BL-17A/2022G642

Influence of Glycan Agents on Protein Crystallization with Ammonium Sulfate

Tyuji Hoshino[1,*] Yan Guo[1]

[1] Graduate School of Pharmaceutical Sciences, Chiba University,
1-8-1 Inohana, Chuo-ku, Chiba 260-8675, Japan

1 Introduction

The information from the high-resolution crystal structures is advantageous for understanding the molecular mechanisms of macromolecules and is essential in rational drug design. A limiting factor is obtaining single protein crystals of good quality since the crystallization process is influenced by many parameters such as precipitant agent, pH, temperature, etc. The predictions of the adequate values of those parameters are difficult. A current orthodox approach is to optimize the initial crystallization conditions obtained with the screening kits. Ammonium sulfate (AS) is a typical salting-type agent and is the most widely applied precipitant in protein crystal growth. Additionally, the protein crystallization with the salt-type precipitants can sometimes be improved by adding specific chemicals like small compounds. The additives serve to increase protein stability and mediate crystal lattice contacts. However, finding the appropriate additives to improve the quality of single crystals is an empirical procedure and a time-consuming process of trial and error.

In this study, two kinds of protein, influenza virus polymerase acidic subunit N-terminal domain (PAN) and chitosanase, were crystallized by the AS-based precipitants. The precipitant solutions contain trehalose and 3α,7α,12α-tri-((O-β-D-glucopyranosyl) ethyloxy)-cholane (Facade-TEG), respectively. Both crystals diffracted better than the crystals obtained in our previous works. Multiple MD simulations were performed with and without glycans to understand the protein–glycan interactions at an atomistic scale and to gain deep insight into the conformational dynamics of glycan-protein complexes. Root mean square deviation (RMSD), radius of gyration (Rg), and B-factors were obtained from simulation trajectories to assess the conformational stability of the complex structures. Principal component analysis (PCA) was employed to examine the conformational dynamics of protein structure. The binding energy between protein and glycan was also estimated using the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) approach. Next, the shapes of electrostatic potentials around the proteins were investigated and compared with the distributions of glycan agents. In the comparison, we examined if the electrostatic potential correlated with the favorable binding sites of glycan agents on the protein surface. Further, the crystal structures obtained by incorporating glycan agents in AS solution were surveyed through the Protein Data Bank (PDB) to assess the relevance of glycan agents to protein crystal growth from a statistical viewpoint.

2 Experiment

The PAN domain, residues 1-197, of the influenza H1N1 strain was cloned into the pET50b plasmid vector. A loop region, residues 51-72, was replaced by an Ala-Ser dipeptide linker to enhance the protein crystallization. The E. coli strain, C41 (DE3), transformed with the plasmid vector, was cultured in an LB medium. The protein was expressed at 17 °C for 48 hours after induction at an OD600 value of 0.8 with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG). A cell pellet obtained by centrifugation was resuspended in the buffer of 50 mM Tris-HCl at pH 8.0, 150 mM NaCl, and 10 mM imidazole. The bacterial cell membrane was disrupted by sonication.

The protein was purified by a Co metal affinity column with a linear gradient of the imidazole concentration. The eluted protein fraction was dialyzed overnight against a buffer containing 50 mM Tris-HCl at pH 8.0 and 150 mM NaCl. The 6×His-fused Nus-tag was cleaved by incubating with human rhinovirus 3C (HRV-3C) protease for 24 hours at 4 °C. The 6×His-fused Nus-tag, HRV-3C protease, and the remaining uncleaved protein were removed by Ni-NTA resin. PAN was further purified by anion exchange column with a starting buffer A (50 mM Tris-HCl at pH 8.0, 50 mM NaCl) and an elution buffer B (50 mM Tris-HCl at pH 8.0 with 1000 mM NaCl), followed by gel filtration. The gel filtration running buffer was of 20 mM Tris-HCl at pH 8.0 and 100 mM NaCl. The protein was finally concentrated at 10.2 mg/mL.

An N-terminus-truncated chitosanase was expressed as a 6×His-fused Nus-tag-conjugated form with a pET50b vector. An Escherichia coli strain, Rosetta (DE3) pLysS, was transformed with the plasmid vector. After pre-culture in an LB medium at 37°C, the protein was expressed at 26°C overnight by induction with 0.2 mM IPTG at an OD600 value of 0.6. The conjugated protein was purified by Co metal-affinity chromatography, followed by the cleavage of Nus-tag with HRV-3C protease. The cleaved Nus-tag, HRV-3C protease, and uncleaved conjugated protein were removed by NTA resin. The protein was further purified by gel filtration with a running buffer of 10 mM Tris-HCl at pH 8.0 and 150 mM NaCl. Finally, the protein was concentrated at 7.3 mg/mL.

A crystal of PAN was grown by vapor diffusion at 18°C in hanging drops containing 1.0 µL of protein solution including 4.0 mM MnCl2 and 1.0 µL of well solution consisting of 0.1 mM MES at pH 5.8, 1.275 M ammonium sulfate, 0.1 M potassium chloride and 9% (v/v) trehalose. The crystal was cryo-protected by brief immersion in a well solution containing 25% (v/v) glycerol, followed by freezing in liquid nitrogen.

A single crystal of chitosanase was grown by the sitting drop vapor diffusion method at 18°C. The protein solution was mixed with a precipitant solution consisting of 0.1 M sodium citrate at pH 5.0, 3.4 M AS, and 5% (v/v) Facade(R)-TEG at an equal volume. The crystal was cryo-protected by immersion in a solution containing 25% (v/v) glycerol before freezing.

X-ray diffraction data were collected at 100 K with a wavelength of 0.98 Å at the BL17A beamline of photon factory (PF), Tsukuba, Japan. The diffraction data were indexed, scaled, and merged by XDS. Intensities were converted into structure factors, and 5% of the reflections were flagged for Rfree calculations. The structures for PAN and chitosanase were determined by molecular replacement by MolRep with the crystal structures of PDB codes: 4ZQQ and 7CJU as search models, respectively. Structure refinement and model building were carried out using Refmac and Phenix. The molecular structure and electron density map were visualized by COOT.

3 Results and Discussion

Trehalose, C12H22O11, is a non-reducing disaccharide composed of two D-glucose units linked by an α-1,1-glycosidic linkage (Figure 1). Trehalose is useful for various applications, including enzyme catalysis, metabolism, life sciences, and medicine. The non-reducing glycosidic bonds of trehalose prevent hydrolysis under weakly acidic and high temperature conditions.32 Based on its high solubility, low reactive nature, and antioxidant properties, trehalose is particularly effective in improving the stability of the protein.

Facade(R)-TEG, C48H84O21, is a sugar-based facial amphiphile consisting of three β-D-glucoside units with uncharged polar groups and a steroid backbone (Figure 1). Three glucoside units are attached to the steroid backbone at the 3α-, 7α- and 12α-hydroxy group positions. Facade(R)-TEG improves protein stability and dispersion. Hence, Facade(R)-TEG is a desirable detergent for solubilizing membrane proteins.

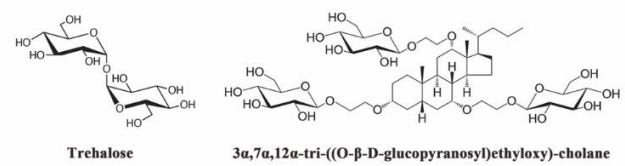

Fig. 1: Chemical structures of trehalose (left) and Facade(R)-TEG (right).

Single crystals of PAN and chitosanase were grown in the presence of glycan agents. The crystals of PAN appeared within 3 days after the setup, and the crystal sizes reached 0.5 mm at harvest. The rod-shaped chitosanase crystals were observed 4–6 days after setup, and the maximum size was 0.15 mm. The X-ray crystal structures were obtained for PAN (Figure 2a) and chitosanase (Figure 3a). The maximum resolutions for the protein crystals were 1.60 Å for PAN and 1.50 Å for chitosanase. These resolutions were 0.15 Å and 0.14 Å better than those obtained in our previous studies [1,2].

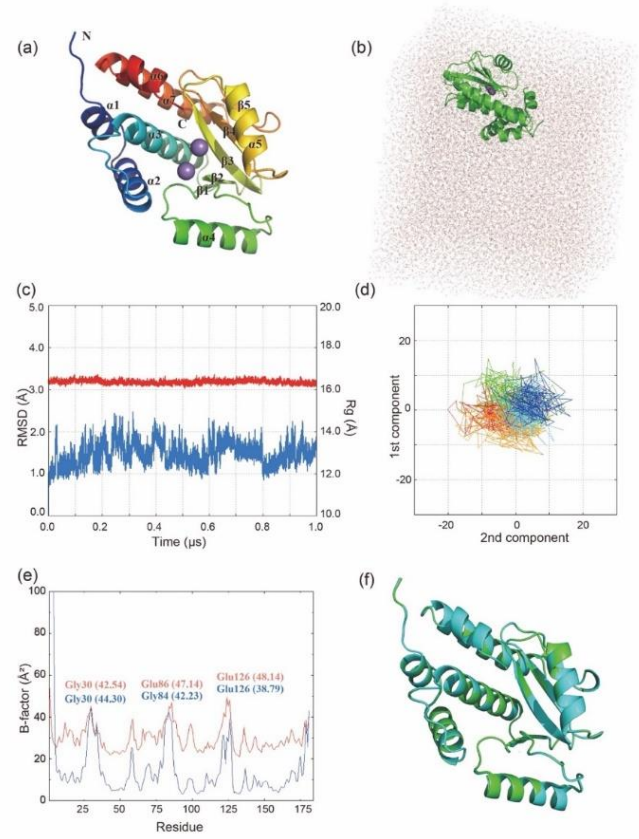

Fig. 2: (a) Crystal structure of PAN. Protein structures are represented in cartoons, where the residues of the N-terminal sides are indicated in blue and the C- terminal sides are shown in red, and manganese ions are shown in purple spheres. (b) Initial calculation model of PA_N without glycan. Protein molecules are represented by green cartoons, water molecules are represented by red (oxygen) and white (hydrogen) thin lines, sodium ions are indicated by purple dots, and chloride ions are indicated by green dots. (c) Time evolution plot of RMSD (blue) and the radius of gyration (red) for the PA_N without glycan. (d) Two-dimensional projection of PCA. The last 0.1 µs trajectory is extracted from the 1 µs MD simulation, where the color graduation from red to blue represents the time evolution for 0.1 µs. (e) B-factors of main-chain atoms of the individual amino acid residues for the last 0.1 µs simulation trajectory (blue) and B-factors of the crystal structure (red). The residues with prominent peaks in B-factor are labeled in blue and red, respectively. (f) Superposition of the crystal structures for PA_N obtained with (cyan) and without (green) trehalose, viewed in the same orientation as (a).

X-ray structure analysis revealed that the molecular packing in the crystals grown with glycan was almost the same as those of the crystals without glycan. Hence, the crystallographic symmetry was equal between the crystals with and without glycan. The number of protein molecules in the asymmetric unit was one for the trehalose-facilitating PAN crystal and two for the Facade(R)-TEG-facilitating chitosanase crystal. The space group was P41212 for the trehalose-facilitating PAN crystal and P212121 for the Facade(R)-TEG-facilitating chitosanase one. These space groups were the same as the crystals grown without glycan. According to our previous work [3], the precipitant utilized for crystallization significantly impacts on the space group of crystals. Since the inclusion of glycan hardly changed the space groups of crystals, the effect of additives on the crystal structure is relatively less than that caused by precipitant agents. A superposition of the crystal structures with and without glycan demonstrated fine overlaps. The root-mean-square deviation (RMSD) was 0.128 Å for PAN and 0.123 Å for chitosanase (Figures 2f and 3f).

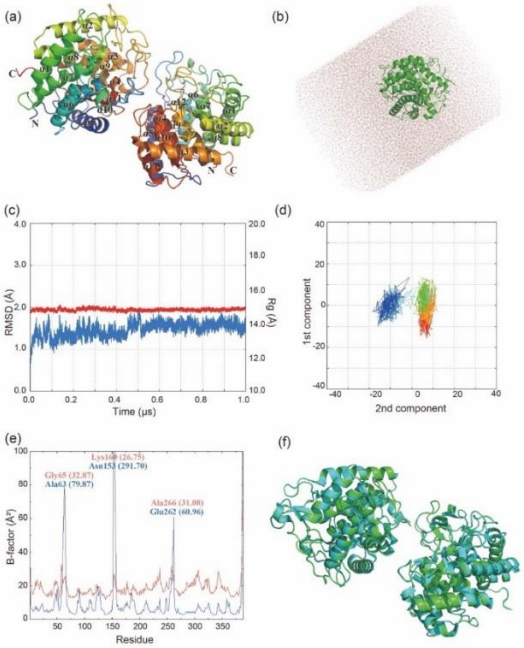

Fig. 3: (a) Crystal structure of chitosanase. Protein structures are represented in cartoons. (b) Initial calculation model of chitosanase without glycan. (c) Time evolution plot of RMSD and radius of gyration for the chitosanase without glycan for the whole simulation time. (d) Two-dimensional projection of PCA. The last 0.1 μs trajectory is extracted from the 1 μs MD simulation, where the color graduation from red to blue represents the time evolution for 0.1 μs. (e) B-factors of main-chain atoms of the individual amino acid residues for the chitosanase without glycan and B-factors of the crystal structure. (f) Superposition of the crystal structure of chitosanase crystallized by AS with and without Facade(R)-TEG, viewed in the same orientation as (a). The color and representation in the drawing are the same as those in Figure 2.

A reasonable initial model is a prerequisite for accurate computer simulation. In this study, the initial protein structures were obtained from the respective crystallographic experiments above. One trehalose molecule was included in the PAN calculation model, and one Facade(R)-TEG was in the chitosanase one. Eight independent 1 μs MD simulations were performed to enhance the reliability of simulation analysis. Hence, the glycan molecule was randomly placed at eight different positions in the solvent around the protein. The total simulation time of 8 μs (8 × 1 μs) will ensure adequate sampling of binding poses. For comparison, one 1 μs MD simulation was also performed for the model without including glycan (Figures 2b and 3b). Therefore, 18 independent trajectories were analyzed to monitor the glycan-protein interaction.

The electrostatic potentials were depicted for PAN and chitosanase (Figure 4). The iso-surface of the electrostatic potentials around PAN showed that the separation of the positive and negative areas is distinct. Their contact area was restricted to the vicinity of protein (Figure 4a). The negative areas had a spherical lobe, whereas the positive area slightly deviated from a spherical shape. Protein molecules are positioned at the boundary area between two spherical lobes. The distributions of glycan molecules in the PAN-trehalose complex showed that trehalose molecules made contact with the protein surface at the boundary region of two areas of the electrostatic potential in five models. In contrast, trehalose was located in the positive area in the three models. No glycan molecule was observed in the negative area of the electrostatic potential. The electrostatic potential for chitosanase in Figure 4b showed that the positive and negative areas are almost equivalently separated in volume. Each area has a spherical lobe shape expanding in the opposite direction. The positive and negative areas clearly split even at the boundary, and the contact region of the two areas is limited in the vicinity of protein. Facade(R)-TEG molecules are likely to be located at the boundary region of the two spherical lobes. With three exceptions, glycan molecules are positioned at the boundary regions for chitosanase, where the absolute value of the electrostatic potential energy is low. No Facade(R)-TEG is positioned at the positive area of the electrostatic potential [4].

The function of ammonium sulfate in protein crystallization is considered to vary salt concentration-dependent solvation with changing the local solute-solvent interactions. Sulfate ion is a strongly hydrated anion, and it competes for water molecule with the solute. The ions tended to gather at the positive and negative areas of the high absolute values of the electrostatic potential.12 A strong anisotropy induced by AS ions is effective for making solute molecules aligned and perturbs their exposure

to the solvent. Then, solute molecules are contacted each other in the areas where the absolute value of the electrostatic potential is low. Glycan molecules were likely to be positioned at the boundary region of the two spherical lobes with a low absolute value of electrostatic potential. Glycan molecules will alter the intramolecular interactions between proteins to promote the formation of a crystal lattice. The loose binding of trehalose to PAN is advantageous for crystal growth because the release of attached molecules will induce the ordered association of macromolecules. Although Facade(R)-TEG is stably bound to chitosanase in MD simulations, the glycan molecule is not observed in the crystallography. Therefore, Facade(R)-TEG is also released from the protein surface in crystal growth, which will assist the regular packing of macromolecules. Furthermore, the lattice order depends on precipitant concentration, in which high concentrations of AS may improve the stabilization of charged residues. The compatibility between the shape of the electrostatic potential and the distribution of glycan molecules suggested their substantial influence on the crystal packing.

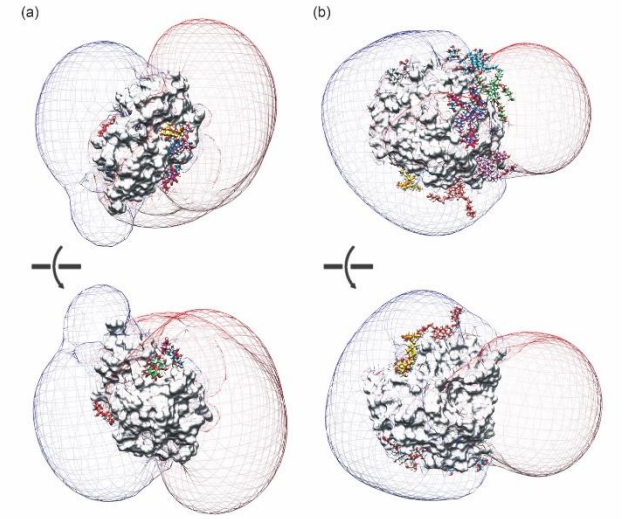

Fig. 4. (a) Positive and negative iso-surfaces of electrostatic potentials for PAN with trehalose molecules. Protein is represented by a white surface, and eight trehalose molecules are in sticks (yellow, magenta, green, white, blue, dark blue, pink, and red). The positive and negative iso-surfaces are drawn in blue and red meshes, with their contour values at +0.5 and −0.5 in units of kT/e, respectively. (b) Electrostatic potentials for the chitosanase with Facade(R)-TEG molecules. Chitosanase is represented by a surface of white, and eight molecules of Facade(R)-TEG are in stick (magenta, pink, green, blue, dark-blue, red, yellow, and white). The drawing for electrostatic iso-surface is the same as in (a).

In conclusion, X-ray structure analysis was carried out for PAN and chitosanase, in which crystals were grown by the AS solution containing glycoside-glycans. The maximum resolutions were 1.60 Å for PAN and 1.50 Å for chitosanase. These resolutions were better than those of the crystal structures obtained in the previous studies. MD simulations were performed for the glycan-included models to examine the behavior of glycans on a protein surface. The shape of the electrostatic potential for both proteins revealed that many glycan molecules were positioned at the boundary region of the positive and negative regions, where the absolute value of electrostatic potential is low. The findings of this work will help the understanding of the role of glycan in protein crystallization and provide a hint for the optimization of additive agents.

Acknowledgement

Calculations were performed at the Research Center for Computational Science, Okazaki, Japan and at the Information Technology Center of the University of Tokyo. A part of this work was supported by a grant for Scientific Research C from the Japan Society for the Promotion of Science.

References

[1] Fudo, S., Qi, F., Nukaga, M., Hoshino, T. *Cryst. Growth Des.* **17**, 534-542 (2017).
[2] Guo, Y., Qu, L., Nishida, N., Hoshino, T. Electrostatic Potentials around the Proteins Preferably Crystallized by Ammonium Sulfate. *Cryst. Growth Des.* **21**, 297– 305 (2021).
[3] Kitahara, M., Fudo, S., Yoneda, T., Nukaga, M., Hoshino, T. *Cryst. Growth Des.* **19**, 11, 6004-6010 (2019).
[4] Guo, Y., Hoshino, T. *Cryst. Growth Des.*, **22**, 6751–6765 (2022).

* hoshino@chiba-u.jp