

# 基于分子力学的蛋白质分子场可视化

汪 莉 万华根 彭群生

(浙江大学 CAD&CG 国家重点实验室 杭州 310058)

**摘 要** 分子场是研究分子结构与功能的重要工具之一,已经成为药物设计和分子对接的常规方法。分子间作用力分布在 3 维空间,现有的蛋白质显示方法,如范德华表面,分子表面,球棍模型等,都难以对分子间作用力进行空间可视化分析。分子间各种相互作用中,静电力和范德华力在许多生物过程中都扮演着重要的角色,对蛋白质静电相互作用和范德华相互作用的正确计算模拟有着重要的理论和实践意义。分子场的计算分析从蛋白质分子结构出发,先经过分子模拟采样计算得到分子场,再对分子场进行可视化分析,提出针对分子场特点的分步式传输函数设计法,直观地反映了分子场在空间的分布,对于分析蛋白质结构功能、了解分子间相互作用机理以及进行药物设计和分子对接都有重要作用。

**关键词** 体绘制 分子场 静电势能 范德华势能 药物设计 分子对接 蛋白质

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1006-8961(2010)03-530-06

## Visualization of Protein Interaction Fields Based on Molecular Mechanics

WANG Li, WAN Huagen, PENG Qunsheng

(State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058)

**Abstract** Molecular interaction field (MIF) is an important tool used in the research of molecular structure and function and has become a conventional method in drug design and molecular docking. Yet it is difficult to visualize and analyze molecular interactions with existing graphical representations of proteins such as VDW, molecular surface and CPK due to three dimensional nature of protein interaction distribution. A novel method for volumetrically visualizing and analyzing the MIF of a protein of known structure is proposed. Beginning from a protein's molecular structure, we compute an MIF after performing a molecular dynamics simulation on the protein and then visualize the MIF volume. A two-pass method for designing transfer functions is proposed specifically for producing informative images that reveal the MIF volume features. The volume rendering results are used to analyze HIV-1 protease with biologically meaningful result being derived.

**Keywords** volume rendering molecular interaction field electrostatics Van der Waals drug design molecular docking protein

## 0 引言

蛋白质是生命的重要分子,几乎在所有的生物学过程中都扮演着重要角色。借助于 X 射线晶体

学、多维核磁共振技术或同源性建模等手段,已经揭示了许多蛋白质的空间 3 维结构。而对于蛋白质功能进行分析的主流方法仍依赖于实验过程,这些实验往往需要耗费大量的资金和时间。同样,药物设计的主流方法也依赖于实验过程。

基金项目:国家自然科学基金重点项目(60533050)

收稿日期:2008-09-09 改回日期:2009-02-08

第一作者简介:汪莉(1982—),女,浙江大学博士研究生。主要研究领域为分子图形学与生物计算、科学计算可视化。E-mail: livang@cad.zju.edu.cn

通讯作者:万华根, E-mail: hgwan@cad.zju.edu.cn

分子对接是分子模拟的重要方法之一,目前的分子对接方法是研究小分子与大分子相互作用模式、生物大分子间识别、分子自组装、超分子结构等课题的常用方法之一,其过程涉及分子之间的空间匹配和能量匹配,空间匹配是分子间发生相互作用的基础,而能量匹配则是分子间保持稳定结合的基础<sup>[1]</sup>。因此,分子间相互作用分析是分子对接的核心技术之一。

实验结果表明,分子相互作用力中,静电相互作用和范德华相互作用是唯一一起决定作用,能产生生物化学效应的两种作用力。静电相互作用和范德华相互作用是生物作用中仅有的两种长程作用力,几乎影响着生物化学作用的各个方面,如维持蛋白质的三级结构和四级结构的稳定性等。静电相互作用和范德华相互作用还决定了分子的结构和功能特征,如 3 维形状、分子定点突变的 pKa 效应、结合能量、催化率等。对静电相互作用和范德华相互作用的准确模拟对基于结构的药物设计及蛋白质折叠都有着重要的理论和实践意义<sup>[2-3]</sup>。因此,以直观的方式显示分析这两种作用能的分布情况有着重要意义。

有很多种表示蛋白质的模型,按其类型可以分为 3 类: 结构模型, 表面模型, 场模型。结构模型主要有骨架模型、棍状模型、球棍模型、CPK (空间填充) 模型、带状模型等, 这些模型显示了蛋白质的 3 维结构。表面模型有等值面模型, 范德华表面模型, 分子表面模型等, 显示了蛋白质稳定状态下的作用面。这两种模型以实验数据为基础, 对蛋白质 3 维结构中原子的空间定位及其连接关系、C<sub>α</sub> 链、二级结构、motif 等做合理抽象, 将蛋白质表示为线、棍、球、范德华表面和条带, 在辅助人们理解蛋白质空间结构及功能上起着重要作用, 成为表达生物大分子的主要手段<sup>[4]</sup>。而分子场则将分子间相互作用表示为 3 维场, 反应了分子间相互作用时能量空间分布的变化, 有利于揭示蛋白质结构功能及分子间相互作用的机理<sup>[5]</sup>, 在药物设计和材料设计领域都有着广泛应用。

目前, 主要有两种方法来模拟显示蛋白质相互作用能的空间分布。一种是将蛋白质相互作用能映射到分子表面上。分子表面是蛋白质分子的一种主要表达方式。各种分子属性, 如原子电荷、静电势能、疏水性、极化度等, 都可以映射到分子表面上, 通过颜色编码和纹理映射来绘制<sup>[6]</sup>。

另一种方法通过绘制蛋白质分子场来反应蛋白

质能量分布<sup>[5-7,11]</sup>。分子场是定义在已知分子结构的蛋白质空间 3 维网格上的一个标量场, 分子场计算的主要思想是将蛋白质周围空间均匀分割为网格并在每个格点上计算蛋白质与探针之间的相互作用。一般分子场包含范德华相互作用和静电相互作用, 有时也包含疏水作用和氢键<sup>[7]</sup>。有各种不同的方法来绘制分子场。如抽取分子场中特定能量值的等值面并绘制。最早 Chan 等人提出将抽取的等值面绘制为点线模型<sup>[8]</sup>, 后来 Goodford 等人提出将抽取等值面绘制为面<sup>[5-9]</sup>, 这一方法现在仍广泛应用于生物化学领域。Goodsell 等人同时提出利用光线跟踪将分子场绘制成围绕分子结构实体模型 (球棍模型等) 的密度云<sup>[9]</sup>。Wallacott 等人将触觉设备引入分子场, 通过力反馈提供更多的信息来分析分子场<sup>[10]</sup>。韩玮等人对分子场进行局部微分计算得到一系列临界点来揭示蛋白质分子的功能区域所在, 并通过点面体等绘制技术来揭示分子结构<sup>[11]</sup>。

分子间相互作用是分布在 3 维空间的动态连续场。以将蛋白质相互作用显示和分析得更直观为目的, 本文提出以体绘制的方法对蛋白质相互作用进行可视化分析。利用直接体绘制技术来绘制蛋白质分子场以显示更多分子场内部信息并提出了针对分子场的传输函数设计方法。针对 HIV-1 蛋白酶 (PDB id 1a30) 的可视化分析结果表明, 本文的方法直观地反应了分子间相互作用在空间的分布, 有利于分析分子结构功能、了解分子间相互作用机理, 进一步加快了发现药物分子和配体分子的速度, 方便了分子对接和药物设计。

## 1 蛋白质分子场

分子场计算的基本原理主要有两种<sup>[4]</sup>: 量子力学和分子力学。量子力学方法计算更准确, 但是计算量很大, 因为计算过程中要进行大量的双电子四中心积分计算, 在目前绝大多数计算条件下, 严格的量子化学方法只能用于处理包含数十个原子的小分子体系。因此, 就量子力学方法而言, 现阶段生物大分子仍是难以处理的复杂体系<sup>[12]</sup>。

分子力学则应用经验势能函数, 即经验力场方法模拟分子的结构。它将分子看做一组靠某种作用力维系在一起的原子集合, 将分子系统的能量看成以各原子空间坐标为参数的经验函数, 由化学键相互作用能和非键相互作用能组成<sup>[1,3]</sup>。分子力学方

法计算的关键是能量计算公式中的参数,即力场参数。力场参数有各种来源,包括从头算法和半经验计算结果、实验观察结果等。最常用的力场有 AMBER 力场、CHARMm 力场、CVFF (consistent valence force field) 力场、CFF (consistent force field) 力场<sup>[4]</sup>等。在许多情况下,分子力学方法的计算结果与量子力学方法计算结果同样准确,而计算量仅与原子数目的平方成正比。

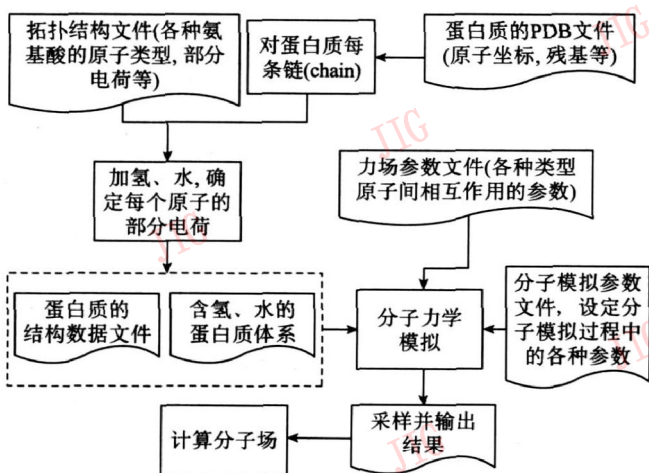


图 1 分子场计算示意图

Fig. 1 Process of protein MIF computation

本文采用分子力学的方法计算蛋白质分子场。由于分子间相互作用比较复杂,一般先分析大分子与各种探针之间的相互作用,综合得出大分子与配体分子的结合模式及亲和力。从本质上说,蛋白质分子的分子场是分布在 3 维空间的动态连续场,为便于 3 维可视化分析,采用离散方法对蛋白质分子场进行建模,将其表示为 3 维空间离散均匀网格上的标量场。如图 1 所示,以蛋白质分子的 X 射线衍射构象为出发点,添加氢原子,计算每个原子上的部分电荷并构造含有水分子的体系。然后采用 CHARMm 力场、在恒温恒压条件下进行平衡计算,模拟蛋白质在体内液体环境的柔性结构,再进行分子力学模拟,同时进行时间维采样计算,采样计算的结果为分子的运动轨迹,即分子随时间变化的坐标集合。上述平衡计算与分子力学模拟都是用 NAMD 实现的。最后,将采样的蛋白质分子所处空间均匀剖分为网格,并在网格点上定义离散函数来计算探针与蛋白质分子间作用力。根据 Lennard-Jones 6-12 与库仑公式计算分子每一时间采样点的分子场。同时为了便于处理,分别存储了蛋白质分子的静电场、范德华场和分子场,尽管这样做增加了存储空间,但

是方便了在几种作用场之间切换和分别处理。

## 2 分子场可视化

给定蛋白质分子场,用直接体绘制技术来提供蛋白质结果、功能以及蛋白质与其他分子相互作用的直观信息。相对面绘制技术来说,直接体绘制能显示更丰富的内部信息。

直接体绘制算法分为图像空间算法和物体空间算法两大类。本文采用光线投射法这一典型的图像空间算法对所计算的蛋白质分子场进行可视分析。首先从屏幕上每一像素点根据视点方向发出一条射线,这条射线穿过数据场;然后沿着这条射线选择  $K$  个等距的采样点,对每个采样点,由距离该采样点最近的 8 个数据点的颜色值和不透明度值作 3 次线性插值,求出该采样点的不透明度值及颜色值;最后,按采样顺序(由前往后或由后往前)将每条射线上所有采样点的颜色值和不透明度值进行合成,得到结果图像<sup>[13-14]</sup>。

体绘制算法的关键是设计传输函数,因为传输函数将 3 维数据场的数值转换为光学成像参数,如颜色、不透明度值等等,决定了投影图像的质量。全自动传输函数生成算法可能丢失了体数据的重要特征,而纯手工设计传输函数的方法则是耗时耗力。因此,提出分两步来指定分子场的传输函数,先用 Kindmann 等人提出的半自动传输函数设计方法<sup>[15]</sup>生成传输函数模板使得用户对被处理的分子场有初步印象。在此基础上,提供用户一个直观的传输函数设计界面,用户可以方便地交互调整第一步产生的传输函数模板直到能绘制出用户感兴趣的分子场特征。下面具体介绍:

第 1 步 本文以体直方图为基础<sup>[15-16]</sup>设计传输函数模板。

设  $x$  表示体素沿梯度方向离边界中心的距离, $f(x)$  表示其函数,即原始数据值。计算每个体素沿其梯度方向的一阶方向导数  $f'(x)$  和二阶方向导数  $f''(x)$ 。

以  $f(x)$ ,  $f'(x)$ ,  $f''(x)$  为坐标轴,建立体直方图数组:遍历体,根据每个体素的数据值、一阶方向导数和二阶方向导数,确定该体素在体直方图中对应的元素,增加该元素的值。

计算一阶方向导数、二阶方向导数的平均值:用  $v$  表示灰度值, $g(v)$  表示对于所有体素  $x$ ,  $f(x) = v$  对应的一阶方向导数的平均值, $h(v)$  表示对于所有点  $x$ ,  $f(x) = v$  所对应的二阶方向导数的平均值(如图 2

(a)、图 2(b)所示)。

位置函数  $p(v)$  表示灰度值到边界的距离 (如图 2(c)所示), 计算公式如下:

$$p(v) = \frac{-\sigma^2 h(v)}{\max(g(v)) - g_{\text{thresh}} - 0} \quad (1)$$

式中,  $g_{\text{thresh}} > 0$  为梯度阈值, 因为物质内部梯度很少为 0  $\sigma$  为边界厚度:

$$\sigma = \frac{2\sqrt{e \max_v(g(v))}}{\max_v(h(v)) - \min_v(h(v))} \quad (2)$$

计算不透明度函数  $a(v)$  (如图 2(d)所示):

$$a(v) = b(p(v)) \quad (3)$$

式中,  $b(x)$  为边界强调函数 (如图 3所示), 将位置映射为不透明度, 控制边界绘制效果。

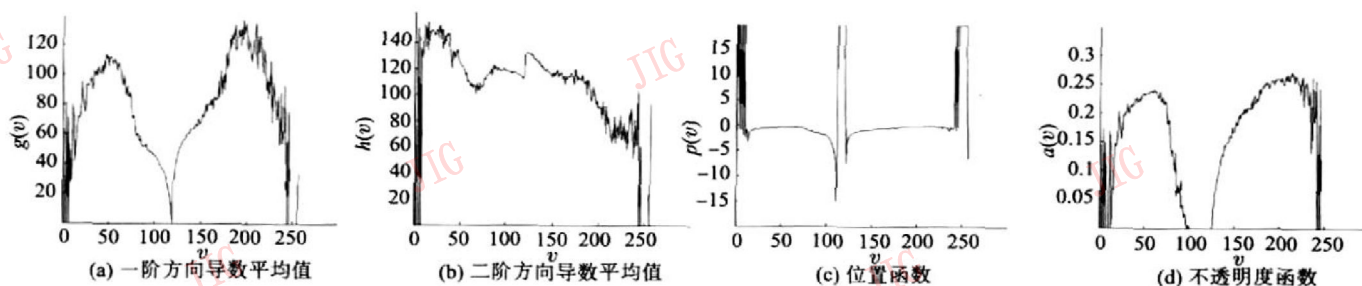


图 2 体绘制部分数据示意图 (HIV-1 蛋白酶)

Fig. 2 Some data in volume rendering (HIV-1 protease)

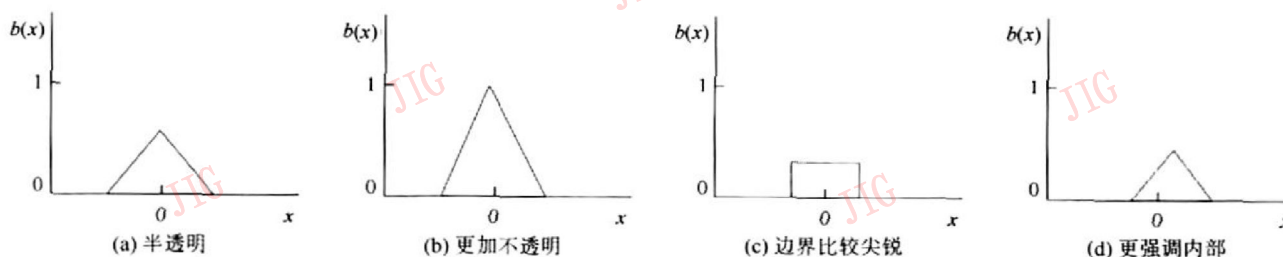


图 3 几种边界强调函数示意图

Fig. 3 Boundary emphasis functions

第 2 步 调整传输函数。对于上面生成的传输函数模板, 设计了一个直观而交互的传输函数调整界面 (如图 4所示) 方便用户调整传输函数。传输函数图中, 水平轴表示体素值而纵轴表示颜色 (RGB) 和不透明度。由于本文采用的传输函数为分段线性函数, 用户可以通过在屏幕上用鼠标拖拽其中一条或几条曲线来反复调整传输函数直到达到满意效果。需要指出的是, 这种调整并不是盲目的。本文为不同蛋白质的分子场设计了各种传输函数, 发现了分子场的一些特点, 并以此指导传输函数的调整:

1) 水平轴的两端分别对应分子场的最大值和最小值, 或者说是分别分子间作用力中排斥和吸引两种力的最大值, 因为分子间相互作用有吸引和排斥两种。

2) 分布在区间  $\pm 0.01$  kcal/mol 内的分子场值 (对应于传输函数图中分子场值统计直方图 (土黄色) 中值最大的区域) 对应于分子场的中性区域, 这

些区域与探针没有作用力, 是生物化学工作者不感兴趣的区域, 因此, 可以将这些值设为完全透明。

3) 作用力 (排斥或吸引) 较大的区域都位于蛋白质内部, 且以这些区域为中心, 作用力从内往外逐渐变小。

4) 作用力大小接近中性区域值的区域都分布在分子表面附近。

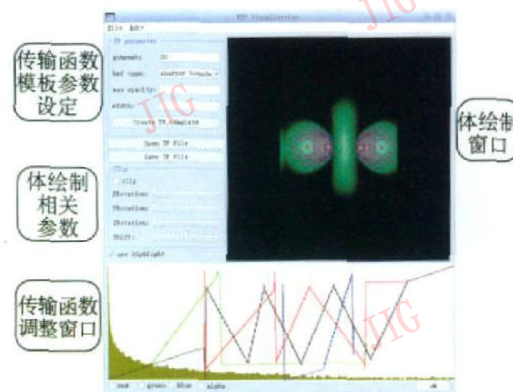


图 4 传输函数设计界面 (以水分子为例)

Fig. 4 The transfer function interface (for molecule  $H_2O$ )

因此,用户可以通过调节传输函数曲线来强调自己感兴趣的区域,突出蛋白质特征。

### 3 结果分析

本文以 HIV-1 蛋白酶为例计算分子场并进行可视化,成功显示了 HIV-1 蛋白酶的排水通道。

HIV 是一种人体免疫缺陷病毒, HIV 有 I 型和 II 型两种。HIV-1 的致命力较强,是引起艾滋病流行的主要病源。HIV-1 蛋白酶是由一个小阻抗剂和两条含 99 个氨基酸的多肽链形成的 C2 对称的均二聚体,含有 3 132 个原子,活性位点位于一个狭长的“通道”的底部<sup>[3]</sup>,对这个排水通道的研究有助于 HIV 抑制剂的研究。图 5 显示了用 VMD (visual molecular dynamics)绘制的分子的各种模型表示<sup>[17]</sup>。

对 HIV-1 蛋白酶做了如下模拟:先做平衡计算模拟分子在水中的柔性结构,再对分子施加外力做分子力学模拟。分子力学模拟中,对距离蛋白质  $-y$  方向  $3\text{\AA}$  ( $1\text{\AA}=10^{-10}\text{m}$ ) 的水分子施加  $y$  方向的常力,对距离蛋白质  $y$  方向  $3\text{\AA}$  的水分子施加  $-y$  方向的常力。平衡计算的时间为 40 ps 分子力学的模拟时间为 16 ps。

以 sp3 杂化的  $C^+$  原子计算了大小为  $64 \times 64 \times 64$  的分子场。绘制程序用 vc++ 编写而界面则是用 Qt 编写。调整传输函数过程中,利用分子场的特点,迅速排除不感兴趣的区域,更快的找到有意义的区域。绘制中,采用了不同的传输函数达到了不同的绘制效果。图 6 显示的 HIV-1 蛋白酶在平衡计算过程中分子场绘制结果,对 HIV-1 蛋白酶的结构功能有了初步认识。图 7 清晰显示了 HIV-1 蛋白酶的排水通道,尤其是排水通道的截面图,这是其他绘制方法所不能显示的。

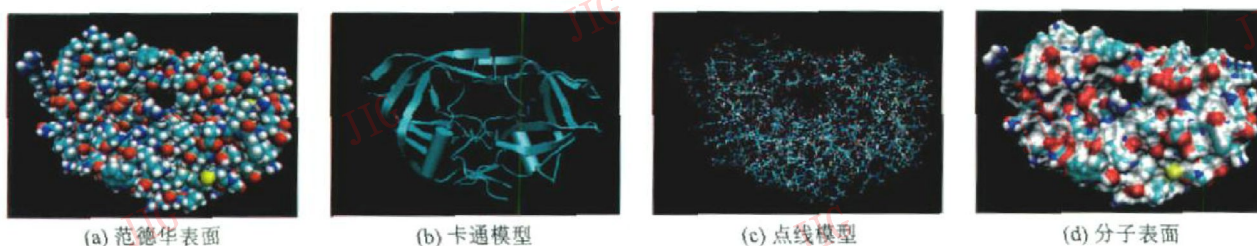


图 5 HIV-1 蛋白酶的各种表示模式 (VMD 绘制)

Fig. 5 Presentations of HIV-1 protease (rendered by software VMD)

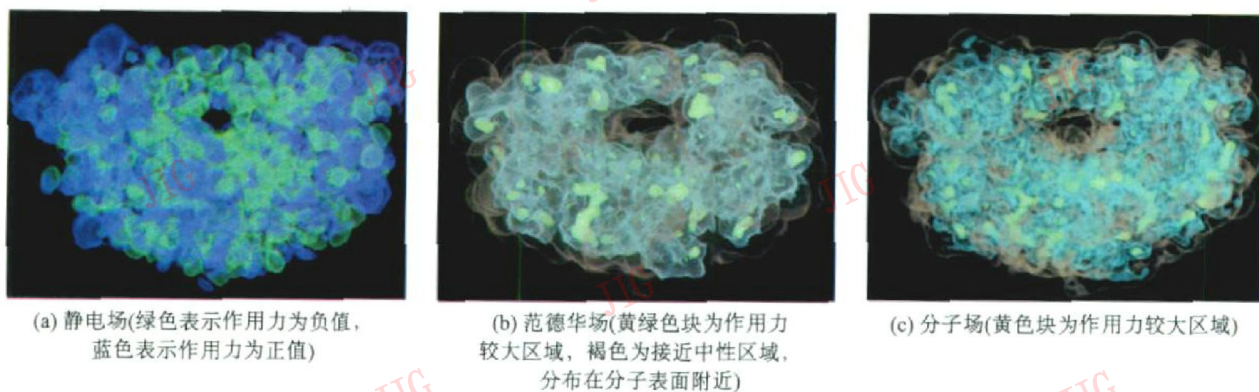


图 6 平衡计算中 HIV-1 蛋白酶的分子场 (第 2 ps)

Fig. 6 MIF of HIV-1 protease during equilibration (the 2nd ps)

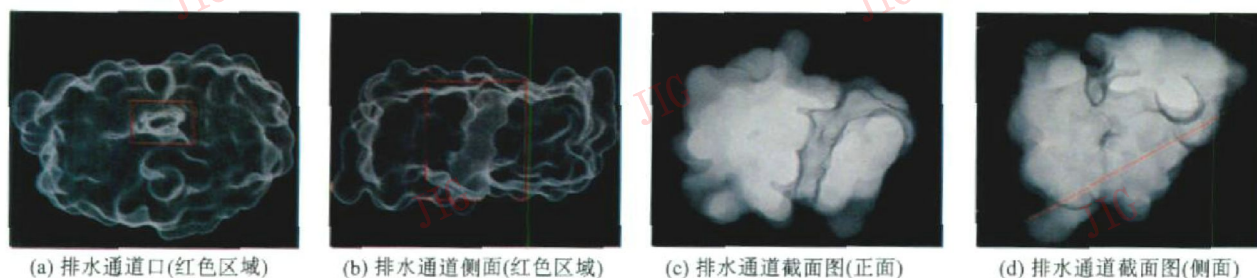


图 7 HIV-1 排水通道 (平衡计算第 2ps, 范德华场)

Fig. 7 The tunnel of HIV-1 protease (Van der Waals part of MIF, the 2nd ps during equilibration)

## 4 结 论

药物设计和分子对接是生物化学的研究重点和热点, 计算机的引入使得二者不再单纯依赖实验, 大大加快了二者发现药物分子和配体分子的速度, 降低了研究成本。分子场反映了蛋白质能量空间分布变化, 是研究蛋白质分子结构功能及分子间相互作用的重要工具, 已经成为药物设计及分子对接的常规方法。本文提出用体绘制的方法对分子场进行绘制分析, 方便了对分子场的理解分析。本文提出针对分子场特点的传输函数设计方法, 提供了丰富的分子场信息, 反映了蛋白质特征。同基于分子表面或等值面抽取的绘制方法相比, 本文方法至少有以下优点: 首先, 避免了构造中间几何图元, 直接由体数据生成图像。其次, 可以生成分子场的整体图像, 包含每一个细节, 图像质量高。最后, 直观底反应了蛋白质分子特征。

蛋白质分子分子场的研究是一项有价值且意义甚远的研究。考虑到蛋白质分子时刻处于运动当中, 对数据场进行时变可视化分析能更加有效地揭露分子在完成其功能过程中的演化情况及多分子之间相互作用的机制。

**致谢** 感谢浙江大学化学系分子设计与分子热力学研究所的吴韬教授、王琦教授、刘迎春博士, 浙江大学生命科学学院的王君晖教授给予的帮助支持和热心指导。

### 参考文献 (References)

- [1] Chen Zhenglong, Xu Weiren, Tang Lida. Molecular Modeling Theory and Practice[M]. Beijing: Chemical Industry Press, 2007. [陈正隆, 徐为人, 汤立达. 分子模拟的理论与实践[M]. 北京: 化学工业出版社, 2007.]
- [2] Branden C, Tooze J. Introduction to Protein Structure[M]. 2nd ed. New York NY, USA: Garland Publishing, Inc, 1998.
- [3] Xu Xiaojie, Hou Tingjun, Qiao Xuebin, et al. Computer Aided Drug Design[M]. Beijing: Chemical Industry Press, 2004. [徐筱杰, 侯廷军, 乔学斌, 等. 计算机辅助药物分子设计[M]. 北京: 化学工业出版社, 2004.]
- [4] Leach A R. Molecular Modelling: Principles and Applications[M]. 2nd ed. Edinburgh, UK: Pearson Education EMA, 2001.
- [5] Goodford P J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules[J]. Journal of Medical Chemistry, 1983, 28(7): 849-857.
- [6] Petrey D, Honig B. GRASP2: Visualization, surface properties and electrostatics of macromolecular structures and sequences[J]. Methods in Enzymology, 2003, 374: 492-509.
- [7] Cruciani G. Molecular interaction fields: applications in drug discovery and ADME prediction[M]. Weinheim, Germany: Wiley-VCH, 2006.
- [8] Chan P L, Dean P M. Molecular recognition: 3D surface structure comparison by gnomonic projection[J]. Journal of Molecular Graphics, 1987, 5(2): 92-96.
- [9] Goodsell D S, Olson A J. Molecular applications of volume rendering and 3-D texture maps[C]//Proceedings of 1989 Chapel Hill Volume Visualization Workshop. Chapel Hill, North Carolina: ACM Press, 1989, 27-31.
- [10] Wollacott A M, Merz K M Jr. Haptic applications for molecular structure manipulation[J]. Journal of Molecular Graphics and Modeling, 2007, 25(6): 801-805.
- [11] Han Weij, Wang Lij, Chen Weij, et al. Feature analysis and visualization of 3D scalar field with the applications to the macromolecule[J]. Journal of Software, 2006, 17(sup): 103-109. [韩玮, 汪莉, 陈为, 等. 面向生物大分子的 3 维数据场特征分析与可视化初探[J]. 软件学报, 2006, 17(增刊): 103-109.]
- [12] Raymond K. Progress in the theory of mixed quantum-classical dynamics[J]. Annual Review of Physical Chemistry, 2006, 57: 129-157.
- [13] Tang Zesheng. Scientific Visualization of 3D Data Set[M]. Beijing: Tsinghua University Press, 1996. [唐泽圣. 3 维数据场可视化[M]. 北京: 清华大学出版社, 1996.]
- [14] Parker S, Parker M, Livnat Y, et al. Interactive ray tracing for volume visualization[C]//Proceedings of the ACM SIGGRAPH 2005 Course Notes. New York, USA: ACM Press, 2005, 1-15.
- [15] Kindhann G, Durkin J W. Semi-automatic generation of transfer functions for direct volume rendering[C]//Proceedings of the 1998 IEEE Symposium on Volume Visualization, Research Triangle Park, North Carolina: ACM Press, 1998, 79-86.
- [16] Kindhann G. Transfer functions in direct volume rendering: Design, interface, interaction[C]//Proceedings of the ACM SIGGRAPH 2002 Course Notes. Singapore: ACM Press, 2002, 1-6.
- [17] William H, Andrew D, Klaus S. VMD: Visual molecular dynamics[J]. Journal of Molecular Graphics, 1996, 14(1): 33-38.