Project 2
Computational Biology (2016 Spring)
(Due May 8, 2016)

Overview: In this project, you are asked to implement an algorithm to reconstruct 3D structure of the molecule from its projection images (particles) in cryo-EM and an initial 3D structure model. For final submission, you need to write a report in *English* to describe your algorithm, evaluate the performance of your algorithm through numerical results, and make a discussion.

Policy: You are allowed to use any reference from papers, books or materials from Internet. If so, please cite these references in your report. If you have discussed with other people (e.g., professors, friends or classmates), please mention these discussions in the acknowledgement section of your report. In a team-based report (with ≤ 3 persons), please describe clearly the contribution of each team member.

Questions: If you have any question about this homework, please contact TA (Wan Fangping, wfp15@mails.tsinghua.edu.cn).

Academic Honor Code: This homework must be done independently. We have zero tolerance on cheating and plagiarism.

1 Problem Description

In this problem, you are given an initial 3D structure model and a number of 2D projection images (particles) of the target molecule. Your goal is to develop an algorithm to reconstruct the 3D structure of this molecule based on these input data. In the end you should compare your result to a reference 3D structure reconstructed by cryo-EM experts (via other existing approaches).

2 Getting started

2.1 Data

The data can be downloaded from `http://pan.baidu.com/s/1eSAXBQa`.

There are two datasets. The protein of one dataset is called TRPV1. The micrographs are stored in .jpg files, while the coordinates of particles in micrographs are stored in the corresponding .star files, where the first two columns are center coordinates of particles. The initial 3D structure is trpv1init.mrc, and the reference 3D structure is trpv1emd5778.mrc. The protein of the other dataset is called β-galactosidase. The micrographs are stored in .mrc files. The coordinates of particles in micrographs are stored in .star files, where the 11th and 12th columns are center coordinates of particles, and the 14th column is the filename of its corresponding micrograph. The initial 3D structure is stored in 3i3elp50A.mrc file. The reference 3D structure is betafinal.mrc. (for more details, see `https://en.wikipedia.org/wiki/MRC\_(file\_format)` and `http://bio3d.colorado.edu/imod/doc/mrc_format.txt`).

2.2 Helpful materials

[1] is a good reference to know about cryo-EM. [2] and [3] are two examples of 3D reconstruction algorithms[4]. We encourage you to design a new algorithm or improve the previous algorithms instead of using the existing ones. The dataset used in this project are from [6].

2.3 Implementation

Implement your algorithm in any programming language that you are familiar with, such as Java, C/C++, Matlab, Python, etc. You are allowed to call any other available public package in your program. If so, please include the library in your final submission.

2.4 Measure your algorithm

When you test your algorithm on the test dataset, you are required to design a rule to compare your results with the reference 3D structure and validate the performance of your method.

3 Requirement of Report

In your final report, you should address the following points:

  (1) Details of your algorithm, such as overview, pseudo-code (or flow chart), etc.
  (2) Performance evaluation of your algorithm.
  (3) Discussion about strength and limitation of your algorithm.
  (4) Give the final 3D structure of the molecule.

4 Final Submission

For final submission, you need to provide: (1) report; (2) source code and binary executable file of your program, and a short readme file that describes how to compile and run your program.

References

[1] Yifan Cheng, Nikolaus Grigorieff, Pawel A. Penczek, and Thomas Walz. A primer to single-particle cryo-electron microscopy. *Cell*, 161(3):438 – 449, 2015.

[2] Sjors H.W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519 – 530, 2012.

[3] Marcus A. Brubaker, Ali Punjani, David J. Fleet Building Proteins in a Day: Efficient 3D Molecular Reconstruction. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[4] Sjors H.W. Scheres. A Bayesian View on Cryo-EM Structure Determination. *Journal of Molecular Biology*, 415(2):406 – 418, 2012.

[5] Maofu Liao, Erhu Cao, David Julius, and Yifan Cheng. Structure of the trpv1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478):107 12, December 2013.

[6] Bartesaghi, Alberto and Matthies, Doreen and Banerjee, Soojay and Merk, Alan and Subramaniam, Sriram. Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences*, 111(3):11709-11714, 2014.