NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCES
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

GRADUATE PROGRAM
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"

Postgraduate Student: Kalamara Angeliki

Thesis Title: Detecting Molecular Boundaries in Density Maps

Abstract

Macromolecular complexes are the building blocks and workhorses of biological organisms. Gaining insight into the structure of such complexes is critical not only to better understand their function but also to comprehend why they , sometimes, fail to function properly. To determine the structure of these complexes electron density maps are used. They derive either from X-Ray Crystallography or Cryo-Electron Microscopy (Cryo-EM) experiments and they are described as a 3-dimensional grid of voxels each containing an electron density value. The advantage of Cryo-EM technique is that we can study large macromolecular assemblies preserved in their native state. However, the main shortcoming of this technique is that the resulting maps are of lower resolution and very fuzzy. Thus, segmentation methods are commonly used to extract structural information from low-resolution density maps.

In this present study we use CGAL, written in C++, and develop a novel method for segmenting density maps by approximating their surface with a triangular mesh. More specifically, we calculate level set surfaces (isosurfaces) from density maps for various level set values (isovalues) and combine them to identify the molecular boundaries within a density map. In other words we define a partition of the map compliant with the number of subunits; if the assembly involves *n* subunits the result of the analysis should consist of *n* simply connected regions. The algorithm is applied to both simulated and experimental Cryo-Electron Microscopy (Cryo-EM) maps. Simulated maps were generated by blurring an atomic structure with a Gaussian kernel at various resolution values (*SITUS* package) while the latter were provided by the EMDBdata Bank. The main advantages of our method are its speed - close to linear time - and the characteristic of shape preservation. In the proposed algorithm, a user chooses a positive integer number which equally splits the $[density_{min}, density_{max}]$ range of values and a plot is produced, revealing the evolution of the connected components as a function of the isovalues. From this we obtain multiple/single valid

isovalue ranges. Boost Graph Library is used for the purpose of returning as output the total number of connected components; in this case the triangulated surface mesh is treated like a graph. Aiming at reliable results, the user may proceed with a refinement process in sub-intervals where noise is observed. Finally, a filtration process of multiple ranges is subsequently followed to eliminate poor volumes or ill-posed components.

The reported results are validated exploiting information from the connected components' and complex structure's volume via the DIOPHANTINE recursive algorithm. To approximate the volume of a component of the map we build a tetrahedral simplicial mesh whereas the structure's volume is calculated by its sequence. Validation results confirm our initial findings.

Fields: Structural Bioinformatics; Computational Geometry
Keywords: Cryo-Electron Microscopy; Image Segmentation; CGAL/C++; 3D Mesh; Density Map