

OPEN

# A host dTMP-bound structure of T4 phage dCMP hydroxymethylase mutant using an X-ray free electron laser

Si Hoon Park<sup>1</sup>, Jaehyun Park<sup>2</sup>, Sang Jae Lee<sup>2</sup>, Woo Seok Yang<sup>1</sup>, Sehan Park<sup>2</sup>, Kyungdo Kim<sup>3</sup>, Zee-Yong Park<sup>3</sup> & Hyun Kyu Song<sup>1\*</sup>

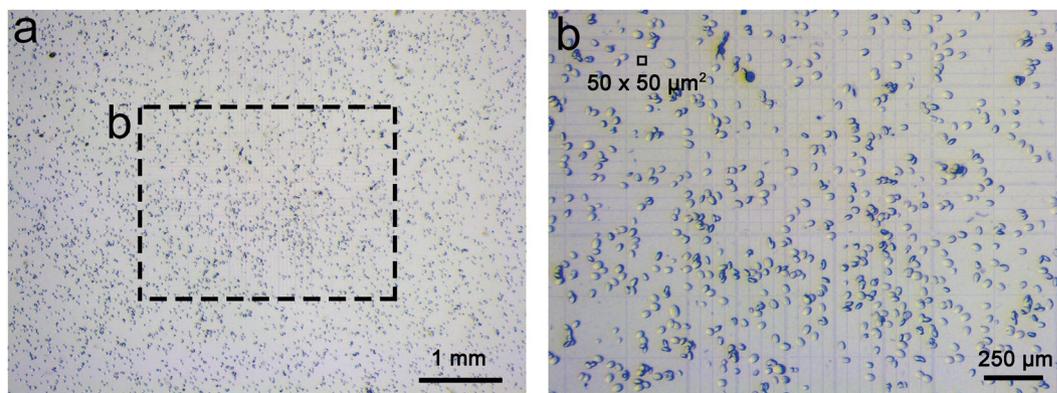
The hydroxymethylation of cytosine bases plays a vital role in the phage DNA protection system inside the host *Escherichia coli*. This modification is known to be catalyzed by the dCMP hydroxymethylase from bacteriophage T4 (T4dCH); structural information on the complexes with the substrate, dCMP and the co-factor, tetrahydrofolate is currently available. However, the detailed mechanism has not been understood clearly owing to a lack of structure in the complex with a reaction intermediate. We have applied the X-ray free electron laser (XFEL) technique to determine a high-resolution structure of a T4dCH D179N active site mutant. The XFEL structure was determined at room temperature and exhibited several unique features in comparison with previously determined structures. Unexpectedly, we observed a bulky electron density at the active site of the mutant that originated from the physiological host (i.e., *E. coli*). Mass-spectrometric analysis and a cautious interpretation of an electron density map indicated that it was a dTMP molecule. The bound dTMP mimicked the methylene intermediate from dCMP to 5'-hydroxymethyl-dCMP, and a critical water molecule for the final hydroxylation was convincingly identified. Therefore, this study provides information that contributes to the understanding of hydroxymethylation.

The X-ray free electron laser (XFEL) is a new fourth-generation light source that produces considerably bright X-ray pulses<sup>1–4</sup>. In contrast to X-rays generated by synchrotron sources, XFELs use femtosecond pulse with a brightness that is a billion times higher. Applications well suited to XFEL include data collection with submicron crystals at room temperature (RT), avoidance of radiation damage, and time-resolved crystallography<sup>1,4–12</sup>.

After the XFEL structures of photosystem I (PDB ID: 3PCQ)<sup>13</sup> and the photosynthetic reaction center (PDB ID: 4AC5)<sup>14</sup> were solved with the Linac Coherent Light Source (LCLS), many structures including G protein-coupled receptors for drug design and photoreactive proteins for molecular motion were determined using XFEL data collected at XFEL facilities<sup>8,15–21</sup>. Interestingly, the *in cellulo* structure of cathepsin B with a natively inhibited propeptide complex (PDB ID: 3MOR) has been reported<sup>22</sup>. Although protein structures using XFELs have been reported consistently, they are an extremely small portion of the overall deposited PDB (0.09%; 143 structures out of 155,618). In addition, the studied proteins are largely limited to some classes of proteins such as photoactivating proteins for time-resolved (TR) experiments, membrane proteins that are stable in the lipidic cubic phase (LCP), and robust model proteins<sup>23,24</sup>. The intermediate structures of the adenine riboswitch (PDB IDs: 5SWD, 5SWE, and 5E54) and  $\beta$ -lactamase with ceftriaxone (PDB IDs: 6B5Y, 6B5X, and 6B6A) have been challenges faced by time-resolved serial X-ray crystallography (TR-SFX)<sup>25,26</sup>. In contrast to the laser-triggered “pump-probe” technique, the TR-SFX at room temperature (RT) based on “mix-and-inject” is more generally applicable for understanding the enzyme mechanism<sup>6–8</sup>.

Herein, we described the XFEL structure of dCMP hydroxymethylase (dCH) from bacteriophage T4 (T4dCH) at 1.9-Å resolution using the Pohang Accelerator Laboratory (PAL) XFEL. The dCH is the first identified viral enzyme from T-even phage, and it gives viral phage the opportunity to escape the restriction-modification system of host bacteria by the hydroxymethylation of a cytosine base<sup>27–30</sup>. It belongs to the thymidylate synthase (TS)

<sup>1</sup>Department of Life Sciences, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea. <sup>2</sup>PAL-XFEL, Pohang Accelerator Laboratory, POSTECH, Pohang, Gyeongbuk, 37673, South Korea. <sup>3</sup>School of Life Sciences, Gwangju Institute of Science and Technology, Gwangju, 61005, South Korea. \*email: [hksong@korea.ac.kr](mailto:hksong@korea.ac.kr)



**Figure 1.** Microcrystallization of T4dCH D179N mutant. (a) Microcrystals observed in hemocytometer by light microscope. (b) Magnified images of microcrystals on Neubauer-improved chamber in hemocytometer. Size of small grid is  $50 \times 50 \mu\text{m}^2$  as drawn in outlined box. See Supplementary Fig. 1 for flow cytometry results using these microcrystals.

superfamily and catalyzes dCMP to hydroxymethyl-dCMP (hmdCMP) using 5,10-methylenetetrahydrofolate (5,10-mTHF) and water as a methyl and hydroxyl donor, respectively<sup>31,32</sup>. Previously, we reported the apo-form and binary and ternary complex structures of T4dCH, and we suggested the unique mechanism of hydroxymethylation<sup>33,34</sup>. However, questions about the cytosine modification catalysis remain to be answered because the methylene intermediate structure has not been elucidated yet.

It is also unknown whether hydroxymethylation is associated with water-mediated hydrogen bond networks in the active site. In this study, we solved the structure of T4dCH D179N mutant by introducing a TS-like character<sup>35,36</sup> using the XFEL technique. Interestingly, the substrate binding site of the mutant was occupied by an unidentified ligand that comes from the natural host (i.e., *E. coli*) for this bacteriophage enzyme. Through a cautious interpretation of the electron-density map and the mass spectrometry results, the unknown ligand was confirmed as dTMP. This dTMP complex structure can mimic the methylene intermediate state. Thus, it can be used to propose the detailed dCMP hydroxymethylation mechanism. From our XFEL-based native dTMP complexed T4dCH structure at RT, we can better understand the hydroxymethylation mechanism of nucleoside monophosphates.

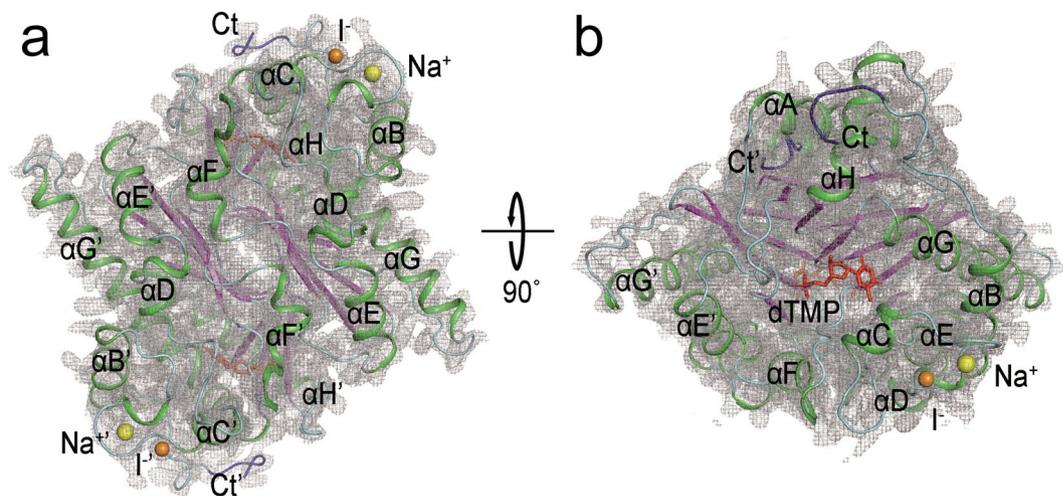
## Results and Discussion

**Microcrystals of T4dCH D179N mutant.** Crystals of T4dCH D179N mutant nucleate faster than those of wild-type (WT) protein. Therefore, we attempted to focus on the crystallization of D179N microcrystals. Initially, we obtained D179N microcrystals by the vapor-diffusion method that was applied to solve the ternary complex structure<sup>34</sup>. However, there were several problems involved in the conventional crystallization method for obtaining the microcrystals. First, it was not sufficient to obtain a large amount of microcrystals owing to the spatial limitation of the drop size (up to  $10 \mu\text{L}$  in one crystallization drop). Second, it was difficult to generate reproducible homogenous and singular microcrystals. Lastly, the diffusion method took several hours to obtain the microcrystals; thus, it was not easy to use excessive crystallization drops at once. Therefore, a batch method in a microtube<sup>37–40</sup> was applied to overcome these difficulties. We increased the protein concentration to approximately ten times that of the previous trials and optimized the crystallization conditions with various trials. Finally, we reproduced millions of microcrystals by simple mixing the proteins and the crystallization solution in the microtube (Fig. 1a).

After mixing the proteins with the precipitants, we observed the crystalline slurry immediately. An hour was sufficient for obtaining the crystals with approximate dimensions of  $20 \times 20 \times 50 \mu\text{m}^3$  (Fig. 1b). The micron-size crystals were easily reproduced homogeneously in a nucleated single form without cracks or disruptions. The hemocytometer indicated that the number of microcrystals was  $1\text{--}2 \times 10^6$  crystals/ml (Fig. 1b). We further checked the homogeneity and population of cellular-size crystals using flow cytometry (Supplementary Fig. S1). The microcrystals exhibited a monodisperse peak and a total of  $2.2 \times 10^6$  crystals/ml in the histogram of forward scattering vs. count. This indicated that the current microcrystals were suitable for SFX experiments.

**Structure determination using XFEL.** We mixed microcrystals in a viscous medium as a carrier matrix and injected using LCP injector. The microcrystals were stable and well diffracted in a monoolein mixture. Then, we successfully collected a total of 232,694 images at RT. The average hit rate was approximately 12.5%, and the number of hit images was 28,966. The diffraction images were indexed by optimizing the detector geometry. Finally, we indexed a total of 22,925 images. The indexing rate was as high as 79.1%. The space group was *I*222, and the distribution of the unit cell parameters was  $a = 55.86 \text{ \AA}$ ,  $b = 74.97 \text{ \AA}$ , and  $c = 157.62 \text{ \AA}$ . Compared with the cell parameters in a cryogenic condition<sup>34</sup>, the *c*-axis was extended to approximately  $3 \text{ \AA}$ .

The number of unique reflections was 27,244, and the redundancy was as high as 595.9. The overall  $R_{\text{split}}$  was 19.14%, and the overall signal-to-noise ratio (SNR) was 4.87. The T4dCH D179N microcrystals were diffracted up to approximately  $1.9 \text{ \AA}$ , and the SNR of the highest-resolution shell ( $1.95\text{--}1.88 \text{ \AA}$ ) was 2.51 near the edge of the detector. At the highest-resolution shell, the Pearson correlation coefficient (CC) and CC\* values were 0.80 and



**Figure 2.** XFEL structure of dimeric T4dCH D179N within electron density map (gray) projected along twofold rotational axis of symmetry. (a) Front and (b) top view rotated along x-axis by 90°. 2Fo-Fc maps (gray) are contoured at 1.0  $\sigma$ . Secondary structural elements are drawn as ribbon diagram in green ( $\alpha$ - and  $3_{10}$ -helix), purple ( $\beta$ -strand), and cyan (loops). C-terminal ends are colored blue and labeled as Ct (N-terminal ends exist in behind side). Bound ligands are shown using ball-and-stick model in red (dTMP), orange (iodide), and yellow (sodium ion).

0.94, respectively. After processing the entire dataset with a completeness of 100%, we solved the XFEL structure of T4dCH D179N at a resolution of 1.9 Å (Fig. 2). At the end of the refinement, the model yielded a  $R_{\text{work}}/R_{\text{free}}$  value of 0.169/0.191. Details of the data collection and refinement are listed in Table 1.

**Overall structure of T4dCH D179N mutant.** The overall model of T4dCH D179N fitted in well with an electron density map generated by XFEL (Fig. 2). The polypeptide chains were clearly visible from the N-terminus of Met1 to the C-terminus of Ala246, except for the additional C-terminal histidine tag (Supplementary Fig. S2a,b). Near the active site, a flexible loop around Lys28 residue for covering the bound substrate fitted in well with the electron densities (Supplementary Fig. S2c). We could also observe the clear electron density for the side chains of catalytic key residues (Supplementary Fig. S2d). Intriguingly, a bulky difference map was detected in the substrate binding site of the XFEL structure even though no substrate-like compounds were added during the purification and crystallization steps (Fig. 3). We assumed that this molecule may be dTMP. This is further detailed in the following section.

We found two blubs around  $3_{10}$ -helix G1 that were interpreted as an iodide and a sodium ion (Supplementary Fig. S3a). As mentioned in our previous report<sup>34</sup>, the iodide ion affected the crystal contacts of space group *I*222, which was clearly implied by the Fo-Fc difference map. The binding of a sodium ion was not discussed in previous studies because the software automatically predicted this position as a water molecule. We carefully explored the XFEL structure and found that this electron density matches with the sodium ion that forms an octahedral geometry with the main chain oxygen atoms of Lys72, Ile74, and Gly76, and side chain hydroxyl group of Thr78, in addition to two water molecules (Supplementary Fig. S3b). The distances between the sodium ion and the oxygen atoms of protein and water molecules were well matched with the observed distance range as reported in the PDB data bank<sup>41,42</sup>.

**Unique features of RT structure.** Next, we compared the XFEL structure of T4dCH D179N with our previous structures of C2-form crystals collected at RT (PDB ID: 1B5E) and *I*222-form crystals under cryogenic conditions (PDB ID: 6A9A). The calculated root-mean-square deviation (RMSD) value was 0.447 Å for 240 matching C $\alpha$  atoms between the XFEL- and C2-form structures. Interestingly, despite having the same space group and similar cell parameters, the C $\alpha$  RMSD between the XFEL- and *I*222- form structures in cryogenic conditions was 0.780 Å, which indicates more deviations than the previous one. The major difference between previous C2- and *I*222-form crystal structures was the presence or absence of hydrogen-bonding interactions of the N-terminus to the other protomer in the XFEL structure<sup>33,34</sup>.

The conformation of the N-terminal region of the XFEL structure was similar to that of the C2 structure collected at RT. The hydrogen-bonding interactions were maintained by Met1-Asp23', Met1-Glu35', Met1-Ile37', and Ser3-Ile37' (apostrophe indicates another subunit of the dimer) (Supplementary Fig. S4). Therefore, there might be minor structural rearrangements in the N-terminal region when the hydrogen-bonding interaction is hindered during the freezing process. This is an excellent example of determining the protein structure at a more natural RT using XFEL technology.

**Molecule derived from host *E. coli* found at substrate binding site.** The unknown electron density was occupied at the substrate binding site of T4dCH D179N (Fig. 3a). In the protein preparation and crystallization conditions, no ligand-like molecules were included. For the molecular replacement (MR) calculation, the template model did not include any ligand-like molecules. Therefore, the omit map seemed to represent the

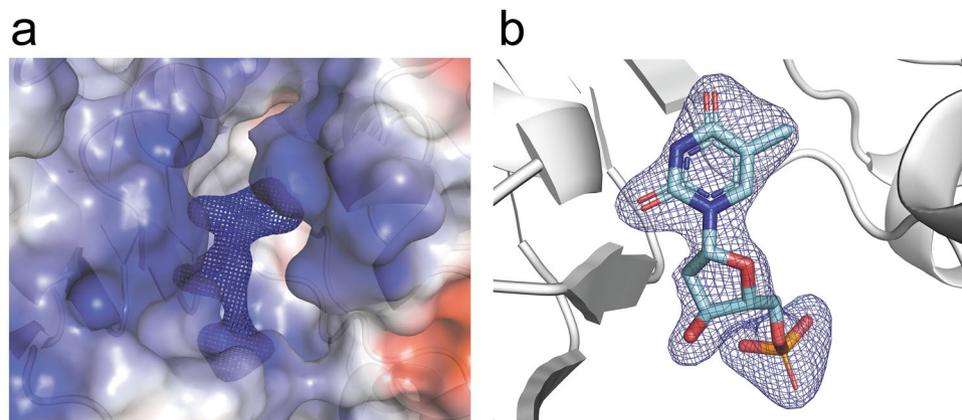
	D179N mutant
<b>Data collection</b>	
X-ray source	PAL-XFEL (SFX)
Energy (eV)	9,700
Pulse width (fs)	20
Space group	I222
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	52.86, 74.97, 157.62
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
No. of total collected images	232,694
No. of diffracted images	28,963
No. of indexed images	22,925
No. of total reflections	16,234,836
No. of unique reflections	27,244
Resolution (Å)	40.0–1.88 (1.95–1.88) <sup>a</sup>
$R_{\text{split}}$ (%) <sup>b</sup>	19.14 (37.95)
$\langle I/\sigma(I) \rangle$	4.87 (2.51)
CC <sup>c</sup>	0.93 (0.80)
CC* <sup>d</sup>	0.98 (0.94)
Completeness (%)	100.0 (100.0)
Redundancy	595.9 (176.2)
<b>Refinement</b>	
Resolution (Å)	37.50–1.90 (1.97–1.90)
No. reflections	26,553
$R_{\text{work}}/R_{\text{free}}$ (%)	16.91/19.05 (21.08/22.90)
No. atoms	
Protein	2,009
Ligand (dTMP, I, Na) / Water	23 (21, 1, 1) / 60
B-factors (Å <sup>2</sup> )	
Protein	58.4
Ligand (dTMP, I, Na)/Water	50.6 (50.6, 50.4, 50.4) / 54.9
RMSDs	
Bond lengths (Å)	1.286
Bond angles (°)	0.019
Ramachandran plot (%)	
Favored	96.72
Allowed	3.28
Outliers	0
Clashscore	1.14
MolProbity score	1.75

**Table 1.** Data collection and refinement statistics. <sup>a</sup>Values in parentheses are for highest-resolution shell.

<sup>b</sup> $R_{\text{split}} = 2^{-1/2} \times \sum |I_{\text{even}} - I_{\text{odd}}| / [1/2 \times \sum (I_{\text{even}} + I_{\text{odd}})]$ . <sup>c</sup>Pearson correlation coefficient. <sup>d</sup> $CC^* = [2CC_{1/2} / (1 + CC_{1/2})]^{1/2}$ .

molecule that originated from *E. coli*, which is the physiological host of the T4 bacteriophage. Because the resolution of our XFEL structure was sufficiently high, we attempted to interpret a ligand molecule based on electron density. First, a bulky electron density was found between the guanidinium groups of Arg123' and Arg124' residue<sup>33</sup>. This is known as a binding site of a phosphate moiety (Supplementary Fig. S5a). Second, an electron density similar to the previous one<sup>33</sup> existed between the  $\epsilon$ -nitrogen of His216 and the hydroxyl group of Tyr218. This is known as the recognition site of a deoxyribose ring (Supplementary Fig. S5b)<sup>33,43</sup>. Lastly, a disc-shaped electron density filled the active cavity near the catalytic cysteine (Cys148) and the pyrimidine-determining residue (D179N) (Supplementary Fig. S5c). Moreover, it was very similar to the electron density of dCMP in the active site of the binary complex (PDB ID: 1B5E), except for the residual density at the C5 position of the pyrimidine base (Fig. 3b and Supplementary Fig. S5d). Thus, we assumed that the unknown ligand was a methylated pyrimidine nucleotide analog.

**Mass spectrometry analysis for ligand identification.** For the accurate identification of the unknown nucleotide analog by qualitative analysis, we employed mass spectrometry for analyzing the analog (Fig. 4). Because the WT protein possessed an empty active site, we compared the mass spectral signals for T4dCH D179N and WT protein. The major difference in the signals between D179N and WT protein was found at 321.0495 m/z



**Figure 3.** Electron density map of initial refinements in substrate binding site of T4dCH D179N XFEL structure. Fo-Fc map is contoured at  $3.0 \sigma$ . (a) Electron density map with electrostatic surface model and ribbon diagram of T4dCH D179N. (b) Fitting dTMP molecule into electron density map.

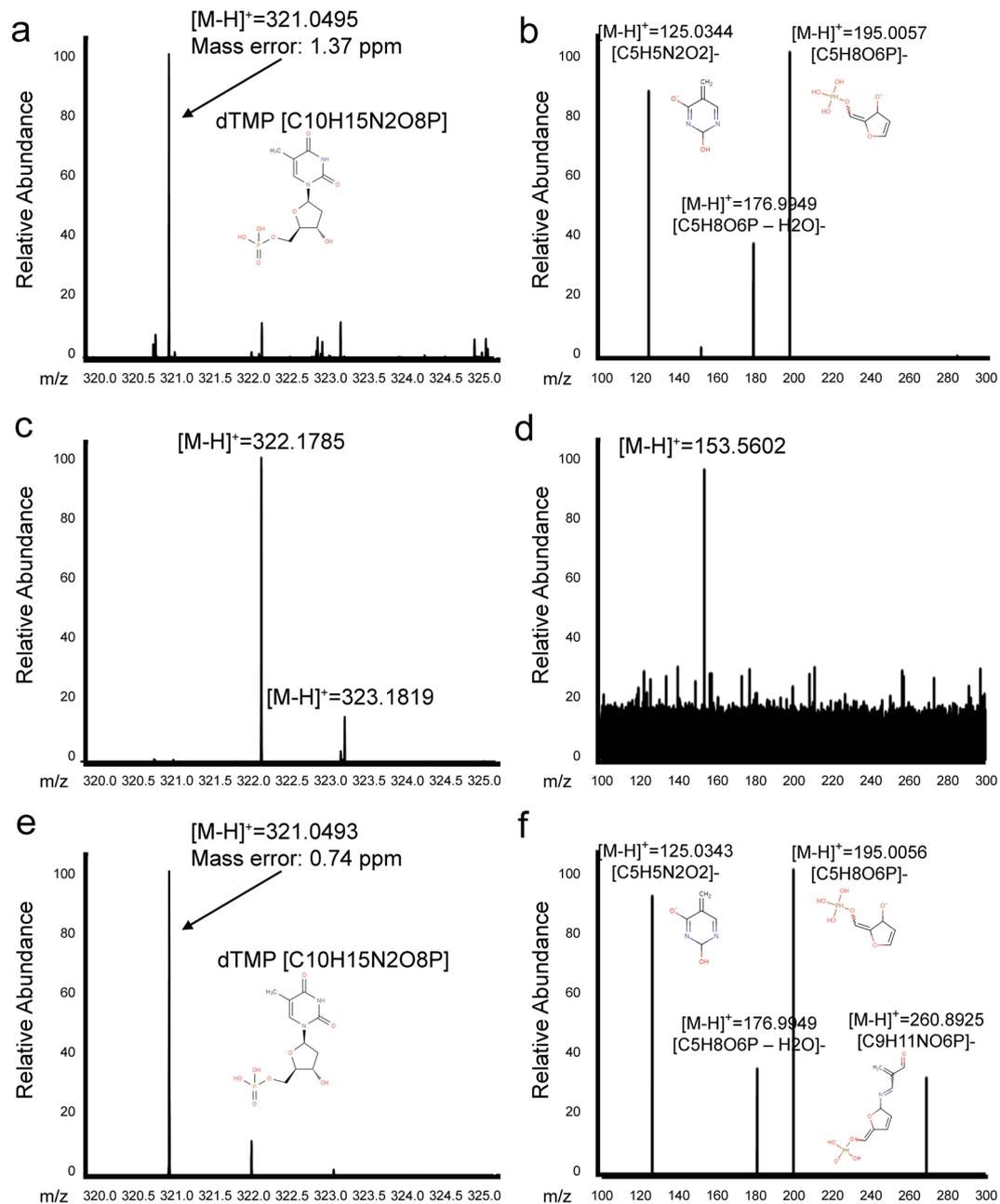
(Fig. 4a,c). Interestingly, the mass spectral signal of dTMP occurred at 321.0493 m/z and exactly matched the m/z ratio of the unknown substance in T4dCH D179N protein (Fig. 4a,e). For further confirmation, high-resolution collisional dissociation (HCD) fragmentation was analyzed. A fragment spectrum of m/z ratio 321.0495 in D179N was detected at the same positions of the dTMP fragments (Fig. 4b,f). Therefore, by combining XFEL crystallography and mass spectrometric analysis, the unknown substance bound to T4dCH D179N was revealed as a dTMP molecule.

**Native dTMP-bound structure of T4dCH D179N.** The T4dCH D179N structure was further refined after incorporating the dTMP model at the active site. The dTMP molecule was well fitted to the electron density map in the active site (Fig. 3b). The overall ligand geometry of dTMP was similar to that of dCMP in the binary complex of T4dCH (Fig. 3b and Supplementary Fig. S5d). The details of the interaction between dTMP and T4dCH D179N are as follows. The phosphate moiety of dTMP was recognized by the side chain atoms of Lys28, Arg168, Ser169, Arg123', Arg124', and a water molecule (Wat603), which formed a hydrogen bond with the main-chain atom of Phe146 (Fig. 5). A hydroxyl group of deoxyribose moiety was recognized by the side chain atoms of His216 and Tyr218.

For the recognition of the thymine base, the carbonyl oxygen at position 2 was recognized by the side-chain nitrogen atom of Asn170 and the main-chain nitrogen atom of Asp171, and the carbonyl oxygen at position 4 was recognized by the side-chain oxygen atom of Tyr96 and side-chain nitrogen atom of Asn179. Interestingly, the Asn residue of the mutant recognized the N3 atom and the carbonyl oxygen at position 4 in the thymine base, whereas the carboxylate of Asp179 of WT interacted with the N3 atom and primary amine group at position 4 of the cytosine base; this is shown in the binary complex structure<sup>33</sup>. In comparison with the ligand geometry of dCMP in the binary complex of T4dCH WT, the ligand geometry of dTMP in the XFEL structure of D179N exhibited a critical difference around a methyl group at the C5 position in the pyrimidine base. Indeed, the C5 position is where the hydroxymethylation occurs, and the previous study suggested that there is a methylene intermediate at this position<sup>34</sup>. Therefore, our dTMP-bound structure provided an insight into the methylene intermediate structure.

**Mimicry of methylene intermediate state.** Our previous studies on the hydroxymethylation mechanism of T4dCH suggested that the hydride transfer found in the enzymatic reaction of TS was blocked because the distance from the C6 of THF to the putative C7 exocyclic methylene position<sup>34</sup> was too large. Instead, the reaction of T4dCH underwent hydroxylation by the potential substrate water molecule near the putative C7 exocyclic methylene position<sup>34</sup>. This C7 methyl group must be transferred, and it must form a covalent bond with the C5 atom of the cytosine base. Our dTMP-bound structure well mimicked the methylene intermediate that was the methylated pyrimidine nucleotide complex (Supplementary Fig. 6). Therefore, we tried to understand the final hydroxylation reaction compared to the dTMP-bound binary complex with ternary complexes of T4dCH and EcTS (PDB IDs: 6A9A and 1KZI)<sup>34,44</sup>.

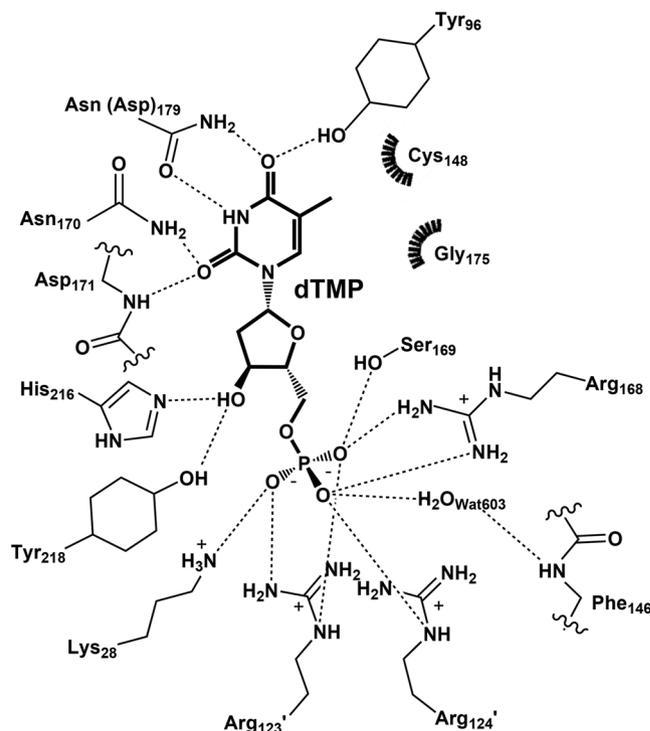
When the dTMP-bound complex was superimposed on the ternary [T4dCH-dCMP-THF] complex, the C7 position of dTMP and the C6 position of THF were at a distance of 3.8 Å from each other (Fig. 6a), whereas the C7 position of dTMP was at a distance of 2.1 Å from a potential substrate water molecule (Wat401 in 6A9A) (Fig. 6b). Indeed, the current structure exhibited an electron density for a water molecule (Wat601) that may correspond to the Wat401 in the ternary complex structure<sup>34</sup> (see next section for more details). In contrast, by superimposing the dTMP-bound complex and the ternary complex [EcTS-dUMP-THF], the C7 position of dTMP and C6 position of THF were within a distance of 2.8 Å, which is much less than 3.8 Å (Fig. 6c). The hydrogen atom attached to the C6 position of THF transferred to the methylene intermediate. This was the rate-limiting hydride transfer step of the TS enzymatic mechanism<sup>44,45</sup>. The main difference was derived from the different orientation of the THF cofactor (Fig. 6d).



**Figure 4.** Identification of dTMP in T4dCH D179N and WT by high-resolution, high-mass-accuracy mass spectrometry. (a) Mass spectrum of T4dCH D179N. (b) Tandem mass spectrum of  $[M-H]^+ = 321.0495$  from T4dCH D179N. (c) Mass spectrum of T4dCH WT. (d) Tandem mass spectrum of  $[M-H]^+ = 321.0495$  from T4dCH WT. (e) Mass spectrum of standard dTMP. (f) Tandem mass spectrum of  $[M-H]^+ = 321.0495$  from standard dTMP.

**Water molecules at dTMP-bound region.** Six water molecules (Wat601–606) were built with reliable electron densities at the active site of the intermediate-mimicking structure (Supplementary Fig. 7). All water molecules were maintained by hydrogen-bonding interactions with the amino acid residues of the active site pocket. Compared with the binary and ternary complex structures (PDB IDs: 1B5E and 6A9A), most of the water molecules were well ordered (Wat602–606); however, a water molecule was newly observed (Wat601) near the Asp145 residue. The position of this water molecule did not overlap that of the previously proposed substrate water molecule (Wat401 in 6A9A).

However, both Wat601 in the current structure and Wat401 in the ternary complex were equidistant (2.8 Å) from the Asp145 residue. The electron density of the new water molecule was relatively weak than those of other solvents because it was stabilized by hydrogen-bonding interactions with the carboxylate of Asp145 only. Despite a weak interaction, Wat601 existed at a very close distance from the previous substrate water molecule in the ternary complex structure (within 2.0 Å). Therefore, the substrate water molecule existed near the carboxyl group



**Figure 5.** Schematic diagram of dTMP interaction. Hydrogen bonds are indicated by dashed lines, while hydrophobic contacts are represented by arc with spokes. Diagram was generated by LIGPLOT<sup>55</sup> and modified for clarity.

of Asp145 residue at the methylene intermediate stage. Then, the molecule was activated and moved to a position closer to the C7 of the exocyclic methylene intermediate by hydrogen-bonding interactions of Asp145 and THF. As noted earlier, all other water molecules were structurally conserved except for Wat601. The previous study proposed that another water molecule (Wat409 in 6A9A) plays an important role in the activation of the dCMP substrate<sup>34</sup>. This step occurs before the formation of the methylene intermediate. The position of Wat602 in the current structure overlapped that of Wat409 in the ternary complex. Thus, a water molecule (Wat602) for substrate activation existed constitutively at the fixed position, whereas a water molecule (Wat601) involved in the final hydroxylation step was located at a slightly different position that could be adjusted during the enzymatic reaction steps of T4dCH.

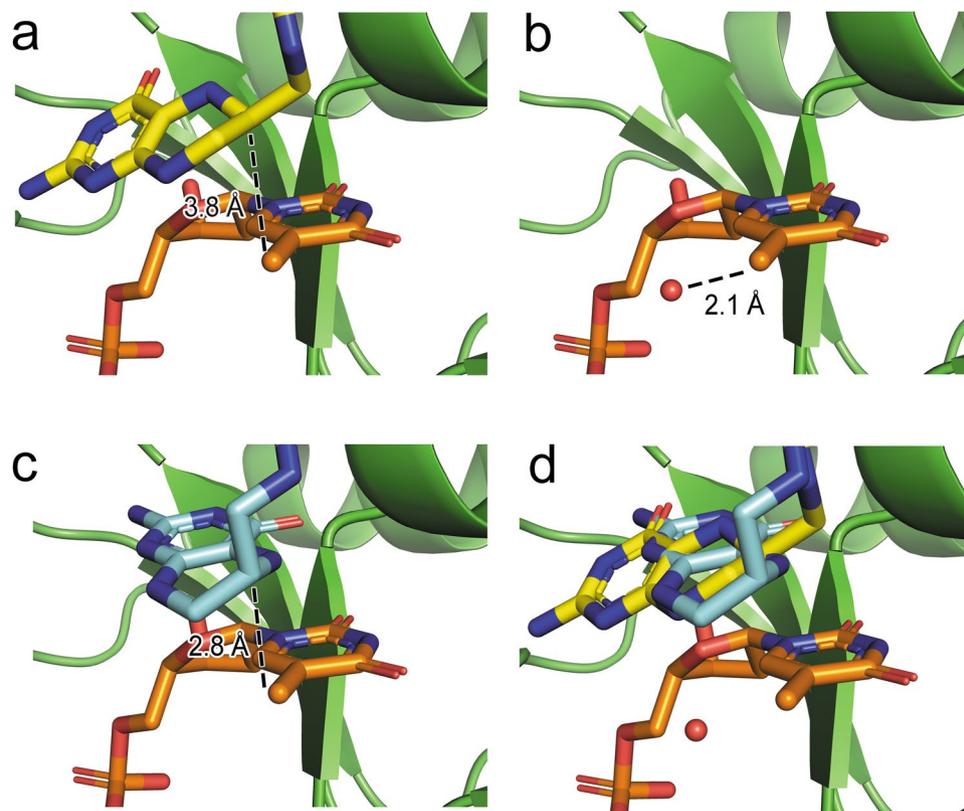
In conclusion, the current dTMP-bound structure provides clear experimental support for the previously proposed hydroxylation step of T4dCH<sup>34</sup>. Owing to the space constraint of the ternary [EcTS-dUMP·THF] complex, no water molecule could be accommodated near the cofactor and substrate, whereas a water molecule found in the ternary complex [T4dCH·dCMP·THF] was located at the optimal position near the methylene intermediate for hydroxylation. Compared to the hydride transfer of TS, the last hydroxylation step of dCH could be directly monitored by the movement of the electron density of oxygen atoms including the substrate water molecule in a crystal structure. Further TR-XFEL study using a rapid mix-and-inject technique will allow us to visualize the molecular motion of the hydroxymethylation mechanism.

## Materials and Methods

**Expression and protein purification.** T4dCH D179N mutant was generated by PCR-based mutagenesis, which has been described in a previous paper in detail<sup>34</sup>. The protein was expressed in *E. coli* BL21(DE3) cells with a C-terminal octa-histidine affinity tag to maximize the protein yield. The cells were grown overnight in LB media and then induced with 0.5 mM of isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) for 4 h at 37 °C. Subsequently, the cells were harvested and resuspended in buffer A [50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 1 mM tris(2-carboxyethyl)phosphine] with 10 mM of imidazole. The resuspended cells were disrupted by sonication and then centrifuged (17,000 rpm for 1 h at 4 °C) to remove the cell debris.

Supernatants were loaded onto a Histrap FF (GE Healthcare, 17-5255-01) and eluted with gradually increasing concentrations of imidazole up to 500 mM in buffer A. The eluents were concentrated using an Amicon Ultra-30K centrifugal filter unit (Millipore, UFC903024) and dialyzed against a dialysis buffer (13.6 mM Na<sub>2</sub>HPO<sub>4</sub>, 8.1 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM ethylenediaminetetraacetic acid, 150 mM NaCl, and 1 mM 1,4-dithiothreitol) using a Slide-A-Lyzer MINI Dialysis Device (Thermo Fisher Scientific, 88405) at 4 °C overnight. The final concentration of purified protein was 40 mg/ml.

**Microcrystal preparation.** Microcrystals of T4dCH D179N mutant were grown by the free-interface diffusion (FID) method in a 1.5-ml microtube (Axygen, MCT-150-C). The crystallization condition was initially



**Figure 6.** Mimicry of methylene intermediate by dTMP-bound T4dCH D179N mutant. Superposition between current T4dCH mutant (green) complexed with dTMP (orange) and (a) THF (yellow) in ternary complex of T4dCH, (b) potential substrate water (red) in ternary complex of T4dCH (Wat401 in PDB 6A9A), and (c) THF in ternary complex of EcTS (cyan). (d) Superimposed models of panels (a–c). Models are shown by ribbon diagram and stick model. In stick model, atoms are colored red (oxygen) and blue (nitrogen).

attempted from the wild-type crystal condition reported previously<sup>33</sup>. The optimized crystallization solution [2  $\mu$ l of 1.0-M Tris-HCl (pH 8.5), 2  $\mu$ l of 1.0-M NaI (Sigma, 383112), and 17  $\mu$ l of  $C_6H_5Na_3 \cdot 2H_2O$  (Hampton Research, HR2-549)] and 10  $\mu$ l of 40 mg/ml D179N mutant protein were mixed gently to avoid air bubbles. Microcrystals of approximately  $20 \times 20 \times 50 \mu m^3$  were grown at 20 °C within 1 h. The homogeneity and morphology were determined by a hemocytometer (Marienfeld-Superior, 0610030 and INCYTO, DHC-N01) and a flow cytometer (BD, BD Accuri C6 Plus). For SFX experiments, 40  $\mu$ l of microcrystal solution was mixed with 60  $\mu$ l of monoolein (Hampton Research, HR2-435) using syringes (Hamilton, 81065-1710RNR and 81165-1725RNR) and an LCP coupler (Mitegen, SKU:M-R-1006905). The crystal population of the monoolein mixture was monitored using a light microscope.

**Data collection and processing.** The diffraction data were collected at a Nanocrystallography and Coherence Imaging (NCI) hutch at PAL-XFEL using an LCP injection system<sup>46</sup>. The X-ray energy was 9.7 keV with a repetition rate of 30 Hz and duration of 20 fs. The pulse energy was approximately 485  $\mu$ J, and the photon flux was approximately  $1-2 \times 10^{11}$  photons per pulse. The incident X-ray beam was focused at  $5 \times 5 \mu m^2$  full width at half maximum (FWHM) by Kirkpatrick-Baez mirrors<sup>47</sup>. The sample was delivered to the X-ray pulse in a helium ambience at RT in a multifarious injection chamber of a molecular structure study (MICOSS) system<sup>48</sup>. The diffraction data were obtained on an MX225-HS (Rayonix LLC) CCD detector with  $4 \times 4$  binning mode corresponding to a pixel size of 156  $\mu m$ . The detector was positioned 108 mm from the sample. The edge of the detector had a resolution of approximately 1.8 Å. The serial X-ray images were monitored in real time using OnDA<sup>49</sup>, and the diffraction images were sorted using the Cheetah software with the ‘peakfinder 8’ algorithm<sup>50</sup>. The filtered images were indexed using *indexamajig* in the CrystFEL package<sup>51</sup>. The indexed images were scaled and merged by *partialator*. The data quality was calculated by *check\_hkl* and *compare\_hkl* in the CrystFEL package. The total number of collected images/hits/indexed images was 232,694/28,963/22,925 (Table 1).

**Structure determination, ligand fitting, and refinement.** The phases were obtained by molecular replacement (MR) using *Phaser*<sup>52</sup> with the previously solved wild-type structure as a search model (PDB ID: 1B5E)<sup>33</sup>. To model the dTMP as a ligand, the eLBOW module in PHENIX was used with ligand geometric constraints. Using the LigandFit module in PHENIX, the 21 atoms of dTMP without hydrogens were well fitted into the difference map with an overall CC and score of 0.852 and 73.10, respectively. Refinement and validation of the dTMP-bound model were performed using COOT<sup>53</sup> and PHENIX<sup>54</sup>. For interface analysis of the ligand-protein

contacts, LIGPLOT<sup>55</sup> and the PISA server in the CCP4 suite<sup>56</sup> were used. All structural figures were prepared using PyMOL (<http://www.pymol.org>).

**Mass spectrometry.** Mass spectrometry was performed with a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, IQLAAEGAAPFALGMAZR). Samples dissolved in 0.1% formic acid were infused with a syringe at a flow rate of 5 µl/min. A mass spectrometer equipped with a microspray ESI source was operated in full mass-spectral-acquisition mode for MS scan data in a mass range of m/z 100–500. A resolving power setting of 70,000 was used in negative ion mode. MS/MS scan data were also acquired in negative ion scan mode; a normalized collision energy (NCE) of 35% was applied to a precursor ion selected, m/z 321.05. The MS and MS/MS scan AGC target value was set at 1e<sup>6</sup>. All data acquisition and analysis were carried out with XCalibur v.2.2 software (Thermo Fisher Scientific).

**Accession code.** Atomic coordinates and structure factor files for T4dCH D179N have been deposited in the Protein Data Bank under the accession code 6L18.

Received: 6 October 2019; Accepted: 22 October 2019;

Published online: 08 November 2019

## References

1. Emma, P. *et al.* First lasing and operation of an ångström-wavelength free-electron laser. *nature photonics* **4**, 641–647 (2010).
2. Boutet, S., Fromme, P. & Hunter, M. S. *X-ray Free Electron Lasers: A Revolution in Structural Biology*. 1–479 (Springer, 2018).
3. Pellegrini, C. The history of X-ray free-electron lasers. *The European Physical Journal H* **37**, 659–708 (2012).
4. Kang, H.-S. *et al.* Hard X-ray free-electron laser with femtosecond-scale timing jitter. *Nature Photonics* **11**, 708–713 (2017).
5. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
6. Van Den Bedem, H. & Fraser, J. S. Integrative, dynamic structural biology at atomic resolution—it's about time. *Nature methods* **12**, 307–318 (2015).
7. Stauch, B. & Cherezov, V. Serial femtosecond crystallography of G protein-coupled receptors. *Annual review of biophysics* **47**, 377–397 (2018).
8. Johansson, L. C., Stauch, B., Ishchenko, A. & Cherezov, V. A bright future for serial femtosecond crystallography with XFELs. *Trends in biochemical sciences* **42**, 749–762 (2017).
9. Madey, J. M. Stimulated emission of bremsstrahlung in a periodic magnetic field. *Journal of Applied Physics* **42**, 1906–1913 (1971).
10. Ackermann, W. A. *et al.* Operation of a free-electron laser from the extreme ultraviolet to the water window. *Nature photonics* **1**, 336–342 (2007).
11. Ishikawa, T. *et al.* A compact X-ray free-electron laser emitting in the sub-ångström region. *nature photonics* **6**, 540–544 (2012).
12. Halavanau, A., Decker, F.-J., Emma, C., Sheppard, J. & Pellegrini, C. Very high brightness and power LCLS-II hard X-ray pulses. *Journal of synchrotron radiation* **26**, 635–646 (2019).
13. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).
14. Johansson, L. C. *et al.* Lipidic phase membrane protein serial femtosecond crystallography. *Nature Methods* **9**, 263–265 (2012).
15. Liu, W. *et al.* Serial femtosecond crystallography of G protein-coupled receptors. *Science* **342**, 1521–1524 (2013).
16. Spence, J. C. X-ray lasers for structure and dynamics in biology. *IUCrJ* **5**, 236–237 (2018).
17. Kern, J. *et al.* Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. *Science* **340**, 491–495 (2013).
18. Ghosh, E., Kumari, P., Jaiman, D. & Shukla, A. K. Methodological advances: the unsung heroes of the GPCR structural revolution. *Nature Reviews Molecular Cell Biology* **16**, 69–81 (2015).
19. Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561–567 (2015).
20. Johansson, L. C. *et al.* XFEL structures of the human MT 2 melatonin receptor reveal the basis of subtype selectivity. *Nature* **569**, 289–292 (2019).
21. Stauch, B. *et al.* Structural basis of ligand recognition at the human MT 1 melatonin receptor. *Nature* **569**, 284–288 (2019).
22. Redecke, L. *et al.* Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an X-ray laser. *Science* **339**, 227–230 (2013).
23. Boutet, S. *et al.* High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 (2012).
24. Botha, S. *et al.* De novo protein structure determination by heavy-atom soaking in lipidic cubic phase and SIRAS phasing using serial synchrotron crystallography. *IUCrJ* **5**, 524–530 (2018).
25. Olmos, J. L. *et al.* Enzyme intermediates captured “on the fly” by mix-and-inject serial crystallography. *BMC biology* **16**, 59 (2018).
26. Stagno, J. *et al.* Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242–246 (2017).
27. Wyatt, G. & Cohen, S. A new pyrimidine base from bacteriophage nucleic acids. *Nature* **170**, 1072–1073 (1952).
28. Cohen, S. S. The Northrop hypothesis and the origins of a virus-induced enzyme. *Protein science* **3**, 150–153 (1994).
29. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8**, 317–327 (2010).
30. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology* **11**, 675–687 (2013).
31. Hardy, L. W., Graves, K. L. & Nalivaika, E. Electrostatic guidance of catalysis by a conserved glutamic acid in *Escherichia coli* dTMP synthase and bacteriophage T4 dCMP hydroxymethylase. *Biochemistry* **34**, 8422–8432 (1995).
32. Graves, K. L. *Studies on the Mechanism of Deoxycytidylate Hydroxymethylase from Bacteriophage T4*, University of Massachusetts Medical School (1994).
33. Song, H. K., Sohn, S. H. & Suh, S. W. Crystal structure of deoxycytidylate hydroxymethylase from bacteriophage T4, a component of the deoxyribonucleoside triphosphate-synthesizing complex. *The EMBO journal* **18**, 1104–1113 (1999).
34. Park, S. H., Suh, S. W. & Song, H. K. A cytosine modification mechanism revealed by the structure of a ternary complex of deoxycytidylate hydroxymethylase from bacteriophage T4 with its cofactor and substrate. *IUCrJ* **6**, 206–217 (2019).
35. Graves, K. L., Butler, M. M. & Hardy, L. W. Roles of Cys148 and Asp179 in catalysis by deoxycytidylate hydroxymethylase from bacteriophage T4 examined by site-directed mutagenesis. *Biochemistry* **31**, 10315–10321 (1992).
36. Graves, K. L. & Hardy, L. W. Kinetic and Equilibrium.  $\alpha$ -Secondary Tritium Isotope Effects on Reactions Catalyzed by dCMP Hydroxymethylase from Bacteriophage T4. *Biochemistry* **33**, 13049–13056 (1994).
37. Pav, S. *et al.* Microtube batch protein crystallization: Applications to human immunodeficiency virus type 2 (HIV-2) protease and human renin. *Proteins: Structure, Function, and Bioinformatics* **20**, 98–102 (1994).
38. Darmanin, C. *et al.* Protein crystal screening and characterization for serial femtosecond nanocrystallography. *Scientific reports* **6**, 25345 (2016).

39. Sugahara, M. *et al.* Hydroxyethyl cellulose matrix applied to serial crystallography. *Scientific reports* **7**, 703 (2017).
40. Kováčsová, G. *et al.* Viscous hydrophilic injection matrices for serial crystallography. *IUCrJ* **4**, 400–410 (2017).
41. Harding, M. M. Metal–ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallographica Section D: Biological Crystallography* **58**, 872–874 (2002).
42. Zheng, H. *et al.* CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallographica Section D: Structural Biology* **73**, 223–233 (2017).
43. Zhao, G. *et al.* Structural basis of the substrate preference towards CMP for a thymidylate synthase MilA involved in mildiomycin biosynthesis. *Scientific reports* **6**, 39675 (2016).
44. Fritz, T. A., Liu, L., Finer-Moore, J. S. & Stroud, R. M. Tryptophan 80 and leucine 143 are critical for the hydride transfer step of thymidylate synthase by controlling active site access. *Biochemistry* **41**, 7021–7029 (2002).
45. Swiderek, K., Kohen, A. & Moliner, V. The influence of active site conformations on the hydride transfer step of the thymidylate synthase reaction mechanism. *Phys Chem Chem Phys* **17**, 30793–30804 (2015).
46. Park, J., Kim, S., Nam, K.-H., Kim, B. & Ko, I. S. Current status of the CXI beamline at the PAL-XFEL. *Journal of the Korean Physical Society* **69**, 1089–1093 (2016).
47. Kim, J. *et al.* Focusing X-ray free-electron laser pulses using Kirkpatrick–Baez mirrors at the NCI hutch of the PAL-XFEL. *Journal of synchrotron radiation* **25**, 289–292 (2018).
48. Park, J., Kim, S. & Nam, K. Multifarious injection chamber for molecular structure study (MICOSS) system: development and application for serial femtosecond crystallography at Pohang Accelerator Laboratory X-ray Free-Electron Laser. *Journal of synchrotron radiation* **25**, 323–328 (2018).
49. Mariani, V. *et al.* OnDA: online data analysis and feedback for serial X-ray imaging. *Journal of applied crystallography* **49**, 1073–1080 (2016).
50. Barty, A. *et al.* Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *Journal of applied crystallography* **47**, 1118–1131 (2014).
51. White, T. A. Processing serial crystallography data with CrystFEL: a step-by-step guide. *Acta Crystallographica Section D* **75**, 219–233 (2019).
52. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658–674 (2007).
53. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126–2132 (2004).
54. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213–221 (2010).
55. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering, design and selection* **8**, 127–134 (1995).
56. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235–242 (2011).

## Acknowledgements

We would like to thank all the members of PAL-XFEL. This study was supported by a National Research Foundation of Korea (NRF) grant from the Korean government (Grant No. 2016R1E1A1A01942623).

## Author contributions

S.H.P. and W.S.Y. prepared the protein samples and microcrystals; S.H.P., J.H.P., S.J.L. and S.P. performed the XFEL experiments; S.H.P. determined the X-ray structure of T4dCH D179N; K.K. and Z.-Y.P. performed the mass spectrometry experiments; S.H.P. and H.K.S. designed the experiments and wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-52825-y>.

**Correspondence** and requests for materials should be addressed to H.K.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019