


Insights into the molecular determinants of thermal stability in halohydrin dehalogenase HheD2

Julia Wessel¹, Giovanna Petrillo², Miquel Estevez-Gay³, Sandra Bosch⁴, Margarita Seeger¹, Willem P. Dijkman¹, Javier Iglesias-Fernández³, Aurelio Hidalgo⁴, Isabel Uson^{2,5}, Sílvia Osuna^{3,5} and Anett Schallmeyer¹ 

1 Institute for Biochemistry, Biotechnology and Bioinformatics, Technische Universität Braunschweig, Germany

2 Molecular Biology Institute of Barcelona, IBMB-CSIC, Barcelona, Spain

3 CompBioLab Group, Institut de Química Computacional i Catàlisi (IQCC), Departament de Química, Universitat de Girona, Spain

4 Centro de Biología Molecular 'Severo Ochoa', UAM-CSIC, Madrid, Spain

5 ICREA, Barcelona, Spain

Keywords

crystal structure; halohydrin dehalogenase; molecular dynamics simulation; structure-function relationship; thermal stability

Correspondence

A. Schallmeyer, Institute for Biochemistry, Biotechnology and Bioinformatics, Technische Universität Braunschweig, Braunschweig, Germany
 Tel: +49 531 391-55400
 E-mail: a.schallmeyer@tu-braunschweig.de
 Website: <https://www.tu-braunschweig.de/bbt/biochem>

Julia Wessel, Giovanna Petrillo and Miquel Estevez-Gay contributed equally to the work

(Received 23 December 2020, revised 7 February 2021, accepted 15 February 2021)

doi:10.1111/febs.15777

Halohydrin dehalogenases (HHDHs) are promising enzymes for application in biocatalysis due to their promiscuous epoxide ring-opening activity with various anionic nucleophiles. So far, seven different HHDH subtypes A to G have been reported with subtype D containing the by far largest number of enzymes. Moreover, several characterized members of subtype D have been reported to display outstanding characteristics such as high catalytic activity, broad substrate spectra or remarkable thermal stability. Yet, no structure of a D-type HHDH has been reported to date that could be used to investigate and understand those features on a molecular level. We therefore solved the crystal structure of HheD2 from *gamma proteobacterium* HTCC2207 at 1.6 Å resolution and used it as a starting point for targeted mutagenesis in combination with molecular dynamics (MD) simulation, in order to study the low thermal stability of HheD2 in comparison with other members of subtype D. This revealed a hydrogen bond between conserved residues Q160 and D198 to be connected with a high catalytic activity of this enzyme. Moreover, a flexible surface region containing two α -helices was identified to impact thermal stability of HheD2. Exchange of this surface region by residues of HheD3 yielded a variant with 10 °C higher melting temperature and reaction temperature optimum. Overall, our results provide important insights into the structure–function relationship of HheD2 and presumably for other D-type HHDHs.

Databases

Structural data are available in PDB database under the accession number [7B73](#).

Introduction

Halohydrin dehalogenases (HHDHs, also called haloalcohol dehalogenases or halohydrin hydrogenhalide-lyases) belong to the enzyme class of lyases (EC 4.5.1.-) and catalyse the reversible dehalogenation of vicinal haloalcohols with formation of the

corresponding epoxides [1]. In the reverse reaction, epoxide ring opening, a number of small anionic nucleophiles such as cyanide, nitrite, azide, cyanate or thiocyanate are accepted resulting in the irreversible formation of novel C-C, C-O, C-N or C-S bonds,

Abbreviations

HHDH, halohydrin dehalogenase; MD, molecular dynamics; RMSF, root mean square fluctuation; SDR, short-chain dehydrogenases/reductases; tICA, time-lagged independent component analysis.

which makes these enzymes very attractive for biocatalytic applications [2,3]. The most prominent example is probably their use in the synthesis of ethyl (*R*)-4-cyano-3-hydroxybutyrate, a chiral synthon for the production of statin side chains [4].

Halohydrin dehalogenases are members of the short-chain dehydrogenases/reductases (SDR) superfamily and share considerable sequence similarity and structural and mechanistic features with SDR enzymes, albeit catalysing chemically different reactions [5]. Based on a database mining approach using HDDH-specific sequence motifs [6], the number of available HDDHs was recently increased from merely a handful to > 70 [1] and is further increasing due to the tremendous growth of sequence information in public databases. Based on phylogenetic analyses, HDDHs are grouped into subtypes A through G [6], with subtype D containing the largest number of HDDH enzymes [1]. Members of different HDDH subtypes share sequence identities of 21–51% with the exception of subtypes B and D reaching up to 57% sequence identity. So far, crystal structures of members of HDDH subtypes A, B, C and G have been determined [7–10], whereas structures for enzymes of subtypes D, E and F are still missing. Structural comparison of available crystal structures of HDDHs revealed an overall similar fold to SDR enzymes and in each case also a homotetrameric assembly. There are, however, also subtype-specific differences observed such as the lack of a β -strand and an α -helix in HheB in comparison with HheA/A2, HheC and HheG [9]. The conformational landscape of some of these representative subclasses of HDDHs has also been computationally studied by means of molecular dynamics (MD) simulations [11]. Substantial deviations in the loops located adjacent to the active site and halide binding pockets were observed among HDDH subclasses, which were found to impact the available tunnels for epoxide binding to the active site. Furthermore, HheC features a C-terminal extension, which protrudes into the active site of an opposing monomer within the tetrameric quaternary assembly [7], which is not observed in other HDDHs. This C-terminal extension in HheC contains a tryptophan residue (W249), that was shown to contribute to the enzyme's stereoselectivity [12,13].

The biochemical characterization of 19 newly identified HDDHs, covering all phylogenetic subtypes, revealed interesting novel features not described for the previously known enzymes HheA/A2, HheB/B2 and HheC [14]. Thus, HheG and HheG2 from *Ilumato bacter coccineus* and *I. nonamiensis*, respectively, are able to convert cyclic epoxide substrates [10] and exhibit high regioselectivity for nucleophilic attack of the

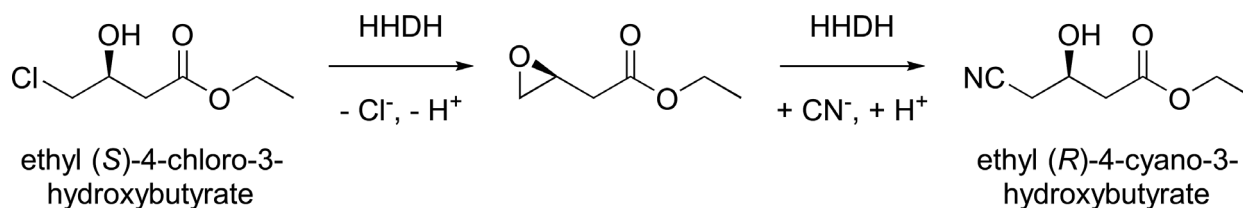
sterically more hindered carbon atom in the ring opening of styrene oxide derivatives [15]. The latter is in contrast to enzymes of all other HDDH subtypes, which display absolute regioselectivity for nucleophilic attack of the sterically less hindered carbon atom in the conversion of terminal epoxides [14]. This regioselectivity, however, varies if nonterminal epoxides are used as substrates. Thus, we could recently demonstrate that HDDHs belonging to subtype E as well as HheC and HheA3 are highly regioselective for nucleophilic attack of the sterically less hindered carbon atom of aliphatic nonterminal epoxides, whereas members of subtypes D and G are unselective forming both possible regioisomeric products in roughly equal amounts [16]. Moreover, D-type HDDHs seem to display a broader substrate scope in the conversion of aliphatic nonterminal epoxides than enzymes of subtypes A, B, C, E and F. Only HheG and HheG2 are even more promiscuous in terms of substrate scope [16].

During biochemical characterization of the new HDDHs, the HDDH HheD2 from *gamma proteobacterium* HTCC2207 caught our attention as it displayed very high catalytic activity in the synthesis of ethyl 4-cyano-3-hydroxybutyrate (**2**) starting from ethyl 4-chloro-3-hydroxybutyrate (**1**) (Scheme 1), whereas the other three characterized members of subtype D, namely HheD, HheD3 and HheD5, were 10-fold less active in this reaction [14]. In contrast, HheD2 exhibited only low thermostability ($T_{50}^{10} = 38$ °C), while respective T_{50}^{10} values of the other characterized D-type HDDHs were 17–24 °C higher [14]. Both observations are remarkable considering the high sequence identity of $\geq 67\%$ among all four enzymes. Hence, we aimed to investigate the thermal stability of HheD2 in comparison with other D-type HDDHs on a molecular level. For this, the crystal structure of HheD2 was determined, enabling a subsequent *in silico* analysis of this enzyme in combination with targeted mutagenesis. This ultimately allowed us to gain insight regarding the structure–function relationship of D-type HDDHs.

Results and Discussion

Overall structure and active site of HheD2

The crystal structure of HheD2 was solved at a resolution of 1.6 Å with the method implemented in ARCIMBOLDO_SHREDDER [17,18] (see [Materials and methods](#)), combining fragment-based molecular replacement with density modification. Despite a relatively low sequence identity among HheD2 and other HDDHs with known structure (see below for more details), both the overall fold and the oligomeric



Scheme 1. HDDH-catalysed conversion of ethyl (*S*)-4-chloro-3-hydroxybutyrate into ethyl (*R*)-4-cyano-3-hydroxybutyrate.

organization of the protein are conserved. Nevertheless, interesting differences in the active site, which have a relevant influence on the catalytic activity, substrate spectra and thermal stability, can be highlighted.

HheD2 exists as a tetramer with 222 point group symmetry (Fig. 1A) in the asymmetric unit. This is consistent with analytical size-exclusion chromatography results (data not shown), confirming that HheD2 also exists as a biological tetramer in solution. Also, the PISA program [19], which evaluated the energy of protein–protein interfaces observed in the tetramer as $-63 \text{ kcal}\cdot\text{mol}^{-1}$ and the buried area as $13\,000 \text{ \AA}^2$, confirms these contacts would be stable in solution. The

four chains (A–B–C–D) of each 223 residues (3–226) superimpose with rmsd between 0.11 and 0.19 for all C α atoms and are all equivalent within the experimental error.

In the HheD2 monomer, displayed in Fig. 1B, six β -strands (β 1– β 6) form a large parallel β -sheet structure, interacting through hydrophobic and electrostatic interactions of their side chains. A total of five α -helices (A, B, C, D and H), all antiparallel to the β -strands (Fig. S1), are distributed on both sides of the β -sheet, arranged in a characteristic Rossmann fold-like architecture (Fig. 1B). Two monomers form a dimer, in which the longest α -helices, α C and α D, and

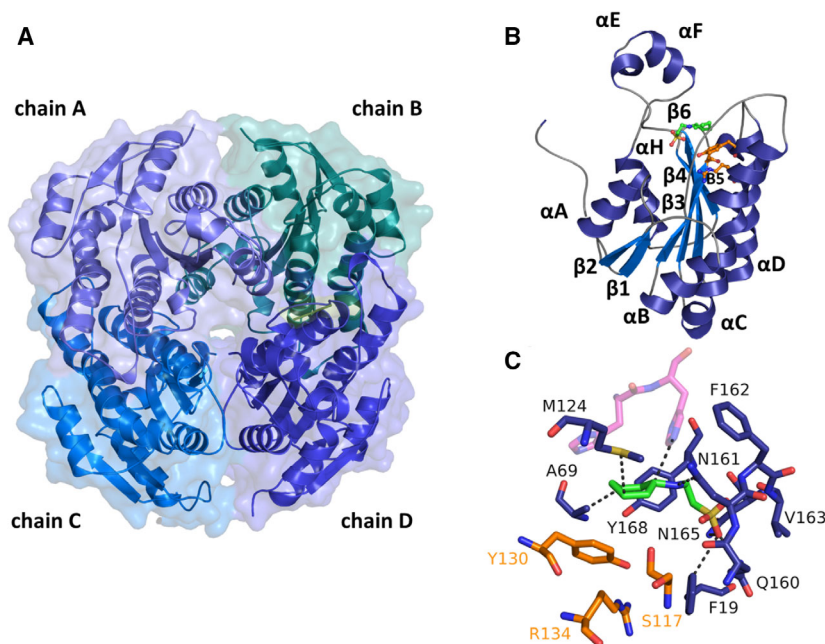


Fig. 1. (A) Cartoon and surface representation of tetrameric HheD2. The tetramer is formed by two kinds of interactions: those formed by two pairs of α -helices that build a four-helix bundle in the central horizontal plane and those in the surface formed along the vertical plane by the aromatic side chains from α H and β 6 and surrounding hydrophobic residues. (B) Cartoon view of the HheD2 monomer, showing the typical Rossmann fold-like architecture. The CHES molecule is shown as green sticks, and the residues constituting the catalytic triad are highlighted as orange sticks to locate the enzyme active site. (C) A detailed overview of the protein active site, showing the amino acids involved in CHES coordination (indicated by grey dotted lines) and the catalytic triad (highlighted in orange). In addition to the active site residues, two histidines (shown in magenta with transparency) from the purification tag of a neighbouring monomer are found. Representations of the structure were generated with PYMOL Molecular Graphics System version 2.4.1 (Schrödinger, LLC, New York, NY, USA).

part of their connecting loops associate into an intermolecular, antiparallel four-helix bundle. Two of these dimers are assembled into the tetrameric structure through interactions of aromatic side chains from α H and β 6 and surrounding hydrophobic residues. The same fold and oligomerization interfaces of HheD2 are seen in other members of the SDR family, which mainly occur as tetramers or dimers [20], as well as in the few functionally related HDDHs that have been structurally characterized so far. Although the amino acid sequence identity of HheD2 with HheA2 [8], HheB [9], HheC [7,13] and HheG [10] is relatively low (30%, 47%, 22% and 23%, respectively), the tertiary structure of HheD2 is very similar to that of these four proteins. When superimposed, the four monomeric structures show rmsd values of 1.32, 0.74, 1.34 and 1.36 Å, respectively, for all C α atoms.

One of the most remarkable differences is the lack of additional secondary structure elements: HheD2, as HheB, lacks a β -strand and an α -helix with respect to HheA2, HheC and HheG. Thus, in case of HheD2 the loop of Asp39 to Ala46 connects β 2 and α B. This region is surface-exposed and does not contribute to tetramer formation (Fig. 2). Another significant difference in HheD2 is the absence of the C-terminal extension that donates a tryptophan residue (Trp249) to the

active site of an adjacent subunit in HheC. The C-terminal end of each HheC monomer binds the twofold symmetry-related subunit that lies opposite in the other dimer [13]. This C-terminal extension is absent also in the HheA2, HheB and HheG structures. It is interesting to note that in our case, a comparable but not functional interaction across tetramers is seen through His -10 and His -11 in the N-terminal His-tag of monomer A. This interaction is induced by the crystallization.

Similar to all the other HDDH structures, the active site of HheD2 contains the Ser117-Tyr130-Arg134 catalytic triad. Tyr130 is hydrogen bonded to Arg134 in the next turn of helix α D. The residues are located in adjacent turns of one of the two helices that form one of the dimer interfaces in the HheD2 tetramer. The third residue of the catalytic triad, Ser117, is located at the tip of β -strand 4. Ser117 does not interact with other residues of the catalytic triad, but rather with the main and side chain of Ser119 through its hydroxymethyl group. This is comparable to the reported interaction of Ser132 and Thr134 in the structure of HheC. In the latter enzyme, exchange of Thr134 by alanine resulted in significantly (up to 11-fold) increased cyanolytic activity of the respective HheC mutant, while dehalogenation activity was positively

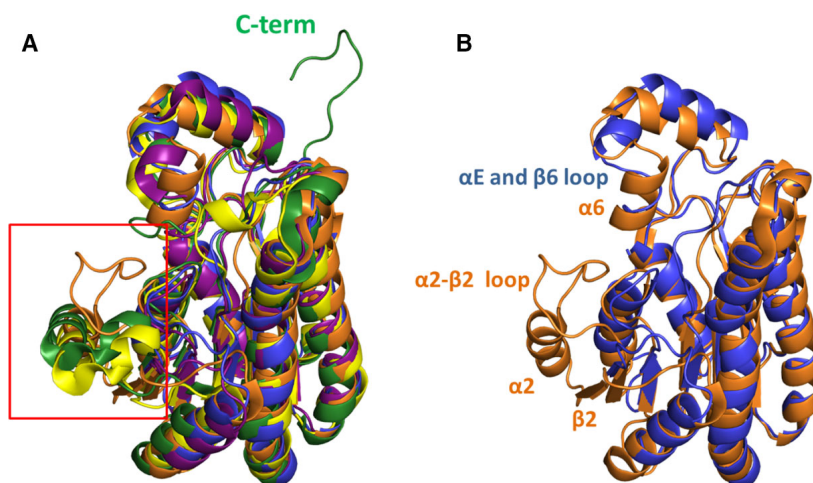


Fig. 2. (A) Structural comparison between HheD2 (blue) and related HDDHs, HheA2 from *Arthrobacter* sp. AD2 (yellow; PDB: 1ZMO), HheB from *Corynebacterium* sp. N-1074 (purple, PDB: 4ZU3), HheC from *Rhizobium radiobacter* (green, PDB: 1ZMT) and HheG from *I. coccineus* YM16-304 (orange, PDB: 5O30). In HheC, a C-terminal extension is present that protrudes into the active site of an opposite monomer within the tetrameric quaternary assembly. In HheD2 and the other related HDDH structures, the C-terminal extension is absent and it contributes to the much open active site conformation in these structures. In the red square is highlighted the position of the secondary structure elements (a β -strand and an α -helix) that are only present in HheA2, HheC and HheG. (B) Superimposition between HheD2 and HheG is shown. The secondary structural differences are depicted: HheD2, with respect to HheG lacks an α -helix (α 2), a β -strand (β 2) and the loop connecting them (β 2- α 2 loop). Moreover, HheG contains an α -helix (α 6), which is substituted by a loop in HheD2 (α E- β 6 loop). Representations of the structures were generated with PYMOL Molecular Graphics System version 2.4.1.

affected as well [21]. Introducing mutation Ser119Ala in HheD2, however, did not improve enzyme activity towards ethyl 4-chloro-3-hydroxybutyrate further (data not shown).

In the HheD2 structure, a CHES molecule from the crystallization buffer is present in each active site (Fig. 1C), with the charged sulfonate group occupying a position equivalent to that of the halide ions in the structures of HheA [9], HheB [9] and HheC [13] and equivalent to that of a water molecule in the structure of HheA2 [8]. In chain A, the CHES molecule forms, through its phenolic ring, interactions with Met124 (3.7 Å) and Tyr168 (4.2 Å), as well as His -10 (3.4 Å) and His -11 (4.3 Å) of the N-terminal His-tag. The amino group of the CHES molecule is H-bonded with the side chain of Asn161 (2.8 Å). Finally, it interacts through one oxygen of the sulfonate group with the side chain NH group of Gln160 (2.5 Å), through another oxygen with the side chain NH group of Asp165 (3.9 Å), and through the third oxygen of the sulfonate group with main chain NH groups of Val163 (3.2 Å) and Phe162 (2.9 Å) (Fig. 1C).

The HheD2 active site architecture is very similar to other HDDHs. The most notable difference regards the more open structure of the substrate-binding pocket of HheD2 in comparison with HheC. In fact, HheD2, lacking the C-terminal extension that contributes in HheC to the active site of the neighbouring subunit, has a more solvent-exposed active site similar to HheA2, HheB and HheG. Another important contribution to the openness of the active site of HheD2 is the presence of Met124, which is replaced by the bulkier Trp139 in HheC (Fig. 3). Trp139 is deemed to be responsible for the high enantioselectivity of HheC [12], precisely because the catalytic site is more restricted. Tyr168 of HheD2 (equivalent to residues Tyr185 in HheA2, Tyr169 in HheB, Phe186 in HheC and Phe203 in HheG), which is present in the loop between α E and β 6, is mainly responsible for delimiting the active site together with the residues of the catalytic triad and an additional conserved aromatic residue, Phe19 (equivalent to residues Phe12 in HheA2 and HheC, Tyr19 in HheB and Tyr18 in HheG), which is located in the loop between α A and β 1. In HheG, Phe203 is located in an α -helix (α 6) (Fig. 2B) instead of a loop, and its side chain is flipped away from the active site by $\sim 180^\circ$ with respect to the corresponding amino acid side chains in the other HDDH structures (Fig. 3), inducing a further opening of HheG's active site in comparison with HheD2.

The volume of the active site pocket of HheD2 was analysed using the software KVFinder [22] and compared with that of HheC and HheG, as shown in

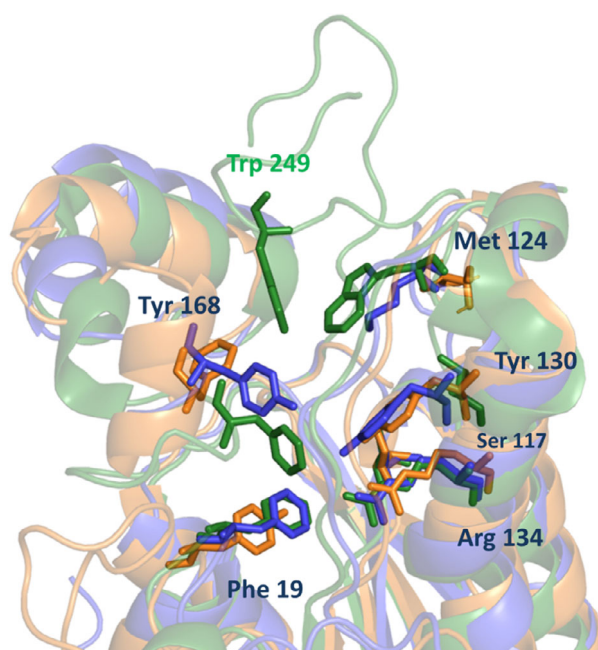


Fig. 3. Close-up of the active sites of HheD2 (blue, PDB: 5O30) and HheC (green, PDB: 1ZMT). The residues of the catalytic triad are shown as sticks (numbered according to the HheD2 sequence); their position and orientation are highly conserved among different HDDHs. The residues that delimit the active site are also depicted as sticks. In HheC, the C-terminal extension of a neighbouring subunit donates a tryptophan residue (Trp249, also in green) to the active site, contributing to delimiting the access to the catalytic pocket. Representations of the structures were generated with PYMOL Molecular Graphics System version 2.4.1.

Fig. 4. The cavity limits were plotted considering the amino acid residues that delimit the active site (Tyr168, Phe19 and the catalytic triad in case of HheD2). This revealed that the active site pocket of HheD2 is almost twice as big as that of HheC (350 vs. 180 Å³, respectively), while it is still a bit smaller compared with that of HheG (510 Å³). The bulky active site of HheD2 can explain why this HDDH, as well as other members of subtype D, displayed activity towards a broader range of nonterminal epoxides compared with most other HDDHs, except for HheG and HheG2 [16].

Targeted mutagenesis and characterization of thermostable variants

Based on HheD2's crystal structure, residues influencing the enzyme's stability were identified by the potential-based stability prediction algorithm FoldX [23]. The AlanineScan command compares the ΔG of each

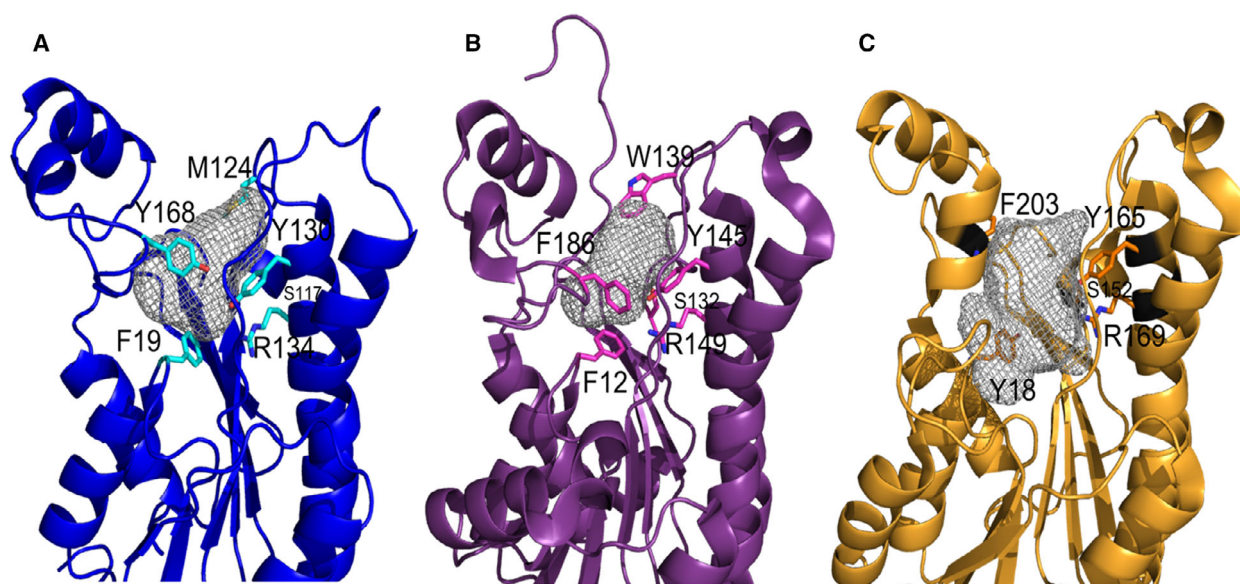


Fig. 4. Comparison of the volumes of the active site pockets (black mesh) in (A) HheD2 (B) HheC and (C) HheG. The structures are displayed as cartoon with the residues delimiting the active sites highlighted as sticks. Representations of the structures were generated with PYMOL Molecular Graphics System version 2.4.1.

residue with an alanine residue. Taking into account the homotetrameric structure of HheD2, the resulting $\Delta\Delta G$ values for each equivalent residue in the homotetramer were averaged. Residue D198 showed the lowest $\Delta\Delta G$ value ($-3.57 \text{ kcal}\cdot\text{mol}^{-1}$) indicating that it affects protein stability. Hence, this residue was replaced by the most abundant amino acids (isoleucine and valine with frequencies of 43.8% and 26.6%, respectively) and the amino acid with the lowest frequency (leucine, 2.3% frequency) at the structurally equivalent position in the HHDH family according to the 3DM database [24] on SDRs (Bio-Product, Nijmegen, the Netherlands).

Based on this approach, three HheD2 single mutants, D198I, D198L and D198V, were obtained exhibiting melting temperature (T_m) increases of up to 21 °C relative to the parental enzyme (Table 1). Interestingly, for all three mutants two distinct peaks were observed in the first derivative of their melting curves: a first minimum at 54 °C in all three cases and a second minimum at 60 °C (D198I), 63 °C (D198L) and 62 °C (D198V). In comparison, the wild-type (WT) enzyme gave only one distinct peak at 42 °C. Based on the homotetrameric assembly of HheD2, these distinct melting events could possibly reflect the dissociation of the tetramer and the denaturation of resulting dimers or monomers. In order to investigate the activity of the mutants at different temperatures, temperature profiles were determined in a temperature range of 20–80 °C for the dehalogenation of *rac*-ethyl-4-chloro-

3-hydroxybutyrate and compared with WT HheD2 (Fig. 5). Thus, mutants D198I and D198V exhibited a 5 °C higher temperature optimum compared with the WT, whereas D198L displayed the same optimum (40 °C) as the WT. For all three HheD2 variants, however, a significantly higher residual activity was found at temperatures ≥ 50 °C compared with the WT enzyme. Whereas only 10% residual activity at 50 °C could be detected for HheD2 WT, the residual activity of the mutants at this temperature ranged between 92% and 98%. This is a direct indication of the increased thermal stability of HheD2 variants D198I, D198L and D198V. Moreover, this improvement was not only observed for replacement of D198 by the most abundant amino acids, isoleucine and valine, at this position in the HHDH family, but also for replacement by the least abundant amino acid leucine. This suggests a common feature of I, V and L to be responsible for increased thermal stability.

Despite the increased thermal stability of the HheD2 variants, their specific activities ($\text{U}\cdot\text{mg}^{-1}$) in the dehalogenation of *rac*-ethyl-4-chloro-3-hydroxybutyrate were significantly reduced. Variant HheD2 D198V displayed the highest activity ($4.2 \text{ U}\cdot\text{mg}^{-1}$) among the three mutants. This dehalogenation activity, however, is already seven times lower in comparison with the WT enzyme (Table 1). Moreover, reduced activity of the mutants was also found in the cascade reaction for the production of ethyl 4-cyano-3-hydroxybutyrate from *rac*-ethyl-4-chloro-3-

Table 1. Overview of determined apparent melting temperatures (T_m) and specific activities of different HheD2 variants in the dehalogenation of *rac*-ethyl-4-chloro-3-hydroxybutyrate (**1**), as well as product formation and enantiomeric excess (ee_p) in the cascade reaction of **1** to ethyl 4-cyano-3-hydroxybutyrate (**2**) obtained after 2 h of reaction. The preferentially formed product enantiomer is given in parentheses. All values are the result of duplicate measurements given as mean ± SD.

HheD2 variant	T_m (°C)	Specific dehalogenation activity (U·mg ⁻¹)	Product 2 ^a (%)	ee _p (%)
WT	42 ± 0.2	31 ± 4.6	87 ± 0.6	1.6 ± 0.1 (S)
D198I	54 ± 0.0 / 60 ± 0.0	3.9 ± 0.6	22 ± 0.1	12 ± 1.1 (S)
D198L	54 ± 0.2 / 63 ± 0.2	2.5 ± 0.5	21 ± 1.8	16 ± 2.0 (S)
D198V	54 ± 0.2 / 62 ± 0.2	4.2 ± 0.1	36 ± 2.1	14 ± 0.0 (S)
D198S	53 ± 0.2	12 ± 1.5	80 ± 3.4	0.6 ± 0.3 (S)
helixD3	51 ± 0.2	18 ± 0.9	81 ± 1.8	1.5 ± 0.2 (S)
helixD5	43 ± 0.0	13 ± 0.7	56 ± 0.2	5.4 ± 0.4 (S)
helixD3-D198V	60 ± 0.0 / 68 ± 0.0	2.0 ± 0.1	7.3 ± 1.0	14 ± 3.1 (S)
helixD3-D198S	60 ± 0.0	5.2 ± 0.7	54 ± 0.7	3.1 ± 0.7 (R)

^aFor all mutants with an amino acid exchange at position 198, the double amount of enzyme compared with WT was applied in the cascade reaction.

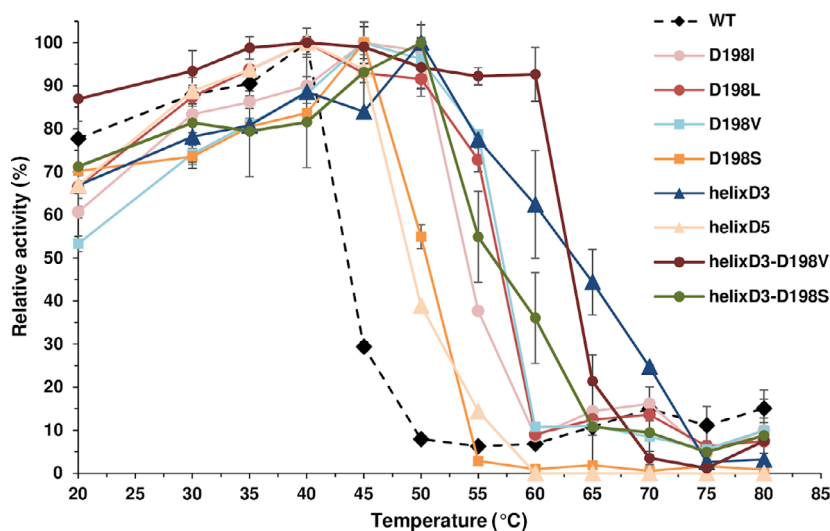


Fig. 5. Temperature profiles of HheD2 WT and generated variants. Dehalogenation reactions were incubated for 30 min at different temperatures, and activity was determined by quantification of released halide ions. The highest activity for each enzyme was set to 100%. All reactions were performed in duplicate ($n = 2$) with error bars indicating SD.

hydroxybutyrate (Table 1). WT HheD2 produced 87% of the cyanoalcohol within 2-h reaction time. In the same time, the mutants yielded three to four times less cyanoalcohol even though twice as much enzyme was used. In contrast, the enzymes' enantioselectivity was hardly affected. In all cases, a slightly higher enantiomeric excess for the (*S*)-cyanoalcohol was observed for the mutants, which is actually due to their lower activity. In fact, after 24 h of reaction using the HheD2 mutants, the determined product enantiomeric excesses were as low as for the WT (data not shown). Inspection of the HheD2 structure around D198 revealed hydrogen bond interactions between the side chains of D198 and Q160, the latter being H-bonded with the sulfate group of the cocrystallized CHES molecule and residues Gly116 and

Ala118, which are neighbouring residues of Ser117 of the catalytic triad (Fig. 6). Hence, this hydrogen bonding network might be important for correct positioning of Ser117 for catalysis. Moreover, the sulfate group of the CHES molecule adopts a similar position as halide ions present in the nucleophile binding pockets of HheA, HheB and HheC crystal structures. Hence, Q160 of HheD2 could be important for binding of the haloalcohol substrate in the dehalogenation reaction as well. In fact, mutagenesis of Q160 had a tremendous negative effect on enzyme activity. The only amino acid exchange at this position that still yielded a mutant with 20% residual activity was Q160N. The exchange of D198 by hydrophobic residues isoleucine, leucine and valine will result in a loss of the hydrogen bond between the side chains of

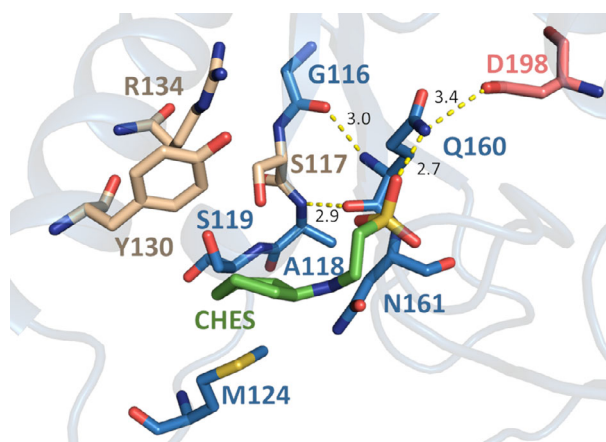


Fig. 6. Hydrogen bonding network between D198 and Q160 of HheD2 and the sulfate group of the CHES molecule and residues G116 and A118, surrounding S117 of the catalytic triad. For better representation of the active site, Y130 and R134 of the catalytic triad and other active site residues are shown. Distances are displayed in angstroms (Å). The figure was generated with PYMOL Molecular Graphics System version 2.4.0.

D198 and Q160, which might account for the significantly reduced activity and increased thermal stability of the corresponding mutants.

To investigate whether other amino acids at position 198 would also yield more thermostable variants, however, without negatively affecting activity, site saturation mutagenesis at position D198 was carried out. Ninety-four random clones were selected and screened in microtiter plate format using the thermofluor assay. Thus, variants D198N, D198S, D198R and D198Y were identified exhibiting a ≥ 10 °C increase in T_m compared with WT HheD2. Of these, only variant D198S displayed a higher dehalogenation activity ($12 \text{ U}\cdot\text{mg}^{-1}$, Table 1) in comparison with the previous mutants at position 198. Interestingly, for this variant only one distinct peak in the melting curve's first derivative was observed. The resulting T_m is almost identical to the first melting point of the previous mutants. Correspondingly, variant D198S is also faster inactivated at higher temperature exhibiting only 55% residual activity at 50 °C (Fig. 5). A higher activity of this variant compared with the previous mutants, D198I, D198V and D198L, was also observed in the cascade reaction from *rac*-ethyl-4-chloro-3-hydroxybutyrate to ethyl 4-cyano-3-hydroxybutyrate (Table 1). With mutant D198S, 80% product formation was achieved within 2-h reaction time and using the double amount of enzyme compared with WT HheD2. Hence, this result is in agreement with a 2.5-fold reduced

specific activity of mutant D198S compared with the WT enzyme. Also, mutation D198S had no influence on the enantioselectivity of the enzyme in this reaction. In comparison with the hydrophobic amino acids leucine, isoleucine and valine, serine carries a terminal hydroxyl group that still could form a hydrogen bond interaction with Q160, which might explain the observed higher activity of this variant compared with D198L, D198I and D198V. In fact, a closer look at a multiple sequence alignment of all currently known D-type HHDHs revealed that D198 and Q160 of HheD2 are strictly conserved among this subtype (Fig. 7). This likely indicates that the observed hydrogen bond between the side chains of D198 and Q160 of HheD2 is conserved among other D-type HHDHs as well (sequence identity within this subtype was determined to be $\geq 44\%$). This suggests that the presence of the hydrogen bond between Q160 and D198 cannot explain the observed lower thermal stability of HheD2 wild-type in comparison with HheD3 and HheD5, but a different molecular/structural feature needs to be responsible.

Another common approach for thermostabilization of enzymes with known crystal structure is to mutate residues exhibiting high B-factor values, which are indicative of a high flexibility within the crystal structure [26]. Inspection of the HheD2 structure revealed a flexible region spanning helices αE and αF (residues 170–186) based on B-factors. Interestingly, residues of this region are less conserved among D-type HHDHs according to the generated multiple sequence alignment (Fig. 7). Instead of separate mutagenesis of individual amino acids, we aimed for a complete exchange of the flexible region of HheD2 with the sequences of HheD3 and HheD5 (yielding variants helixD3 and helixD5, respectively), as these enzymes were previously shown to exhibit much higher thermal stability than HheD2 [14]. The resulting HheD2 variants helixD3 and helixD5 contained six (E171P-T173V-K174Q-N176D-K178R-D186E) and nine mutations (E171P-T173P-K174Q-N176T-K178A-Q180K-E181D-R185W-D186Q), respectively. Both variants were successfully produced in *E. coli* and purified. While exchange of the flexible region to the sequence of HheD5 resulted in a mutant (HheD2 helixD5) showing only insignificantly improved thermal stability compared with HheD2 WT, the specific activity of this mutant was reduced by half (Table 1). In contrast, introducing the HheD3 sequence at residues 170–186 had a great impact on the melting temperature ($\Delta T_m = +9$ °C) and the temperature optimum (10 °C higher compared with WT HheD2) of this variant (Table 1, Fig. 5). Moreover, variant helixD3 still displayed

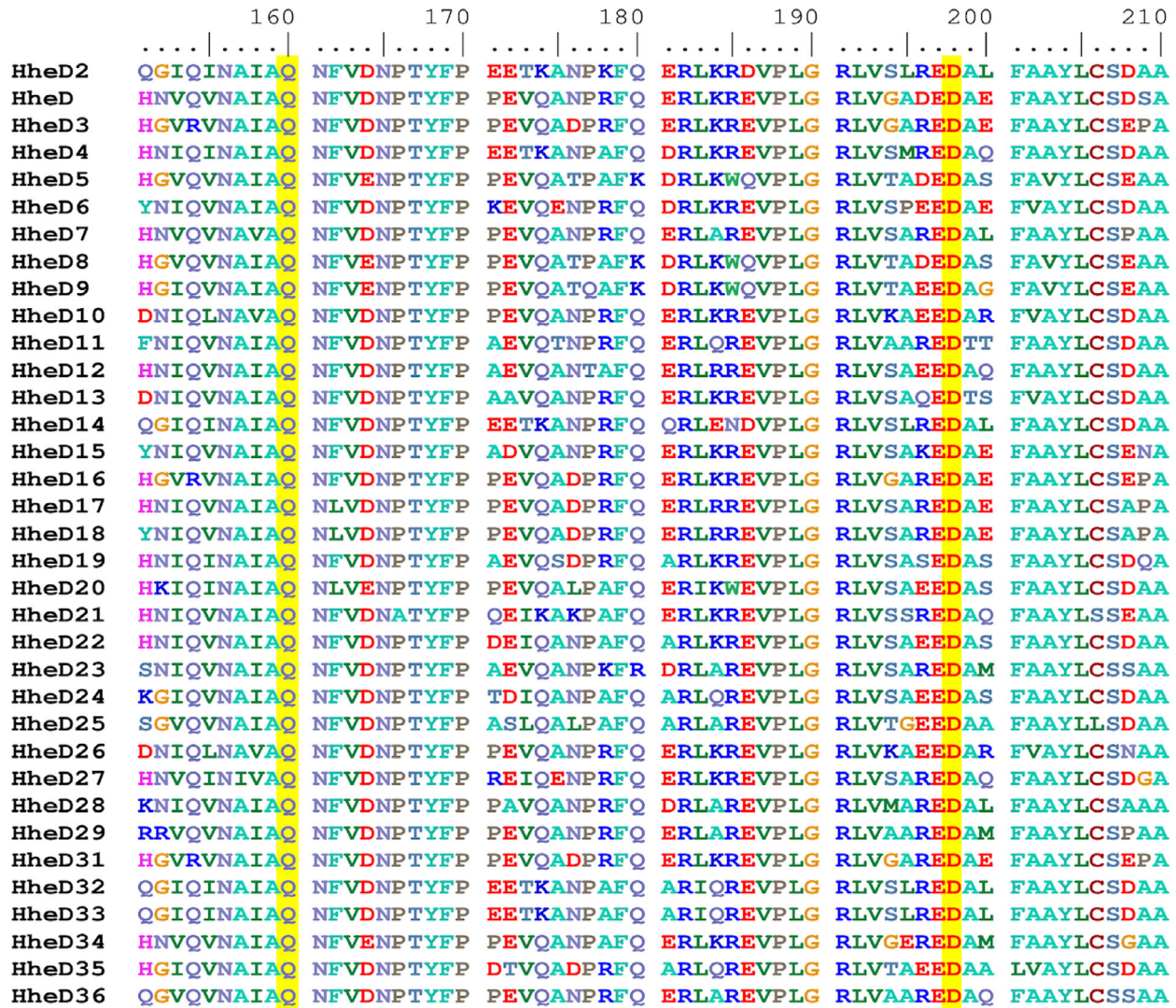


Fig. 7. Excerpt from a multiple sequence alignment of all currently known members of HHDH subtype D covering residues 151 to 210 (residue numbering according to HheD2). The multiple sequence alignment was generated with webPrank [25]. Conserved residues Q160 and D198 of HheD2 and the other D-type HHDHs are highlighted with a yellow background. Enzymes are named according to Schallmey *et al.* [1]; source organisms are as follows with respective GenBank accession numbers given in parentheses: HheD from *Dechloromonas aromatica* RCB (WP_011285856), HheD2 from *gamma proteobacterium* HTCC2207 (WP_007233072), HheD3 from *Methylobium petroleiphilum* PM1 (WP_011828277), HheD4 from marine metagenome (ECY18578), HheD5 from *Thauera* sp. MZ1T (WP_012585440), HheD6 from *Marinobacter nanhaiticus* D15-8W (WP_004579485), HheD7 from *Thauera* sp. 27 (WP_002926105), HheD8 from *Thauera aminoaromatica* S2 (WP_004302136), HheD9 from *Thauera phenylacetica* B4P (WP_004355811), HheD10 from *Limnohabitans* sp. Rim28 (WP_040506033), HheD11 from *Thiothrix disciformis* DSM 14473 (WP_020394200), HheD12 from *Pseudomonas pelagia* CL-AP6 (WP_022962804), HheD13 from *Betaproteobacteria* bacterium MOLA814 (WP_023472742), HheD14 from *Gammaproteobacteria* bacterium MOLA455 (WP_035490777), HheD15 from *Candidatus Competibacter denitrificans* Run_A_D11 (WP_048670059), HheD16 from *Methylobium* sp. T29 (WP_036236554), HheD17 from *Curvibacter gracilis* ATCC BAA-807 (WP_051443321), HheD18 from *Curvibacter lanceolatus* ATCC 14669 (WP_051107362), HheD19 from *Idiomarina salinarum* ISL-52 (WP_034774738), HheD20 from *Rhodocyclaceae* bacterium PG1-Ca6 (AJP48347), HheD21 from *Desulfatitalea* sp. BRH_c12 (KJS31884), HheD22 from marine sediment metagenome (KKN91043), HheD23 from marine sediment metagenome (KKN99199), HheD24 from marine sediment metagenome (KKO03046), HheD25 from *Ideonella sakaiensis* 201-F6 (GAP36263), HheD26 from *Limnohabitans planktonicus* II-D5 (WP_053169862), HheD27 from *Desulfatitalea tepidiphila* S28bF (WP_054029282), HheD28 from *beta proteobacterium* AAP51 (WP_054155488), HheD29 from *beta proteobacterium* AAP65 (WP_054139362), HheD30 from *Limnohabitans* sp. 63ED37-2 (ALK88134), HheD31 from *Methylobium* sp. Root1272 (WP_056325236), HheD32 from SAR92 bacterium BACL16 MAG-120619-bin48 (KRP19042), HheD33 from SAR92 bacterium BACL16 MAG-120322-bin99 (KRP26526), HheD34 from *Noviherbaspirillum* sp. Root189 (WP_057289066), HheD35 from *Pseudohongiella spirulinae* (WP_058022908) and HheD36 from *Pseudacidovorax intermedius* NS331 (WP_058643388).

remarkable activity in the dehalogenation reaction at 60 and 65 °C. Hence, it performed even better than the most stable single mutants D198L, D198I and D198V. Even though the specific dehalogenation activity of helixD3 was still lower than for HheD2 WT, a similar product yield as for WT was obtained with the helixD3 mutant in the cascade reaction from *rac*-ethyl-4-chloro-3-hydroxybutyrate to ethyl 4-cyano-3-hydroxybutyrate (Table 1). Interestingly, HheD3 was previously found to display the highest temperature optimum and a better residual activity at higher temperatures among all characterized D-type HDDHs [14]. Hence, our data suggest that the surface region spanning helices α E and α F is important for thermal stability.

MD simulation of thermostabilized HheD2 variants

To investigate on a molecular and structural level how position 198 and residues 170–186 influence the enzyme's thermal stability, MD simulations have been performed. D198V and helixD3 variants of HheD2 and the WT enzyme were each simulated at 27 and 57 °C for 2500 ns (five replicas of 500 ns), to yield a total simulation time of 15 μ s (7.5 μ s at each temperature). As a first approach to study the effect of the introduced mutations on the overall flexibility of the enzymes, the root mean square fluctuation (RMSF) (Fig. 8) was calculated for each enzyme system individually. This parameter suggests an increased flexibility at high temperatures of helices α E and α F, and the loop containing residues 68–74 for the helixD3 variant, with no other major differences found.

We further analysed the MD dataset by applying the dimensionality-reduction technique time-lagged independent component analysis (tICA) [27]. This method identifies those conformational changes that most rarely occur, thus allowing a better characterization of the slow conformational motions characteristic of the enzyme systems under study. The whole MD dataset coming from all enzyme trajectories was analysed together, and a common conformational space based on the two slowest tIC movements was constructed (see grey regions on Fig. 9). The different conformations visited along the individual MD trajectories for each enzyme were then projected into the common conformational space using a different colouring scheme. Initial MD conformations were coloured in purple, whereas final conformations were coloured in yellow (check the purple to yellow coloured dots in Fig. 9 for WT and helixD3 variant). This allowed us to characterize the common

conformational behaviour of all enzyme variants and to evaluate the different conformational states visited for each simulation (Fig. 9).

Our conformational analysis indicates that at high temperatures, the WT HheD2 system mainly explores conformations close to minima A, which is characterized by a large displacement towards the solvent of α E and α F halide-binding pocket helices (Figs 9A and 10D). Although the HheD2 catalytic triad arrangement remains unaltered, this conformational transition completely changes the halide-binding pocket shape towards nonoptimal substrate-binding conformations, thus explaining the dramatic decrease in activity observed experimentally at high temperatures. High temperature simulations of the helixD3 variant showed a much wider range of sampled conformations for this system, covering almost the full conformational landscape (Figs 9 and 10F). Similar to the WT system, this engineered variant is also able to explore minima A, thus explaining its reduced activity. Apart from minima A, an additional conformational state, characterized by the slow motion of the α E and α F helices towards residues 68–74, is observed (minima B). A prerequisite for this transition to happen is D198-Q160 hydrogen bond breakage, increasing loop flexibility and allowing the formation of new polar and hydrophobic interactions between residues 166–170 and 68–74 that stabilize conformation B. It is worth mentioning the new hydrogen bond interaction between residues P170 and T72 identified with the Arpeggio program [28]. The increase in thermal stability in energy minima B can be explained by a partial cover of the active site by residues 170–186, which increases the buried surface area of the enzyme [29,30]. Despite the huge rearrangement of helices α E and α F in minima B, the residues that compose the halide-binding site and the catalytic triad maintained similar positions than those observed in the X-ray structure, making possible the reaction to occur at 57 °C (Fig. 9). This is in line with the experimental activities observed for the helixD3 variant at high temperatures. Overall, our simulations indicate that the additional mutations introduced in the loop region 170–186 in variant helixD3 not only enabled the formation of new active site stabilizing interactions, but also completely remodelled the loop conformational dynamics.

In the D198V variant, the exchange of D198 by a hydrophobic valine residue disrupts the hydrogen bond interaction D198-Q160 observed in the crystal structure. Mutation D198V slightly increases the flexibility of the loop region 170–186 of HheD2, but is not able to stabilize energy minima B as in the case of variant helixD3 (Fig. 10B,E). The latter explains the reduced

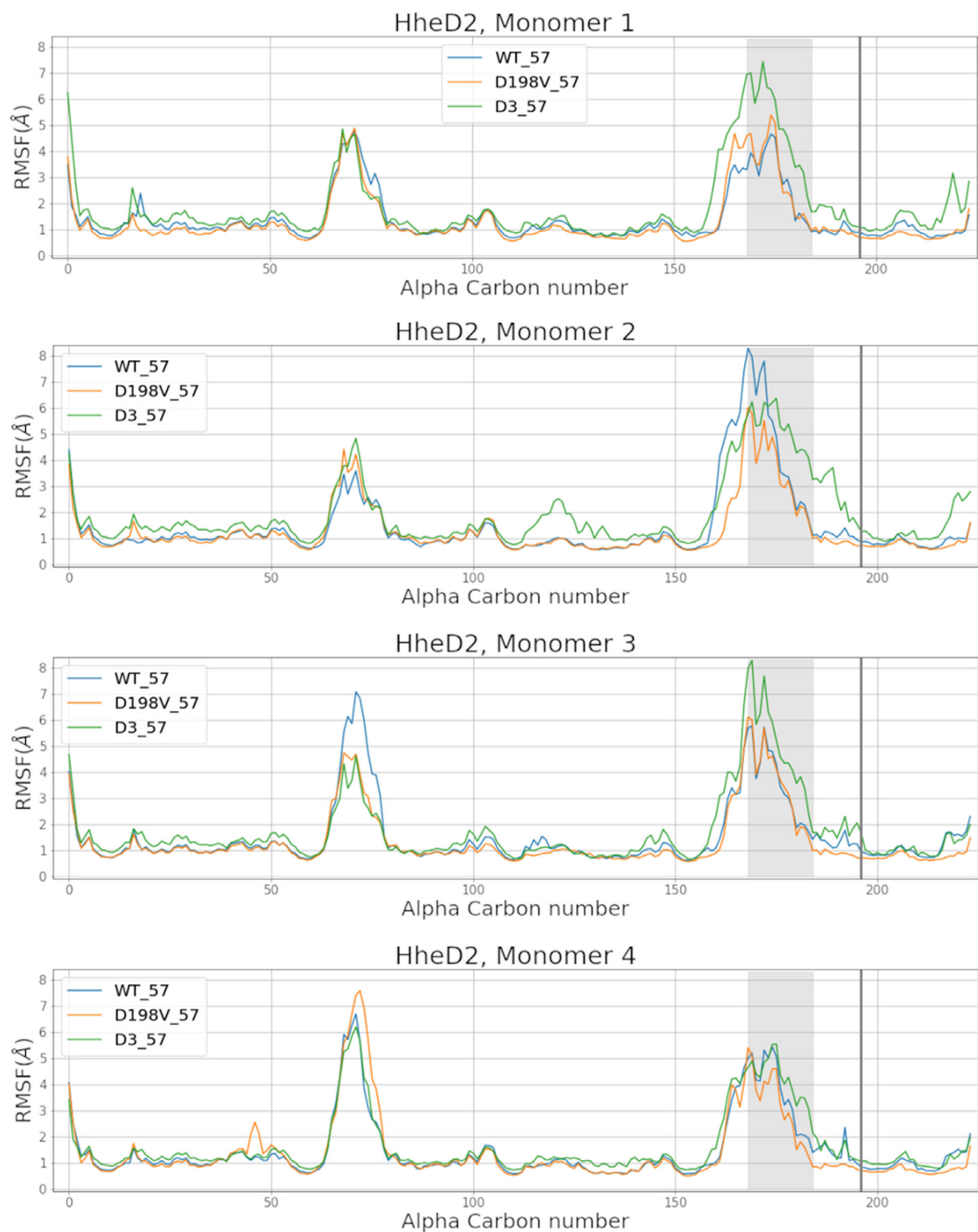


Fig. 8. RMSF of all α -carbons of the HheD2 systems at 57 °C during the simulations for all monomers. In blue is represented the WT data, in orange the D189V variant data and in green the helixD3 variant data. The position of residue 198 is marked using a black vertical line, and the position of α E and α F helices is marked using a grey shadow. No important differences between monomers are observed in this analysis. For α E and α F helices and the loop containing residues 68-74, a different mobility is shown.

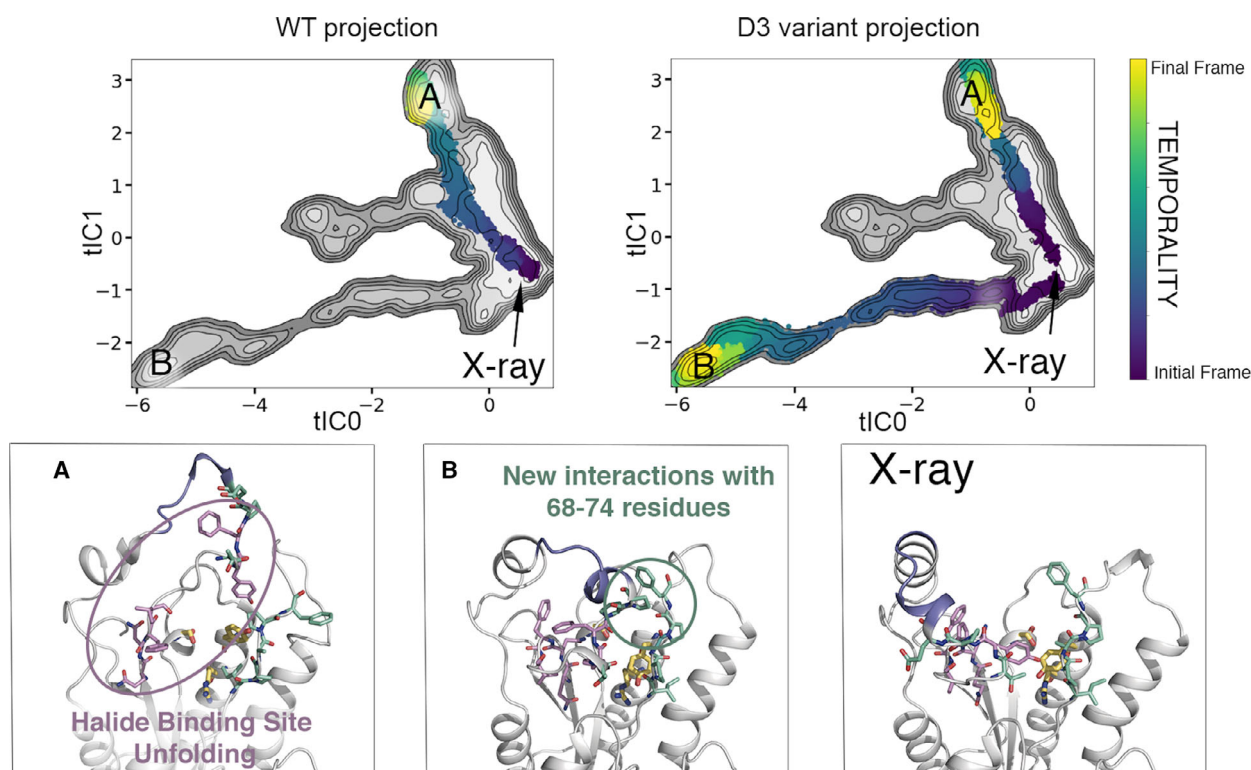


Fig. 9. (Top) Common conformational space (shown in grey) for WT and helixD3 variant reconstructed by applying the dimensionality-reduction technique tICA to the combined (WT and helixD3) MD trajectories at 57 °C (five replicas of 500 ns, i.e. 2500 ns for each system). The *x*- and *y*-axis correspond to the two first identified tIC components (tIC0 and tIC1), which describe the slowest kinetically relevant conformational changes observed along the MD runs. In particular, tIC0 describes the unfolding of the halide binding pocket, whereas tIC1 describes the slow motion of the α E and α F helices towards residues 68–74. The different conformations visited along one representative 500-ns MD simulation for WT (top left) and two 500-ns MD simulations for helixD3 (top right) at high temperature are projected on the common conformational space using coloured dots (colour range: initial MD frames are shown in purple, whereas final ones are shown in yellow). All simulations start at the X-ray structure (marked with a black arrow). In the WT trajectory (top left), the simulation evolves towards minima A. For helixD3 variant, one of the MD simulations also progresses towards A, whereas in the other MD trajectory, minima B is explored. (Bottom) Representative conformations extracted from energy minima A and B compared with the starting X-ray structure. Catalytic triad residues are displayed in yellow, the halide-binding site residues in purple, and in teal the residues that establish new interactions in the thermostable conformation. Representations of the structures were generated with PYMOL Molecular Graphics System version 2.4.0.

thermal stability of variant D198V at temperatures ≥ 60 °C as compared to helixD3 resulting in very low residual activity (Fig. 5). At the same time, variant D198V is not able at higher temperatures to sample energy minima A that presents catalytically nonproductive conformations of the halide binding pocket (Fig. 10E). This is in line with the higher catalytic activity of D198V as compared to HheD2 WT at temperatures between 45 and 60 °C.

As rearrangement of the loop containing helices α E and α F in variant helixD3 at higher temperatures requires the hydrogen bond between D198 and Q160 to break, we studied also the combination of mutation D198V and variant helixD3. The resulting combinatorial variant helixD3-D198V indeed displayed an even

higher melting temperature than variants helixD3 or D198V alone (Table 1), and exhibited still 90% relative activity at 60 °C (Fig. 5). This gain in stability, however, was again at the expense of catalytic activity resulting in a (slightly) less active enzyme than HheD2 variants D198V, D198I and D198L (Table 1). Interestingly, the temperature profile of the corresponding combinatorial variant helixD3-D198S, carrying a serine at position 198, indicated a lower stability for this variant compared with helixD3 under reaction conditions (Fig. 5), whereas the T_m value of helixD3-D198S with 60 °C is higher than that of helixD3. Moreover, the catalytic activity of this combinatorial variant is significantly reduced as well, but still 2.5 times higher than for helixD3-D198V (Table 1). This is in line with

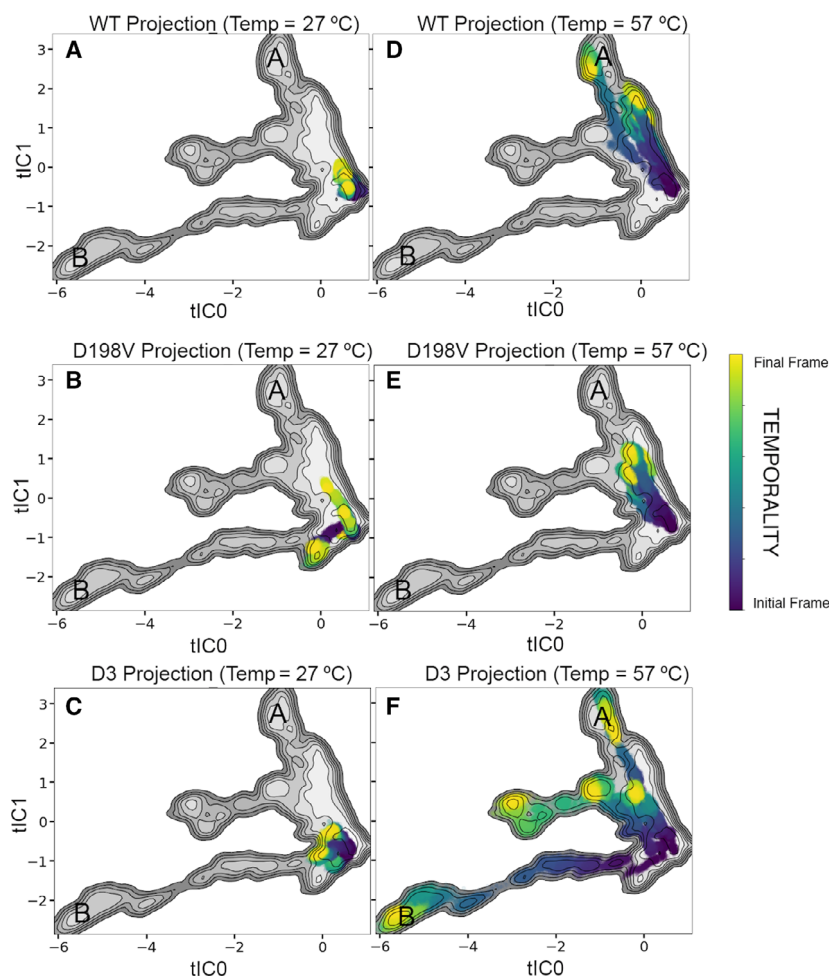


Fig. 10. Reconstructed conformational population analysis obtained considering all replicas from all studied systems (WT, D198V, and helixD3 variant), which is shown in grey scale. Time evolution along all MD simulations corresponding to all replicas at 27 °C for: (A) WT, (B) D198V and (C) helixD3; and at higher temperature (57 °C) for: (D) WT, (E) D198V and (F) helixD3. A colour range is used to denote the time evolution for all replicas: initial MD frames are shown in purple, whereas final ones are shown in yellow. Minima A and B are also indicated.

the observed higher activity of variant D198S compared with HheD2 D198V.

Conclusions

With our study, we provide important insights into the structure–function relationship of D-type HDDHs with special focus on thermal stability of HheD2. The herein reported crystal structure of this enzyme will serve as template for homology modelling and structure elucidation of other D-type HDDHs in the future, to facilitate further protein engineering campaigns of those enzymes. Moreover, the crystal structure of HheD2 confirmed the presence of a more spacious active site in enzymes of subtype D in comparison with HheC. This is in agreement with the reported broader substrate scope of characterized D-type HDDHs, enabling them to convert also various nonterminal epoxide substrates [16].

The observed hydrogen bond between side chains of Q160 and D198 in HheD2 is expected to be present in

other D-type HDDHs as well, due to the conservation of both residues in all known members of subtype D. This hydrogen bond in HheD2 was found to be important for catalytic activity as mutagenesis of either D198 or Q160 resulted in variants with significantly reduced catalytic performance. Moreover, MD simulations of HheD2 WT at higher temperature revealed a fast unfolding of the halide binding site and displacement of a surface loop region, explaining the comparably low thermal stability of this enzyme. Exchange of this surface region (spanning residues 170–186 in HheD2) by residues of HheD3 yielded a variant with significantly improved thermal stability. In this case, the loop conformational dynamics of variant helixD3 was remodelled in comparison with WT HheD2, permitting new stabilizing interactions between residues 166–170 and 68–74, while maintaining catalytic function. Considering the higher thermal stability of HheD3 compared with HheD2 WT, this result suggests that the surface loop region might impact enzyme thermal stability in other D-type HDDHs as well.

Hence, our data provide a good starting point for further thermostabilization campaigns of other members of HHDH subtype D.

Materials and methods

Chemicals

All used chemicals were of analytical grade or higher quality. Ethyl (*R*)-4-chloro-3-hydroxybutyrate and ethyl (*S*)-4-chloro-3-hydroxybutyrate were purchased from Sigma-Aldrich (Steinheim, Germany). All other chemicals were purchased from Sigma-Aldrich, AppliChem (Darmstadt, Germany), abcr GmbH or Carl Roth (Karlsruhe, Germany).

Bacterial strains and plasmids

Escherichia coli strains DH5 α and BL21 (DE3) Gold (Thermo Fisher Scientific, Darmstadt, Germany) were used as hosts for cloning and heterologous protein production, respectively. HheD2 WT (GenBank accession number: [KU501245](#)) and mutant genes were expressed from pET28a (+)-based vectors, utilizing a T7 promoter and resulting in an N-terminal hexahistidine (His₆) tag fusion [6].

Rational protein design

The crystal structure of HheD2 was used to analyse the influence of each residue on the structure stability of the enzyme using the FoldX server [23]. The structure was pre-processed using the RepairPDB command to minimize the energy of the structure. Then, each residue was mutated *in silico* to alanine and the energy difference for each mutation was calculated using the AlanineScan command. Based on the results, residue 198 yielding the lowest $\Delta\Delta G$ was further analysed in the 3DM database of the SDR superfamily provided by Bio-Product. A structural alignment of the HHDH subset within the 3DM database was used to identify the frequency of each amino acid at position 198.

Generation of HheD2 mutants

Single amino acid exchanges were introduced into the HheD2- or HheD2_helixD3-encoding gene of vector pET28a(+)/hheD2 and pET28a(+)/hheD2_helixD3, respectively, using the QuikChange™ Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA) according to the standard protocol provided by the manufacturer. The incorporation of these mutations was carried out using primers listed in Table S1.

The site saturation mutagenesis library at position D198 was prepared using the PfuUltra II Hotstart PCR Master Mix (Agilent Technologies) following the manufacturer's instructions. Hundred nanogram parental vector DNA,

125 ng respective forward (HheD2_D198SSM_fw; Table S1) and reverse primer (HheD2_D198SSM_rv; Table S1) were added to 25 μ L of PfuUltra II Hotstart master mix, and ddH₂O was added to a final volume of 50 μ L. After DNA amplification, 10 U DpnI restriction enzyme was added to the reaction mixture to digest the parental vector DNA and incubated up to 8 h at 37 °C. Chemically competent *E. coli* DH5 α cells were transformed with the PCR product. To ensure complete coverage of the generated diversity, > 4000 colonies from the transformation plates were resuspended in water and the plasmid was isolated using the E.Z.N.A. Plasmid DNA Mini Kit I (Omega Bio-tek, Inc., Norcross, GA, USA). Chemically competent *E. coli* BL21 (DE3) Gold cells were transformed with the resulting plasmid DNA, and 94 single colonies were picked in a 96-well microtiter plate together with a WT standard and an empty vector control. Cells were grown in 200 μ L TB media supplemented with 50 mg·L⁻¹ kanamycin at 37 °C and 900 r.p.m. overnight. For storage, 100 μ L of this overnight culture was mixed with an equal amount of 50% glycerol and stored at -80 °C.

Exchange of the flexible region of HheD2 (amino acids 170–186) by the respective sequences of HheD3 and HheD5 was carried out using the forward primers (HheD2_helixD3_fw and HheD2_helixD5_fw, respectively; Table S1) in combination with the same reverse primer (HheD2_helixD3/5_rv; Table S1). To generate a megaprimer for megawhop PCR, 30 ng template DNA (pET28a(+)/hheD2) and 125 ng of the respective forward and reverse primers were used in combination with the Q5® Site-Directed Mutagenesis Kit (NEB, Ipswich, MA, USA) following the standard protocol. After amplification, megaprimers were purified using the E.Z.N.A.® MicroElute Cycle Pure Kit (Omega Bio-tek, Inc.). Afterwards, megawhop PCRs were carried out using the purified megaprimers in a concentration of 200 ng, each 20 ng of template (pET28a(+)/hheD2) and the PfuUltra II Hotstart PCR Master Mix. The resulting vectors pET28a(+)/hheD2_helixD3 and pET28a(+)/hheD2_helixD5 were again purified using the E.Z.N.A.® MicroElute Cycle Pure Kit. 10 U DpnI was added and incubated up to 8 h at 37 °C to digest the parental vector DNA. Chemically competent *E. coli* DH5 α cells were transformed with 100–300 ng of this DNA. Correct sequence exchange was confirmed by sequencing. Chemically competent *E. coli* BL21 (DE3) Gold cells were transformed with both vectors, and heterologous expression of both HheD2 mutants was carried out as described above.

Expression and purification

For heterologous production of HheD2 WT and generated defined mutants, 500 mL TB media (4 mL·L⁻¹ glycerol, 12 g·L⁻¹ peptone, 24 g·L⁻¹ yeast extract, 0.17 M KH₂PO₄, 0.74 M K₂HPO₄) supplemented with 50 mg·L⁻¹ kanamycin was inoculated using 10% (v/v) of the respective overnight

culture, and protein expression was directly induced by adding 0.2 mM IPTG. After 24-h incubation (20 °C, 200 r.p.m.), cells were harvested by centrifugation (4400 g, 20 min at 4 °C) and cell pellets were stored at -20 °C until further use. Purification of resulting enzymes was performed by immobilized metal affinity chromatography using a 5 mL HisTrap HP column (GE Healthcare, Freiburg, Germany) and an Äkta FPLC system (GE Healthcare) according to a published protocol [14].

For protein expression of site saturation mutagenesis library in 96-deep-well plates, each 900 µL TB media supplemented with 50 mg·L⁻¹ kanamycin was inoculated with 10% (v/v) of a preculture. Expression was induced with 0.2 mM IPTG, and plates were incubated at 20 °C and 800 r.p.m. for 24 h. Cells were harvested (4400 g, 20 min at 4 °C) and stored at -20 °C until further use. For protein purification of the mutant library in MTP format, a His MultiTrap FF plate (GE Healthcare) was used. Cell pellets were resuspended in buffer A (50 mM Tris·SO₄, 500 mM Na₂SO₄, 25 mM imidazole, pH 7.9) containing 1 mg·mL⁻¹ lysozyme and 100 µM phenylmethylsulfonyl fluoride as protease inhibitor and incubated for 1 h at 30 °C and 950 r.p.m. The cell suspension was frozen once (-20 °C, 20 min), and afterwards, 50 µL DNase solution (0.1 mg·mL⁻¹ DNase in 20 mM MgSO₄) was added, followed by a thawing step at 30 °C (10 min, 950 r.p.m.). The resulting crude cell lysate was centrifuged (4400 g, 45 min, 4 °C) to remove cell debris. Purification of His-tagged HheD2 mutants in MTP format was carried out according to the manufacturer's instructions. Bound proteins were eluted using buffer B (50 mM Tris·SO₄, 500 mM Na₂SO₄, 500 mM imidazole, pH 7.9).

Protein concentration of purified enzymes was determined based on absorbance at 280 nm and the respective extinction coefficient.

Thermofluor assay and temperature optima

Apparent melting temperatures (T_m) of HheD2 WT and its mutants were determined by thermal unfolding using the thermofluor assay [31]. A mixture of 5× SYPRO Orange (Thermo Fisher Scientific, Waltham, MA, USA) and 30 µL purified but not desalted protein in buffer B (in case of HheD2 library screening) or 0.2 mg·mL⁻¹ purified and desalted protein in TE buffer (for defined HheD2 mutants) was measured in an iQ 96-well real-time PCR plate (Bio-Rad Laboratories, Munich, Germany) using a q-PCR machine (Bio-Rad Laboratories, CFX96™ Real-Time PCR Detection Systems) over a linear gradient from 10 to 90 °C in 0.5 °C steps. The final T_m was obtained from the local minimum of the negative first derivative of the measured relative fluorescence plotted versus the temperature.

Temperature profiles of HheD2 and selected mutants were determined using 25 µg·mL⁻¹ (HheD2 WT) or 50 µg·mL⁻¹ (HheD2 mutants) of purified enzyme and 5 mM racemic ethyl

4-chloro-3-hydroxybutyrate in 25 mM Tris SO₄ buffer, pH 7.0. The reaction mixture was incubated at temperatures ranging from 10 to 80 °C, and halide release was monitored after 30 min using the halide release assay (see below). Specific activities (µmol of formed product per mg of purified enzyme) were calculated as described below. The highest measured value per enzyme was set to 100%, and relative activities for the other temperatures were calculated accordingly. Negative control reactions without HHDH enzyme were performed in parallel using the same conditions.

Activity and stereoselectivity determination

To determine initial specific activities of HheD2 WT and its mutants in the dehalogenation of ethyl 4-chloro-3-hydroxybutyrate, the halide release assay was used as described previously [14]. In short, 5 mM *rac*-ethyl-4-chloro-3-hydroxybutyrate (prepared by mixing equal amounts of the commercially available enantiomers) and 25 µg (HheD2 WT and mutant D198S) or 50 µg (other HheD2 mutants) of enzyme were incubated at 30 °C in a total reaction volume of 1 mL in 25 mM Tris SO₄, pH 7. Hundred microlitre samples were taken after 10, 25, 40, 60, 90 and 120 s and mixed with 100 µL halide release assay reagent (mixture of equal volumes of solution I [0.25 M NH₄Fe(SO₄)₂ in 9 M HNO₃] and solution II [saturated solution of Hg(SCN)₂ in pure ethanol]). Absorbance of the samples at 460 nm was recorded, and the amount of released chloride was calculated using a calibration curve. The respective calibration curve for Cl⁻ was prepared in triplicate using KCl in a range of 0–2 mM. Specific dehalogenation activity of each enzyme was calculated in units per milligram protein (U·mg⁻¹) with one unit defined as the amount of enzyme that converts 1 µmol of substrate per minute.

Additionally, product formation for HheD2 WT and its mutants in the cascade reaction from *rac*-ethyl-4-chloro-3-hydroxybutyrate to ethyl 4-cyano-3-hydroxybutyrate was determined. Reactions were carried out in 1 mL total volume using 10 mM substrate, 25 µg (HheD2 WT and mutants helixD3/D5) or 50 µg (other HheD2 mutants) enzyme and 20 mM NaCN as nucleophile in 50 mM Tris SO₄, pH 8. After 2 and 24 h, each 400 µL sample was taken and extracted using 400 µL *tert*-butylmethylether (TBME) containing 0.1% v/v dodecane as internal standard. The resulting organic phases were dried over anhydrous MgSO₄ prior to injection on GC. Samples were analysed on achiral GC for the calculation of ethyl 4-cyano-3-hydroxybutyrate formation and on chiral GC for enantioselectivity determination.

Analytical methods

Achiral and chiral GC analysis was performed on a GC 2010 plus gas chromatograph (Shimadzu, Duisburg, Germany) equipped with an FID detector. An Optima 5ms

and a Lipodex E column (both from Macherey-Nagel, Düren, Germany) were used for achiral and chiral separation, respectively, of ethyl-(*R/S*)-4-chloro-3-hydroxybutyrate and ethyl-(*R/S*)-4-cyano-3-hydroxybutyrate. Please refer to Table S2 for applied temperature programmes and resulting retention times.

Crystallization and data collection

Crystals of HheD2 protein were grown at 20 °C by the sitting-drop vapour-diffusion method after mixing 2 μL of protein (13 $\text{mg}\cdot\text{mL}^{-1}$) with 2 μL of mother liquor (500 μL in total), containing 0.2 M lithium sulfate, 0.1 M CHES, pH 9.5, and 1 M Na/K tartrate. Bipyramidal crystals appeared within 3–4 days and grew to a size of 0.1–0.2 mm in the largest dimension. They were soaked with a cryoprotectant solution (crystallization solution plus 20–25% glycerol), before flash cooling to 100 K.

Suitable crystallization conditions were found by extensive screening of commercial chemical conditions (Hampton Research, Aliso Viejo, CA, USA) employing automated dispensing methods. For this, 96-well plates were dispensed by a nanodrop-dispensing robot (Cartesian model by Genomic solutions, Ann Arbor, MI, USA; or Phoenix by Rigaku Europe SE, Neu-Isenburg, Germany). Successful crystallization conditions were scaled up to μL volumes using 24-well plates, also in sitting-drop set-up.

Two data sets of HheD2 with a total of 900 images spanning 180° were collected in frames of 0.2° rotation and with an exposure time of 0.1 min, using a PILATUS 6 M detector and a fixed wavelength. The data sets were collected from the beamline XALOC at the ALBA synchrotron in Barcelona, Spain [32]. The protein crystallized in P212121 space group with unit-cell parameters $a = 76.2$, $b = 93.9$, $c = 140.5$, $\alpha = \beta = \gamma = 90.0^\circ$. Diffraction data and refinement statistics are summarized in Table 2.

Structure determination and refinement

Diffraction data were processed with the XDS program [33]. Intensities were scaled with XPREP [34] and truncated to structure factors F with C-truncate from the CCP4 suite [35]. The HheD2 structure was solved at a resolution of 1.6 Å using the homologous HHDH HheB (PDB 4ZD6; 47% sequence identity) [9], which was identified by HHPRED [36], as a template for ARCIMBOLDO_SHREDDER [18]. ARCIMBOLDO_SHREDDER performed eLLG-guided location of template-derived fragments with Phaser [37]. Additional degrees of freedom modified the fragments during molecular replacement through gyre refinement against the rotation function and gimble refinement after placement [38]. Consistent fragments were combined in reciprocal space with ALIXE [39]. The best scored phase sets were subject to density modification and auto-tracing with SHELXE [40] leading to a main chain trace

Table 2. Crystallographic statistics on data collection and refinement. Values within parentheses correspond to the last resolution shell.

	HheD2 WT
Data collection	
Source	XALOC
Wavelength (Å)	0.9699
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	78.0–94.8–140.2
α, β, γ (°)	90.0–90.0–90.0
A.U. content	one tetramer
Resolution (Å)	78.57–1.6 (1.7–1.60)
Unique reflections	137 444 (13 590)
Completeness (%)	99.95 (99.96)
Multiplicity	3.6 (3.6)
$I/\sigma I$	1.59
CC $\frac{1}{2}$ (%)	83 (47)
R _{meas} (%)	8 (79)
R _{pim} (%)	3 (33)
Refinement	
Resolution (Å)	47.4–1.60
No. of reflections	137 433 (13 589)
$R_{\text{work}} / R_{\text{free}}$	0.18/0.22
No. of atoms	
Protein	6967
Ligand	64
Water	722
B -factors	
Protein	26.88
Water	40.98
Ligand/ion	32.27
Rmsd	
Bond lengths (Å)	0.015
Bond angles (°)	1.87
PDB code	7B73

characterized by a CC of 34%. Finally, the model was built manually with COOT [41] and automatically refined with Phenix.refine [42] and BUSTER (Global Phasing Ltd., Cambridge, UK) [43]. The stereochemistry of the refined model is good, with all overall MolProbity scores above average [44].

The structure incorporates four molecules of CHES from the crystallization buffer, two molecules of glycerol from the cryo-buffer and 722 ordered water molecules. Electron density is clear and continuous from amino acids 3–226 in all four monomers, from which a complete structural model could be built; the first few N-terminal residues, presumably disordered, are missing. Residues labelled from –16 to –1, corresponding to the N-terminal His-tag introduced for purification, were also well ordered and thus traced in the structure of monomer A; in contrast, only residues from –14 to –9 and –12 to –9 are well ordered and traced in monomers B and C, respectively. The Ramachandran plot

shows 96% of all residues in the core and extended regions [45]. Outliers regard amino acid residues close to the active site; this hints at the flexibility required by the active site to undergo changes during catalysis and stabilize the reaction intermediates.

HheD2 diffraction data and coordinates were deposited in the Protein Data Bank under PDB: 7B73.

Computational MD methods

MD simulations in explicit water were performed using AMBER 16 package [46] in our in-house GPU cluster GALATEA (Nvidia GTX1080). Amino acid protonation states were predicted using the PropKa server. Then, the enzyme was solvated in a pre-equilibrated cubic box with a 10-Å buffer of TIP3P water molecules using the AMBER16 leap module. The systems were neutralized by addition of explicit counterions (Na⁺). All subsequent calculations were done using the David E. Shaw modification of the Amber 99SB force field (ff99SB-ILDN) [47].

A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on solute, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems are gently heated using six 50-ps steps, incrementing the temperature 50 K each step (0–300 K) under constant volume and periodic boundary conditions. Extra heating step of 30K was performed for the 330K MD simulations.

In order to control the temperature, Langevin thermostat was used. All systems were equilibrated without restraints for 2 ns at a constant pressure and temperature. After system equilibration, all MD simulations were performed under NVT ensemble, performing at 60 ns per day. In particular, five replicas of 500 ns were carried out for each system and temperature, adding up to 15 μs (7.5 μs at each temperature) of accumulated MD simulation time.

All analysis done was carried out in Jupyter-notebook environment (python), using *mdtraj*, *pytraj* (Ambertools 16) and *pyemma* libraries. tICA features included in this manuscript are alpha carbon coordinates.

Acknowledgements

We thank Marilyn Schlager for generation of the HheD2 site saturation mutagenesis library and Marcus Schallmey for help with the megaprimer PCR. Bio-Product is gratefully acknowledged for providing access to their 3DM database on SDR enzymes. A generous allocation of computing time at the Scientific Computation Center of the UAM (CCC-UAM) is also acknowledged.

This work has received funding from the European Union's Horizon 2020 research and innovation

programme under grant agreement No 635595 (CARBAZYMES). This work was further supported by grants PGC2018-101370-B-I00 (the Spanish Ministry of Science and Innovation and EU FEDER funds) and the Generalitat de Catalunya (2017SGR-1192). We also thank the Generalitat de Catalunya for the emerging group CompBioLab (2017 SGR-1707) and Spanish MINECO for project PGC2018-102192-B-I00. JIF was supported by the European Community for Marie Curie fellowship (H2020-MSCA-IF-2016-753045) and Juan de la Cierva-Incorporación fellowship (IJCI-2017-34129). SO is grateful to the funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (ERC-2015-StG-679001).

This communication reflects only the beneficiary's view, and the European Commission is not responsible for any use that may be made of the information it contains

Conflict of interest

The authors declare no conflict of interest.

Author contributions

AS, AH, IU and SO designed the study and planned experiments. JW, GP, ME-G, SB, MS and WPD performed experiments. JW, GP, ME-G, SB, MS, WPD and JI-F analysed data. JW, GP, ME-G and AS wrote the paper. AH, IU, SO and AS revised it critically.

Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1111/febs.15777>.

References

- Schallmey A & Schallmey M (2016) Recent advances on halohydrin dehalogenases - from enzyme identification to novel biocatalytic applications. *Appl Microbiol Biotechnol* **100**, 7827–7839.
- Hasnaoui-Dijoux G, Majeric Elenkov M, Lutje Spelberg JH, Hauer B & Janssen DB (2008) Catalytic promiscuity of halohydrin dehalogenase and its application in enantioselective epoxide ring opening. *ChemBioChem* **9**, 1048–1051.
- Findrik Blažević Z, Milčić N, Sudar M & Majeric Elenkov M (2020) Halohydrin dehalogenases and their potential in industrial application – a viewpoint of enzyme reaction engineering. *Adv Synth Catal* **362**, 388–410.

- 4 Ma SK, Gruber J, Davis C, Newman L, Gray D, Wang A, Grate J, Huisman GW & Sheldon RA (2010) A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chem* **12**, 81–86.
- 5 van Hylekama Vlieg JET, Tang L, Lutje Spelberg JH, Smilda T, Poelarends GJ, Bosma T, van Merode AEJ, Fraaije MW & Janssen DB (2001) Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases. *J Bacteriol* **183**, 5058–5066.
- 6 Schallmey M, Koopmeiners J, Wells E, Wardenga R & Schallmey A (2014) Expanding the halohydrin dehalogenase enzyme family: Identification of novel enzymes by database mining. *Appl Environ Microbiol* **80**, 7303–7315.
- 7 de Jong RM, Tiesinga JJW, Rozeboom HJ, Kalk KH, Tang L, Janssen DB & Dijkstra BW (2003) Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD(P)H binding site. *EMBO J* **22**, 4933–4944.
- 8 de Jong RM, Kalk KH, Tang L, Janssen DB & Dijkstra BW (2006) The X-ray structure of the haloalcohol dehalogenase HheA from *Arthrobacter* sp. strain AD2: insight into enantioselectivity and halide binding in the haloalcohol dehalogenase family. *J Bacteriol* **188**, 4051–4056.
- 9 Watanabe F, Yu F, Ohtaki A, Yamanaka Y, Noguchi K, Yohda M & Odaka M (2015) Crystal structures of halohydrin hydrogen-halide-lyases from *Corynebacterium* sp. N-1074. *Proteins Struct Funct Bioinforma* **83**, 2230–2239.
- 10 Koopmeiners J, Diederich C, Solarczek J, Voß H, Mayer J, Blankenfeldt W & Schallmey A (2017) HheG, a halohydrin dehalogenase with activity on cyclic epoxides. *ACS Catal* **7**, 6877–6886.
- 11 Estévez-Gay M, Iglesias-Fernández J & Osuna S (2020) Conformational landscapes of halohydrin dehalogenases and their accessible active site tunnels. *Catalysts* **10**, 1403.
- 12 Tang L, van Merode AEJ, Lutje Spelberg JH, Fraaije MW & Janssen DB (2003) Steady-state kinetics and tryptophan fluorescence properties of halohydrin dehalogenase from *Agrobacterium radiobacter*. Roles of W139 and W249 in the active site and halide-induced conformational change. *Biochemistry* **42**, 14057–14065.
- 13 de Jong RM, Tiesinga JJW, Villa A, Tang L, Janssen DB & Dijkstra BW (2005) Structural basis for the enantioselectivity of an epoxide ring opening reaction catalyzed by haloalcohol dehalogenase HheC. *J Am Chem Soc* **127**, 13338–13343.
- 14 Koopmeiners J, Halmschlag B, Schallmey M & Schallmey A (2016) Biochemical and biocatalytic characterization of 17 novel halohydrin dehalogenases. *Appl Microbiol Biotechnol* **100**, 7517–7527.
- 15 An M, Liu W, Zhou X, Ma R, Wang H, Cui B, Han W, Wan N & Chen Y (2019) Highly α -position regioselective ring-opening of epoxides catalyzed by halohydrin dehalogenase from *Ilumatobacter coccineus*: a biocatalytic approach to 2-azido-2-aryl-1-ols. *RSC Adv* **9**, 16418–16422.
- 16 Calderini E, Wessel J, Süß P, Schrepfer P, Wardenga R & Schallmey A (2019) Selective ring-opening of di-substituted epoxides catalysed by halohydrin dehalogenases. *ChemCatChem* **11**, 2099–2106.
- 17 Sammito M, Meindl K, de Ilarduya IM, Millán C, Artola-Recolons C, Hermoso JA & Uson I (2014) Structure solution with ARCIMBOLDO using fragments derived from distant homology models. *FEBS J* **281**, 4029–4045.
- 18 Millán C, Sammito MD, McCoy AJ, Ziem Nascimento AF, Petrillo G, Oeffner RD, Domínguez-Gil T, Hermoso JA, Read RJ & Usón I (2018) Exploiting distant homologues for phasing through the generation of compact fragments, local fold refinement and partial solution combination. *Acta Cryst* **D74**, 290–304.
- 19 Krissinel E (2015) Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res* **43**, W314–W319.
- 20 Jörnvall H, Persson B, Krook M, Atrian S, González-Duarte R, Jeffery J & Ghosh D (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **34**, 6003–6013.
- 21 Schallmey M, Jekel P, Tang L, Majerić Elenkov M, Höffken HW, Hauer B & Janssen DB (2015) A single point mutation enhances hydroxynitrile synthesis by halohydrin dehalogenase. *Enzyme Microb Technol* **70**, 50–57.
- 22 Oliveira SHP, Ferraz FAN, Honorato RV, Xavier-Neto J, Sobreira TJP & de Oliveira PSL (2014) KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinf* **15**, 197–204.
- 23 Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F & Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Res* **33**, W382–W388.
- 24 Kuipers RK, Joosten H-J, van Berkel WJH, Leferink NGH, Rooijen E, Ittmann E, van Zimmeren F, Jochens H, Bornscheuer U, Vriend G *et al.* (2010) 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins Struct Funct Bioinforma* **78**, 2101–2113.
- 25 Löytynoja A & Goldman N (2010) webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* **11**, 579.
- 26 Sun Z, Liu Q, Qu G, Feng Y & Reetz MT (2019) Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chem Rev* **119**, 1626–1665.
- 27 Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G & Noé FJ (2013) Identification of slow molecular

- order parameters for Markov model construction. *J Chem Phys* **139**, 015102.
- 28 Jubb HC, Higuero AP, Ochoa-Montaña B, Pitt WR, Ascher DB & Blundell TL (2017) Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J Mol Biol* **429**, 365–371.
- 29 Schallmeyer M, Floor RJ, Hauer B, Breuer M, Jekel PA, Wijma HJ, Dijkstra BW & Janssen DB (2013) Biocatalytic and structural properties of a highly engineered halohydrin dehalogenase. *ChemBioChem* **14**, 870–881.
- 30 Solarczek J, Klünemann T, Brandt F, Schrepfer P, Wolter M, Jacob CR, Blankenfeldt W & Schallmeyer A (2019) Position 123 of halohydrin dehalogenase HheG plays an important role in stability, activity, and enantioselectivity. *Sci Rep* **9**, 5106.
- 31 Ericsson UB, Hallberg BM, DeTitta GT, Dekker N & Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289–298.
- 32 Juanhuix J, Gil-Ortiz F, Cuni G, Colldelram C, Nicolás J, Lidón J, Boter E, Ruget C, Ferrera S & Benach J (2014) Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the Alba Synchrotron. *J Synchrotron Radiat* **21**, 679–689.
- 33 Kabsch W (2010) XDS. *Acta Cryst D* **66**, 125–132.
- 34 Sheldrick GM (2008) A short history of SHELX. *Acta Cryst A* **64**, 112–122.
- 35 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy AJ *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Cryst D* **67**, 235–242.
- 36 Söding J (2005) Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**, 951–960.
- 37 Oeffner RD, Afonine PV, Millán C, Sammito M, Usón I, Read RJ & McCoy AJ (2018) On the application of the expected log-likelihood gain to decision making in molecular replacement. *Acta Cryst D* **74**, 245–255.
- 38 McCoy AJ, Oeffner RD, Millán C, Sammito M, Usón I & Read RJ (2018) Gyre and gimble: a maximum-likelihood replacement for Patterson correlation refinement. *Acta Cryst D* **74**, 279–289.
- 39 Millán C, Sammito M, Garcia-Ferrer I, Goulas T, Sheldrick GM & Uson I (2015) Combining phase information in reciprocal space for molecular replacement with partial models. *Acta Cryst D* **71**, 1931–1945.
- 40 Usón I & Sheldrick GM (2018) An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features. *Acta Cryst D* **74**, 106–116.
- 41 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst D* **60**, 2126–2132.
- 42 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* **66**, 213–221.
- 43 Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek P, Roversi P, Sharff A, Smart OS, Vornrhein C *et al.* (2017) BUSTER Version 2.11.2. Global Phasing Ltd., Cambridge.
- 44 Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D* **66**, 12–21.
- 45 Kleywegt GJ & Brünger AT (1996) Checking your imagination: applications of the free R value. *Structure* **4**, 897–904.
- 46 Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W *et al.* (2016) AMBER 16. University of California, San Francisco, CA.
- 47 Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO & Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct Funct Bioinforma* **78**, 1950–1958.

Supporting information

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

Table S1. Mutagenic primers used in this study.

Table S2. Conditions used for GC analyses.

Fig. S1. HheD2 secondary structure topology diagram.