The "Molecular Grabber" method: Development of new crystalline sponge method New Idea for structural analysis of compounds using protein

Takashi Matsumoto*, Akihito Yamano*, Ryosuke Nakashima** and Kunihiko Nishino**

1. Introduction

Molecular structure determination is very useful for the development of medicines, aroma chemicals and agrochemicals. Single crystal X-ray diffraction (SC-XRD) analysis is the most powerful technique for molecular structure determination. However, SC-XRD analysis requires good quality crystals. In fact, the biggest hurdle for SC-XRD analysis is crystallization. Crystallization trials require a large amount of high-purity target compounds. Moreover, despite performing tedious and time-consuming trials, sometimes good quality crystals for SC-XRD analysis may not be obtained. In this case, we have to give up on structure determination. As one way to address this situation, Fujita *et al*. have reported the crystalline sponge method (CS method) for the structure determination of small molecules [(1)].

With this method, crystallization of the target compound is not required. The CS method uses a metal-organic framework (MOF). The target compounds are incorporated into the CS crystal by soaking and are oriented in the porous coordination network of the MOF. Then, the structure can be determined by SC-XRD analysis. As a result, the CS method allows the SC-XRD analysis of many compounds that cannot be crystallized. However, as with other analysis techniques, the CS method has some limitations. The method is not applicable to all types of compounds.

The CS method uses the MOF as the "container" for the compounds. Looking at CS method figures, we came up with a new idea—the container does not have to be an MOF. We wanted to prepare a container with the ability to bind to a wide variety of compounds; therefore, we focused on proteins because they can bind to organic compounds. Some proteins can bind to a wide variety of compounds, such as anionic, cationic and neutral compounds. Therefore, we started to develop a new crystalline sponge method that is different from the existing method [(2)].

2. Selection of Protein as "Container"

We thought that using proteins as the "container" had four requirements: the ability to bind to a wide variety of compounds, stability, a simple purification

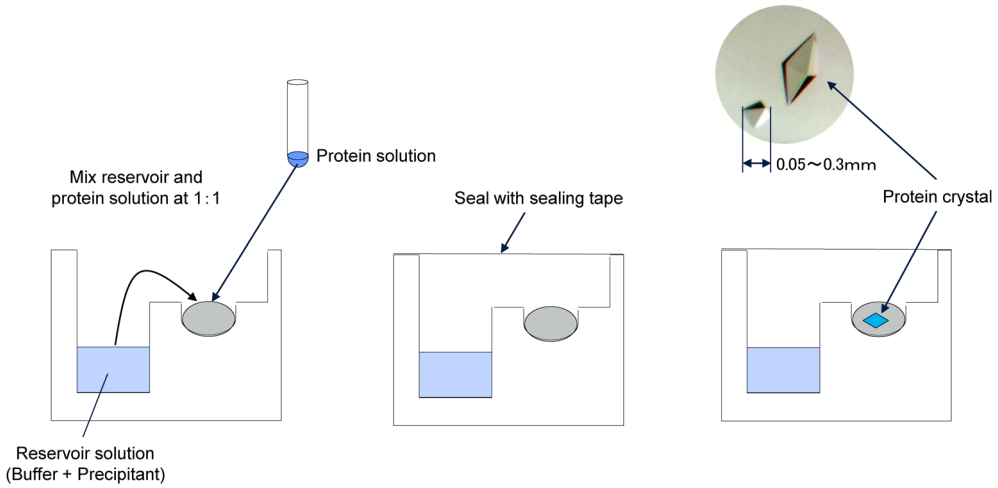

Fig. 1. Sitting drop vapor diffusion method.

*Application Laboratories, Rigaku Corporation
**Institute of Scientific and Industry Research, Osaka University

procedure, and ease in obtaining good quality crystals. Therefore, as a result of searching for a protein that fulfills these requirements, we thought the transcriptional repressor RamR was suitable for the container [2]. RamR controls the expression of the multidrug efflux transporter AcrB and can bind to a wide variety of compounds. Moreover, RamR is a stable protein and is easily purified and crystallized. For these reasons, we decided to explore the possibility of using RamR as the container to determine the structures of unknown compounds instead of using MOFs.

3. Experimental Methods

3.1. Crystallization of RamR-compound complex

In the CS method, the compounds are incorporated by the soaking method but, in this method, the co-crystallization method was used. The compound was added to the RamR solution and the mixture was incubated overnight. The crystallization was carried out by the sitting drop vapor diffusion method at 25°C (Fig. 1) and, after 1 week, block-like crystals appeared. Two crystallization conditions with good reproducibility were used. Using only these two conditions, it is possible to minimize the amount of target compound consumed. The crystallization conditions are described below.

I. 10%–15% (w/v) polyethylene glycol (PEG) 3350,
0.2 M ammonium sulfate,
0.1 M sodium citrate pH 5.6

II. 15%–20% (w/v) PEG3350,
0.2 M sodium acetate,
0.1 M 2-morpholinoethanesulfonic acid monohydrate (MES)-Na pH 6.5

3.2. Structure determination

All diffraction data were collected on a XtaLAB Synergy Custom at 100 K. The diffraction data were processed using CrysAlis^{Pro} [3]. The structure of each complex was solved by molecular replacement, which is an easy phase determination method.

In general, synchrotron radiation is considered to be essential for X-ray diffraction data collection of protein crystals, and laboratory X-ray systems are used for preliminary measurements for selecting crystals to send to the synchrotron. However, modern technologies—such as X-ray generators, detectors, goniometers, and software—have dramatically evolved, and a state-of-the-art system such as the XtaLAB Synergy enables you to collect high-quality data on protein crystals.

4. Results

4.1. Structure of RamR

RamR consists of nine α-helices (α1–α9), and has 2 domain structures: the *N*-terminal DNA binding domain and the *C*-terminal compound binding domain. There is a binding pocket in the center of this compound binding domain (Fig. 2).

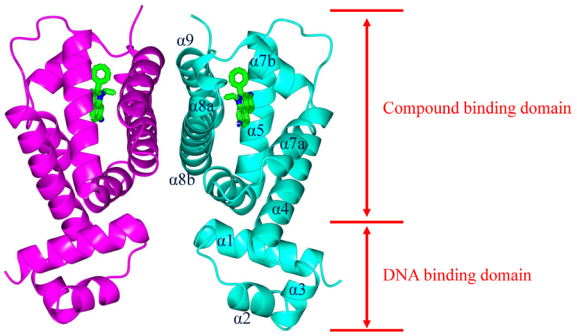

Fig. 2. Crystal structure of RamR. RamR consist of two domains, an *N*-terminal DNA binding domain and a *C*-terminal compound binding domain.

4.2. Structure of RamR-ethidium complex

First, structural analysis of the complex with ethidium, which is famous as a substrate for RamR, was performed. Ethidium is used as a nucleic acid stain and is familiar in the field of molecular biology. As a result of structure analysis, the electron density map of the bound compound clearly appeared (Fig. 3). From the shape of the density map, ethidium is easily identified. RamR binds one ethidium molecule per monomer (Fig. 4). The ethidium molecule binds to amino acids in the binding pocket via hydrogen bonds and π–π stacking interactions.

4.3. Structure of RamR-substrate complexes (rhodamin 6G, crystal violet, berberine)

Figure 5 shows the structures of three RamR-substrate complexes (rhodamine 6G: PDB code 3VVZ, crystal violet: PDB code 3VW1, and berberine: PDB code 3VW2) solved by Yamasaki *et al.* [4]. These compounds were bound by hydrogen bonds and π–π stacking interactions in the same way as ethidium. Each compound molecule interacted with different sets of amino acid residues in the binding pocket. This binding mechanism is called multisite binding. It is expected that RamR can bind to a wide variety of compounds by this multisite binding mechanism.

4.4. Structure of RamR-cholic acid complex

We investigated whether RamR can bind large, rigid cholic acid ($C_{24}H_{40}O_5$), which has a size and shape that would not be accommodated in the binding pocket. Cholic acid has a steroid skeleton and is a component of digestive juices to help digestion and absorption of lipids. The structure of the RamR–cholic acid complex was determined at 1.55 Å resolution. As a result of structure analysis, a fine electron density map corresponding to cholic acid was obtained (Fig. 6). RamR bound the cholic acid by different amino acid residues compared with other compounds. This complex used hydrogen bonds and hydrophobic interactions without π–π stacking interactions. Surprisingly, the structure of the binding pocket of the RamR–cholic acid complex was significantly different from that of

Camera length (mm)	40
Exposure time (sec.)	**25**
Resolution (Å)	Inf. – **1.70** (1.76 – 1.70)
*R*int (%)	**7.2** (35.0)
Completeness (%)	98.7 (95.8)
R-factor / *R*-free (%)	16.85 / 23.48

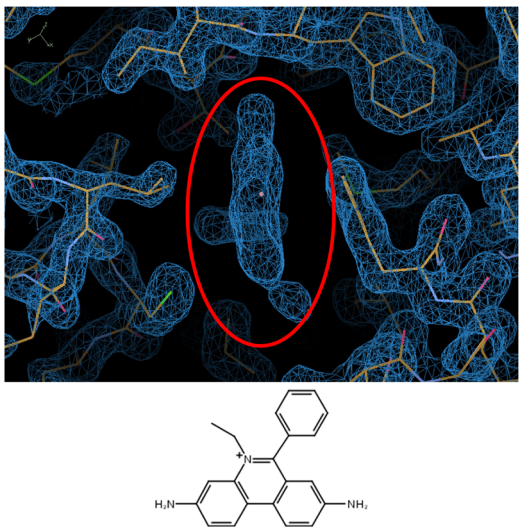

Fig. 3. Electron density map of RamR–ethidium complex. As a result of structural analysis at 1.70 Å, a clear electron density corresponding to ethidium appeared.

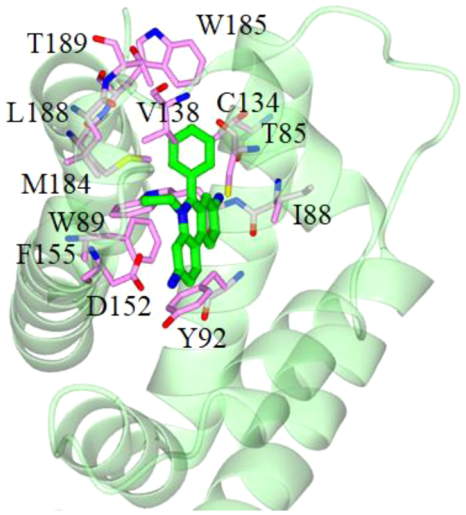

Fig. 4. Structure of the RamR–ethidium complex. The key residues for ethidium binding are shown using a pink representation.

the RamR–ethidium complex. The α7b and α8a helices of the RamR–cholic acid complex were completely uncoiled. Moreover, these uncoiled regions shifted outward, resulting in expansion of the binding pocket volume to accommodate the bulky, rigid cholic acid molecule (Fig. 7). As a result, it was revealed that the voluminous binding pocket of RamR is flexible and can increase its volume to accommodate a large molecule. This flexibility is also a feature of RamR.

4.5. Structure of RamR-dequalinium complex

Figure 8 shows the crystal structures of the RamR–dequalinium complex (PDB code 3VW0) solved by Yamasaki *et al.*[4]. Dequalinium has a long, flexible linear alkane (10 carbon atoms) chain. This compound is too long to be accommodated in the binding pocket. The α6, α7b, α8a, α8b, and α9 helices can shift outward to expand and reshape the binding pocket to accommodate a long flexible compound. It is concluded that RamR can flexibly change the shape of the binding pocket

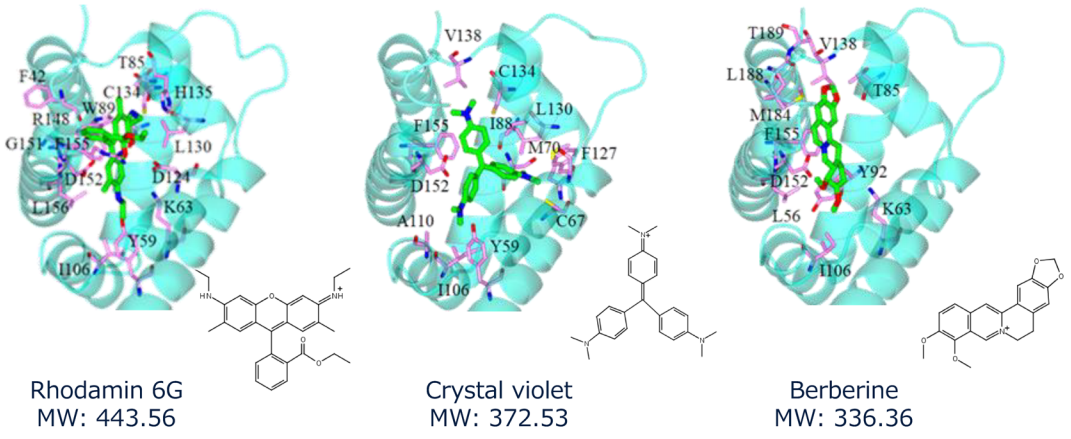

Fig. 5. Structure of the RamR–substrate complexes (rhodamine 6G, crystal violet, berberine). The key residues for ethidium binding are shown using a pink representation.

Camera length (mm)	40
Exposure time (sec.)	**29**
Resolution (Å)	Inf. – **1.55** (1.61 – 1.55)
*R*int (%)	**3.1** (47.1)
Completeness (%)	99.5 (97.6)
R-factor / *R*-free (%)	14.61 / 20.38

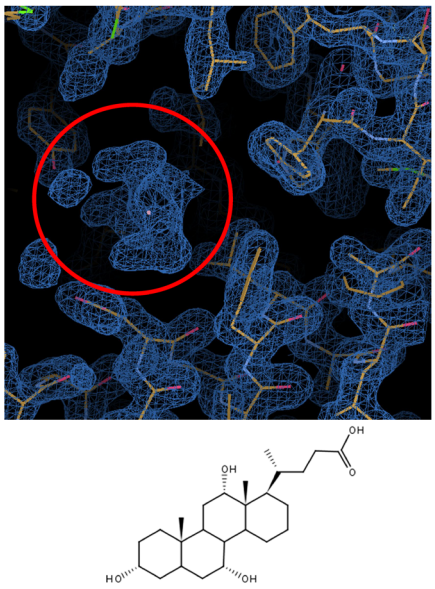

Fig. 6. Electron density map of RamR–cholic acid complex. As a result of structural analysis at 1.55 Å, a clear electron density corresponding to cholic acid appeared.

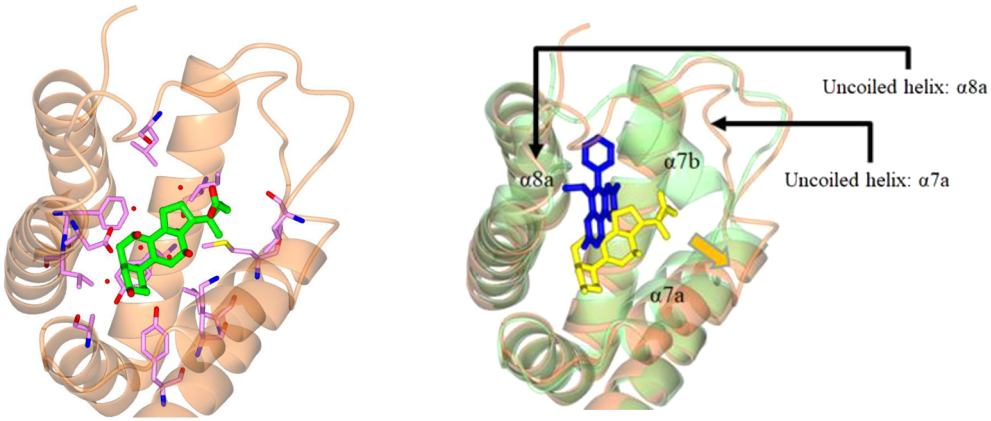

Fig. 7. Structure of the RamR–cholic acid complex. The key residues for cholic acid binding are shown using a pink representation. The *α*7b and *α*8a helices were completely uncoiled and shifted outward to expand the binding pocket.

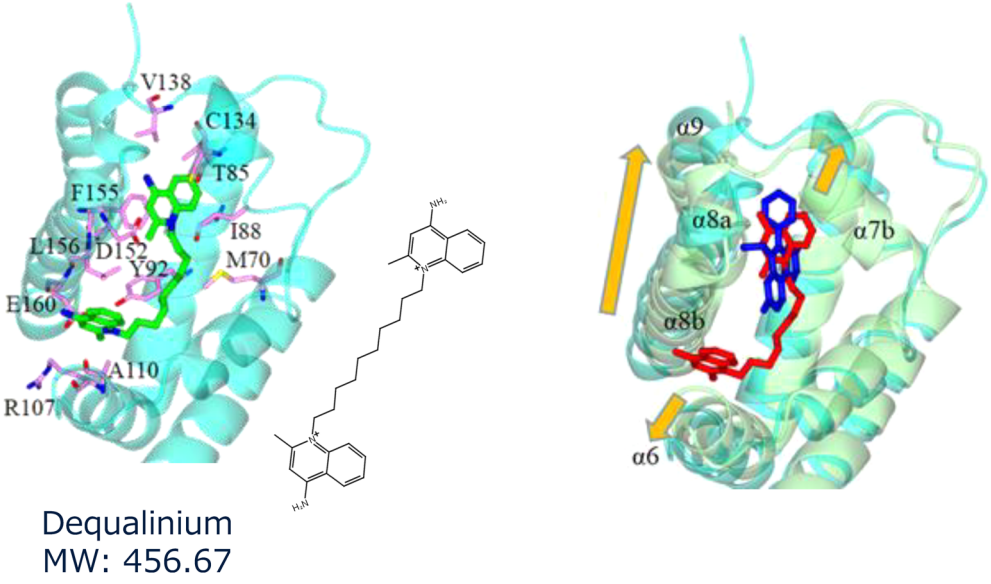

Fig. 8. Structure of the RamR–dequalinium complex. The key residues for dequalinium binding are shown using a pink representation. The *α*6, *α*7b, *α*8a, *α*8b, *α*9 helices shifted outward to expand and elongate the binding pocket.

Camera length (mm)	40
Exposure time (sec.)	**45**
Resolution (Å)	Inf. – **2.20** (2.28 – 2.20)
*R*int (%)	**9.7** (46.1)
Completeness (%)	99.3 (99.0)
R-factor / *R*-free (%)	21.96 / 29.92

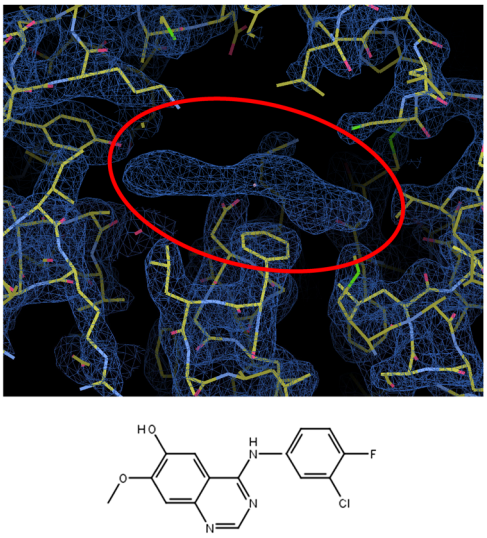

Fig. 9. Electron density map of RamR–gefitinib intermediate 1 complex. As a result of structural analysis at 2.20 Å, a good electron density corresponding to gefitinib intermediate 1 appeared. From density map shape, it was revealed that gefitinib intermediate 1 had a flat structure.

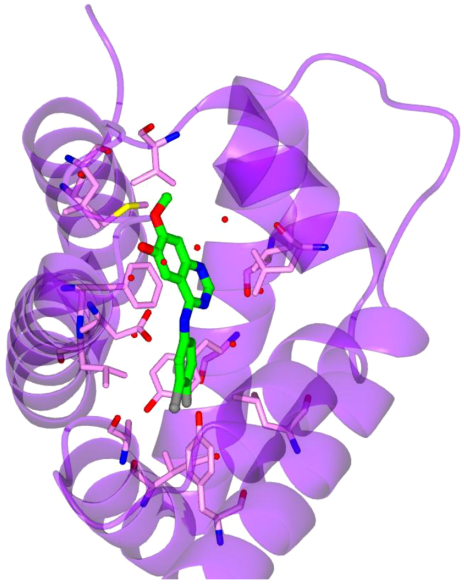

Fig. 10. Structure of the RamR–gefitinib intermediate 1 complex. The key residues for gefitinib intermediate 1 binding are shown using a pink representation.

and optimize it to the shape of the guest compounds by rearranging the position of the helices.

4.6. Structure of RamR-structure unknown compound complex (gefitinib intermediate 1)

The structures of various known compounds have been revealed using RamR so far. Finally, we investigated whether this method could be used to determine the structures of unknown compounds that are not RamR substrates. Gefitinib intermediate 1 is a synthetic intermediate of gefitinib, a well-known anti-cancer medicine. As a result of the structure analysis, a good electron density map appeared and, from the electron density, it was revealed that gefitinib intermediate 1 has a flat structure (Fig. 9). It was bound by hydrogen bonds and π–π stacking interactions in the same way as ethidium (Fig. 10). Consequently, it was proved that this method can reveal the structure of a compound that is not a substrate.

Surprisingly, this method succeeded in capturing the structure of this compound despite 10 times lower concentration (0.24 mM) than that used for ethidium or cholic acid (2.4 mM). It was suggested that RamR–organic compound complexes crystallize preferentially over the apo RamR under these crystallization conditions. This result indicates that this new structural analysis method using RamR is effective even for a limited amount of target compounds.

5. Conclusion

It was revealed that RamR flexibly changes the volume and shape of its binding pocket according to the target compound, and then grabs the organic compound. As shown in Fig. 11, the structure of the grabbed organic compounds can be revealed by a simple molecular replacement method. We named this method the “Molecular Grabber” method because it utilizes the characteristic property that RamR has the ability to grab molecules.

It is possible that the Molecular Grabber method can be used for the X-ray structure analysis of difficult-to-crystallize organic compounds, as with the crystalline sponge method. In addition, this method has the potential to be used to identify the functional groups of target organic compounds that interact with amino acid residues of proteins by analyzing the interactions between the compounds and RamR. Furthermore, based on the result with gefitinib intermediate 1, the Molecular Grabber method can be used for structural

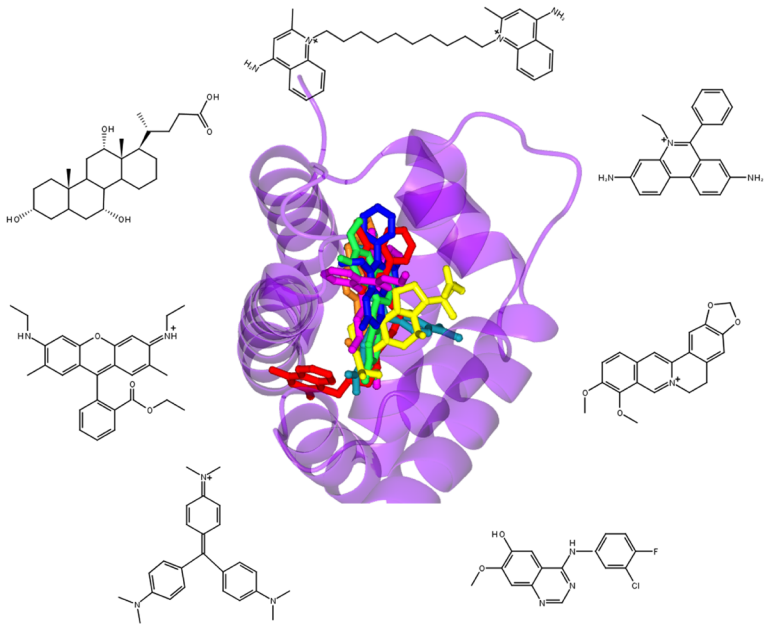

Fig. 11. Superimposition of the various compounds bound to RamR. The compounds ethidium, rhodamine 6G, crystal violet, berberine, cholic acid, dequalinium, and gefitinib Intermediate 1 are shown in blue, magenta, dark cyan, orange, yellow, red, and green, respectively.

analysis of trace amounts of intermediates in tablets, capsules, or powdered medicines, aroma chemicals and agrochemicals in many cases.

We hope that the Molecular Grabber method can contribute to the development of innovative new medicines, aroma chemicals, agrochemicals and the identification of residual intermediates.

This method does not require any license to use, so it is also a significant feature that any interested researcher can try it.

References

(1) Y. Inokuma, S. Yoshioka, J. Ariyoshi, T. Arai, Y. Hitora, K. Takada, S. Matsunaga, K. Rissanen and M.Fujita: *Nature*, **495** (2013), 461–466.

(2) T. Matsumoto, R. Nakashima, A. Yamano, K. Nishino: *Biochem. Biophys. Res. Commun.,* **518** (2019), 402–408.

(3) T. Matsumoto, A. Yamano, T. Sato, J. D. Ferrara, F. J. White, M. Meyer, *J. Chem. Crystarogr.*, https://doi.org/10.1007/s10870-020-00867-w

(4) S. Yamasaki, E. Nikaido, R. Nakashima, K. Sakurai, D. Fujiwara, I. Fujii, K. Nishino: *Nat. Commun*., **4** (2013), 1–7.