Usage Report for Fiscal Year 2018

Project Title: Protein Structure Prediction and Design

Name:
Kam Zhang (1), Aditya Padhi (1), Francois Berenger (1)
Laboratory at RIKEN:
(1) Laboratory for Structural Bioinformatics, Center for Biosystems Dynamics Research

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

β-Propeller proteins form one of the largest families of protein structures, with a pseudo-symmetrical fold made up of subdomains called blades. They are not only abundant but are also involved in a wide variety of cellular processes, often by acting as a platform for the assembly of protein complexes. WD40 proteins are -propeller proteins that have limited sequence conservation as a group overall, but that possess a conserved pair of tryptophan and aspartic acid residues within each blade of about 40 amino-acid residues. The uniform structure of most known WD40 proteins suggests that they are an excellent starting point for the design of an artificial ring protein with perfectly conserved internal symmetry. Tandem-repeat proteins are believed to have arisen from an evolutionary process involving duplication and fusion events, creating a perfectly symmetrical intermediate protein which later diversified under evolutionary pressure. We have use a method that reverse-engineers the evolutionary process of gene duplication, fusion and diversification using a computational design procedure to Tako8 and Ika4, which are fourfold or eightfold repeated proteins. The designed structures have been validated by X-ray crystallography and numerous biophysical methods.

2. Specific usage status of the system and calculation method

We started with the crystal structure of the S-phase kinase-associated protein 1A (PDB 2ovp) as the template to create the eight-fold repeated protein, Tako8. The repeat structures were used to construct an eightfold-symmetrical backbone template by utilizing Rosetta symmetric docking with C8 symmetry. Backbone energy minimization and subdomain linking were carried out with MOE. 2000 possible ancestral sequences were predicted by the FastML server. Mapping of the ancestrally reconstructed consensus sequences onto the symmetrical protein backbones was performed using a PyRosetta script, followed by energy minimization and scoring using the Talaris2014 scoring function. The eightfold identical repeat aminoacid sequence was reverse-translated into a DNA sequence with silent restriction sites for further cloning. A four-fold repeat protein called Ika4 was created sing a computational protein-design tool built on top of the recent artificial intelligence prover ToulBar 2. This tool uses a fixed-backbone representation of proteins, and computes the optimal nature and orientation of side-chain rotamers for all possible amino acids. It is able to identify the global minimum energy sequence and conformation of proteins of as large as 100 residues using the energy function defined by Rosetta.

3. Result

The Tako8 and Ika4 proteins expressed at a very high level, were soluble and could readily be purified and concentrated for crystallization and biophysical characterization. Both SEC and AUC indicated a monodisperse monomeric species with the expected molecular weight. The protein crystallized under a variety of conditions. The eightfold symmetry of the Tako8 protein is reflected in the water structure within the central cavity, which is 11Å across. Tako8 was found to have highest stability around pH 8, but to be extremely unstable in the absence of salt. Increasing the salt concentration led to increased stability over the tested pH range. Ika4 was designed as a fourfold-symmetrical protein, allowing

the search algorithm to identify compensating interactions between adjacent blades. SEC and AUC revealed that each Ika4 construct exists as a single eight-bladed propeller in solution, indicating that oligomerization can occur. The Cα RMSD between the designed and experimentally observed Tako8 and Ika4 structures were 1.0Å and 1.7Å respectively, clearly verifying the proteins have folded into the structures as designed. The melting/denaturation curves show that the Ika4 protein are more stable than Tako8, especially under low-salt conditions. Thermal stability measurements by CD, which reflect the loss of secondary structure upon protein unfolding, demonstrated that the **T**m value of Ika4 is roughly 50°C higher than that of Tako8. The higher stability of Ika suggests that the increased design space offered by the lower fourfold symmetry was well exploited by global optimization using the Talaris2014 energy function. The highly symmetric nature of the designer proteins is not only apparent in the crystal symmetry, as discussed above, but can also be observed at the tertiary structural level. The individual repeats of the Tako8 protein can be perfectly aligned. In comparison, only the alternating repeats of the Ika4 protein can be superposed, in contrast to the template structure where the overlap of the individual blades is much poorer.

4. Conclusion

We have successfully designed and structurally characterized a perfectly symmetrical WD40 protein with eightfold symmetry (Tako8). From this template, a derivative fourfold-symmetrical protein (Ika4) was created. The latter was designed using a novel computational procedure and is able to reassemble from individual repeating units. Both proteins are stable and can be purified in high yields, but the altered charge distribution in Ika4 allows it to remain folded under low-salt conditions. Applications have yet to be demonstrated for artificial symmetrical proteins, although the designer lectin Mitsuba has been shown to bind Raji cancer cells selectively. Similarly, the central channel of Tako8 and Ika8 may be able to nucleate catalytic metal clusters, with potential uses in chemistry and medicine. It is hoped that these proteins will prove to be valuable building blocks for various bionanotechnological applications and in evolutionary studies of the WD40 protein family.

5. Schedule and prospect for the future

Protein design is a burgeoning field; only recently we begin to witness the success of creating proteins that do not exist in nature with computational methods. Our ability to design non-natural proteins will not only allow us to explore the universe of protein folds but also enable to create new materials as more efficient enzymes, for drug delivery, and as therapeutics. We plan to use the method of reverse engineering evolution to design more repeat proteins that do not exist in nature to as building blocks to create bionanomaterials to benefit human health and our society.

Usage Report for Fiscal Year 2018

Fiscal Year 2018 List of Publications Resulting from the Use of the supercomputer

[Paper accepted by a journal]

1. Simoncini, D., Zhang, K. Y. J., Schiex, T., Barbe, S. (2019) A Structural Homology Approach for Computational Protein Design with Flexible Backbone. *Bioinformatics*, https://doi.org/10.1093/bioinformatics/bty975.
2. Noguchi, H., Addy, C., Simoncini, D., Wouters, S., Mylemans, B., Van Meervelt, L., Schiex, T., Zhang, K. Y. J., Tame, J. R. H., Voet, A. R. D. (2019) Computational design of symmetrical 8-bladed ß-propeller proteins. *IUCrJ*, 6, 46-55. doi:10.1107/S205225251801480X

[Others (Book, Press release, etc.)]

1. Simoncini, D., Zhang, K. Y. J. (2018) Population Based Sampling and Fragment Based De Novo Protein Structure Prediction. In: Guenther, R. and Steel, D. (eds.), *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1, pp. 774–784. Oxford: Elsevier. https://doi.org/10.1016/B978-0-12-811414-8.20507-5