Computational Structural Biology Research Unit

1. Team 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.

A variety of experimental techniques exist toward this goal. X-ray crystallography has been the primary tool to study protein conformations, providing high-resolution structures. Even though cryo electron microscopy (EM) provided lower resolution data, it has provided critical information on structure and dynamics of large biological molecules. More recently, efforts like 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, at the current experimental condition, it cannot achieve atomic level resolution such as obtained by X-ray crystallography.

Computational approaches are valuable under this situation, since they could provide information that is not accessible from experiments. 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. In particular, we are developing hybrid analysis that combines low-resolution experimental data as obtained from, such as, cryo-EM and XFEL, with computational methods that utilizes high performance computers, such as K computer. These tools help us to acquire knowledge on the structure of physiologically important protein complexes that are unattainable with existing experimental techniques, and in the longer term 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. In previous years, we had developed a new method, Hybrid Electron Microscopy Normal Mode Analysis (HEMNMA) that utilizes molecular mechanics algorithms to simulate conformational dynamics and use the resulting conformations to extract the dynamical information from cryo-EM images. In order to make this method more accessible to the general public and to facilitate its usage, a friendly graphical interface was developed with the Xmipp framework. Three modules (normal mode analysis, flexible alignment and advanced results analysis) are available (see Figure 1) and all HEMNMA analysis can be performed using this new graphical interface.

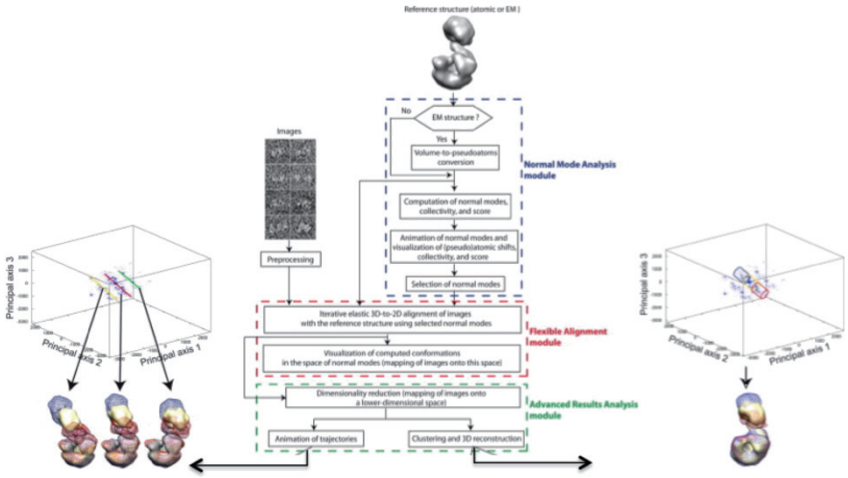

Figure 1: Modules available in the HEMNMA graphical interface and example of analysis that can be performed. (Adapted from CO. Sorzano et al. J. Struct. Biol. (2014))

In addition, we discussed the application of this method to the dynamics of the Tomato Bushy Stunt Virus (TBSV), which is a small icosahedral plant virus known to undergo conformational rearrangements upon change in pH conditions. 4000 particles images and three symmetric normal modes, which reflect the icosahedral arrangement of the virus, were considered for HEMNMA approach. The analysis showed that images could be classified into 5 classes representing different conformation of the virus, with the primary conformational change consisting in a radial expansion of the virus particle consistent with previous experiments. Such conformations were used to discuss in more details the conformational change.

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. Numerous computational tools have been developed to interpret conformational change observed in cryo-EM data (flexible fitting).

Using multiple flexible fitting methods can be a useful approach to evaluate the accuracy of models by comparing difference in their resulting model conformation. While flexible fitting methods can produce accurate atomic models, there are not always successful. In some particular cases, several of these methods consistently fail. In light of such results, to obtain a better understanding of the behaviors observed in these fittings and evaluate the limitation of the flexible methods, a more comprehensive survey of the fitting performance is performed using multiple initial fitting conditions.

Our flexible fitting approach was implemented in GENESIS (Dr. Sugita's team). Running multiple simulations (320 different conditions) revealed that accurate models are not consistently obtained (Figure 2, No REUS). Therefore we implemented a new scheme to perform flexible fitting, which relies on replica exchange simulation, a method that enables better conformational sampling. In such approach, *n* simulations are run concurrently with *n* different force constant. Conformations are allowed to exchange between these simulations, therefore experiencing different force constant and enhancing sampling (REUS). Without REUS (No REUS in Figure 2) most models have low accuracy (RMSD ~6Å) while with REUS, all models have an RMSD around 2.5Å RMSD. Therefore, appropriate conformational sampling is critical to obtain accurate model and using REUS can improve accuracy of the models.

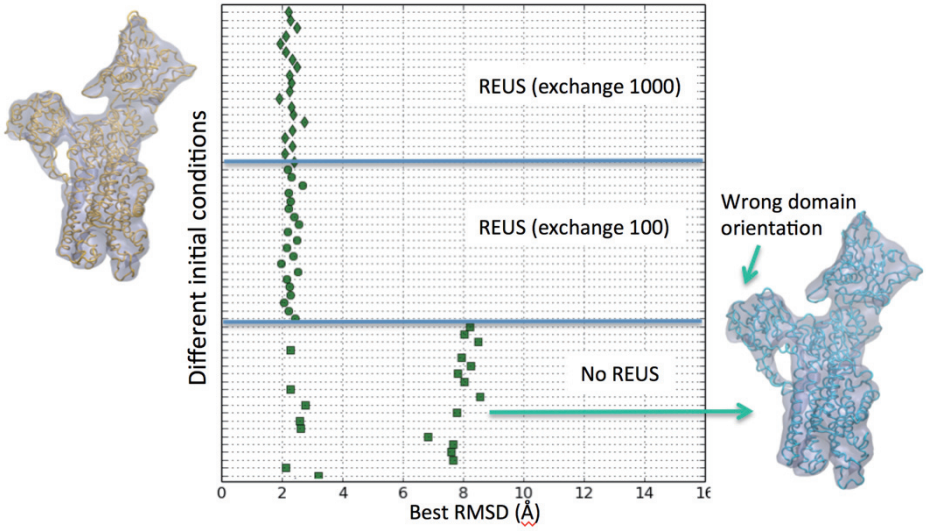

Figure 2: Best RMSD with the target structure obtained from multiple simulations. Without REUS, atomic models might not be accurate (high RMSD observed). Using REUS, atomic models within 2. 5Å are consistently obtained.

3.3. Collaboration on X-ray diffraction imaging

In this fiscal year, we have started a collaboration with Dr. Song (RIKEN SPring-8 Center, Imaging Development Team) on RNA sponges, which are assembled from small interfering RNA molecules. RNA sponges are biologically relevant as they can silence gene expression. Dr. Song and collaborators collected data on RNA sponges using coherent diffraction imaging. Our group was in charged of reconstructing the overall shape of the molecules using computational tools. The first overall 3D structure of a RNA sponge was described (Figure 3). Even though the structure is low resolution, it revealed a core with higher density suggesting a higher organization or a core more packed than the surface.

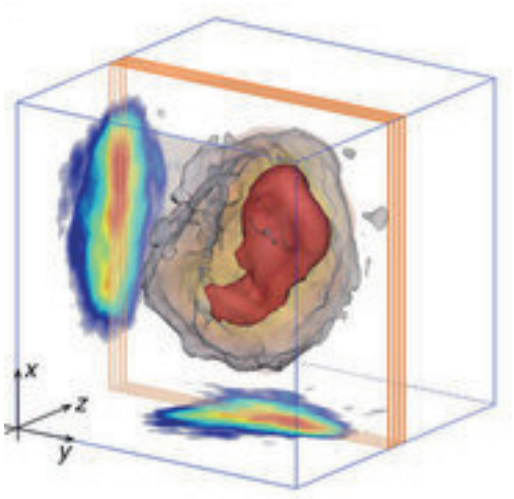

Figure 3: Complete 3D structure of a microsponge is shown. The overall shape is displayed transparently (grey) in order to visualize the internal dense region (red). Projection densities along xz-plane and yz-plane are superposed. Adapted from M. Gallagher-Jones et al. (2014) Nat Commun. May 2;53798

3.4. Computational tools to analyze XFEL experimental data

We are also developing tools to analyze XFEL data in order to obtain structural information of biological molecules. In particular, we aim to develop computational algorithms that would provide the shape of the biological systems. 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. Such algorithm is better suited to current XFEL single particle experiments targeted on μm scale systems, such as organelles and small cells.

Previously, we performed "fitting" studies i.e. an atomic model is created by deforming other known structures. However, for these μm scale systems, such an approach would not be suitable. Therefore, we have started to work on ab initio structure modeling from XFEL diffraction patterns, which can propose a structure model, though at low-resolution, from the data alone. We chose gold particles as a model system for the first test case. Such model system is often used for initial testing in experiments and many data are available, therefore it is a good candidate to develop a new approach. Preliminary data on a small system composed of three gold colloids shows that actual arrangements can be obtained by comparing the target X-ray diffraction pattern with the patterns from candidate model structures (Figure 4).

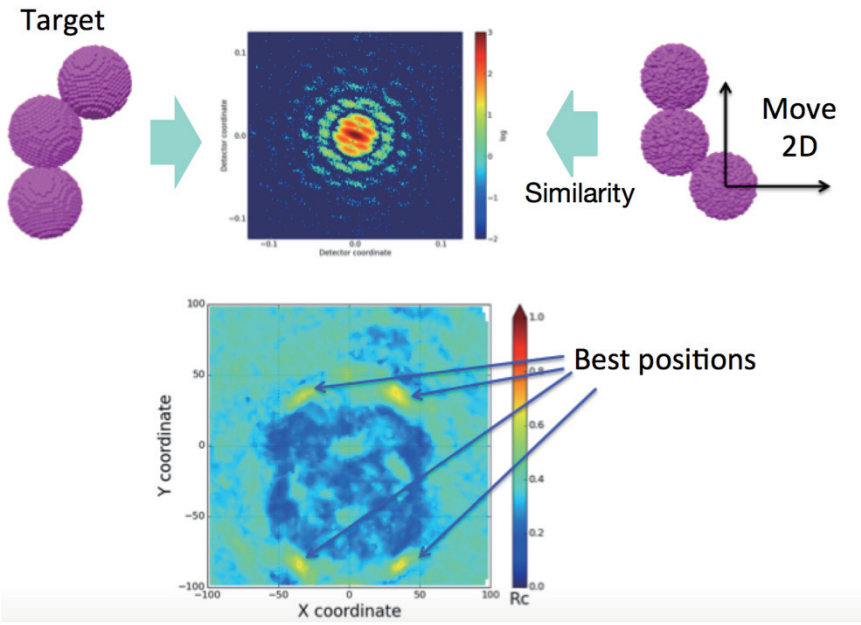

Figure 4: Comparison between target diffraction pattern and candidate models as one gold colloid is moved in X and Y directions.

3.5. Collaboration

We are collaborating with Dr. Sugimoto's group (FIBER, Konan University) and Dr. Tanaka (Kobe University) to study the effects of solvent properties on DNA thermal stability. Detailed atomic molecular dynamics simulations on DNA molecules adopting different conformations with different

co-solute concentrations were performed. Such studies revealed new insights (effect of co-solute, water-DNA and co-solute-DNA interactions) for a better understanding of the thermal stability of DNA structures in a molecular crowding environment, which is critical to understand physiology in vivo.

4. Schedule and Future Plan

We are planning to continue to develop tools to analyze XFEL data in order to obtain structural information of 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.

One emphasis of our research is the description of the dynamics of biological molecules through the development of computational tools to characterize low-resolution experimental data. We intend to continue and establish new collaborations with experimental groups in Japan and abroad in order to study structure, function and dynamics of biological molecules.

Finally, on the longer term we plan to establish a computational framework to build structures from low-resolution structural data without a priori knowledge of the overall structure of the molecular complexes. Such approach would integrate multiple types of experiments, multiple types of computational methods (multi-scale modeling and simulations, protein structure prediction, protein-protein interactions…). Such framework would be important to study structure/function and dynamics of biological molecules.

5. Publication, Presentation and Deliverables

(1) Journal Papers

1. M. Gallagher-Jones, Y. Bessho, S. Kim, J. Park, S. Kim, D. Nam, C. Kim, Y. Kim, Y. Noh do, O. Miyashita, F. Tama, Y. Joti, T. Kameshima, T. Hatsui, K. Tono, Y. Kohmura, M. Yabashi, SS. Hasnain, T. Ishikawa, C. Song. (2014) Macromolecular structures probed by combining single-shot free-electron laser diffraction with synchrotron coherent X-ray imaging. **Nat Commun**. 2:5:3798
2. S. Patel, E. Vierling and F. Tama. (2014) Replica exchange molecular dynamics simulations provide insight into substrate recognition by small heat shock proteins. **Biophys J**. 10:2644-55
3. Hybrid Electron Microscopy Normal Mode Analysis graphical interface and protocol. (2014) CO. Sorzano, J.M. de la Rosa-Trevín, F. Tama, S. Jonić. **J. Struct. Biol**. 188:134-41

(2) Conference Papers

1. Q. Jin, C.O. Sanchez Sorzano, I. Callebaut, F. Tama, S. Jonic (2014). Elastic image registration to fully explore macromolecular dynamics by electron microscopy. In: Image Processing (ICIP), IEEE International Conference on, 27-30 Oct. 2014, 2075-2079

(3) Invited Talks

1. Computational tools to characterize structure of biological molecules from low-resolution data. Novel measurement techniques for visualizing 'live' protein molecules at work - Kickoff Symposium. October 2014. Kyushu University, Japan.
2. Structure modeling with EM and XFEL data. Osamu Miyashita, Atsushi Tokuhisa, Florence Tama. Coarse-Grained Modeling of Structure and Dynamics of Biomacromolecules 3. August 2014. Telluride workshop. USA

(4) Posters and presentations

1. Examination of *ab* initio structural modeling for the pattern matching method using X-ray free electron laser. Atsushi Tokuhisa, Osamu Miyashita, Florence Tama. Annual Meeting of the Biophysical Society of Japan. September 2014. Sapporo, Japan.
2. Hybrid Approach for X-Ray Free Electron Laser Single Particle Analysis of Biomolecular Systems Osamu Miyashita, Atsushi Tokuhisa, Florence Tama. Keystone Symposia. Hybrid Methods in Structural Biology. March 2015. Tahoe City, USA

(5) Patents and Deliverables