BL-5A, AR-NW12A/2016G554

# Formation of Chain-like Arrays of Ammonium Sulfate Ions in Protein Crystallization

Tyuji Hoshino[1,*] Mariko Kitahara[1] Satoshi Fudo[1] Tomoki Yoneda[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

Protein structure is fundamental information to understand the property, shape, and function of a protein. X-ray crystal analysis is one of the most popular approaches to determine the protein structure. Growth of single crystal is a critical step in the X-ray crystallography. Ammonium sulfate (AS) is one of the most popular precipitant agents to induce the crystallization of protein molecules. AS is, however, not always effective for all kinds of proteins in crystal growth. Some proteins are readily crystallized by utilizing AS as a precipitant but some proteins are not. The reason why the probability of crystal growth by AS is different among proteins is unknown. Similar to AS, poly-ethylene glycols (PEGs) are frequently utilized in protein crystallization. Crystal growth of some kinds of proteins are effectively induced by PEGs, instead of AS. The cause for the dependency of protein crystallization on precipitant has not been studied in the atomistic level so far.

In our previous study [1], a single kind of protein was grown with three different precipitant agents, including AS and PEG. Molecular dynamics (MD) simulations were carried out for the protein in the presence of precipitants in the concentrations equivalent to the respective crystallization conditions. As a result, the distribution of precipitant molecules was found to be not isotropic around the protein. In the simulations with AS, AS ions play roles not only of decreasing the protein solubility but also of restricting the contact sites on the protein surface, which is reflected in the molecular packing in crystal growth. The molecular packing is linked to the space group of crystals. Therefore, the relationship between space group and precipitant agent was also investigated for several kinds of proteins, by surveying more than a thousand of crystal structures.

In this work, we performed X-ray structure analysis, by growing single crystals for four kinds of proteins; human carbonic anhydrase II (CAII), horse myoglobin (Mb), hen egg white lysozyme (HEWL), and human serum albumin (HSA). The former two proteins, CAII and Mb, are known to be easily crystallized by AS. On the other hand, the latter ones, HEWL and HSA, are not. The crystal growth of HEWL by AS is so difficult that no crystal structure has been deposited in protein data bank (PDB). Only a photo image of HEWL crystal by AS is available in literature. The crystal growth of HSA by AS is also difficult and then no PDB entry is available and no crystallization has been reported. It should be noted that a large number of crystal structures are available in PDB both for HEWL and HSA, but all of them were obtained by the crystals grown with the precipitants other than AS.

## 2 Experiment

Recombinant CAII was produced by an *E. coli* strain transformed with the expression vector. Mb, HEWL, and HSA were purchased from suppliers who provided non-recombinant proteins extracted from natural products. All of the proteins were purified before crystallization. X-ray diffractions were acquired at synchrotron beamlines. Atom coordinates were determined by molecular replacement method.

In this work, we also performed MD simulations for the four proteins with and without AS ions in the calculation models. Motion of AS ions, representative AS distribution around the proteins, and atom fluctuation with respect to protein residues and AS ions were deduced from the MD trajectories. Electrostatic potentials were drawn from the protein structures. To clarify the reason for readiness and difficulty in crystal growth by AS, these computational results were compared among the proteins. MD simulations were executed by a graphic processing unit (GPU) executable code of pmemd module of AMBER16. Three 200 ns MD runs were performed for every protein in the presence of AS ions and a single 200 ns run was also performed for comparison.

## 3 Results and Discussion

Protein single crystals were grown for CAII, Mb, HEWL, and HSA in the AS concentrations of 2.7, 3.2, 1.2, and 1.2 M, respectively. Details in crystallographic data are shown in Table S1. The max resolutions for the respective proteins were 1.27, 1.56, 1.13, and 2.92 Å. The major precipitant reported in literature for crystallization of CAII and Mb is AS and the space groups are P 1 21 1 for CAII and P 6 or P 1 21 1 for Mb. Therefore, the crystals obtained in this work reproduced the previous studies. The space group of HEWL crystal grown by AS was P 1, which was different from the most popular space group of HEWL crystals, P 43 21 2. The max resolution for HEWL, 1.13 Å, was better than the average resolution of the crystal structures deposited in PDB, 1.73 Å. The space group of HSA by AS was C 1 2 1 and the max resolution was 2.92 Å. The major space group of the HSA crystals deposited in PDB is C 1 2 1 and the average resolution is 2.60 Å. In terms of resolution, the quality of

the HSA crystal grown by AS is inferior to the crystals obtained by other precipitants.

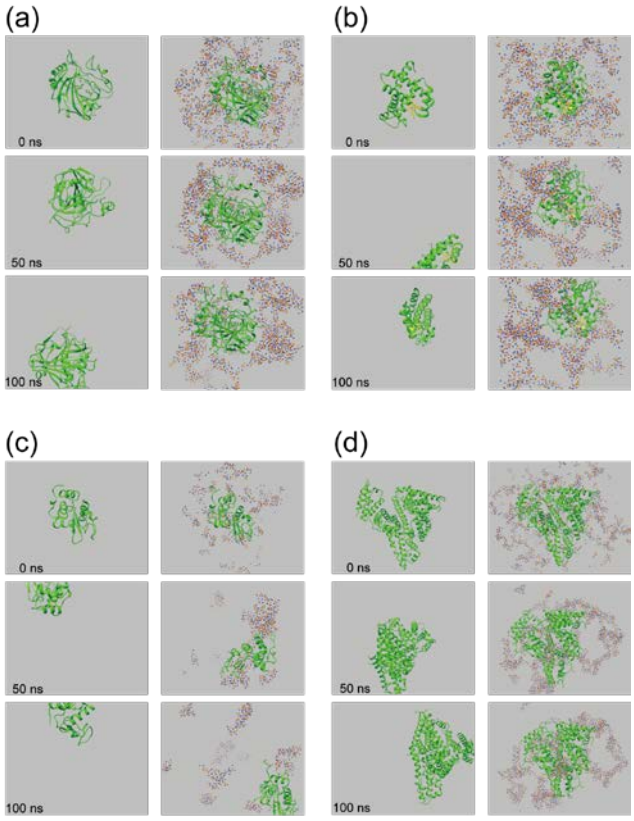


Fig. 1: Snapshot structures extracted at 0, 50, and 100 ns from the simulation trajectories for (a) CAII, (b) Mb, (c) HEWL, and (d) HSA. The right and left columns represent the simulations with and without AS precipitants, respectively. The simulations were executed in triplicate in the presence of AS and the snapshots shown are of the first one for every protein. Protein molecules are represented by green cartoons. N, S, and O atoms of AS ions are coloured blue, yellow, and red, respectively. Hem molecule in Mb is depicted in a yellow stick representation. Sodium and chloride ions and water molecules are not shown for clarity.

Calculation models for CAII, Mb, HEWL, and HSA were built from the obtained crystal structures. MD simulations were performed for 200 ns in the presence of AS and also for 200 ns in the absence of AS. The simulations with AS were executed in triplicate for reproducibility. The concentrations of AS ions in the calculation models were set to be equal to the initial droplet in the experimental setup for protein crystallization and other minor ingredients were omitted from the models.

The motions of protein and precipitant molecules were monitored all through the 200 ns simulation. The snapshot structures at 0, 50, 100 ns were extracted both for the models with and without AS (Fig. 1). In every simulation without AS, protein molecule largely moved in the solvent and the direction of the principal molecular axis was ceaselessly changed. Since the molecular weights of Mb and HEWL are small compared to those of CAII and HSA, the displacement of center of mass was considerably large for Mb and HEWL. Due to the presence of AS ions, the motion of protein molecule was markedly reduced for every protein. In CAII (Fig. 1a), not only the center of mass but also the direction of molecular axis was rarely altered. The initial setup of AS positions, in which AS ions were randomly placed in solvent, was reflected in the AS distribution at 0 ns. At 50 ns, AS ions were relocated not to cover all over the protein surface but to make local contacts with CAII. The AS distribution is not uniform and instead chain-like arrays were formed. The chain-like distribution became clear at 100 ns. The motions of AS precipitants were unexpectedly small. In Mb (Fig. 1b), AS distribution at 0 ns was random. AS ions were arranged to form chain-like distribution at 50 ns and the AS dense areas were connected by the chains. The chain-like distribution was stable because AS positions were scarcely changed at 100 ns. In HEWL (Fig. 1c), while AS ions were randomly distributed at 0 ns, AS ions were strongly localized at a few areas at 50 ns. The AS localization is, however, instable because the distribution was rearranged at 100 ns. In HSA (Fig. 1d), the AS initial random distribution at 0 ns was rearranged at 50 ns. The chain-like AS formation was stable and kept at 100 ns. A comparison of aminations among CAII, Mb, and HSA suggested that most of the AS ions moderately moved in CAII and Mb but a considerable number of AS ions largely fluctuated in HSA.

To quantify the motions of AS ions in simulation, root mean square fluctuation (RMSF) were calculated for the protein residues and AS ions (Fig. 2). In CAII, RMSFs of most of the protein residues were below 20 $Å^2$ while three peaks over 30 $Å^2$ were observed in the simulation without AS. The three large peaks were decreased due to the presence of AS. RMSFs of AS ions were large compared with the protein residues and the average RMSF was 38.7 $Å^2$. The number of the peaks beyond 150$Å^2$ was zero. In Mb, most of the protein residues showed RMSFs below 10 $Å^2$. The average RMSF of AS ions was 49.9 $Å^2$ and none of the peaks was beyond 150$Å^2$. In contrast, there were many peaks over 150$Å^2$ in HEWL and HSA. These RMSFs reflect the situations that the restraint on AS ions are weak in HEWL and HSA. RMSFs in Fig. 2 are the results of the first simulation of the three calculation trials for each protein. In CAII, RMSFs of the AS ions in the second simulation were considerably low and the average is below 25 $Å^2$. In Mb, RMSFs in the second simulation were quite low. In contrast, RMSFs in the second and third simulations were quite large for HEWL. RMSFs in the second and third simulations in HSA were also large. These results clearly suggest that the mobility of AS ion is small in CAII and Mb but that is large in HEWL and HSA. In HSA, the decrease of RMSFs of the protein residues due to the presence of AS was remarkable.

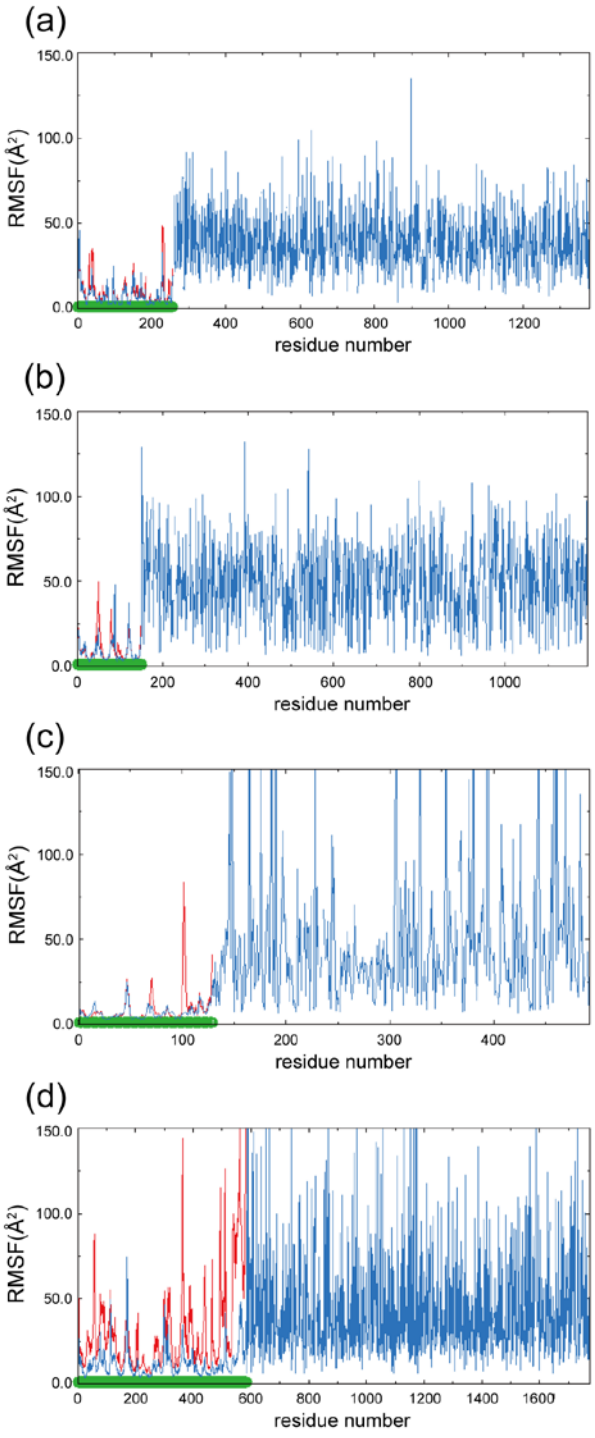


Fig. 2: RMDFs with respect to the protein residues and the AS ions in the simulations for (a) CAII, (b) Mb, (c) HEWL, and (d) HSA. The simulations with and without AS are represented by blue and red lines, respectively. The green bar on the x-axis corresponds to the protein residues. RMSF with AS in every protein is of the first one of the three simulations in the presence of AS.

In order to explain the chain-like distribution of AS ions, the iso-surfaces of electrostatic potential around the proteins were depicted (Fig. 3). In CAII and Mb, the electrostatic potential was clearly separated into the positive and negative areas almost in half and both areas had hemispherical shapes. AS chains are positioned to connect the positive and negative areas. Hence, AS ions are arranged along the electric field lines that run perpendicular to the iso-surfaces of electrostatic potential. In contrast, the polarity of the electrostatic potential is seriously in partial in HEWL and HSA and there is a marked difference between the positive and negative areas in volume. In HEWL, the whole protein is completely covered with the positive potential surface and then the chain-like arrays of AS ions are unlikely to be formed. In HSA, the electrostatic potential is greatly biased to negative and then many of the chain arrays are entirely within the negative potential area and only a short array manages to connect the positive and negative areas. Consequently, the distribution of AS ions are stabilized in the situation that the electrostatic potential around the protein is equally separated into the positive and negative areas and the AS chain arrays are likely to be formed to connect the two areas as seen in CAII and Mb.

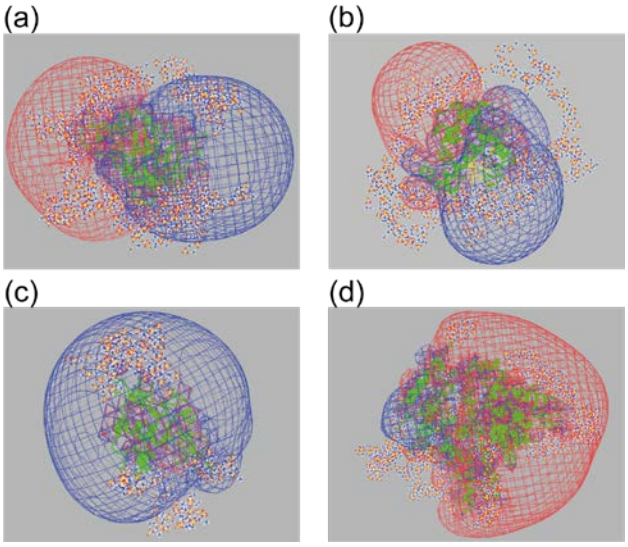


Fig. 3: Positive and negative iso-surfaces of electrostatic potential around the proteins for (a) CAII, (b) Mb, (c) HEWL, and (d) HSA. The positive and negative iso-surfaces are shown in the blue and red mesh representation. The contouring values of iso-surfaces are (a) +0.43 and -0.14 for CAII, (b) +0.36 and -0.13 for Mb, (c) +1.75 and -0.004 for HEWL, and (d) +0.43 and -1.96 for HSA, respectively. Colorings for molecules are the same as Figure 1. Every electrostatic potential was obtained by the presentative structure of the first one of the three simulations in the presence of AS.

It will be expected from the findings of this study that an introduction of amino mutations can enhance the

crystallization even for a protein whose experimental condition for the crystal growth with AS is not well-established. The optimal zone of precipitant concentration is narrow for crystallization with AS. The electrostatic potential will be utilized as a guide to modify recombinant proteins adequate for crystal growth with AS. For example, when a protein crystal is obtained with precipitant other than AS but the resolution of X-ray diffraction is not satisfactorily high, a better quality of protein crystal is sometimes required. Modeling of the protein structure from the low-resolution X-ray diffraction, modification of the model by introducing amino mutations, and calculation of the electrostatic potential for the modified model will provide a good suggestion on the choice of amino mutations to promote the crystal growth with AS. The electrostatic potential significantly depends on the pH of solution. Multiple calculations of the electrostatic potentials for the models changing the protonation states of the titratable residues will provide the information on the appropriate pH condition to promote the protein crystallization by AS. Therefore, the analysis of the electrostatic potential enables us to take the rational strategy in experiments.

In summary, protein crystals were grown for CAII and Mb, which are easily crystallized by AS, and also for HEWL and HSA, which are not. X-ray diffractions were acquired from the respective protein crystals and structures were determined by the molecular replacement. Based on the determined protein structure, 200 ns MD simulations were performed in triplicate with the model including AS ions in a concentration same as the experimental condition in crystal growth for every protein. A 200 ns simulation was also performed without AS ions for comparison. For all the four proteins, the presence of AS ions greatly reduced the motion of protein during simulation. RMSF analysis indicated that the motions of AS ions were also restricted especially for CAII and Mb. This means that the AS ions are likely to be stabilized around the proteins that are easily crystallized by AS. AS ions are not randomly distributed and instead they are arranged in the chain-like formation. The chain-like formation is caused by the electrostatic potential around the protein, in which AS ions are aligned along the electric field lines. In CAII and Mb, the positive and negative areas of the electrostatic potential are almost equally separated and both areas have hemispherical shape. Hence, a large electric field line connecting both areas can be generated. In contrast, one of the positive and negative areas is enlarged in HEWL and HSA. Hence, long electric field line is difficult to be generated. That is, the shape of the iso-surface of the electrostatic potential is strongly responsible for the readiness in crystal growth with AS. If the shape of the electrostatic potential can be controlled by adjusting pH condition or introducing amino mutations, the quality of crystallization by AS will be improved even for the proteins which are not easily crystallized by AS.

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. 2016G554).

References

[1] S. Fudo *et al*., *Cryst. Growth Des.*. **17**, 534 (2017).

* hoshino@chiba-u.jp