A family of native amine dehydrogenases for the asymmetric reductive amination of ketones

Ombeline Mayol, Karine Bastard, Lilian Beloti, Amina Frese, Johan P Turkenburg, Jean-Louis Petit, Aline Mariage, Adrien Debard, Virginie Pellouin, Alain Perret, et al.

A family of Native Amine Dehydrogenases for the Asymmetric Reductive Amination of Ketones

Ombeline Mayol[1], Karine Bastard[1], Lilian Beloti[2], Amina Frese[2], Johan P. Turkenburg[2], Jean-Louis Petit[1], Aline Mariage[1], Adrien Debard[1], Virginie Pellouin[1], Alain Perret[1], Véronique de Berardinis[1], Anne Zaparucha[1], Gideon Grogan*[2], Carine Vergne-Vaxelaire*[1]

[1] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France
[2] York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, YO10 5DD, UK.

The asymmetric reductive amination of ketones enables the one-step synthesis of chiral amines from readily available starting materials. Here we report the discovery of a family of native NAD(P)H-dependent Amine Dehydrogenases (nat-AmDHs) competent for the asymmetric reductive amination of aliphatic and alicyclic ketones, adding significantly to the biocatalytic toolbox available for chiral amine synthesis. Studies of ketone and amine substrate specificity and kinetics reveal a strong preference for aliphatic ketones and aldehydes, with activities of up to 614.5 mU mg^{-1} for cyclohexanone with ammonia and 851.3 mU mg^{-1} for isobutyraldehyde with methylamine as amine donor. Crystal structures of three nat-AmDHs (AmDH4, *Msme*AmDH and *Cfus*AmDH) reveal the active site determinants of substrate and cofactor specificity and enable the rational engineering of AmDH4 for generated activity towards pentan-2-one. Analysis of the 3D-catalytic site distribution among bacterial biodiversity revealed a superfamily of divergent proteins with representative specificities ranging from amino acid substrates to hydrophobic ketones.

The asymmetric reductive amination of ketones is one of the most significant reactions in the industrial pharmaceutical synthesis of chiral amines. Conventional catalytic methods usually require expensive transition metal complexes, which are difficult to completely remove, or chiral ligands, rendering industrial processes ultimately non-sustainable. Given the urgent demand for alternative catalysts,[1] enzymatic methods of amine synthesis have been extensively investigated in recent years.[2-7] The asymmetric reductive amination of ketones can be accomplished using engineered Amine Dehydrogenases (AmDHs), firstly developed from native amino-acid dehydrogenases,[8-15] and imine reductases (IREDs), which catalyze the reduction of preformed secondary imines, but have been shown to enable the reductive amination of ketones when supplied with large excesses of amine precursor.[16] In addition, Codexis have engineered opine dehydrogenases (OpDHs), which natively couple α-amino acids with α-keto acids, to broaden their substrate range and convert non-functionalized ketones to chiral amines with ammonia.[17] Native enzymes catalyzing this activity have been reported first from *Streptomyces virginiae* by Itoh and coworkers, but with low enantioselectivity[18] and later by Wang *et al.* in crude extracts of *Pseudomonas kilonensis and P. balearica,*[19] but the genes encoding the activities were not identified. More recently, a subset of IREDs has also been shown to catalyze the reductive amination of selected ketones with amine partners provided in an equimolar ratio,[20-21] but this was primarily directed toward synthesis of secondary amines, as the activity with ammonia as a donor was low. These reductive aminases (RedAms) catalyze both the formation of the imine intermediate and its subsequent reduction.

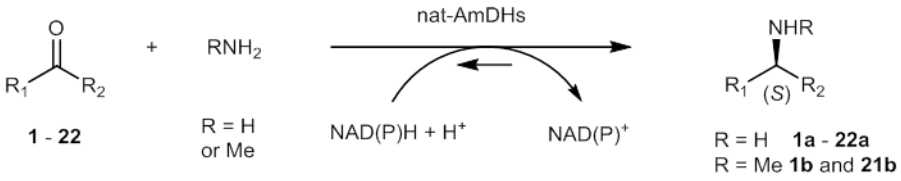
We have recently used as a first reference set the sequence of a (2*R*, 4*S*)-2,4-diaminopentanoate dehydrogenase from *Clostridium sticklandii* [22] to identify among bacterial biodiversity the first native AmDHs (4OP-AmDH) catalyzing the amination of 4-oxopentanoic acid (4OP) with ammonia, in addition to their metabolic product (2*R*)-2-amino-4-oxopentanoic acid (2A4OP).[23] These enzymes were inactive towards ketones such as acetophenone, cyclohexanone or pentan-2-one, suggesting that a carboxylic acid was a prerequisite for activity.

We have now used the 4OP-AmDH sequences in turn to further screen the available sequences within bacterial biodiversity in order to discover additional enzymes with native activity towards ketones and aldehydes without a carboxylic acid group. We report here the discovery of five native AmDHs (nat-AmDHs) with significant activities toward such substrates, together with their application to the semi-preparative scale synthesis of four amines. The 3D-structure determination of three of them provides insight into the catalytic mechanism of reductive amination by this characterized family of AmDHs, and expands the biocatalytic toolbox of available enzymes for asymmetric reductive amination reactions.

Results

Screening of bacterial biodiversity

As 4OP-AmDHs do not share significant sequence homology with the engineered AmDHs and RedAms (Supplementary Figure 11), they were used to search for distant homologs with activity toward non-functionalized ketones, using a sequence-driven approach.[24] 23 candidate enzymes were screened as cell-free extracts against two known substrates of 4OP-AmDHs [(2*R*,4*S*)-2,4-diaminopentanoate (2,4-DAP) and 4-aminopentanoate (4-AP)], and also three non-functionalized amines (pentan-2-amine, α-methylbenzylamine and cyclohexylamine) (Supplementary Methods and Supplementary Table 1). One enzyme from *Mycobacterium smegmatis* (*Msme*AmDH, Uniprot ID: A0A0D6I8P6) displayed an activity of 6 mU mg^{-1} against cyclohexylamine, although no activity was detected against 2,4-DAP and 4-AP. This suggested that this target has a biological role far from those of diaminopentanoate dehydrogenases constituting our first reference set.[23] No activity was detected against α-methylbenzylamine. A 100 µL scale reaction on pentan-2-one and cyclohexanone using a cofactor regeneration system and purified enzyme confirmed the presence of the expected amine). Further homologs with amino-acid sequence identity ranging from >70 % to 35 % were produced as purified enzymes and screened against the same substrates (Supplementary Methods and Supplementary Figure 1). All those with identity above 38 % displayed AmDH activity and were named nat-AmDH (Fig. 1).



Name	Uniprot identifier	Organism	Uniprot reported annotation	% identity vs *Msme*AmDH
*Msme*AmDH	A0A0D6I8P6	*Mycobacterium smegmatis*	Dihydrodipicolinate reductase	100
*Mvac*AmDH	K0UKT5	*Mycobacterium vaccae*	Dihydrodipicolinate reductase	89.9
*Myco*AmDH	A0A101AWU7	*Mycobacterium sp.*	Dihydrodipicolinate reductase	75.6
*Micro*AmDH	C3UMY1	*Microbacterium sp.*	Dihydrodipicolinate reductase	66.6
*Apau*AmDH	E3CZE3	*Aminomonas paucivorans*	Dihydrodipicolinate reductase	38.9
*Cfus*AmDH	S9Q235	*Cystobacter fuscus*	Uncharacterized protein	38.2

Fig. 1| Amine Dehydrogenases identified in this study. Sequence identity of *Msme*AmDH homologs are presented relatively to *Mmse*AmDH. A similarity matrix is available in Supplementary Figure 11.

Carbonyl acceptor and amine donor spectra

Four purified nat-AmDHs (*Msme*AmDH, *Cfus*AmDH, *Micro*AmDH and *Apau*AmDH) (Supplementary Methods and Supplementary Figure 2) were screened against a range of carbonyl substrates (**1** – **22**) with ammonia as the amine donor and NADH or NADPH as reducing agents. AmDH activity was assayed in the reductive amination direction by monitoring the spectrophotometric decrease in [NAD(P)H] (Table 1A). *Msme*AmDH, *Cfus*AmDH and *Micro*AmDH showed a similar carbonyl acceptor spectrum, with a preference for cycloalkanones and aliphatic aldehydes, the latter being the only type of substrates transformed by *Apau*AmDH. Interestingly, branched acyclic substrates seemed to be superior substrates to linear ones. Little or no activity was detected against aromatic ketones such as acetophenone, 4-phenylbutan-2-one, phenoxypropan-2-one or 4-methylphenylacetone, which were the preferred substrates of engineered AmDHs, themselves with little or no activity against cycloalkanones.[10,13-14,25] *Msme*AmDH, *Cfus*AmDH and *Micro*AmDH accepted both NADH and NADPH, but the highest specific activities for each cofactor were substrate-dependent. *Apau*AmDH was specific for NADH. Importantly, no conversion to alcohol product could be detected for any of the tested enzymes with carbonyl substrates in the absence of ammonia.

These purified nat-AmDHs were then tested with a limited number of amine donors (methylamine, ethylamine, benzylamine, cyclopentylamine and 3-pentanamine) with cyclohexanone **1** and isobutyraldehyde **21** as carbonyl acceptors. 100 equivalents of amine were used to maximize the chance of identifying hits. Only methylamine was accepted as an alternative amine donor, particularly for *Micro*AmDH, which showed even better specific activity with methylamine than ammonia (Table 1B). The activity with methylamine and **1** (10.1 - 90.4 mU mg^{-1}) distinguishes these enzymes from engineered AmDHs, which typically displayed strict specificity for ammonia as the amine nucleophile, but was not as high as that recorded for *Asp*RedAm towards the same substrates.[20] A pH study revealed that the specific activity of *Micro*AmDH was only slightly dependent on pH, with 70 % of the activity maintained between pH 6.7 and 12.2. The specific activity was also of the same order of magnitude, with either 50 or 2 equivalents of methylamine (Supplementary

Methods and Supplementary Figure 3). These results suggest that the activity was independent of the amount of preformed imine in solution,[26] thus reflective of the catalytic formation of the imine by *Micro*AmDH, in a mode similar to that of RedAms.

Table 1. Carbonyl/amine substrate spectrum of nat-AmDHs.

Biochemical characterization

According to gel filtration experiments, *Msme*AmDH, *Cfus*AmDH and *Micro*AmDH are all homodimers (Supplementary Methods). The kinetic parameters for *Msme*AmDH and *Cfus*AmDH (Table 2) indicated that both enzymes are slightly more efficient with NADPH than NADH, as estimated by the ratios k_{cat}/K_m NADPH and k_{cat}/K_m NADH (2.5 and 2.7 times, respectively). This promiscuous behavior appears to be different from that of Phe-AmDH or Leu-AmDH on one hand[8-9] and RedAms on the other hand,[20-21] which are specific for NADH and NADPH, respectively. The K_m values obtained for ammonia were similar to what was described for engineered AmDH.[13] The enzymes were subject to inhibition by their carbonyl substrates. *Msme*AmDH displayed higher catalytic efficiency toward **1** compared to **2** and **21**, while *Cfus*AmDH was more efficient with **21**.

Table 2. Kinetic data of *Msme*AmDH and *Cfus*AmDH

Stereoselectivity

Amines were obtained with e.e.s of between 68 % to ≥ 98.5 %, depending on the substrate and the enzyme (Supplementary Methods and Supplementary Table 4). As for 4OP-AmDH, the (*S*)-amine was preferentially formed, in contrast to engineered AmDH,[8,15] *Asp*RedAm[20] and IRED-Sr,[27] each of which gave the (*R*)- enantiomer in the case of reaction with ammonia. From this point of view, nat-AmDHs are comparable with IRED IR_7, IR_10-11, IR_15, IR_20 and IR_25 reported by Wetzl *et al*. [16] In the case of methyl-substituted cyclohexanones **2** and **3**, the (*R*)- configuration was only slightly preferentially transformed.

A family of native AmDHs

Our discovered native AmDHs are evolutionarily unrelated to engineered AmDHs, IREDs and RedAms as illustrated by a heatmap based on sequence identity (Supplementary Figure 12).To analyze the sequence diversity and relationships between proteins of this family (Supplementary Figure 14), we created a sequence similarity network (SSN) built with 5313 protein homologs to AmDH4, *Cfus*AmDH and *Msme*AmDH (Fig. 2, Supplementary Figure 13), in which the most similar proteins (above 40 % of sequence identity) are connected. *Msme*-, *Myco*-, *Mvac*- and *Micro*AmDH are highly related and are found in the same cluster as *Cfus* and *Apau*AmDH, but in a different sub-cluster. By contrast, 4OP-AmDHs are found in a distant cluster (Fig. 2a). Over the family, 70 % of the proteins are annotated as Dihydrodipicolinate reductase (EC 1.17.1.8). Interestingly, the experimentally validated 4-

hydroxytetradihydrodipicolinate reductase DapB[28] is not present in this set. The nat-AmDHs have been probably over-annotated due to the presence of similar NADPH domains, as indicated on their Uniprot/TrEMBL automatic annotation records (Supplementary Figure 13). About 1 % are annotated as 2,4-diaminopentanoate dehydrogenase (EC 1.4.1.12), including the experimentally validated ones[22] (*ord* genes, Uniprot ID: C1FW05 and E3PY99) (Fig. 2b).

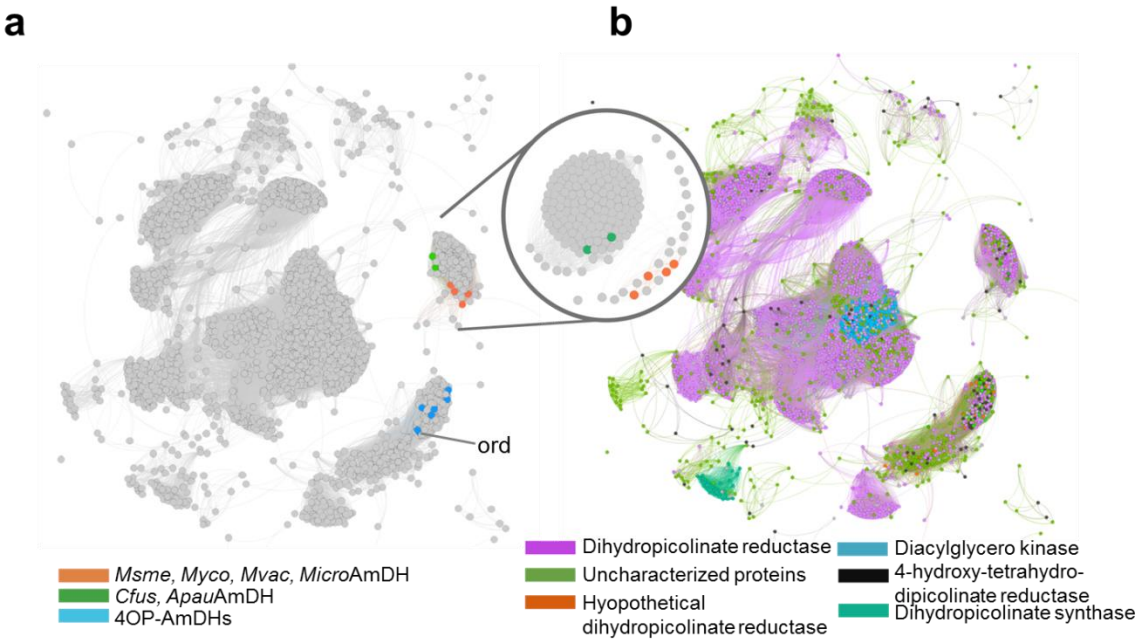


Fig.2| Sequence similarity network (SSN) of the native AmDH family. Edges between proteins (represented by nodes) are represented if the alignment score between two proteins is above 62, which correspond to 30 % of sequence identity. Mapping of: (**a**) Discovered native AmDHs. The only two proteins with experimental data listed in Swissprot are indicated on the graphics (ord for 2,4-diaminopentanoate dehydrogenase). The circle on the right shows a zoom on the group containing nat-AmDHs. The SSN have been cut here for an alignment score of 80 (i.e. at least 40 % of sequence identity). (**b**) Annotations retrieved from Uniprot.

The lack of knowledge about the mechanism of reaction of these enzymes, combined with the biocatalytic interest in AmDHs, prompted us to examine the structure of some of them using X-ray crystallography. Our efforts focused on AmDH remote homologs, *i.e.* belonging to different SSN clusters or subclusters, but belonging to the same phyla (Supplementary Figure 13): AmDH4, *Msme*AmDH and *Cfus*AmDH.

Structure and mutagenesis of nat-AmDHs

The structure of AmDH4, without substrate but bound to the cofactor NAD^+, was determined in two crystal forms, one of which contained one monomer (open) and the other form two dimers (open/closed) within the asymmetric units (Fig. 3a, Methods). The AmDH4 monomer consisted of two domains; an N-terminal Rossmann fold domain and a six-strand C-terminal sheet domain, the whole structure bearing a superficial resemblance to dihydropicolinate reductase (PDB ID: 5KT0) and some natural amino acid dehydrogenases such as *meso*-diaminopimelate dehydrogenases (DAPDHs)

(PDB ID: 1F06[29] and 3WBF[30]), with which AmDH4 does not share significant sequence homology. Within these structures, the Rossmann fold domain was well-conserved, but significant differences were observed in the C-terminal beta-sheet domain, which featured more and longer strands than in DAPDHs, and also in the structure of the loops that connect the strands. The structure with two dimers revealed that one monomer within each dimer was significantly more closed over the cofactor, changing the nature of the active site significantly (Fig. 3b). A similar domain closure upon substrate binding has been recognized as important for catalysis in DAPDHs.[30]

A model of the ternary complex was constructed with the natural product 2,4-DAP in the closed conformation of AmDH4 using *in silico* docking (Fig. 3c, Supplementary Methods). The active site features residues Glu102, Val134, Asn135, Arg161, Asn163, Phe168, Val172, Gln176, His197, Ile198, His264, Gln266, Gly299 and Thr303. Although there are many significant differences between the active sites of DAPDHs and AmDH4, a conserved feature is Glu102, which is replaced by Asp92 in *m*-DAPDH (PDB ID: 3WBF),[30] which forms a close interaction with the amine of the *meso*-diaminopimelate (*m*-DAP) substrate. This is strongly suggestive of a role for Asp92 in the activation of ammonia for the reductive amination reaction in this enzyme; a role which can also be considered for Glu102 in AmDH4. The K_m for the substrate 2A4OP was only slightly affected by the mutation of Glu102 to alanine (326 *vs* 894 μM). However, NH_3 saturation could not be reached within the experimental limit of 8 M, indicating a much weaker affinity for the mutant, compared to the K_m of 320 mM for the wild type. These results, again, underline the importance of Glu102 for both the fixation and activation of NH_3 (Table 3). Modeling studies with the docked keto derivative and NH_4^+ (Fig. 3d) suggests that, when ammonia is present, the 2A4OP is positioned at the bottom of the active site, its carboxylate being then maintained through electrostatic interactions with Arg161, Asn163 and His264. Indeed, variants R161M, N163V, H264L displayed highly reduced activities toward 2A4OP (Supplementary Methods and Supplementary Table 2-3), data supported by energy and ranking of poses issues from docking 2,4-DAP on AmDH4 variants (Supplementary Methods, Supplementary Discussion and Supplementary Table 8). Gln266, situated near the bottom of the active site, and Phe198, which belongs to the flexible lid which closes up over the cofactor in presence of the substrate, form the ideal active site configuration for catalyzing the reaction.

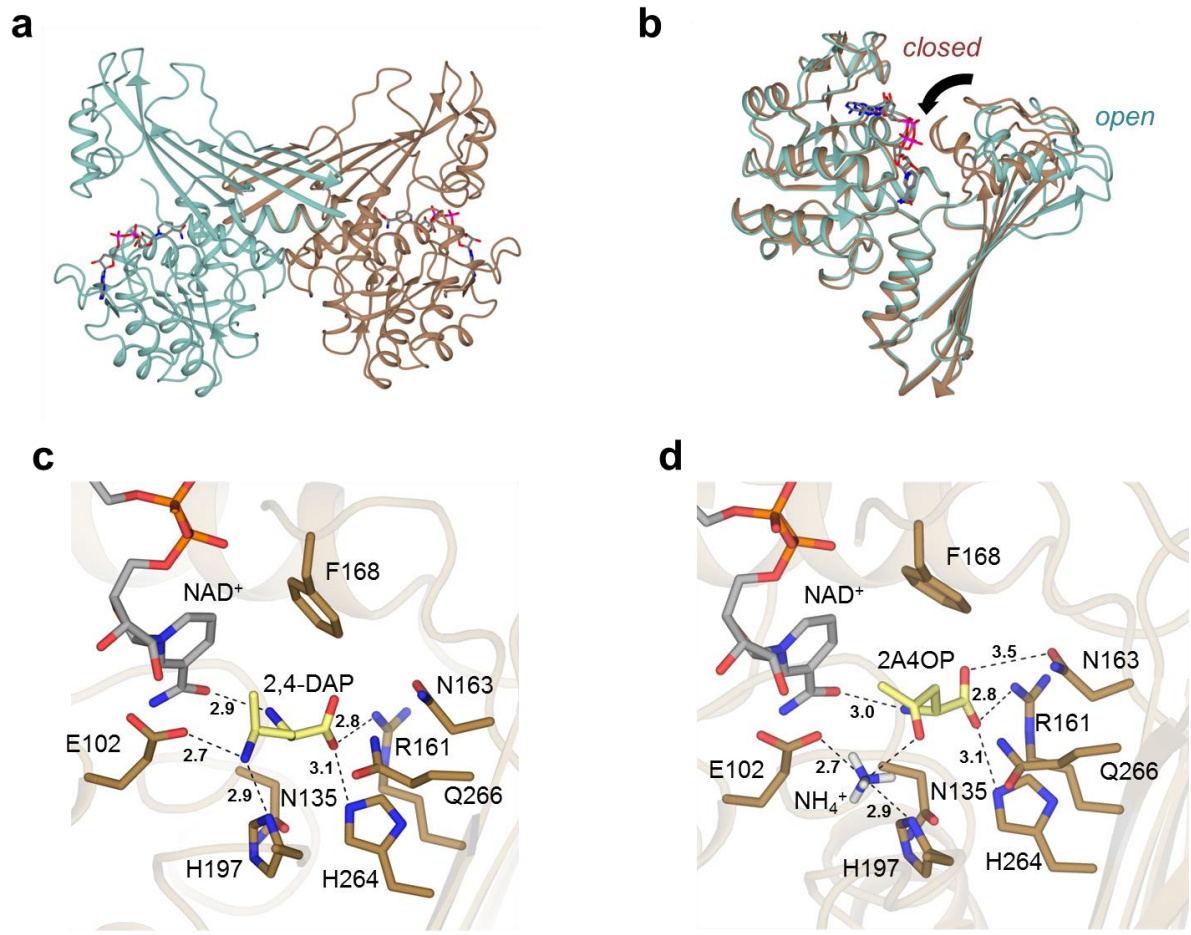


Fig.3| Structural and mutagenesis data of AmDH4. (**a**): Structure of dimer of AmDH4 in complex with NAD^+ consisting of one open and one closed form (PDB ID: **6G1M**); (**b**): Superimposition of open and closed monomers of AmDH4 in blue and brown respectively; (**c**): Active site within closed form of AmDH4 with product 2,4-DAP modelled into it. Selected interactions are shown with black dashed lines; distances are in Ångstroms; (**d**). Model of AmDH4 (closed form) in complex with NAD^+, ammonia and substrate (2*R*)-2A4OP.

Table 3. **Kinetic parameters of AmDH4 and AmDH4 variant E102A.**

The structures of *Cfus*AmDH and *Msme*AmDH also featured dimers (Methods and Supplementary Figure 9) as suggested by gel filtration experiments. In one monomer of *Cfus*AmDH, electron density for the cofactor, $NADP^+$ and the amine product, cyclohexylamine **1a**, each of which had been included in the protein solution for crystallization, was apparent (Fig. 4a). The structure of *Msme*AmDH was also obtained with $NADP^+$ at the active site, and, although the crystals of the enzyme were grown in the presence of racemic 2-methylcyclohexanone **2**, density consistent only with the cryoprotectant ethylene glycol was observed at the active site (Fig. 4b). The structures of *Cfus*AmDH and *Msme*AmDH were closely related to that of AmDH4, with the monomers of each enzyme trapped in the closed form, providing further evidence for this conformation as crucial in the catalytic cycle of this class of dehydrogenases. The active sites of *Cfus*AmDH and *Msme*AmDH

resembled that of AmDH4 in that a cage is formed for the substrate by the nicotinamide ring of $NADP^+$ and an aromatic residue, either Tyr168 (*Cfus*AmDH) or Trp164 (*Msme*AmDH) on the other. Tyr173 (*Cfus*AmDH) and Phe169 (*Msme*AmDH) were present in place of Phe198 in AmDH4, closing over the substrates to complete the hydrophobic pocket as the gates to the active sites. In the case of *Cfus*AmDH, the amine group of the cyclohexylamine product interacts with the side chain of a glutamate residue Glu108 homologous with Glu102 in AmDH4 and Glu104 in *Msme*AmDH. The coordination of the amine group to Glu108 holds the tetrahedral carbon bearing the amine at a distance of 2.8 Å from the C4 atom of $NADP^+$, from which hydride is delivered. The interaction through space completes the tetrahedral geometry of this atom, was would be expected if hydride were to be delivered from C4 of the cofactor. As expected, neither enzyme featured a residue homologous to Arg161 in AmDH4, as neither was active towards ketone substrates with a carboxylic acid group.

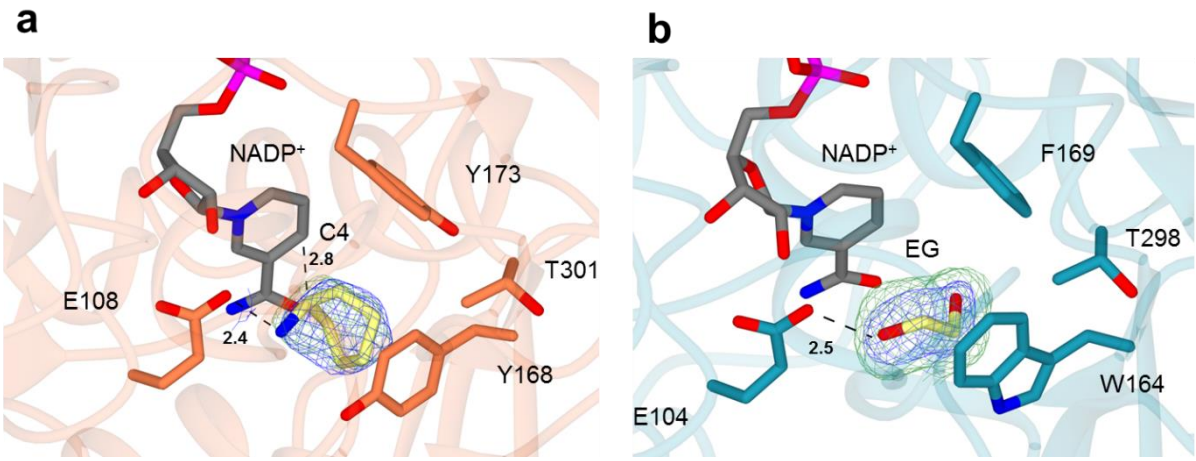


Fig. 4| Structure of active site of *Cfus*AmDH and *Msme*AmDH. (**a**): Structure of active site of *Cfus*AmDH complexed with $NADP^+$ and **1a** (PDB ID: **6IAU**). (**b**): Structure of active site of *Msme*AmDH complexed with $NADP^+$ and ethylene glycol (PDB ID: **6IAQ**). Electron density in blue and green corresponds to the 2*F*o-*F*c and *F*o-*F*c [31] omit maps at levels of 1σ and 3σ respectively. These were obtained prior to modeling of the ligand. Selected interactions are shown with black dashed lines; distances are in Ångstroms.

Based on the structures of nat-AmDH enzymes, coupled with kinetic analysis of mutants and tertiary and quaternary models (Supplementary Methods and Supplementary Table 3), a mechanism similar to the already reported biocatalytic reductive amination can be proposed (Supplementary Figure 10). The binding of the amine groups of the diaminopimelate and cyclohexylamine products by Asp92 in *m*-DAPDH and by Glu108 in *Cfus*AmDH respectively suggests a role for the activation of ammonia by this residue in the nat-AmDH enzymes. Ammonia would attack the electrophilic carbon of the ketone carbonyl secured by H-bond donors within the active site - possibly Tyr168 (*Cfus*AmDH) or Trp164 (*Msme*AmDH) (Supplementary Figure 18), or by electrostatic interactions via Gln266 (AmDH4, Fig. 3d), to form a carbinolamine which would be dehydrated to form the amine product after the attack of the hydride of the cofactor on the *re* face in case of a prochiral substrate. Thus a

mechanistic convergence between this enzymatic reductive amination reaction, the mechanism of phenylalanine dehydrogenase[32], the mechanism of fungal RedAms[33] and the imine formation reaction catalyzed within the Pictet-Spenglerase enzyme norcoclaurine synthase[34] may be hypothesized.

Engineering AmDH4 for the conversion of 2-pentanone

The structure of the active site of AmDH4 suggested that substitutions of polar residues Asn135, Arg161, Asn163 and His264, the last three being implicated in carboxylate binding in the enzyme, for non-polar ones, may alter substrate specificity with regard to the recognition of non-carboxylate substituted substrates such as pentan-2-one. These four residues were therefore mutated into non- polar residues to create single and multiple mutations at these positions. An activity was detected against pentan-2-one **14** for mutant containing at least the mutation R161M, and a specific activity of 104.8 $mU.mg^{-1}$ of enzyme was obtained for the quadruple mutant N135V/N163V/R161M/H264L (Supplementary Table 3 and Supplementary Figure 4). A model of this mutant with substrate pentan-2-one **14** and NH_4^+ docked into the engineered active site showed a preferential positioning of this substrate with favorable H-bond distances between H197/ Q266 and the carbonyl function (3.1 and 3.2 Å respectively). As with the wild-type, Glu102 is positioned favorably for the activation of ammonia (Supplementary Methods and Supplementary Figure 16). The (*S*)- stereochemistry was unchanged as confirmed by GC analysis (Supplementary Methods and Supplementary Figure 8).

Structural analysis of homolog active sites

Using the Active Site Modeling and Clustering method (ASMC),[35] 2029 members of the here defined family were modelled, using AmDH4, *Cfus*AmDH and *Msme*AmDH as structural templates. Residues defining the reference active site were chosen according to crystallographic and biochemical data obtained in this study on AmDH4 and are represented by the 3D-positions *P1* to *P20* (See Methods). Figure 5 illustrates the resulting Active Site hierarchical tree and the five groups defined, on which experimental activities mentioned in this study and in Mayol *et al*.,[23] are mapped. The G1 group contains enzymes for which 2,4-DAP activity could not be detected whereas the G2 group is populated with enzymes characterized as diaminopentanoate dehydrogenases, including AmDH4 (for further experimental screening of 2,4-DAP activity, see Supplementary Discussion, Supplementary Table 7 and Supplementary Figure 15). *Cfus*AmDH and *Msme*AmDH are found in the G3 and G4 groups respectively, whereas the G5 group is mainly composed of remote homologs. The three dimensional superposition of active sites belonging to the same group are projected linearly to form conservation patterns represented by logo sequences.

The catalytic glutamate (in position *P3* of the logo) in the three solved structures is highly conserved in G1 to G4 proteins supporting its critical catalytic role for enzymes with AmDH activity. Its good orientation may be secured by the conserved serine (*P1*). The absence of Glu in *P3* and Ser in *P2* in G5 group suggests a divergent activity for enzymes of this cluster. They might not perform imine formation but might use NADP as threonine (*P20*), which interacts with the nicotinamide moiety, is highly conserved. The volumes of the modeled active site pockets are much smaller in G3 and G4 groups, due to the presence of Trp in *P7*, Tyr/Trp in *P9*, Tyr/Phe in *P11* and Tyr/Phe in *P16*. These residues are very conserved in the active site despite these groups are heterogeneous in term of sequence and microbial diversity (Supplementary Discussion). With regard to the postulated

substrates for the different groups, Arg161 of AmDH4 (*P8*), which is suggested to form an electrostatic interaction with the carboxylate of 2,4-DAP/4-AP, is conserved in G1 and G2, suggesting that metabolic substrates for enzymes of these groups would also bear a carboxylic acid function. Its absence in groups G3 and G4 suggest that a compound with a charged group cannot be transformed by enzymes from either of these two groups, such as *Cfus*AmDH (G3) and *Msme*AmDH (G4). The lack of activity of these enzymes toward α-amino acids or 4OP-like substrates corroborates this hypothesis. Nine proteins from G1 were tested toward these two substrates, but none of them showed activity, confirming our hypothesis (Supplementary Table 7). The structural analysis of nat-AmDHs allows to define the following Prosite motif[36] [ST]-x(23)-E-x(30)-G-x(28)-[WYF]-N-x(3)-[YF]-x(133)-T which can be used as sequence signature to identify proteins of this family (Supplementary Methods).

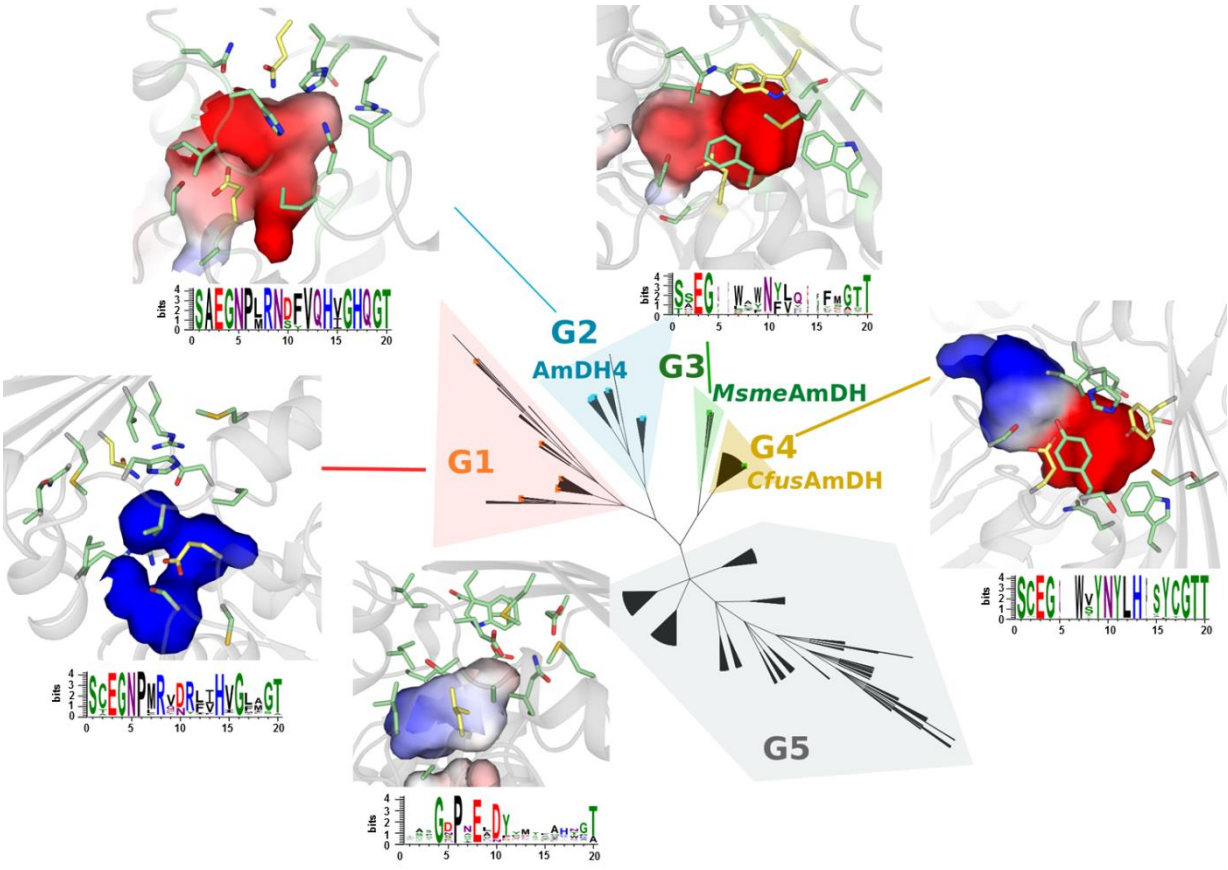


Fig.5| Dividing the AmDH family into groups with similar active sites. The hierarchical tree of active sites, generated by ASMC, contains 2011 proteins. The number of sequences (seq.) for each ASMC group is indicated under each logo, which represents the conservation of the active site residues. Electrostatic potentials for a representative structure of each ASMC group obtained using an adaptive Poisson-Boltzmann solver (APBS software) are displayed. The electrostatic potential maps are scaled from −15 to +15 kbT/ec, the negative electrostatic potential is highlighted in red, and the positive potential is highlighted in blue. The residues of the logo are represented in stick format. All these residues are colored in green except for P3, P9 and P16 which correspond to the catalytic

glutamate, and the hydrogen donors respectively, as observed in AmDH4, *Cfus*AmDH and *Msme*AmDH.

Semi-preparative scale reductive amination

To test the broad synthetic applicability of our nat-AmDHs,[37] 20 mL- reactions were performed with *Msme*AmDH, *Cfus*AmDH and *Micro*AmDH using racemic ketones **2**, **3** and **14** and aldehyde **21** and glucose/glucose dehydrogenase cofactor regeneration system (See Methods). Moderate to good conversions (45 - 76 %) were achieved at 50 mM of carbonyl compound in reasonable reaction times (6 – 24 h) with moderate enzyme loading (0.1 - 0.5 mg mL^{-1}). Products (2*S*)-pentan-2-amine **14a**, (1*S*, 2*R*)-2-methylcyclohexylamine **2a**, (1*S*, 3*R*)-3-methylcyclohexylamine **3a** and *N*-methylisobutylamine **21b** were obtained in 35 to 56 % isolated yields with *ee*/*de* > 97 % and good cis/trans ratio (90/10 for **2a** and 92/8 for **3a**) (Supplementary Table 9). These reactions validate these enzymes as interesting alternatives for the synthesis of chiral aliphatic amines and also as candidates for the reductive amination of aldehydes, leading to the bulk synthesis of terminal alkylamines, among other polymer precursors.

Conclusion

A family of native enzymes (nat-AmDHs) capable of catalyzing the reductive amination of ketones and aldehydes with ammonia and methylamine has been described. The illustrated biocatalytic synthesis on semi-preparative scale of three primary chiral amines and one terminal alkylated amine demonstrate their potential as biocatalysts for producing value-added chiral amines or bulk amines. Directed evolution may be employed to improve their performance, particularly with respect to increasing substrate concentration and substrate scope, as illustrated here with the quadruple AmDH4 variant. Nat-AmDHs constitute a type of AmDHs distinct from the already reported RedAms, engineered AmDHs and IREDs. A 3D-active site exploration of the bacterial biodiversity revealed various clusters to be explored for finding other valuable biocatalysts for reductive amination. A panel of active sites are available, ranging from one specialized for α-amino acid-bearing substrates to much more hydrophobic ones, which suggests a phylogenetic evolution with divergence inside this superfamily.[38]

Methods

General. For details of chemicals, strains, various procedures and all data, see Supplementary Information and Supplementary Data 1-7.

Specific activities measurements. All the reactions were conducted from duplicate at 20 °C in a spectrophotometric cell (10 mm light path) in a final reaction volume of 100 µL. For the screening of carbonyl-containing compounds **1**-**22** with ammonia as amine source, the procedure was as follow: to a mixture of ammonium formate buffer (2 M NH_4HCO_2/NH_4OH, pH 9.5) with NADH or NADPH (0.2 mM) was added an appropriate amount of purified enzyme (0.01 – 0.5 mg/ml). Carbonyl-containing compound (10 mM) was then added to initiate the reaction. The initial slope measured at 340 nm determined the specific activity of the enzyme according to Beer–Lambert's law and the molar absorptivity of β-NADH (ε = 6220 M^{-1} cm^{-1}) after subtraction of the slope obtained under the same conditions except without substrate. In the case of methylamine as amine source, the procedure was

as follows: to a mixture of TRIS.HCl buffer (50 mM, pH 9.5), NADH/ NADPH (0.1 mM each), and methylamine (500 mM) was added an appropriate amount of purified enzyme (0.02 – 0.2 mg/ml). Carbonyl-containing compound **1** or **21** (5 mM) was then added to initiate the reaction (reaction final volume 100 µL). The specific activity was determined as described above.

Determination of kinetic parameters. Kinetic parameters were determined by spectrophotometric NAD(P)H-monitoring in the forward (reductive amination) direction at 340 nm. All the reactions were performed in ammonium formate buffer pH 8.5-9.5 in a final volume of 100 µL at room temperature (approximately 23 °C) or at 50 °C when specified, in spectrophotometer cell with optical paths of 0.6 or 1 cm, depending on the saturated concentration of NADH or NADPH. Initial rates of the reaction were measured with various concentrations of substrate and saturated or optimal concentrations of the other substrates. The uncertainties are those generated by the fitting and the data are averages of 3 experiments. Details regarding the fitting (Sigma Plot software) and plots are provided in Supplementary Methods and Supplementary Figures 4-6.

Biocatalytic synthesis. The detailed procedures of analytical biocatalytic reactions carried out to validate the AmDH activities, including calibration curves (Supplementary Figure 7), the determination of enantio - and diastereoselectivities, are described in Supplementary Methods. For semi-preparative scale reactions, rapid optimizations of the reaction conditions are detailed in Supplementary Discussion and Supplementary Figures 20-22. Typical procedures are as follows: 20 mL-reactions with ammonia were conducted in a 50 mL-Greiner tube equipped with a screw cap containing 50 mM ketone, 60 mM D-glucose, 0.2 mM $NADP^+$, 60 U of GDH (Codexis GDH 105) and the indicated amount of purified AmDH in 1-2 M ammonium formate buffer pH 9.0. Reactions were stirred at 30°C 400 rpm for 6 - 24 h and then basified to pH 12 with 10 M KOH solution. The products were extracted with diethylether (3 x 20 mL), the combined organic layers were dried ($MgSO_4$) and concentrated to approximately 10 mL before addition of 1.2 eq of a solution of 2 M HCl in diethylether. In case of precipitation, the resulting solid was filtered, washed with cold diethylether and dried to afford the desired amine as monohydrochloride salt. Otherwise, 10 mL of distilled water were added and the product extracted with 2 x 20 mL of water. The combined aqueous phases were washed with diethylether (3 x 10 mL) to remove the unreacted ketone. The water phase was then lyophilized to afford the desired product as monohydrochloride salt. The semi-preparative scale reaction with methylamine was conducted in a 50 mL-Greiner tube equipped with a screw cap containing 50 mM isobutyraldehyde, 60 mM D-glucose, 0.2 mM $NADP^+$, 60 U of GDH (Codexis GDH 105), 0.1 $mg.mL^{-1}$ of purified *Micro*AmDH in 50 mM sodium phosphate buffer pH 8.0 at 30°C 400 rpm for 10 h. The product was extracted from the reaction mixture as described above in the case of no precipitation. For further details including GC and UHPLC chromatograms and NMR spectra, see Supplementary Figures 23-32.

Crystallization. See Supplementary Methods for production and purification of AmDHs for crystallization. AmDH4, *Cfus*AmDH and *Msme*AmDH were subjected to crystallization trials using a range of commercial screens in 96 well-plate sitting-drop format dispensed by a Mosquito robot (TTP Labtech). Crystals were optimized in hanging drops in 24 well Linbro dishes in 2 mL drops by the vapor diffusion method, using the best crystallization conditions from the initial screen. The best AmDH4 crystals were obtained in drops containing either 0.1 M bis-TRIS pH 6.5 with 0.8 M Li_2SO_4 (open form) or 0.1 M TRIS-HCl pH 6.5, with 0.2 M Li_2SO_4 and 25 % (w/v) PEG 2000 MME (open/closed form). The best *Cfus*AmDH crystals were obtained in 0.1 M Tris pH 8.5 with 2.0 M Ammonium sulfate

with 10 mM $NADP^+$ and 20 mM cyclohexylamine added to the protein before crystallization. The best *Msme*AmDH crystals were obtained in 0.1 M bis-TRIS pH 6.5 with 0.2 M magnesium chloride, 25 % (w/v) PEG 3350 with 10 mM $NADP^+$ and 10 mM racemic 2-methylcyclohexanone added to the protein before crystallization. In the case, of AmDH4 and *Msme*AmDH, the best crystals were transferred into a cryoprotectant solution of 15 % (w/v) ethylene glycol in the mother liquor before flash-cooling with liquid nitrogen for testing at the synchrotron. *Cfus*AmDH crystals were flash-cooled from their mother liquor without the addition of further cryoprotectant.

AmDH4 crystals were obtained using a number of crystallization conditions, but extensive molecular replacement experiments using a range of models and automated MR programs did not prove successful in solving the structure. SAD phasing was then attempted using sulfur as the target atom and was successful in phasing the first structure of AmDH4. The crystal was in the $P6_122$ space group, and featured one molecule in the asymmetric unit, although an examination of the symmetry-related neighbors strongly suggested a dimeric assembly, which was confirmed by gel filtration studies.

Despite the ease of crystallization of *Cfus*AmDH and *Msme*AmDH, molecular replacement strategies using either PDB database models, or indeed the recently acquired structure of the related AmDH4, with which *Cfus*AmDH and *Msme*AmDH share 22 and 26 % sequence homology respectively, were unsuccessful. The structure of *Cfus*AmDH was therefore solved using a selenomethionine derivative of the protein. The successful determination of the structure of *Cfus*AmDH gave a model which was then used to solve the structure of *Msme*AmDH by molecular replacement. The structure of *Cfus*AmDH was obtained from crystals in space group $P2_12_12_1$ with two molecules in the asymmetric unit, representing one dimer, whereas *Msme*AmDH was obtained in space group $P2_1$ with two dimers. The dimeric association of two monomers for each enzyme, which was suggested by gel filtration experiments, was clear in the crystal structures.

Data Collection and Refinement. Datasets described herein were collected at the Diamond Light Source, Didcot, Oxfordshire, U.K. The data on the open form of AmDH4 were collected on beamline I04-1. Data on the open/closed form of AmDH4, and also for *Cfus*AmDH and *Msme*AmDH were collected on beamline I04. Data were processed and integrated using XDS[39] and scaled using SCALA as part of the Xia2 processing system.[40] Data collection statistics are given in Supplementary Tables 5-6. The open structure of AmDH4 was initially solved using SAD with sulfur as the target atom in the following manner. Vitrified crystals of AmDH4 were taken to beamline I04 at Diamond Light Source. Data were collected at a wavelength of 1.7A (chosen to maximise the anomalous signal from sulphur while mitigating the effect on crystal lifetime of the increased absorption when going to longer wavelengths). High multiplicity of the data was achieved by attenuating the incoming beam to 8 % and the collection of 999 degrees of data (no significant crystal decay observed; average anomalous multiplicity just over 50). Data were integrated using XDS via xia2, and scaled and merged using the data reduction pipeline in CCP4I2. Structure solution and model building were performed using the CRANK2 pipeline in CCP4I2 (using SHELXD/E, refmac, parrot and buccaneer). SHELXD found 18 peaks with an occupancy over 25 % of the highest peak. The resulting model contained 338 residues and had R/R_{free} of 0.26/0.29. Final refinement was carried out using coot/refmac. These coordinates were subsequently used to solve further datasets of superior quality, including the one for which data are presented in Supplementary Table 5. In this case, the space group was $P6_122$ and featured one molecule in the asymmetric unit.

The structure of the open/closed form, in the *C*2 space group was subsequently solved using molecular replacement MOLREP[41] and the 'open' form of AmDH4 as the model. Structures in all cases were built and refined using iterative cycles using Coot[42] and REFMAC,[43] employing local NCS restraints with the *C*2 crystal form. Following building of the protein backbone and side chains in the case of both structures, clear residual density was observed in the omit maps at the domain interfaces, and this could be modelled as the cofactor NAD^+. Although additional density was observed in the omit maps in closed monomers of the *C*2 dataset, this was not deemed sufficient for modeling as the added ligand (2*R*)-2-amino-4-oxopentanoate. The final structures exhibited % R_{cryst} and R_{free} values of 15.9/19.6 (open form) and 24.8/29.5 (open/closed form). Refinement statistics for all structures are given in Supplementary Table 5. The Ramachandran plot for the open form showed 96.1 % of residues in the most favored regions, 3.3% in additional allowed and 0.6% residues in outlier regions. The corresponding values for the open/closed form were 94.5 %, 4.8 % and 0.7 %. Coordinates and structure factors for the open and open/closed forms have been deposited in the Protein Data Bank (PDB) with accession codes **6G1H** and **6G1M** respectively.
The structure of *Cfus*AmDH was solved using a selenomethionine derivative, which was prepared using a method previously described.[44] The structure obtained was used to solve native data from crystals which were obtained in space group $P2_12_12_1$ with two molecules in the asymmetric unit. Following building of the protein backbone and side chains, clear residual density was observed in the omit maps at the domain interfaces, and this could be modelled as the cofactor $NADP^+$. Additional density adjacent to the nicotinamide ring could then be modelled as the added ligand cyclohexylamine. The final structure exhibited R_{cryst} and R_{free} values of 19.4 % and 21.9 %. Refinement statistics for all structures are given in Supplementary Table 6. The Ramachandran plot showed 96.6 % of residues in the most favored regions, 3.1 % in additional allowed and 0.3 % residues in outlier regions. Coordinates and structure factors for *Cfus*AmDH have been deposited in the Protein Data Bank (PDB) with accession code **6IAU**.
The structure of *Msme*AmDH was solved using the structure of *Cfus*AmDH as a model. The structure obtained was used to solve native data from crystals which were obtained in space group $P2_1$ with four molecules in the asymmetric unit. Following building of the protein backbone and side chains, clear residual density was observed in the omit maps at the domain interfaces, and this could be modelled as the cofactor $NADP^+$. Additional density adjacent to the nicotinamide ring could be modelled as the cryoprotectant ethylene glycol. The final structure exhibited R_{cryst} and R_{free} values of 20.0 % and 22.1 %. Refinement statistics for all structures are given in Supplementary Table 6. The Ramachandran plot showed 96.9 % of residues in the most favored regions, 2.8 % in additional allowed and 0.4 % residues in outlier regions. Coordinates and structure factors for *Msme*AmDH have been deposited in the Protein Data Bank (PDB) with accession code **6IAQ**.

Active Site Modeling and Clustering (ASMC). All members of the AmDH family (5942 sequences), was clustered using 100 % of similarities to obtain a non-redundant set of 5313 proteins. This set was submitted to ASMC software[45] whose three main steps are described below. Homology modeling was run using as template the structures of AmDH4 (PDB ID: **6G1M**: chain B), *Cfus*AmDH (PDB ID: **6IAU**: chain A), and *Msme*AmDH (PDB ID: **6IAQ**: chain A). We have prior checked that no other proteins from the family had a structure available in the Protein Data Bank. Only proteins sharing at least 23 % of sequence identity with AmDH4, *Cfus*AmDH or *Msme*AmDH were modelled. In total 2011 proteins were used for the active site classification. Secondly, detection of pockets in the structure of AmDH4 was performed using Fpocket software.[46] The reference pocket is composed of

20 residues: 78, 101, 102, 133, 135, 136, 140, 161, 163, 164, 168, 172, 176, 197, 198, 199, 264, 266, 299, 303 (see Supplementary Methods for definition of the reference pocket). Thirdly, the 2011 sequences were structurally aligned on AmDH4 structure. Residues aligned with the 20 residues of the reference pocket were extracted to form a multiple sequence alignment of 2011 sequences. WEKA, a sorting algorithm was used to classify the families and to generate a hierarchical tree of active sites. The tree was cut manually as a function of the root, except for two branches, which were divided into two groups (G1 / G2, and G3 / G4). See Supplementary Figure 17 for the corresponding numbering of binding sites residues in the logo sequence and Supplementary Figure 19 for the logo representation of the multiple alignments of the whole sequences of G3 and G4 groups.

References.

1 Bornscheuer, U.T. Biocatalysis: Successfully Crossing Boundaries. *Angew. Chem. Int. Ed.* **55**, 4372-4373 (2016).

2 Ghislieri, D. & Turner, N.J. Biocatalytic Approaches to the Synthesis of Enantiomerically Pure Chiral Amines. *Top. Catal.* **57**, 284-300 (2014).

3 Sharma, M., Mangas-Sanchez, J. & Grogan, G. NAD(P)H-Dependent Dehydrogenases for the Asymmetric Reductive Amination of Ketones: Structure, Mechanism, Evolution and Application. *Adv. Synth. Catal.* **359**, 2011-2025 (2017).

4 Grogan, G. Synthesis of chiral amines using redox biocatalysis. *Curr. Opin. Chem. Biol.* **43**, 15-22 (2017).

5 Vidal, L.S., Kelly, C.L., Mordaka, P.M. & Heap, J.T. Review of NAD(P)H-dependent oxidoreductases: Properties, engineering and application. *Biochim. Biophys. Acta (BBA) – Prot. Proteomics* **1866***,* 327-347 (2017).

6 Höhne, M. & Bornscheuer, U.T. Biocatalytic Routes to Optically Active Amines. *ChemCatChem* **1**, 42-51 (2009).

7 Patil, M.D., Grogan, G., Bommarius, A.S., Yun, H. Oxidoreductase-Catalyzed Synthesis of Chiral Amines. *ACS Catal.* **8**, 10985–11015 (2018)

8 Abrahamson, M.J. et al. Development of an amine dehydrogenase for synthesis of chiral amines. *Angew. Chem. Int. Ed.* **51**, 3969-3972 (2012).

9 Abrahamson, M.J., Wong, J.W. & Bommarius, A.S. The Evolution of an Amine Dehydrogenase Biocatalyst for the Asymmetric Production of Chiral Amines. *Adv. Synth. Catal.* **355**, 1780-1786 (2013).

10 Bommarius A.S., Abrahamson M.J. & Bommarius, B. Engineered amine dehydrogenases and methods of use thereof. US 8835136 B2 (2014).

11 Au, S.K., Bommarius, B.R. & Bommarius, A.S. Biphasic Reaction System Allows for Conversion of Hydrophobic Substrates by Amine Dehydrogenases. *ACS Catal.* **4**, 4021-4026 (2014).

12 Pushpanath, A. et al. Understanding and Overcoming the Limitations of *Bacillus badius* and *Caldalkalibacillus thermarum* Amine Dehydrogenases for Biocatalytic Reductive Amination. *ACS Catal.* **7**, 3204-3209 (2017).

13 Ye, L.J. et al. Engineering of Amine Dehydrogenase for Asymmetric Reductive Amination of Ketone by Evolving *Rhodococcus* Phenylalanine Dehydrogenase. *ACS Catal.* **5**, 1119-1122 (2015).

14 Lowe, J., Ingram, A.A. & Groger, H. Enantioselective synthesis of amines via reductive amination with a dehydrogenase mutant from *Exigobacterium sibiricum*: Substrate scope, co-solvent tolerance and biocatalyst immobilization. *Bioorg. Med. Chem.* **26**, 1387-1392 (2017).

15 Chen, F. et al. Reshaping the Active Pocket of Amine Dehydrogenases for Asymmetric Synthesis of Bulky Aliphatic Amines. *ACS Catal.* **8**, 2622-2628 (2018).

16 Wetzl, D. et al. Asymmetric Reductive Amination of Ketones Catalyzed by Imine Reductases. *ChemCatChem* **8**, 2023-2026 (2016).

17 Agard N.J. et al. Engineered imine reductases and methods for the reductive amination of ketone and amine compounds. US2015132807 A1 (2015).
18 Itoh, N., Yachi, C. & Kudome, T. Determining a novel NAD^+-dependent amine dehydrogenase with a broad substrate range from *Streptomyces virginiae* IFO 12827: purification and characterization. *J. Mol. Catal. B: Enz.* **10**, 281-290 (2000).
19 Wang, S. & Fang, B. Method for preparing chiral amine through asymmetric reduction under catalysis of marine strain OTI. CN103224963B (2013).
20 Aleku, G.A. et al. A reductive aminase from *Aspergillus oryzae*. *Nat. Chem.* **9**, 961-969 (2017).
21 France, S.P. et al. Identification of Novel Bacterial Members of the Imine Reductase Enzyme Family that Perform Reductive Amination. *ChemCatChem* **10**, 1-6 (2018).
22 Fonknechten, N. et al. A Conserved Gene Cluster Rules Anaerobic Oxidative Degradation of l-Ornithine. *J. Bacteriol.* **191**, 3162-3167 (2009).
23 Mayol, O. et al. Asymmetric reductive amination by a wild-type amine dehydrogenase from the thermophilic bacteria *Petrotoga mobilis*. *Catal. Sci. Technol.* **6**, 7421-7428 (2016).
24 Zaparucha, A., de Berardinis, V. & Vaxelaire-Vergne, C. *Chapter 1. Genome Mining for Enzyme Discovery, in Modern Biocatalysis: Advances Towards Synthetic Biological Systems.* (RSC, Cambridge, UK, 2018).
25 Knaus, T., Bohmer, W. & Mutti, F.G. Amine dehydrogenases: efficient biocatalysts for the reductive amination of carbonyl compounds. *Green Chem.* **19**, 453-463 (2017).
26 Godoy-Alcántar, C., Yatsimirsky, A.K. & Lehn, J.M. Structure-stability correlations for imine formation in aqueous solution. *J.Phys. Org. Chem.* **18**, 979-985 (2005).
27 Scheller, P.N. et al. Imine Reductase-Catalyzed Intermolecular Reductive Amination of Aldehydes and Ketones. *ChemCatChem* **7**, 3239-3242 (2015).
28 Reddy, S.G., Sacchettini, J.C. & Blanchard, J.S. Expression, Purification, and Characterization of *Escherichia coli* Dihydrodipicolinate Reductase. *Biochemistry* **34**, 3492–3501 (1995).
29 Cirilli, M. et al. The three-dimensional structure of the ternary complex of *Corynebacterium glutamicum* diaminopimelate dehydrogenase-NADPH-L-2-amino-6-methylene-pimelate. *Protein Sci.* **9**, 2034-2037 (2000).
30 Liu, W. et al. Structural and mutational studies on the unusual substrate specificity of meso-diaminopimelate dehydrogenase from *Symbiobacterium thermophilum*. *Chembiochem* **15**, 217-222 (2014).
31 Tomita, H., Katsuyama, Y., Minami, H. & Ohnishi, Y. Identification and characterization of a bacterial cytochrome P450 monooxygenase catalyzing the 3-nitration of tyrosine in rufomycin biosynthesis. *J. Biol. Chem.* **292**, 15859-15869 (2017).
32 Vanhooke, J. L. et al. Phenylalanine Dehydrogenase from *Rhodococcus sp*. M4: High-Resolution X-ray Analyses of Inhibitory Ternary Complexes Reveal Key Features in the Oxidative Deamination Mechanism. *Biochemistry* **38**, 2326-2339 (1999).
33 Sharma, M. et al. A Mechanism for Reductive Amination Catalyzed by Fungal Reductive Aminases (RedAms). *ACS Catal.* **8**, 11534-11541 (2018).
34 Lichman, B. R. et al. Structural Evidence for the Dopamine-First Mechanism of Norcoclaurine Synthase. *Biochemistry* **56**, 5274-5277 (2017).
35 Bastard, K. et al. Revealing the hidden functional diversity of an enzyme family. *Nat. Chem. Biol.* **10**, 42-49 (2014).
36 Sigrist, C. J. et al. New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**, D344-347 (2013).
37 Cosgrove, S. C. et al. Imine Reductases, Reductive Aminases, and Amine Oxidases for the Synthesis of Chiral Amines: Discovery, Characterization, and Synthetic Applications. *Methods Enzymol.* **608**, 131-149 (2018).
38 Furnham, N. et al. Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies. *J. Mol. Biol.* **428**, 253-267 (2016).
39 Kabsch, W. Xds. *Acta Cryst. D Biol. Cryst.* **66**, 125-132 (2010).

40 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43**, 186-190 (2009).
41 Vagin, A. & Teplyakov, A. MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Cryst.* **30**, 1022-1025 (1997).
42 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst. D Biol. Cryst.* **60**, 2126-2132 (2004).
43 Murshudov, G.N., Vagin, A. & Dodson, E.J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst.* **D53**, 240-255 (1997).
44 Nestl, B.M. et al. Structural and functional insights into asymmetric enzymatic dehydration of alkenols. *Nat. Chem. Biol.* **13**, 275-281 (2017).
45 de Melo-Minardi, R.C., Bastard, K. & Artiguenave, F. Identification of subfamily-specific sites based on active sites modeling and clustering. *Bioinformatics* **26**, 3075-3082 (2010).
46 Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics.* **10**: 168 (2009).

Acknowledgements. The authors would like to thank Marcel Salanoubat for supporting the project, Nuria Fonknechten for fruitful discussions and assistance with genomic context analysis, Christine Pelle and Peggy Sirvain for large scale purification and analytical gel filtration of the described enzymes, Olek Maciejak (University Val d'Essonne) for NMR assistance, the Region Ile de France for financial support of the 600 MHz spectrometer, Mr Sam Hart and Dr Jon Agirre for assistance with X-ray data collection and data analysis respectively, and the Diamond Light Source for access to beamlines I04 and I04-1 under proposal number mx-9948. This work was supported by Commissariat à l'énergie atomique et aux énergies alternatives (CEA), the CNRS and the University of Evry Val d'Essonne. We thank GlaxoSmithKline for award of a part-studentship to A.F., the Brazilian Government for a fellowship to L.B. under the Coordination for the Improvement of Higher Education Personnel (CAPES) scheme and the COST Action CM1303 "System Biocatalysis" for STSM of O.M in G.G.'s laboratory.

Author contributions. C.V.V. conceived the project and directed it with G.G.. C.V.V., G.G., O.M., A.Z. and V.d.B. supervised the project. V.d.B. and J.-L.P. performed the candidate enzyme selection. A.D., V.P. carried out the gene cloning, the protein expression and purification on a small scale and the enzymatic screening with input from A.M. and J.-L.P.. O.M., V.P. and C.V.V. conducted the specific activity measurements. G.G., J.P.T., A.F. and L.B. conducted all the structural resolution. K.B. conceived and conducted the structural bioinformatics analysis with input from O.M., C.V.V. and G.G. O.M. carried out the biochemical experiments under the supervision of A.P.. O.M. and C.V.V. performed the analytical and semi-preparative scale reactions. C.V.V., G.G., K.B. and O.M. wrote the manuscript with input from A.P., A.Z. and V.d.B.

Competing interest.
The authors declare no competing interests.

Additional information.
Supplementary information is available for this paper at

Data availability statement: Crystallographic data that support the findings of this study have been deposited in the Protein Data Bank (PDB) with accession codes **6G1H**, **6G1M**, **6IAU** and **6IAQ**. All the other data supporting the findings of this study are available within the paper and its Supplementary

Information file and Supplementary Data or from the corresponding authors upon reasonable request.