BL-5A, BL-1A /2016G554

Influence of precipitants on molecular arrangement and space group of protein crystals

Tyuji Hoshino[1,*] Satoshi Fudo[1] Yuji Komukai[1] and Michiyoshi Nukaga[2]
[1] Graduate School of Pharmaceutical Sciences, Chiba University,
1-8-1 Inohana, Chuo-ku, Chiba 260-8675, Japan
[2] Faculty of Pharmaceutical Sciences, Josai International University,
Gumyo 1, Togane-shi Chiba 283-8555, Japan

1 Introduction

In our previous work, we carried out the cluster analyses of crystal structures for a single same kind of protein, exemplifying human immune-deficiency virus type 1 (HIV-1) protease, hemoglobin, myoglobin, human serum albumin, and hen egg-white lysozyme. In each kind of proteins, crystal structures were separated into several groups based on their structural similarities. From this analysis, it was found that the crystal structures belonging to the same group were crystallized by using the same kind of precipitants. It was also shown that the crystal structures belonging to the same group almost always had a common space group. These results suggest that precipitants strongly affect how each protein molecule makes contacts with other protein molecules in crystal and the molecular contacts in crystal are reflected in the space group.

In this study, we obtained three crystal structures for a single same kind of protein with three different crystallizing agents. The aim of the present study is to analyse the behaviour of precipitants in terms of the interaction with protein molecule in solution and to clarify the influence of precipitants on the manner of molecular arrangement of proteins in crystal. The protein used for this study was influenza virus polymerase acidic subunit N-terminal domain one-loop deletion recombinant ($PA_N^{\Delta loop}$), which was handled in our previous works.[1, 2] First, we crystallized this protein using the three different precipitants; ammonium sulfate, potassium sodium tartrate, and PEG 8000. Next, molecular dynamics (MD) simulations were carried out for the protein solution with each of these precipitants. Then the relationship between the molecular packing in crystal and the molecular geometry in simulation was compared.

2 Experiment

Truncated PA_N protein, $PA_N^{\Delta loop}$, from influenza virus A/Puerto Rico/8/34 (PR8) strain (H1N1) was expressed and purified as follows. The *Escherichia coli* strain transformed with the pET50b(+) vector containing the $PA_N^{\Delta loop}$ gene was cultured in LB medium. The protein was expressed at 17 °C for 48 h after induction with 0.2 mM IPTG. The protein was purified by a His-affinty column, followed by the cleavage of 6×His-fused Nus-tag by HRV 3C protease. The cleaved protein was again purified by Ni-NTA resin. The protein was further purified by gel filtration with a running buffer of 20 mM Tris-HCl at pH 8.0 and 100 mM NaCl. Finally, the protein was concentrated to 9.9 mg/mL.

Three crystal structures for $PA_N^{\Delta loop}$ were obtained with three different crystallizing agents **1**, **2**, and **3**. Crystals of $PA_N^{\Delta loop}$ were grown by the vapor diffusion method with hanging drops consisting of 1.0 μL of 9.9 mg/mL protein solution containing 4.0 mM $MnCl_2$ and 1.0 μL of crystallizing agent **1**, **2**, or **3** at 18°C. The composition of crystallizing agent **1** is 100 mM MES at pH 5.8, 1.1 M ammonium sulfate, 0.1 M potassium chloride and 9% (v/v) trehalose. The crystallizing agent **2** includes 100 mM MES at pH 6.0, 0.8 M potassium sodium tartrate, and the agent **3** includes 100 mM Tris-HCl at pH 7.5, 28% PEG 8000, 0.2 M calcium acetate, 10% Jeffamine M-600. The crystals were cryoprotected by brief immersion into well solution containing 22.5% (v/v) glycerol, followed by flash-freezing in liquid nitrogen.

X-ray diffraction data for the crystal structures were collected at 100 K on PF BL-5A or BL-1A beamline. The diffraction data were indexed, scaled and merged with HKL2000. Intensities were converted into structure factors and 5% of the reflections were flagged for R_{free} calculations. Atom coordinates for crystal structures were determined by the molecular replacement with MOLREP program. Structure refinement and model building were carried out using PHENIX and COOT. The PDB codes of crystal structures **1**, **2**, and **3** are 4ZQQ, 5JHT, and 5JHV.

The crystal structure **1** was used as a template for building the models for MD simulation. The template was placed in a rectangular periodic boundary box filled with TIP3P waters, and counter ions were added to neutralize the model system using the LEaP module in AMBER 14. The model at this step was used as an initial structure for MD simulation without any precipitant (simulation **0**). To build the models which include ammonium sulfate as a precipitant, three water molecules were chosen randomly in the model for simulation **0** and then two of them were replaced with ammonium ions and the other one with sulfate ion. This operation was repeated until the concentration of ammonium sulfate was the same as that in the crystallizing droplet in experiments. The procedure for replacing water molecules with ammonium sulfate was done twice to make two independent initial models for MD simulations with ammonium sulfate (simulations **1-a** and **1-b**). Similarly, the water molecules chosen randomly in the model for simulation **0** were replaced with either potassium, sodium, or tartrate ion, and the procedure was duplicated to make two independent initial

models for MD simulations with potassium sodium tartrate (simulations **2-a** and **2-b**). Since the length of PEG 1000 in the linear conformation is almost the same as that of the longest edge of the rectangular box, the random generation, as done for the other precipitants, was not applicable. Hence, PEG 1000 molecules were generated to surround the protein in the model for simulation **0**, and the water molecules which overlapped with a PEG molecule were deleted. The coordinates of PEG molecules were changed between two initial models for MD simulations, by rotating every other PEG by 180° around the axis perpendicular to the primary molecular axis (simulations **3-a** and **3-b**). It should be emphasized again that the number of precipitant molecules in each model was set to reflect the initial concentration of precipitant in the hanging drop in experiments. Therefore, the concentration of precipitant in each model is a half of that in well solution.

3 Results and Discussion

The crystal structures **1**, **2**, and **3** were obtained by using the different precipitants for crystallizing agents (Table 1). The numbers of $PA_N^{\Delta loop}$ molecules in an asymmetric unit were 1, 1, and 2 for the crystal structures **1**, **2**, and **3**, respectively. The structures of these protein molecules are almost the same among them as shown in Figure 1a. For the crystal structure **3**, the protein molecule of only chain A is displayed because structural difference between chain A and B is quite small. The space groups of the crystal structures **1**, **2**, and **3** are $P4_12_12$, $P4_12_12$, and $C2$, respectively. According to our previous works, a precipitant used for crystallization strongly affects the space group of the crystal. Because the space groups for crystal structures **1** and **2** are the same, ammonium sulfate and potassium sodium tartrate, which were used as precipitants for the crystals, are considered to have similar influence on the protein. Figure 1b - 1d shows how protein molecules in multiple asymmetric units are packed in crystals. Obviously the packing modes in the crystal structures **1** and **2** are same since they have the same space group and almost same cell constants. There are 8 protein molecules per one unit cell in all of the three crystal structures, while the size of the unit cell of the crystal structure **3** is much smaller than those of the others, which means the proteins in the crystal structure **3** are tightly packed.

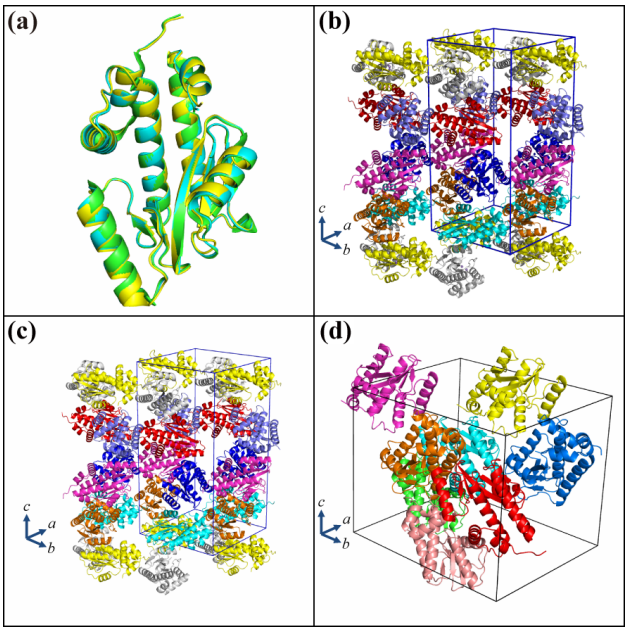

Fig. 1: Crystal structures **1**, **2**, and **3**. (a) Superimposition of the three crystal structures **1** (green), **2** (cyan), and **3** (yellow, chain A only). (b-d) Crystal packing of structures **1** (b), **2** (c), and **3** (d). Unit cells are depicted in lines, and *a*, *b*, and *c* axes are depicted in arrows. In all of the three crystals, one unit cell contains eight protein molecules, each of which is depicted in a different color.

Table 1: Crystallographic properties

Crystal structure	**1**	**2**	**3**
Precipitant	Ammonium sulfate	Potassium sodium tartrate	PEG 8000
Space group	$P4_12_12$	$P4_12_12$	$C2$
a, *b*, *c* (Å)	66.33, 66.33, 127.34	66.86, 66.86, 128.00	64.24, 87.40, 66.50
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 94.31, 90.00
Resolution (Å)	29.41-1.80	32.34-1.75	36.49-2.75
No. reflections (R_{free} set)	25617 (1316)	29882 (1519)	9578 (514)
R_{work}/R_{free}	0.224/0.274	0.194/0.238	0.197/0.234

To clarify the influences of precipitants on protein molecules and crystal structures, we carried out MD simulations for a $PA_N^{\Delta loop}$ molecule without any precipitant (simulation **0**) and with ammonium sulfate (simulations **1-a** and **1-b**), potassium sodium tartrate (simulations **2-a** and **2-b**), and PEG 1000 (simulations **3-a** and **3-b**). PEG 1000 is a substitute for PEG 8000 in simulation. These models (except for that for simulation **0**) were constructed so as to reproduce the concentrations of the precipitants in the hanging drops at the moment when they were set up by mixing the protein solution and one of the crystallizing agents. In order to focus on the precipitants, the other minor chemicals included in the crystallizing agents were ignored. In simulations **3-a** and **3-b**, simulation time was set to 200 ns, while the others to 100 ns, because the movement of PEG 1000 was

relatively slow and the equilibration enough for the reliable analysis needed longer simulation time.

The averaged distributions of the precipitant molecules were analyzed for all of the simulations using the trajectories of the last 20 ns. Figure 2a shows the distributions of ammonium ions in dots and sulfate ions in mesh for simulation **1-a**. The distributions of these two ions are similar to each other and, to our surprise, the distributions are not isotropic around the protein. To clarify what these anisotropic distributions mean, the molecular arrangement in crystal structure was contrasted to the distribution. In Fig. 1b, each protein molecule has the same surrounding environment and has direct contacts with four surrounding protein molecules. One protein molecule with its four surrounding molecules were extracted as shown in Fig. 2b. Then the distributions of the two kinds of ions (Fig. 2a) and the extracted protein molecules (Fig. 2b) were superimposed (Fig. 2c, backside views in Fig. 2d). Intriguingly, there were almost no distributions of precipitants at the areas where the protein molecule has contacts with other molecules in the crystal structure. In other words, there was almost no overlap between the distribution areas of precipitants (result of the simulation) and the contact sites between proteins (result of the crystal structure analysis).

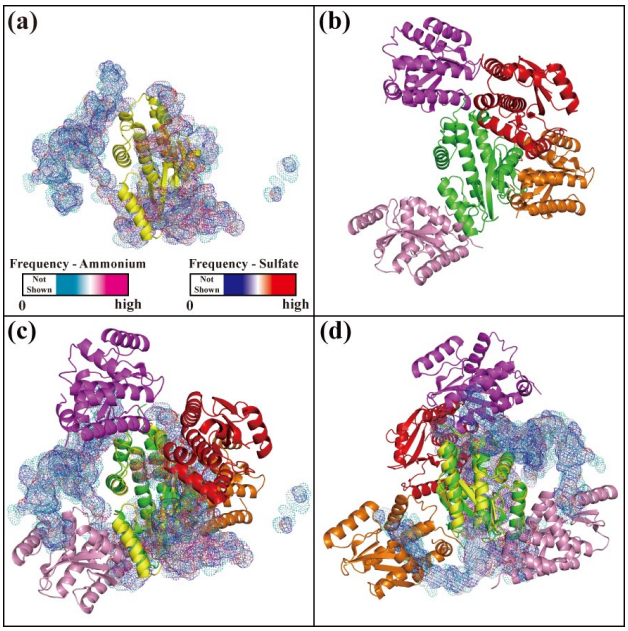

Fig. 2: (a) Distributions of ammonium and sulfate ions in simulation **1-a**. The distribution of ammonium ions is depicted in dots and that of sulfate ions is in mesh. They are colored according to the rate of the presence of those ions in each point over the last 20 ns in the simulation (Color scales for both ions are depicted in the bottom.). Protein molecule is depicted in cartoon (yellow). (b) Direct contacts of protein molecules in the crystal structure **1**. A protein molecule (green) has direct contacts with four other protein molecules (red, orange, magenta, and pink). (c, d) Superimposition of (a) and (b), viewed from two different angles.

Figure 3a shows the averaged distributions of the potassium and sodium ions in dots (as a sum of the two ions) and the tartrate ion in mesh for the last 20 ns of simulation **2-a**. The distributions of cations and tartrate ions were similar to each other but not isotropic around the protein. In the same manner as ammonium sulfate, a protein molecule and four protein molecules making direct contact with the protein were extracted from the crystal structure **2** (Fig. 3b) and superimposed on the distributions of cations and tartrate ions in simulations **2-a** and **2-b** (Fig. 3c, backside view in Fig. 3d). Again, there were no overlap between the distribution areas of precipitants and the contact sites of proteins.

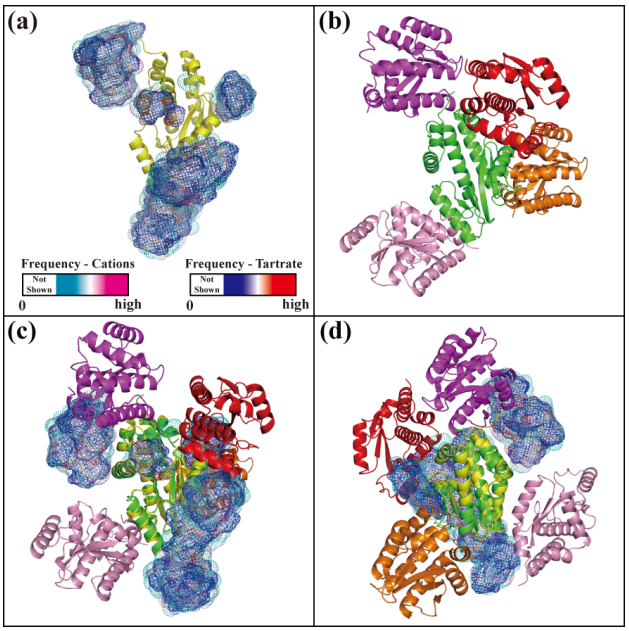

Fig. 3: (a) Distributions of cations and tartrate ions in simulation **2-a**. The distribution of cations (as a sum of sodium and potassium ions) is depicted in dots and that of tartrate ions is in mesh. They are colored according to the rate of the presence of those ions. Protein molecule is depicted in cartoon (yellow). (b) Direct contacts of protein molecules in the crystal structure **2**. A protein molecule (green) has direct contacts with four other protein molecules (red, orange, magenta, and pink). (c, d) Superimposition of (a) and (b), viewed from two different angles.

Based on the findings in this study, we propose the mechanism of the protein crystallization under the first two precipitants as follows. First, precipitant molecules are attracted to the protein surface due to the electrostatic potential and are distributed heterogeneously around the protein. This causes the electrostatic screening of the protein surface charge, which leads to the suppression of the electrostatic repulsion of protein molecules, and the stabilization of the conformational fluctuation of protein molecules, both of which make the protein solubility lower. Then protein molecules are promoted to make contacts with each other. Precipitant molecules covering some sites of the protein surface prevent the other protein

molecules from accessing to the covered sites. Hence, the contact sites on the proteins are restricted, which helps protein molecules to form regularly ordered clusters, that is, to crystallize with a specific space group.

Figure 4a shows the averaged distributions of PEG molecules for the last 20 ns of simulations **3-a**. PEGs cover most parts of the protein surface, which is considerably different from the results of the other two precipitants. In the crystal structure **3**, each protein molecule has direct contacts with other 10 molecules (Fig. 4b). The protein molecule of chain A in the crystal structure was used as a center molecule in these figures (colored green) since the two molecules in an asymmetric unit are almost identical in structure and contact geometry. In a similar manner to the other two precipitants, the distributions of PEG molecules (Fig. 4a) and the extracted crystal structure (Fig. 4b) were superimposed and compared (Fig. 4c, d). Although many parts of the protein surface are covered by PEGs or surrounding protein molecules, only the sites that are not covered by PEGs in the simulations have no contact with other protein molecules in the crystal structure (Those sites are indicated by red arrows in Fig. 4c, d).

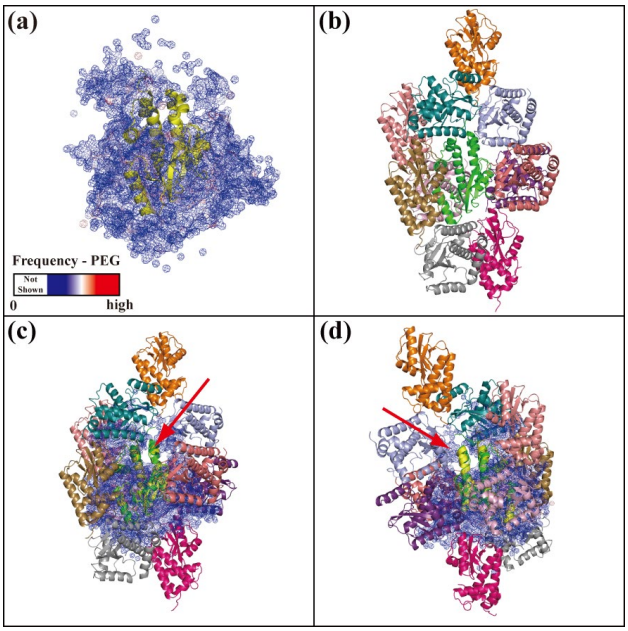

Fig. 4: (a) Distribution of PEGs in simulation. The distribution is shown in mesh. Mesh is colored according to the rate of PEG atom presence at each point over the last 20 ns in the simulation. Protein molecule is depicted in cartoon (yellow). (b) Direct contacts of protein molecules in the crystal structure **3**. A protein molecule (green) has direct contacts with 10 other protein molecules (10 different colors). (c, d) Superimposition of (a) and (b), viewed from two different angles.

Based on the findings that protein molecules preferentially make contacts with each other at the sites covered by PEGs in simulation, we propose the crystallization process using PEG as a precipitant as follows. In the initial stage of crystallization just after the mixture of protein solution and crystallizing agents for droplet, PEGs are bound to some specific sites of protein molecules. The molecular contacts of proteins are formed with the progress of crystallization, accompanying the release of PEG molecules that were originally bound to the contact sites. The release of PEG molecules leads to the increase of the entropy of the system, which efficiently drives the crystallization. This can be regarded as a kind of depletion forces, in which the decrease of the depletion zone by the increase of contacts of macromolecules is promoted by the entropic gain of PEG molecules in solution. Accordingly PEGs play a role of guiding protein molecules for making adequate contacts and packing, thereby leading to a specific space group in crystal.

In conclusion, three kinds of crystals were obtained with three different precipitants for a single same kind of protein. The space group of one of the crystals was different from that of the others. The relationship between the precipitants used in the protein crystallization and the molecular packing in the crystals was investigated by combining crystallographic structure analysis and MD simulation. It was found that precipitant molecules were not randomly distributed around the protein and, instead, the distribution showed a strong anisotropy. MD simulations with precipitants in a high-concentration equivalent to the crystallizing condition enabled us to clarify molecular interactions of precipitants with the protein. Our calculation suggested that heterogeneous distribution of precipitants played a role in restricting the contact sites of the protein, thereby causing a specific molecular packing in crystal and leading to the growth in the corresponding space group. Accordingly, the precipitant that can sufficiently restrict the contact sites of protein molecules will be one of the critical elements for successful crystallization.

Acknowledgement

Calculations were performed at Research Center for Computational Science, Okazaki, Japan and at Information Technology Center of the University of Tokyo. This work has been performed under the approval of the Photon Factory Program Advisory Committee (proposal no. 2014G563, 2015G554).

References

[1] S. Fudo *et al.*, *Bioorg. Med. Chem.* **23**, 5466 (2015).
[2] S. Fudo *et al.*, *Biochemistry* **55**, 2646 (2016).

* hoshino@chiba-u.jp