



## 冷冻电镜单颗粒三维重构密度掩模的自动生成方法研究\*

宿锐\*\*、\*\*\* 张叔阳\*\* 王佳伟\*\*\*

(清华大学生命科学学院, 北京 100084)

**摘要** 冷冻电镜单颗粒三维重构技术是用来解析生物大分子三维结构的常用方法. 然而目前在单颗粒三维重构过程中, 溶剂平滑操作还存在一定缺陷: 没有一款主流的单颗粒三维重构程序能够自动寻找掩模 (mask) 三维密度图, 使得三维重构过程难免受到噪音统计学模型计算偏差的干扰. 为解决这一问题, 本研究借鉴X射线晶体学中解析优化相位所广泛采用的溶剂平滑方法, 采用高斯滤波、坎尼边缘检测、最小误差阈值处理等方法处理重构所得三维密度图, 优化溶剂平滑操作, 发展在单颗粒三维重构过程中自动寻找 mask 三维密度图的方法. 运用三维密度图傅里叶壳层相关系数 (fourier shell correlation, FSC) 曲线图、模拟颗粒数据重构角度误差散点图等指标评估此方法的效果. 结果表明, 自动寻找 mask 密度图的方法能够较好地找到涵盖分子结构信号区域的 mask 密度图, 较为明显提高三维重构所得密度图分辨率.

**关键词** 单颗粒, 冷冻电子显微学, 三维重构, 溶剂平滑

**中图分类号** Q-31

**DOI:** 10.16476/j.pibb.2019.0102

结构生物学领域中, 冷冻电子显微镜技术 (cryo electron microscopy, cryoEM), 特别是单颗粒冷冻电子显微镜 (single particle cryoEM) 技术, 近几年发展迅速, 并且因其无需获得蛋白质晶体、所需蛋白质样品量少、三维重构时不存在求解相位问题等相对于X射线晶体学的巨大优势, 解析生物大分子数量和分辨率不断提升, 已经成为结构生物学领域解析生物大分子结构非常重要的方法<sup>[1-3]</sup>.

在单颗粒冷冻电子显微镜的三维密度图重构方法中, 贝叶斯方法拥有巨大优势且为目前采用最为广泛的三维重构参数优化方法, 其主要为通过引入计算噪音的统计学模型, 在三维重构的过程中, 通过多轮迭代, 寻找每个颗粒三维重构最佳的参数<sup>[4-8]</sup>.

RELION<sup>[9]</sup> 为目前单颗粒三维重构所广泛采用的软件之一<sup>[10]</sup>, 其也采用贝叶斯方法进行三维重构, 搜索每个颗粒的最佳三维重构欧拉角. RELION 程序在完成每一迭代轮三维重构之后, 会进行溶剂平滑操作 (solvent flattening), 优化重构所得三维密度图, 增强信号区域所占比重. 目前溶

剂平滑操作主要采用两种方法: a. 程序设定一个固定半径的球体, 球体内部为分子区域, 保留信号, 球外面的溶剂区域, 设为0; b. 人为提供一个固定的 mask 三维密度图, 在每一轮重构的最后, 将 mask 密度图叠加至重构所得三维密度图, 只保留三维密度图对应 mask 内部区域的图像信号, 将其所对应的 mask 外部区域的灰度值抹平为0.

溶剂平滑操作最早应用于X射线晶体学在相位解析和优化过程, 是其密度图修正 (density modification) 方法中非常重要的手段, 溶剂平滑操作通常和直方图匹配方法 (histogram matching)、非晶体学对称平均、直接法、最大熵法等密度图修正方法联合采用来解析和优化相位<sup>[11-15]</sup>, 且其能

\* 国家重点研发计划蛋白质机器与生命过程调控重点专项资助项目(2016YFA0501103).

\*\* 并列第一作者.

\*\*\* 通讯联系人.

宿锐. Tel: 010-62781873, E-mail: sur16@mails.tsinghua.edu.cn

王佳伟. Tel: 010-62782124, E-mail: jwwang@tsinghua.edu.cn

收稿日期: 2019-06-19, 接受日期: 2019-07-23

根据已经搭建的原子模型和电子云密度图自动寻找掩模 (mask) 密度图, 将 mask 密度图外部的区域认定为溶剂区并将其灰度值设定为 0<sup>[16]</sup>.

但是, 目前在 RELION 软件的三维重构的溶剂平滑操作中, 还无法像 X 射线晶体学中的相位优化方法一样, 在每一轮三维重构过程结束之后, 根据当前重构所得三维密度图, 自动寻找 mask 三维密度图, 并消除分子结构信号区域外的噪音, 这就为颗粒噪音模型的计算带来一定的偏差, 对每个颗粒三维重构最优欧拉角的计算搜索造成一定偏差, 最终对重构所得三维密度图分辨率的提升产生不利影响.

本研究借鉴 X 射线晶体学相位优化中溶剂平滑方法和直方图匹配方法, 优化 RELION 程序重构过程中的溶剂平滑操作, 发展在自动精修 (auto-refine) 三维重构过程中自动寻找出能够较好覆盖分子信号区域的 mask 密度图方法, 帮助提升三维重构密度图的分辨率.

## 1 材料与方法

### 1.1 EMDB 发表的三维密度图

选取 EMDB<sup>[17]</sup> 数据库中 4 个已经发表的高分辨率, 即 3.0 Å 以内的三维密度图, EMDB ID 分别为 8 743<sup>[18]</sup>、4 243<sup>[19]</sup>、4 116<sup>[20]</sup>、7 039<sup>[21]</sup>, 对其采用本研究发展的自动寻找 mask 密度图的方法, 自动寻找 mask 三维密度图.

### 1.2 三维重构测试采用的数据集

优化 RELION 程序的自动精修三维重构过程的溶剂平滑操作, 并根据重构所得三维密度图自动寻找 mask 密度图, 主要采用以下原始单颗粒数据进行自动精修三维重构测试.

#### 1.2.1 中低分辨率的数据

采用 EMAN2 软件<sup>[22]</sup> 发布时的 workshop 中所采用的分子伴侣家族 GroEL 蛋白单颗粒原始数据. 经过 EMAN2 程序的自动精修三维重构过程, 所报告的分辨率为 8.4 Å, 由 X 射线晶体学解析结构计算所得傅里叶壳层相关系数 (fourier shell correlation, FSC) 曲线值为 0.5 的分辨率为 8.7 Å.

#### 1.2.2 高分辨率的数据

整体分辨率为 4.1 Å 的耻垢分枝杆菌 (*Mycobacterium smegmatis*) 核糖体 (Smeg ribosome) 整体结构<sup>[23]</sup> 的原始单颗粒冷冻电子显微镜照片数据.

整体分辨率为 3.8 Å 的红藻 (*Cyanidioschyzon*

*merolae*) 附带有 3 根天线分子的光系统 I - 捕光复合物复合体 (PS I - LHCR supercomplex with three Lhcr antennae attached, PS I -3Lhcr) 整体结构<sup>[24]</sup> 的原始单颗粒冷冻电子显微镜照片数据.

### 1.3 使用软件

单颗粒冷冻电子显微镜三维重构软件 RELION1.4<sup>[9]</sup>; 数字图像处理 C++ 库 ITK<sup>[25]</sup>.

### 1.4 实验方法

#### 1.4.1 三维重构密度图自动寻找 mask 密度图的方法

调用 ITK 数字图像处理库, 对重构所得三维密度图采用高斯平滑滤波消减高频噪音, 并计算平滑后三维密度图的灰度分布直方图, 同时结合 minimum error thresholding<sup>[26]</sup>、Otsu thresholding<sup>[27]</sup>、Canny 边缘搜寻<sup>[28]</sup>、watershed 分割<sup>[29]</sup> 等方法, 搜索得到 mask 二值密度图, 并对所得 mask 密度图进行边缘扩展 (edge extension) 及软化 (soft edge), 产生最终的 mask 密度图.

#### 1.4.2 RELION 软件自动精修三维重构过程的改进方法

在自动精修最后一轮前的每一轮迭代过程中溶剂平滑操作中, 对当前轮重构所得三维密度图采用前文 1.4.1 部分中自动寻找 mask 密度图的操作, 将搜索得到的 mask 密度图加到当前自动精修重构所得三维密度图上, 仅保留 mask 密度图边缘之内的三维密度图灰度值, 产生自动精修三维重构过程下一迭代轮的参考 (reference) 三维密度图.

#### 1.4.3 模拟 (simulate) 投影颗粒数据以及三维重构方法

从 PDB 数据库<sup>[30]</sup> 中获取蛋白质分子的三维原子模型, GroEL 和 Smeg ribosome 数据对应的原子模型分别为 1XCK<sup>[31]</sup> 和 5XYI<sup>[32]</sup>, 使用 EMAN2 转换为三维密度图, 并调节功率谱 (power spectrum) 尺度和自动精修所得三维密度图功率谱一致; 将由原子模型产生的三维密度图, 依照对应原始颗粒数据自动精修最终计算所得三维重构概率最大的欧拉角, 采用 RELION 中的投影 (project) 功能, 进行投影, 并按照自动精修计算所得噪音模型加入高斯随机噪音, 得到和原始颗粒数据相同数量的投影模拟颗粒数据; 再对模拟所得颗粒数据采用如前所述的普通的和自动寻找 mask 密度图的自动精修三维重构方法, 进行三维重构操作, 并计算每个颗粒数据三维重构最终搜索重构欧拉角和投影欧拉角之间的误差.

#### 1.4.4 初始模型错误情况下RELION自动精修三维重构过程实验方法

采用非球形样品数据，即红藻捕光复合物原始单颗粒冷冻电子显微镜照片数据，作为此步实验的数据。

采用EMAN2生成直径为240 Å（刚好能将整个复合物全部覆盖）的球形三维密度图（球体内部值为1，球体外部值为0），作为自动精修三维重构过程错误的初始模型，提供给RELION的自动精修程序，并对实验所用数据采用如前所述的普通自动寻找mask密度图的自动精修三维重构方法，进行三维重构操作。

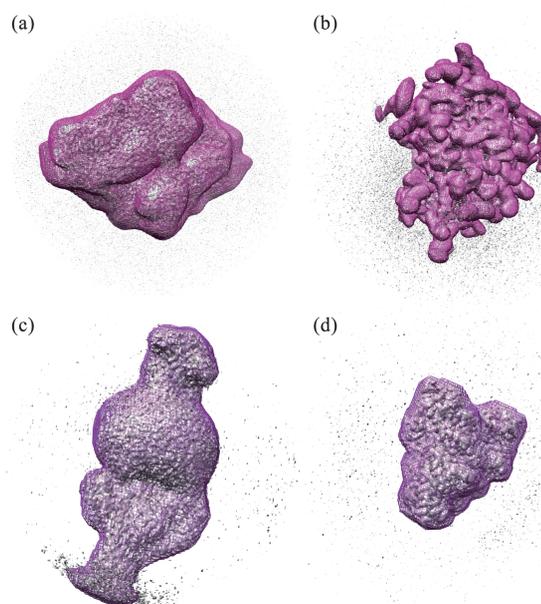
## 2 结果与分析

### 2.1 EMDB发表高分辨率三维密度图寻找mask密度图结果与分析

选取EMDB中已发表的高分辨率三维密度图，以及自动寻找得到的mask三维密度图的结果如图1所示。

图1a为β半乳糖苷酶数据重构所得三维密度图及自动寻找得到的mask密度图，该分子为同源四聚体，形成D2对称性复合体结构，单颗粒重构报告分辨率为2.2 Å，且由重构所得三维密度图可见，复合体整体形状较为规则，形成近似菱形体形状的结构；图1b为人80 S核糖体数据所得三维密度图及自动寻找所得mask密度图，该分子为大小亚基所组成的复合体，为C1对称性结构，单颗粒三维重构报告分辨率为2.9 Å，密度图可见核糖体三维密度图无论大亚基还是小亚基都呈现较无规则的形状；图1c为人激动剂结合的GLP-1受体-Gs复合体（agonist-bound human GLP-1 receptor-Gs complex）数据所得三维密度图及自动寻找所得mask密度图，该分子为5条链组成的异源五聚体，为C1对称性，单颗粒三维重构报告分辨率为3.3 Å，由重构所得密度图可见，该分子复合体呈长棒状形状，两端略扁，中间稍凸；图1d为兔肌肉醛缩酶（rabbit muscle aldolase）数据所得三维密度图及自动寻找所得mask密度图，该分子为同源四聚体组成复合体结构，为D2对称性，单颗粒重构报告分辨率为2.6 Å，图中可见该分子复合体形状较为规则，呈近似于稍微扭曲的方形。

综合图1自动寻找mask密度图结果，对于所选取的EMDB数据库中高分辨率三维密度图，无论是对称性较高的D2对称性的密度图，还是没有



**Fig. 1 Selected EMDB database released high resolution density map and corresponding automatically searched mask map**

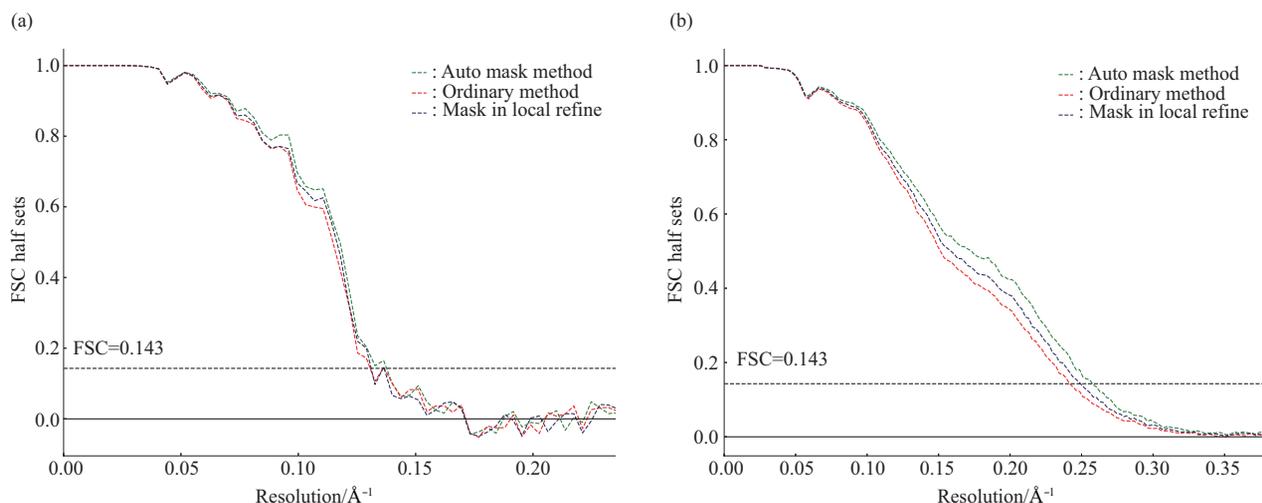
Density map in gray color is selected EMDB released high resolution density map displayed at recommended contour gray level. Density maps shown in pink grid shape are corresponding soft mask map, generated by auto-mask method reported in this paper. (a) EMD-4116 3D density map and soft mask map generated by our auto-mask method; (b) EMD-4243 3D density map and soft mask map generated by our auto-mask method; (c) EMD-7039 3D density map and soft mask map generated by our auto-mask method; (d) EMD-8743 3D density map and soft mask map generated by our auto-mask method.

对称性的C1对称性密度图，自动寻找mask密度图方法，能够较为准确地找到分子结构的边缘区域及核心的分子结构信号区域，将三维密度图中的分子结构信号区域基本上完全涵盖。但是，因为分子某些运动性较强的结构域或者末端柔性较强的区域，其重构三维密度图部分区域信号较为模糊，信噪比较低，自动寻找mask密度图的方法所得到的mask密度图无法将这些区域很好地覆盖。

### 2.2 原始颗粒数据自动精修三维重构结果与分析

颗粒数据经过自动精修三维重构过程后，三维重构报告FSC曲线如图2所示。

GroEL原始单颗粒数据经过自动精修三维重构之后，报告分辨率为7.8 Å，使用自动寻找mask密度图方法进行自动精修三维重构之后，报告分辨率为7.3 Å，在RELION程序自动精修三维重构过程进入局部精修过程（local refine）后，采用固定不



**Fig. 2 Half sets reconstructed density map FSC curves of different three-dimensional reconstruction methods from original particles dataset**

(a) Half sets reconstructed density map FSC curves of GroEL original particles dataset after auto-refine process; (b) Half sets reconstructed density map FSC curves of Smeg ribosome original particles dataset after auto-refine process. FSC curves in green color are from auto-refine reconstruction process using auto-mask method of original particles dataset; FSC curves in red color are from auto-refine reconstruction process using ordinary spherical mask method of original particles dataset; FSC curves in dark blue color are from auto-refine reconstruction process using solvent mask, generated by our auto-discovering method, when RELION auto-refine process entered in local refine.

变的mask密度图三维重构的方法, 报告分辨率为7.8 Å. Smeg ribosome原始单颗粒数据经过自动精修三维重构, 报告的分辨率为4.1 Å, 采用自动寻找mask密度图方法进行自动精修三维重构之后, 报告的分辨率为3.9 Å, 在RELION程序的自动精修三维重构过程进入局部精修过程之后, 采用固定不变mask密度图三维重构的方法, 重构报告分辨率为4.0 Å.

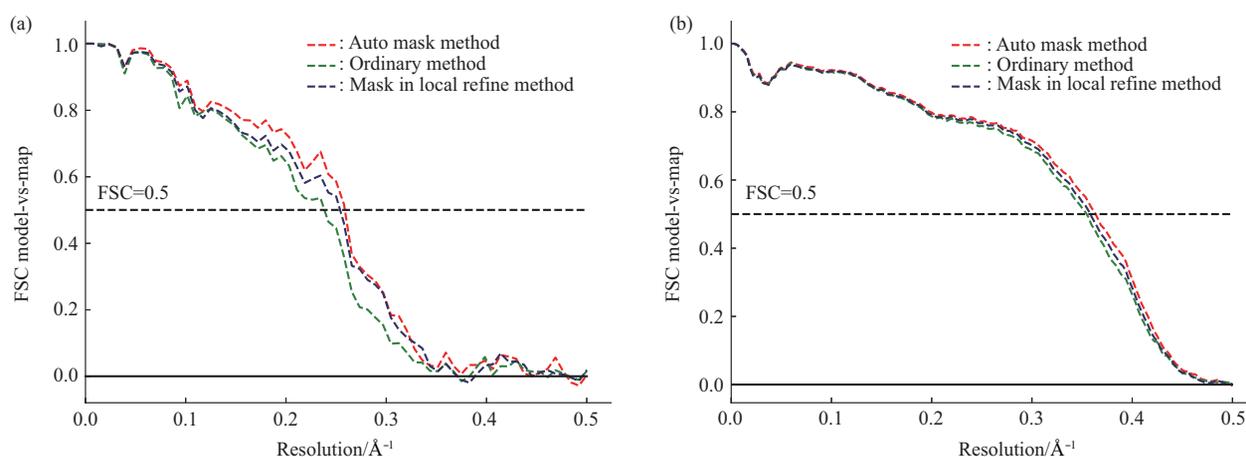
从图2的FSC曲线图中的绿色曲线以及最终的报告分辨率可以看出: 在自动精修三维重构过程中采用自动寻找mask密度图的方法, 相对于在自动精修三维重构过程中不采用自动寻找mask密度图的方法, 能够比较明显地提升三维重构所得密度图在FSC值为0.143时的分辨率, 即报告分辨率; 在RELION自动精修三维重构进入局部精修过程后加入固定的mask三维重构方法, 相对于不采用mask密度图三维重构的方法, 能够在一定程度上提升报告分辨率; 并且, 自动精修重构过程中采用自动寻找mask密度图的方法, 相对于在局部精修加入固定不变mask的方法, 重构结果报告分辨率也略有提升.

从整体来看, 无论是较低分辨率的GroEL单颗粒原始数据, 还是高分辨率的核糖体单颗粒原始

数据, 通过自动精修三维重构过程所得FSC曲线图, 无论是低分辨率区间还是靠近FSC为0.143的较高分辨率区间, 图2中的绿色FSC曲线中对应的点及数值都较为明显地高于红色FSC曲线中的点及数值, 整体较为显著的高于红色FSC曲线; 虽然自动寻找mask密度图方法报告分辨率和局部精修中加入固定的mask密度图方法报告分辨率非常接近, 但从整体上看, 绿色FSC曲线相对于深蓝色FSC曲线略有提升. 可见自动精修三维重构过程中自动寻找mask三维密度图的方法, 能够从低分辨率至高分辨率提高随机分为两份的原始数据(half sets)重构所得三维密度图的相关性.

因此, 从图2所示三维重构不同方法所得不同FSC曲线图以及报告分辨率可见, 在自动精修三维重构过程中采用自动寻找mask密度图方法, 从低分辨率区域至高分辨率区域都能提高三维重构的精度以及重构所得三维密度图的整体分辨率, 并且相对于在局部优化重构过程中加入固定的mask密度图的方法, 整体分辨率也略有提升.

原始颗粒数据经过自动精修三维重构过程后, PDB中对应的原子模型和重构所得三维密度图之间的FSC曲线如图3所示.



**Fig. 3 FSC curves of PDB model vs reconstructed density map by different auto-refine methods**

(a) Atomic model against density map FSC curves after different auto-refine method of original GroEL particles dataset; (b) Atomic model against density map FSC curves after different auto-refine method of original Smeg ribosome particles dataset. FSC curves in green color are from auto-refine reconstruction process using auto-mask method of original particles dataset; FSC curves in red color are from auto-refine reconstruction process using ordinary spherical mask method of original particles dataset; FSC curves in dark blue color are from auto-refine reconstruction process using solvent mask, generated by our auto-discovering method, when RELION auto-refine process entered in local refine.

GroEL原始单颗粒数据经过自动精修三维重构之后,原子模型和三维密度图FSC曲线为0.5时的分辨率为4.3 Å,在自动精修三维重构过程中采用自动寻找mask密度图的方法,所得三维密度图和原子模型FSC曲线为0.5时的分辨率为3.9 Å,对于在自动精修三维重构的局部精修过程中加入固定的mask三维重构方法,所得三维密度图和原子模型FSC曲线为0.5时的分辨率为4.0 Å;Smeg ribosome原始单颗粒数据经过自动精修三维重构之后,三维密度图和原子模型FSC曲线为0.5时的分辨率为2.9 Å,自动精修三维重构过程中采用自动寻找mask密度图的方法,所得三维密度图和原子模型FSC曲线为0.5时的分辨率为2.8 Å,自动精修三维重构中局部精修加入固定的mask密度图三维重构方法,所得三维密度图和原子模型FSC曲线为0.5时的分辨率也为2.8 Å.

从图3所示的FSC曲线图及原子模型和三维密度图FSC值为0.5时的分辨率可以看到,采用自动寻找mask密度图的自动精修三维重构过程,相对于不采用自动寻找mask密度图的自动精修三维重构过程,能够较为显著地提升原子模型和三维密度图FSC为0.5时的分辨率.并且,无论在较低分辨率区域还是较高分辨率区域,红色FSC曲线及对应的数值均较为显著地高于绿色FSC曲线及数值,整体较为明显的高于绿色FSC曲线;虽然红色FSC

曲线及对应的数值和蓝色FSC曲线及对应数值非常接近,而且自动寻找mask方法自动精修三维重构方法相对于局部精修加入固定mask重构方法,原子模型和密度图FSC为0.5时的分辨率是一致的,但从整体上看,红色FSC曲线仍然相对于蓝色FSC曲线略有改善.

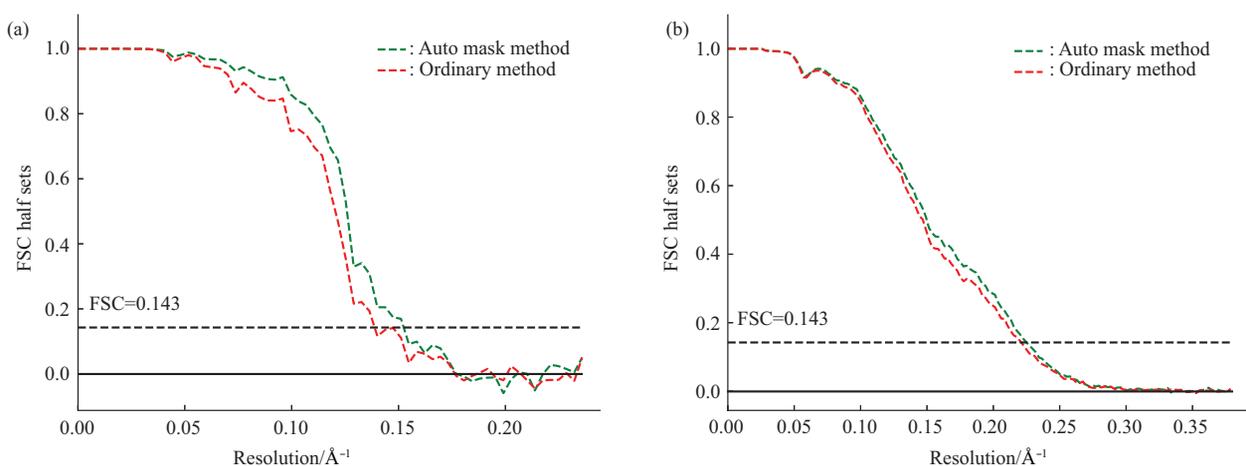
因此,图3所示FSC曲线图结果表明,自动精修三维重构过程中采用自动寻找mask密度图的方法,从低分辨率区域至较高分辨率区域,皆能提升搭建原子模型和重构所得三维密度图间的吻合程度,从而为后期原子模型的搭建提供更好的帮助.综合原始单颗粒数据经过自动精修三维重构过程的结果,可以看出,不论中低分辨率的GroEL单颗粒原始数据还是高分辨率的Smeg ribosome单颗粒原始数据,在自动精修三维重构过程中采用自动寻找mask密度图的方法,相对于普通的自动精修三维重构过程,能够从低分辨率区域至高分辨率区域提升三维重构的精度,重构所得三维密度图的整体分辨率,并且提升所搭建的生物大分子原子模型和三维重构所得密度图的吻合程度,进而能够更好地帮助后续依据三维重构所得密度图的原子模型搭建;而且,虽然自动寻找mask密度图方法相对于在局部精修过程中加入固定不变mask的重构方法,报告分辨率几乎一致,且三维重构密度图和原子模型FSC为0.5时的分辨率也几乎一致,但自动寻找

mask 的自动精修三维重构方法报告 FSC 曲线以及结果三维密度图和原子模型 FSC 曲线均略有改善, 表明自动寻找 mask 密度图三维重构方法相对于在局部精修过程中加入固定的 mask 三维重构方法, 对三维重构的精度、三维重构结果密度图的分辨率以及密度图于原子模型的吻合程度均有一定程度的

提升.

### 2.3 模拟单颗粒数据自动精修三维重构结果与分析

模拟单颗粒数据自动精修三维重构过程后, 三维重构报告 FSC 曲线如图 4 所示.



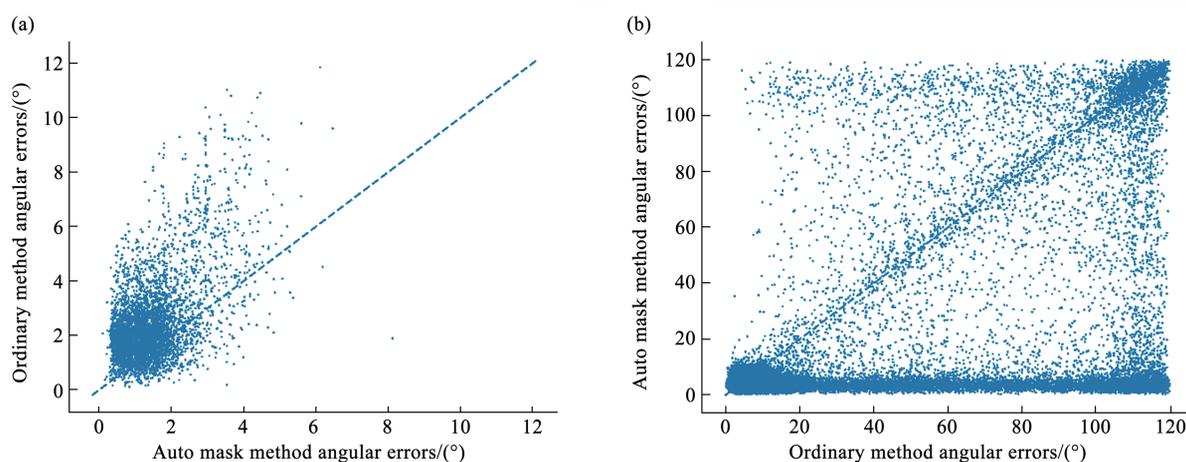
**Fig. 4** FSC curves of half sets reconstructed density maps generated by different auto-refine methods of simulated dataset

(a) Half sets reconstructed density map FSC curves after process of different auto-refine methods of GroEL simulated particles dataset; (b) Half sets reconstructed density map FSC curves after process of different auto-refine methods of Smeg ribosome simulated particles dataset. FSC curves in green color are from auto-refine reconstruction process using auto-mask method of simulated projected particles dataset; FSC curves in red color are from auto-refine reconstruction process using ordinary spherical mask method of simulated projected particles dataset.

GroEL 模拟颗粒数据经过自动精修的三维重构过程, 采用自动寻找 mask 密度图方法所报告分辨率为  $6.6 \text{ \AA}$ , 而不采用自动寻找 mask 密度图方法的报告分辨率为  $7.4 \text{ \AA}$ . Smeg ribosome 模拟颗粒数据通过普通自动精修三维重构过程, 报告分辨率为  $4.6 \text{ \AA}$ , 采用自动寻找 mask 密度图方法自动精修三维重构报告分辨率为  $4.4 \text{ \AA}$ . 并且, 从图 4 中的 FSC 曲线可见, 对于中低分辨率的 GroEL 模拟单颗粒数据和高分辨率的 Smeg ribosome 的模拟单颗粒数据, 自动精修三维重构过程所报告的 FSC 曲线图, 无论在低分辨率区域还是在高分辨率区域, 绿色 FSC 曲线整体均高于红色 FSC 曲线, 即自动寻找 mask 密度图方法能够提高模拟颗粒数据重构的 FSC 曲线, 可见自动寻找 mask 密度图的方法能够较为显著地提高从低分辨率区域至高分辨率区域的重构精度. 因此, 在自动精修三维重构过程中采用自动寻找 mask 密度图的方法, 也能够较明显地提高模拟颗粒数据从低分辨率区域至高分辨率区域的重构精度以及重构所得三维密度图的整体分辨率.

模拟颗粒数据采用不同方法的自动精修三维重构过程计算所得重构欧拉角误差散点图如图 5 所示.

从图 5 的三维重构欧拉角误差散点图可以看出, 自动精修三维重构过程中, 自动寻找 mask 密度图方法计算重构欧拉角误差小于普通自动精修三维重构方法计算角度误差的模拟颗粒数, 在 GroEL 和 Smeg ribosome 的模拟单颗粒数据中, 均超过模拟颗粒总数的 90%. 从图 5a 中可见: 对于中低分辨率的 GroEL 模拟颗粒数据, 自动精修三维重构中采用自动寻找 mask 密度图方法的欧拉角角度误差集中于  $0^\circ$  至  $3^\circ$ ; 对于高分辨率 Smeg ribosome 模拟颗粒数据, 自动精修三维重构中采用自动寻找 mask 密度图方法所计算的重构欧拉角角度误差集中于  $0^\circ$  至  $18^\circ$ , 且因为 Smeg ribosome 模拟颗粒数量较多, 有 57 162 个模拟单颗粒, 加之 Smeg ribosome 分子仅有 C1 的对称性, 可以明显看出普通自动精修三维重构方法计算的欧拉角角度误差无明显集中角度区间, 较均匀地分散于  $0^\circ$  至  $120^\circ$  区



**Fig. 5 Scatter plots of Euler angular errors of computed optimal reconstructed angles from true projected angles of simulated projected particles dataset**

(a) Scatter plot of Euler angular errors after different auto-refine methods of GroEL simulated particles dataset; (b) Scatter plot of Euler angular errors after different auto-refine methods of Smeg ribosome simulated particles dataset. One of the coordinate of each scatter point is Euler angular error from true projection angle by auto-refine process using auto-mask method in solvent flattening operation, the other coordinate of each scatter point is Euler angular error from true projection angle by auto-refine process using ordinary spherical mask in solvent flattening operation.

间内，并且普通方法的角度误差显著高于自动寻找mask密度图方法的角度误差。

因此，不论对于中低分辨率的GroEL模拟单颗粒数据，还是高分辨率的Smeg ribosome模拟单颗粒数据，自动精修三维重构过程中采用自动寻找mask密度图的方法，经过三维重构，最终搜索得到的最大概率三维重构欧拉角和真实投影欧拉角度的误差，显著小于自动精修三维重构过程中不采用自动寻找mask密度图方法计算所得三维欧拉角误差。

通过GroEL和Smeg ribosome模拟单颗粒数据，经过不同方法的自动精修三维重构过程计算所得重构三维密度图分辨率、FSC曲线图和重构欧拉角度误差散点图，可见自动精修三维重构过程中采用自动寻找mask密度图的方法，能够提高原始颗粒数据在自动精修三维重构过程中三维重构欧拉角搜索的准确性，从低分辨率区域至高分辨率区域提高三维重构的精度和重构所得三维密度图的分辨率。

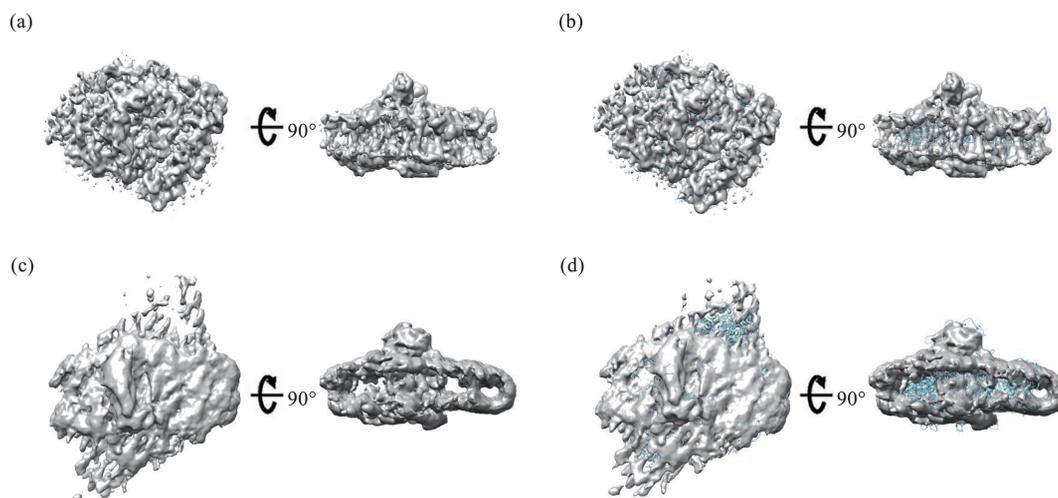
#### 2.4 原始单颗粒数据在初始模型错误情况下自动精修三维重构结果与分析

非球形样品原始单颗粒数据在提供的初始模型错误的情况下，经过自动精修三维重构过程后，所得三维密度图以及原子模型互相匹配如图6所示。

图6的结果表明，在初始模型错误的情况下，不加入mask密度图的自动精修三维重构过程，重构所得三维密度图，整体分辨率较低，仅能大致吻合原子模型，匹配程度较低，且仅提供分子结构的大致轮廓、形状信息，无法提供充足的二级结构信息，并且有较多的重构错误信号；而在自动精修三维重构过程中采用自动寻找mask密度图的方法，重构所得三维密度图，整体分辨率较高，除了能够提供分子结构轮廓和形状信息之外，分子结构的二级结构信号更强，并且能够比较好地与原子模型相吻合，肉眼可见重构的错误信息更少。

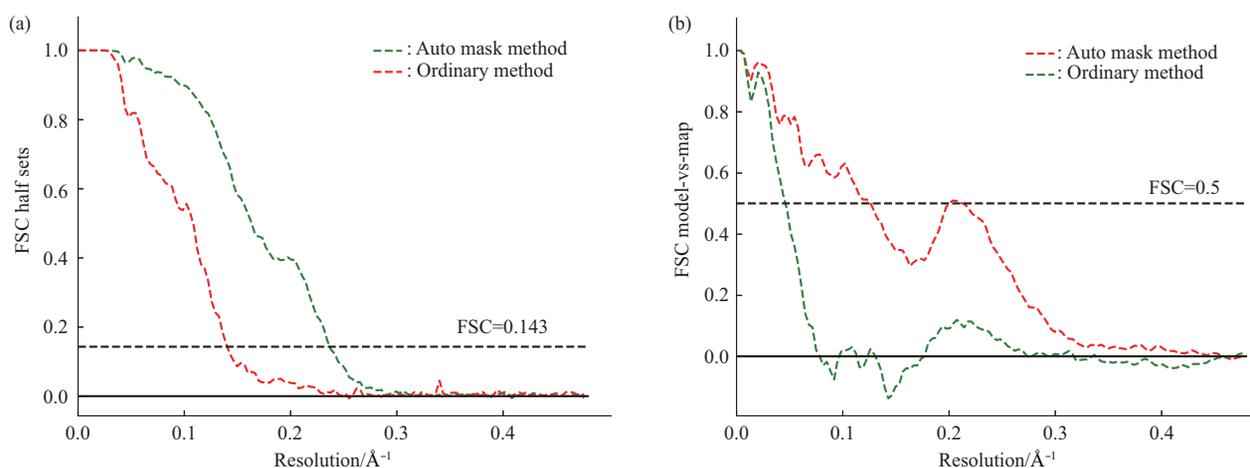
在RELION自动精修过程初始模型错误的情况下，经过自动精修三维重构过程，程序报告FSC曲线图以及PDB中对应的原子模型(5ZGH)和重构所得三维密度图之间的FSC曲线图如图7所示。

捕光复合物原始单颗粒数据在初始模型错误的情况下，经过自动精修三维重构，报告分辨率为7.2 Å，使用自动寻找mask密度图方法进行自动精修三维重构之后，报告分辨率为4.2 Å。而经过不采用自动寻找mask密度图方法自动精修三维重构过程之后，重构所得三维密度图和原子模型FSC曲线为0.5时的分辨率为22.6 Å，而在自动精修三维重构过程中采用自动寻找mask密度图的方法，所得三维密度图和原子模型FSC曲线为0.5时的分辨



**Fig. 6 Reconstructed three-dimensional density maps from RELION auto-refine process when given wrong initial model, by different auto-refine methods**

(a) Reconstructed density map of RELION auto-refine process using auto mask method when initial model is wrong; (b) The same reconstructed density map as in subfigure (a), with fitted in corresponding PDB atomic model; (c) Reconstructed density map of RELION auto-refine process, without using any solvent mask during the reconstruction process, when initial model is wrong; (d) The same reconstructed density map as in subfigure (c), with fitted in corresponding PDB atomic model.



**Fig. 7 Half sets reconstructed density map FSC curves and PDB model vs reconstructed density map FSC curves of different three-dimensional reconstruction methods of RELION auto-refine process from original particles dataset, when initial model is wrong**

(a) Half sets reconstructed density map FSC curves of PSI-3Lhcr original particles dataset after auto-refine process; (b) Atomic model against density map FSC curves after different auto-refine method of original PSI-3Lhcr particles dataset. For subfigure (a), FSC curve in green color is from auto-refine reconstruction process using auto-mask method, FSC curve in red color is from auto-refine reconstruction process using ordinary spherical mask method of original particles dataset; For subfigure (b), FSC curve in red color is from auto-refine reconstruction process using auto-mask method, FSC curve in green color is from auto-refine reconstruction process using ordinary spherical mask method of original particles dataset.

率为 7.9 Å.

从图7所示的FSC曲线图可以看出,在提供的初始模型错误的情况下,采用自动寻找mask密度图方法的自动精修三维重构过程,相对于不采用自动寻找mask密度图的自动精修三维重构过程,能够较为显著地提升低分辨率至高分辨率区域的三维重构的精度以及三维重构所得密度图的分辨率,并且也能够较为显著地提升搭建原子模型和重构所得三维密度图的吻合程度.

综合原始单颗粒数据在提供初始模型错误的情况下,经过RELION程序自动精修三维重构过程的结果,可以看出,在自动精修三维重构过程中采用自动寻找mask密度图的方法,相对于普通的自动精修三维重构过程,也显示出一定的优势:能够较为显著地提升三维重构从低至高分辨率区域的精度;重构所得三维密度图的整体分辨率,显示出更多分子中的二级结构的信息;并且提升所搭建的生物大分子原子模型和重构所得密度图的吻合程度,更好地帮助后续原子模型的搭建.

### 3 讨 论

本研究对单颗粒冷冻电子显微镜技术的自动精修三维重构过程中的溶剂平滑操作进行了优化,发展了在自动精修三维重构当前迭代轮的溶剂平滑操作中自动寻找mask三维密度图的方法.该方法能够在自动精修三维重构过程的每一轮中,比较好地自动寻找到三维重构所得密度图中,涵盖分子结构核心信号区域的mask密度图,并平滑三维密度图中分子结构信号区域之外的噪音区域.

本研究所发展的在自动精修三维重构过程中自动寻找mask三维密度图的方法,相对于在自动精修三维重构过程中不采用自动寻找mask的方法,能够从低分辨率信号区域至高分辨率信号区域,提高颗粒在三维重构过程中噪音模型估计的精度,降低搜索重构欧拉角的误差,提高三维重构的精度,进而提高自动精修过程重构所得三维密度图的分辨率,同时也使经过重构得到的三维密度图和原子模型更加匹配,增加对三维重构之后生物大分子结构的原子模型搭建的帮助,以更好、更精确地解释通过单颗粒冷冻电子显微镜技术所解析的生物大分子的结构及相关生物学功能,即使相对于在局部精修过程中加入固定mask密度图三维重构方法,自动精修重构过程中采用自动寻找mask密度图的方法,对三维重构的精度、自动精修重构过程重构所得三

维密度图的分辨率以及原子模型和重构所得密度图的匹配程度,也均略有改善.

此外,即使提供给RELION自动精修三维重构过程的初始模型是错误的,本研究所发展的自动寻找mask三维密度图的方法,也能够重构得到和PDB中对应的已搭建的原子模型较为吻合的三维密度图,并且提高自动精修三维重构的精度以及重构所得密度图的分辨率,减少重构产生的噪音,以及增加原子模型搭建的帮助.

虽然本研究发展的三维重构过程中,溶剂平滑操作中的自动寻找mask三维密度图的方法能够提高三维重构的分辨率及为模型搭建提供帮助,但是自动寻找mask密度图的方法目前还存在一些需要改进的地方:最主要的是对于分子中一些活动性较强的结构域或片段,三维重构所得密度图对应区域较为模糊不清,信噪比较低,自动寻找得到的mask密度图无法很好地涵盖这些信号不太显著、看上去密度图模糊不清楚的区域,对溶剂平滑操作以及整体的三维重构和分辨率的计算造成偏差;并且,本研究针对溶剂平滑操作发展的自动寻找mask三维密度图目前仅用于自动精修三维重构过程,并未将此方法应用到三维分类(3D classification)过程中.

未来我们将着力提升自动寻找mask密度图对活性较强、信噪比较低区域寻找的准确性;我们还将把自动寻找mask三维密度图的方法应用于三维分类(3D classification)重构过程中,自动寻找经过分类重构所得,处于不同构象状态,即不同类别(class)的三维密度图,在每一轮分类中的不同mask三维密度图,进而为提升三维分类的准确性提供帮助.

对于自动寻找mask三维密度图的研究,未来还将采用深度学习的方法,根据数据库中已经发表的三维密度图,训练能够准确寻找到三维mask密度图的深度学习模型,为在自动精修和三维分类重构过程中,增加单颗粒重构三维欧拉角搜寻准确度以及提高重构所得三维图像分辨率提供更多帮助.

### 参 考 文 献

- [1] Cheng Y. Single-particle cryo-EM—How did it get here and where will it go. *Science*, 2018, **361**(6405): 876-880
- [2] Kühlbrandt W. The resolution revolution. *Science*, 2014, **343**(6178): 1443-1444
- [3] Fernandez-Leiro R, Scheres S H W. Unravelling biological macromolecules with cryo-electron microscopy. *Nature*, 2016,

- 537(7620): 339-346
- [4] Sigworth F J. A maximum-likelihood approach to single-particle image refinement. *Journal of Structural Biology*, 1998, **122**(3): 328-339
- [5] Sigworth F J, Doerschuk P C, Carazo J-M, *et al.* Chapter ten - an introduction to maximum-likelihood methods in Cryo-EM [M]// JENSEN G J. *Methods in Enzymology*. Academic Press. 2010: 263-294
- [6] Scheres S H. A Bayesian view on cryo-EM structure determination. *J Mol Biol*, 2012, **415**(2): 406-418
- [7] Scheres S H W. Classification of structural heterogeneity by maximum-likelihood methods. *Methods in Enzymology*, 2010, **482**: 295-320
- [8] Elmlund H, Elmlund D, Bengio S. PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure*, 2013, **21**(8): 1299-1306
- [9] Scheres S H W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 2012, **180**(3): 519-530
- [10] Patwardhan A. Trends in the electron microscopy data bank (EMDB). *Acta Crystallographica Section D*, 2017, **73**(6): 503-508
- [11] Terwilliger T. Reciprocal-space solvent flattening. *Acta Crystallographica Section D*, 1999, **55**(11): 1863-1871
- [12] Abrahams J P, Leslie A G W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica Section D*, 1996, **52**(1): 30-42
- [13] Giacovazzo C, Siliqi D. Improving direct-methods phases by heavy-atom information and solvent flattening. *Acta Crystallographica Section A*, 1997, **53**(6): 789-798
- [14] Cowtan K D, Main P. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallographica Section D*, 1996, **52**(1): 43-48
- [15] Zhang K Y, Cowtan K, Main P. Combining constraints for electron-density modification. *Methods in Enzymology*. Academic Press. 1997, **277**: 53-64
- [16] Podjarny A D, Rees B, Urzhumtsev A G. Density modification in X-ray crystallography [M]// Jones C, Mulloy B, Sanderson M R. *Crystallographic Methods and Protocols*. Totowa, NJ; Humana Press. 1996: 205-226
- [17] Hudson B P, Berman H M, Westbrook J D, *et al.* EMDatabank unified data resource for 3DEM. *Nucleic Acids Research*, 2015, **44**(D1): D396-D403
- [18] Herzik M A, Wu M, Lander G C. Achieving better-than-3-Å resolution by single-particle cryo-EM at 200 keV. *Nat Methods*, 2017, **14**(11): 1075-1078
- [19] Natchiar S K, Myasnikov A G, Kratzat H, *et al.* Visualization of chemical modifications in the human 80S ribosome structure. *Nature*, 2017, **551**(7681): 472-477
- [20] Kimanius D, Forsberg B O, Scheres S H W, *et al.* Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife*, 2016, **5**: e18722
- [21] Liang Y-L, Khoshouei M, Glukhova A, *et al.* Phase-plate cryo-EM structure of a biased agonist-bound human GLP-1 receptor-Gs complex. *Nature*, 2018, **555**(7694): 121-125
- [22] Tang G, Peng L, Baldwin P R, *et al.* EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 2007, **157**(1): 38-46
- [23] Li Z, Ge X, Zhang Y, *et al.* Cryo-EM structure of Mycobacterium smegmatis ribosome reveals two unidentified ribosomal proteins close to the functional centers. *Protein & Cell*, 2018, **9**(4): 384-388
- [24] Pi X, Tian L, Dai H-E, *et al.* Unique organization of photosystem I - light-harvesting supercomplex revealed by cryo-EM from a red alga. *Proc Natl Acad Sci USA*, 2018, **115**(17): 4423-4428
- [25] Ackerman M J, Yoo T S. The Visible human data sets (VHD) and insight toolkit (ITk): experiments in open source software. *AMIA Annual Symposium Proceedings AMIA Symposium*, 2003, **2003**: 773
- [26] Kittler J, Illingworth J. Minimum error thresholding. *Pattern Recognition*, 1986, **19**(1): 41-47
- [27] Otsu N. A Threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, **9**(1): 62-66
- [28] Canny J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, **PAMI-8**(6): 679-698
- [29] Cousty J, Bertrand G, Najman L, *et al.* Watershed cuts: minimum spanning forests and the drop of water principle. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(8): 1362-1374
- [30] Berman H M, Westbrook J, Feng Z, *et al.* The protein data bank. *Nucleic Acids Research*, 2000, **28**(1): 235-242
- [31] Bartolucci C, Lamba D, Grazulis S, *et al.* Crystal structure of wild-type chaperonin GroEL. *Journal of Molecular Biology*, 2005, **354**(4): 940-951
- [32] Li Z, Guo Q, Zheng L, *et al.* Cryo-EM structures of the 80S ribosomes from human parasites *Trichomonas vaginalis* and *Toxoplasma gondii*. *Cell Research*, 2017, **27**(10): 1275-1288

## Automatically Generating Mask for Mapping Out The Density Map in cryoEM Single-particle Reconstruction\*

SU Rui<sup>\*\*\*</sup>, ZHANG Shu-Yang<sup>\*\*</sup>, WANG Jia-Wei<sup>\*\*\*</sup>

(School for Life Sciences, Tsinghua University, Beijing 100084, China)

**Abstract** CryoEM single particle reconstruction technique is a widely used method for solving macromolecular structures in structural biology. However, there are still some shortcomings in the current processes of single particle three-dimensional reconstruction. None of the current single particle reconstruction software has implemented the function of automatically finding soft mask for binary mapping out the reconstructed map, which can import bias in the computation of statistical model of noise, influencing effectiveness of three-dimensional reconstruction. To solve this problem and to improve the solvent flattening operation in single particle reconstruction, here, we report a new method for automatically finding soft mask for mapping out the reconstructed map in single particle reconstruction, by using Gaussian filtering, Canny edge detection, minimum error thresholding *etc.*, for processing reconstructed map. This new method is derived from the solvent flattening method, which is widely used in X-ray crystallography for solving and improving macromolecular phases. In the experiment, we implemented FSC between reconstructed maps, Euler angular error scatter plot *etc.* methods to assess effect of auto-mask method. It was shown that mask map generated by auto-mask method can soundly cover molecular signal region of reconstructed map, and our new auto-mask method in three-dimensional reconstruction can help to improve resolution of three-dimensional reconstruction.

**Key words** single particle, cryo electron microscopy, three-dimensional reconstruction, solvent flattening

**DOI:** 10.16476/j.pibb.2019.0102

---

\* This work was supported by a grant from Ministry of Science and Technology of China (2016YFA0501103).

\*\* These authors contributed equally to this work.

\*\*\* Corresponding author.

SU Rui. Tel: 86-10-62781873, E-mail: sur16@mails.tsinghua.edu.cn

WANG Jia-Wei. Tel: 86-10-62782124, E-mail: jwwang@tsinghua.edu.cn

Received: June 19, 2019 Accepted: July 23, 2019