






# Multimeric structure of a subfamily III haloalkane dehalogenase-like enzyme solved by combination of cryo-EM and x-ray crystallography

Klaudia Chmelova<sup>1,2</sup> | Tadeja Gao<sup>1,2</sup> | Martin Polak<sup>3</sup>  |  
 Andrea Schenk Mayerova<sup>1,2</sup>  | Tristan I. Croll<sup>4</sup> | Tanvir R. Shaikh<sup>3,5</sup> |  
 Jana Skarupova<sup>1</sup> | Radka Chaloupkova<sup>1,2</sup> | Kay Diederichs<sup>6</sup>  |  
 Randy J. Read<sup>4</sup> | Jiri Damborsky<sup>1,2</sup> | Jiri Novacek<sup>3</sup>  | Martin Marek<sup>1,2</sup> 

<sup>1</sup>Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Brno, Czech Republic

<sup>2</sup>International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic

<sup>3</sup>Central European Institute of Technology, Masaryk University, Brno, Czech Republic

<sup>4</sup>Department of Hematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

<sup>5</sup>Institute of Neuropathology, University Medical Center Göttingen, Göttingen, Germany

<sup>6</sup>Department of Biology, University of Konstanz, Konstanz, Germany

## Correspondence

Martin Marek, Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Kamenice 5, Bld. C13, 625 00 Brno, Czech Republic.  
 Email: [martin.marek@recetox.muni.cz](mailto:martin.marek@recetox.muni.cz)

## Funding information

Czech Science Foundation, Grant/Award Number: GA22-09853S; Czech Ministry of Education, Grant/Award Numbers: e-INFRA LM2018140, RECETOX RI LM2018121, INBIO CZ.02.1.01/0.0/0.0/16\_026/0008451; European Union's Horizon 2020 Research and Innovation Program, Grant/Award Numbers: 814418, 857560; MEYS CR, Grant/Award Number: LM2023042; European Regional Development Fund-Project "UP CIISB", Grant/Award Number: CZ.02.1.01/0.0/0.0/18\_046/0015974; Wellcome Trust Principal Research Fellowship, Grant/Award Number: 209407/Z/17/Z

## Abstract

Haloalkane dehalogenase (HLD) enzymes employ an S<sub>N</sub>2 nucleophilic substitution mechanism to erase halogen substituents in diverse organohalogen compounds. Subfamily I and II HLDs are well-characterized enzymes, but the mode and purpose of multimerization of subfamily III HLDs are unknown. Here we probe the structural organization of DhmeA, a subfamily III HLD-like enzyme from the archaeon *Haloferax mediterranei*, by combining cryo-electron microscopy (cryo-EM) and x-ray crystallography. We show that full-length wild-type DhmeA forms diverse quaternary structures, ranging from small oligomers to large supramolecular ring-like assemblies of various sizes and symmetries. We optimized sample preparation steps, enabling three-dimensional reconstructions of an oligomeric species by single-particle cryo-EM. Moreover, we engineered a crystallizable mutant (DhmeA<sup>ΔGG</sup>) that provided diffraction-quality crystals. The 3.3 Å crystal structure reveals that DhmeA<sup>ΔGG</sup> forms a ring-like 20-mer structure with outer and inner diameter of ~200 and ~80 Å, respectively. An enzyme homodimer represents a basic repeating building unit of the crystallographic ring. Three assembly interfaces (dimerization, tetramerization, and multimerization) were identified to form the supramolecular ring that displays a negatively charged exterior, while its interior part harboring

Klaudia Chmelova, Tadeja Gao, Martin Polak, and Andrea Schenk Mayerova contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

Review Editor: John Kuriyan

catalytic sites is positively charged. Localization and exposure of catalytic machineries suggest a possible processing of large negatively charged macromolecular substrates.

#### KEYWORDS

catalysis, cryo-EM, DhmeA, haloalkane dehalogenase, *Haloferax mediterranei*, multimerization, x-ray crystallography

## 1 | INTRODUCTION

Haloalkane dehalogenases (EC 3.8.1.5, HLDs) are a group of enzymes that catalyze the hydrolytic cleavage of a carbon–halogen bond in a range of halogenated organic compounds. The HLD family is thoroughly studied with a wide range of biotechnological applications including industrial biocatalysis, toxic by-product recycling, decontamination of chemical warfare agents, biosensing and bioremediation of environmental pollutants, and protein tagging for cell imaging (Koudelakova et al., 2013). They can be found in all domains of life; prokaryotic archaea and bacteria as well as eukaryotic organisms (Chovancova et al., 2007). However, their fundamental biological role in cellular processes remains unclear to this day.

Structurally, HLDs are categorized under the  $\alpha/\beta$ -hydrolase superfamily, a diverse group of hydrolytic enzymes (Dimitriou et al., 2017). The canonical  $\alpha\beta$ -sandwich is an integral part of the haloalkane dehalogenase architecture. Another vital component of the HLD molecule is the helical cap domain, the structure of which has an impact on the substrate specificity of many  $\alpha/\beta$ -hydrolase fold enzymes (Dimitriou et al., 2017). A distinct catalytic pentad forms the active site of HLD enzymes: a nucleophile aspartic acid, a histidine base, a catalytic acid (aspartic or glutamic acid) and two halide stabilizing residues (two tryptophans or tryptophan + asparagine) (Janssen, 2004). The active site of HLDs is situated in a hydrophobic pocket between the  $\alpha/\beta$ -fold core and the cap domain, connected to the surrounding solvent via a main tunnel and a slot tunnel (Markova et al., 2020; Pavlova et al., 2009).

The most thoroughly described HLD function is the conversion of alkyl halides to their corresponding alcohols, halides, and a proton. This two-step catalytic process consists of an  $S_N2$  nucleophilic substitution followed by hydrolysis of the ester intermediate (Verschueren et al., 1993). Water is the only required co-factor for this reaction (Damborsky & Koča, 1999). The substrate array of HLDs encompasses a large variety of halogenated aliphatic hydrocarbons, such as alkanes, cycloalkanes, alkenes, ethers, alcohols, ketones, and cyclic dienes

(Damborsky et al., 2001). The HLD family has therefore been divided into four substrate specificity groups (SSG). The preference for certain substrates seems to be influenced by the architecture of the access tunnels and the active site, rather than the sequence similarity across the HLD family (Koudelakova et al., 2013; Pavlova et al., 2009). This categorization of HLDs is vital for further bioengineering efforts. Nevertheless, many of the tested organohalogen substrates were synthetic, and therefore the SSG classification does not necessarily reflect the *in vivo* function of these enzymes.

Another approach to classify HLDs utilizes an evolutionary perspective. Phylogenetic analyses of the haloalkane dehalogenase family revealed three distinct subfamilies: HLD-I, HLD-II, and HLD-III. These three subfamilies vary mainly by the composition of their catalytic pentad and the structure of their cap domain (Chovancova et al., 2007). The first two subfamilies contain the three archetypal HLDs: Dh1A from HLD-I, and DhA and LinB from HLD-II. These three renowned enzymes have been identified decades ago and paved the way for the engineering and biotechnology utilization of HLDs (Nagata et al., 2015). Numerous crystal structures of the members of the HLD-I and HLD-II families have been determined, partially due to their ease of handling and overall stability (Chaloupkova et al., 2014; Franken et al., 1991; Gehret et al., 2012; Lahoda et al., 2014; Marek et al., 2000; Mazumdar et al., 2008; Oakley et al., 2004; Verschueren et al., 1993).

Comparatively, subfamily III (HLD-III) is the least studied of the three HLD subfamilies. Numerous members originate from pathogenic or extremophilic organisms and display low dehalogenase activity compared to the other two subfamilies (Vanacek et al., 2018; Vasina et al., 2022). Therefore, haloalkane dehalogenase-like enzymes is a more accurate term for the members of the HLD subfamily III. Prominent examples that have been previously characterized include DrbA from *Rhodospirillum rubrum*, DmbC from *Mycobacterium bovis*, and DmrB extracted from *Mycobacterium rhodesiae* (Fung et al., 2015; Jesenska et al., 2009). While maintaining low dehalogenase activity, there are promiscuous members of the HLD-III subfamily that have been identified. A

notable example of such a putative HLD-III is OleB, an enzyme from *Xanthomonas campestris* that acts as a  $\beta$ -lactone decarboxylase in the biosynthetic pathway of long-chain olefinic hydrocarbons (Christenson, Jensen, et al., 2017).

Attempts in experimental characterization revealed many practical limitations to working with the HLD-III subfamily members, namely low-expression yields, low solubility of the recombinant proteins, multimerization, and high polydispersity of the protein samples (Fung et al., 2015; Jesenska et al., 2009; Koudelakova et al., 2013; Vanacek et al., 2018). These circumstances undeniably contribute to the main challenge of studying the HLD-III enzymes; no high-resolution structure has yet been published for this subfamily of enzymes.

DhmeA is yet another member of the elusive HLD-III subfamily (Chovancova et al., 2007). DhmeA originates from the extremophilic archaeon *Haloferax mediterranei* that inhabits hypersaline water ecosystems, such as salt-water evaporation ponds (Han et al., 2012; Oren & Hallsworth, 2014). Previous experimental characterizations confirmed many of the commonly observed features of the HLD-III family: low-expression yield and poor solubility, oligomerization tendency and polydisperse heterogeneity of the final product, as well as low dehalogenase activity (Vanacek et al., 2018; Vasina et al., 2022). One of its unique features is the melting temperature of 71°C, currently making it the most natively thermostable wild-type HLD (Vanacek et al., 2018).

In this work, we report insight into the structural organization of the DhmeA protein. Collectively, we present an integrative structural biology approach, combining cryo-EM and x-ray crystallography, to obtain the structural snapshots and multimerization potential of an HLD-III enzyme. This was accomplished via optimization of the sample preparation process while considering its halophilic origin. Numerous oligomeric and multimeric states of wild-type DhmeA were visualized by cryo-electron microscopy (cryo-EM), where the protein dimers were repeating units in the building of multimeric assemblies. A similar mode of multimerization was also observed in a 3.3 Å crystal structure of an engineered DhmeA variant (DhmeA<sup>ΔGG</sup>), which adopts a ring-like 20-mer structure. The ring exhibits a highly negatively charged exterior and positively charged interior, where catalytic machinery is located. Additionally, every second monomeric unit of the ring has an exposed catalytic machinery by a partially unfolded cap domain which is responsible for a substrate specificity and substrate entrance. DhmeA<sup>ΔGG</sup> ring anatomy and charge distribution show a certain resemblance to enzymes binding DNA and/or nucleotides (recombinases). Therefore, we

anticipate that the biological function of DhmeA might be the catalysis of bulky, negatively charged substrates. Although there is little known about halogenation/dehalogenation mechanisms in bacteria/archaea, halogenated nucleotides were found even in non-halophilic cells, for example, human cells (Henderson et al., 2001, 2003; Valinluck et al., 2005). Therefore, there might be a need in many organisms for enzymes specialized in dehalogenation of nucleotides/DNA.

## 2 | RESULTS

### 2.1 | Wild-type DhmeA is a multimeric protein with a tendency to form ring-like structures

We used a single-particle cryo-EM technique to probe the multimerization potential of wild-type DhmeA. As shown in Figure 1, the sample of recombinantly produced and affinity-purified wild-type DhmeA is structurally highly heterogeneous, containing oligomeric and multimeric assemblies of various sizes. We thus picked each particle semi-manually and created a dataset containing 6,294 particles in multiple orientations or conformations. After that, similar images were computationally aligned to each other and averaged in order to obtain the images with a better signal-to-noise ratio. Finally, classification was performed to group similar particles into different

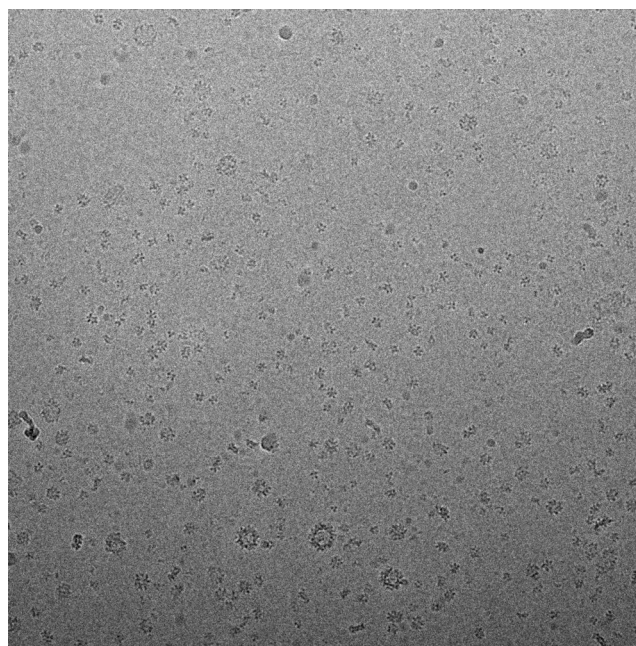


FIGURE 1 Cryo-EM micrograph of full-length wild-type DhmeA, showing structural heterogeneity and high multimerization potential of the protein (magnification: 50,000 $\times$ ).

classes, resulting in 29 different classes of 2D projections of wild-type DhmeA quaternary structure assemblies (Table 1). Among them, we could observe DhmeA homomultimers of various symmetries, ranging from trimers, tetramers, pentamers, hexamers, heptamers, and octamers to high-ordered ring-like supramolecular structures. Collectively, the preliminary cryo-EM imaging showed that the wild-type DhmeA is a multimerization-prone protein that forms diverse and polydisperse macromolecular assemblies, preventing its structural characterization by high-resolution techniques such as x-ray crystallography and single-particle cryo-EM.

## 2.2 | Optimized purification of wild-type DhmeA

We noted that a high yield and solubility of DhmeA positively correlate with increasing ionic strength of the purification buffer. Specifically, we demonstrated that the highest solubility of DhmeA was achieved in the buffer containing 2 M NaCl (Figure 2a), which may reflect its extremophilic origin. However, 2 M NaCl is not only too high for the efficient use of metal affinity chromatography, but it could as well interfere with data collection in cryo-EM. Therefore, in our next experiments, we used the purification buffer with 1 M NaCl as a compromise.

DhmeA was purified as high-molecular-weight and often aggregated multimers when eluted from a Superdex S200 gel filtration column following automated FPLC-based immobilized metal affinity chromatography (IMAC). EM imaging revealed that these complexes of wild-type DhmeA have various sizes and shapes, indicating that the sample was still too heterogenous for further structural analysis (Figure 2b). We found out that the use of automated FPLC-based IMAC contributes to the formation of these highly heterogenous and polydisperse DhmeA complexes due to high pressure. For this reason, we switched to a manual, bench-based gravity flow IMAC setup that provides lower pressure and therefore fewer heterogenous DhmeA complexes of lower molecular weights, better suited for structural characterization (Figure 2b). Biophysical characterization of the wild-type DhmeA complexes purified by this optimized protocol showed a good polydispersity (<20%), proper folding and high thermal stability ( $T_m = 76^\circ\text{C}$ ) (Figure 2c–e). DLS experiments estimated an average particle radius of wild-type DhmeA complexes of  $\sim 5.2$  nm, with an approximate molecular weight (MW) of  $\sim 161.5$  kDa. With the theoretical MW of the monomeric DhmeA,  $\sim 35.8$  kDa, the DLS measurements suggest that DhmeA could form homotetramers (theoretical MW for DhmeA tetramer is  $\sim 143$  kDa). Here, it is important to highlight that the

elution peak upon the gel filtration (Superdex S200 column) is rather broad, indicating the presence of multiple oligomeric species.

The correct folding and the secondary structure of wild-type DhmeA were verified by CD spectroscopy in the far-UV spectral region. The enzyme exhibited similar CD spectra to other previously characterized HLDs harboring the  $\alpha/\beta$ -hydrolase fold (Babkova et al., 2017; Franken et al., 1991; Ollis et al., 1992; Shan et al., 2021). Together, the optimized purification procedure provides wild-type DhmeA enzyme in quantity and quality suitable for structural studies.

## 2.3 | Cryo-EM structure of the wild-type DhmeA complexes

The cryo-EM data collected on optimized DhmeA sample showed particles with tentative threefold symmetry (Figure 3). The initial data analysis revealed significant orientational preference of the particles in the ice. Therefore, we have used a strategy addressing the issue of non-uniform particle orientation distribution by combining data collected at different sample tilt (Tan et al., 2017). The cryo-EM map reconstructed from the major class of the particles present in the data showed six DhmeA molecules arranged into a D3 assembly. The resolution of the final model was limited to 7–8 Å which limited further data analysis. We have tested both focused refinement and focused classification of a single DhmeA cryo-EM density. None of the focused data refinement strategies led to improvement of model resolution which suggests that wild-type DhmeA forms a highly dynamic assembly.

No individual secondary structure elements were clearly defined in the cryo-EM map, with the exception of a central eight-stranded  $\beta$ -sheet that could be unambiguously located. The structure is built from three DhmeA dimers, which are arranged around a 3-fold symmetry axis. The dimerization interface seems to be tightly packed, while inter-dimer contacts are less tight. We therefore deduce that the dimeric unit is a common building block of large assemblies formed by wild-type DhmeA.

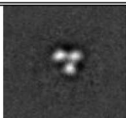
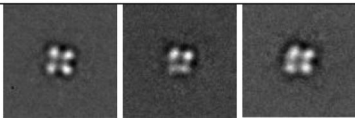
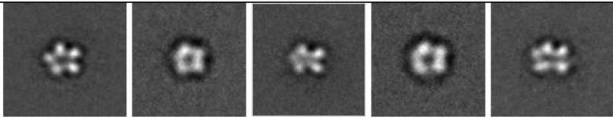
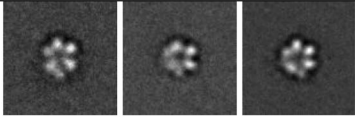
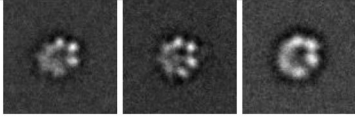
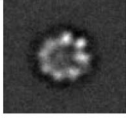
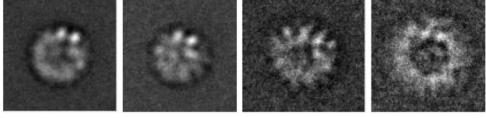
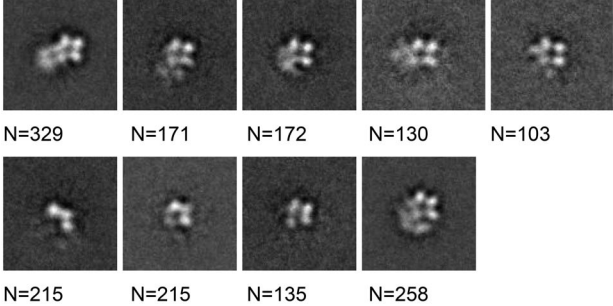
## 2.4 | Engineering the DhmeA<sup>AGG</sup> to favor its crystallization

Since cryo-EM did not yield a high-resolution structure, we initiated crystallization experiments with DhmeA. Attempts to crystallize the full-length wild-type DhmeA yielded only irreproducible, and poorly diffracting

crystals. We thought that this was mainly due to (i) the high heterogeneity of the sample as observed in cryo-EM and (ii) the high conformational flexibility of the DhmeA protein. In our previous work we learned that a helical cap domain is the most flexible and malleable part of the HLD-fold proteins (Markova et al., 2021;

Schenk Mayerova et al., 2023). We noted that the solvent-exposed L10 loop, connecting  $\alpha 4$  and  $\alpha 5'$  helices in the cap domain, displays a unique feature containing two glycine residues (G173, G174) in DhmeA. We hypothesized that this glycine-rich loop could be responsible for the high conformational flexibility of the wild-type

**TABLE 1** Identified classes of wild-type DhmeA quaternary structure assemblies and their populations.

Class	Population	No. particles	%
Trimer	 N=291	291	4.6
Tetramers	 N=384    N=211    N=272	867	13.8
Pentamers (intermediate to hexamer)	 N=429    N=192    N=289    N=169    N=378	1457	23.1
Hexamers	 N=128    N=163    N=342	633	10.1
Heptamers	 N=160    N=157    N=157	474	7.5
Octamers	 N=313	313	4.9
Multimeric states	 N=238    N=203    N=52    N=38	531	8.4
Isoforms	 N=329    N=171    N=172    N=130    N=103 N=215    N=215    N=135    N=258	1728	27.4

DhmeA. We thus constructed a DhmeA mutant lacking these two glycines, hereafter referred to as DhmeA<sup>ΔGG</sup>. The DhmeA<sup>ΔGG</sup> turned out to be more soluble when recombinantly expressed in *E. coli*, and more importantly, it was purified in higher homogeneity than its wild-type counterpart. Importantly, the DhmeA<sup>ΔGG</sup> protein could be reproducibly crystallized. Although diffraction was generally weak, x-ray data were collected, and several full data sets reached up to 3.3 Å resolution.

## 2.5 | Crystal structure of DhmeA<sup>ΔGG</sup> reveals a 20-mer ring-like assembly

The structure was solved by a multi-step molecular replacement procedure (see Methods for details), using a modeled DhmeA<sup>ΔGG</sup> dimer. Due to the large unit cell ( $a = 172.76$  Å,  $b = 289.9$  Å,  $c = 168.31$  Å;  $\alpha = \beta = \gamma = 90^\circ$ ) and the low resolution of crystallographic data ( $\sim 3.3$  Å), the molecular replacement process was challenging, but the availability of the cryo-EM map made this possible. In the first step, we generated a homology model of DhmeA<sup>ΔGG</sup> that was then docked into the low-resolution cryo-EM map, and further optimized with ISOLDE (Croll, 2018) help, to retrieve a homodimeric unit. With the correctly positioned homodimeric model structure, we could then locate all five homodimers in the asymmetric unit through rotational and translation searches. The initial model was further refined by several cycles of manual building and automated refinement. The final model contains 10 DhmeA molecules in the asymmetric unit and has a low deviation from the ideal geometry (Table S1).

The crystal structure of the DhmeA<sup>ΔGG</sup> reveals molecular details of its multimerization potential (Figure 4). Two monomeric units associate with each other, forming a homodimeric unit. Ten homodimers assemble into a ring-like structure. The resulting ring-like structure is built up from 20 DhmeA<sup>ΔGG</sup> molecules. The 20-mer ring displays five-fold symmetry, with an inner diameter of the ring of  $\sim 80$  Å, an outer diameter of  $\sim 200$  Å, and a ring height is  $\sim 95$  Å (Figure 4). The crystallographic ring contains three types of assembly interfaces, namely di-, tetra-, and multimerization interface. The dimerization interface connects two monomeric units into a homodimer. The crystallographic ring represents five-fold symmetry (Figure 4d), meaning that two homodimers can be presented as a homotetramer. The interface between two homodimers is referred to as a tetramerization interface. Five homotetramers finally form a crystallographic ring-like structure via a so-called multimerization interface. One should note that the above-mentioned order of assembling (dimer  $\rightarrow$  tetramer  $\rightarrow$  ring-like structure) does not necessarily represent the actual order of assembling *in vivo*. The supramolecular ring-like arrangement

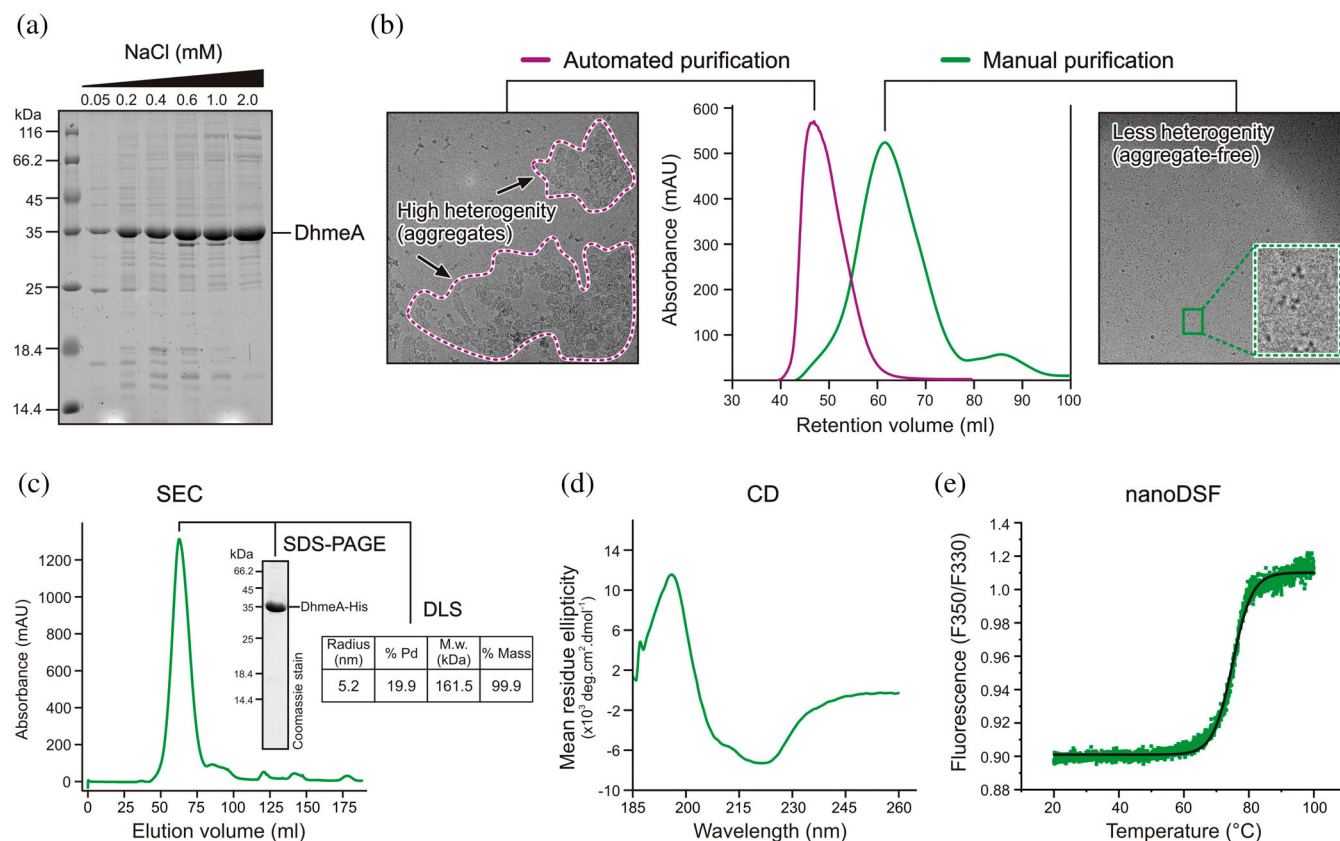
found in the lattice of DhmeA<sup>ΔGG</sup> crystals resembles some multimeric ring-like structures observed by cryo-EM imaging with the full-length wild-type DhmeA (Table 1). We therefore think that the structural features observed in the x-ray structure of DhmeA<sup>ΔGG</sup> reflect to some extent a behavior of the wild-type protein.

## 2.6 | Zooming in on monomeric units: Buried versus exposed active site

A monomeric unit, or a single chain, of DhmeA<sup>ΔGG</sup> canonically consists of two domains: the main  $\alpha/\beta$ -hydrolase core and the helical cap domain, with a typical organization for the HLD family (Chovancova et al., 2007). The main domain consists of eight-stranded  $\beta$ -sheets with a parallel orientation, except the antiparallel oriented  $\beta 2$ . The  $\beta$ -strands are surrounded by eight  $\alpha$ -helices: four on one side ( $\alpha 2$ ,  $\alpha 3$ ,  $\alpha 8$ , and  $\alpha 9$ ) and four on the other side ( $\alpha 0$ ,  $\alpha 1$ ,  $\alpha 10$ , and  $\alpha 11$ ). The cap domain, consisting of the residues 153–216, is formed by five  $\alpha$ -helices ( $\alpha 4$ ,  $\alpha 5'$ ,  $\alpha 5$ ,  $\alpha 6$ , and  $\alpha 7$ ) and five connecting loops. The protein sequence of DhmeA, accompanied by the topology of secondary structure elements, is shown in Figure 5a. The catalytic pentad of DhmeA<sup>ΔGG</sup> consists of five residues, typical for the subfamily III HLDs (Chovancova et al., 2007). These residues are two halide-binding residues (N63 and W130), a nucleophilic aspartate (D129), a catalytic acid (D258), and a histidine base (H286), as depicted in the Figure 5b.

An unusual striking feature of the DhmeA<sup>ΔGG</sup> organization is that the first monomeric unit in the homodimer has a complete, properly folded helical cap domain, while the cap domain of the second monomeric unit is partially unfolded and distorted. Notably, the electron density map between residues R181 and D196, encompassing  $\alpha 5'$  and  $\alpha 5$  helices, is missing, and this part could not be built into the model. Due to this, a neighboring part, encompassing residues L171 and E180, is strongly distorted when compared to the canonical helical cap domain observed in the first monomeric unit. Although the distorted region can be a consequence of the introduced mutation in the cap domain, the correctly folded cap domain of the first monomer gives us confidence that the folding is not completely impaired. Therefore, we suggest that the cap domain of DhmeA protein may naturally unfold, which leads to an uncovering of the active site. As a consequence of this cap domain unfolding, the exposed catalytic machinery underneath could be accessible to bulky and large substrate molecules, which is never possible when the cap domain is fully folded.

As shown in Figure 5b, despite these structural changes in the cap domain, the positioning of catalytic pentad residues is very similar, if not identical, in both



**FIGURE 2** Optimized purification and characterization of wild type DhmeA. (a) Mini-scale expression and purification of DhmeA using buffers with increasing salt concentration (from 0.05 to 2 M NaCl). SDS-PAGE analysis shows that the solubility of DhmeA correlates with increased salt concentration. (b) Automated FPLC-based versus manual gravity flow affinity chromatography effects on quality of DhmeA sample. Size-exclusion chromatography (Sephadex S200 gel filtration column) experiments show a difference in sample sizes. EM analyses of corresponding fractions show that the automated FPLC-based affinity chromatography yields highly heterogenous DhmeA complexes of various sizes and shapes, which is not the case for manual, gravity flow chromatography purification. (c,d) Quality control of DhmeA purified through the optimized manual purification and assayed by SDS-PAGE and DLS measurement (c), CD spectroscopy (d) and differential scanning fluorimetry (e).

monomeric units within the homodimeric association. Moreover, the catalytic pentad residues perfectly superpose with their counterparts from DhaA and DhIA, the well-characterized members of the subfamily I and II HLDs, respectively (Figure S1).

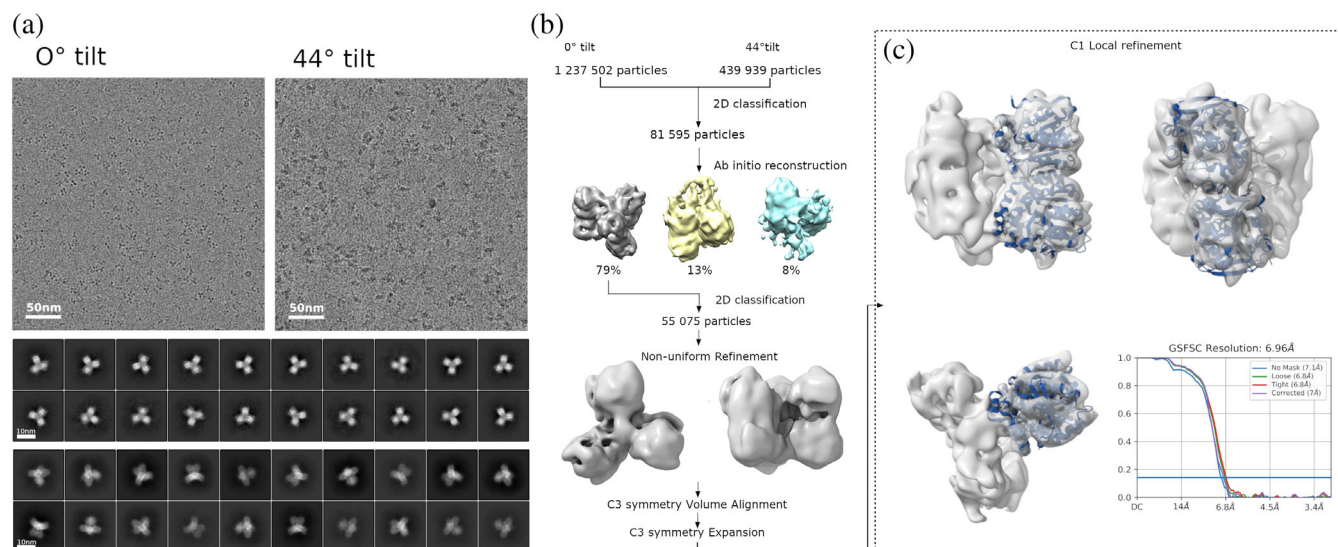
## 2.7 | Molecular details of assembly interfaces

The supramolecular ring-like structure found in the lattice of DhmeA<sup>ΔGG</sup> crystals is composed of 20 monomeric units (chains), built through homodimerization, homotetramerization, and homomultimerization interfaces (Figure 4). The dimerization interface is symmetrical, and it is located at the C-terminal part of the main domain, encompassing α8, α9, β8, L16, and L18, although a small part of the cap domain is involved too (L9 loop). The protein–protein interactions at the dimer interface are formed mainly via H-bonds, hydrophobic and non-polar contacts (Figure 6a).

The detailed interactions at the dimerization interface are comprehensively mapped in Supplementary Note S1.

A homotetrameric assembly of DhmeA<sup>ΔGG</sup> is made by the association of two homodimers, as shown in Figure 6b. The first dimer consists of chains A and B, and the second dimer is from chains C and D. The tetramerization interface is formed between the chains A and C, and the chains B and D. The interactions between A and C chains are identical to those made by B and D chains. The tetramerization interface of DhmeA<sup>ΔGG</sup> encompasses mainly the cap domain region (α4, α5', α5, α6, and L14). The interfacial residues lying inside the main domain are part of α8, L5, and L16. The tetramerization interface is formed via H-bonds and hydrophobic or non-polar interactions (Figure 6b). The detailed description of molecular interactions at the tetramerization interface is provided in Supplementary Note S2.

The multimerization interface is formed between two interacting homotetrameric units (Figure 6c). The first tetramer is built from chains A to D, while the second



**FIGURE 3** Cryo-EM structure of the full-length DhmeA. Cryo-EM data were collected at 0° and 44° stage tilt with 1,237,502 and 439,939 particles extracted, respectively (a). After the 2D classification, 81,595 particles were selected for cryoSPARC *ab initio* reconstruction with three models. The model with 64,146 particles was processed with another 2D classification to balance the distribution of differently oriented particles continued with C1 Non-uniform Refinement. The dataset was subsequently aligned and expanded with C3 symmetry to effectively increase the dataset size (b). The local refinement resulted in the final cryo-EM density of the DhmeA dimer at 7 Å (EMDB ID: EMD-17015) (c).

one contains chains E to H. The oligomerization interface between these two tetramers is formed between the chains C and E, and between the chains D and F as depicted in Figure 6c, where the interactions between C and E are identical to the interactions between D and F. The multimerization interface is secured by the main  $\alpha/\beta$ -hydrolase core, where  $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$ ,  $\alpha 8$ ,  $\alpha 11$ , L2, and L5 play important roles. The interface is formed via H-bonds, salt bridges, hydrophobic, and non-polar interactions with a detailed description in Supplementary Note S3. Finally, five tetramers are assembled into a supramolecular ring-like 20-mer structure through the multimerization interface.

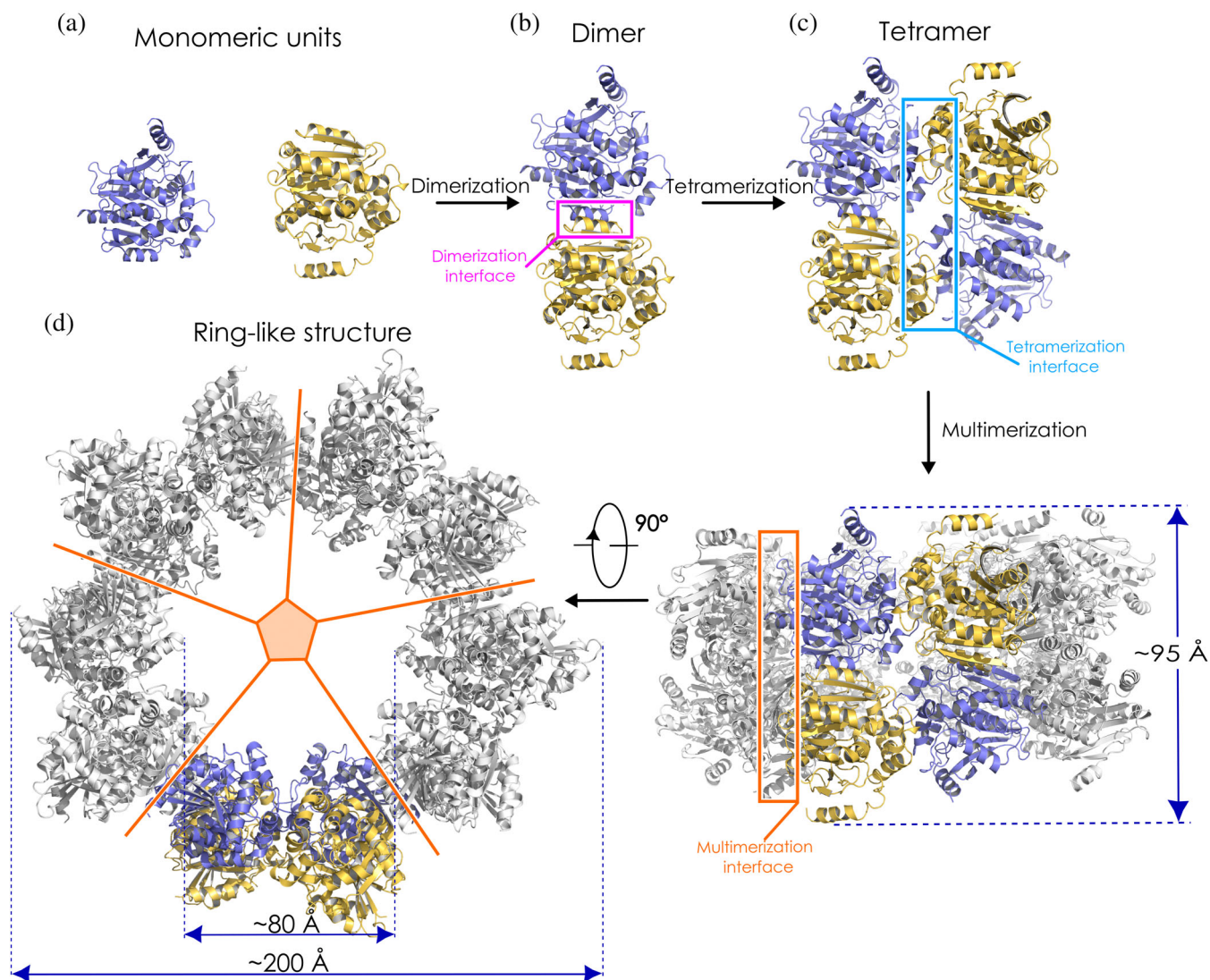
The interaction interfaces were quantitatively analyzed using the PDBePISA server (Krissinel & Henrick, 2007) (Table 2). The percentage of the total solvent accessible area per chain is the highest in the case of the dimerization and the tetramerization interface (6.4% and 6.5%, respectively) followed by the multimerization interface with 4.1% coverage. The solvation free energy gain upon formation of the interface ( $\Delta^i G$ ) is the most negative for the tetramerization interface, approx.  $-20$  kcal/mol, followed by the dimerization interface with the  $\Delta^i G$  value around 0, and the multimerization interface having slightly positive  $\Delta^i G$  (around 2 kcal/mol). The  $p$  value of  $\Delta^i G$  for the tetramerization interface is close to 0, suggesting that no other interface of the observed area may have lower  $\Delta G$ . Therefore, such an interface is a unique spot at the protein surface. The  $p$  values of the dimerization and the multimerization

interfaces are similar, approx. 0.5, meaning that the interfaces are not “surprising” at all.

## 2.8 | Analysis of conservation at self-interaction hot spots

Here we explored the conservation level of residues responsible for the multimerization of the DhmeA protein. To this end, we generated a multiple sequence alignment between DhmeA and its close relatives from the subfamily III HLD, including DhcA from *Hahella chejuensis* KCTC, DmbC from *Mycobacterium bovis* 5033/66, and XcOleB from *X. campestris*. As shown in Figure 7, the catalytic pentad residues are conserved in the subfamily III, while the sequence encompassing the cap domain, the protein part that is responsible for the substrate specificity, is less conserved than the main  $\alpha/\beta$ -hydrolase domain. Although it has been shown experimentally that the relatives from subfamily III also form oligomeric and multimeric assemblies (Jesenska et al., 2009), their assembly interfaces are not known. Therefore, we focused solely on the conservation of the assembly residues in DhmeA<sup>ΔGG</sup>. The residues of the dimerization interface in DhmeA<sup>ΔGG</sup> are not conserved, however, four of them (W157, P158, M256, and L172) have similar physicochemical properties compared to the aligned residues from the other three proteins. The tetramerization interface has one conserved residue (Y102) and several residues with similar physicochemical





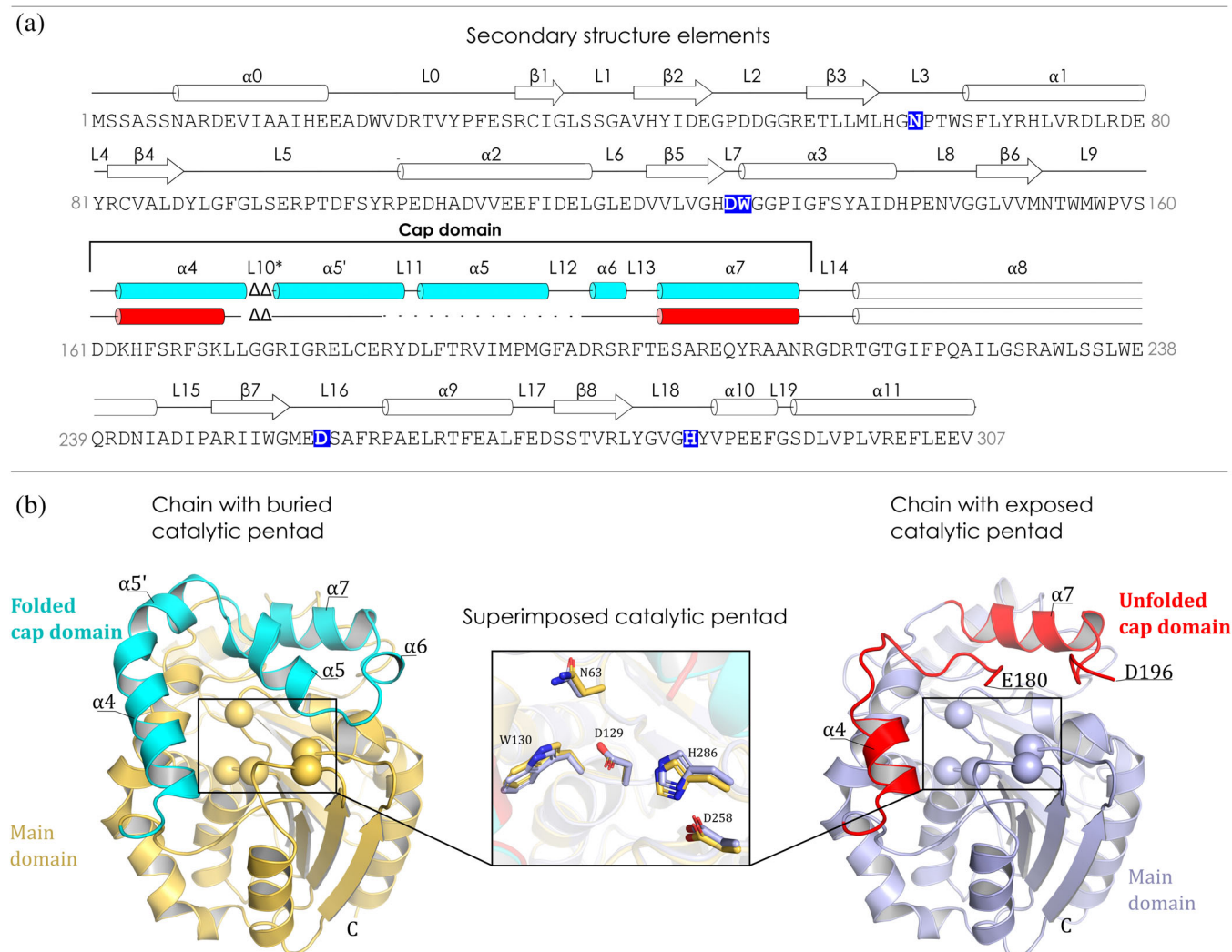
**FIGURE 4** Cartoon representation of the overall crystal structure of DhmeA<sup>ΔGG</sup>. Two monomers (a) form a homodimer (b). Two homodimers associate into a homotetramer (c). Five homotetramers are arranged around fivefold symmetry axis, assembling crystallographic ring-like 20-mer structure (d). The inner diameter of the ring measures ~80 Å, outer diameter ~200 Å, while the height of the ring is ~95 Å. The monomeric units are colored as yellow and blue. PDB ID accession code: [8CKP](#).

properties (F168, L171, L180, Y184, L186, I191, and M194). The multimerization interface is not conserved, but three residues have similar physicochemical properties (D106, P298, and E302).

The sequence and structure of DhmeA were also compared with members of subfamily II whose dimerization interfaces have been already reported (DbjA from *B. japonicum* (Prokop et al., 2010), DbeA from *Bradyrhizobium elkanii* (Chaloupkova et al., 2014), HanR from *Rhodobacteraceae* sp. (Novak et al., 2014), and DmxA from *Marinobacter* sp. (Chrast et al., 2019) and with DpaA from *Paraglaciecola agarilytica* NO2 (Mazur et al., 2021), a member of subfamily I, forming a tetramer (a non-typical behavior for a HLD-I member) (Figure S2). The tetramerization interface of DhmeA<sup>ΔGG</sup> partially

overlaps with the dimerization interface of DpaA in one region (conserved residues: I191 and P193) and vice versa, the dimerization interface of DhmeA<sup>ΔGG</sup> partially overlaps with the tetramerization interface of DpaA in another region (conserved residues: M256 and A264). However, the dimerization interface of DhmeA<sup>ΔGG</sup> does not overlap with the other members of subfamily II. The residues at the multimerization interface of DhmeA<sup>ΔGG</sup> are not conserved between subfamily I and II HLDs.

In addition, we employed HotSpot Wizard (Sumbalova et al., 2018) for a prediction of the putative mutations that might have happened during the evolution of DhmeA. In total, 28 mutation hot spots were identified in DhmeA (Table S2). However, only two of them participate in the interface formation: T266 and D106, forming dimerization



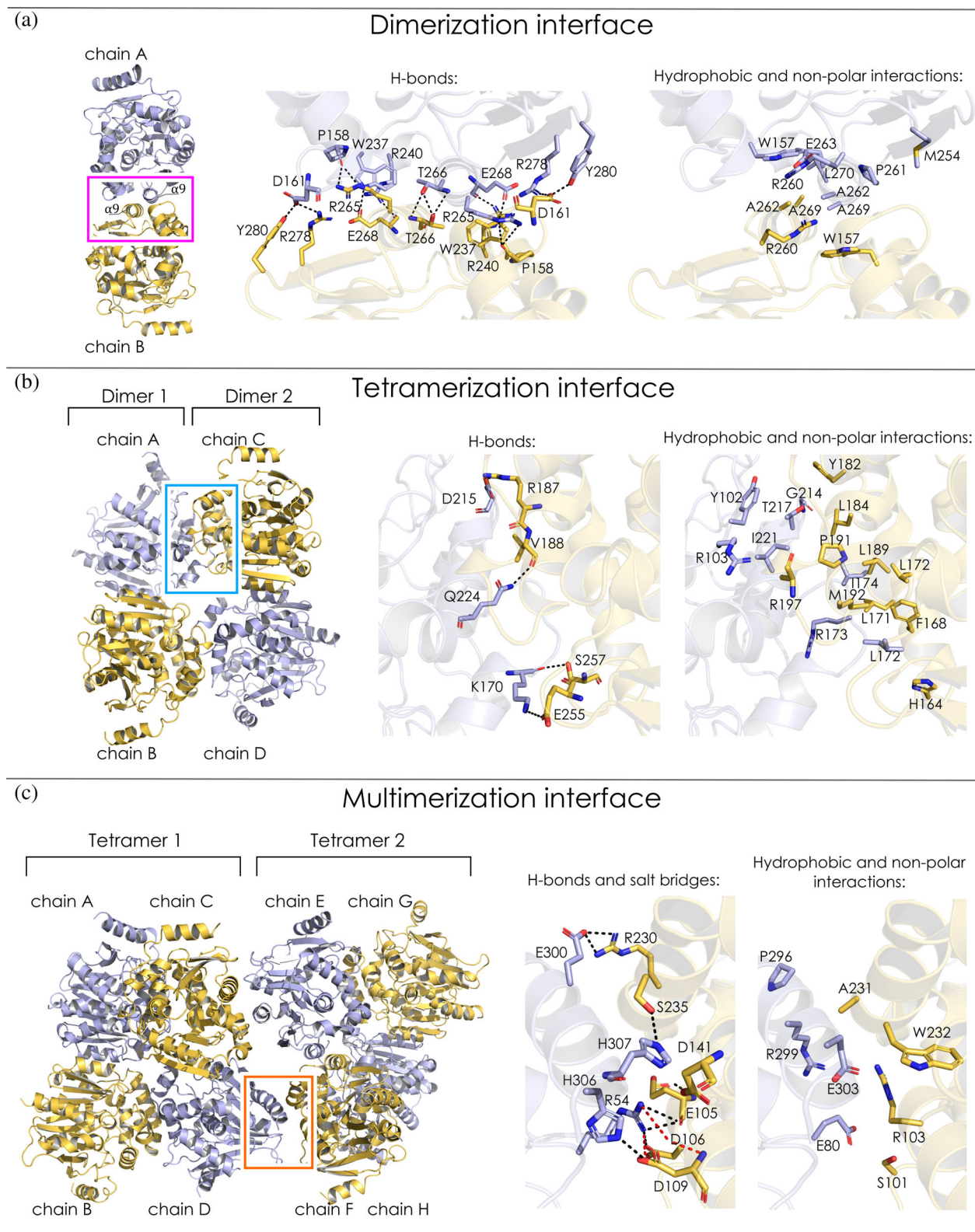
**FIGURE 5** The structural features of DhmeA<sup>AGG</sup> monomeric units. (a) The amino acid sequence of DhmeA<sup>AGG</sup> with the secondary structure elements. The cylinders depict  $\alpha$ -helices, the arrows  $\beta$ -strands, and the lines depict loops (L). The elements of the cap domain are colored in cyan (folded) and red (unfolded). The dotted line in the cap domain region denotes the missing residues of the incomplete cap domain. “L10\*” stands for a putative loop that is not present in DhmeA<sup>AGG</sup> variant. Catalytic residues are marked as white letters on a blue background. (b) Cartoon representation of DhmeA<sup>AGG</sup> monomeric units. The left part of the panel depicts the chain with folded cap domain (cyan) and the catalytic pentad (spheres) being buried in the hydrophobic core of the main domain (yellow). On the right is represented the chain with unfolded cap domain (red) and the exposed catalytic pentad (spheres). The main domain is depicted in light blue. The superimposition of the catalytic pentad from both chains is presented in the middle.

and multimerization interfaces, respectively. The observations from the multiple sequence alignments and HotSpot Wizard suggest that the self-associating interfaces of DhmeA are poorly conserved and they are difficult to predict based solely on the amino acid sequence.

## 2.9 | Electrostatic properties, cap domain unfolding, and consequence for catalysis

We were surprised when we displayed electrostatic potential maps on the surface of the DhmeA<sup>AGG</sup>

structure. The monomeric unit of DhmeA<sup>AGG</sup> shows an extremely negative surface charge distribution on one side, while the other side displays a positively charged surface (Figure S3). The comparison of the DhmeA<sup>AGG</sup> monomer with other HLDs shows a unique surface-charge pattern for DhmeA<sup>AGG</sup> (Figure S3). This intriguing feature becomes even more obvious in the context of the ring-like 20-mer structure. The exterior of the ring together with the multimerization interface has a highly negatively charged surface, while the interior is decorated with a positive charge with the highest positive surface charge density around the tetramerization interface (Figure 8a).



**FIGURE 6** Assembly interfaces of the DhmeA<sup>ΔGG</sup> crystal structure. (a) Molecular interactions at the dimerization interface between chains A and B. (b) Molecular interactions at the tetramerization interface. The tetramerization interface is formed between chains A and C and between B and D. Here, only the interface between A and C is zoomed since B and D form an analogous symmetrical interface pattern. (c) Molecular interactions at the multimerization interface. The multimerization interface is formed between chains C and E, and chains D and F. Here, only the interface between D and F is zoomed for the same reason mentioned in the panel (b). In the zoomed regions the amino acid residues involved in the interface are presented as sticks. The interfaces are formed via H-bonds (black dashed lines), salt-bridges (red dashed lines) and hydrophobic or non-polar interactions.

Interface	Area (Å <sup>2</sup> ) <sup>a</sup>	% of total SASA <sup>b</sup>	Δ <sup>1</sup> G (kcal/M) <sup>c</sup>	p value <sup>d</sup>
Dimerization	808 ± 12	6.4	-0.10 ± 0.94	0.47 ± 0.06
Tetramerization	1,650 ± 30	6.5	-20.2 ± 0.8	0.019 ± 0.006
Multimerization	1,028 ± 30	4.1	2.4 ± 1.8	0.57 ± 0.05

Note: The error was estimated as an average of all corresponding interfaces inside the ring-like structure.

<sup>a</sup>The average interface area.

<sup>b</sup>The percentage of the total solvent accessible area per chain covered by the assembly interface.

<sup>c</sup>The solvation free energy gain upon formation of the interface.

<sup>d</sup>p Value of Δ<sup>1</sup>G.

TABLE 2 Quantitative analysis of DhmeA<sup>ΔGG</sup> assembly interfaces using the PDBePISA web server (Krissinel & Henrick, 2007).

The anatomy and distribution of charges of the DhmeA<sup>ΔGG</sup> 20-mer ring show a certain resemblance to supramolecular ring-like structures of some nucleic-acid-modifying enzymes or proteins that bind DNA, RNA, or nucleotides. As shown in Figure S4, the surface charge distributions of ring-like structures of two recombinases, RadA from *Pyrococcus furiosus* (PDB ID: 1pzn<sup>43</sup>) and human Dmc1 (PDB ID: 2zjb<sup>44</sup>) display similar patterns. Additionally, the multiple sequence alignment of DhmeA with the archaeal recombinases (RadA from *Haloferax volcanii* [RadA<sub>hv</sub>], RadA from *Pyrococcus furiosus* [RadA<sub>pf</sub>], and RadA from *Saccharolobus solfataricus* P2 [RadA<sub>ss</sub>]) shows that approximately 15% of all DhmeA residues are conserved or have similar physicochemical properties (Figure S5).

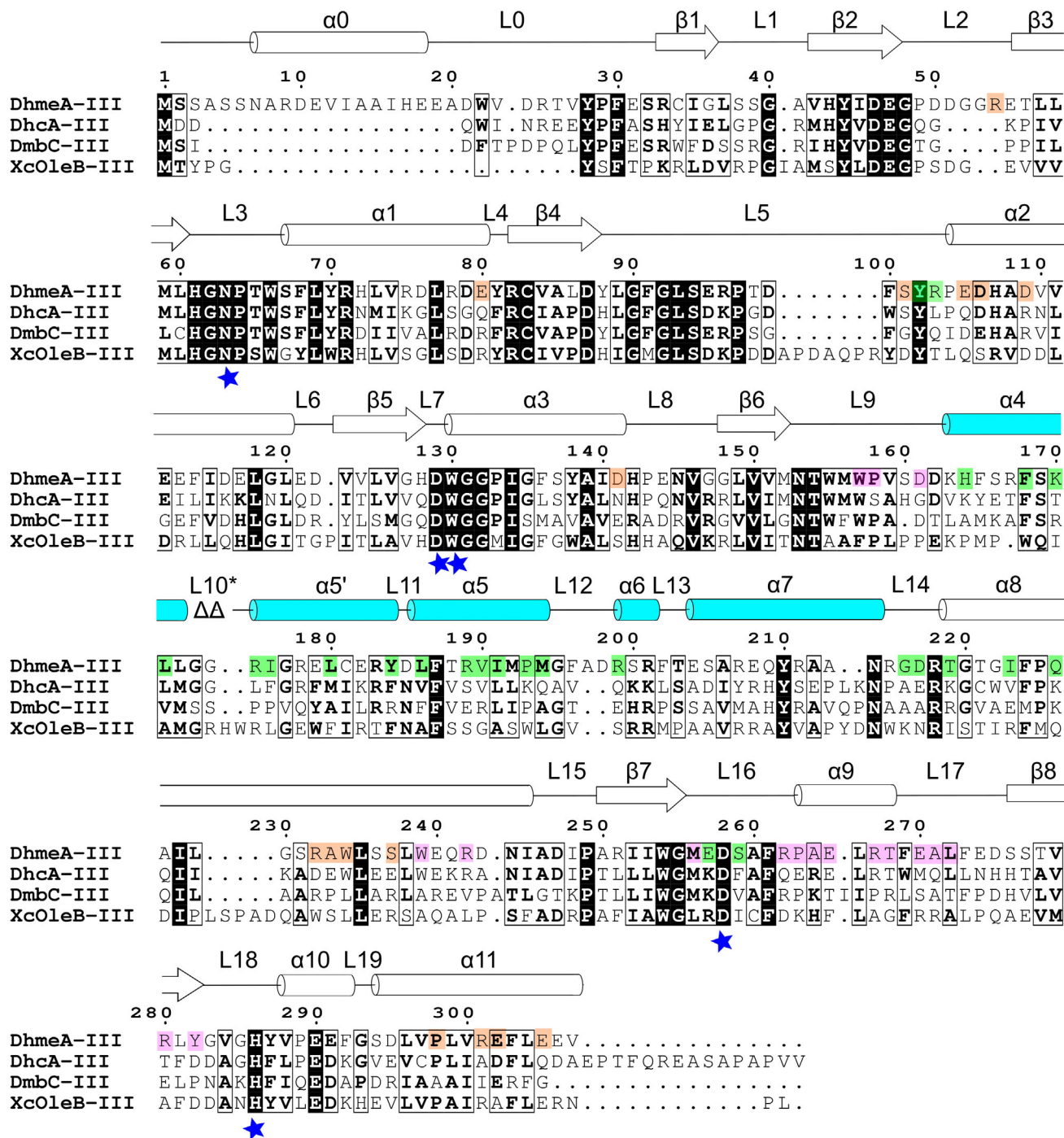
Taken together, the crystal structure of DhmeA<sup>ΔGG</sup> reveals two intriguing features: (i) the mode of self-interactions secures the positively charged enzymatic pocket (Figure 8a), and (ii) the partial unfolding of the helical cap domain exposes the catalytic machinery to the solvent (Figure 8b). Our results thus suggest that DhmeA could accommodate and catalyze processing of negatively charged large substrate molecules, such as for example nucleic acids or their precursors, however, this is only our hypothesis with no actual evidence.

### 3 | DISCUSSION

The members of subfamily III are considered outliers within the HLD enzyme family, based on the following aspects: (i) their sequence annotation is low, (ii) their dehalogenase enzymatic activity is generally low, and (iii) they often form higher-ordered quaternary structures (Vanacek et al., 2018). The mode and purpose of protein multimerization of the subfamily III HLDs are still under debate (Christenson, Jensen, et al., 2017; Christenson, Robinson, et al., 2017; Fung et al., 2015; Vanacek et al., 2018; Vasina et al., 2022). For example, XcOleB enzyme, a subfamily III member from *X. campestris*, has been shown to retain a low-level dehalogenase activity, but it predominantly works as a β-lactone decarboxylase within large multi-enzyme assemblies synthesizing long-

chain olefinic hydrocarbons (Christenson, Jensen, et al., 2017; Christenson, Robinson, et al., 2017). In general, most proteins function as homodimers and/or higher-ordered homo- or hetero-assemblies. Only around one-quarter of all human enzymes are monomeric (Chang et al., 2021; Hochberg et al., 2018). Protein multimerization may have some advantageous effects, among others, for instance, a decrease of surface-to-volume ratio, an increase of protein stability by reducing internal motions, a reduction of denaturation propensity and promiscuous interactions, a decrease of aggregation propensity by blocking aggregation-prone hot-spot regions, or enabling cooperativity (Gwyther et al., 2019; Kiss-Szeman et al., 2022; Lynch, 2013; Marianayagam et al., 2004). East and co-workers showed that monomeric variants of cyclohexadienyl dehydratase exhibit lower catalytic activity compared to trimeric variants, due to conformational restriction, which forces the chains to sample more compact Michaelis conformations in the trimeric form. Some of the variants even exhibited the equilibrium between the dimers and trimers, and the equilibrium was shifted by a small number of mutations at the trimer-forming interfaces (East et al., 2022).

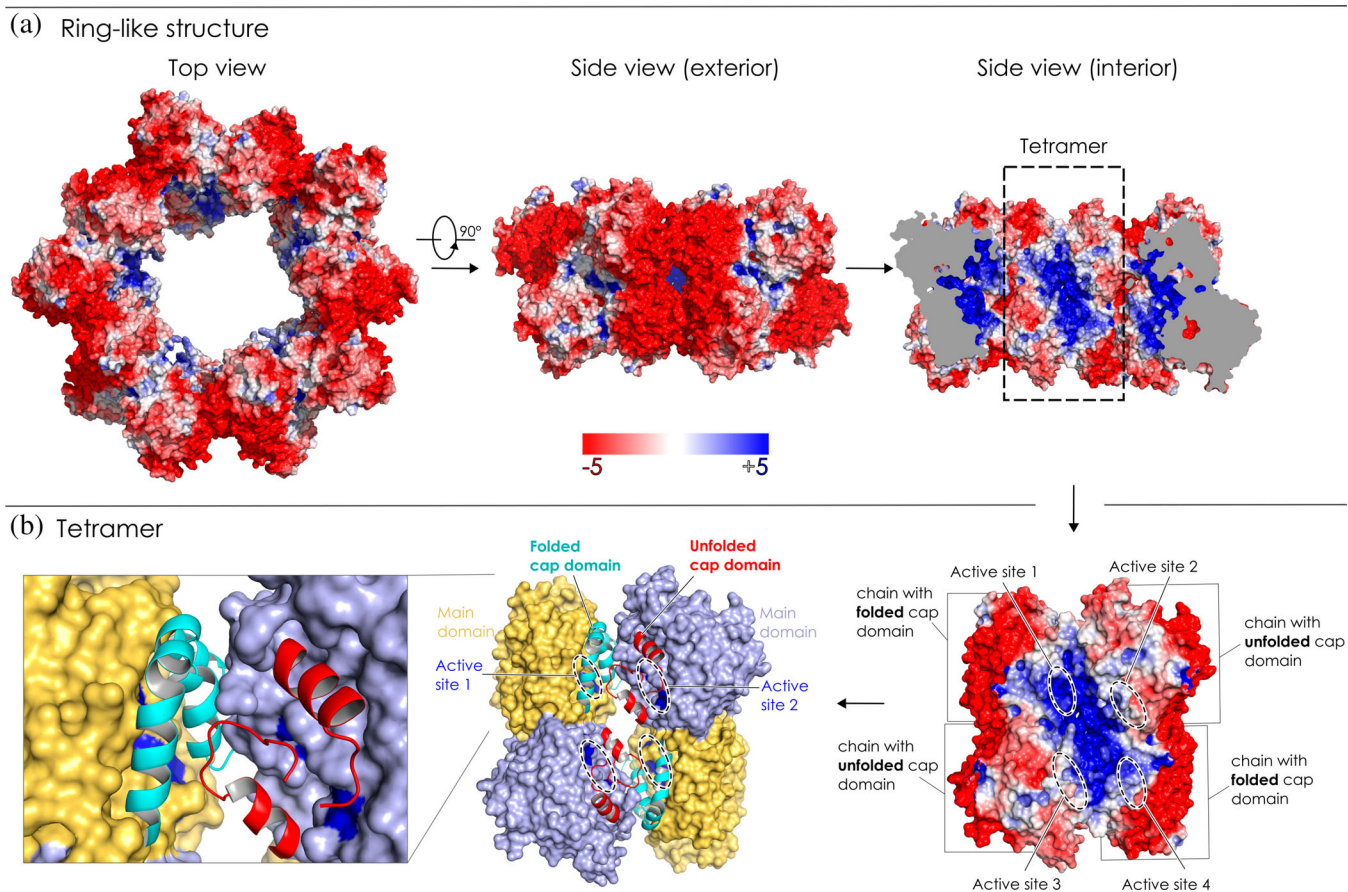
In this work, we probed the structural organization of a prone-to-multimerize DhmeA, a member of the subfamily III HLDs from the halophilic archaeon *H. mediterranei*. To date, the subfamily III is structurally poorly explored as its members have complex quaternary structures. Yet, it has been reported that a homotetrameric association can be a common feature of this multimerization (Christenson, Jensen, et al., 2017; Christenson, Robinson, et al., 2017; Fung et al., 2015). A subfamily III DmrB enzyme from *Mycobacterium* strain JS60 could be previously crystallized, but its crystals diffracted poorly (~8 Å), preventing its high-resolution structural characterization (Fung et al., 2015). Our experiments with DhmeA from *H. mediterranei* showed that wild-type DhmeA self-associates into diverse protein quaternary structures, ranging from small oligomers to large supramolecular ring-like assemblies of various sizes and symmetries. The most intriguing is the latter feature. The formation of macromolecular assemblies displaying various symmetries suggests the existence of a flexible



**FIGURE 7** Sequence comparison between DhmeA from *Haloflex mediterranei* and its close relatives from the subfamily III HLD: DhcA from *Hahella chejuensis* KCTC, DmbC from *Mycobacterium bovis* 5033/66, and XcOleB from *Xanthomonas campestris*. The identical residues are presented as bold-white letters on a black background, and similar residues as bold-black letters on a white background. The secondary-structure elements of DhmeA<sup>AGG</sup> are shown above the sequences with the cap domain colored in cyan. The residues forming di-, tetra-, and multimerization interfaces are marked with the pink, green, and orange background, respectively. The catalytic residues, characteristic for subfamily III, are marked by blue stars. \*L10 is not present in case of DhmeA<sup>AGG</sup>, because of the deletion of two glycines G173 and G174 (marked as ΔΔ). The alignment was generated with MAFFT (Kato & Standley, 2013) and visualized using ESPrnt 3.0 (Robert & Gouet, 2014).

protein element that can structurally re-arrange, enabling self-interactions in diverse symmetric modes. Recently, we demonstrated that the cap domain is the most flexible

and malleable part of HLD-fold proteins (Markova et al., 2021; Schenkmyerova et al., 2023). We, therefore, hypothesized that the cap domain of DhmeA could be



**FIGURE 8** (a) Surface charge distribution of DhmeA<sup>ΔGG</sup>. Ring structure from the top (left panel) and from the side (exterior of the ring—middle panel and interior of the ring—right panel). Red color denotes negative charge and blue, positive charge. (b) Tetramer of DhmeA<sup>ΔGG</sup> as seen from interior of the ring. Right panel: Surface charge distribution. Middle panel: Tetramer with the cap domain depicted as a cartoon (cyan for folded and red for unfolded cap domain) and main domain as a surface (yellow for a chain with folded and light blue for a chain with unfolded cap domain). Dark blue surface represents the catalytic pentad. Left panel: the zoomed view of two cap domains and catalytic pentad.

this conformationally rich enigmatic element. Indeed, this hypothesis was verified by the construction of the DhmeA<sup>ΔGG</sup> mutant, where we removed two glycine residues from the solvent-exposed L10 loop. While crystallization experiments with the wild type DhmeA have been shown to be challenging, the DhmeA<sup>ΔGG</sup> protein crystallized reproducibly and provided diffraction-quality crystals. However, the crystals diffracted poorly (up to 3.3 Å) and the unit cell was large ( $a = 172.76$  Å,  $b = 289.9$  Å,  $c = 168.31$  Å;  $\alpha = \beta = \gamma = 90^\circ$ ), which complicated structure determination by molecular replacement. Key in this process was a combination of cryo-EM data and the use of molecular modeling and advanced molecular dynamics simulations implemented in ISOLDE (Croll, 2018). The low-resolution cryo-EM reconstructions ( $\sim 7$  Å) allowed us to find orientations of two protomers in the homodimeric unit that was then successfully used for comprehensive multi-cycle molecular replacement searches, to locate all 10 DhmeA<sup>ΔGG</sup> molecules in the

asymmetric unit. The crystal structure of DhmeA<sup>ΔGG</sup> reveals a mode of its multimerization, where a homodimeric unit is a key repeating building block. The two homodimers self-interact to form a homotetrameric association, and then five homotetramers assemble into a 20-mer ring-like structure. Note, this sequential order of the assembling presented here might not truly represent the assembly order *in vivo*.

The DhmeA protein is encoded by *H. mediterranei*, the halophilic archaeon that inhabits environments with high salt (1–4 M) concentrations (Cui et al., 2017; Han et al., 2012; Oren & Hallsworth, 2014). Therefore, the cellular components of the halophilic archaeon must be adapted to extremely high salt concentrations. It has been shown that halophilic proteins have adapted to these extremes by exhibiting the negatively charged amino-acid residues on their surface, which allows the binding of water and salt to build up a hydrated network on the surface of the proteins (Lanyi, 1974). Structural studies have

shown that the binding of the hydrated cations (provided by the excess of salt and water molecules) around the negatively charged residues on the protein surface reduces the electrostatic repulsion. At lower concentrations of salt, the protective effect is lost, and the repulsive forces between the acidic residues lead to the unfolding and the inactivation of the protein (Lanyi, 1974). Here we demonstrate that the halophilic DhmeA is no exception. The expression trials with DhmeA showed that the solubility and yield are positively correlated with increasing ionic strength of the purification buffer.

The exterior of DhmeA<sup>ΔGG</sup> 20-mer ring-like structure is negatively charged, while its interior is decorated with a positive charge. Notably, the structural organization of the DhmeA<sup>ΔGG</sup> supramolecular ring shows a resemblance to enzymatic ring-like assemblies that bind DNA and/or nucleotides, such as DNA recombinases (Hikiba et al., 2008; Shin et al., 2003). Recently, a supramolecular ring-like assembly of RadA-RadB DNA recombinase complex from *Haloflex volcanii*, a moderate halophilic archaeon, was visualized by transmission electron microscopy (Patoli et al., 2017). However, its atomic-level structure is not known yet. Our structural observations suggest that multimeric DhmeA could be involved in catalytic modifications of large, negatively charged substrate molecules (DNA, RNA, or nucleotides), although there is no experimental evidence for this. We know very little about the halogenation/dehalogenation processes of DNA (RNA) molecules in halophilic organisms. However, it is known that the halogenation of pyrimidines in DNA occurs even in non-halophilic cells, e.g., human cells, during peroxidase-mediated inflammatory processes (Henderson et al., 2001, 2003; Valinluck et al., 2005). Therefore, we anticipate that there might exist in many organisms a need for some halogenation enzymatic erasers.

The structure of DhmeA solved in this work is the first experimentally determined structure of any subfamily III HLD enzyme at atomic resolution. We believe that the strategy presented herein for structural characterization of multimerization-prone protein, combining cryo-EM and x-ray crystallography, will serve as a template for future structural studies.

## 4 | CONCLUSIONS

In this work, the structural organization of DhmeA, a subfamily III HLD-like enzyme from the archaeon *H. mediterranei*, was determined through a combination of single-particle cryo-EM and x-ray crystallography. After optimization of sample preparation steps, three-dimensional reconstructions of oligomeric species were acquired with ~7 Å resolution by single-particle cryo-

EM. Due to the low-resolution gained by cryo-EM, we engineered a crystallizable DhmeA mutant (DhmeA<sup>ΔGG</sup>). The crystal structure was determined due to the availability of cryo-EM reconstruction, which allowed us to place two protomers in the homodimeric unit that was then applied in multi-cycle molecular replacement searches, to locate all protomers in a large asymmetric unit. The 3.3 Å crystal structure revealed that DhmeA<sup>ΔGG</sup> forms a ring-like 20-mer structure, with an outer and inner diameters of ~200 and ~80 Å, respectively. The supramolecular ring displays a negatively charged exterior, while its interior part harboring catalytic sites was positively charged. Localization and exposure of catalytic machinery suggest a processing of large, negatively charged macromolecular substrates.

## 5 | MATERIALS AND METHODS

### 5.1 | Molecular cloning and mutagenesis

For cryo-EM studies, two expression vectors including full-length (wild-type) *dhmeA* were generated. The *dhmeA* sequence was cloned between the *NdeI* and *BamHI* restriction sites of pET21b (Sigma-Aldrich, 69741) and pNEA/3CH<sup>56</sup> vectors. The pET21b::*dhmeA* codes the protein tagged by a carboxy-terminal hexahistidine (His<sub>6</sub>) tag, while the protein expressed from the pNEA/th3CH-*dhmeA* codes the wild-type DhmeA enzyme with an attached 3C protease cleavage site, followed by a His<sub>6</sub> tag sequence, at its C-terminal end. The latter plasmid, pNEA/th3CH-*dhmeA*, was used to produce the wild-type DhmeA for cryoEM analyses.

To stabilize DhmeA for a crystallographic study, a sequence coding two glycine residues (G173 and G174) was deleted from a putative solvent-exposed loop L10 in the *dhmeA* sequence. This was achieved via two-step fusion PCR-based mutagenesis, using two oligonucleotides shown in Table S3. The resulting PCR product *dhmeA*<sup>ΔGG</sup> was cloned into the pET21b plasmid between the *NdeI* and *BamHI* restriction sites. It is important to note that the mutant *dhmeA* vector also contained the sequence coding for the C-terminal poly-histidine (His<sub>6</sub>) affinity purification tag. The final, error-free expression plasmid pET21b::*dhmeA*<sup>ΔGG</sup> was verified by DNA sequencing.

### 5.2 | Mini-scale protein expression assays

The chemo-competent *Escherichia coli* BL21(DE3) cells (New England Biolabs, USA) were transformed with pNEA/3CH-*dhmeA* by heat shock. Transformed bacteria were seeded in a 6-well plate with Luria-Bertani (LB;

Sigma-Aldrich, USA) agar supplemented with 100 µg/mL ampicillin. The plates were incubated overnight at 37°C. Ampicillin-resistant clones were selected and inoculated into a 24-well plate containing 2 mL of 2 × LB (40 g LB/L), 0.5% glucose, and 100 µg/mL ampicillin. After 6 hours of incubation at 37°C and 275 rpm, the expression of the recombinant protein was induced by the addition of 2 mL of 2 × LB medium with 0.6% lactose, 0.5 mM IPTG, 20 mM HEPES with pH 7.4, and 100 µg/mL ampicillin. Plates were subsequently incubated at 22°C overnight (275 rpm). Cells were harvested by centrifugation (3700 g, 10 min) and re-suspended in a lysis buffer (10 mM Tris-HCl with pH 8.0, 5 mM imidazole) with various concentrations of NaCl: 0.05, 0.20, 0.40, 0.60, 1.00 or 2.00 M. The re-suspended cultures were lysed by sonication, and the lysates were centrifuged (20,000 g, 20 min). The supernatants were loaded onto separate 25 µL Talon Superflow Metal Affinity Resins (Clontech, USA), pre-equilibrated in the respective lysis buffers. The samples were subsequently incubated on a roller shaker (4°C, 2 h). After incubation, each resin-bound sample was washed twice with the buffer volume 1.2 mL during the centrifugation (130 g, 2 min) with the corresponding lysis buffer. The resins were re-suspended in 40 µL of Laemmli buffer (50 mM Tris-HCl at pH 6.8, 2% SDS, 4% β-mercaptoethanol, 30% glycerol, 0.001% Coomassie Brilliant Blue R250), and the eluted proteins were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

### 5.3 | Protein overproduction and unoptimized protein purification

The chemo-competent *E. coli* BL21(DE3) cells were transformed with the pET21b::dhmeA expression plasmid using heat shock. A single ampicillin-resistant colony was selected and inoculated into 10 mL of LB medium (100 µg/mL ampicillin). The preculture grew 6–8 h (37°C, 150 rpm). Prior to the cultivation the EnPresso B medium (BioSilta, Finland) was carefully prepared and pre-tempered based on manufactural protocol: five tablets of EnPresso B medium were added into 250 mL of sterile Milli-Q water in an Ultra Yield Flask (Ukkonen et al., 2011) and dissolved. Then ampicillin (in final concentration 100 µg/mL), antifoam agent Struktol SB2020 and 125 µL of the Reagent A (glucoamylase) were added. Finally, 10 mL of preculture was used to initialize cultivation (30°C, 250 rpm). After 15–18 h cultivation, five booster tablets were added, together with 375 µL of Reagent A and 125 µL of IPTG. After 24 h of expression, cells were harvested using centrifugation (3700 g, 10 min, 4°C) and washed by 20 mM phosphate buffer (pH 7.5)

with 10% glycerol. The pellets were resuspended in purification buffer A (20 mM potassium phosphate, 20 mM imidazole, 0.5 M NaCl, pH 7.5), then homogenized by the sonication and finally centrifuged (20,000 g, 1 h, 4°C) to separate DhmeA from insoluble proteins and cell debris.

The supernatant of wild-type DhmeA enzyme with the His<sub>6</sub> tag was purified in a HiTrap HP chelating column charged by Co<sup>2+</sup> (Qiagen, Germany), pre-equilibrated by the purification buffer A, and attached to the FPLC purification system (Bio-Rad, USA). The protein was eluted from the column by purification buffer B (20 mM potassium phosphate, 500 mM imidazole, 0.5 M NaCl, pH 7.5) using a combination of linear-gradient (0%–30% of purification buffer B) followed by a gradient with isocratic steps (30%, 60%, and 100% of purification buffer B). The target protein was eluted in 60% of buffer B, which corresponds to 300 mM imidazole.

Finally, the protein samples were concentrated and dialyzed against Tris-HCl buffer (10 mM, pH 8) overnight. The molecular weight and homogeneity of collected fractions were further assessed by SDS-PAGE.

### 5.4 | Optimized preparation of wild-type DhmeA for cryo-EM

*E. coli* BL21(DE3) cells were transformed by the heat shock method with the plasmid pNEA/3CH-dhmeA. Ampicillin-resistant colonies were inoculated in 10 mL of LB medium (1xLB medium, ampicillin 100 µg/mL), which was incubated (200 rpm) overnight at 37°C. The next day, the bacterial culture was used to inoculate 5-L Erlenmeyer flasks containing 1 L of 2 × LB medium with ampicillin (100 µg/mL) where cells were grown (200 rpm, 37°C) until the culture reached OD<sub>600</sub> = 0.8–1.2. Induction of expression was done at 22°C by adding 0.5 mL of 1 M IPTG, and the culture was then incubated overnight (typically 12–16 h) at 22°C and 200 rpm. The next day, the bacterial biomass was harvested by centrifugation (3,600 g, 15 min, 4°C) and re-suspended in a lysis buffer (1 M NaCl, 10 mM Tris-HCl pH 8.0) and lysed by sonication (50% amplitude, 6 min) using a sonicator Ultrasonic Processor Hielscher UP200S (Germany). The sonicated lysate was clarified by centrifugation (20,000 g, 55 min, 4°C). The supernatant was loaded onto Talon Superflow Metal Affinity Resin (Clontech, USA) pre-equilibrated with the lysis buffer. The His<sub>6</sub>-tagged DhmeA protein was released from the Talon resin by overnight treatment with home-made 3C protease (Marek et al., 2021) (60 µg/mL), and subsequently loaded onto a 16/60 Superdex 200 gel filtration column (GE Healthcare, UK) pre-equilibrated with the lysis buffer supplemented with



0.5 mM TCEP (Hampton, USA). The recombinant DhmeA protein was concentrated with an Amicon Ultra centrifugal filter unit (Millipore, USA), and protein concentration was assayed by a DS-11 Spectrophotometer (DeNovix, USA).

## 5.5 | Optimized preparation of DhmeA<sup>ΔGG</sup> for crystallization

*E. coli* BL21(DE3) cells were transformed with the recombinant plasmid pET21b::dhmeA-ΔGG, plated on agar plates with ampicillin (100 μg/mL), and grown overnight at 37°C. The obtained colonies were used for the inoculation of 10 mL of LB medium with ampicillin (100 μg/mL). The cells grew overnight at 37°C and 200 rpm. The overnight culture (4–5 mL) was used to inoculate 1 L of 2 × LB medium containing ampicillin (100 μg/mL). The culture was incubated at 37°C and 150 rpm for 5–6 h, up to the point of reaching OD<sub>600</sub> > 1.5. Prior to induction, the culture was cooled to 20°C for 30 min. The expression of DhmeA<sup>ΔGG</sup> was induced by the addition of IPTG with a final concentration of 0.5 mM. After overnight incubation at 20°C and 150 rpm, the cells were harvested by centrifugation (20 min; 3,300 g, 4°C) and immediately resuspended in a purification buffer A (1 M NaCl, 10 mM imidazole, 10 mM Tris-HCl, pH 8). Sonication was performed to lyse the cells and the cell-free supernatant was extracted after centrifuging at 4°C, 20,000 g for 1 hour.

The supernatant containing DhmeA<sup>ΔGG</sup> was used for the purification by metallo-affinity chromatography via a C-terminal His<sub>6</sub>-tag, using the 5 mL Ni-NTA Superflow column (Qiagen, Germany) attached to the FPLC purification system BioLogic DuoFlow (Bio-Rad, USA). A gradient of 0%, 10%, 60%, and 100% of purification buffer B (1 M NaCl, 500 mM imidazole, and 10 mM Tris-HCl, pH 8) was used for the elution process. The target protein was eluted by 60% of the buffer B, corresponding to 300 mM imidazole.

Protein fractions eluted at 60% of buffer B were collected for further separation by GPC. The samples were loaded onto a HiLoad™ 16/600 Superdex™ 200 column (GE Healthcare, Sweden) on an Äkta Purifier FPLC system (GE Healthcare, Sweden). The column was pre-equilibrated by separation buffer (1 M NaCl, 0.5 mM TCEP, and 10 mM Tris-HCl, pH 8), and the constant flow rate was set at 1 mL/min. The protein eluted in two major peaks: the first peak contained high-molecular-weight complexes, while the second peak contained lower-molecular-weight complexes. The fractions containing the latter peak were collected, concentrated with an Amicon Ultra centrifugal filter unit (Merck Millipore Ltd), and re-separated with GPC to yield a

monodisperse protein sample suitable for crystallization experiments. The second GPC step was essential to obtain crystallization-quality protein. The enzyme concentration was determined by DS-11 Spectrophotometer (DeNovix, USA) with an extinction coefficient of DhmeA 1.73 mL (mg cm)<sup>-1</sup> at 280 nm, and the product purity was validated by SDS-PAGE.

## 5.6 | Differential scanning fluorimetry

The thermal stability of the DhmeA (2 mg/mL) in a buffer containing 1 M NaCl, 10 mM Tris (pH 8.0), and 0.5 mM TCEP was analyzed by a label-free differential scanning fluorimetry (DSF) using a Prometheus NT.48 instrument (NanoTemper Technologies, Germany). The temperature-triggered unfolding was measured in the temperature range of 20–100°C. The measurements were carried out in nanoDSF-grade high-sensitivity glass capillaries (NanoTemper Technologies, Germany) at a heating rate of 1°C/min. Melting temperatures (*T*<sub>m</sub>) were inferred from the first derivative of the ratio of tryptophan fluorescence emission intensities at 330 and 350 nm.

## 5.7 | Circular dichroism spectroscopy

Circular dichroism (CD) spectra of the DhmeA (2 mg/mL) in a buffer containing 1 M NaCl, 10 mM Tris (pH 8.0), and 0.5 mM TCEP were recorded at room temperature using a Chirascan CD Spectrometer equipped with a Peltier thermostat (Applied Photophysics, UK). CD data were expressed as mean residue ellipticity ( $\Theta_{MRE}$ ).

## 5.8 | Dynamic light scattering

Protein solutions were centrifuged (16,000 g, 10 min) prior to dynamic light scattering (DLS) measurement to remove the impurities. DLS experiments were conducted with wild-type DhmeA solution (1 mg/mL) in a buffer containing 1 M NaCl, 10 mM Tris (pH 8.0), and 0.5 mM TCEP, using the Delsa Max Core (Beckman Coulter, USA).

## 5.9 | Cryo-EM data acquisition

The full-length DhmeA (3.5 μL, 0.035 mg/mL) was applied to freshly glow-discharged (Quorum, SC7620 Sputter Coater) TEM grids (Quantifoil, Cu, 200 mesh, R2/1) and vitrified in liquid ethane using a Vitrobot mark

IV (Thermo Scientific) with single blotting, 15 s wait time,  $-2$  blot force, 4.5 s blot time and no drain time. Grids were subsequently mounted into Autogrid cartridges and loaded to the FEI Tecnai F20 and Titan Krios (Thermo Fisher Scientific) transmission electron microscopes for screening and data acquisition, respectively.

Two datasets with the sample tilt of  $0^\circ$  and  $44^\circ$ , respectively, were collected for the data analysis using the Titan Krios microscope. The experimental details of the data acquisition are summarized in Table S4.

## 5.10 | Cryo-EM data analysis

Both data sets were analyzed using cryoSPARC (Punjani et al., 2017) software. MotionCor2 (Zheng et al., 2017) and Gctf (Zhang, 2016) were used for motion correction and CTF estimation as it is implemented in cryoSPARC. First, particles from 20 micrographs were manually selected and classified to generate templates for template picking (cryoSPARC). The overall number of 1,237,502 and 439,939 particles was extracted from the data collected at  $0^\circ$  and  $44^\circ$  stage tilt, respectively. After 2D classification, 81,595 particles were selected for cryoSPARC *ab initio* reconstruction with 3 models followed by multiple rounds of 3D classification (heterogeneous refinement job). Finally, 64,146 particles were processed for Non-uniform Refinement (cryoSPARC) with a final resolution of 7 Å.

## 5.11 | Preparation of structural models

Models for cryo-EM reconstruction were prepared for the docking calculations using the same workflow as for molecular replacement. A search with HHpred (Zimmermann et al., 2018) was used to find and align the sequences of homologs of known structure. Using this alignment, the program Sculptor (Bunkóczi & Read, 2011) was used to edit the top three hits from HHpred, PDB entries 3g9x (Stsiapanava et al., 2010), 4f0j, and 4psu, to truncate non-identical side chains and remove unaligned portions of the structure. Then the program Ensembler (Bunkóczi & Read, 2011) was used to superimpose the three sculpted models and to trim regions deviating among the structures to leave an ensemble model of the conserved core.

As an alternative, the Robetta server was used to construct a model of DhmeA, using the trRosetta deep-learning algorithm (Yang et al., 2020). The first 25 residues of this model were judged to be unreliable and were omitted, as the modeling assigned them expected RMS errors ranging from 2 to 12 Å.

## 5.12 | Docking models into cryo-EM reconstruction

Docking calculations were performed using the real-space molecular replacement features of Phaser (McCoy et al., 2007) providing the Fourier terms corresponding to the cryo-EM reconstruction as target structure factors. In this approach, the Phaser likelihood-based rotation function is computed to obtain a set of trial orientations, then the phased translation function (Colman et al., 1976; Read & Schierbeek, 1988) finds the translation of the oriented model that maximizes its correlation with the target map.

Initial attempts to place the ensemble model into the full cryo-EM reconstruction failed to yield a clear answer. To reduce the noise in the calculation, the program phenix.map\_box (Lieschner et al., 2019) was used to extract the unique portion of the map representing a dimer. When this was used as the target map, two unambiguous placements were found, generating a dimer model. A full hexamer was then generated by applying the remaining threefold symmetry to the dimer model.

## 5.13 | Optimizing the hexamer model to fit the cryo-EM reconstruction

The trimmed ensemble comprises only the conserved core of wild-type DhmeA, containing only 126 of the 307 residues in the full-length protein. A more complete model could be constructed from the model derived using Sculptor from the top HHpred hit, PDB entry 3g9x, which contains 279 residues that match in the sequence alignment. The additional residues largely fit within the reconstruction. A rigid-body optimization in ChimeraX (Goddard et al., 2018) of the correlation between one chain of the model and the cryo-EM reconstruction gave a correlation coefficient of 0.903. However, a hexamer constructed from the trimmed trRosetta model provided an even better fit to the map, with a correlation coefficient of 0.936. Subsequent work therefore focused on the trRosetta model.

ISOLDE (Croll, 2018) was used to modify the hexamer model to optimize the fit to the cryo-EM reconstruction. First, residues making intractable severe clashes between chains (W239 and L272) were pruned back to the beta-carbon. Clashes between R262 and the opposing D161 were resolved by adjusting the rotamers. Overfitting was avoided by applying torsion restraints to maintain the backbone geometry of all  $\alpha$ -helices and  $\beta$ -strands. During the fitting simulation, the temperature was gradually reduced from 100 to 0 K. The most obvious result of the fitting was an increase in the spacing

between monomers by about 2.7 Å (centroid-centroid distance), combined with subtle changes to the relative positions of secondary structure elements.

### 5.14 | Crystallization of DhmeA<sup>ΔGG</sup> enzyme

A freshly purified sample of DhmeA<sup>ΔGG</sup> was concentrated to 12–14 mg/mL and set up for crystallization in 24-well crystallization plates (Hampton Research, USA) at 10°C using the sitting-drop vapor-diffusion method. The crystallization mother liquor contained 0.1 M MES/imidazole buffer system (pH 6.5), 0.1 M amino acids, 10% PEG 4,000, and 20% glycerol, originating from the Morpheus screen (Molecular Dimensions Ltd, UK). Each drop consisted of 2 μL of the enzyme and 2 μL of the mother liquor equilibrated against 500 μL of the reservoir solution. The crystal growth was monitored by an Olympus-SZX10 (Olympus) stereomicroscope. The crystals formed within 2–5 days. The obtained crystals were fished out and cryo-cooled in liquid nitrogen.

### 5.15 | X-ray data processing

Diffraction data were collected at the PXIII beamline at SLS Synchrotron (Villigen, CH) at a wavelength of 0.999 Å. Multiple data sets from the best diffracting crystals were processed using XDS (Kabsch, 2010) Version February 5, 2021, scaled via XSCALE, and merged by XDSCONV. The selection of datasets that were merged was guided by the program XDSCC12 (Assmann et al., 2020).

### 5.16 | Molecular replacement calculations

Based on considerations of solvent content (Kantardjieff & Rupp, 2003; Matthews, 1968), up to 12 copies of the DhmeA<sup>ΔGG</sup> monomer were expected in the asymmetric unit of the crystal. Molecular replacement searches with Phaser (McCoy et al., 2007) failed with all monomer models, including the individual models processed with Sculptor, the trimmed ensemble model, and the trRosetta model. This was not surprising given the relatively low resolution, the limited sequence identity of available templates, and the small fraction of the asymmetric unit content accounted for by these models. The hexamer model constructed by docking the trRosetta model into the cryo-EM reconstruction was tested next, searching for two copies, but this also failed

to yield any convincing solutions. Given the possibility that the hexamer could sit on a crystallographic 2-fold axis in the space group C2221, a search for four copies of a trimer model was also attempted, again unsuccessfully.

It was noted that the deletion of glycine residues G173 and G174 would be located near the 3-fold axis of the hexamer, raising the possibility that the quaternary structure had been perturbed. Therefore, a dimer extracted from the hexamer was tested as a potential molecular replacement model. This failed initially, but when a dimer was extracted from the hexamer model that had been optimized in ISOLDE, a search for six copies gave a solution in which the first five dimers were placed with excellent signal, but the placement of the sixth copy failed to increase the log-likelihood-gain score. Inspection of this potential solution showed that there were in fact only five dimers in the asymmetric unit of the crystal. When a sixth copy was placed, it had poor electron density and clashed badly with other chains, but a model comprising the first five dimers packed well in the unit cell. Notably, residues 173–189 (beginning with the G173 and G174 deleted in the crystallization construct and modeled by trRosetta as an α-helix) clashed severely with adjacent dimers and were removed prior to initial rebuilding.

### 5.17 | Model building and refinement

Initial rebuilding of the model was performed in ISOLDE, aided markedly by a 10-fold NCS-averaged map generated using Resolve density modification in Phenix (Terwilliger, 2002). Rebuilding one chain in this map, reimposing strict symmetry on the remaining chains and refining with torsion-angle NCS restraints in phenix.refine reduced R<sub>work</sub>/R<sub>free</sub> from 0.38/0.42 to 0.32/0.35. Further inspection revealed clear differences between the two monomers in each dimer, so further rebuilding was performed in non-averaged maps.

### 5.18 | Crystal structure representation

The crystal structure of DhmeA<sup>ΔGG</sup> was visualized by PyMOL (Schrödinger, LLC, 2019). Protein surface charge distribution was calculated by APBS Electrostatics (Jurrus et al., 2018) which was used as a plugin of PyMOL. The default settings of PDBE/PISA server (Krissinel & Henrick, 2007) were used to (i) identify the interactions at assembly interfaces and their corresponding distances, (ii) calculate the interface area, solvation free energy gain upon formation of the interface ( $\Delta^{\ddagger}G$ ), and *p* value.

## AUTHOR CONTRIBUTIONS

Klaudia Chmelova, Tadeja Gao, Martin Polak, and Andrea Schenk Mayerova contributed equally to this study. Klaudia Chmelova and Martin Marek prepared and cloned the DNA constructs and produced the protein samples for structural and biochemical experiments. Klaudia Chmelova, Radka Chaloupkova, Tanvir R. Shaikh, and Jiri Novacek performed initial cryo-EM imaging. Martin Polak and Jiri Novacek collected and processed EM data. Klaudia Chmelova and Martin Marek carried out the crystallization screenings, and optimized crystallization hits. Andrea Schenk Mayerova, Tristan I. Croll, Kay Diederichs, Randy J. Read, and Martin Marek collected diffraction data and solved the protein crystal structure. Radka Chaloupkova, Jiri Damborsky, and Martin Marek designed the project, supervised research, and interpreted data. Tadeja Gao, Jana Skarupova, Andrea Schenk Mayerova, and Martin Marek wrote the manuscript with contributions from all authors. All authors have approved the final version of the manuscript.

## ACKNOWLEDGMENTS

The authors would like to express their thanks to the Czech Science Foundation (GA22-09853S) and the Czech Ministry of Education (INBIO CZ.02.1.01/0.0/0.0/16\_026/0008451, RECETOX RI LM2018121, e-INFRA LM2018140). This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement No 857560 (CETOCOEN Excellence) and No 814418 (Sinfonia). This publication reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains. We acknowledge CF Biomolecular Interactions and Crystallography and CF Cryoelectron microscopy and tomography of CIISB, Instruct-CZ Centre, supported by MEYS CR (LM2023042) and European Regional Development Fund-Project "UP CIISB" (No. CZ.02.1.01/0.0/0.0/18\_046/0015974). R.J.R. and T.I.C. were supported by a Wellcome Trust Principal Research Fellowship (grant No. 209407/Z/17/Z to R.J.R.). The crystallographic data were collected at the PXIII beamline at the Swiss Light Source (SLS) in Villigen (Switzerland). We are grateful to the members of the SLS synchrotron for the use of their beamline and help during data collection. A.S. would like to thank Ronan Keegan and Andrey Lebedev, the lecturers of the CCP4 Crystallographic School in South Africa 2021, for their efforts in solving the DhmeA<sup>ΔGG</sup> crystal structure.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Atomic coordinates and structure factors have been saved in the Protein Data bank ([www.wwpdb.org](http://www ww p d b . o r g)) (Berman et al., 2000) under PDB ID accession code: 8CKP and EMDB ID accession code: EMD-17015. The authors will release the atomic coordinates and experimental data upon article publication.

## ORCID

Martin Polak  <https://orcid.org/0000-0001-6724-6222>

Andrea Schenk Mayerova  <https://orcid.org/0000-0003-0883-9866>

Kay Diederichs  <https://orcid.org/0000-0001-5360-0905>

Jiri Novacek  <https://orcid.org/0000-0003-4013-3898>

Martin Marek  <https://orcid.org/0000-0001-7220-5644>

## REFERENCES

- Assmann GM, Wang M, Diederichs K. Making a difference in multi-data-set crystallography: simple and deterministic data-scaling/selection methods. *Acta Crystallogr D Struct Biol.* 2020; 76:636–52.
- Babkova P, Sebestova E, Brezovsky J, Chaloupkova R, Damborsky J. Ancestral haloalkane dehalogenases show robustness and unique substrate specificity. *Chembiochem.* 2017;18:1448–56.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000;28:235–42.
- Bunkoczi G, Read RJ. Improvement of molecular-replacement models with sculptor. *Acta Crystallogr D Biol Crystallogr.* 2011;67:303–12.
- Bunkóczi G, Read RJ. Phenix. *Ensembler: a tool for multiple superposition.* *Comput Crystallogr Newsl.* 2011;2:8–9.
- Chaloupkova R, Prudnikova T, Rezacova P, Prokop Z, Koudelakova T, Daniel L, et al. Structural and functional analysis of a novel haloalkane dehalogenase with two halide-binding sites. *Acta Crystallogr D Biol Crystallogr.* 2014;70:1884–97.
- Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, et al. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res.* 2021;49:D498–508.
- Chovancova E, Kosinski J, Bujnicki JM, Damborsky J. Phylogenetic analysis of haloalkane dehalogenases. *Proteins.* 2007;67:305–16.
- Chrast L, Tratsiak K, Planas-Iglesias J, Daniel L, Prudnikova T, Brezovsky J, et al. Deciphering the structural basis of high thermostability of dehalogenase from psychrophilic bacterium *Marinobacter* sp. ELB17. *Microorganisms.* 2019;7:498.
- Christenson JK, Jensen MR, Goblirsch BR, Mohamed F, Zhang W, Wilmot CM, et al. Active multienzyme assemblies for long-chain olefinic hydrocarbon biosynthesis. *J Bacteriol.* 2017;199:e00890.
- Christenson JK, Robinson SL, Engel TA, Richman JE, Kim AN, Wackett LP. OleB from bacterial hydrocarbon biosynthesis is a beta-lactone decarboxylase that shares key features with haloalkane dehalogenases. *Biochemistry.* 2017;56:5278–87.
- Colman PM, Fehlhhammer H, Bartels K. (1976). *Crystallographic Computing Techniques.* In: Bartels FR, Huml K, Sedlacek B, editors. Munksgaard: Copenhagen; 1976. p. 248–58.
- Croll TI. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol.* 2018;74:519–30.

- Cui Y-W, Gong X-Y, Shi Y-P, Wang Z. Salinity effect on production of PHA and EPS by *Haloferax mediterranei*. RSC Adv. 2017;7:53587–95.
- Damborsky J, Koča J. Analysis of the reaction mechanism and substrate specificity of haloalkane dehalogenases by sequential and structural comparisons. Protein Eng. 1999;12:989–98.
- Damborsky J, Rorije E, Jesenska A, Nagata Y, Klopman G, Peijnenburg WJGM. Structure-specificity relationships for haloalkane dehalogenases. Environ Toxicol Chem. 2001;20:2681–9.
- Dimitriou PS, Denesyuk A, Takahashi S, Yamashita S, Johnson MS, Nakayama T, et al. Alpha/beta-hydrolases: a unique structural motif coordinates catalytic acid residue in 40 protein fold families. Proteins. 2017;85:1845–55.
- East NJ, Clifton BE, Jackson CJ, Kaczmarek JA. The role of oligomerization in the optimization of cyclohexadienyl dehydratase conformational dynamics and catalytic activity. Protein Sci. 2022;31:e4510.
- Franken SM, Rozeboom HJ, Kalk KH, Dijkstra BW. Crystal structure of haloalkane dehalogenase: an enzyme to detoxify halogenated alkanes. EMBO J. 1991;10:1297–302.
- Fung HK, Gadd MS, Drury TA, Cheung S, Guss JM, Coleman NV, et al. Biochemical and biophysical characterisation of haloalkane dehalogenases DmrA and DmrB in *Mycobacterium* strain JS60 and their role in growth on haloalkanes. Mol Microbiol. 2015;97:439–53.
- Gehret JJ, Gu L, Geders TW, Brown WC, Gerwick L, Gerwick WH, et al. Structure and activity of DmmA, a marine haloalkane dehalogenase. Protein Sci. 2012;21:239–48.
- Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci. 2018;27:14–25.
- Gwyther REA, Jones DD, Worthy HL. Better together: building protein oligomers naturally and by design. Biochem Soc Trans. 2019;47:1773–80.
- Han J, Zhang F, Hou J, Liu X, Li M, Liu H, et al. Complete genome sequence of the metabolically versatile halophilic archaeon *Haloferax mediterranei*, a poly(3-hydroxybutyrate-co-3-hydroxyvalerate) producer. J Bacteriol. 2012;194:4463–4.
- Henderson JP, Byun J, Mueller DM, Heinecke JW. The eosinophil peroxidase-hydrogen peroxide-bromide system of human eosinophils generates 5-bromouracil, a mutagenic thymine analogue. Biochemistry. 2001;40:2052–9.
- Henderson JP, Byun J, Takeshita J, Heinecke JW. Phagocytes produce 5-chlorouracil and 5-bromouracil, two mutagenic products of myeloperoxidase, in human inflammatory tissue. J Biol Chem. 2003;278:23522–8.
- Hikiba J, Hirota K, Kagawa W, Ikawa S, Kinebuchi T, Sakane I, et al. Structural and functional analyses of the DMC1-M200V polymorphism found in the human population. Nucleic Acids Res. 2008;36:4181–90.
- Hochberg GKA, Shepherd DA, Marklund EG, Santhanagoplan I, Degiacomi MT, Laganowsky A, et al. Structural principles that enable oligomeric small heat-shock protein paralogs to evolve distinct functions. Science. 2018;359:930–5.
- Janssen DB. Evolving haloalkane dehalogenases. Curr Opin Chem Biol. 2004;8:150–9.
- Jesenska A, Monincova M, Koudelakova T, Hasan K, Chaloupkova R, Prokop Z, et al. Biochemical characterization of haloalkane dehalogenases DrbA and DmbC, representatives of a novel subfamily. Appl Environ Microbiol. 2009;75:5157–60.
- Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018;27:112–28.
- Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66:125–32.
- Kantardjiev KA, Rupp B. Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. 2003;12:1865–71.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
- Kiss-Szeman AJ, Straner P, Jakli I, Hosogi N, Harmat V, Menyhard DK, et al. Cryo-EM structure of acylpeptide hydrolase reveals substrate selection by multimerization and a multi-state serine-protease triad. Chem Sci. 2022;13:7132–42.
- Koudelakova T, Bidmanova S, Dvorak P, Pavelka A, Chaloupkova R, Prokop Z, et al. Haloalkane dehalogenases: biotechnological applications. Biotechnol J. 2013;8:32–45.
- Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372:774–97.
- Lahoda M, Mesters JR, Stsiapanava A, Chaloupkova R, Kutý M, Damborsky J, et al. Crystallographic analysis of 1,2,3-trichloropropane biodegradation by the haloalkane dehalogenase DhaA31. Acta Crystallogr D Biol Crystallogr. 2014;70:209–17.
- Lanyi JK. Salt-dependent properties of proteins from extremely halophilic bacteria. Bacteriol Rev. 1974;38:272–90.
- Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using x-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D. 2019;75:861–77.
- Lynch M. Evolutionary diversification of the multimeric states of proteins. Proc Natl Acad Sci USA. 2013;110:E2821–8.
- Marek J, Vevodova J, Smatanova IK, Nagata Y, Svensson LA, Newman J, et al. Crystal structure of the haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26. Biochemistry. 2000;39:14082–6.
- Marek M, Ramos-Morales E, Picchi-Constante GFA, Bayer T, Norstrom C, Herp D, et al. Species-selective targeting of pathogens revealed by the atypical structure and active site of *Trypanosoma cruzi* histone deacetylase DAC2. Cell Rep. 2021;37:110129.
- Marianayagam NJ, Sunde M, Matthews JM. The power of two: protein dimerization in biology. Trends Biochem Sci. 2004;29:618–25.
- Markova K, Chmelova K, Marques SM, Carpentier P, Bednar D, Damborsky J, et al. Decoding the intricate network of molecular interactions of a hyperstable engineered biocatalyst. Chem Sci. 2020;11:11162–78.
- Markova K, Kunka A, Chmelova K, Havlasek M, Babkova P, Marques SM, et al. Computational enzyme stabilization can affect folding energy landscapes and lead to catalytically enhanced domain-swapped dimers. ACS Catal. 2021;11:12864–85.
- Matthews BW. Solvent content of protein crystals. J Mol Biol. 1968;33:491–7.
- Mazumdar PA, Hulecki JC, Cherney MM, Garen CR, James MN. X-ray crystal structure of *Mycobacterium tuberculosis* haloalkane dehalogenase Rv2579. Biochim Biophys Acta. 2008;1784:351–62.
- Mazur A, Prudnikova T, Grinkevich P, Mesters JR, Mrazova D, Chaloupkova R, et al. The tetrameric structure of the novel

- haloalkane dehalogenase DpaA from *Paraglaciecola agarilytica* NO2. *Acta Crystallogr D Struct Biol.* 2021;77:347–56.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Cryst.* 2007;40:658–74.
- Nagata Y, Ohtsubo Y, Tsuda M. Properties and biotechnological applications of natural and engineered haloalkane dehalogenases. *Appl Microbiol Biotechnol.* 2015;99:9865–81.
- Novak HR, Sayer C, Isupov MN, Gotz D, Spragg AM, Littlechild JA. Biochemical and structural characterisation of a haloalkane dehalogenase from a marine Rhodobacteraceae. *FEBS Lett.* 2014;588:1616–22.
- Oakley AJ, Klvana M, Otyepka M, Nagata Y, Wilce MCJ, Damborsky J. Crystal structure of haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26 at 0.95 Å resolution: dynamics and catalytic residues. *Biochemistry.* 2004;43:870–8.
- Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, et al. The  $\alpha/\beta$  hydrolase fold. *Protein Eng.* 1992;5:197–211.
- Oren A, Hallsworth JE. Microbial weeds in hypersaline habitats: the enigma of the weed-like *Haloferax mediterranei*. *FEMS Microbiol Lett.* 2014;359:134–42.
- Patoli BB, Winter JA, Patoli AA, Delahay RM, Bunting KA. Co-expression and purification of the RadA recombinase with the RadB paralog from *Haloferax volcanii* yields heteromeric ring-like structures. *Microbiology.* 2017;163:1802–11.
- Pavlova M, Klvana M, Prokop Z, Chaloupkova R, Banas P, Otyepka M, et al. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat Chem Biol.* 2009;5:727–33.
- Prokop Z, Sato Y, Brezovsky J, Mozga T, Chaloupkova R, Koudelakova T, et al. Enantioselectivity of haloalkane dehalogenases and its modulation by surface loop engineering. *Angew Chem Int ed Engl.* 2010;49:6111–5.
- Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods.* 2017;14:290–6.
- Read RJ, Schierbeek AJ. A phased translation function. *J Appl Cryst.* 1988;21:490–5.
- Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 2014;42:W320–4.
- Schenkmyerova A, Toul M, Pluskal D, Baatallah R, Gagnot G, Pinto GP, et al. Catalytic mechanism for *Renilla*-type bioluminescence. *Nat Catal.* 2023;6:23–38.
- Schrödinger, LLC. The PyMOL molecular graphics system, version 2.3.2. 2019.
- Shan Y, Yu W, Shen L, Guo X, Zheng H, Zhong J, et al. Conjugation with inulin improves the environmental stability of haloalkane dehalogenase DhaA. *Enzyme Microb Technol.* 2021;149:109832.
- Shin DS, Pellegrini L, Daniels DS, Yelent B, Craig L, Bates D, et al. Full-length archaeal Rad51 structure and mutants: mechanism for RAD51 assembly and control by BRCA2. *EMBO J.* 2003;22:4566–76.
- Stsiapanava A, Dohnalek J, Gavira JA, Kutty M, Koudelakova T, Damborsky J, et al. Atomic resolution studies of haloalkane dehalogenases DhaA04, DhaA14 and DhaA15 with engineered access tunnels. *Acta Crystallogr D Biol Crystallogr.* 2010;66:962–9.
- Sumbalova L, Stourac J, Martinek T, Bednar D, Damborsky J. Hot-Spot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Res.* 2018;46:W356–62.
- Tan YZ, Baldwin PR, Davis JH, Williamson JR, Potter CS, Carragher B, et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods.* 2017;14:793–6.
- Terwilliger TC. Statistical density modification with non-crystallographic symmetry. *Acta Crystallogr D Biol Crystallogr.* 2002;58:2082–6.
- Ukkonen K, Vasala A, Ojamo H, Neubauer P. High-yield production of biologically active recombinant protein in shake flask culture by combination of enzyme-based glucose delivery and increased oxygen transfer. *Microb Cell Fact.* 2011;10:107.
- Valinluck V, Liu P, Kang JI Jr, Burdzy A, Sowers LC. 5-halogenated pyrimidine lesions within a CpG sequence context mimic 5-methylcytosine by enhancing the binding of the methyl-CpG-binding domain of methyl-CpG-binding protein 2 (MeCP2). *Nucleic Acids Res.* 2005;33:3057–64.
- Vanacek P, Sebestova E, Babkova P, Bidmanova S, Daniel L, Dvorak P, et al. Exploration of enzyme diversity by integrating bioinformatics with expression analysis and biochemical characterization. *ACS Catal.* 2018;8:2402–12.
- Vasina M, Vanacek P, Hon J, Kovar D, Faldynova H, Kunka A, et al. Advanced database mining of efficient haloalkane dehalogenases by sequence and structure bioinformatics and microfluidics. *Chem Catal.* 2022;2:2704–25.
- Verschuere KHG, Seljee F, Rozeboom HJ, Kalk KH, Dijkstra BW. Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature.* 1993;363:693–8.
- Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci U S A.* 2020;117:1496–503.
- Zhang K. Gctf: real-time CTF determination and correction. *J Struct Biol.* 2016;193:1–12.
- Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods.* 2017;14:331–2.
- Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol.* 2018;430:2237–43.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Chmelova K, Gao T, Polak M, Schenkmyerova A, Croll TI, Shaikh TR, et al. Multimeric structure of a subfamily III haloalkane dehalogenase-like enzyme solved by combination of cryo-EM and x-ray crystallography. *Protein Science.* 2023;32(10):e4751. <https://doi.org/10.1002/pro.4751>