

基于高斯曲率的蛋白质嵌入到对称低温电镜密度图的方法

路建波¹, 高华方¹, 卢本卓², 马旭¹

(1. 国家卫生计生委科学技术研究所人类遗传资源中心, 北京 100081;

2. 中国科学院数学与系统科学研究院计算数学与科学工程计算研究所, 北京 100190)

摘要: 为了解决低温电镜三维重构得到的对称生物大分子三维结构分辨率低的问题, 提出一种新的基于高斯曲率计算的自动化的方法. 通过计算生物分子表面的高斯曲率和优化模型等数学方法把单个蛋白质或其他亚基嵌入到对称的低温电镜密度图中, 通过计算蛋白质和对称低温电镜密度图的离散高斯曲率来定义特征点, 对低温电镜密度图, 利用 K-means 分类来定义相应的特征点. 接下来用均方根偏差值和曲率的交叉关联系数和优化模型来匹配对应的特征点. 然后, 用最小二乘法计算旋转矩阵和平移向量. 最后用同样的方法把其他部分也嵌入到对称的生物大分子中. 给出了椭球体 $x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$ 的解析高斯曲率和数值计算的高斯曲率的比较、蛋白质 1DOR 的高斯曲率计算和特征点的计算结果、蛋白质 1DOR 和 2REC 的分子装配结果可视化. 数值实验结果表明本文所提出的一系列的方法是准确和有效的.

关键词: 大分子结构; 低温电镜; 表面高斯曲率; 优化模型

中图分类号: Q 819

文献标志码: A

文章编号: 0254-0037(2017)12-1881-07

doi: 10.11936/bjtxb2017040025

Method of Protein Fitting into the Symmetric Cryo-EM Density Maps Based on Gaussian Curvature

LU Jianbo¹, GAO Huafang¹, LU Benzhuo², MA Xu¹

(1. Human Genetics Resource Center, National Research Institute for Family Planning, Beijing 100081, China;

2. LSEC, Institute of Computational Mathematics, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: To solve the problem of low resolution of three-dimensional symmetric biomacromolecule structure obtained by Cryo-electron microscopy (cryo-EM) density maps, a new method based on Gaussian curvature calculation was proposed. The new feature points were defined by calculating the discrete Gaussian curvature of the protein and the symmetrical cryo-EM density maps. For the symmetrical cryo-EM density maps, the K-means method was used to define the corresponding feature points. The corresponding feature points were matched by the cross correlation score of root mean square deviation (RMSD) values and Gaussian curvature and the optimization model. The rotation matrix and the translation vector were then calculated using the least squares method. Finally, the other parts were

收稿日期: 2017-03-31

基金项目: 国家重点研究发展计划资助项目(2016YFC1000307, 2016YFB0201304); 国家自然科学基金资助项目(21573274), 国家重点研究发展计划子课题资助项目(2016YFC1000307-10); 国家卫生计生委科学技术研究所科技创新基金面上项目(2017GJM04)

作者简介: 路建波(1982—), 男, 助理研究员, 主要从事生物信息学、计算生物学、计算机辅助小分子药物设计方面的研究, E-mail: jblu@lsec. cc. ac. cn

fitted into the symmetrical cryo-EM density map in the same way. The Gaussian curvature of ellipsoid was calculated, $x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$, and the analyzed Gaussian curvature was compared with the numerical Gaussian curvature. The Gaussian curvature of the protein 1DOR was calculated and the feature point is also calculated. The fitting results of the macromolecules assembly of protein 1DOR and 2REC were visualized. The numerical experiments show that this method is accurate and effective.

Key words: macromolecular structure; cryo-EM; gaussian curvature; optimization model

低温电镜技术最近受到科研人员的重点关注, 该技术不仅能解析蛋白质和病毒的三维结构, 而且可以处理细胞器和超分子. 其优点是不需经过繁复的步骤制备样品, 就可以通过大量采集其投影图像、图像处理技术和三维重构方法得到蛋白或细胞器在近生理状态的三维结构. 目前低温电镜三维重构得到的三维结构分辨率还很低, 仍然需要其内部的独立部分的高分辨率的三维结构, 他们可以通过结晶、NMR、或者蛋白质结构比较等方法^[1-3]得到. 这些高分辨率的独立的结构包括: 蛋白质、结构域、核糖体等. 把这些高分辨率的独立结构装配到低温电镜的整体密度图中^[4-9], 就可以得到一个完整的生物大分子三维结构^[10-11]. 近几年研究人员提出了一系列的将高分辨率的晶体结构、低温电镜密度图和计算生物相结合的建模方法^[12-19]. 本文提出一种新的自动化的方法解决对称生物大分子的装配问题.

1 研究现状

2007年, SALI 课题组^[3]采用多种方法详细研究了酵母核膜孔复合体的三维结构. 该方法综合利用低温电镜三维重构技术、超速离心技术、定量免疫分析技术、亲和层析技术、质谱技术、免疫电镜技术, 再结合约束条件和计算生物学, 建立模型, 最终得到其高分辨率三维结构. 1999年, Wriggers 等^[20]发表了将高分辨率晶体结构嵌入到低分辨率低温电镜密度图的向量量化的方法. 该方法用向量量化的特征点代替低温电镜密度图的图像和高分辨率的蛋白质或者其他亚基, 并完成了相应的 Situs 程序包. Hugo 等^[21]采用蛋白质和低温电镜密度图的表面的方法, 并需要考虑到各个蛋白质(或者亚基)之间的连接关系. Pintilie 等^[22]采用分水岭和尺度空间滤波的方法即首先利用分水岭方法对低温电镜密度图像进行分割, 然后用尺度空间滤波的分组, 最后把单独的每一部分填充到低温电镜密度图中. Kawabata^[23]于2008在 Biophysical Journal 上发表的文章, 针对对称的生物大分子, 利用高斯混合模型和 EM 算法建立

了新的模型. Topf 等^[24]综合利用蒙特卡洛搜索、共轭梯度最小化、模拟退火分子动力学模拟等方法, 提出一种新的嵌入和优化模型. Zhang 等^[25]于2010年提出了二次规划的方法, 该方法用向量量化分别把密度图和蛋白质(或者亚基)转化为几个特征点^[26-27], 建立优化模型.

在大分子细胞器的分子装配过程中, 大部分研究者都会用到特征点的方法. 本文针对对称的生物大分子结构, 提出一种基于蛋白质高斯曲率表面的方法, 结合优化模型来定义特征点模型.

2 研究方法和结果

2.1 曲率的相关概念

曲率是刻画曲面弯曲程度的重要几何概念. 其定义是, 在曲面 S 的任意一点 $x(u, v)$ 的切平面上定义线性变换 W , 称之为 Weingarten 变换, 使得

$$W(x_u) = -n_u, \quad W(x_v) = -n_v \quad (1)$$

那么对于切平面上的任何一个向量 $t = ax_u + bx_v$, 有

$$W(t) = -an_u - bn_v \quad (2)$$

曲面参数的选择与这个变换无关, 并且 Weingarten 变换是线性的自共轭的变换. 曲面的主曲率是 Weingarten 变换的 2 个实特征值 k_1, k_2 .

由 k_1, k_2 可以得到高斯曲率的定义

$$K_G = k_1 k_2 \quad (3)$$

和平均曲率的定义

$$K = \frac{k_1 + k_2}{2} \quad (4)$$

2.2 数据的初步处理

首先, 用 Situs 软件包中的 pdb2vol 软件, 在 linux 系统下, 把高分辨率的蛋白质结构转化为低分辨率的低温电镜密度图三维结构或从 EMDB 数据库下载低温电镜密度图. 这里体素的取值均是 0.3 nm, 用的核函数均是高斯光滑核函数

$$N(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

式中 $\sigma = 0.8826$ nm. 用三线插值法转化为正方体

网格. 用面绘制的显示软件 Chimera^[28] 生成结构体等值面网格. 该网格包括: 1) 三角形网格的顶点编号; 2) 网格的顶点; 3) 网格顶点的法向量. 接着提取顶点和三角形的顶点编号.

2.3 不规则几何体表面的曲率计算

根据 Gauss-Bonnet 定理可以将高斯曲率离散化^[29-30], 离散的高斯曲率表示为

不规则曲面在点 P 处的高斯曲率

$$K_G(P) = \frac{3(2\pi - \sum \theta_i)}{A(x_i)} \quad (6)$$

式中: θ_i 为角度 $\angle Q_i P Q_{i+1}$; $A(x_i)$ 为围绕点 P 的表面三角形网格的面积总和; Q_i, Q_{i+1} 为三角形网格的 2 个顶点.

曲面的特殊参数可表示为 $z = z(x, y)$, 高斯曲率的准确计算公式为

$$K_G = \frac{rt - s^2}{(1 + p^2 + q^2)^2} \quad (7)$$

式中 p, q, r, s, t 分别表示为 $p = \frac{\partial z}{\partial x}, q = \frac{\partial z}{\partial y}, r = \frac{\partial^2 z}{\partial x^2}, s = \frac{\partial^2 z}{\partial x \partial y}, t = \frac{\partial^2 z}{\partial y^2}$.

文中曲率如果没有说到高斯曲率还是平均曲率, 则曲率指的是高斯曲率. 椭球体 $x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$ 的解析高斯曲率分布图和数值计算高斯曲率分布图分别由图 1、2 给出. 图 1 是椭球体 $x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$ 的解析高斯曲率分布图, 可以看出, 整个椭球体的真实高斯曲率在 0.2 ~ 1.4. 高斯曲率比较大的顶点分布在椭球体的最长轴顶端, 高斯曲率比较小的顶点分布在椭球体最短轴的顶端附近. 从曲率最小到最大是逐渐递增的过程. 图 2 是椭球体 $x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$ 的数值计算高斯曲率分布图, 仔细观察不难发现数值计算的高斯曲率在 0 ~ 2. 同样是曲率比较大的顶点分布在最长轴的顶端附近, 曲率比较小的顶点分布在最短轴的顶点附近. 从曲率最小到最大也是逐渐递增的过程. 比较图 1、2, 可以得到, 高斯曲率的分布规律相同, 但是曲率范围不一样. 数值计算的曲率值相对较大, 但是差别不是很大. 从计算的结果可以发现差别在准确曲率的 2 倍以内.

2.4 低温电镜密度图的高斯曲率

由 Chimera 软件得到的三角形网格有很大噪音, 例如有的三角形网格内 3 条边长度差距很大, 从而高斯曲率也会很大. 本文中出現这种情况, 就令

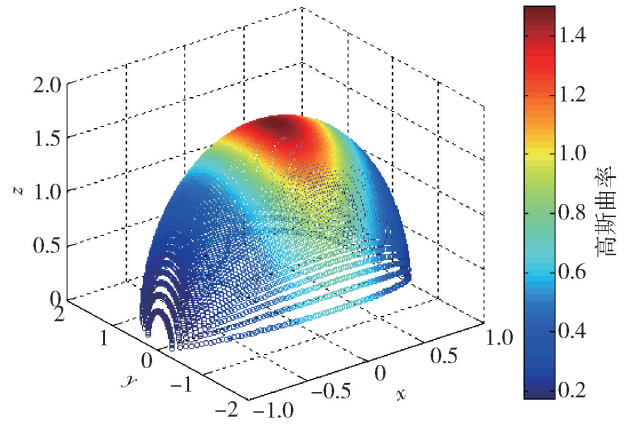


图 1 椭球体 $x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$ 的解析高斯曲率分布图

Fig. 1 Exact Gaussian curvature value of an ellipsoid:

$$x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$$

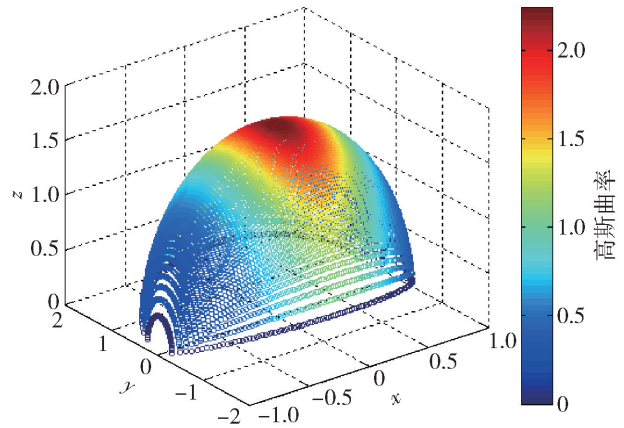


图 2 椭球体 $x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$ 的数值计算高斯曲率分布图

Fig. 2 Numerical Gaussian curvature value of an

$$\text{ellipsoid: } x^2 + \frac{y^2}{2} + \frac{z^2}{3} = 1$$

其高斯曲率为 0, 或者删掉这个三角形网格.

为了减小误差带来的曲率计算不准确, 采用如下方法: 首先计算蛋白质或者低温电镜密度图的每一个点的高斯曲率, 然后按照曲率值递减顺序排序. 取前 r_n 个顶点, 然后搜索每个顶点的球形领域的顶点, 半径 $r = 4.5 \times 10^{-10}$ m. 得到 2 个矩阵 C 和矩阵 G , C_{ij} 是顶点在搜索范围内的顶点编号, G_{ij} 是对应顶点的曲率. 顶点 P 的高斯曲率定义为其邻域的所有顶点的曲率平均值. 图 3 为计算的低温电镜下的几何体密度图的曲率分布. 把小于 0 的高斯曲率都设置为 0, 图 3 显示的是蛋白质 1DOR 的低温电镜密度图的高斯曲率分布图. 可以发现高斯曲率比较大

的顶点都分布在相对较尖的地方. 通过计算得到对于 1DOR 的低温电镜密度图的高斯曲率分布在 $-0.02 \sim 0.18$, 比较接近于 0. 高斯曲率数值为正的顶点表明这个顶点向外凸, 反之高斯曲率为负的顶点表明这个顶点向内凹.

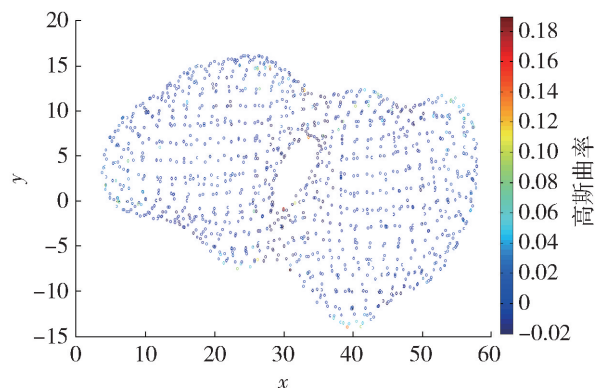


图3 蛋白质 1DOR 的高斯曲率分布图
Fig. 3 Distribution of Gaussian curvature of protein 1DOR

2.5 对称生物大分子特征点的分割

由于研究对象是对称的生物大分子, 所以可以用 K-Means 分类. 该方法是一种动态聚类的方法, 它是由麦卡因 (MacQueen) 在 1967 年提出来的. 经 K-means 分类后的每一类的特征点记为 N_{em} , 和它对应的曲率集合记为 G_{em} .

用以上方法得到的低温电镜下蛋白质 1DOR 的一个亚基的特征点分布如图 4 所示. 可以看到, 对于对称的生物大分子, 采用 K-means 分类以后得到的特征点的分割基本符合本文对于特征点的要求. 红色的顶点表示特征点, 绿色的顶点表明是其他的顶点. 经过计算得到 7 个特征点, 但是有两个特征点不是需要的特征点, 因为它们在几何体的内部凹陷地方, 这是由于误差导致的错误, 可以利用它们附近的高斯曲率为负值排除它们. 另一部分的特征点没有显示, 计算的过程发现与显示的特征点几乎一样, 这是因为它们是对称的缘故.

2.6 特征点的匹配

这里把分割后 Target 的特征点与 Probe 的特征点匹配的问题抽象成一个优化问题:

把 Target 上的经过分类的特征点记为集合 T_{em} , Probe 上的特征点记为集合 P_{pr} .

记集合 T_{em} 和 P_{pr} 的特征点个数为 $|T_{em}| = m$, $|P_{pr}| = n$, 显然 $m < n$. 对集合 N_{em} 和 N_{pr} 中的特征点按照高斯曲率递减顺序排序. 接下来在集合 T_{em} 中

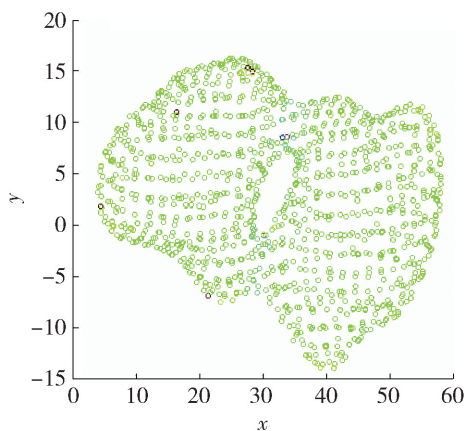


图4 低温电镜下蛋白质 1DOR 的一个亚基的特征点分布
Fig. 4 Feature points distribution of one subunit of protein 1DOR

选 3 个点, 共有 $|T_{em}|^3$ 种情况. 假设所选的顶点为 $n_{em}^1, n_{em}^2, n_{em}^3$, 对应的曲率为 $g_{em}^1, g_{em}^2, g_{em}^3$. 然后在集合 P_{pr} 中逐一与他们的组合计算最小 RMSD 值和曲率的交叉关联系数. 得到曲率的交叉关联系数

$$cc(G) = \frac{\mathbf{u}\mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \frac{\sum_{i=1}^n u_i v_i}{\sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n v_i^2}} \quad (8)$$

式中: \mathbf{u}, \mathbf{v} 分别表示 target 和 probe 上的特征点的曲率组成的向量.

最小化 RMSD 可通过解下面的优化模型, 集合 P_{pr} 中顶点高斯曲率较大或者较小的顶点可删掉.

$$\begin{aligned} \min \quad & \sum_{i=1}^m |k_{x_i} - k_{a_i}| + \sum_{i=1}^m \sum_{j=1}^m |x_{ij} - a_{ij}| \quad (9) \\ \text{s. t.} \quad & x_{ij} = |x_i - x_j|, \quad x_i, x_j \in P \\ & a_{ij} = |a_i - a_j|, \quad a_i, a_j \in T \end{aligned}$$

式中: k_{x_i} 为点 x_i 的曲率; k_{a_i} 为点 a_i 的曲率.

通过求解, 就得到了与 $n_{em}^1, n_{em}^2, n_{em}^3$ 对应的在 probe 上的特征点不妨记为 $n_{pr}^1, n_{pr}^2, n_{pr}^3$.

2.7 求旋转矩阵 \mathbf{R} 和平移向量 \mathbf{T}

有了对应的特征点, 下一步用最小二乘法^[31-32] 求出 T_{em} 上特征点 $n_{em}^1, n_{em}^2, n_{em}^3$ 到 P_{pr} 上的特征点 $n_{pr}^1, n_{pr}^2, n_{pr}^3$ 的旋转矩阵 \mathbf{R} 和平移向量 \mathbf{T} . 然后, 把旋转矩阵 \mathbf{R} 和平移向量 \mathbf{T} 作用到整个蛋白质或者低温电镜密度图中.

2.8 旋转矩阵和平移向量的最小二乘求解法

设 X_1, X_2, \dots, X_n 和 Y_1, Y_2, \dots, Y_n 分别是 X 和 Y 对应蛋白质的原子的三维坐标. 在下面的目标函数中, \mathbf{R} 为一个 3×3 的旋转矩阵, \mathbf{T} 为平移向量, $S_{i,j}$

为分配变量.

$$\min \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} S_{i,j} |X_i - \mathbf{R}(Y_j - \mathbf{T})|^2 \quad (10)$$

$$\text{s. t. } \sum_{i=1}^{n_x} S_{i,j} \leq 1, j=1, \dots, n_y$$

$$\sum_{j=1}^{n_y} S_{i,j} \leq 1, i=1, \dots, n_x$$

$$S_{i,j} \in \{0, 1\}$$

\bar{X} 和 \bar{Y} 定义为

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (11)$$

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j \quad (12)$$

式(10)中的平移向量 \mathbf{T} 可表示为

$$\mathbf{T} = \bar{Y} - \mathbf{R}'\bar{X}$$

旋转矩阵 \mathbf{R} 可用欧拉角 Φ, θ, ψ 表示为

$$\mathbf{R} = \begin{pmatrix} \cos\psi\cos\Phi - \cos\theta\sin\Phi\sin\psi & \cos\psi\sin\Phi + \cos\theta\cos\Phi\sin\psi & \sin\psi/\sin\theta \\ \sin\psi\cos\Phi - \cos\theta\sin\Phi\cos\psi & -\sin\psi\sin\Phi + \cos\theta\cos\Phi\cos\psi & \cos\psi/\sin\theta \\ \sin\theta\sin\Phi & -\sin\theta\cos\Phi & \cos\theta \end{pmatrix} \quad (13)$$

由文献[33]可以通过表达式定义四元数参数

ξ, η, ζ, χ :

$$\xi = \sin \frac{\theta}{2} \sin \frac{\psi - \Phi}{2}$$

$$\eta = \sin \frac{\theta}{2} \cos \frac{\psi - \Phi}{2}$$

$$\zeta = \cos \frac{\theta}{2} \sin \frac{\psi - \Phi}{2}$$

$$\chi = \cos \frac{\theta}{2} \cos \frac{\psi - \Phi}{2}$$

则旋转矩阵 \mathbf{R} 可以表示为

$$\mathbf{R} =$$

$$\begin{pmatrix} -\xi^2 + \eta^2 - \zeta^2 + \chi^2 & 2(\xi\chi - \eta\zeta) & 2(\xi\chi + \eta\zeta) \\ -2(\xi\eta + \zeta\chi) & \xi^2 - \eta^2 - \zeta^2 + \chi^2 & 2(\eta\chi - \xi\zeta) \\ 2(\eta\zeta - \xi\chi) & -2(\xi\zeta + \eta\chi) & -\xi^2 - \eta^2 + \zeta^2 + \chi^2 \end{pmatrix} \quad (14)$$

四元数并不是独立的, 它们有如下关系

$$\xi^2 + \eta^2 + \zeta^2 + \chi^2 = 1 \quad (15)$$

定义如下协方差矩阵 \mathbf{C}

$$\mathbf{C} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})' \quad (16)$$

目标是最小化式(10), 可通过计算 \mathbf{R} 和 \mathbf{T} 使目标函数最小. 通过计算得到 $(\xi, \eta, \zeta, \chi)'$ 是矩阵 \mathbf{P} 的最大特征值对应的特征向量.

$$\mathbf{P} = \begin{pmatrix} -C_{11} + C_{22} - C_{33} & -C_{12} - C_{21} & -C_{23} - C_{32} & C_{13} - C_{31} \\ -C_{12} - C_{21} & C_{11} - C_{22} - C_{33} & C_{13} + C_{31} & C_{23} - C_{32} \\ -C_{23} - C_{32} & C_{13} + C_{31} & -C_{11} - C_{22} + C_{33} & C_{12} - C_{21} \\ C_{13} - C_{31} & C_{23} - C_{32} & C_{12} - C_{21} & C_{11} + C_{22} + C_{33} \end{pmatrix} \quad (17)$$

可以用矩阵 SVD 分解求最大的特征值和特征向量.

3 实验结果可视化

实验用的蛋白质 1DOR 和 2REC 的数据由蛋白质数据库 PDB 给出. 蛋白质 1DOR 有 2 个完全相同的蛋白质链, 这里首先对 1DOR 蛋白质用 Situs 软件包中的 pdb2vol 软件把它转化为低温电镜下的密度图. 首先拿出一条蛋白质链, 进行平移和旋转. 经过计算得到了比较相近的结果, 图 5 为计算的一个 1DOR 链经过旋转平移后嵌入到低温电镜密度图的计算结果展示. 为了和原来的蛋白质相比较, 这里把低温电镜密度图换成了蛋白质 1DOR. 通过仔细观察计算的结果和原来的蛋白质的一条链位置有一定的差别, 但是差别不是很大. 由于问题的计算非常复杂, 前面一个小小的误差就会导致结果差别很大.

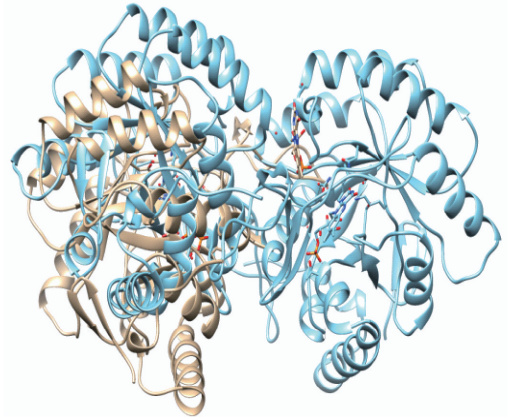


图 5 经过平移旋转后的一条链与 1DOR 的一条链的结果比较

Fig. 5 Comparison of one chain after transformation and rotation to one chain of the protein 1DOR

同样, 蛋白质 2REC 有 6 条完全相同的蛋白质链, 这里对蛋白质 2REC 用 Situs 软件包的 pdb2vol 软件把它转化为低温电镜下的密度图. 首先, 拿出一条蛋白质链, 把它做平移和旋转. 其目的是把它正确的放回去. 经过计算可以得到比较相近的结果, 如图 6 所示. 可以看到, 蛋白质 2REC 的一条链

嵌入到相应的低温电镜密度图中的结果展示. 为了和图 5 相区别, 图 6 用低温电镜的密度图和蛋白质的一条链显示结果. 可以看出, 这条链基本嵌入正确位置. 但是由于误差的存在, 嵌入没有完全正确. 这里的误差主要还是在特征点的选择, 下一步的工作将会继续完善特征点的定义和选择.

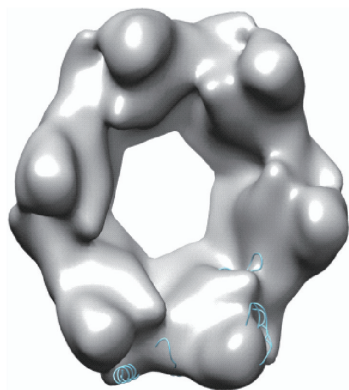


图 6 2REC 的一条链嵌入到其对应低温电镜密度图的情况

Fig. 6 Fitting of one chain of the protein 2REC to its cryo-EM density map

4 结论

本文通过计算蛋白质和对称低温电镜密度图的离散高斯曲率来定义特征点. 直接根据高斯曲率定义蛋白质的特征点; 对于低温电镜密度图利用 K-means 分类来定义它的某一个亚基的特征点, 进而与蛋白质的特征点匹配. 特征点的匹配用到 RMSD 值和曲率的交叉关联系数和优化模型. 并用最小二乘法找到旋转矩阵 R 和平移向量 T . 用同样的方法把其他部分也嵌入到对称的生物大分子中. 并用真实的数据通过实验数值说明了该方法的有效性. 通过可视化的结果展示来说明给出的一系列方法的准确性和可靠性.

参考文献:

[1] JIANG W, LUDTKE S J. Electron cryomicroscopy of single particles at subnanometer resolution [J]. *Curr Opin Struct Biol*, 2005, 15(5): 571-577.

[2] CHIU W, BAKER M L, JIANG W, et al. Electron cryomicroscopy of biological machines at subnanometer resolution [J]. *Structure*, 2005, 13, 363-372.

[3] ESWAR N, WEBB B, MARTI-RENO M A, et al. Comparative protein structure modeling using MODELLER [J]. *Curr Protoc Protein Sci*, 2007, 50(2), 912-931.

[4] ZHANG K, ZHANG Y, HU Z J, et al. Development and

frontier of electron microscopy 3d Reconstruction [J]. *Acta Biophysica Sinica*, 2010, 26(7): 533-559.

- [5] GABASHVILI I S, AGRAWAL R K, SPAHN C M, et al. Solution structure of the e. coli 70 s ribosome at 11.5 Å resolution [J]. *Cell*, 2000, 100(5): 537-549.
- [6] FABIAN K, MAREIKE K, MARKUS G G, et al. Projection structure of the secondary citrate/sodium symporter CitS at 6 Å resolution by electron crystallography [J]. *Journal of Molecular Biology*, 2012, 418: 117-126.
- [7] HENDERSON R. The potential and limitations of neutrons, electrons and x-rays for atomic resolution microscopy of unstained biological molecules [J]. *Q Rev Biophys*, 1995, 28(2), 171-193.
- [8] WANG D N, WERNER K. High-resolution electron crystallography of light-harvesting chlorophyll a/b-protein complex in three different media [J]. *Journal of Molecular Biology*, 1991, 217(4): 691-699.
- [9] STEFAN B, WILLY W. Multi-resolution anchor-point registration of biomolecular assemblies and their components [J]. *Journal of Structural Biology* 2007, 157: 271-280.
- [10] ROSSMANN M G, MORAIS M C, LEIMAN P G, et al. Combining x-ray crystallography and electron microscopy [J]. *Structure*, 2005, 13: 355-362.
- [11] FABIOLA F, CHAPMAN M S. Fitting of high-resolution structures into electron microscopy reconstruction images [J]. *Structure*, 2005, 13: 389-400.
- [12] SALI A, GLAESER R, EARNEST T, et al. From words to literature in structural proteomics [J]. *Nature*, 2003, 422(6928): 216-225.
- [13] FRANK J. Single-particle imaging of macromolecules by cryo-electron microscopy [J]. *Annu Rev Biophys Biomol Struct*, 2002, 31: 303-319.
- [14] WRIGGERS W, BIRMANNS S. Using situs for flexible and rigid-body fitting of multiresolution single-molecule data [J]. *J Struct Biol*, 2001, 133(2): 193-202.
- [15] TOPF M, SALI A. Combining electron microscopy and comparative protein structure modeling [J]. *Curr Opin Struct Biol*, 2005, 15: 578-585.
- [16] ALBER F, DOKUDOVSKAYA S, VEENHOFF L M, et al. Determining the architectures of macromolecular assemblies [J]. *Nature*, 2007, 450(7170): 683-694.
- [17] ROBERT M G, KENNETH H D. High-resolution electron crystallography of protein molecules [J]. *Ultramicroscopy*, 1993, 52: 478-486.
- [18] DEMPSTER A P, LAIRD N M, RUBIN D B. Sur la division des corps matériels en parties [J]. *Bull Acad*

- Polon Sci, 1957, 4(12): 801-804.
- [19] MACKAY D J C. Information theory, inference and learning algorithms [M]. Cambridge: Cambridge University Press, 2003: 284-292.
- [20] WRIGGERS W, MILLIGAN R A, MCCAMMON J A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy [J]. J Struct Biol, 1999, 125(2): 185-195.
- [21] HUGO C, ROBERT B R. Fast fitting of atomic structures to low-resolution electron density maps by surface overlap maximization [J]. J Mol Biol, 2004, 338: 783-793.
- [22] PINTILIE G D, ZHANG J, GODDARD T D, et al. Quantitative analysis of cryo-em density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions [J]. Journal of Structural Biology, 2010, 170: 427-438.
- [23] KAWABATA T. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model [J]. Biophysical Journal, 2008, 95: 4643-4658.
- [24] TOPF M, LASKER K, WEBB B, et al. Protein structure fitting and refinement guided by cryo-em density [J]. Structure, 2008, 16: 295-307.
- [25] ZHANG S H, DAVEB V, MIN X, et al. A fast mathematical programming programming procedure for simultaneous fitting of assembly components into cryoem density maps [J]. Bioinformatics, 2010, 26: 261-268.
- [26] JI Z, SHI J. A robust algorithm for feature point matching [J], Computers and Graphics, 2002, 26, 429-436.
- [27] LOWE D G. Distinctive image features from scale-invariant key points [J], International Journal of Computer Vision, 2004, 60(2): 91-110.
- [28] GODDARD T D, HUANG C C, FERRIN T E, Visualizing density maps with ucsf chimera [J]. J Struct Biol, 2007, 157: 281-287.
- [29] ZHANG W, KAUFMANN B, CHIPMAN P R, et al. Membrane curvature in flaviviruses [J]. Journal of Structural Biology, 2013, 183: 86-94.
- [30] ZHANG W, BARBEL K, PAUL R C, et al. Membrane curvature if flaviviruses [J], Journal of Structural Biology, 2013, 183: 86-94.
- [31] LU J, XU G, ZHANG S, et al. An effective sequence-alignment-free superpositioning of pairwise or multiple structures with missing data [J]. Algorithms Mol Biol. 2016, 11(18): 1-10.
- [32] DOUGLAS L T, PHILLIP A S. Optimal simultaneous superpositioning of multiple structures with missing data [J]. Bioinformatics, 2012, 28: 1972-1979.
- [33] EVANS D J. On the representation of orientation space [J]. Mol Phys, 1977, 34(2): 317-325.

(责任编辑 杨开英)