Usage Report for Fiscal Year 2017

Project Title:

Protein structure prediction and design

Name:

Kam Zhang
Pierre Thevenet
Daiki Terada
Aditya Padhi

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

Computational protein design has advanced very rapidly over the last decade, but there remain few examples of artificial proteins with direct medical applications. MytiLec-1 is a protein isolated from the Mediterranean mussel Mytilus galloprovincialis, and found to bind sugar chains with α-D-galactose at the reducing end. The polypeptide chain has three well-conserved repeats of a roughly 50-residue sequence, and adopts a β-trefoil fold. Recently we have developed a new method of designing symmetric proteins.

We have use a symmetric protein design method and applied it to MytiLec-1, to create a related protein with three identical subdomains, that retains sugar binding activity and the ability to bind selected cell types. MytiLec-1 is strongly stabilized by forming a tight dimer, and mutating the dimerization interface yields unstable monomers9. Symmetrizing the β-trefoil eliminated this interface to create a new monomeric form. We have refined the X-ray crystallographic structure of the symmetrical lectin to high resolution, and show that this artificial protein is significantly more stable than the parent protein, despite the loss of the dimer interface.

2. Specific usage status of the system and calculation method

Crystal structures of MytiLec-1 (both with and without ligands) were previously refined to high resolution, and the structure of the apo-protein (PDB 3WMU) was chosen as the template to create Mitsuba. Backbone models were produced using Rosetta Symmetric Docking, working from the crystal structure of MytiLec-1 (PDB 3WMV). Backbone energy minimization and subdomain linking were carried out with MOE. 2000 possible ancestral sequences were predicted by the FastML server, and mapped onto each symmetrical backbone model with Rosetta. Three different backbone structures were used for modelling with these sequences, one built from the MytiLec-1 subdomain A alone, and two others incorporating either 6 or 9 residues from Threefoil in each subdomain. The backbone using 6 Threefoil residues gave models with the best energy scores, including Mitsuba-1, the overall top scoring solution.

3. Result

A DNA coding sequence was designed for each chosen protein by back-translating with an in-house program. Codon usage was optimized for expression in E. coli and the synthesized genes were inserted into the standard expression vector pET28, allowing the protein to be expressed and purified using a thrombin cleavable histidine tag. Analytical ultracentrifugation (AUC) shows that Mitsuba-1 is a

monomer in solution, a result confirmed by size-exclusion column chromatography. Circular dichroism indicated that the protein adopted a stable fold, rich in ß structure, allowing the melting temperature to be determined to be 55 °C. Unfolding experiments were also carried out by observing the change in circular dichroism at 228 nm with added increments of guanidinium hydrochloride. The complex of Mitsuba-1 and NAcGal was crystallized, and a 1.54 Å resolution dataset was collected. Preliminary phases could be determined directly by molecular replacement with the designed model. One complete monomer is found in the asymmetric unit, with no breaks in the 2mFo-DFc electron density map of the chain, and there are no Ramachandran outliers. A histidine residue, from the linker peptide after cleavage of the His-tag, is visibly attached to the methionine residue at the N-terminus of the designed sequence. The structure is well ordered with each subdomain forming a β-sheet of four strands. Overlaying 143 Cα atoms of Mitsuba-1 in the crystal structure and the designed model gives an RMSD value of 0.86 Å, showing the atomic level accuracy of our design.

4. Conclusion

This study describes a new artificial β-trefoil lectin that recognizes Burkitt's lymphoma cells, and which was designed with the intention of finding a basis for novel cancer treatments or diagnostics. The new protein, called "Mitsuba", is based on the structure of the natural shellfish lectin MytiLec-1, a member of a small lectin family that uses unique sequence motifs to bind α-D-galactose. The three subdomains of MytiLec-1 each carry one galactose binding site, and the 149-residue protein forms a tight dimer in solution. Mitsuba was created by symmetry constraining the structure of a MytiLec-1 subunit, resulting in a 150-residue sequence that contains three identical tandem repeats. Mitsuba-1 was expressed and crystallized to confirm the X-ray structure matches the predicted model. Mitsuba-1 recognizes cancer cells that express globotriose (Galα(1,4)Galβ(1,4)Glc) on the surface, but the cytotoxicity is abolished.

5. Schedule and prospect for the future

The successful design of Mitsuba-1 has demonstrated again the power of symmetric protein design method based on ancestral reconstruction. The ability of creating a symmetric protein that retains cancer cell surface polysaccharide recognition and binding without cytotoxicity is very encouraging. In the future, we plan to take advantage of the decoupling of cancer cell recognition and cytotoxicity to design magic bullet anti-cancer medicine by engineering a chimera of Mitsuba-1 and cytotoxic agent, such as ricin to be used to kill cancer cells. Moreover, we plan to future improve the polysaccharide binding affinity of Mitsuba-1, from the current micromolar level to that of nanomolar level, using a computational protein design approach.

Fiscal Year 2017 List of Publications Resulting from the Use of the supercomputer

[Publication]

1. Berenger, F., Simoncini, D., Voet, A., Shrestha, R., ***Zhang, K. Y. J.** (2017) Fragger: a protein fragment picker for structural queries. *F1000Research*, **6**, 1722. doi:10.12688/f1000research.12486.1.
2. Terada, D., Voet, A. R. D., Noguchi, H., Kamata, K., Ohki, M., Addy, C., Fujii, Y., Yamamoto, D., Ozeki, Y., *Tame, J. R. H., ****Zhang, K. Y. J.** (2017) Computational design of a symmetrical ß-trefoil lectin with cancer cell binding activity. *Sci. Rep.*, **7**, 5943. DOI:10.1038/s41598-017-06332-7.

[Others (Press release, Science lecture for the public)]

腫瘍細胞の糖鎖と結合する人工レクチンを計算機科学で設計

1. 日本経済新聞プレスリリース 2017 年 7 月 27 日
2. 科学新聞 2017 年 8 月 3 日