Chapter 7:

Stabilization of Cyclohexanone Monooxygenase by Computational and Experimental Library Design

Maximilian J. L. J. Fürsta, Marjon Boonstraa, Selle Bandstraa, and Marco W. Fraaijea*

aMolecular Enzymology Group, University of Groningen, Nijenborgh 4, 9747AG, Groningen, The Netherlands

*Corresponding author

Published in:

Biotechnology and Bioengineering, 2019

Reprinted with permission. ©2019 John Wiley & Sons, Inc.
Abstract

Enzymes often by far exceed the activity, selectivity and sustainability achieved with chemical catalysts. One of the main reasons for the lack of biocatalysis in the chemical industry is the poor stability exhibited by many enzymes when exposed to process conditions. This dilemma is exemplified in the usually very temperature-sensitive enzymes catalyzing the Baeyer-Villiger reaction (BVMOs), which display excellent stereo- and regioselectivity and offer a green alternative to the commonly employed, explosive peracids. Here we describe a protein engineering approach applied to cyclohexanone monooxygenase from Rhodococcus sp. HI-31, a substrate promiscuous enzyme that efficiently catalyzes the production of the nylon-6 precursor ε-caprolactone. We employed a framework for rapid enzyme stabilization by computational libraries (FRESCO), which predicts protein-stabilizing mutations. From 128 screened point mutants, approximately half had a stabilizing effect, albeit mostly to a small degree. To overcome incompatibility effects observed upon combining the best hits, an easy shuffled library design strategy was devised. The most stable and highly active mutant displayed an increase in unfolding temperature of 13 °C, and an approximately 33x increase in half-life at 30 °C. In contrast to the wild-type enzyme, this thermostable 8x mutant is an attractive biocatalyst for biotechnological applications.
Introduction

The oxidation of ketones with peroxides and peracids is known as the Baeyer-Villiger (BV) reaction, named after the scientists who discovered it 120 years ago.\(^1\) The credit for the “invention” goes to nature, however, where mechanistically analogously working enzymes called Baeyer-Villiger monooxygenases (BVMOs) were presumably already present in the last universal common ancestor.\(^2\) BVMOs gained attention in the field of biocatalysis, because several representatives accept a very wide range of substrates, and catalysis often occurs chemo-, regio-, and stereoselectively.\(^3\) The BV reaction mechanism commences with a nucleophilic attack of the peroxide on the carbonyl carbon. In BVMOs, the catalytic entity is a peroxylflavin, generated from the reaction of molecular oxygen with reduced flavin adenine dinucleotide (FAD), which in turn received the electrons from NADPH. After the addition step, the so called “Criegee-intermediate” rearranges, and generally the carbonyl-adjacent carbon that is better at stabilizing a positive charge migrates to the incorporated oxygen.\(^4\) Enzymes can violate this “migratory aptitude” by binding the substrate such that the required antiperiplanar conformation of the canonically migrating bond is prevented, thus forcing the arrangement to the “abnormal” BV product.\(^5\) An additional advantage of biocatalysis is the use of O\(_2\) as oxidant, in contrast to the requirement of stoichiometric amounts of a peroxide reactant in the chemical reaction. This circumvents the peril of highly explosive concentrated peracids, as well as the generation of hazardous waste.\(^6\) Since these problems are above all relevant for industrial processes, biocatalytic alternatives should also be applicable at this scale.

One major hurdle to that end has efficiently been addressed in recent years: the prohibitively expensive nicotinamide cofactor consumed in the reaction is now routinely recycled by using whole-cell systems, or by using a redox partner enzyme.\(^7\) Various alternatives have been described, making for example use of dehydrogenases that regenerate NADPH from NADP\(^+\), either at the expense of a cheap sacrificial substrate, or with concurrent oxidation of a cascade reaction-participating substrate.\(^8\) Molecular fusions of the recycling enzymes facilitate handling and seldom affect enzyme activity.\(^9-10\) A remaining issue, however, is the poor stability of most BVMOs. In an industrial process, high temperatures and the use of co-solvents often are a prerequisite, or greatly increase efficiency or suitability. These conditions are incompatible with many BVMOs, which can inactivate within minutes, even in buffer and at room temperature. The only truly thermostable BVMO, phenylacetone monooxygenase (PAMO) from *Thermobifida fusca*, unfortunately presents a strictly limited substrate scope and poor selectivities.\(^11\) PAMO was the first
Engineering a Thermostable CHMO

BVMO whose crystal structure was solved,\textsuperscript{12} and despite intensive subsequent protein engineering,\textsuperscript{13-16} catalytic efficiencies with most substrates remain too low to achieve industrially required productivities. One of the most important BV transformation, the conversion of cyclohexanone to $\varepsilon$-caprolactone, a precursor to nylon-6, is particularly difficult to achieve with PAMO. There is, however, a subgroup of BVMOs that is highly active on this substrate, the so-called cyclohexanone monoxygenases (CHMOs). In fact, the first BVMO isolated was a CHMO stemming from \textit{Acinetobacter} sp. NCIMB 9871 (AcCHMO). Although this enzyme is particularly unstable, it has become a prototype in BVMO research because of its high activities and substrate promiscuity, while often showing excellent selectivities. Three studies have been reported in which AcCHMO’s robustness could be improved: two studies introduced disulfide bridges,\textsuperscript{17-18} and the best mutants displayed an increase in the protein’s melting temperature ($T_m$), defined as the midpoint of the thermal denaturation curve, of 5 °C and 6 °C, respectively. In an attempt to target the enzyme’s oxidative stability, the third study introduced point mutations at sulfur-containing residues.\textsuperscript{19} While not measuring $T_m$, the authors reported on the temperature at which 50% of activity remained after incubation at various temperatures, which shifted by 6.8 °C in the best mutant. While these engineering efforts were greatly hampered by the unavailability of a crystal structure, structures are now available of two close homologs of AcCHMO: RhCHMO from \textit{Rhodococcus} sp. HI-31,\textsuperscript{20} and TmCHMO from \textit{Thermocrispum municipale}.\textsuperscript{21} The various structures of RhCHMO advanced mechanistical BVMO research, as it allowed intricate insight into both structural and catalytic features of CHMOs.\textsuperscript{20,22-23} TmCHMO, on the other hand, can be seen as a big progression toward applicable CHMOs, and with a $T_m$ of 48 °C is considerably more stable than both AcCHMO and RhCHMO.\textsuperscript{21} Yet, TmCHMO also shows a significantly lower activity on cyclohexanone, with a $k_{cat}$ of 2 s\textsuperscript{-1} being approximately 5 times slower than the two homologs. Because RhCHMO remains a prototype in BVMO research, and its substrate spectrum is highly similar to the by far best-studied BVMO AcCHMO, we attempted to engineer RhCHMO towards increased thermostability. To that end, we applied our previously described framework for rapid enzyme stabilization by computational libraries (FRESCO).\textsuperscript{24-25} This method uses two independent algorithms to calculate the difference in folding free energy, $\Delta\Delta G_{\text{Fold}}$, between a wild-type structure and a single point mutant. After an \textit{in silico} screening step supported by molecular dynamics simulations, selected mutants are expressed and tested for improved stability, and successful hits are combined to a final stabilized mutant. So far, we applied the FRESCO procedure to five enzymes, and achieved considerable improvements, with $T_m$ increases up to 35 °C.\textsuperscript{24,26-29}
Here we describe the results gained by applying FRESCO to RhCHMO, where we could increase the $T_m$ by up to 15 °C in the most stable mutant. Although numerous stabilizing point mutations were encountered and confirmed the high success rate of the computational predictions, several complications needed to be overcome: many stability increases were small and not additive upon combination, and some mutations caused insoluble or inactive enzymes. Besides investigating potential causes of these effects, we also devised a novel methodology to easily create random combinatorial mutant libraries, which aided in obtaining the most stable variant.

**Results**

**In silico predictions**

There are five different crystal structures of RhCHMO in the protein data bank (PDB), all of which have been suggested to represent a distinct conformation adopted by the enzyme throughout its catalytic mechanism. Although all structures contain both the FAD and the NADP cofactor, previous computational analysis of RhCHMO was based on the “closed” structure (PDB ID 3GWF), because it was assumed to represent the catalytically more relevant conformation in which the substrate is converted. While the “open” structure (3GWF) was proposed to represent the enzyme in the ligand free form, the lack of electron density for a large loop on the protein’s surface already impedes a reliable computational analysis. This so-called control-loop, stretching from residue 488 to 504, is also lacking in the “rotated” (3UCL) and “loose” (4RG4) conformations.

For the FRESCO calculations, we used the fifth, “tight” RhCHMO structure (4RG3), which presents a “closed” conformation with an ordered control loop, contains a ligand in the active site, and is the structure with the highest resolution. We applied the in silico protocols as previously described in detail, which predict stabilizing mutations by applying the Rosetta-DDG and FoldX algorithms in combination with MD simulations. Instead of performing the calculations on all of the 10,260 possible point mutants ($19 \times 540$ residues), mutagenesis of catalytically relevant positions was avoided by excluding residues in a distance smaller than 5 Å from the active site. As BVMOs stably bind to FAD and NADP$,^4$ we also had to exclude residues in the vicinity of these two cofactors (Figure S7.1). Although the preservation of cofactor affinity is an obvious aim, the main reason is the Rosetta and FoldX algorithms are not parameterized to work with cofactors. Any calculations for these residues would therefore be unreliable. After ranking the calculations of the remaining 8,350 mutations by energy (Figure S7.1), the 853 variants that
showed a $\Delta \Delta G^{\text{fold}} \leq -5$ kJ mol$^{-1}$ underwent short MD simulations, which included the cofactors. The generated in silico structures were then visually inspected. After discarding chemically unreasonable mutations and those where MD simulations suggested an increase in flexibility of the protein, we chose initially 114 point mutations for experimental verification (Figures S7.1, 1).

Figure 1. Tm increase of RhCHMO for the first (A) and second (B) set of point mutants. Stabilizing mutations with a $T_m$ increase of at least 1 °C are shaded pink, stabilizing mutations not considered for mutant combinations are indicated by a striped pattern.

Stability determination of point mutants

We performed the experimental stabilization procedure (see Figure S7.1 for an overview of the workflow) from DNA mutagenesis to $T_m$ measurements of purified enzymes following a protocol recently published as a step-by-step description. A 96-well format was maintained throughout nearly the entire procedure. First, we exchanged the targeted codon using the QuikChange method on a plasmid harboring the RhCHMO gene fused to a NADPH-recycling phosphite dehydrogenase (PTDH)-encoding gene. After transformation into E. coli NEB 10 beta cells, we expressed the proteins on a 2 mL scale. The cell-free extract (CFE) of the lysed cells was then subjected to affinity-chromatography that took advantage of an N-terminal 6x-histidine tag. The $T_m$ of the purified protein was determined in a flavin-fluorescence-exploiting thermal shift assay called ThermoFAD. As it is convenient to avoid full purification and perform $T_m$ determinations using cell-free extracts (CFE), but the results may vary significantly, we measured $T_m$ for both the CFE and the purified proteins. With a coefficient of determination $R^2$ of only 0.68 (Figure S7.2), we deemed the $T_m$ of RhCHMO to apparently be too strongly influenced by the soluble cellular components in the CFE, and decided for $T_m$ measurements to rely on purified proteins only.

Out of 114 of the RhCHMO single mutants initially selected from the FRESCO calculations, 18 repeatedly gave no detectable ThermoFAD signal, likely caused by insufficient expression of soluble protein and/or a lack of FAD
incorporation. Although BVMOs with lowered cofactor affinity can be reconstituted to the holoenzyme by adding an excess of FAD during purification, this is an unattractive feature for application and no reconstitution attempt was made. Of the remaining 96 mutants, 34 had a decreasing effect, 13 had no effect and 50 had an increasing effect on $T_m$ (Figure 1A). Although this success rate compares favorably with previous FRESCO stabilizations, where typically 20-40% were found stabilizing, most of the $T_m$ improvements we found for RhCHMO were very small. When we selected 13 mutations with a $\Delta T_m$ of at least 1 °C and produced the proteins on a larger scale for verification, we nevertheless confirmed the stabilizing effect for all but one mutation (E293A). The remaining 12 were subjected to the next step in the FRESCO procedure, the combination of mutations in a multiple mutant.

**Mutant combination libraries**

Initially, we combined the most stabilizing mutations by successive rounds of QuikChange. When we tested the mutant proteins, however, we noticed that the $T_m$ increases were not additive, and that the addition of one mutation (A11M) caused expression of insoluble protein. The most stable mutant with a $T_m$ increase of 6.3 °C was the 4x mutant $Q409P/N431Y/A115V/A455V$ (M4). As we also observed unsatisfactory results when adding mutations in another order, we decided to change strategy and devised a more systematic method. We previously created a combinatorial library of wild-type and mutant residues by golden gate shuffling using synthetic genes. To explore a method that avoids DNA synthesis, we now applied a PCR- and Gibson assembly-based method. We decided to use the M4 mutant as a starting point and to exclude the A11M mutation, while randomizing the remaining seven mutations. As two mutations targeted position E91, the library size was 96 ($2^7 \times 3$). Initially, we created the library using an altered method described for the simultaneous mutation of several positions on a plasmid. Details of our adapted procedure, which is based on the direct generation of mixed wild-type and mutated PCR fragments and Gibson assembly, are described in the experimental section. DNA sequencing of 95 clones of the library confirmed a shuffled library, but revealed an uneven distribution of wild-type and mutant in some cases (Figure S7.4A). When we expressed three 96-well plates of individual clones (to achieve ~95% library coverage if there was no bias) and measured the purified protein’s $T_m$, we found only a few hits with marginally increased stability (data not shown). This seemed to suggest that our starting point was a dead-end combination that had reached a thermostability plateau. Noticing also the absence of a signal in ~70% of the clones, we cross-referenced the results with their sequences. We found that the mutants giving
a signal always had the wild-type residue at two positions (D10Y and D119R), which we thus deemed as incompatible for combinations (Figure S7.4B). We decided to create a second library without these mutations and tried to escape the stability dead-end by randomizing A115V and A455V, using the double mutant Q409P/N431Y (M2) as the new template. To furthermore create the resulting library of again 96 possibilities without a bias, we also adapted the method for generating the library (Figure 2). We now first created a fully mutated plasmid and then generated mixed wild-type and mutated PCR fragments with primers remote from the mutated site. We sequenced 48 clones of the assembled and transformed library and again found successful randomization. Furthermore, a $\chi^2$ test on the data found no significant deviation from the expected 50% (or 33% in the case of position 91) ratio (Figure S7.5, Table S7.1). We also sequenced a culture of a mix of all colonies, and as the mixed signals in the electropherogram reflected the same distribution, we conclude that this suffices in verifying such libraries quickly (Figure S7.6).

![Figure 2](image.png)

**Figure 2.** Method to create the second shuffled library without bias. The wild-type plasmid was used as a template in PCRs with mutated primers to generate DNA fragments containing two mutations on each end. The purified fragments were assembled by Gibson cloning to obtain a plasmid with all mutations. An equal mix of this mutated and the wild-type plasmid was used in the next round of PCR where primers were used that bound outside the mutated region to generate a mix of fragments with and without mutations. Upon Gibson assembly of this mix, the shuffled library was obtained.

We then produced the proteins from three 96-well plates of individual clones and measured $T_m$. Although we found a much higher signal rate (81% of all clones), again no combination of mutations increased the thermostability of RhCHMO to a large extent (data not shown). At this point, the strongest increase in $T_m$ with respect to the wild type was 7.3 °C found for a 7x mutant (M7), only 1 °C more stable than the M4 mutant (Figure 3).
Engineering a Thermostable CHMO

Acknowledging the fact that we seemed to have found the largest possible $T_m$ increase achievable with our initially selected mutations but being unsatisfied with the result, we decided to test an additional small set of point mutations. Revisiting the computationally predicted list of single mutants, we therefore selected 14 more mutations to be tested experimentally. Following the same procedure as before, we found no signal for four, a destabilizing or no effect for five, and a significant increase in $T_m$ for three mutations (Figure 1B). At this point, an independent study found indications for an implication of one of the stabilizing residues, located in the substrate tunnel, on the substrate scope of a related CHMO.36 For this reason we excluded this mutation (L509F) from further combination. The other two mutations were directly added to our most stable mutant M7. Gratifyingly, we found a larger than additive effect on $T_m$: while the single mutants D95H and Q191M increased the $T_m$ by 4.5 °C and 1.5 °C, respectively, addition to M7 increased the $T_m$ by 6.4 °C and 3.2 °C, respectively. When we added both mutations, the effect remained additive in the resulting mutant M9, which showed a $T_m$ increase of 15 °C with respect to wild type.

As the FRESCO protocol foresees to exclude residues in a 5 Å radius from the active site, changes in activity are rarely observed in stabilized mutants, with the exception of a common upshift of the temperature optimum. There are indications, however, that the activity of BVMOs is not purely determined by active-site residues.36 When we checked the most stable mutants for activity on the native cyclohexanone substrate, we found a highly diminished activity for the M9 mutant, but not for M7. We then looked at the activity of the 8x mutant with the second highest $T_m$ (49 °C), and found this mutant (M8B) to be highly active again. In comparison to M9, M8B only lacks mutation Q191M, which was thus identified as the activity-abolishing mutation. We also determined the temperature optimum of the two enzyme variants and found a shift that was very much in line with their melting temperatures (Figure 4A). The kinetic constants were then measured at standard conditions and we
found a $k_{cat}$ of 4.9 s$^{-1}$ for the mutant compared to 9.2 s$^{-1}$ for the wild type (Figure S7.7, Table 1). The high affinity for cyclohexanone was retained in the mutant, and the $K_m$ could only be estimated to be < 1 µM for both wild type and mutant (Table 1). To estimate the monooxygenases-inherent degree to which NADPH oxidation is uncoupled from substrate oxygenation, we also measured rates in the absence of substrate. We found a very low uncoupling rate of 0.05 s$^{-1}$ in the wild type, which only slightly increased in the mutant (0.13 s$^{-1}$).

Figure 4. Comparison of the stabilized mutants with wild-type RhCHMO. A) Temperature optimum of wild-type RhCHMO and M8B mutant. B) Activity over time upon incubation at 30 °C. C) Comparison of $T_m$ and half-life.

Previous FRESCO projects also demonstrated that the proteins’ $T_m$ increase was accompanied by higher operational stability, i.e. their ability to retain activity over time upon exposure to process conditions. In order to determine this characteristic for our most stable RhCHMO mutants as well, we measured residual activity of the wild type, the M8B, and the intermediate M7 mutant after incubation at 30 °C and 37 °C (Figures 4B, S7.8). These measurements demonstrated a strong correlation between $T_m$ and operational stability, as the enzyme’s half-life of activity at 30 °C was improved 33-fold from 22 minutes in the wild type to 12.3 hours. For the M7 mutant, we measured a half-life of ~5 h, which lies approximately in the middle between wild type and M8B mutant, as does the $T_m$ of this variant (Figure 4C). Similarly, the half-life at 37 °C increased from 8.6 min in the wild type to 135 min in the M8B mutant (Figure S7.8). We thus again demonstrated the high correlation between a protein’s unfolding temperature and its stability over time. Considering the largely retained catalytic properties (Figure S7.7) and the high increase in operational stability, the RhCHMO M8B mutant is a much more suitable biocatalyst for application than both its wild-type template and the very similar AcCHMO (Table 1).

Structural analysis of mutations

In order to rationalize some of the effects observed by the mutations, we analyzed the location and environment of the respective residues in the protein’s structure.

The two most stabilizing mutations of the first set, combined in the M2 mutant, were N431Y and Q409P, which are located very closely together (Figure 5A).
For the N431Y mutation, a combination of improved protein packing and hydrogen-bonding network seems to occur: in the wild-type structure, the asparagine fringes a cavity, which is filled by the bulky tyrosine in the mutant. Furthermore, the carbonyl of wild-type N431 is unable to hydrogen bond and this is prevented in the mutant, where van der Waals interactions with the nearby V505 occur instead. The asparagine’s amine group coordinates two water molecules, one of which is replaced by the tyrosine’s hydroxyl group in the mutant, and which constitutes the link to Q409. For this residue, the effect of the mutation stems from elsewhere: Q409 is located on a surface-exposed loop and, unlike proline, can hydrogen bond to the solvent waters. Analysis of the MD simulations indicate that the more rigid torsion angles forced upon the backbone by proline decrease the flexibility of the loop and nearby regions, which likely is the cause of the improved stability (Figure S7.9).

D95 is located at the center of a negatively charged surface patch (Figure S7.10A, 5B). As no obvious hydrogen bonding impairments or flexibility alterations are observed in the histidine mutant, the stabilizing effect might be related to the charge change. One possibility is that the mutated residue forms a salt bridge, assuming that the solvent-exposed histidine surrounded by acidic residues is mostly in the protonated form. Such an interaction is, however, not observed in the predicted structure. The stabilizing effect could therefore simply result from preventing a thermodynamically unfavorable clustering of anionic residues (Figure S7.10).

The activity-abolishing effect of mutation Q191M could be related to a hydrogen bond that this residue establishes with a strictly conserved glycine that is part of the NADP-binding Rossmann fold motif (Figure 5C). Although this interaction occurs through Q191’s backbone amine, even small perturbations of the backbone entailed by other sidechains may be enough to critically impair folding and/or NADP binding. It is also noteworthy that Q191 is part of the NADP domain, but connects to the BVMO’s so-called helical domain through a hydrogen bond of the amide with S214. The residue thus might also be critical for domain orientation and/or movement, which may occur in these proteins during catalysis. Performing a multiple sequence alignment, we also found Q191 to be highly conserved in BVMOs, further suggesting an important functional role (Figure S7.11).

Interestingly the final stabilized mutant contains most mutations in the FAD domain (Figure 5D). While some mutations were found in the NADP domain in the single mutant screen, they proofed incompatible in the combination mutants. Also one of the two mutations in the helical domain was discarded in this step.
Figure 5. Mutations introduced in RhCHMO to increase thermostability. The crystal structure of RhCHMO (PDB ID 4RG3) shown in cartoon representation and colored yellow, green and cyan for FAD, NADP and helical domain, respectively. FAD is shown in yellow and NADP+ as green sticks. Figures 5A-C) show the close up view of important residues, where the computationally predicted mutations are superimposed and shown as sticks with pink carbons. Alpha carbons are indicated by balls, hydrogen bonds as yellow dashed lines, and water molecules from the crystal structures as red balls. In Figure 5A, a transparent sphere indicates a cavity observed in the wild type, which is filled by the tyrosine mutation. Figure 5D shows the overall protein structure and all mutations experimentally shown to be stabilizing are shown as balls in shades of red indicating the degree of stabilization.

The fact that we found a thermostability plateau in the combinations of the first single mutant set let us to believe that we were missing out on stabilizing an early-unfolding region. If such a region triggers irreversible unfolding at a fixed temperature, this could mask a stabilization towards higher temperatures in another region. This hypothesis is also substantiated by the fact that the mutations of set 2 gave a more than additive effect. Possibly, these mutations prevented the early unfolding, thereby unmasking the less than additive effects observed for the previous mutants.
Conclusions

Cyclohexanone monooxygenase has been a frequent target of protein engineering, despite its complex catalytic mechanism: in order to overcome the challenge of coordinating oxygen, two cofactors, and the organic substrate, these enzymes employ protein domain and loop movements as well as cofactor rearrangement. This hampers many computational design approaches including the here applied FRESCO predictions, as molecular dynamics simulations are unable to cover the timescales of large protein motions. As FRESCO’s core algorithms are restricted to the proteinogenic component of the enzyme, the cofactor-dependency was an additional limitation and required to exclude ~19% of the enzyme’s residues from the calculations. Our subsequent screening confirmed the reliability of those predictions, however, with nearly half of the tested mutations exhibiting a stabilizing effect. An innovative shuffled library design method then aided in identifying the most stable mutant, which displayed a 13 °C higher $T_m$ accompanied by a vastly extended lifetime. Our best mutant shows a slightly higher $T_m$ than a recently discovered naturally thermotolerant CHMO while performing catalysis at a roughly 2.5x higher maximum rate. The stabilizing effects of the here-identified mutations were context-dependened and we found that the full potential of initially non-additive mutations can emerge after stabilizing other protein regions.

Materials and methods

General

All chemical reagents were purchased from Sigma-Aldrich or TCI Europe, unless otherwise stated. Oligonucleotide primers were synthetized by Sigma-Aldrich (Haverhill, UK) or Eurofins Scientific (Ebersberg, Germany). DNA sequencing was performed by GATC (Konstanz, Germany) and Eurofins Scientific.

Computational methods

The computational predictions by FRESCO were carried out by following the detailed protocol described in Wijma et al., 2018.

Structure preparation

Using the YASARA software, the RhCHMO structure from PDB ID 4RG3 was prepared by adding hydrogens and removing alternative side chain configurations, buffer and active site ligand molecules. After using a YASARA script to identify residues with less than 5 Å distance from the active site ligand, the FAD and NADP+ cofactor, the remaining residues were subjected to Rosetta-DDG and FoldX.

Single mutant predictions

Rosetta’s dgd_monomer version 2015.25.57927 was run using the “row 3 protocol” and these settings: dgd::weight_file soft_rep_design, dgd::iterations 50,
ddg::local_opt_only true, ddg::min_cst false, ddg::mean true, ddg::min false, ddg::sc_min_only false, ddg::ramp_repulsive false, ddg::opt_radius 8.0, ddg::dump_pdb false.

FoldX version 3.0b6 was used in the BuildModel mode with these settings: Temperature=298K, pH=7, IonStrength=0.05, moveNeighbours=true, VdWDesign=2, numberOfRuns=5.

**MD screening**

The $\Delta \Delta G^{\text{Fold}}$ energy predictions of both programs were sorted and all mutations with a predicted energy of $<-5$ kJ mol$^{-1}$ were subjected to MD simulations. For each mutant, five MD simulations with a random initial velocity seed were performed using a YASARA script employing Yamber3 as the force field. The system was warmed up from 0 to 298 K in 30 ps, followed by 20 ps equilibration, and a subsequent production run of 50 ps, where snapshots were sampled every 5 ps.

**Visual inspection**

All mutants were visually inspected using YASARA with a custom plugin. The average structure of each MD simulation was superimposed and the structure with the most notable differences eliminated, while the remaining four served as the basis for the flexibility alteration visual screening. For experimental screening only those were chosen, which showed no unfavorable hydrogen bonding alterations, no drastic hydrophobic surface exposure, and no increased local or overall flexibility. YASARA was also used to create the structural figures, except for Figure S7.10, which was created using the adaptive Poisson–Boltzmann Solver (APBS) web server with standard settings and visualizing the results using Chimera 1.10.2.

**Molecular biology methods**

The experimental procedure of screening targeted libraries were carried out by following the detailed protocol described in Fürst et al., 2018.

**QuikChange mutagenesis**

High throughput mutagenesis was conducted using a modified QuikChange protocol for which the primers were designed using AAscan. PCRbs were conducted in 96-well thin-walled plates (Bio-Rad) using the Pfu Ultra II Hotstart MasterMix (Agilent). As template served the native sequence of the *Rhodococcus* sp. Hi-31 gene (genbank accession number BAH56677) cloned into the previously described pCRE vector, a pBAD-NK derivative where the monooxygenase is N-terminally fused to a 6x-His-tagged, thermostable phosphite dehydrogenase. The PCR program was 95 °C – 3 min; (95 °C – 30 s; 55 °C – 30 s; 72 °C – 3 min) x 26; 72 °C – 12 min. Subsequently, 10 U DpnI were added to each well to digest the template plasmid. After at least 4 h reaction, 3.5 µl were transformed in chemically competent *E. coli* NEB10beta cells via 30 s heat shock at 42 °C. Plating was conducted on 24 well plates containing LB-agar with ampicillin. A single colony was picked and grown on a 96-well LB-agar plate and sent for plasmid extraction and sequencing to Eurofins Genomics.
**Randomly shuffled library creation**

The first library, which used the M4 mutant as a template, was based on the multichange isothermal mutagenesis (MISO) method, in which one Gibson-assembles overlapping PCR-fragments amplified using mutated primers. As seven mutations were to be randomized and two targeted the same residue (E91K and E91Q), our library size was $2^5 \times 3 = 96$ possible combinations. We designed overlapping primers according to MISO, but to randomize, we performed the PCR with an oligonucleotide mix, combining a wild-type sequence-containing primer with a mutation-containing primer, or two, in case of E91. We assembled the resulting fragment mix to full plasmids by Gibson cloning and analyzed 95 clones of the transformed library by DNA sequencing. The clones were mixed and each targeted position was either wild type or mutant, but in one case the ratio was strongly biased toward wild type (84%), and position 91 favored one of the mutations (60%) (Figure S7.4A). This bias can originate from the PCR with mixed primers, where template-binding may be altered by the mutation. Additionally, efficiency variations can occur in the assembly, where sequence-dependent secondary structures impair fragment annealing. Although we anticipatory attempted to mitigate these effects through silent mutations in the wild-type primers and the use of secondary structure prediction software (Geneious), this precaution proved to be only partly successful.

For the second library, the M2 mutant served as a template, and mutations D10Y and D119R were excluded. To create the resulting library of again 96 possibilities without a bias, we first created a plasmid with six mutations by MISO. Then, PCRs were performed with primers that generated fragments where the mutated site lies centrally. To randomize, we used a template mix combining the wild-type plasmid and the 6x mutated plasmid (plus the single mutant plasmid with the third mutation for the fragment with position 91). With both the primer binding site and the assembly overlap now occurring at areas remote from the mutation, we expected to avoid a bias. We sequenced 48 clones of the assembled and transformed library and performed a $\chi^2$ test on the data (Microsoft Office Excel), using an equal 1:1 distribution as expected value range. We also sequenced a culture of a mix of all colonies by inoculating a single 5 ml culture tube with all of the 96 clones. The mixed signals in the electropherogram were analyzed using Geneious.

**Biochemical methods**

**Expression and cell free extract preparation**

NEB10beta cells transformed with the pCRE-RhCHMO construct were grown overnight in 200 µL LBamp with vigorous shaking at 37 °C in two identical 96-well plate. 800 µL TBamp were added and the plates were shifted to 24 °C with vigorous shaking. After ~36 h, an OD$_{600}$ of ~20 was reached and the cells were harvested by centrifugation. Cells were resuspended in 50 mM Tris/HCl pH 7.5 buffer containing 10 mM MgCl$_2$, 0.5 mg/mL DNase I (NEB), and 1 mg/mL lysozyme (Sigma-Aldrich). After 30 min incubation at 20 °C, the cells were shock frozen in liquid N$_2$ and thawed in a water bath. Cell debris was removed by centrifugation at 4 °C and 2250 g for 45 min.
Protein purification
The cell free extract (CFE) of the two identical plates was combined and subjected to affinity chromatography using 100 µl Ni²⁺-Sepharose (GE Healthcare) in a 96-well filter plate. The Ni²⁺ resin was allowed to incubate with the CFE for 1 h, before removing the flow through by centrifugation. The resin was washed with 200 µl 50 mM Tris/HCl pH 7.5 and the same buffer with additionally 5 mM imidazole added. The pure protein was then eluted by adding 100 µl of 50 mM Tris/HCl pH 7.5 + 500 mM imidazole. Desalting was performed by using a PD MultiTrap G-25 96-well desalting plate (GE Healthcare) according to manufacturer’s instructions.

Thermostability screen
The purified protein’s $T_m$ was determined by the ThermoFAD method. 20 µl of the desalted protein solution was transferred to a qPCR plate (iQ, Bio-Rad) and subjected to a melt curve program in a qPCR machine (Touch Thermal Cycler, Bio-Rad), which ramped the temperature from 20 °C to 99 °C in 0.5 °C increments after 10 s incubation and a fluorescent read step. For excitation, a 450-490 nm filter was used and for emission, a 515-530 nm filter.

Enzyme kinetics
To determine the kinetic constants, reactions of RhCHMO with cyclohexanone and NADPH were followed spectrophotometrically (V-660, Jasco) at 340 nm ($\varepsilon_{340} = 6.22$ mM⁻¹ cm⁻¹). The 200 µl reaction mixture contained 50 mM Tris/HCl pH 7.5, 0.05 µM RhCHMO, 100 µM NADPH, and 0-500 µM substrate. The mix was quickly transferred to a 0.1 ml quartz cuvette after addition of NADPH and the absorbance at 340 nm was measured for 45 s. The kinetic constants were calculated by fitting the data to the Michaelis-Menten equation using the GraphPad Prism 6.

Temperature optimum
The temperature optimum was determined by following the same procedure as for the enzyme kinetics determination, but using always 100 µM of cyclohexanone and varying the temperature. 180 µL of enzyme-substrate mix were incubated at the specified temperature for 2 min, to allow the solution’s temperature to equilibrate. Addition of 20 µl NADPH then started the reaction, and the mix was transferred to a 0.1 mL quartz cuvette, which was placed in the preheated sample holder of the spectrophotometer.

Stability over time
The enzyme was incubated at the specified temperature, after which a sample was taken and the residual activity determined by following the same procedure as specified above.

Acknowledgements
The authors would like to gratefully acknowledge Dr. Hein Wijma, developer of the FRESCO method, who provided the scripts and gave advice on the computational procedures. The research for this work has received funding from the European Union (EU) project ROBOX (grant agreement n° 635734).
Engineering a Thermostable CHMO under EU’s Horizon 2020 Programme Research and Innovation actions H2020-LEIT BIO-2014-1.

References
17. van Beek, HL; Wijma, HJ; Fromont, L; Janssen, DB; Fraaije, MW. Stabilization of Cyclohexanone Monooxygenase by a Computationally Designed Disulfide Bond Spanning Only One Residue. FEBS Open Bio 2014 (4) 168-174.
Engineering a Thermostable CHMO

18 Schmidt, S; Genz, M; Balke, K; Bornscheuer, UT. The Effect of Disulfide Bond Introduction and Related Cys/Ser Mutations on the Stability of a Cyclohexanone Monooxygenase. J. Biotechnol. 2015 (214) 199-211.


20 Mirza, IA; Yachnin, BJ; Wang, S; Grosse, S; Bergeron, H; Imura, A; Iwaki, H; Hasegawa, Y; Lau, PC; Berghuis, AM. Crystal Structures of Cyclohexanone Monooxygenase Reveal Complex Domain Movements and a Sliding Cofactor. J. Am. Chem. Soc. 2009 (131) 8848-8854.


22 Yachnin, BJ; McEvoy, MB; MacCuish, RJ; Morley, KL; Lau, PC; Berghuis, AM. Lactone-Bound Structures of Cyclohexanone Monooxygenase Provide Insight into the Stereochemistry of Catalysis. ACS Chem. Biol. 2014 (9) 2843-2851.

23 Yachnin, BJ; Sprules, T; McEvoy, MB; Lau, PC; Berghuis, AM. The Substrate-Bound Crystal Structure of a Baeyer–Villiger Monooxygenase Exhibits a Criegee-Like Conformation. J. Am. Chem. Soc. 2012 (134) 7788-7795.


26 Floor, RJ; Wijma, HJ; Colpa, DI; Ramos-Silva, A; Jekel, PA; Szymanski, W; Feringa, BL; Marrink, SJ; Janssen, DB. Computational Library Design for Increasing Haloalkane Dehalogenase Stability. ChemBioChem 2014 (15) 1660-72.

27 Wu, B; Wijma, HJ; Song, L; Rozeboom, HT; Poloni, C; Tian, Y; Arif, MI; Nuijens, T; Quaedflieg, PJ; Szymanski, W. Versatile Peptide C-Terminal Functionalization Via a Computationally Engineered Peptide Amidase. ACS Catal. 2016 (6) 5405-5414.

28 Arabnejad, H; Dal Lago, M; Jekel, PA; Floor, RJ; Thunnissen, A-MWH; Terwisscha van Scheltinga, AC; Wijma, HJ; Janssen, DB. A Robust Cosolvent-Compatible Haloalkane Dehalogenase by Computational Library Design. Protein Engineering, Design and Selection 2017 (30) 175-189.

29 Martin, C; Ovalle Maqueo, A; Wijma, HJ; Fraaije, MW. Creating a More Robust 5-Hydroxymethylfurfural Oxidase by Combining Computational Predictions with a Novel Effective Library Design. Biotechnology for Biofuels 2018 (11) 56.


34 Forneris, F; Orru, R; Bonivento, D; Chiarelli, LR; Mattevi, A. Thermofad, a Thermofluor-Adapted Flavin ad hoc Detection System for Protein Folding and Ligand Binding. FEBS J. 2009 (276) 2833-2840.
Mitchell, LA; Cai, Y; Taylor, M; Noronha, AM; Chuang, J; Dai, L; Boeke, JD. Multichange Isothermal Mutagenesis: A New Strategy for Multiple Site-Directed Mutations in Plasmid DNA. *ACS Synth. Biol.* 2013 (2) 473-477.

Fürst, MJLJ; Romero, E; Gómez Castellanos, JR; Fraaije, MW; Mattevi, A. Side-Chain Pruning Has Limited Impact on Substrate Preference in a Promiscuous Enzyme. *ACS Catal.* 2018 (8) 11648-11656.


Jurrus, E; Engel, D; Star, K; Monson, K; Brandi, J; Felberg, LE; Brookes, DH; Wilson, L; Chen, J; Liles, K. Improvements to the Apbs Biomolecular Solvation Software Suite. *Protein Sci.* 2018 (27) 112-128.


Sun, D; Ostermaier, MK; Heydenreich, FM; Mayer, D; Jaussi, R; Standfuss, J; Veprintsev, DB; Aascan, Pcrdesign and Mutantchecker: A Suite of Programs for Primer Design and Sequence Analysis for High-Throughput Scanning Mutagenesis. *PLoS One* 2013 (8) e78878.

Torres Pazmiño, DE; Riebel, A; de Lange, J; Rudroff, F; Mihovilovic, MD; Fraaije, MW. Efficient Biooxidations Catalyzed by a New Generation of Self-Sufficient Baeyer-Villiger Monoxygenases. *ChemBioChem* 2009 (10) 2595-8.
Section 4

Understanding Substrate Selectivity and Product Specificity of BVMOs