

文章编号:1004-0374(2005)01-0082-06

计算方法在蛋白质相互作用研究中的应用

曹建平¹, 马义才¹, 李亦学², 石铁流^{2*}

(1 电子科技大学生命科学与技术学院 成都 610054;

2 中国科学院上海生命科学研究院生物信息中心 上海 200031)

摘要: 计算方法在蛋白质相互作用研究的各个阶段扮演了一个重要的角色。对此, 作者将从以下几个方面对计算方法在蛋白质相互作用及相互作用网络研究中的应用做一个概述: 蛋白质相互作用数据库及其发展; 数据挖掘方法在蛋白质相互作用数据收集和整合中的应用; 高通量方法实验结果的验证; 根据蛋白质相互作用网络预测和推断未知蛋白质的功能; 蛋白质相互作用的预测。
关键词: 蛋白质相互作用数据库; 蛋白质功能预测; 蛋白质相互作用网络; 蛋白质间相互作用
中图分类号: Q-332; TQ937 文献标识码: A

The application of the computational methods in protein-protein interaction study

CAO Jian-Ping¹, MA Yi-Cai¹, LI Yi-Xue², SHI Tie-Liu^{2*}

(1 College of Life Science and Technology, UESTC, Chengdu 610054, China; 2 Bioinformation Center, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China)

Abstract: Computational methods play an important role at all stages of the process of determining protein-protein interactions. Here, computational methods dealing with protein-protein interactions and interaction network are discussed from five aspects: interaction databases with their construction and development; automated data mining techniques developed to extract information about interactions from the published literature; computational methods used for the assessment of the results of high-throughput approaches; exploitation of the information provided by protein interaction networks for predicting functional features of the proteins; computational methods adopted for predicting protein-protein interactions.

Key words: protein interaction database; protein function prediction; protein interaction networks; protein-protein interaction

1 引言

细胞的生命紧密地依赖于其自身与周边环境的交流能力, 而细胞的代谢、信号传导以及基因表达调控等都与蛋白质的功能密切相关, 生物分子必须参与到错综复杂的相互作用的网络中行使其功能, 这也是目前所有生物学研究的基础。随着基因组工

作的不断进展, 我们发现人类和老鼠在基因组上有很大的结构和序列相似性。因此, 我们有理由认为, 人类和老鼠的主要区别在于基因所编码的蛋白质之间的相互作用及其网络的不同, 并且更高级的物种应该有更复杂的相互网络。然而, 直到近些年来, 随着实验技术的发展和高通量方法的应用, 大

收稿日期: 2004-12-03

基金项目: 国家“863”项目(2002AA231051)

作者简介: 曹建平(1978—), 男, 硕士; 马义才(1956—), 男, 副教授; 李亦学(1955—), 男, 研究员, 博导; 石铁流(1964—), 研究员, 博导; * 通讯作者。

范围的蛋白质相互作用及其网络的研究才成为可能。

研究蛋白质相互作用的最终目标是建立模式细胞系统中全部蛋白质相互作用的网络,这将为研究蛋白质功能及其细胞全局特征构筑一个框架。基于蛋白质相互作用及其网络的实践应用也已经得到发展,例如药物开发中的药靶研究,相关的详细描述见 Archakov 等^[1]的综述。

研究蛋白质相互作用的手段,包括已经建立的一系列传统的实验方法,如酵母双杂交系统、质谱仪方法和蛋白质芯片等^[2]。近年来,随着计算机科学的发展,计算方法已经成为生物科学研究中一个有力的工具,面向于整个基因组范围内研究蛋白质相互作用的理论策略已经被提出。基于实验数据的蛋白质相互作用数据库已是相关研究领域的重要基础,蛋白质相互作用数据和蛋白质注释信息整合在一起是目前蛋白质相互作用数据库的主要特征。从蛋白质相互作用的科学文献中自动化地挖掘信息,是蛋白质相互作用数据库发展的主要推动技术之一;而用计算方法验证高通量方法得到的蛋白质相互作用数据,已成为目前蛋白质相互作用验证的一个必要补充。根据数理统计知识,开发和应用各种算法进行蛋白质相互作用及其网络的预测,成为该领域研究的热点。

2 蛋白质相互作用数据库

随着科学技术的发展,生物信息学在现代生物研究工作中显得越来越重要,而数据库是一切生物信息学工作的出发点,很多生物软件的开发和应用都需要数据库的支撑,计算机科学家和生物信息学家通过数据库的构建和维护为生物学家提供服务。随着实验数据的积累和实验手段的快速发展,蛋白质相互作用的数据不断增加,以及对整个基因组的蛋白质相互作用的网络分析的需求,蛋白质相互作用的数据库应运而生,如 DIP、BIND 等。蛋白质相互作用的数据库不仅仅是相互作用的蛋白质对的列表,而且包括了一些相关的注释信息和附加证据等。在蛋白质相互作用研究中常用的一些数据库见表 1。

关于表 1 各数据库的数据描述和收集方法及开发工具等在相关的网站都有详尽的描述,用户可以根据自己的不同需要选择相关的数据。通常情况下,各数据库对学术用户支持免费的下载和服务。随着蛋白质相互作用实验数据的积累和相关数据库的不断发展,这些数据提供了用于表达、储存、查

询和浏览蛋白质相互作用信息的机制,并且还包括了部分的相互作用的实验条件和过程的信息。这些资源将为开发新的算法和验证已有的算法提供材料。

3 文献的数据挖掘

常规注释生物学数据库和详细注释蛋白质相互作用数据库的任务令人望而生畏。这包括广泛审阅文献以确定和提取相关的信息。由于数据库注释类似于手册的特性,提取的信息大都引用有良好文档、可重复结果。事实上,大量有关蛋白质相互作用和蛋白质功能的信息已经存在于已发表的文章中。美国国立医学图书馆(NLM)在 MEDLINE 数据库中收集并维护大量的生物学文摘,这为文本数据挖掘方法提供了基础。

在蛋白质相互作用数据库的建立和维护更新中,从已发表的文献内容中自动地挖掘生物信息是最为复杂的一环。最近,一些基于文本数据挖掘并且围绕这个概念的方法被提出^[3]。这些方法的中心思想是采用自然语言处理(NLP)技术来提取相应的信息片段,并按预先定义的模型将其收录在数据库中。这是个信息提取的过程,可能包括非常精密或尖端的处理算法以识别文本复杂的语法和语义结构。有关处理和发现蛋白质相互作用的文本信息的尝试已有报道^[4]。文本挖掘的其他过程包括文本聚类^[5],该方法根据相似的术语将相似的文档聚集到相关的文档组。这些文档组收集了可能包含关于某个主题,如蛋白质相互作用的特异性信息。

目前,一些蛋白质相互作用数据库直接应用了文本数据挖掘的方法,比如 DIP, BIND 等。最新的 MEDLINE 数据库包括了 1 400 万多条的论文和摘要,以及 4 800 多种期刊出版物。该方法通过扫描 MEDLINE 数据库的标题和摘要,得到那些可能与蛋白质相互作用相关的条目。然而,该自动化获取系统仍然仅仅被用做基本的过滤,得到 MEDLINE 数据库中蛋白质相互作用相关的子集,其信息仍然需要人们的进一步确认。

文本数据挖掘方法的困难在于科学文本的复杂特性和人类语言的不确定性,这些算法承受着很高的噪声水平。此外,面对自动化的数据挖掘程序相关的单个单词的词义模糊匹配,基因名和蛋白质名称的同义或多义,使其成为一个极具挑战性的任务。目前自动化的获取系统仅仅停留在扫描文献数据库中的标题和摘要,是因为大部分的出版刊物的全文并不是免费可用的,这涉及到作者的知识产权

表 1 主要的蛋白质数据库和蛋白质相互作用数据库

英文名	中文名	网址
Swiss-Prot	蛋白质序列数据库	http://www.expasy.ch/cgi-bin/sprot-search-ful
NCBI	蛋白质序列数据库	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein
PIR	蛋白质序列数据库	http://pir.georgetown.edu/
PDB	生物大分子结构数据库	http://www.rcsb.org/pdb/
BIND	生物分子相互作用数据库	http://bind.ca/
DIP	蛋白质相互作用数据库	http://dip.doe-mbi.ucla.edu/
GRID	蛋白质相互作用数据库	http://biodata.mshri.on.ca/grid/servlet/Index
IntAct	蛋白质相互作用数据库	http://www.ebi.ac.uk/intact/index.html
INTERACT	蛋白质相互作用数据库	http://bioinf.man.ac.uk/interactpr.htm
InterDom	结构域相互作用数据库	http://interdm.lit.org.sg/
MINT	生物分子相互作用数据库	http://mint.bio.uniroma2.it/mint/
PimRider	蛋白质相互作用数据库及工具	http://pim.hybrigenics.com/pimriderext/common/
PRONet	蛋白质相互作用数据库	http://pronet.doublet.wist.com/
STRING	蛋白质相互作用网络数据库	http://string.embl.de/
Pawson Lab	相互作用蛋白结构域信息数据库	http://www.mshri.on.ca/pawson/domains.html
Predictome	功能相关和相互作用数据库	http://predictome.bu.edu/
PPI Server	复合体相互作用界面分析数据库	http://www.biochem.ucl.ac.uk/bsm/PP/server/
PathCalling	相互作用工具和数据库	http://portal.curagen.com/pathcalling_portal/index.htm
PIM	基因杂交相互作用数据库和工具	http://pim.hybrigenics.com/pimriderext/common/
CYGD	酵母相互作用数据库	http://mips.gsf.de/proj/yeast/CYGD/interaction/
YPD	酵母蛋白质相互作用数据库	http://proteome.incyte.com/
HIV IDB	HIV蛋白质相互作用数据库	http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions/
HPRD	人类蛋白质参考数据库	http://www.hprd.org/
HPID	人类蛋白质相互作用数据库	http://wlab.inha.ac.kr/hpid/
OPHID	在线预测人类相互作用数据库	http://ophid.utoronto.ca/
InterPreTS	三级结构预测相互作用数据库	http://www.russell.embl.de/interprets/
MPP1	哺乳动物相互作用数据库	http://fantom21.gsc.riken.go.jp/PPI/
FusionDB	细菌和古细菌的基因融合数据库	http://igs-server.cnrs-mrs.fr/FusionDB/
ADVICE	共进化相互作用数据库	http://advice.i2r.a-star.edu.sg/
InterWeaver	蛋白质相互作用在线证据	http://interweaver.i2r.a-star.edu.sg/
PathBLAST	蛋白质相互作用网络算法	http://www.pathblast.org/
ClusPro	蛋白质对接全自动化算法	http://nrc.bu.edu/cluster/
COGS	直系同源, 种系发生谱	http://www.ncbi.nlm.nih.gov/COG/
WIT	直系同源, 种系发生谱	http://wit.mcs.anl.gov/WIT2/
ALLFUSE	基因融合	http://www.ebi.ac.uk/research/cgg/allfuse
KEGG	代谢/调控通路分析和重建	http://www.genome.ad.jp/kegg/
ECOCYC	代谢通路分析	http://ecocyc.pangeasystems.com/ecocyc/
ExpProfiler	基因表达谱分析	http://ep.ebi.ac.uk/EP/
Brown Lab	基因表达数据	http://genome-www4.Stanford.edu/MicroArray/SMD
Church Lab	基因表达数据	http://twod.med.harvard.edu/ExpressDB/

问题, 另外一个问题, 也可能存在计算机能够读取的文件格式问题, 比如国内的许多文献并没采用国际标准的 PDF 格式。

4 蛋白质相互作用的验证

目前, 许多实验技术已经被用来检测大范围蛋

白质相互作用网络, 包括酵母双杂交实验、GST 融合蛋白质鉴定、免疫共沉淀、蛋白质复合体的质谱技术、合成裂解、荧光检测、Biacore 表面等离子体共振分析、mRNA 表达相关分析、液相分析、蛋白质芯片、哺乳动物双杂交实验等^[2]。然而, 令

人惊奇的是在同一物种中,用不同的高通量实验方法得到的蛋白质相互作用数据之间很难彼此覆盖(coverage)。最近,Aloy和Russell^[6]比较了用酵母双杂交实验、液相分析、质谱方法以及用他们自己的生物信息学方法得到酵母蛋白相互作用的数据,发现各种方法之间的覆盖难以置信得低,其结果见表2。另外有研究指出在酵母中用高通量方法得到的蛋白质相互作用网络的覆盖也很有限,用酵母双杂交方法获得的相互作用的网络估计会有近50%的假阳性^[7-8]。

表2 酵母中用不同的高通量方法得到的蛋白质相互作用数据的交叉重叠数目及百分比^[6]

	Aloy	Ito	Uetz	Ho	Gavin
Aloy	499	0.2%	1.2%	5.4%	4.6%
Ito	1	4 475	13.8%	2.4%	2.5%
Uetz	6	199	1 447	6.4%	6.7%
Ho	27	106	92	72 690	8.6%
Gavin	23	113	97	4 197	48 751

最近几年,评价和验证相互作用网络结果的一些计算方法开始引进来:Satio等^[9]引入了“相互作用的一般性”方法,用该方法评价实验方法得到结果的可靠性。该方法通过定义蛋白质相互作用的网络拓扑特性,根据相互作用的拓扑图的连接数量特征,监控相互作用水平,从而鉴定相互作用的可靠性;Mrowka等^[7]根据蛋白质在mRNA的表达水平相关性来评价两个蛋白质相互作用的可靠性。该方法通过比较高通量方法得到数据全局的 D^2 分布和一个随机样本分布,得到蛋白质相互作用表达谱可靠性索引(expression profile reliability, EPR) 据此鉴定高通量方法得到蛋白质相互作用数据的可靠性;Deane等^[8]则用相似性检验方法(paralogous verification method, PVM)来估计蛋白质相互作用的可靠性。在实际研究中,上述几种方法经常一并应用。

5 基于蛋白质相互作用预测蛋白质的功能

利用蛋白质相互作用信息预测蛋白质的功能及相关特性是蛋白质相互作用研究的一个重要方面。根据Brown等^[10]的研究表明相互作用的蛋白质具有相似的生化功能、细胞角色和亚细胞定位等的概率更高。在进行鉴定蛋白质的功能实验中,预先用批量方法预测蛋白质的功能,将对实验的设计有很好的指导意义。Lee等^[11]通过理论分析蛋白质相互作

用网络中的蛋白质功能信息,得到一些“功能模块”(相互作用网络中的一些子集),根据模块中已知功能蛋白质预测未知蛋白质功能。Deng等^[12]则根据蛋白质相互作用网络中蛋白质对其注释信息应用马尔可夫随机理论成功地预测了蛋白质的功能,并用贝叶斯方法给出了所预测蛋白质功能的置信概率。最近,Brun等^[13]提出一种基于蛋白质相互作用网络预测蛋白质功能的聚类方法,该方法通过相互作用网络中蛋白质进行聚类,依据所得类中蛋白质的注释信息,预测其未知蛋白质。他们用该方法成功地注释了37个未知蛋白质的功能。

在所有的利用蛋白质相互作用预测蛋白质功能的方法中,蛋白质相互作用的数据决定了其预测的可靠性,数据量的增加无疑将会提高其预测的准确性。另外一方面,很多蛋白质在细胞中具有多种功能,如何成功地预测蛋白质的多种功能将会是蛋白质功能预测中的下一个研究方向。

6 蛋白质相互作用的预测

预测蛋白质的相互作用是目前生物信息学中最具挑战性的目标,而通过模拟和计算方法要比大部分的实验方法快得多,同时花费更少。在最近几年,一些生物信息学算法已经用来预测蛋白质的相互作用与否,这些方法总结起来可以分成四类:(1)基于基因组信息方法;(2)基于进化关系方法;(3)基于蛋白质序列的从头预测方法;(4)需要三维结构信息方法。在比较不同方法时,其主要难题是各种方法采用了不同的数据库,同时这些数据库的数据,其可靠性相当不一致。

基于基因组信息的预测方法包括系统发育谱(phylogenetic profile)、基因邻接(gene neighborhood)、基因融合(gene fusion event)以及镜像树(mirror tree)等方法。

基于进化信息的方法包括突变关联(correlated mutation)、保守的蛋白质相互作用(interologs)、进化速率关联(correlated evolutionary-rate)等方法。关于这些方法的概述见朱新宇和沈百荣^[14]的综述。以上的各种方法均不可避免的存在一定的局限性,它们都需要一些蛋白质的先验知识,如基因组信息、进化信息等,从而限制了其适用性。

基于蛋白质的一级结构的预测方法首先由Bock和Gough^[15]提出,该方法不需要基因组的或进化的信息,仅仅需要单个蛋白质的序列信息。该方法从DIP的数据库中提取相互作用的蛋白质的序列数据,

根据蛋白质对的序列信息,包括氨基酸残基的理化特性、电荷以及疏水特性等,用支撑向量机的方法训练,其交叉验证的结果表明了该方法具有很高的准确率,大约80%;但他们的方法仅仅能鉴定真实蛋白质对和“假蛋白质”对,不能解决实际问题。最近在我们的研究工作中,用真实蛋白质序列对训练数据的负集进行重新构建,并利用了基于蛋白质一级结构预测的二级结构信息及氨基酸溶解性等新的特性,用支撑向量机的方法训练,进行交叉验证测试,预测的准确率达到了70%左右。

最近,有人提出了利用蛋白质的三维结构信息进行蛋白质相互作用的预测的方法,即同源结构复合物(homologous structural complexes)方法^[16]。该方法构建了一种全新的策略,并提供了在线Web服务(<http://www.russell.embl.de/interprets>)。该数据库包括了429对非冗余相互作用结构域和1131个已知三维结构的复合体,用提交的两个序列对数据库搜索序列的同源性,如果找到一个同源序列,即可证明该提交的蛋白质序列存在对相互作用有利的经验潜能。另外,在一对随机序列背景的基础下,一种用来估计蛋白质潜在相互作用的具有统计意义的方法也已经被提出。Lichtarge等^[17]据此设计了一种在进化中有意义的重要氨基酸的聚类方法,用来在三维空间里进行蛋白质的功能位点的预测。

各种预测蛋白质的结构域和结构域相互作用的方法基于不同的背景知识,但这些方法的最终目标是预测蛋白质的相互作用,在每个蛋白质中这些结构域是与蛋白质的相互作用相关的。Ng等^[18]根据Pfam数据库、PDB数据库以及Cell-zome数据库获得结构域的相关信息,如复合体或序列的融合关系等,对两个结构域的相互作用进行概率统计,得到一个分值,该分值与两个结构域相互作用的概率是成比例的。另外,一个最大似然值的方法^[19]最近也被提出来,两个蛋白质的相互作用的概率应该和任何与在他们中的结构域相互作用的概率成比例。该方法估计了两个结构域相互作用的概率,用这样的比例来描述,即包括这两个相互作用结构域的相互作用蛋白质的数量与那些包括这两个结构域蛋白质的总数。

7 结论

计算方法在蛋白质相互作用应用已经极其广泛,但一种简单广泛有效的方法依然值得期待,如在相互作用的基因组分析方面,某一物种中适用的

方法,期待能够在别的物种中也能应用。各种实验技术的发展和高通量方法以及基于文献数据挖掘方法在蛋白质相互作用中的应用,极大地丰富了相互作用蛋白质数据库。用计算方法对高通量结果进行验证,为蛋白质相互作用数据的可靠性提供保障。蛋白质相互作用是蛋白质功能的重要方面,将为蛋白质组学研究的最终目标“提高生活质量”的实现奠定基础。各种预测蛋白质相互作用方法的计算方法有着不同的适用范围,而相应的评价标准还未建立,在实际应用中一个可行的考虑是组合不同的方法进行预测。此外,这些方法与实验方法的结合可以挖掘网络中更多的相互作用节点,尤其重要的是这将能够更完整地描述生物学过程,深刻理解细胞的全局特征。运用各种计算方法,对基因的、基因组的以及蛋白质组的信息进行整合分析,将成为细胞生物学和分子生物学一个新的研究方向。

[参 考 文 献]

- [1] Archakov A I, Govorun V M, Dubanov A V, et al. Protein-protein interactions as a target for drugs in proteomics. *Proteomics*, 2003, 3 (4): 380-391
- [2] Erica Golemis 编著. 贺福初, 钱小红, 张学敏, 等译. 分子克隆手册: 蛋白质-蛋白质相互作用[M]. 北京: 中国农业出版社, 2004. 74
- [3] Craven M, Kumlien J. Constructing biological knowledge bases by extracting information from text sources[A]. *Proc Int Conf Intell Syst Mol Biol* [C]. California: AAAI Press, 1999, 77-86
- [4] Blaschke C, Andrade M A, Ouzounis C, et al. Automatic extraction of biological information from scientific text: protein-protein interactions[A]. *Proc ISBM Heidelberg* [C]. California: AAAI Press 1999, 60-67
- [5] Iliopoulos I, Enright A J, Ouzounis C A. Textquest: document clustering of Medline abstracts for concept discovery in molecular biology. *Pac Symp Biocomput*, 2001, 384-395
- [6] Aloy P, Russell R B. The third dimension for protein interactions and complexes. *Trends Biochem Sci*, 2002, 27 (12): 633-638
- [7] Mrowka R, Patzak A, Herzel H. Is there a bias in proteome research? *Genome Res*, 2001, 11 (12): 1971-1973
- [8] Deane C M, Salwinski L, Xenarios I, et al. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics*, 2002, 1 (5): 349-356
- [9] Saito R, Suzuki H, Hayashizaki Y. Interaction generality, a measurement to assess the reliability of a protein-protein interaction. *Nucleic Acids Res*, 2002, 30 (5): 1163-1168
- [10] Brown M P, Grundy W N, Lin D, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA*, 2000, 97 (1): 262-267

- [11] Lee T I, Rinaldi N J, Robert F, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 2002, 298 (5594): 799-804
- [12] Deng M, Zhang K, Mehta S, et al. Prediction of protein function using protein-protein interaction data. *J Comput Biol*, 2003, 10 (6): 947-960
- [13] Brun C, Herrmann C, Guenoche A. Clustering proteins from interaction networks for the prediction of cellular functions. *BMC Bioinformatics*, 2004, 5 (1): 95
- [14] 朱新宇, 沈百荣. 生物技术通讯. 2004, 15 (1): 70-73
- [15] Bock J R, Gough D A. Predicting protein-protein interactions from primary structure. *Bioinformatics*, 2001, 17 (5): 455-460
- [16] Aloy P, Russell R B. InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics*, 2003, 19 (1): 161-162
- [17] Lichtarge O, Yao H, Kristensen D M, et al. Accurate and scalable identification of functional sites by evolutionary tracing. *J Struct Funct Genomics*, 2003, 4 (2-3): 159-166
- [18] Ng S K, Zhang Z, Tan S H. Integrative approach for computationally inferring protein domain interactions. *Bioinformatics*, 2003, 19 (8): 923-929
- [19] Deng H M, Mehta S, Sun F Z, et al. Inferring domain-domain interactions from protein-protein interactions. *Genome Res*, 2002, 12 (10): 1540-1548

美科学家说现有人类胚胎干细胞系已受污染

新华网洛杉矶 2005 年 1 月 23 日电(记者陈勇)美国加利福尼亚大学圣迭戈分校的科学家 23 日报告说,他们发现目前所有的人类胚胎干细胞系都被一种人体内不存在的分子所污染,这可能危及胚胎干细胞的医学应用。

加州大学圣迭戈分校的瓦尔基等人在 2005 年第 1 期 *Nature Medicine* 杂志上发表论文说,这一名为 N - 羟乙酰神经氨酸(Neu5Gc)的硅铝酸通常存在于其他动物细胞的表面,但人类在遗传上是不能合成的。

研究人员说,现有干细胞的培养和诱导分化,都离不开动物源培养基,比如从实验鼠身上分离出来的结缔组织细胞和牛胚胎血清等,这些含有 N - 羟乙酰神经氨酸的细胞污染了人类胚胎干细胞。在实验中,被 N - 羟乙酰神经氨酸污染的干细胞很容易被人类免疫系统识别和攻击,因此胚胎干细胞的医学应用潜力受到威胁。

瓦尔基说,他们在实验中用传统方法培育和分化一个胚胎干细胞系,并借助电喷雾质谱法测定 N - 羟乙酰神经氨酸的含量。结果发现,胚胎干细胞中 N - 羟乙酰神经氨酸占硅铝酸总量的 2.5% 至 10.5%,而在干细胞初步分化的胚状体中, N - 羟乙酰神经氨酸占硅铝酸总量的 5% 至 17% 之间。

研究人员同时在论文中提出一系列改进方法。比如,培育不产生 N - 羟乙酰神经氨酸的转基因实验鼠,用其结缔组织细胞来做培养基,再用牛胚胎血清替代品来培育干细胞。但他们指出,由于多数牛胚胎血清替代品也是从动物组织中提取,而且现有干细胞系都是以前用传统手段培养的干细胞的“后代”,污染恐怕难以根除。

瓦尔基认为,从人类胚胎中提取新的干细胞系,然后用人类自身的组织细胞来培养,才能从根本上避免 N - 羟乙酰神经氨酸的污染。

摘自新华网