Project Title:

Quality assessment of predicted protein models by fragment molecular orbital quantum energy

Name:Kam Zhang, 中田 浩弥

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

The ability to predict the 3D structure of proteins from their sequences has critical impact on the understanding of their cellular function. We have developed a new method called EdaFold that uses an estimation of distribution algorithm to efficiently sample the conformational space. Our method has been shown to generate more accurate models than the current state-of-the-art method, Rosetta. However, the selection of the best model from a large pool of predicted models is still a very challenging and an unsolved problem. We propose to use the quantum energy calculated by the fragment molecular orbital (FMO) method as a means to select the best model generated by EdaFold. In the FMO method, the protein model is divided into small pieces of fragment, and the total energy is calculated with the embedded electrostatic potentials. After self-consistent electrostatic potential is obtained, fragment pair calculations are performed. Although FMO method is one of the most efficient way of computing quantum energies, the RICC supercomputing resource is necessary to conduct this research.

2. Specific usage status of the system and calculation method

We have chosen a benchmark dataset of eight targets to test whether FMO can facilitate model selection. These targets represent varying degrees of challenge for model selection based on the empirically derived energy functions, such as that of Rosetta energy score. Top scoring models selected by Rosetta score are then subjected to FMO calculations. The ranking of these models according to either Rosetta scores or energies estimated by FMO are compared against TM scores as a measure of the accuracy of each predicted models from the native structure.

The total energies of proteins were calculated by FMO with GAMESS, parallelized with a generalized distributed data interface. The 6-31G(d) basis set was used with spherical harmonics (ISPHER=1) in this study. All the calculations were performed by Hartree-Fock with Grimme's dispersion correction. We used separated dimer approximations, and point charge approximation for ESP in this study. Before calculating the FMO total energies, the position of hydrogen atoms in predicted structures were optimized using AMBER software package. Each fragment consisted of two amino acid residues generated using FragIt program.

3. Result

We found that the model ranking based on FMO energies is better than that based on empirically derived energies when there is sufficient diversity among these models. This model diversity can be estimated prior to the FMO energy calculations. The average pairwise TM score of all versus all protein models gives a rough estimation of the diversity in the data set. The lower the score reveals the higher the diversity. We have computed the gain in Pearson correlation of calculated FMO energy for each target by subtracting the Pearson's R coefficients obtained with both scoring schemes (Rosetta score and FMO energy). The gain is a positive value if the

FMO energy scoring scheme is better, and negative otherwise. We have shown that their Pearson correlation is 0.88 with a p-value of 0.003.

4. Conclusion

Our result demonstrates that the FMO energy calculated by the fragment molecular orbital method is a practical and promising measure for the assessment of protein model quality and the selection of the best protein model among many generated. Our test has also revealed the importance of including the solvation effect into the FMO calculations.

5. Schedule and prospect for the future

We will improve the effectiveness of our method to make it capable of ranking less diverse set of models. The computational time required is very large at the moment. We will reduce the computing time while maintaining the accuracy of model ranking.

.

Fiscal Year 2014 List of Publications Resulting from the Use of RICC

[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*, DOI: 10.1002/minf.201400108.