VIP *Very Important Paper*

Stabilizing AqdC, a *Pseudomonas* Quinolone Signal-Cleaving Dioxygenase from Mycobacteria, by FRESCO-Based Protein Engineering

Sandra C. Wullich,[a] Hein J. Wijma,*[b] Dick B. Janssen,[b] and Susanne Fetzner*[a]

The mycobacterial PQS dioxygenase AqdC, a cofactor-less protein with an α/β-hydrolase fold, inactivates the virulence-associated quorum-sensing signal molecule 2-heptyl-3-hydroxy-4(1*H*)-quinolone (PQS) produced by the opportunistic pathogen *Pseudomonas aeruginosa* and is therefore a potential anti-virulence tool. We have used computational library design to predict stabilizing amino acid replacements in AqdC. While 57 out of 91 tested single substitutions throughout the protein led to stabilization, as judged by increases in T_m^{app} of $>2\,^{\circ}C$, they all impaired catalytic activity. Combining substitutions, the proteins AqdC-G40K-A134L-G220D-Y238W and AqdC-G40K-G220D-Y238W showed extended half-lives and the best trade-off between stability and activity, with increases in T_m^{app} of 11.8 and 6.1 °C and relative activities of 22 and 72%, respectively, compared to AqdC. Molecular dynamics simulations and principal component analysis suggested that stabilized proteins are less flexible than AqdC, and the loss of catalytic activity likely correlates with an inability to effectively open the entrance to the active site.

Introduction

An arsenal of virulence factors, combined with a variety of antibiotic resistance mechanisms and high adaptability to diverse environments, makes *Pseudomonas aeruginosa* a very potent bacterial pathogen and a major cause of nosocomial infections.[1] Like many other opportunistic pathogens, *P. aeruginosa* uses cell-to-cell communication ("quorum sensing") systems, relying on the production and detection of signal molecules, to regulate and coordinate virulence-associated processes such as biofilm formation or the production of virulence factors. A complex QS network that integrates the *N*-acylhomoserine lactone dependent *las* and *rhl* systems and the 2-alkyl-4-quinolone-based *pqs* system tunes the expression of hundreds of target genes involved in motility, biofilm formation, cytotoxicity, and virulence factor production.[2–5] The *pqs* system uses 2-heptyl-3-hydroxy-4(1*H*)-quinolone (the *"Pseudomonas* quinolone signal", PQS) as a major signal molecule. PQS plays an important role in regulating the synthesis of a series of virulence factors and antimicrobials, furthermore, it has iron-chelating, pro-oxidant and immunomodulatory properties.[3,6]

Tools to interfere with QS include small-molecule inhibitors of signal biosynthesis or signal detection, as well as enzymes mediating the inactivation of signal molecules by catalysing chemical modification or cleavage reactions.[7–10] These so-called quorum quenching enzymes include a small group of cofactor-less dioxygenases which catalyse the 2,4-dioxygenolytic cleavage of PQS to form carbon monoxide and *N*-octanoylanthranilate, thereby attenuating virulence factor production of *P. aeruginosa*.[11–13] PQS dioxygenases belong to the HQD subfamily within the α/β-hydrolase (ABH) fold superfamily, which comprises dioxygenases acting on the 3-hydroxy-4(1*H*)-quinolone ring[14] (Figure 1A). These enzymes are of potential therapeutic use by suppressing *P. aeruginosa* virulence. To explore such applications, long-term stability of the enzymes at body temperature is an important prerequisite. The PQS dioxygenase AqdC of *Mycobacteroides abscessus* unfortunately is rapidly inactivated at 37 °C. However, the crystal structure of this enzyme has been solved,[15] enabling structure-based enzyme engineering approaches. The crystal structure of AqdC revealed a core domain consisting of the typical ABH fold comprising eight β-strands, of which the second one is antiparallel, surrounded by α-helices. A four helices insertion into the core domain results in a cap domain covering the upper part of the core (Figure 1B). However, the canonical catalytic triad of classical ABHs, consisting of a strictly conserved histidine, an acidic residue, and a nucleophile, is not complete in AqdC as it lacks a nucleophile at the respective position, and catalysis depends on the His/Asp dyad.[15]

Both the core and the cap domain contribute to the formation of a bipartite tunnel that traverses the enzyme, providing two entrances from the surface of the protein to the substrate cavity. One funnel-shaped opening, leading to the catalytic histidine, is most probably the main entrance to the

[a] *S. C. Wullich, Prof. Dr. S. Fetzner*
Institut für Molekulare Mikrobiologie und Biotechnologie
WWU Münster
Corrensstraße 3, 48149 Münster (Germany)
E-mail: fetzner@uni-muenster.de
[b] *Dr. H. J. Wijma, Prof. Dr. D. B. Janssen*
Department of Biochemistry
Groningen Biomolecular Sciences and Biotechnology Institute
University of Groningen
Nijenborgh 4, 9747 AG Groningen (The Netherlands)
E-mail: h.j.wijma@rug.nl

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202000641

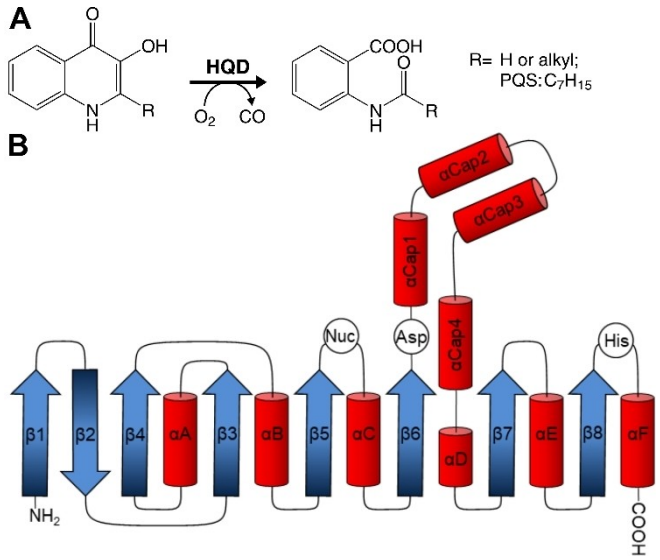

Figure 1. Reaction and structure of 3-hydroxy-4(1*H*)-quinolone dioxygenases (HQDs). A) Cleavage of (2-alkyl)-3-hydroxy-4(1*H*)-quinolones as catalysed by HQDs. B) Schematic representation of the secondary structure of a typical HQD. α-Helices and β-strands are depicted as red barrels and blue arrows, respectively. The positions of the HQD catalytic dyad (a conserved histidine-aspartate charge relay) and of the canonical nucleophile of the classical α/β-hydrolases are indicated by circles.[14]

active site that provides access for the substrate to the cavity. The second, "distal" part of the tunnel, shown to harbour the alkyl side chain of substrate and product and having a truly tunnel like character, leads to another opening that might serve as an exit for product release.[15] Surrounding the substrate cavity, a series of hydrophobic residues tightly enclose the substrate PQS, which additionally is stabilized by hydrogen bond interactions. AqdC thus is highly specific for PQS, with a Michaelis constant (K_M) in the low-micromolar range.[14,15] Because the concentration of PQS in, for example, the sputum of cystic fibrosis patients is around 2 μM or less,[16] besides adequate thermal stability a high affinity for PQS is indispensable for the use of PQS dioxygenases as quorum quenching enzyme.

To improve the robustness of AqdC for potential application as a quorum quenching enzyme, we aimed at enhancing its stability against thermal denaturation. We used the FRESCO workflow, which combines well-established computational prediction methods and molecular dynamics-based screening to avoid excessive laboratory testing[17] to predict and engineer more stable AqdC variants, and investigated their stability and catalytic activity.

Results

Computational library design

FoldX, Rosetta, and a sequence consensus protocol were used to pick stabilizing single substitutions at 246 positions within AqdC. Of the 266 residues of AqdC, 20 were excluded as active site residues since they were within a short distance of a substrate modelled into the active site (5.5 Å was used as a criterion based on visual inspection). Replacements by cysteines were excluded to prevent the unexpected formation of disulfide bonds. Therefore, maximally 4428 (i.e., 246×18) amino acid substitutions were possible. With FoldX and Rosetta, standard protocols were used to predict stabilizing substitutions.[17–19] The consensus method predicted the change in folding energy based on the frequency of the existing amino acid in AqdC versus alternative residues at that position in homologues (Eq. (1), Experimental Section). With these protocols, a total of 650 substitutions were selected as potentially stabilizing. Of these variants, 266 were selected by FoldX, 492 by Rosetta, and 27 by the phylogenetic method.

MD simulations were carried out for all of these proteins, followed by visual inspection to eliminate variants with problematic features. The visual inspection was done according to the recently published protocol.[17] As a result, 537 substitutions were eliminated because of the introduction of a hydrophobic group on the protein surface (156), appearance of unsatisfied H-bond donors or acceptors (89), increase in protein flexibility (159), or because of other reasons such as blocking of the entrance to the active site (17). Almost a quarter of the eliminated variants (114) were dismissed because they occurred at positions where basically every substitution was predicted to be stabilizing, suggesting a systematic calculation error at certain positions. Of the remaining 113 single amino acid substitutions (Table S2 in the Supporting Information), spread over the whole structure of AqdC, 51 variants were initially selected by FoldX, 79 by Rosetta, and 15 by the phylogenetic method (some substitutions were selected by more than one prediction method).

Screening of AqdC proteins with single amino acid substitutions

From the 113 AqdC variants with single replacements that resulted from computational library design, 91 were successfully constructed and expressed as soluble proteins (Table S3). All proteins were prepared following the high-throughput (HT) protocol and subsequently subjected to thermal shift assays to determine their apparent melting temperature (T_m^{app}). Stabilizing substitutions were found throughout the protein, without a preference for highly flexible positions, and with a very satisfactory success percentage of 63%. The highest ΔT_m^{app} values were +8–8.5 °C for four different mutant proteins (Table S3 and 1). Variants having a lower, equal, or $\leq +2$ °C higher T_m^{app} compared to wild-type (WT) AqdC (T_m^{app}: 44.1 °C) were excluded in order to reduce the number of candidates. For the remaining 57 variants, which after preparation by the HT protocol showed an increase in T_m^{app} of $> +2$ °C, the catalytic activity was determined (Table S3). From the variants with the highest stability and catalytic activity, ten substitutions were selected for combination (Table 1, Figure 2). The AqdC-E20G protein has a lower T_m^{app} than the WT protein but was the most

Table 1. Selected AqdC variants for combination of substitutions.

	Substitution	T_m^{app} [°C][a]	Specific activity[b] [U/mg] at 30 °C	$\Delta\Delta G_{fold}$ [kJ/mol] [c] as predicted by Rosetta	FoldX	Sequence consensus	B-fitter rank
WT	–	44.0 ± 0.5	13.9 ± 1.8	–	–	–	–
–	E20G[d]	43.0 ± 0.4	22.3 ± 0.8	−3.2	−1.0	**−6.4**	29
1	G40K	52.0	3.4 ± 0.1	−2.3	**−4.9**	3.0	201
2	A134L	52.0 ± 0.0	5.8 ± 1.0	**−8.1**	**−5.5**	−1.9	67
3	G137V	51.0	3.9 ± 0.5	**−6.7**	−2.5	11.1	41
4	E146A	50.0 ± 1.1	3.5 ± 0.9	**−6.4**	1.3	−1.0	4
5	G147A	47.5 ± 0.4	10.4 ± 0.6	**−6.3**	−2.0	−1.4	17
6	E189A	48.5	5.6 ± 1.0	**−6.3**	−0.8	−4.1	24
7	E202Q	55.8 ± 4.9	2.1 ± 0.8	**−4.8**	1.4	13.9	63
8	H216R	51.5 ± 0.4	1.6 ± 0.4	**−6.3**	**−13.0**	−2.0	35
9	G220D	48.0 ± 1.8	9.2 ± 1.4	8.8	9.1	**−8.0**	83
10	Y238W	48.0	10.3 ± 0.9	**−13.1**	2.2	6.7	177

[a] Depending on the clarity of the melting curve, a single experiment was performed, or 2 to 4 biological replicates were analysed. [b] Specific activities were determined for enzymes enriched using the HT protocol ($n = 2$ technical replicates). All errors represent standard deviations. [c] Bold $\Delta\Delta G_{fold}$ values indicate that the particular substitution was selected because it met the initial selection criteria of the particular (set to −3.5 kJ/mol for Rosetta and FoldX, and < −5 kJ/mol for the sequence consensus predictions). [d] Used as a control variant in MD simulation.

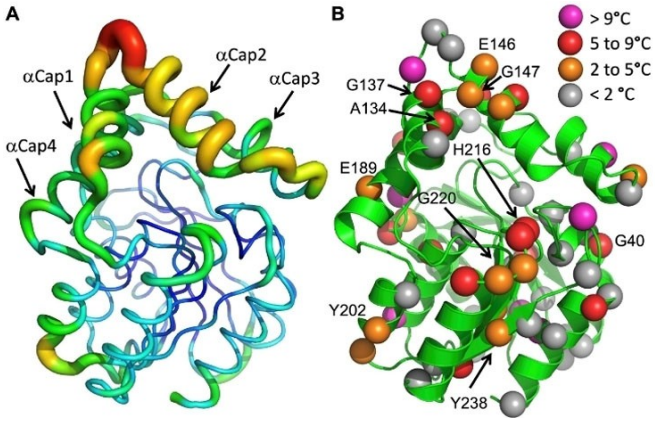

Figure 2. Location of substituted residues in AqdC and their effect on protein stability. A) Flexibility, as indicated by *B*-factors. Blue corresponds to the lowest *B*-factors and red to the highest. α-Helices of the cap domain are indicated. B) Location of all substituted residues and the effect of the best substitutions at those positions. The labelled positions are those of which the stabilizing substitutions were subsequently recombined.

active variant among all proteins tested (Table S2) and therefore served as a control in further MD simulations.

The ten substitutions of the selected proteins originated from different computational sources. Rosetta selected eight of these replacements, FoldX three, and the phylogenetic protocol two (Table 1). Two substitutions were selected both by FoldX and Rosetta. It is also noteworthy that from these top substitutions in terms of T_m^{app} increase and catalytic activity, only three had a B-fitter rank within the top-30 (Table 1). Thus, eight out of the ten top substitutions would have been missed by using the B-fitter approach to select the most promising positions for stabilization. Furthermore, eight out of the ten substitutions were predicted to be destabilizing by only one method (Table 1). These would have been missed by a computational selection of only elite-ranked mutations, predicted to be stabilizing by all methods.

Combination of single substitutions

Amino acid substitutions of the ten variants, which ranked top according to T_m^{app} and activity, were scheduled for combination in the order listed in Table 2, as suggested by MD calculations. The combination variants (Table 2, entries I–X), as well as the WT protein, the most active single-substitution variant (AqdC-E20G), and the variant we started the combination with (AqdC-A134L), were again purified in a larger scale, using gravity flow columns. Kinetic data and melting temperatures of each pure protein were determined (Table 2).

The first substitution, A134L (positioned in αCap1), caused a significant increase of T_m^{app} by 7 °C compared to the WT protein, but decreased the specific activity by 40 U/mg (−67%), due to both a decrease of k_{cat} (−60%) and an increase of K_M (+45%; Table 2). While the T_m^{app} was not affected, the specific activity was again decreased by additionally introducing G137V (αCap1), mainly due to a 75% decrease in k_{cat} (variant I). Similarly, re-introducing G137V into proteins V and VIII to generate VI and IX, respectively, reduced the catalytic activity, without significant stabilization. The substitution G220D (loop, at the beginning of α5) had an increasing effect on both k_{cat} and K_M without affecting T_m^{app} (variant AqdC-A134L to II, and III to IV). The introduction of Y238 W to variant II decreased the k_{cat}, but also, more importantly, decreased the K_M without changing the T_m^{app} (variant IV). Adding the substitution G40K (α1) to variant IV had a somewhat negative effect on K_M and k_{cat} but stabilized the protein by 5.2 °C (variant V). Introducing E189A into variant IV to generate protein VII increased K_M by 11.5 μM, resulting in the highest K_M (24.9 μM) of all proteins generated in this study. Revoking A134L, the initial substitution, in variant V, VI, and VII to generate proteins VIII, IX, and X, respectively, led to a loss in T_m^{app} by 4.7–5.7 °C, while the specific activities increased dramatically by 21.6–30.4 U/mg (≈46–232%). The effects caused by this substitution exemplarily illustrate a common trend. Indeed, while a series of substitutions tested led to stabilization of the AqdC protein as judged by the increases in T_m^{app}, they all impaired catalytic activity.

Table 2. Melting temperatures and kinetic data of combination variants, WT-AqdC, combination starting variant (AqdC-A134L), and most active variant (AqdC-E20G).

	Substitution(s)	T_m^{app} [°C]	Specific activity [U/mg] at 30 °C	K_M [μM]	k_{cat} [s^{-1}]	k_{cat}/K_M
WT	–	44.1 ± 0.1	60.2 ± 2.2	5.8 ± 0.4	41.9 ± 1.1	7.3 ± 0.6
–	E20G	42.3 ± 0.4	56.2 ± 2.8	3.3 ± 0.7	35.6 ± 2.6	11.0 ± 1.6
–	A134L	51.1 ± 0.9	20.2 ± 2.6	8.4 ± 0.2	16.5 ± 1.9	2.0 ± 0.3
I	A134L-G137V	51.0 ± 2.5	5.2 ± 2.5	11.2 ± 1.2	4.2 ± 1.4	0.4 ± 0.2
II	A134L-G220D	51.0 ± 0.4	23.3 ± 2.1	18.8 ± 1.2	27.4 ± 4.2	1.5 ± 0.1
III	A134L-Y238W	50.3 ± 1.0	20.0 ± 1.5	11.5 ± 0.5	17.6 ± 1.4	1.5 ± 0.2
IV	A134L-Y238W-G220D	50.7 ± 0.4	22.4 ± 2.4	13.4 ± 1.7	19.7 ± 0.9	1.5 ± 0.3
V	G40K-A134L-G220D-Y238W	55.9 ± 0.4	13.1 ± 0.2	22.1 ± 4.1	15.9 ± 1.0	0.7 ± 0.1
VI	G40K-A134L-G137V-G220D-Y238W	53.3 ± 8.1 [a]	4.7 ± 0.4	21.1 ± 14.7	5.7 ± 0.9	0.3 ± 0.2
VII	A134L-E189A-G220D Y238W	50.1 ± 0.8	19.5 ± 0.3	24.9 ± 0.6	24.4 ± 0.6	1.0 ± 0.1
VIII	G40K-G220D-Y238W	50.2 ± 0.5	43.5 ± 4.7	3.7 ± 0.3	27.0 ± 2.8	7.4 ± 0.1
IX	G40K-G137V-G220D-Y238W	51.5 ± 2.3	26.3 ± 0.5	4.6 ± 1.3	17.6 ± 2.2	4.0 ± 0.7
X	E189A-G220D-Y238W	45.4 ± 0.1	43.6 ± 5.5	3.1 ± 0.9	26.2 ± 2.9	9.0 ± 3.6

All data were determined from high-purity enzymes using $n = 3$ biological and ≥ 3 technical replicates each. All errors represent standard deviations. [a] Ambiguous results caused by an unusual melting curve.

Among the ten combination variants, proteins V (AqdC-G40K-A134L-G220D-Y238W) and VIII (AqdC-G40K-G220D-Y238W) showed the best trade-off between stability and activity, with increases in T_m^{app} by 11.8 and 6.1 °C, and relative activities (compared to WT-AqdC) of 22 and 72%, respectively. Their catalytic half-life was determined at 37 °C to assess their potential applicability at the human body temperature. Remarkably, protein V showed an about 30-fold prolonged half-life compared to the WT protein, which in the long term may outweigh its 3.7-fold decrease in activity at 37 °C (Table 3). The half-life of protein VIII was about five times higher than that of WT-AqdC. However, the highly active AqdC-E20G protein was even less stable than the WT protein (Table 3).

Computational analysis of flexibility and conformational preferences

MD simulations were carried out to investigate the differences in flexibility of some AqdC variants. While selecting a protocol for these MD simulations, it was considered that multiple relatively short MD simulations, on a single protein, provide a much better conformational sampling than a single much longer MD simulation.[20–23] Therefore, for each of the ten selected (single and multi-site) variants, as well as for AqdC-E20G, for comparison, 20 independent MD simulations (of 12 ns) were carried out (giving a total simulation time of 2.6 μs = 11 × 20 × 12 ns). To avoid as much as possible simulation artefacts caused by starting with a possibly wrongly modelled variant structure, mainly proteins carrying single substitutions were selected for MD simulation (no X-ray structures were available for the variants). Additionally, three multi-site substitution variants were selected, including AqdC-A134L-G137V showing poor catalytic activity.

Table 3. Specific activity and half-life ($t_{1/2}$) of selected AqdC proteins at 37 °C.

#	Substitution(s)	Specific activity [U/mg] at 37 °C[a]	$t_{1/2}$ [h] at 37 °C[b]
WT	-	60.6 ± 1.0	0.32 ± 0.09
–	E20G[c]	49.2 ± 2.8	0.11 ± 0.01
V	G40K-A134L-G220D-Y238W	16.5 ± 2.7	9.71 ± 0.02
VIII	G40K-G220D-Y238W	46.9 ± 2.3	1.63 ± 0.28

All errors represent standard deviations. [a] $n = 3$ biological replicates (3 technical each), [b] $n = 3$ biological replicates (1 technical each). [c] Unstable but highly active variant for comparison.

The MD simulations predicted that most mutant proteins were slightly less flexible than WT-AqdC (see Figure 3 for examples; Figure S1 shows all the tested proteins). The most flexible areas were located in the cap domain (residues 124–195) and at a loop in the core domain (residue 61–67; Figure S2).

The flexibility at the loop was not strongly affected in any of the tested mutant proteins. The most flexible parts of the cap domain were around residues 129 and 172. The former is located at the start of the first α-helix of the cap domain (αCap1) and close to the funnel-like entrance to the active site. Residue 172, located in a loop between αCap3 and αCap4, is part of the distal, alkyl-tail tunnel. Only a few tested variants were more flexible in the cap domain than the WT protein, namely AqdC-E20G (the only less thermostable variant analysed), AqdC-G220D (higher flexibility around position 172 but lower flexibility around position 129), and the four-fold mutant protein V (AqdC-G40K-A134L-G220D-Y238W; Figure S1). The latter was more rigid at positions 129 and 172, but much more flexible around position 160 (which is located at the loop between αCap2 and αCap3 and does not directly border the tunnel). Thus, most stabilized variants display during the MD simulations a noticeable decrease in flexibility near the assumed entrance part of the tunnel.

For the variant with the lowest activity, AqdC-A134L-G137V, the RMSF dropped from 1.5 to 1.0 Å around residue 129 and from 2.1 to 1.8 Å at residue 172. Somewhat surprisingly, the AqdC-A134L protein displayed a very similar drop in flexibility while it is almost 4-fold more active than AqdC-A134L-G137V. Because the effects on flexibility were not very dramatic and

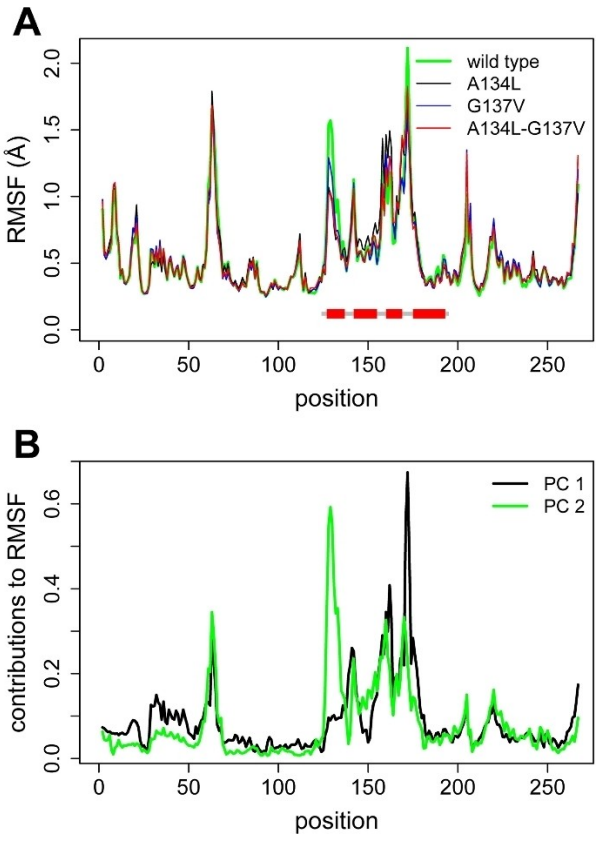

Figure 3. Comparison of mutant protein conformational flexibility and its main components. A) Protein flexibility, as observed during MD simulation plotted against residue number. The structure of the cap domain is indicated with red for α-helices and grey for loops. B) Contributions of principal components (PC) 1 and 2 to protein flexibility.

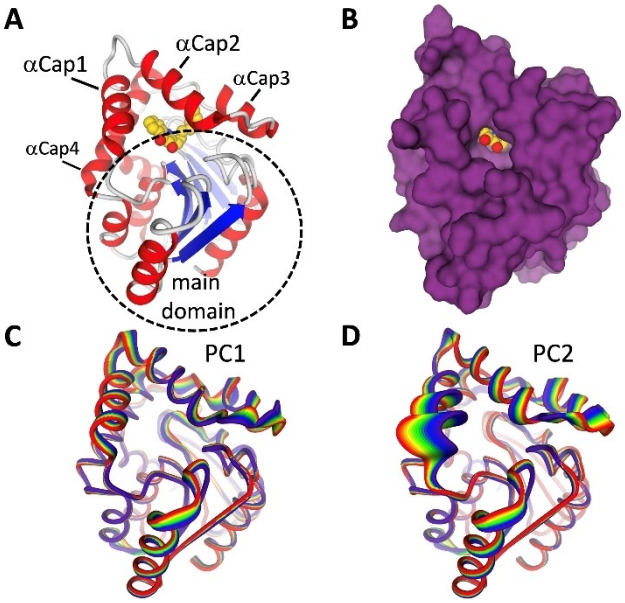

Figure 4. Funnel-like entrance to the active site of AqdC and the first principal component of conformations observed during MD simulation. A) Structural elements of AqdC with product bound (PDB ID: 6RB3[15]) visible while looking into the active site through the entrance; α-Helices are shown in red, β-strands in blue, and grey represents turns and coils. All helices in the cap domain are labelled as "αCap" with their number. The strands and helices of the main domain are not labelled. B) Molecular surface representation with the protein in magenta and the bound product (van der Waals spheres) with yellow carbon atoms, red oxygen atoms, and blue nitrogen. The view shows the tunnel to the AqdC active site with the enzyme in the same orientation as in the other three panels. C) The conformations belonging to principal component 1. The colour scale corresponds to a range for principal component 1 from −20 (red) to +20 (purple). D) Conformations corresponding to principal component 2. Here the colour scale ranges from −22 (red) to +22 (purple).

there was no apparent relation to catalytic activity, as often proposed, we wondered whether it was not so much the changes in flexibility that diminished catalytic activity but rather the conformational preferences of the enzyme.

To identify trends in the conformations visited by AqdC during MD trajectories and the effect of substitutions, a principal component (PC) analysis was carried out. PC analysis can reveal the main correlated conformational changes occurring in a protein and suggest functional implications.[24–26] The first two PCs found with WT-AqdC both concerned motions in the cap domain, in the same regions where also the highest flexibility was observed (Figure 3). Visualization of the involved structural changes showed that along both PC axes, the shape of the proposed active site entrance channel was changed dramatically, making it vary from narrower to broader (Figure 4 and movies M1 and M2 in the Supporting Information). At negative values of both PCs, this part of the tunnel, surrounded by αCap1 and αCap2, opened up due to the increased distance between the two helices.

A visualization of the conformational preferences of the different AqdC variants (Figures 5 and S3) revealed that from all these investigated variants, the WT protein displayed the most open conformations (i.e., with the most negative values for PC 1 and PC 2). The variant with the lowest activity, AqdC-A134L-G137V, displayed the most dynamic shift away from these open conformations, with the edges of the observed conformations shifting roughly 5 units away from the most open conformations along PC 1 and approximately 12 units along PC 2 (Figure 5). Mutant proteins with intermediate activity, such as AqdC-A134L and AqdC-G137V, displayed intermediate characteristics (Figures 5 and S3). AqdC-A134L displayed open conformations along PC 1 to a similar extent as WT-AqdC but retreated by roughly 6 units along PC 2. AqdC-G137V exhibited fewer open conformations than the WT protein for both PC 1 and PC 2, but to a lesser extent than AqdC-A134L-G137V. For the complete set of nine variants with intermediate catalytic activity, only AqdC-G40K-G137V-G220D-Y238W (which was 2.3-fold less active than the WT, but 4-fold more active than AqdC-A134L-G137V), displayed similar retreats from the most open conformations as the variant with the lowest activity AqdC-A134L-G137V (Figure S3). Thus, with only one exception amongst ten tested variants, the ability to display open conformations at the active site tunnel agreed with the catalytic activity of these variants. From these results, it appears that the mutant proteins are less catalytically active because they have lost the ability to open the active site entrance funnel

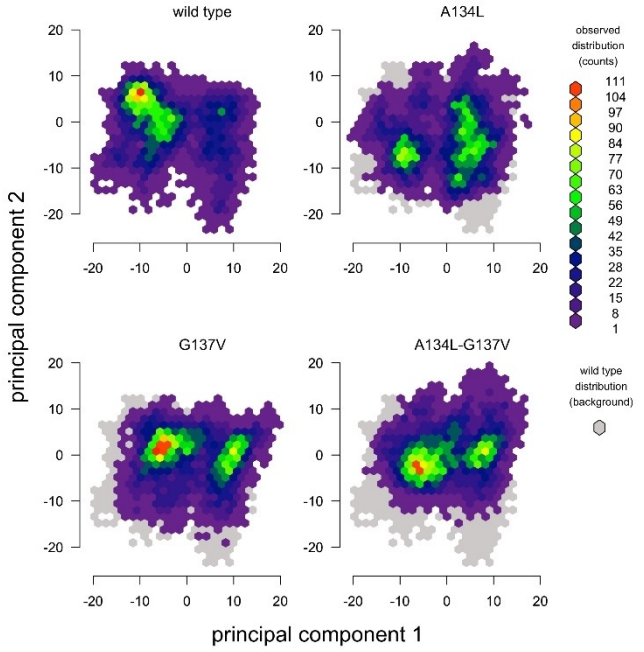

Figure 5. Examples of closure of the funnel-like entrance to the enzyme's active site by substitutions that decrease catalytic activity. For both principal components, the more negative values correspond to more open conformations at the main entrance to the active site. A grey background, which corresponds to the distribution of conformations observed for the WT protein, is shown behind each variant to facilitate comparison between mutant proteins and WT-AqdC. For every mutant protein, more closed conformations are observed than for WT-AqdC, which is the most active protein. Variant AqdC-A134L has an activity corresponding to 39% of the WT activity (based on k_{cat}), AqdC-G137V 28% (based on specific activity), and AdC-A134L-G137V 14% (based on k_{cat}).

effectively. The most likely effect caused by the impaired opening of the cap domain, a reduced substrate entry, would be rate-limiting for the catalytic activity of stabilized variants.

The effects at the second, alkyl-tail part of the bipartite tunnel were also analysed (Figure S2, movies M3 and M4 in the Supporting Information). More negative values for PC 1 clearly corresponded to a more closed distal opening, the opposite of what was found for the funnel-shaped active site entrance. Thus, for the distal part of the tunnel, it does not appear that more open conformations along PC 1 are related to the catalytic process. Negative values of PC 2 corresponded with a more open tunnel, the same trend as found for the presumed "substrate entry" part of the tunnel, but the opening up of the active site was much smaller than the closing by PC 1. Thus, for this part of the tunnel, there is no clear trend showing that a decreased tendency of the protein to adopt open conformations can explain a lower catalytic activity. Therefore, substrate entrance or product release through the distal part of the tunnel (if any) is likely not rate-limiting.

Discussion

In this study, the PQS dioxygenase AqdC was engineered with the goal of obtaining a more stable enzyme, because robustness is a prerequisite for the potential use of AqdC as quorum quenching enzyme *in vivo*. The melting temperature (T_m^{app}) of an enzyme is, in general, a good descriptor for stability, and the reported improvements by individual protein engineering projects are typically in the range of 2 to 15 °C.[27–29] For AqdC, the combination variant V (AqdC-G40K-A134L-G220D-Y238W) that showed the best stability gain while preserving reasonable catalytic activity, displayed a T_m^{app} increase of 11.8 °C relative to the WT protein. The k_{cat} of this protein was 2.6-fold lower than that of WT-AqdC, and its K_M was 3.8 times higher. However, this activity loss was more than offset by its 30-fold increase in half-life ($t_{1/2}$; from about 30 min to almost 10 h). Another most promising combination variant is protein VIII (AqdC-G40K-G220D-Y238W), with its T_m^{app} increased by 6.1 °C compared to WT-AqdC, and a half-life at 37 °C of 1.6 h. The stabilizing effect of the A134 L substitution as present in protein V is remarkable. However, the substitution significantly impairs catalytic activity, mainly due to inducing an increase in K_M. Interestingly, the K_M value of 3.7 μM of protein VIII for PQS is among the lowest observed for all AqdC proteins tested, and its catalytic efficiency (k_{cat}/K_M) corresponds to that of the WT protein. For potential use as a quorum quenching enzyme, a low K_M value for the signal molecule is an important feature.

While it was possible to engineer more stable variants of AqdC, it seems that gains in thermostability were traded against losses in catalytic activity. When engineering enzyme stability, it can occur that (some of) the stabilizing substitutions cause activity loss[30] even though it is more often observed that catalytic activity is well preserved at low temperature.[31–33] It is certainly unusual that virtually all stabilizing substitutions give a decrease in activity, as observed here for AqdC (Table S3). As a result, for AqdC, it was only possible to combine four single amino acid substitutions while maintaining a reasonable catalytic activity. For comparison, very recently, a ketoreductase was engineered using the same FRESCO method resulting in a ninefold combination variant that featured a 45 °C higher T_m^{app} and no loss in catalytic activity at low temperatures.[34]

The decreases in catalytic activity as a consequence of stabilizing amino acid substitutions are probably related to decreased flexibility in the cap domain of AqdC. For another ABH-fold enzyme with a cap domain, the dehalogenase DhlA from *Xanthobacter autotrophicus*, it was found that the flexibility in its cap domain influenced the catalytic activity.[35–37] Mutational, kinetic, and thermodynamic evidence confirmed that a considerable opening up of the active site entrance of DhlA is required to facilitate the release of halides, which probably need to be solvated by multiple water molecules to become released from their binding site in the active site.[38–41] When applying FRESCO stabilization to another ABH-fold dehalogenase (LinB),[42] it was observed that several stabilizing substitutions, located in the cap domain and near the active site entrance, decreased catalytic activity. Furthermore, the Damborsky group has found that mutations in the entrance tunnel

of various related dehalogenases can strongly affect both catalytic activity and stability.[43–49]

Most importantly, the MD simulation results for AqdC corroborate the hypothesis that the stabilizing substitutions – which all led to impairments in catalytic activity to varying extents – consistently result in a more rigid cap domain. The substrate PQS is relatively large and bulky, which would require at least one entrance to the active site to open up during the catalytic cycle. A more rigid cap domain could result in a slower overall reaction rate because substrate entry or product release becomes rate-limiting. Also, while the WT enzyme has the same catalytic activity at 30 and 37 °C, the stabilized variants feature increased catalytic activities at the higher temperature (Tables 2 and 3), probably caused by increased flexibility at higher temperatures.

The principal component analysis revealed that changes in conformational preferences, namely towards more closed states, can also cause the observed decrease in catalytic activity. Both a higher rigidity and increased prominence of more closed conformations could relate to the increased stability of the enzyme. For natural variants of the same enzyme, it is, in general, observed that the thermostable variants are more rigid and less catalytically active at lower temperatures than their mesophilic analogues.[50–54] In larger proteins, thermal inactivation often starts with a local unfolding.[51] Based on the current results with AqdC and previous results with stability engineering of various dehalogenases,[42,43,48] it seems reasonable to assume that for these enzymes, this early unfolding takes place in the cap domain. The cap domain, in any case, needs to open up to allow the substrate to enter and product to leave, and the opened form could be the starting point for early unfolding. Therefore, substitutions that lead to more closed conformations in the cap domain could result in stabilization.

Principal component analysis on the distal opening to the active site of AqdC showed no apparent correlation to activity (Figure S2). It is noteworthy that this part of the bipartite tunnel is not conserved; it is present in AqdC but not in the structurally highly homologous 3-hydroxy-2-methyl-4(1*H*)-quinolone 2,4-dioxygenase Hod.[15,55] Together, this suggests that the second, distal opening of the tunnel of AqdC is not involved in a step that is rate-limiting for the overall reaction.

Alternative explanations for the consistent decrease of catalytic activity by stabilizing substitutions are possible but appear less likely. Perhaps, with AqdC, the substrate needs to reorient after entering the active site to adopt a catalytically productive orientation, and this could become rate-limiting due to small increases in protein rigidity. Such a mechanism could explain the consistent drop in activity with stabilizing substitutions. Alternatively, stabilization may change the active site shape such that the enzyme is less reactive, but it is unclear why this should happen for all the different substitutions.

To decrease the number of variants that need to be screened experimentally, there are methods available to prioritize substitutions by predicting the early unfolding region of a protein and only mutate at positions with the highest priority. The B-fitter approach[56–60] is based on the idea that highly flexible parts of the protein are more likely to be the first sites of unfolding. While the usefulness of such methods is often reported and stabilized enzymes are indeed obtained, hardly ever controls are presented in which also substitutions at positions that are predicted to be unimportant, are characterized. Earlier results with FRESCO stabilized enzymes consistently revealed that stabilizing substitutions also occurred in relatively rigid parts of the enzyme that would not have been targeted with the B-fitter approach. Also, for AqdC, it is now found that most stabilizing substitutions occur in regions that would not have been targeted by the B-fitter approach (Table 1, Figure 2), confirming that stabilizing mutations frequently occur at relatively rigid positions and that approaches excluding these positions, such as the B-fitter approach[59,60] will miss stabilizing substitutions.

Conclusion

In this study, computational library design, MD simulations, and biochemical studies were combined to design and analyse variants of the mycobacterial PQS dioxygenase AqdC with improved thermostability. The study not only yielded a more stable variant of AqdC with a much longer half-life, but also new insights on activity-stability relationships of ABH-fold enzymes were gained. The phenomenon that an increase in enzyme stability is associated with a decrease in catalytic activity is a mild-to-severe problem for a significant part of the enzymes that are targeted by stability engineering. AqdC turns out to demonstrate this phenomenon to an extreme degree. Also, because of its potential medical relevance, AqdC could serve as a model system to better understand this phenomenon. The results here indicate that not only increased rigidity can cause a decrease in activity (as is well established), but that also an increased closing of the entrances to the active site could be the source of the activity drop. This insight has the potential to inspire new design techniques to generate substitutions that increase stability without affecting catalytic activity.

Experimental Section

Computational library design: The X-ray structure of AqdC (PDB ID: 6RA2[15]) was used for molecular modelling. Protein molecule E, which is the only monomer in the asymmetric unit that does not contain gaps, was chosen for all modelling. Residues predicted to be involved in substrate positioning and catalysis were not allowed to mutate. At the beginning of this study, an X-ray structure of AqdC with bound substrate was not available. To define the substrate-binding site, the AqdC structure was superimposed on the closely related structure of Hod (PDB ID: 2WJ4), which has the substrate 3-hydroxy-2-methyl-4(1*H*)-quinolone bound.[55] Residues within 5.5 Å from the substrate were excluded from mutagenesis. Single substitutions were selected based on their predicted effect on the change in Gibbs energy of folding ($\Delta\Delta G_{fold}$).

A new addition to the FRESCO method was to predict the $\Delta\Delta G_{fold}$ of variants with a phylogenetic protocol. Sequence alignment was performed with MAFFT[61] using sequences that were at least 25% identical to AqdC. The sequences were collected from the UNIREF90

database.[62] The abundance of amino acids at a given position was translated into an energy scale using

$$\Delta\Delta G_{\text{fold}}^{\text{predicted}} = RT \ln \frac{[\text{alternative AA}] + 0.01}{[\text{current AA}] + 0.01} \quad (1)$$

in which [current AA] equals the percentage of occurrence of an existing amino acid at that position in AqdC, and [alternative AA] the considered alternative.[63] R is the gas constant, and T the absolute temperature (298 K was used). The 0.01 % is added to the nominator and denominator to prevent both the logarithm of zero and division by zero. If, with this method, the predicted $\Delta\Delta G_{\text{fold}}$ for a substitution was < -5 kJ/mol, the substitution was selected for further evaluation using short molecular dynamics (MD) simulations.

The rest of the procedure to select stabilizing substitution was standard, as described in detail elsewhere.[17] User-friendly scripts can be downloaded from https://groups.google.com/forum/#!forum/fresco-stabilization-of-proteins, and are also available upon request. The $\Delta\Delta G_{\text{fold}}$ values were also predicted using Rosetta[19] and FoldX.[18] For the latter, the standard settings were applied with five repeat calculations, of which the outcomes were averaged. For Rosetta, the accurate and computationally cheap Row3 settings[19] were used. If the predicted $\Delta\Delta G_{\text{fold}}$ with either of these two methods was < -3.5 kJ/mol for a substitution, then the resulting variant was also selected for short MD simulations.

The short MD simulations consisted of five independent MD simulations of 100 ps for each protein variant. Snapshots were recorded every 5 ps, and the last 50 ps were analysed. The precise protocols were described in detail elsewhere.[17,64] The averaged structure from each MD trajectory was visually inspected and compared to that of the WT AqdC. Proteins that displayed an increase in flexibility or recognizable destabilizing features (like increased water exposure of hydrophobic side chains or the introduction of unsatisfied H-bond donors or acceptors) were eliminated. All variants that did not reveal such problems were selected for experimental characterization.

Computational analysis of flexibility and conformational preferences: After experimental characterization, further computational analysis was carried out for a small subset of mutant proteins with single and multisite substitutions. The three-dimensional structures of these variants were modelled using the standard FRESCO script "CombineMutations.mcr", which is available via the forum above. MD simulations were done with YASARA[65] using the Yamber3 force-field, which is an Amber99 derivative that was optimized for structural accuracy.[66] After an energy minimization to remove strain from the structure, the temperature was gradually increased from 5 to 298 K in 30 ps. The MD simulation was subsequently run for an additional 11970 ps, of which the last 8 ns were analysed for RMSF and principal components. Snapshots were recorded every 25 ps. In total, 20 MD simulations were started for each variant, each with a different seeding for the initial atom velocities (all atom velocities did obey a Maxwell-Boltzmann distribution), resulting in different trajectories.[20–23] The same 20×12 ns protocol was used before to compare the flexibility of an engineered and much more stable peptide amidase variant, which had retained the same catalytic activity as the WT protein.[67] The predictions for that system were that the thermostable variant was just as flexible as the WT. The peptide amidase is larger (490 amino acids) than the currently studied AqdC (266 amino acids). Together, this confirms that the protocol provides for sufficient conformational sampling to estimate the relative flexibility of the mutants.

The principal component analysis was done within R software[68] using the Bio3d package.[69] The 11 variants were analysed simultaneously to make the principal components of the different variants directly comparable. The analysis was done using all heavy backbone atoms (the C, O, N, and Cα, instead of using only the Cα), which facilitated three-dimensional visualization. The principal component analysis took only 100 min on a single CPU and required only 28 Gigabytes of memory even though 75 million atom positions were involved (=70620 [snapshots]×266 [residues]×4 [atoms/residue]). It was verified that when the mutant proteins were analysed individually, the first two principal components were the same. Two-dimensional histograms (heatmaps) were created using the Hexbin package[70,71] within R. Structural visualizations, including movies, were created using YASARA.[72]

The B-fitter approach is a commonly used method to predict which positions in a protein should be mutated to stabilize an enzyme.[59,60] To retrospectively compare positions where stabilizing substitutions were found with positions that would have been suggested by the B-fitter method, the same subunit E of 6RA2 was used. As is standard for the B-fitter method, the B-factors of all non-hydrogen atoms were collected for each residue and averaged (manual at http://www.kofo.mpg.de/en/research/biocatalysis). The residues with the highest average B factors received the best ranking.

Chemicals and other materials: All chemicals were obtained from commercial sources at the highest purity available. *Escherichia coli* TOP10 was used for cloning purposes as well as for the heterologous production of AqdC proteins. Incubation of cultures was performed in a New Brunswick Innova 44 orbital shaker (Eppendorf).

Mutagenesis of *aqdC*: The *E. coli* codon-optimized *aqdC* gene from *M. abscessus* subsp. *abscessus* (DSM 44196), attached to a sequence encoding an N-terminal TEV-protease cleavable octa-histidine tag, was amplified from pET28b::his8-TEV-*aqdC*[73] using the primers pair aqdC–N-NcoI_fw/ aqdC–C-HindIII_rv (Table S1) and inserted into the pBAD/His A vector (Invitrogen) by restriction cloning, resulting in plasmid pBAD::his8-TEV-*aqdC*.

Individual amino acids in AqdC were substituted by site-directed mutagenesis using pBAD::his8-TEV-*aqdC* as a template with overlapping mutagenic primers as indicated in Table S2. The PCR reactions (20 μL) contained 10 μL 2×PfuUltra II Hotstart Master Mix (Agilent), 5 ng template plasmid, and 0.5 μM of each primer. A touchdown PCR-protocol was used: 2 min at 95 °C, 17 cycles of 20 s at 95 °C, 20 s at 65 °C (−1 °C/cycle), 2.5 min 72 °C, 10 cycles of 20 s at 95 °C, 20 s at 45 °C, 2.5 min 72 °C and a final 10 min-step at 72 °C. After DpnI digestion (10 U, 2 h, 37 °C), 5 μL of DpnI-digested PCR product were used to transform chemically competent *E. coli* TOP10 (Life Technologies) cells. From each transformation, one of the overnight grown colonies was picked and sent for plasmid extraction and sequencing (Eurofins Genomics) in order to confirm the mutations. Transformants containing verified mutant plasmids were cultivated in LB medium with ampicillin (100 μg mL^{-1}) supplemented with 15 % glycerol and stored at −80 °C.

Protein production: For the first high-throughput (HT) protein production, frozen stocks of each recombinant strain were used to inoculate 1 mL LB supplemented with ampicillin (100 μg mL^{-1}) in 96-deep-well plates. After overnight incubation (37 °C, 200 rpm), 150 μL of these cultures were used to inoculate 5–10 mL of TB in 24-deep-well plates supplemented with ampicillin (100 μg mL^{-1}). For the high-purity (HP) purification of single proteins, 150 mL TB was inoculated to an $OD_{600} \approx 0.05$–0.1 using 10 mL overnight cultures inoculated with frozen stocks. All cultures were incubated at 37 °C, with 200 rpm (HT) or 150 rpm (HP) orbital shaking. At

$OD_{600} \approx 0.9$ (HT) or 0.6–0.8 (HP), the expression of the variants was induced by the addition of 0.2% arabinose. After 4 h at 24 °C and 200 rpm (HT) or 150 rpm (HP), the cells were harvested by centrifugation (HT: 4 °C, 2530 *g*, 15 min; HP: 4 °C, 8820 *g*, 15 min).

High-throughput (HT) protein preparation: Cells were thawed, washed (2 mL, 50 mM Tris·HCl pH 7.5, 150 mM NaCl, 10% (*v*/*v*) glycerol; centrifuged at 4 °C, 2530 *g*, 15 min), and lysed by resuspension in 200 μL lysis buffer (50 mM Tris·HCl pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, 1 mg/mL lysozyme, 0,5 mg/mL DNase I, 10% (*v*/*v*) glycerol, 1 tablet of „cOmplete™, EDTA-free Protease Inhibitor Cocktail"/10 mL) and incubation at 20 °C for 25 min. After a freeze (−80 °C) and thaw (4 °C) cycle, the cell-free crude extract was obtained by centrifugation (4 °C, 2530 *g*, 15 min). The crude extract was diluted using two volumes of binding buffer (50 mM Tris·HCl pH 7.5, 150 mM NaCl, 10% (*v*/*v*) glycerol, 20 mM imidazole), loaded on a 96-well filter plate (AcroPrep, 1.2 μm Supor, PALL) filled with 100 μL Ni–NTA resin pre-equilibrated in binding buffer, and incubated (4 °C, 40 min, nutating shaker). The plate was centrifuged (4 °C, 500 *g*, 1 min) and washed with 800 μL binding buffer. Elution was carried out by centrifugation, using 130 μL HT elution buffer (50 mM Tris·HCl pH 7.5, 150 mM NaCl, 10% (*v*/*v*) glycerol, 250 mM imidazole). A desalting plate (PD Multitrap G-25, GE) was used to remove the imidazole. The eluted fractions were stored in storage buffer (50 mM Tris·HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 10% (*v*/*v*) glycerol) at −20 °C.

High-purity (HP) protein preparation: Cells were thawed, resuspended in binding buffer, and lysed by sonication (Bandelin Sonopuls, amplitude 100%, pulse 1 s, break 0.5 s, up to 3 times 4 min cycles). The soluble fraction was obtained by centrifugation (18000 *g*, 15 min, 4 °C) and applied on a 2 mL Ni–NTA gravity-flow column pre-equilibrated in binding buffer. The flow-through was collected and – after washing the column with 5 mL binding buffer – re-applied. After a 10 mL washing step with binding buffer, 6 mL HP elution buffer 1 (50 mM Tris·HCl pH 7.5, 150 mM NaCl, 10% (*v*/*v*) glycerol, 66 mM imidazole) was applied to wash away unspecifically bound proteins. The next elution step (using 6 mL HP elution buffer 2 (50 mM Tris·HCl pH 7.5, 150 mM NaCl, 10% (*v*/*v*) glycerol, 250 mM imidazole)) was collected, the imidazole was removed by buffer exchange using ultrafiltration, SDS-PAGE was performed to check the purity (Figure S4), and the protein was stored in storage buffer at −80 °C.

Activity assay: The catalytic activity of AqdC and its variants was determined spectrophotometrically by measuring PQS conversion at 337 nm at 30 or 37 °C (as indicated), using an initial PQS concentration of 20 μM. The assay buffer contained 50 mM Tris·HCl pH 8.0, 2 mM EDTA, 4% (*v*/*v*) DMSO, 10% (*w*/*v*) PEG 1500 (DMSO and PEG 1500 solubilize and stabilize PQS). The extinction coefficient of PQS, determined at these conditions, is 10169 $M^{-1} cm^{-1}$. The steady-state kinetic constants were estimated by fitting the Michaelis–Menten equation to the activity data, measured using different PQS concentrations.

Thermal-shift assay: The apparent melting temperature (T_m^{app}) of each protein was deduced from thermal shift assays. For the first HT assays, 20 μL of desalted protein in storage buffer was transferred to a qPCR plate and mixed with 5 μL of 100-fold diluted SYPRO Orange (5000× fluorescent dye, Merck) to an end concentration of 10x SYPRO Orange. For the analysis of HP purified proteins, 20 μL samples consisting of 2 μM protein in storage buffer, and 12.5x SYPRO Orange were prepared. After air bubbles were removed by centrifugation (short spin), the fluorescence of the samples was measured in a qPCR cycler (Bio-Rad, CFX96 C1000 Touch Thermal Cycler) over a range from 20 to 99 °C (0.5 °C increments, 10 s measurement delay at each temperature step). The fluorescence signal was plotted as a function of the temperature (BIORAD CFX Manager 3.1). The T_m^{app} of each variant was identified by plotting the first derivative of the fluorescence emission as a function of temperature ($-dF/dT$ vs. T [°C]); T_m is represented by the minimum of the curve.

Thermostability at 37 °C: Purified proteins were incubated at 37 °C at a concentration of 20 ng/μL in storage buffer. The residual activity was determined at appropriate time points. The resulting plot of activity over time was used to calculate the half-life ($t_{1/2}$; period required for the enzymatic activity to decrease to half of its initial value).

Acknowledgments

S.C.W. thanks Misun Lee (Groningen) for helpful advice. Computational work was performed at the Peregrine cluster of the high-performance computing centre at the University of Groningen. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (FE 383/25-1 to S.F.). Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: alpha/beta hydrolase fold · protein engineering · signal molecule · structure-activity relationships · thermostability

[1] M. F. Moradali, S. Ghods, B. H. A. Rehm, *Front. Cell. Infect. Microbiol.* **2017**, *7*, 39.
[2] M. Schuster, C. P. Lostroh, T. Ogi, E. P. Greenberg, *J. Bacteriol.* **2003**, *185*, 2066–2079.
[3] S. Heeb, M. P. Fletcher, S. R. Chhabra, S. P. Diggle, P. Williams, M. Cámara, *FEMS Microbiol. Rev.* **2011**, *35*, 247–274.
[4] M. Schuster, E. P. Greenberg, *Int. J. Med. Microbiol.* **2006**, *296*, 73–81.
[5] G. Girard, G. V. Bloemberg, *Future Microbiol.* **2008**, *3*, 97–106.
[6] J. Lin, J. Cheng, Y. Wang, X. Shen, *Front. Cell. Infect. Microbiol.* **2018**, *8*, 230.
[7] S. Fetzner, *J. Biotechnol.* **2015**, *201*, 2–14.
[8] M. E. Mattmann, H. E. Blackwell, *J. Org. Chem.* **2010**, *75*, 6737–6746.
[9] A. Thomann, A. G. G. de Mello Martins, C. Brengel, M. Empting, R. W. Hartmann, *ACS Chem. Biol.* **2016**, *11*, 1279–1286.
[10] D. Maura, S. L. Drees, A. Bandyopadhaya, T. Kitao, M. Negri, M. Starkey, B. Lesic, S. Milot, E. Déziel, R. Zahler, M. Pucci, A. Felici, S. Fetzner, F. Lépine, L. G. Rahme, *ACS Chem. Biol.* **2017**, *12*, 1435–1443.
[11] C. Müller, F. S. Birmes, C. Rückert, J. Kalinowski, S. Fetzner, *Appl. Environ. Microbiol.* **2015**, *81*, 7720–7729.
[12] C. Pustelny, A. Albers, K. Büldt-Karentzopoulos, K. Parschat, S. R. Chhabra, M. Cámara, P. Williams, S. Fetzner, *Chem. Biol.* **2009**, *16*, 1259–67.
[13] F. S. Birmes, R. Säring, M. C. Hauke, N. H. Ritzmann, S. L. Drees, J. Daniel, J. Treffon, E. Liebau, B. C. Kahl, S. Fetzner, *Infect. Immun.* **2019**, *87*, e00278–19.
[14] S. C. Wullich, A. Arranz San Martín, S. Fetzner, *Appl. Environ. Microbiol.* **2020**, *86*, e00279–20.
[15] S. C. Wullich, S. Kobus, M. Wienhold, U. Hennecke, S. H. J. Smits, S. Fetzner, *J. Struct. Biol.* **2019**, *207*, 287–294.
[16] D. N. Collier, L. Anderson, S. L. McKnight, T. L. Noah, M. Knowles, R. Boucher, U. Schwab, P. Gilligan, E. C. Pesci, *FEMS Microbiol. Lett.* **2002**, *215*, 41–46.
[17] H. J. Wijma, M. J. L. J. Fürst, D. B. Janssen, *Protein Eng. Methods Protoc.*, **2018**, pp. 69–85.

[18] R. Guerois, J. E. Nielsen, L. Serrano, *J. Mol. Biol.* **2002**, *320*, 369–387.
[19] E. H. Kellogg, A. Leaver-Fay, D. Baker, *Proteins* **2011**, *79*, 830–838.
[20] L. S. D. Caves, J. D. Evanseck, M. Karplus, *Protein Sci.* **1998**, *7*, 649–666.
[21] L. Monticelli, E. J. Sorin, D. P. Tieleman, V. S. Pande, G. Colombo, *J. Comput. Chem.* **2008**, *29*, 1740–1752.
[22] S. Genheden, U. Ryde, *J. Comput. Chem.* **2011**, *32*, 187–195.
[23] H. J. Wijma, S. J. Marrink, D. B. Janssen, *J. Chem. Inf. Model.* **2014**, *54*, 2079–2092.
[24] A. Amadei, A. B. M. Linssen, H. J. C. Berendsen, *Proteins* **1993**, *17*, 412–425.
[25] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, V. S. Pande, *Biophys. J.* **2015**, *109*, 1528–1532.
[26] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, L. S. D. Caves, *Bioinformatics* **2006**, *22*, 2695–2696.
[27] H. J. Wijma, R. J. Floor, D. B. Janssen, *Curr. Opin. Struct. Biol.* **2013**, *23*, 588–594.
[28] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.
[29] A. S. Bommarius, M. F. Paye, *Chem. Soc. Rev.* **2013**, *42*, 6534–65.
[30] J. E. Diaz, C.-S. Lin, K. Kunishiro, B. K. Feld, S. K. Avrantinis, J. Bronson, J. Greaves, J. G. Saven, G. A. Weiss, *Protein Sci.* **2011**, *20*, 1597–1606.
[31] B. Van den Burg, G. Vriend, O. R. Veltman, G. Venema, V. G. H. Eijsink, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2056–2060.
[32] L. Serrano, A. G. Day, A. R. Fersht, *J. Mol. Biol.* **1993**, *233*, 305–312.
[33] J. C. Williams, J. P. Zeelen, G. Neubauer, G. Vriend, J. Backmann, P. A. M. Michels, A. M. Lambeir, R. K. Wierenga, *Protein Eng.* **1999**, *12*, 243–250.
[34] F. S. Aalbers, M. J. Fürst, S. Rovida, M. Trajkovic, J. R. Gómez Castellanos, S. Bartsch, A. Vogel, A. Mattevi, M. W. Fraaije, *eLife* **2020**, *9*, e54639.
[35] J. P. Schanstra, A. Ridder, J. Kingma, D. B. Janssen, *Protein Eng.* **1997**, *10*, 53–61.
[36] J. P. Schanstra, I. S. Ridder, G. J. Heimeriks, R. Rink, G. J. Poelarends, K. H. Kalk, B. W. Dijkstra, D. B. Janssen, *Biochemistry* **1996**, *35*, 13186–13195.
[37] J. P. Schanstra, D. B. Janssen, *Biochemistry* **1996**, *35*, 5624–5632.
[38] D. B. Janssen, *Curr. Opin. Chem. Biol.* **2004**, *8*, 150–159.
[39] G. H. Krooshof, R. Floris, A. W. J. W. Tepper, D. B. Janssen, *Protein Sci.* **1999**, *8*, 355–360.
[40] K. H. G. Verschueren, F. Seljée, H. J. Rozeboom, K. H. Kalk, B. W. Dijkstra, *Nature* **1993**, *363*, 693–698.
[41] S. M. Franken, H. J. Rozeboom, K. H. Kalk, B. W. Dijkstra, *EMBO J.* **1991**, *10*, 1297–1302.
[42] R. J. Floor, H. J. Wijma, D. I. Colpa, A. Ramos-Silva, P. A. Jekel, W. Szymański, B. L. Feringa, S. J. Marrink, D. B. Janssen, *ChemBioChem* **2014**, *15*, 1660–1672.
[43] T. Koudelakova, R. Chaloupkova, J. Brezovsky, Z. Prokop, E. Sebestova, M. Hesseler, M. Khabiri, M. Plevaka, D. Kulik, I. Kuta Smatanova, P. Rezacova, R. Ettrich, U. T. Bornscheuer, J. Damborsky, *Angew. Chem. Int. Ed.* **2013**, *52*, 1959–1963; *Angew. Chem.* **2013**, *125*, 2013–2017.
[44] R. Chaloupková, J. Sýkorová, Z. Prokop, A. Jesenská, M. Monincová, M. Pavlová, M. Tsuda, Y. Nagata, J. Damborský, *J. Biol. Chem.* **2003**, *278*, 52622–52628.
[45] L. Biedermannová, Z. Prokop, A. Gora, E. Chovancová, M. Kovács, J. Damborský, R. C. Wade, *J. Biol. Chem.* **2012**, *287*, 29062–29074.
[46] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R. C. Wade, M. Tsuda, Y. Nagata, J. Damborsky, *Nat. Chem. Biol.* **2009**, *5*, 727–733.
[47] M. Klvana, M. Pavlova, T. Koudelakova, R. Chaloupkova, P. Dvorak, Z. Prokop, A. Stsiapanava, M. Kuty, I. Kuta-Smatanova, J. Dohnalek, P. Kulhanek, R. C. Wade, J. Damborsky, *J. Mol. Biol.* **2009**, *392*, 1339–1356.
[48] V. Liskova, D. Bednar, T. Prudnikova, P. Rezacova, T. Koudelakova, E. Sebestova, I. K. Smatanova, J. Brezovsky, R. Chaloupkova, J. Damborsky, *ChemCatChem* **2015**, *7*, 648–659.
[49] J. Brezovsky, P. Babkova, O. Degtjarik, A. Fortova, A. Gora, I. Iermak, P. Rezacova, P. Dvorak, I. K. Smatanova, Z. Prokop, R. Chaloupkova, J. Damborsky, *ACS Catal.* **2016**, *6*, 7597–7610.
[50] P. Zavodszky, J. Kardos, A. Svingor, G. A. Petsko, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7406–7411.
[51] V. G. H. Eijsink, A. Bjørk, S. Gåseidnes, R. Sirevåg, B. Synstad, B. Van Den Burg, G. Vriend, *J. Biotechnol.* **2004**, *113*, 105–120.
[52] M. Wolf-Watz, V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, D. Kern, *Nat. Struct. Mol. Biol.* **2004**, *11*, 945–949.
[53] M. Vihinen, *Protein Eng.* **1987**, *1*, 477–80.
[54] P. L. Wintrode, F. H. Arnold, *Adv. Protein Chem.*, **2001**, pp. 161–225.
[55] R. A. Steiner, H. J. Janssen, P. Roversi, A. J. Oakley, S. Fetzner, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 657–662.
[56] H. Yu, Y. Yan, C. Zhang, P. A. Dalby, *Sci. Rep.* **2017**, *7*, 41212.
[57] Z. Sun, Q. Liu, G. Qu, Y. Feng, M. T. Reetz, *Chem. Rev.* **2019**, *119*, 1626–1665.
[58] S. Mao, X. Cheng, Z. Zhu, Y. Chen, C. Li, M. Zhu, X. Liu, F. Lu, H.-M. Qin, *Enzyme Microb. Technol.* **2020**, *132*, 109441.
[59] M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem. Int. Ed.* **2006**, *45*, 7745–7751; *Angew. Chem.* **2006**, *118*, 7909–7915.
[60] W. Augustyniak, A. A. Brzezinska, T. Pijning, H. Wienk, R. Boelens, B. W. Dijkstra, M. T. Reetz, *Protein Sci.* **2012**, *21*, 487–497.
[61] K. Katoh, D. M. Standley, *Mol. Biol. Evol.* **2013**, *30*, 772–780.
[62] B. E. Suzek, Y. Wang, H. Huang, P. B. McGarvey, C. H. Wu, *Bioinformatics* **2015**, *31*, 926–932.
[63] B. Steipe, B. Schiller, A. Plückthun, S. Steinbacher, *J. Mol. Biol.* **1994**, *240*, 188–192.
[64] H. J. Wijma, R. J. Floor, P. A. Jekel, D. Baker, S. J. Marrink, D. B. Janssen, *Protein Eng. Des. Sel.* **2014**, *27*, 49–58.
[65] E. Krieger, G. Vriend, *J. Comput. Chem.* **2015**, *36*, 996–1007.
[66] E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, G. Vriend, *Proteins* **2004**, *57*, 678–83.
[67] B. Wu, H. J. Wijma, L. Song, H. J. Rozeboom, C. Poloni, Y. Tian, M. I. Arif, T. Nuijens, P. J. L. M. L. M. Quaedflieg, W. Szymanski, B. L. Feringa, D. B. Janssen, *ACS Catal.* **2016**, *6*, 5405–5414.
[68] "R Core Team (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*," can be found under http://www.r-project.org/., **2013**.
[69] L. Skjærven, X.-Q. Yao, G. Scarabelli, B. J. Grant, *BMC Bioinf.* **2014**, *15*, 399.
[70] D. Carr, N. Lewin-Koh, M. Maechler, "'hexbin: Hexagonal binning routines' R package version 1.26.0," can be found under http://github.com/edzer/hexbin, **2011**.
[71] D. B. Carr, R. J. Littlefield, W. L. Nicholson, J. S. Littlefield, *J. Am. Stat. Assoc.* **1987**, *82*, 424–436.
[72] E. Krieger, G. Vriend, *Bioinformatics* **2014**, *30*, 2981–2982.
[73] F. S. Birmes, T. Wolf, T. A. Kohl, K. Rüger, F. Bange, J. Kalinowski, S. Fetzner, *Front. Microbiol.* **2017**, *8*, 339.

Manuscript received: September 9, 2020
Revised manuscript received: October 13, 2020
Accepted manuscript online: October 14, 2020
Version of record online: November 16, 2020