High-throughput single-particle analysis in 3D electron microscopy

Renata Gimenez and Tiina Leiponen

The research group of Carlos Oscar Sánchez Sorzano from the National Center of Biotechnology in Madrid aims at optimizing the steps in the 3D analysis of single-particles, with optimized XMIPP, high throughput application, in further reduced computing time. This will increase the productivity of structural biologists and optimize the use of available.

The research group of Slavica Jonic from IMPMC, CNRS-University Pierre and Marie Curie, in Paris, aims at developing new image analysis methods for visualizing the conformational changes undertaken by macromolecular complexes. The collaborative efforts of these two groups open a new door towards high throughput 4D Electron Microscopy.

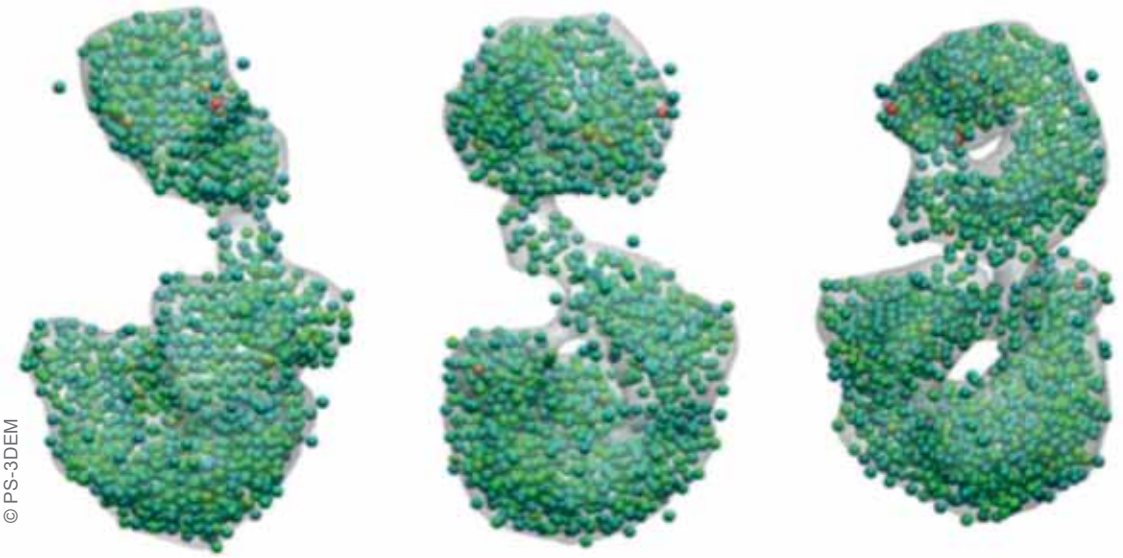
Structural biology aims at the elucidation of the three-dimensional (3D) structure of biological macromolecular complexes in order to fully understand its function in the live cell. An approach to collect such structural information is imaging tens of thousands of copies of the same complex in an electron microscope. Images have to be intensively processed to obtain the 3D structure at high resolution that is essential for studying the relation structure-function. This field is usually called 3D Electron Microscopy (3DEM). Currently, solving a structure at high (subnanometer) resolution typically takes 50.000 CPU-hours. The time would be even longer without the parallelization of the most time consuming steps.

Increasing productivity and optimizing

Over the last 20 years, the National Centre of Biotechnology (CSIC) in Spain has developed an open-source software package to address this 3D reconstruction problem. The group of researchers in the PS-3DEM project group aim at optimizing the steps involved in the 3D analysis of single-particles. "Optimized XMIPP results a high throughput application and further reduced computing time. This will increase the productivity of structural biologists and optimize the use of available", tells professor Sánchez from the National Centre of Biotechnology, Spain.

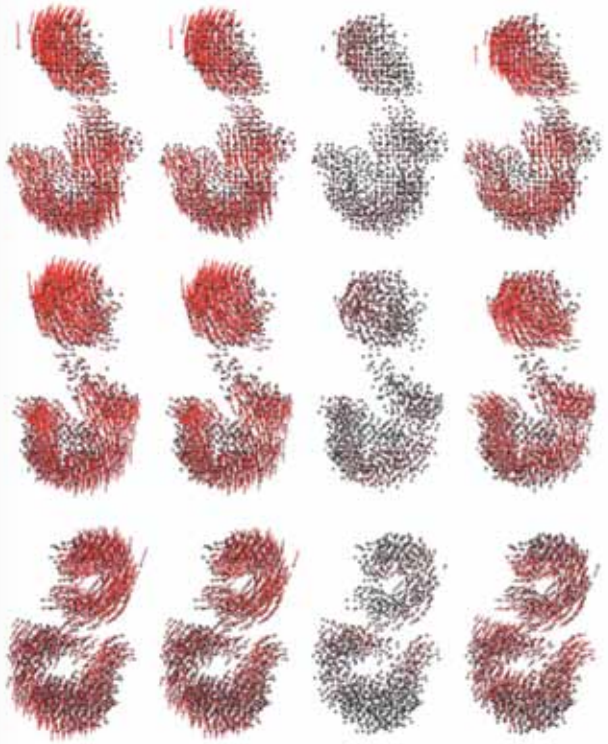
Macromolecular complexes perform their biological functions in the cell thanks to their structural flexibility and their biochemical properties. The transmission electron microscope provides two-dimensional projections of frozen isolated complexes and allows simultaneous imaging of thousands of copies of the same complex but in different conformations. "The problem addressed in this work aims at identifying the deformation and projection direction of each imaged complex. By doing so, we are able to reconstruct the deformation paths used by complexes which allows filming molecules in action using transmission electron microscopy", explain Sánchez and Jonic.

A hybrid approach that integrates Normal Mode Analysis (NMA) into a rigid-body, projection matching method has been developed to study conformational changes of macromolecular assemblies. The novelty lies in allowing the determination of a whole set of intermediate structural conformations for studying continuous-type conformational variability. To isolate gradual structural transformations, this approach uses NMA of high-resolution or low-resolution



© PS-3DEM

Figure 1a. The DNA Polymerase α is an enzyme responsible for the initiation of DNA replication in eukaryotic cells. In this figure we show three different views of a static three-dimensional structure represented by a coarse-grained model obtained from an EM structure .



© PS-3DEM

Figure 1b. Dynamics of different regions of the DNA Polymerase α with one trajectory per view (same three views as in the previous figure). Arrows indicate the direction and amplitude of the displacement. Movies have also been produced showing the dynamics.

reference structure within an iterative process involving atomic or pseudoatomic reference-structure deformation, volume building, and volume-to-image alignment.

"The new method can be generally used to study large-scale structural rearrangements (slow collective motions of atoms) of any biomolecule that can be imaged by transmission electron microscopy (TEM)" says professor Sánchez.

Macromolecular flexibility

Single-Particle Analysis (SPA) is a widely used to collect structural information by imaging tens of thousands of copies of the same complex in an electron microscope. Studying the specimen using SPA, structural variability of the sample is very often the main factor limiting the resolution. Macromolecules with structural flexibility may adopt different conformations to perform different molecular functions. However, reconstruction algorithms assume that images come from a homogeneous population (identical copies of a macromolecule). Several methods have been developed to study structural heterogeneity. These methods either separate projections in homogeneous classes using multi-reference classification or multivariate statistical classification, compute elastic geometric transformations between (pseudo)atomic coordinates of two provided 3D structural conformations, explicitly estimate the structural variability present in projections or identify regions that are the most likely to vary from one image to another.

Towards 4D Electron Microscopy

"We propose to follow a different strategy to analyse heterogeneity. We first compute a pseudo-atomic representation of the structure being analysed. We, then, study the most important (regarding low-frequency movements and collectiveness of the movements) modes of vibration of the structure through normal mode analysis (NMA). NMA defines a movement space, such that the macromolecule is supposed to be deformed in the subspace spanned by the NMA basis. For each experimental image we determine its projection orientation as well as the deformation in the NMA subspace. Finally, we reduce the dimensionality of the deformation coordinates by Principal Component Analysis (PCA). PCA defines a new coordinate basis in which the deformation paths of a particular macromolecule can be identified", explains Dr Jonic (see Figs. 1a and 1b).

The determination of the projection orientation and the corresponding macromolecule deformation is a formidable task that, in some cases such as complex conformational changes and large macromolecules, may require 1 day per image on a single CPU. " In a typical 3DEM study we use around 10,000 projection images, meaning that a single CPU might require around 30 years to fully analyse the flexibility of the complex" clarifies Sorzano. "Supercomputing infrastructures are needed to reduce the computation time. The Barcelona Supercomputing Center (BSC) participated in the analysis of the software source code performing these studies. We identified the code inefficiencies and reduced the computation time by a factor of 2. We also parallelized the code in order to benefit from the large computing infrastructures provided by supercomputing centres like the BSC", says Sanchez Sorzano. "The optimized code was successfully ported to two more European supercomputing centres (IDRIS/Orsay and CINES/Montpellier in France) where the results showed in this article were produced (Figs. 1a and 1b)" concludes Jonic.

FURTHER INFORMATION

http://www.deisa.eu/science/deci/projects2010-2011/PS-3DEM

http://xmipp.cnb.csic.es

Compute resources used at BSC.
C.O.S. Sorzano, J.M. de la Rosa Trevín (coss@cnb.csic.es).
National Center of Biotechnology (CSIC) – Madrid – Spain.
S. Jonic (Slavica.Jonic@impmc.jussieu.fr).
Institut de Minéralogie et de Physique des Milieux Condensés (UMR 7590 CNRS-Université Pierre et Marie Curie-IRD) – Paris – France.

- **Centro Superior de Investigaciones Científicas (CSIC), National Centre of Biotechnology, Spain**
- **Institut de Minéralogie et de Physique des Milieux Condensés (UMR 7590 CNRS-Université Pierre et Marie Curie-IRD), Paris, France**