Computational and Structural Biotechnology Journal 18 (2020) 1497–1508

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Structures of hyperstable ancestral haloalkane dehalogenases show restricted conformational dynamics

Petra Babkova [a,b,1], Zuzana Dunajova [a,1], Radka Chaloupkova [a], Jiri Damborsky [a,b], David Bednar [a,b], Martin Marek [a,1,*]

[a] Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Kamenice 5, Bld. A13, 625 00 Brno, Czech Republic
[b] International Clinical Research Center, St. Anne's University Hospital Brno, Pekarska 53, 656 91 Brno, Czech Republic

ARTICLE INFO

Article history:
Received 7 April 2020
Received in revised form 8 June 2020
Accepted 10 June 2020
Available online 19 June 2020

Keywords:
Enzyme
Haloalkane dehalogenase
Ancestral sequence reconstruction
Thermostability
X-ray crystallography
Protein simulations
Conformational flexibility
Protein design

ABSTRACT

Ancestral sequence reconstruction is a powerful method for inferring ancestors of modern enzymes and for studying structure–function relationships of enzymes. We have previously applied this approach to haloalkane dehalogenases (HLDs) from the subfamily HLD-II and obtained thermodynamically highly stabilized enzymes (ΔT_m up to 24 °C), showing improved catalytic properties. Here we combined crystallographic structural analysis and computational molecular dynamics simulations to gain insight into the mechanisms by which ancestral HLDs became more robust enzymes with novel catalytic properties. Reconstructed ancestors exhibited similar structure topology as their descendants with the exception of a few loop deviations. Strikingly, molecular dynamics simulations revealed restricted conformational dynamics of ancestral enzymes, which prefer a single state, in contrast to modern enzymes adopting two different conformational states. The restricted dynamics can potentially be linked to their exceptional stabilization. The study provides molecular insights into protein stabilization due to ancestral sequence reconstruction, which is becoming a widely used approach for obtaining robust protein catalysts.

1. Introduction

A detailed understanding of the structure–function relationships is important for the engineering of novel protein catalysts with enhanced thermostability and improved catalytic properties. Ancestral sequence reconstruction (ASR) is a valuable tool for experimental testing of structure-dynamics-function relationships between predicted ancestral enzymes and their modern-day descendants [1–2]. This powerful approach enables reconstruction of ancient enzymes based on the sequence data inferred by phylogenetic analyses, which allows interpretation of their molecular evolution [3]. In the past two decades, ASR was successfully applied for example: (i) to stabilize thioredoxin enzymes [4–5], 3-isopropylmalate dehydrogenase [6], serum paraoxonase [7], (ii) to enhance catalytic performance of copepod luciferase [8] and polysaccharide lyases [9], and (iii) to modify substrate specificity of β-lactamases [10] and lactate dehydrogenase [11].

Haloalkane dehalogenases (HLDs) are predominantly microbial enzymes that catalyze the hydrolysis of halogenated hydrocarbons to the corresponding alcohol, halide ion and proton [12]. For this biocatalytic activity, HLDs are recognized as key protein catalysts in many industrial and biotechnological processes [13]. Based on sequence and phylogenetic analyses HLDs are clustered into three subfamilies HLD I-III [14]. HLDs structurally belong to the α/β-hydrolase fold superfamily [15]. HLD structure consists of (i) a conserved main domain formed by an eight-stranded β-sheet sandwiched by six α-helices (αβα-sandwich architecture), and (ii) a structurally variable and flexible cap domain that is typically formed by five α-helices [16]. The HLD active site is buried at an interface between the main and the cap domain, and is connected with the exterior environment via several access tunnels [17]. The catalytic pentad, composed of amino acid residues essential for the enzymatic reaction, is formed by the nucleophile Asp, the base His, the catalytic acid Asp/Glu and two halide stabilizing residues, often formed by Trp and Trp/Asn pair [12].

Recently, we applied ASR to infer putative ancient HLDs within the subfamily HLD-II [18]. The study resulted in a successful reconstruction of five ancestral HLDs, hereafter referred to as AncHLD1-5, which were recombinantly expressed and biochemically and

* Corresponding author.
E-mail address: martin.marek@recetox.muni.cz (M. Marek).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2020.06.021

biophysically characterized [18]. These ancestral enzymes exhibit higher specific activities than their corresponding descendants with a preference for short multi-substituted halogenated substrates. Importantly, their thermal stabilities markedly increased compared to the modern-day enzymes (ΔT_m up to 24 °C) [18]. Since thermal stability is a key parameter in the usage of protein catalysts in industrial and technological settings, reconstructed AncHLDs represent industrially promising biocatalysts. Therefore, the comprehensive structural analysis of these promising biocatalysts can help us to better understand enhanced thermal stability and inspire future protein engineering efforts.

Here we used a combination of X-ray crystallography and molecular dynamics (MD) simulations to explore the structural and dynamical features distinguishing ancestral and contemporary HLDs. The crystal structures of three ancestral HLDs – AncHLD2, AncHLD3 and AncHLD5 – reveal an overall similarity in their architecture with modern-day HLDs, but structural variability in two different loops. MD simulations carried out with the obtained crystal structures indicate that the enhanced thermal stability of the ancestral HLDs could be related to decreased conformational dynamics, due to their strong preference for a single conformation.

2. Results

2.1. Crystallization and structure determination of ancestral HLDs

Despite intensive efforts, we were not able to obtain diffraction-quality crystals for all five reconstructed ancestral enzymes, AncHLD1-5. Well-diffracting crystals could be grown for AncHLD2, AncHLD3 and AncHLD5, and high-resolution crystallographic data were collected at 1.7 Å, 1.26 Å and 1.75 Å resolutions, respectively (Table 1). The structures were solved by molecular replacement using homology-derived 3D structures as search models. The initial models were further refined through multiple cycles of manual building and automatic refinement. The final models show good statistics (Table 1). Most of the residues could be built into the densities, with the exception of unstructured flexible N- and C-terminal ends. All newly-solved AncHLDs structures (AncHLD2, AncHLD3 and AncHLD5) show an overall structural similarity with other characterized modern-day HLDs [19–23], whose structures were reported previously. The topologies of secondary structure elements are comparable with those observed for present-day HLDs (Fig. 1). Sequence and structural comparisons, including root mean squared deviation (RMSD) on Cα atoms and Dali Z-scores [24] between reconstructed ancestral and modern HLD enzymes are summarized in Fig. 1B-D. Specifically, AncHLDs adopt a canonical α/β-hydrolase fold with a conserved core, which consists of a central eight-stranded β-sheet, with β2 lying in an antiparallel orientation with respect to the direction of the β-sheet that is sandwiched between six α-helices. In addition, there is a typical cap domain present, which makes a shield over the main core domain (Figs. 2 and 3). The active site of all three AncHLDs contains the catalytic pentad in canonical catalytically relevant conformations (Figs. 4 and 5).

2.2. Structural analysis of AncHLD2

The crystals of AncHLD2 were found to belong to the $P2_12_12$ space group (Table 1), where six enzyme molecules are present in the asymmetric unit, with the root mean squared deviation (RMSD) of Cα's ranging from 0.1 to 0.3 between individual monomers. The AncHLD2 enzyme is structurally similar to its descendant DbeA (RMSD = 0.4 Å) and DbjA (RMSD = 0.5 Å) enzymes (Fig. 2). The only exception is a different conformation of the L9 loop of AncHLD2, due to a lack of DbjA-specific extension (Fig. 1B and Fig. 2C). Interestingly, we observe unambiguous peak densities for two chloride anions, in all six AncHLD2 monomers. We designated these sites chloride-binding site 1 and 2 (CBS-1 and CBS-2). CBS-1 is located in the bottom of the active site cavity, thus corresponding to the product-binding site, where the chloride anion (Cl1) is bound between two halide-stabilizing residues, N38 (~3.5 Å) and W104 (~3.2 Å), and is in close contact with both the pyrrolidine ring of P205 (~3.4 Å) and a solvent water molecule (~3.0 Å) (Fig. 4). This water molecule is a part of the hydrogen-bonded water molecule network that is lined up in the main (p1) access tunnel, which connects the buried active site with the exterior environment (Figure S1A). The second chloride anion (Cl2) is located approximately 10.2 Å from CBS-1 and is coordinated by G37 (~3.3 Å), T40 (~3.1 Å), Q102 (~3.2 Å) and Q274 (~3.5 Å) (Fig. 4). While CBS-1 is common for all members of HLD family [12], CBS-2 was previously observed only in the crystal structure of DbeA enzyme from *Bradyrhizobium elkanii* USDA94 [20], the most related descendant to AncHLD2 (Fig. 1D and Fig. 2A) [18]. We also identified several molecules of 2-Methyl-2,4-pentanediol (MPD), 2-Amino-2-(hydroxymethyl)-1,3-propanediol (TRIS) and Ethylene glycol (EDO) in the crystal structure of AncHLD2, which originated from purification buffer and/or crystallization mother liquor, and predominantly bound in enzyme surface binding pockets or making crystal packing contacts.

Table 1
Crystallographic data collection and refinement statistics.

Data collection*	AncHLD2	AncHLD3	AncHLD5
Wavelength (Å)	0.97625	0.97625	0.97625
Space group	$P2_12_12$	$P6_522$	$P12_11$
Cell dimensions			
a, b, c (Å)	109.58, 166.62, 100.29	71.42, 71.42, 225.76	63.871, 51.365, 78.419
α, β, γ (°)	90, 90, 90	90, 90,120	90, 94.067, 90
Resolution (Å)	67.62 – 1.7 (1.76 – 1.7)	61.85 – 1.26 (1.305 – 1.26)	39.99 – 1.748 (1.81 – 1.748)
Total reflections	1,316,441 (95,050)	1,575,603 (54,070)	165,995 (16,069)
Unique reflections	201,178 (19,903)	92,642 (6,558)	50,931 (4,897)
Rmerge	11.5 (151.4)	5.1 (111.9)	5.2 (44.5)
I / σI	9.9 (1.1)	25.7 (1.8)	14.9 (2.7)
Completeness (%)	99.9 (99.9)	99.8 (97.7)	98.8 (95.8)
Multiplicity	6.5 (6.4)	17 (8.2)	3.3 (3.3)
CC(1/2)	99.7 (55.9)	100 (56.4)	99.8 (80.2)
Wilson B-factor	19.947	14.776	21.52
Refinement			
Resolution (Å)	67.62 – 1.7 (1.76 – 1.7)	61.85 – 1.26 (1.305 – 1.26)	39.99 – 1.748 (1.81 – 1.748)
No. reflections	201,178 (19,853)	92,502 (8,867)	50,931 (4,897)
Rwork (%) / Rfree (%)	18.15 / 21.83	14.66 / 16.87	18.31 / 21.54
Number of atoms			
Protein	13,869	2,316	4,552
Ligand	212	26	178
Water	1,046	272	570
B-factors			
Protein	28.5	22.8	25.9
Ligand	43.4	43.5	–
Water	35.0	39.8	29.6
R.m.s. deviations			
Bond lengths (Å)	0.007	0.005	0.009
Bond angles (°)	0.89	0.79	1.02
Ramachandran favored (%)	96.41	95.8	96.63
Ramachandran allowed (%)	3.59	4.2	3.19
Ramachandran outliers (%)	0	0	0.18
PDB ID code	6Y9E	6Y9F	6Y9G

*Values in parentheses are for the highest-resolution shell.

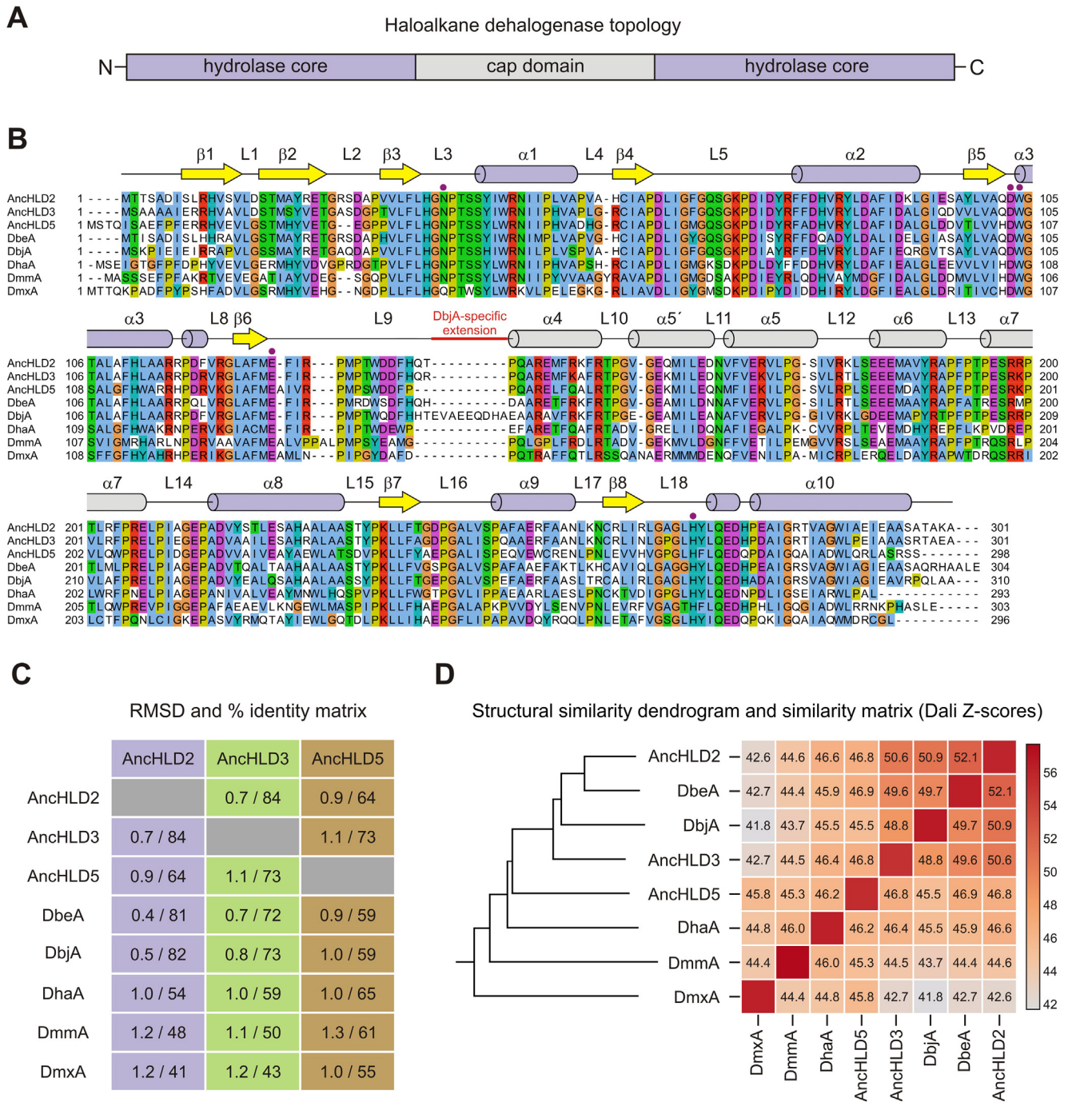

Fig. 1. Sequence and structural comparison between ancestral and modern HLD enzymes. (A) Schematic representation of the domain topology of haloalkane dehalogenase enzymes. (B) Sequence alignment of the in-lab resurrected ancestral enzymes (AncHLD2, AncHLD3 and AncHLD5) and modern enzymes (DbeA, DbjA, DhaA, DmmA and DmxA). The alignment was constructed using the MUSCLE, and coloring was performed in JalView. Secondary structure elements found in AncHLD2 are shown above the alignment. The residues of the catalytic pentad are labelled with purple circles. (C) Pairwise structural comparisons between ancestral and modern HLD enzymes. Values of RMSD on Cα atoms and percent (%) structural identity are noted. (D) Structural similarity dendrogram and structural similarity matrix. The dendrogram was derived by average linkage clustering of the structural similarity matrix (Dali Z-scores). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Structural analysis of AncHLD3

The crystals of AncHLD3 were found to belong to the $P6_522$ space group, where one monomer is present in the asymmetric unit (Table 1 and Fig. 2B). We observe unambiguous density of 2-(cyclohexylamino)ethanesulfonic acid (CHES), a buffer molecule originating from crystallization mother liquor, which is bound in the active site pocket (Fig. 4B). The sulphate group of the CHES molecule is deeply buried in the bottom of the active site, where it occupies the product-binding site. Here, the sulphate group interacts with two halide-stabilizing residues, N38 (~2.9 Å) and W104 (~2.5 Å), which mimics halide stabilization. The nitrogen atom of the CHES ethylamine linker is hydrogen-bonded with a carboxyl moiety of D103 (~2.7 Å) (Fig. 4B). Finally, the capping

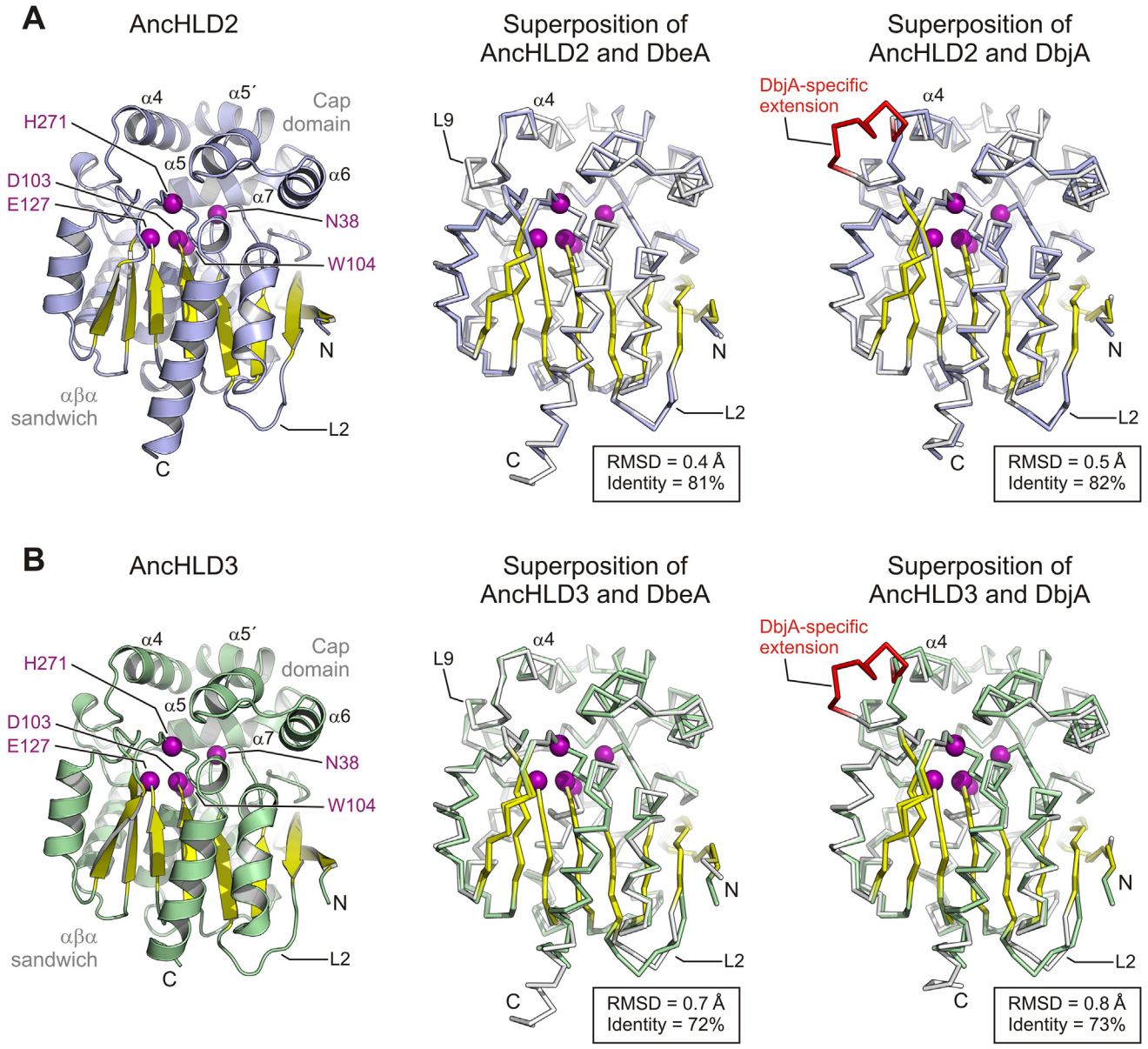

Fig. 2. Structures of AncHLD2 and AncHLD3. (A) Cartoon representation of AncHLD2 structure with the central eight-stranded β-sheet (yellow), α/β-hydrolase helices and helical cap domain (light blue) (left panel). Ribbon representations of structural comparison between AncHLD2 and DbeA (middle panel) and AncHLD2 and DbjA (right panel) are shown. The positions of catalytic pentad residues are shown as purple spheres. (B) Cartoon representation of AncHLD3 structure with the central eight-stranded β-sheet (yellow), α/β-hydrolase helices and helical cap domain (light green) (left panel). Ribbon representations of structural comparison between AncHLD3 and DbeA (middle panel) and AncHLD2 and DbjA (right panel) are shown. The positions of catalytic pentad residues are shown as purple spheres. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cyclohexyl group adopts a twist-boat conformation, and it forms multiple non-polar contacts with active site cavity-lining residues, namely F148, F167, I171, H271 and L245. Additionally, we observe a second molecule of CHES bound to the enzyme. The second CHES molecule is bound via its hydrophobic cyclohexyl moiety in the enzyme's shallow surface pocket. The cyclohexyl group is perfectly resolved as it is tightly bound in the surface pocket, formed by predominantly hydrophobic residues I7, V22, Y43, R46, I49, and A60 (Figure S1B). In contrast, the remaining part carrying the sulphate group is solvent-exposed, and therefore the density for this moiety is poorly resolved.

2.4. Structural analysis of AncHLD5

The crystals of AncHLD5 were found to belong to the $P12_11$ space group, where two enzyme monomers are present in the asymmetric unit (Table 1). Although chloride anions were present in the crystallization mother liquor, there is no halide anion bound in the AncHLD5 enzyme, which is quite unusual for the apo structure of HLDs. Instead, we observe two water molecules bound in the active site. The first water molecule is bound at the product-binding site, where it is coordinated by two halide-stabilizing residues, N40 (~3.1 Å) and W106 (~3.6 Å) (Fig. 4A). The second water molecule makes simultaneous hydrogen-bonding with the former water molecule (~2.6 Å) and with the carboxyl group of D105 (~2.6 Å) (Fig. 4A). Whilst the access tunnel entries in AncHLD2 and AncHLD3 enzymes are narrow and rather occluded, the corresponding surface entries in AncHLD5 are spatially larger (Fig. 4B). This observation may be in agreement with the fact that AncHLD5 exhibits lower dehalogenation activity, when compared to ancestral proteins AncHLD2 and AncHLD3 [18].

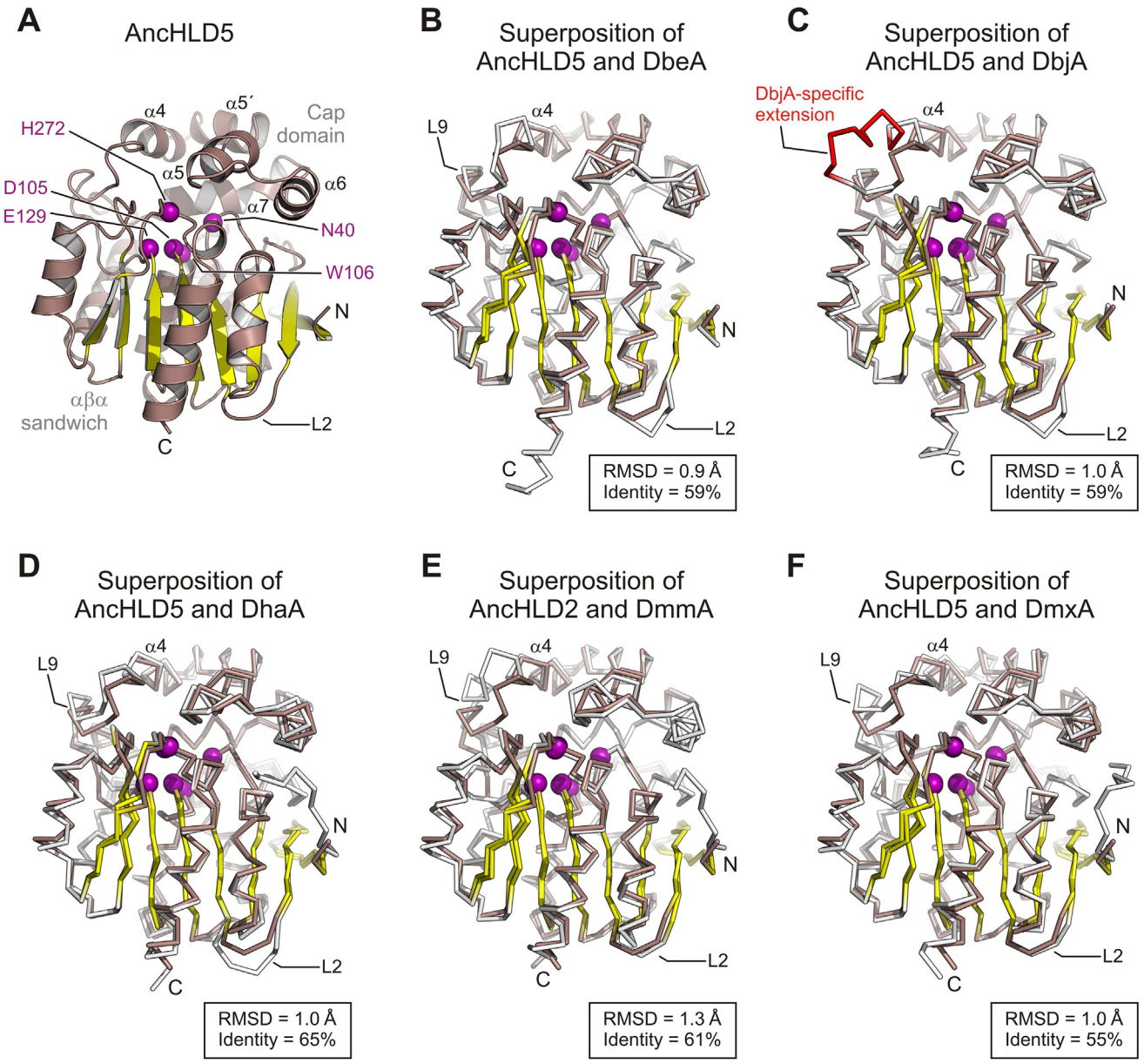

Fig. 3. Structure of AncHLD5. (A) Cartoon representation of AncHLD5 structure with the central eight-stranded β-sheet (yellow), α/β-hydrolase helices and helical cap domain (salmon). The positions of catalytic pentad residues are shown as purple spheres. (B-F) Structural comparison between AncHLD5 and DbeA (B), AncHLD5 and DbjA (C), AncHLD5 and DhaA (D), AncHLD5 and DmmA (E) and AncHLD5 and DmxA (F). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.5. Structural comparisons between ancestral and descendant enzymes

All reconstructed ancestral enzymes AncHLD2, AncHLD3, and AncHLD5 exhibited large sequence and structural similarities. Specifically, the sequence and structural similarities expressed by RMSD, % structural identity and Dali Z-score values are summarized in Fig. 1C-D, while the structural superposition of AncHLD2, AncHLD3 and AncHLD5 is shown in Figure S2. The structural conservation of five catalytically essential residues, the so-called catalytic pentad, is shown in Fig. 5.

The most significant structural deviations between presented ancestral structures are observed (i) in the L2 loop connecting the β2 and β3 strands and (ii) in the L9 loop between the β6 strand and α4 helix, which anchors a helical cap domain to a main α/β-hydrolase domain (Figs. 2 and 3). The reconstructed ancestors AncHLD2 and AncHLD3 displayed the highest structural and sequence similarities with the contemporary HLDs, DbjA and DbeA (Fig. 2). The topologies of the secondary structure elements of AncHLD2 and AncHLD3 are almost identical with the structure of DbeA. The most significant difference was observed in the comparison with the structure of DbjA, which contains extended α4 helix in the cap domain and the L9 loop connecting the main and a cap domain (Fig. 2). The structure of AncHLD5 exhibits the highest percent structural identity with DhaA (65%) and DmmA (61%) (Fig. 1C). The main structure deviations between AncHLD5 and modern HLDs are again mainly in the flexible L9 loop connecting the cap and the main domain and in the L2 loop between the β2 and β3 strands (Fig. 3). The superposition of AncHLD5 and DhaA reveals the deviation of 2.9 Å at these loops (Fig. 3D).

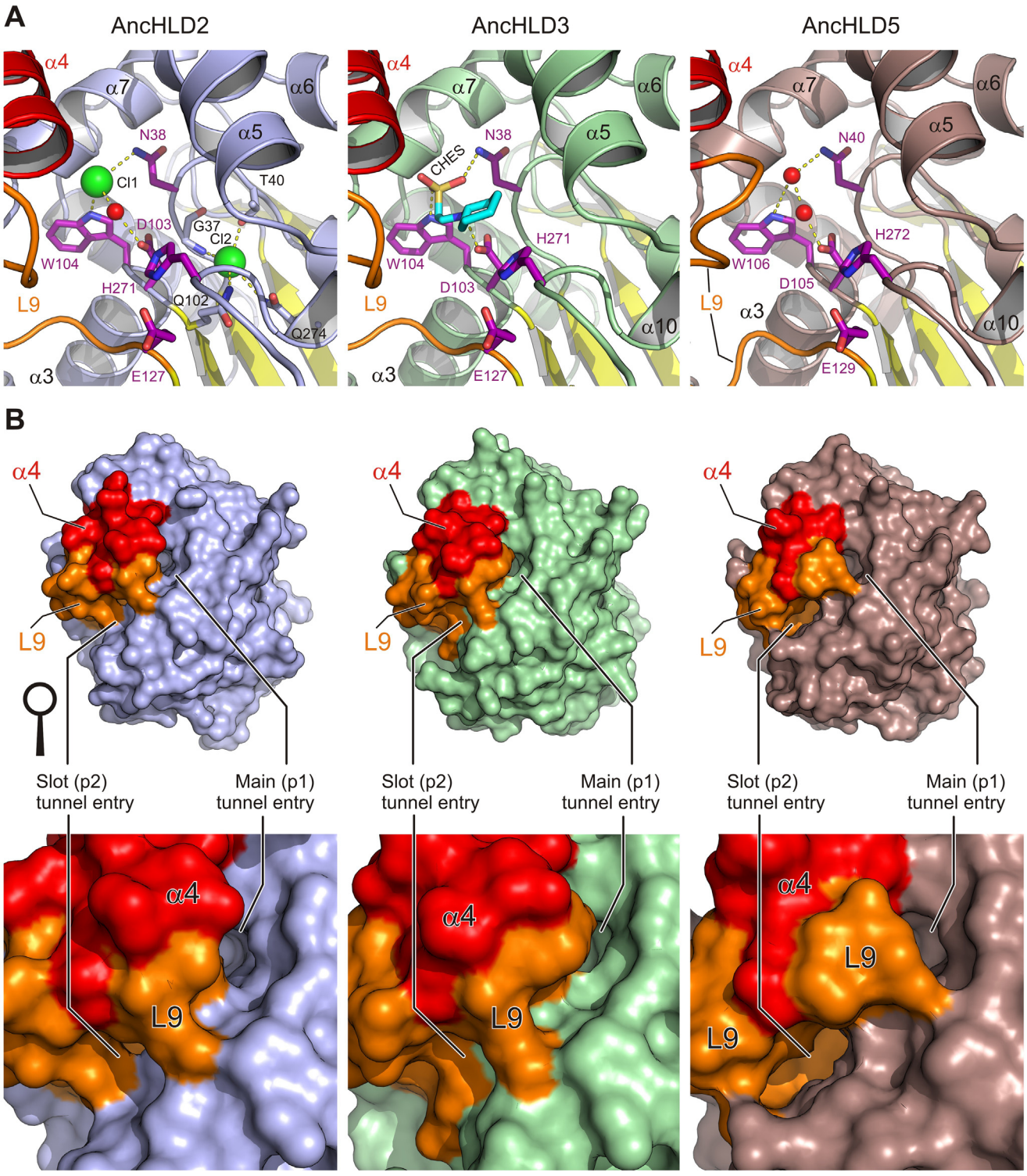

Fig. 4. Structural comparison of AncHLD2, AncHLD3 and AncHLD5. (A) Close-up views of the enzyme active site with key catalytic residues (purple sticks), bound ligand and water molecules. The left panel; close-up view of the AncHLD2 active site with shown residues participating in catalysis and chloride binding. Two chloride ions (Cl1 and Cl2) are shown as green spheres, and a water molecule as red sphere. The middle panel; close-up view of the AncHLD3 active site with shown residues participating in catalysis, CHES molecule is shown as cyan sticks. The right panel; close-up view of the AncHLD5 active site with shown residues participating in catalysis, two water molecules are shown as red spheres. Yellow dashed lines represent coordinating interactions. (B) Surface representations of AncHLD2 (left panels), AncHLD3 (middle panels) and AncHLD5 (right panels) enzymes. Main (p1) and slot (p2) tunnel entries are depicted and zoomed in (bottom panels). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.6. Conformational dynamics of ancestral and descendant enzymes

Computational molecular dynamics (MD) simulations were used to investigate the dynamics of stabilized ancestral HLDs and to compare them with those of modern-day HLDs. We used adaptive sampling method, which is an enhanced-sampling method for efficient exploration of the under-sampled regions of the conformational space using knowledge obtained from past MD simula-

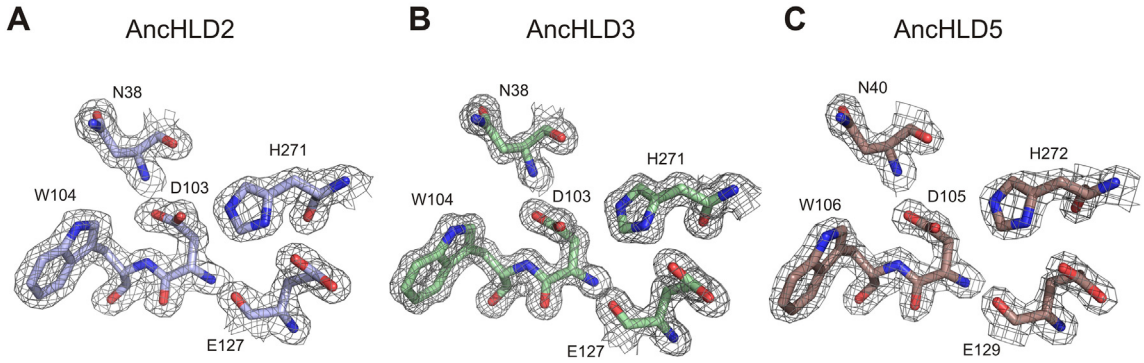

Fig. 5. Structural conservation of active site residues. The 2Fo − Fc electron-density map contoured at 1.5 σ for five catalytically-essential active site residues in AncHLD2 (A), AncHLD3 (B) and AncHLD5 (C).

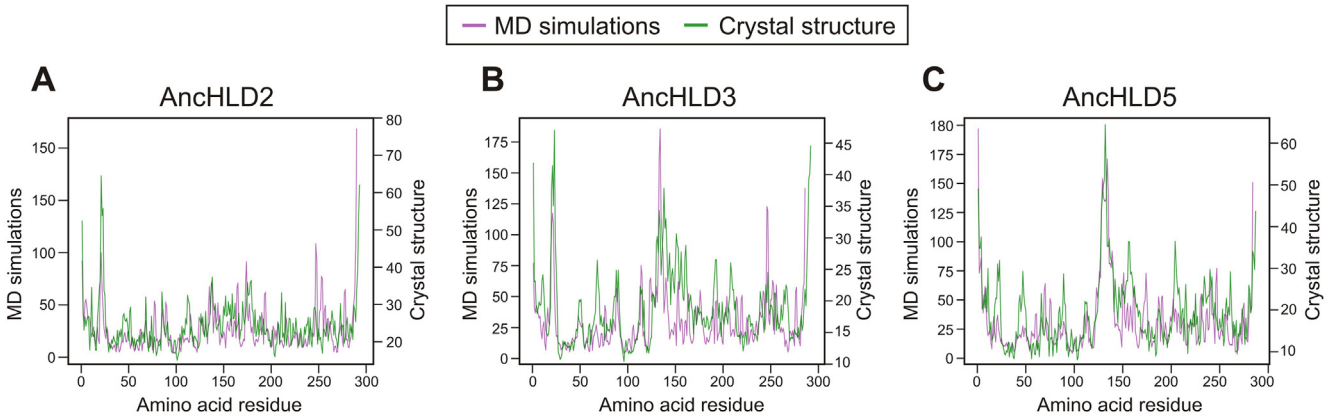

Fig. 6. Comparison of residue-wise Cα B-factors calculated by MD simulations (violet lines) and retrieved from crystal structures (green lines) for AncHLD2 (A), AncHLD3 (B) and AncHLD5 (C). Note that the numerical datasets are smoothed with moving average n = 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tions. By iterative launching of multiple short runs in the least sampled regions, this approach achieves accelerated sampling compared to nonadaptive MD simulations [25] (Figure S3 and Table S1).

The analysis of B-factors from MD simulations revealed that the most flexible part of the studied enzymes is the L9 loop connecting the β6 strand with α4 helix, encompassing the residues of ~ 130–150 (Figure S4). The lowest B-factors in this region were exhibited by AncHLD2 and DhaA, whereas the highest values were found in simulations with AncHLD5, DbjA, and DmmA. Ancestral HLDs that are closer to the root showed a general increase in the B-factors of the studied region compared to more recent ones (Figure S4). We also compare the B-factors from MD simulations with the refined B-factors of Cα atoms of the crystal structures of reconstructed enzymes (Fig. 6). Even though the numerical B-factors exhibit generally higher residue-to-residue fluctuations than the experimental ones, the two datasets covary, mainly in the regions with high fluctuations (L9 loop in AncHLD3 and AncHLD5, and L2 loop in AncHLD2 and AncHLD3; Figs. 6 and 7). Interestingly, we found significant correlations in all three variants demonstrating a good agreement between simulation and experimental data (Pearson's correlation coefficient = 0.40 – 0.57, p-value = $8.4 \cdot 10^{-13}$ – $2.5 \cdot 10^{-19}$; Fig. 8 and Table S2). Removing the noise of numerical B-factors by moving average increased the correlation up to r = 0.63 ($p = 8.4 \cdot 10^{-35}$). The consistency of experimental and numerical B-factors is very good when considering the methodology differences and can be viewed as a verification of the conformational sampling in MD simulations.

In the next step, we analysed Markov State Models (MSM) based on RMSD values of Cα atoms of investigated enzymes. The calculated parameters describing the enzyme dynamics are summarized in Fig. 9 and Table S3. Importantly, our simulations sampled only one conformational state of AncHLD2, despite using the enhanced sampling scheme. This result, together with low B-factor values from MD simulations, imply very low flexibility of this enzyme with low probability of finding other conformations. The remaining enzymes adopted two conformations (Figure S5), usually with a varying position of the L9 loop connecting the main and cap domain and the α4 helix of the cap domain. Due to the consequence of this conformational change to the active site cavity, we refer to the identified states as "closed" and "open". The movements of this region in DbjA were fast, causing high deviations for the two states. On the other hand, the conformational changes in DbeA were subtle and involved mainly the movement of the L9 loop connecting the main and cap domain. Different dynamics were observed in DhaA and AncHLD3. The DhaA enzyme was rigid at the α4 helix, but highly flexible at the L5 loop between the β4 strand and the α2 helix located on the opposite side of the enzyme main domain. AncHLD3 was flexible at the L9 loop connecting the main and the cap domain and α4 helix, but also at the L16 loop between β7 strand and α9 helix located below the above-mentioned mobile region. The Markov state models analysis showed that the two distinct states of modern-day HLDs were energetically close to each other with low kinetic barriers (low ΔG), whereas the ancestral enzymes were adopting preferentially only a single state. This suggests strikingly different conforma-

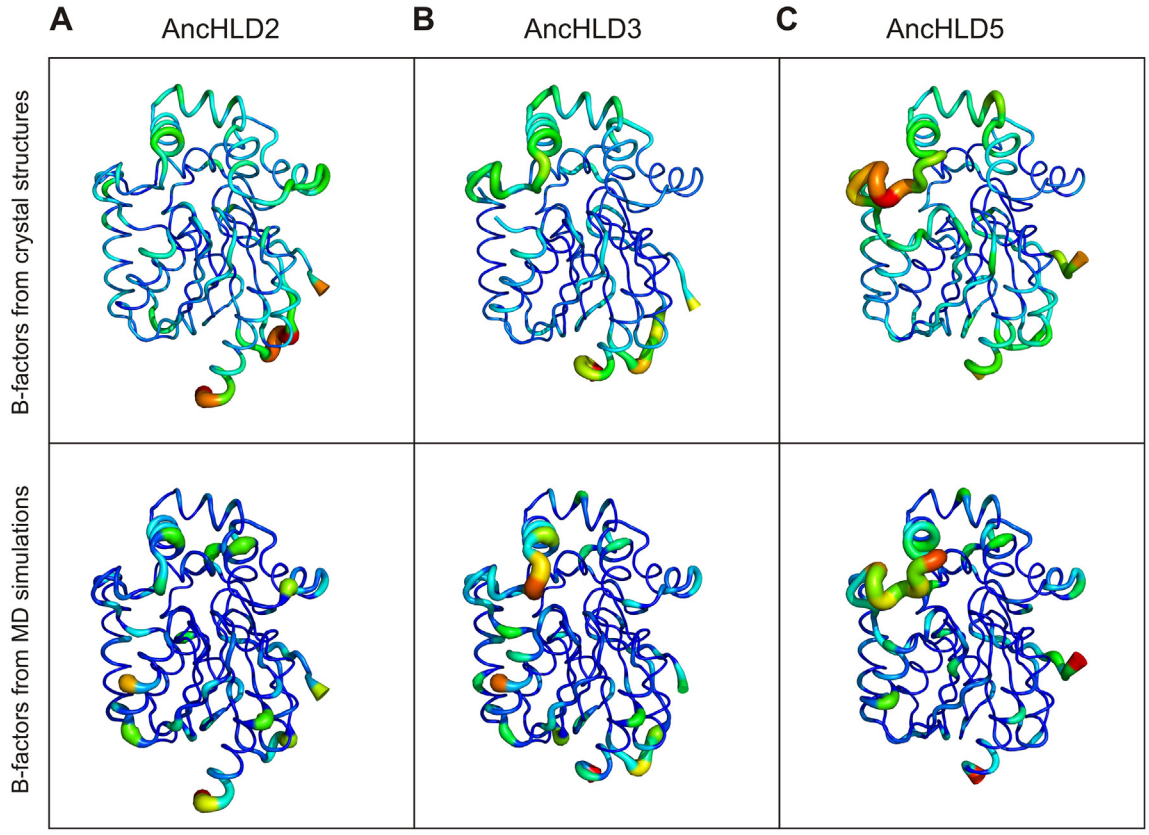

Fig. 7. A visual comparison of ancestral HLDs Cα B-factors magnitude for AncHLD2 (A), AncHLD3 (B) and AncHLD5 (C) enzymes retrieved from static crystal structures (top panels) and MD simulations (bottom panels).

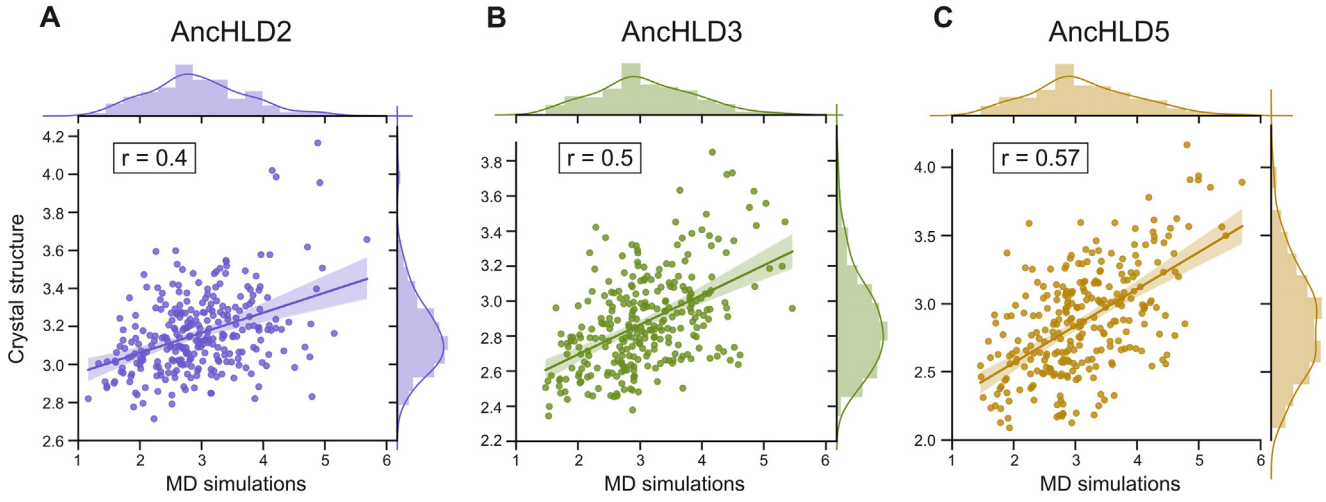

Fig. 8. Correlation plots of Cα B-factor values derived from MD simulations and crystal structures for AncHLD2 (A), AncHLD3 (B) and AncHLD5 (C). The logarithmic transformation of both datasets was performed to achieve normally distributed data.

tional dynamics of ancestral and modern-day haloalkane dehalogenases (Fig. 9B and Table S3).

2.7. Computational analysis of catalytic residues

The RMSD analysis of the catalytic pentad in AncHLD2, DbeA and DmmA revealed a single reactive conformation for the entire simulations (Table S4). The remaining enzymes were also found to adopt a distorted orientation of their catalytic residues with varying equilibrium probability and kinetics. The catalytic pentad of DhaA adopted active conformations with only 52% probability, which is surprising considering high catalytic activity of this enzyme. The active conformation of the catalytic pentad strongly dominated in AncHLD5 with a 96% probability. The pentads of DbjA and AncHLD3 adopted active conformations, with probabilities of 72% and 66%, respectively. The kinetic barriers in AncHLD3 were high, suggesting that the distorted conformation cannot easily be reached (Table S4). Overall, we note that there is no rela-

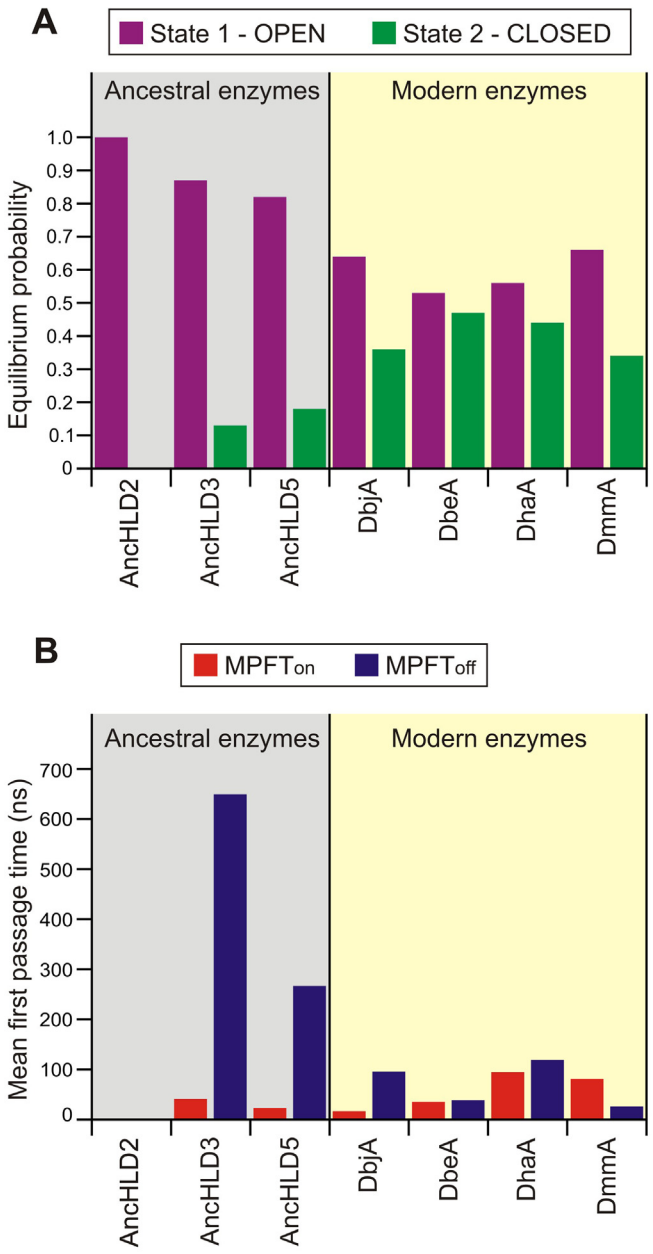

Fig. 9. Restricted conformational dynamics of ancestral enzymes revealed by molecular dynamics simulations. (A) The equilibrium probabilities of the conformational states (State 1 – open and State 2 – closed) of resurrected ancestral HLDs (AncHLD2, AncHLD3 and AncHLD5) and their modern descendants (DbjA, DbeA, DhaA and DmmA). (B) The calculated energy barriers illustrated by mean first passage time between conformational states ($MFPT_{on}$ is time to get into the open state and $MFPT_{off}$ is time to get into the closed state).

tionship between the probabilities of the active conformations of catalytic pentad and the experimentally determined activities of individual enzymes towards the set of halogenated substrates. In other words, our analysis shows that there is no significant difference between the orientation of catalytic residues in ancestral and modern enzymes. Therefore, the catalytic properties should not be compromised.

2.8. Computational analysis of enzyme access tunnels

HLDs contain several internal tunnels connecting the exterior environment with the buried active site [26]. The tunnel parameters for studied enzymes obtained by MD simulations are summarized in Table S5. The main (p1) tunnel is located between the α4 and α5 helices of the cap domain, the slot (p2) tunnel is in between the L9 loop connecting the main and the cap domain and the L16 loop, which connects the β7 strand and α9 helix. We analysed the access tunnels separately for both conformational states in each enzyme. The differences in access pathways between the "closed" and "open" states were much more significant in ancestral enzymes compared to modern day HLDs. The average bottleneck radius of the p1 and p2 tunnels continuously increase from AncHLD2, exhibiting two access pathways with a size of the average bottleneck 1.7 Å and 1.1 Å, to the AncHLD5 displaying access tunnels with the average bottlenecks 2.5 Å and 2.2 Å, respectively. Modern-day HLDs such as DbjA and DbeA contain two access tunnels with the comparable size of the average bottleneck as AncHLD2 and AncHLD3. The only exception is DhaA which exhibited the narrowest pathway of all analysed enzymes with the average bottleneck size 1.3 Å and 1.0 Å (Table S5).

3. Discussion

ASR is a valuable method for studying structure-function relationships in enzymes and for the design of robust protein catalysts. Here, we present the structural and computational analysis of reconstructed HLDs from the subfamily HLD-II, which exhibit high thermodynamic stability and improved specific activities with short brominated and iodinated halogenated substrates [18]. All AncHLDs structures show a structural topology of the α/β-hydrolase fold similar to modern HLDs (Fig. 2,3) [19–22]. The structure conservation is closely related to the high sequence similarity between ancestral and modern HLDs (Fig. 1) [18]. The structure conservation of ancestral enzymes has been previously demonstrated for seven Precambrian thioredoxins [27].

A number of ions and small ligands were found in the structures of ancestral HLDs. The structure of AncHLD2 revealed the presence of two halide-binding sites, both occupied by chloride anions (Fig. 4A). The first halide-binding site, present in all HLDs, is responsible for the stabilization of the released halide ion after the dehalogenation reaction [21,28–29]. The second halide binding site is unique to DbeA and has not previously been observed in the crystal structure of any related enzymes [20]. The residues lining the second site (Gly37, Thr40, Ile44, Gln102, and Gln274) are also present in AncHLD3 and in modern enzymes DbjA, DmlA, DhmA [19–20], but the second halide ion has not been detected in their structures yet. The structure of AncHLD3 uncovered two CHES molecules originating from the crystallization buffer (Figure S1B). The presence of the CHES molecule in the active site has been already observed in the structure of DatA [30]. In the CHES-bound AncHLD3 structure, the sulphate group of the ligand is coordinated by halide stabilizing residues (N38 and W104), while the nitrogen atom of the CHES ethylamine linker interacts with the carboxyl group of the catalytic aspartate, D103. Guan and colleagues examined the catalytic activity of haloalkane dehalogenase DatA in presence of CHES and revealed that this molecule acts as an uncompetitive inhibitor [30]. Equally to the CHES molecule, the similar chemical compound 2-(N-morpholino)ethanesulfonic acid (MES) is bound into the active site of DhaA101 enzyme variant [31]. Analogously, the MES molecule bound in DhaA101 is also stabilized by the interactions with the halide-stabilizing residues (N41 and W107) and nucleophilic aspartate (D106) [31].

Both structural and computational analysis revealed that the most variable element of the HLD structures is the solvent-exposed L9 loop connecting the main and the cap domain and the adjacent α4 helix in the cap domain. The architecture and dynamics of this L9-loop/α4-helix element has a direct impact on

the enzyme access tunnels, influencing the entry of the substrates to the active site, the release of the reaction products, as well as access of water solvent to the active site (Fig. 4B) [26]. Moreover, we speculate that the dynamical behaviour of the neighbouring L16 loop between the strand β7 and helix α9 in AncHLD3 enables the access of the bulkier compounds into the active site as for example the observed CHES molecule in the AncHLD3 crystal structure (Fig. 4A).

The enhanced thermostability of ancestral enzymes was broadly described in several previous studies. However, our understanding of the structural basis of this thermodynamic stability is still limited. The long-standing hypothesis for the stabilization of ancestral proteins is their hyperthermophilic origin [32] The other hypothesis relates the high stability to other factors like oxidative stress, radiation, or high mutability due to lack of chaperons, reparation mechanisms, and transcriptional and translational error rate [7,33] Moreover, the elimination of rare mutations detrimental to structural stability during the maximum likelihood method could cause an improvement in the stability due to an artefact of the reconstruction method [34–35] Our MD simulations indicated that the thermostable ancestral HLDs show more restricted conformational dynamics than their mesophilic descendants. So far we have no experimental data that could confirm our MD simulations results. One option would be to perform a stopped-flow analysis to probe conformational space of ancestral protein forms. In general, our results are in line with the proposal that thermophilic enzymes are more rigid than their mesophilic homologues and that the rigidity should be a prerequisite for high protein thermostability [36]. The reduced flexibility of ancestral proteins has been recently observed by Okafor and co-workers, who studied the thermostable elongation factors Tus [37]. On the other hand, it has been shown by experimental studies combined with computer simulations that the thermal tolerance of a protein is not strictly correlated with the suppression of internal fluctuations and mobility [38].This finding has been further confirmed by the existence of highly stable and flexible proteins [39–40]. Several studies explicate that local flexibility is crucial for high catalytic activity of thermostable enzymes [36,41]. The importance of the conformational flexibility of ancestral enzymes for the emergence of new enzymatic function confirm the studies of Precambrian β-lactamase and ancestral guanylate kinase [42–43]. Our previous study revealed higher specific activities of more rigid ancestral HLDs with small halogenated substrates [18], which contradicts those observations. Our observation can be related to the simplicity of studied halogenated compounds, which do not require extensive conformational changes for substrate binding, transition state stabilization, and product release.

In summary, we present the comparison between newly solved X-ray structures of three reconstructed ancestral enzymes (AncHLD2, AncHLD3 and AncHLD5) and their evolutionary related descendants. Despite the differences in stability and activity of ancestral and extant HLDs, all ancestors possess high structural similarity with extant enzymes. Conformation of the catalytic pentad is preserved during entire simulations in all studied enzymes, explaining the good catalytic activity of reconstructed ancestors. Interestingly, the dynamics of ancestral proteins are specific irrespective of their descendants. The mutations in reconstructed ancestors lead to high thermostability and seem to be connected with restricted backbone dynamics and preferred sampling of only one conformational state. Future experimental work will aim at obtaining of detailed molecular insights into the catalytic mechanism of the wild type and the reconstructed ancestral enzymes. For example, transient kinetic experiments monitoring a change in fluorescence of the halide-binding tryptophan residue, upon the substrate binding and its conversion, could help to explain a molecular basis underlying the increased thermostability of reconstructed ancestral enzymes.

4. Methods

4.1. Protein expression and purification

The gene sequences of predicted ancestral enzymes AncHLD2, AncHLD3 and AncHLD5 were synthesized (GenArt, Thermo Fisher), then subcloned into pET21b (Merck) for expression by using restriction endonucleases NdeI and BamHI (New England Biolabs) and T4 DNA ligase (Promega). *Escherichia coli* chemo-competent BL21(DE3) cells were transformed with prepared recombinant plasmids, plated on agar plates with ampicillin (100 $\mu g \cdot ml^{-1}$) and grown overnight. The obtained colonies were used for inoculation of 10 ml of Luria-Bertani (LB) medium with ampicillin (100 $\mu g \cdot ml^{-1}$), and the cells were grown overnight at 37 °C at 200 rpm. The overnight culture (10 ml) was used to inoculate 1 l of LB medium (ampicillin, 100 $\mu g \cdot ml^{-1}$). When the culture reached OD_{600} of 0.5–0.6 at 37 °C and 115 rpm, the enzyme expression was induced by the addition of IPTG to a final concentration of 100 μM. The cells were harvested and disrupted by sonication in aUP200S ultrasonic processor (Hielscher Ultrasoniscs, Teltow, Germany). The supernatant was collected after centrifugation (4000 g, 1h, 4 °C). The crude extract was further purified on a HiTrap IMAC HP 5-mL column charged with Ni^{2+} (for AncHLD2-3) or Co^{2+} (for AncHLD5) ions (GE Healthcare). His-tagged enzymes in equilibrating buffer (potassium phosphate buffer, 20 mM, pH 7.5) containing sodium chloride (0.5 M) and imidazole (10 mM) were bound to the resin. Unbound and non-specifically bound proteins were washed out by buffer containing imidazole (10 mM). The bound enzymes were eluted by buffer containing imidazole (300 mM). The enzyme fractions were pooled and dialyzed against 100 mM Tris-HCl buffer, pH 7.5 at 4 °C and concentrated to 6–10 $mg \cdot ml^{-1}$. The enzyme concentration was determined by using a Bradford assay (Merck) with bovine serum albumin as a standard. Enzyme purities were checked by SDS-PAGE analysis.

4.2. Crystallization, data collection, and structure determination

The crystallization trials were performed by using the sitting-drop vapor-diffusion method at 23 °C. The crystals of AncHLD2, AncHLD3 and AncHLD5 were grown during the initial screening in 24-well plates (Hampton research, USA) in ratio 1:1 between enzymes and precipitants within 3 to 7 days. The triangular prism shaped crystals of AncHLD2 with dimensions 0.5 × 0.09 × 0.08 mm and the hexagonally shaped crystals of AncHLD5 with average dimension 0.02 × 0.04 × 0.07 mm were observed in condition No. 51 of Wizard classic (equivalent to No. 42 of JCSG-plus screen) containing 20% (w/v) PEG 8000, 200 mM magnesium chloride and 100 mM Tris base/hydrochloric acid, pH 8.5. The hexagonally shaped crystals of AncHLD3 with dimensions 0.2 × 0.1 × 0.04 m m appeared in condition No. 64 of the Wizard classic consisting of 1 M sodium citrate tribasic and 100 mM CHES/ Sodium hydroxide, pH 9.5. Selected protein crystals were mounted in nylon cryoloops (Hampton Research, USA), immersed in originated conditions containing 25% MPD for few seconds, and then flash frozen in liquid nitrogen.

Complete diffraction data sets were collected at the Diamond Light Source beamline I03 (Didcot, UK). Data of AncHLD2 and AncHLD5 were indexed and integrated by XDS [44] and scaled and merged by Aimless of the CCP4 program suite [45]. Data of AncHLD3 were processed with xia2 [46] using XDS [44] for indexing and integration, and Aimless [45] for scaling and merging. The crystal structures were solved by molecular replacement using

Phaser [47] implemented in Phenix [48]. The structure of AncHLD3 was solved by molecular replacement using software Balbes [49]. The homology-derived structural models were built on the template structure of HLDs DbjA [50] used as a search model. The initial model was refined through several cycles of a manual building using COOT [51], and automated refinement with Phenix [48]. The final model was validated using tools provided in COOT [51] and MolProbity [52]. The crystallographic data and final model characteristics are summarized in Table 1.

4.3. Protein structures preparation for computational studies

The crystal structures of ancestral HLDs and modern HLDs were downloaded from the RCSB protein data bank (PDB IDs 4K2A for DbeA, 3A2M for DbjA, 4E46 for DhaA, and 3U1T for DmmA). The hydrogen atoms were added and the structure was prepared with HTMD in pH 7.5 [53]. The catalytic histidine was set to be doubly protonated for these simulations. The structure was then solvated in a 10 Å box of TIP3P water molecules with the sodium chloride concentration of 0.1 M.

4.4. System equilibration and adaptive sampling

The systems were equilibrated using the Equilibration_v1 module of High Throughput Molecular Dynamics (HTMD) [53]. The system was first minimized using the conjugate-gradient method for 500 steps. Then the system was heated and minimized as follows: (I) 500 steps (2 ps) of NVT thermalization with the Berendsen barostat to 298 K with constraints on all heavy atoms of the protein, (II) 625 000 steps (2.5 ns) of NPT equilibration with Langevin thermostat with 1 kcal·mol^{-1}·Å^{-2} constraints on all heavy atoms of the protein and (III) 625 000 steps (2.5 ns) of NPT equilibration with the Langevin thermostat without constraints. During the equilibration simulations, holonomic constraints were on all hydrogen-heavy atom bond terms and the mass of hydrogen atoms was scaled with factor 4, enabling the 4 fs timestep [54–56]. The simulations employed periodic boundary conditions, using the particle mesh Ewald method for treatment of interactions beyond 9 Å cut-off, electrostatic interactions suppressed for more than 4 bond terms away from each other and the smoothing and switching van der Waals and electrostatic interaction cut-off at 7.5 Å [54]. HTMD was used to perform adaptive sampling of the RMSD of the Cα atoms. The 20 ns production runs were started with the files resulting from the equilibration and they employed the same settings as the last step of the equilibration. The trajectories were saved every 0.1 ns.

4.5. B-factors and Markov state model construction

The simulations were made into a simulation list using HTMD and water was filtered out and crashed simulations with length less than 20 ns were omitted resulting in 3–6 μs of simulation time (Table S1). To analyze the B-factors, the fluctuation of the Cα atoms of all residues was computed by Metric-Fluctuation module of HTMD and converted to root-mean-square fluctuation (RMSF) using Eq. (1).

$$B = \frac{8\pi^2}{3} RMSF^2 \quad (1)$$

The B-factors were calculated from RMSF. To compare the numerical and experimental B-factors we computed the Pearson's correlation. Logarithmic transformation of both numerical and experimental datasets was performed to ensure the normal distribution before calculating the correlation. The Pearson's correlation was also computed with smoothed numerical B-factors using a moving average algorithm with window n = 2. The reactive orientation of the the overall dynamics of the enzymes and catalytic residueswas analyzed using 2-state Markov state models (MSMs). Before the MSM construction, we computed the implied timescales in order to estimate the number of observed transitions and choose the lagtime for subsequent models. The metric for building the first MSM was the RMSD of backbone atoms of the whole enzyme apart from the highly flexible ends (residues 10 to 284, chosen according to the B-factors). The data were clustered using MiniBatchKmeans algorithm to 100 clusters. 12 ns lag time was used in the models to construct the Markov states and the Chapman-Kolmogorov test was performed to validate the model (Figure S6). The RMSD of the heavy atoms of the catalytic residues excluding the equivalent oxygen atoms ($OD_{1,2}$ and $OE_{1,2}$) was used as the metric for analysis of the second MSM model. The data were clustered using MiniBatchKmeans algorithm to 100 clusters. 12 ns lag time was used in the models to construct the Markov states and the Chapman-Kolmogorov test [57] was performed to validate the model (Figure S7).

4.6. Analysis of access tunnels

A set of 100 structures of the resulting states were generated and used for the tunnel analysis. CAVER 3.02 [58], was used to calculate the tunnels. The tunnels were calculated using a probe radius of 0.9 Å, a shell radius of 3 Å and shell depth 4 Å. The starting point for the tunnel calculation was defined as OD1 of nucleophilic aspartate. The clustering was performed by the average-link hierarchical Murtagh algorithm, with weighting coefficient of 1 and clustering threshold of 3.5. Only the first and second tunnels identified by CAVER and ranked by the priority were analysed.

Data and code availability

Atomic coordinates and structural factors of AncHLD2, AncHLD3 and AncHLD5 have been deposited in the Protein Data Bank (www.wwpdb.org) under PDB accession codes: 6Y9E, 6Y9F and 6Y9G. Authors will release the atomic coordinates and experimental data upon article publication.

CRediT authorship contribution statement

Petra Babkova: Methodology, Investigation. **Zuzana Dunajova:** Methodology, Investigation. **Radka Chaloupkova:** Conceptualization. **Jiri Damborsky:** Conceptualization, Funding acquisition. **David Bednar:** Methodology, Investigation. **Martin Marek:** Conceptualization, Supervision, Visualization, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Czech Ministry of Education (02.1.01/0.0/0.0/18_046/0015975, LM2018121, LM2015047, LM2018140) and the European Commission (857560, 720776 and 814418) for financial support. M.M. acknowledges the financial support from Marie Skłodowska-Curie Actions (MSCA) project no. 792772 and Grant Agency of the Masaryk University (GAMU) project no. MUNI/H/1561/2018. D.B. is supported by the Czech Grant Agency (20-15915Y). The authors also thank Diamond Light Source

for beamtime and to Dr. Jitka Waterman and the staff of I03 beamline for technical support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.06.021.

References

[1] Harms MJ et al. Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. Proc Natl Acad Sci 2013;110:11475–80.
[2] Harms MJ, Thornton JW. Analyzing protein structure and function using ancestral gene reconstruction. Curr Opin Struct Biol 2010;20:360–6.
[3] Skovgaard M et al. Using evolutionary information and ancestral sequences to understand the sequence-function relationship in GLP-1 agonists. J Mol Biol 2006;363:977–88.
[4] Perez-Jimenez R et al. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. Nat Struct Mol Biol 2011;18:592–6.
[5] Romero-Romero ML, Risso VA, Martinez-Rodriguez S, Ibarra-Molero B, Sanchez-Ruiz JM. Engineering ancestral protein hyperstability. Biochem J 2016;473:3611–20.
[6] Hobbs JK et al. On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of *Bacillus*. Mol Biol Evol 2012;29:825–35.
[7] Trudeau DL, Kaltenbach M, Tawfik DS. On the potential origins of the high stability of reconstructed ancestral proteins. Mol Biol Evol 2016;33:2633–41.
[8] Takenaka Y et al. Computational analysis and functional expression of ancestral copepod luciferase. Gene 2013;528:201–5.
[9] McLean R et al. Functional analyses of resurrected and contemporary enzymes illuminate an evolutionary path for the emergence of exolysis in polysaccharide lyase family 2. J Biol Chem 2015;290:21231–43.
[10] Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-lactamases. J Am Chem Soc 2013;135:2899–902.
[11] Boucher, J. I., Jacobowitz, J. R., Beckett, B. C., Classen, S. & Theobald, D. L. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. eLife 3, e02304 (2014).
[12] Janssen DB. Evolving haloalkane dehalogenases. Curr Opin Chem Biol 2004;8:150–9.
[13] Koudelakova T et al. Haloalkane dehalogenases: biotechnological applications. Biotechnol J 2013;8:32–45.
[14] Chovancová E, Kosinski J, Bujnicki JM, Damborský J. Phylogenetic analysis of haloalkane dehalogenases. Proteins Struct Funct Bioinf 2007;67:305–16.
[15] Nardini M, Dijkstra BW. α/β Hydrolase fold enzymes: the family keeps growing. Curr Opin Struct Biol 1999;9:732–7.
[16] Verschueren KHG, Seljée F, Rozeboom HJ, Kalk KH, Dijkstra BW. Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. Nature 1993;363:693–8.
[17] Klvana M et al. Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations. J Mol Biol 2009;392:1339–56.
[18] Babkova P, Sebestova E, Brezovsky J, Chaloupkova R, Damborsky J. Ancestral haloalkane dehalogenases show robustness and unique substrate specificity. ChemBioChem 2017;18:1448–56.
[19] Sato Y et al. Crystallization and preliminary crystallographic analysis of a haloalkane dehalogenase, DbjA, from *Bradyrhizobium japonicum* USDA110. Acta Crystallogr, Sect F: Struct Biol Cryst Commun 2007;63:294–6.
[20] Chaloupkova R et al. Structural and functional analysis of a novel haloalkane dehalogenase with two halide-binding sites. Acta Crystallographica Section D 2014;70:1884–97.
[21] Newman J et al. Haloalkane dehalogenases: a structure of a *Rhodococcus* enzyme. Biochemistry 1999;38:16105–14.
[22] Gehret JJ et al. Structure and activity of DmmA, a marine haloalkane dehalogenase. Protein Sci 2012;21:239–48.
[23] Chrast L et al. Deciphering the structural basis of high thermostability of dehalogenase from psychrophilic bacterium *Marinobacter* sp. ELB17. Microorganisms 2019;7:498.
[24] Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545–9.
[25] Doerr S, De Fabritiis G. On-the-fly learning and sampling ligand of ligand binding by high-throughput molecular simulations. J Chem Theory Comput 2014;10:2064–9.
[26] Damborsky, J., Chaloupkova, R., Pavlova, M., Chovancova, E. & Brezovsky, J. in Handbook of Hydrocarbon and Lipid Microbiology (ed. Timmis, K. N.) 1081-1098 (Springer, Berlin, Heidelberg, 2010).
[27] Ingles-Prieto A et al. Conservation of protein structure over four billion years. Structure 2013;21:1690–7.
[28] Pikkemaat MG, Linssen ABM, Berendsen HJC, Janssen DB. Molecular dynamics simulations as a tool for improving protein stability. Protein Eng Des Sel 2002;15:185–92.
[29] Monincova M, Prokop Z, Vevodova J, Nagata Y, Damborsky J. Weak activity of haloalkane dehalogenase LinB with 1,2,3-trichloropropane revealed by X-ray crystallography and microcalorimetry. Appl Environ Microbiol 2007;73:2005–8.
[30] Guan L, Yabuki H, Okai M, Ohtsuka J, Tanokura M. Crystal structure of the novel haloalkane dehalogenase DatA from *Agrobacterium tumefaciens* C58 reveals a special halide-stabilizing pair and enantioselectivity mechanism. Appl Microbiol Biotechnol 2014;98:8573–82.
[31] Beerens K et al. Evolutionary analysis as a powerful complement to energy calculations for protein stabilization. ACS Catal 2018;8:9420–8.
[32] Harvey RB. Enzymes of thermal algae. Science 1924;60:481–2.
[33] Goldsmith M, Tawfik DS. Potential role of phenotypic mutations in the evolution of protein expression and stability. Proc Natl Acad Sci 2009;106:6197–202.
[34] Akanuma S et al. Experimental evidence for the thermophilicity of ancestral life. PNAS 2013;110:11067–72.
[35] Williams PD, Pollock DD, Blackburne BP, Goldstein RA. Assessing the accuracy of ancestral protein reconstruction methods. PLoS Comput Biol 2006;2. e69 e69.
[36] Vieille C, Zeikus GJ. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 2001;65:1–43.
[37] Okafor CD et al. Structural and dynamics comparison of thermostability in ancient, modern, and consensus elongation factor Tus. Structure 2018;26:118–29.
[38] Karshikoff A, Nilsson L, Ladenstein R. Rigidity versus flexibility: the dilemma of understanding protein thermal stability. The FEBS Journal 2015;282:3899–917.
[39] Tsuboyama K et al. A widespread family of heat-resistant obscure (Hero) proteins protect against protein instability and aggregation. PLoS Biol 2020;18:e3000632.
[40] Swulius MT, Waxham MN. Ca^{2+}/calmodulin-dependent protein kinases. Cell Mol Life Sci 2008;65:2637–57.
[41] Kamerzell TJ, Middaugh CR. The complex inter-relationships between protein flexibility and stability. J Pharm Sci 2008;97:3494–517.
[42] Risso VA et al. *De novo* active sites for resurrected Precambrian enzymes. Nat Commun 2017;8:16113.
[43] Whitney DS, Volkman BF, Prehoda KE. Evolution of a protein linteraction domain family by tuning conformational flexibility. J Am Chem Soc 2016;138:15150–6.
[44] Kabsch W. XDS. Acta Crystallographica Section D 2010;66:125–32.
[45] Winn MD et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 2011;67:235–42.
[46] Winter G. xia2: an expert system for macromolecular crystallography data reduction. J Appl Crystallogr 2010;43:186–90.
[47] McCoy AJ et al. Phaser crystallographic software. J Appl Crystallogr 2007;40:658–74.
[48] Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D 2010;66:213–21.
[49] Long F, Vagin AA, Young P, Murshudov GN. BALBES: a molecular-replacement pipeline. Acta Crystallogr D Biol Crystallogr 2008;64:125–32.
[50] Prokop Z et al. Enantioselectivity of haloalkane dehalogenases and its modulation by surface loop engineering. Angew Chem Int Ed 2010;49:6111–5.
[51] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 2010;66:486–501.
[52] Williams CJ et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci 2018;27:293–315.
[53] Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: high-throughput molecular dynamics for molecular discovery. J Chem Theory Comput 2016;12:1845–52.
[54] Harvey MJ, Giupponi G, De Fabritiis G. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. J Chem Theory Comput 2009;5:1632–9.
[55] Feenstra KA, Hess B, Berendsen HJC. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem 1999;20:786–98.
[56] Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-time-step molecular dynamics through hydrogen mass repartitioning. J Chem Theory Comput 2015;11:1864–74.
[57] Prinz JH et al. Markov models of molecular kinetics: generation and validation. J Chem Phys 2011;134:174105.
[58] Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol 2012;8. e1002708 e1002708.