CRYSTAL CONTACT/FREE CONFORMATION OF AN INTRINSICALLY FLEXIBLE LOOP IN PROTEIN CRYSTAL: TIM21 AS THE CASE STUDY

斯琴巴拉

CRYSTAL CONTACT-FREE CONFORMATION OF AN INTRINSICALLY FLEXIBLE LOOP IN PROTEIN CRYSTAL: TIM21 AS THE CASE STUDY

by

SIQINBALA

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY (Ph.D.) IN SYSTEM LIEF SCIENCES

Graduate School of Systems Life Sciences

KYUSHU UNIVERSITY
Fukuoka, Japan

Feb, 2020

Supervised by:

Professor Daisuke Kohda

Table of Contents

Acknowledgments

Upon the completion of this thesis, I am grateful to those who have offered me encouragement and support during the course of my study. Firstly, I would like to express my sincere gratitude to my supervisor, Prof. Daisuke Kohda, whose patient instruction and constructive suggestions are beneficial to me a lot. I also thank the associate professor, Dr. Atsushi Shimada, for teaching me the data collection at the synchrotrons as well as the analysis of X-ray diffraction data of proteins. I thank Dr. Daisuke Fujinami, Dr. Yuya Taguchi, Mr. Takahiro Yamasaki, and Mr. Hajime Motomura for technical supports in the experiments and computer works. Also, I thank Miss Ishikawa for the construction of the MBP-Tim21 expression plasmids. I would like to thank all lab members, especially Secretary Mrs. Otsu, Dr. Taguchi, and Mr. Yamasaki, whose work of laboratory management gave us a secure and supportive environment for research. I am very grateful to Drs. Shoko Shinya, Naohiro Kobayashi, and Chojiro Kojima for the NMR analysis, and Mr. Arpita Srivastava, and Drs. Florence Tama and Osamu Miyashita for the MD simulation. The NMR data and MD simulation data gave us strong supports for the results obtained by the CCFS crystallography. I would like to thank the Technical Support of Medical Institute of Bioregulation, Kyushu University for DNA sequencing. I thank the staff at the Photon Factory and SPring-8 for helping me to use the beamlines for the collection of X-ray diffraction data. The experiments at the Photon Factory were performed with the approval of the Photon Factory Program Advisory Committee, as Proposal 2017G009, and those at SPring-8 were performed under the Cooperative Research Program of the Institute for Protein Research, Osaka University, Osaka, Japan, as Proposals 20166617, 20176718, and 20186815. Finally, I would like to express the most heartfelt gratitude to my family members: my parents, whenever I contacted them, always asked me when I could get the Ph.D. degree, thanks for giving me so much love and care; my husband, who encouraged me in my every depression periods and gave me not just the good advice in life but also the spiritual support; and my sons, whose love is always the source of my strength.

Abbreviations used in this thesis:

CCFS, crystal-contact free space;

NMR, nuclear magnetic resonance;

NOE, nuclear Overhauser effect;

NOESY, nuclear Overhauser effect spectroscopy;

RDC, residual dipolar coupling;

2006_crystal, the crystal structure of the intermembrane space fragment of yeast Tim21 (PDB 2CIU);

2014_NMR, NMR structure of the intermembrane space fragment of yeast Tim21 (PDB 2MF7);

2019_NMR, new NMR structure of the intermembrane space fragment of yeast Tim21 determined in the present study;

MBP, maltose-binding protein;

MD, molecular dynamics;

MDS, multidimensional scaling;

RMS, root-mean-square;

RMSD, root-mean-square distance;

Tim21, Translocase of Inner Mitochondrial membrane 21, in this article, also indicates intermembrane space domain of Tim21.

Abstract

In protein crystals, flexible loops are frequently deformed by crystal contacts, whereas in solution, the large motions result in the poor convergence of such flexible loops in NMR structure determinations. We need an experimental technique to characterize the structural and dynamic properties of intrinsically flexible loops of protein molecules. We designed an intended crystal contact-free space (CCFS) in protein crystals and arranged the flexible loop of interest in the CCFS. The yeast Tim 21 protein was chosen as the model protein because one of the loops (loop 2) is distorted by crystal contacts in the conventional crystal. Yeast Tim21 was fused to the MBP protein by a rigid α-helical linker. The space created between the two proteins was used as the CCFS. The linker length provides adjustable freedom to arrange loop 2 in the CCFS. We re-determined the NMR structure of yeast Tim21 and conducted MD simulations for comparison. Multidimensional scaling was used to visualize the conformational similarity of loop 2. We found that the crystal contact-free conformation of loop 2 is located close to the center of the ensembles of the loop 2 conformations in the NMR and MD structures. We concluded that loop 2 of yeast Tim21 in the CCFS adopts a representative, dominant conformation in solution. No single powerful technique is available for the characterization of flexible structures in protein molecules. NMR analyses and MD simulations provide useful, but incomplete information. CCFS crystallography offers a third route to this goal.

Chapter 1. Introduction

1.1. The conformation of flexible loops in a protein

Elucidation of protein structure is essential to unravel the protein's physiological functions at the molecular level. It is generally believed that the native structure of protein is spontaneously at the global minimum of free energy. But, in response to the alteration of the surrounding environment, or through the interactions with other protein/molecules, the structures of proteins are changeable [1-3].

The interaction of proteins with other molecules requires the structural flexibility. The changes of form/state of proteins are accompanied with the structural changes. In many cases, it is the flexible loop that is directly involved in the recognition and interactions with other molecules [4, 5]. The correct structural information on the flexible loops of a protein is critical for the analysis of protein functions and the elucidation of interaction mechanisms.

1.2. 3D structure determination methods

1.2.1. X-ray crystallography

X-ray crystallography is a discipline that uses X-rays to study the arrangement of atoms in a crystal. More precisely, using the scattering effects of electrons on an X-ray beam, X-ray crystallography can determine the distribution of electron density in the crystal and thus provide the position information of the atoms, that is, the crystal structure. The X-rays used for the 3D structure determination of molecules are approximately 0.5 to 1.5 angstroms long, which are just the right size to measure the distance between atoms in a molecule.

The general procedure for the 3D protein structure determination using X-ray crystallography is as follows. Growing high-quality crystals of a purified protein, measuring the directions and intensities of X-ray beams diffracted by the crystal(s) while rotating at different orientations, Fourier transformation to convert the diffracted intensities with estimated phases to a three-dimensional electron density within the asymmetric unit of the crystal, and building a molecular model consistent with the electron density (Fig. 1-1).

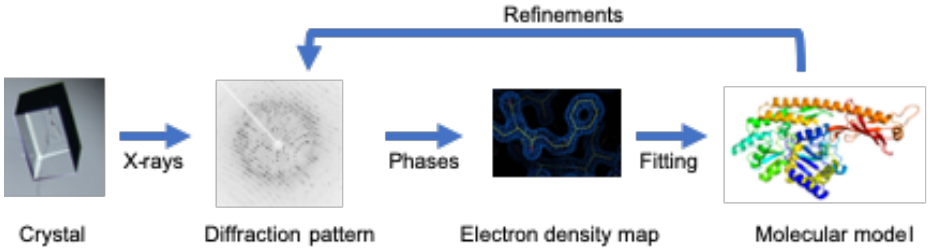


Fig.1-1. General procedure of the 3D structure determination of proteins with X-ray crystallography.

Among several techniques, X-ray crystallography is the most reliable and hence popular method for protein structure determination at atomic resolutions [6]. 88% of the PDB entries released in 2018 were X-ray crystal structures [7]. In general, the overall structure of a protein molecule determined by X-ray crystallography is assumed to be identical to that in solution. This long-established empirical fact is now widely accepted, based on numerous comparisons between structures in crystals and in solution [8-14]. Intermolecular contacts are necessary for crystal formation, but the interactions are not always biologically relevant. In particular, a flexible loop located on the protein surface may adopt a distorted conformation caused by the crystal contacts, in some cases [15-18]. Even if the conformation of such a flexible loop in the crystal is very close to that in solution, the contacts with neighboring molecules effectively trap the flexible loop into one possible conformation, like a snapshot. Thus, the true picture of a flexible loop in a protein molecule cannot be obtained by conventional X-ray crystallography.

1.2.2. Nuclear magnetic resonance (NMR) spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is the method that can determine the 3D structures of protein molecules at the atomic resolution in aqueous solution. NMR is a method to investigate electronic environments around nuclei. The atomic nuclei, such as ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P that have net nuclear spins, adopt preferred orientations in a magnetic field. The energy transfer takes place at pulses of electromagnetic radiation in the radio-frequency (RF) range and when the spin returns to its basal level, energy is emitted at the same frequency. The signal that matches this transfer is measured and processed to yield an NMR spectrum for the nucleus concerned.

NMR spectroscopy provides geometric restraints, which include the internuclear distance restraints from nuclear Overhauser effect (NOE), the local angular restraints from scalar coupling, and the global orientational restraints from residual dipolar coupling (RDC). Nuclear Overhauser effect spectroscopy (NOESY) is a through-space correlation NMR spectrum. It establishes correlations between two nuclei that are physically close to each other in space, regardless of a covalent bond between them. It uses the nuclear Overhauser effect by which nearby atoms (within about 5 Å) undergo cross-relaxation by a mechanism related to the spin-lattice relaxation. If a sufficient number of distance constraints are collected, the three-dimensional structure of a protein can be determined uniquely.

NMR spectroscopy is generally considered to be the best method for the analysis of solution structures, but the large motions of flexible loops result in the insufficient acquisition of the restraints derived from NOEs, and hence result in the poor convergence of the loop structures [19]. It is still possible to calculate an "averaged" structure, but the highly non-linear distance dependency of NMR information (e.g., NOE is inversely proportional to the sixth power of the inter-proton distance) distorts such average structures.

1.2.3. Molecular dynamics (MD) simulation

Molecular dynamics is a computer simulation method for studying the physical movements of atoms in a molecule. MD simulation can provide information about mobility or flexibility of various regions of a molecule and also can be used to test the accuracy of a structure model [20]. However, MD simulation techniques are still inadequate for the reliable and precise prediction of the dynamic behaviors of a flexible segment [21].

1.3. Crystal contact-free space (CCFS) crystallography

In the three methodologies, X-ray crystallography, NMR spectroscopy, and MD simulation, each approach has its own limitations in studying the dynamic behaviors of the flexible loops in solution as mentioned above. Although single-particle cryo-electron microscopy has explosively been developed in recent years, the visualization of the intrinsically disordered or flexible regions in a target protein is still the major limitation of this technique [22]. The resolutions of other solution methods,

such as X-ray small-angle scattering and fluorescence resonance energy transfer (FRET), are inferior, in comparison with the spatial resolution of solution NMR. In summary, it is a fundamental challenge to obtain the conformations of intrinsically flexible loops of proteins in solution. Another methodology based on a different physicochemical basis is needed, to complement the existing solution methods and MD simulations.

The electron density obtained by X-ray crystallography is proportional to the atom occupancy. This fact allows the distortionless summation of the electron densities of individual loop conformations, even if large motions occur. In this respect, X-ray crystallography is superior to solution NMR. If a flexible loop is fortuitously located at a site lacking direct contacts in protein crystals, then the electron density becomes somewhat faint due to dynamic or static disorder in the crystal, but useful information about the solution conformation can be obtained.

In 2016, our laboratory proposed the rational design of "crystal contact-free space" (CCFS) in protein crystals, and the arrangement of a flexible segment of a protein in the CCFS in an intended fashion [23]. The CCFS method consists of two key techniques. First, a target protein was fused to a tag protein. The space created between the tag protein and the target protein in the same molecule is used as the CCFS. Second, a rigid α-helical linker was used to connect the tag and target proteins firmly, to ensure that space exists between them. Since the crystal packing mode (*i.e.*, space group and cell dimensions) cannot be controlled, the CCFS must be created within a molecule, using the rigid framework of the fusion protein. The length of the α-helical linker inserted between the tag and target proteins determines the distance and rotation of the target protein relative to the tag protein. This single degree of freedom permits the rational design of the CCFS within the fusion protein molecule and the proper positioning of the flexible loop of the target protein in the CCFS.

Within the CCFS, the undisturbed conformation of a flexible segment was expected to be observed. Also, the amplitude of the motions can be estimated as large B-factor values (alias, temperature factors or atomic displacement factors), if the range of motion from the mean position is up to 1.5 Å [23]. A smeared of faint electron density prevents modeling, and it even disappears if the range of motion exceeds 1.5 Å.

In our study in 2016, the CCFS method was applied to a highly mobile α-helical presequence peptide bound to the mitochondrial import receptor, Tom20, for the analysis of the spatial distribution of motions of the presequence peptide in the bound state [23]. In this study, we planned to test the applicability of the CCFS method to the analysis of the intrinsic conformation of flexible loops in protein molecules.

1.4. Tim21

The Tim21 protein of *Saccharomyces cerevisiae* is a 21-kDa subunit of the presequence translocase of the inner mitochondrial membrane, the TIM23 complex [24]. The TIM23 complex accepts preproteins from the outer membrane TOM complex and mediates the insertion into or transfer across the mitochondrial inner membrane. In recent researches, some outer membrane proteins are processed by the TIM23 complex [25, 26]. Tim21 consists of a single transmembrane segment in the inner membrane, and an intermembrane space domain [24, 27]. The intermembrane space domain of Tim21 interacts with the TOM complex via the Tom22 [27].

The three-dimensional structures of the intermembrane space fragment of *Saccharomyces cerevisiae* Tim21 were determined by X-ray crystallography (PDB 2CIU) [28] and solution NMR (PDB 2MF7) [29]. Hereafter, for simplicity, we refer to the intermembrane space fragment of the yeast Tim21 as just Tim21. As expected, the overall Tim21 structures determined by X-ray crystallography and NMR are almost identical to each other, except that the conformations of loop 2 are quite different (Fig. 1-2). In the crystal structure, loop 2 adopts a β-hairpin conformation with low B-factors, suggesting that the atoms in loop 2 are in fixed positions in the crystal. In contrast, loop 2 in the NMR structure ensemble has little convergence in the atomic positions, suggesting the substantial mobility of loop 2 in solution. The structural difference is attributable to the extensive crystal contacts around loop 2, with three symmetrically related molecules (Fig. 1-2C). Thus, the arrangement of loop 2 in CCFS offers a suitable test case to study the adequacy of the CCFS method.

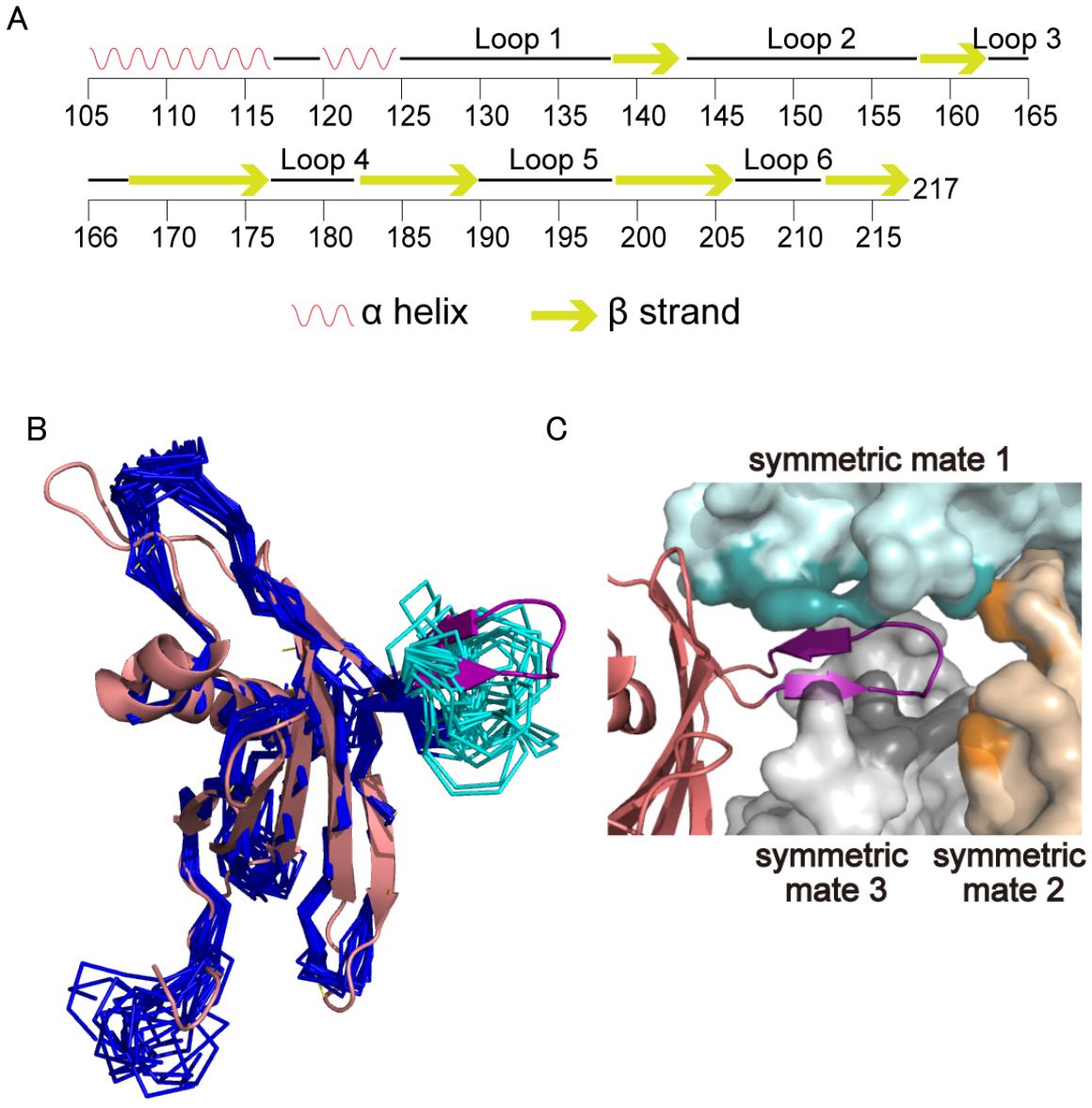


Fig.1- 2. Summary of the yeast Tim21 structures in the previous crystallographic and NMR studies.

A. Secondary structure elements of the intermembrane space fragment (D105-K217) of yeast Tim21.

B. Superimposition of the previous X-ray structure (salmon, PDB 2CIU, referred to as 2006_crystal) and the previous NMR structure (blue, PDB 2MF7, referred to as 2014_NMR). Loop 2 in the 2014_NMR structure is highlighted in cyan and in the 2006_crystal is highlighted in purple

C. Loop 2 has many crystal contacts with three symmetrically related molecules in the 2006_crystal structure. The residues that contact loop 2 in the symmetrically related molecules are highlighted in darker surface colors.

1.5. Summary of research ideas and methods

The structural information of flexible loops is important for the analysis of the mechanism of protein functions. The traditional methodologies of three-dimensional structure determination at the atomic level, X-ray crystallography and NMR, are not perfect for the conformation determination of intrinsically flexible loops (Fig. 1-3). MD simulation just provides support information for the experimental data. It is the CCFS crystallographic method that can provide a solution for this problem (Fig. 1-3). We selected one of the flexible loops of Tim21, whose conformation is distorted due to the crystal contacts in the traditional X-ray crystallography [28], as the target loop to test the applicability of the CCFS method for the conformational determination of intrinsically flexible loops.

For this purpose, we fused Tim21 to the *Escherichia coli* maltose-binding protein (MBP). In the created CCFS, the electron density of loop 2 did not adopt a β-hairpin structure. Contrary to our expectations, the electron density in CCFS was inconsistent with the loop 2 conformations in the previously reported NMR structure [29]. To address this difficult situation, we measured the NMR spectra and recalculated a new NMR structure of Tim21. We found that the conformations of loop 2 in the two NOE-based NMR structures are quite different from each other, even though the core structures are almost identical. In combination with the results from the MD simulation of Tim21 in solution [30], we validated the idea that the crystal contact-free conformation of loop 2 can be regarded as the intrinsic conformation in solution. We also discuss the merits and limitations of the CCFS crystallographic method.

For convenience, in this article, we refer to the conventional crystal structure (2CIU) and the previous NMR structure (2MF7) of Tim21 as 2006_crystal and 2014_NMR, respectively. We also refer to our new NMR structure of Tim21 as 2019_NMR.

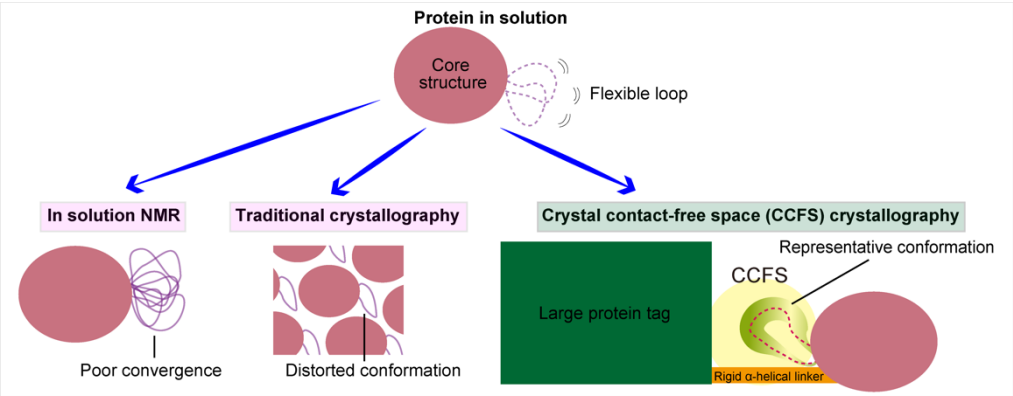


Fig. 1-3. The summary of research ideas and methods.

Chapter 2. Design of CCFS

2.1. Tim21 and MBP for the MBP-Tim21 fusion proteins

2.1.1. Sequence selection from Tim21

The intermembrane space (IMS) sequence from D105 to K217 residues of Tim21 of *Saccharomyces cerevisiae* (strain ATCC 204508 / S288c) was selected as the primary structure of the target Tim21 protein. The N-terminal α-helix is necessary for the creation of CCFS. Although S103 and G104 are in intermembrane space, S103 is not in the α-helical structure and G104 is at the edge of the α-helical structure in the 2006_crystal structure, therefore, we selected the Tim21 sequence starting from D105. The C-terminal sequence of P_{218}KLHPVSNSKGFLGIRWGPRKD$_{239}$ did not form any secondary structures in the 2006_crystal structure, so this fragment was also deleted. We saved the coordinates of the segment of D105 to K217 of 2006_crystal as Tim21_delN.pdb in PyMOL.

MSSSLPRSLLRLGHRKPLFPRYNTFVNSSVITHTSLLRTRLYSNGTGATSGKK
DDKTRNKPKPLWPQVKSASTFTFSGILVIGAVGISAIVIYLILSELFSPSGD_{105}TQL
FNRAVSMVEKNKDIRSLLQCDDGITGKERLKAYGELITNDKWTRNRPIVSTKKL
DKEGRTHHYMRFHVESKKKIALVHLEAKESKQNYQPDFINMYVDVPGEKRYY
LIK_{217}PKLHPVSNSKGFLGIRWGPRKD

The signal sequence of Tim21 is in blue, the transmembrane sequence is in purple, the intermembrane space sequence is in orange.

2.1.2. Sequence selection from MBP

The C-terminal α-helix of MBP is planned to be used as a part of the rigid connection with Tim21. The final two residues of C-terminus, T369K370, are not in α-helical structure, so the sequence from K1 to I368 of the maltose-binding protein sequence from *Escherichia coli* (strain K12) was selected as our tag protein, and we saved the coordinates of the selected part of 4MBP structure as 4MBP_delC.pdb in PyMOL.

K_1IEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVA
ATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLI

AYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTW
PLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDY
SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLS
AGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRI
AATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRI$_{368}$
TK

2.2. Design of α-helical linkers

Firstly, an α-helix was created in PyMOL with the procedure: Selecting menu of Build>Residue>Helix; pressing the buttons of "OPTION" and "A" on the keyboard at the same time for 41 times to create an α-helix comprising 41 Alanines with the sequence from A2-A42; and saving the created α-helix as a file named ala.pdb. Secondly, the created α-helix was superimposed with MBP (4MBP_delC.pdb) using the command of *pair_fit ala///2-8/CA, 4MBP_delC///362-368/CA* to align the 7 residues of the C-terminal α-helix of 4MBP_delC and N-terminal residues of the α-helical linker to form pseudo-MBP_α-helical linker molecule. RMSD = 0.333 Å (7 atoms). Then, the Tim21_delN was superimposed with MBP_α-helical linker using the command of p*air_fit Tim21_delN//A/105-111/CA, ala///11-17/CA* to align the 7 residues of the N-terminal α-helix of Tim21_delN and α-helical linker, thus, the model of MBP-Tim21 fusion protein with a ten-residue linker between MBP and Tim21 was formed. By adjusting the superimposing positions of residues in α-helix, the length of the linker can be adjusted, to control the position of loop 2 of Tim21 relative to MBP, and to place the loop 2 in the crystal contact-free space created between MBP and Tim21.

Finally, the four lengths of α-helical linkers that could locate the loop 2 of Tim21 in CCFS were selected by the modeling in PyMOL: 16-residue linker, 17 residue-linker, 20 residue-linker, and 24 residues linker. The amino acid residues were alanine-based sequences according to the method of Susan Marousee [31]:

16 residues linker: EAAAKEAAAKEAAAKA
17 residues linker: EAAAKEAAAKEAAAKAA
20 residues linker: EAAAKEAAAKEAAAKEAAAK
24 residues linker: EAAAKEAAAKAEAAAKAEAAAKAA

2.3. Design of CCFS in the MBP-Tim21 crystals

E. coli MBP (K1-I368) was fused to yeast Tim21 (D105-K217), as a large tag protein. The two proteins are rigidly connected by a long α-helix, which consist of the C-terminal α-helix of the MBP protein, an inserted linker α-helix, and the N-terminal α-helix of the Tim21 protein (Fig. 2-1). The linker is a poly-alanine-based sequence containing glutamate-lysine pairs, to maximize the probability of the α-helical conformation. We use the systematic names of the fusion proteins as MBP_{apo}<n>Tim21 and MBP_{holo}<n>Tim21, where the apo/holo subscript represents the absence or presence of a maltose molecule in the binding site of MBP, and n represents the length of the α-helical linker (n = 16, 17, 20 or 24). The targeting of loop 2 in the CCFS can be controlled by the length of the inserted α-helical linker. The open conformation of the apo form and the closed conformation of the holo form of MBP are the second degree of freedom in designing the CCFS in the MBP-Tim21 fusion protein.

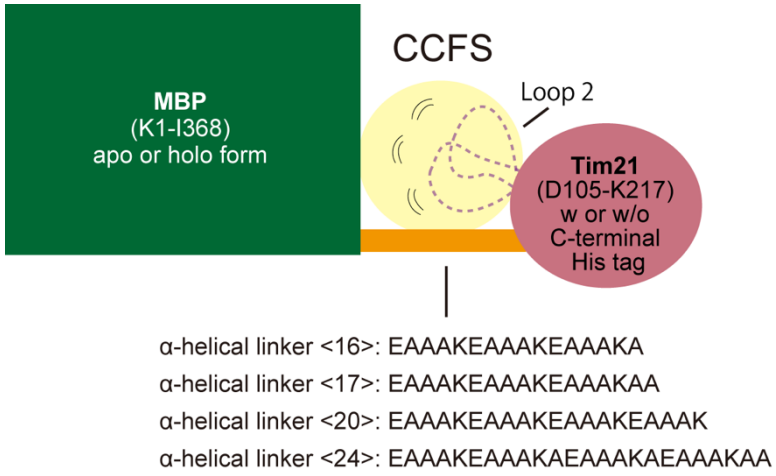


Fig. 2-1. The design concept for the creation of crystal contact-free space (CCFS) in a protein crystal. A tag protein (*E. coli* MBP protein, green) is fused to a target protein (yeast Tim21, salmon) *via* an α-helix forming sequence. Space (light yellow) is created in the fusion protein crystal by the use of the rigid α-helical linker (orange). The length of the linker and the binding and nonbinding of maltose to the MBP are two degrees of freedom in designing the protein constructs, to arrange a flexible loop segment of the target Tim21 protein in the created CCFS. For fusion proteins with a C-terminal His tag, affinity elution with imidazole from Ni-NTA resin maintains the MBP in the apo form, whereas for fusion proteins without a His tag, affinity elution with maltose from amylose resin results in the maltose-bound (holo) form of MBP.

Chapter 3. X-ray crystallography

3.1. Crystallographic analyses for eight MBP-Tim21 fusion proteins

3.1.1. Plasmid construction for the expression of fusion proteins

The DNA sequences encoding the C-terminal part (N349-I368) of *E. coli* maltose binding protein (MBP), a 16-residue linker (EAAAKEAAAKEAAAKA), a 17-residue linker (EAAAKEAAAKEAAAKAA), a 20-residue linker (EAAAKEAAAKEAAAKEAAAK), a 24-residue linker (EAAAKEAAAKAEAAAKAEAAAKAA), and the intermembrane space fragment (D105-K217) of *Saccharomyces cerevisiae* Tim21 were custom synthesized (Genscript). The DNA sequences corresponding to the linkers and Tim21 were codon optimized for *E. coli* host cell expression. The DNA sequence encoding MBP (K1-T366) was amplified by PCR with the pMAL-c5x plasmid (New England Biolabs) as the template. The two DNA fragments were directly cloned into the NdeI/XhoI digested pET-31b(+) vector (Novagen), using an In-Fusion Advantage PCR Cloning Kit (Clontech). For the construction of fusion proteins with and without a C-terminal His tag, primers containing six histidine codons and a stop codon (TGA) were used in the In-Fusion reaction. As a result, we constructed eight MBP-Tim21 fusion proteins expression vectors:
MBP<16>Tim21_pET31b(+), MBP<16>Tim21His6_pET31b(+),
MBP<17>Tim21_pET31b(+), MBP<17>Tim21His6_pET31b(+),
MBP<20>Tim21_pET31b(+), MBP<20>Tim21His6_pET31b(+),
MBP<24>Tim21_pET31b(+), MBP<16>Tim21His6_pET31b(+)

3.1.2. Proteins expression and purification

The MBP-Tim21 fusion proteins were expressed in *E. coli* BL21(DE3) cells (Agilent). The *E. coli* cells were grown in LB medium containing 100 mg L^{-1} ampicillin at 310 K, until the OD_{600} reached 0.6-0.8. After overnight induction with 1 mM isopropyl-1-thio-β-D-thiogalactopyranoside (IPTG) at 289 K, the cells were harvested by centrifugation and then disrupted by sonication. The MBP-Tim21 fusion proteins without a His tag were adsorbed to amylose resin (New England Biolabs) in TS buffer (20 mM Tris-HCl, pH 8.0, containing 0.1 M NaCl) and affinity eluted with 10 mM maltose, whereas the MBP-Tim21 fusion proteins with a His tag were adsorbed to

Ni-NTA resin (High-performance grade, GE Healthcare) in TS buffer, and affinity eluted with 500 mM imidazole (Fig. 3-1A). Note that the His tag was not removed for crystallization. The eluted proteins were further separated by gel filtration chromatography on a Superdex 200 10/300 GL column (GE Healthcare) in TS buffer (Fig. 3-1B, C). The purified proteins were concentrated to 10 - 35 mg mL^{-1} for crystallization, with an Amicon Ultra-4 centrifugal filter unit (Millipore, 30 kDa NMWL), without changing the buffer composition.

In our experimental scheme, the optional C-terminal His tag determines the apo or holo state of MBP. For fusion proteins, with a His tag, affinity elution with imidazole from Ni-NTA resin maintains the MBP in the apo form, whereas for fusion proteins without a His tag, affinity elution with maltose from amylose resin results in the maltose-bound (holo) form of MBP.

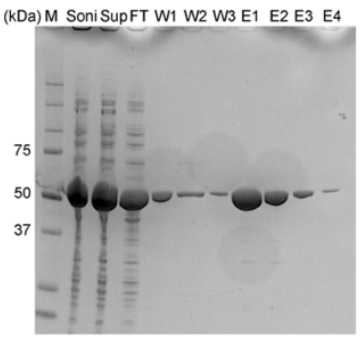
Fig. 3-1. Purification of the eight MBP-Tim21 fusion proteins (from the next page)

A. SDS-PAGE (10-20% gradient gels) images of the analysis of the fractions of affinity purification. MBP-Tim21 fusion proteins with His tag were purified with Ni-NTA resin and the proteins of MBP_{apo}<n>Tim21 were obtained. MBP-Tim21 fusion proteins without His tag were purified with Amylose resin and the proteins of MBP_{holo}<n>Tim21 were obtained. M: Marker, Soni: Sonicate, Sup: Supernatant of sonicated solutions, FT: Flow through of affinity column, W: Wash fraction, E: Elute fraction.

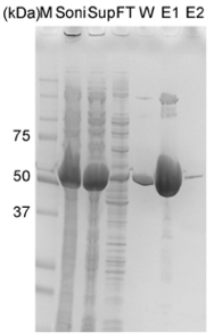
B. Gel filtration chromatograms (Superdex200 10/300GL column) of the eluted fractions of the MBP-Tim21 fusion proteins from the affinity column.

C. SDS-PAGE electrophoresis (10-20 % gradient gels) images of the gel filtration fractions of the MBP-Tim21 fusion protein.

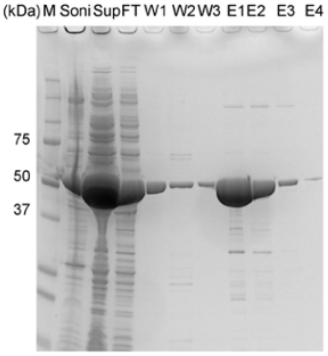
A



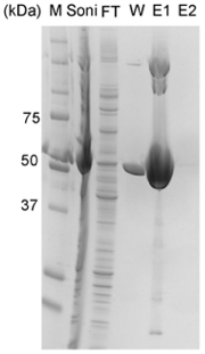
MBP_{holo}<16>Tim21



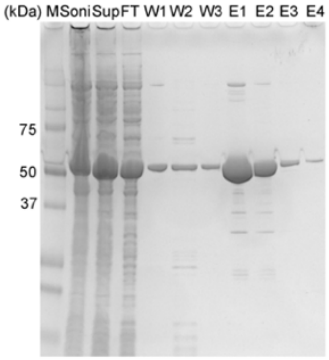
MBP_{apo}<16>Tim21



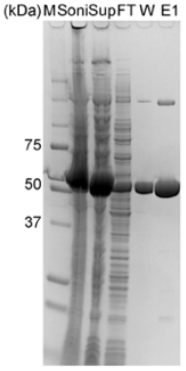
MBP_{holo}<17>Tim21



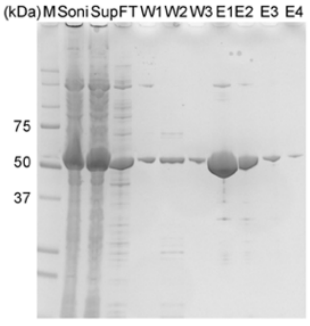
MBP_{apo}<17>Tim21



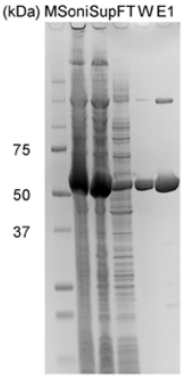
MBP_{holo}<20>Tim21



MBP_{apo}<20>Tim21

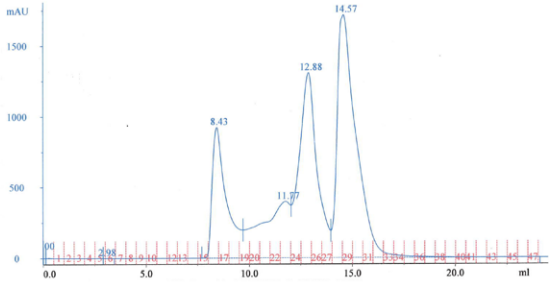


MBP_{holo}<24>Tim21

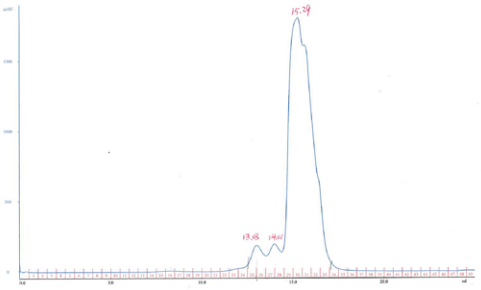


MBP_{apo}<24>Tim21

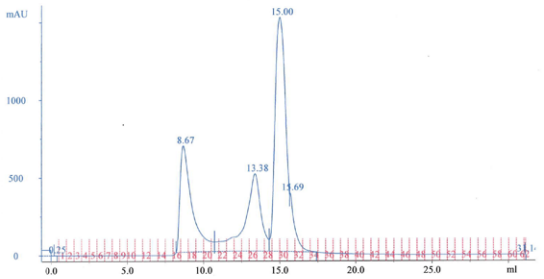
B



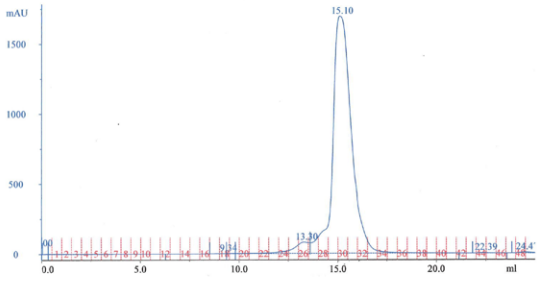
MBP_{holo}<16>Tim21



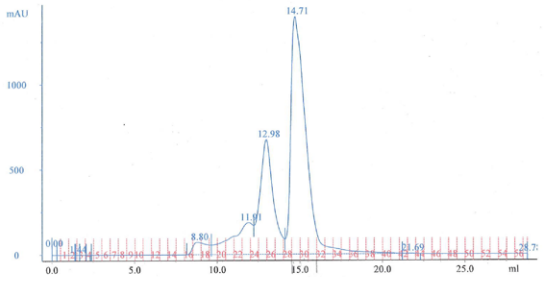
MBP_{apo}<16>Tim21



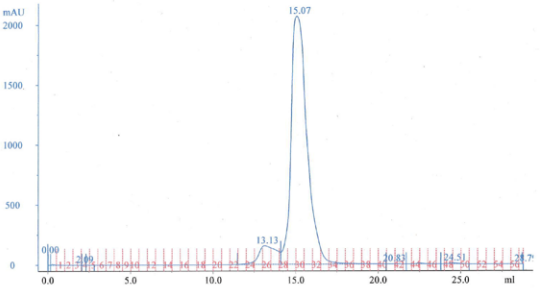
MBP_{holo}<17>Tim21



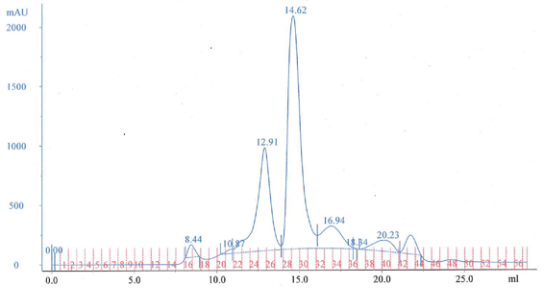
MBP_{apo}<17>Tim21



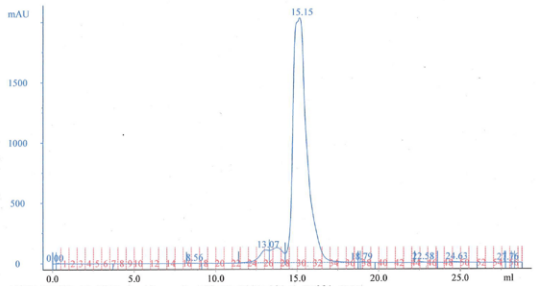
MBP_{holo}<20>Tim21



MBP_{apo}<20>Tim21

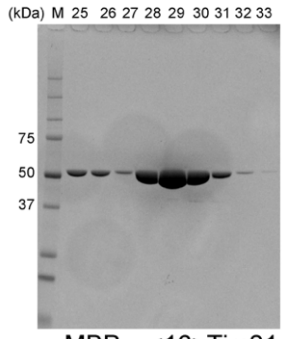


MBP_{holo}<24>Tim21

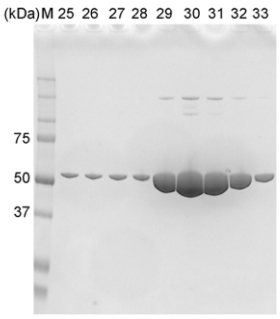


MBP_{apo}<24>Tim21

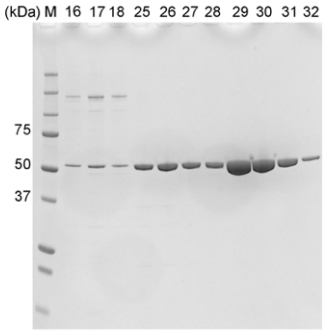
C



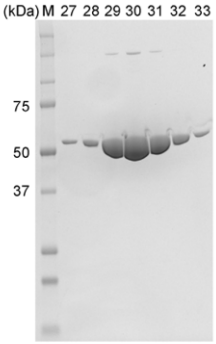
MBP_{holo}<16>Tim21



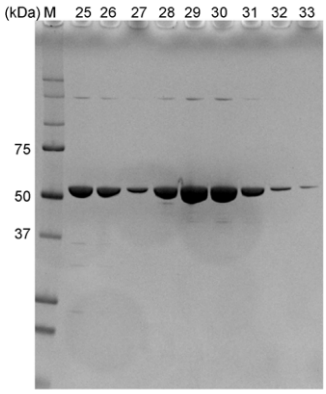
MBP_{apo}<16>Tim21



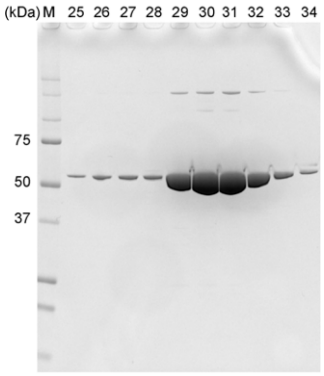
MBP_{holo}<17>Tim21



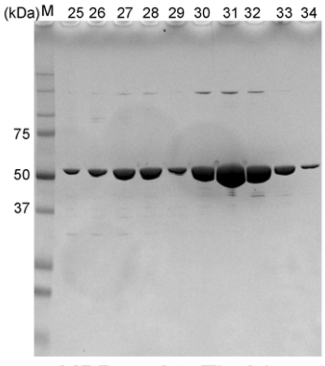
MBP_{apo}<17>Tim21



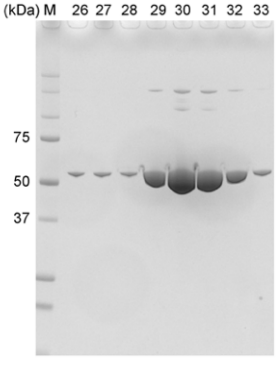
MBP_{holo}<20>Tim21



MBP_{apo}<20>Tim21



MBP_{holo}<24>Tim21



MBP_{apo}<24>Tim21

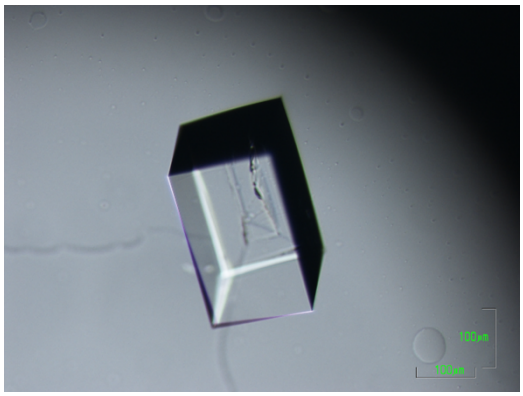
3.1.3. Crystallization, data collection, and structure determination

Initial screening was performed using the sitting-drop vapor-diffusion method in 96-well plates (INTELLI-PLATE, Art Robbins Instruments), using JCSG+ Suite and PACT Suite kits (Qiagen). The crystallization conditions were optimized by a grid screening strategy, using the hanging-drop vapor-diffusion method in 24-well plates (VDX Plate with sealant, Hampton Research). Additive screening (Hampton Research HR2-428) and microseeding crystallization were performed when necessary. Sitting drops were set up by mixing equal volumes (0.2 µl each) of the protein solution and the reservoir solution, using an automated dispenser (Gryphon, Art Robbins Instruments). Hanging drops were prepared manually, by mixing equal volumes (1 µl each) of the protein solution and the reservoir solution. Each sitting drop was placed over 0.09 ml reservoir solution and each hanging drop was placed over 0.4 ml reservoir solution. All crystallizations were performed at 293 K. After optimization, crystals of the fusion proteins grew from a hanging drop under the conditions listed in Table 3-1 within 7 days (Fig. 3-2). Microseeds for crystallization of MBP_{holo}<17>Tim21 were prepared by serial dilutions of the same protein crystals, which grew in 0.1 M MMT buffer, pH 6.0, 25% (w/v) PEG1500 during the initial screening.

The crystals were briefly dipped into cryoprotectant solutions, containing 40% PEG1500 for MBP_{apo}<16>Tim21, 20% glycerol for MBP_{apo}<17>Tim21, and 40% PEG3350 for MBP_{holo}<17>Tim21, and then flash-cooled in liquid nitrogen. X-ray diffraction data were collected at the beamline BL44XU of SPring-8 (Harima, Japan) and the beamline BL1A of the Photon Factory (Tsukuba, Japan). The diffraction data were processed with HKL2000, version 714 [32], and XDS, version Jan 26, 2018 [33]. Initial phases were obtained by the molecular replacement method with Phaser-MR and AutoMR in PHENIX, version 1.8 [34]. The atomic coordinates of MBP (PDB 1ANF for the maltose-bound form [35], and PDB 1PEB for the apo form [36]) and Tim21 (PDB 2CIU) were used as the search models. Manual modeling was performed using COOT, version 0.7 [37, 38]. The coordinates of the fusion proteins plus water molecules were refined with AutoBuild and phenix.refine in PHENIX. The data collection and refinement statistics are summarized in Table 3-1. The figures of the structures were drawn with PyMOL, version 2.1.0 (https://pymol.org/2/, last accessed in June, 2019).

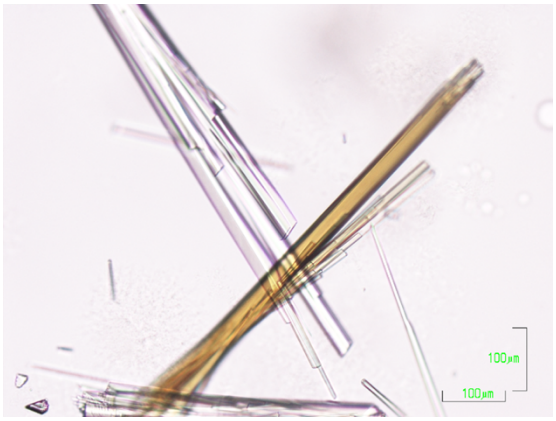
A. Crystal of MBP_{holo}<17>Tim21.

Crystallization method was Hanging Drop Vapor Diffusion; Protein solution was 17 mg/ml MBP_{holo}<17>Tim21 in 20 mM Tris-HCl pH 8.0, 100 mM NaCl; Reservoir solution was 0.2 M K/Na Tartrate, 16% PEG 3350; incubation at 293 K; Cryoprotectant was 40% (w/v) PEG 3350.



B. Crystals of MBP_{apo}<16>Tim21.

Crystallization method was Hanging Drop Vapor Diffusion; Protein solution was 9 mg/mL in 20 mM Tris-HCl pH 8.0, 100 mM NaCl; Reservoir solution was 0.1M MMT pH 6.0, 24% (w/v) PEG1500; 10 mM Betaine hydrochloride was added into crystallization drop with 1:1 volume ratio and total volume was 2 uL; incubation at 293 K; Cryoprotectant was 40% (w/v) PEG 1500.



C. Crystal of MBP_{apo}<17>Tim21.

Crystallization method was Hanging Drop Vapor Diffusion; Protein solution was 20 mg/mL protein in 20 mM Tris-HCl pH 8.0, 100 mM NaCl; Reservoir solution was 0.2 M $CaCl_2$, 0.1 M HEPES pH 7.0, 20% PEG6000; incubation at 293 K; Cryoprotectant was 20% (v/v) Glycerol and 16% (w/v) PEG6000.

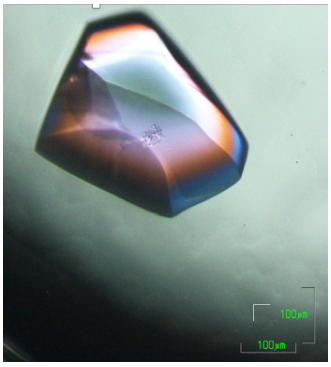


Fig. 3-2. Crystals of the three MBP-Tim21 fusion proteins.

Table 3-1. Design of MBP-Tim21 fusion proteins and X-ray data collection and refinement statistics.

	MBP_{apo}<16>Tim21	MBP_{apo}<17>Tim21	MBP_{holo}<17>Tim21
Protein design			
Linker	EAAAKEAAAKEAAAKA	EAAAKEAAAKEAAAKAA	EAAAKEAAAKEAAA
Tag	N-terminal MBP	N-terminal MBP	N-terminal MBP
	C-terminal His_6[a]	C-terminal His_6[a]	No His_6
Crystallization conditions	0.1 M MMT[b], pH 6.0	0.2 M $CaCl_2$	0.2 M K/Na Tartrate
	24% PEG1500 (w/v)	0.1 M HEPES, pH 7.0	16% PEG3350 (w/v)
	0.01 M Betaine hydrochloride (additive)	20% PEG6000 (w/v)	microseeds
Data collection statistics			
Software	HKL2000	HKL2000	XDS
Cryoprotectant	40% PEG1500 (w/v)	20% Glycerol	40% PEG3350 (w/v)
Beamline	Photon Factory BL-1A	SPring8 BL44XU	SPring8 BL44XU
Wavelength (Å)	1.1000	0.9000	0.9000
Resolution range (Å)	50.00 – 2.00 (2.03 – 2.00)	50.00 – 1.53 (1.56 – 1.53)	50.00 – 1.80 (1.91 –
Space group	$P2_12_12_1$	R 3	$P2_12_12_1$
Unit cell, a, b, c (Å), α, β, γ(°)	47.01, 88.84, 119.364	156.81, 156.81, 67.39, 90°, 90°, 120°	40.97, 69.41, 171.36
Observed reflections	219,364	945,347	304,581
Unique reflections	34,759 (1,709)	92,372 (4,565)	46,285 (7,366)
Multiplicity	6.3	10.2	6.6
Completeness (%)	99.9 (100.0)	99.9 (100.0)	99.90 (99.80)
R_{merge} [c]	0.118 (0.887)	0.118 (1.361)	0.047 (0.794)
$I/sigma$ (I)	15.33 (2.3)[d]	17.13 (2.06)[d]	19.26 (2.02)[e]
Refinement statistics			
Resolution range (Å)	41.63 - 2.00 (2.07 - 2.00)	39.20 - 1.53 (1.59 - 1.53)	35.28 - 1.80 (1.87 - 1.80)
Wilson B-factor (Å^2)	27.67	24.1	34.67
R_{work}	0.1754 (0.2240)	0.1656 (0.2575)	0.1745 (0.2674)
R_{free}[f]	0.2290 (0.2701)	0.1893 (0.3084)	0.2163 (0.2861)
Number of atoms	4,245	4,700	4,264
macromolecules	3,926	3,913	3,922
ligands	N/A	N/A	23
water	319	787	319
Protein residues	499	498	500
Rms deviation of bonds (Å)	0.007	0.019	0.010
Rms deviation of angles (°)	1.10	1.72	1.32
Ramachandran plot (%)			
favored	96.4	97.6	96.6
allowed	2.6	2.2	2.8
outliers	1.0	0.2	0.6
Clashscore	8.92	6.01	8.24
Mean B-factors, overall (Å^2)	37.4	22.0	39.5
PDB ID	6K7D	6K7E	6K7F

Values in parentheses are for highest-resolution shell.

a. The His tag was not removed.

b. MMT Buffer System: Malic Acid, MES, and TRIS (1:2:2 molar ratio).

c. *R*merge = $\Sigma\Sigma|Ii - \langle I\rangle|/\Sigma\Sigma Ii$, where *Ii* is the intensity of the *i*th observation and $\langle I\rangle$ is the mean intensity.

d. *I*/*sigma* (*I*) = (average of intensity) / (average of Sigma (*I*)) of unique reflections.

e. *I*/*sigma* (*I*) = mean of (intensity / Sigma (*I*)) of unique reflections.

f. The *R*free value was calculated for the *R* factor by using only a test set (5 %) of reflections not used in the refinement.

3.1.4. CCFS around loop 2 in three MBP-Tim21 fusion proteins crystals

We roughly evaluated the CCFS around loop 2 with B-factor putty figures of loop 2 (Fig. 3-3). We found that the CCFS of MBP_{holo}<17>Tim21 is the best. For convenience, we named MBP_{holo}<17>Tim21 as CCFS1 which has the best CCFS, MBP_{apo}<16>Tim21 as CCFS2, which has the second best CCFS, and MBP_{apo}<17>Tim21 as CCFS3, which has the worst CCFS in the three MBP-Tim21 fusion proteins.

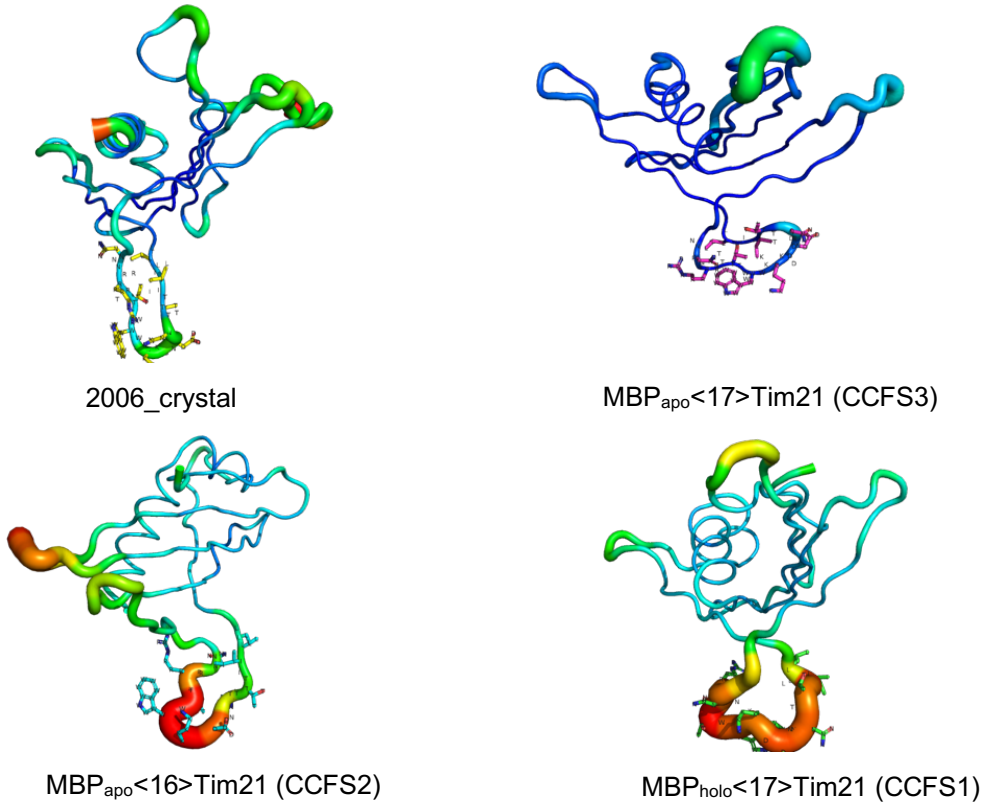


Fig. 3-3. B-factor-putty drawings of 2006_crystal and three MBP-Tim21 fusion proteins. The side chains of loop 2 are shown by the sticks.

3.1.5. F*o*-F*c omit maps of three CCFS1 crystals structures*

We analyzed the X-ray diffraction data from three CCFS1 crystals (Table 3-2) to check the reproducibility of conformation of loop 2 in CCFS1. *F*o-*F*c difference maps were calculated with phenix.refine in the program PHENIX. The atoms of loop 2 were deleted before the map calculation (Fig. 3-4). The low-pass filter was used to increase the signal-to-noise ratio of the electron density maps (Fig. 3-4.B) [23].

Table 3-2 X-ray data collection statistics of the three MBP_{holo}<17>Tim21 (CCFS1) crystals.

	CCFS1	**CCFS1-1**	**CCFS1-2**
Software	XDS	HKL2000	XDS
Date	Nov 2, 2018	Oct 22, 2016	Jun 22, 2018
Beamline	SPring-8 BL44XU	SPring-8 BL44XU	SPring-8 BL44XU
Detector	DECTRIS EIGER X 16M	RAYONIX MX300HE	DECTRIS EIGER X 16M
Collected frames	1800 frames	360 frames	360 frames
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell, *a*, *b*, *c* (Å)	40.97, 69.41, 171.36	40.83, 69.27, 173.98	40.92, 69.75, 170.53
Resolution range (Å)	50.00 – 1.80 (1.91 – 1.80)	50.00 - 1.90 (1.93 - 1.90)	50.00 - 1.30 (1.38 - 1.30)
Multiplicity	6.6 (6.7)	15.3 (12.7)	6.4 (6.6)
Completeness (%)	99.9 (99.8)	100.0 (100.0)	99.8 (99.1)
***R*merge**	0.047 (0.794)	0.149 (>1)	0.042(0.933)
***I*/*sigma* (*I*)**	19.26 (2.02) [a]	18.22 (1.75) [b]	19.26（1.62) [a]

Values in parentheses are for highest-resolution shell.

a. *I*/*sigma* (*I*) = mean of (intensity / sigma (*I*)) of unique reflections.

b. *I*/*sigma* (*I*) = (average of intensity) / (average of sigma (*I*)) of unique reflections.

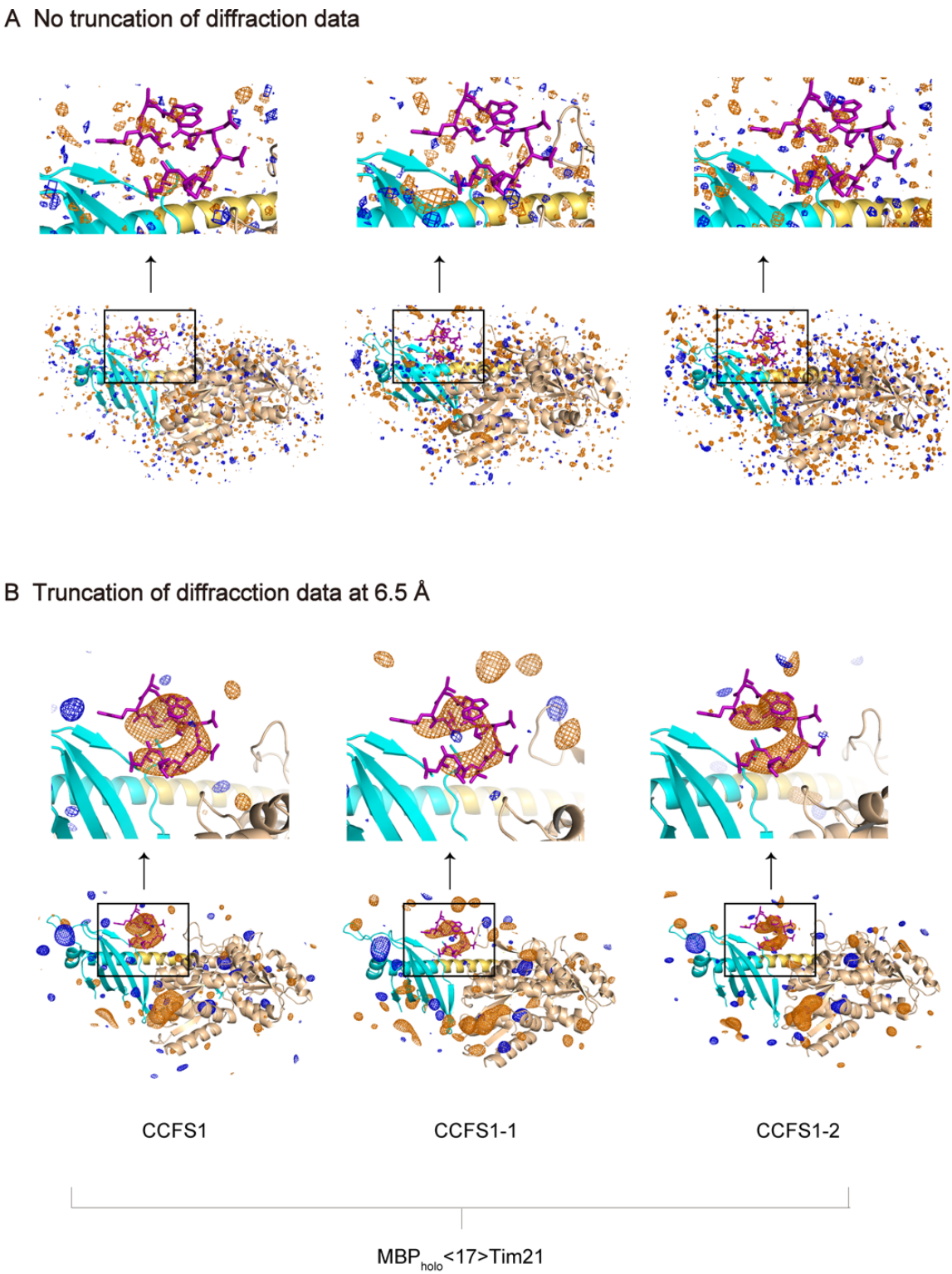


Fig.3-4. *F*o-*F*c simulated-annealing omit maps of the MBP_{holo}<17>Tim21 protein crystals.

A. Omit electron density maps, contoured at +3σ (orange) and -3σ (blue). The diffraction data to the highest resolution (1.3 – 1.9 Å) were used for the map generation.

B. Omit electron density maps with truncation of high-resolution reflections at 6.5 Å prior to Fourier transformation, contoured at +3σ (orange) and -3σ (blue).
MBP and Tim21 are shown in the cartoon representation and colored wheat and cyan, respectively. The 17-residue spacer is colored yellow orange. The structure of the deleted loop 2 is depicted as magenta sticks with the side chains. Note that the same X-ray diffraction data set was used for map generation in A and B. To assess the reproducibility, other two omit maps from different crystals of the same protein were shown (CCFS1-1 and CCFS1-1).

3.1.6. Considerations about the three CCFS crystal structures

We obtained three CCFSs protein crystal structures, MBP_{holo}<17>Tim21 (CCFS1), MBP_{apo}<16>Tim21 (CCFS2), and MBP_{apo}<17>Tim21 (CCFS3). In CCFS1, the loop 2 is the most flexible. Due to the flexibility, electron density of loop 2 in CCFS1 is difficult to be observed in the usual *F*o-*F*c simulated-annealing omit maps. With a low-pass filter which cuts off the diffracted data of the resolution higher than 6.5 Å before the calculation of the *F*o-*F*c simulated-annealing omit maps made the electron density of loop 2 in CCFS1 more visible. We superimposed the *F*o-*F*c simulated-annealing omit maps of the three CCFSs with the determined loop 2 conformations in CCFS1, and found the model of loop 2 overlapped with the electron density of the three CCFSs. This proved our loop 2 model was determined correctly, it also verified the reproducibility of the loop 2 conformation in CCFS1 (Fig. 3-4).

3.2. Evaluation of CCFS created in the crystal structures of MBP-Tim21

3.2.1. B-factors of loop 2 of Tim21 in CCFS

To evaluated the quality of the CCFS in these crystals, we analyzed the B-factor values of loop 2 of Tim21 (Table 3-3), and showed the states of loop 2 in different crystal lattices with figures of B-factor putty/sausage in PyMOL (Fig. 3-5A).

Table 3-3. B-factors of the Cα atoms in the loop 2 region

Residue of loop 2	2006_crystal (2CIU)	MBPapo<17>Tim21 (CCFS3)	MBPapo<16>Tim21 (CCFS2)	MBPholo<17>Tim21 (CCFS1)
L144	15.2	9.9	41.5	73.4
I145	17.0	14.4	51.8	108.9
T146	16.1	16.8	60.2	97.4
N147	24.3	28.6	73.6	103.7
D148	26.5	21.0	105.1	108.2
K149	24.7	19.5	121.4	97.7
W150	17.9	14.5	124.9	104.5
T151	17.8	13.6	117.7	118.3
R152	17.1	12.5	95.2	100.3
N153	17.7	19.7	60.9	98.2
Average	19.4	17.1	85.2	101.1

The B-factors have some biases that are connected with the resolution. To reduce the effects of the resolution differences among the different protein crystal structures, normalized B-factors were used for comparison (Table 3-4, Fig. 3-5B):

$$B' = (B - \langle B \rangle)/\sigma(B)$$

where B is the B-factor of the Cα atoms, and $\langle B \rangle$ and $\sigma(B)$ represent the mean and standard deviation of B [39].

Table 3-4. The normalized B-factors in the residue range from 140 to 160. Loop 2 corresponds to residues 144-153.

Residue Number	2006_crystal (2CIU)	MBP_{apo}<17>Tim21 (CCFS3)	MBP_{apo}<16>Tim21 (CCFS2)	MBP_{holo}<17>Tim21 (CCFS1)
140	−0.67	−1.21	-0.55	-0.48
141	−0.91	-1.12	-0.70	-0.27
142	−0.50	-1.08	-0.64	-0.03
143	−0.65	-0.70	0.06	0.85
144	−0.40	-1.18	-0.11	1.46
145	0.03	-0.55	0.45	3.07
146	-0.19	-0.21	0.83	2.55
147	1.79	1.45	1.61	2.84
148	2.32	0.38	2.75	3.04
149	1.89	0.16	3.37	2.56
150	0.26	-0.53	3.54	2.88
151	0.23	-0.66	3.10	3.50
152	0.06	-0.81	2.10	2.68
153	0.20	0.20	0.79	2.59
154	0.56	-0.70	-0.09	1.93
155	0.90	-0.15	0.57	1.78
156	0.34	-1.05	0.07	0.50
157	-0.67	-1.03	0.21	0.03
158	-0.93	-0.59	-0.21	-0.54
159	-0.56	-0.29	0.08	-0.62
160	-0.39	-0.61	0.38	-0.55

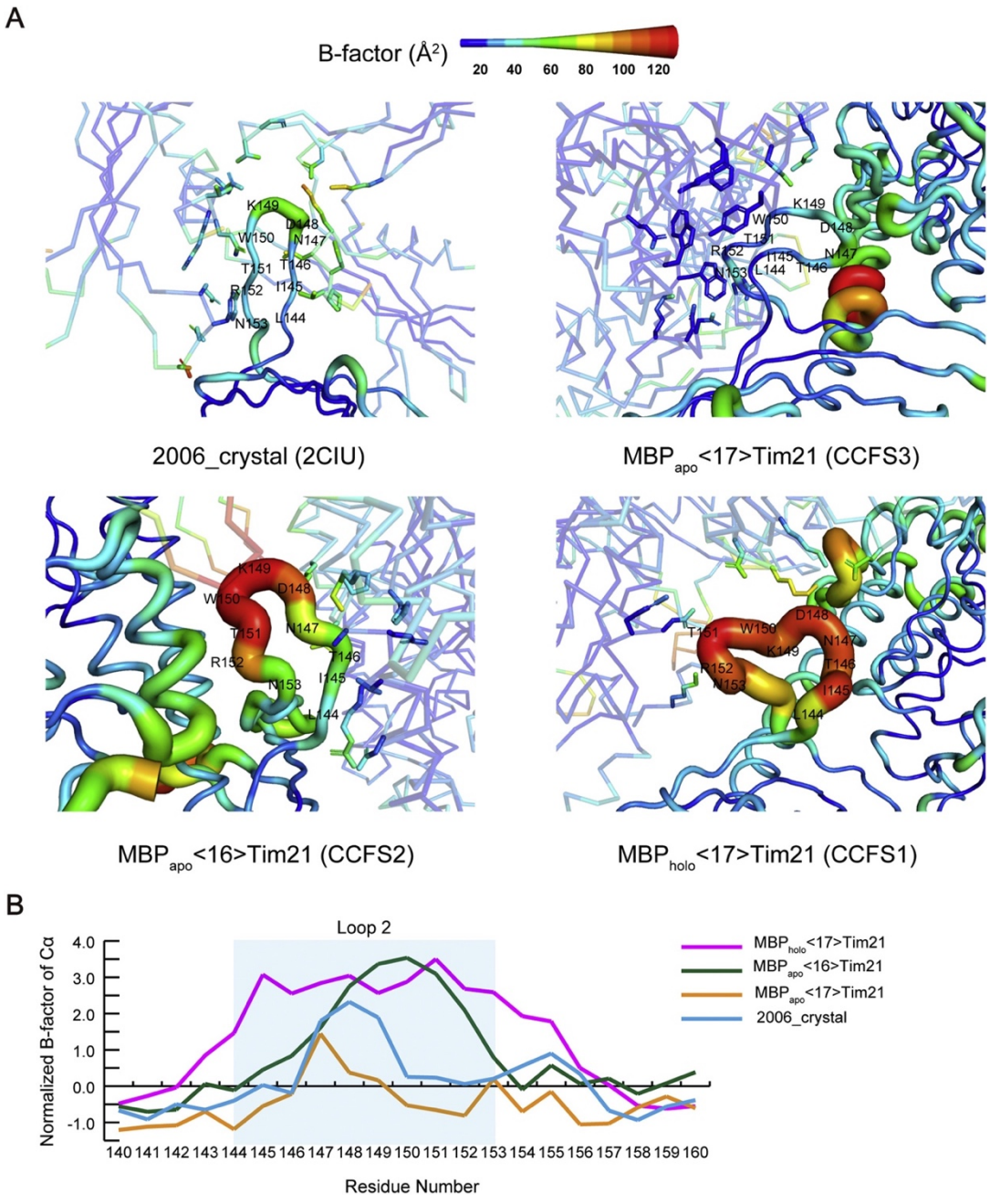


Fig. 3-5. Status of loop 2 in different crystals.

A. B-factor putty/sausage representations of the Cα atoms of loop 2 in the four crystals. The backbones of the symmetrically related molecules are shown in stick representations and colored according to the *B*-factor values. The side chains that form intermolecular contacts with loop 2 are shown in stick representations.

B. The normalized B-factors in the residue range from 140 to 160. Loop 2 corresponds to residues 144–153.

3.2.2. Potential protein crystal contacts of loop 2 in CCFS

The distance threshold of 8 Å was used to search atoms pairs of the Cα atoms of loop 2 and the atoms of surrounding residues which include symmetric protein molecules, and the MBP and the linker parts of the same molecule. The measurement/distances function in PyMOL was used to calculate atom-pair distances. Table 3-5 shows the number of the atom pairs which have potential interactions of loop 2 with other regions in crystals.

Table 3-5. The number of potential interactions of loop 2 with other regions in crystals[a].

		L144	I145	T146	N147	D148	K149	W150	T151	R152	N153	Total
CCFS1	Symmetric molecule 1	0	0	0	2	3	1	2	1	0	0	20
	Symmetric molecule 2	0	0	0	0	0	0	1	4	0	2	
	MBP<17>	0	0	3	1	0	0	0	0	0	0	
CCFS2	Symmetric molecule 1	0	1	4	4	3	1	0	0	0	0	32
	Symmetric molecule 2	4	3	5	3	1	0	0	0	0	2	
	MBP<16>	0	0	0	0	0	0	0	1	0	0	
CCFS3	Symmetric molecule 1	2	3	0	0	2	4	6	1	4	5	27
	MBP<17>	0	0	0	0	0	0	0	0	0	0	
2006_crystal	Symmetric molecule 1	4	6	5	4	3	0	1	3	2	0	68
	Symmetric molecule 2	0	0	0	2	3	3	0	0	0	0	
	Symmetric molecule 3	2	0	1	0	1	5	8	5	6	4	

a. The distance threshold of 8 Å was used to search atoms pairs of the Cα atoms of loop 2 and the side-chain atoms of surrounding residues, which include the symmetric protein molecules, and the MBP and linker parts of the same molecule.

3.2.3. Estimation of the volume of CCFS

Pseudo atoms were arranged in a grid at intervals of 3 Å inside a sphere of radius R. R was set 18 Å after trial and error. The sphere was centered on the mean position of the Cα atoms in the two residues (E143 and R154) situated on the borders of loop 2. In PyMOL, the pseudo atoms within 3 Å of the MBP-Tim21 fusion protein (without loop 2) and its symmetrically related molecules were removed. Then, some pseudo atoms were removed manually by visual inspection. Finally, the number of the remained pseudo atoms was counted and multiplied by 27 $Å^3$ to calculate the volume of the CCFS. The size of the three CCFSs estimated was showed in the Fig. 3-6.

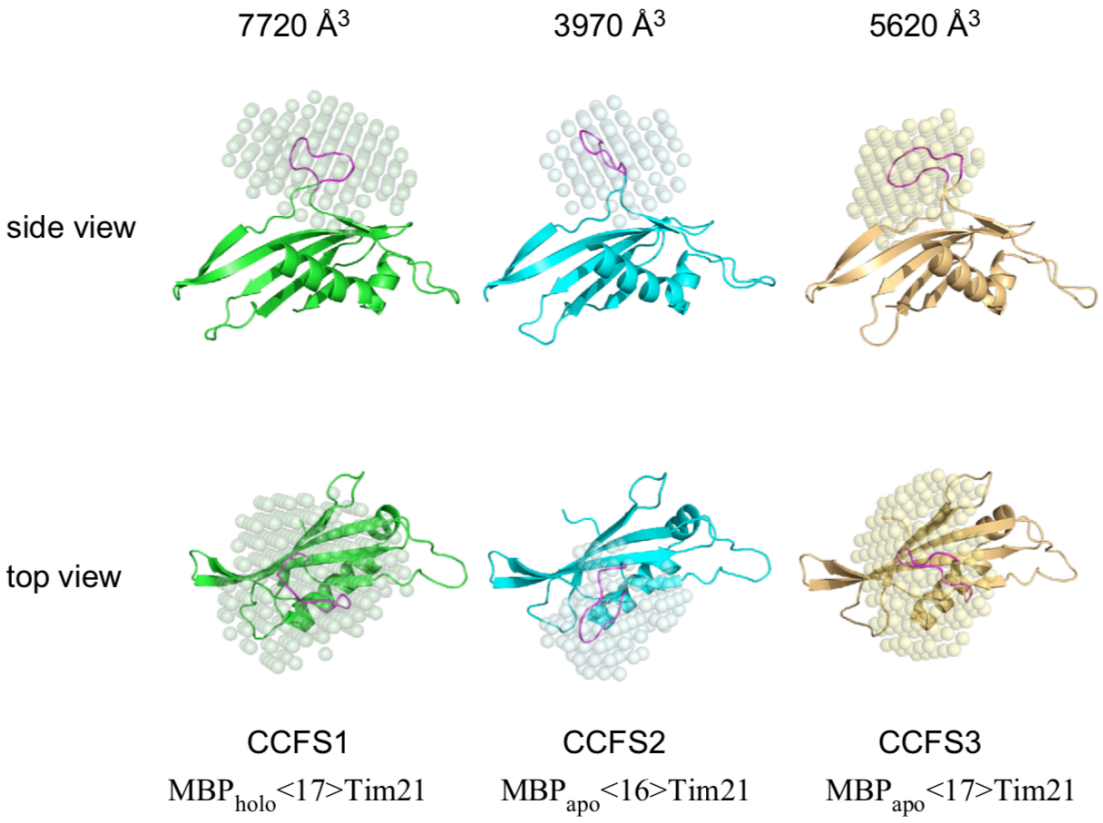


Fig. 3-6. Estimation of the volume of the CCFS formed in the three crystal structures of the MBP-Tim21 fusion proteins.

The yeast Tim21 portions of the MBP-Tim21 fusion proteins were depicted in the cartoon representation with different colors. Loop 2 is colored magenta. The CCFSs formed in the crystals are represented as a cluster of pseudo atoms arranged in a grid at intervals of 3 Å.

3.2.4. Summary of CCFS evaluation

To evaluate the quality of the CCFS in these crystal structures, we analyzed the B-factor values (Tables 3-3, 3-4). The flexible loop 2 segment has a large amplitude of motion in solution. In an ideal CCFS, the electron density is faint (Fig. 3-4A), and the B-factors of the modeled atoms are large (Fig. 3-5), reflecting the large amplitude of motion (or more strictly, static disorder in frozen crystals) in the CCFS. As we described previously [23], the truncation of diffraction data at 6.5 Å improved the signal-to-noise ratio of the difference maps at the expense of the map resolution (Fig. 3-4B). The backbone atoms of the loop 2 models coincide well with the signal-to-noise ratio enhanced electron densities. The B-factor putty/sausage representations of loop 2 (Fig. 3-5A) and the uniform distribution of the normalized B-factor values (Fig. 3-5B) clearly show the minimal restrictions on loop 2 in the CCFS formed in the MBP_{holo}<17>Tim21 crystal. Thus, we selected the MBP_{holo}<17>Tim21 crystal as the best CCFS. In contrast, in the other three crystals, the entire (2006_crystal and MBP_{apo}<17>Tim21) and the N-terminal half (MBP_{apo}<16>Tim21) of loop 2 have small B-factor values, indicating that the unwanted crystal contacts with the symmetrically related molecules in the crystal lattice restricted the motion of loop 2.

We investigated the potential interactions of loop 2 with other regions in the CCFSs. The distance threshold of 8 Å was used to search atoms pairs of the Cα atoms of loop 2 and the side-chain atoms of surrounding residues, which include the symmetric protein molecules, and the MBP and linker parts of the same molecule (Table 3-5). As expected, the number of interactions is minimum for the CCFS1, and the largest for the 2006_crystal structure.

We estimated the size of the CCFSs (Fig. 3-6). The estimated volumes of the CCFS1, CCFS2 and CCFS3 are 7,720 $Å^3$, 3,970 $Å^3$, and 5,260 $Å^3$, respectively. As expected, the CCFS1 has the largest size. The CCFS3 has the second largest size, but loop 2 is located off-center.

3.3. Comparing the conformation of loop 2 in CCFS1 and 2014_NMR

We compared the loop 2 conformations of 2014_NMR and CCFS1 by superposition the two structures with pair_fit the core residues of 105-131, 134-143, 154-193, and 196-217 in PyMOL. The RMSD between the two structures is 1.95 Å. Contrary to our

expectation, the conformation of loop 2 in CCFS1 is different from it in 2014_NMR (Fig.3-7).

Fig. 3-7. Comparison of the conformation of loop 2 in CCFS1 and 2014_NMR.
Superimposition of the CCFS1 structure (wheat) and the 2014_NMR structure (blue). Loop 2 in the 2014_NMR structure is highlighted in cyan and in CCFS1 is highlighted in magenta.

3.4. Summary of the X-ray crystallization and the next work

A preliminary modeling of the MBP-Tim21 fusion protein with PyMOL suggested that 16, 17, 20, and 24 amino acid residues were suitable for the linker. We thus expressed and purified the MBP-Tim21 fusion proteins with the four linker lengths. We also prepared the same four MBP-Tim21 fusion proteins with a C-terminal His tag. After crystal screening and diffraction measurements, we obtained three crystal structures at resolutions in the range of 1.6–2.0 Å. The successful protein constructs are MBP_{apo}<17>Tim21, MBP_{apo}<16>Tim21, and MBP_{holo}<17>Tim21 (Table 3-1).

We performed the B-factors analysis, potential crystal contacts of loop 2 in the CCFS, and calculation of the volume of the three CCFSs and found the CCFS1 (MBP_{holo}<17>Tim21) has the best CCFS around loop 2, CCFS2 (MBP_{apo}<16>Tim21) has the second best CCFS, and CCFS3 (MBP_{apo}<17>Tim21) has the worst CCFS. The omit maps obtained from the different crystals of the same protein (MBP_{holo}<17>Tim21) are shown for the CCFS1 case, to confirm the reproducibility of the electron density and the model of loop 2 in the CCFS crystals (Fig. 3-4, Table 3-2).

Although the loop 2 in CCFS1 shows the high flexibility, it is still a question whether the loop 2 conformation in CCFS1 is its representative/dominant conformation in solution. The structures determined by NMR and calculated by MD simulation are good choices for the assessment. Unexpectedly, the conformation of loop 2 in CCFS1 is different from the conformation of loop 2 in 2014_NMR (Fig.3-7). New NMR data is necessary for the assessment of the loop 2 conformation in CCFS1.

Chapter 4. NMR spectroscopy

4.1. Motivation for performing NMR spectroscopy

At the beginning of the project, we had expected to observe an almost identical conformation of loop 2 in the CCFS to those in the solution NMR structure reported previously [29], but this was not the case (Fig.3-7). Therefore, we measured the NMR spectra of Tim21 and recalculated the NMR structure. For convenience, we refer to the previous NMR structure as 2014_NMR, and our new structure as 2019_NMR.

4.2. Plasmid construction for the expression of the Tim21 protein

The DNA sequence encoding the intermembrane space fragment (D105-K217) of yeast Tim21was amplified by PCR with 5' primer of GTG CCG CGC GGC AGC GAT ACC CAG CTG TTC AAT CGT and 3' primer of AGC AGC CGG ATC TCA TTT GAT CAG GTA GTA GCG TTT CTC, using the plasmid for the expression of the MBP-Tim21 fusion proteins as the template. The linearized pET28b(+) vector was prepared by inverse PCR with 5' primer of GCT GCC GCG CGG CA and 3' primer of TGA GAT CCG GCT GCT AAC AAA GC. The PCR product encoding Tim21 was inserted into the linearized pET28b(+) vector, using an In-Fusion HD Cloning Kit (Clontech).

4.3.Protein expression and purification

The constructed plasmid was transformed into *E. coli* BL21(DE3). For stable isotope labeling, each liter of M9 medium contained 1 g of [^{15}N]NH_4Cl for the ^{15}N-labeled sample, or 1 g of [^{13}C]glucose and 1 g of [^{15}N]NH_4Cl for the double labeled sample, supplemented with 30 mg L^{-1} kanamycin. The cells were grown at 310 K until the OD600 reached 0.6–0.8. After overnight induction with 1 mM IPTG at 289 K, the cells were harvested by centrifugation and disrupted by sonication. After centrifugation, the supernatant was loaded on Ni-NTA resin (High-performance grade, GE Healthcare) in 20 mM Tris-HCl, pH 8.0, containing 300 mM NaCl. The adsorbed proteins were affinity eluted with 250 mM imidazole (Fig. 4-1A), and further purified by gel filtration chromatography on a Superdex 75 10/300 GL column (GE Healthcare) in 20 mM Tris-HCl, pH 8.0, containing 150 mM NaCl (Fig. 4-1B). The eluted protein was digested with thrombin (10 units per mg protein, Sigma-Aldrich) at room temperature

overnight. A MALDI-TOF-mass spectrometer Autoflex (Bruker) was used to monitor the digestion progress (Fig. 4-1C). The His-tag-free protein was isolated by gel filtration chromatography on a Superdex 75 10/300 GL column (GE Healthcare) in 20 mM Tris-HCl, pH 8.0, containing 150 mM NaCl (Fig. 4-1D, E). The purified proteins were buffer exchanged into 20 mM sodium phosphate buffer, pH 6.5, containing 50 mM NaCl, and then concentrated with a VIVASPIN TURBO-15 centrifugal filter unit (Sartorius, 5 kDa MWCO).

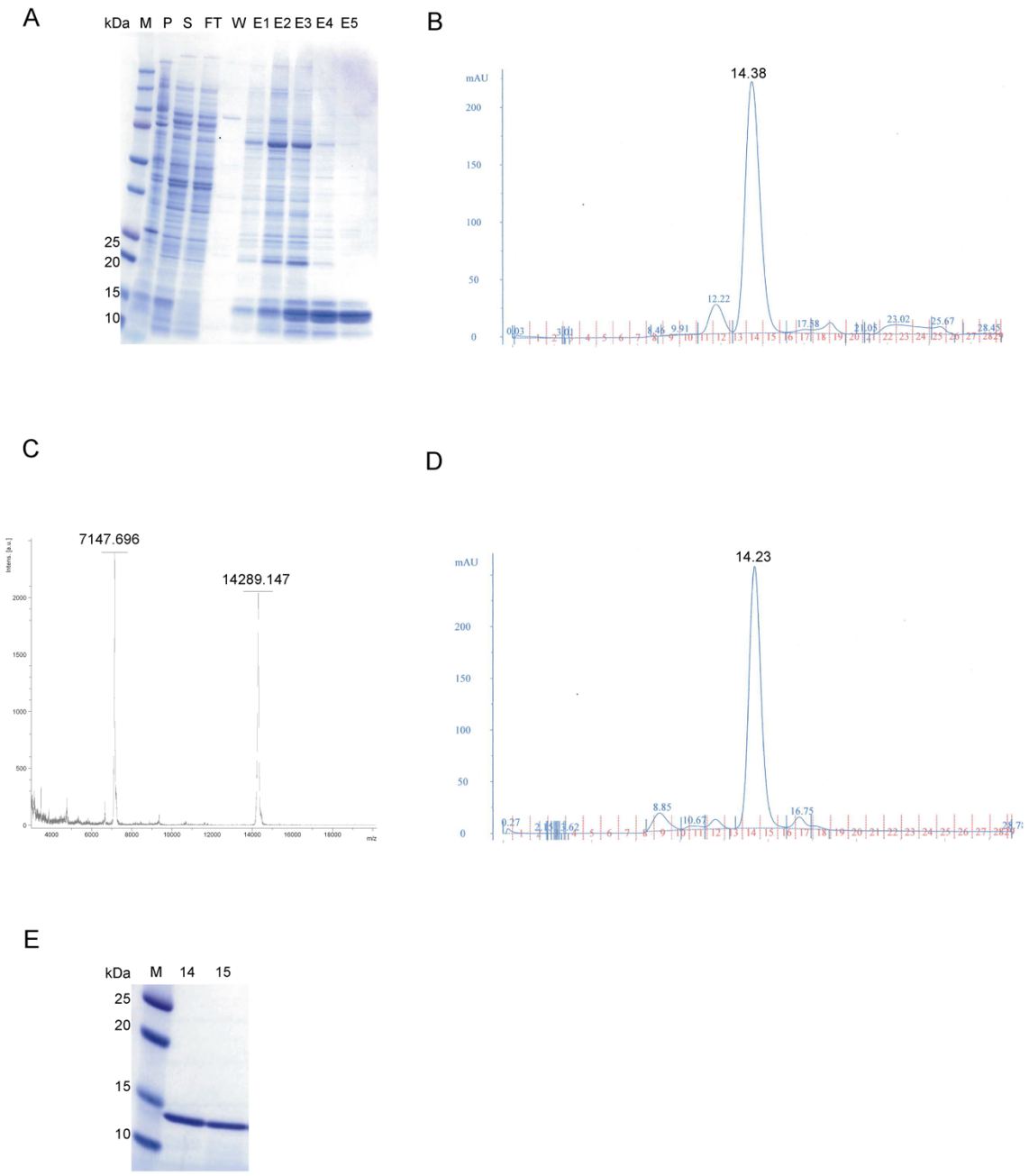


Fig. 4-1. Purification of ^{13}C, ^{15}N-labeled Tim21 for NMR spectroscopy.

A. SDS-PAGE (10-20% gradient gel) image of the affinity elution from Ni-NTA resin. M: protein molecular weight marker, P: sonicate precipitate, S: sonicate supernatant, FT: flow thorough from Ni-NTA resin, W: wash out faction, E: elute fraction.

B. Gel filtration chromatograph after Ni-NTA affinity separation.

C. MALDI-TOF MS measurement to check the thrombin digestion.

D. Gel filtration chromatograph of the digested sample, the elution was harvested with 1 ml per

fraction .

E. SDS-PAGE gel image of the analysis of the second gel filtration of the digested sample. The 14th and 15th fractions of gel filtration were analyzed with SDS-PAGE electrophoresis.

4.4. NMR spectroscopy

NMR measurements were performed at 300 K on Bruker AVANCE III and AVANCE III HD instruments, operated at 950 and 600 MHz and equipped with TCI and QCI-P cryogenic probes, respectively. The NMR sample for resonance assignment and structure determination contained 0.32 mM [^{13}C, ^{15}N] Tim21, dissolved in 20 mM sodium phosphate buffer, pH 6.5, containing 50 mM NaCl and 7% $^{2}H_2O$, and was placed in a 5-mm microtube (Shigemi, Japan). The 2D ^{1}H–^{15}N HSQC, ^{1}H–^{13}C CT (constant time)-HSQCs for aliphatic and aromatic regions, and 3D HNCO, CBCA(CO)NH, HNCACB, HBHA(CO)NH, and HCCH-TOCSYs for aliphatic and aromatic regions were recorded at 600 MHz. The 3D ^{15}N-edited NOESY (with a mixing time of 120 ms) and ^{13}C-edited NOESY (with a mixing time of 120 ms) for aliphatic regions were recorded at 950 MHz. The total measurement times were 1 week at 600 MHz and 6 days at 950 MHz. The {^{1}H}- ^{15}N steady-state heteronuclear Overhauser effects (hetNOE) of the backbone amide groups were measured at 600 MHz, according to [40]. NMR data were processed with nmrPipe and displayed with nmrDraw [41]. The 1H chemical shift was referenced to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) at 0 ppm. The ^{13}C and ^{15}N chemical shifts were indirectly referenced to the ^{1}H-frequency of DSS, using the gyromagnetic ratios of the ^{13}C and ^{15}N nuclei compared to that of the ^{1}H nucleus, respectively [42].

4.5. NMR structure determination

NMR spectra were analyzed with MagRO-NMRView, version 2.01 [43, 44]. Noise peaks were eliminated using the convolutional neural networks implemented in the program package Filt_Robot, with a positional mask based on the ^{1}H–^{15}N HSQC and ^{1}H–^{13}C HSQC peak positions [45]. Automatic resonance assignment in MagRO-NMRView was performed using the FLYA algorithm [46] implemented in CYANA, version 3.98 [47], by combining the CYANA structure calculations using the chemical shift tables from FLYA, the peak lists from the convolutional-neural-network-filtered NOESY, and the backbone dihedral angle restraints from TALOS+ [48]. An ensemble of the 20 lowest-target-function structures

was obtained with CYANA, using 1731 upper distance restraints and 173 backbone dihedral angle restraints, and then subjected to the final refinement with XPLOR-NIH, version 2.46 [49]. Finally, the 20 lowest-energy structures were selected from 100 refined structures. The quality of the final structures was analyzed with PROCHECK-NMR, version 3.5.4 [50]. FitRobot was used to identify the core residues of the 2019_NMR structures [51]. To calculate the average positions of the Cα atoms, 19 structures (model 2-model 20) were fitted to model 1, using the Cα atoms of the core residues 105–131, 134–143, 154–193, and 196–217, and the geometric center of the Cα atom of each residue was calculated. The means and standard deviations of the distances between the Cα atoms and the average positions were calculated as a function of the residue number.

At the beginning of the project, we had expected to observe an almost identical conformation of loop 2 in the CCFS to those in the solution NMR structure reported previously [29], but this was not the case. Therefore, we measured the NMR spectra of Tim21 and recalculated the NMR structure (Table. 4-1). For convenience, we refer to the previous NMR structure as 2014_NMR, and our new structure as 2019_NMR. The two NMR experiments are compared in (Fig. 4-2). The distributions of the NOE-derived distance restraints and the TALOS-derived torsion angle restraints as a function of residue number are similar to each other (Fig. 4-2A). In accordance, the rms deviations of the Cα atoms from the mean positions as a function of residue number and the overall three-dimensional structures are also similar to each other (Fig. 4-2B and D). In contrast, it is evident that the loop 2 conformations in the two NMR structures are significantly different, although the structural convergences of loop 2 are poor in the two NMR structures (cyan, Fig. 4-2D). We measured the $^{1}H–^{1}5N$ heteronuclear NOEs of the backbone amide groups, to assess the mobility of the individual residues in solution (Fig. 4-2C). The dip corresponding to the loop 2 region confirms that loop 2, as well as loop 1, loop 3, and loop 5, is actually mobile in solution.

To analyze the reason for the discrepancy in the loop 2 conformations in the two NMR structure calculations, we compared the distance restraints (*insets* of Fig. 4-2D). In our 2019_NMR structure calculation, there are only two tertiary distance restraints (red lines between residues 145 and 147), among the 93 distance restraints in the loop 2 region. No distance restraints were found between loop 2 and the outside region. In the previous 2014_NMR calculation, there are 39 tertiary distance restraints, consisting of 31 within the loop 2 segment and 8 between loop 2 and the outside region, among the

99 distance restraints. We used a high magnetic field at 950 MHz for the NOESY measurements and applied an automatic NOE assignment procedure to perform the accurate NOE assignments in the 2019_NMR structure calculation. We thus suggest that the non-automatic NOE assignments at 700 MHz in the previous 2014_NMR calculation led to erroneous assignments.

Table 4-1. NMR structure calculation statistics.

Total NOE distance restraints	1731
Short range ($\|i-j\| \leq 1$)	1019
Medium range ($2 \leq \|i-j\| \leq 4$)	232
Long range ($\|i-j\| \geq 5$)	480
TALOS torsion angle restraints	173
Backbone RMSD[a] (Å)	1.00 ± 0.20
Heavy atom RMSD[a] (Å)	1.79 ± 0.23
Ramachandran plot[a] (%)	
Most favored	82.5
Additional allowed	13.2
Generously allowed	3.3
Disallowed region	1.0
BMRB and PDB IDs	36264, 6K8Q

a. Calculated for the core residues, 105–131, 134–143, 154–193, and 196–217.

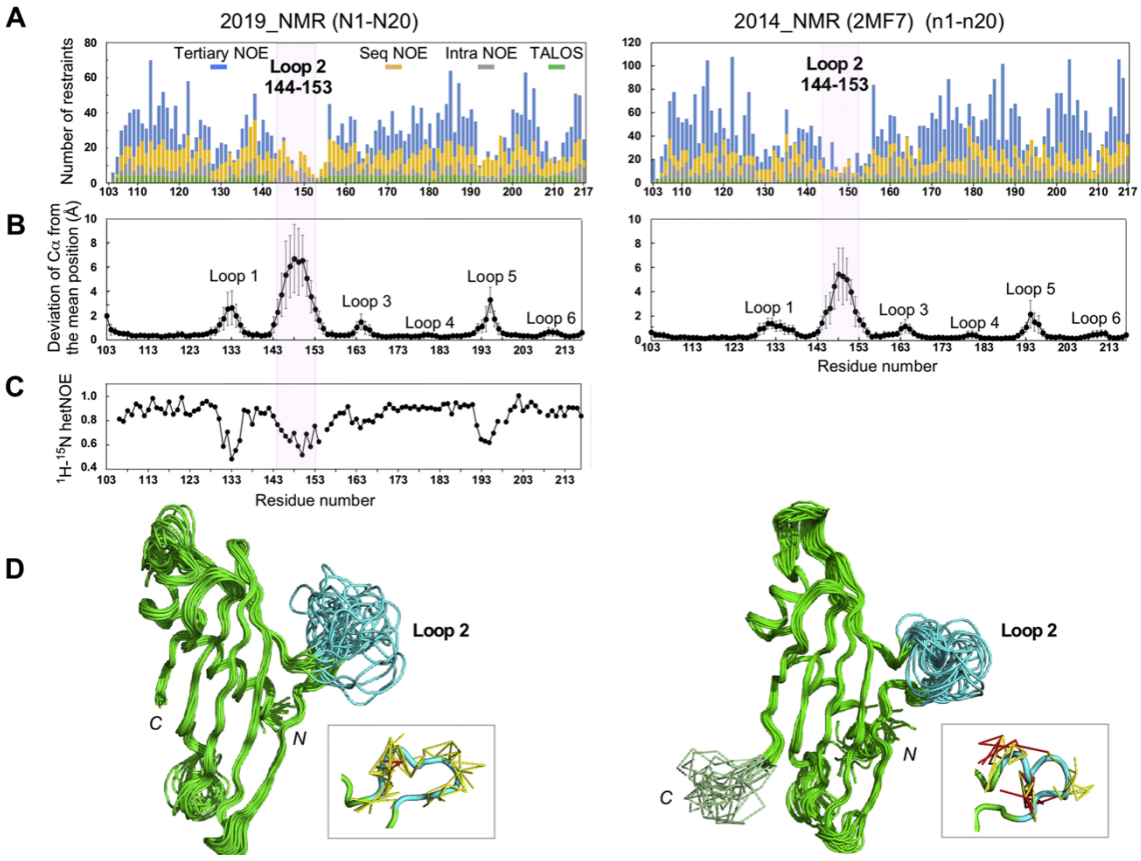


Fig. 4-2. Summary of the NMR spectroscopy of Tim21.

The previous 2014_NMR structure calculation (right) and our new 2019_NMR structure calculation (left).

A. Number of NOE-derived distance restraints and TALOS-derived torsion angle restraints, as a function of residue number. The intra-residue distance restraints ($|i-j| = 0$, gray), sequential distance restraints ($|i-j| = 1$, yellow), and tertiary distance restraints ($|i-j| > 1$, cyan), and torsion angle restraints (green). The 2014_NMR (2mf7.mr) restraints were obtained from the PDB.

B. Mean and standard deviation of the distance of the Cα atoms from the mean positions as a function of residue number.

C. ^{1}H–^{15}N heteronuclear NOEs as a function of residue number.

D. Superimpositions of the new 2019_NMR structure (left) and the previous 2014_NMR structure (right). The average rms deviations of the Cα atoms in the core region are 1.0 ± 0.2 Å for 2019_NMR, 0.80 ± 0.12 Å for 2014_NMR, and 2.5 ± 0.2 Å between the two NMR structures. The loop 2 residues 144–153 are colored cyan. The insets show the distribution of the distance restraints in the loop 2 region of model N6 (left) and model n2 (right). The yellow lines denote the sequential distance restraints, and the red lines depict the tertiary distance restraints.

4.6. Assessment of NMR structure with RDC experiment

4.6.1. The reason to assess the quality of NMR structures of Tim21 with RDC experiment

NMR restraints for the determination of protein structures includes: the internuclear distance restraints from NOE, the local angular restraints from scalar coupling, and the global, orientational restraints provided by RDCs. The NMR structures of Tim21 were determined with the NOE distance restraints and torsion angles restraints. The RDC information was not used in the NMR structure determination of Tim21. Orientational restraints provided by RDCs are independent from NOE distance restraints and torsion angles restraints, therefore we used RDC information to assess the quality of the two NMR structures of Tim21 determined in 2014 and 2019.

4.6.2. Methods of RDC experiment

First, we tested the addition of the filamentous Pf1 phage as an alignment medium, but Tim21 aggregated in the presence of 50 mM and 300 mM NaCl. We then tried polyethylene glycol (PEG)/hexanol-based medium [52]. PEG/hexanol bicelles were prepared as described in http://www.nmr2.buffalo.edu/nesg.wiki/Alignment_Media_Preparation, last accessed in June, 2019. The NMR sample for the residual dipolar coupling (RDC) experiments consisted of 50 μM ^{15}N-labeled Tim21, 4.0% PEG (C12E5, pentaethylene glycol monododecyl ether, Sigma-Aldrich), and about 5% (v/v) 1-hexanol, in 50 mM sodium phosphate buffer, pH 6.5, containing 50 mM NaCl and 9% $^{2}H_2O$. About ±12-Hz splitting of the ^{2}H lock signal was observed. The 2D ^{1}H–^{15}N HSQC and ^{1}H–^{15}N IPAP-HSQC spectra were recorded at 600 MHz in the absence and presence of PEG/hexanol bicelles [53]. Peak picking was performed with nmrDraw, to obtain the $^{1}J_{HN}$ values of the backbone amide groups. In addition to the uniform chemical shift changes (+0.045 ppm according to $\Delta\delta_{obs} = ((\Delta\delta H)^2 + (\Delta\delta N/5)^2)^{1/2}$) caused by locking on one of the two ^{2}H lines, some amide groups showed additional chemical shift variations, suggesting specific interactions with the PEG/hexanol bicelles. We selected the following 33 residues, which reside in the secondary structures and have few specific interactions with the PEG/hexanol bicelles: residues 106–108, 111–113, 115, 117, 118, 120, 122, 123, 126, 159, 161–163, 166, 167, 174, 177, 178, 181, 182, 185, 190, 199, 200–204, and 206. The program Pales, version 2.1, was used to obtain the best-fit of the

measured RDC values (RDCobs) of the 33 residues to each model of the 2014_NMR and 2019_NMR structures [54]. The correlation coefficient between RDCobs and RDCcalc was calculated, to assess the quality of the NOE-based NMR structures.

4.6.3. Results of RDC experiment

The residual dipolar couplings (RDCs) of the backbone amide groups provide orientation information related to the angle between the $^{1}H–^{15}N$ bond vectors and the magnetic field [55]. These measurements require the nonisotropic orientation of protein molecules in alignment media. We employed PEG/hexanol bicelles for this purpose, and used the RDC data to assess the quality of the NOE-based structures. The correlation of the experimental RDC data (RDC_{obs}) with the predicted RDC data (RDC_{calc}) from the protein structures is shown in Fig. 4-3. The plot was constructed using 33 selected residues, which are in the secondary structures and have no specific interactions with the bicelles (see details at *4.6.2.*). The correlation coefficient, *R*, between RDC_{obs} vs. RDC_{calc} is 0.85 for the 2019_NMR structure, and 0.87 for the 2014_NMR structure (Fig. 4-3). The comparable *R* values indicate that the core structures of the two NMR structures have similar quality. This result confirms that the non-automatic assignment of the NOE peaks does not generate practical problems in the structure determination of the stable cores of proteins. In contrast, the small chemical shift dispersion of the flexible segments could make the non-automatic NOE assignment error-prone, and may lead to incorrect local structures.

Fig. 4-3. RDC analysis of Tim21. Correlation between the observed RDCs vs. the RDCs calculated for the 2019_NMR structure (left) and the 2014_NMR structure (right).

Chapter 5. Evaluation the conformations of loop 2 determined by CCFS crystallography and NMR spectroscopy

5.1. Identification of models with severe steric clashes of loop 2 in CCFS, assuming the NMR structure models are in the CCFS

Each model of the 2019_NMR ensemble was fitted into the lattice of the MBP_{holo}<17>Tim21 crystal by the pair_fit command in PyMOL, using the C_α atoms of the Tim21 core residues, 105–131, 134–143, 154–193, and 196–217. A model is considered as "severely clashed" with the symmetrically-related MBP_{holo}<17>Tim21 molecules, if at least one C_α atom in loop 2 is close to the C_α atoms in the surrounding molecules. As contact partners, the atoms belonging to MBP and the linker parts in the same MBP_{holo}<17>Tim21 molecule were also considered. The threshold of the distance was set to 3.8 Å, which is justified as the sum of the van der Waals radii of two carbon atoms *plus* 0.4 Å. Only the C_α atoms were used for the distance calculation, considering the large mobility of the side-chain atoms of loop 2.

In the twenty 2019_NMR ensembles of N1 to N20, the ensembles of N5, N13, N14, N16, N19 and N20 were identified as the models whose loop 2 have severe steric clashes, assuming the NMR structure models are in the CCFS.

5.2. Visualization of the conformational similarity of loop 2 in different structures

We used multidimensional scaling (MDS) to visualize the conformational similarity of loop 2 in the different structures, including the 2006_crystal structure (2CIU), the 2014_NMR structure (20 models, 2MF7), three CCFS crystal structures, and the 2019_NMR structure (20 models). First, the residue numbering was adjusted to set the first Asp residue in the Tim21 sequence as 105. The 44 structures were bundled as a single PDB file, using pairs of the MODEL and ENDMDL lines. The intra_fit command in PyMOL was used to superimpose all structures, using the C_α atoms in the core residues 105–131, 134–143, 154–193, and 196–217. The rms deviations of the C_α atoms of loop 2 (residues 144–153) between all pairs of structures were then calculated, using the intra_rms_cur command in PyMOL. The resultant size 44 square matrix was input to the cmdscale command in MATLAB 2018a (MathWorks). The relative values of the first four eigen values are 1, 0.30, 0.15, and 0.11. The two largest eigen values in the MDS calculation have larger magnitudes than the remaining eigen

values, and so the first two axes are sufficient for a reasonable reproduction of the original distance matrix. The second map was created in the same way, using 21 models consisting of one selected CCFS structure and 20 models of the 2019_NMR structure. The relative values of the first four eigen values are 1, 0.29, 0.23, and 0.09, and the first two axes were used for the generation of the MDS map.

The input for multidimensional scaling (MDS) is an $N \times N$ matrix consisting of the N(N-1)/2 rms distances of loop 2, calculated using the C_α atoms between each pair of N structures. The Tim21 structures were superimposed on each other, using the C_α atoms in the core region before the rmsd calculation. MDS places each loop 2 conformation into an abstract Cartesian space to preserve the distance information as well as possible, to visualize the conformational similarity of loop 2 in various structures. We performed the MDS using a total of 44 loop 2 conformations, from the three CCFS structures, one 2006_crystal structure, twenty 2014_NMR structure models, and twenty 2019_NMR structure models. The axes corresponding to the two largest eigen values were used for plotting a 2-dimensional MDS map displaying the conformational similarity of loop 2 (Fig. 5-1A). The computed R^2, a benchmark for the extraction of variance in the original data, is as good as 0.95 using the first two dimensions. The conformations of loop 2 in the 2019_NMR structure models, N1-N20, are well separated from those of the 2014_NMR structure models, n1-n20, in the MDS map, confirming the different loop 2 conformations in the two NMR ensembles. The co-localization of the best CCFS1 and the second best CCFS2 with the 2019_NMR structure models strengthens the notion that the solution conformation of loop 2 is better represented in the 2019_NMR structure. In contrast, the conventional 2006_crystal structure and the worst CCFS3 are located between the two NMR groups. We then generated a second MDS map, using a total of 21 loop 2 conformations in the CCFS1 and the twenty 2019_NMR models, to investigate the conformational similarity in more detail. Again, the first two axes were sufficient for a reasonable reproduction of the original distance matrix (Fig. 5-1B). The computed R^2 is 0.94, using the first two dimensions. The resultant map is similar to the parent map, after rotation and inversion.

To examine the fitness in the CCFS1, each 2019_NMR structure model was superimposed on the Tim21 portion of the MBP_{holo}<17>Tim21 structure. We then counted the virtual steric clashes involving loop 2 (see 2.2.6 for details). We used a small, conservative cut-off value for the detection, to identify the steric clashes that would be unavoidable in CCFS1. Six structures were identified as being incompatible

with the CCFS1 in the MBP_{holo}<17>Tim21 crystal. Their positions in the MDS map are labeled in gray in Fig. 5-1B. For the loop 2 conformations in the six 2019_NMR models, two interpretations are possible. These six conformations actually exist in solution, but the limited volume of the CCFS in the MBP_{holo}<17>Tim21 structure eliminates the possibility of observation in the CCFS. Alternatively, the insufficient NMR information in the loop 2 region simply generates a maximal range of the loop 2 conformation, which is much wider than the true distribution in solution; i.e., the six conformations do not exist in solution. The two possible scenarios are difficult to experimentally differentiate from each other.

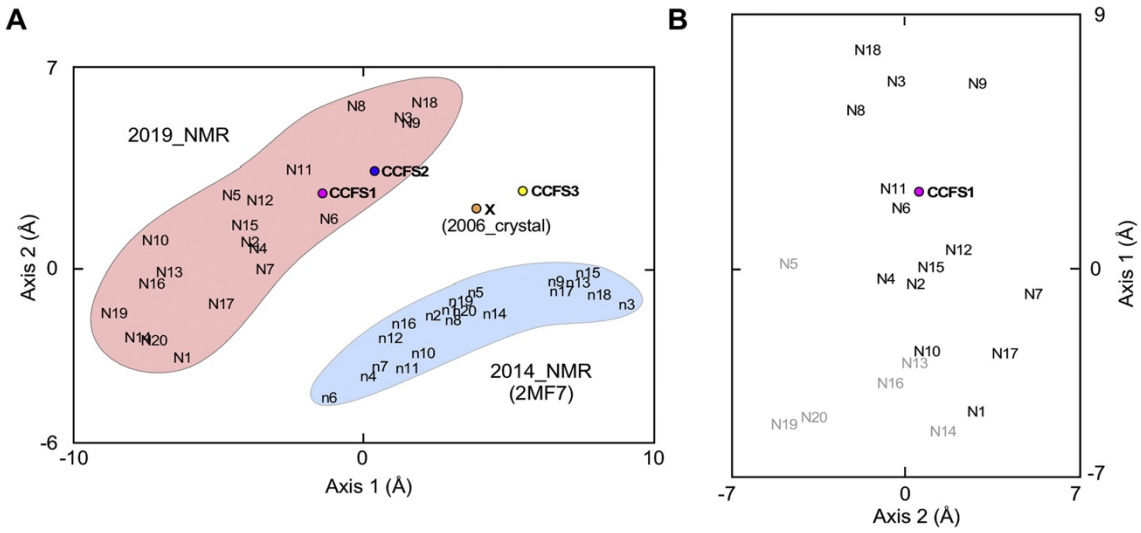


Fig. 5-1. Multidimensional scaling maps to visualize the conformational similarity of loop 2 in different structures.

A. MDS map constructed from the loop 2 conformations in the 44 structures, including the three CCFS structures (CCFS1–CCFS3), one 2006_crystal structure (X), twenty 2014_NMR structure models (n1–n20), and twenty 2019_NMR structure models (N1-N20). The loop 2 conformation in the CCFS in the MBP_{holo}<17>Tim21 crystal is referred to as CCFS1, that in MBP_{apo}<16>Tim21 is referred to as CCFS2, and that in MBP_{apo}<17>Tim21 is referred to as CCFS3.

B. MDS map constructed from the conformation of loop 2 in the selected 21 structures. The six models of the 2019_NMR structure, in which loop 2 has at least one severe intermolecular contact when virtually fitted into the CCFS1 in the MBP_{holo}<17>Tim21 crystal, have gray labels.

5.3. MD simulation of Tim21 in solution to get insight into the solution conformation of loop 2

5.3.1. Methods of MD simulation

In order to explore the conformational space of loop 2, we first constructed 10 loop models using MODELLER [56] and used them as the starting structures for molecular dynamics (MD) simulations. No structural information from crystallography or NMR was used for modeling loop 2, to eliminate any bias towards the experimental data. MD simulations were performed using Gromacs 2018.3 [57], with an AMBER ff14SB7 [58] force field. The details of the MD simulation are described in this issue [30]. From each trajectory of four independent production runs of 0.2–1.1 μs, 492 frames were recovered, and a total of 19,680 frames (i.e., 492 frames * 4 runs * 10 starting conformations) were collected. The CCFS1 and 2019_NMR structure (20 models) were added to the MD frames, and the similarity of the loop 2 conformations in the different structures was evaluated by calculating the rms deviations of the C_α atoms of loop 2, after superimposing the core residues. The definition of the core residues is described in this issue [30]. The resultant square distance matrix, with a size of about 20,000, was used in the MDS analysis. The relative values of the first four eigen values are 1, 0.72, 0.32, and 0.22. The two largest eigen values in the MDS calculation have larger magnitudes than the remaining eigen values, and so the first two axes are sufficient for a reasonable reproduction of the original distance matrix. The count of the similar frames, N, per 0.45 Å × 0.45 Å of the two-dimensional map was converted to a pseudo energy F by using the Boltzmann equation, $F = -\mathrm{RT} \ln N$, where $\mathrm{RT} = 0.593$ kcal mol^{-1}. A heat map was generated and color-coded according to the value of F, and the positions of the CCFS1 and the 2019_NMR structure (20 models) were marked on the map.

5.3.2. Results of MD simulation

We conducted the MD simulation of Tim21 in solution. The details are provided elsewhere in [30]. From the MD trajectories, a total of 19,680 frames were collected and used for the next analysis. After the addition of the CCFS1 and the 2019_NMR structures (20 models) to the MD frames, we performed MDS to visualize the conformational similarity of loop 2. The density of the structure/frame distribution in the two-dimensional MDS map was converted to an energy landscape style, in which the more frequently visited mesh unit (0.45 Å × 0.45 Å) has a bluer color (Fig. 5-2). On

the resultant heat map, the positions of CCFS1 and the twenty NMR models are marked. The loop 2 conformation in CCFS1 is located close to the center of the ensemble of the MD frames, suggesting that loop 2 in CCFS1 adopts the representative, dominant conformation of loop 2 in solution.

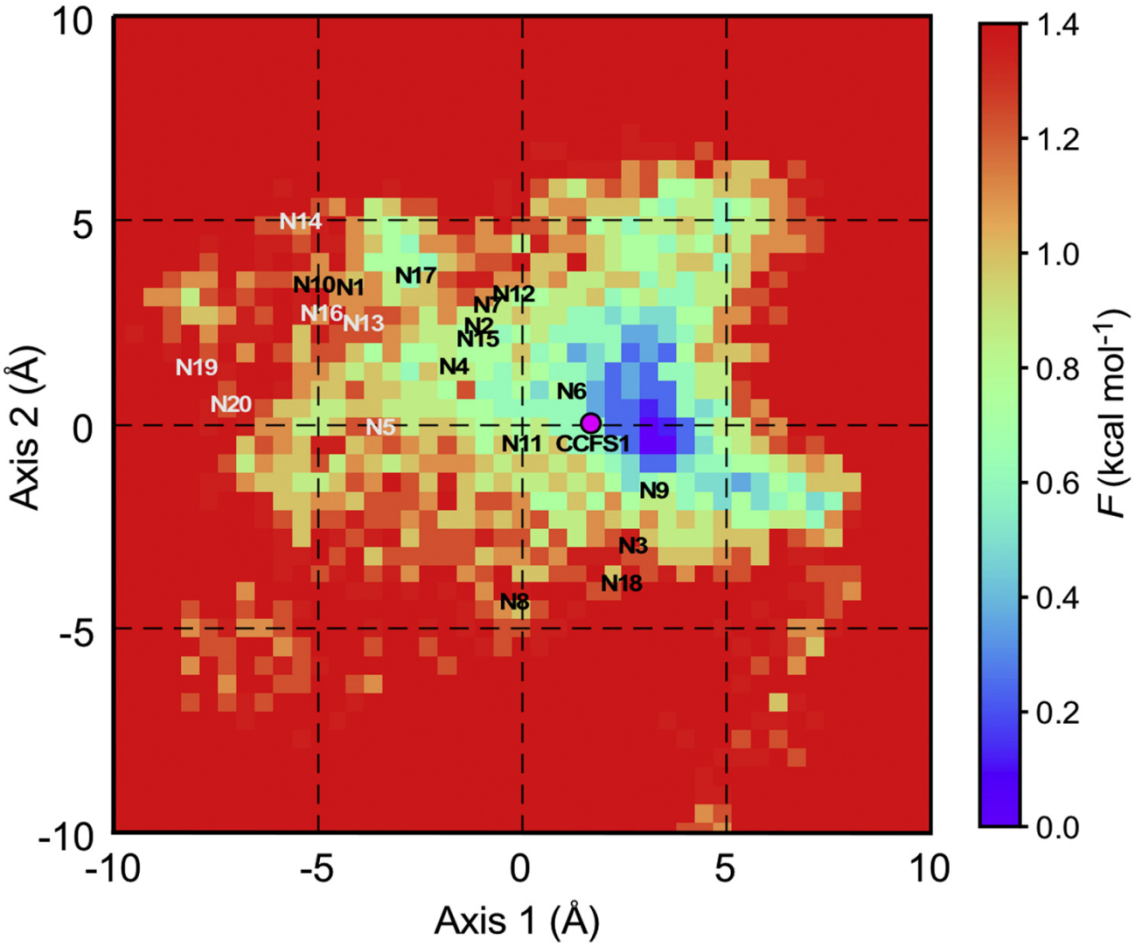


Fig. 5-2. Energy landscape of the loop 2 conformation estimated from the MD simulation of yeast Tim21. A two-dimensional MDS map was generated, to represent the structural similarity of the loop 2 conformation during the MD simulation. The distribution of the frames (i.e., conformations) in the MDS map was converted into an energy landscape style with rainbow coloring. A mesh (0.45 Å × 0.45 Å) with bluer color indicates a more frequently visited conformation of loop 2 during the MD simulation. The loop 2 conformations in the CCFS1 and the twenty 2019_NMR models are overlaid on the energy landscape.

Chapter 6. General discussion and conclusions

6.1. Discussion

The creation of the crystal contact-free space (CCFS) was proposed by our group in 2016, to study the spatial distribution of the motions of a flexible segment within a protein molecule or a mobile ligand in the bound state [23]. Here, we applied the CCFS method to estimate the crystal contact-free conformation of an intrinsically flexible loop of a protein molecule. This is not trivial, since such flexible loops tend to be deformed by crystal contacts. A 10-residue loop of yeast Tim21 adopts different conformations in the conventional crystal structure (PDB 2CIU) and the NMR structure (PDB 2MF7) (Fig. 1-2B). The discrepancy in the conformations is attributable to the crystal contacts in the crystal structure (Fig. 1-2C). Thus, we used loop 2 in the Tim21 protein as a test case for the CCFS application. First, we designed several MBP-Tim21 fusion protein constructs for the arrangement of loop 2 in the CCFS, which is expected to form between MBP and Tim21 (Fig. 2-1). We expressed a total of eight fusion protein constructs in *E. coli* host cells, containing four different lengths of the α-helical linker that directly connects MBP with Tim21, and with an optional C-terminal His tag. Among them, we successfully determined three crystal structures (Table 3-1). The comparison of the B-factor values revealed that the CCFS formed in the MBP_{holo}<17>Tim21 crystal has enough space to accommodate the loop 2 segment (Fig. 3-5 and Table 3-3, 3-4).

Contrary to our expectations, the conformation of loop 2 in the best CCFS (referred to as CCFS1) is different from those in the previously reported NMR structure (referred to as 2014_NMR). Therefore, we re-determined the NMR solution structure (referred to as 2019_NMR), and measured the residual dipolar coupling (RDC) data, which provide structural information independent of the NOE data. The good correlation of the experimental RDCs and the simulated RDCs calculated from the structures confirmed the comparable good qualities of the core structures in the two NOE-based NMR calculations (Fig. 4-3). As for loop 2, however, the conformations in the 2014_NMR structure are apparently distorted due to the incorrect assignment of the NOEs, whereas the conformation of loop 2 in the 2019_NMR structure is not properly constrained due to the absence of effective distance and torsion angle restraints (Fig. 4-2D). In general, NMR is a powerful tool to study the solution structures of globular proteins, but it has some limitations for studies of flexible segments. NOE data of such highly flexible

segments are hard to collect, due to the very small fraction of short distances between the atoms in the minor conformations [19]. Therefore, it is likely that the insufficient NMR information for the loop 2 region simply generates a maximal spatial range of the loop 2 conformation. In other words, the solution conformations of loop 2 are located somewhere in this range.

Two-dimensional MDS maps were constructed to visualize the conformational similarity of loop 2 in the different structures (Fig. 5-1). The loop 2 conformations in the previous 2014_NMR structure models, n1-n20, are well separated from those in our 2019_NMR structure models, N1-N20, in the MDS map generated from the 44 structures (Fig. 5-1A). The loop 2 conformations in the best CCFS1 and the second best CCFS2 are located in the 2019_NMR region. The loop 2 conformation (actually, a β-hairpin conformation stabilized by the crystal contacts) in the conventional 2006_crystal structure and that in the imperfect CCFS3 are both located between the two NMR regions. The second MDS map was generated from the selected 21 structures, to analyze the conformation of loop 2 in more detail (Fig. 5-1B). We examined the fitness of the individual 2019_NMR models with the CCFS1. Six structures were identified as severely clashed and incompatible with the CCFS1 (gray labels in Fig. 5-1B). As mentioned earlier, there are two possibilities. First, these six conformations exist in solution, but the limited size of the CCFS1 removes them from the CCFS. A second possibility is that the six conformations do not exist in solution, and are simply false-positives due to the lack of NMR information in the loop 2 region (Fig. 4-2A).

To complement the insufficient experimental information, we performed the MD simulation of Tim21. Structures (frames) were sampled from the MD trajectories, and the energy landscape of the loop 2 conformation was created (Fig. 5-2). Since the loop 2 in the CCFS1 is positioned roughly in the center of the energy landscape, we concluded that the crystal contact-free conformation of the loop 2 in the CCFS1 is a representative conformation of the intrinsically flexible loop 2 in solution. The MD simulation could also shed on light on the unanswered question raised by the NMR experiment. The six loop 2 conformations that are incompatible with the CCFS1 are located in the peripheral region of the energy landscape of loop 2 (gray labels in Fig. 5-2). This implies that these six conformations, as well as the other conformations in the proximity (N1, N10, N17) and on the opposite border (N3, N8, N18), are only minor and do not actually exist in solution.

In the practical use of the CCFS method, the two degrees of freedom in the fusion protein design, the length of the linker and the apo/holo forms of the MBP tag, increase the chance of successful CCFS formation in the crystals. Prior to the actual protein production, computer modeling is helpful to narrow down the possible good designs of the protein constructs. In addition, the rigid connection via a long α-helical linker guarantees the high-quality diffraction data, typically to a resolution of 2 Å or better. The use of the α-helical linker, however, limits the application to proteins that contain an α-helix at either the N- or C-terminus. Since the crystal packing of protein molecules cannot be controlled, the proper formation of CCFS in the crystal lattice must be examined experimentally. If the amplitude of motions of the flexible segment is moderate (rms displacement of an atom from the mean position in the range of 1.0–1.5 Å), then the diffusive electron density in CCFS should be observed in the *F*o-*F*c difference map and the modeling of atoms is possible. In such a case, as in the present study, the B-factor putty/sausage representation is useful to judge the formation of CCFS (Fig. 3-5). Intriguingly, the atoms surrounding the CCFS have low B-factors, suggesting the absence of stable crystal contacts with the flexible segment. If the amplitude of motions of the flexible segment exceeds the detection limit (rms displacement of an atom from the mean position >1.5 Å), then the electron density will not be observed in the *F*o-*F*c map. Another technique is required to characterize such large amplitude motions.

6.2. Conclusions

The aim of this research is to verify the feasibility of the CCFS method for determining the crystal contact-free conformation of a flexible segment. The comparison with the solution NMR structure suggested that the conformation of loop 2 of yeast Tim21 was deformed in the conventional crystal by the crystal contact effects (Fig. 1-2B and C), and thus it serves as a good test case. Here, we successfully created a CCFS around loop 2 in the MBP-Tim21 fusion protein crystal, and determined the crystal contact-free conformation of loop 2. To interpret the obtained conformation in the CCFS, we conducted the NMR structure determination and MD simulation of Tim21. The MDS analysis, visualizing the conformational similarity of loop 2, supported the notion that the crystal contact-free conformation of loop 2 is actually a representative conformation of the intrinsically flexible loop 2 in solution. Although electrostatic and water-mediated interactions with symmetrically related molecules are

unavoidable in CCFS, these unwanted effects are virtually negligible in the CCFS1 created in the MBP_{holo}<17>Tim21 crystal.

Finally, we emphasize that no single powerful technique is currently available for the structural and dynamic characterization of flexible structures in protein molecules. Solution NMR and MD simulations provide useful but incomplete information. Here, we have shown that CCFS crystallography is a third route that offers complementary information about the structural and dynamic properties of flexible structures in protein molecules.

The atomic coordinates of the crystal structures and structure factors have been deposited in the Protein Data Bank, as entries 6K7D, 6K7E, and 6K7F. The NMR data were deposited in the Biological Magnetic Resonance Data Bank under the accession number 36264, and the atomic coordinates of the NMR structures were deposited in the Protein Data Bank, as entry 6K8Q.

Appendix: Crystallographic analysis of CCFS1 mutations

1. Motivation of mutating CCFS1

We wondered whether the CCFS of MBP_{holo}<17>Tim21 (CCFS1) can be enlarged by the replacement of amino acid residues, which have potential crystal contacts with the residues of loop 2, by alanine.

2. Selection of residues for mutation

Firstly, for the consistent sequence numbering in MBP_{holo}<17>Tim21 and 2006_crystal, we renumbered the sequence of MBP_{holo}<17>Tim21 (Fig. A-1).

```
      -272       -262       -252       -242       -232       -222
MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ VAATGDGPDI

      -212       -202       -192       -182       -172       -162
IFWAHDRFGG YAQSGLLAEI TPDKAFQDKL YPFTWDAVRY NGKLIAYPIA VEALSLIYNK

      -152       -142       -132       -122       -112       -102
DLLPNPPKTW EEIPALDKEL KAKGKSALMF NLQEPYFTWP LIAADGGYAF KYENGKYDIK

       -92        -82        -72        -62        -52        -42
DVGVDNAGAK AGLTFLVDLI KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK

       -32        -22        -12         -2          8          18
VNYGVTVLPT FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG LEAVNKDKPL

        28         38         48         58         68         78
GAVALKSYEE ELVKDPRIAA TMENAQKGEI MPNIPQMSAF WYAVRTAVIN AASGRQTVDE

        88         98        108        118        128        138
ALKDAQTRIE AAAKEAAAKE AAAKAADTQL FNRAVSMVEK NKDIRSLLQC DDGITGKERL

       148        158        168        178        188        198
KAYGELITND KWTRNRPIVS TKKLDKEGRT HHYMRFHVES KKKIALVHLE AKESKQNYQP

       208        217
DFINMYVDVP GEKRYYLIK
```

Fig. A-1. Renumbered sequence of MBP_{holo}<17>Tim21 (CCFS1). The sequence from residue -281 to 87 corresponds to MBP, that from residue 88 to 104 is the 17-residue linker, and that from residue 105 to 217 is the Tim21 protein.

We selected the residues in the symmetrical molecules which located in the positions within 5 Å of the residues of loop 2 in CCFS1. The residues K-193, N1, E165, D199, F200, and I201 were selected, assuming potential crystal contacts with loop 2 (Fig. A-2).

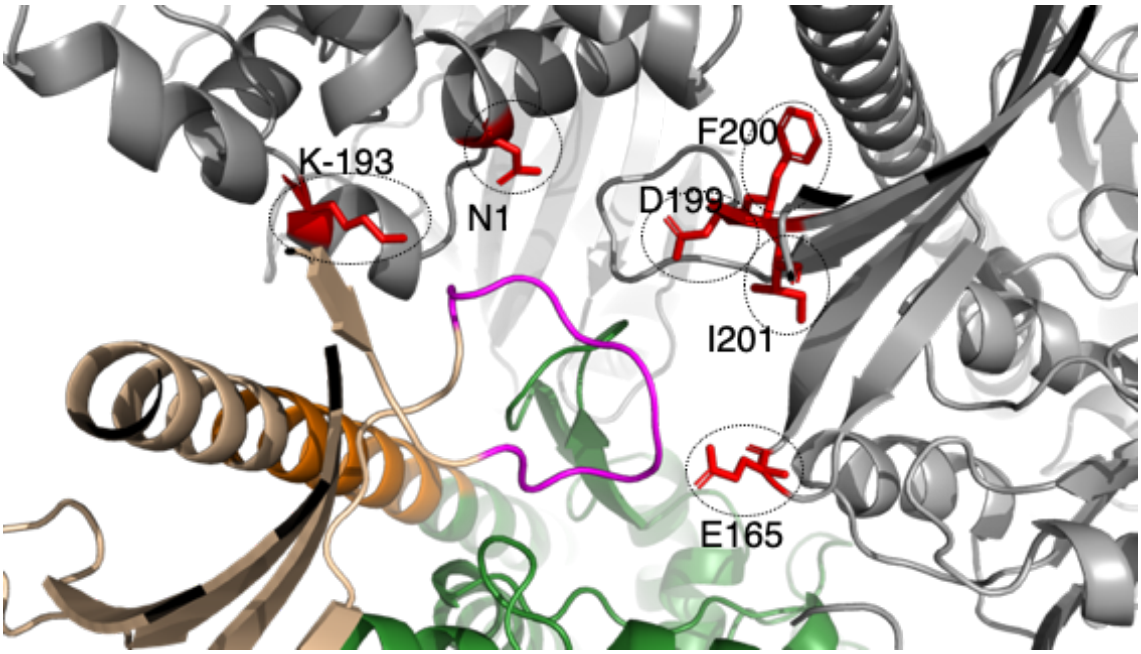


Fig. A-2. Amino acid residues in the symmetrical molecules (highlighted by red sticks) within the distance of 5 Å of the atoms in residues of loop 2 (magenta).

3. Plasmid construction

Mutations were introduced into the MBP<17>Tim21 gene with PCR. The template for PCR was the plasmid of MBP<17>Tim21_pET31b(+). KOD plus (TOYOBO) system was used in the PCR reaction, and the PCR reaction condition was thermal cycle of 11 times in the sequential condition of 367 K, 30 sec, 328 K, 30 sec and 341 K, 10 min. Below are the primers for PCR (The uppercase letters represent mutations):

K-193A: 5'primer GCg ctg tat ccg ttt acc tgg gat gcc g, 3'primer gtc ctg gaa cgc ttt gtc cgg ggt gat ttc agc c;
N1A: 5'primer GCc tat ctg ctg act gat gaa ggt ctg gaa gcg g, 3'primer ttc gag gaa ctc ttt tgc cag ctc ttt gtt cgg;

E165A: 5'primer GCa ggt cgt acc cat cac tac atg cgc ttt cat gtc g, 3'primer ttt atc cag ttt ctt ggt aga cac aat cgg;
D199A: 5'primer gCc ttt atc aac atg tat gtg gat gtt cca ggc, 3'primer cgg ttg ata gtt ctg ttt gct ctc ttt cgc;
F200A: 5'primer GCt atc aac atg tat gtg gat gtt cca ggc, 3'primer gtc cgg ttg ata gtt ctg ttt gct ctc ttt cgc;
I201A: 5'primer GCc aac atg tat gtg gat gtt cca ggc, 3'primer aaa gtc cgg ttg ata gtt ctg ttt gct ctc ttt cgc;
D199A, F200A, I201A(DFI): 5'primer gCc GCt GCc aac atg tat gtg gat gtt cca ggc, 3'primer cgg ttg ata gtt ctg ttt gct ctc ttt cgc.

The amplified liner plasmid with mutated nucleotides was ligated with Ligation High (TOYOBO) after digestion with Dpn I (TOYOBO).

By combining the mutations with the same methods described above, we constructed MBP<17>Tim21_ pET31b(+) with mutates of K-193A, N1A, E165A, D199A, F200A, I201A, D199A_F200A_I201A, K-193A_N1A, K-193A_N1A_E165A, E165A_D199A_F200A_I201A, K-193A_N1A_ D199A_F200A_I201A, and K-193A_N1A_E165A _ D199A_F200A_I201A.

4. Protein expression and purification

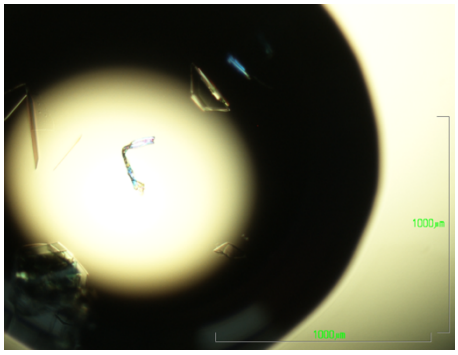
The procedure was the same as that for the MBP_{holo}<17>Tim21. The mutants of the MBP_{holo}<17>Tim21 fusion proteins were expressed in *E. coli* BL21 (DE3) cells (Agilent). The *E. coli* cells were grown in LB medium containing 100 mg L^{-1} ampicillin at 310 K, until the OD_{600} reached 0.6-0.8. After overnight induction with 1 mM isopropyl-1-thio-β-D-thiogalactopyranoside (IPTG) at 289 K, the cells were harvested by centrifugation and then disrupted by sonication. The supernatant of sonicate were adsorbed to amylose resin (New England Biolabs) in TS buffer (20 mM Tris-HCl, pH 8.0, containing 0.1 M NaCl) and affinity eluted with 10 mM maltose. The eluted proteins were further separated by gel filtration chromatography on a Superdex 200 10/300 GL column (GE Healthcare) in TS buffer. The purified proteins were concentrated to 10 - 35 mg mL^{-1} for crystallization, with an Amicon Ultra-4 centrifugal filter unit (Millipore, 30 kDa NMWL), without changing the buffer composition.

5. Crystallization, data collection, and structure determination

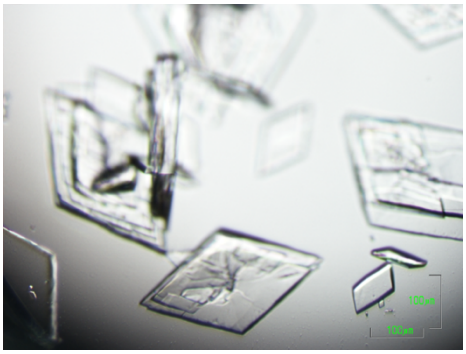
The procedure was identical to the MBP_{holo}<17>Tim21. Initial screening was performed using the sitting-drop vapor-diffusion method in 96-well plates (INTELLI-PLATE, Art Robbins Instruments), using JCSG+ Suite and PACT Suite kits (Qiagen). The crystallization conditions were optimized by a grid screening strategy, using the hanging-drop vapor-diffusion method in 24-well plates (VDX Plate with sealant, Hampton Research). Microseeding crystallization were performed. Sitting drops were set up by mixing equal volumes (0.2 µl each) of the protein solution and the reservoir solution, using an automated dispenser (Gryphon, Art Robbins Instruments). Hanging drops were prepared manually, by mixing equal volumes (1 µl each) of the protein solution and the reservoir solution. Each sitting drop was placed over 0.09 ml reservoir solution and each hanging drop was placed over 0.4 ml reservoir solution. All crystallizations were performed at 293 K. After optimization, crystals of the fusion proteins grew from a hanging drop under the conditions of 0.2 M K/Na Tartrate with16% PEG 3350 (w/v) and microseeds within 7 days. Microseeds for crystallization of mutations were prepared by serial dilutions of the same protein crystals, which grew in 0.1 M MMT buffer, pH 6.0, 25% (w/v) PEG1500 during the initial screening.

The crystals were briefly dipped into cryoprotectant solutions, containing 40% PEG3350, and then flash-cooled in liquid nitrogen. X-ray diffraction data were collected at the beamline BL44XU of SPring-8 (Harima, Japan). The diffraction data were processed with HKL2000, version 714 [32]. Initial phases were obtained by the molecular replacement method with Phaser-MR and AutoMR in PHENIX, version 1.8 [34]. The atomic coordinates of MBP_{holo}<17>Tim21 determined by us were used as the search models. Manual modeling was performed using COOT, version 0.7 [37, 38]. The coordinates of the fusion proteins plus water molecules were refined with AutoBuild and phenix.refine in PHENIX. The data collection and refinement statistics are summarized in Table A-1. The figures of the structures were drawn with PyMOL, version 2.1.0 (https://pymol.org/2/, last accessed in June, 2019).

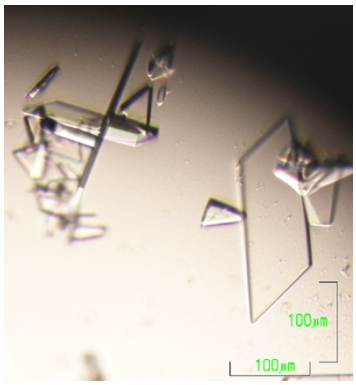
We obtained the crystals of the five mutated fusion proteins (Fig. A-3), and the diffraction data of the crystals were collected (Table A-1).



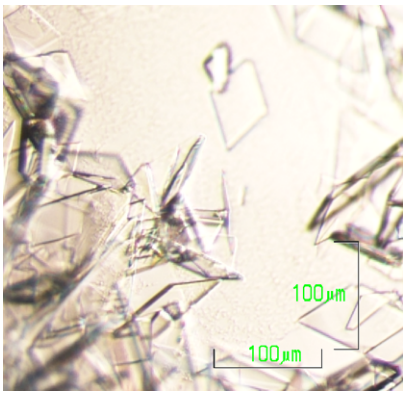
MBP_{holo}<17>Tim21_K-193A



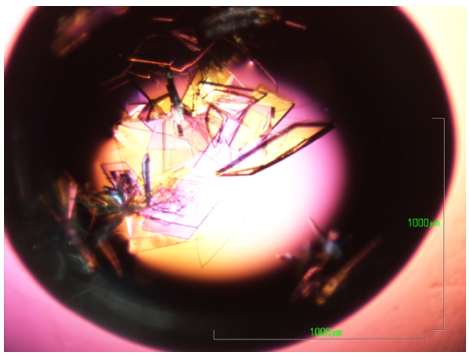
MBP_{holo}<17>Tim21_D199A



MBP_{holo}<17>Tim21_F200A



MBP_{holo}<17>Tim21_K-193A_ N1A_ E165A



MBP_{holo}<17>Tim21_K-193A_ N1A

Fig. A-3. Crystals of the five mutations of MBP_{holo}<17>Tim21. Crystallization method was Hanging Drop Vapor Diffusion; Protein solutions were 10-40 mg/ml mutation of MBP_{holo}<17>Tim21 in 20 mM Tris-HCl pH 8.0, 100 mM NaCl; Reservoir solutions were 0.2 M K/Na Tartrate, 16% (w/v) PEG 3350; 293 K incubated; Cryoprotectant was 40% (w/v) PEG 3350.

Table A-1 . X-ray data collection and refinement statistics of mutations of MBP_{holo}<17>Tim21.

MBP_{holo}<17>Tim21mutations	**K-193A**	**D199A**	**F200A**	**K-193A_N1A**	**K-193A_N1A_E165A**
Data collection statistics					
Software	HKL2000	HKL2000	HKL2000	HKL2000	HKL2000
Beamline	SPring8 BL44XU	SPring8 BL44XU	SPring8 BL44XU	SPring8 BL44XU	SPring8 BL44XU
Wavelength (Å)	0.9000	0.9000	0.9000	0.9000	0.9000
Resolution range (Å)	50.00 – 1.80	50.00 – 1.85	50.00 – 1.93	50.00 – 1.85	50.00 – 1.70
	(1.83– 1.80)	(1.88 – 1.85)	(1.96– 1.93)	(1.88 – 1.85)	(1.73 – 1.70)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell, *a*, *b*, *c*	40.776, 69.357, 169.99	40.992, 69.707, 171.590	39.030, 70.703, 170.782	41.088, 69.529, 172.424	40.993, 69.350, 171.042
Observed reflections	285732	314836	258011	295956	365880
Unique reflections	43525	43189	36497	43239	54900
Multiplicity	6.6	7.3	7.1	6.8	6.7
Completeness (%)	95.3 (93.2)	99.8 (100)	99.5 (100)	99.6 (100)	99.6 (99.9)
R_{merge} [a]	0.135 (>1)	0.088 (>1)	0.155 (>1)	0.085 (0.937)	0.082 (>1)
I/*sigma* (*I*) [b]	11.84 (2.16)	21.68 (2.05)	12.68 (2.08)	13.95 (1.68)	20.93 (2.20)
Refinement statistics					
Resolution range (Å)	39.65 – 1.80	36.53 – 1.85	38.05 – 1.93	39.97 - 1.85	42.76 – 1.70
	(1.86 – 1.80)	(1.91 – 1.85)	(2.00 -1.93)	(1.91 - 1.85)	(1.76 – 1.70)
Wilson B-factor (Å^2)	26.55	28.62	25.70	34.16	26.65
R_{work}	0.1772 (0.2664)	0.2009 (0.2975)	0.1798 (0.2469)	0.1798 (0.2621)	0.1734 (0.2512)
R_{free} [c]	0.2200 (0.2925)	0.2509 (0.3704)	0.2340 (0.3113)	0.2052 (0.3056)	0.2060 (0.2902)
Number of atoms	4348	4263	4254	4291	4508
macromolecules	3827	3797	3767	3869	3911
ligands	23	23	23	23	23
water	498	443	464	399	574
Protein residues	490	486	483	495	500
Rms deviation of bonds (Å)	0.007	0.007	0.007	0.007	0.007
Rms deviation of angles (°)	1.14	1.11	1.16	1.06	1.12
Ramachandran plot (%)					
favored	99	99	98	99	98
allowed	1	0.79	1.16	1	1
outliers	0	0.21	0.84	0	1
Clashscore	9.48	7.06	9.24	6.68	8.39
Mean B-factors, overall (Å^2)	44.80	30.00	23.10	46.10	31.40

Values in parentheses are for highest-resolution shell.

a. $R_{merge} = \Sigma\Sigma|Ii - \langle I\rangle|/\Sigma\Sigma Ii$, where *Ii* is the intensity of the *i*th observation and <*I*> is the mean intensity.

b. *I*/*sigma* (*I*) = (average of intensity) / (average of Sigma (*I*)) of unique reflections.

c. The R_{free} value was calculated for the *R* factor by using only a test set (5 %) of reflections not used in the refinement.

6. Structural similarities of the CCFS1 protein and its mutated proteins

The similarities of the crystal structures between the CCFS1 protein and the mutated proteins were examined. We aligned the Cα atoms of the core residues in CCFS1 and the mutations with pair_fit command in PyMOL. Our first thought was to align the core residues of 105-131, 134-143, 154-193, and 196-217. However, because some residues in the mutations were not modeled due to the poor electron density, we compared the structural similarities of CCFS1 and the mutations with the alignment of the Cα atoms of residues of 105-131, 134-143, 155-191, 197-215. The rmsd values between CCFS1 and each of the mutations (K-193A , K-193A_N1A, K-193A_N1A _E165A, D199A, and F200A) is 0.231, 0.145, 0.162, 0.157, 1.200 Å respectively.

7. F*o*-F*c omit maps of the mutated crystals*

*F*o-*F*c simulated-annealing omit maps were calculated with phenix.refine in PHENIX (Fig. A-4). The atoms of loop 2 were deleted before the map calculation. The low-pass filter was used to increase the signal-to-noise ratio of the omit electron density maps [23].

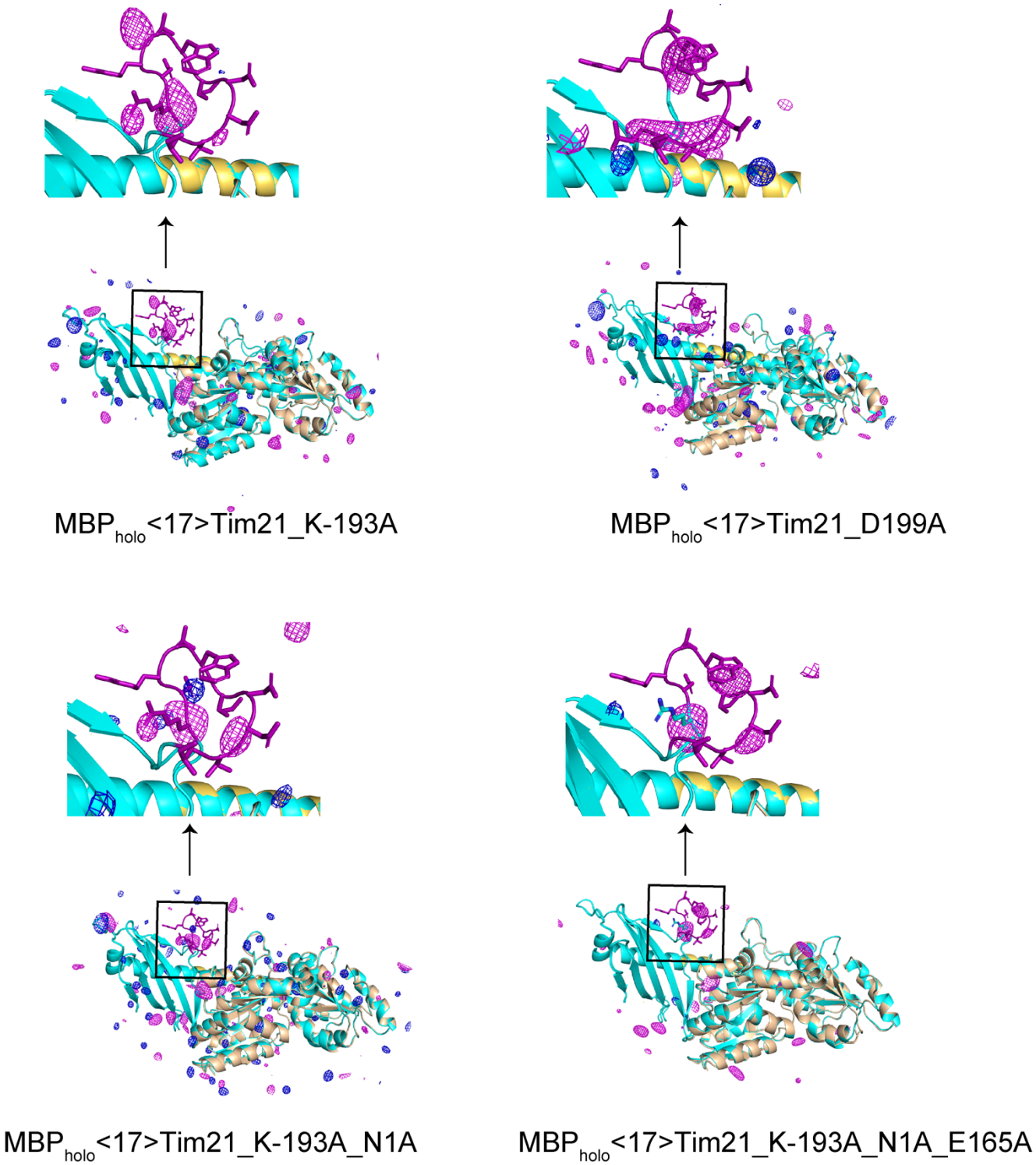


Fig. A-4. *Fo*-*Fc* simulated-annealing omit maps of CCFS1 and the mutant crystals.

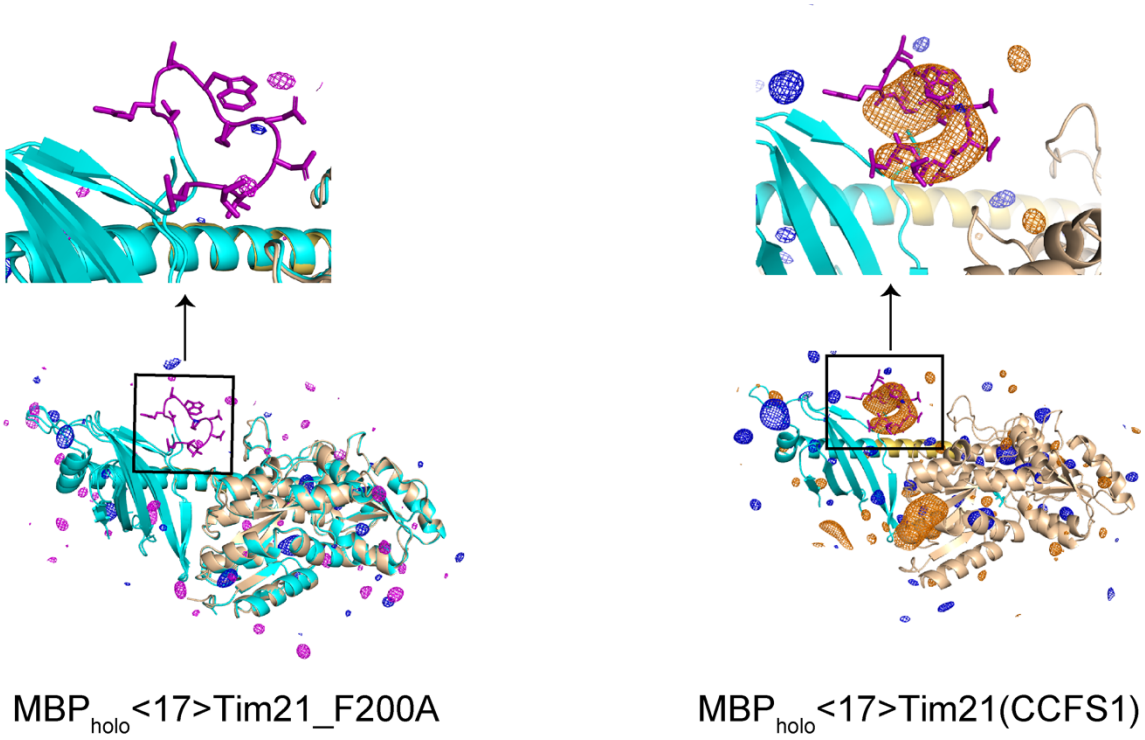


Fig. A-4. *Fo*-*Fc* simulated-annealing omit maps of CCFS1 and its mutations crystals structures.

Omit electron density maps with truncation of high-resolution reflections at 6.5 Å prior to Fourier transformation, contoured at +3σ (orange for CCFS1, magenta for mutations) and -3σ (blue).

In the figure for MBP_{holo}<17>Tim21 (CCFS1), MBP and Tim21 are shown in the cartoon representation and colored wheat and cyan, respectively. The 17-residue spacer is colored yellow orange. In the figures of the mutations, the mutant model (in cyan) without loop 2 were superposed with the model of CCFS1. For all of the figures, the structure of the deleted loop 2 is the model of loop 2 in CCFS1 and is depicted as magenta sticks with the side chains.

8. Arguments

The mutation of F200A has the worst quality of X-ray diffraction data, this might be the reason for the electron density of loop 2 in the mutation of F200A is completely invisible in the *Fo*-*Fc* omit map (Fig. A-4). For the other four mutations, the quality of X-ray diffraction data is comparable as CCFS1 (Table A-1), but the electron density of loop 2 becomes more invisible than that in the wild type (CCFS1). Especially, for the mutation of K-193A_N1A_E165A, which has a better quality of the X-ray diffraction data (Table A-1), however, has fainter electron density than the wild type does (Fig. A-4). This indicates that loop 2 becomes more flexible in the CCFS in these mutations.

Furthermore, the model of loop 2 in CCFS1 is superimposable with the electron density corresponding to loop 2 in these four mutations (Fig. A-4). It proved again that the conformation of loop 2 in CCFS1 can represent the conformation of loop 2 in solution.

9. Future work

The faint electron density of loop 2 split into pieces in the mutated CCFS1 (Fig. A-4), probably due to the large amplitude of motions of loop 2 within the enlarged CCFS. If the quality of the diffraction data is much improved, we can expect to visualize the electron density of loop 2 in the enlarged CCFS. Therefore, our future work is to acquire the high signal-to-noise ratio electron density maps, by means of screening of the crystallization conditions for larger crystals, remeasurement of the diffraction data with greater accuracy, and merge data from many crystals.

References

[1] M. Ikura, G.M. Clore, A.M. Gronenborn, G. Zhu, C.B. Klee, A. Bax, Solution structure of a calmodulin-target peptide complex by multidimensional NMR, Science, 256 (1992) 632-638.

[2] D.S. Carney, B.A. Davies, B.F. Horazdovsky, Vps9 domain-containing proteins: activators of Rab5 GTPases from yeast to neurons, Trends Cell Biol, 16 (2006) 27-35.

[3] A.E. King, M.A. Ackley, C.E. Cass, J.D. Young, S.A. Baldwin, Nucleoside transporters: from scavengers to novel therapeutic targets, Trends Pharmacol Sci, 27 (2006) 416-425.

[4] K. Teilum, J.G. Olsen, B.B. Kragelund, Functional aspects of protein flexibility, Cell Mol Life Sci, 66 (2009) 2231-2247.

[5] F. Del Cano-Ochoa, A. Grande-Garcia, M. Reverte-Lopez, M. D'Abramo, S. Ramon-Maiques, Characterization of the catalytic flexible loop in the dihydroorotase domain of the human multi-enzymatic protein CAD, J Biol Chem, 293 (2018) 18903-18913.

[6] H. Zheng, K.B. Handing, M.D. Zimmerman, I.G. Shabalin, S.C. Almo, W. Minor, X-ray crystallography over the past decade for novel drug discovery - where are we heading next?, Expert Opin Drug Discov, 10 (2015) 975-989.

[7] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank, Nat Struct Biol, 10 (2003) 980.

[8] G.T. Robillard, C.E. Tarr, F. Vosman, H.J. Berendsen, Similarity of the crystal and solution structure of yeast tRNAPhe, Nature, 262 (1976) 363-369.

[9] G. Williams, N.J. Clayden, G.R. Moore, R.J. Williams, Comparison of the solution and crystal structures of mitochondrial cytochrome c. Analysis of paramagnetic shifts in the nuclear magnetic resonance spectrum of ferricytochrome c, J Mol Biol, 183 (1985) 447-460.

[10] M. Billeter, A.D. Kline, W. Braun, R. Huber, K. Wuthrich, Comparison of the high-resolution structures of the alpha-amylase inhibitor tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals, J Mol Biol, 206 (1989) 677-687.

[11] M. Billeter, Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals, Q Rev Biophys, 25 (1992) 325-377.

[12] M. Billeter, J. Vendrell, G. Wider, F.X. Aviles, M. Coll, A. Guasch, R. Huber, K. Wuthrich, Comparison of the NMR solution structure with the X-ray crystal structure of the activation domain from procarboxypeptidase B, J Biomol NMR, 2 (1992) 1-10.

[13] F.J. Blanco, A.R. Ortiz, L. Serrano, 1H and 15N NMR assignment and solution structure of the SH3 domain of spectrin: comparison of unrefined and refined structure sets with the crystal structure, J Biomol NMR, 9 (1997) 347-357.
[14] B. Mohanty, P. Serrano, B. Pedrini, K. Jaudzems, M. Geralt, R. Horst, T. Herrmann, M.A. Elsliger, I.A. Wilson, K. Wuthrich, Comparison of NMR and crystal structures for the proteins TM1112 and TM1367, Acta Crystallogr Sect F Struct Biol Cryst Commun, 66 (2010) 1381-1392.
[15] J.J. Birktoft, G. Rhodes, L.J. Banaszak, Refined crystal structure of cytoplasmic malate dehydrogenase at 2.5-A resolution, Biochemistry, 28 (1989) 6065-6081.
[16] S.M. Pascal, T.A. Cross, Polypeptide conformational space. Dynamics by solution NMR disorder by X-ray crystallography, J Mol Biol, 241 (1994) 431-439.
[17] C.S. Rapp, R.M. Pollack, Crystal packing effects on protein loops, Proteins, 60 (2005) 103-109.
[18] E. Eyal, S. Gerzon, V. Potapov, M. Edelman, V. Sobolev, The limit of accuracy of protein modeling: influence of crystal packing on protein structure, J Mol Biol, 351 (2005) 431-442.
[19] W.F. Vranken, NMR structure validation in relation to dynamics and structure determination, Prog Nucl Magn Reson Spectrosc, 82 (2014) 27-38.
[20] S.A. Hollingsworth, R.O. Dror, Molecular Dynamics Simulation for All, Neuron, 99 (2018) 1129-1143.
[21] D. Song, R. Luo, H.F. Chen, The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins, J Chem Inf Model, 57 (2017) 1166-1178.
[22] E. Nwanochie, V.N. Uversky, Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) is the Limit, Int J Mol Sci, 20 (2019).
[23] R. Matsuoka, A. Shimada, Y. Komuro, Y. Sugita, D. Kohda, Rational design of crystal contact-free space in protein crystals for analyzing spatial distribution of motions within protein molecules, Protein Sci, 25 (2016) 754-768.
[24] D. Mokranjac, D. Popov-Celeketic, K. Hell, W. Neupert, Role of Tim21 in mitochondrial translocation contact sites, J Biol Chem, 280 (2005) 23437-23440.
[25] J. Song, Y. Tamura, T. Yoshihisa, T. Endo, A novel import route for an N-anchor mitochondrial outer membrane protein aided by the TIM23 complex, EMBO Rep, 15 (2014) 670-677.
[26] M. Sinzel, T. Tan, P. Wendling, H. Kalbacher, C. Ozbalci, X. Chelius, B. Westermann, B. Brugger, D. Rapaport, K.S. Dimmer, Mcp3 is a novel mitochondrial outer membrane

protein that follows a unique IMP-dependent biogenesis pathway, EMBO Rep, 17 (2016) 965-981.
[27] A. Chacinska, M. Lind, A.E. Frazier, J. Dudek, C. Meisinger, A. Geissler, A. Sickmann, H.E. Meyer, K.N. Truscott, B. Guiard, N. Pfanner, P. Rehling, Mitochondrial presequence translocase: switching between TOM tethering and motor recruitment involves Tim21 and Tim17, Cell, 120 (2005) 817-829.
[28] R. Albrecht, P. Rehling, A. Chacinska, J. Brix, S.A. Cadamuro, R. Volkmer, B. Guiard, N. Pfanner, K. Zeth, The Tim21 binding domain connects the preprotein translocases of both mitochondrial membranes, EMBO Rep, 7 (2006) 1233-1238.
[29] R. Bajaj, L. Jaremko, M. Jaremko, S. Becker, M. Zweckstetter, Molecular basis of the dynamic structure of the TIM23 complex in the mitochondrial intermembrane space, Structure, 22 (2014) 1501-1511.
[30] A. Srivastava, S. Bala, H. Motomura, D. Kohda, F. Tama, O. Miyashita, Conformational ensemble of an intrinsically flexible loop in mitochondrial import protein Tim21 studied by modeling and molecular dynamics simulations, Biochim Biophys Acta Gen Subj, (2019) 129417.
[31] S. Marqusee, V.H. Robbins, R.L. Baldwin, Unusually stable helix formation in short alanine-based peptides, Proc Natl Acad Sci U S A, 86 (1989) 5286-5290.
[32] Z. Otwinowski, W. Minor, [20] Processing of X-ray diffraction data collected in oscillation mode, Methods Enzymol, 276 (1997) 307-326.
[33] W. Kabsch, Xds, Acta Crystallogr D Biol Crystallogr, 66 (2010) 125-132.
[34] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr D Biol Crystallogr, 66 (2010) 213-221.
[35] F.A. Quiocho, J.C. Spurlino, L.E. Rodseth, Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor, Structure, 5 (1997) 997-1015.
[36] P.G. Telmer, B.H. Shilton, Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants, J Biol Chem, 278 (2003) 34555-34567.
[37] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr D Biol Crystallogr, 60 (2004) 2126-2132.
[38] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr D Biol Crystallogr, 66 (2010) 486-501.

[39] S. Parthasarathy, M.R. Murthy, Analysis of temperature factor distribution in high-resolution protein structures, Protein Sci, 6 (1997) 2561-2567.
[40] L.E. Kay, D.A. Torchia, A. Bax, Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, Biochemistry, 28 (1989) 8972-8979.
[41] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J Biomol NMR, 6 (1995) 277-293.
[42] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, J Biomol NMR, 5 (1995) 67-81.
[43] B.A. Johnson, Using NMRView to visualize and analyze the NMR spectra of macromolecules, Methods Mol Biol, 278 (2004) 313-352.
[44] N. Kobayashi, Y. Harano, N. Tochio, E. Nakatani, T. Kigawa, S. Yokoyama, S. Mading, E.L. Ulrich, J.L. Markley, H. Akutsu, T. Fujiwara, An automated system designed for large scale NMR data deposition and annotation: application to over 600 assigned chemical shift data entries to the BioMagResBank from the Riken Structural Genomics/Proteomics Initiative internal database, J Biomol NMR, 53 (2012) 311-320.
[45] N. Kobayashi, Y. Hattori, T. Nagata, S. Shinya, P. Guntert, C. Kojima, T. Fujiwara, Noise peak filtering in multi-dimensional NMR spectra using convolutional neural networks, Bioinformatics, 34 (2018) 4300-4301.
[46] E. Schmidt, P. Guntert, A new algorithm for reliable and general NMR resonance assignment, J Am Chem Soc, 134 (2012) 12817-12829.
[47] P. Guntert, Automated NMR structure calculation with CYANA, Methods Mol Biol, 278 (2004) 353-378.
[48] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, J Biomol NMR, 44 (2009) 213-223.
[49] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, The Xplor-NIH NMR molecular structure determination package, J Magn Reson, 160 (2003) 65-73.
[50] R.A. Laskowski, J.A. Rullmannn, M.W. MacArthur, R. Kaptein, J.M. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J Biomol NMR, 8 (1996) 477-486.
[51] N. Kobayashi, A robust method for quantitative identification of ordered cores in an ensemble of biomolecular structures by non-linear multi-dimensional scaling using inter-atomic distance variance matrix, J Biomol NMR, 58 (2014) 61-67.

[52] M. Rückert, G. Otting, Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments, Journal of the American Chemical Society, 122 (2000) 7793-7797.
[53] M. Ottiger, F. Delaglio, A. Bax, Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra, J Magn Reson, 131 (1998) 373-378.
[54] M. Zweckstetter, NMR: prediction of molecular alignment from structure using the PALES software, Nat Protoc, 3 (2008) 679-690.
[55] K. Chen, N. Tjandra, The use of residual dipolar coupling in studying proteins by NMR, Top Curr Chem, 326 (2012) 47-67.
[56] B. Webb, A. Sali, Comparative Protein Structure Modeling Using MODELLER, Curr Protoc Bioinformatics, 54 (2016) 5 6 1-5 6 37.
[57] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1 (2015) 19-25.
[58] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J Chem Theory Comput, 11 (2015) 3696-3713.