

Rational design of crystal contact-free space in protein crystals for analyzing spatial distribution of motions within protein molecules

松岡, 礼

<https://doi.org/10.15017/1654679>

出版情報：九州大学, 2015, 博士（システム生命科学）, 課程博士
バージョン：
権利関係：全文ファイル公表済



**Rational design of crystal contact-free space in
protein crystals for analyzing spatial distribution
of motions within protein molecules**

Graduate school of system life sciences,

Kyushu University

Rei Matsuoka

CHAPTER1

Rational design of crystal contact-free space in protein crystals for analyzing spatial distribution of motions within protein molecules.

<u>Abstract</u>	4
<u>Introduction</u>	5
<u>Materials and Methods</u>	7
• Protein expression and purification	
• ALDH presequence peptide and tethering to Tom20	
• Crystallization, data collection and structure determination	
• Electron density map to locate the mobile segment/peptide	
• Calculation of electron density from molecular dynamics simulation	
<u>Results</u>	13
• Rational design of the crystal contact-free space in the Tom20 crystals	
• Tethering for the full occupancy of the ligands	
• Deletion in the MBP sequence to expand the CCFS	
• Defining the location of the mobile presequence in the conventional electron density maps	
• Low-pass filter effectively located the mobile presequence in the difference electron density maps	
• Interpretation of the electron density in the CCFS	
• Design of CCFS in the AgIB crystal	
<u>Discussions</u>	21

<u>Figures</u>	24
----------------------	----

<u>Supplemental materials</u>	33
-------------------------------------	----

<u>References</u>	51
-------------------------	----

CHAPTER2

For the X-ray crystal structural analysis of *Leishmania major* STT3

<u>Introduction</u>	56
---------------------------	----

<u>Methods and Results</u>	57
----------------------------------	----

- Contraction and Protein expression
- Protein Purification
- Crystallization and diffraction experiment
- Structural analysis

<u>Conclusions</u>	66
--------------------------	----

<u>References</u>	67
-------------------------	----

<u>Abbreviations</u>	69
----------------------------	----

<u>Acknowledgements</u>	70
-------------------------------	----

CHAPTER1

Rational design of crystal contact-free space in protein crystals for analyzing main-chain dynamics of proteins

ABSTRACT:

Contacts with neighboring molecules in protein crystals inevitably restrict the internal motions of intrinsically flexible proteins. The resultant clear electron densities permit model building, as crystallographic snapshot structures. Although these still images are informative, they could provide biased pictures of the protein motions. If the mobile parts are located at a site lacking direct contacts in rationally designed crystals, then the amplitude of the movements can be experimentally analyzed. We propose a fusion protein method, to create crystal contact-free space (CCFS) in protein crystals and to place the mobile parts in the CCFS. The electron density obtained by X-ray crystallography is very poor when large amplitude motions exist. In this study, the mobile parts appear as smeared electron densities in the CCFS, by suitable processing of the X-ray diffraction data. We applied the CCFS method to a large mobile presequence peptide bound to the mitochondrial import receptor, Tom20, and a catalytically relevant flexible segment in the oligosaccharyltransferase, AglB. These two examples demonstrated the general applicability of the CCFS method to analyze the spatial distribution of motions within protein molecules.

Introduction

The dynamical view of protein molecules is essential for understanding their biological functions.¹⁻³ Although NMR spectroscopy is generally considered to be the method of choice to analyze protein dynamics,⁴ the strong non-linear distance dependency of NMR data hinders the proper analysis of internal protein motions when large amplitude motions exist.⁵ An average structure inevitably skews when a single structure is assumed. In contrast, an unbiased spatial view of motions is expected by crystallography. From a practical standpoint, however, the molecular contacts in crystals restrict the internal protein motions, unless the mobile parts are fortuitously located at a site lacking direct contacts with neighboring molecules in the crystal. Consequently, the structures determined by protein crystallography should be regarded as snapshots of fluctuating proteins. A mobile part of a protein or a moving ligand in a bound state is trapped in one possible state/conformation in the protein crystal lattice, and generates a clear electron density. These crystallographic snapshots provide valuable information on the plasticity, if multiple snapshots are available, but could be misleading by overlooking the dynamic aspects due to the limited sampling of states/conformations.^{6,7} X-ray crystallography at room temperature has recently emerged as a new method to access protein conformational ensembles.^{8,9} The contact network analysis of the room-temperature structures focuses on the analysis of the rotamer distribution of the amino-acid side chains. To investigate the motions of protein backbones, another experimental methodology is required.

It is an interesting situation when the mobile protein segment or ligand is located at a site lacking direct contacts with neighboring molecules in the crystal (Supporting Information Fig. S1). Such crystal contact-free space (CCFS) occasionally forms in protein crystals, but CCFS has not been used proactively as an analytical tool to study protein motions so far. Here we propose a fusion protein method to create CCFS in

protein crystals, and to place the fluctuating segment/ligand in the created CCFS. The idea of using fusion proteins in crystallography is not new; for example, they are used in protein crystallization as crystallization chaperones,^{10,11} but we applied the fusion protein-based approach for a distinct purpose. When few contacts with neighboring molecules exist, the electron densities of the mobile elements will become weak or even invisible. In traditional practice, such electron densities have simply been ignored, because an enforced model building for the faint electron density will not be supported by a simulated annealing-omit map,¹² or an averaged kick-omit map construction.¹³ Here, we showed that the dynamic aspects could be recovered from the abandoned information as smeared electron densities in CCFS, when combined with suitable data processing.

We applied the CCFS method to visualize a presequence peptide moving in the binding site of the Tom20 protein. Tom20 resides in the mitochondrial outer membrane, where it functions as a receptor for presequences (mitochondrial signal sequences) for efficient import of mitochondrial matrix and inner membrane proteins into mitochondria.¹⁴⁻¹⁶ The analysis of the binding mode of a presequence to Tom20 is particularly suitable as a test case, because the large amplitude motions of a presequence in the bound state are considered to be the structural basis for the promiscuous recognition of mitochondrial presequences.¹⁷ We also applied the CCFS method to a catalytically relevant flexible segment in the AglB protein. AglB is an archaeal oligosaccharyltransferase that catalyzes oligosaccharide chain transfer to asparagine residues in glycoproteins.¹⁸ These two examples demonstrated the general applicability of the CCFS method, as an analytical tool for investigating the spatial aspects of protein internal motions.

Materials and Methods

Protein expression and purification

The DNA sequence encoding maltose binding protein (MBP, residues K¹-T³⁶⁶) was amplified by PCR, using the plasmid pMAL-c5x (New England Biolabs) as the template. The reverse primer contained codons encoding Arg-Ile, to restore the original MBP sequence (K¹-T³⁶⁶R³⁶⁷I³⁶⁸), and a 20-bp overlap with the 5' region of the Tom20 sequence. The DNA sequence of the cytosolic domain (A⁶⁵-L¹²⁶) of Tom20 from *Rattus norvegicus* (UniProt: TOM20_RAT, AC: Q62760) was amplified by PCR from the expression plasmid encoding GST-Tom20, generated in the previous study.³⁵ The two DNA fragments were combined by the SOEing PCR method.³⁶ The final PCR product was cloned into *NdeI-SalI* digested pET-41b(+) (Novagen), using an In-Fusion Advantage PCR Cloning Kit (Clontech). The resultant expression plasmid was used for the production of MBP<0>Tom20. Then, one to eight amino acid residues were inserted or one to four residues were deleted at the junction of the two proteins, by the inverse PCR method.³⁷ The amino acid sequences at the junction site are summarized in Supporting Information Table S1. The deletion of $\Delta 5$ (A⁵¹ATGD⁵⁵) in MBP was also generated by the inverse PCR method. The DNA sequence (codon-optimized for *E. coli* expression) of the C-terminal globular domain (W⁴⁹¹-H⁹⁶⁷ plus a C-terminal 8×His) of AglB from *Pyrococcus furiosus* was synthesized by Life Technologies. The amplified PCR product was cloned into *SmaI-SalI* digested pET-47b(+) (Novagen). The expression plasmid thus obtained was used for the expression of MBP<0>sAglB. One (Alanine), 16, and 20 spacer residues were inserted at the junction of the two proteins by the inverse PCR method. The amino acid sequences of the long spacers are listed in Supporting Information Table S4.

The fusion proteins were expressed in the *E. coli* host strain BL21(DE3) (Stratagene). The *E. coli* cells were grown at 310 K in LB medium, supplemented with 30 mg L⁻¹

kanamycin. When the A_{600} reached 0.3-0.5, isopropyl-1-thio- β -D-thiogalactopyranoside was added at a final concentration of 0.5 mM. After 3 - 12 h induction at 289 K, the cells were harvested by centrifugation and disrupted by sonication. The MBP-Tom20 fusion proteins containing the wild-type MBP were purified by affinity chromatography on amylose resin (New England Biolabs) in TS buffer (50 mM Tris-HCl, pH 8.0, 0.1 M NaCl, with 10 mM maltose added for elution), gel filtration chromatography on a Superdex 200 10/300 column (GE Healthcare) in TS buffer, and anion-exchange chromatography on a MonoQ 5/50 column (GE Healthcare) in 20 mM Tris-HCl, pH 8.0, with a salt gradient from 0 to 1 M NaCl. In spite of the absence of maltose in the crystal structures, the MBP-Tom20 fusion proteins with the $\Delta 5$ deletion weakly adsorbed to the amylose resin. Thus, the affinity chromatography on amylose resin was also used for the $\Delta 5$ MBP-containing MBP-Tom20 fusion proteins. The eluted proteins were desalted and concentrated with an Amicon Ultra-15 centrifugal filter unit (Millipore, 30 kDa NMWL) to 10 mg ml⁻¹ in 20 mM MES buffer, pH 6.5, for crystallization. To purify MBP-sAgIB, the fusion proteins eluted from the amylose resin were denatured by 8 M urea, refolded by gel filtration chromatography on a Superdex 200 HiLoad 26/60 column in TS buffer, and finally purified by anion-exchange chromatography on a MonoQ 5/50 column in 20 mM Tris-HCl, pH 8.0, with a salt gradient from 0 to 1 M NaCl. The eluted proteins were desalted and concentrated with an Amicon Ultra-15 unit (50 kDa NMWL) to 10 mg/ml in 20 mM MES buffer, pH 6.5, for crystallization.

ALDH presequence peptides and tethering to Tom20

The presequence (G¹²PRLSRLLS²⁰) is derived from rat mitochondrial aldehyde dehydrogenase (ALDH, AC: P11884). The residue numbering starts at 12, since the sequence corresponds to the C-terminal half of the ALDH presequence. A cysteine residue, Cys²⁴, was attached to the C-terminus via a three-residue spacer, Y²¹A²²G²³.²¹ The peptide and its sequence variants were synthesized with an N-terminal acetyl group

and a C-terminal amide group, by Hokkaido System Science (Sapporo, Japan). An intermolecular disulfide bond was formed between the Cys²⁴ residue of the presequence and a single cysteine residue (Cys¹⁰⁰) of Tom20 in high pH buffer (0.1 M Tris-HCl, pH 9.0). Note that MBP contains no cysteine residues. The formation of the disulfide bond was monitored by native-PAGE and reverse-phase HPLC analyses in the absence and presence of 1 mM dithiothreitol (DTT).

For the preparation of an N-terminally iodine-labeled pALDH peptide, *N*-succinimidyl-4-iodobenzoate was synthesized by carbodiimide coupling.³⁸ Briefly, to a solution of 4-iodobenzoic acid and *N*-hydroxysuccinimide in dry dichloromethane, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride was added. The mixture was refluxed for two hours. After extraction with a saturated bicarbonate solution, the organic layer was concentrated and the resulting residue was washed with a small volume of chloroform. The product was obtained as a white powder. The purity was checked by ¹H and ¹³C-NMR. The pALDH peptide for iodine labeling was synthesized without an acetyl group at the N-terminus. The α -amino group of the peptide was reacted with 0.1 M *N*-succinimidyl-4-iodobenzoate in 40 % dimethylsulfoxide. The modified peptide was purified by reverse-phase HPLC. The modification was confirmed by monitoring the increase of the molecular mass corresponding to the 4-iodobenzoyl group, IC₆H₄CO, using MALDI-TOF-MS. The iodinated pALDH peptide was tethered to Δ 5MBP<+4>Tom20 through a disulfide bond in the high pH buffer. The resultant Δ 5MBP<+4>Tom20-SS-(*N*-iodo)pALDH was purified by anion exchange chromatography.

Crystallization, data collection, and structure determination

Initial crystallization screening was performed by the sitting drop vapor diffusion method. To search for optimized crystallization conditions, we performed grid screening and additive screening (Hampton Research). X-ray diffraction data were

collected at beamlines BL32XU, BL44XU, and BL26B1 of SPring-8 (Harima, Japan), and beamlines PF-BL1A, PF-BL5A, PF-BL17A, and AR-NW12A of the Photon Factory (Tsukuba, Japan). Diffraction data were collected with a wavelength of 0.9000 or 1.000 Å, and for the detection of the anomalous iodine signal, with a 1.6000 Å wavelength. The crystals were cooled to 95 - 100 K. The final crystallization conditions and the cryo-conditions are summarized in Supporting Information Tables S2, S3, and S4.

The diffraction data were processed with the program *HKL2000*.³⁹ Initial phases were obtained by the molecular replacement method with the program *PHASER* in *CCP4*,⁴⁰ using the structures of MBP (PDB code 1ANF for the maltose-bound form of MBP,⁴¹ and PDB code 3PUY for the apo form of MBP,⁴²), Tom20 (PDB code 2V1T, without the presequence peptide,¹⁷), and sAglB (PDB code 2ZAI, without the THL segment,⁴³) as the search models. The coordinates of the fusion proteins plus water molecules bound to MBP were refined with the program *REFMAC5* ver. 5.7 in *CCP4*.⁴⁴ The final manual modeling was performed using the program *COOT*.⁴⁵ To create the anomalous difference Fourier map, the phases were calculated from the model of $\Delta 5\text{MBP}<+4>\text{Tom20}$ in the $\Delta 5\text{MBP}<+4>\text{Tom20-SS-pALDH}$ crystal, by the program *SFALL* in *CCP4*, and the map was generated by the program *FFT* in *CCP4*. For the tethered complex crystals with mutated presequence peptides, pALDH(L18S) and pALDH(L19S), pseudomerohedral twinning was detected using the program *Xtriage* in the *PHENIX* program package.⁴⁶ *Xtriage* indicated that the twin law was a mirror index ($h,-k,-h-l$) for $P2_1$. The estimated twin fractions were 0.18 for L18S and 0.25 for L19S. The detwinned data from *DETWIN* in the *CCP4* program package,⁴⁷ were used for the refinement calculations. Data collection and refinement statistics are summarized in Supporting Information Tables S2, S3, and S4.

Electron density map to locate the mobile peptide/segment

The 20 refinement calculations were rerun with 20 different free test sets, using the refined model of the CCFS scaffolds (fusion protein plus water molecules bound to MBP). No model was placed for the smeared electron density corresponding to the presequence peptide or the THL segment, and no additional water molecules were picked. A sigma A-weighted F_o-F_c (mF_o-DF_c) difference electron density map was calculated, using Fourier amplitudes (DELFWT) and phases (PHDELWT) outputted by the program *REFMAC5*. The program *SFTOOLS* in *CCP4* was used to perform the truncation and zero padding of the high-resolution structure factor amplitudes, by using the keywords, “SELECT RESOL < 7” and “CALC COL DELFWT = 0”. Since the truncation of high-resolution reflections results in a coarse mesh size of the electron density map, zero padding maintains the original mesh spacing size simply for cosmetic reasons. The truncation threshold was empirically selected. This value, $r_{\min} = 7 \text{ \AA}$, seems suitable for visualizing the α -helical structures. *FFT* was used for map generation.

The figure generation was performed with the program *PyMOL* (Schrödinger). *PyMOL* was also used for the model building. An ideal α -helical conformation was assumed for the connector helix between MBP (PDB code 4MBP, ⁴¹) and Tom20/sAglB. The structure of Tom20 was adopted from one of the three crystallographic snapshot structures (PDB code 2V1T, ¹⁷). The structure of sAglB was the C-terminal globular domain of the *Pyrococcus furiosus* AglB (PDB code 2ZAI, ⁴³). Multiple structure fitting and rmsd calculations in batch mode were performed with the program *ProFit*, version 3.1 (Martin, A.C.R. and Porter, C.T., <http://www.bioinf.org.uk/software/profit/>, 2009).

Calculation of electron density from molecular dynamics simulation

The 50,000 snapshots of the Tom20-pALDH complex obtained from the previous replica-exchange MD simulation, ²⁷ were classified into four major clusters with different probabilities (I: 4.76%, II: 5.13%, III: 80.0%, IV: 4.23%). Based on the probabilities, 100 structures were randomly chosen and superimposed to fit the 40 Ca

atoms (F⁶⁹-E⁷⁹+Y⁸⁶-L¹¹⁴) of the first three α -helices of Tom20. The coordinates of the presequence were extracted and combined into one file. The occupancy and isotropic B factor of all atoms were uniformly set to 0.01 and 1.00, respectively. The program *SFALL* was used to calculate the structure factors from the coordinate file, and the structure factors were used to generate a simulated map with a resolution limit of 7 Å, using the program *FFT*. The GRID SAMPLE parameter was set to 8 for fine grid spacing.

Results

Rational design of the crystal contact-free space in the Tom20 crystals

We developed a special protein design to create CCFS intentionally in protein crystals. The basic idea is the fusion of a target protein with a tag protein via a rigid linker (Fig. 1A). Rigid linkers are usually used to effectively reduce unfavorable inter-domain interactions,¹⁹ but in this study, they were used to create isolated space within the framework formed by the rigid fusion protein. We refer to the fusion protein (except for the flexible segment/ligand) as the ‘CCFS scaffold’ (Fig. 1B). Since the crystal packing mode of protein molecules is uncontrollable, the CCFS scaffold must be designed to ensure the formation of an inaccessible space to other molecules within a protein molecule.

We selected *Escherichia coli* maltose binding protein (MBP) as the fusion partner of the Tom20 protein (Fig. 1C). We focused on the long C-terminal α -helix of MBP, which was seamlessly fused to the N-terminal α -helix in the cytosolic domain of Tom20 to form a rigid helical connector structure. To place the presequence binding site in the CCFS, the number of amino acid residues inserted or deleted at the junction site in the connector helix provides one degree of freedom, which determines the relative orientation between MBP and Tom20 within the same molecule (Fig. 1A). Prior to the protein expression experiment, we performed a preliminary computer modeling study to estimate the optimal spacer length in the connector helix, by assuming an ideal α -helical conformation. A sufficient amount of space was predicted to form between MBP and Tom20 with the spacer lengths of -4, 0, +4, and +8.

We constructed a series of MBP-Tom20 fusion proteins with spacers containing different numbers of inserted/deleted residues (-4 to +8) at the junction site (Supporting Information Table S1). The basic spacer sequence, KEALQELA, was designed according to the solvent *exposed-buried* pattern, *eeebeebe*, based on the helical propensity and hydrophobicity of the amino acid residues. All of the constructs were

successfully expressed in the recombinant *E. coli* cells (Supporting Information Fig. S2). We purified the proteins and performed crystallization screening. Protein crystals were obtained for the constructs with spacer lengths of 0, +2, +4, +7, and +8. Among them, the crystal structures were determined for the constructs with spacer lengths of +2 and +4 to resolutions of 2.9 and 1.95 Å, respectively (Fig. 2A and Supporting Information Table S2). We examined the formation of CCFS in the crystals of MBP<+2>Tom20 and MBP<+4>Tom20, where <+2> and <+4> denote the 2-residue and 4-residue insertions in the connector helix, respectively. In the MBP<+2>Tom20 crystal, an adjacent molecule blocked the Tom20 binding site, whereas the volume of space around the Tom20 binding site appeared to be sufficient to accommodate a presequence peptide in the MBP<+4>Tom20 crystal (Fig. 2B). Thus, we selected the +4 spacer length for further analysis. This result also confirmed the benefit of performing the computer model building study before the actual experiments.

Tethering for the full occupancy of ligands

In the cases of weak-affinity ligands, the problem of partial ligand occupancy must be considered. The dissociation constant of the presequence peptide derived from rat aldehyde dehydrogenase (pALDH) for Tom20 is about 250 μM.²⁰ The pALDH peptide was tethered onto Tom20, to ensure the full occupancy of the presequence in the binding site (Fig. 1C). We added a cysteine residue at the C-terminus of pALDH, to form an intermolecular disulfide bond with the single cysteine residue in the fusion protein. The spacer length (3 residues) between pALDH and the C-terminal cysteine was optimized in the previous peptide library experiment.²¹ The tethered complex is referred to as MBP<+4>Tom20-SS-pALDH. The crystals diffracted to a resolution of 1.9 Å (Supporting Information Table S2).

Deletion in the MBP sequence to expand the CCFS

The structure of the CCFS scaffold, MBP<+4>Tom20, was determined by the molecular replacement method. The extra electron density corresponding to the presequence was clearly observed near the binding site of Tom20, and so a model of the presequence peptide was built (Supporting Information Fig. S3). Close examination of the structure suggested the presence of undesired intramolecular interactions between the presequence peptide and MBP. To enlarge the CCFS, we generated a new construct, $\Delta 5$ MBP<+4>Tom20, to delete the 5-residue segment, A⁵¹ATGD⁵⁵, in the MBP sequence (Fig. 1C). The crystal of $\Delta 5$ MBP<+4>Tom20-SS-pALDH diffracted to a resolution of 1.8 Å (Supporting Information Table S3). During the purification, we noticed that the $\Delta 5$ deletion reduced the maltose binding to MBP. The MBP in the crystal did not contain a maltose molecule after the structure refinement of the CCFS scaffold. The apo form of MBP adopts a more open conformation than the maltose-bound form, which enlarged the CCFS, in addition to the simple expansion effect of the $\Delta 5$ deletion.

Defining the location of the mobile presequence in the conventional electron density maps

When the difference map was generated ²², however, we did not see any significant electron densities in the binding site of Tom20 (Fig. 3A). For confirmation, we attempted the placement of a helical model in the binding site of Tom20 and the subsequent refinement. As expected, the B factors of the presequence atoms exceeded 100 Å², and were significantly larger than those in the crystallographic snapshots (15 – 50 Å²).^{17,20} After simulated annealing-omit map construction for removing model bias, no meaningful electron densities remained in the binding site of Tom20. Taken together, the invisibility suggested the large mobility of the presequence peptide in the binding site, beyond the limit of the conventional treatment.

To find the location of the presequence, an iodine atom was attached to the

N-terminus of the pALDH peptide. *N*-Succinimidyl-4-iodobenzoate was synthesized and used to modify the N-terminal α -amino group of the presequence (Supporting Information Fig. S5). An anomalous difference map clearly showed the blurred electron density of the iodine atom close to the expected N-terminal position, likely reflecting the movement of the presequence in the binding site (Fig. 3B). This suggests that the presequence peptide was actually present, at least partly, in the binding site of Tom20. Our interpretation is that an extra, weak interaction of the 4-iodobenzoate group with the Tom20 protein increased the probability of the iodine atom existing at a particular position close to Tom20 (inset of Fig. 3B). Owing to the enhanced detection ability of the anomalous map, the iodine atom was located at the partially-occupied site.

Low-pass filter effectively located the mobile presequence in the difference electron density maps

We then examined various data-processing techniques, to determine whether they would be effective to find the location of mobile ligands/segments. We tested the map sharpening to locate the unmodeled presequence (Fig. 4). To our disappointment, only the noise pattern changed due to the over-sharpening of the residual amplitudes at high resolution, and ripples resulted from the Fourier series termination (Fig. 4B). This is because the blurred local electron density corresponding to a mobile region, such as the presequence, cannot be enhanced effectively using the current overall sharpening method, which assumes one rigid body in the asymmetric unit.²³ Thus, for the selective enhancement of the mobile part, map sharpening (with a negative *b* value) was applied for the low-resolution reflections, and map blurring (with a positive *b* value) was used for the high-resolution reflections for noise suppression. We found that the simultaneous sharpening and blurring improved the quality of the electron density in the binding site of Tom20 (Fig. 4C). A simple blurring was also effective (Fig. 4D). Here, we adopted a simpler form of map blurring using the truncation (to be precise, zero

padding) of the high-resolution reflections to avoid unnecessary amplitude modifications, which would distort the shape of the smeared electron density in CCFS (Fig. 4E). The improvement in the signal-to-noise ratio depends on the choice of the resolution limit (r_{\min}) of the truncation for map generation. We found that $r_{\min} = 7 \text{ \AA}$ was optimal for visualization of the α -helical structure of the presequence peptide (Supporting Information Fig. S6). The optimal value of r_{\min} may vary, depending on the secondary/local structures of the target. Lastly, as expected, the truncation had no obviously favorable effects on the visualization of the immobile helix in MBP (Supporting Information Fig. S6).

Interpretation of the electron density in the CCFS

Tom20 recognizes a 5-residue consensus motif, $\phi\chi\chi\phi\phi$, where ϕ means a hydrophobic residue and χ means any amino acid residue, embedded at various positions within the mitochondrial presequences.^{21,26} We performed a mutagenesis study to validate the electron density in the CCFS. For each mutation, one of the three hydrophobic leucine residues at the ϕ positions was substituted with a hydrophilic serine residue. These substitutions resulted in affinity reductions ranging from 40- to 200-fold.²⁰ The crystals of the three mutant complexes diffracted to resolutions of 2.1 \AA , 1.8 \AA , and 1.6 \AA (Supporting Information Table S3). Although the space group and the number of molecules in the asymmetric unit changed, the CCFS invariably formed. As expected, the electron density corresponding to the presequence part almost disappeared, but that of the tether part remained visible in the binding site (Fig. 5B). Even though the mutated peptides were still located in the CCFS, without the interactions with Tom20, their large motions were beyond the detection limit of the truncated difference map.

We then compared the electron density in the CCFS with the three crystallographic snapshots of the presequence previously obtained by conventional crystallography.^{17,20} The electron density in the CCFS, excluding the tether region, superimposed well on the

three snapshots and appeared to correspond to the crossover volume of the moving presequence in the binding site (Fig. 6A).

The molecular dynamics (MD) simulations of the Tom20-pALDH complex were previously performed in water without tethering.²⁵ The presequence peptide stably bound to Tom20 during the replica-exchange MD simulation. The electron density of pALDH, calculated from the MD simulation, overlapped well with the electron density (with the exclusion of the tether region) in the truncated difference map, but a small displacement existed between the two densities (Fig. 6B). A likely explanation is the non-negligible effects of the tethering or a cryo-induced shift (see below in discussion).

From the MD simulation, we calculated the amplitude of motions of the presequence in the binding site, with respect to each C α atom. The average displacement from the mean position is about 1.3 Å (isotropic B-factor value of 130 Å²) for the Tom20-binding consensus and tether residues. The fewer direct crystal lattice contacts may increase the rigid body disorder of the target protein. In fact, the average B-factor value of the Tom part (70 Å² for C α atoms) is larger than that of the MBP part (30 Å²). However, considering the significantly higher value of the estimated average B-factor of the presequence (130 Å²), the CCFS method is useful to extract the aspects of the presequence dynamics in the binding site of Tom20.

Design of CCFS in the AglB crystal

To demonstrate the general applicability of our approach, we applied the CCFS method to another protein. Oligosaccharyltransferase (OST) is a membrane-bound enzyme that catalyzes oligosaccharide transfer to asparagine residues in glycoproteins.¹⁸ The C-terminal globular domain of OST possesses a binding pocket that recognizes Ser and Thr residues in the N-glycosylation consensus sequence, Asn-X-Ser or Asn-X-Thr, where X is a nonproline residue.^{26,27} The Ser/Thr pocket is part of a segment called the “Turn-Helix-Loop (THL)” in the globular domain.²⁸ A large variety of snapshot

conformations of the THL segment were observed in the crystals of different OST proteins, suggesting its intrinsic flexibility in solution. In fact, the THL segment of one OST was shown to be mobile in solution by an NMR relaxation study.²⁸ The restriction of the flexibility of the segment by an engineered disulfide bond in another OST reversibly suppressed the enzymatic activity, suggesting that the flexibility of the THL segment is essential for the N-glycosylation sequon recognition.²⁸

We selected an archaeal oligosaccharyltransferase protein, AglB, from *Pyrococcus furiosus*. The C-terminal globular domain of AglB (soluble AglB or sAglB) contains a long N-terminal α -helix, which was used as part of the rigid connector helix (Fig. 7A). We first performed a model building study to deduce the appropriate length of the spacer between MBP and sAglB. The spacer lengths of +1, +16, and +20 were chosen for the proper placement of the THL segment in the CCFS. The amino acid sequences of the spacers were poly-alanine-based sequences containing glutamate-lysine pairs.²⁹ Crystals were obtained for MBP<+16>sAglB and MBP<+20>sAglB, which diffracted to resolutions of 2.6 Å and 2.1 Å, respectively (Supporting Information Table S4). Close inspection of the MBP<+16>sAglB structure revealed that the relative orientation of MBP and sAglB largely deviated from the expected arrangement (Supporting Information Fig. S7 and S8), due to the disruption of the helical structure in the long spacer. In contrast, the MBP<+20>sAglB structure had similar geometry to the model, and was used for further study.

In the conventional difference Fourier map of MBP<+20>sAglB, there were no continuous electron densities corresponding to the THL segment in the CCFS (Fig. 7B). In the truncated and FreeR-averaged difference map, clear electron densities corresponding to the turn (3 res) and subsequent helix (9 res) were visible, but the loop (12 res) remained invisible except for 1 or 2 C-terminal residues (Fig. 7C). The CCFS formed at the interface between symmetrically related MBP-sAglB molecules appeared to have sufficient volume, as planned (Fig. 7D). For reference, the r_{\min} dependency and

the effect of the choice of the free data set are shown in Supporting Information Figs. S9 and S10. These results showed that the turn and helix parts have narrower spatial distributions than that of the loop.

Discussion

We have proposed a fusion protein method for the creation of crystal contact-free space (CCFS) in protein crystals (Figs. 5B and 7D), independent of the protein molecule packing mode in crystals (Fig. 1). We selected MBP as a fusion partner among many tag proteins, to construct a rigid CCFS scaffold. We successfully used α -helical spacers up to 20-residues in length, to fuse the C-terminal α -helix of MBP and the N-terminal α -helix of the target proteins firmly. The present study has confirmed that the CCFS in protein crystals can be rationally created with reasonable efforts. In addition, a preliminary model building study prior to protein expression is quite useful to reduce the number of experimental trials.

The presequence peptide in an α -helical conformation with relatively large amplitude motions in the binding site of Tom20 was hardly visible in the created CCFS in the conventional difference electron density map (Fig. 3A), but was detected as a continuous, elongated electron density in the difference map with the truncation of high-resolution reflections (Fig. 5A). The dynamical segment in the C-terminal globular domain of the AglB protein in the created CCFS was also visible in the truncated difference map (Fig. 7C). We obtained an averaged rmsd value of 1.3 Å from the mean positions during the MD simulation, as the amplitude of the presequence peptide movement (Fig. 6B). This estimated amplitude value suggested that the difference map with truncation has the potential to analyze motions up to 1.5 Å rmsd from the mean positions. Obviously, without any proactive CCFS designs, the truncation of high-resolution reflections prior to map generation. In fact, the truncated difference map revealed the mobile C-terminal tail segment of sAglB, in addition to the THL segment (Fig. 7C). The size of the electron density suggested that it corresponds to the first 4 or 5 residues of the invisible 12-residue segment (4 sAglB residues and 8 His-tag residues). Finally, although we used truncation at $r_{\min} = 7$ Å in this study, the resolution limit of the diffraction measurement (d_{\min}) should be as high as possible, for the accurate

molecular replacement and refinement of the CCFS scaffold. In our case, d_{\min} was 1.6 - 2.6 Å (Supporting Information Tables S3 and S4).

The bulk solvent is the disordered water and ions filling the intermolecular space between protein molecules, and it attenuates low-resolution reflections. Bulk solvent correction is necessary to account for the attenuation effect.^{30,31} The bulk solvent correction may have significant impacts on the quality and accuracy of the electron density in the CCFS, since the truncated difference maps are generated only from low-resolution data. We compared the map generated by the program *REFMAC5* (Fig. 5A) with those generated by the programs *CNS* and *PHENIX*, using the same CCFS scaffold (MBP-Tom20) structure refined by the program *REFMAC5* (Supporting Information Fig. S9). The three programs produced similar shapes and volumes of the electron density in the CCFS. This comparison indicated that different implementations of the bulk solvent correction did not cause significant differences in the electron density in the CCFS. Improvements of bulk solvent correction methods will benefit CCFS crystallography in the future.

Our current working hypothesis is that “a rapid equilibrium of multiple states with partial recognitions” is the molecular basis for the promiscuous binding of the Tom20 receptor to diverse mitochondrial presequences with nearly equal affinities^{17,20}. The smeared electron density in the difference electron density map corresponded to the partially-overlapped volume among the multiple poses of the presequence helix (Fig. 6A), but did not allow us to estimate the amplitude of the presequence helix motions in the binding site. Thus, the results presented in this work are mostly confirmatory of our previous structural studies. In the future, better diffraction measuring and data processing will improve the signal-to-noise ratio of the difference electron density maps, and reveal the spatial distribution of the moving α -helical presequence peptide experimentally. Obviously, the influences of the non-physiological solution conditions required for crystallization must be considered when interpreting the electron density in

the CCFS.

In protein crystallography, protein crystals are routinely frozen prior to X-ray diffraction measurements, to avoid radiation damage. Thus, the presequences in the CCFS do not literally ‘move’, but thermal motions in solution are observed as static disorder in a glassy state in frozen crystals. Consequently, the electron density in the CCFS obtained at a cryogenic temperature does not fully represent the kinetically trapped ensemble in a room-temperature equilibrium, because a substate population shift is unavoidable during the flash-cooling of crystals, due to unit cell contraction and water remodeling.^{32,33} In fact, a small displacement was found between the experimental electron density in the CCFS and the simulated electron density (Fig. 6B). A promising solution is room-temperature X-ray diffraction measurement. A detailed comparison of the high-resolution electron density maps of dihydrofolate reductase (DHFR), measured at room and cryogenic temperatures, revealed that cryocooling induced artificial biases in the rotamer distributions of the side chains.³³ At the main-chain level, a narrower distribution of a loop structure was also detected after cryocooling, but the movement of the averaged position was not evident. Room-temperature X-ray diffraction measurements of protein crystals with CCFS will reveal the unbiased movements of protein segments in the future.

Finally, the concept of CCFS is not limited to the analysis of protein dynamics. Other potential applications include correcting the distorted protein conformations induced by the crystallographic contact effects, allowing ligands to soak into otherwise occluded sites in protein crystals, and studying the time-resolved large motions that are normally constrained in conventional crystal lattices.

Figures

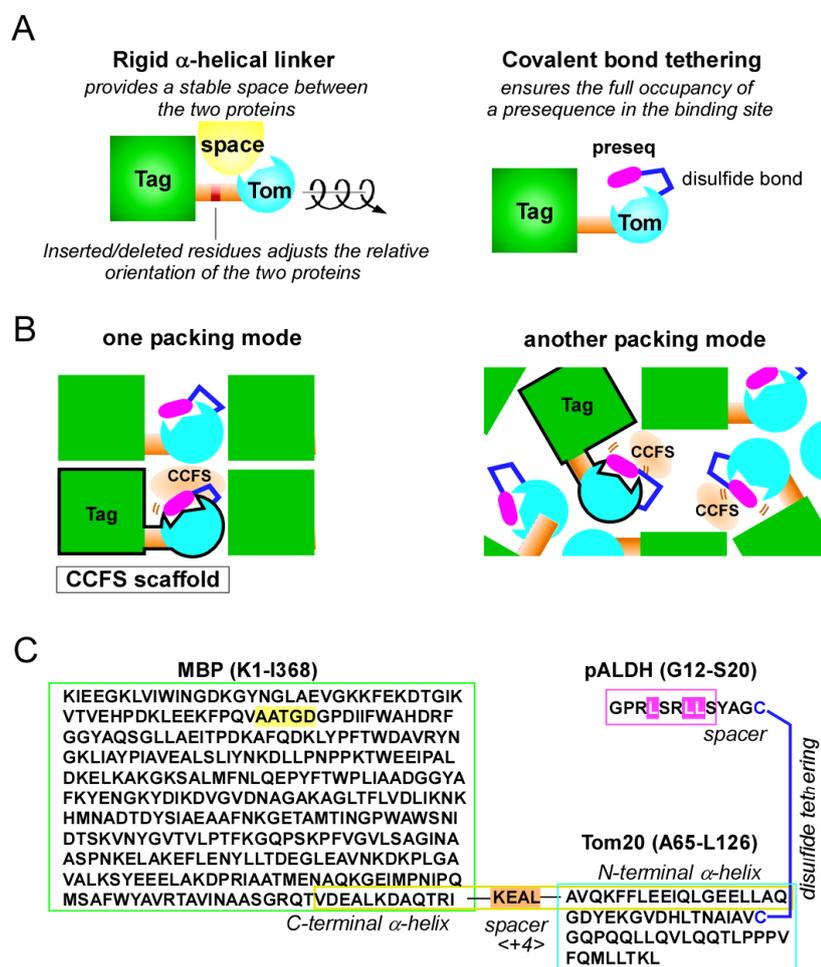


Figure 1. Key techniques to create CCFS in protein crystals. (A) Design of a fusion protein with a rigid α -helical connection. The relative orientation of the tag and target proteins can be adjusted as a function of the number of residues inserted/deleted at the middle of the connector helix. Covalent-bond tethering ensures the full occupancy of a weak-affinity presequence in the binding site. (B) Concept of the CCFS scaffold. The CCFS can be intentionally created in the rigid framework formed by the ‘CCFS scaffold’, independent of the packing modes of the protein molecules. The CCFS scaffold (surrounded by black borders) is a fusion protein connected by a rigid linker. (C) Amino acid sequence of MBP<+4>Tom20-SS-pALDH. The connector helix is enclosed by the yellow box, and the 4-residue insertion as a spacer is highlighted in orange. A disulfide bond was formed between the cysteine residue attached to the C-terminus of the pALDH peptide and the single cysteine residue in MBP-Tom20. The

5-residue deletion, $\Delta 5$, in MBP and the three essential hydrophobic leucine residues in pALDH are highlighted in yellow and magenta, respectively.

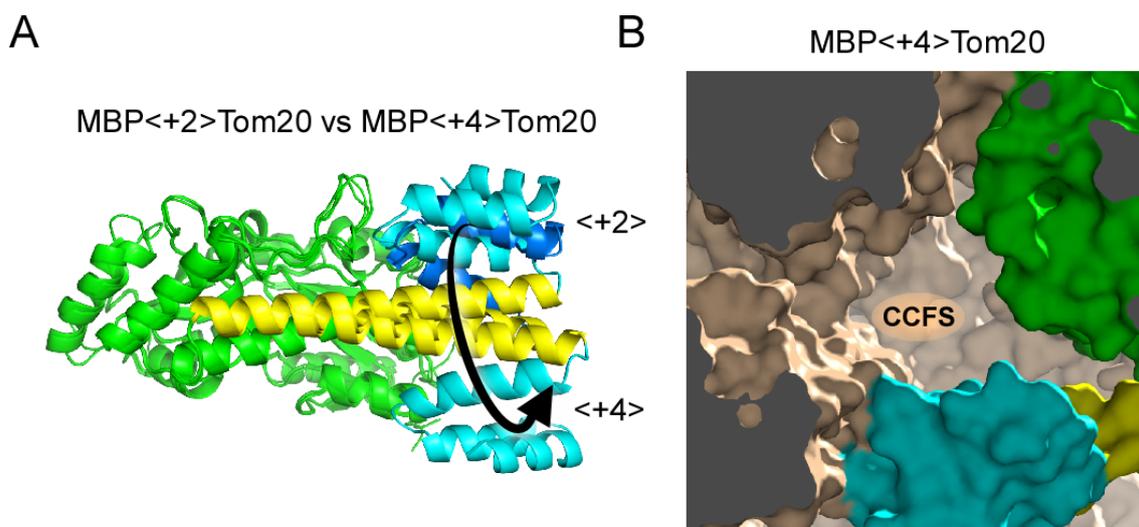


Figure 2. Effects of the different spacer lengths in the connector helix on the structure and the formation of CCFS. (A) Structural comparison of two fusion proteins with spacer lengths of +2 and +4. A long, continuous α -helical structure (consisting of the C-terminal α -helix of MBP, the spacer, and the N-terminal α -helix of Tom20) was formed in the two crystal structures (yellow). The Tom20 proteins are located on nearly opposite sides around the helical axis of the connector helix in the two structures, as intended. (B) Formation of CCFS in the crystal of MBP<+4>Tom20.

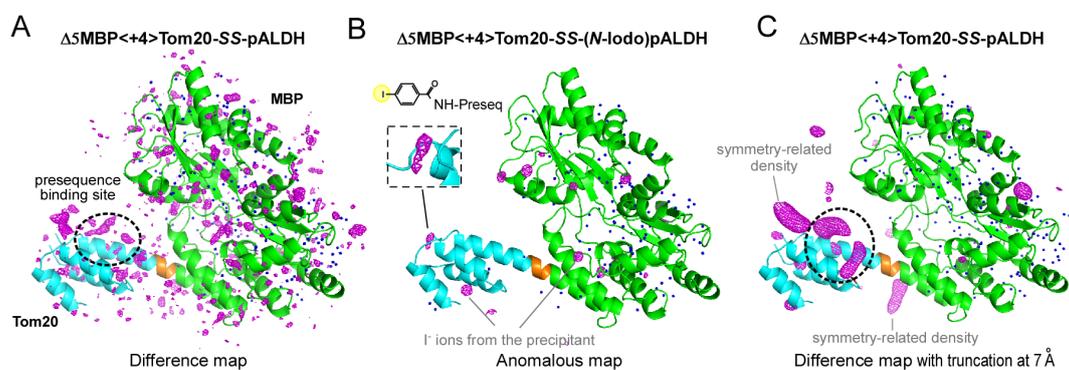


Figure 3. Electron density maps of the disulfide-bond tethered MBP-Tom20 complexes. (A) F_0-F_c difference electron density map of $\Delta 5\text{MBP}<+4>\text{Tom20-SS-pALDH}$, contoured at $+3\sigma$. (B) Anomalous difference map of $\Delta 5\text{MBP}<+4>\text{Tom20-SS-(N-iodo)pALDH}$, contoured at $+5\sigma$ to locate iodine atoms/ions. The inset shows an enlarged view of the electron density of the iodine atom in the 4-iodobenzoyl group attached to the N-terminus of pALDH. The electron densities of the bound iodine ions from the precipitant solution are also clearly visible, but their spherical shapes are easily discriminated from the iodine atom at the N-terminus. (C) Difference map with truncation of high-resolution reflections at 7 \AA prior to Fourier transformation, contoured at $+3\sigma$. Note that the same X-ray diffraction data set was used for map generation in (A) and (C).

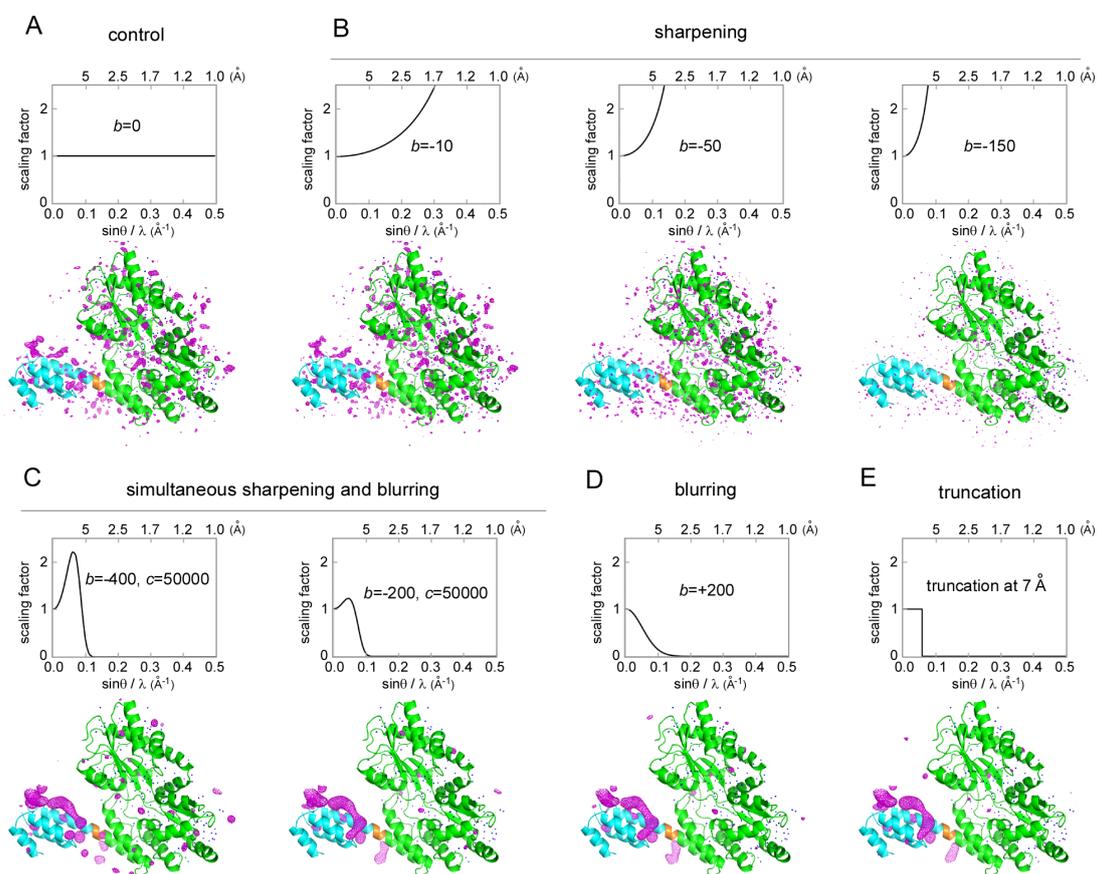


Figure 4. Effects of map sharpening/blurring and truncation on the difference electron density map, contoured at $+3\sigma$. (A) Control: no sharpening or blurring was applied. (B) The observed diffraction data (F_o) are scaled up using the equation, $F_{\text{sharpened}} = F_o \exp[-b(\sin\theta/\lambda)^2]$, where b is a sharpening factor, θ is the scattering angle, and λ is the X-ray wavelength. (C) Simultaneous application of map sharpening for the low-resolution reflections and map blurring for the high-resolution reflections. The bell-shaped function for scaling is $F_{\text{sharpened}} = F_o \exp[-b(\sin\theta/\lambda)^2 - c(\sin\theta/\lambda)^4]$. (D) Simple

blurring was also effective. (E) Truncation of the high-resolution reflections, as a simplified version of the simultaneous sharpening and blurring.

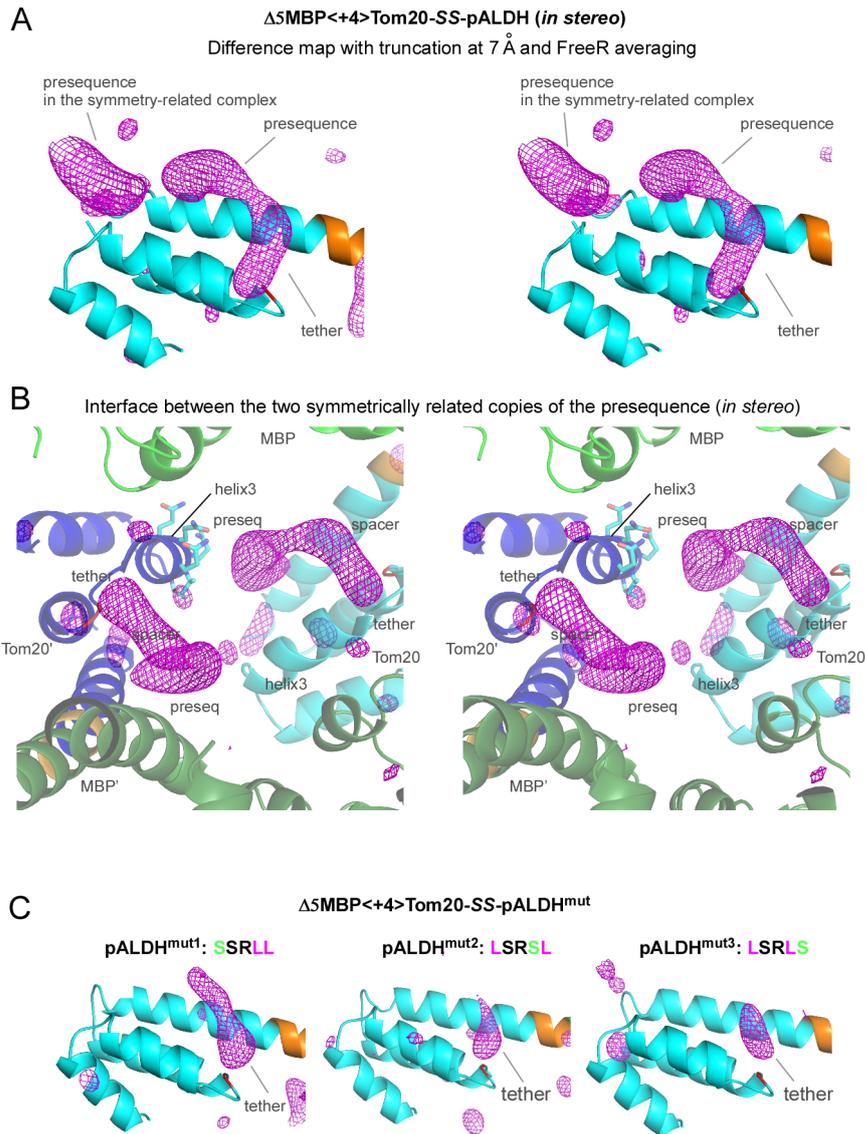


Figure 5. Electron densities in the CCFS corresponding to the presequence peptide and disulfide tether in the binding site of Tom20. (A) Stereo view of the truncated and

FreeR-averaged difference map of $\Delta 5\text{MBP}^{<+4>}\text{Tom20-SS-pALDH}$, contoured at $+3\sigma$. The electron densities in the binding site of Tom20 were improved by the averaging of 20 maps calculated with different free test sets, as compared with the map without FreeR averaging in Fig. 3C. Tom20 and the 4-residue spacer are colored cyan and orange, respectively. (B) Stereo close-up view of the interface between two symmetrically related copies of the elongated electron density. MBP and Tom20 are colored green and cyan in one molecule, and dark green and blue in another molecule, respectively. The side chains of the helix3 of one Tom20 structure are shown in a stick model. (C) Effects of the essential leucine substitutions in pALDH on the electron density in the CCFS. These difference maps were generated and drawn in exactly the same way as in (A). Only the tether parts remained visible in the CCFS.

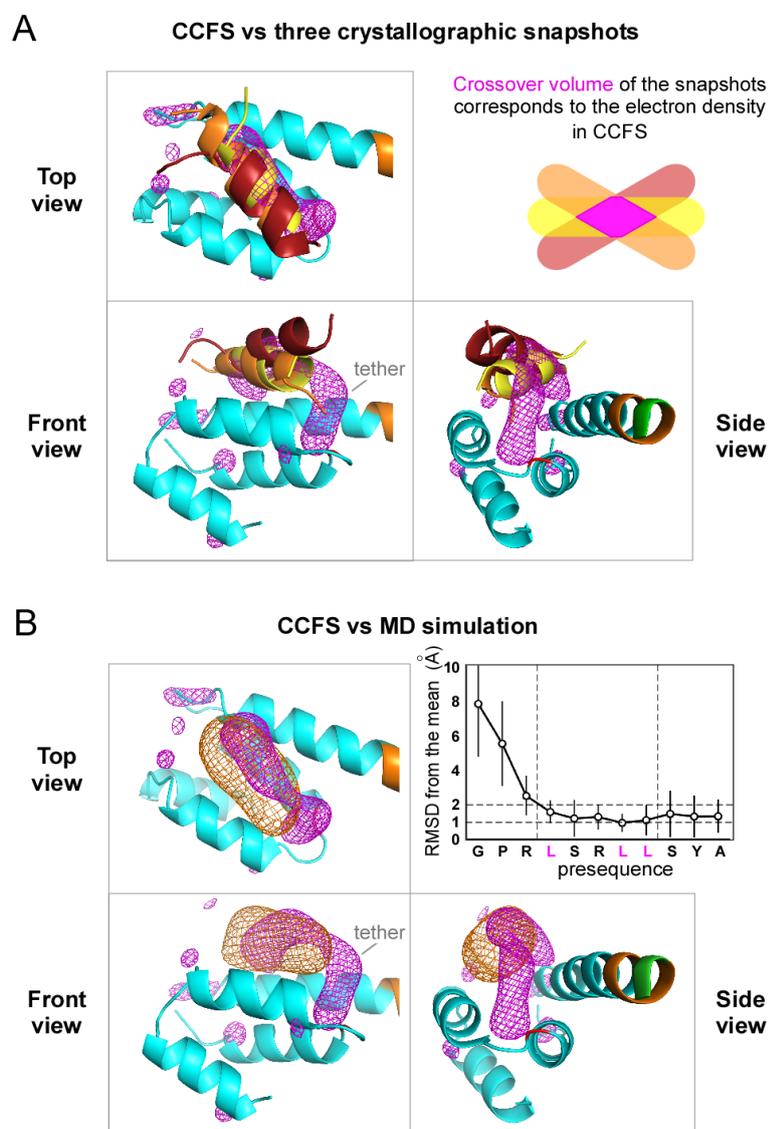


Figure 6. Validation of the electron density in the CCFS of the $\Delta 5\text{MBP}^{<+4>}\text{Tom20-SS-pALDH}$ crystal. (A) Superimposition with the three crystallographic snapshots of the pALDH presequence (PDB codes 2V1S, 2V1T, 3AX3). The 44 Ca atoms ($\text{A}^{65}\text{-E}^{79}+\text{Y}^{86}\text{-L}^{114}$) of the first three α -helices of Tom20 were superimposed. (B) Superimposition with the electron density calculated from the MD

simulation (orange mesh). The graph shows the rmsds of the C α atoms from the mean positions during the MD simulation.

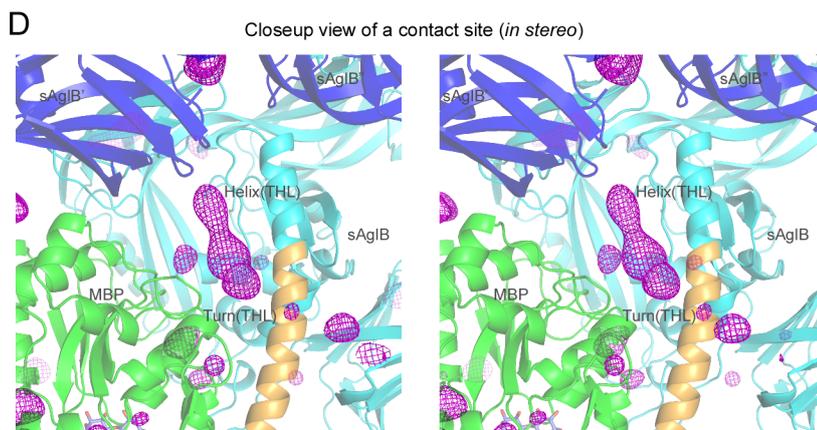
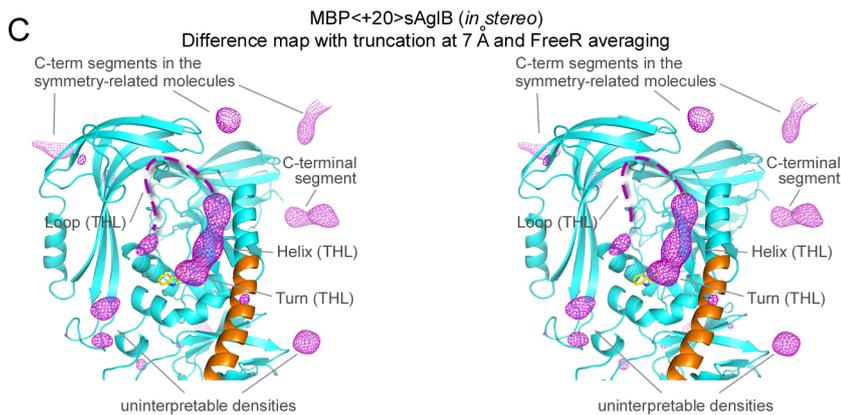
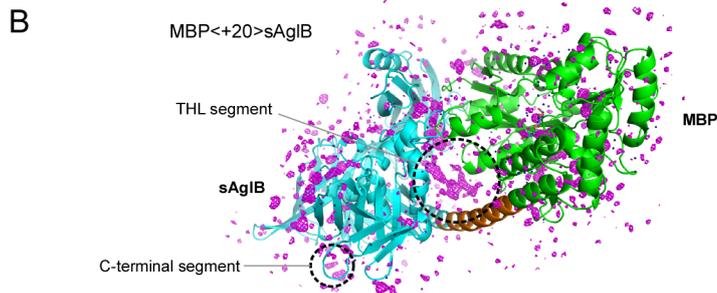
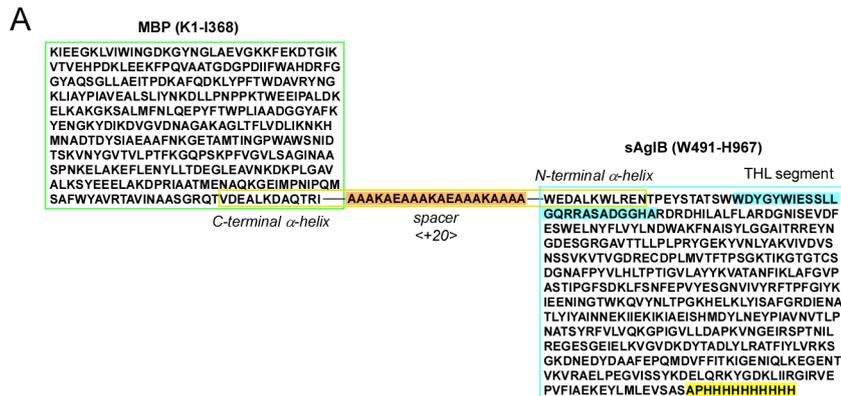


Figure 7. Electron density maps of the MBP-sAglB fusion protein. (A) Amino acid sequence of MBP<+20>sAglB. The connector helix is enclosed by the yellow box, and the +20 spacer is highlighted in orange. The THL segment and the C-terminal segment, which were not included in the CCFS scaffold, are highlighted in cyan and yellow, respectively. (B) F_o-F_c difference map, (C) truncated and FreeR-averaged difference map (in stereo), and (D) close-up view of a crystal contact site (in stereo) of MBP<+20>sAglB, all contoured at $+3\sigma$. Note that the same X-ray diffraction data set was used for map generation in (B), (C), and (D). The 3-residue turn structure of the THL (turn-helix-loop) segment consists of the underlined residues in the well-conserved WWDYG motif. The first Trp residue of the WWDYG motif is part of the CCFS scaffold, and its side chain is depicted as a yellow stick model in (C). The invisible, highly mobile loop of the THL segment is shown as dashed lines. MBP, sAglB, and the 20-residue spacer are colored green, cyan, and orange in one molecule, respectively. sAglBs in other fusion protein molecules are colored blue, and labeled with sAglB' and sAglB''.

Supplementary Material

Rational design of crystal contact-free space in protein crystals for analyzing main-chain dynamics of proteins

Table of Contents

Supplementary Figures

Figure S1. Concept of CCFS in protein crystals.

Figure S2. Native PAGE of the soluble fractions of *E. coli* cell lysates, to monitor the expression of the MBP-Tom20 fusion proteins with different spacer lengths.

Figure S3. Stereoview of the structure of MBP<+4>Tom20-SS-pALDH.

Figure S4. Principle of Fo-Fc difference electron density map in real and reciprocal spaces.

Figure S5. Reaction scheme of *N*-succinimidyl-4-iodobenzoate with the amino groups in proteins.

Figure S6. Effects of the truncation of high-resolution reflections prior to Fourier transformation on the quality of the electron density in the difference map in the case of the $\Delta 5$ MBP<+4>Tom20-SS-pALDH crystal.

Figure S7. Comparison of the experimental and model structures of the CCFS

scaffolds.

Figure S8. Effects of the truncation of the high-resolution reflections prior to Fourier transformation on the quality of the electron density in the difference map in the case of the MBP<+20>sAglB crystal.

Figure S9. Effects of the bulk solvent corrections implemented in the different programs on the shape and volume of the electron density in the truncated difference map in the case of the Δ 5MBP<+4>Tom20-SS-pALDH crystal.

Figure S10. Data-processing workflow for the generation of the Fo-Fc difference electron density map with truncation and FreeR averaging.

Supplementary Tables

Table S1. Design, expression, and crystallization of the MBP-Tom20 fusion proteins

Table S2. Protein design, X-ray data collection, and refinement statistics (1)

Table S3. Protein design, X-ray data collection, and refinement statistics (2)

Table S4. Protein design, X-ray data collection, and refinement statistics (3)

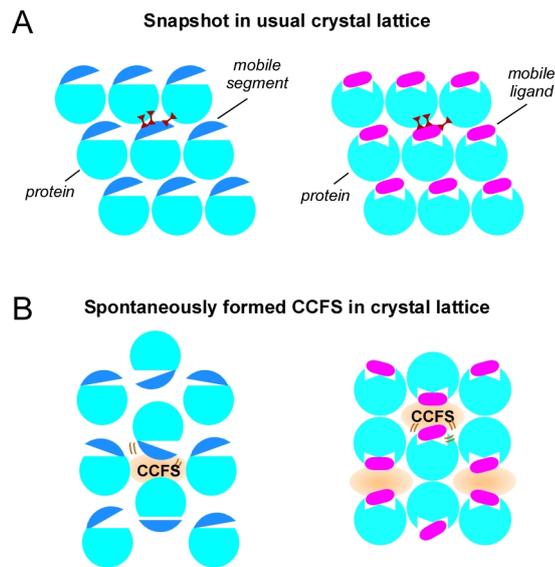


Figure S1. Concept of CCFS in protein crystals. Two cases, a mobile segment in a protein molecule and a mobile ligand in the binding site of a protein molecule, are shown. (A) In a usual protein crystal lattice, crystal contacts with neighboring molecules inhibit the movements of the segment/ligand, and thus a snapshot is obtained. (B) If the mobile segment/ligand is fortuitously located in the CCFS, then the motions of the moving segment/ligand are preserved in the protein crystals.

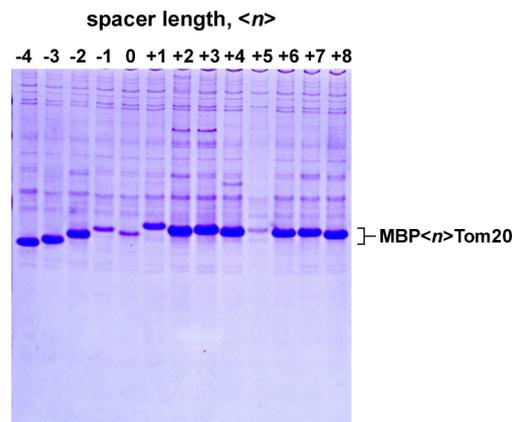


Figure S2. Production of recombinant MBP-Tom20 fusion proteins. Native PAGE of the soluble fractions of *E. coli* cell lysates, to monitor the expression of the MBP-Tom20 fusion proteins with different spacer lengths.

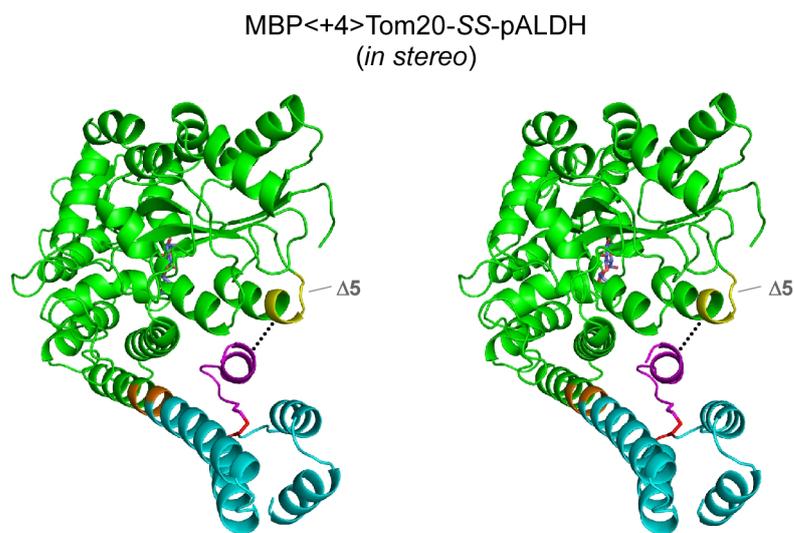


Figure S3. Stereoview of the structure of MBP<+4>Tom20-SS-pALDH. $\Delta 5$ in the crystal structure of MBP<+4>Tom20-SS-pALDH denotes the deleted segment (yellow), A⁵¹ATGD⁵⁵, in $\Delta 5$ MBP<+4>Tom20 to enlarge the CCFS. Due to the undesired interaction between pALDH and the $\Delta 5$ segment (the dotted line), the electron density of the pALDH presequence was clearly seen outside of the binding site, and so its model was constructed (magenta).

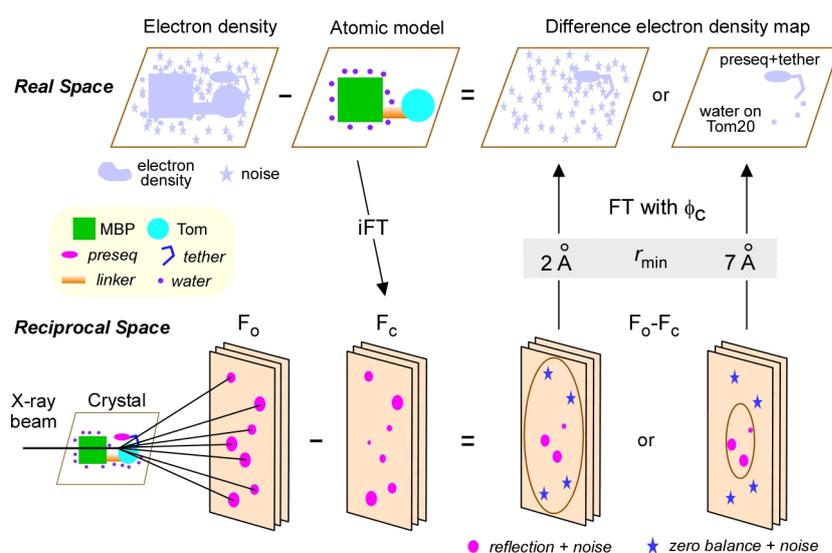


Figure S4. F_o - F_c difference electron density map in real and reciprocal spaces. The subtraction of calculated diffraction spots (F_c) from experimental diffraction spots (F_o) in reciprocal space cancels out the high-angle reflections derived from the immobile CCFS scaffold, but retains the low-angle reflections mainly attributable to the mobile presequence. The difference map calculated by Fourier transformation contains noise originating from the inclusion of high-angle reflections, which should have zero intensity. Thus, the truncation of high-angle reflections in reciprocal space works as a low-pass filter to reduce noise in real space.

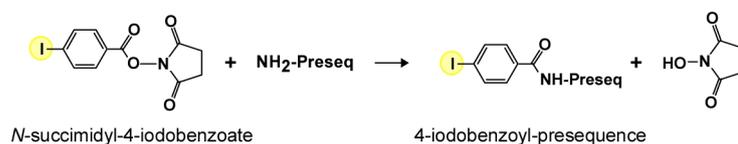
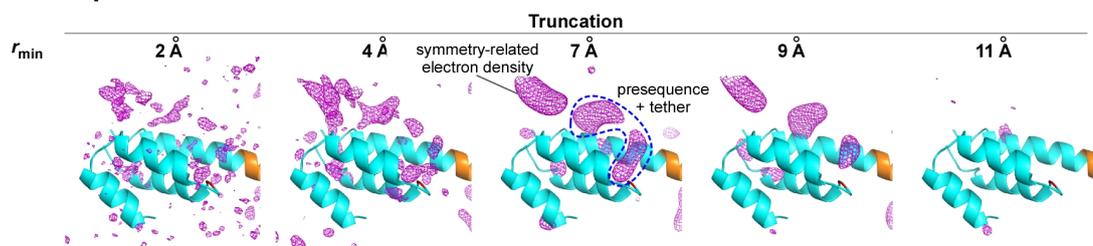


Figure S5. Reaction scheme of *N*-succinimidyl-4-iodobenzoate with the amino groups in proteins. *N*-succinimidyl-4-iodobenzoate was used to label the N-terminus of the pALDH peptide with an iodine atom.

A Presequence in CCFS



B Deleted helix in MBP (control)

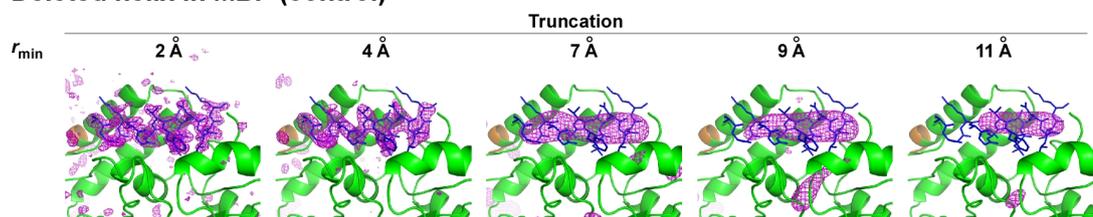
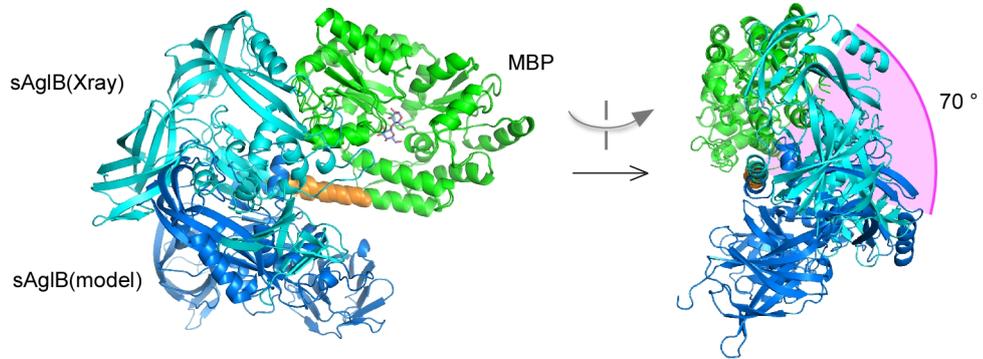
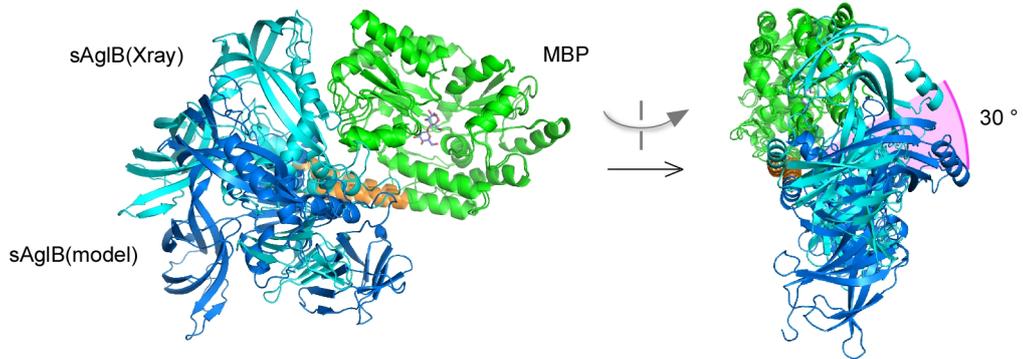


Figure S6. Effects of the truncation of high-resolution reflections prior to Fourier transformation on the quality of the electron density in the difference map, contoured at $+3\sigma$. (A) The electron density of the pALDH presequence peptide in CCFS, and (B) that of an arbitrarily chosen α -helical segment ($A^{186}GAKAGLTFLVDLIK^{200}$) in MBP, as a control for a fixed segment, are shown as a function of the truncation threshold, r_{\min} . In (B), the MBP-Tom20 structure lacking the α -helix was used as the search model in the *PHASER* molecular replacement and the subsequent *REFMAC5* refinement. The structure of the deleted α -helix is depicted as blue sticks with the side chains. The comparison between (A) and (B) indicates that the low-pass filter by truncation improves the quality of the electron densities of atoms with large B-factors, but as expected, has rather deteriorative effects on the electron densities of atoms with small B-factors. Note that the maps were generated by the *REFMAC5* run with the free test set of default Set 0 (See Fig. S7).

A MBP<+16>sAgIB



B MBP<+20>sAgIB



C Δ 5MBP<+4>Tom20-SS-pALDH

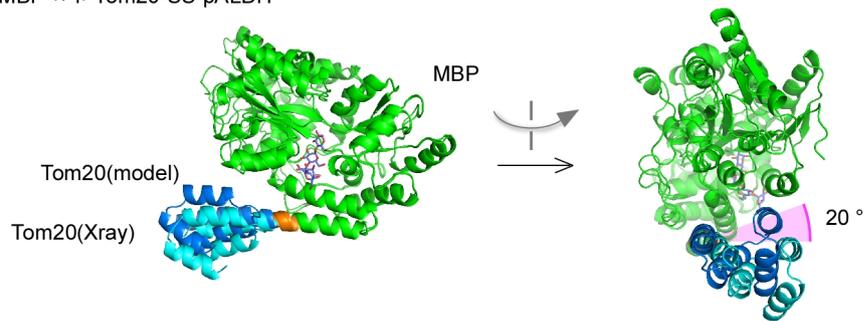


Figure S7. Comparison of the experimental and model structures of the CCFS scaffolds. The model structures were constructed by assuming an ideal α -helical conformation of the spacer in the connector helix. The MBP structures (green) of the fusion proteins were superimposed on each other. (A) The large deviation of the relative orientation of sAglB (cyan) from the model structure (blue) was seen in MBP<+16>sAglB. (B) The deviation of the relative orientation was smaller and acceptable in MBP<+20>sAglB, in spite of the longer spacer length. (C) The deviation of the relative orientation of Tom20 in Δ 5MBP<+4>Tom20-SS-pALDH was also small. Taken together, the model building is effective, but must be tested experimentally, particularly for long spacers (> 10 residues).

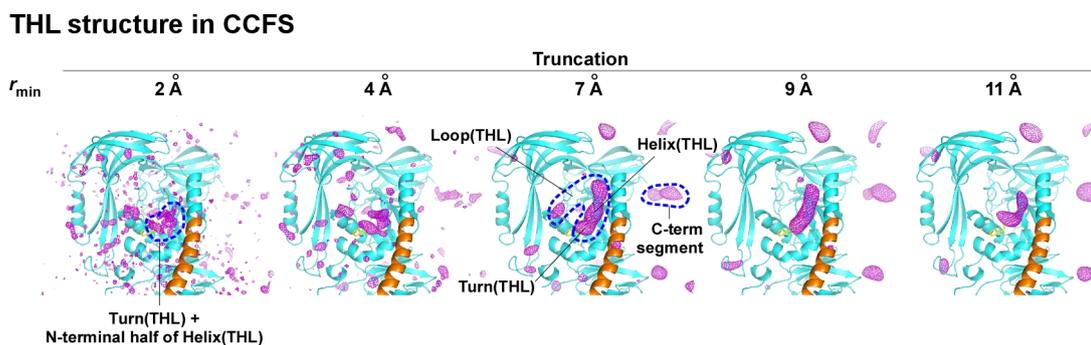


Figure S8. Effects of the truncation of the high-resolution reflections prior to Fourier transformation on the quality of the electron density in the difference map, contoured at $+3\sigma$. The electron density of the THL segment, in the CCFS created in the MBP<+20>sAglB crystal, is shown as a function of the truncation threshold, r_{\min} . The maps were generated by the *REFMAC5* run with the free test set of default Set 0 (See Fig. S10).

One-half (3-residue turn + 9-residue helix) of the mobile THL segment was clearly visualized in the map at $r_{\min} = 7 \text{ \AA}$. The other half (12-residue loop) except for its C-terminal 1 or 2 residues remained invisible, indicating that the movement was larger than the detection limit ($>1.5 \text{ \AA}$ rmsds from the mean positions). Close inspection of the

map at $r_{\min} = 2 \text{ \AA}$ revealed that the electron density corresponding to the turn part and the N-terminal half of the helical part of the THL segment marginally appeared. This indicated that the visualized half of the THL segment at $r_{\min} = 7 \text{ \AA}$ is further divided into two parts with respect to the mobility: The spatial distributions of the turn part and the N-terminal half of the helical part are narrower than that of the C-terminal half of the helical part, which is invisible in the map truncated at $r_{\min} = 2 \text{ \AA}$. Note that the C-terminal tail segment is also invisible in the same map at $r_{\min} = 2 \text{ \AA}$.

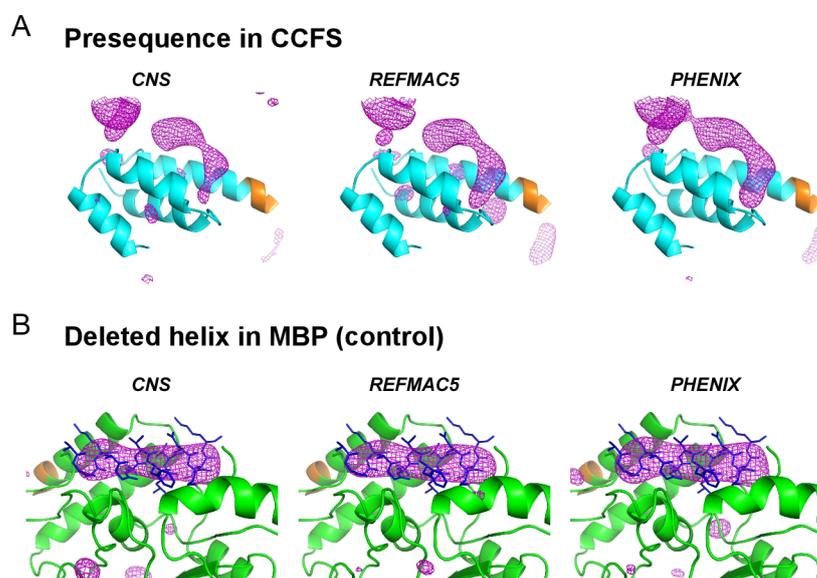


Figure S9. Effects of the bulk solvent corrections implemented in the different programs on the shape and volume of the electron density in the truncated difference map, contoured at $+3\sigma$. (A) The electron density of the pALDH presequence peptide in CCFS, and (B) that of an arbitrarily chosen α -helical segment (A¹⁸⁶GAKAGLTFLV²⁰⁰) in MBP, as a control for a fixed segment, are shown. In (B), the MBP-Tom20 structure lacking the α -helix was used as the search model in the *PHASER* molecular replacement and the subsequent *REFMAC5* refinement. The structure of the deleted α -helix is depicted as blue sticks with the side chains. A *REFMAC5* map with the flat mask-based bulk solvent correction was generated, by using the keywords “SOLVENT YES” and “SCALE TYPE SIMPLE”. The default setting of *REFMAC5* (“SOLVENT YES” and “SCALE TYPE BULK”) provides a map with the two types of corrections (the exponential type correction + the flat mask-based bulk solvent correction), but the generated map (Fig. 5A) was almost identical to the map obtained with the flat mask-based bulk solvent correction only. The CCFS scaffold (MBP-Tom20) structure refined by the program *REFMAC5* was used as input for the programs *CNS* ver. 1.3 and *PHENIX* ver. 1.9, with the default settings of the flat mask-based bulk solvent corrections. In the *CNS* calculation, *model_map.inp* was used to generate a sigma-A weighted Fo-Fc difference map with truncation by using the keyword, “high_res=7.0”. In the *PHENIX* calculation, *phenix.maps* was used to

generate the MTZ file that contained the sigma-A weighted Fourier coefficients. Then, *SFTOOLS* and *FFT* in *CCP4* were used for truncation and map generation.

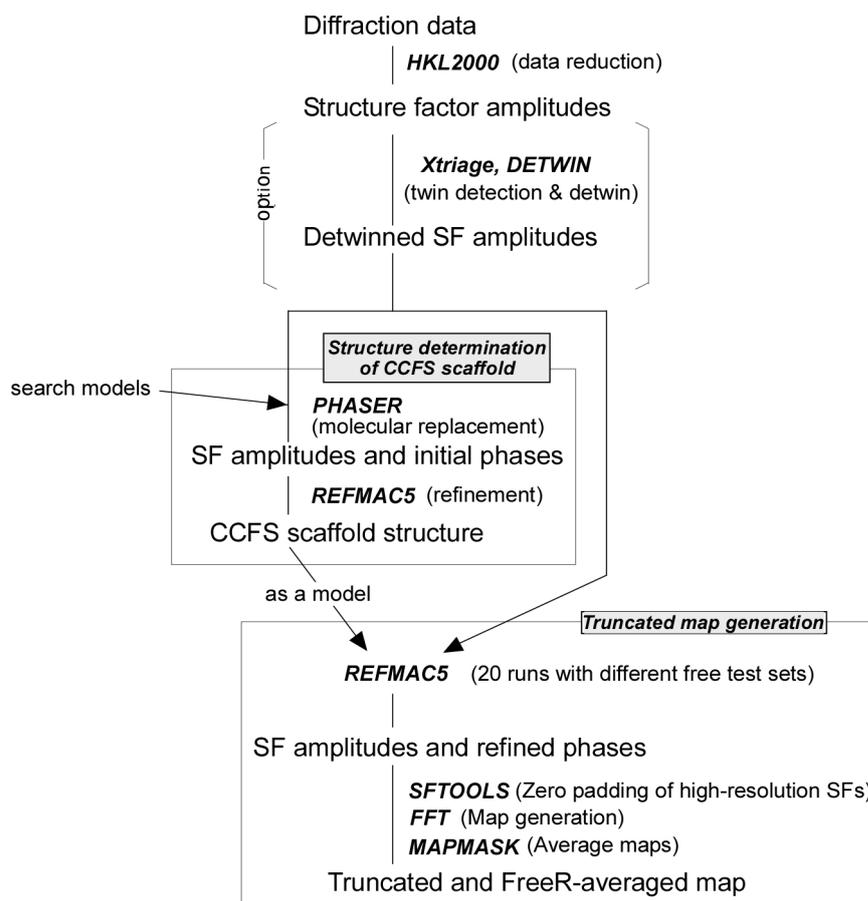


Figure S10. Data-processing workflow for the generation of the Fo-Fc difference electron density map with truncation and FreeR averaging.

Table S1. Design, expression, and crystallization of the MBP-Tom20 fusion proteins

$\langle n \rangle^a$	MBP $\langle n \rangle$ Tom20			Expression of soluble protein ^c	Crystallization ^d	Crystal structure ^e
	spacer junction ^b	sequence	at the			
-4	AQTRI- F			+		
-3	AQTRI- . . . KF			+		
-2	AQTRI- . . QKF			+		
-1	AQTRI- . VQKF			±		
0	AQTRI- AVQKF			±	√	
+1	AQTRI- <i>K</i> - AVQKF			±		
+2	AQTRI- <i>KE</i> - AVQKF			++	√	√
+3	AQTRI- <i>KEA</i> - AVQKF			++		
+4	AQTRI- <i>KEAL</i> - AVQKF			+	√	√
+5	AQTRI- <i>KEALQ</i> - AVQKF			±		
+6	AQTRI- <i>KEALQE</i> - AVQKF			+		
+7	AQTRI- <i>KEALQEL</i> - AVQKF			+	√	
+8	AQTRI- <i>KEALQEELA</i> - AVQKF			+	√	

^a $\langle n \rangle$ denotes the length of the spacer.

^b The MBP sequence is shown in plain text, the spacer sequence is italicized, and the Tom20 sequence is in bold text. The basic spacer sequence is KEALQEELA.

^c ±, +, and ++ indicate rough estimates of the protein amounts in the soluble fractions of *E. coli* cell lysates, based on the Coomassie Brilliant Blue stained native PAGE gel in Fig. S2.

^d √, Crystallized. Crystallization conditions: for $\langle 0 \rangle$, 8% PEG400, 0.1 M sodium acetate (pH 4.6); for $\langle +2 \rangle$ and $\langle +4 \rangle$, in Table S2; $\langle +7 \rangle$, 20% PEG4000, 5% isopropanol, 0.1 M sodium citrate; $\langle +8 \rangle$, 16% PEG6000, 0.2 M calcium chloride, 0.1 M HEPES (pH 7.0).

^e √, Structures of the fusion proteins were determined. See Table S2.

Table S2. Protein design, X-ray data collection, and refinement statistics (1)

	MBP<+2>Tom20	MBP<+4>Tom20	MBP<+4>Tom20-SS-pA LDH
Protein design			
Fusion protein	MBP-KE-Tom20	MBP-KEAL-Tom20	MBP-KEAL-Tom20
pALDH peptide	none	none	Ac-GPRLSRLLSYAGC-NH ₂
Crystallization conditions			
	16% PEG3350 0.2 M disodium malonate 0.1M HEPES (pH 7.3)	50% PEG400 0.1 M phosphate citrate (pH 4.4)	24% PEG8000 20% glycerol 0.04 M potassium thiocyanate
Data collection statistics			
Beamline	SPring-8 BL32XU	PF BL5A	PF BL17A
Cryo conditions	20% glycerol	20% glycerol	15% glycerol
Wavelength, Å	1.0000	1.0000	1.0000
Oscillation range, °	180	180	180
Space group	<i>C</i> 2	<i>I</i> 222	<i>P</i> 4 ₃ 2 ₁ 2
Cell dimensions	<i>a</i> = 129.9 Å, <i>b</i> = 81.2 Å, <i>c</i> = 109.9 Å, β = 106.4°	<i>a</i> = 63.1 Å, <i>b</i> = 113.9 Å, <i>c</i> = 125.8 Å	<i>a</i> = <i>b</i> = 69.0 Å, <i>c</i> = 212.6 Å
No. of molecules in AU	2	1	1
Resolution range, Å	50.0 – 2.56 (2.60 – 2.56)	50.0 – 1.97 (2.00 – 1.97)	50.0 – 1.70 (1.73 – 1.70)
Observed reflections	135990	188262	784858
Unique reflections	35889	30663 (1040)	56782 (2824)
Completeness, %	99.9 (99.9)	93.1 (64.4)	98.4 (99.6)
<i>R</i> _{merge} (<i>I</i>) ^a	0.076 (0.436)	0.106 (0.591)	0.128 (0.734)
<i>I</i> / σ(<i>I</i>)	11.0 (3.65)	14.6 (1.35)	22.1 (3.38)
Refinement statistics			
Resolution range, Å	24.7– 2.56 (2.61 – 2.55)	42.2 – 1.96 (2.01 – 1.96)	42.1 – 1.70 (1.74 – 1.70)
No. of protein atoms	6752	3367	3367
No. of water molecules	20	101	407
<i>R</i> _{work} / <i>R</i> _{free} ^b	0.220/0.279	0.207/0.254	0.177/0.211
rmsd ^c from ideality bond length, Å	0.009	0.019	0.018
angles, °	1.175	1.939	1.800
Ramachandran plot ^d , %			
Favored	93.6	97.7	98.4

region			
Allowed	5.1	2.1	1.2
region			
Outlier region	1.3	0.2	0.1
PDB code	5AZ6	5AZ7	5AZ8

Values in parentheses are for the highest resolution shell.

^a $R_{\text{merge}}(I) = (\sum \sum |I_i - \langle I \rangle|) / \sum \sum I_i$, where I_i is the intensity of the i th observation and $\langle I \rangle$ is the mean intensity.

^b $R_{\text{work}}/R_{\text{free}} = \sum |F_o - F_c| / \sum |F_o|$. R_{work} was calculated from the working set (95% of the total reflections). R_{free} was calculated from the test set, using 5% of the total reflections. The test set was not used in the refinement.

^crmsd, root mean square deviation.

^dCalculated using the program *RAMPAGE* in the CCP4 program suite.

Table S3. Protein design, X-ray data collection, and refinement statistics (2)

	$\Delta 5\text{MBP}<+4>\text{To}$	$\Delta 5\text{MBP}<+4>\text{To}$	$\Delta 5\text{MBP}<+4>\text{To}$	$\Delta 5\text{MBP}<+4>\text{To}$	$\Delta 5\text{MBP}<+4>\text{To}$
	m20-SS-pALD	m20-SS-(N-iodo	m20-SS-pALD	m20-SS-pALD	m20-SS-pALD
	H)pALDH	H(L15S)	H(L18S)	H(L19S)
Protein design					
Fusion protein	$\Delta 5\text{MBP-KEAL-Tom20}$				
pALDH peptide	Ac-GPRLSRLLSYAGC-NH ₂	4-iodobenzoyl-GPRLSRLLSYAGC-NH ₂	Ac-GPRSSRLLSYAGC-NH ₂	Ac-GPRLSRSLLSYAGC-NH ₂	Ac-GPRLSRLLSYAGC-NH ₂
Crystallization conditions					
	20% PEG3350 0.2 M potassium nitrate	20% PEG3350 0.2 M potassium iodide	20% PEG3350 0.2 M ammonium nitrate	20% PEG6000 0.2 M lithium chloride	24% PEG3350 0.2 M sodium iodide
Data collection statistics					
Beamline	SPring-8 BL44XU	SPring-8 BL26B1	SPring-8 BL44XU	SPring-8 BL26B1	SPring-8 BL44XU
Cryo conditions	15% glycerol	20% glycerol	20% glycerol	20% glycerol	20% glycerol
Wavelength, Å	0.9000	1.6000	0.9000	1.0000	0.9000
Oscillation range, °	180	720	180	180	180
Space group	<i>C2</i>	<i>C2</i>	<i>C2</i>	<i>P2</i> ₁	<i>P2</i> ₁
Cell dimensions	$a = 160.5 \text{ \AA}, b = 71.4 \text{ \AA}, c = 47.5 \text{ \AA}, \beta = 104.9^\circ$	$a = 160.6 \text{ \AA}, b = 71.4 \text{ \AA}, c = 47.6 \text{ \AA}, \beta = 105.2^\circ$	$a = 157.1 \text{ \AA}, b = 70.8 \text{ \AA}, c = 86.1 \text{ \AA}, \beta = 102.5^\circ$	$a = 70.4 \text{ \AA}, b = 83.5 \text{ \AA}, c = 86.5 \text{ \AA}, \beta = 113.9^\circ$	$a = 70.8 \text{ \AA}, b = 84.0 \text{ \AA}, c = 86.5 \text{ \AA}, \beta = 113.9^\circ$
No. of molecules in AU	1	1	2	2	2
Resolution range, Å	39.6 – 1.82 (1.86 – 1.82)	50.0 – 2.71 (2.76 – 2.71)	50.0 – 2.36 (2.40 – 2.36)	50.0 – 1.78 (1.81 – 1.78)	50.0 – 1.62 (1.65 – 1.62)
Observed reflections	169218	188174	132708	314494	430860
Unique	46512 (4633)	14001 (704)	37578 (1833)	87479 (4387)	116783 (5429)

reflections					
Completeness, %	99.8 (99.8)	98.2 (98.7)	99.1 (99.1)	99.8 (100.0)	98.2 (92.0)
$R_{\text{merge}}(I)^a$	0.083 (0.765)	0.077 (0.146)	0.165 (0.728)	0.064 (0.716)	0.082 (0.502)
$I / \sigma(I)$	14.0 (1.8)	29.9 (16.8)	9.25 (1.96)	18.7 (2.23)	13.9 (2.45)
Refinement					
statistics					
twinning	no	no	no	Pseudomerohed ral, (<i>h</i> , <i>-k</i> , <i>-h-l</i>), $\alpha = 0.18$	Pseudomerohed ral, (<i>h</i> , <i>-k</i> , <i>-h-l</i>), $\alpha = 0.25$
Resolution range, Å	39.6 – 1.82		41.5 – 2.13	30.0 – 1.78	39.5 – 1.62
No. of protein atoms	3323		6696	6696	6696
No. of water molecules	172		224	234	260
$R_{\text{work}}/R_{\text{free}}^{b,e}$	0.195±0.0005/0. 232±0.006		0.199±0.0007/ 0.251±0.005	0.232±0.0004/ 0.276±0.005	0.229±0.0006/ 0.265±0.004
rmsd ^{c,e} from ideality bond length, Å	0.019±0.0008		0.016±0.0003	0.018±0.0004	0.019±0.0006
angles, °	1.90±0.04		1.72±0.03	1.89±0.02	1.89±0.04
Ramachandran plot ^{d,e} , %					
Favored region	96.9±0.24		97.3±0.34	96.0±0.40	97.7±0.20
Allowed region	2.4±0.21		2.0±0.32	3.2±0.35	1.9±0.18
Outlier region	0.6±0.13		0.6±0.08	0.8±0.24	0.5±0.06
PDB code	5AZ9	No structure	-	-	-

^{a, b, c, d} See the legends of Table S2.

^e Average and standard deviation of the 20 refinement calculations with different free test sets.

Table S4. Protein design, X-ray data collection, and refinement statistics (3)

	MBP<+16>sAglB	MBP<+20>sAglB
Protein design		
Spacer	AEAAAKAEAAAKAAAA	AAAKAEAAAKAEAAAKAAAA
Crystallization conditions		
	14% PEG4000	17% PEG10000
	0.1 M sodium acetate, pH 4.5	0.1 M ammonium acetate
	0.1M manganese(II) chloride	0.1 M bis-Tris, pH 5.5
		1.0 M lithium chloride
Data collection statistics		
Beamline	SPring-8 BL44XU	SPring-8 BL32XU
Cryo conditions	20% glycerol	20% glycerol
Wavelength, Å	0.9000	1.0000
Oscillation range, °	360	300
Space group	<i>P1</i>	<i>P2₁2₁2₁^f</i>
Cell dimensions	$a = 54.85 \text{ Å}, b = 62.56 \text{ Å}, c = 136.4 \text{ Å}, \alpha = 102.0^\circ, \beta = 98.3^\circ, \gamma = 92.9^\circ$	$a = 66.5 \text{ Å}, b = 100.5 \text{ Å}, c = 140.9 \text{ Å}$
No. of molecules in AU	2	1
Resolution range, Å	50.0 – 2.60 (2.64 – 2.60)	50.0 – 2.08 (2.12 – 2.08)
Observed reflections	208533	694036
Unique reflections	53073 (2606)	57642 (2807)
Completeness, %	99.3 (95.1)	100.0 (100.0)
$R_{\text{merge}}(I)^a$	0.406 (0.102)	> 1.00 (0.142)
$I / \sigma(I)$	12.8 (3.4)	18.7 (2.0)
Refinement statistics		
Resolution range, Å	40.57 – 2.60	43.61 – 2.08
No. of protein atoms	13018	6580
No. of ligands	48	23
No. of water molecules	61	106
$R_{\text{work}}/R_{\text{free}}^{b,e}$	0.191±0.0006/0.257±0.005	0.196±0.0006/0.245±0.005
rmsd ^{c,e} from ideality	0.013±0.0004	0.017±0.0004

bond length, Å		
angles, °	1.696±0.027	1.902±0.030
Ramachandran plot ^{d,e} , %		
Favored region	95.7±0.23	97.0±0.20
Allowed region	3.4±0.20	2.6±0.26
Outlier region	0.9±0.11	0.4±0.15
PDB code	-	5AZA

^{a, b, c, d} See the legends of Table S2.

^e Average and standard deviation of the 20 refinement calculations with different free test sets.

^f After three micro-seeding steps, the space group changed from *P1* to *P2₁2₁2₁*.

References

1. Singh P, Abeysinghe T, Kohen A (2015) Linking protein motion to enzyme catalysis. *Molecules* 20:1192-1209.
2. Ramanathan A, Savol A, Burger V, Chennubhotla CS, Agarwal PK (2014) Protein conformational populations and functionally relevant substates. *Acc Chem Res* 47:149-156.
3. van den Bedem H, Fraser JS (2015) Integrative, dynamic structural biology at atomic resolution--it's about time. *Nat Methods* 12:307-318.
4. Gobl C, Madl T, Simon B, Sattler M (2014) NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog Nucl Magn Reson Spectrosc* 80:26-63.
5. Vranken WF (2014) NMR structure validation in relation to dynamics and structure determination. *Prog Nucl Magn Reson Spectrosc* 82:27-38.
6. Zhang XJ, Wozniak JA, Matthews BW (1995) Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J Mol Biol* 250:527-552.
7. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KF, Becker S, Meiler J, Grubmuller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320:1471-1475.
8. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462:669-673.
9. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc Natl Acad Sci U S A* 108:16247-16252.
10. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Sci* 12:1313-1322.
11. Zhan Y, Song X, Zhou GW (2001) Structural analysis of regulatory protein domains using GST-fusion proteins. *Gene* 281:1-9.
12. Hodel A, Kim S-H, Brunger AT (1992) Model bias in macromolecular crystal structures. *Acta Crystallogr A* 48:851-858.

13. Praznikar J, Afonine PV, Guncar G, Adams PD, Turk D (2009) Averaged kick maps: less noise, more signal... and probably less bias. *Acta Crystallogr D Biol Crystallogr* 65:921-931.
14. Schmidt O, Pfanner N, Meisinger C (2010) Mitochondrial protein import: from proteomics to functional mechanisms. *Nat Rev Mol Cell Biol* 11:655-667.
15. Endo T, Yamano K, Kawano S (2011) Structural insight into the mitochondrial protein import system. *Biochim Biophys Acta* 1808:955-970.
16. Abe Y, Shodai T, Muto T, Mihara K, Torii H, Nishikawa S, Endo T, Kohda D (2000) Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. *Cell* 100:551-560.
17. Saitoh T, Igura M, Obita T, Ose T, Kojima R, Maenaka K, Endo T, Kohda D (2007) Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. *EMBO J* 26:4777-4787.
18. Jarrell KF, Ding Y, Meyer BH, Albers SV, Kaminski L, Eichler J (2014) N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol Mol Biol Rev* 78:304-341.
19. Chen X, Zaro JL, Shen WC (2013) Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* 65:1357-1369.
20. Saitoh T, Igura M, Miyazaki Y, Ose T, Maita N, Kohda D (2011) Crystallographic snapshots of Tom20-mitochondrial presequence interactions with disulfide-stabilized peptides. *Biochemistry* 50:5487-5496.
21. Obita T, Muto T, Endo T, Kohda D (2003) Peptide library approach with a disulfide tether to refine the Tom20 recognition motif in mitochondrial presequences. *J Mol Biol* 328:495-504.
22. Henderson R, Moffat JK (1971) The difference Fourier technique in protein crystallography: errors and their treatment. *Acta Crystallogr B Struct Sci* 27:1414-1420.
23. Liu C, Xiong Y (2014) Electron density sharpening as a general technique in crystallographic studies. *J Mol Biol* 426:980-993.
24. Muto T, Obita T, Abe Y, Shodai T, Endo T, Kohda D (2001) NMR identification of the Tom20 binding segment in mitochondrial presequences. *J Mol Biol* 306:137-143.
25. Komuro Y, Miyashita N, Mori T, Muneyuki E, Saitoh T, Kohda D, Sugita Y

- (2013) Energetics of the presequence-binding poses in mitochondrial protein import through Tom20. *J Phys Chem B* 117:2864-2871.
26. Lizak C, Gerber S, Numao S, Aebi M, Locher KP (2011) X-ray structure of a bacterial oligosaccharyltransferase. *Nature* 474:350-355.
 27. Matsumoto S, Shimada A, Nyirenda J, Igura M, Kawano Y, Kohda D (2013) Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc Natl Acad Sci U S A* 110:17868-17873.
 28. Nyirenda J, Matsumoto S, Saitoh T, Maita N, Noda NN, Inagaki F, Kohda D (2013) Crystallographic and NMR evidence for flexibility in oligosaccharyltransferases and its catalytic significance. *Structure* 21:32-41.
 29. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng* 14:529-532.
 30. Koide S, Huang X, Link K, Koide A, Bu Z, Engelman DM (2000) Design of single-layer beta-sheets without a hydrophobic core. *Nature* 403:456-460.
 31. Afonine PV, Grosse-Kunstleve RW, Adams PD, Urzhumtsev A (2013) Bulk-solvent and overall scaling revisited: faster calculations, improved results. *Acta Crystallogr D Biol Crystallogr* 69:625-634.
 32. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2:2728-2733.
 33. Dunlop KV, Irvin RT, Hazes B (2005) Pros and cons of cryocrystallography: should we also collect a room-temperature data set? *Acta Crystallogr D Biol Crystallogr* 61:80-87.
 34. Keedy DA, van den Bedem H, Sivak DA, Petsko GA, Ringe D, Wilson MA, Fraser JS (2014) Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR. *Structure* 22:899-910.
 35. Igura M, Maita N, Obita T, Kamishikiryo J, Maenaka K, Kohda D (2007) Purification, crystallization and preliminary X-ray diffraction studies of the soluble domain of the oligosaccharyltransferase STT3 subunit from the thermophilic archaeon *Pyrococcus furiosus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 63:798-801.
 36. Horton RM, Cai ZL, Ho SN, Pease LR (1990) Gene splicing by overlap

- extension: tailor-made genes using the polymerase chain reaction. *Biotechniques* 8:528-535.
37. Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* 120:621-623.
 38. Shell TA, Mohler DL (2005) Selective targeting of DNA for cleavage within DNA-histone assemblies by a spermine-[CpW(CO)₃Ph]₂ conjugate. *Org Biomol Chem* 3:3091-3093.
 39. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307-326.
 40. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.
 41. Quioco FA, Spurlino JC, Rodseth LE (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* 5:997-1015.
 42. Oldham ML, Chen J (2011) Crystal structure of the maltose transporter in a pretranslocation intermediate state. *Science* 332:1202-1205.
 43. Igura M, Maita N, Kamishikiryo J, Yamada M, Obita T, Maenaka K, Kohda D (2008) Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J* 27:234-243.
 44. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355-367.
 45. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126-2132.
 46. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213-221.
 47. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011)

Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235-242.

CHAPTER 2

For the X-ray crystal structural analysis of *Leishmania major* STT3

Introduction

Oligosaccharyl transferase (OST) transfers N-linked sugar chain to a nitrogen of asparagine residues in acceptor proteins which have consensus sequence (NXT; N is asparagine, X is any amino acid without proline and T is threonine).[1, 2] This step is one of post-translational modifications, and this process is important in all of living things.[3] In prokaryote or bacteria, STT3 which is a subunit of OST catalyzes the sugar chain transfer by itself. STT3 is composed 13-transmembrane domain (13TM) of N-terminus and globular domain (soluble domain) of C-terminus. In contrast, in eukaryote, OST consists of several subunits included in STT3 and forming the complex is essential to expression the functions.[4] As the exception, protoctist in eukaryote catalyzes by single subunit STT3 itself, such as *Trypanosoma* or *Leishmanis*. [5, 6] The constitution of OST complex is known well by biochemical experiment, but the structural aspects are not revealed yet. The aim of this study is to determine the crystal structure of OST complex and to reveal the mechanism.

First of all, we focused on a protoctist in eukaryote of OST which is *Leishmania major* OST (LmSTT3). This is because that the LmSTT3 protein catalyzes the sugar chain transfers by itself in spite of eukaryote. To express the protein, we used *E. coli*, yeast, insect cells (S2 cell) or human cells (HEK293T). As the selection of expression system, we used yeast expression system, since we confirmed the protein expression levels using western blot. The overexpressed LmSTT3 protein was purified by affinity and gel-filtration chromatography and crystallized. We obtained crystals under some conditions and performed X-ray diffraction experiment. The crystals were diffracted well. To solve the structure, we used template protein (PDB ID: 2LGZ) for molecular replacement (MR). We could not solve the structure using MR methods, so we performed single anomalous dispersion (SAD) using selenomethionine-labeled LmSTT3 protein. Now, we are analyzing the LmSTT3 structure.

Methods and Results

Contraction and Protein expression

To express the soluble domain and full length of STT3 from *Leishmania major* (*LmSTT3*), first of all, we selected E.coli expression system. The *stt3* gene was accumulated by PCR and inserted into pET52b at restriction enzyme site between *sma1* and *sal1*. The plasmid was transformed in *E.coli* BL21(DE3) and culture using LB medium. After the $OD_{600} = 0.35 \sim 0.50$, 0.5 mM IPTG was added in the culture and the cell was grown at 16°C for 12 hours. The cell pellet from the culture was resuspended in buffer (50mM Tris-HCl pH8.0 and 100mM NaCl) and lysed by sonication. The supernatant was collected and centrifugated at 4°C. To make membrane fraction, the supernatant was recollected and ultracentrifugated at 100,000g. The pellet was dissolved by 1.0% n-Dodecyl β -D maltoside (DDM) and ultracentrifugated at 100,000g.

As other expression systems, we used the insect cells (*Drosophila* S2 cells) and yeast (*Pichia pastoris*). To express the recombinant protein using the insect cell, the *stt3* gene was accumulated by PCR and inserted into pMT V5 His B (for expression in the cell) and pMT V5/Bip HisB (secretion) (from *Invitrogen*) at restriction enzyme site between *spe1* and *Age1*. The plasmid and pT371 vector (which has resistant gene of puromycine) were co-transfected to S2 cells using effective transfection agent kit (Qiagen) which was grown in Schneider's medium (Lonza) added 10% FCS. After the transfection, the S2 cells adapted the Insect Xpress (Lonza) added 10 μ g/ml puromycine by the step of 1:0.25, 1:0.5, 1:0.75, 1:1: mixture of Schneider's medium and Insect Xpress. Expression of the protein was induced by the addition of 0.5 mM CuSO₄ to the Insect Xpress medium and the recombinant protein was expressed in the cells and secreted outside the cells. In the case of expression in the cells, the cells were harvested by

centrifugation and the recombinant protein was extracted using BagBuster (Novagen). The supernatant was collected and centrifugated at 4°C. To make membrane fraction, the supernatant was recollected and ultracentrifugated at 100,000g. The pellet was dissolved by 1.0% n-Dodecyl β -D maltoside (DDM) and ultracentrifugated at 100,000g.

To express the recombinant protein using *Pichia pastoris*, the *stt3* gene was accumulated by PCR and inserted into pPICZ_A at restriction enzyme site EcoRI. The 10 μ g of DNA was linearized in two hours at 37°C with Bgl II (Fermentas). Electrocompetent GS115 strain and KM71H strain were prepared following the protocol in the *Pichia* EasySelect™ Expression kit (Invitrogen). Linearized plasmid DNA was incubated with 80 μ L of electrocompetent GS115 and LM71H on ice and electroporated using a BioRad Gene Pulser at 2000V, 25 μ F, 100 Ω cuvette size 2 mm. Cells were plated on YPDS (1% yeast extract, 2% peptone, 2% glucose and 1M sorbitol) media containing 50 μ g/mL zeocin, and incubated at 30°C until colonies appeared (approximately 2~4 days). Colonies were inoculated into 5 mL of BMGY (1% yeast extract, 2% peptone 100 mM potassium phosphate buffer pH 6.0, 1.34% YNB and 0.001% biotin, 1% glycerol) and grown overnight at 30°C. Cultures were sub-inoculated into 100 mL of BMGY and grown for 24 h at 30°C at 120 rpm. When the cells had reached full growth, they were spun down and induced by resuspension of the cell pellet in 100 mL of BMMY (1% yeast extract, 2% peptone 100 mM potassium phosphate buffer pH 6.0, 1.34% YNB and 0.001% biotin, 1% methanol) media and incubated overnight at 30°C.

Protein purification

All purification steps were carried out at room temperature. The first step, the dissolved solution was loaded onto Ni-resin (GE Healthcare), washed with buffer

(50 mM Tris-HCl pH8.0, 100 mM NaCl, 0.1% DDM and 20 mM Imidazole) and eluted with buffer (50 mM Tris-HCl pH8.0, 100 mM NaCl, 0.1% DDM and 500 mM Imidazole) (Figure.1).

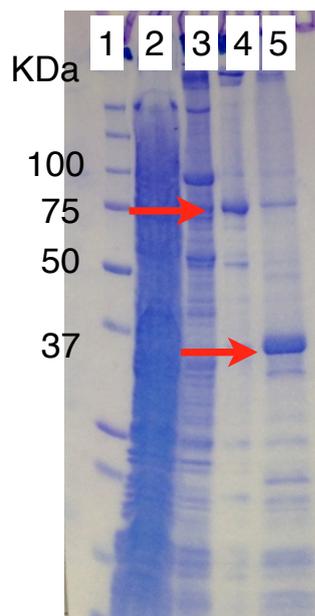


Figure 1. SDS-PAGE in Affinity chromatography (Ni-resin)

1 line shows molecular weight maker, 2 shows flow through, 3 shows wash without imidazole, 4 shows wash with imidazole and 5 shows elution, Red arrows show the LmSTT3_FL (high molecular weight) and LmSTT3_SD (low molecular weight) protein.

The next step, the solution was loaded onto Superdex200 10/300 column (GE Healthcare) for gel-filtration chromatography with buffer (20 mM Tris-HCl pH8.0, 100 mM NaCl, 0.01% DDM and 5 mM DTT) (Figure 2 and 3).

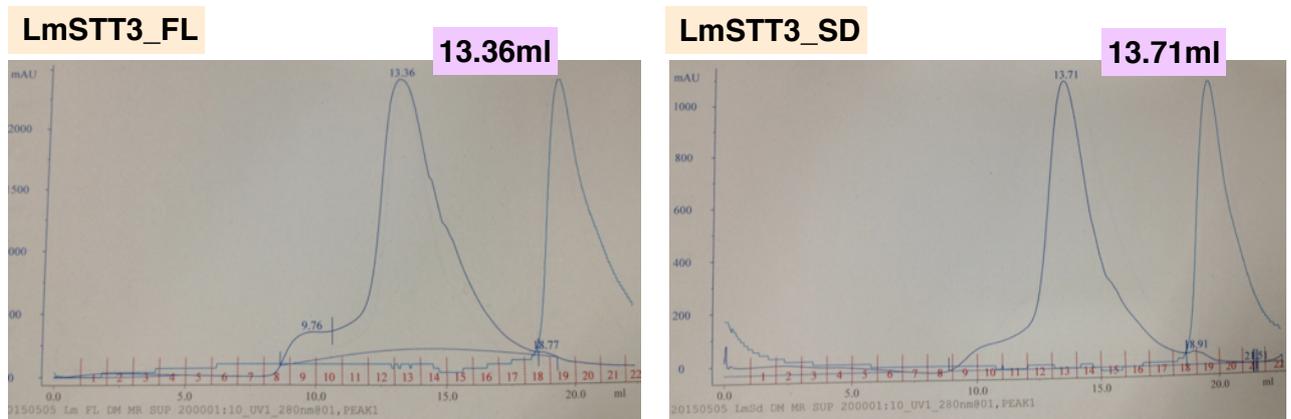


Figure 2. Chromatogram from gel-filtration chromatography
 Left panel shows LmSTT3_FL and right panel shows LmSTT3_SD and the retention time are 13.36 ml and 13.71 ml, respectively.

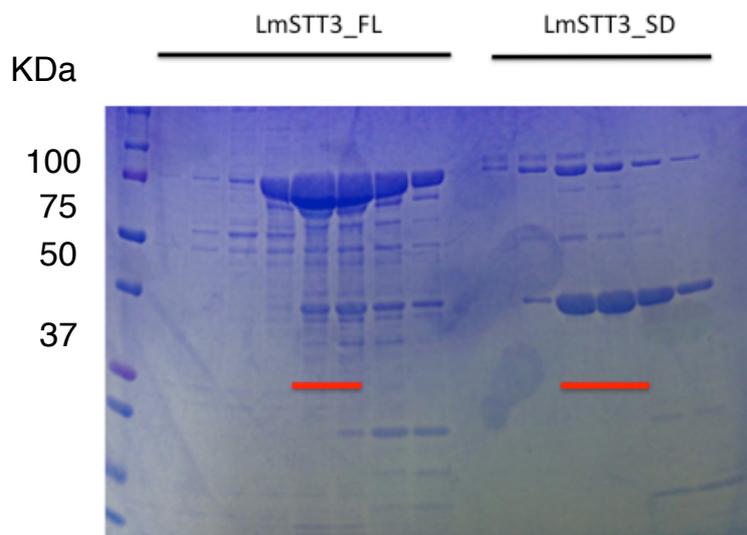


Figure 3. SDS-PAGE from gel-filtration chromatography
 The red underline shows the peak fractions for crystallization.

Preparation for Slenomthionine-labeled derivative of LmSTT3_SD and data collection

The cells inserted LmSTT3 gene were inoculated into 5 mL of YPD (1% yeast extract, 2% peptone, 2% glucose) and grown overnight at 30°C. Cultures were sub-inoculated into 1L of YPD and grown for 24 h at 30°C at 120 rpm. When the cells had reached full growth, they were spun down and induced by resuspension of the cell pellet in 1 L of culture media (23.4 g selenomethionine core media (Wako) , 100 mM potassium phosphate buffer pH 6.0, 1.34% YNB and 10µg inositol, 1% methanol) and incubated overnight at 30°C.

Crystallization and X-ray diffraction experiment

The all crystallization steps were held at 20°C using sitting drop vapor diffusion technique. The crystal conditions of soluble domain (LmSTT3_SD) grew under crystal conditions in 0.2M Ammonium phosphate monobasic, 0.05M PIPES pH7.0 and 20% PEG2000 from 1:1 mixture of 20mg/ml protein solution and reservoir. To improve the quality of crystals, crystal conditions were optimized and performed streak seeding (Figure 4 and 5).

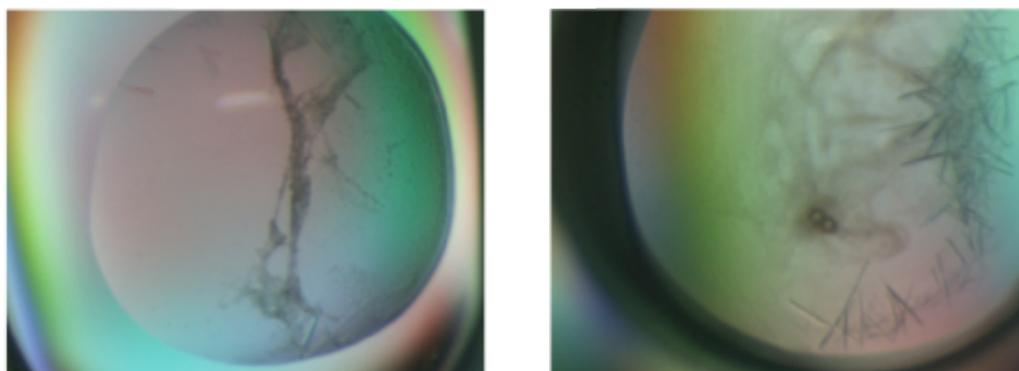


Figure 4. Crystals of LmSTT3_SD

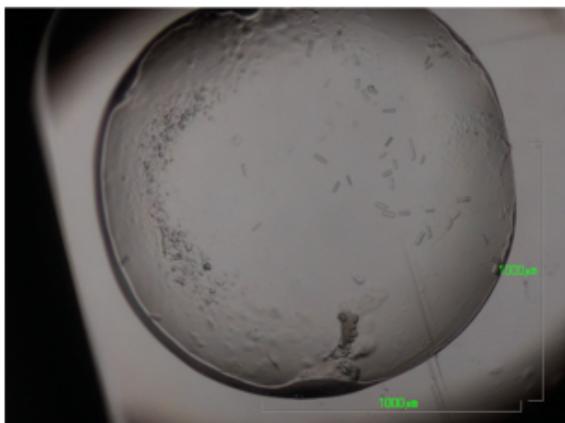


Figure 5. Crystals of LmSTT3_FL

For the X-ray diffraction experiment, the crystal was cryoprotected in the corresponding to the reservoir with by 20% glycerol (v/v) in liquid nitrogen. Diffraction data were collected at the 100K on the BL32XU and BL44XU at SPring-8 (Hyogo in Japan) and on the BL1A and BL5A at Photon Factory (Tsukuba in Japan) (Figure 6).

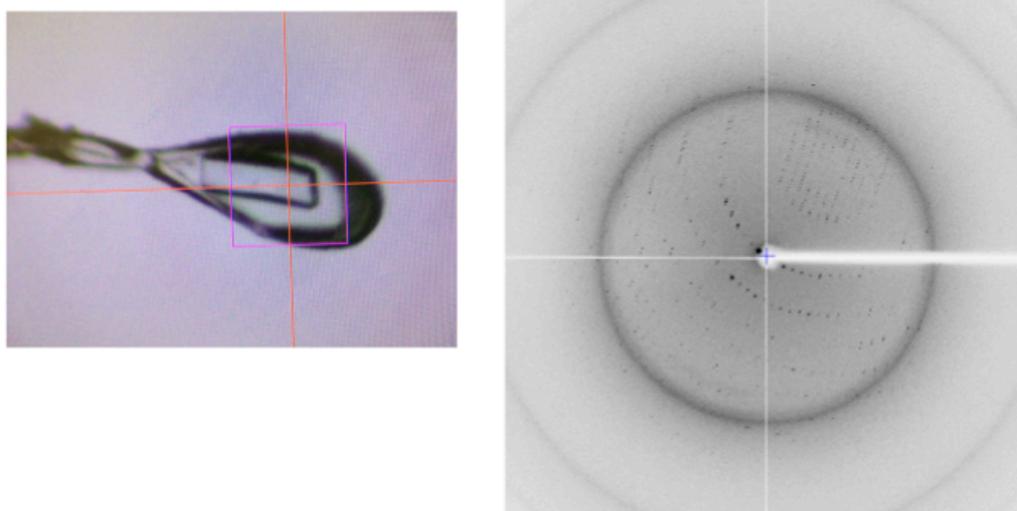


Figure 6. X-ray diffraction experiment

Left panel shows a crystal of LmSTT3_SD and pink square is X-ray beam position. Right panel is a diffraction pattern.

The diffraction data was integrated and scaled with HKL200 [7] or XDS [8, 9]. To observe the anomalous dispersion of selene, we selected the wavelength at 0.978 Å for the anomalous peak (Figure 7).

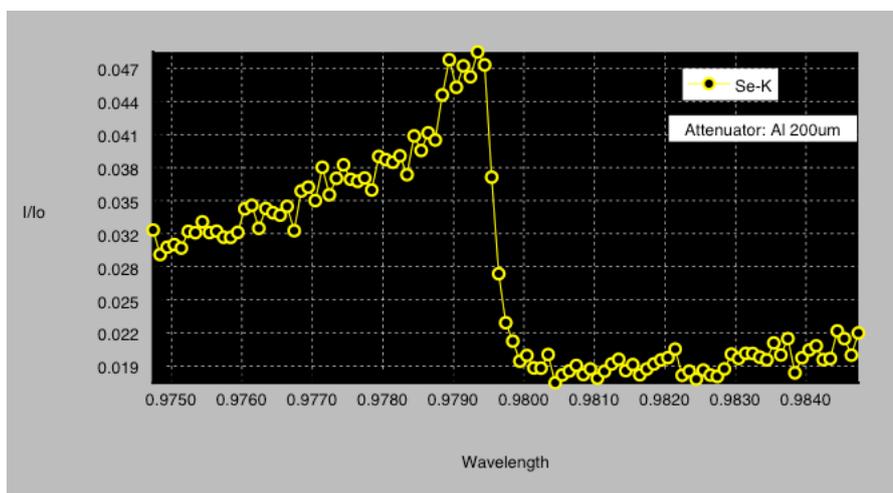


Figure 7. XAFS of LmSTT3_SD crystal

Structure determination of LmSTT3_SD

The structure determination was performed by molecular replacement. We used the phaser and molrep in CCP4 suites[10] [11] using template (PDB ID: 2LGZ). This template structure was determined by NMR spectroscopy and the solution was added SDS. This NMR structure is not appropriate to solve our crystal structure. Therefore, we used automatic search template software BALBES, but the BLABES did not find any solutions.

The selene structure for single anomalous dispersion (SAD) was searched using the HKL2MAP[12] with SHELXC, SHELXD and SHELXE. [13] SHELXC showed anomalous signal extending to 3.0 Å (Figure 8).

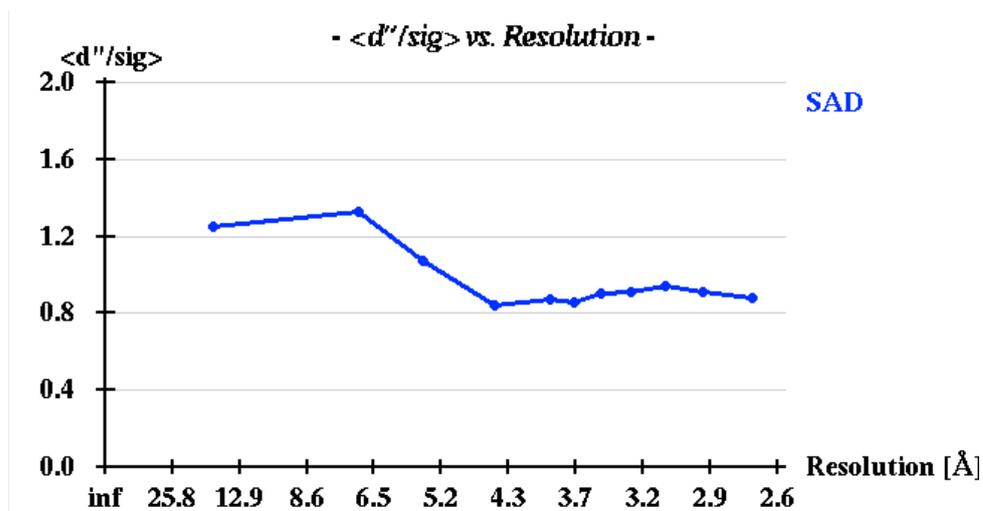


Figure 8. HKL2MAP graphical interface with SHELXC

It was difficult to locate any selene sites with SHELXD. After performing multiple runs (10000 trials per run) with different number of heavy-atom sites and by optimizing the parameter, the heavy-atom sites may be found by SHELXD.

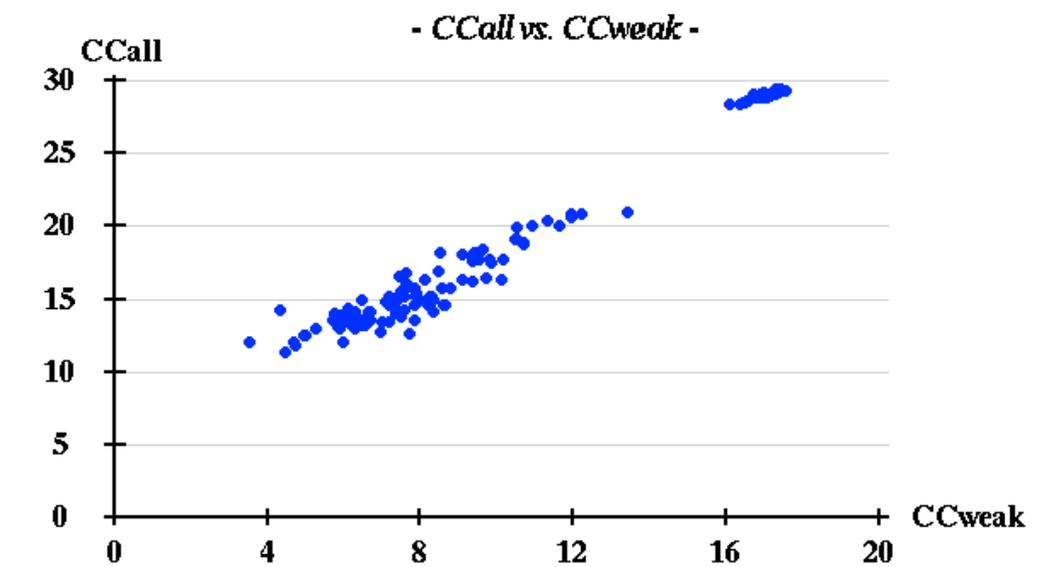


Figure 9. HKL2MAP graphical interface with SHELXD

By using the right solution, SHLEXE is performed for the electron density modification and creates initial maps (Figure 10 and 11). We didn't place the

some amino acid model into the initial electron density by COOT. [14]

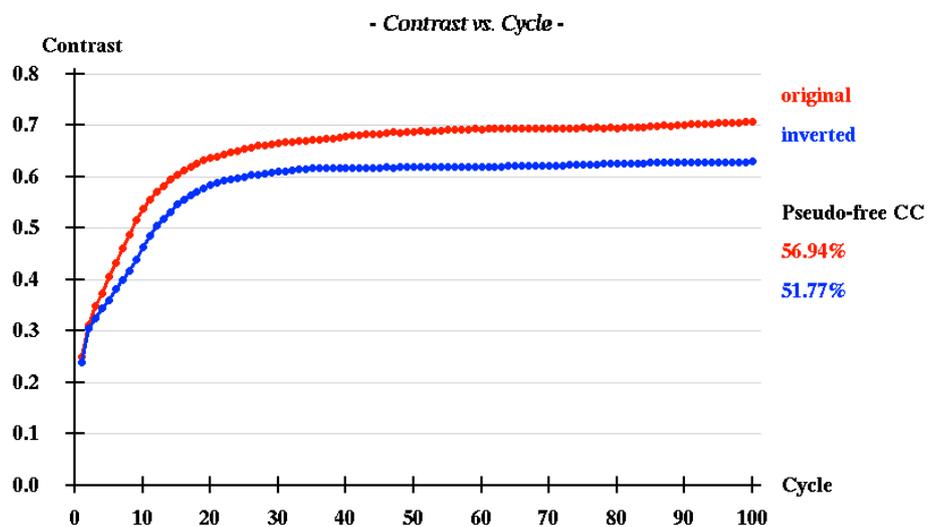


Figure 9. HKL2MAP graphical interface with SHELXE

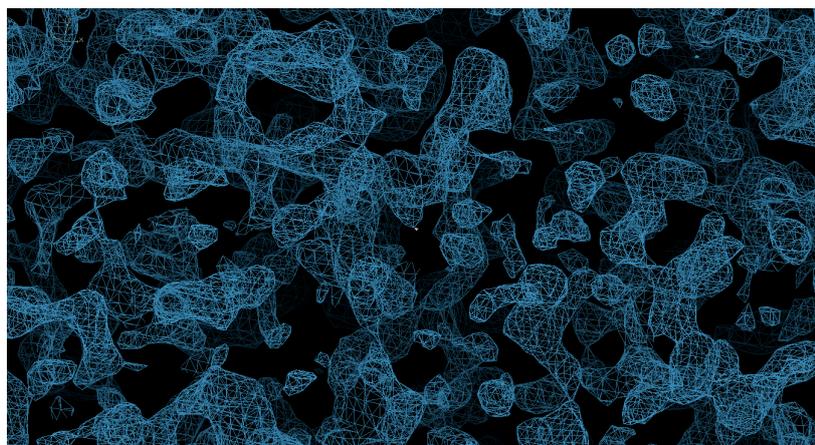


Figure 10. The initial electron density map by COOT

Conclusions

LmSTT3 protein was expressed purified and crystallized. At the X-ray diffraction experiment, we measured 3.0Å resolution of the crystal. There, we performed molecular replacement (MR) to determine the phase and solve the crystal structure. As the template for MR, we used the solution NMR structure of yeast STT3P soluble domain (PDB ID: 2LGZ). However, MR didn't work well. So, as the next step, we performed single anomalous dispersion (SAD) measurement. Selenomethionine-labeled protein of LmSTT3_SD was crystallized and diffracted at 3.0 Å resolution, which is same with non-labeled crystal. In addition, the result of XAFS experiment indicated that the selenomethionine-labeled was diffracted with anomalous signals. To solve the phase problem, we used HKL2MAP graphical interface with SHLEXC, SHELXD and SHELXE. As the result, the anomalous signal from selenomethionine was weak by SHELXC and SHELXD, and it was difficult to solve the structure by SHELXE.

To solve the structure, we have to improve the efficiency of the anomalous signal. The occupancies of selene are able to increase by optimizing the culture condition of *P.Pastoris*. The recently Sulfer-SAD (S-SAD) study succeed in weak anomalous signal of S atoms. The key of success is to determine the S atoms number and trial number for searching. The technique is needed to measure the high redundant data set from multi crystals. The advanced software is developed to determine the crystal structure in the case of weak anomalous signal, so these software helps to solve the phase problems.

References

1. Spirig, U., et al., *The STT3 protein is a component of the yeast oligosaccharyltransferase complex*. Mol Gen Genet, 1997. **256**(6): p. 628-37.
2. Zufferey, R., et al., *STT3, a highly conserved protein required for yeast oligosaccharyl transferase activity in vivo*. EMBO J, 1995. **14**(20): p. 4949-60.
3. Yan, Q. and W.J. Lennarz, *Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process*. J Biol Chem, 2002. **277**(49): p. 47692-700.
4. Gutierrez, A., et al., *Cloning of the Entamoeba histolytica STT3 gene, a subunit of the oligosaccharyltransferase complex*. Arch Med Res, 2000. **31**(4 Suppl): p. S162-4.
5. Hese, K., et al., *The yeast oligosaccharyltransferase complex can be replaced by STT3 from Leishmania major*. Glycobiology, 2009. **19**(2): p. 160-71.
6. Nasab, F.P., et al., *All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae*. Mol Biol Cell, 2008. **19**(9): p. 3758-68.
7. Otwinowski, Z. and W. Minor, *[20] Processing of X-ray diffraction data collected in oscillation mode*, in *Methods in Enzymology*. 1997, Academic Press. p. 307-326.
8. Kabsch, W., *Xds*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
9. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 133-44.
10. McCoy, A.J., et al., *Phaser crystallographic software*. J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
11. Potterton, E., et al., *A graphical user interface to the CCP4 program suite*. Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 7): p. 1131-7.
12. Pape, T. and T.R. Schneider, *HKL2MAP: a graphical user interface for*

- macromolecular phasing with SHELX programs*. Journal of Applied Crystallography, 2004. **37**(5): p. 843-844.
13. Sheldrick, G.M., *Experimental phasing with SHELXC/D/E: combining chain tracing with density modification*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 479-85.
 14. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

Abbreviations

List

- CCFS (Crystal contact-free space)
- MBP (Maltose Binding Protein)
- pALDH (presequence peptide derived from rat aldehyde dehydrogenase)
- THL (Turn-Helix-loop)
- NMR (Nuclear Magnetic Resonance)
- MD (Molecular Dynamics)
- OST (Oligosacharyl transferase)
- DHFR (Dihydrofolate reductase)
- PCR (Polymerase Chain Reaction)
- GST (Glutathione S-transferase)
- LB (Luria Broth)
- IPTG (isopropyl-1-thiol- β -D-thiogalactopyranoside)
- DTT (Dithiothreitol)
- SDS, Native-PAGE (SDS, Native- Poly-acrylamide Gel Electrophoresis)
- HPLC (High performance liquid chromatography)
- MALDI (Matrix Assisted Laser Desorption/ Ionization)
- TOF (Time of flight)
- MS (mass spectroscopy)
- MR (molecular replacement)

Acknowledgements

I appreciate Prof. Daisuke Kohda (from Div. of structural biology, Med. Inst. of bioregulation, Kyushu university) to help our experiments and suggest to improve the quality of the researches. I also thank Dr. Atsushi Shimada (from Div. of structural biology, Med. Inst. of bioregulation, Kyushu university and Institute of Physical and Chemical Research) to help X-ray experiments and advice structural analysis and validation of the structure. In addition, I thank all Beam line staffs in Photon Factory (PF, Tsukuba city in Ibaraki prefecture) and SPring-8 (Harima city in Hyogo prefecture). They help our X-ray experiments to do smoothly and give me some valuable advises. I think that we could not success our experiments without their helps. I thank Dr. Hiroaki Komuro and Yuji Sugita. They performed MD simulation in our study.

I thank our lab members from Div. of structural biology, Kyushu university and I appreciate Ms. Miki Otsu, especially. She took care of my travel to PF, SPring-8 or some conferences. I thank Dr. Satoru Yuzawa (Graduate School of Medical Sciences, Kyushu university), Dr. Hajime Niwa (Graduate School of System Life Sciences, Kyushu university) and Dr. Takao Hashiguchi (Graduate school of Medical Sciences, Kyushu university). I shared their Beam Time for X-ray experiment at PF or SPring-8.

Finally, I thank my parents to help my life and support my way, and my old brother encouraged me to perform valuable researches and give me some skills of computer programing. I'd like to apology to them that it was late to get the Ph.D.

Thank for everyone and everything that related me.

I was supported by a Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science. In addition, this work was supported by JSPS KAKNEHI Grant number 13J04493.