19.Computational Structural Biology Research Unit

19.1. Team members

Florence Tama (Unit Leader)

19.2. 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. 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, it cannot achieve atomic level resolution such as obtained by X-ray crystallography. Indeed, theoretical work has shown that electron density maps of biological molecules could be obtained with resolution around 5 Å. Therefore; additional analyses on the XFEL data would be necessary to obtain atomic level structures.

Computationally, methods have been developed to predict structures from low-resolution data such as cryo-electron microscopy (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 too bring experimental data as obtained from X-ray 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.

19.3. Research Results and Achievements

The computational structural biology unit was established full-time in January 2013. The results presented here reflect the period from January to March 2013.

1. Cryo-EM qualitative fitting

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 low-resolution, in order to obtain higher-resolution information, know X-Ray structure are often combined with cryo-EM data which is 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.

We have designed a new protocol for automating this process in order to obtain qualitative data on the accuracy of the fitting. Several starting rigid body fitting orientations are considered. In order to automatically identify a best orientation among the candidates obtained by rigid-body fitting, we performed flexible fittings using four different approaches starting with each of the candidate orientation. Models obtained from each fitting were then compared using a Root Mean Square Deviation (RMSD) as a measure. It is found that different flexibly fitting models only show consensus, i.e. low RMSD values, in conformation when the fitting procedures are started with a viable orientation (Figure 1).

This approach for identifying best orientation is an important step towards making the flexible fitting procedure fully automated. It is important to point out that the flexible fitting approaches used are themselves fully automated – they do not require any domain segmentations or human interactions. However, they require a suitable guess of the starting orientation, which requires human interactions or additional knowledge of the system. Our results shows that this requirement can now be eliminated and the whole fitting procedure can be automatized. As a result, the fitting procedure would only require initial structure and the target map. We are currently writing a paper describing those results.

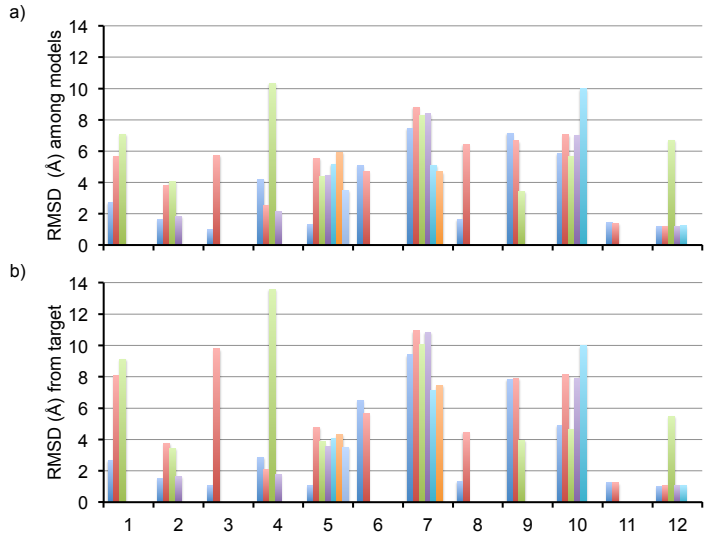

Figure 1: Mean of the pairwise RMSD among the four flexibly fitted models (panel a) and mean RMSD between the target structure and the four flexibly fitted models (panel b) using the different rigid body fitted orientations (representatives in colors) of the initial structure for each of the 12 proteins.

2. XFEL

During this fiscal year, the unit has also started discussions with researchers involved with XFEL experiments. Our unit has attended a workshop held a Keio University on XFEL. In addition, our unit visited the Harima site to observe XFEL experiments and discussed possible collaborations. These initial discussions have enabled us to identify several areas where our experience with cryo-EM data analysis could be pertinent with XFEL data analysis.

19.4. Schedule and Future Plan

We are planning on developing tools to analyze XFEL data, with special emphasis on the dynamics that can be extracted from the data. During the next fiscal year, we plan to make flexible fitting programs available to the community. XFEL data should provide low-resolution to medium resolution structural data. For an atomic description of the biological molecules, algorithms will be needed to deform known X-ray/NMR structures to fit correctly with the data from XFEL. The methods we have been developing for cryo-EM data will lay the foundation for this work. Such methods would allow investigating the role of the structural changes responsible for a protein function. Algorithms will be implemented within the Genesis program suite being developed specifically for K computer by Dr. Sugita (AICS RIKEN).

In addition, we plan to study extend of dynamics within raw low-resolution data. 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.

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 as potential targets of the structural analysis by XFEL 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 XFEL, 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.

19.5. Publication, Presentation and Deliverables

(1) Journal Papers

-None

(2) Conference Papers

-None

(3) Invited Talks

Understanding functions of biological molecules using hybrid methods, Feb 15th 2013 Department of Physics, Nagoya University

(4) Posters and presentations

Characterization of large biological complexes using computational approaches and low-resolution experimental data. AICS 3rd International Symposium

(5) Patents and Deliverables

-None