

· 专论与综述 ·

蛋白质分子表面高质量网格构建技术的研究进展*

闫玉娇 荆 艳 宋晓峰[△]

(南京航空航天大学生物医学工程系 江苏 南京 210016)

摘要 :分子表面即分子边界,在一定程度上蕴含了分子的生物化学属性信息,对分子表面进行分析将有助于理解分子对接、识别和相互作用等问题。由于蛋白质分子表面的构造相对复杂,尤其是分子表面的网格化,因此寻求高效的算法构建高质量的蛋白质分子表面网格对生成光滑的分子表面、分子可视化及分子模拟都有着重要的意义。本文主要根据现有定义的蛋白质分子表面,针对近年来几种高质量分子表面网格构建的新技术进行了阐述,同时介绍了几款蛋白质分子表面可视化软件,并对它们的性能进行了简单的分析。

关键词 :蛋白质分子表面 网格构建 分子可视化软件

中图分类号 :Q71/TP181 **文献标识码** :A **文章编号** :1673-6273(2012)21-4140-04

Progress in High-Quality Protein Molecular Surface Mesh Generation*

YAN Yu-jiao, JING Yan, SONG Xiao-feng[△]

(Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China)

ABSTRACT: Molecular surface is also called molecular boundary, it contains the biological and chemical information of the protein molecule. The analysis of the molecular surface will help to understand the molecular docking, recognition and molecular interaction. While the construction of the protein molecular surface is so complicated, especially the molecular surface meshing, therefore, the efficient algorithm for generating the high-quality protein molecular surface meshes has the great significance to smooth molecular surface generation, molecular visualization and molecular simulation. According to the current definitions of the protein molecular surface, several new methods of the high-quality molecular surface mesh generation have been reviewed. Also, this paper introduces several software of the protein molecular surface visualization and makes a simple analysis about their performance.

Key words: Protein molecular surface; Meshing; Molecular visualization

Chinese Library Classification (CLC): Q71/TP181 **Document code**: A

Article ID: 1673-6273(2012)21-4140-04

引言

蛋白质分子表面对研究蛋白质的功能和蛋白质分子间的相互作用都有着非常重要的作用。关于构建光滑的蛋白质分子表面,人们提出了不同的模型和算法来表示、计算和分析分子形状^[1-4]。随着量子化学理论在分子表面建模中的广泛应用^[5-9],为了构建分子表面模型,每一个构成分子的原子都被理想化为一个球体。在有效的判断组成蛋白质边界原子的基础上,利用这些抽象为球体后的表面原子集合和它们之间的拓扑关系,计算出蛋白质分子表面上一个完整的光滑曲面。

由于分子表面的构造非常复杂,尤其是分子的表面网格化一直是一个难题,网格质量直接影响计算结果的精度和计算时间,因此生成高质量、可控性好且具有良好的曲面适应能力的网格对构建光滑分子表面及其可视化有着重要的意义。分子表面网格生成主要是基于 Delaunay 方法^[10,11]和波前法(Advancing Front Method, AFM)^[12,13],两种方法时间复杂度相当,但 AFM

生成的网格质量高而且可控性好,表现出良好的曲面适应能力,因此是曲面网格生成研究的热点^[12,13]。本文主要根据现有定义的蛋白质分子表面,对几种构建高质量分子表面网格技术进行了讨论,并介绍了几款蛋白质分子表面可视化软件。

1 概述

蛋白质分子表面对计算蛋白质分子之间的相互作用和分子可视化有着重要的作用,因此定义分子表面将有助于理解和预测蛋白质的功能。

1.1 蛋白质分子表面的定义

蛋白质分子表面的概念由 Lee 和 Richards^[14,15]首先提出。它包括三种不同类型表面:范德华表面(Vander Waals Surface, VDW),它是由一组表示原子的范德华球堆积形成的表面(图 1a),其中粗黑色轮廓表示定义的分子表面;Lee 和 Richards^[14]于 1971 年首次定义了基于探测球的溶剂可及表面(Solvent Accessible Surface, SAS)^[14],它是通过给定一个蛋白质分子和

* 基金项目:国家自然科学基金资助项目(61171191);江苏省自然科学基金资助项目(BK2010500)

作者简介:闫玉娇(1987-),女,硕士研究生,研究方向:生物信息学。E-mail: yanyuquan2009@163.com

[△]通讯作者:宋晓峰 E-mail: xfsong@nuaa.edu.cn

(收稿日期:2011-12-06 接受日期:2011-12-31)

探测球,探测球在蛋白质分子表面滚动所形成的探测球球心的轨迹集合(图 1b);1977 年 Richard 进一步提出了溶剂外部曲面(Solvent Excluded Surface, SES),它是探测球在蛋白质分子表面滚动时空间里所有可能出现位置合集的补集边界(图 1c),

通常也称 Connolly 表面^[16]。此外,也出现了高斯表面(Gaussian Surface)的定义^[17],目前它已经被广泛应用于计算生物学中,如分子对接问题^[18]、SAS 面积计算^[19]和分子形状比较^[20]等。

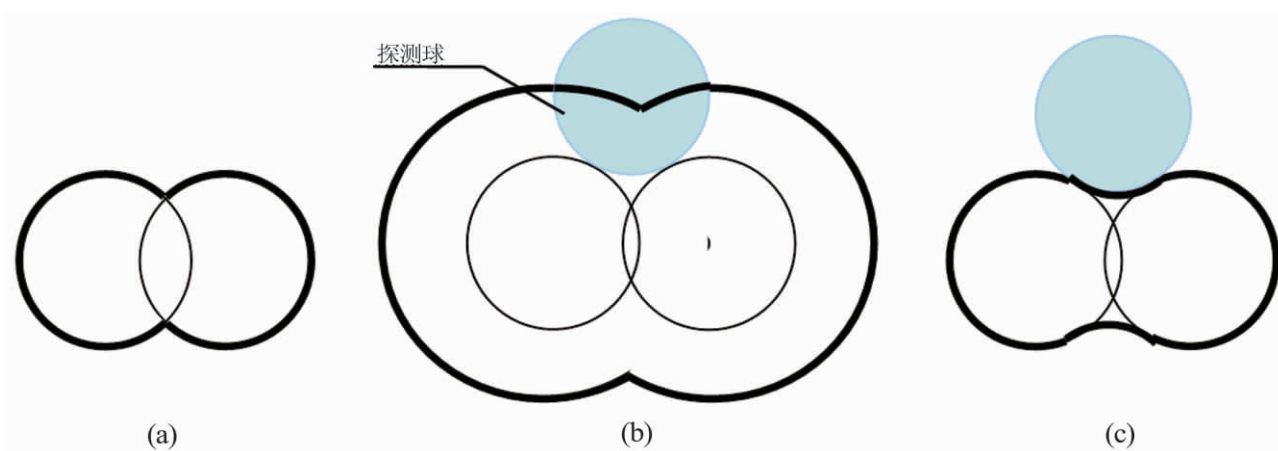


图 1

Fig. 1

(a) 范德华表面
(a) Vander Waals Surface

(b) 溶剂可及表面
(b) Solvent Accessible Surface

(c) 溶剂外部曲面
(c) Solvent Excluded Surface

1.2 SES 表面

SES 表面主要包括接触表面和凹陷表面^[21]。接触表面是指探测球与蛋白质分子 VDW 表面直接接触的凸球面,凹陷表面是探测球不能进入蛋白质分子中过于狭窄的凹陷时形成的分子表面。其中根据探测球与表面原子接触的个数可将凹陷表面分为凹球面和马鞍面。凹球面是探测球与三个原子相接触形成的球面,而马鞍面是探测球与两个原子相接触形成的环面。

实际上,SES 可以通过 SAS 减掉探测球所占的体积获得。对 SES 表面的凸球面、凹球面和马鞍面进行三角化,可以获取完整的 SES 表面。

1.3 蛋白质分子表面原子的定义

位于蛋白质表面边界的原子决定了蛋白质分子的主要功能^[16,22]。根据 SES 表面的定义,蛋白质分子的表面原子可以定义为探测球在蛋白质分子表面滚动时能够接触到所有原子的集合。

2 蛋白质分子表面的比较

VDW 表面形式简单,计算方便,但由于探测球半径的限制,使其不能接触到死体积内的分子表面,从而引起“死体积”出现。此外,VDW 表面在蛋白质分子静电作用的研究中会引起静电能量计算的过拟合。近期的一些研究主要致力于寻找高效的方法构建基于 SAS 的三角网格^[23-27],但 SAS 表面会导致静电能量计算的欠拟合,而 SES 对于稳定能量的计算非常合适,这对蛋白质分子静电作用的研究有着重要的意义。但 SES 有一个重要的缺点,即由滚动的探测球的自相交产生的尖点进而导致在生成半径的计算以及相应的微分计算时出现奇异点。因此在构建基于 SES 的分子表面模型时,要避免奇异点的出现。此外,SES 表面的构造相对繁琐,尤其是表面的三角网格化。对于高斯表面,可以通过选择不同的参数去较好的逼近 VDW、SAS

和 SES 表面^[28,29],因此它的应用更加灵活方便。

3 蛋白质分子表面网格构建技术

分子表面网格生成技术广泛应用于分子可视化和分子模拟,其中网格质量对求解精度有着重要的影响。定义的蛋白质分子表面不同,构建蛋白质表面分子网格所应用的技术也不相同。

3.1 基于任意表面的网格构建技术

该技术可以针对任意分子表面生成高质量网格,具体算法如下:(1)首先利用边界表示法获取分子表面信息,即一系列面和原子的拓扑关系(对于 SES,主要是指分子边界和凸球面、凹球面及马鞍面之间的拓扑关系);(2)针对每一个面进行合适的参数化;(3)结合 AFM 算法,通过使用广泛流行的 Delaunay 三角网格生成算法,在参数化表面基础上实现二维空间三角网格的构建;(4)最终将网格投影到整个蛋白质分子表面上。

该技术只是应用于单个分子(蛋白质、RNA、蛋白核糖核酸)结构,不适用于较大的复合蛋白质分子。

3.2 基于 SAS 的有限球集合的网格构建技术

该技术利用有限球集合的增量网格生成算法进而生成高质量的分子表面网格。具体算法为:(1)将位于分子表面的每个原子抽象成一个圆心为 p ,半径为 γp 的球体,其中 $0 < \gamma < 1$;(2)给定一个边 ab ,其中 a, b 分别是半径为 $\gamma p(a)$ 和 $\gamma p(b)$ 的边界球球心,加入新的采样点,利用增量网格算法^[30]构建 Delaunay 三角网格;(3)提取候选的分子表面三角形,其中候选三角形满足两个条件,即小的外接圆半径($R_{abc} < \frac{\varepsilon}{1-\varepsilon} p_{abc}$)和有限元 Delaunay 三角形;(4)通过检查候选三角形的每条边,对之前的三角网格进行更新直到有限球集合覆盖了整个分子表面,这样就可以获得 ε 采样和候选三角形的集合包括有限元 Delaunay 三角剖分。

由于需要计算分子表面所有采样点的 Delaunay 三角剖分, 所以导致计算效率相对较低。该技术可以通过选择较小的 γ 和 ε 来获得更好的网格质量, 但是这将会导致分子表面三角网格数量的增多。因此要在网格质量和数量之间做一个很好的折中。

3.3 基于 SES 的收缩网格构建技术

该算法从分布在一个包含整个分子表面的椭球上的三角网格开始, 逐步收缩网络直到所有的三角形顶点到达分子表面, 从而获得三角分割的分子表面。具体实现步骤为: (1) 分析蛋白质分子表面, 即将分子表面分为凸球面、凹球面和马鞍面; (2) 构造一个包含整个分子表面的起始椭球, 在椭球表面生成一个均匀分布的三角包络网格; (3) 利用收缩技术逐步向内收缩包络表面, 使得所有网格到达分子表面; (4) 进行光照渲染, 获得光滑的分子表面。

引进局部包络球可以大大减少计算时间。对于凸球面, 局部包络球即为对应原子的范德华球; 对于凹球面三角, 局部包络球球心即为探测球的球心在形成凹面的三个原子中心所定义平面上的正交投影, 半径为从局部包络球球心到球面三角各顶点的最大距离; 对于马鞍面, 局部包络球球心为形成马鞍面的两原子球相切的探测球球心所经历的圆弧中心, 半径为球心到两原子的最大距离。

收缩技术改为扩张技术, 起始包络椭球改为空腔内含椭球, 就可方便的将该算法应用于蛋白质空腔的表示。

3.4 基于 SES 的分段样条模型的网格构建技术

该技术主要是利用了一种新的方法生成基于 SES 的光滑分段样条模型去逼近分子表面。具体步骤: (1) 通过提取 PDB 文件中原子坐标信息, 利用 Delaunay 三角网格构建方法获得较为粗糙的分子表面网格; (2) 在此基础上构建一个三角棱镜框架 P_s , 然后利用分段多项式 BB (Bernstein-Bezier) 样条函数在 P_s 中逼近曲面, 同时计算近似误差和采样点收敛的约束边界; (3) 对生成的曲面进行参数化, 从而利于更加高效的进行曲面的采样和简化静电能量计算时积分准确性的评估。

由于密集的三角面片会引起分子表面的过采样从而降低效率, 而此技术能够利用很少的三角形面片来获得简单的结果, 此外它也可以避免奇异点的出现, 进而构建一个准确、光滑和方便计算的分子表面模型去更好的逼近分子表面。

3.5 基于高斯表面的跟踪表面的网格构建技术

该技术用跟踪分子表面的方法来对蛋白质分子的高斯表面进行网格构建。该算法主要包括两步: (1) 通过一个非线性核

函数 $\varphi(x) = t_0^{1/2} \sum_{i=1}^N e^{-\frac{d(\|x-c_i\|)^2}{r_i^2-1}}$ 计算出分子表面的点, 其中 $\varphi(x) = \sum_{i=1}^N e^{-\frac{d(\|x-c_i\|)^2}{r_i^2-1}}$;

(2) 通过连接这些生成的点对高斯表面进行多边形化; (3) 在对表面多边形化基础上, 应用标准的多边形三角剖分方法^[31]实现分子表面三角网格划分。

该技术对任意大小的蛋白质分子都能进行高质量的表面网格构建, 它克服了有限元和边界元方法运用于生物大分子模拟的一个瓶颈问题, 并进一步为整个分子立体网格的构建提供了一个可行的方案。

4 蛋白质分子表面可视化软件

蛋白质分子表面可视化软件已经发展了数年, 然而现在大部分的软件只是应用于小分子, 很少有软件对较大的复合蛋白、几何融合和容积可视化表示进行操作的或者高效描述分子灵活性和动态性^[32]。目前应用较为广泛的分子表面可视化软件主要有 ASMS、MSMS、LSMS 和 TMSmesh。

4.1 ASMS

ASMS 是在 TexMol^[32] 分子可视化软件基础上通过构建光滑分段样条模型逼近分子表面发明的软件。该软件通过读入 PDB 文件对蛋白质的 SES 表面进行可视化。由于 ASMS 利用较少的三角网格进行分子表面计算, 因此计算效率高。

4.2 MSMS 和 LSMS

MSMS 在分子表面建模领域作为一个典型和高效的软件被广泛应用。它通过读入 PQR 文件 (包括 PDB 文件、原子净电荷和原子半径信息) 进行 SES 表面可视化。LSMS 是通过利用 Level Set 方法, 基于立方体网格点逼近分子表面的。由于 LSMS 和 MSMS 不需要进行多边形化, 因此 CPU 运行时间较短。但是分子表面奇异点的存在会导致不正确的结果产生, 所以生成的分子表面没有 TMSmesh 生成的表面光滑。

4.3 TMSmesh

TMSmesh 是为生成任意大小分子体系的高质量表面网格而设计的, 它已在超过一百万个原子的病毒分子上测试成功。而且 TMSmesh 产生的分子表面网格也经测试可直接用于边界元方法的计算模拟中, 同时它生成的分子表面与其他软件相比更加光滑, 但是它在网格构建过程中不能区分空腔的内表面和外表面。

表 4 是四种分子表面可视化软件在算法复杂度、内存开销和适用范围上的比较, 其中 N 表示原子个数。

表 1 四种分子表面可视化软件性能比较

Table 1 The performance comparison of the four molecular surface visualization software

Software	Algorithm complexity	Memory use	Applicability
ASMS	$O[N]$	Very large	Arbitrary size protein molecules
LSMS	$O[N \log(N)]$	Large	Only a single and small size protein molecule
MSMS	$O[N \log(N)]$	Small	Only a single and small size protein molecule
TMSmesh	$O[N]$	Very small	Arbitrary size protein molecules

5 总结

蛋白质分子功能和结构的研究一直是生命科学研究的热点, 其中蛋白质分子表面原子结构对功能影响极大, 对蛋白质分子表面的分析将有助于我们寻找能与底物结合的口袋, 预测可能的结合位点以及进行药物设计。本文根据定义的蛋白质分

点, 其中蛋白质分子表面原子结构对功能影响极大, 对蛋白质分子表面的分析将有助于我们寻找能与底物结合的口袋, 预测可能的结合位点以及进行药物设计。本文根据定义的蛋白质分

子表面,重点介绍了几种构建分子表面网格的新技术,这些技术针对不同的分子表面进行高质量的网格构建,从而有助于分子表面模型的可视化。一方面,我们需要研究更加高效、稳定的网格构建算法从而确保分子表面模型的收敛性和准确性;另一方面,分子表面的光滑程度需要进一步提高。对于目前的蛋白质表面分子可视化软件,大部分适用于小分子蛋白质,很少应用于大分子蛋白质,而ASMS和TMSmesh这些应用于大分子蛋白质的软件执行效率低、内存开销大,因此开发支持多平台运行的复合型大分子蛋白质表面可视化软件将成为近几年的发展趋势。

6 展望

生成高质量、可控性好且具有良好的曲面适应能力的网格构建技术将是未来研究的热点。目前大多数蛋白质分子表面高质量网格构建技术都是基于三角网格模型的,由于三角网格模型存在存储量大、光滑性差和不易修改等缺点,因此在对蛋白质分子表面进行网格构建时,可以考虑将三角网格面进一步表示为分片光滑的曲面或者使用立体网格模型。随着网格构建技术的不断发展,一些稳定、高效的网格构建算法不仅提高了网格构建的质量和效率,更重要的是有助于大分子蛋白质可视化软件的进一步完善以及实现对更为复杂的蛋白质分子的高性能计算的结构优化。在获取高质量网格的基础上,我们可以对蛋白质空间结构形态进行分析计算,利用这些计算结果进而研究蛋白质分子表面三维结构对其功能的影响。

参考文献(References)

- [1] Perrot G, Maigret B. New determinations and simplified representations of macromolecular surfaces [J]. J Mol Graph, 1990, 8: 141-144
- [2] Sanner M F, Olson A J, Spehner J C. Reduced surface: an efficient way to compute molecular surfaces [J]. Biopolymers, 1996, 38 (3): 305-320
- [3] Cossi M, Menucci B. Analytical first derivatives of molecular surfaces with respect to nuclear coordinates [J]. J Comput Chem, 1996, 17 (1): 51-73
- [4] You T, Bashford D. An analytical algorithms for the rapid determination of the solvent accessibility of points in a three dimensional lattice around a solute molecule[J]. J Comput. Chem, 1995, 16 (6): 743-757
- [5] Akkiraju, N., Edelsbrunner, H. Triangulating the surface of a molecule [J]. Discrete Applied Mathematics, 1996, 71: 5-22
- [6] Connolly, M. L. (1996) Molecular Surfaces: A Review. <http://www.netsci.org/Science/Compchem/feature14.html>
- [7] MathMol Library. <http://www.nyu.edu/pages/mathmol>.
- [8] RasMol Molecular Visualisation Program. <http://klaatu.oit.umass.edu:80/microbio/rasmol>
- [9] Sanner, M. F. (1996) Molecular Surface Computation. <http://www.scripps.edu/pub/olsonweb/people/sanner/html/msmsFhome.html>
- [10] 赵岩, 张子平. 一种动态构建 Delaunay 三角网的算法 [J]. 测绘工程, 2008, 17(3): 24-27
Zhao Yan, Zhang Zi-ping. One dynamic algorithm of constructing Delaunay triangulation [J]. Engineering of Surveying and Mapping, 2008, 17(3): 24-27
- [11] 刘学军, 龚健雅. 约束数据域的 Delaunay 三角剖分与修改算法[J]. 测绘学报, 2001, 30(1): 82-87
Liu Xu-jun, Gong Jian-ya. Delaunay Triangulation of Constrained Data Set[J]. Acta Geodetica et Aertographica Sinica, 2001, 30(1): 82-87
- [12] Lee C K. Automatic adaptive mesh generation using metric advancing front approach [J]. Engineering and Computations, 1999, 16 (2): 230-263
- [13] Cuilliere J C. An adaptive method for the automatic triangulation of 3D parametric surfaces [J]. Computer-Aided Design, 1998, 30 (2): 139-149
- [14] Lee B, Richards F M. The interpretation of protein structures: estimations of static accessibility[J]. J Mol Biol, 1971, 55: 379-400
- [15] Richards F M. Areas, Volumes, Packing and protein structure [J]. Annu Rev Biophys Bioeng, 1977, 6: 151-176
- [16] ML Connolly. Solvent-accessible surfaces of proteins and nucleic acid [J]. Science, 1983, 221:709-713
- [17] Minxin Chen, Benzhuo Lu. TMSmesh: A Robust Method for Molecular Surface Mesh Generation Using a Trace Technique [J]. Theory Comput, 2011, 7: 203-212
- [18] McGann M R, Almond H R, Nicholls A, Grant A J, Brown F K. Gaussian docking functions [J].Biopolymers, 2003, 68: 76-90
- [19] J.Weiser, Shenkin P, Still W J Comput. Chem, 1999, 20: 688-703
- [20] Grant J. A., Gallardo M. A., Pickup B. T. J. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape[J]. Journal of Computational Chemistry, 1996, 17 (14): 1653-1666.
- [21] W Zhao, G Xu, C Bajaj. An algebraic spline model of molecular surfaces[J]. IEEE, 2007: 297-302
- [22] Xiong Wang. Mining Protein Surfaces. ACM SIGMOD Workshop on Research Issues in Data Mining and knowledge Discovery[C].Santa Barbara California, 2001: 20-24
- [23] Nataraj Akkiraju and Herbert Edelsbrunner. Triangulating the surface of a molecule[J]. Discrete Applied Mathematics, 1996, 71(1-3): 5-22
- [24] Chandrajit L. Bajaj, Valerio Pascucci, Ariel Shamir, Robert J. Holt, and Arun N. Netravali. Dynamic maintenance and visualization of molecular surfaces [J]. Discrete Applied Mathematics, 2003, 127(1): 23-51
- [25] Michel F. Sanner, Arthur J. Olson, and Jean-Claude Spehner. Fast and robust computation of molecular surfaces. In Proceedings of the eleventh annual symposium on Computational geometry [C].ACM Press pages, 1995: 406-407
- [26] Michel F, Sanner, Arthur J. Olson and Jean-Claude Spehner. Reduced surface: an efficient way to compute molecular surfaces [J]. Biopolymers, 1996,38(3): 305-320
- [27] Amitabh Varshney, Frederick P. Brooks. Fast analytical computation of Richards's smooth molecular surface [C]. In Proceedings of the 4th conference on Visualization, 1993, 93:300-307
- [28] Duncan B S, Olson A J. [J]. Biopolymers, 1993, 33:231-238
- [29] Blinn J F. A generalization of algebraic surface drawing [J]. ACM Transactions on Graphics, 1982, 1: 235-256
- [30] Charles L. LAWSON. Software for c1 surface interpolation [M]. Mathematical Software III, 1977: 161-194
- [31] De Berg, M, Van Kreveld M,Overmars M, Schwarzkopf O. Computational Geometry: Algorithms and Applications, 2nd ed [M]. Springer-Verlag: New York, 2000:45-61
- [32] C L Bajaj, P Djeu, V Siddavanahalli, A Thane. TexMol: Interactive Visual Exploration of Large Flexible Multi-component Molecular Complexes [J]. 15th IEEE Visualization 2004 Conference (VIS 2004), 2004: 243-250