様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月　 ４ 日現在

機関番号：１３９０１
研究種目：基盤研究(C)（一般）
研究期間：2017～2019
課題番号：１７Ｋ０７３０５

研究課題名（和文）Database of Molecular Shapes and Diffraction Patterns for X-ray Free Electron Laser Data Analysis

研究課題名（英文）Database of Molecular Shapes and Diffraction Patterns for X-ray Free Electron Laser Data Analysis

研究代表者
ＴＡＭＡ　ＦＬＯＲＥＮＣＥ（Tama, Florence）

名古屋大学・理学研究科・教授

研究者番号：２０６４８１９１

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：本研究ではクライオ電子顕微鏡やX線自由電子レーザーにより得られる少数の画像データから大まかな構造を推定するアルゴリズムを開発した。この手法では既知の生体分子の形に関するデータベースを構築し、その中からインプット画像と一致度の高い構造を選択、提案する。
まず、電子顕微鏡データによりこの手法の実現性を検討し、少数の画像から比較的正確に正しい形の情報を選択できることを示した。次に、X線自由電子レーザーデータの解析を高速化するためにGaussian mixture modelで粗視的にモデルを表現する手法を開発した。また、画像比較の精度を高めるために最適な画像領域を自動的に選択する手法を開発した。

研究成果の学術的意義や社会的意義
生体分子の構造に関する情報はそれらの機能を理解して医薬などの応用に活用するために重要である。本研究ではデータベースを活用することによりクライオ電子顕微鏡やX線自由電子レーザーによる観測データからターゲットの形を素早く推定するための手法を開発した。

研究成果の概要（英文）：In this project, we have been developing new efficient approaches to find 3D biological shapes from a few EM or XFEL images to serve as a starting point for further data analyses. In this approach, databases of known molecular shapes are assembled and numerical algorithms are used to identify the shapes that are consistent with a few query images.
We had first developed the protocol for using EM real space images as inputs to test the feasibility of the approach. We showed that a small number of images can be sufficient as query images to identify similar 3D shapes. Then we have been developing algorithms to use XFEL diffraction patterns as inputs. Since simulations of XFEL diffraction patterns is time-consuming, we have developed a new approach to use Gaussian mixture model to model the structure. We have also the algorithms to automatically identify the region with strong information in the diffraction patterns to identify matches.

研究分野：Computational Biophysics

キーワード：2D images　XFEL　Cryo-EM

１．研究開始当初の背景

Developments of X-ray free-electron laser (XFEL) light sources offer a new possibility for imaging biological systems. Its extremely strong X-ray laser allows imaging of biological systems without crystallization and therefore it could be applied to a wider variety of systems under various physiological conditions. In addition, such a strong light enables “single shot” imaging, i.e., one pulse of photons can obtain the image of “damage free” samples. Currently sub-μm systems are actively studied, and one recent example is “live cell imaging”, in which a whole bacterial cell in solution was observed (Nishino, et. al. Nature Comm. 2014). In these experiments, typically a few good diffraction patterns can be used to determine shapes and inside densities of the samples. The measured diffraction intensity is the “absolute square” of the Fourier transform of the electron density (complex numbers). Thus, to reconstruct the real image of the sample through inverse Fourier transform, the phase for each diffraction point needs to be computationally estimated. However, current procedure for phase recovery is not straightforward. If the procedure is not successful, the real image cannot be obtained, or the results are unreliable. Especially, biological systems have weak X-ray diffraction power, and often there are not enough diffraction points to achieve reliable phase recovery.

２．研究の目的

We propose here an alternative hybrid approach that provides an intuitive and discovery driven interface to interpret XFEL diffraction patterns when only a few data are available. The standard approach tries to recover the real image solely from the diffraction data, and it poses difficulty when the diffraction is weak. Here, instead, we pre-compute a large number of hypothetical models and select the models that are in good agreement with the experimental data. We have shown that such an approach can be used to identify conformational transitions of proteins even from very weak diffraction patterns (Tokuhisa et. al., 2016). The goal of this project is to provide an intuitive tool to propose low-resolution models (shapes) that matches XFEL diffraction data, which could be used as an “idea generator” for initial interpretation of XFEL diffraction patterns.

３．研究の方法

In order to develop the proposed tool, we need to perform the following tasks

1. Create database of hypothetical shapes. We need a large collection of shapes that biological systems can possibly have, so that candidate models can be proposed for many experimental data that utilizes the program. Obviously, construction of such a dataset cannot be complete since we do not know the structures of all biological systems, however with sufficient number of shapes, we should be able to provide matches for new data.

A. Data Source - As the initial dataset, we will utilize the structures from electron microscopy. In addition, we will also utilize PDB to generate hypothetical low-resolution structures.

B. Volume Scales – This tool proposes only candidate shapes, and the size of the system is to be estimated from the experimental data. For example, a shape generated from a small protein may match the diffraction from a system much bigger but with the same shape. Thus, in the dataset, all shapes will be normalized to have similar dimensions

C. Redundancy Reduction – To remove redundancy of the data (shapes), all normalized shapes (will be compared in pair-wise manner. We have used Gaussian Mixture Model (GMM) (Kawabata 2008) for a similar purpose.

４．研究成果

1) To test the feasibility of retrieving shapes from a database, we started with 2D images in real space as obtained from cryo-EM. In addition, it could also provide information for EM data as recent works have also shown that cryo-EM experiments can be performed on cell extracts and therefore many single particle images of different molecules are collected. In such cases, analysis of 2D images could still be used to identify biomolecules observed from the cell extracts as well as to infer their structures from limited data. The protocol developed assembles a non-redundant set of 3D shapes for generating a 2D image library, and to retrieve 3D shapes that potentially match 2D experimental images. We tested the strategy using images from three EM models as query images for searches against a library of 22,750 2D projection images generated from 250 random EM models (Figure 1). We found that our ability to identify 3D shapes that match the query images depends on how complex the outline of the 2D shapes are and whether they are represented in the search image library.

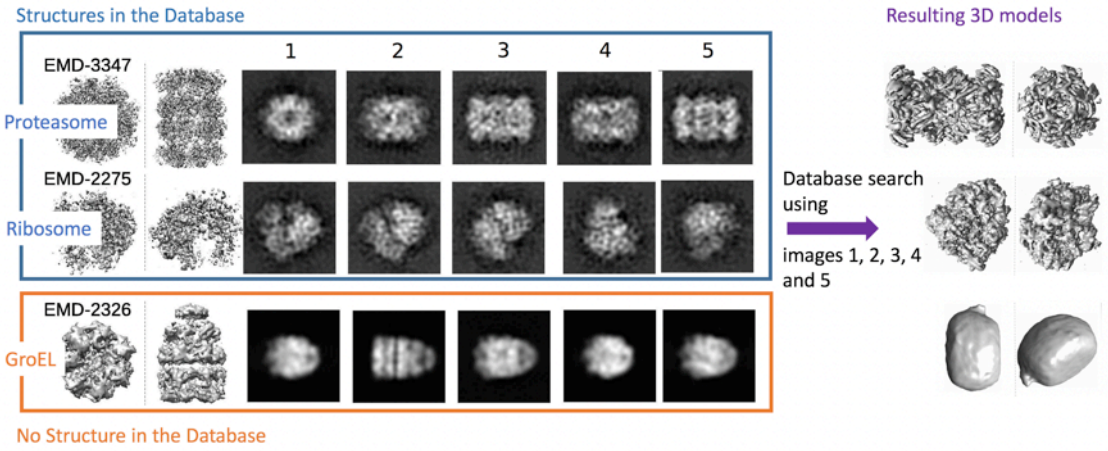

Figure 1: Five random 2D projection images used as input for testing 3D initial model search from EMD-3347, EMD-2275 and EMD-2326. Two views of each EM model are displayed below the model name (left) and the input projection images numbered 1 to 5 are displayed in the same row (right). The resulting top search hits is also shown. In the case of EMD-3347 and EMD-2275, we were able to retrieve the most similar 3D models within the first five hits for each. Although the hits for EMD-2326 are less consistent in their shapes and reflect the fact that there is no true match, retrieved molecular shape captures essential features of the query images.

2) For our database, XFEL diffraction patterns from a large number of models with many incident beam angles have to be calculated which is time-consuming. Therefore, we have explored the utilization of Gaussian mixture model (GMM) as a coarse-grained model for structure modeling from XFEL data. GMM approximates a biomolecular shape by the superposition of Gaussian distributions. As the Fourier transformation of GMM can be quickly performed, we can efficiently simulate XFEL diffraction patterns from approximated structure models. We have shown that the resolution accurately reproduced by GMM is proportional to the cubic root of the number of Gaussians used in the modeling. Furthermore, our study showed that GMMs can successfully be used to identify the orientations of the molecule and to detect conformational variation. These results demonstrate that GMMs serve as useful coarse-grained models for hybrid approach in XFEL single particle experiments.

3) XFEL database: We had first developed the protocol for using EM real space images as inputs and then have been developing algorithms to use XFEL diffraction patterns as inputs. Furthermore, we are improving the algorithms so that it can deal with the diffraction patterns from actual experimental data. Diffraction patterns from XFEL experiments contain limited amount of signals due to the weak diffraction intensity. 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 a numerical algorithm, in which the approximate size of the sample in the input image is estimated via matching against a theoretical model and used to estimate the ROI. Using the estimated ROI, match-finding algorithms to identify plausible candidate 3D models from a few XFEL diffraction patterns has been improved (Figure 2).

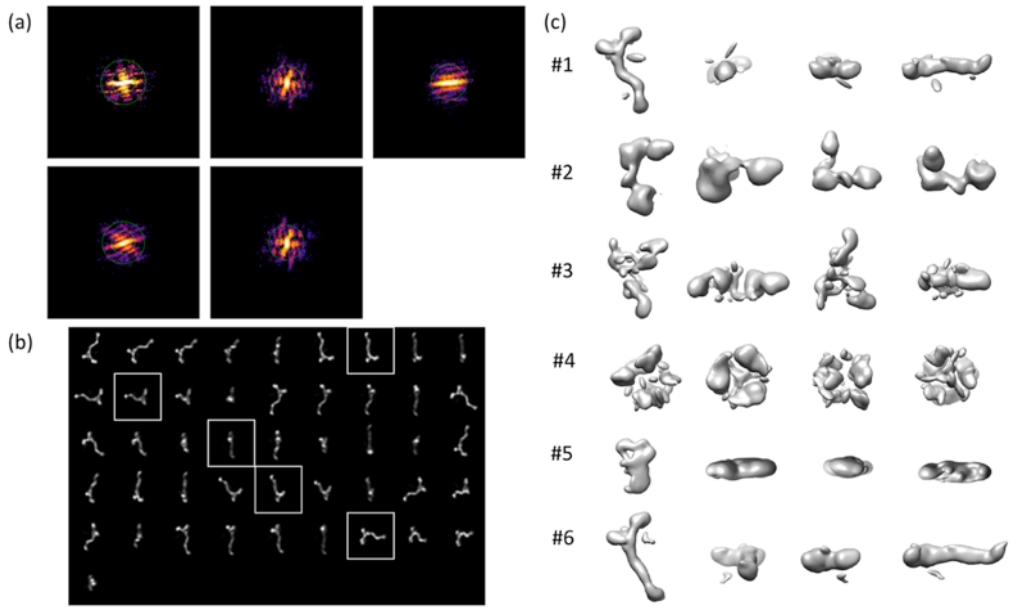

Figure 2: (a) Five simulated diffraction patterns that are used as the query inputs. Two circles on each pattern indicate the definition of Region of Interest that are determined using the newly developed algorithm. (b) Projection images of the "answer" structure. Five input patterns correspond to the input diffraction patterns. (c) Six cryo-EM maps that are identified as the models consistent with the input diffraction patterns (shown in (a)).

５．主な発表論文等

〔雑誌論文〕　計10件（うち査読付論文　10件／うち国際共著　3件／うちオープンアクセス　4件）

１．著者名 Tiwari Sandhya P.、Chhabra Sahil、Tama Florence、Miyashita Osamu	４．巻 -
２．論文標題 Computational Protocol for Assessing the Optimal Pixel Size to Improve the Accuracy of Single-particle Cryo-electron Microscopy Maps	５．発行年 2020年
３．雑誌名 Journal of Chemical Information and Modeling	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jcim.9b01107	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Srivastava Ashutosh、Tiwari Sandhya Premnath、Miyashita Osamu、Tama Florence	４．巻 432
２．論文標題 Integrative/Hybrid Modeling Approaches for Studying Biomolecules	５．発行年 2020年
３．雑誌名 Journal of Molecular Biology	６．最初と最後の頁 2846～2860
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jmb.2020.01.039	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Dasgupta Bhaskar、Miyashita Osamu、Tama Florence	４．巻 1864
２．論文標題 Reconstruction of low-resolution molecular structures from simulated atomic force microscopy images	５．発行年 2020年
３．雑誌名 Biochimica et Biophysica Acta (BBA) - General Subjects	６．最初と最後の頁 129420～129420
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.bbagen.2019.129420	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Srivastava Ashutosh、Nagai Tetsuro、Srivastava Arpita、Miyashita Osamu、Tama Florence	４．巻 19
２．論文標題 Role of Computational Methods in Going beyond X-ray Crystallography to Explore Protein Structure and Dynamics	５．発行年 2018年
３．雑誌名 International Journal of Molecular Sciences	６．最初と最後の頁 3401～3401
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/ijms19113401	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Tiwari SP, Tama F, Miyashita O.	４．巻 12
２．論文標題 Searching for 3D structural models from a library of biological shapes using a few 2D experimental images.	５．発行年 2018年
３．雑誌名 BMC Bioinformatics	６．最初と最後の頁 320
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s12859-018-2358-0	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Mori Takaharu、Kulik Marta、Miyashita Osamu、Jung Jaewoon、Tama Florence、Sugita Yuji	４．巻 27
２．論文標題 Acceleration of cryo-EM Flexible Fitting for Large Biomolecular Systems by Efficient Space Partitioning	５．発行年 2019年
３．雑誌名 Structure	６．最初と最後の頁 161～174.e3
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.str.2018.09.004	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Nakano Miki、Miyashita Osamu、Jonic Slavica、Tokuhisa Atsushi、Tama Florence	４．巻 25
２．論文標題 Single-particle XFEL 3D reconstruction of ribosome-size particles based on Fourier slice matching: requirements to reach subnanometer resolution	５．発行年 2018年
３．雑誌名 Journal of Synchrotron Radiation	６．最初と最後の頁 1010～1021
掲載論文のDOI（デジタルオブジェクト識別子） 10.1107/S1600577518005568	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Nagai Tetsuro、Mochizuki Yuki、Joti Yasumasa、Tama Florence、Miyashita Osamu	４．巻 26
２．論文標題 Gaussian mixture model for coarse-grained modeling from XFEL	５．発行年 2018年
３．雑誌名 Optics Express	６．最初と最後の頁 26734～26734
掲載論文のDOI（デジタルオブジェクト識別子） 10.1364/OE.26.026734	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Ashutosh Srivastava, Tsuyoshi Hirota, Stephan Irle and Florence Tama	4．巻 86
2．論文標題 Conformational dynamics of human protein kinase CK2α and its effect on function and inhibition	5．発行年 2018年
3．雑誌名 Proteins Structure Function and Genetics	6．最初と最後の頁 344-353
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1002/prot.25444	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 宮下 治, タマ フロハンス	4．巻 57
2．論文標題 構造解析へのハイブリッドアプローチ：電子顕微鏡とX線自由電子レーザーからの実験データの活用	5．発行年 2017年
3．雑誌名 生物物理	6．最初と最後の頁 90-94
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.2142/biophys.57.090	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計12件（うち招待講演　12件／うち国際学会　8件）

1．発表者名 Florence Tama
2．発表標題 Hybrid approaches to reveal structure and dynamics of large biological complexes from single molecule experiments
3．学会等名 American Chemical Society National Meeting（招待講演）（国際学会）
4．発表年 2019年

1．発表者名 Florence Tama
2．発表標題 Flexible fitting methods and applications
3．学会等名 EMBO workshop, CEM3DIP 2020: Single particle cryoEM of macromolecular-assemblies and cellular tomography（招待講演）（国際学会）
4．発表年 2020年

１．発表者名
Florence Tama

２．発表標題
Modeling conformational transitions of biomolecules from cryo-EM data

３．学会等名
French Electron Microscopie Society meeting（招待講演）（国際学会）

４．発表年
2019年

１．発表者名
Florence Tama

２．発表標題
Integrative modeling to characterize structure and dynamics of biomolecules from single molecule experiments

３．学会等名
Integrative structural biology meeting（招待講演）（国際学会）

４．発表年
2019年

１．発表者名
Florence Tama

２．発表標題
Hybrid modeling approaches to study structures and dynamics of biological systems.

３．学会等名
63rd Annual Meeting of the Biophysical Society.（招待講演）（国際学会）

４．発表年
2019年

１．発表者名
Florence Tama

２．発表標題
Computational tools to characterize structure and dynamics of biomolecular systems from single molecule experiments.

３．学会等名
American Chemical Society National Meeting（招待講演）（国際学会）

４．発表年
2018年

１．発表者名
Florence Tama

２．発表標題
Hybrid modeling approaches to study structures and dynamics of biological systems

３．学会等名
Second workshop on Advances in Theory and Computation of Complex Systems - Biological Systems. Nanjing, China（招待講演）（国際学会）

４．発表年
2018年

１．発表者名
Florence Tama

２．発表標題
Hybrid modeling approaches to study structures and dynamics of biological systems.

３．学会等名
Coarse-Grained Modeling of Structure and Dynamics of Biomacromolecules Workshop. Telluride USA（招待講演）（国際学会）

４．発表年
2018年

１．発表者名
Florence Tama

２．発表標題
Hybrid modeling approaches to study structures and dynamics of biological systems.

３．学会等名
Annual Meeting of Biophysical Society of Japan（招待講演）

４．発表年
2018年

１．発表者名
Florence Tama

２．発表標題
Computational tools to characterize structure and dynamics of biomolecular systems from single molecule experiments

３．学会等名
Biophysical Society of Japan Meeting（招待講演）

４．発表年
2017年

１．発表者名 Florence Tama
２．発表標題 Computational tools to characterize structure and dynamics of biomolecular systems from single molecule experiments
３．学会等名 Laue Langevin Institute, Grenoble, France（招待講演）
４．発表年 2018年

１．発表者名 Florence Tama
２．発表標題 Computational tools to characterize structure and dynamics of biomolecular systems from single molecule experiments
３．学会等名 Statistics and Algorithms for Biology, INRA, Toulouse, France（招待講演）
４．発表年 2017年

〔図書〕　計1件

１．著者名 O Miyashita, F. Tama	４．発行年 2018年
２．出版社 Springer, Singapore	５．総ページ数 19
３．書名 Hybrid Methods for Macromolecular Modeling by Molecular Mechanics Simulations with Experimental Data. In: Nakamura H., Kleywegt G., Burley S., Markley J. (eds) Integrative Structural Biology with Hybrid Methods. Advances in Experimental Medicine and Biology, vol 1105	

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考