Chapter 18

Computational Structural Biology Research Team

18.1 Members

Florence Tama (Team Leader)

Osamu Miyashita (Senior Scientist - October-March)

Bhaskar Dasgupta (Research Scientist - May-March)

Miki Nakano (Post-doctoral Researcher)

Sandhya Tiwari (Post-doctoral Researcher)

Loic Broyer (Intern – 3months from University of Grenoble)

Yumeno Kusahara (Assistant)

18.2 Overview of Research Activities

Biological molecular complexes of proteins and RNAs are of great interest in the area of molecular biology, as they are responsible for core biological functions such as cell replication, gene transcription, protein synthesis, regulation of cellular transport and numerous others. Those systems undergo large conformational transitions to achieve functional processes. Therefore, characterization of dynamical 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. Cryo-electron microscopy (EM) is becoming an important technique due to development of experimental apparatus as well as data analysis software. Although the structural resolution of the cryo-EM data tends to be at lower resolution, it has revealed 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 biomolecules 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. High-speed atomic force microscopy (HS-AFM) is also a technique that is seeing growing number of biological applications. This technique enables observation of biological molecular complexes in motions in near-native environment, providing unique information regarding functional dynamics. However, due to the limitation of manufacturing process of the observation device, the structural resolution of the images are relatively low. Each experimental technique has its own strengths and weaknesses, and thus integration of computational simulations with experimental data is beneficial to obtain new and detailed information on the structure and dynamics of biological molecules.

Computationally, methods have been developed to predict structures from low-resolution experimental data such as from 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, which could also be used for the modeling of new biomolecular complexes. Such hybrid approaches need to be extended further to integrate various available experimental data to determine the structures of important biomolecular complexes.

The ultimate line of our interdisciplinary research is development and applications of computational tools, through high performance computing, to integrate experimental data as obtained from various techniques such as X-ray crystallography, cryo-EM and XFEL with molecular modeling and simulations to acquire knowledge on the structure of a physiologically important protein complexes that are unattainable with existing experimental techniques.

18.3 Research Results and Achievements

18.3.1 Developments of computational tools for XFEL experimental data

Recent development of intense X-ray free-electron laser (XFEL) light sources offers a new possibility to image single biological macromolecules. Since crystallization is not necessary for such structure analysis, it would be possible to investigate the structure of macromolecular complexes under various physiological conditions or to observe elementary steps of a biochemical function. SPring-8/SACLA in Japan is one of the pioneering XFEL facilities in the world.

However, with current experimental conditions, atomic level resolution such as obtained by X-ray crystallography cannot yet be achieved. As XFEL experiments are very recent and still undergoing further development for routine study of biological molecules, computational algorithms and tools to understand and analyze experimental data need to be developed simultaneously. One focus of our research is the development of such computational tools. We have started to tackle these issues from multiple angles described below.

18.3.1.1 3D Reconstruction from XFEL Diffraction Patterns – Applications to Experimental Data

In order to apply the XFEL single particle structure analysis algorithms we have been developing to experimental data, we are conducting joint research with Professor Yoshinori Nishino's research group at Hokkaido University. Pulsed coherent X-ray solution scattering (PCXSS), which is being developed by Nishino group, is a method that can observe samples in aqueous solution using XFEL, and with the continuous development of experimental methods, it is becoming possible to observe nano-scale samples.

This year, we have been analyzing the XFEL diffraction data of gold nanoparticles with a size of about 40-50 nanometers, aiming to reconstruct the three-dimensional structure the sample. Since experimental data include measurements in various situations, such as sample hits, multiple-sample hits, and misses, and furthermore the intensities of the diffraction patterns are not homogeneous due to fluctuations in the beam intensity during the measurements, we had to develop algorithms for selecting data that can be used for three-dimensional structure recovery (Figure 18.1). Filtered diffraction patterns are then assembled into 3D space using the algorithms we have been developing, with additional improvement to take into account the fluctuation of diffraction intensity strength resulting from beam intensity fluctuations. The model of nano-particle can be obtained from the obtained 3D diffraction intensity volume through phase-recovery procedure. The suitable condition for the procedure is currently investigated.

18.3.1.2 "Idea generator" from 2D XFEL data of biological systems

Data analysis for XFEL data remains challenging. XFEL diffraction pattern is an unintuitive representation of the projection image of the sample in Fourier space. For biological systems, the current standard approach to reconstruct a real-space image of the sample, phase recovery, often fails due to the low diffraction power of biological samples. Therefore, we are developing a new hybrid approach to interpret diffraction patterns that utilizes image analysis with database search. More specifically, for a given set of XFEL diffraction patterns, we identify a plausible structures through searching database of possible 3D low-resolution shapes. We have previously shown the feasibility of such an approach with real space images from EM.

In the XFEL adaptation, we have had to overcome the challenge of aligning 2D image patterns that are in Fourier space, considering they do not contain phase information. Being in Fourier space, low-frequency (center)

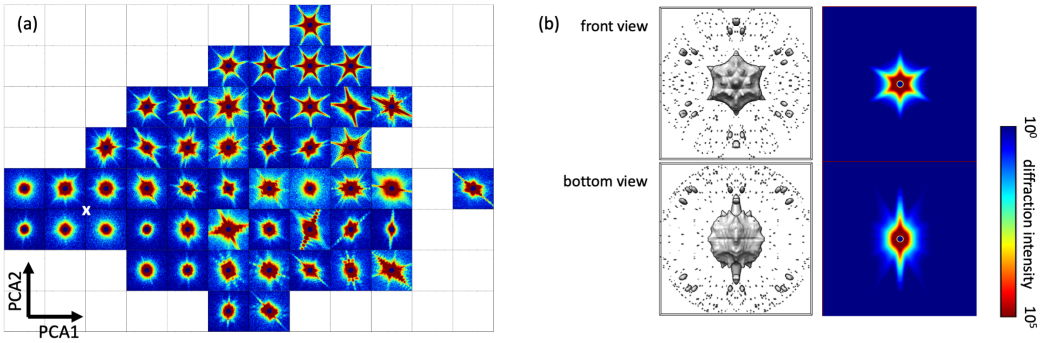

Figure 18.1: (a) Classification of experimental data. (b) A 3D diffraction intensity distribution reconstructed by slice matching algorithm. Isosurface plots from the top and the bottom view and the corresponding slice images are shown.

regions of diffraction patterns reflect the overall shapes of the molecule, while the details of the molecular structure is represented in high-frequency (outer) regions. Since our goal is to identify the low-resolution shapes that are consistent with target experimental data, we need to focus on the low-frequency regions. Furthermore, we are improving the algorithms so that it can deal with the diffraction patterns from actual experimental data. Diffraction patterns from biomolecules contain limited amount of signals due to the weak diffraction intensity, particulary being weak at high and strong at low wavenumber pixels, requiring careful selection of the matching region when aligning the diffraction images in Fourier space. Thus, only a certain region in the diffraction pattern (Region of Interest, ROI) can be used for the proposed match-finding algorithms, which poses a significant challenge in comparison to EM images. To automate the identification of ROI, we have developed an algorithm, in which the approximate size of the sample in the input image is estimated via fitting against a theoretical model, spherical form factors, and used to estimate the appropriate ROI. Using the estimated ROI, the match-finding algorithm to identify plausible candidate 3D models from a few XFEL diffraction patterns has been improved (Figure 18.2). In general, we find that the success of the strategy relies on the complexity of the shape and the diversity of input diffraction patterns. Nevertheless, the features of the shape in the input images are still captured within in the top matching hits.

18.3.2 Developments of computational tools for analyzing cry-EM experimental data

We have been developing a method to obtain detailed structural models from low resolution experimental data such as EM data using a molecular dynamics simulation. The three-dimensional image obtained from the cryo-EM is given as data on a three-dimensional lattice. Here, the parameter that specifies the actual size of each lattice determines the size of the whole molecule and thus affects the modeling results. However, it is known that there can be an error of a few % in the estimation of this parameter. Therefore, in order to improve the accuracy of structural modeling, we have developed an algorithm that can estimate an accurate value for this lattice size parameter.

In this algorithm, many EM maps with different grid sizes are prepared and large number of models are created by performing structure refinement against these maps . Then, the correct lattice size is estimated by evaluating the "quality" of each of these structures. Here, evaluation of the "quality" of the obtained structures needs to be performed independent of the fitting procedure. The algorithm was tested using the maps created by simulation for several model systems, and various "quality" evaluation approaches were examined. First, we showed that the general correlation coefficient, which measures the degree of agreement between the structure and the map as corresponding electron densities, is not sufficient to detect correct grid size (Figure 18.3). It was also found that the Molprobity index, which is commonly used to evaluate X-ray crystal structure models, is insufficient because there is no physicochemical aspect in the evaluation of interatomic interactions. Finally, we found that GOAP, which is an index used for protein structure prediction, is effective in evaluating the local structure and interatomic distance in proteins, which is the "quality" of the structure necessary to identify correct grid size. We applied the developed algorithm to experimental data and demonstrated the usability of the algorithm.

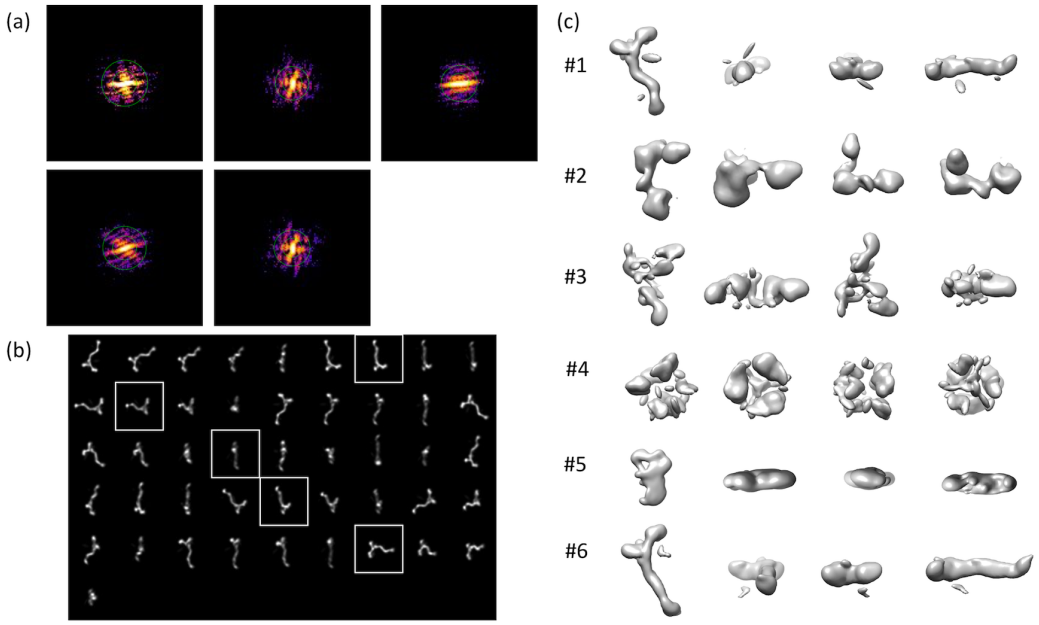

Figure 18.2: An example of the model finding algorithm outputs (a) Five simulated diffraction patterns that are used as the query inputs. Two circles on each pattern indicate the definition of Region of Interest that are determined using the newly developed algorithm. (b) Projection images of the "answer" structure. Five input patterns correspond to the input diffraction patterns. (c) Six cryo-EM maps that are identified as the models consistent with the input diffraction patterns.

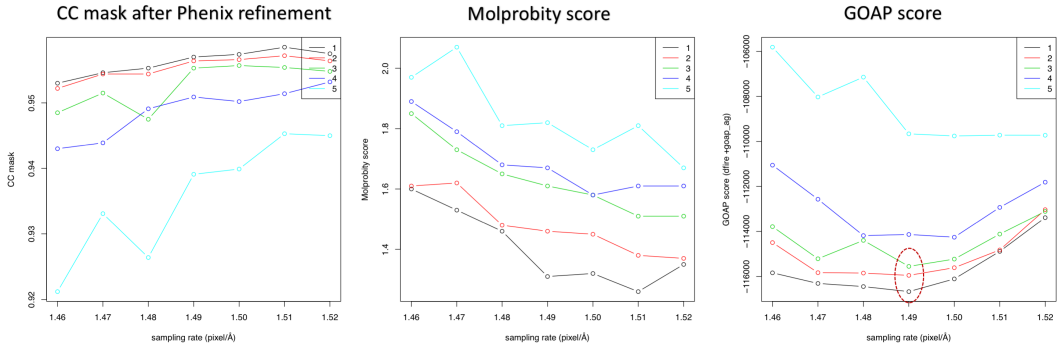

Figure 18.3: A variation of target EM maps were generated by changing the size of the grid lattice (X axis) and multiple structure refinement trials were performed against these EM maps. Then the "quality" of those structures was evaluated by several indicators (correlation coefficient CC, Molprobity score, GOAP score), and the GOAP score was most consistent with the ground truth (1.49). It was found that the structural model (circle) based on the cubic lattice size can be identified as the most correct.

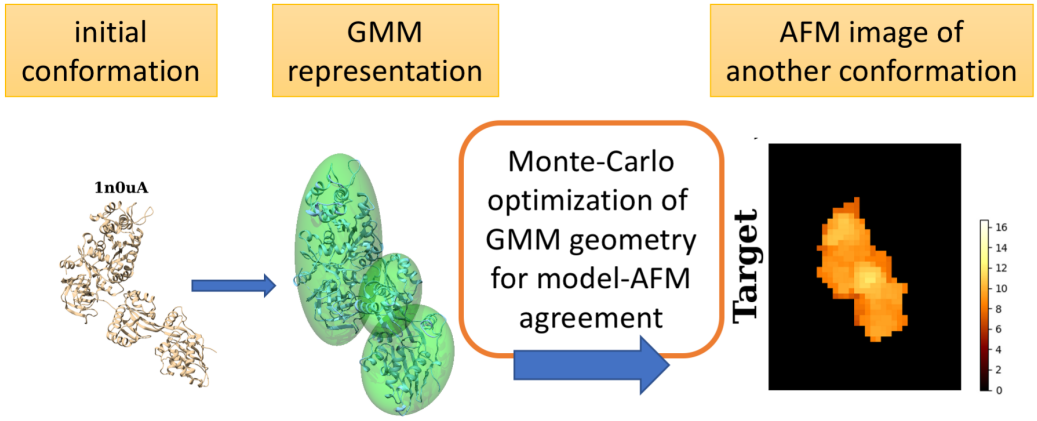

Figure 18.4: Overview of the new approach for conformational modeling of proteins based on AFM images. A structure of the target protein (initial conformation) is represented using Gaussian mixture model (GMM representation). The positions of the Gaussian kernels are optimized via Monte-Carlo optimization so that the agreement with the target AFM experimental image is maximized.

18.3.3 Modeling from AFM images

Atomic Force Microscopy (AFM) is an experimental technique which enables observation of biomolecules in near native condition. AFM uses mechanical device (cantilever) to trace the molecular surface and provides images of biomolecules at nanometer resolution. In particular, High-speed AFM experiments produce a series of images following live dynamics of biomolecules, which may provide a wealth of information regarding protein dynamics and functions. However, the information in the data is very limited, 2-dimensional and low-resolution. To further understand biomolecular functions, information on three-dimensional (3D) structures is beneficial.

We are developing algorithms to recover 3D information from AFM images by computational modeling. The AFM image includes only low-resolution representation of a molecule; therefore we represent the structures by a coarse grained model (Gaussian mixture model). Using Monte-Carlo sampling, candidate models are generated to increase the agreement between AFM images simulated from the models and target AFM image. We have tested the algorithm on two proteins which undergo large conformational transitions. AFM images were simulated from one conformation and used as synthetic target AFM image (Figure 18.4). We showed that, starting from another conformation, the algorithm can produce a low-resolution 3D model of the target molecule. Effect of molecular orientations captured in AFM images on the 3D modeling performance was also examined and it is shown that similar accuracy can be obtained for many orientations. The proposed algorithm can generate 3D low-resolution protein models, from which conformational transitions observed in AFM images can be interpreted in more detail.

18.3.4 Molecular dynamics simulations of Tim21 protein for interpretation of low-resolution X-ray crystallographic data

X-ray crystallography has been the major source of information on protein structures. It can provide detailed atomic level information, yet the sample needs to be crystallized and generally cryo-cooled, and the obtained data represents a frozen snapshot of proteins in crystal. Recently, new approaches to obtain dynamical information of biological molecules are actively studied. Professor Kohda group at Kyushu University proposed a method called CCFS, in which protein crystal is designed to have sufficient space inside so that molecules can still move inside the crystal. The electron density map from CCFS is at much lower resolution than conventional crystallography, due to the conformational flexibility, and as a result, analysis of such data also requires new computational approaches.

We have performed molecular dynamics simulations in order to interpret CCFS experimental data of a protein, Tim21, particularly regarding the conformation of a loop, L2. Tim21 is a subunit of a highly dynamic translocase of the inner mitochondrial membrane complex. A loop segment in Tim21, which is in close proximity of the binding site of Tim23, was recently studied by CCFS and the obtained electron density indicates the loop conformation that is different conformations obtained form conventional X-ray and NMR experiments, making

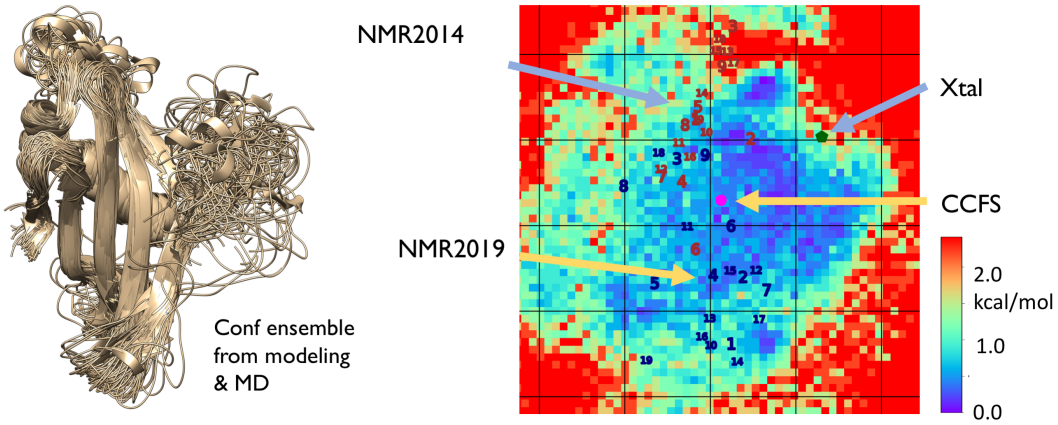

Figure 18.5: Conformational ensemble of loop L2 of Tim21 protein obtained by molecular dynamics simulation in solution environment (left). Free energy surface estimated from the results show that CCFS structural model is most consistent with the predicted solution conformations.

the analysis of the new experimental data difficult. Using MD simulations we can provide information on the structure and dynamics of the loop in solution to provide interpretation of these data.

We obtained the conformational ensemble of the loop using ab-initio loop modeling techniques and molecular MD simulations (Figure 18.5). MD simulations confirmed mobility of the loop. Multidimensional scaling and clustering were used to characterize the dynamic conformational ensemble of the loop. Free energy landscape showed that the CCFS crystal structure occupied a low energy region as compared to the conventional X-ray crystal structure. Analysis of crystal packing indicates that the CCFS provides larger conformational space for the motions of the loop. These methods in integration with experimental techniques such as CCFS has the potential to transform the studies on flexible regions of proteins.

18.4 Schedule and Future Plan

Cryo-EM experiments are quickly becoming an essential tool for studying biomolecular complexes. New XFEL facilities are getting into operation every year in the world, providing opportunities for new experiments. The amount of data from such experiments will continue to grow in numbers and analysis of such big datasets will increase the necessity of high-performance computing. Time-resolved experiments using high-speed AFM and XFEL serial femto-second crystallography also provides additional information of molecular movies. We aim to utilize high performance computer such as Fugaku to break the limitation of current data processing and hybrid computational modeling approaches to obtain a new level of structural information of biological complexes from various experimental data. For this goal, we plan to develop algorithms and software to analyze large dataset to obtain not only structural models but also dynamical information which is essential for understanding of the mechanisms of biomolecular functions. With these techniques, through collaborations with experimental groups, we aim to contribute to the structural biology community.

18.5 Publications

18.5.1 Articles/Journal

[1] Bala S, Shinya S, Srivastava A et al. Crystal contact-free conformation of an intrinsically flexible loop in protein crystal: Tim21 as the case study. Biochim Biophys Acta Gen Subj. 2020;1864:129418.
[2] Srivastava A, Bala S, Motomura H, Kohda D, Tama F, Miyashita O. Conformational ensemble of an intrinsically flexible loop in mitochondrial import protein Tim21 studied by modeling and molecular dynamics simulations. Biochim Biophys Acta Gen Subj. 2020;1864:129417.
[3] Srivastava A, Bala S, Motomura H, Kohda D, Tama F, Miyashita O. Conformational ensemble of an intrinsically flexible loop in mitochondrial import protein Tim21 studied by modeling and molecular dynamics simulations. Biochim Biophys Acta Gen Subj. 2020;1864:129417.

[4] Dasgupta B, Miyashita O, Tama F. Reconstruction of low-resolution molecular structures from simulated atomic force microscopy images. Biochim Biophys Acta Gen Subj. 2020;1864:129420.
[5] Nakano M, Miyashita O, Tama F. Parameter optimization for 3D-reconstruction from XFEL diffraction patterns based on Fourier slice matching. Biophys Physicobiol. 2019;16:367-376.
[6] Nagai T, Tama F, Miyashita O. Cryo-Cooling Effect on DHFR Crystal Studied by Replica-Exchange Molecular Dynamics Simulations. Biophys J. 2019;116:395-405.
[7] Srivastava A, Tama F, Kohda D, Miyashita O. Computational investigation of the conformational dynamics in Tom20-mitochondrial presequence tethered complexes. Proteins. 2019;87:81-90.

18.5.2 Invited Talks

[8] Tama F, "Hybrid approaches to reveal structure and dynamics of large biological complexes from single molecule experiments", American Chemical Society National Meeting, 2019/3/31-4/3, Orlando, US
[9] Miyashita O, "Hybrid Structure Modeling of Biomolecules from", Chem-Bio Informatics Society(CBI) Annual Meeting 2019, 2019/6/17-19, Tokyo, Japan
[10] Tama F, "Modeling conformational transitions of biomolecules from cryo-EM data", French Electron Microscopie Society meeting, 2019/7/2-5, Poitier, France
[11] Tama F, "Kinase inhibitors for modulation of circadian rhythms", American Chemical Society National Meeting, 2019/8/25-29, San Diego, US
[12] Tama F, "Molecular mechanisms involved in the regulation of the Circadian Clock", The 10th Toyota RIKEN international workshop on Science of Life Phenomena Woven by Water and Biomolecules, 2019/9/4-6, Nagoya, Japan
[13] Tama F, "Integrative modeling to characterize structure and dynamics of biomolecules from single molecule experiments", Integrative structural biology meeting, 2019/10/7-11, Toulouse, France
[14] Miyashita O, "Hybrid Approach for Structural Biology: Simulation and Experimental Data", CBI-society meeting 2019, 2019/10/22-24, Tokyo, Japan
[15] Miyashita O, "Hybrid Structure Modeling of Biomolecules using Cryo-EM Data", KINKA Chemical Society 106th Open Lecture - Structure Analysis of Biomolecules using Cryo-electron Microscopy, 2019/10/30, Osaka Japan
[16] Miyashita O, "Reconstruction of Low-resolution (Bio) Molecular Structures from Simulated Atomic Force Microscopy Images", The 1st International Conference on Big Data and Machine Learning in Microscopy, 2020/01/15-17, Kanazawa, Japan
[17] Tama F, "Flexible fitting methods and applications", EMBO workshop, CEM3DIP 2020: Single particle cryoEM of macromolecular-assemblies and cellular tomography, 2020/1/19-30, Kolkota, India

18.5.3 Oral Talks

[18] Nakano M, Miyashita O, Tama F, "Effect of binning size of XFEL Diffraction Patterns on the resolution of reconstructed 3D-molecular structure", Chem-Bio Informatics Society(CBI) Annual Meeting 2019, 2019/10/22-24, Tokyo, Japan

18.5.4 Poster Presentations

[19] Tiwari SP, Miyashita O, Tama F, "COMPUTATIONAL PROTOCOL FOR OPTIMIZING THE PIXEL SIZE PARAMETER TO IMPROVE SINGLE PARTICLE CRYO-ELECTRON MICROSCOPY MAPS.", Biophysical Society 64th Annual Meeting, 2020/2/17, San Diego, California, USA
[20] Miki N, Tama F, Miyashita O, "Effect of binning size of XFEL Diffraction Patterns on the resolution of reconstructed 3D-molecular structure", Chem-Bio Informatics Society(CBI) Annual Meeting 2019, 2019/10/22-24, Tokyo, Japan
[21] Dasgupta B, Miyashita O, Tama F, "Hybrid Structural Modeling of Proteins Based on Atomic Force Microscopy (AFM) to Recover Conformational Transition Included in AFM Images", The 2nd R-CCS International Symposium, 2020/2/17-18, Kobe, Japan