

分子对接与全局极小化方法*

李炜疆**

(江南大学工业生物技术教育部重点实验室, 无锡 214036)

摘要 全局极小化方法及其在结构生物学中的应用近年来取得了显著的进展. 适当简化的分子对接问题是全局极小化方法的一个很好目标, 并且是当前一个相当活跃的研究领域. 对接可分为两类: 主要用于从头配体设计的细致对接和用于已知化合物数据库筛选以发现药物的粗略对接, 它们对全局极小化算法的要求是不同的. 简要评述了新出现的适合于对接问题的随机和确定性全局极小化算法, 其中势能平滑算法看来很有希望, 值得密切关注.

关键词 分子对接, 全局极小化, 势能平滑

学科分类号 Q615

分子的相互作用与识别是生物化学非常广泛研究的课题. 由于在发现药物方面的潜力, 理论预测小分子与生物大分子靶之间相互作用的方法有着永恒的兴趣. 药物的作用是干预与疾病相关的生物化学过程, 这种干预通过低分子质量的药物分子与特定的大分子靶的相互作用实现. 利用给定靶的结构数据, 发现有这样相互作用能力的化合物并对它们进行优化是基于结构的药物设计的基础.

可利用的受体结构数据的快速增加, 刺激了直接药物设计方法的加速发展. 自动分子对接 (automated molecular docking) 方法是基于结构的药物设计的重要环节^[1,2], 其目标之一是在数值计算基础上预测小分子如何与大分子靶相互作用, 即确定给出的小分子与蛋白质是否和如何相互作用. 由于涉及许多的物理化学因素以及计算技术本身的限制, 不容易取得细致准确的结果. 有实际应用价值的对接方法必须解决两个主要问题: 合理简化的物理化学模型和高效可靠的优化算法.

对接可以由下列方式实现, 通过配体与其受体的几何匹配或通过极小化相互作用能量. 几何匹配途径^[3,4]的计算量相对较小, 而能量极小化途径有更好的理论基础, 但必须开发合适算法克服问题本身的多极小特性产生的计算困难.

全局极小化方法及其在分子对接的应用在最近的几年中取得了长足的进步, 快速可靠的算法不断涌现, 算法性能提高很快. 本文将对该领域一些突出的进展进行简要的方法论的概念性介绍, 不涉及具体的计算机程序. 关于对接计算的一些细节, 可

参阅文献 [4].

1 分子对接及其对全局极小化方法的要求

分子对接问题就是要寻找相互作用分子间的最稳定结合状态, 严格地说, 这需要寻找相互作用系统的自由能全局极小, 但计算自由能需要相互作用能量函数的整体性质, 其计算相当困难^[5], 目前尚缺乏有效快速的计算方法, 所以将自由能简化为相互作用能量而忽略熵的作用. 实际计算表明, 在适当考虑溶剂作用以及小分子柔性的情况下, 这种近似是可以接受的^[6].

分子对接系统的全局最优化目标函数称为计分函数 (scoring function). 通常, 计分函数是系统自由能的某种近似, 包含三个方面的相互作用, 即分子内、分子间以及溶剂相互作用, 其具体函数形式可仿照一般的分子力场函数形式赋值^[5].

对接问题的最简单情形是刚性对接, 即参与对接的分子结构保持不变. 此时, 分子内相互作用部分为常数, 系统只有 6 个独立自变量. 刚性分子各自处于能量极小构象, 这并不一定有利于降低整个系统的总能量. 事实上在结合态的受体和配体构象可以处于相对较高能量状态^[7], 所以高质量的对接必须考虑分子构象的可变性, 即柔性. 在柔性对接中, 系统除刚体运动自变量以外, 还有若干描述

* 国家自然科学基金资助项目 (39760027).

** 通讯联系人.

Tel: 0510-5800968, E-mail: wjlee@pub.wx.jsinfo.net

收稿日期: 2000-06-16, 接受日期: 2000-08-23

两个分子各自内部自由度的变量。柔性对接不仅意味着要考虑配体分子的构象可变性,也要考虑受体分子在对接过程中产生诱导形变。一般地,柔性自由度越大,对接的结果就越客观准确,但不可避免地增加了计算复杂度^[8]。

基于结构的药物设计有两个基本策略:从头设计 (*de novo design*) 和化合物数据库筛选 (*database screening*)。与此相对应,对接算法也可分为两类,可分别称为细致对接 (*detailed docking*) 算法和粗略对接 (*rough docking*) 算法^[2]。

粗略对接用于对候选配体的化合物数据库进行对接,从中筛选出可能与受体靶结合的化合物,通常需要对 10^5 个小分子的数据库进行对接过滤筛选。要在合理时限内完成这样规模的计算,必须以很快的速度完成对单个小分子化合物的对接。当然,这需要以牺牲一定的精度,如考虑较低程度的分子柔性为代价。假定对接整个数据库的可接受合理时限为 10^6 s (约两周),则对典型单个小分子的对接时间就只有 10 s。

细致对接用于改造或设计配体分子,对精度要求较高,如考虑较多的分子柔性,溶剂作用等,有时还需要对亲和性作出估计。细致对接对速度的要求不很严格,通常一次对接计算应在 10^4 s (几个小时) 内完成。

对接算法的 CPU 时间消耗主要用于反复计算相互作用能量函数上,其计算复杂度可用目标函数的调用次数来计量。按典型数量估算,每次能量调用在目前最快的单 CPU 工作站上约需 0.1 s。据此,细致对接算法的能量函数调用上限约为 10^5 ,而筛选算法的调用上限仅为 10^2 。根据已有的报道,目前的对接算法的能量函数调用量在 10^4 次以上。因此,筛选算法必须采用简化的近似能量计算(如插值法)大幅度减少计算量。当靶蛋白采取刚性构象近似时,确实可以采用网格插值法将势能计算速度提高两个数量级^[9]。当然,预先在网格上的计算还是要耗费大量 CPU 时间,也正因为如此,这种技术只有在针对固定靶的数据库对接过滤搜索时才有实质性的效果。

2 适合于对接问题的全局极小化方法

由于力场势能函数与蛋白质折叠问题相似,所以,适用于蛋白质折叠问题的全局极小化技术一般也同样适用于对接问题。蛋白质折叠和分子对接一

般都要化为显式的无约束极小化问题,大体有随机算法和确定性(解析)算法两类方法。

随机算法具有适应范围广、实现简单、通常不需要计算导数等优点,是当前蛋白质折叠理论研究中普遍应用的方法^[10],同时也是对接方法使用最多的搜索引擎。典型方法有:模拟退火算法 (*simulated annealing*, SA),遗传算法 (*genetic algorithm*, GA),以及这些算法的各种修正变种。最近也出现了一些其他随机算法,如随机聚点搜索算法^[11],平均计算量有所降低。还有一些其他的受控随机搜索 (*controlled random search*, CRS) 算法^[12]尚未应用于对接问题,但由于较好的全局极小搜索性能,具有相当潜力。

已有不同随机算法在对接问题中的性能评价报道,结论并不完全一致^[9, 13],不同方法各有优缺点。新近出现一种随机近似平滑 (*stochastic approximation with smoothing*, SAS) 方法,采用随机近似势能平滑技术,在对接测试对比计算中有很好的表现^[9]。随机算法的主要缺点是需要充分探索能量地形后才能找到全局极小,因而计算量大,找到一次全局极小平均需要计算计分函数在 10^5 次以上。但同时,以此为代价,这类方法也有可能根据计算过程中系统在能量地形上的运动情况给出结合亲和性的某种估计。

解析算法属于确定性算法,需要计算计分函数的导数甚至更复杂的解析计算,其突出优点是速度快。近来引起特别注意的势能平滑搜索 (*potential smoothing search*, PSS) 算法^[14, 15]是有可能在 10^4 次以内能量调用找到全局极小的最有希望的快速算法,在对接问题中已有可喜进展^[16]。

典型的 PSS 算法首先将势能函数 $f(x)$ 用某种方式变换为 $F(x; t)$,其中 t 为决定平滑程度的控制参数。当 $t = 0$ 时, $F(x; 0)$ 就是原始的势能函数 $f(x)$ 。随着 t 的增大,较浅的次要极小逐渐消失,势能超曲面变得越来越平滑, $F(x; t)$ 的极小数目越来越少,只有那些较深且较宽的极小保留下来。当 t 足够大时,势能超曲面上将只剩一个极小。我们希望这个极小是原始势能函数全局极小在平滑过程中保留下来的遗迹。然后这个最终保留下来的极小出发,逐步恢复势能曲面的本来面目,追踪极小在此过程中位置的变化情况,直至原始未变形的势能曲面。

PSS 算法的关键是平滑方案,由于实际问题都是多维的,所以不能采用数值平滑方法。一个可行

的平滑方案必须能够以较少的额外代价计算 $F(x; t)$, 这就要求 $F(x; t)$ 必须有比较简单便于计算的解析表达式. 平滑方案可以有多种^[17], 但物理意义清晰且计算效果好的是平滑积分变换^[18, 19], 以这类平滑方案为基础的 PSS 算法已应用于对接问题^[14, 16, 20], 并取得了很好的计算结果. 目前主要问题有: 平滑方案的改进, 平滑后的势能函数解析计算困难, 极小追踪技术的改进等.

3 展 望

使用全局最优化方法进行分子对接已经有了实质性的进展, 算法性能与药物设计的要求之间的距离已显著缩小. 寻找便于计算的计分函数更好地近似系统自由能, 以及进一步改进优化算法, 提高搜索效率, 使得算法能够考虑更多的柔性, 仍是该领域的重要课题. 就优化算法而言, PSS 方法是极有希望的快速柔性对接算法.

参 考 文 献

- Blaney J M, Dixon J S. A good ligand is hard to find: automated docking methods. *Perspectives in Drug Discovery and Design*, 1993, **1** (3): 301~ 319
- Lengauer T, Raey M. Computational methods for biomolecular docking. *Curr Opin Struct Biol*, 1996, **6** (2): 402~ 406
- Sobolev V, Wade R C, Vriend G, *et al.* Molecular docking using surface complementarity. *Proteins*, 1996, **25** (1): 120~ 129
- Shoichet B K. Docking ligands to proteins. In: Sternberg M J E, ed. *Protein Structure Prediction: A Practical Approach*. Oxford: Oxford University Press, 1996. 263~ 290
- Tame J R H. Scoring functions: a view from the bench. *J Comput Aided Mol Des*, 1999, **13** (1): 99~ 108
- Morris G M, Goodsell D S, Halliday R S, *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*, 1998, **19** (14): 1639~ 1662
- Nicklaus M C, Wang S, Driscoll J S, *et al.* Conformational changes of small molecules binding to proteins. *Bioorganic & Medicinal Chemistry*, 1995, **3** (2), 411~ 428
- Wooley J C. Trends in computational biology: a summary based on a RECOMB plenary lecture, 1999. *J Comput Biol*, 1999, **6** (3/4): 459~ 474
- Diller D J, Verlinde C L M J. A critical evaluation of several global optimization algorithms for the purpose of molecular docking. *J Comput Chem*, 1999, **20** (16): 1740~ 1751
- Hansmann U H E, Okamoto Y. New monte carlo algorithms for protein folding. *Curr Opin Struct Biol*, 1999, **9** (1): 177~ 183
- 李炜疆, 张有广. 改进的随机聚点搜索算法及其在分子对接问题中的应用. *内蒙古大学学报*, 2000, **31** (3): 257~ 260
Li W J, Zhang Y G. *Acta Scientiarum Naturalium Universitatis Intramogolicae*, 2000, **31** (3): 257~ 260
- Ali M M, Torn A, Viitanen S. A numerical comparison of some modified controlled random search algorithms. *Journal of Global Optimization*, 1997, **11** (4): 377~ 385
- Vieth M, Hirst J D, Dominy B N, *et al.* Assessing search strategies for flexible docking. *J Comput Chem*, 1998, **19** (14): 1623~ 1631
- Pappu R V, Hart R K, Ponder J W. Analysis and application of potential energy smoothing for global optimization. *J Phys Chem B*, 1998, **102** (48): 9725~ 9742
- Hart R K, Pappu R V, Ponder J W. Exploring the similarities between potential smoothing and simulated annealing. *J Comput Chem*, 2000, **21** (7): 531~ 552
- Pappu R V, Marshall G R, Ponder J W. A potential smoothing algorithm accurately predicts transmembrane helix packing. *Nature Struct Biol*, 1999, **6** (1): 50~ 55
- Hart R K, Pappu R V, Ponder J W. Generalized potential function smoothing algorithms for conformational energy optimization. *Biophys J*, 1998, **74** (2): A176
- 李炜疆, 罗辽复. 全局最优化的平滑积分变换方法. I. 基本思想和理论框架. *内蒙古大学学报*, 1997, **28** (2): 157~ 161
Li W J, Luo L F. *Acta Scientiarum Naturalium Universitatis Intramogolicae*, 1997, **28** (2): 157~ 161
- Li W J, Luo L F. Finding global minimum by removal of minor minima: the smoothing integral transformation method for global optimization. In: Luo L F, eds. *Proceedings of the International Symposium on Theoretical Biophysics and Biomathematics*. Hohhot: Inner Mongolia University Press, 1997. 113~ 119
- 李炜疆, 李春华. 分子对接模型中的平滑积分变换方法. *内蒙古大学学报*, 2000, **31** (1): 12~ 16
Li W J, Li C H. *Acta Scientiarum Naturalium Universitatis Intramogolicae*, 2000, **31** (1): 12~ 16

Molecular Docking and Global Minimization Methods*

LI Wei Jiang**

(Key Laboratory of Industrial Biotechnology of Ministry of Education, Sothern Yangtze University, Wuxi 214036, China)

Abstract Global minimization methods and their applications to structural molecular biology have made significant progresses in recent years. Appropriately simplified molecular docking problem is a good object of global minimization, which is now a rather active research area. Molecular docking can be divided into two categories. The detailed docking for *de novo* ligand design and the rough docking for known chemical database screening for drug discovery. Their demands for the global minimization algorithms are different. New stochastic and deterministic global minimization algorithms that are suitable for docking problems are briefly reviewed. Those algorithms with potential smoothing techniques seem to be promising and are worth close noting.

Key words molecular docking, global minimization, potential smoothing

* This work was supported by a grant from National Natural Science Foundation of China (39760027).

** Corresponding author. Tel: 86-510-5800968, E-mail: wjlee@pub.wx.jsinfo.net

Received: June 16, 2000 Accepted: August 23, 2000