

# Progress in Computing Technology of Molecular Docking

XU Chen-yan<sup>1</sup>, LUO Zhi-gang, ZHANG Xiang-lilan

School of Computer Science, National University of Defense Technology, Hunan Changsha 410073

Email: 1.xuchenyan@-163.com

**Abstract:** Molecular docking is an important method of drug design based on receptors structure. Molecular docking processing must search for complex's conformations as much as possible, and evaluate these conformations searched in last step as accurate as possible. Molecular docking requires strong computing ability to finish the docking process. The existing molecular docking methods usually simplify the characteristics of molecular flexibility, scoring functions and physical model to achieve fast and effective computation. Parallel computing is the means which has provided technical support to large-scale scientific computing, so parallel computing makes efficient computing of molecular docking possible. This paper gives a brief introduction about the principles and concept of molecular docking, summarizes typical molecular docking algorithms and some molecular docking software which are widely used in the field of drug design in detail. Then it introduces applications of parallel computing technology in the field of molecular docking studies, and analyzes problems and trends of molecular docking studies.

**Keywords:** computer applications; molecular docking; parallel computing

## 分子对接计算技术的研究进展

徐琛彦<sup>1</sup>, 骆志刚, 张湘莉兰

国防科技大学计算机学院, 湖南 长沙 410073

Email: 1.xuchenyan@163.com

**【摘要】**分子对接是基于受体结构进行药物设计的重要方法。分子对接必须尽可能多地搜索复合物可能的构象, 必须尽可能准确、快速地完成所有搜索到的构象的打分评估。分子对接计算需要强大的计算能力来支持。现有的分子对接计算通常采取简化分子柔性特点和打分评估的模型等方法, 实现有效的计算, 而并行计算为分子对接提供了技术支持, 是一种实现高效分子对接的有效手段。本文在对分子对接的典型算法和相关软件进行详细总结的基础上, 对并行计算技术在分子对接领域的应用研究进行介绍, 分析了分子对接研究的问题及可能的发展趋势。

**【关键词】** 计算机应用; 分子对接; 并行计算

### 1 引言

近年来, 随着计算机辅助药物设计研究的不断深入, 各种理论计算方法和分子模拟技术不断涌现。而X射线衍射等技术的广泛应用, 有越来越多生物大分子的三维结构被测定, 使得基于受体结构的药物设计更为现实。分子对接是基于受体结构进行药物设计的重要方法。通过分子对接可以进行全新药物的设计, 还可以对已有的药物分子库进行筛选, 寻找合适的药物分子, 避免了烦琐的合成过程。

分子对接中寻找复合物低能构象的过程是一个NP难问题<sup>[1]</sup>。高效的分子对接要求过程快速、结果精确。随着并行计算技术和计算机体系结构的发展, 计

算机计算能力迅速提高, 使得进行更高精度、更快速度的分子对接成为可能。

### 2 分子对接概念

#### 2.1 分子对接原理

分子对接是将两个或多个分子通过几何匹配和能量匹配相互识别的过程。分子对接过程首先在受体表面产生一个可以对其进行填充的凹陷的球集, 据此确定一系列活性位点, 然后计算配体和受体的结合模式和亲和力, 并评判其结合程度。由此可见, 分子对接包含了两个部分内容: 构象搜索和构象打分评估。构象搜索是通过既定策略寻找配体和受体在相互识别过程中为相互“适应”所产生的复合物的不同构象; 而构

象打分评估是对搜索到的复合物不同构象进行结合程度的评价。

通过计算机模拟的分子对接过程，会在物理模型或计算过程中存在的一定程度的简化。根据分子构象处理的简化程度不同，可以将分子对接分成刚性对接、半柔性对接、柔性对接三类。

刚性对接是指在计算过程中，配体和受体的构象均不发生变化，仅改变其空间位置与姿态。刚性对接方法的简化程度最高，计算量相对较小，适合于处理大分子之间的对接。例如，Stoddard等人<sup>[2]</sup>把麦芽糖和蛋白质皆当作刚性体，实现两者的刚性对接。

半柔性对接则是指在对接过程中配体构象发生一定程度的变化，而受体构象不变的对接。该方法兼顾了计算量与模型预测能力，是应用比较广泛的对接方法之一，适用于小分子配体和大分子受体之间的对接<sup>[3]</sup>。

柔性对接是指在对接过程中配体和受体的构象可以自由变化的对接。1999年，Mangoni等人<sup>[4]</sup>用柔性的配体对柔性的受体进行了对接。柔性对接方法的计算量非常大，消耗计算时间很多，以柔性小分子为例，粗略估计其搜索的合理构象至少有 $10^{30}$ 个<sup>[5]</sup>。因此，柔性对接更适合精确考察分子间识别情况。

## 2.2 构象搜索方法

分子对接的构象搜索方法依据构象采样方式的不同可以分成系统方法、随机方法和确定性方法

### 2.2.1 系统方法

系统方法是指以固定的系统采样方式对配体的构象空间进行搜索，从而进行对接。系统方法还可分成构象搜索方法、片段生长法、构象库方法。所谓构象搜索方法是指以一定的增量对配体做 $0-360^\circ$ 的旋转，获得配体所有可能的构象。使用构象搜索方法时，构象的数目与增量有关，且呈指数增长。片断生长方法是将配体分割成多个刚性片断，在将配体的核心片断与受体进行对接之后，通过生长的方式连接各个片断，重新构成配体，常用的对接软件DOCK所使用的搜索方法就是片断生长法。构象库方法，首先搜集配体可能的构象，建立配体的构象库，通过对构象库的检索实现配体的柔性处理。

### 2.2.2 随机方法

随机方法是指用事先定义好的概率函数，对构象空间进行随机的采样，最后得到近似最优的对接结果。采样过程中接受的构象则可以用于随机产生新的构

象。常用的随机方法包括遗传方法、蒙特卡罗方法和模拟退火算法。

遗传算法是以生物界自然选择和自然遗传机制为基础的随机化搜索算法。遗传算法模拟生命进化机制，通过一组遗传算子的操作，按一定的随机规则，在解空间进行迭代搜索，直到求得问题的最优解。

蒙特卡罗方法以固定的概率模型为基础，计算该模型的统计特征，最后给出问题的近似解。在构象搜索过程中，首先给出初始构象，在该构象上进行随机的微小改变产生新的构象，计算相应的能量变化，当变化值小于等于阈值时接受新构象，并将其作为下一步的初始构象，否则由随机数决定是否接受该构象。蒙特卡罗方法对低能构象取样几率大，但其计算量也相当大。

模拟退火算法是将物理退火过程与组合优化相结合的一种随机迭代寻优算法。它与蒙特卡罗方法不同的是当新构象能量大于当前构象能量时，改用概率函数判断，是否接受新构象。算法从选定初始构象开始，以 $t = kT$ 为递减的控制参数 ( $k$ : 波尔兹曼常数,  $T$ : 温度) 进行迭代过程，当 $t$ 趋于0时，获得最优构象。相对蒙特卡罗法，模拟退火算法由于采用了重新采样法，计算量显著减少，且尽可能地避免陷入局部最优解。

### 2.2.3 确定性方法

确定性搜索方法包括了禁忌搜索方法和分子动力学方法。

禁忌搜索算法是一种通过局部搜索机制和严格的禁忌准则来避免重复搜索的确定性迭代优化算法。它会记录已搜索过的局部最优解，并在之后的搜索中避免对其的搜索。由于禁忌搜索采用了严格的禁忌准则，搜索途径单一化，搜索结果容易陷入局部最优。为此常通过“破禁”来释放一些被禁忌的优良状态，从而最终实现全局最优化。

分子动力学方法<sup>[6]</sup>通过数值积分方法由各个粒子不同时刻的受力情况来寻求它可能的运动轨迹，最终完成对接的方法。分子动力学方法所需计算时间过长，目前仍不能广泛的应用。

## 2.3 打分评估方法分类

通过构象搜索产生了众多的候选构象，要获得近似最优的构象，则必须对其进行打分，评价配体和受体之间的结合程度。常用的打分函数分成基于力场的打分函数、基于经验的打分函数、基于知识的打分函数和一致性的打分函数四类。

基于力场打分函数计算的是分子力学各项能量的总和。计算时,通过计算配体-受体之间、配体内部和受体内部的各种不同类型力学作用所产生的能量,并对其进行求和,从而获得复合物的能量。基于力场的能量打分通常过分强调分子的极性作用,忽略或简化处理溶解等作用,原子电荷的计算快速但不准确<sup>[7]</sup>。

基于经验的打分函数通过对分子对接中结合自由能不同作用项(包括氢键、范德华等作用)的加和,来获得打分结果。其中,各项的系数是由训练集训练得到的。假使来自一个训练集中的样本结构差异很大,所获得各项系数并不准确,则能量计算的准确性受到影响。

基于知识的打分函数是将原子间距离的概率分布转化为与距离有关的受体-配体原子对间的作用能,实现受体-配体原子对作用的加和。它简化的原子间的作用,可实现化合物数据库的有效搜索,减少计算成本,但该类打分计算难以预测配体的行为,且其模拟的是重复试验的结构而非结合能。

一致性的打分函数是将不同打分方法获取的信息进行综合,互相补充的方法。在对接时采用多种打分函数分别进行评价打分,取多数评价结果优秀的小分子作为对接结果,由此提高了寻找到最优对接结果的概率。

### 3 分子对接的应用及常用分子对接软件

目前,已经有多个分子对接软件被广泛地使用,包括:DOCK、AUTODOCK、HEX、MVD等。这些软件采用了傅里叶变换、遗传算法、模拟退火等方法,且各有所长,在实际使用中,往往可选择多个对接软件配合着使用,实现对接结果的互补,以获取较高的对接效果。

DOCK<sup>[8,9]</sup>是Kuntz研究小组研究开发的开源分子对接程序,是目前应用最为广泛的分子对接程序之一。DOCK先生成受体表面所有凹陷的负像,并对其进行分类分析,完成活性位点确定;接着,在活性位点上,进行配体分子和凹陷的匹配;之后,通过能量打分函数等多种打分函数来评价配体和受体之间的结合情况,选择最优的结果。DOCK考虑了配体的柔性,利用锚优先搜索或同时搜索进行配体的构象搜索,实现半柔性分子对接,并且DOCK提供了并行筛选功能。目前,DOCK已经发布到了6.4版本。

AUTODOCK<sup>[10,11]</sup>是由Scripps的Olson科研小组开发的开源的对接软件。AUTODOCK采用了模拟退火

算法和拉马克遗传算法(LGA)<sup>[11]</sup>进行构象的搜索。LGA方法结合了遗传算法和局部搜索方法<sup>[12]</sup>。遗传算法通过对基因型进行遗传算子的操作来实现全局搜索;局部搜索通过对表现进行优化实现自适应的搜索。AUTODOCK采用基于力场的半经验自由能计算方法来评价配体和受体之间结合程度。

Hex是一款由法国INRIA开发的免费分子对接软件,可以进行蛋白质-DNA、蛋白质-蛋白质、蛋白质-小分子的对接计算。Hex<sup>[13]</sup>利用球极傅里叶(SPF)相关性计算进行分子的对接,并实现了对接结果的图形显示。目前,Hex已发布到6.1版,该版本还在对接中引入了对接GPU加速技术<sup>[13]</sup>。

MVD(Molegro Virtual Docker)是一款由丹麦的Molegro ApS公司开发的预测小分子-蛋白质相互作用的软件。该软件使用了新的启发式算法,在活性位点的预测的算法中融合多种不同的进化方法,采用了包含氢键和静电作用的扩展的分段线性电位打分函数。软件的精准度达到87%<sup>[14]</sup>,但其计算时间较长。

## 4 分子对接的并行计算

快速、高效的分子对接需要以现代计算机强大的计算能力为基础,同时还需要合理的算法设计和计算策略。在单个的4核处理器(Intel, 3GHz)上,一对分子的柔性对接约需7个小时的时间<sup>[15]</sup>。然而目前通常基于分子对接的药物筛选需要将一个或多个受体与数据库中数十万的配体分子进行对接,假设配体数量为50万,用上述的处理器需约400年的时间才能完成。传统串行计算在进行的单对分子对接或进行筛选是很难满足快速、高效的要求。在现有的大规模计算条件下,采用并行计算技术实现更快速、更精确的分子对接计算已成为现实,与之相应的研究成果也不断涌现。分子对接的并行计算研究主要可归结为两个方面:一个是基于集群计算机的分子对接并行计算,另一个是基于加速部件的分子对接并行对接。

### 4.1 基于集群计算机的分子对接并行计算

集群计算机凭借其出色计算能力解决了众多科学计算问题。在分子对接的并行计算上,集群计算机仍旧有着出色的表现。基于集群计算机的分子对接研究已经深入展开,尤其是在配体筛选方面,取得了众多的研究成果<sup>[15-24]</sup>。DOCK6.0实现了主从式并行的配体筛选功能,Amanda Peters等人<sup>[18]</sup>在IBM的BLUE GENE/L上测试的结果表明,DOCK6.0的筛选可获得

线性的加速比。新版DOVIS 2.0利用AutoDock4.0在拥有256个CPU的集群计算机上实现了每个CPU每天670个配体的筛选<sup>[19]</sup>。爱丁堡大学实现了分子对接并行系统,将原本需要8天完成的对接任务缩小到62分钟<sup>[20]</sup>。

在单对分子对接方面,基于集群计算机的并行计算研究也有一定的进展。北京工业大学生命科学与生物工程学院的研究人员提出了基于MPI的单对分子对接的并行计算方法,实现LGA中局部搜索的并行,从而提高了单次LGA计算的时间<sup>[21]</sup>。大连理工大学的研究人员<sup>[22]</sup>将低能构象搜索转化为求解约束极小化问题,并采用带有空间收缩的多种群并行遗传算法进行求解,在并行机曙光3000上进行模拟计算,大大提高了分子构象搜索速度,获得较高的加速比,且保证了对接结果的准确性。

## 4.2 基于加速部件的分子对接

近年来,图形处理器(GPU)和现场可编程门阵列(FPGA)凭借其速度、并行计算能力和可编程功能为通用计算提供了良好的并行计算平台。与CPU相比,两者突显出低功耗、低成本等优势,逐渐应用于科学计算。在生物信息学领域,基于GPU和FPGA的算法层出不穷<sup>[23-30]</sup>。将FPGA加速技术和GPU计算技术应用到分子对接中,提高了计算的并行粒度,从而加快分子对接和筛选的速度。美国波士顿大学Martin C. Herbordt等人在Xilinx XC2VP70-5的FPGA芯片上,通过扩展三维相关性结构,进行多个相关性函数的计算,并结合其他加速方法,实现了分子对接的加速,与单核CPU相比获得36的加速比。同时,他们利用NVIDIA的Tesla C1060 GPU卡实现了刚性对接的加速<sup>[15,24,25]</sup>。此外,Hex软件实现GPU上基于快速傅里叶变换的分子对接<sup>[13]</sup>,Martin Prokop等<sup>[26]</sup>用GPU实现了AutoDock中AutoGrid部分加速,Silicon Informatics公司<sup>[27]</sup>利用NVIDIA的Tesla GPU实现AutoDock的加速。

## 5 小结

目前的分子对接技术主要存在二个方面的局限:第一,对接精确度的局限;第二,单对分子的并行对接计算欠成熟。

已有的对接方法简化分子柔性特点和打分评估过程,且搜索方法存在陷入局部最优的问题,对接的精确度造成了影响,应用广泛的DOCK、AutoDock等对接软件对接精确度能仅达到70-80%<sup>[14]</sup>。

为获取准确的结果,搜索过程尽可能多的覆盖所

有三维构象,并完成所有候选构象的评估。正因如此,单对分子的精确对接计算耗时长。缩短搜索和评估的时间,无疑可以加快分子对接的速度。如何利用现有的并行计算技术,对单对分子的对接计算加速已经成了分子对接计算研究的必要组成部分。

分子对接计算技术发展总是围绕着最大化对接分子相关预测评估精度,最小化计算时间这一目标展开的。为此,分子对接技术的发展将会使用精细的物理模型<sup>[31]</sup>,完善打分函数<sup>[32,33]</sup>,改进现有遗传算法等搜索方法<sup>[34]</sup>,并借助于现有的并行计算技术,从而实现提高对接计算效率的目的。

## 致谢

本文是在总结分子对接相关研究基础上形成的,在此对从事相关研究的工作人员表示诚挚的敬意;同时,感谢导师的细心指导与帮助。

## References(参考文献)

- [1] R Unger. Finding the lowest free energy conformation of a protein is an NP-hard problem:Proof and implications. *Bulltin of Mathematical Biology*,1993,55(6):1183-1198.
- [2] Stoddard B L, *et al.* Prediction of the structure of a receptor-protein complex using a binary docking method[J]. *Nature*,1992,358(27):774-776.
- [3] Oshiro c M, *et al.* Flexible ligand docking using a genetic algorithm[J]. *J Comput-Aided Mol Design*,1995,9(2):113-130.
- [4] Mangoni M, *et al.* Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation[J]. *Protein Struct Funct Genet*,1999,35(2):153-162.
- [5] Westhead R D,*et al.* A comparison of heuristic search algorithms for molecular docking[J]. *J Comput-Aided Mol Design*,1997. 11:209-228.
- [6] Wang Ting. Flexible docking by molecular dynamics simulation [J]. *Computer and Applied Chemistry*,2007,24 (1): 91-93.
- [7] Montserrat Vaqué, *et al.* Protein-ligand Docking: A Review of Recent Advances and Future Perspectives[J]. *Current Pharmaceutical Analysis*,2008,4:1-19.
- [8] Kuntz I D, Blaney J M, *et al.* A geometric approach to macromolecule-ligand interactions[J]. *J Mol Biol*,1982,161: 269-288.
- [9] Todd J. A. Ewing, Shingo Makino, *et al.* DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases[J]. *Journal of Computer-Aided Molecular Design*,2001, 15: 411-428.
- [10] Morris GM, *et al.* AutoDock4 and AutoDock Tools4: Automated docking with selective receptor flexibility[J]. *J Comput Chem*, 2009, 30(16):2785-2791.
- [11] Morris G M, *et al.* Automated docking using a lamarckian genetic algorithm and empirical binding free energy function[J]. *J Comput Chem*,1998,19:1639-1662.
- [12] FJ Solis, RJB Wets. Minimization by random search techniques[J]. *Mathematics of Operations Research*,1981,1 (6): 19-30.
- [13] D Ritchie, *et al.* Hex6.1 User Manual[EB]. [http://www.loria.fr/~ritchied/hex/manual/hex\\_manual.pdf](http://www.loria.fr/~ritchied/hex/manual/hex_manual.pdf), 2010,7,25.
- [14] Rene Thomsen, *et al.* MolDock: A New Technique for High-Accuracy Molecular Docking[J]. *J. Med. Chem*, 2006, 49:3315-3321.
- [15] Bharat Sukhwani, *et al.* Fast Binding Site Mapping using GPUs

- and CUDA[EB]. [http://www.bu.edu/caadlab/GPU\\_FT\\_Map\\_TR2010\\_1.pdf](http://www.bu.edu/caadlab/GPU_FT_Map_TR2010_1.pdf),2010,7,25.
- [16] Khodade, P, *et al.* Parallel Implementation of Autodock Subsequences[J]. *J. Appl. Cryst*,2007,40:598-599.
- [17] Zhang S, *et al.* DOVIS: an implementation for high-throughput virtual screening using AutoDock Subsequences[J]. *BMC Bioinformatics*,2008,9:126-129.
- [18] Amanda Peters, Marcus E. Lundberg, *et al.* High Throughput Computing Validation for Drug Discovery Using the DOCK Program on a Massively Parallel System[EB]. <http://www.redbooks.ibm.com/redpapers/pdfs/redp4410.pdf>, 2010,7, 2 5.
- [19] Jiang X, *et al.* DOVIS 2.0: an efficient and easy to use parallel virtual screening tool based on AutoDock 4.0[J]. *Chem Cent J*,2008,2(1):18-24.
- [20] S. R. Shave, *et al.* Ligand discovery on massively parallel systems[J]. *IBM J. RES. & DEV*,2008,52: 57-67.
- [21] Chang Shan, Kong Ren, *et al.* Molecular Docking Parallel Approaches Based on MPI[J]. *Chinese Journal of Computational Physics*,2008,25(2):241-246.  
常珊,孔韧,等.基于 MPI 的分子对接并行算法[J].*计算物理*,2008,25(2):241-246.
- [22] Li Chunlian, Wang Xicheng, *et al.*. Computer Aided Drug Molecular Docking Design Based on Parallel Evolution[J]. *Journal of Computer-Aided Design and Computer Graphics*, 2005,1(17):168-173.  
李纯莲,王希诚等. 计算机辅助药物分子对接并行演化设计[J].*计算机辅助设计与图形学学报*,2005,1(17):168-173.
- [23] Guo Quan, Wang Xicheng, *et al.* Grid Computing Technology Applied in Drug Molecule Docking[J].*Computer Engineering*, 2005,31(22):48-50.  
郭权,王希诚等. 网格计算技术在药物分子对接中的应用[J].*计算机工程*,2005,31(22):48-50.
- [24] Bharat Sukhwani, *et al.* GPU acceleration of a production molecular docking code[A]. Washington,D.C. ACM International Conference Proceeding Series; Vol. 383[C]. NY: USA: ACM,2009.19-27.
- [25] Bharat Sukhwani, *et al.* FPGA acceleration of rigid-molecule docking codes[J]. *IET Comput. Digit. Tech*,2010 ,4(3):184-195.
- [26] Olšák, *et al.* FastGrid--The Accelerated AutoGrid Potential Maps Generation for Molecular Docking[A].MEMICS 2009 [C]. Brno, Czech Republic: Masaryk University and Technical University of Brno, 2009.160-167.
- [27] Silicon Informatics. Announcing the DockStar™ Solution[OL]. <http://www.siliconinformatics.com/products.html>,2010.7,25.
- [28] Svetlin A Manavski, Giorgio Valle. CUDA compatible GPU cards as efficient hardware accelerators for Smith-Waterman sequence alignment. *BMC Research Notes*,2008,9 (Suppl 2):S10-S18.
- [29] Mark S Friedrichs,*et al.* Accelerating Molecular Dynamic Simulation on Graphics Processing Units[J]. *J Comput Chem*,2009, 30(6):864-872.
- [30] Nicholas A Sinnott-Armstrong, *et al.* Accelerating epistasis analysis in human genetics with consumer graphics hardware[J]. *BMC Research Notes*,2009,2(1):149-155.
- [31] Kang Ling, Li Honglin, *et al.*. A Refined Molecular Docking Method Considering Protein Flexibility[J]. *Journal of Basic Science and Engineering*,2008,16(6):801-810.  
康玲,李洪林等. 一种精细药物分子对接模型和优化方法[J].*应用基础与工程科学学报*,2008,16(6):801-810.
- [32] Nidhi Singh, Arieh Warshel. Absolute Binding Free Energy Calculations: On the Accuracy of Computational Scoring of Protein-ligand Interactions. *Protein Struct Funct Genet*,2010,78 (7):1705-1723.
- [33] Zsolt Bikadi, Eszter Hazai. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock[J]. *Journal of Cheminformatics*, 2009,10: 1-15.
- [34] Liu Min, Zeng Tao, *et al.* A Molecular Docking's Conformational Search Strategy Based on Immune Genetic Algorithm[J]. *Journal of Computer Research and Development*, 2009, 46(Suppl.):222-226.  
刘敏,曾涛,徐开阔等. 一种基于免疫遗传算法的分子对接构象搜索策略[J]. *计算机研究与发展*,2009,46(Suppl.):222-226.