scientific reports

OPEN

Protein engineering of the aldoxime dehydratase from *Bacillus* sp. OxB-1 based on a rational sequence alignment approach

Keiko Oike[1], Jens Sproß[1], Daisuke Matsui[2], Yasuhisa Asano[2✉] & Harald Gröger[1✉]

Recently, the program INTMSAlign_HiSol for identifying aggregation hotspots in proteins only requiring secondary structure data was introduced. We explored the utility of this program further and applied it for engineering of the aldoxime dehydratase from *Bacillus* sp. OxB-1. Towards this end, the effect of inverting the hydropathy at selected positions of the amino acid sequence on the enzymatic activity was studied leading to 60% of our constructed variants, which showed improved activity. In part, this activity increase can be rationalised by an improved heme incorporation of the variants. For example, a single mutation gave a 1.8 fold increased enzymatic activity and 30% improved absolute heme incorporation.

Aldoxime dehydratases (Oxds; E.C. 4.99) are special heme containing lyases involved in the "aldoxime–nitrile-pathway" of microbes (Fig. 1). These enzymes catalyse the dehydration of aldoximes to nitriles, which are metabolised to carboxylic acids by nitrile hydratases and amidases or by nitrilases[1–4].

The first discovered enzyme of this class is the Oxd from *Bacillus* sp. OxB-1 (OxdB), which was isolated from soil[1]. This enzyme naturally catalyses the dehydration of *E*- and *Z*-phenylacetaldoxime (*E*- and *Z*-PAOx) to phenylacetonitrile (PAN). It was purified as wild-type enzyme and expressed heterologously in *Escherichia coli* (*E. coli*)[5,6]. However, only a handful of these Oxds have been identified, isolated and characterised yet[7–13].

Oxds can be applied for the synthesis of nitriles[13–20] and dinitriles[21], compound classes with diverse applications in the chemical and pharmaceutical industry[22]. Common chemical syntheses of nitriles usually require harsh reaction conditions such as the use of the highly toxic hydrogen cyanide and salts thereof or high temperature and pressure[23,24]. Therefore, Oxds are very interesting biocatalysts for substituting these reactions by a biocatalytic process under mild conditions avoiding cyanide. Furthermore, we recently showed that the aldoxime dehydration is enantioselective[25] and that Oxds have a preference for one isomer of the aldoximes and can switch the enantioselectivity in dependency on the *E*- and *Z*-aldoxime as a substrate[26]. Chiral nitriles are of general interest as, e.g., such structural motif can be also found in drugs such as, for example, Saxagliptin[27] and Vildagliptin[28].

Although the enzymes show a high potential for synthetic purposes, only a few works on engineering these types of proteins are reported[29]. The recombinant Oxd from *Rhodococcus* sp. YH3-3 was targeted by directed evolution via random mutagenesis in order to improve the catalytic efficiency for the dehydration of 2-furfurylaldoxime to 2-furfuryl nitrile, a biorenewable feedstock-based nitrile with several applications in the chemical industry[29]. Other mutagenesis studies on aldoxime dehydratases so far were focused on the understanding of the substrate binding, catalytic activity and mechanism but not the improvement of catalytic properties[30–34]. The crystal structures of the Oxds from *Rhodococcus* sp. N-771 (OxdRE)[33] and *Pseudomonas chloraphis* B23 (OxdA)[34] are the only ones known up to now.

Even though the crystal structure is still unknown, OxdB is by far the most well characterised and investigated Oxd enzyme, in particular with respect to synthetic applications. However, engineering of this enzyme has not been carried out yet. Here we present a method for the improvement of the catalytic efficiency of Oxds

[1]Chair of Industrial Organic Chemistry and Biotechnology, Faculty of Chemistry, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany. [2]Biotechnology Research Center and Department of Biotechnology, Toyama Prefectural University, 5180 Kurokawa, Imizu, Toyama 939-0398, Japan. ✉email: asano@pu-toyama.ac.jp; harald.groeger@uni-bielefeld.de
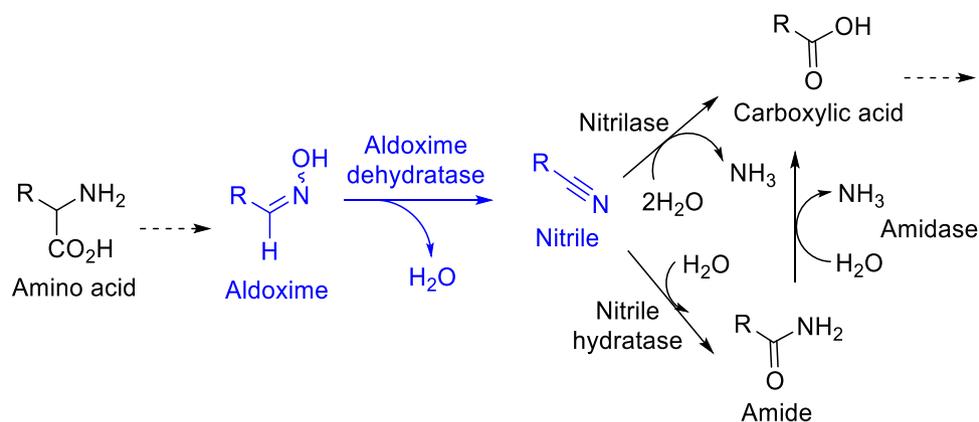
Figure 1. Scheme of the "aldoxime–nitrile-pathway" in microbes containing an Oxd-catalysed aldoxime dehydration step under formation of nitriles, which are further metabolised to carboxylic acids[2].

by site-directed mutagenesis without the necessity of structural knowledge of the enzyme. We used a rational method for the calculation of these mutation sites based on multi-sequence alignment under consideration of the amino acid hydropathy.

Results

Assay design. First, we decided on the method being used for the engineering of the target protein which is strongly related to the library size we are able to screen. While for the dehydration of 2-furfuryl aldoxime a large library could be screened by spectroscopy, this method is not suitable for arylaliphatic aldoximes. Thus, we decided to use an already established assay monitoring the product formation of the nitrile in reversed-phase HPLC measurements according to a method developed in our previous studies[25,26]. As our intention was the identification of variants with a generally increased activity independent from a specific substrate, we performed the screening with the native substrate of OxdB, Z-PAOx.

Identification of mutation sites. The program INTMSAlign was originally developed by the Asano group for the identification of novel hydroxynitrile lyases based on multi-sequence alignment[35]. In a later work, the program was modified by implementation of a tool for amino acid hydropathy calculation (INTMSAlign_HiSol)[36]. This additional tool is suitable to compare the amino acid hydropathy of each position in the target protein with the conserved hydropathy in the protein family. The difference of this hydropathy is defined as the so called HiSol score. A higher absolute value of the HiSol score indicates that opposite hydropathy is highly conserved in the selected position. This calculation method has been utilised for identification of aggregation hotspots in proteins. When the hydropathy of amino acid residues located in α-helices was inverted by exchange with highly conserved amino acids, several proteins could be solubilised for recombinant expression in *E. coli*[36]. We assumed that the hydropathy switch induces structural changes on the protein constitution, which can also influence the catalytic efficiency.

Toward this end, we decided to utilise the hydropathy calculation function implemented in INTMSAlign_HiSol in order to identify potential mutation sites in the sequence of OxdB. At the same time, we could limit the number of variants for our library while still finding positions with a significant difference compared to the aligned sequences. When the INTMSAlign_HiSol software was applied to the sequence of OxdB, twenty residues exhibited the highest absolute HiSol scores (Table 1). The amino acids, which are targets for the replacement by more hydrophilic residues, gave positive HiSol scores (Entries 1–10). The amino acids, which are targets for the replacement by more hydrophobic residues, gave negative HiSol scores (Entries 11–20). We did not further investigate the residues with lower absolute HiSol scores. Besides the HiSol score, the result sheet of INTMSAlign_HiSol displays the conservation rates of each amino acid in each position in the sequence of the target protein (Supplementary Information, Table S1). We chose the amino acid with the highest conservation rate for construction of variants bearing single-point mutations. The conservation of the identified residues in the gene family was spread out between low values of under 20% (Entries 8 and 11) and up to 90% (Entry 1). With this approach, we could construct our screening library by a rational approach which requires only the amino acid sequence of the protein. According to a secondary structure determination with the PSIPRED tool, the majority of the twenty identified residues with high absolute HiSol scores was located in strand (45%) or coil (30%) structures[37,38]. The other five identified residues (25%) were located in α-helices. A comparison of these mutation sites with a homology model structure of OxdB suggested that these twenty residues were not located in the active site of OxdB[32].

Generation of the mutant library and initial screening. The identified point mutations were introduced by site-directed mutagenesis with the corresponding primers (Supplementary Information, Table S2) and the vector pET22b harbouring the OxdB gene as template DNA, thus yielding OxdB-M1 to OxdB-M20. This vector was chosen because of its C-terminal hexahistidine-tag, which simplifies protein purification. In an ini-

Entry	Position	Original	Changed	Similarity [%]	HiSol score	Secondary structure
1	319	I	E	90.0	2.656	Strand
2	226	I	Q	26.5	2.411	Helix
3	220	L	E	39.7	2.281	Helix
4	149	V	E	65.6	2.278	Coil
5	105	I	R	32.6	2.252	Coil
6	267	L	R	40.0	2.226	Coil
7	316	I	Q	67.9	2.175	Strand
8	180	I	A	13.9	1.890	Coil
9	65	L	H	24.2	1.880	Strand
10	262	V	G	43.8	1.852	Strand
11	317	E	L	17.8	−1.452	Strand
12	190	K	R	34.5	−1.471	Strand
13	135	D	L	35.0	−1.496	Coil
14	67	Q	T	32.2	−1.533	Strand
15	49	K	V	24.2	−1.582	Helix
16	63	R	L	34.6	−1.882	Strand
17	296	H	L	56.7	−1.943	Coil
18	229	N	M	80.1	−1.974	Helix
19	250	Q	I	29.3	−2.153	Strand
20	283	K	L	57.8	−2.485	Helix

Table 1. Predicted mutation sites in the sequence of OxdB by the program INTMSAlign_HiSol[36]. The structure elements were predicted by PSIPRED[37,38].

tial screening, wild-type OxdB and the twenty variants were overexpressed in *E. coli* BL21(DE3) according to a protocol reported in literature[32]. We expressed the variants on small scale in shaking flasks. Here we wanted to make sure that the expression conditions were comparable to the actual conditions as oxygen saturation during expression of OxdB plays an important role[6]. It should be added that (in analogy to the wild-type enzyme OxdB) after expression in *E. coli* all of the 20 variants have been obtained in soluble form, which is illustrated by the SDS-PAGEs (see Supplementary information, Figure S2).

We then tested if the generated variants were active and compared them to the wild-type enzyme in an assay carried out with whole-cell catalysts (see Supplementary Information, Figure S1). This assay is closely related to our typical application of Oxds in catalysis. As a model substrate for this study, we chose the native substrate of OxdB, *Z*-PAOx, since it was our intention to achieve activity improvement for a substrate, which is already smoothly converted by OxdB, and no specific, "unusual" substrate with very low overall activity. In this initial screening, nearly all generated variants were active. We observed an improvement of the enzymatic activity for seven mutants. SDS-PAGE analysis (Supplementary information, Figure S2) showed that all variants were expressed in soluble form, we did not observed the formation of inclusion bodies for wild-type OxdB or any of its variants. The relative expression level compared to the wild-type enzyme was in overall slightly lower for several variants. At this stage, we did not further evaluate the whole-cell catalyst activity as we were more interested in the characterisation of the variants inter alia regarding their specific activity. In addition to the 20 variants with single point mutations, we constructed three further variants combining the most efficient variants of the initial whole-cell screening leading to promising single mutations, thus yielding Oxd-D1: L65H_I180A, OxdB-D2: I180A_K190R, and OxdB-D3: L65H_K190R.

Activity determination of the purified OxdB variants. The wild-type enzyme and all 23 constructed variants were purified according to a literature known protocol by affinity chromatography on a Talon affinity resin loaded column[32]. The purity of all mutants was considered as sufficient for the subsequent experiments (see Supplementary Information, Figures S3 to S5). Again, the specific enzymatic activity was determined by measuring the initial conversion rate of *Z*-PAOx to PAN. In their purified form, a majority of 60% of the variants showed an improved specific activity compared to the wild-type enzyme (Fig. 2). The highest specific activities observed were 24.7 U mg^{-1} and 24.0 U mg^{-1} for OxdB_I319E (M1) and OxdB_Q250I (M19), respectively. These
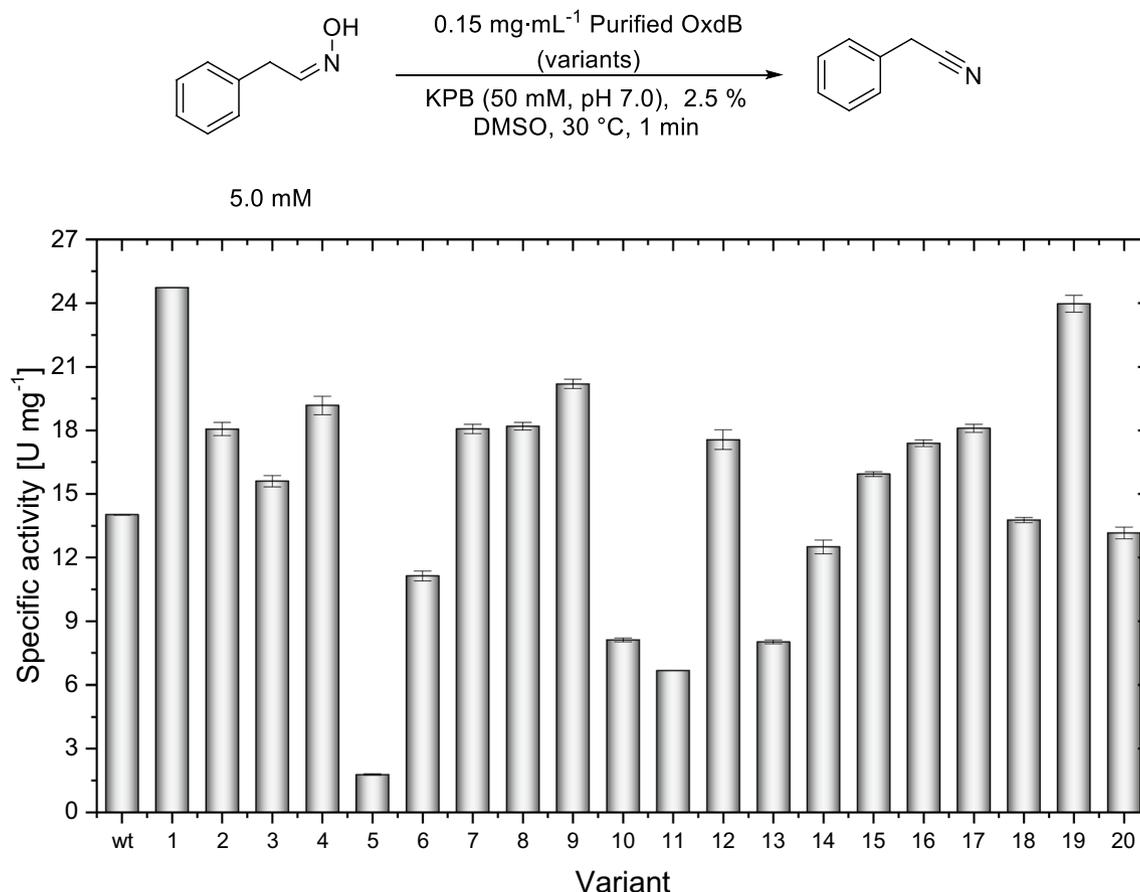
 3

Figure 2. Specific activities of wild-type OxdB and its variants (purified enzyme) for the standard substrate Z-PAOx. The numbers of the variants refer to the mutations displayed in Table 1.

Entry	Specific activity (U mg⁻¹)	Activity relative to wt (%)	Activity relative to parent 1 (%)	Activity relative to parent 2 (%)
1	26.4 ± 0.4	+ 189 ± 3	+ 31 ± 2	+ 51 ± 2
2	19.1 ± 0.2	+ 136 ± 1	+ 5.1 ± 1	+ 8.8 ± 1
3	28.4 ± 0.3	+ 203 ± 2	+ 41 ± 2	+ 62 ± 2

Table 2. The specific activities of the variants OxdB-D1 to OxdB-D3 and the relative activities compared to the wild-type protein and the corresponding parent variants.

variants were no hits in the whole-cell screening, despite of their improved catalytic efficiency, which can be explained by their lower expression level (Figure S2). Five variants (OxdB-I105R (M5), V262G (M10), E317L (M11), D135L (M13)) showed activities lower than 11 U mg⁻¹, corresponding to a relative activity below 79% of the wild-type enzyme. Interestingly, a different migration behaviour (slightly faster than the other proteins) was observed for OxdB-M13 in the SDS-PAGE analysis (see Supplementary Information, Figure S4). The specific activity of OxdB-M5 (I105R) was particularly low with 13% relative to the wild-type enzyme. In addition, the amount of produced protein was the lowest as well. Seven variants (OxdB_I226Q (M2), I316Q (M7), I80A (M8), K190R (M12), K49V (M15), R63L (M16), and H296L) exhibited a slightly improved activity between 17.4 and 18.2 U mg⁻¹, which corresponds to a relative increase of up to 30% compared to the wild-type enzyme. OxdB_V149E (M4) and OxdB_L65G (M9) showed an activity increase of 36 and 44%, respectively. The other variants (OxdB-L220E (M3), Q67T (M14), N229M (M18), and K283L (M20)) showed no significant deviation of the specific activity compared to the wild-type enzyme (± 10%). We decided not to investigate reaction kinetics, as the K_m-value for the substrate Z-PAOx is or can be expected to be in sub-millimolar range and, thus, a certain change of the K_m-value should not alter significantly the applicability in synthetic transformations which run at significantly higher substrate concentration[5]. Furthermore, the mutations are not located in the active site.

For the variants with two combined point mutations, activities outperforming their "parents" were observed (Table 2). The double mutants OxdB_L65H_I180A (D1) and OxdB-D3_L65H_K190R (D3) showed a specific activity of 26.4 U mg⁻¹ and 28.4 U mg⁻¹, respectively (Entries 1 and 3). This corresponds to a 1.9- and 2.0-fold increase compared to the wild-type. It is noteworthy that the variants M8, M9 and M12 turned out not to be
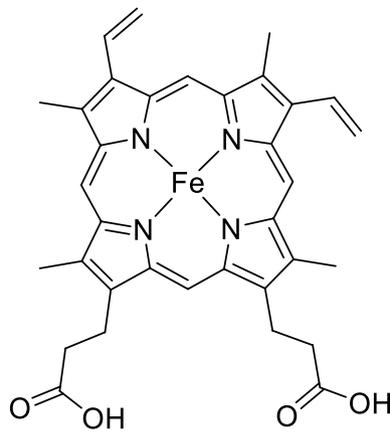
Figure 3. Structure of protoheme IX (heme b)[39].

the most active single variants, even though they showed promising properties in the whole-cell screening. In contrast to the most active OxdB-M1 and M19 variants, however, these three variants were expressed better. Nevertheless, the observed specific activities of OxdB-D1 and OxdB-D3 were exceeding the best single mutations OxdB-I319E and OxdB-Q250I (M1). The double mutant OxdB-D2 (I180A_K190R) also showed an activity improvement, however, only exceeding its parents' activities by less than 10% (Entry 2). The expression levels of the double variants were comparable to those of the parents. Their whole-cell activity was not tested.

Mass spectrometric analysis. Surprisingly, a majority of the OxdB variants showed a higher activity compared to the one of the wild-type enzyme. In order to find an explanation for these observed increased activities, we speculated about the influence of the heme content of the Oxd variants on the specific activity. Oxds are dependent on the cofactor protoheme IX (Fig. 3), which is bound via coordination to a proximal histidine residue to the protein backbone[33,34]. The incorporation of the cofactor is probably necessary for the protein folding, as OxdB expressed in the absence of iron is produced exclusively in insoluble form[5,6]. Oxds typically do not contain the cofactor in all protein molecules. The heme content varies for each Oxd and is, e.g., influenced by the expression system and the position of a histidine tag[9]. For instant, OxdB typically contains 33% heme when expressed in *E. coli* JM109 or 35% when expressed homologously in *Bacillus* sp. OxB-1 while OxdA contains 69% heme when being expressed in *E. coli*. However, a rationalisation of the quantitative cofactor incorporation in Oxds has not been done until now. An influence of our mutations on the heme content might be likely as our mutation point were chosen based on a sequential alignment and similarity considerations.

Thus, in order to gain more insight into the impact of mutations and sequence changes of Oxds on the quantitative heme content, we analysed the heme content of the wild-type OxdB and of selected variants by mass spectrometry. Toward this end, we conducted static nanoESI-MS measurements under denaturing conditions for the desired determination of the protein masses (Table 3). The following constructs were analysed: the wild-type enzyme, the most active variants (OxdB-M1, M2, M4, M8, M9, M12, M17, and M19) and OxdB-M13 because of its different gel electrophoresis behaviour (see Supplementary Information, Figure S4). All variants with two point mutations were investigated as well. Besides the charge state envelope of the protein ions, protoheme IX was identified with a high mass accuracy (m/z 616.1773, m/z_{theo} 616.1774, Δmass 0.1 mmu; deviation: 0.17 ppm) in measurements under denaturing conditions. This confirmed that the heme cofactor is bound non-covalently to the enzyme as previously reported[5,31]. Under native conditions, the static nanoESI-MS spectra contained a charge state envelope at significantly higher m/z values with a lower amount of charge states (Fig. 4). For most variants, each charge state consisted of two peaks with a typical mass shift of 615 ± 40 Da (after charge state clean up). This is in good agreement for two enzyme species, one with and the other without the cofactor protoheme IX.

The determined molecular masses of wild-type OxdB and its variants exhibit deviations compared to the theoretical molecular weight calculated from the sequence, which exceeded the usual error range of such an analysis. These differences can be explained by residual amounts of salt, which lead to broad signals in the mass spectrum, causing the inaccuracy of the molecular masses. It should be added that these effects are quite common in measurements under native conditions as ions of the buffer (here: ammonium acetate) form salt adducts with the protein. However, in the case of denaturing conditions protein clean-up might not be sufficient. Interestingly, the variants OxdB-M9 and OxdB-M19 yielded a lower protein mass compared to the theoretical mass derived from the sequence. An explanation for this finding is difficult without further data.

Nevertheless, we were pleased to find that by overlaying both spectra of the denatured and the native protein (or an excerpt of a few charge states) the protein species with bound heme can be easily differentiated from the protein species without heme. An example for this approach is visualised for the wild-type enzyme (Fig. 4). By comparing the peak intensities of the protein signals of apo- and holoprotein in the native mass spectra, a rough estimate about the relative heme content can be made (Fig. 4). For the wild-type enzyme, the ratio of the integrals of the peaks of apoprotein to holoprotein was estimated to be 30:70, thus referring to 70% holoprotein formed. For this analysis method, we assumed that the relative ionization efficiency of the different protein

Variant	M$_{sequence, theo}$ (Da)	M$_{Denatured, exp}$ (Da)	ΔMass to sequence (Da)	M$_{Native, exp}$ (Da)	ΔMass to denatured protein (Da)
Wild-type	41,216.78	41,255.8 ± 9.0	+ 39.02	41,901.8 ± 6.0	646.0
M1 (I319E)	41,232.73	41,248.21 ± 1.97	+ 15.45	41,863.76 ± 0.37	615.55
M2 (I226Q)	41,231.75	41,236.9 ± 4.8	+ 5.15	41,844.5 ± 2.6	607.6
M4 (V149E)	41,246.76	41,302.7 ± 6.4	+ 55.94	41,862.4 ± 2.1	559.7
M8 (I180A)	41,174.69	41,208.0 ± 6.8	+ 33.51	41,824.1 ± 6.7	615.9
M9 (L65H)	41,240.76	41,218.7 ± 4.8	− 22.06	41,829.4 ± 12.4	610.7
M12 (K190R)	41,244.79	41,291.3 ± 4.6	+ 46.51	41,863.8 ± 0.4	572.5
M13 (D135L)	41,214.85	41,216.4 ± 0.7	+ 1.55	41,830.60 ± 0.04	615.0
M17 (H296L)	41,192.79	41,229.0 ± 8.0	+ 36.21	41,791.2 ± 3.3	562.2
M19 (Q250I)	41,201.80	41,182.9 ± 7.8	− 18.90	41,801.6 ± 1.0	618.7
D1 (L65H_I180A)	41,198.68	41,215.9 ± 2.6	+ 17.22	41,829.9 ± 0.4	614.0
D2 (I180A_K190R)	41,202.71	41,217.82 ± 0.16	+ 15.09	41,824.02 ± 0.91	606.2
D3 (L65H_K190R)	41,268.77	41,298.9 ± 1.1	+ 30.13	41,889.8 ± 18.8	590.8

Table 3. Protein masses determined by static nanoESI-MS for wild-type OxdB and selected variants. The protein masses after charge state clean-up in the denatured state and the native state are given in Da. For the native state, two masses were determined, only the result with the higher molecular mass is displayed in this table. For the original data see the processed spectra displayed in Figures S7–S19.
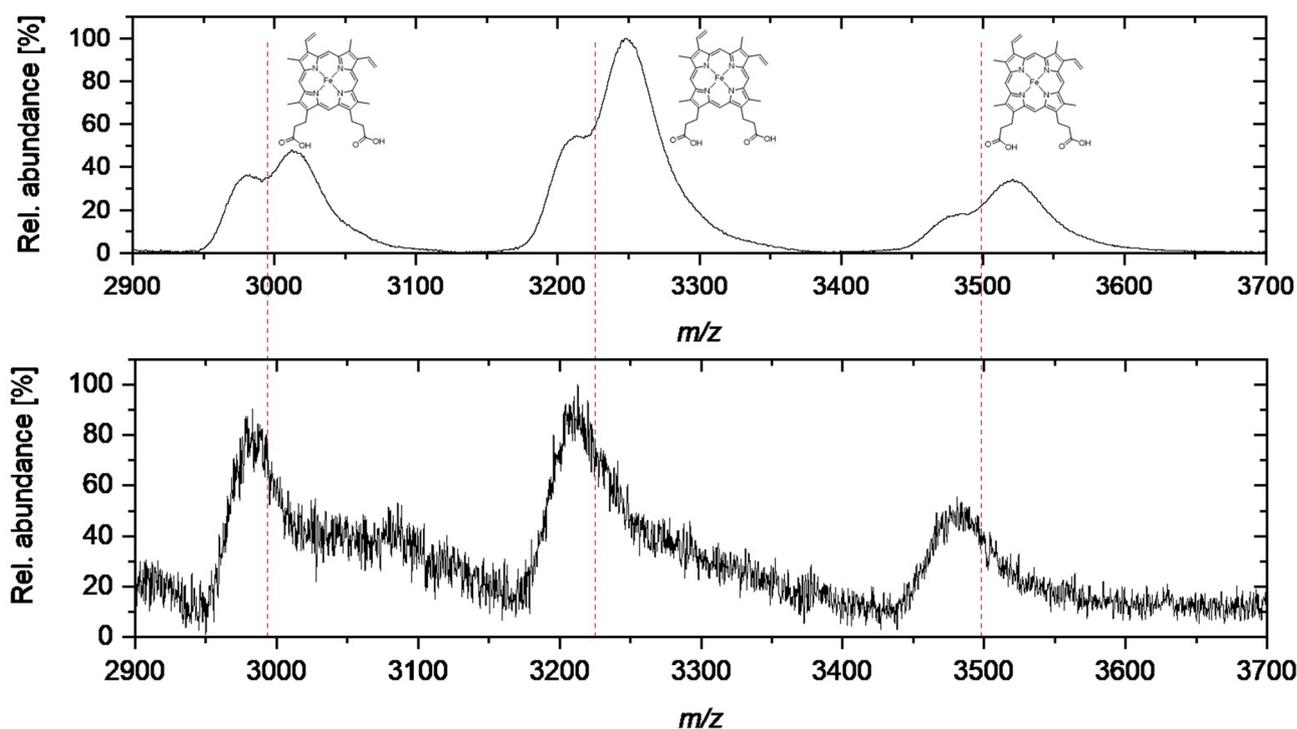

Figure 4. Overlay of excerpts of native (upper) and denatured (lower) wild-type oxdb mass spectra. The shown region refers to 2930 and 3520 m/z, in which the peaks refering to the charge states + 11 to + 13 are displayed.

species is not altered by the presence of the cofactor. For all variants investigated via native MS measurements, differences in the ratio of the relative peak integrals of a selected charge state were observed (e.g., Fig. 5). For all mutants with an improved specific activity, we hypothesised that the amount of heme incorporated during the expression had been increased.

Correlation between heme content and specific enzymatic activity. We plotted the specific enzymatic activity of the wild-type enzyme and all investigated mutants against the heme content in order to visualise the correlation between these values (Fig. 6). If an increased heme content is the only reason for a higher activity, one would obtain a linear increase of the specific activity with a higher heme content. For nearly all investigated
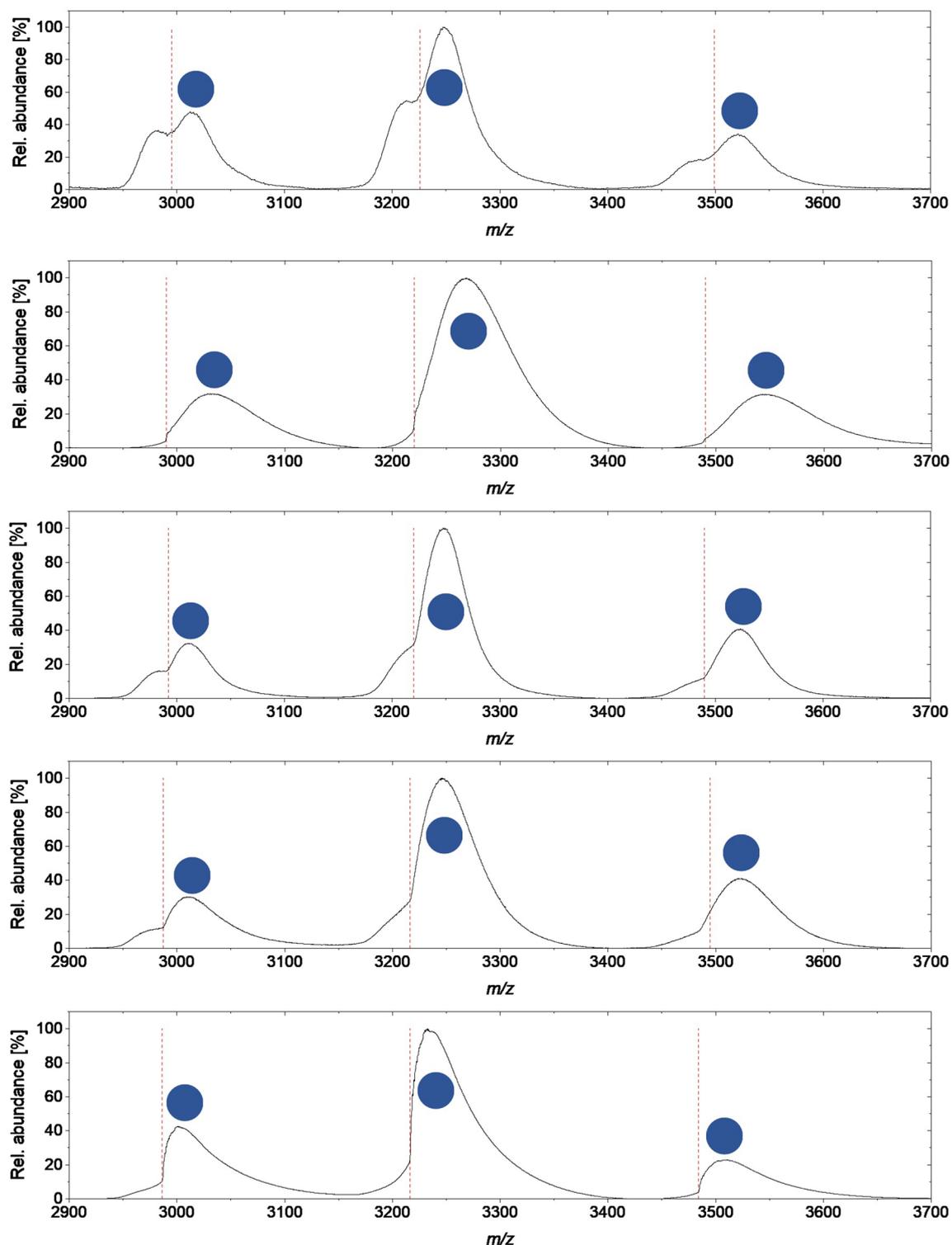
Figure 5. Excerpt of native mass spectra of wild-type OxdB and selected variants (from top to bottom: wild-type OxdB, OxdB_I319E (M1), OxdB_L65H (M9), OxdB_Q250I (M19), OxdB_L65H_I180A (D1). The shown region refers to 2930 and 3520 m/z, in which the peaks referring to the charge states + 11 to + 13 are displayed. The dashed red lines indicate the separation of the peaks of apoprotein and holoprotein. The blue dots indicate the peaks of the holoprotein.

variants, a higher heme content resulted in a higher specific activity. The variant OxdB-M13 is an exception to this finding. Despite of the higher heme content of 91%, the activity was notably lower compared to wild-type OxdB.

For example, the single mutation variants OxdB-M1, OxdB-M2, OxdB-M4, OxdB-M8, OxdB-M12, and OxdB-M13, and all double mutation variants exhibited heme contents above 90%; the apoprotein was barely present in this analysis. The observed heme incorporation for these variants refers to an increase of at least 20%
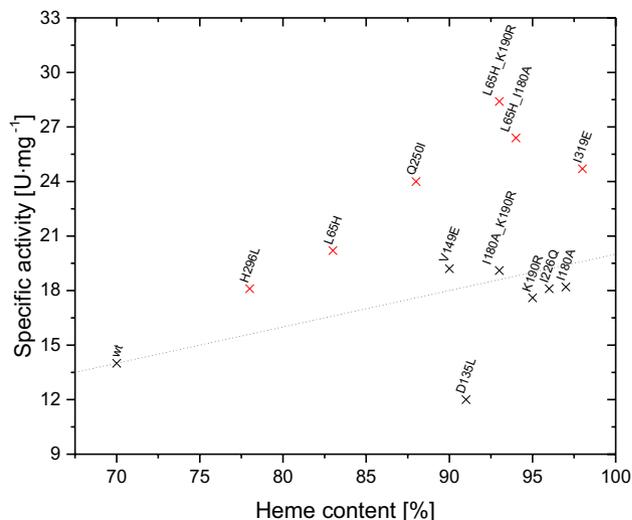
Figure 6. Correlation of the estimated heme content of selected OxdB variants and their specific activity. The dotted line refers to a linear dependence of the specific activity and the heme content compared to the wild-type enzyme.

compared to wild-type OxdB. The specific activities of OxdB-M2, OxdB-M4, OxdB-M8 and OxdB-M12 as well as the double mutation variant OxdD2 correspond nearly ideally to a linear correlation with the heme content. Four single mutant variants (OxdB-M1, OxdB-M9, OxdB-M17 and OxdB-M19) do not correspond to the ideal linear dependence. A higher activity than expected from the heme content was observed in these cases. While the heme content of OxdB-M9 was relatively low with 83%, this variant showed the third highest activity amongst the single mutant variants. This effect was even more significant for OxdB-M19 and OxdB-M1, which were the most active single mutant variants with heme contents of 88% and 98%, respectively. For these variants, effects apart from the increased heme incorporation probably lead to the higher activity. It is very notable that the heme content could be raised from 70% to nearly total incorporation by exchanging a single amino acid. The double mutants OxdB-D1 and OxdB-D3 showed even higher catalytic efficiency than all single mutation variants. Their heme content was approximately located in between those their parents. Both variants also showed a correlation between their heme content and specific activity exceeding the linear dependence.

Discussion

We used the algorithm of INTMSAlign_HiSol to identify mutation sites by multi-sequence alignment. The algorithm itself is a suitable tool to identify positions, in which the target protein has significantly different hydropathy and distribution compared to the consensus. By introducing highly conserved amino acid residues being available from the output file of INTMSAlign_HiSol, we constructed variants which exhibit inverted hydropathy. In addition, the high conservation of the introduced amino acids makes it more likely that the variants exhibit activity. This is in contrast to random mutagenesis-based approaches, for which the observation of inactive variants is common. The variants from our library exhibited two significant improvements compared to the wild-type protein: higher specific activity and higher cofactor incorporation. Naturally, there cannot be a direct correlation and, thus, predictability between an abstract value such as the HiSol score and higher catalytic efficiency. We assume that the likelihood of a positive influence on the protein structure is obtained by this algorithm. The method was suitable to generate small libraries by picking interesting positions without any structural knowledge about the protein apart from the amino acid sequence.

Paying respect to the deviations in the mass spectra, the heme content of the wild-type OxdB is compared to references where OxdB is expressed in *E. coli* JM109 harboring a pUC18 plasmid encoding the OxdB gene[6]. The wild-type OxdB as a reference for our mutagenesis showed a specific activity of 14.0 U mg^{-1} for Z-PAOx compared to a specific activity 8.33 U mg^{-1} reported in literature[6]. At the same time, the heme content was higher with 70% compared to 33% in literature, where the heme content was determined by precipitation with acetone/hydrochloric acid. The activity of 0.20 U mg^{-1} per % heme is lower compared to 0.25 U mg^{-1} per % heme reported in literature. This can be explained be the fusion of the hexahistidine-tag and the lower purity (single affinity chromatography vs several purification steps). The reason for the higher heme incorporation is probably the fusion of the tag, which is also reported to influence the heme content of OxdRE[9]. These considerations indicate that the MS measurements are a suitable method for estimation of the heme content and our assumption of a similar ionisation efficiency of the apoprotein and holoprotein is in good agreement with the obtained results.

With our rational approach of exchanging amino acids in residues with a hydropathy mismatch to highly conserved amino acids with the opposite hydropathy, we generated 12 mutants with an improved specific activity. The hit rate for an improved activity is therefore 60%. There are only a few strategies for improving the enzymatic activity by rational methods, especially if there is no knowledge about the structure. Typically, random mutagenesis is the method of choice for such studies[29]. However, also other multi-sequence alignments were suitable to

identify mutations improving the catalytic efficiency[40]. An advantage of our method is the small library size, which is especially helpful when high-throughput screening methods are not available. Thus, the results from our study show an additional utility of the program INTMSAlign_HiSol pointing out the strength of sequence alignment. Originally, INTMSAlign_HiSol was developed for solubilisation of proteins for expression in *E. coli* also under higher expression temperatures. Therefore, positions located in helices are chosen for the construction of variants. In contrast to our results, a notable improvement of the catalytic efficiency was not achieved[36]. We could partly rationalise the improved catalytic efficiency of our constructed variants by showing that the absolute heme content was increased. However, we could not find improved expression levels for the majority of our variants. At the same time a major difference in our study is that we chose the already soluble protein OxdB as a target for engineering instead of insoluble proteins. It is yet to be tested if our variants show an improved expression level at higher expression temperatures. The wild-type OxdB is usually forming insoluble aggregates at this temperature. The optimised pUC18 vector encoding the OxdB gene would be more suitable for expression of our variants when considering synthetic applications[6]. The improved heme incorporation certainly represents a beneficial factor for enhanced activity besides the expression level.

As visualised in Fig. 6, some variants exceeded the specific activity expected with respect to their heme content. The three most active variants with a single mutation (OxdB-M1, OxdB-M9 and OxdB-M19) had in common that they are located in β-strand structures. In the related OxdRE, most β-strand structures form a single β-barrel consisting of eight β-strands, which also contains the catalytically active amino acid residues[33]. From the sequence alignment (Figure S6) and secondary structure determination (Table 1), we hypothesised that OxdB forms this structure element in analogy to OxdRE. β-Barrels are structures formed by hydrogen bonds of typically at least eight antiparallel β-sheets connected by loops[41]. As our mutations inverted the hydropathy of the amino acids residues, thus changing the possibility for hydrogen bond formation, a deformation of the β-barrel structure is likely. Such a deformation could relocalise the catalytically active residues contained in the β-barrel structure to a position being advantageous for substrate binding and, thus, causing a higher catalytic efficiency.

It is noteworthy that by means of introducing a single mutation, the heme content could be increased from 70% up to 98%. This is especially surprising as no mutation was located in the active site or in positions where the interaction of the protein with the heme side chains is conserved, e.g. hydrogen bond formation (see the alignment of the sequences of OxdB, OxdRE and OxdA in the Supplementary Material, Figure S6)[33,34]. Thus, it appears that another factor has an impact on the heme incorporation. In addition, we suppose that our mutations induce structural changes in the protein, which improve the interaction with the cofactor, thus, facilitating its incorporation. Still, this is a new example for another utilisation of sequence alignment as a tool for rational protein design. For comparison, there are other possibilities for improving the heme incorporation during protein expression. The most common strategy for improving the heme incorporation is the addition of the heme precursor δ-amino levulinate to the culture medium[42], which also could be applied for Oxds. Other strain optimisation strategies use, e.g., the *E. coli* strain Nissle 1917, which has the natural ChuA heme receptor[43]. Alternatively, the strain *E. coli* RP523, which contains a porphobilinogen synthase gene disruption and a an uncharacterised permeability mutation that renders the bacteria heme-permeable, can be utilised as well[44]. All these methods, however, have in common that they require the addition of expensive heme (precursors) to the culture medium, which is a cost factor when bulk chemicals or fine chemicals should be synthesised by biocatalytic reactions. A different approach is the co-expression of a ferrochelatase, an enzyme catalysing the complexation of iron-ions into protoporphyrin IX[45]. In contrast, our method for improved heme incorporation is based on protein engineering and, thus, does not require expensive additives to the culture medium or the co-expression of another enzyme. This should be advantageous regarding the application of the enzyme. It is yet to be proved if this method also improves the heme incorporation of other Oxds and related hemoproteins or if this observation is limited to the OxdB enzyme.

Conclusion

In summary, we engineered OxdB by a rational multi-sequence alignment approach utilising the program INT-MSAlign_HiSol. It is noteworthy that 60% of our variants constructed by switching the hydropathy of amino acids showed an improved activity. For most variants, this activity correlated to an improved heme incorporation compared to the wild-type enzyme. With a single mutation, nearly full cofactor incorporation was achieved. A few variants amongst the variants with higher heme incorporation showed an improved activity not only related to the heme content. For these variants, we hypothesise that a structural change in a β-barrel structure might be related to the higher activity. We also expect that a structure elucidation of these variants will be facilitated in the future, when the crystal structure of OxdB will be solved. Such findings should then also enable a rationalisation of the beneficial impact of the individual mutations. A further task of future work is the application of the constructed variants for various synthetic purposes.

Methods

General information.   Protein concentrations were determined according to the method from *Bradford*[46] or using a Nanodrop One (Thermo Fisher Scientific, Massachusetts, USA). Gel electrophoresis was performed according to a protocol from *Laemmli*[47]. All chemicals used for biochemical experiments were of buffer/biochemical grade; solvents of HPLC-grade, PAOx, and PAN were self-synthesised. The protein was handled, if not stated differently, at 4 °C and stored at the same temperature. NMR-measurements were conducted on a Bruker Avance III 500HD. ^{1}H-NMR measurements were performed at 500 MHz and ^{13}C-NMR at 126 MHz.

Identification of the mutation sites.   The mutation sites were identified sequence of OxdB in accordance to Matsui and coworkers using the program INTMSAlign_HiSol applied to the amino acid sequence of OxdB[36].

The output file was used to calculate the HiSol score (as dimensionless value) by the following formula using the normalised hydropathy indices (I_{hyp}, z-transformed) according to Kyte and Doolittle[48]

$$\text{HiSol score (j)} = (I_{hyp})_{j,\,OxdB(i)} - \sum_{i=1}^{20} (I_{hyp})_i \left\{ (R_{cons})_{ij}/100 \right\}$$

The variable "i" represents the amino acid residues numbered in alphabetical order of their one-letter abbreviations; e.g. the i values of Ala, Cys and Asp are 1, 2, and 3, respectively. Thus, $(I_{hyp})_{j,OxdB(i)}$ refers to the hydropathy index of the amino acid residue "i" at the jth residue in the sequence of OxdB. $(R_{cons})_{ij}$ represents the appearance rate of that amino acid at the jth residue of OxdB in the library. We considered the amino acids with the 20 highest absolute HiSol scores as promising mutation sites.

The program INTMSAlign_HiSol reported in Ref.[36] is based on the software INTMSAlign, which is described in Ref.[35]. As outlined in this Ref.[35], the software INTMSAlign is available from the corresponding author Prof. Dr. Yasuhisa Asano upon request via E-mail, and furthermore this software is planned to become opened to be used by means of a web site for downloading the software.

Generation of the mutant library.   Site-directed mutagenesis of the plasmid pET22b-OxdB(C)6His was carried out with the QuikChange Lightning Site-directed Mutagenesis Kit (Stratagene, California, USA). A pET22b vector containing the OxdB gene fused to a C-terminal hexahistidine-tag (pET22b-OxdB(C)6His) was used as a dsDNA template in site-directed mutagenesis. The reaction mixture comprised of 0.5 μL of 10×reaction buffer, 0.2 μL of Quick Solution reagent, 0.2 μL of dNTP mix, 0.2 μL of QuikChange Lightning Enzyme, 125 ng of each the forward and reverse primers (see Supplementary Information, Table S2), 20–30 ng of the plasmid as template DNA, and sterilised water to a final volume of 5 μL. After initial denaturation at 95 °C for 2 min, 18 cycles were repeated as subsequently stated: A denaturing step at 95 °C for 20 s, an annealing step at 50 °C for 10 s, and an elongation step at 68 °C for 3.5 min. The product was treated with 0.2 μL of DpnI at 37 °C for 2 h, and then transformed to chemocompetent cells of *E. coli* BL21(DE3).

Protein expression and purification.   Expression and purification were performed according to a slightly modified literature known protocol[32]. A pre-culture of *E. coli* BL21(DE3) harbouring pET22b-OxdB(C)6His or its variants was inoculated from a LB-agar-plate and grown for 13 h in test-tubes containing 5 mL LB-medium (Carl Roth GmbH + Co. KG, Karlsruhe, Germany) and 50 μg/mL Carbenicillin (Carl Roth GmbH + Co. KG, Karlsruhe, Germany) at 37 °C at 160 rpm rotary shaking. Subsequently, a main culture of 350 mL Terrific-broth medium (Carl Roth GmbH + Co. KG, Karlsruhe, Germany) containing 50 μg/mL Carbenicillin in a 500 mL Erlenmeyer flask was inoculated with 1 mL of the pre-culture. The culture was incubated for approximately 7 h at 30 °C and 150 rpm rotary shaking until the OD_{600} reached approximately 0.8 and induced with 1 mM isopropyl-β-D-thiogalactopyranoside (Gerbu Biotechnik GmbH, Heidelberg, Germany). After induction, the incubation conditions were changed to 26 °C and 120 rpm rotary shaking. After 21 h, the cells were harvested by centrifugation (4000×g, 4 °C, 15 min) and washed twice with equilibration buffer (Tris–HCl buffer (20 mM, pH 8.0) containing 300 mM sodium chloride and 10 mM imidazole).

The cells from two cultures (~ 1.0 g) were resuspended with equilibration buffer (1:5 m,v) disrupted with sonotrode, and the cell debris and insoluble protein was removed via centrifugation (18,000×g, 4 °C, 60 min). The obtained cell-free extract was applied to a column containing TALON Metal Affinity Resin (Takara Biotechnology (DALIAN) Co., LTD., Dalian, PR China) (4.0 mL) equilibrated with wash buffer. After the column had been washed thoroughly (10 cv) with the same buffer, the enzyme was eluted with equilibration buffer (Tris–HCl buffer (20 mM, pH 8.0) containing 300 mM sodium chloride and 150 mM imidazole). The purification step was performed at < 20 °C. The combined active fractions were dialysed in Amicon Ultra centrifugal filters (Merck KGaA, Darmstadt, Germany) towards Tris–HCl buffer (20 mM, pH 7.0).

Activity assay.   The activity was determined in modified protocol according to literature known methods[32]. The assay solution contained 437.5 μL potassium phosphate buffer (50 mM, pH 7.0), 12.5 μL of *Z*-PAOx (200 mM solution in dimethyl sulfoxide, final concentration 5.0 mM), 50.0 μL of purified OxdB(C)6His or its variants (0.15 mg mL^{-1}; 3.75 μM), in a total volume of 500 μL. The reaction was carried out for 1 min at 30 °C with shaking at 900 rpm. Simultaneous addition of 400 μL of acetonitrile (ACN) and 100 μL of hydrochloric acid (0.1 M) stopped the reaction. After centrifugation (21,500×g, 10 min, 4 °C), the supernatant was transferred into HPLC vials and the conversion to PAN was measured by reversed phase-HPLC (Jasco, Nova Scotia, Canada) in comparison to a calibration curve. Measurements were conducted on a Nucleodur C18 HTec column (Macherey–Nagel, Düren, Germany) at 40 °C isocratic with water/ACN (70:30 v/v) as mobile phase and UV detection at 210 nm. Assays were conducted in triplicates for each variant.

Mass spectrometry.   Accurate nanoESI measurements of proteins were performed using a Q-IMS-TOF mass spectrometer Synapt G2Si (Waters GmbH, Manchester, UK) in resolution mode, interfaced to nano-ESI ion source. Nitrogen served as both the nebuliser gas and the dry gas for nanoESI. Nitrogen is generated by a nitrogen generator NGM 11. Samples (10 μM protein) were dissolved in 0.1 M ammonium acetate (pH 7.0) for native state measurements or 50% acetonitrile containing 0.1% formic acid for measurements under denaturing conditions. The samples were filled into *in-house* made ESI-emitters. The mass axis was externally calibrated with Agilent tune mix as calibration standard. Scan accumulation and data processing (baseline substraction, smoothing and centroidation) was performed with MassLynx 4.1 (Waters GmbH, Manchester, UK) on a PC

Workstation. The spectra shown here were generated by the accumulation and averaging of several single spectra. Determination of the accurate mass of protoheme IX and protein masses were performed using centroided data. The protein masses were determined using the software ESIprot from Winkler[49].

Substrate and reference compound synthesis. *Synthesis of Z-phenylacetaldoxime.* Hydroxylamine hydrochloride (10.3 g, 148 mmol) and sodium carbonate (17.3 g, 163 mmol) were dissolved in H_2O at room temperature. Phenylacetaldehyde (12.04 g, 100 mmol) was added and the reaction mixture was stirred vigorously for 18 h until complete conversion according to TLC analysis (cyclohexane/ethyl acetate) was achieved. The solution was extracted thrice with ethyl acetate (1:1 *v/v*), the combined organic phases were washed with brine (1:3 *v/v*) and dried over $MgSO_4$. The solvent was removed *in vacuo* gave a crude product, which was purified by recrystallisation (diethyl ether/di*iso*-propylether, − 80 °C) yielding Z-phenylacetaldoxime (5.60 g, 41.4 mmol, 41%) with > 99% isomerical purity (according to 1H-NMR) as colourless crystals.

1H-NMR (500 MHz, CD_2Cl_2): δ [ppm] = 8.88 (s, 1H, CH = NO*H*), 7.37–7.30 (m, 2H,Ph*H*), 7.29–7.22 (m, 3H, Ph*H*), 6.91 (t, ^3J = 5.4 Hz, 1H, C*H* = NOH), 3.74 (d, ^3J = 5.3 Hz, 2H, Ph-C*H$_2$*). 13C-NMR (500 MHz, CD_2Cl_2): δ [ppm] = 151.2, 129.1, 129.1, 129.0, 126.9, 31.9. RP-HPLC: *Macherey–Nagel* Nucleodur C_{18} HTec, water/ACN 70:30, 1.0 mL min^{-1}, 40 °C, 210 nm, R_{t1} = 7.54 min, R_{t2} = 8.82 min. The analytical data corresponds with the literature[50].

Synthesis of phenylacetonitrile. The synthesis was carried out in analogy to Ma et al.[51]. Copper(II) acetate (8.0 mg, 44 µmol) was dissolved in ACN at room temperature. Z-PAOx (51 mg, 380 µmol) was added and the solution was heated to reflux for 90 min. Complete conversion was determined via TLC (cyclohexane/ethyl acetate) and the solvent was removed in vacuo. The crude product was suspended in cyclohexane/ethyl acetate (2:1 *v/v*) and filtered over a short plug of silica. Removal of the solvent in vacuo yielded phenylacetonitrile (30 mg, 250 µmol, 66%) as a yellow oil.

1H-NMR (126 MHz, $CDCl_3$): 7.40–7.37 (m, 2H, Ar*H*), 7.35–7.33 (m, 3H, Ar*H*), 3.76 (s, 2H, N≡C–C*H$_2$*). 13C-NMR (126 MHz, $CDCl_3$): δ [ppm] = 130.1, 129.3, 128.2, 128.1, 118.0, 23.8. RP-HPLC: *Macherey–Nagel* Nucleodur C_{18} HTec, water/acetonitrile 70:30, 1.0 mL min^{-1}, 40 °C, 210 nm, R_t = 12.1 min. The analytical data corresponds with the literature[25].

References

1. Asano, Y. & Kato, Y. Z-phenylacetaldoxime degradation by a novel aldoxime dehydratase from *Bacillus* sp. strain OxB-1. *FEMS Microbiol. Lett.* **158**, 185–190 (1998).
2. Kato, Y., Ooi, R. & Asano, Y. Distribution of aldoxime dehydratase in microorganisms. *Appl. Environ. Microbiol.* **66**, 2290–2296 (2000).
3. Kato, Y., Yoshida, S. & Asano, Y. Polymerase chain reaction for identification of aldoxime dehydratase in aldoxime- or nitrile-degrading microorganisms. *FEMS Microbiol. Lett.* **246**, 243–249. https://doi.org/10.1016/j.femsle.2005.04.011 (2005).
4. Kato, Y., Ooi, R. & Asano, Y. Isolation and characterization of a bacterium possessing a novel aldoxime-dehydration activity and nitrile-degrading enzymes. *Arch. Microbiol.* **170**, 85–90 (1998).
5. Kato, Y., Nakamura, K., Sakiyama, H., Mayhew, S. G. & Asano, Y. Novel heme-containing lyase, phenylacetaldoxime dehydratase from *Bacillus* sp. strain OxB-1: purification, characterization, and molecular cloning of the gene. *Biochemistry* **39**, 800–809. https://doi.org/10.1021/bi991598u (2000).
6. Kato, Y. & Asano, Y. High-level expression of a novel FMN-dependent heme-containing lyase, phenylacetaldoxime dehydratase of *Bacillus* sp. strain OxB-1, in heterologous hosts. *Prof. Expr. Purif.* **28**, 131–139. https://doi.org/10.1016/S1046-5928(02)00638-1 (2003).
7. Xie, S.-X., Kato, Y., Komeda, H., Yoshida, S. & Asano, Y. A gene cluster responsible for alkylaldoxime metabolism coexisting with nitrile hydratase and amidase in *Rhodococcus globerulus* A-4. *Biochemistry* **42**, 12056–12066. https://doi.org/10.1021/bi035092u (2003).
8. Oinuma, K.-I. *et al.* Novel aldoxime dehydratase involved in carbon-nitrogen triple bond synthesis of *Pseudomonas chlororaphis* B23. Sequencing, gene expression, purification, and characterization. *J. Biol. Chem.* **278**, 29600–29608. https://doi.org/10.1074/jbc.M211832200 (2003).
9. Kato, Y., Yoshida, S., Xie, S.-X. & Asano, Y. Aldoxime dehydratase co-existing with nitrile hydratase and amidase in the iron-type nitrile hydratase-producer *Rhodococcus* sp. N-771. *J. Biosci. Bioeng.* **97**, 250–259. https://doi.org/10.1016/S1389-1723(04)70200-5 (2004).
10. Kato, Y. & Asano, Y. Purification and characterization of aldoxime dehydratase of the head blight fungus, *Fusarium graminearum*. *Biosci. Biotechnol. Biochem.* **69**, 2254–2257 (2005).
11. Kato, Y. & Asano, Y. Molecular and enzymatic analysis of the "aldoxime-nitrile pathway" in the glutaronitrile degrader *Pseudomonas* sp. K-9. *Appl. Microbiol. Biotechnol.* **70**, 92–101. https://doi.org/10.1007/s00253-005-0044-4 (2006).
12. Rädisch, R. *et al.* Overproduction and characterization of the first enzyme of a new aldoxime dehydratase family in *Bradyrhizobium* sp.. *Int. J. Biol. Macromol.* **115**, 746–753. https://doi.org/10.1016/j.ijbiomac.2018.04.103 (2018).
13. Choi, J. E. *et al.* Cyanide-free synthesis of an aromatic nitrile from a biorenewable-based aldoxime: Development and application of a recombinant aldoxime dehydratase as a biocatalyst. *Biocatal. Biotransfor.* **37**, 414–420. https://doi.org/10.1080/10242422.2019.1591376 (2019).
14. Kato, Y., Ooi, R. & Asano, Y. A new enzymatic method of nitrile synthesis by *Rhodococcus* sp. strain YH3–3. *J. Mol. Catal. B Enzym.* **6**, 249–256 (1999).
15. Xie, S.-X., Kato, Y. & Asano, Y. High yield synthesis of nitriles by a new enzyme, phenylacetaldoxime dehydratase, from *Bacillus* sp. Strain OxB-1. *Biosci. Biotechnol. Biochem.* **65**, 2666–2672 (2001).
16. Miki, Y. & Asano, Y. Biosynthetic pathway for the cyanide-free production of phenylacetonitrile in *Escherichia coli* by utilizing plant cytochrome P450 79A2 and bacterial aldoxime dehydratase. *Appl. Environ. Microbiol.* **80**, 6828–6836. https://doi.org/10.1128/AEM.01623-14 (2014).

17. Plass, C. *et al.* Approaching bulk chemical nitriles from alkenes: a hydrogen cyanide-free approach through a combination of hydroformylation and biocatalysis. *ACS Catal.* **9**, 5198–5203. https://doi.org/10.1021/acscatal.8b05062 (2019).

18. Hinzmann, A., Glinski, S., Worm, M. & Gröger, H. Enzymatic synthesis of aliphatic nitriles at a substrate loading of up to 1.4 kg/L: a biocatalytic record achieved with a heme protein. *J. Org. Chem.* **84**, 4867–4872. https://doi.org/10.1021/acs.joc.9b00184 (2019).

19. Hinzmann, A., Adebar, N., Betke, T., Leppin, M. & Gröger, H. Biotransformations in pure organic medium: organic solvent-labile enzymes in the batch and flow synthesis of nitriles. *Eur. J. Org. Chem.* **2019**, 6911–6916. https://doi.org/10.1002/ejoc.201901168 (2019).

20. Betke, T. *et al.* Biocatalytic synthesis of nitriles through dehydration of aldoximes: the substrate scope of aldoxime dehydratases. *ChemBioChem* **19**, 768–779. https://doi.org/10.1002/cbic.201700571 (2018).

21. Betke, T., Maier, M., Gruber-Wölfler, H. & Gröger, H. Biocatalytic production of adiponitrile and related aliphatic linear α, ω-dinitriles. *Nat. Commun.* **9**, 5112. https://doi.org/10.1038/s41467-018-07434-0 (2018).

22. Subramanian, L. R. In *Science of Synthesis* (ed. Murahashi, S.-I.) (Georg Thieme Verlag, 2004).

23. Gröger, H. Catalytic enantioselective Strecker reactions and analogous syntheses. *Chem. Rev.* **103**, 2795–2828. https://doi.org/10.1021/cr020038p (2003).

24. Murahashi, S.-I. (ed.) *Science of Synthesis* (Georg Thieme Verlag, 2004).

25. Metzner, R., Okazaki, S., Asano, Y. & Gröger, H. Cyanide-free enantioselective synthesis of nitriles: synthetic proof of a biocatalytic concept and mechanistic insights. *ChemCatChem* **6**, 3105–3109. https://doi.org/10.1002/cctc.201402612 (2014).

26. Betke, T., Rommelmann, P., Oike, K., Asano, Y. & Gröger, H. Cyanide-free and broadly applicable enantioselective synthetic platform for chiral nitriles through a biocatalytic approach. *Angew. Chem. Int. Ed.* **56**, 12361–12366. https://doi.org/10.1002/anie.201702952 (2017).

27. Savage, S. A. *et al.* Preparation of saxagliptin, a novel DPP-IV inhibitor. *Org. Process Res. Dev.* **13**, 1169–1176. https://doi.org/10.1021/op900226j (2009).

28. Pellegatti, L. & Sedelmeier, J. Synthesis of vildagliptin utilizing continuous flow and batch technologies. *Org. Process Res. Dev.* **19**, 551–554. https://doi.org/10.1021/acs.oprd.5b00058 (2015).

29. Choi, J. E., Shinoda, S., Asano, Y. & Gröger, H. Aldoxime dehydratase mutants as improved biocatalysts for a sustainable synthesis of biorenewables-based 2-furonitrile. *Catalysts* **10**, 362. https://doi.org/10.3390/catal10040362 (2020).

30. Kobayashi, K., Yoshioka, S., Kato, Y., Asano, Y. & Aono, S. Regulation of aldoxime dehydratase activity by redox-dependent change in the coordination structure of the aldoxime-heme complex. *J. Biol. Chem.* **280**, 5486–5490. https://doi.org/10.1074/jbc.M410474200 (2005).

31. Konishi, K. *et al.* Identification of crucial histidines involved in carbon-nitrogen triple bond synthesis by aldoxime dehydratase. *J. Biol. Chem.* **279**, 47619–47625. https://doi.org/10.1074/jbc.M407223200 (2004).

32. Miao, Y., Metzner, R. & Asano, Y. Kemp elimination catalyzed by naturally occurring aldoxime dehydratases. *ChemBioChem* **18**, 451–454. https://doi.org/10.1002/cbic.201600596 (2017).

33. Sawai, H. *et al.* X-ray crystal structure of michaelis complex of aldoxime dehydratase. *J. Biol. Chem.* **284**, 32089–32096. https://doi.org/10.1074/jbc.M109.018762 (2009).

34. Nomura, J. *et al.* Crystal structure of aldoxime dehydratase and its catalytic mechanism involved in carbon-nitrogen triple-bond synthesis. *Proc. Natl. Acad. Sci. USA* **110**, 2810–2815. https://doi.org/10.1073/pnas.1200338110 (2013).

35. Nakano, S. & Asano, Y. Protein evolution analysis of S-hydroxynitrile lyase by complete sequence design utilizing the INTMSAlign software. *Sci. Rep.* **5**, 8193. https://doi.org/10.1038/srep08193 (2015).

36. Matsui, D., Nakano, S., Dadashipour, M. & Asano, Y. Rational identification of aggregation hotspots based on secondary structure and amino acid hydrophobicity. *Sci. Rep.* **7**, 9558. https://doi.org/10.1038/s41598-017-09749-2 (2017).

37. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

38. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, 402–407. https://doi.org/10.1093/nar/gkz297 (2019).

39. Zhang, J., Kang, Z., Chen, J. & Du, G. Optimization of the heme biosynthesis pathway for the production of 5-aminolevulinic acid in *Escherichia coli*. *Sci. Rep.* **5**, 8584. https://doi.org/10.1038/srep08584 (2015).

40. Xiao, Z. *et al.* Improvement of the thermostability and activity of a pectate lyase by single amino acid substitutions, using a strategy based on melting-temperature-guided sequence alignment. *Appl. Environ. Microbiol.* **74**, 1183–1189. https://doi.org/10.1128/AEM.02220-07 (2008).

41. Wimley, W. C. The versatile β-barrel membrane protein. *Curr. Opin. Struc. Biol.* **13**, 404–411. https://doi.org/10.1016/S0959-440X(03)00099-X (2003).

42. Kery, V., Elleder, D. & Kraus, J. P. δ-Aminolevulinate increases heme saturation and yield of human β-cystathionine synthase expressed in *Escherichia coli*. *Arch. Biochem. Biophys.* **316**, 24–29 (1995).

43. Fiege, K., Querebillo, C. J., Hildebrandt, P. & Frankenberg-Dinkel, N. Improved method for the incorporation of heme cofactors into recombinant proteins using *Escherichia coli* Nissle 1917. *Biochemistry* **57**, 2747–2755. https://doi.org/10.1021/acs.biochem.8b00242 (2018).

44. Woodward, J. J., Martin, N. I. & Marletta, M. A. An Escherichia coli expression-based method for heme substitution. *Nat. Methods* **4**, 43–45. https://doi.org/10.1038/nmeth984 (2007).

45. Sudhamsu, J. *et al.* Co-expression of ferrochelatase allows for complete heme incorporation into recombinant proteins produced in *E. coli*. *Prof. Expr. Purif.* **73**, 78–82. https://doi.org/10.1016/j.pep.2010.03.010 (2010).

46. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

47. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1979).

48. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

49. Winkler, R. esiprot: a universal tool for charge state determination and molecular weight calculation of proteins from electrospray ionization mass spectrometry data. *Rapid Commun. Mass Spectrom.* **24**, 285–294. https://doi.org/10.1002/rcm.4384 (2010).

50. Ferreira-Silva, B., Lavandera, I., Kern, A., Faber, K. & Kroutil, W. Chemo-promiscuity of alcohol dehydrogenases: reduction of phenylacetaldoxime to the alcohol. *Tetrahedron* **66**, 3410–3414. https://doi.org/10.1016/j.tet.2010.03.050 (2010).

51. Ma, X.-Y., He, Y., Lu, T.-T. & Lu, M. Conversion of aldoximes into nitriles catalyzed by simple transition metal salt of the fourth period in acetonitrile. *Tetrahedron* **69**, 2560–2564. https://doi.org/10.1016/j.tet.2013.01.059 (2013).

Acknowledgements

Author contributions

D.M. conducted the calculations, K.O. and J.S. conducted the experiments. K.O. and J.S. analysed the results. H.G. and Y.A. managed the research. K.O., J.S. and H.G. wrote and revised the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-92749-0.

Correspondence and requests for materials should be addressed to Y.A. or H.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.