BL-1A, AR-NE3A/2015G085

Rational Design of Crystal Contact-Free Space in Protein Crystals for Analyzing Spatial Distribution of Motions within Protein Molecules

Rei Matsuoka, Atsushi Shimada and Daisuke Kohda*
Medical Institute of Bioregulation, Kyushu University, Higashi-ku, Fukuoka 812-8582, Japan

A dynamical view of protein molecules is essential for understanding the biological functions. However, in protein crystals, the crystal contacts with neighboring protein molecules block such motions. We propose the rational design method of a fusion protein, to create 'crystal contact-free space' (CCFS) in protein crystals. The created CCFS can be used to analyze the spatial distribution of the motions of a segment/ligand that is intentionally placed in the CCFS by careful fusion-protein design.

1 Introduction

Contacts with neighboring molecules in protein crystals inevitably restrict the internal motions of intrinsically dynamic proteins. The resultant clear electron densities due to the crystal contacts permit model building as crystallographic snapshot structures. Although these still images are informative, they could provide biased pictures of the protein motions. If the mobile parts are located at a site lacking direct contacts, then the amplitude of the movements can be experimentally analyzed. Here, we propose a fusion protein method, to create crystal contact-free space (CCFS) in protein crystals and to place the mobile parts in the CCFS. Conventional model building will fail due to the very weak electron densities when the mobile parts have large amplitude motions. We used suitable data processing of the X-ray diffraction data to make the mobile parts in the CCFS appear as smeared electron densities in the Fo-Fc difference Fourier electron density map.

2 Design of Fusion Proteins

The basic idea is the fusion of a target protein with a tag protein, via a rigid linker (Fig. 1A). The rigid linker is used to create isolated space within the framework formed by the rigid fusion protein. We refer to the fusion protein (except for the flexible segment/ligand) as the 'CCFS scaffold'. Since the crystal packing mode of protein molecules is uncontrollable, the CCFS scaffold must be designed to ensure the formation of a space within a protein molecule that is inaccessible to other molecules.

We selected *Escherichia coli* maltose binding protein (MBP) as the fusion partner of the Tom20 protein. Tom20 resides in the mitochondrial outer membrane, where it functions as a receptor for presequences (mitochondrial signal sequences) for efficient import of mitochondrial matrix and inner membrane proteins into mitochondria. The analysis of the binding mode of a presequence to Tom20 is particularly suitable as a test case, because the large amplitude motions of a presequence in the bound state are considered to be the structural basis for the promiscuous recognition of mitochondrial presequences. We focused on the long C-terminal α-helix of MBP, which was seamlessly fused to the N-terminal α-helix in the cytosolic domain of Tom20 to form a rigid helical connector structure. To place the presequence binding site in the CCFS, the number of amino acid residues inserted or deleted at the junction site in the connector helix provides one degree of freedom, which determines the relative orientation between MBP and Tom20 within the same molecule. In the cases of ligands with weak affinity, the problem of partial ligand occupancy must be considered. The pALDH presequence peptide was tethered onto Tom20, to ensure the full occupancy of the presequence in the binding site (Fig. 1B). We added a cysteine residue at the C-terminus of pALDH, to form an intermolecular disulfide bond with the single cysteine residue in the fusion protein. The spacer length (3 residues) between pALDH and the C-terminal cysteine was optimized in the previous peptide library experiment. After several round of protein design and diffraction experiments, we created a fusion protein referred to as Δ5MBP<+4>Tom20-SS-pALDH, where Δ5, <+4>, and SS denote a 5-residue deletion in MBP, 4-residue spacer bewteen MBP and Tom20, and intermolecular disulfide bond for tethering, respectively.

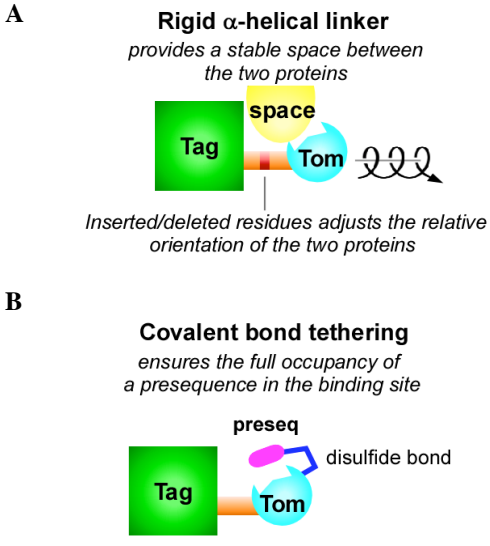

Fig. 1: Fusion protein with a rigid α-helical connection.

3 Special Data Processing for Visulalization

The crystal of Δ 5MBP<+4>Tom20-SS-pALDH diffracted to a resolution of 1.8 Å. The structure of the CCFS scaffold, Δ 5MBP<+4>Tom20, was determined by the molecular replacement method. We did not build a

model for the pALDH presequence, to avoid model bias. To locate the presequence peptide in the binding site, we used the Fo-Fc difference Fourier electron density map. However, we did not see any significant electron densities in the binding site of Tom20 in the difference map (Fig. 2). The invisibility suggested the large mobility of the presequence peptide in the binding site, beyond the limits of conventional treatment.

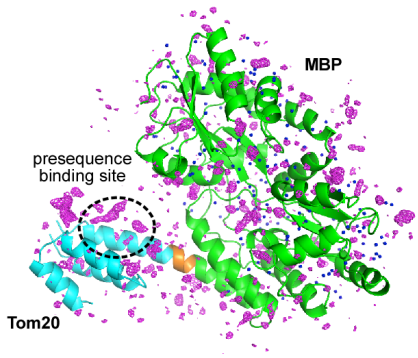

Fig. 2: Conventional Fo-Fc difference electron density map, contoured at +3σ.

We then examined various data-processing techniques, to determine whether they would be effective to find the location of mobile ligands/segments. We used map blurring for the high-resolution reflections for noise reduction. Here, we adopted a simple form of map blurring, using the truncation (to be precise, zero padding) of the high-resolution reflections. The improvement in the signal-to-noise ratio depends on the choice of the resolution limit (r_{min}) of the truncation for map generation. We found that r_{min} = 7 Å was optimal to locate the α-helical structure of the presequence peptide.

The use of a "free" test data set for cross-validation caused the quality of the electron densities of the mobile atoms to deteriorate. The severity of the deterioration depended on the choice of the free test set. We adopted the averaging of maps generated with different free test sets, in a procedure referred to as "FreeR averaging" in this study. For a free test set assigned to 5 % of the structure factors, twenty maps were calculated and averaged. The specially processed difference map, with the truncation and FreeR averaging, clearly revealed an elongated electron density in the binding site of Tom20 (Fig. 3A), even though the same diffraction data set shown in Fig. 2 was used.

4 Mutagenesis of pALDH for Validation

We performed a mutagenesis study to validate the electron density in the CCFS. For each mutation, one of the three hydrophobic leucine residues in the presequence was substituted with a hydrophilic serine residue. These substitutions resulted in affinity reductions ranging from 40- to 200-fold. The crystals of the three mutant complexes diffracted to resolutions of 2.1 Å, 1.8 Å, and 1.6 Å. Although the space group and the number of molecules in the asymmetric unit changed, the CCFS invariably formed. As expected, the electron density corresponding to the presequence part almost disappeared, but that of the tether part remained visible in the binding site (Fig. 3B). Even though the mutated peptides were still located in the CCFS, without the interactions with Tom20, their large motions were beyond the detection limit of the truncated difference map.

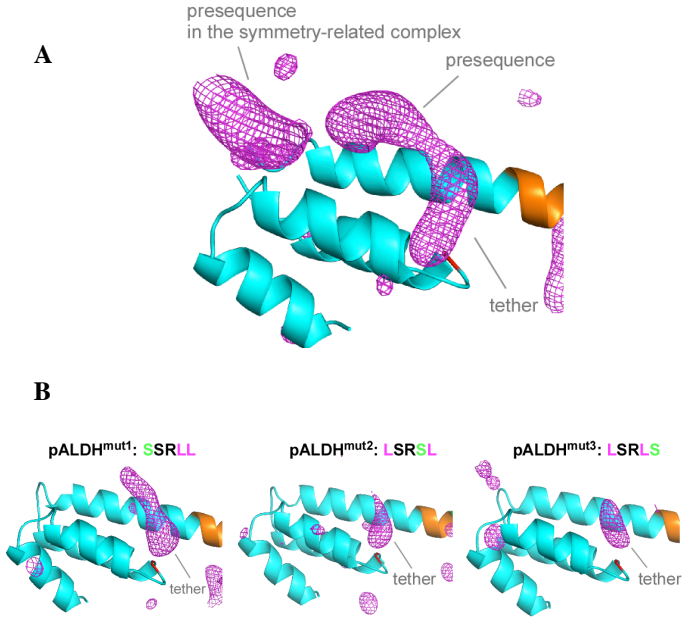

Fig. 3: Truncated and FreeR-averaged difference maps, contoured at +3σ. WT pALDH (*A*) and three mutated pALDHs (*B*).

5 Conclusion

We have proposed a fusion protein method for the creation of crystal contact-free space (CCFS) in protein crystals, independent of the protein molecule packing mode in crystals (Fig. 1). We selected MBP as a fusion partner among many tag proteins to use its long C-terminal α-helix. The concept of CCFS is not limited to the analysis of protein dynamics. Other potential applications include correcting the distorted protein conformations induced by the crystallographic contact effects, allowing ligands to soak into otherwise occluded sites in protein crystals, and studying the time-resolved large motions that are normally constrained in conventional crystal lattices.

Acknowledgement

We thank Drs. Yasuaki Komuro and Yuji Sugita (RIKEN Wako) for the MD simulation of the Tom20-pALDH complex in solution. We also thank Dr. Shunsuke Matsumoto (Kyoto Sangyo University) for the plasmid constructions and pilot expression experiments.

References

[1] R. Matsuoka, A. Shimada, Y. Komuro, Y. Suigta, D. Kohda, *Protein Sci*. **25**, 754-768 (2016).

* kohda@bioreg.kyushu-u.ac.jp