

学校编码: 10384

分类号_____密级_____

学号: 20120051302139

UDC_____

厦 门 大 学

硕 士 学 位 论 文

生物分子相互作用动力学参数数据库及动力学网络的构建

Construction of Kinetic Data of Bio-molecular Interactions
database (KDBI) and Kinetic network

韩步聪

指导教师姓名: 纪志梁 副教授

专业名称: 生物化学与分子生物学

论文提交日期: 2008 年 4 月 25 日

论文答辩时间: 2008 年 6 月 3 日

学位授予日期:

答辩委员会主席: _____

评 阅 人: _____

2008 年 6 月

厦门大学学位论文原创性声明

兹提交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文而产生的权利和责任。

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

- 1、保密（ ），在 年解密后适用本授权书。
- 2、不保密（ ）

（请在以上相应括号内打“√”）

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

目 录

摘 要.....	1
ABSTRACT	2
1. 前言	3
1. 1 蛋白质相互作用及其研究意义	4
1. 2 蛋白质相互作用的研究方法	5
1. 2. 1 传统方法.....	5
1. 2. 2 生物信息学方法	8
1. 2. 3 传统方法和生物信息学的比较.....	12
1. 3 构建蛋白质相互作用网络的意义.....	12
1. 3. 1 蛋白质相互作用网络的意义.....	12
1. 3. 2 构建蛋白质相互作用网络的生物信息学方法.....	13
1. 4 本论文的思路、目的和意义.....	16
2. KDBI 数据库的更新.....	18
2. 1 数据库平台配置.....	18
2. 1. 1 硬件平台.....	18
2. 1. 2 软件平台.....	19
2. 1. 3 整个平台的协调配置.....	20
2. 2 数据库设计.....	21
2. 2. 1 需求分析.....	21
2. 2. 2 实体关系设计.....	21
2. 3 数据处理.....	23
2. 3. 1 数据收集.....	23
2. 3. 2 数据整理.....	27
2. 4 构建数据库.....	27
2. 4. 1 建立空表.....	28
2. 4. 2 数据导入.....	28
2. 4. 3 建立视图.....	29
2. 5 数据库展示.....	30
2. 5. 1 网页设计.....	30
2. 5. 2 网页代码编写.....	31
2. 5. 3 调试和功能实现.....	31
3. 基于 KDBI 的蛋白质相互作用的动力学网络构建.....	37
3. 1 数据来源	37
3. 2 数据处理	37
3. 3 建立网络	38

3. 4	优化拓扑.....	39
3. 5	flash 作图.....	40
4.	结果与讨论.....	44
4. 1	结果与分析.....	44
4. 1. 1	数据库更新显著.....	44
4. 1. 2	动力学网络的构建.....	47
4. 2	构建蛋白质相互作用的动力学网络的应用.....	48
4. 3	展望.....	48
	参考文献.....	50
	致 谢.....	54
	附录一：攻读硕士研究生期间发表的文章.....	55

厦门大学博硕士论文摘要库

CONTENTS

Abstract in Chinese	1
Abstract in English	2
1. Preface	3
1.1 Protein protein interaction and significance of reseach	4
1.2 Procedure of protein protein interaction research	5
1.2.1 Tradition method	5
1.2.2 Bioinformatics method.....	8
1.2.3 Comparison between the two.....	12
1.3 Significance of protein protein interaction kinetic network construction	12
1.3.1 Significance of protein protein interaction network.....	12
1.3.2 Bioinformatics in construction of protein protein interaction network	13
1.4 Idea, objective and significance of this article	16
2. Improvement of KDBI	18
2.1 Configuration for database platform	18
2.1.1 Hardware platform.....	18
2.1.2 Software platform.....	19
2.1.3 Configuration between hardware and software.....	20
2.2 Database Design	21
2.2.1 Analysis of requirement.....	21
2.2.2 Design for entities relationship.....	21
2.3 Procedure of data process	23
2.3.1 Data collection	23
2.3.2 Data process.....	27
2.4 Database Construction	27
2.4.1 Tables creation	28
2.4.2 Data import.....	28
2.4.3 Views creation.....	29
2.5 Visualization of database	30
2.5.1 Web design.....	30
2.5.2 Search engine coding.....	31
2.5.3 Function and debugging.....	31
3. Construction of protein protein kinetic network based on KDBI	37
3.1 Data Resource	37
3.2 Data Process	37
3.3 Network construction	38

3.4 Optimization for topology.....	39
3.5 Procedure by flash.....	40
4. Results and discussion.....	44
4.1 Results and analysis	44
4.1.1 Significant improvement of database.....	44
4.1.2 Construction of kinetic network.....	47
4.2 Application for protein protein interaction kinetic network	48
4.3 Prospect	48
References.....	50
Acknowledgements	54
Appendix 1: Publication during graduate study.....	55

厦门大学博硕士论文摘要库

摘 要

分子间的相互作用特别是蛋白质间相互作用几乎存在于机体每个细胞的生命活动过程，生物路径及其底层分子事件。蛋白质相互作用网络的定性和定量研究在探查和研究细胞行为、发展治疗疾病的新疗法过程中具有重要作用。构建蛋白质相互作用网络，特别是动力学网络不论是在科学认识上还是在实验技术上以及在药物设计等领域中，都有重大意义。本文详细描述了构建生物分子相互作用动力学参数数据库（KDBI）。该数据库收集了文献中报道的实验得出的有关蛋白与蛋白、蛋白与核酸、蛋白与配体、核酸与配体之间相互反应的动力学参数信息。并采用生物信息学方法建立了 27 条超过 2 层关系的蛋白质相互作用动力学网络作为这个数据库动力学研究细胞过程的全新应用。

关键词： 蛋白质相互作用 动力学网络 生物信息

ABSTRACT

Interactions between molecules, especially protein-protein interactions are implicated in almost every cellular process, biological pathways and their underlying molecular events. Qualitative and quantitative study of protein interaction networks is believed to play key roles in the exploration and engineering of cell behavior and in the development of novel therapeutics to combat diseases. Developing the protein-protein interaction network especially Kinetic data interaction network is not only important to research on science but also to experiment technology even drug design. In this article, we construct kinetic protein-protein interaction networks based on Kinetic Data of Bio-molecular Interaction database (KDBI) to facilitate both qualitative and quantitative study of cellular processes. KDBI is a collection of experimentally determined kinetic data of protein-protein, protein-nucleic acids, protein-ligand, nucleic acids-ligand binding or reaction events described in the literature. Based on KDBI, totally 27 kinetic protein-protein interaction maps of more than 2 levels are created with bioinformatics method as its novel application for quantitative study of cellular processes.

Key words: protein-protein interaction kinetic network bioinformatics

1. 前言

人类基因组计划发现了大量的新基因，然而生命现象的阐明仅仅依靠基因组序列是远远不够的。相对于静态的基因，其编码的产物——蛋白质是动态的，具有时空性和调节性，是生物功能的重要体现者和执行者。蛋白质是生命功能的物质基础，其存在方式、表达水平、以及相互作用等，都与生物功能密切相关。蛋白相互作用一直以来都是现代生物学领域研究的热点及难点。随着基因组学的发展，蛋白质组学逐渐成为分子生物学的研究重点，蛋白质这个与人类疾病最直接相关的基因产物受到更多人的重视。蛋白质相互作用网络的破坏和失稳，可能引发细胞功能障碍，因此，蛋白质相互作用网络的建立有助于发展网络动态模型，寻找合适的药物作用靶点，将为新药开发提供理论依据。

现代生物技术为蛋白质的研究提供了强有力的手段。电泳层析等技术以后，一些新的技术如 X 光衍射、质谱、核磁共振等在蛋白研究中的重要作用日益显现。一些高通量的技术也相继出现，能在更短的时间里鉴定出更多的蛋白质及其相互作用关系，如双杂交系统 (two-hybrid system)、蛋白芯片 (protein chip)、以质谱为基础的蛋白质复合物纯化技术、相关 mRNA 表达谱等技术。这些实验产生了大量的生物数据，急需用高效的手段存储、处理、分析。数据库存储成了最有效的手段。于是随着生物技术和信息技术的发展，产生了大量的生物数据库，蛋白质研究甚至包括整个生命科学的研究都正经历着一场前所未有的变革。利用电脑存储和处理数据的强大功能，可以把实验数据有序保存起来，并与基础理论结合起来，探索实验及人的分析能力难以解决的问题。另一方面，电脑分析后产生的新理论或新的模型又可以通过实验得以验证。当基因组学发展成为蛋白质组学，蛋白质组学进一步发展到更复杂的系统生物学时，有时真实的实验变得难以操作，就要依靠计算机模型模拟了。这就促进了交叉学科——生物信息学的兴起和发展。

生物信息学 (Bioinformatics)——指用计算机辅助数据管理的方法来收集、分析、再现生物的信息以了解生命过程的学科 (Computer-assisted data management discipline that helps us gather, analyze, and represent biological information in order to understand life's processes)。该学

科以信息科学、计算机科学、数学、比较生物学、统计学等学科的观点和方法，以计算机和生物电子设备为工具，研究生命中物质的组成、进化、结构与功能的规律、以及这些物质在生命体中能量和信息的交换或传递，用信息理论技术及数学和物理的理论方法去理解和阐述生物大分子的存在和生命价值，最终对它们进行各种处理与应用。通过这些处理和应用，不仅能理解已有的核酸和蛋白质序列及其功能，而且预测新的基因和蛋白序列及其功能。

1.1 蛋白质相互作用及其研究意义

基因数量的有限和基因结构的相对稳定与生命现象的复杂多变之间存在着巨大反差。这种反差揭示了基因只是遗传信息的载体。要研究生命现象，阐释生命规律，只了解基因组的结构是远远不够的，需要对生命活动的直接执行者——蛋白质的有了更深入的研究。蛋白质是生命功能体现者和执行者，是一切生命现象和生命活动的前提和基础。因此，全面和深入地研究蛋白质的表达量、结构、性质、相互关系和生物学功能等已成为生命科学研究的迫切需要和重要任务。

生命体是一个由大量特异的细胞过程所构成的复杂系统。详尽研究这些细胞过程以及它们之间所进行的信息及能量传递的分子机理是生命科学中至关重要的问题。随着对生物大分子 DNA/RNA 和蛋白质功能的深入研究，尤其是后基因组时代新兴学科如基因组学、蛋白质组学和生物信息学等在大范围内对生物大分子的功能分析，大大地加速了对细胞过程的了解。

在分子水平上，生命的复杂性主要依赖于特异的蛋白质——蛋白质相互作用推动各色的生物组织、信号转导、酶活性和产物调节等细胞活动而产生的。这些细胞功能的调节很大程度上是依靠平衡不同蛋白质作用的结合能力，以及蛋白质与其它生物分子如核酸分子、配体小分子、金属离子等的结合能力而实现的。因此，了解蛋白质相互作用的方式、作用强度、上下游的关系将极大地有助于研究蛋白质的功能、细胞的生命过程和生命发育的探索，并促使我们对疑难杂症的病理研究和新的药物靶蛋白的发掘。

在所有生命活动中，蛋白质之间的相互作用是生物体许多生命活动过程的重要组成部分，是生物体生化反应的基础。它是细胞进行一切代谢活动的基础。蛋白质间相互作用存在于机体每个细胞的生命活动过程中，生物学中的许多现象如复制、转录、翻译、剪切、分泌、细胞的增殖、分化和死亡、细胞周期调控、信

号转导和中间代谢等均受蛋白质间相互作用的调控^[1]。

蛋白相互作用及它们如何一起行使功能是理解生命运动的基础。通过蛋白质间相互作用，可改变细胞内蛋白质的动力学特征，如底物结合特性、催化活性；也可产生新的结合位点，改变蛋白质对底物的特异性；还可失活其它蛋白质，调控其它基因表达，改变蛋白质对其作用底物的专一性等等^[1]。因此，只有使蛋白质间相互作用顺利进行，细胞的正常生命活动过程才有保障。因此，建立相互作用关系网络图，已成为蛋白质组学研究中的热点。

1. 2 蛋白质相互作用的研究方法

研究蛋白质相互作用传统实验方法包括如酵母双杂交系统 (yeast two-hybrid system)^[2]、质谱法 (mass spectrometry method)^[3-6]和蛋白芯片 (protein chips)^[5]等；近年来，随着计算机科学的发展，计算机模拟 (in silico modeling) 成为蛋白质相互作用中一个有力的工具^[6-7]，同实验方法相比，计算生物学方法不仅是实验方法的有价值的补充，且能够弥补目前实验方法难以克服的缺点^[8]，而且能扩展实验方法的预测范围；同时，一些重要的分子进化和分子生物学概念也是在开发这些方法的过程中建立的。

1. 2. 1 传统方法

1. 2. 1. 1 生物物理学方法

酶联免疫吸附检测法 (ELISA)^[9]

体外验证两蛋白质的相互作用是最简单的研究两蛋白质相互作用的方法。如验证所制备的抗体是否与抗原特异性地结合，要求该方法能：一检测到结合作用，二排除非特异性结合，于是建立了酶联免疫吸附检测法 (ELISA)。

蛋白质亲和层析法 (protein affinity chromatography)

基本原理是：将感兴趣蛋白偶联到基质上，如 Sepharose 基质上，并用它从合适的抽提液中筛选能与基质上蛋白结合的蛋白，大部分蛋白通过这样的柱子后，易在低盐条件下被洗脱掉，然后用高盐、辅助因子、溶剂或十二烷基硫酸钠 (SDS) 洗脱保留的蛋白，保留结合能力强的蛋白比弱的更可能与感兴趣蛋白相互作用^[10]。虽然亲和层析是研究蛋白质相互作用的有利工具，但也可能有假阳性结果：蛋白间的相互作用可能不是直接的，而是通过第三方面间接相互作用的；检测到的相互作用可能是蛋白质静电作用；两种在体内不可能相遇却有着极强的亲和

力的蛋白，如 DNase I 与肌动蛋白之间的相互作用^[11]。

亲和印迹 (affinity blotting)

类似于 Western blotting^[12]，将经过 SDS-PAGE 电泳分离后的蛋白样品转移到硝酸纤维膜上，然后检测哪种蛋白能与标记了的“诱饵”蛋白发生作用。由于实验中所用含有 SDS 的变性胶将失活大多数蛋白，并使复合物的亚单位分离，可用非变性胶来解决。影响因素包括蛋白结合到膜上的生物学活性，制备的蛋白探针等(放射性或生物素标记的蛋白及抗体)。

免疫共沉淀法 (immunoprecipitation)

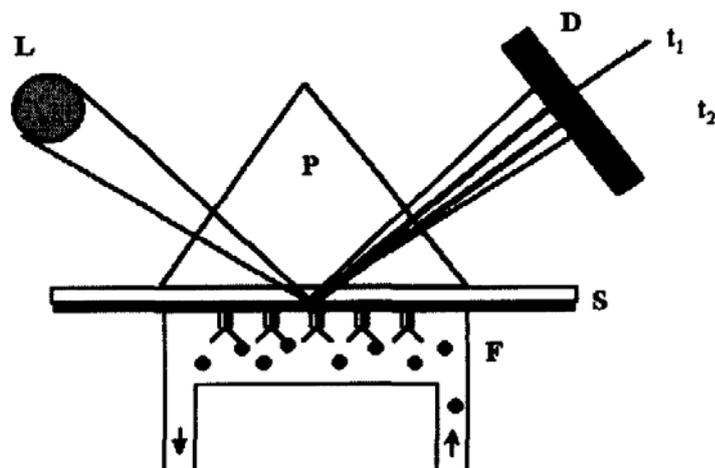
免疫共沉淀法是以抗体和抗原之间的专一性作用为基础的蛋白质间相互作用的经典方法，也是常用方法。其基本原理是：细胞裂解物中加入抗体，与抗原形成特异免疫复合物，经过洗脱后收集免疫复合物，然后进行 SDS-PAGE 及 Western blotting 分析。例如：Filhol 等利用该方法发现酪蛋白激酶 E 与抑癌基因 p53 相关^[13]。

交联技术 (cross-linking)

用于蛋白质相互作用的交联技术可以检测体内或体外的蛋白质相互作用。研究主要有两方面：(1) 分析鉴定相互作用蛋白质复合物的组成；(2) 检测能发生相互作用的蛋白。体内进行的交联技术采用了具有膜渗透性的交联试剂，然后配合免疫共沉淀技术检测到相互作用蛋白。此种方法比单独使用免疫共沉淀技术的检测灵敏度高。

表面等离子共振技术 (Surface Plasmon Resonance, SPR)

表面等离子体共振 (Surface Plasmon Resonance) 技术，如图 1.1 所示，是一种全新的研究蛋白质之间相互作用的手段。该技术是利用表面等离子体共振现象和 SPR 谱峰对金属表面上电介质变化敏感的特点，通过将受体蛋白固定在金属膜上，检测受体蛋白与液相中配体蛋白的特异性结合。



L:光源, D:检测器, P:棱镜, S:传感器表面, F:检测液

图 1.1 表面等离子共振检测系统

Fig 1.1 Surface Plasmon Resonance system

荧光共振能量转移技术 (FRET)

荧光共振能量转移是指两个荧光发色基团在足够靠近(< 100 埃)时, 它们之间可发生能量转移的现象。由于绿色荧光蛋白 (green fluorescence protein, GFP) 受紫外线激发可以发出绿色荧光, 因此它可以作为标记蛋白被广泛用于生命研究领域。

1.2.1.2 分子生物学方法

分子生物学方法的特点是从大量的基因或基因片段产物中筛选能相互作用的编码蛋白, 并且一旦得到目的蛋白, 就有可能直接得到该蛋白的编码基因。

蛋白质探针技术 (Protein probing)

基本原理是: 用目的蛋白为探针筛选表达文库, 在硝酸纤维素膜上探测与该蛋白相互作用的蛋白的编码序列。例如: Blackwood 等利用该方法发现 Myc 与 Max 可相互作用^[14]。

噬菌体展示技术 (Phage display)

噬菌体呈现技术是目前研究生物分子间相互作用、寻找新靶点、研究开发新药的极为有用的方法。其原理是将表达多肽的基因与噬菌体表面蛋白的编码基因融合后, 以融合蛋白的形式表达在噬菌体表面的一种方法。最常用的表达系统是 M13 噬菌体。将 cDNA 文库插入噬菌体载体进行表达后, 所得到的总体为噬菌体

展示库。为了得到与诱饵蛋白相作用的目的蛋白，可将该展示库与固定化的诱饵蛋白相互作用，目的蛋白被吸附下来，被吸附的重组噬菌体可经由再感染而扩增，从而得到目的蛋白的插入基因。还有不依赖于宿主细胞的完全在体外展示的技术，所构建的文库大大超过完整细胞或病毒^[15]。

双杂交技术 (Two hybrid system)

酵母双杂交系统又称蛋白阱捕获系统，是由 S. Fields 和 O-K Song 等人于 1989 年根据真核转录调控的特点创建的。近几年来，该方法成了研究蛋白质间相互作用的一种非常有效的分子生物学方法^[16]。真核细胞的转录因子的 DNA 结合域 (DB) 和转录激活域 (AD) 共同作用时才能使转录正常进行，来自不同转录激活因子的两种结构域也能使转录正常。可分别使 DB 与 AD 同“诱饵”蛋白 (X) 和“猎物”蛋白 (Y) 形成融合蛋白，并在真核细胞内同时表达。如果两者可以发生作用，就能使 AD 与 DB 在空间上充分接近，从而激活报告基因的转录。

1.2.1.3 遗传学方法

Criekingel W V 和 Beyaert R^[17]将蛋白质相互作用的研究比作人们对汽车发动机理的探索，而遗传学方法就像是使某一部件发生缺损，然后检测其对汽车发动的影响。这就意味着在实验中需要建立特殊的表型。遗传学方法通常用于研究两个已知蛋白之间的相互作用。

基因外抑制子 (Extragenic Suppressors)

抑制子有基因内和基因外之分，基因内抑制子是指通过基因自身的一个突变来抑制其原有的突变，而基因外抑制子是通过另外一个基因的突变来弥补原有基因的突变的。

合成致死筛选 (Synthetic Lethal Screening)

合成致死效应是指两个基因同时发生突变时会产生致死效应，而当每个基因单独发生突变时，则无致死效应。

这种遗传学方法可以用于分析两个具有相同重要功能的蛋白质之间的相互作用。合成致死筛选所得到的两种相互作用蛋白可能是同一复合物的两种组分或是一种蛋白对另一种蛋白有调节作用。

1.2.2 生物信息学方法

生物信息学是采用计算机技术和信息论方法研究生命科学中各种生物信息

的表达、采集、储存、传递、检索、分析和解读的科学，是现代生命科学与信息科学、计算机科学、数学、统计学、物理学、化学等学科相互渗透和高度交叉形成的学科。生物信息学提供了蛋白质相互作用研究的高效方法

1.2.2.1 根据编码蛋白的基因预测蛋白间相互作用

系统发育谱(phylogenetic profile)

这个方法基于假定：功能相关的(functionally related)基因，在完全测序的基因组中预期同时存在或不存在，这种存在或不存在的模式(pattern)称作系统发育谱；如果没有序列同源性两个基因的系统发育谱一致或相似，可以推断它们在功能上是相关的^[18-19]。

基因邻接(gene neighborhood)

在细菌基因组中，功能相关的基因紧密连锁地存在于一个特定区域，这种基因之间的邻接关系，在进化过程中具有保守性，可以作为基因产物之间功能关系的指示^[20-22]。这种指示只能适用于进化早期的结构简单的微生物。

基因融合事件(gene fusion event)

这个方法基于假设：由于在物种演化过程中发生了基因融合事件，一个物种的两个(或多个)相互作用的蛋白，在另一个物种中融合成为一条多肽链，因而基因融合事件可以作为蛋白质功能相关或相互作用的指示。Marcotte 等^[23]和 Enright 等^[24]几乎同时建立了这个方法。这个方法的限制是，不能判断发生融合的蛋白是否“物理”上直接接触，此外，基因融合的机制也可能是复杂多样的。

1.2.2.2 根据进化过程中蛋白发生的变化预测蛋白间相互作用

Valencia 等人对发育谱系方法作了进一步改进，发展了 mirror tree 和 i2h 法。他们认为相互作用的蛋白由于进化压力可能会具有相似的进化关系。首先找到被考查蛋白在不同组织的基因组中的同源蛋白，然后分别比较它们之间的进化关系，用距离矩阵或相关系数分布进行量化。最后根据蛋白进化的相似程度预测其相互作用关系。

镜像树(mirrotree)

功能相关的蛋白质或同一个蛋白的域之间，受功能约束，其进化过程应该保持一致，即呈现共进化(co-evolution)特征^[25-26]。通过构建和比较它们的系统发育树，如果发现树的拓扑结构显示相似性，这种相似的树被称作镜像树。那么，

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库