

Project Title:

Protein structure prediction and design

Name: Kam Zhang, Rojan Shrestha, Daiki Terada, Pierre Thevenet

Laboratory at RIKEN:

**Structural Bioinformatics Team,
Center for Life Science Technologies**

1. Background and purpose of the project, relationship of the project with other projects

How the primary sequence of a protein determines its tertiary structure is one of the greatest challenges in computational biology. Advancement in our ability to predict the structures of proteins from their sequences will improve our understanding of protein function in the cell. Our ability to predict protein structure can also benefit the design of new proteins with desired structure and thus new function.

Recent advancement in computational methods for protein structure prediction has made it possible to generate high quality *de novo* models required for *ab initio* phasing of crystallographic diffraction data using molecular replacement. Despite those encouraging achievements in *ab initio* phasing using *de novo* models, its success is limited only to those targets for which high quality *de novo* models can be generated. In order to increase the scope of targets for which the *ab initio* phasing with *de novo* models can be successfully applied, we propose to use known structures of remote homologs as templates to solve new crystal structures.

2. Specific usage status of the system and calculation method

We propose a divide and conquer approach to solve new structures from the fragments of remote homologs. Starting from a protein sequence, sensitive sequence homology search algorithms will be used to identify all distantly related homologs with known 3D structures. Although these

structures share the same fold as the query sequence but they cannot be used directly as template for structure solution using molecular replacement due to large differences between the structures. We will break them into small structural fragments and find their position in the unit cell by molecular replacement. Subsequently, we will assembly these fragments together to form the complete structure.

3. Result

We have developed a fragment assembly phasing method that starts from an ensemble of remote homology models, disassembles them into fragments, places them independently in the crystallographic unit cell by molecular replacement, and then reassembles them into a whole structure that can provide sufficient phase information to enable the complete structure determination by automated model building. Tests on ten protein targets have shown that our method can solve structures for the majority of these targets, although the remote homologs cannot be used as templates for successful molecular replacement since they are far away from the native structure. Our method has extended the applicability of the *ab initio* phasing by *de novo* models approach. Our method can be used to solve structures when only remote homologs are available.

4. Conclusion

Remote homologs may not be close enough to the native structure to enable them to be used as templates for successful molecular replacement.

Usage Report for Fiscal Year 2015

However, these large differences are sometimes caused by rotation or shift between domains or regions. When broken into fragments, it is more likely to find matches that are close enough to the native structure for successful structure solution. The fragmentation and reassembly method has increased the range of remote homologs to be used as templates to solve protein crystal structures by molecular replacement.

5. Schedule and prospect for the future

We will continue to explore new resampling methods for improved de novo structure prediction. The fragmentation and reassembly approach to *ab initio* phasing will be improved to increase its ability to use low quality templates.

6. If no job was executed, specify the reason.

N/A.

Fiscal Year 2015 List of Publications Resulting from the Use of the supercomputer

[Publication]

1. Simoncini, D., Nakata, H., Ogata, K., Nakamura, S., Zhang, K.Y.J. (2015) Quality assessment of predicted protein models using energies calculated by the fragment molecular orbital method. *Molecular Informatics*, **34**, 97-104. doi:10.1002/minf.201400108.
2. Shrestha, R., Zhang, K. Y. J. (2015) A fragmentation and reassembly method for *ab initio* phasing. *Acta Cryst. D71*, 304-312. doi:10.1107/S1399004714025449.

[Oral presentation at an international symposium]

1. The 13th Conference of the Asian Crystallographic Association, Dec. 5-8, 2015, Kolkata, India. Kam Zhang, “RE₃volutionary Design of Symmetric Proteins that Biomineralize Nanocrystals”.
2. JCUP-V, June 4-5, 2015, Tokyo, Japan. Arnout Voet, Hiroki Noguchi, Christine Addy, Kam Zhang, Jeremy Tame, “RE3Volutionary Computational Design of symmetric proteins”.
3. JCUP-V, June 4-5, 2015, Tokyo, Japan. Rojan Shrestha, Kam Zhang, “FRAP: fragmentation and reassembly method for *ab initio* phasing with *de novo* models”.
4. The British Crystallographic Association Spring Meeting, Mar. 30 – Apr. 2, 2015, York, UK. Rojan Shrestha, Kam Zhang, “A fragmentation and reassembly method for the *ab initio* phasing of protein crystallographic data”.