Crystal Structure Determination of Ubiquitin by Fusion to a Protein That Forms a Highly Porous Crystal Lattice

Nobuo Maita*.

Laboratory of X-ray Crystallography, Institute of Advanced Medical Sciences, Tokushima University, 3-18-15, Kuramotocho, Tokushima, 770-8503, Japan.

Supporting Information Placeholder

ABSTRACT: The protein crystallization process requires screening a large number of conditions using a large quantity of high-purity protein, which makes crystal structure analysis difficult. Thus, the development of easy and versatile protein crystallization techniques is both extremely desirable and highly challenging. Here, I demonstrate the crystallization and structure determination of ubiquitin by genetic fusion to the highly porous honeycomb lattice of R1EN. I successfully crystallized and collected X-ray data from three R1EN-ubiquitin constructs with varied linker lengths under the same conditions as the original R1EN. The crystals diffracted to 1.7-2.4 Å resolution, and the ubiquitin structures were determined with results almost identical to the previously published structure. Moreover, the ubiquitin structure could be solved by molecular replacement using R1EN alone. This method may reduce the effort required for crystallization screening and is applicable to de novo protein structure determination.

The crystal structure of a protein reveals its functions, architecture, and interactions with binding substrates or ligands, thereby revealing information on fundamental biology or the causes of disease; however, producing high-quality crystals is extremely difficult[1,2]. Thus, the development of easy and versatile protein crystallization techniques is both extremely desirable and highly challenging. A major practical technique to facilitate protein crystallization is to use antibodies[3], small compounds[4,5], or fusion tags[6-8] as a glue of the lattice. However, these methods cannot control the crystallization conditions, so that crystallization screenings are indispensable. Hence, crystallization without crystallization screening has been proposed. In 1982, Seeman suggested a strategy for protein crystallization in which a crystal lattice made of DNA origami is used as a scaffold and a DNA-binding protein is arranged therein, leading to the solution of the crystal structure[9], however, no applications of this method have been reported until now. More recently, Fujita and colleagues reported a smart strategy for crystal structure determination, called the crystalline sponge method, that uses a self-assembled mesocrystalline metal-organic framework (MOF) to organize a small molecule inside the lattice and consequently obtains the X-ray structure[10,11]. However, this method cannot be applied to proteins because of the small pore size of MOFs. Therefore, the development of mesoporous MOFs which have large pore sizes have been carried out[12]. In addition, Fujita's group succeeded in encapsulating ubiquitin inside an MOF cage and obtained the x-ray data[13], however, the ubiquitin was randomly arranged in the crystal so that the structure of ubiquitin could not obtained. Another promising crystallization technique is to use *in silico* designed proteins[14-20]. This method is also applicable to structure determination by CryoEM[21]. Although there are several reports of the crystal structure of *in silico* designed protein cages[15-17] or a lattice[18], no structure of guest proteins has been reported.

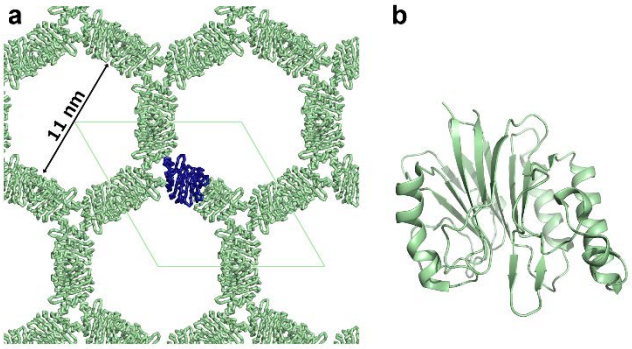


Figure 1. (a) The crystal lattice of R1EN viewed along the c-axis. A monomer is highlighted in blue. Rhombi represent the unit cell of the R1EN crystal. The inner diameter of the honeycomb is approximately 11 nm. (b) Crystal structure of R1EN (PDB:2EI9).

The use of a protein that assembles into a highly porous crystal lattice as a building block instead of MOFs provides a simple and easy way to incorporate the protein of interest into the crystalline framework by genetic fusion, and 100% occupancy can be achieved. Several examples of proteins that naturally form a highly porous (>70% of solvent volume) crystal lattice are found in the Protein Data Bank[22] and these lattices are capable of enabling the determination of the crystal structure of a fusion guest protein. Here, I focused on R1EN, an endonuclease domain that is encoded at the ORF2 N-terminus of the non-LTR retrotransposon found in silkworms, which consists of 243 residues and has a molecular weight of 23 kDa[23,24]. R1EN forms a hexagonal rod crystal with unit cell dimensions of $a=b=141.3$, $c=37.5$ Å in the space group $P321$ and assembles into a hollow honeycomb lattice with an inner diameter of ~110 Å (Figure 1)[24]. In addition, the N- and C-termini are exposed in the pores; hence, a fusion guest is expected to be positioned in the solvent without disrupting the crystal lattice (Figure S1).

Human ubiquitin was selected as a fusion guest protein to test this method. As the C-terminal region of the R1EN structure was disordered[24], a series of R1EN-Ubiquitin fusion proteins with different linker lengths were generated based on the C-terminal

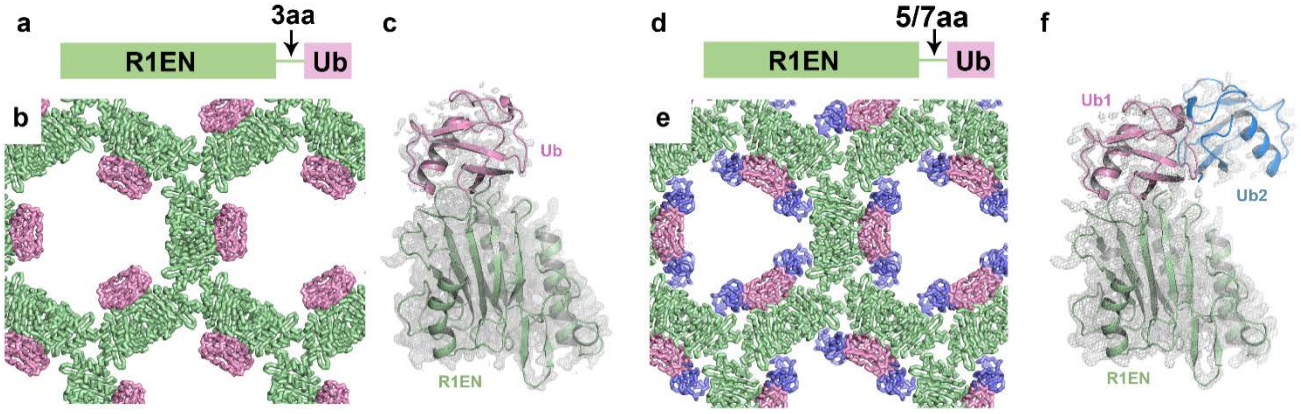


Figure 2. (a, d) Diagrams of the fusion proteins R1EN223-Ub (a) and R1EN225-Ub/R1EN227-Ub (d). (b, e) Crystal lattices of R1EN223-Ub (b) and R1EN225-Ub/R1EN227-Ub (e) viewed along the c-axis. (c, f) Crystal structures of R1EN223-Ub and R1EN225-Ub/R1EN227-Ub. The 2mFo-DFc electron densities (> 1.0σ) are overlaid. In R1EN225-Ub/R1EN227-Ub, two conformers of ubiquitin were observed (Ub1 and Ub2).

deletion R1EN constructs with sequences extending to Gly223 for R1EN223-Ub, Thr225 for R1EN225-Ub Glu227 for R1EN227-Ub and Asp231 for R1EN231-Ub (Figure S2). All constructs were purified and crystallized with microseeding method. The X-ray diffraction data set of each construct was collected using synchrotrons and applied to structural refinement (Table S1). In R1EN231-Ub, which has an 11-residue linker, no electron density for Ub was observed, presumably due to the flexible linker. In contrast, in R1EN223-Ub, R1EN225-Ub and R1EN227-Ub, electron density corresponding to ubiquitin was observed in the central cavity (Figure S3a,c). As the electron density was not clear enough to build the model in a straightforward manner, I performed molecular replacement using ubiquitin (PDB: 1UBQ; ref. 25) and R1EN together as the search models. In the R1EN223-Ub data, I could obtain the correct solution of ubiquitin. The electron density was much weaker than that of the R1EN moiety but was clearly defined (Figure 2c, S3b). As expected, the ubiquitin molecule lies on the inside wall and interacts weakly with the N-terminus of R1EN (Figure 2b). There are two contact interfaces between ubiquitin molecules in different symmetric units (Figure 3a): one appears to represent a steric collision between the symmetry-related ubiquitin molecules at helix α1 (contact A), while the other is a natural contact between the two molecules (contact B); a contact similar to contact B is also observed in the crystal packing of ubiquitin in PDB entry 5DK8 (ref. 26; Figure S4). The steric collision in contact A was not eliminated even when the space group was converted to *P*1. In the refined R1EN223-Ub structure, the average B factors of R1EN and ubiquitin are 28.6 Å^2 and 80.3 Å^2, respectively. These results suggest that not all ubiquitin molecules are fixed in the crystal, and thus, the occupancy of ubiquitin is low.

In R1EN225-Ub and R1EN227-Ub, I observed another ubiquitin (Figure 2e,f, S3c) that was considered to represent an alternate conformer. For clarity, the ubiquitin in the position observed in R1EN223-Ub is referred to as Ub1, and the ubiquitin in the position newly observed in R1EN225-Ub and R1EN227-Ub is referred to as Ub2 (Figure 2f). Analysis of the average B factor and electron density suggested that Ub2 is less stable than Ub1 (Table S1, Figure 2f). The distance between Gly223 in R1EN and Met1 in Ub2 is 9.4 Å, which is too far to allow ubiquitin to occupy the Ub2 position in R1EN223-Ub. Ub2 has no definite interaction with cis-R1EN; however, the Gln31-Gly35 region of Ub2 exhibits water-mediated interactions with a symmetry-related R1EN, and Lys6 of Ub2 interacts with Asn60 and Gln62 in a symmetry-related Ub1 (Figure S5).

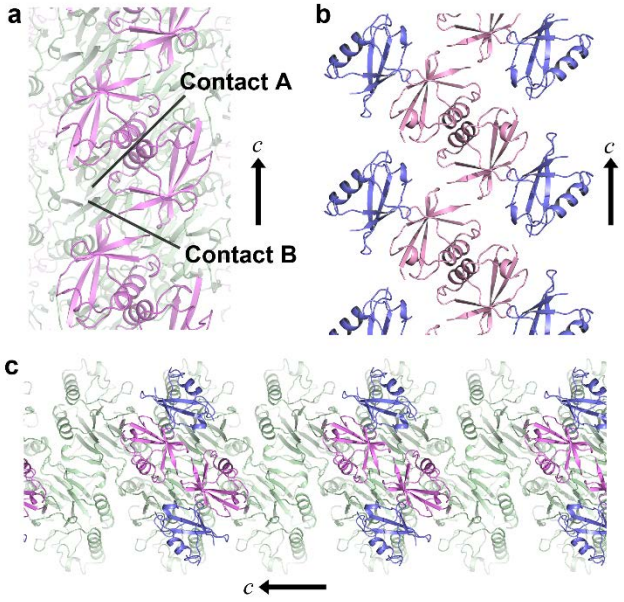


Figure 3. (a) Crystal packing of ubiquitin in R1EN223-Ub viewed from the central pore. The ubiquitin molecule has two contact interfaces between symmetry-related ubiquitin molecules. (b) Crystal packing of R1EN225-Ub/R1EN227-Ub. In R1EN225-Ub/R1EN227-Ub, two ubiquitin conformers were observed. (c) Predicted ubiquitin packing of R1EN225-Ub/R1EN227-Ub. The R1EN-Ub tetramer composes one unit. The direction of the c-axis is indicated in each figure.

In summary, the Ub1-Ub1 contact at α1 (contact A) is in collision, while the other Ub1-Ub1 contact (contact B) and the Ub2-Ub1 interface are consistent with other Ub crystal structures. These observations imply that in the crystal lattice, four R1EN-Ub moieties compose one unit: two Ub1s are in contact B relationships, and the other two are Ub2s (Figure 3c). The conformation of the four R1EN-Ubs completes the *c*-axis arrangement and composes one side of the hexagon; however, each side of the hexagon is assembled independently during the assembly of the honeycomb structure, resulting in mixed electron density and deterioration at α1 of Ub1. Similarly, in R1EN223-Ub, two Ub1s may be in a contact B relationship, while the other two ubiquitins are disordered. The refined ubiquitin structures in all fusion pro-

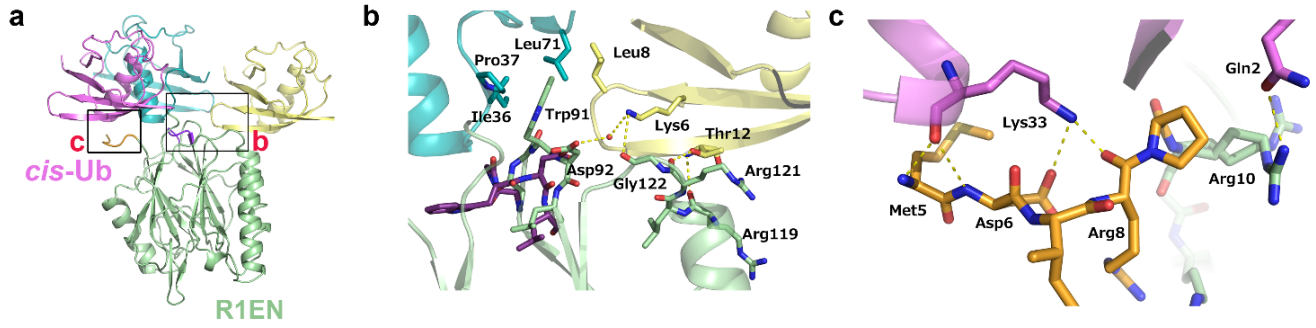


Figure 4. The details of the ubiquitin interactions in R1EN223-Ub. (a) Protomers of R1EN223-Ub are colored in green (R1EN) and pink (ubiquitin). Symmetry-related ubiquitins with contact A (cyan) and contact B (yellow) interactions are also shown. The original and changed conformations of the R1EN 88-93 loop region are colored in purple and green, respectively. Squares represent the positions of panels b and c. (b) R1EN interaction between symmetrical ubiquitins. The 88-93 loop region of R1EN forms multiple conformers; Trp91 is flipped outward and participates in hydrophobic interactions with Leu71, Pro37, Ile36 and Leu8. The original and changed conformations are colored in purple and light green, respectively. (c) Cis-ubiquitin (pink) interaction with the N-terminal region (orange) of R1EN. Lys33 in cis-Ub interacted with Met5-Arg8 by polar contacts.

teins are almost identical to the reported structure (PDB: 1UBQ), with the root-mean-square deviation (RMSD) of the main-chain Cα positions ranging from 0.449-0.531 Å; thus, the structure determined by the R1EN-fusion method is quite reliable (Figure S6).

Next, I asked whether the ubiquitin structure can be solved with only the phases calculated from the molecular replacement of R1EN. Using a refined R1EN model, the phases were improved and the ubiquitin model was automatically built with the R1EN223-Ub data. After three cycles of calculation, the resulting ubiquitin model contained 65 of the 72 residues in the main chain and 36 residues with correctly modeled side chains (Figure S7). Notably, however, Ub2 was not found in the molecular replacement or autobuilding trials, and no ubiquitin models based on the R1EN225-Ub and R1EN227-Ub data could be built.

In R1EN223-Ub, the R1EN moiety had almost the same structure as in the previous model except in the Ile88-Leu93 loop region (Figure 4a,b). Nevertheless, electron density that corresponded to an unchanged conformation was observed at the same time, suggesting that this region has two conformations. The Ile88-Leu93 loop is located at the interface of two neighboring ubiquitins, and the side chain of Trp91 is flipped outward and contributes to the stabilization of the hydrophobic interactions with Ile36, Pro37, Leu71 and Leu8 in ubiquitin (Figure 4b). In addition, electron density was observed for the Met5-Pro9 region, which was disordered in the previous R1EN structure. There are four hydrogen bonds between ubiquitin and Met5-Pro9; presumably, the Met5-Pro9 region interacts with ubiquitin, and the two moieties stabilize each other (Figure 4c).

In this study, I showed that ubiquitin fused to R1EN can assemble into the same crystal lattice under the same crystallization conditions as R1EN and determine the crystal structure without experimental phasing. This method has several advantages over previously reported methods. First, the crystal lattice of R1EN is kept intact, the fusion protein can be positioned in the void space, and the crystallization conditions are the same as those used for R1EN. Second, the R1EN crystal produces high-resolution (~1.7 Å) diffraction data; thus, the fusion protein is expected to produce high-resolution diffraction data. Moreover, the high-resolution data may be able to overcome the low-occupancy issues as well as the de novo phasing difficulties.

Of course, several problems remain be overcome. So far, the success rate of this method is depends on fusion guest proteins. I have tested seven proteins or domains, which have molecular weight ranges between 10 to 23 kDa. Most of them yielded no crystals. Of these trials, crystals of two constructs, including Sumo-R1EN, were obtained but no electron densities of the fusion guest were observed. Even in the R1EN-Ubs, a suitable linker length is important to avoid disorder. In addition, unexpected but novel interactions were observed between ubiquitin and the N-terminus of R1EN, and these interactions seemed to be important for the fixation of ubiquitin. These results indicate that designing the N-terminus of R1EN as a "glue peptide" that interacts with the fusion protein and screening various linker lengths are necessary for this method to work more effectively. As ubiquitin has six protein-binding patches[27], it is likely that the specific molecular orientation is induced in the crystalline framework more efficiently than other proteins. In the R1EN-Ub, three binding patches, Ile36 patch (Leu8, Ile36 and Leu71), TEK box (Lys6, Lys11 and Thr12) and Flexible loop (Leu8, Thr9 and Gly10), were involved in the interactions (Figure 2b), which is a key to the success of structural determination of the fused moiety. On the other hand, a strong self-association of fusion guest may interfere with the lattice assembly[21]. Therefore, it is supposed that the strength of the interaction between the ubiquitin and R1EN or other ubiquitin molecules is sufficient to fix their orientations, but at the same time, not to interfere with the crystal lattice assembly of R1EN.

The limit of the molecular weight of the fusion protein is another issue. The volume of asymmetric unit is 125,913 Å^3, and considering the average density of proteins in a crystal (2.69 Å^3/Da, ref. 28), the mean molecular weight of the R1EN-fusion protein is up to 46.8 kDa. Thus, upper limit for the fusion guest protein is ~21.9 kDa when using R1EN227. To apply the R1EN-fusion method to larger proteins, I have constructed a stable dimer of R1EN by introducing a cysteine residue on the *C*2 symmetry axis (Figure S8). When an $(R1EN)_2$-fusion chimera is available, the molecular weight limit of the fusion guest protein will increase to 43.8 kDa, which covers almost 50% of all proteins[29].

In conclusion, the effectiveness of protein crystallization and structure determination by fusion to a protein that forms a highly porous lattice is demonstrated.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Experimental procedures, figures and crystallographic data (PDF).

AUTHOR INFORMATION

Corresponding Author

*nmaita@tokushima-u.ac.jp

ORCID

Nobuo Maita: 0000-0003-2260-5783

Notes

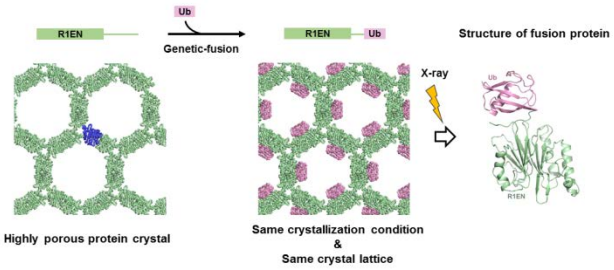
A patent application related to this work has been filed (application number PCT/JP2018/7697). Sharing of materials described in this patent application will be subject to standard material transfer agreements.

ACKNOWLEDGMENT

I thank the beamline staff at the Photon Factory (Proposal No. 2013G075 and 2017G615) and SPring-8 BL44XU (2015A6537, 2015B6537, 2016A6637 and 2016B6637) for supporting the data collection. I also thank Dr. Mariko Ariyoshi (Osaka University) for proofreading of the manuscript, Prof. Haruhiko Fujiwara (University of Tokyo) for providing the pET16b-R1EN (5-243), and Dr. Ayako Ohno (Tokushima University) for providing the human ubiquitin plasmid. This work was supported by KAKENHI (15K13747) from JSPS.

REFERENCES

(1) McPherson, A. *Methods Mol. Biol.* **2017**, *1607*, 17–50.
(2) Holcomb, J.; Spellmon, N.; Zhang, Y.; Doughan, M.; Li, C.; Yang, Z. *AIMS Biophys.* **2017**, *4*, 557–575.
(3) Hunte, C; Michel H. *Curr. Opin. Struct. Biol.* **2002**, *12*, 503–508.
(4) McGovern, R. E.; Fernandes, H.; Khan, A. R.; Power, N. P.; Crowley, P. B. *Nat. Chem.* **2012**, *4*, 527–533.
(5) Sakai, F.; Yang, G.; Weiss, M. S.; Liu, Y.; Chen, G.; Jiang, M. *Nat. Commun.* **2014**, *5*, 4634.
(6) Kobe, B.; Ve, T.; Williams, S. J. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2015**, *71*, 861–869.
(7) Waugh, D. S. *Protein Sci.* **2016**, *25*, 559–571.
(8) Bokhove, M.; Sadat Al Hosseini, H.; Saito, T.; Dioguardi, E.; Gegenschatz-Schmid, K.; Nishimura, K.; Raj, I.; de Sanctis, D.; Han, L.; Jovine, L. *J. Struct. Biol.* **2016**, *194*, 1–7.
(9) Seeman, N. C. *J. Theor. Biol.* **1982**, *99*, 237–247.
(10) Inokuma Y.; Yoshioka, S.; Ariyoshi, J.; Arai, T.; Hitora, Y.; Takada, K.; Matsunaga, S.; Rissanen, K.; Fujita, M. *Nature*, **2013**, *495*, 461–466.
(11) Yoshioka, S.; Inokuma, Y.; Hoshino, M.; Sato, T.; Fujita, M. *Chem. Sci*. **2015**, *6*, 3765–3768.
(12) Deng, H.; Grunder, S.; Cordova, K. E.; Valente, C.; Furukawa, H.; Hmadeh, M.; Gándara, F.; Whalley, A. C.; Liu, Z.; Asahina, S.; Kazumori, H.; O'Keeffe, M.; Terasaki, O.; Stoddart, J. F.; Yaghi, O. M. *Science* **2012**, *336*, 1018–1023.
(13) Fujita, D.; Suzuki, K.; Sato, S.; Yagi-Utsumi, M.; Yamaguchi, Y.; Mizuno, N.; Kumasaka, T.; Takata, M.; Noda, M.; Uchiyama, S.; Kato, K.; Fujita, M. *Nat. Commun*. **2012**, 3, 1093.
(14) Padilla, J. E.; Colovos, C.; Yeates, T. O. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2217–2221.
(15) King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D. *Nature* **2014**, *510*, 103–108.
(16) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D. *Science* **2012**, 336, 1171–1174.
(17) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. *Science* **2016**, *353*, 389–394.
(18) Lai, Y. T.; Reading, E.; Hura, G. L.; Tsai, K. L.; Laganowsky, A.; Asturias, F. J.; Tainer, J. A.; Robinson, C. V.; Yeates, T. O. *Nat. Chem.* **2014**, *6*, 1065–1071.
(19) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P. S.; Ravichandran, R.; Yi, S.; Davis, T. N.; Gonen, T.; King, N. P.; Baker, D. *Nature* **2016**, *535*, 136–139.
(20) Sinclair, J. C.; Davies, K. M.; Vénien-Bryan, C.; Noble, M. E. M. *Nat. Nanotech.* **2011**, *6*, 558–562.
(21) Liu, Y.; Gonen, S.; Gonen, T.; Yeates, T. O. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 3362–3367.
(22) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(23) Xiong, Y.; Eickbush, T. H. *Mol. Cell. Biol.* **1988**, *8*, 114–123.
(24) Maita, N.; Aoyagi, H.; Osanai, M.; Shirakawa, M.; Fujiwara, H. *Nucleic Acids Res.* **2007**, *35*, 3918–3927.
(25) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531–544.
(26) Camara-Artigas, A.; Plaza-Garrido, M.; Martinez-Rodriguez, S.; Bacarizo, J. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2016**, *72*, 29–35.
(27) Kulathu, Y.; Komander, D. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 508–523.
(28) Kantardjieff, K. A.; Matthews, R. B. *Protein Sci.* **2003**, *12*, 1865–1871.
(29) Brocchieri, L.; Karlin, S. *Nucleic Acids Res.* **2005**, *33*, 3390–3400.


Ub
R1EN
Genetic-fusion
R1EN
Ub
Structure of fusion protein
X-ray
Ub
R1EN
Highly porous protein crystal
Same crystallization condition
&
Same crystal lattice