Chapter 2.
Biocatalytic and Structural Properties of a Highly Engineered Halohydrin Dehalogenase

Marcus Schallmey, Robert J. Floor, Bernhard Hauer, Michael Breuer, Peter A. Jekel, Hein J. Wijma, Bauke W. Dijkstra and Dick B. Janssen

Two highly engineered halohydrin dehalogenase variants were characterized in terms of their performance in dehalogenation and epoxide cyanolysis reactions. Both enzyme variants outperformed the wild-type enzyme in the cyanolysis of ethyl (S)-3,4-epoxybutyrate, a conversion yielding ethyl (R)-4-cyano-3-hydroxybutyrate, an important chiral building block for statin synthesis. One of the enzyme variants, HheC2360, displayed catalytic rates for this cyanolysis reaction enhanced up to tenfold. Furthermore, the enantioselectivity of this variant was the opposite of that of the wild-type enzyme, both for dehalogenation and for cyanolysis reactions. The 37-fold mutant HheC2360 showed an increase in thermal stability of 8 °C relative to the wild-type enzyme. Crystal structures of this enzyme were elucidated with chloride and ethyl (S)-3,4-epoxybutyrate or with ethyl (R)-4-cyano-3-hydroxybutyrate bound in the active site. The observed increase in temperature stability was explained in terms of a substantial increase in buried surface area relative to the wild-type HheC, together with enhanced interfacial interactions between the subunits that form the tetramer. The structures also revealed that the substrate binding pocket was modified both by substitutions and by backbone movements in loops surrounding the active site. The observed changes in the mutant structures are partly governed by coupled mutations, some of which are necessary to remove steric clashes or to allow backbone movements to occur. The importance of interactions between substitutions suggests that efficient directed evolution strategies should allow for compensating and synergistic mutations during library design.

This chapter has been published:

Chembiochem 2013, (14), 870–881
doi:10.1002/cbic.201300005
Introduction

Halohydrin dehalogenases (EC 4.5.1.–) catalyze the elimination of halides from vicinal haloalcohols, resulting in epoxide ring formation[1]. Although the natural role of these enzymes is in the dehalogenation of halogenated xenobiotics such as epichlorohydrin[2], they can also be used for the preparation of enantiopure haloalcohols and epoxides[3]. The halohydrin dehalogenase HheC from Agrobacterium radiobacter AD1 can be recombinantly produced in Escherichia coli in high yields, and this has allowed detailed biochemical and mechanistic investigation[4], as well as engineering studies[4–5]. The enzyme is active as a homotetramer[6] and belongs, together with other halohydrin dehalogenases, to the short-chain dehydrogenase/reductase (SDR) enzyme superfamily[7].

Halohydrin dehalogenases each possess a proton-abstracting catalytic triad (Ser-Tyr-Arg), which has a function similar to that of the commonly observed catalytic triad in SDR enzymes (Ser-Tyr-Lys). Halohydrin dehalogenases also exhibit the Rossmann-fold structural unit present in dehydrogenases, but lack the typical Gly-rich motif required for nucleotide cofactor binding[6]. Instead, halohydrin dehalogenases each contain a spacious anion binding pocket formed by a loop that—in SDR proteins—participates in substrate and cofactor binding. Mechanistic[8], structural[6, 9], and computational[10] studies have explained the observed high regioselectivity and enantiopreference of the HheC-type halohydrin dehalogenase for R-configured substrates.

HheC catalyzes haloalcohol dehalogenation in equilibrium with the reverse reaction of halide-mediated epoxide ring opening. As a promiscuous activity, HheC also catalyzes nucleophilic epoxide ring opening with various small anions such as cyanide, azide, nitrite, and cyanate[11]. Some examples of these biotechnologically interesting reactions are the preparation of optically active oxazolidinones[12] and (R)-azidoalcohols[13] as valuable optically active intermediates.

The formation of carbon-carbon bonds through enzyme-catalyzed epoxide cyanolysis was one of the first promiscuous activities observed for halohydrin dehalogenases and has been exploited for the production of optically pure β-hydroxynitriles such as (R)-γ-chloro-β-hydroxybutyronitrile[14]. Other aliphatic and aromatic epoxides have also been shown to be converted to the corresponding β-hydroxynitriles with good yields and selectivities[11a, 15]. In particular, chiral C3 and C4 compounds, such as 4-chloro-3-hydroxybutyric acid (1; Scheme 2-1), are very important building blocks for high-value pharma- and
nutraceuticals[^16]. Compound (R)-1 can be used as a precursor for the production of L-carnitine, whereas (S)-1 is the direct precursor for the preparation of the hydroxynitrile (R)-4-cyano-3-hydroxybutyric acid [(R)-2], which is used in the production of statins[^17].

![Scheme 2-1](image)

**Scheme 2-1.** Precursors for the statin side chain [compounds (S)-1 and (R)-2] and other halohydrin dehalogenase substrates or products (7 to 18).

Statins are a class of drug compounds that inhibit 3-hydroxy-3-methylglutaryl-coenzyme A reductase, the key enzyme that regulates the cholesterol biosynthesis pathway[^18]. One of the most widely prescribed drugs for the treatment of elevated blood cholesterol levels is Lipitor (atorvastatin calcium), with global annual sales exceeding US$ 10 billion[^19]. Most routes for the synthesis of the (3R,5S)-3,5-dihydroxyhexanoate (also as lactone) statin side chain employ (R)-2 as a chiral building block[^17,20]; this results in a demand for more than 100 000 kg (R)-2 each year[^19]. A variety of chemoenzymatic routes for the synthesis of optically pure (R)-2 have been explored[^17-18,20]. These routes involve whole cells or enzymes such as dehydrogenases, lipases, and nitrilases.

One possible process employs variants of a ketoreductase and a HheC-derived halohydrin dehalogenase in a three-step/two-enzyme process (Scheme 2-2) to produce optically pure ethyl (R)-4-cyano-3-hydroxybutyrate [(R)-6] from prochiral ethyl 4-chloroacetoacetate (3)[^19,21]. In this process, 3 is reduced by a ketoreductase variant from *Candida magnolia* to ethyl (S)-4-chloro-3-hydroxybutyrate [(S)-4]. In a second step, (S)-4 is converted into the corresponding ethyl (S)-3,4-epoxybutyrate [(S)-5] by use of an engineered HheC variant. Subsequently, (R)-6 is produced from (S)-5 by cyanolysis with use of the same HheC variant, yielding the product with 99.5% purity and in an enantiomeric excess exceeding 99.9%[^21a].

HheC variants have been tailored to the needs of this process by multiple rounds of directed evolution[^21a]. After 18 rounds of directed evolution, driven by establishing protein sequence activity relationships (ProSAR), the best variant contained at least 35 mutations[^19,21a]. Detailed biochemical properties of
individual process variants have not been published, except for patent disclosures of sequences encoding variants with 25 and 37 mutations and at least 1.5- to twofold activity improvement over the wild-type enzyme for the conversion of \((S)-4\) into \((R)-6\)\textsuperscript{[21b]}. Because the ProSAR algorithm applied by Fox and co-workers discards neutral or deleterious mutations\textsuperscript{[21a]} the substitutions present in the obtained best multisite mutants are highly likely to contribute to the observed phenotype.

**Scheme 2-2.** Biocatalytic route to optically pure hydroxynitrile \((R)-6\)\textsuperscript{[19,21b]}.

Knowledge of the biochemical and structural properties of a highly evolved enzyme variant would provide useful information on functional and structural changes that occur during extensive directed evolution. However, structures of laboratory-evolved enzymes containing large numbers of mutations are rare. Most structural data for enzymes improved by directed evolution studies are for mutants with no more than four amino acid substitutions per protein chain. Some exceptions are represented by the *Aspergillus terreus* acyltransferase variant LovD G5\textsuperscript{[22]}, the *Bacillus subtilis* lipase variant LipA 4D3\textsuperscript{[23]} the *Bacillus megaterium* P450 monooxygenase BM-3 variant 139-3\textsuperscript{[24]}, the *Aspergillus* sp. carboxylesterase variant NylB’ Hyb-S4M94\textsuperscript{[25]}, the p-nitrobenzyl esterase variant 8g8 from *B. subtilis*,\textsuperscript{[26]} and an *E. coli* aspartate aminotransferase variant ATB17\textsuperscript{[27]}, which contain between six and 17 amino acid exchanges per mutant. Structural information on such heavily engineered enzymes can also provide explanations for their modified properties, such as higher catalytic rates or increased thermal stabilities, and might contribute to the design of directed evolution libraries through, for example, adoption of strategies that enhance the possible occurrence of cooperative effects of mutations.

In this study, we have biochemically characterized two highly engineered HheC variants, each containing 37 mutations\textsuperscript{[21a]}. One variant, HheC2360, shows improved catalytic rates, enhanced temperature stability, and opposite enantioselectivity relative to its parent HheC. Furthermore, its X-ray structure was elucidated to assess the influence of structural changes on its kinetic properties and temperature stability.
Results and Discussion

Sequence analysis of improved HheC variants

HheC was tailored by Fox and co-workers for the improved conversion of (S)-4 to (R)-6 under industrial process conditions\[21a\]. As a result, a set of 1151 unique HheC sequences was described, with a maximum of 42 substitutions per sequence\[21b\].

To obtain insight into the sequence variation responsible for the improved catalytic properties, the substitutions occurring in the reported sequence population were analyzed. This analysis revealed that 153 of the 254 HheC residues are mutated in less than 1% of the cases (i.e., in 11 or fewer of the sequences) and that a further 49 residues are mutated in less than 5% of the reported HheC sequences. Of the remaining HheC residues, two subsets of 25 and 27 positions were classified either as frequently mutated (in >66% of the sequences) or moderately mutated (altered in 5–37% of the sequences), respectively. It was expected that the set of most frequent mutations would include the substitutions that are of major importance for the relevant changes in properties of HheC, such as the reported improvement of at least 1.5-fold in the conversion of (S)-4 to (R)-6\[21b\].

The 25 most frequent mutations (Q37H, K38Q, K52I, Y70L, Q72H, F82A, A83P, P84V, F86W, Q87R, G99D, A100T, R107K, T134A, T146S, C153S, T154A, G174A, E181G, F186Y, T189S, N195S, K203R, A222T, M245V) map to different parts of the enzyme structure. Only two mutations target buried residues (A100T, G174A), whereas the other 23 affected residues each have a solvent-exposed area of ≥5 Å². Furthermore, of these frequent substitutions, only F186Y is among the six first-shell residues within a 5 Å distance from the catalytic triad Y145-OH. Another 34 residues lie within 5–10 Å of the catalytic triad Y145-OH (second shell), and eight of these are frequent substitutions (F82A, A83P, P84V, F86W, A100T, T134A, T146S, G174A). The most frequently occurring mutations were analyzed further to obtain insight into the structural basis of the properties of the enzyme variants under study.

HheC2360 and HheC2656 were selected out of the reported sequences for detailed investigation because they each contain at least 24 of the mutations strongly enriched in the ProSAR procedure, but originate from different branches of the diversity that was generated\[21\]. Each of these variants contains 37 mutations in total. Variant HheC2360 contains all of the frequent mutations mentioned above with the exception of two different amino acid substitutions (A100M, T146A). It also contains 12 less frequent mutations (A60V, V75I,
V112A, K121R, P135S, Y166H, Y177G, L178V, H179D, V201W, V205Y, I246V). Variant HheC2656 differs from HheC2360 in its sequence at 18 positions (K10L, E95G, A100, V112, P135, T146S, A152T, Y166, S180T, A152T, Y177F, L178, S180T, H201, V205, W238T, G251S, M252V). As in HheC2360, the majority of the less frequently occurring mutations in HheC2656 are positioned in or flank structural regions that interact either with neighboring monomers (E95G, K121R, A152T, Y177F, H179D, W238T) or with the C-terminal extension (Y177F, G251S, M252V) of neighboring subunits in the tetrameric assembly.

**Biochemical characterization of HheC2360 and HheC2656**

Synthetic genes for HheC2360 and HheC2656 were expressed in a recombinant fashion in *E. coli* MC1061 from a pBAD-based expression vector at similar levels as HheC. After purification of the enzymes, yielding samples that gave single bands on SDS-PAGE, catalytic activities for dehalogenation and cyanolysis were measured. Most data were compared with those for the variant HheC-C153S, which is wild-type HheC containing a single mutation that makes the enzyme less susceptible to oxidative damage, but otherwise has similar properties\[4,8\].

The dehalogenation of the chloroalcohol ester \((S)^{-4}\), which yields the corresponding \((S)^{-5}\) as an intermediate, is the first HheC-catalyzed reaction in the production of \((R)^{-6}\). The catalytic constants for this dehalogenation reaction with \((R)^{-}\) and \((S)^{-4}\), as well as those with several other haloalcohols, were obtained by measuring initial halide release rates. Data were measured with both mutants and were compared with data for HheC-C153S (Table 2-1). The HheC-C153S enzyme exhibited steady-state parameters similar to previously determined constants for the wild-type enzyme. The \(k_{\text{cat}}\) values were in the range of 2.7 to 35.1 s\(^{-1}\) per active site for all tested substrates. For achiral 1,3-dichloropropan-2-ol \(7\) and 1,3-dibromopropan-2-ol \(8\), as well as for \((R)^{-4}\), the \(K_M\) values were lower than the lowest tested haloalcohol concentration of 0.05 mM. For the other haloalcohols, the \(K_M\) values were in the 0.16 to 4 mM range. Because of the very low \(K_M\) values with the two achiral dihalopropanols \(7\) and \(8\) and with \((R)^{-4}\), the highest catalytic efficiencies were found with these substrates, whereas the lowest catalytic efficiency was observed for rac-3-chloropropene-1,2-diol \((\text{rac-11})\), which has the highest \(K_M\) value. Surprisingly, both enantiomers of \(4\) showed low \(K_M\) values; this means that the wild type already recognizes both enantiomers of this substrate quite well. Previously, the HheC wild type was applied in the conversion of the methyl ester of \(1\) with stoichiometric yields but no catalytic constants were reported\[28\].
Table 2-1. Comparison of kinetic constants for halide release.

<table>
<thead>
<tr>
<th>Haloalcohol</th>
<th>HheC-C153S</th>
<th></th>
<th></th>
<th>HheC2360</th>
<th></th>
<th>HheC2656</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_{cat}$</td>
<td>$K_M$</td>
<td>$k_{cat}/K_M$</td>
<td>$k_{cat}$</td>
<td>$K_M$</td>
<td>$k_{cat}/K_M$</td>
</tr>
<tr>
<td>(R)-4-chloro-3-hydroxybutyrate</td>
<td>4.0 ± 0.3</td>
<td>&lt;0.05[^a]</td>
<td>&gt;80[^a]</td>
<td>6.3 ± 0.2</td>
<td>2.8 ± 0.2</td>
<td>2.3</td>
</tr>
<tr>
<td>([R]-4)</td>
<td></td>
<td>(&lt;0.05[^a])</td>
<td>(&gt;80[^a])</td>
<td>6.3 ± 0.2</td>
<td>2.8 ± 0.2</td>
<td>2.3</td>
</tr>
<tr>
<td>(S)-4-chloro-3-hydroxybutyrate</td>
<td>2.68 ± 0.06</td>
<td>0.16 ± 0.01</td>
<td>16.8</td>
<td>8.3 ± 0.3</td>
<td>1.9 ± 0.2</td>
<td>4.4</td>
</tr>
<tr>
<td>([S]-4)</td>
<td></td>
<td>(0.16 ± 0.01)</td>
<td>(16.8)</td>
<td>8.3 ± 0.3</td>
<td>1.9 ± 0.2</td>
<td>4.4</td>
</tr>
<tr>
<td>1,3-dichloro-2-propanol</td>
<td>6.0 ± 0.3</td>
<td>&lt;0.05[^a]</td>
<td>&gt;120[^a]</td>
<td>4.8 ± 0.4</td>
<td>10.4 ± 1.9</td>
<td>0.5</td>
</tr>
<tr>
<td>(7)</td>
<td></td>
<td>(&lt;0.05[^a])</td>
<td>(&gt;120[^a])</td>
<td>4.8 ± 0.4</td>
<td>10.4 ± 1.9</td>
<td>0.5</td>
</tr>
<tr>
<td>1,3-dibromo-2-propanol</td>
<td>6.6 ± 0.3</td>
<td>&lt;0.05[^a]</td>
<td>&gt;132[^a]</td>
<td>7.8 ± 0.2</td>
<td>0.59 ± 0.05</td>
<td>13.2</td>
</tr>
<tr>
<td>(8)</td>
<td></td>
<td>(&lt;0.05[^a])</td>
<td>(&gt;132[^a])</td>
<td>7.8 ± 0.2</td>
<td>0.59 ± 0.05</td>
<td>13.2</td>
</tr>
<tr>
<td>rac-1-chloro-2-propanol</td>
<td>18.2 ± 0.6</td>
<td>0.7 ± 0.1</td>
<td>26.0</td>
<td>0.30 ± 0.01</td>
<td>19.6 ± 1.3</td>
<td>0.02</td>
</tr>
<tr>
<td>(rac-9)</td>
<td></td>
<td>(0.7 ± 0.1)</td>
<td>(26.0)</td>
<td>0.30 ± 0.01</td>
<td>19.6 ± 1.3</td>
<td>0.02</td>
</tr>
<tr>
<td>1-chloro-2-methyl-2-propanol</td>
<td>35.1 ± 0.5</td>
<td>0.53 ± 0.02</td>
<td>66.2</td>
<td>6.5 ± 0.1</td>
<td>16.5 ± 0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>(10)</td>
<td></td>
<td>(0.53 ± 0.02)</td>
<td>(66.2)</td>
<td>6.5 ± 0.1</td>
<td>16.5 ± 0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>rac-3-chloro-1,2-propanediol (rac-11)</td>
<td>3.17 ± 0.05</td>
<td>4.0 ± 0.2</td>
<td>0.8</td>
<td>&gt;0.2[^b]}</td>
<td>&gt;&gt;85.5[^b]}</td>
<td>0.002[^b]}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.0 ± 0.2)</td>
<td>(0.8)</td>
<td>&gt;0.2[^b]}</td>
<td>&gt;&gt;85.5[^b]}</td>
<td>0.002[^b]}</td>
</tr>
<tr>
<td>rac-3-bromo-1,2-propanediol (rac-12)</td>
<td>6.1 ± 0.1</td>
<td>0.16 ± 0.02</td>
<td>38.1</td>
<td>5.7 ± 0.2</td>
<td>22.6 ± 1.8</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.16 ± 0.02)</td>
<td>(38.1)</td>
<td>5.7 ± 0.2</td>
<td>22.6 ± 1.8</td>
<td>0.3</td>
</tr>
<tr>
<td>rac-2-chloro-1-phenylethanol (rac-13)</td>
<td>12.6 ± 0.3</td>
<td>1.3 ± 0.1</td>
<td>9.7</td>
<td>3.8 ± 0.4</td>
<td>5.3 ± 1.4</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.3 ± 0.1)</td>
<td>(9.7)</td>
<td>3.8 ± 0.4</td>
<td>5.3 ± 1.4</td>
<td>0.7</td>
</tr>
</tbody>
</table>

[^a] The $K_M$ was lower than the lowest measured haloalcohol concentration.
[^b] No substrate saturation according to Michaelis-Menten kinetics was observed at the highest measured haloalcohol concentration.

In conclusion, both HheC mutants exhibited elevated $K_M$ values relative to HheC-C153S for all tested haloalcohols (Table 2-1). These elevated $K_M$ values are not surprising because the mutants were obtained after screening for conversion of substrate concentrations as high as 20 mM (first and second tier) and >600 mM (third tier) ([S]-4[^21a]). The screening conditions thus lacked selective pressure for the evolution of HheC mutants with high affinities.

Mutant HheC2360 exhibited the highest increase in $k_{cat}$ (3.1-fold) for the target haloalcohol substrate ([S]-4), whereas for the opposite enantiomer ([R]-4) it showed only a 1.6-fold increase in $k_{cat}$. This increase results in a $k_{cat}$ value for ([S]-4) similar to those found with other tested haloalcohols ([7, 8, and rac-3-bromopropane-1,2-diol (rac-12)], the $k_{cat}$ values of which were little affected (0.8- to 1.2-fold changes). For rac-1-chloropropan-2-ol (rac-9), 1-chloro-2-methylpropan-2-ol (10), and rac-2-chloro-1-phenylethanol (rac-13) 3.3- to 60-fold drops in $k_{cat}$ relative to HheC-C153S were observed.

The $K_M$ values of mutant HheC2360 corresponded to affinities for the tested haloalcohols that were reduced four- to >200-fold relative to HheC-C153S.
For chloroalcohol rac-11, the increase in $K_M$ with the HheC2360 mutant enzyme was so high that no saturation according to Michaelis–Menten kinetics was observed, even at the highest substrate concentration tested (85.5 mM). As a consequence, the catalytic efficiencies ($k_{cat}/K_M$) of mutant HheC2360 for all haloalcohols tested were lower than those with HheC-C153S. The reduction in catalytic efficiency was lowest, however, for the target chloroalcohol (S)-4.

Mutant HheC2656 also showed increases in $k_{cat}$ values with most haloalcohols. Absolute values were 1.6 to 4.1 times higher than for mutant HheC2360. The $K_M$ values for mutant HheC2656 were also 2.8 to more than 350 times higher than those found with HheC-C153S. Here, too, catalytic efficiencies for all haloalcohols were lower than with HheC-C153S and, as with HheC2360, the target chloroalcohol (S)-4 exhibited the lowest relative drop in $k_{cat}/K_M$.

The enantiopreference of HheC-C153S with 4 is in favor of the $R$ enantiomer, with an apparent enantiomeric ratio of $E_R>4.8$. The mutant HheC2656 hardly discriminated between the enantiomers of 4 ($E_R=1.3$). In contrast, mutant HheC2360 exhibited an enantiomeric ratio of $E_S=1.9$, corresponding to a modest preference for (S)-4 over (R)-4. Even though the observed values are low, the enantiopreference of mutant HheC2360 is the opposite of those of HheC-C153S or mutant HheC2656. Enantiopreference towards $S$-configured haloalcohols and epoxides has not been observed in previous biochemical studies on HheC and its variants, but mutants with a preference for $S$-configured substrates were found in mutants of Arthrobacter sp. strain AD2[29].

The cyanolysis of epoxide ester (S)-5 by a HheC mutant is the last enzymatic step in the production of (R)-6. Apparent kinetic constants for cyanide with fixed initial concentrations of (S)-5 and several aliphatic epoxides were obtained for all three enzymes (Table 2-2). Enzyme HheC-C153S displayed modest cyanolysis apparent $k_{cat}$ values in the range 0.5 to 4.8 s$^{-1}$. Interestingly, the highest rate was obtained with target epoxide (S)-5. The apparent $K_M$ values for cyanide were in the 0.3 to 1.8 mM range for all epoxides, so the catalytic efficiencies did not exceed 3.7 s$^{-1}$mM$^{-1}$. Both 37-fold mutant enzymes displayed higher cyanolysis rates than HheC-C153S for all epoxides tested. The apparent $K_M$ values for cyanide were also increased for all tested epoxides, except for mutant HheC2656 converting epoxide (S)-5.

The apparent $k_{cat}$ values of epoxide cyanolysis observed with mutant HheC2360 were ten to 120 times greater than those observed with HheC-C153S. Whereas the highest absolute rate was 110 s$^{-1}$ for chiral (S)-1,2-epoxybutane [(S)-14]- the greatest relative increase was observed with rac-1,2-epoxyhexane
(rac-16), with the bulkiest aliphatic side chain. The apparent \( K_M \) values of mutant HheC2360 were 2.2 to 17.3 times larger than the HheC-C153S values. Accordingly, catalytic efficiencies of HheC2360 were 1.9 to 55 times larger than those of HheC-C153S for all tested epoxides.

<table>
<thead>
<tr>
<th>Epoxide</th>
<th>HheC-C153S</th>
<th>HheC2360</th>
<th>HheC2656</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( k_{cat} ) [s(^{-1})]</td>
<td>( K_M ) [mM]</td>
<td>( k_{cat}/K_M ) [s(^{-1}) mM(^{-1})]</td>
</tr>
<tr>
<td>ethyl (S)-3,4-epoxybutyrate [(S)-5]</td>
<td>4.8 ± 0.4</td>
<td>1.8 ± 0.4</td>
<td>2.7</td>
</tr>
<tr>
<td>rac-1,2-epoxybutane (rac-14)</td>
<td>1.1 ± 0.1</td>
<td>0.6 ± 0.1</td>
<td>1.8</td>
</tr>
<tr>
<td>(R)-1,2-epoxybutane (R)-14</td>
<td>1.1 ± 0.1</td>
<td>0.3 ± 0.1</td>
<td>3.7</td>
</tr>
<tr>
<td>(S)-1,2-epoxybutane (S)-14</td>
<td>1.3 ± 0.1</td>
<td>0.5 ± 0.1</td>
<td>2.6</td>
</tr>
<tr>
<td>rac-1,2-epoxy-2-methylbutane (rac-15)</td>
<td>2.5 ± 0.2</td>
<td>0.9 ± 0.2</td>
<td>2.8</td>
</tr>
<tr>
<td>rac-1,2-epoxyhexane (rac-16)</td>
<td>0.51 ± 0.03</td>
<td>1.4 ± 0.2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 2-2. Apparent kinetic constants for cyanide obtained with various epoxides in ring-opening reactions.\(^{(a)}\)

\(^{(a)}\) All kinetic constants for the cyanolysis reaction were determined using a fixed starting concentration (20 mM) of epoxide.

Mutant HheC2656 showed apparent \( k_{cat} \) values 3.6 to 58 times higher than those obtained for HheC-C153S. The apparent \( K_M \) values with the mutant HheC2656 and rac-14 to rac-16 were 1.2 to 7.0 times higher than those with HheC-C153S. For target epoxide (S)-5, a slightly reduced apparent \( K_M \) (1.3-fold lower) was even observed. Variant HheC2656 displays higher affinities than mutant HheC2360 towards all tested epoxides; this seems to come at the expense of catalytic rates, which are higher for HheC2360.

To explore the enantiopreferences of all three enzymes, catalytic efficiencies of cyanolysis reactions were determined with (R)- and (S)-14. Mutant HheC2360 shows an apparent \( E_S \) of 1.2, indicating a very small preference for S-configured 14. In contrast, the determined apparent enantiomeric ratios of \( E_R \)=1.4 and \( E_R \)=1.8 indicated slight preferences for (R)-14 for HheC-C153S and HheC2656, respectively. The directed evolution of HheC-C153S towards HheC2360 thus also inverted the enantiopreference in cyanolysis, whereas both HheC-C153S and HheC2656 exhibit the \( R \) enantiopreference for dehalogenation shown earlier for the cyanolysis of the methyl ester of 1 by HheC\(^{[28]}\).
In summary, mutant HheC2656 appears to be a rather generalist enzyme, giving higher dehalogenation rates than HheC2360. Although the cyanolysis rates of mutant HheC2656 were not as high as those observed for HheC2360, HheC2656 is still a good catalyst for various cyanolysis reactions, as illustrated by the low apparent $K_m$ values for cyanide. In contrast, mutant HheC2360 displayed significantly higher cyanolysis activities and exhibited an enantiopreference for $S$-configured molecules both for the dehalogenation and for the cyanolysis reactions. In the statin side chain production process described above, the chirality of $(S)$-5 is introduced by reduction of 3 to $(S)$-4 with an $S$-selective ketoreductase\textsuperscript{19, 21a}. This reaction is followed by dehalogenation of $(S)$-4 to yield $(S)$-5 with the HheC2360 mutant halohydrin dehalogenase, which displays catalytic efficiency for $(S)$-4 about twice that for $(R)$-4. Because the $k_{cat}$ and $k_{cat}/K_m$ values of mutant HheC2360 in cyanolysis of $(S)$-5 to $(R)$-6 are more than 5.6 and 2.6 times higher than the corresponding values for the conversion of $(S)$-4 to epoxide $(S)$-5, the epoxide intermediate should be efficiently removed from a chloroalcohol/epoxide reaction mixture by rapid conversion to the desired product $(R)$-6. The better $S$ activity and a higher rate of cyanolysis that come from the combination of an $S$-selective ketoreductase variant and the HheC variant HheC2360 thus allow three enzymatic reaction steps that lead to formation of $(R)$-6 from prochiral 3 (Scheme 2-2). In view of its better catalytic properties, HheC2360 was selected for structural analysis.

**Crystal structure of HheC2360**

To explain the greatly enhanced epoxide cyanolysis rates and the observed inversion of enantiopreference, structures of mutant HheC2360 were solved by X-ray crystallography (Table 2-3). Three structures of the enzyme were obtained: HheC2360 with chloride bound (PDB ID: 4IY1), HheC2360 with both epoxide $(S)$-5 and chloride bound (PDB ID: 4IXW), and HheC2360 with hydroxynitrile $(R)$-6 bound in the active site (PDB ID: 4IXT). The structures were compared with wild-type HheC crystal structures.

The structures of the three mutant tetrameric proteins were superimposed on that of wild-type HheC (PDB ID: 1PWX), giving RMSD values of less than 0.7 Å for the equivalent $Cα$ atom positions. These low RMSD values indicate that the general structure of the tetrameric enzyme assembly is unaltered. Both wild-type HheC and mutant HheC2360 are tetramers, and the tetrameric assembly can be regarded as a dimer of A+B and C+D dimers. In the tetramer, subunit A is located opposite C, and B opposite D. Backbone differences
were only observed in the loops between A83 and W86 and between T134 to 
W139 and in the C-terminal loop, including the last C-terminal residues visible in 
the electron density. The three mutant protein structures were highly similar, 
with RMSD values for Cα atoms <0.3 Å. A similarly high degree of similarity 
between structures with substrate or product bound was also observed when 
comparing wild-type enzyme with (R)-styrene oxide [(R)-17] and chloride 
bound (PDB ID: 1PWZ) and with (R)-p-nitrophenyl-2-azidoethanol [(R)-18] 
bound in the active site (PDB ID: 1PX0)[6].

The HheC2360 structure with hydroxynitrile (R)-6 shows for the first 
time how a cyanate moiety is bound in the active site of a halohydrin 
dehalogenase. The cyanate binding mode in the mutant is very similar to that of 
the azido moiety of (R)-18 in the wild-type enzyme (PDB ID: 1PX0) and also to 
the chloride binding mode in HheC (PDB ID: 1PWZ). The nitrogen of the (R)- 
6 cyano group is thus at the same location as a chloride ion in the wild-type 
anion binding site. Furthermore, a water molecule that interacts with the halide 
in the wild-type is at the same position in the HheC2360 structure, where it is 
involved in a hydrogen-bonding interaction (2.9 Å) with the nitrogen atom of the 
cyano group. Such equivalence was suggested in previous studies,[6,10a] but the 
HheC2360 mutant structure 4IXT confirms these assumptions and provides 
structural information on cyanate binding by halohydrin dehalogenases.

Structural changes near the active site
Comparison of the catalytically important residues S132, Y145, and R149 in the 
mutant and wild-type enzyme structures showed no differences in their side-
chain orientations (RMSD<0.2 Å). The active-site cavity of HheC can be 
subdivided into three sites: 1) the site that interacts with the substrate hydroxy 
group (or epoxide oxygen), 2) the binding site for the anionic leaving group (or 
incoming nucleophile), and 3) the binding site for the epoxide ring substituent 
(e.g., phenyl in Figure 2-1A). Whereas only minor structural changes occur in the 
epoxide oxygen and anion binding sites (1 and 2), the epoxide ring-substituent 
binding region (3) has undergone significant changes in the HheC2360 mutant 
structure.

The shape of the HheC2360 active site is altered by a combination of 
several, sometimes coupled, mutations (Figure 2-1B). This change is largely 
caused by mutation P84V (see below) and concomitant backbone changes that 
sterically hinder binding of the epoxide ring substituent. Furthermore, mutation 
F86W affects the rotational freedom of W139, which creates room for a longer 
aliphatic substrate in the site that binds the epoxide ring substituent (site 3). The
highest relative increase in catalytic efficiency was found for the cyanolysis of the longest tested linear aliphatic epoxide 16 (50-fold); this suggests that the mutations create an active site geometry optimized for longer aliphatic substrates.

Table 2-3. Data collection and refinement statistics for HheC2360

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Chloride</th>
<th>(S)-3,4-epoxybutanoate ethylester and chloride</th>
<th>(R)-4-cyano-3-hydroxybutanoate ethylester</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam line (ESRF)</td>
<td>ID 23-2</td>
<td>ID 14-4</td>
<td>ID 14-4</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>40.4-2.10 (2.21-2.1)</td>
<td>63.51-2.47 (2.61-2.47)</td>
<td>73.36-2.47 (2.61-2.47)</td>
</tr>
<tr>
<td>Reflections</td>
<td>194915 (28224)</td>
<td>128124 (18553)</td>
<td>187079 (27496)</td>
</tr>
<tr>
<td>Space group</td>
<td>P432121</td>
<td>P432121</td>
<td>P432121</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>104.91, 104.91, 121.33</td>
<td>105.13, 105.13, 122.17</td>
<td>103.74, 103.74, 120.93</td>
</tr>
<tr>
<td>( R_{sym} )</td>
<td>0.10 (0.58)</td>
<td>0.07 (0.267)</td>
<td>0.09 (0.20)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.8726</td>
<td>0.9395</td>
<td>0.9395</td>
</tr>
<tr>
<td>( I/\sigma (I) )</td>
<td>8.5 (2.4)</td>
<td>14.1 (4.9)</td>
<td>14.3 (7.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.7 (98.7)</td>
<td>98.5 (99.7)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.9 (5.0)</td>
<td>5.2 (5.2)</td>
<td>5.9 (6.0)</td>
</tr>
</tbody>
</table>

**Refinement**

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>40.4-2.10</th>
<th>63.51-2.47</th>
<th>73.36-2.47</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number reflections</td>
<td>37428</td>
<td>23346</td>
<td>22989</td>
</tr>
<tr>
<td>( R_{work}/R_{free} ) (%)</td>
<td>0.21/0.26</td>
<td>0.21/0.28</td>
<td>0.19/0.26</td>
</tr>
</tbody>
</table>

**Number of atoms**

<table>
<thead>
<tr>
<th>Number of residues</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>604</td>
<td>3966</td>
</tr>
<tr>
<td>528</td>
<td>3901</td>
</tr>
<tr>
<td>563</td>
<td>3930</td>
</tr>
</tbody>
</table>

**RMS deviations**

| Bond lengths (Å) | 0.023 | 0.019 | 0.019 |
| Bond angles (°) | 1.961 | 1.857 | 1.841 |
| Ramachandran     | 94.5/1.0 | 93.5/1.4 | 94.7/1.2 |
| Favored/outliers (%) | 94.5/1.0 | 93.5/1.4 | 94.7/1.2 |

Values for the highest resolution shell are shown between parentheses.

The mutations F82A and A83P, which border the epoxide ring substituent binding region (3), change the shape of the active site in HheC2360 in such a way (Figure 2-1B) that the cavity can now bind (S)-5 in a catalytically productive mode, with the epoxide ring oxygen oriented towards the catalytic triad residues S132 and Y145. This productive binding mode of (S)-5 by the mutant is very different from the catalytically unproductive binding of (S)-p-nitrostyrene oxide [(S)-19] by the wild-type enzyme[9]. In the latter case, the epoxide ring oxygen of
(S)-19 was oriented away from the catalytic triad residues, due to inverted positions of the epoxide oxygen and the terminal carbon of the substrate relative to the (R)-19 binding mode. This explained why the wild-type enzyme exhibits enantiopreference towards R-configured substrates. For mutant HheC2360, the mutations F82A and A83P lead to changes in the active site shape that allow productive binding of an S-configured epoxide.

Neither HheC2360 nor HheC-C153S displays high enantioselectivity in the cyanolysis of aliphatic oxiranes such as (R)- and (S)-19. For wild-type HheC, this low enantioselectivity with aliphatics has been observed before and contrasts with the enantioselective positioning of chiral aromatic substrates such as 19. The low enantioselectivity with 14 is likely the result of a lack of specific interactions between the enzyme and the ethyl substituent on the oxirane ring, which would otherwise fix the orientation of the substrate in the active site.

**Figure 2-1.** Comparison of active site cavities of HheC and HheC2360. A) Active site of HheC with bound (R)-17 (yellow) and chloride (green; PDB ID: 1PWZ). B) Active site of HheC2360 with (S)-5 (orange) and chloride (PDB ID: 4IXW). The most important changes in the active sites are a rotation of W139 and the intrusion of residue 84 into the active site.

The only mutation that affects the epoxide ring oxygen binding region (site 1) is T134A, which is present in both mutants. In wild-type HheC, residue T134 forms a hydrogen bond with the catalytic triad residue S132. The role of S132 is to provide a hydrogen bonding interaction with the oxygen of the substrate hydroxy group, which must be deprotonated by Y145 during the epoxide formation reaction. In more than 90% of the sequences of HheC mutants, this hydrogen bond cannot be formed, due to a T134A substitution.
Because this is the most frequently observed substitution, it is presumably of high importance for the evolved target activity.

Lastly, in mutant HheC2360 the anion binding site (2) is also slightly modified. In the wild-type structure, anions are stabilized by interactions with the backbone amide groups of Y177 and L178, the side chains of P175 and L178, the slightly positive edges of the aromatic rings of F12, F186, and Y187, and through a bound water molecule\[6\]. The loop between residues P175 to Y185 forms part of the spacious halide-binding loop. The L178V mutation increases the available space in the anion-binding site. This allows the position of the chloride to be shifted by 0.8 to 1.1 Å relative to HheC structures with bromide (PDB ID: 1PWX) and chloride (PDB ID: 1PWZ), respectively. As a consequence, the distance between chloride and the primary epoxide ring carbon atom (which is the site of nucleophilic attack) increases to 4 Å in structure 4IXW. In wild-type structures (PDB ID: 1PWZ, 1ZMT) this distance is 3.2 Å. Apparently, this increase does not have a negative effect on the catalytic activity in ring-opening reactions with cyanide. Similar distances, in the 3.5 to 4.5 Å range, were observed in quantum chemical models for cyanolysis and azidolysis reactions catalyzed by wild-type HheC\[10a\]. Interestingly, mutations L178A and L178F have been described before but produced inactive, dimeric mutant enzyme assemblies in a wild-type context\[5b\].

The mutation F186Y stabilizes a rare non-proline cis-peptide bond, which is present at the anion-binding site (2) both in the wild-type and in the mutant structures. When the cis-peptide bond was discovered upon elucidation of the wild-type enzyme structure, the authors proposed that it has a stabilizing role on the anion binding architecture. In the wild type, the cis-peptide bond is only stabilized by a single hydrogen bond with the backbone amide of F12\[6\]. The substitution F186Y further stabilizes this cis-peptide bond by introducing an extra hydrogen bond between the hydroxy group of Y186 and the backbone amide group of F82A. The mutation F186Y was the second most frequently observed substitution in the mutant sequence population (89 %),\[21b\] which highlights the importance of this cis-peptide bond stabilization for the evolved target activity.

From inspection of the structure it is clear that substitutions flanking the active site pocket are likely to contribute to the observed changes in catalytic properties, such as the more than 80- and 120-fold enhanced activities with epoxides rac-14 and rac-16, respectively, and the inversion of enantioselectivity with chloroalcohol 4. Such replacements around active site residues are often reported in directed evolution studies directed towards improving catalytic
activities or expansion of substrate scope,[30] and have in fact been reported in most structural investigations of highly mutated proteins[22, 25-27]. For this reason, efficient directed evolution protocols such as iterative saturation mutagenesis selectively target residues close to the active site.[31] Mutations of a few key residues close to an active site can drastically alter substrate specificities. A carboxylesterase [25] and an N-acetylneuramic acid lyase,[32] for example, required only mutations of a few key residues to produce variants that hydrolyze an artificial dimeric substrate or condense non-natural substrates, respectively.

**Improved thermal stability**

As well as shifts in catalytic activity, the HheC mutants were also reported to show enhanced stabilities. Indeed, mutant HheC2360 displayed an increased thermal stability of at least 8°C. In measurements of residual dehalogenase activity, the mutant exhibits an apparent T_M (temperature at which 50% of the activity is lost after ten minutes of incubation) of 63 °C, in contrast with a value of 55 °C for HheC-C153S (Figure 2-2).

![Figure 2-2](image-url)

**Figure 2-2.** Residual dehalogenase activity after 10 min incubation of enzyme at different temperatures. HheC-C153S (open diamonds) and mutant HheC2360 (closed squares) exhibited initial activities at 30 °C (100 %) of 0.26 U mg⁻¹ and 0.11 U mg⁻¹, respectively.

Thermal unfolding was monitored by determining changes in fluorescence upon heating in the presence of the fluorophore SYPRO Orange.[33] The maximum change in fluorescence is a measure of the apparent unfolding or dissociation temperature (T_m,app). The T_m,app was 55.5 °C for HheC-C153S and 66.5 °C for HheC2360 (Figure 2-3).
Figure 2-3. First derivative of fluorescence change (dF/dT) caused by binding of SYPRO Orange to hydrophobic parts of HheC-C153S (black) and mutant HheC2360 (gray) due to thermal unfolding in a temperature gradient from 30 to 90 °C.

The tetrameric assembly of HheC is stabilized by a large hydrophobic surface area that is buried upon formation of the oligomer. This buried surface was calculated to be 17,840 Å² for the wild type and 18,340 Å² for the mutant (Table 2-4). The surface area buried upon formation of the tetramer is thus increased by 500 Å² in mutant HheC2360, which is predicted to cause an increase in the dissociation energy of 8 kcal mol⁻¹. It is known that such an increase in buried surface area can stabilize a protein. An increase in buried surface area was found to lead, for example, to a 6 °C increase in the melting temperature of cocaine esterase mutants from Rhodococcus sp. strain MB1.[34] An increase in buried surface area was also found to be responsible for the 40 °C difference in melting temperature between a mesostable and thermostable homologue of a Rhodococcus sp. naphthalene 1,2-dioxygenase.[35]

<table>
<thead>
<tr>
<th>Structure</th>
<th>Ligand</th>
<th>Surface area (Å²)</th>
<th>Buried surface area (Å²)</th>
<th>ΔG_int (kcal mol⁻¹)</th>
<th>ΔG_diss (kcal mol⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1PWX</td>
<td>Bromide</td>
<td>30,870</td>
<td>17,840</td>
<td>-126</td>
<td>87</td>
</tr>
<tr>
<td>4IY1</td>
<td>Chloride</td>
<td>32,430</td>
<td>18,340</td>
<td>-185</td>
<td>95</td>
</tr>
</tbody>
</table>

Another contribution to the increased stability comes from intra-subunit van der Waals interactions. A new intra-subunit van der Waals interaction is created by the relocation of residue F136 induced by the nearby P135S mutation. This mutation shifts the position of the phenyalanine side chain by more than 11 Å (Figure 2-4A). The distance between the two F136 residues in the wild type (5.9 to 7.5 Å) was too large for van der Waals interactions, whereas in HheC2360...
the residues are at favorable interaction distance (4.3 Å to 4.8 Å). It is therefore reasoned that this mutation stabilizes the tetrameric ensemble by creating an additional intramolecular van der Waals interaction between neighboring monomers.

Figure 2-4. Enhanced inter-subunit interactions observed in mutant HheC2360 (magenta) relative to wild-type HheC (cyan). A) The relocation of F136, which leads to an increase of 330 Å² in buried surface area. B) The enlarged aromatic network in HheC2360 contains five tryptophan residues from two different subunits; these stabilize the C-terminal W249' of the opposite monomer in the active site. C) Mutation V205Y introduces an additional hydrogen bond between two adjacent subunits.

Figure 2-5. The structure around residue 87 in A) HheC (Q87), and B) HheC2360 (Q87R). The mutation extends an existing hydrogen bond to residues E85, K91 and D96.
The mutations H201W and F86W further enhance the oligomeric stability by extending an aromatic interaction network that is not present as such in the wild-type enzyme (Figure 2-4B). Tryptophan residues from opposing monomers interact in this cluster located at the inter-subunit interface. Earlier research showed that residue W192 is essential for the formation of the tetramer and that the W192F mutation leads to inactive, monomeric or dimeric enzyme. In HheC2360, the tryptophan residues W192, W86, W139, W201, and W249′ (protruding into the active site from the opposite monomer) participate in multiple tryptophan–tryptophan interactions. Whereas in the wild-type structure the distance between H201 and W139 was too large for aromatic interactions (5.3 Å), the larger aromatic side chains introduced by the F86W and H201W mutations allow for closer contacts, especially with residue W249′. Thus, in the HheC2360 enzyme a partially new network of aromatic interactions affects the shape of the active site and stabilizes the positioning of the C-terminal extension into the active site of an opposite subunit, thereby stabilizing a region in the protein that is involved in tetramer formation.

Inter-subunit hydrogen bonds also contribute to the oligomeric stabilization of HheC2360 by, for example, mutation V205Y. This interfacial mutation creates a new inter-subunit hydrogen bond with residue V205Y′ from the opposite monomer (2.8 Å) (Figure 2-4C). Furthermore, the aromatic ring of Y205 participates in van der Waals interactions with the hydrophobic carbon atoms of the K204 side chain (3.6 Å), as well as in an edge-to-face interaction with W201 (3.8 Å). Other possibly stabilizing substitutions include the surface-exposed mutation Q87R, which leads to an additional hydrogen bond with the peptide bond nitrogen of K91 (Figure 2-5), and Y166H, which extends the existing hydrogen bonding network between E165 and K140′ from a neighboring monomer to residue K123 (Figure 2-6).

These results suggest that the stabilization of HheC is to a large extent governed by improved inter-subunit interactions. Consequently, protein stabilization experiments might profitably focus, in cases of oligomeric proteins, on improving hydrophobic and hydrogen-bonding interactions between subunits. Improved interactions of this type have also been found in other thermostable mutants, such as L-isoaspartyl-O-methyltransferase from Sulfolobus tokodaei, and in comparisons of thermostable and mesostable alcohol dehydrogenases.
Figure 2-6. Effect of the mutation at position 166 by comparing A) HheC (Y166) B) HheC2360 (Y166H). The mutation creates an inter-subunit hydrogen bond network between residues Y166, E165, K140 and residue K123 from another molecule.

Other mutations in the HheC2360 mutants suggest further stabilizing effects. Both mutation M245V and the previously found substitution C153S\(^{[4]}\) remove oxidation-prone sulfur-containing residues from the protein. The majority of mutant sequences generated by Fox et al. contain the substitutions C153S (87 %) or M245V (82 %).\(^{[21b]}\) Removal of sulfoxidation-susceptible sites is indeed important for protein stability.\(^{[39]}\) Introduction of charged groups on the protein surface, causing formation of new surface salt bridges or extension of existing salt-bridge networks, can also provide stabilization.\(^{[40]}\) This was observed in, for example, the crystal structure of a thermostable variant of B. subtilis lipase LipA.\(^{[23]}\) Formation of new salt bridges does not appear prominent from the HheC2360 crystal structures, however, even though six surface mutations influence electrostatics (Q37H, K38Q, Q78R, G99D, K121R, K203R).

**Proline mutations**
The modifications of the HheC2360 active site are not caused merely by side chain substitutions, but also by protein backbone changes resulting from the introduction or replacement of proline residues. Proline units stabilize protein structures by decreasing the difference in entropy between the folded and unfolded state. The presence of proline residues also restricts the local backbone $\phi$ dihedrals to a relatively narrow area of the Ramachandran plot. This reduces rotational freedom and might restrict backbone conformations, including conformations that occur as early intermediates in an unfolding pathway. Structures of evolved enzymes containing mutations involving prolines are rare, but have been described in cases of a p-nitrobenzyl esterase...
variant\textsuperscript{[26]} and a glycine oxidase variant.\textsuperscript{[41]} Because changes in backbone conformation could influence active site geometry, we inspected the structural effects of the HheC2360 mutations that involve prolines.

The HheC2360 37-fold mutant contains one substitution through which a proline residue is introduced (A83P) and two mutations through which proline is removed (P84V and P135S). All three mutations produce significant local backbone changes relative to the wild type, whereas no large differences in backbone structure are present in regions that do not contain proline mutations. The A83P and P84V mutations lead to C\(\alpha\) positional changes (relative to the wild-type enzyme) of up to 2.5 Å in the loop between A83 and W86. The P135S mutation is accompanied by a drastic change in the \(\phi\) dihedral of residue 135 from \(-60^\circ\) to \(-30^\circ\), which in turn results in a 3.4 Å shift of the position of C\(\alpha\) of F136 and an average shift of 2.8 Å in the position of the C\(\alpha\) atoms between T134A and W139. The altered backbone conformations in the 134 to 139 region are accompanied by a shift of the F136 C\(z\) atom by 11 Å (Figure 2-4A). This relocation of the side chain of F136 enhances the inter-subunit interactions and leads to an increase in buried surface area, which contributes to enhanced stability as discussed above.

Similar effects are caused by the A83P and P84V mutations, which are located in a loop between residues A83 and W86, surrounding the active site. The \(\psi\) and \(\phi\) dihedrals at the position of the introduced proline residue are changed in relation to the wild-type enzyme, which results in different backbone conformations between residues 83 and 86, but wild-type backbone conformations are re-established from position W86. Apparently, the introduced proline locally modifies the backbone, but nearby strands and helices fix the backbone back to its wild-type conformation outside the affected loop. This is similar to what has been described for an evolved \textit{p}-nitrobenzyl esterase variant.\textsuperscript{[26]} Although not located in the active site, the A83P mutation significantly alters the shape of the active site by inducing backbone changes, which in turn cause a different orientation of the side chain of V84. This reorientation makes atoms C\(\gamma_1\) and C\(\gamma_2\) protrude into the active site (Figure 2-1), which results in a narrowing of the binding region for the R group connected to the epoxide ring. The modified binding pocket is more suitable for binding an aliphatic group than an aromatic ring. These results suggest that experiments directed towards changing enzyme activity can be benefit from the introduction of proline residues in loops that flank residues shaping active site. These resulting backbone changes of first- or second-shell residues can induce side chain reorientations, which ultimately modify substrate binding and catalysis.
Synergistic mutations

The structures of the HheC2360 enzyme reveal several coupled mutations that mainly alleviate steric clashes that would result from some mutations when present individually. For example, the indole group introduced by substitution H201W would sterically clash with the side chain of Y177 (0.6 Å). Mutation Y177G resolves this clash (Figure 2-7A). Another steric clash would occur between the Cγ2 atom of wild-type I246 and the side chain of P138 (0.9 Å), resulting from the 3 Å backbone shift of the loop between residues 134 and 139, which is in turn caused by the P135S substitution discussed above. This clash is resolved by the I246V mutation and associated backbone changes in the C-terminal loop, which increase the distance between the side chains of P138 and the Cγ2 atom of V246 to 3.4 Å. Furthermore, the relocation of the loop harboring F136, caused by the P135S substitution discussed above, would cause a steric clash between the side chains of T146 and T154, but the substitutions T146A and T154A prevent this.

Figure 2-7. Synergistic mutations in the evolved HheC2360 enzyme. A) The synergistic mutations H201W and Y177G prevent a steric clash between residues W201 and Y177. B) The substitution F82A creates a destabilizing hydrophobic cavity in the protein, which is filled by the methionine side chain introduced by the A100M mutation. The structures of HheC2360 are in magenta and those of wild-type HheC in cyan.

Another example of paired mutations is given by the F82A+A100M couple. Mutation F82A removes an aromatic ring from the hydrophobic interior of the protein, creating an unfavorable cavity. The substitution A100M compensates for this detrimental effect by positioning the new large methionine side chain into this cavity (Figure 2-7B).

Whereas many directed evolution studies have targeted residues individually, such as by single-site saturation mutagenesis, these observations and other reports\textsuperscript{[26, 42]} suggest that it is more efficient to introduce mutations in
a combined fashion. Cooperative effects of mutations have been found to be responsible for large changes in activity in a p-nitrobenzyl esterase,[26] in glucose dehydrogenase,[42a] and in lipase [42b-d] mutants. If synergistic mutations are allowed by design of pairwise or three-site libraries, with use of structural information, a larger set of substitutions at each site will be allowed, resulting in a much larger number of possibly functional variants than is accessed by the sum of all functional single-site mutants.

To confirm that certain mutation pairs act in a cooperative manner, the changes in folding energy (ΔΔG\text{Fold}) were computationally predicted for two pairs of mutations, both for the separate mutations and their combination. Both for the F82A+A100M and the Y177G+H201W pairs, the ΔΔG\text{Fold} values calculated for the individual mutations in a wild-type background were very large and unfavorable (Table 2-5). When the ΔΔG\text{Fold} values were calculated for the pairwise combinations, however, the destabilizing effects were much smaller. The calculations with FoldX predicted that the effects of the combined mutations on ΔΔG\text{Fold} are negligible when the error of the calculation method is taken into consideration. The calculations thus confirm that the F82A and A100M mutations act in a cooperative manner, as do the Y177G and H201W substitutions.

Table 2-5. Computationally predicted cooperative effect of mutations.

<table>
<thead>
<tr>
<th>Mutation pairs</th>
<th>Predicted ΔΔG\text{Fold} [kJ mol(^{-1})]</th>
<th>FoldX calculation</th>
<th>Rosetta calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum single mutations</td>
<td>Combined mutations</td>
<td>Sum single mutations</td>
</tr>
<tr>
<td>F82A+A100M</td>
<td>125 ±27</td>
<td>11 ±27</td>
<td>172 ±40</td>
</tr>
<tr>
<td>Y117G+H201W</td>
<td>118 ±27</td>
<td>11 ±27</td>
<td>240±40</td>
</tr>
<tr>
<td>P135S+I246V</td>
<td>72 ±27</td>
<td>72 ±27</td>
<td>53±40</td>
</tr>
<tr>
<td>T146A+T154A</td>
<td>-15 ±27</td>
<td>-15 ±27</td>
<td>4±40</td>
</tr>
</tbody>
</table>

For the P135S+I246V and the T146A+T154A mutations, the calculations predicted no compensating effects. This might be the result of the large backbone changes that accompany the differences in side chain structures and conformations (see above), which make predictions of the effect on ΔΔG\text{Fold} inaccurate with computational methodology that was used (Table 2-5). These observations underscore the necessity of improving backbone remodeling protocols[43] if the aim is computational design of stabilizing mutations based on shifts in backbone conformations.
Conclusions

Directed evolution experiments can yield highly evolved enzyme variants with dramatic improvements in stability and catalytic efficiency, without requiring much insight into the molecular basis of the target properties. The investigation of the highly engineered halohydrin dehalogenase variant HheC2360 described here provides insight into the structural effects of the mutations that are introduced during directed evolution. We show that the increases, of up to 100-fold, in rates of cyanide-mediated epoxide ring opening, as well as the modified enantiopreference occurring in HheC2360, can be explained by inspection of crystal structures. Specifically, side chain substitutions that allow formation (F186Y) or removal (T134A) of hydrogen bonds affect essential features of the HheC active site, such as the cis-peptide bond in the anion binding loop and the hydrogen bonding pattern of the catalytic triad residue S132, respectively. Furthermore, the mutations reshape the active site cavity for improved binding of S-configured target substrates, not only through altered side chains (F86W, L178V), but also through backbone modifications that result from the introduction and removal of proline residues (A83P, P84V). The enhanced thermal stability of the evolved dehalogenase is explained by several mutations that improve oligomeric interface contacts (V205Y, H201W), and also by a substantial increase in buried surface area that further stabilizes the tetrameric enzyme assembly. The latter is again due to changes in the side chain location and orientation (F136) resulting from the removal of a proline (P135S). Several of the mutations that contribute to the change in functional properties appear to be facilitated by or are dependent on additional coupled mutations.

From the structure–function analysis of HheC2360 it can be concluded that backbone changes to loop regions caused by mutations that introduce or remove proline residues give access to additional functional changes in active site geometry, relative to mutations that only influence side chains. Targeting of proline residues can thus be highly useful when engineering enzymes to obtain modified catalytic properties. Through cooperative and synergistic effects, combinations of mutations allow greater functional diversity than individual mutations. Simultaneous introduction of mutations at two, three, or four structurally close sites in libraries can increase the functional diversity explored in directed evolution without strongly diluting libraries with inactive variants.

The enantiopreference of Hhe-C153S with 4 is in favor of the \( R \) enantiomer, with an apparent enantiomeric ratio of \( E_R > 4.8 \). The mutant HheC2656 hardly discriminated between the enantiomers of 4 \( (E_R = 1.3) \). In contrast, mutant HheC2360 exhibited an enantiomeric ratio of \( E_S = 1.9 \),
corresponding to a modest preference for (S)-4 over (R)-4. Even though the observed values are low, the enantiopreference of mutant HheC2360 is the opposite of those of HheC-C153S or mutant HheC2656. Enantiopreference towards S-configured haloalcohols and epoxides has not been observed in previous biochemical studies on HheC and its variants,[8-9] but mutants with a preference for S-configured substrates were found in mutants of Arthrobacter sp. strain AD2[29].

Materials and methods

Chemicals and commercially available enzymes: The hydroxynitrile ethyl (R)-4-cyano-3-hydroxybutyrate [(R)-6], the haloalcohols ethyl (R)-4-chloro-3-hydroxybutyrate [(R)-4], ethyl (S)-4-chloro-3-hydroxybutyrate [(S)-4], 1,3-dichloropropan-2-ol (7), 1,3-dibromopropan-2-ol (8), rac-1-chloropropan-2-ol (rac-9), 1-chloro-2-methylpropan-2-ol (10), rac-3-chloropropane-1,2-diol (rac-11), rac-3-bromopropane-1,2-diol (rac-12), and rac-2-chloro-1-phenylethanol (rac-13), as well as the epoxides ethyl (S)-3,4-epoxybutyrate [(S)-5], rac-1,2-epoxybutane (rac-14), (R)-1,2-epoxybutane [(R)-14], (S)-1,2-epoxybutane [(S)-14], rac-1,2-epoxy-2-methylbutane (rac-15), and rac-1,2-epoxyhexane (rac-17), were purchased in the highest available purity from Sigma–Aldrich (St. Louis, MO, USA). Restriction enzymes NdeI and XhoI and T4 ligase were ordered from New England Biolabs (Ipswich, MA, USA). SYPRO Orange was obtained from Life Technologies (Carlsbad, CA, USA).

Cloning, expression, and enzyme preparation: The directed evolution study by Fox and co-workers yielded 1422 HheC variants,[21b] a multiple sequence alignment of which was prepared by using ClustalX v2.1.[44] After removal of identical protein sequences from the alignment, the observed mutations were inspected and mapped to the wild-type crystal structure (PDB ID: 1PWZ). Sequences HheC2360 and HheC2656 were obtained as synthetic DNA constructs from Geneart AG (Regensburg, Germany). Synthetic genes for HheC2360 and HheC2656 were cloned into expression vector pBADNk[45] by standard protocols. After transformation of constructed plasmids into E. coli MC1061, expression and enzyme purification were carried out as described before.[46] As a control, HheC-C153S was expressed in recombinant fashion from plasmid pBAD (C. Tarabiono, unpublished results) and purified in a similar way.[46] Enzyme was stored at −20 °C in TEMG buffer [Tris·SO₄ (10 mM), ethylenediaminetetraacetic acid (EDTA, 1 mM), β-mercaptoethanol (1 mM), glycerol (10%), pH 7.5]. Protein concentrations were determined spectrophotometrically after staining with Coomassie Brilliant Blue.

Dehalogenase assays: Dehalogenase assays were performed in duplicate by monitoring halide release by the colorimetric assay described by Bergmann and Sanlik.[47] Briefly, initial halide release was monitored at 460 nm in 96-well plates on mixing of assay reagent (150 μL) with sample (150 μL). The assay reagent was prepared freshly before measurements, by mixing equal volumes of solution I [NH₄Fe(SO₄)₂ (0.25 M) in HNO₃ (9 M)] and solution II [saturated solution of Hg(SCN)₂ in absolute ethanol]. Reactions were carried out in Tris·SO₄ (50 mM, pH 7.5 in a total volume of 2 mL) at 30°C. Reactions were initiated by addition of enzyme, and halide concentrations were monitored over time. Kinetic parameters were obtained from initial rates by use of the SimFit v6.2.3 package (http://www.simfit.man.ac.uk/) for curve fitting by Michaelis-Menten kinetics. The apparent enantiomeric ratio E was calculated from the ratio of catalytic efficiencies (kcat/KM) for (R)- and (S)-4 as described by Straathof and Jongejan.[48]
Cyanolysis assays: Cyanide consumption was monitored by utilizing the specific absorption (at 267 nm) of the [(CN)₄Ni]²⁻ complex ion, formed after mixing of a cyanide-containing sample and Ni²⁺ in aqueous NH₃ solution. In detail, samples (150 μL) were mixed with NiCl₂ (2 mM, 150 μL) in an aqueous solution of NH₃ (1 M), and the absorbance at 267 nm was read in 96-well UV-Star microplates (Greiner Bio One, Frickenhausen, Germany). Standard assays for cyanide-mediated epoxide ring opening contained epoxide (20 mM) in potassium phosphate buffer (50 mM, pH 8.0) and varying concentrations of sodium cyanide. The reactions were performed in duplicate in a total volume of 2 mL at 30 °C after addition of enzyme. Apparent kinetic constants and apparent catalytic efficiencies were obtained as described for the dehalogenase reactions.

Stability assays: The thermal stabilities of variant HheC2360 and wild-type HheC were determined by measuring residual dehalogenase activity after 10 min incubation of enzyme at temperatures between 30 and 95 °C with subsequent storage on ice (less than 10 min, to minimize potential enzyme refolding). After heat treatment, halide release was measured at 30 °C in reaction mixtures with 8 (10 mM) and enzyme (5 μg mL⁻¹ final concentration) as described for the dehalogenase assays. Initial activities were determined, and relative dehalogenase activities were plotted as a function of the corresponding temperature. The reported residual activities thus represent T₅₀ values (that is, the temperatures at which 50% of the activity is lost upon 10 min of incubation).

In addition, thermal unfolding of enzyme was monitored by a fluorescence-based thermal unfolding assay (thermofluor). This method is based on fluorescence increase upon binding of SYPRO Orange to hydrophobic protein surfaces that become exposed upon thermal protein unfolding or multimer dissociation. The fluorescence increase was monitored by excitation at 490 nm and recording of the emission at 575 nm with a MyiQ real-time PCR machine (Bio-Rad, Hercules, CA, USA) while increasing block temperature from 30 to 90 °C by 1 °C min⁻¹. In iQ 96-well real-time PCR plates (Bio-Rad), protein solutions containing 300-fold diluted SYPRO Orange (7.5 μL) in MilliQ water and purified protein (17.5 μL, 0.4 mg mL⁻¹) in TEMG buffer were prepared. After sealing (iQ 96-well PCR plate seals, Bio-Rad), the temperature gradient was started. The first derivative of measured fluorescence versus temperature was calculated. It gives the apparent melting temperature at the local maximum.

Crystallization and structure determination of HheC2360: The HheC2360 mutant did not crystallize under wild-type conditions. Instead, HheC2360 crystals were obtained by sitting drop vapor diffusion at room temperature. Purified protein (1 μL, 14.7 mg mL⁻¹ in HEPES buffer [4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid, 50 mM, pH 8.0] supplemented with NaCl (300 mM)) was mixed with reservoir solution [1 μL, MgCl₂ (200 mM), Tris-Cl (pH 8.5, 100 mM), and PEG (polyethylene glycol) 8000 (12.5%, w/v)]. After two weeks very small crystals had appeared, and these could be used as seeds for the growth of larger 150×150×150 μm crystals. Crystals were flash-frozen after having been transferred to a cryo solution (similar to the mother liquor, except that the PEG 8000 concentration was increased to 45%). To obtain substrate- and product-bound structures, either (S)-5 or (R)-6 (7 mM) was added to the cryo solution, and crystals were soaked for 30 min prior to flash freezing. Diffraction data were collected at beam lines 1D-4 and 1D-4 at the ESRF in Grenoble, France. The diffraction data were processed by use of iMOSFLM and scaled by use of SCALA. No phasing was necessary, because a rigid body refinement using the wild-type structure of HheC (PDB ID: 1PWZ) showed clear electron density. The model was further refined by use of Refmac5 and manually improved by use of Coot. Finally, the model was validated by use of MolProbity. Structures with bound chloride,
chloride and (S)-5, and (R)-6 as ligands have been deposited at the PDB with entry codes 4IY1, 4IXW, and 4IXT, respectively (Table 2-3).

**Computational methods:** Buried surface areas, free energy gains upon formation of the tetrameric assembly ($\Delta G_{\text{int}}$), and free energies of tetrameric assembly dissociation ($\Delta G_{\text{diss}}$) were calculated by use of the PISA webserver at PDB with structures 4IXW and 1ZMT.\[38] A solvent-accessible surface area of <5 Å² was used as a cut-off to discriminate buried from solvent-accessible residues in tetrameric assemblies. The predicted changes in folding energies ($\Delta\Delta G_{\text{Fold}}$) were calculated both with FoldX\[56] and with Rosetta.\[57] With Rosetta an established protocol that employs a soft-repulsion packing of all side chains within 8 Å of the mutated residue was used.\[58] FoldX has a standard error of 3.4 kJ mol⁻¹ (excluding 5% outliers). For the Rosetta method, no standard error has been reported, but its correlation coefficient versus experimental data ($r=0.68$) was reported to be better than for FoldX ($r=0.5$), so assumption of an error of ±5 kJ mol⁻¹ appears prudent. Because these errors are for single mutations in monomeric proteins, whereas the current protein is tetrameric and double mutations are involved, we used values eight times higher (27.2 kJ mol⁻¹ for FoldX, 40 kJ mol⁻¹ for Rosetta) to analyze whether predicted changes in folding energy were significant.

**Acknowledgements**

HJW was funded by NWO ECHO Grant 08.B3.051, and RJF was supported by the Metaexplore (22625) project, funded by the European Union (7th Framework KBBE-2007-3-3-05).

Author contributions: MS constructed and characterized the enzyme variants and PAJ also characterized the enzymes. RJF elucidated the structures of the enzyme variant. HJW calculated the effects of the mutations on the unfolding energies. MS, RJF, BH and DBJ designed the laboratory experiments. MS, RJF and DBJ wrote the manuscript and HJW, DBJ, BWD, BH and MB corrected it.
References


Chapter 2


[38] Y. Korkhin, A. Joseph Kalb, M. Peretz, O. Bogin, Y. Burstein, F. Frolow, Protein Sci. 1999, 8, 1241-1249.


Biocatalytic and Structural Properties of a Highly Engineered Halohydrin Dehalogenase


