; P10636: tau protein; Q71U36: Tubulin alpha - 3 chain; Q14192: Four and a half LIM domains protein 2; P05218: Tubulin beta - 2 chain; Q08211: ATP - dependent RNA helicase A; P11940: Polyadenylate-binding protein 1; P46940: Ras GTPase-activating-like protein IQ GAP1)

分析第一条路径,首先 AATF,作为细胞凋亡阻抗因子,该过程与磷酸化密切相关,所以紧随其后的是 Q99547, M 相磷酸蛋白;然后与核糖体蛋白 40s 亚基发生作用;接下来与 O75369 发生作用,该蛋白与 FLNA 结合后将允许成神经细胞从心室区转移至大脑皮层,这一步的含义并不是很清楚;最后通过 P29353,实际上每条路径最后都是通过该蛋白与 APP 发生作用,该蛋白起到了将激活生长因子受体与信号通路耦合的信号适配作用。同理也可发现, P29353、Q14192 和 O75369 这几个蛋白的位置也是很重要的,它们在左右两个子网络之间起到了桥梁作用。

H. Jeong 等研究了一个真实的酵母的蛋白相互作用网络^[21]。通过酵母蛋白的随机移除,发现酵母基因组的随机突变不会影响到整个网络的拓扑结构。同时作为对比,当最大联结的那些蛋白被移除时,网络迅速解体。换言之,那些拥有大联结数目的蛋白时细胞的生存时最重要的。所以,一些在我们网络中高联结数的蛋白也被挑选了出来。它们对于 AD 网络是至关重要的,决定着网络的属性,在某种程度上,

如果即使在现在它们还未被实验证实与 AD 相关,但从网络的性质可推断,很可能将是会相关,像 P29353 和 Q92870,至少在 AD 的调控上起着重要作用(在表 1 中,被标记上星号)。如上所述, P05067、P29353、Q14192、O75369、P10636、Q9NY61、Q8IZP0 和 Q06481 被挑选为中心节点蛋白。

由表 1 可见,除 Q14192 和 P29353 之外,其他的中心节点都多少的和神经功能相关,同时都在神经元或者大脑神经干细胞中表达。所以,它们确实与神经系统相关,而从组织表达的特异性上说,与 AD 的发病机制也许相关。

2 蛋白质相互作用预测

2.1 预测算法 鉴于实际计算能力,不可能对所有的蛋白进行预测,所以,如前文所述,中心节点蛋白在网络中的作用要远远的超出一般的蛋白,且由于优先连接 (preferential attachment) 规则的介入,使得那些具有大 k 的节点,即节点蛋白更倾向于接受新的蛋白相互作用,尽管它们只是占了少数,所以优先检测中心节点蛋白的相互作用蛋白。

预测方法是基于序列比对。根据 Anfinsen 的蛋白质序列决定其高级结构理论,在蛋白质之间存在相互作用时,可以看作是二个蛋白的结合区域的三维结构正好相契,所以当有另一蛋白的某一区域的三维结构与其中某一蛋白的结构相似,那么它就很有可能与另一蛋白发生作用。如图 2 所示,蛋白 A 是某个在网络中的中心节点,其中 AR 区域是与蛋白 B 的 BR 区域发生相互作用,这里的相互作用已由实验确认。蛋白 C 与蛋白 D 代表未知的需要预测的蛋白序列,而其中的深色区域表示可能存在的结合位点。选取那些已经经过实验验证的相互作用蛋白序列作为标准序列,则 AR 是靶序列,而 BR 则为标准序列。当输入待测序列时,通过比对找出与 BR 匹配的序列片段。

表 1 一些重要的节点及其相关性质

Table 1 Some vital nodes in the network and their properties

SwissProt accessional No.	Name	Function	Tissue specificity
P05067	Amyloid beta A4 protein	Functions as a cell surface receptor and performs physiological functions on the surface of neurons relevant to neurite growth, neuronal adhesion and axonogenesis. Involved in cell mobility and transcription regulation through protein-protein interactions. Appicans elicit adhesion of neural cells to the extracellular matrix and may regulate neurite outgrowth in the brain	Expressed in all fetal tissues examined with highest levels in brain, kidney, heart and spleen
P29353	SHC transforming protein 1	Signaling adapter that couples activated growth factor receptors to signaling pathway	Expressed in neural stem cells but absent in mature neurons
Q14192	Four and a half LIM domains protein 2	Unknown	Expressed in skeletal muscle and heart
O75369	Filamin-B	Connects cell membrane constituents to the actin cytoskeleton. May promote orthogonal branching of actin filaments and links actin filaments to membrane glycoproteins. Anchors various transmembrane proteins to the actin cytoskeleton. Interaction with FLNA may allow neuroblast migration from the ventricular zone into the cortical plate	Expressed Ubiquitous
P10636	Microtubule -associated protein tau	Promotes microtubule assembly and stability, and might be involved in the establishment and maintenance of neuronal polarity	Expressed in neurons
Q9N Y61	Apoptosis- antagonizing transcription factor	May function as a general inhibitor of the histone deacetylase HDAC1, Also antagonizes PAWR mediated induction of aberrant amyloid peptide production in Alzheimer disease	Expressed at high levels in brain, heart, kidney, placenta and thymus.
Q8IZP0	Abl interactor 1	May act in negative regulation of cell growth and transformation by interacting with nonreceptor tyrosine kinases ABL1 and/or ABL2	Widely expressed, with highest expression in brain
Q92870	Fe65-like protein	May modulate the internalization of beta-amyloid precursor protein	Unknown

标 的表明该蛋白质虽然未被实验证实与 AD 相关,但从图中基本可看出实际上是密切相关的。

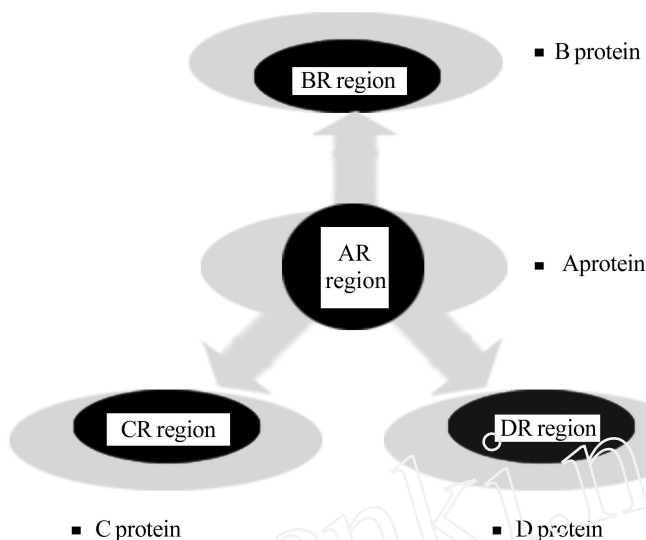


图 2 预测过程

Fig. 2 The process of prediction

使用滑窗法来找出最佳匹配的序列片段,待测序列和标准序列分别为 m 和 n 长度,然后待测序列上的每个 n 长度的窗口都会通过 BLOSUM62 矩阵来做比对记分, BLOSUM62 矩阵是氨基酸之间相互替换的相似性矩阵,即越相似,分值越大,若氨基酸完全一致,则分值最大。最后,最大分值的窗口将被挑选出来,判断是否会发生作用。

具体的过程如图 3 所演示。最后上面的待测的序列中的浅色被选择出来,因为通过 BLOSUM62 矩阵的比对,其具有最大分值,所以被认为与标准序列具有最大的相似性。

预测可靠性分析:现阶段的标准还仅限于半经验的方式,一旦待测序列的最大片段分值大于标准序列的分值的 10%,就认为它们之间存在着相互作用,因为待测序列的结构可能和标准序列可能相似。

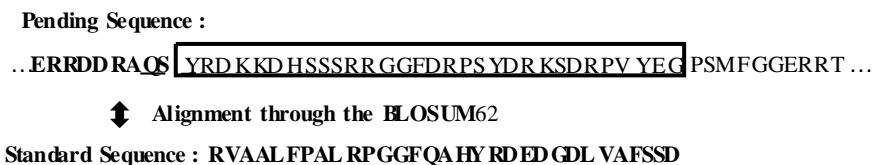


图 3 比对示意图

Fig. 3 The illustration of alignment

标准序列,即那些于中心节点蛋白相互作用的蛋白的作用序列—已有实验确定的一列举在表 2^[20,22~24]。根据上述的算法和序列,使用 Visual C++ 开发了蛋白相互作用程序 Protein-Protein Interaction Prediction program (PIIP),通过它,可以大致上确认待测序列能否于中心节点蛋白产生作用,如能发生作用,则

大致上的位置是在哪儿,从而将其纳入人类 AD 相关蛋白相互作用网络中。

2.2 预测结果

2.2.1 可能的相互作用蛋白 所有的孤立蛋白,其有可能相互作用的蛋白和可能的位点都列在表 3 中,其中前文所述之孤立而无法联入网络中的蛋白由表 3 之第 3 列所示。

在文献中人工逐个的检查了上述所有的潜在相互作用,发现 O96018 和 P05067 的相互作用是真实存在的,已有 Tanahashi 和 Tabira 在

1999 年的实验所验证^[25]. 从某种程度上说,当标准序列足够长时预测的效果会更好,理由是长的序列更可能形成相似的结构.

表 2 用来作为标准序列的蛋白质及其比对部分的序列

Table 2 The proteins and their sequences for standard sequence

Hub proteins			Proteins interaction with hub proteins	
Hub proteins	Domains of hub proteins	Target domain from experiment	Sequences	
P05067 (APP) ^[20]	Cytoplasmic domain (via YENPTY motif)	PID domain	DGIIFAAN YL GSTQLL SERNPS KNIRMMQAQE AVSRV KRMQ KAA KIKKKANSE GDAQ TL TEV DL FISTQRIKVL NAD TQ ETMMD HAL RTIS YIA DIGNRRRM PRSA SQDCIET TPGA QEGKKQ YK MICHV FESEDAQL IAQSIGQAFSVA YQEFLRA NGIN PEDL SQ KE YSDIIN TQ EM YND	
Q9NY61 (AATF) ^[22,23]	Leucine Zipper Domain	Leucine Zipper Domain	GKL KEEIDLL NRDLDIED EN EQL KQEN KTLL KVV GQL TR	
	The first region spans from amino acid in position 1 to 163, the second region spans from amino acid portion 371 to 470	The region spanning from amino acid position 44 to 230	ESPLQ TPTED GSEEP GSETSDA KSTPTA EDV T APL VDEGAPGKQAAA QPH TEIPEGTTAEEA G IGD TPSL EDEAA GHV TQEPESGKVVQ EGFLRE PGPPGL SHQL MSGMP GALL PEGPREA TRQPS GTGPEDTEGGR HAPELL KHQLL GDL HQEGPP L KGA GGKERPGS KEEVDEDRDVDES SPQD RVAAL FPAL RPPGFQA HYRDED GDL VAFSSD	
Q96IZ0 (PAWR) ^[22]	Unknown	Region 50 - 80		
O75369 (FLNB) ^[22,24]	The region from 1128 to 1511	Filamin Binding	MASKPEKRVASSV FITL APPRRDVAVAEVVR QAVCEARR GRPWEA PAMP KTPEA GLA GRPS PWTT PGRAA	
	Actin Binding Domain	Loop Region	CPKGPLRML VETAQ ERNETL FPALIYSSTMV WLVNMA EGDPEA QRRVS KNSKYNA ESTERE SQD TVAENDD GGFSEWEAQ RDSHL GPHRS TPESRAAVQEL SSSILA GEDPEERGV KL GL GD FIFYSVL V GKASA TASGDWNTTIACF	

2.2.2 序列模式 序列模式(Motif)是指核酸或者蛋白质序列中具有保守性的序列片断.

现在的蛋白相互作用数据主要是来自于序列作用. 随着规模的扩大,这种相互作用预测的

可能性将会以指数的形式增长. 然而我们知道,在这些蛋白质序列中,经常包含若干保守的区域,即序列模式. 由于从序列到结构的进化,这种保守区域将存在于蛋白相互作用网络中,它

们可以被描述为模式保守相互作用网络. 许多生物功能都是通过一些高度相互作用的细胞内的部件,即功能模块来完成^[26~28]. 序列模式被

认为是一种在拓扑结构上很明确的复杂网络中的作用模式,它是最基础的功能模块. 因而有必要发展能够预测这种序列模式(motif)的方法.

表 3 PPIP 程序的预测结果

Table 3 The results of PPIP program

Hub proteins			Proteins interaction with hub proteins	
Hub proteins	Domains of hub proteins	Proteins verified through PPIP	Potential interaction proteins from prediction	sequence
P05067 (APP)	Cytoplasmic domain		O96018	EVPGPCDHEDLLDGVIFGAR YL GST QLVSERNPPTSTRMAQAREAMDRV KAPDGETQPMTEVDL FVSTKRIKVL TADSQEAMMDHAL HTISYTADIGCV LVL MARRRLARRPAPQDH GRRL YK MLCHVFYAEDAQLIAQAIGQAFAAA YSQFL RESGIDPSQV GV HPSPGAC HL HN GDLD HFSNSDNC
Q9N Y61 (AATF)	Leucine Zipper Domain	Q9 Y5Z0 Q9 Y2W7 Q9 Y287 Q9U KY0		
		Q9U HC1 Q8IZP0	Q12830	C KIL EEMREEIHR HMDITEDL TN KAR GSN KSFL AAANEEL
		Q9N TX7 Q93045		
		Q92876 P14136	P49756	IEMEED KRDL ISREIS KFRD TH KKL EE
		Q8N3V7 Q8IZD9		EKGKKE KERQIEK
		Q15392 Q15078	P10909	SKYVN KEIQNA VN GV KQIKTL IEKTN
		Q15011 Q14114		EERKTLL SNL EEA KK
		Q12830 Q04609		LAA YRQEAD EATL ARLDL ER KIESL E
		P68036 P56817	P14136	EEIRFL R KIHEEEVR
		P49770 P49757		IRSL LQDAAIEEVEMEDFDANIEEQ KE
		P49755 P49754	P49750	EKKDA EEESEL GY
		P49753 P49750		ENEL KEMIE TLARKL NEKS KEQMEL
		P37840 P25713	P78380	HHQNLNLQETL KRVAN
		P25021 P15374		VGKYK KTH KNIMINL MDIL EVD T KK
Q96IZO (PAWR)	Unknown	P10909 P09936	Q15392	QIVRVEPL V TMGQV TA
		P08174 P07339		RD KKD HSSSRR GGFDRPS YDR KSD
		P05451 P04156	P49750	RPV YEGP
		P04080 P03891		ITEAQVAENRPGA FIKQ GRKL DIDFG
		P02795 P02787	Q9U KY0	A EGNR
		P02649 P01011		EWGQSSGAASPG L FQA GHRR TPSE
O75369 (FLNB)	Loop Region	O75173 Q8IV G9	P49757	ADRWL EE
		O96018 P78380		RV TGL FPGN YV ESIM H YTD
		P49756 P11137	Q8IZP0	VTKSPEKRSSL PRPSSIL PPRRGVSG
			P11137	DRDENSFSLNSSISSAARR TTRSEPIR
				RAGKSGTSTPTTPGSTA
			O96018	AAATTPTKV YSEV HFTLA KPPSVVN
				RTARPFGIQAPGGTSQMERSPML ER
				RHFGEKAPAPQPPSL PDRSP

我们尝试下述方法,表 3 中的潜在的能与 AATF (Q9N Y61) 通过其 Leucine Zipper 域而发生作用的 7 条局部序列被挑选出来,分别是 P10909、P78380、P14136、Q15392、Q12830、P49750 和 P49756 的潜在位点. 通过 Pratt 的计算和 Prosite 数据库的检索^[29,30],发现一个很好的序列模式,其正则表达式为 $I-x(0,1)-E-x(2)-[AENT]-x-[EK]$. 式中每个横线的左右均是表示一个残基,该序列模式共有 8 个残基. 第 6 位的 [AENT] 表明该位置上只可能出现 A、E、N、T 中的一种,而它们的残基性质都是相似的;同理,最后的 [EK] 表明只可能式 E 或者 K 中的一种,它们的性质亦相似. 简单的比对如下,其中 * 表示位空格,该序列模式在 P49750 中出现两次,所以用下标标出.

```

P10909    I *E k t N e E
P78380    I *E t l A r K
P14136    I *E s l E e E
Q15392    I l E v d T k K
Q12830    I t E d l T n K
P49750_1  I *E e v E m E
P49750_2  I e E q k E e K
P49756    I *E m e E d K

```

图 4 模体比对

Fig. 4 Alignment of motif

如前所述,我们希望这个看来不起眼的保守序列能在复杂网络中起着某种作用,也许是整个网络大厦中的一块地基,也许涉及到能与 AATF 蛋白产生相互作用的微观机制,比如在产生相互作用的识别中起着作用,当然这还需要进一步的实验验证.

3 讨 论

本文通过建立一个人类 AD 相关蛋白质相互作用网络,从宏观角度来分析 AD 可能的发病机理和蛋白质的调控路径. 找出网络中的中心节点蛋白,其对于网络的整体性质有着至关重要的作用,因此,深入的了解这些中心节点蛋白将会有助于解开 AD 的发病机理. 同时,这些中心节点蛋白也很可能是开发治疗 AD 药物

的好的靶点,因为它们构成了整个网络的骨架,一旦这些节点移除,那么网络可能会崩溃,从而对治疗产生某种作用.

随后开发了基于序列比对的蛋白质相互作用预测程序,并在未知情的情况下成功复检出一蛋白相互作用. 本文还提炼出一个可能涉及到 AATF 蛋白识别的保守模体片断.

由于采用的为单序列比对思想,但该方法在结果的可靠性分析上有所欠缺. 国际上对单序列比对的可靠性分析至今尚无良好的对策,我们暂时还缺乏大量应用而积累的结论,所以可靠性分析需得等积累比较多的结论后才能进行. 我们正在做一个关于但序列比对方法的发展,简单的说,是想赋予 BLOSUM 矩阵的各个氨基酸之间的替换以合理的权重,同时结合氨基酸残基的疏水性和残基偏好性,即也使这两者具有合理的权重,通过这种权重,能够使得各个残基对于整体结构的贡献量化,从而对结果进行科学的可靠性分析. 该工作正在进行中,在后续文章将详细论述.

References

- [1] Burke R E. Apoptosis in degenerative diseases of the basal ganglia. *Neuroscientist*, 1998, 4: 301 ~ 311.
- [2] Jellinger K A. General aspects of neurodegeneration. *Journal of Neural Transmission - Supplement*, 2003, 65: 101 ~ 144.
- [3] Martin J B. Molecular basis of the neurodegenerative diseases. *New England Journal of Medicine*, 1999, 340: 1970 ~ 1980.
- [4] Arispe N, Rojas E, Pollard H B. Alzheimers disease amyloid beta-protein forms calcium channels in bilayer membranes: Blockade by tromethamine and aluminum. *Proceedings of the National Academy of Science*, 1993, 90: 567 ~ 571.
- [5] Hardy J. Framing beta-amyloid. *Nature Genetics*, 1992, 1: 233 ~ 234.
- [6] Cascante M, Boros L G, Comin-Anduix B, et al. Metabolic control analysis in drug discovery and disease. *Nature Biotechnology*, 2002, 20: 243 ~ 249.

- [7] Huang S. Rational drug discovery: What can we learn from regulatory networks? *Drug Discovery Today*, 2002, 7: S163 ~ S169.
- [8] Jones S, Thornton J M. Analysis of protein-protein interaction sites using surface patches. *Journal of Molecular Biology*, 1997, 12, 272(1): 121 ~ 132.
- [9] Korn A P, Burnett R M. Distribution and complementarity of hydrophobicity in multisubunit proteins. *Proteins*, 1991, 9(1): 37 ~ 55.
- [10] Young L, Jernigan R L, Covell D G. A role for surface hydrophobicity in protein-protein recognition. *Protein Science*, 1994, 3: 717 ~ 729.
- [11] Hermjakob H, Montecchi-Palazzi L, Lewington C. IntAct - an open source molecular interaction database. *Nucleic Acids Research*, 2004, 32: D452 ~ D455.
- [12] Albert R, Albert L, Barabási A L. Statistical mechanics of complex networks. *Review of Modern Physics*, 2002, 74: 47 ~ 97.
- [13] Newman M E J. The structure and function of complex networks. *SIAM Review*, 2003, 45: 167 ~ 256.
- [14] Cohen R, Erez K, Avraham D, *et al.* Breakdown of the internet under intentional attack. *Physics Review Letters*, 2001, 86(16): 3 682 ~ 3 685.
- [15] Cohen R, Erez K, Avraham D, *et al.* Resilience of the internet to random breakdowns. *Physics Review Letters*, 2000, 85(21): 4 626 ~ 4 628.
- [16] Albert R, Jeong H, Barabási A L. Attack and error tolerance of complex networks. *Nature*, 2000, 406: 378 ~ 382.
- [17] Arispe N, Rojas E, Pollard H B. Alzheimer's disease amyloid beta-protein forms calcium channels in bilayer membranes: Blockade by tromethamine and aluminum. *Proceedings of the National Academy of Science*, 1993, 90: 567 ~ 571.
- [18] Hardy J. Framing beta-amyloid. *Nature Genetics*, 1992, 1: 233 ~ 234.
- [19] Hardy J, Selkoe D J. The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science*, 2002, 297, 353 ~ 356.
- [20] Jun X, Qing G. PAR-4 is involved in regulation of α -secretase cleavage of the Alzheimer amyloid precursor protein. *Journal of Biological Chemistry*, 2005, 280(14): 13 824 ~ 13 832.
- [21] Jeong H, Mason S P, Barabási A L, *et al.* Lethality and centrality in protein networks. *Nature*, 2002, 411: 41 ~ 42.
- [22] Barbato C, Corbi N, Canu N, *et al.* Rb binding protein Che-1 interacts with Tau in cerebellar granule neurons. *Molecular and Cellular Neuroscience*, 2003, 24: 1 038 ~ 1 050.
- [23] Zhang W, Han S W, McKeel D W, *et al.* Interaction of presenilins with the filamin family of actin-binding proteins. *Journal of Neuroscience*, 1998, 18: 914 ~ 922.
- [24] Chow N, Korenberg J R, Chen X N, *et al.* APP-BP1, a novel protein that binds to the carboxyl-terminal region of the amyloid precursor protein. *Journal of Biological Chemistry*, 1996, 271: 11 339 ~ 11 346.
- [25] Tanahashi H, Tabira T. X11L2, a new member of X11 protein family interacts with Alzheimer's beta-amyloid precursor protein. *Biochemical and Biophysical Research Communications*, 1999, 255: 663 ~ 667.
- [26] Tong A H, Drees B, Nardelli G, *et al.* A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science*, 2002, 295: 321 ~ 324.
- [27] Wagner A, Fell D A. The small world inside large metabolic networks. *Proceedings of the Royal Society of London - B*, 2001, 268: 1 803 ~ 1 810.
- [28] Sherr-Orr S S, Milo R, Mangan S, *et al.* Network motifs in the transcriptional regulation network of *E. coli*. *Nature Genetics*, 2002, 31: 64 ~ 68.
- [29] Jonassen I. Efficient discovery of conserved patterns using a pattern graph. *Computer Applications in the Bioscience*, 1997, 13: 509 ~ 522.
- [30] Jonassen I, Collins J F, Higgins D G. Finding flexible patterns in unaligned protein sequences. *Protein Science*, 1995, 4: 1 587 ~ 1 595.