

Computational Structural Biology Research Unit

1. Unit members

Florence Tama (Unit Leader)

Atsushi Tokuhisa (Research Scientist)

Sachiko Kikumoto (Assistant)

2. Research Activities

Biological molecular complexes of such as proteins and RNAs are of great interest in the area of molecular biology as they are involved in cell replication, gene transcription, protein synthesis, regulation of cellular transport and other core biological functions. Those systems undergo large conformational transitions to achieve functional processes. Therefore characterization of structures of these macromolecular complexes is crucial to understand their functional mechanisms, and play an important role in the development of new drugs to treat human disease.

Experimentally, X-ray crystallography has been the primary tool to study protein conformations, providing high-resolution structures. Even though it is at low resolution, cryo electron microscopy (EM) has been providing critical information on structure and dynamics of large biological molecules. More recently, large efforts, such as in Riken/Spring 8 have focused on developing intense X-ray free-electron laser (XFEL) light sources, which offer a new possibility to image single biological macromolecules. Since crystallization is not necessary for such a protein structure analysis, it would be possible to investigate the structure of macromolecular complexes and proteins under various physiological conditions or to observe elementary steps of a biochemical function. However, the technology is still at early stage and it cannot yet achieve atomic level resolution such as obtained by X-ray crystallography.

Computationally, methods have been developed to predict structures from low-resolution data such as cryo-EM either using rigid body fitting or flexible deformations of known atomic structures. In addition, even when structures of the molecules are unknown, atomic models can be predicted using homology modeling and ab initio predictions. While, ab initio prediction still remains difficult for large proteins, success in predicting small proteins have been observed. Finally, algorithms to analyze protein/proteins interactions also have shown success in predicting proteins complexes.

Our research focuses on the development of computational tools to study biological systems,

more specifically, to help in their 3D structural determination using various experimental techniques and to analyze their potential interactions with small molecules in order to design new drugs.

The ultimate line of our interdisciplinary research is to bring experimental data as obtained from X-ray, cryo-EM and XFEL with development and applications of computational tools through the K computer to acquire knowledge on the structure of a physiologically important protein complexes that are unattainable with existing experimental techniques, and to contribute to development of drug design and medical treatment in collaboration with pharmaceutical companies.

3. Research Results and Achievements

3.1. Dynamical information embedded into cryo-EM 2D raw data

Cryo-EM Single-Particle Analysis (SPA) is a method to study the structure and dynamics of macromolecular assemblies. Three-dimensional (3D) structures are computed from a large number of two-dimensional (2D) images collected by transmission electron microscopy. SPA has shown to be promising in capturing heterogeneous conformations of the same macromolecular complex. To study dynamics, a logical solution is thus to acquire EM images of a heterogeneous population of conformations of the same macromolecular complex and then, to use image analysis methods to separate the mixed particles into classes (2 or 3) with similar conformations of particles and, finally, to analyze the 3D structures computed from each of the image classes. This approach is suited if the system is expected to have a few distinct conformations (for example, with and without ligand). However, if the conformational transition is continuous, such approaches cannot capture the intermediate conformations. We developed a new hybrid approach that utilizes molecular mechanics algorithms to simulate conformational dynamics and use the resulting conformations to extract the dynamical information from cryo-EM images. More specifically, a conformational ensemble is generated by molecular mechanics approaches such as normal mode analysis. For each EM image, a conformation that matches most closely is

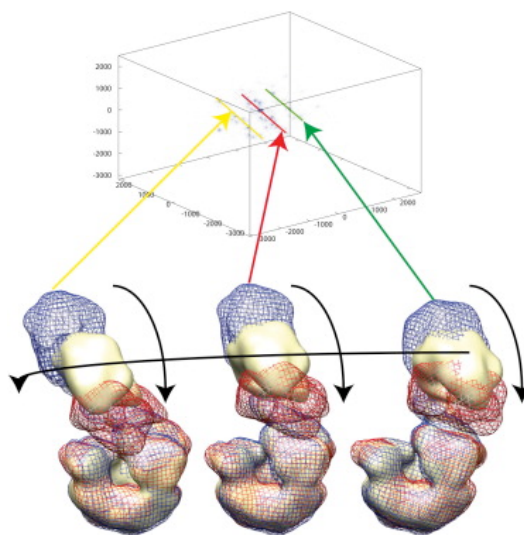


Figure 1. Computed deformation amplitudes projected on to the three most important principal with reconstructed volumetric form of the pseudoatomic structure displaced along the lines yellow, red and green line.

identified (see Figure 1). Through clustering of image/conformation pairs, the 3D conformational variations embedded in EM images are revealed. The performance of the method was shown using synthetic data of two different molecular complexes (*E. coli* 70S ribosome and Tomato Bushy Stunt Virus (TBSV)) and using experimental cryo-EM data of the TBSV.

3.2. Annotating cryo-EM low resolution structure with high-resolution X-ray data

Cryo-EM experiments produces low-to-medium resolution structures (usually in the range between 20 and 4 Å) but allows studying large (diameter larger than 10 nm and molecular weight sometimes of several mega-Daltons) and flexible macromolecular complexes inaccessible to X-ray and NMR techniques. Because the data is at low-resolution, in order to obtain higher-resolution information, known X-Ray structures are often combined with cryo-EM data which in some cases requires X-ray structure deformation (flexible fitting). Flexible fitting of X-ray structure into cryo-EM maps requires first a rigid body fitting of the X-ray structure followed by deformation to fit the density map. In collaboration with Dr. Sugita's team, we have been implementing the algorithms in GENESIS to perform flexible fitting of atomic structures using either full atom or coarse grained model description of the molecule into such low-resolution data. Using generalized ensemble algorithms embedded in GENESIS, such as replica exchange (REMD), the accuracy and efficiency of fittings can be enhanced.

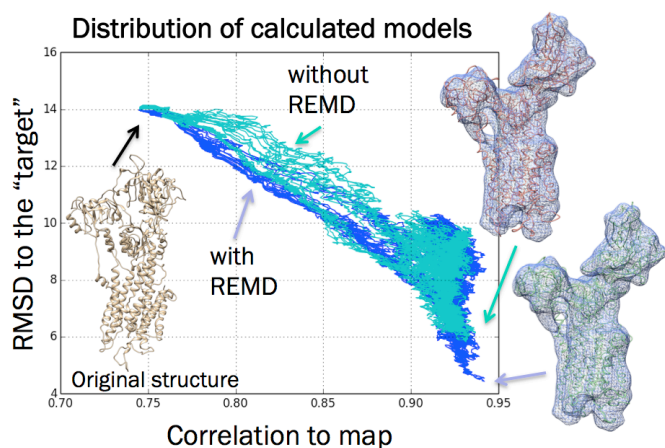


Figure 2. A result of flexible fitting using GENESIS. The original structure is deformed using molecular dynamics simulation to fit the low-resolution data. Fitting with REMD provides structure with better agreement (lower RMSD) with the target structure.

3.3. Computational tools to analyze XFEL experimental data

Since XFEL measurement can be at a single molecular level, and not averaged, its data contains the information regarding the dynamics and conformational variations. In this regard, it is similar to cryo-EM single particle analysis, yet XFEL has a potential to provide more detailed information; it has higher transmissibility and no aberration issue and high resolution is attainable with "diffraction before destruction" effect. We started to explore new methodologies and develop computational tools to analyze XFEL data to extract dynamical

information. This work is in collaboration with Dr. Jonic, CNRS, who has been developing algorithms to analyze 2D images from cryo-EM experiments. The goal is to examine how well XFEL experiments can reveal conformational variations in samples. Simulated diffraction data were generated for incident beams with random orientation for several biological systems adopting several conformations (elongation factor, ribosome, cowpea chlorotic mosaic virus). Similarities in conformation were analyzed for both 2D real image data recovered from

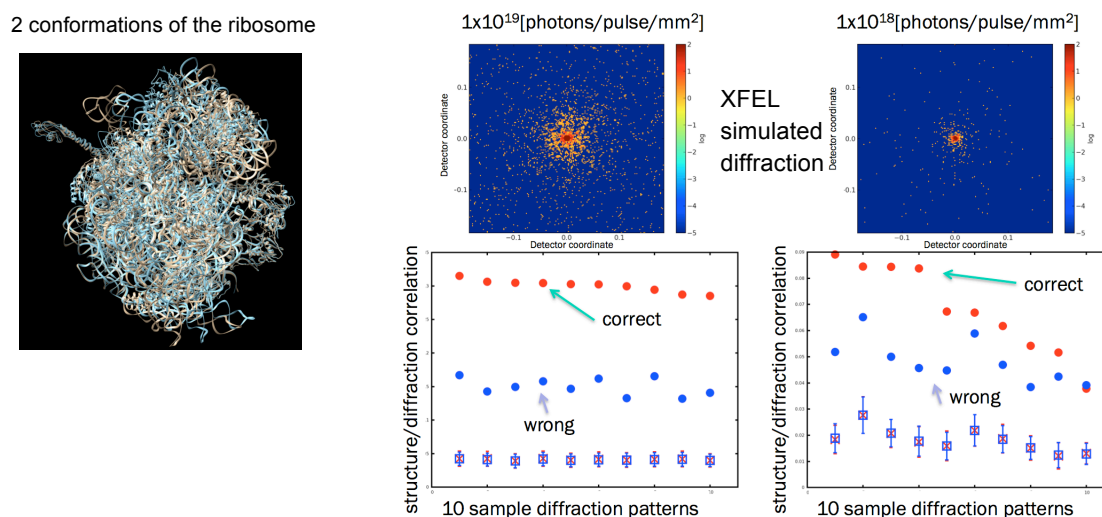


Figure 3. For each 10 sample, the correct conformation shows a higher score (red point) than the wrong conformation/orientation.

diffraction pattern and from diffraction pattern directly. We found that XFEL can differentiate conformation with a 2 Å resolution but that the laser intensity affects the accuracy of detections.

3.4. Structure function of proteins, study of small heat shock proteins

The small heat shock proteins (sHSPs) are a virtually ubiquitous and diverse group of molecular chaperones that can bind and protect unfolding proteins from irreversible aggregation. It has been suggested that intrinsic disorder of the N-terminal arm (NTA) of sHSPs is important for substrate recognition. To investigate conformations of the NTA that could recognize substrates, we performed replica exchange molecular dynamics simulations. Behavior at normal and stress temperatures of the dimeric building blocks of dodecameric HSPs from wheat (Ta16.9) and pea (Ps18.1) were compared because they display high sequence similarity, but Ps18.1 is more efficient in binding specific substrates. In our simulations, the NTAs of the dimer are highly flexible and dynamic, however, rather than exhibiting highly extended conformations they retain considerable α -helical character and contacts with the conserved α -crystallin domain (ACD). Network analysis and clustering methods defined two major NTA conformational forms, designated either “open” or “closed” based on the relative position of the two NTAs in a dimer (see Figure 4).

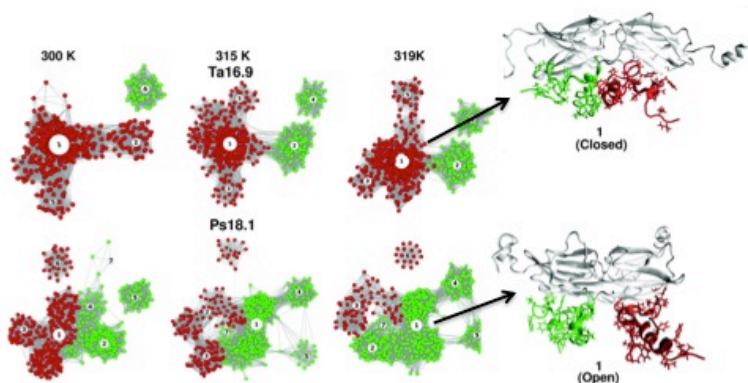


Figure 4. Network representation of open (green) and closed (red) clusters defined by conformations of Ta16.9 and Ps18.1 dimers at 300, 315 and 319 K. Representative structure of the prevalent open conformation of Ps18.1 and close conformation of Ta16.9 are also shown.

The equilibrium constant for a closed to open transition displays significant difference between Ta16.9 and Ps18.1 with the latter one showing more open forms at elevated temperature. The Ps18.1 NTAs are predominantly open and have more hydrophobic solvent accessible surface than the Ta16.9 NTAs, correlated with more effective chaperone activity. NTA hydrophobic patches are comparable in size to the area buried in many protein-protein interactions, which would enable sHSPs to bind early unfolding intermediates. Reduced dimeric interactions of the Ps18.1 NTAs and with the ACD contribute to the differences in dynamics and hydrophobic surface area between the two sHSPs. These data support a major role for the conformational equilibrium of the NTA in substrate binding and indicate features of the NTA that contribute to sHSP chaperone efficiency.

4. Schedule and Future Plan

We plan to continue to develop tools to analyze XFEL data in order to get structural information of biological molecules. In particular, we aim to develop computational algorithms that would provide the shape of the biological molecules. Such algorithms will require the use of simplified representation of the biological molecules as well as a multi-step optimization procedure to build a shape that would be in agreement with the diffraction pattern obtained from XFEL experiments.

In addition, we aim to characterize dynamical information embedded within raw low-resolution data obtained from cryo-EM. This research is an on-going collaboration with Dr. Slavica Jonic (CNRS, Paris). Macromolecular structure determination by cryo-electron microscopy (EM) and single particle analysis are based on the assumption that imaged molecules have identical structure. With the increased size of processed datasets it becomes apparent that many complexes coexist in a mixture of conformational states or contain flexible regions. Algorithms have been developed to yield estimates of voxel-by-voxel variance of a structure reconstructed from the set of its projections. Such variances will be compared from enhanced sampling

molecular dynamics simulations of biological molecules. Such type of approach could later on be extended to data from XFEL experiments as well.

As our research focuses on developing computational tools to analyze low-resolution experimental data, we intend to establish collaborations with experimental groups in Japan and abroad in order to study structure, function and dynamics of biological molecules.

On the longer term we plan to establish methodology to build structure from low-resolution structural data without a priori knowledge of the overall structure of the molecular complexes, since potential targets of the structural analysis by low resolution experimental techniques are multiprotein/RNA complexes. Although it is difficult to acquire the crystal structure of the whole complex, the atomic structure of each component protein and RNA may be known. Moreover, for small proteins, even when there is no structure, their structures can be predicted in relatively high precision using homology modeling. Therefore, if such structures are correctly combined into a model of the complex that fits the three-dimensional electron density map obtained from low-resolution experimental techniques, the atomic structure of a complex could be obtained. A computational framework using multiscale simulations, which would combine the representations at different resolution from all the atoms to coarse-grained representations as well as protein-protein docking algorithms will be developed for such purpose.

5. Publication, Presentation and Deliverables

(1) Journal Papers

- [1] R. Nogales-Cadenas, S. Jonic, F. Tama, A. Arteni, D. Tabas-Madrid, M. Vazquez, A. Pascual-Montano, CO. Sorzano, “3DEM Loupe analysis of macromolecular dynamics using structures from electron microscopy”, *Nucl. Acid. Res.*, (2013) 41:W363-7.
- [2] D. Lyumkis, H. Talley, A. Stewart, S. Shah, CK. Park, F. Tama, CS Potter, B. Carragher, NC. Horton, “Allosteric Regulation of DNA Cleavage and Sequence-Specificity through Run-On Oligomerization”, *Structure*, (2013) 21:1848-58.
- [3] Ahlstrom LS, Baker JL, Ehrlich K, Campbell ZT, Patel S, Vorontsov II, Tama F, Miyashita O., “Network visualization of conformational sampling during molecular dynamics simulation”, *J Mol Graph Model*, (2013) 46: 140-149.
- [4] Q. Jin, COS. Sorzano, JM. de la Rosa-Trevín, JR Bilbao-Castro, R. Núñez-Ramírez, O. Llorca, F. Tama and Slavica Jonić, “Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes”, *Structure*, (2014) 22:496–506.

(2) Invited Talks

- [5] Tama F, “Development of computational tools to characterize structure of biological molecules from low-resolution data”, RIKEN joint symposium with INSA, Lyon, France May 16th 2013.
- [6] Tama F, “Structural modeling from low-resolution experimental data”, Workshop on “modeling of biomolecular systems in cellular environments” Kyoto University Oct 31st - Nov 1st, 2013.
- [7] Tama F, “Structural modeling of biological molecules from low-resolution experimental data”, AICS international symposium, Kobe Dec 2nd-3rd 2013.
- [8] Miyashita O, Tokuhisa A, Tama F, “An Overview of Single Biomolecular Imaging by X-ray Free Electron Laser”, Institut de Minéralogie et de Physique des Milieux Condensés, France. Dec. 2013.

(3) Posters and Presentations

- [9] Q. Jin, C. O. S. Sorzano, J.M. Delarosa-Trevin, J.R. Bilbao-Castro, R. Núñez-Ramírez, O. Llorca, F. Tama, and S. Jonić, “HEMNMA: Hybrid Electron Microscopy Normal Mode Analysis to fully explore continuous-type conformational changes”, Cryo-EM 3D Image Analysis Symposium 2014, Lake Tahoe, USA.
- [10] Atsushi Tokuhisa, Osamu Miyashita, Florence Tama, “Methodology of a single bio-molecular structure determination for low-resolution data set obtained by X-ray Free Electron Laser”, The 51th Annual Meeting of the Biophysical Society of Japan, Kyoto, October 28–30, 2013.