

蛋白质 - 蛋白质分子对接方法研究进展 *

李春华 马晓慧 陈慰祖 王存新 **

(北京工业大学生命科学与生物工程学院, 北京 100022)

摘要 蛋白质分子间相互作用与识别是 21 世纪生命科学研究的前沿和热点. 分子对接方法是研究这一课题有效的计算机模拟手段. 通常, 蛋白质 - 蛋白质分子对接包括四个阶段: 搜索受体与配体分子间的结合模式, 过滤对接结构以排除不合理的结合模式, 优化结构, 用精细的打分函数评价、排序对接模式并挑选近天然构象. 结合国内外研究蛋白质 - 蛋白质分子对接方法进展和本研究小组的工作, 对以上四个环节做了详细的综述. 另外, 还分析了目前存在的主要问题, 并提出对未来工作的展望.

关键词 蛋白质 - 蛋白质分子对接, 过滤, 打分函数, 分子柔性

学科分类号 Q617

蛋白质分子间相互作用与识别是 21 世纪生命科学研究的前沿和热点. 细胞内各种重要的生理过程, 如基因的复制、转录、翻译以及细胞周期调控、信号转导、免疫反应等, 都是以蛋白质分子间相互作用为纽带进行的. 随着计算机处理能力的不断增强, 理论模拟方法得到迅速发展和广泛的应用. 采用分子对接方法研究蛋白质分子间相互作用与识别, 进而预测复合物结构, 已成为研究该问题的有力的计算机模拟手段, 并可为实验工作提供有益的理论指导.

为推动分子对接方法的发展, 欧洲生物信息学中心举办了 CAPRI (Critical Assessment of Prediction of Interactions) 蛋白质 - 蛋白质复合物结构预测竞赛^[1]. 自 2001 年开始, 迄今已成功进行了 9 届国际对接竞赛, 大大促进了分子对接方法学的研究^[2,3]. 目前, 一些新的分子对接算法相继推出^[4-7], 并取得了一系列成功的预测范例^[2,3]. 2002 年, 我们小组作为中国的唯一代表参加了第 2 届 CAPRI 竞赛, 并取得了较好的成绩^[8], 之后连续参加了 7 届^[9].

所谓分子对接就是已知两个分子的三维结构, 考察它们之间是否可以结合, 并预测复合物的结合模式. 根据对接中两分子所处的状态, 可将分子对接分为三种类型: 第一种, 复合态分子对接(bound docking), 对接中两分子均直接取自复合物晶体结构中的相应部分, 即复合态(bound)结构; 第二种,

自由态分子对接(unbound docking), 顾名思义, 这种对接起始于复合物形成前的两个分子的单体结构, 即自由态(unbound)结构; 第三种, 同源分子对接(homology docking), 通常, 已知一个分子的三维结构以及另一个分子的一级序列和与其同源的分子结构, 需要先模建后一分子的结构, 然后才能进行对接. 显然, 以上三种分子对接类型的困难程度是依次增加的. 在复合态分子对接中, 因不需要考虑分子的柔性, 结构预测相对容易. 但在实际复合物形成中, 受体和配体分子的构象是要发生变化的, 因此, 模拟真实情况中分子结合的自由态对接就要困难得多. 而同源分子对接, 由于一个分子结构的不确定性, 就更加具有挑战性.

一般情况下, 分子对接包括四个阶段: a. 搜索受体与配体分子的结合模式; b. 利用生物学信息或简单的分子表面几何互补性等评价标准过滤排除不合理的结构; c. 对获得的结构进行能量优化, 允许氨基酸残基侧链和骨架的运动; d. 用精细的打分函数评价、排序对接模式并挑选近天然构象. 下面, 结合国内外研究蛋白质 - 蛋白质分子对接方法的进展情况和本小组的工作, 对以上四个环节做详细的

*国家自然科学基金资助项目(30400087,10574009), 北京市自然科学基金(5042003)和国家教育部博士点基金(2004005013)资助项目.

** 通讯联系人. Tel: 010-67392724, Fax: 010-67392837

E-mail: cxwang@bjut.edu.cn

收稿日期: 2006-01-17, 接受日期: 2006-02-28

综述.

1 复合物构象空间的搜索

分子对接首先要进行受体与配体分子结合模式的空间搜索. 对于蛋白质 - 蛋白质分子对接, 数目庞大的原子数和自由度数使系统搜索法已不能胜任全空间的构象采集. 目前常用的全空间搜索算法有快速傅立叶变换(fast Fourier transform, FFT)算法、遗传算法(genetic algorithm, GA)和蒙特卡罗(Monte Carlo, MC)算法.

Katchalski-Katzir 等^[10]首次将 FFT 用于分子对接方法中. 受体和配体分子被投影到三维空间网格

中, 只要过滤评价函数可表示为 $\sum_{ij} p_i q_j$ 的形式(其中 p_i 和 q_j 分别为受体和配体分子某种特性的离散函数), 那么就可以利用 FFT 加快函数的计算, 从而提高采样速度. FTDock^[10]方法对受体和配体分子的几何形状进行离散化, 以几何互补性标准筛选结合模式. 之后, Sternberg 小组对该方法进行了改进, 发展了 3D-Dock 程序^[11]. 改进后的方法不仅考虑了受体和配体分子表面几何互补性, 而且增加了静电互补性评价标准. 静电相互作用能表示为受体分子在格点产生的静电势与配体分子格点所带电荷的乘积. 由于 FFT 算法的高效性, 它被广泛地使用, 产生了一系列蛋白质 - 蛋白质分子对接程序, 如 ZDOCK^[4]、GRAMM^[12]、DOT^[13]、Modified FT_DOCK^[14]、SmoothDock^[6]、ClusPro^[7]和 MolFit^[15]. 其中, ZDOCK^[4]是目前复合物结构预测较为成功的方法之一. 它不仅考虑了受体和配体分子表面几何互补性和静电相互作用能, 还考虑了去水化自由能的贡献.

Gardiner^[16]和 Taylor^[17]则利用 GA 算法进行构象搜索. Gardiner^[16]采用溶剂可接近表面来描述蛋白质分子, 并标有法向矢量、曲率和氢键特性, 以表面几何匹配性来筛选对接结构. Taylor^[17]则采用分子势能作为适应性函数来淘汰或保留对接构象. RosettaDock 程序^[5,18]采用 MC 算法进行全空间构象搜索, 整个过程包括低分辨率采样和高分辨率优化两个阶段. 在前一阶段, 用氨基酸骨架原子和侧链质心来构建分子模型, 用 MC 算法进行空间搜索, 用一个低分辨率的势函数来判断结构保留与否. 在第二阶段, 所有的重原子和极性氢原子被复原, 再一次用 MC 算法来优化并安装侧链, 此时目标函数是更加完整的全原子势函数. RosettaDock 也是目前

较为成功的复合物结构预测算法之一, 该思想还被成功地用在蛋白质折叠结构预测中^[19].

除此之外, 还有一些其他的方法用于构象搜索, 如球极傅立叶相关(spherical polar Fourier correlation)^[20]、构象空间退火(conformational space annealing)^[21]和分子相互作用场(molecular interaction field)^[22]方法. 另外, 值得一提的是, 面对更加具有挑战性的蛋白质聚集体复合物结构预测, 国际上一些研究小组已经在发展同源多聚体的分子对接方法了^[23].

2 过滤方法

构象采集是分子对接程序的初始环节, 获得大量有效的复合物结构对后期打分排序至关重要. 目前主要在对接过程中的两个阶段(采样期和采样后期), 通过引入生物学信息或简单的评价函数来排除一些不合理的结合模式, 提高结构预测的成功率.

在采样后期, 3D-Dock 程序^[11]将活性位点信息加入到几何约束上进行构象过滤, 这需要知道确切的界面残基对, 或者至少是位于复合物界面上的残基. ZDOCK^[4]用一个经验的函数来筛选复合物结合模式, 其中考虑了受体与配体分子表面几何互补性、静电相互作用能和去水化自由能的贡献. 我们小组将复合物界面残基成对偏好性引入到几何匹配过滤中, 发展了双重过滤技术来筛选合理的结合模式, 获得了更多的近天然结构数^[24]. 另外, 又提出了依赖于复合物类型的过滤方案^[25]. 根据复合物的界面特性, 对蛋白酶 - 抑制剂类、抗原 - 抗体类和其他类复合物采用不同的组合过滤方法, 提高了结构预测的成功率. 我们研究小组采用该方法在 CAPRI 第 2 届结构预测竞赛上取得了较好的成绩^[8].

采样后期的过滤处理对于近天然构象采集不足的局面则不能给予任何改善. 因此进行结合位点附近的集中采样成为分子对接中的主要采样策略. Ben-Zeev 和 Eisenstein 等^[26]用权重几何对接方法, 通过赋给分子间结合位点残基有接触的构象更大的几何互补分值来进行偏好性采样. 在 BESGDock^[9]中, 我们采用 FFT 算法进行结合模式的搜索, 在对蛋白质三维格点进行几何离散化时, 将受体或配体分子结合位点中心一定半径以外的格点设置为 0, 其他格点的赋值与 FTDock 方法相同. 采样中, 只有球内部的格点对几何互补打分有贡献, 而球外部的格点则无贡献. 这样, 会有更多的受体或配体

分子结合位点处于复合物界面的结构被搜索到,从而大大提高了采样有效性.用该方法对 2003~2004 年 CAPRI 国际竞赛上提供的 10 个蛋白质复合物分子(<http://capri.ebi.ac.uk>)进行测试^[9],结果表明,把生物信息加入到早期采样阶段较后期过滤阶段更有助于提高采样获得的近天然结构数,使预测成功率明显提高.

当过滤完成之后,对接模式的数量将大大减少,在用精细的能量函数对它们进行打分之前,优化体系的结构是非常必要的^[27].其中涉及到分子柔性的问题将在第 4 部分介绍.

3 打分函数的设计

成功的分子对接除了需要一个有效的算法进行构象空间搜索外,还需要有一个合理敏感的打分函数用于近天然构象的挑选.目前,对接打分函数主要分为以下三类:基于物理的能量函数;经验函数;基于知识的函数.基于物理的能量函数是利用“热力学主方程”进行自由能预测.这种方法已被 DARWIN^[17]、Guided Docking^[28]、TSCF^[29]、Smooth-Dock^[6]、Northwestern DOCK^[30]和 Softdock^[25]等对接方法所采纳.经验函数法则考虑了多种因素的贡献,如残基成对偏好性、几何互补性及静电、氢键、疏水相互作用能等.当然,求和中各项会加以适当的权重. DOT^[13]、BIGGER^[31]、GRAMM^[12]、ZDOCK^[4]、ICM-DISCO^[32]、RosettaDock^[5]、BESGDock^[9]和 PPD^[33]均采用了这类方法.基于知识的函数是利用 Boltzmann 分布对已有的蛋白质结构数据库进行统计分析而获得的,诸如残基-残基接触势^[34]、残基成对偏好性^[35]和原子-原子接触势^[36]等.

尽管目前打分函数已经取得了较好的进展.但是迄今仍无一个打分函数能很好地适应于所有体系的分子对接.这主要是由于不同类型复合物界面在氨基酸组成、疏水性和静电性等方面有明显的差异^[37-39].我们以前的研究工作^[25]也说明,针对不同复合物类型采用不同的过滤标准在一定程度上能改善对接结果.因此,我们基于蛋白质复合物的生物学功能,首次提出了蛋白质-蛋白质复合物类型依赖的组合打分函数^[27,40].该打分函数组合了统计势和基于知识的势函数,应用多元线性回归方法来拟合各项权重.结果显示,对蛋白酶-抑制剂、抗原-抗体和其他酶-抑制剂类复合物,该打分函数较一些常用的打分函数在近天然结构的区分能力上有明显

的改善和提高.

4 蛋白质分子柔性的考虑

对接中分子柔性的处理是每次 CAPRI^[2,3]竞赛所关注的焦点.目前考虑蛋白质分子柔性的方法归纳起来有五种.第一种最简单的柔性处理方法是,通过在分子表面定义一个柔软壳层来允许分子表面原子间可以发生一定程度的交叠. FTDock^[10]在对蛋白质受体三维空间网格进行几何离散化时,对表面一定厚度内的格点赋予与内部不同的值,从而在计算几何互补性时,允许一定的分子间交叠.我们建立的 Softdock^[25]和 Palma 小组发展的 BIGGER^[31],进一步考虑了蛋白质表面氨基酸残基 Arg、Lys、Asp、Glu 和 Met 的侧链柔性,允许这些残基比其他表面残基有更大程度的可交叠性.所不同的是 Softdock 用简化蛋白质模型(一个氨基酸残基用一个球来表示),而 BIGGER 采用格点模型来模拟分子的表面.在这些算法中,分子表面软的程度是以一个常数的形式设置的.因此不难想到,其对接结果的好坏在很大程度上依赖于复合物形成过程中受体和配体分子构象变化的程度.

第二种考虑蛋白质分子柔性的方法是采用分子走向(trace)模型^[5,41].在对接采样中,分子模型仅使用 C α 原子或蛋白质骨架或骨架加部分侧链原子来搭建,搜索结束后再安装并优化侧链.这种低分辨率的对接考虑了侧链原子的运动,运算速度快,其思想引起了人们很大的兴趣和关注. RosettaDock^[5]就属于这类方法.

第三种柔性处理方法是 Abagyan 和其同事在 Scheraga 等^[42]提出的方法基础上设计的,是 ICM (internal coordinate mechanics) 和 ICM Pseudo-Brownian 算法的组合^[32].该方法没有考虑蛋白质分子大的骨架运动,但具有一定的预测成功率^[43].

第四种方法考虑了分子铰链弯曲(hinge-bending)的运动^[15,44]. MolFit^[15]和 PatchDock^[44]分子对接程序就采用了这种方法,主要处理结构域间发生运动的情况.根据一定的判断,分子被分为几个部分,各部分间由预先定义的铰链相连接,并可以绕铰链发生相对运动.然后分别对接各部分结构并进行组装.这种方法需要预先设定分子铰链的位置,而通常这一信息是不知道的.

第五种方法是采用多构象模型.利用侧链旋转异构体库在结构优化中考虑氨基酸残基侧链的柔性^[5,19,45]就属于这一类方法. Mutidock 程序^[45]在采用

该方案时还引入了平均场和概率加权平均的思想. RosettaDock 程序^[9]采用骨架依赖的侧链旋转异构体库, 结合模拟退火蒙特卡罗算法来安装并优化侧链, 是侧链结构预测中较成功的方法之一^[19]. 用多构象模型考虑分子柔性的思想也被用在蛋白质主链柔性的处理中^[9,14,46,47]. Bates^[14]和 Bonvin^[46]小组在其开发的分子对接算法中, 通过分子动力学模拟和主成分分析等方法产生多个受体和配体的构象, 然后进行多次分子对接. ED-Hex 方法^[47]采用 Essential Dynamics 和 Polar Fourier Transform 方法来对接多个分子构象. 在我们发展的 BESGDock 程序^[9]中, 采用多构象叠落的方法考虑蛋白质结合位点区域主链的柔性. 该方法对结合位点处 Loop 区有构象变化的复合物结构预测显示了一定的优势.

5 问题及展望

目前, 蛋白质 - 蛋白质对接方法中主要存在三大问题. 首先, 分子模型过于简单, 柔性的考虑仍然不够充分, 还不能成功地处理复合物形成中较大的构象变化, 如蛋白质分子域间类似铰链的运动. 其次, 势函数对近天然结构的区分能力有待提高. 常用的势函数把一个多粒子体系量子力学水平上的相互作用, 以原子对间相互作用的和来表示, 这显然过于粗略了. 半经验力场虽然可以处理包含几千到一万个原子的体系, 但结果仍然有一定的近似性. 另外, 对多自由度的复合物体系, 如何找到结合自由能曲面上的全局极小, 迄今仍未发现有效的办法来解决.

长期以来, 科学家们正在努力解决这些问题. 对物理模型的改进主要制约因素是计算能力, 量子力学理论对描述多粒子体系是严格的, 后来发展的 Hartree-Fock-Roothaan 方法、密度泛函方法等, 在处理分子体系时虽有一定近似, 但物理模型是合理的. 只是由于生物分子粒子数过多, 计算能力不允许才不得不使用分子的简单模型和半经验势函数. 当前速度达约 10^{14} Flops/s 的计算机已研制成功. 随着计算机能力的快速提高和实验数据的不断增加, 使用复杂力场和精确势参数的可能性越来越大, 分子模型及柔性的考虑会越来越真实. 关于极值搜寻问题, 目前对它的讨论与改进是很多的. 要根本解决这一问题, 必须在理论和算法上有实质性的突破. 当然, 了解真实发生在生物体内的事件也是很重要的. 蛋白质尽管自由度极大, 但在生物体中最终都形成了确定的天然态, 可以说多极值问题在活

体中已得到解决. 所以要确定蛋白质复合物的构象, 应该把生物学信息加到理论模拟中去, 这也许是解决这一问题的办法. 现在蛋白质分子设计中非常有效的同源建模方法, 就是依赖于结构生物学提供的信息实现的. 这一点令很多科学家感到鼓舞并充满希望.

参考文献

- Janin J, Henrick K, Moult J, *et al.* CAPRI: a critical assessment of predicted interactions. *Proteins*, 2003, **52** (1): 2~9
- Méndez R, Leplae R, Maria L D, *et al.* Assessing of blind predictions of protein-protein interactions: current status of docking methods. *Proteins*, 2003, **52** (1): 51~67
- Méndez R, Leplae R, Maria L D, *et al.* Assessment of CAPRI predictions in rounds 3~5 shows progress in docking procedures. *Proteins*, 2005, **60** (2): 150~169
- Chen R, Li L, Weng Z P. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, 2003, **52** (1): 80~87
- Gray J J, Moughan S E, Wang C, *et al.* Protein-protein docking with simultaneous optimization of rigid body displacement and side chain conformations. *J Mol Biol*, 2003, **331** (1): 281~299
- Camacho C J, Gatchell D W. Successful discrimination of protein interactions. *Proteins*, 2003, **52** (1): 92~97
- Comeau S R, Gatchell D W, Vajda S, *et al.* ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, 2004, **20** (1): 45~50
- Li C H, Ma X H, Chen W Z, *et al.* A soft docking algorithm for predicting the structure of antibody-antigen complexes. *Proteins*, 2003, **52** (1): 47~50
- Ma X H, Li C H, Shen L Z, *et al.* Biologically-enhanced sampling geometric docking and backbone flexibility treatment with multi-conformational superposition. *Proteins*, 2005, **60** (2): 319~323
- Katchalski-Katzir E, Shariv I, Eisenstein M, *et al.* Molecular surface recognition: determination of geometric fit between protein and their ligands by correlation techniques. *Proc Natl Acad Sci USA*, 1992, **89** (6): 2195~2199
- Aloy P, Querol E, Aviles F X, *et al.* Automated structure based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. *J Mol Biol*, 2001, **311** (2): 395~408.
- Vakser I A. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins*, 1997, **1**(Suppl): 226~320
- Mandell J G, Roberts V A, Pique M E, *et al.* Protein docking using continuum electrostatics and geometric fit. *Prot Eng*, 2001, **14** (2): 105~113
- Smith G R, Sternberg M J E, Bates P A. The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with application to protein-protein docking. *J Mol Biol*, 2005, **347** (5): 1077~1101

- 15 Berchanski A, Shapira B, Eisenstein M. Hydrophobic complementarity in protein-protein docking. *Proteins*, 2004, **56** (1): 130~142
- 16 Gardiner E J, Willett P, Artymiuk P J. Protein docking using a genetic algorithm. *Proteins*, 2001, **44** (1): 44~56
- 17 Burnett R M, Taylor J S. DARWIN: a program for docking flexible molecules. *Proteins*, 2000, **41** (2): 173~191
- 18 Wang C, Schueler-Furman O, Baker D. Improved side-chain modeling for protein-protein docking. *Prot Sci*, 2005, **14** (5): 1328~1339
- 19 Rohl C A, Strauss C E, Misura K M, *et al.* Protein structure prediction using Rosetta. *Methods Enzymol*, 2004, **383** (1): 66~93
- 20 Ritchie D W, Kemp G J L. Protein docking using spherical polar fourier correlations. *Proteins*, 2000, **39** (2): 178~194
- 21 Lee K, Czaplewski C, Kim S Y, *et al.* An efficient molecular docking using conformational space annealing. *J Comput Chem*, 2005, **26** (1): 78~87
- 22 Ritchie D W. Evaluation of protein docking predictions using Hex 3.1 in CAPRI Rounds 1 and 2. *Proteins*, 2003, **52** (1): 98~106
- 23 Pierce B, Tong W, Weng Z P. M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. *Bioinformatics*, 2005, **21** (8): 1472~1478
- 24 Li C H, Ma X H, Chen W Z, *et al.* A soft docking algorithm for predicting the structures of protein-protein complexes. *Acta Biochim Biophys Sin*, 2003, **35** (1): 35~40
- 25 Li C H, Ma X H, Chen W Z, *et al.* A protein-protein docking dependent on the type of the complexes. *Prot Eng*, 2003, **16** (4): 265~269
- 26 Ben-Zeev E, Eisenstein M. Weighted geometric docking: incorporating external information in the rotation-translation scan. *Proteins*, 2003, **52** (1): 41~46
- 27 Shen L Z, Ma X H, Li C H, *et al.* Complex-type-dependent scoring functions in protein-protein docking. *European Biophysics Journal with Biophysics Letters (EBSA)*, 2005, **34** (6): 745
- 28 Fitzjohn P W, Bates P A. Guided docking: first step to locate potential binding sites. *Proteins*, 2003, **52** (1): 28~32
- 29 Komatsu K, Kurihara Y, Iwadata M, *et al.* Evaluation of the third solvent clusters fitting procedure for the prediction of protein-protein interactions based on the results at the Capri blind docking study. *Proteins*, 2003, **52** (1): 15~18
- 30 Lorber D M, Udo M K, Shoichet B K. Protein-protein docking with multiple residue conformations and residue substitutions. *Prot Sci*, 2002, **11** (6): 1393~1408
- 31 Palma P N, Krippahl L, Wampler J E, *et al.* BIGGER: a new soft docking algorithm for predicting protein interactions. *Proteins*, 2000, **39** (4): 372~384
- 32 Fernandez-Recio J, Totrov M, Abagyan R. Identification of protein-protein interaction sites from docking energy landscapes. *J Mol Biol*, 2004, **335** (3): 843~865
- 33 Norel R, Sheinerman F, Petrey D, *et al.* Electrostatics contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Prot Sci*, 2001, **10** (11): 2147~2161
- 34 Zhang C, Liu S, Zhou Y. Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential. *Prot Sci*, 2004, **13** (2): 391~399
- 35 Moont G, Gabb H A, Sternberg M J E. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins*, 1999, **35** (3): 364~373
- 36 Zhang C, Vasmatzis G, Cornette J L, *et al.* Determination of atomic desolvation energies from the structures of crystallized proteins. *J Mol Biol*, 1997, **267**(3): 707~726
- 37 Conte L L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol*, 1999, **285** (5): 2177~2198
- 38 Jones S, Thornton J M. Principles of protein-protein interactions. *Proc Natl Acad Sci USA*, 1996, **93** (1): 13~20
- 39 Jackson R M. Comparison of protein-protein interactions in serine protease-inhibitor and antibody-antigen complexes: implications for the protein docking problem. *Prot Sci*, 1999, **8** (3): 603~613
- 40 Shen L Z, Li C H, Ma X H, *et al.* Scoring function for the other-type protein complexes. *Acta Phys Chim Sin*, 2006, **22** (5): 623~627
- 41 Tovchigrechko A, Vakser I A. How common is the funnel-like energy landscape in protein-protein interactions?. *Prot Sci*, 2001, **10** (8): 1572~1583
- 42 Li Z. Scheraga H A. Monte carlo-minimization approach to the multiple minima problem in protein folding. *Proc Natl Acad Sci USA*, 1987, **84** (19): 6611~6615
- 43 Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme-antibody complex with a 1.6 Å accuracy. *Nat Struct Biol*, 1994, **4** (2): 259~263
- 44 Schneidman-Duhovny D, Inbar Y, Nussinov R, *et al.* Geometry-based flexible and symmetric protein docking. *Proteins*, 2005, **60** (2): 224~231
- 45 Jackson R M, Gabb H A, Sternberg M J E. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J Mol Biol*, 1998, **276** (1): 265~285
- 46 Dominguez C, Boelens R, Bonvin A M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, 2003, **125** (7): 1731~1737
- 47 Mustard D, Ritchie D W. Docking essential dynamics eigenstructures. *Proteins*, 2005, **60** (2): 269~274

Progress in Protein-protein Docking Approaches*

LI Chun-Hua, MA Xiao-Hui, CHEN Wei-Zu, WANG Cun-Xin**

(College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100022, China)

Abstract Protein-protein interactions and recognition are the focal and hot themes of the life science field in the 21st century. Molecular docking approach is the effective computer modeling technology in the study of this topic. Generally, the protein-protein docking procedure is composed of four stages: searching of the binding modes of the receptor and the ligand, filtering of docked modes to eliminate the irrational docked structures, optimizing the structures, evaluating the docked modes with the refined scoring function and ranking them to obtain the near-native structures. Combining the research group's works, in terms of the international and national progress of protein-protein docking approaches, the detailed review was made about the four stages mentioned above. Additionally, the existing major questions are analyzed and the prospects of the future study are made.

Key words protein-protein docking, filtering, scoring function, molecular flexibility

*This work was supported by grants from The National Natural Sciences Foundation of China (30400087, 10574009), Beijing National Science Foundation (5042003) and The National Research Foundation for the Doctoral Program of Ministry of Education of China (2004005013)..

**Corresponding author . Tel: 86-10-67392724, Fax: 86-10-67392837, E-mail: cxwang@bjut.edu.cn

Received: January 17, 2006 Accepted: February 28, 2006