Article

Single active-site mutants are sufficient to enhance serine: pyruvate α-transaminase activity in an ω-transaminase

Dawid Deszcz[1], Pierre Affaticati[1], Nadine Ladkau[2], Alex Gegel[1], John M. Ward[1], Helen C. Hailes[2], Paul A. Dalby[1]

[1] Department of Biochemical Engineering, Gordon Street, University College London, WC1H 0AH, UK

[2] Department of Chemistry, 20 Gordon Street, University College London, WC1H 0AJ, UK

Corresponding Author

Prof. Paul A. Dalby
Department of Biochemical Engineering, Gordon Street, University College London, WC1H 0AH, UK
Tel: +44 (0) 207 6799566
Email: p.dalby@ucl.ac.uk

Running title:

Directed evolution of CV2025 ω-transaminase

Abbreviations:

SPAT, serine-pyruvate aminotransferase; CV2025 *Chromobacterium violaceum* ω-transaminase; PLP, pyridoxal 5'-phosphate; MBA , α-methylbenzylamine

Keywords:

Aminotransferase; biocatalysis; directed evolution; substrate specificity; transaminase

Abstract

We have analysed the natural evolution of transaminase structure and sequence between an α-transaminase serine-pyruvate aminotransferase, and an ω-transaminase from *Chromobacterium violaceum* with <20% sequence identity, and identified the active-site regions which are least conserved structurally. We also show that these structural changes correlate strongly with transaminase substrate specificity during evolution and therefore might normally be presumed to be essential determinants of substrate specificity. However, key residues are often conserved spatially during evolution and yet come from within a different region of the sequence via structural reorganisations. Here we also show that α-transaminase-type serine-pyruvate aminotransferase activity, can be engineered into the CV2025 ω-transaminase scaffold with any one of many possible single point mutations at three key positions, without the requirement for significant backbone remodeling, or repositioning of the residue from a different region of sequence. This finding has significant implications for enzyme redesign in which solutions to substrate specificity changes may be found that are significantly more efficient than by engineering in all sequence and structure determinants identified by correlation to substrate specificity

Introduction

Transaminases (**EC 2.6.1.**) have become increasingly useful industrially and preparatively for catalysing the production of chiral amino-moieties including amino-acids, and drug molecules such as imagabalin [1], sitagliptin [2], norephedrine and pseudoephedrine [3]. The *Chromobacterium violaceum* (CV2025) ω-transaminase was also recently engineered into a *de novo* pathway with transketolase, for the synthesis of chiral amino-diols [4-8].

Transaminases are well suited to industrial biotransformations. They require the pyridoxal phosphate (PLP) cofactor, but this is regenerated during the catalytic cycle. The reaction is also fully reversible and typically ends at a thermodynamic equilibrium of K=1, and so amine synthesis must be driven to completion for an efficient process. This can be achieved using various reaction engineering methods such as an excess of cheap donor amine, *in* situ product removal using resins or spontaneous coupled reactions, removal of a volatile product, or using a coupled enzyme system [9-13].

The protein structures of transaminases can be grouped into the type-I folds of PLP-dependent enzymes comprising the *S*-selective enzymes, and then also the type-IV folds which comprise D-amino acid aminotransferases, L-branched chain amino acid aminotransferases, and the *R*-selective aminotransferases [14-16]. Transaminases are also split into six sub-groups in Pfam [17], which broadly mirror their substrate preferences [18]. Sub-groups I and II contain the α-transaminases (αTAms) and sub-group III contains the ω-transaminases (ωTAms). The αTAms catalyse the transfer of amino moieties from the α-carbon of amino acids to the ketone moiety of an α-ketoacid, whereas the ωTAms mainly utilise donors with amines distal to the carboxyl moiety [19, 20], as well as other amine donors and their respective non-keto-acid acceptors [21, 22].

The structures of several type-I fold transaminases have been solved crystallographically, including the *C. violaceum 2025* ωTAm (CV2025) [23] and *S. solfataricus* serine:pyruvate αTAm (SPAT) [24] which we compare in this study. Two identical active-sites are positioned at the interface between two monomers, with one monomer contributing the majority of the substrate binding site. While the structures of two different ω-transaminases have been

compared directly [25], specific differences between the available CV2025 and SPAT structures have not been reported in detail previously.

CV2025 accepts a wide range of amines, notably *rac*-1-aminoindane, 1-methyl-3-phenylpropylamine, and (*S*)-α-methylbenzylamine (MBA) with 183%, 149% and 100% relative activity respectively [5]. It also accepts the amino-acid L-α-alanine with 165% relative activity to MBA, and yet accepts L-α-serine with only 9% activity relative to that of MBA. Crystal structures for both apo- and holo-enzyme forms of CV2025 were reported previously [23]. More recently, a structure of CV2025 bound to the inhibitor gabaculine has also been determined through co-crystallisation [25]. This identified a significant rearrangement of an active-site loop (residues 81-93), and other residues that interact with the inhibitor of potential importance for substrate specificity.

The *S. solfataricus* SPAT accepts a range of α-amino-acids [24], in particular methionine, phenylalanine and tryptophan. It also accepts histidine, asparagine, glutamine and serine. By contrast, aromatic aminotransferases do not accept serine, hence the classification of the SPAT enzyme as a serine:pyruvate TAm (Scheme 1), even though the specific activity with serine (0.015 μmol min^{-1} mg^{-1}, 5 mM substrates) was 4.5-fold lower than for methionine. Comparison of the active-site residues of SPAT to those of a distantly related yeast alanine:glyoxylate transaminase, suggested that F88 and Y240 in SPAT restricted the aromatic substrates to one rotamer, and also sterically excluded threonine [24]. The considerably hydrophobic substrate-binding pocket of SPAT was also suggested to explain the preference away from charged substrates such as aspartate, glutamate and α-ketoglutarate. Finally, a shortened loop between strands 9 and 10 in SPAT moved residue V329 further from the PLP and created a larger active-site pocket which potentially favoured the larger α-amino-acids.

Enzyme engineering efforts have been reported for several ωTAm types previously. For example, an ωTAm from *Arthrobacter citreus* was subjected to error-prone PCR to yield a variant, designated CNB05-01, with improved thermostability and 250-fold improved specific activity towards a substituted tetralone [26]. Active-site residues within an ωTAm from *Caulobacter crescentus* were identified through structural homology modeling, and then targeted by site-directed mutagenesis to investigate their roles in defining or broadening substrate specificity [27]. Homology modeling of the *Arthrobacter citreus* CNB05-01 ωTAm variant was similarly used to predict three active-site residues interacting with the aromatic group of 4-fluorophenylacetone as modeled to give the *R*-configuration of the amine [28]. A small series of rationally designed mutants yielded some with improved conversion of 4-fluorophenylacetone, and one which reversed the enantiopreference to an *R*-selective ωTAm for the reaction with isopropylamine and either 4-nitroacetophenone or 4-fluorophenylacetone. Homology modeling and substrate docking for ωTAm from *Vibrio fluvialis* also led to the design of two mutants, W57G and W147G, which each altered the substrate specificity towards a range of amines [29]. A related W60C mutation in CV2025 enhanced its specificity towards 4-substituted acetophenones and (*S*)-MBA [12]. More extensive site-directed mutagenesis of the *Vibrio fluvialis* ωTAm, guided by an analysis of homologous sequence alignments, yielded mutants with improved production of an intermediate for imagabalin synthesis [1]. Homology modeling and substrate docking were also used to guide saturation mutagenesis of active-site residues in the *R*-selective ωTAm ATA-117. This yielded improved activity towards progressively larger substrates for the synthesis of sitagliptin [2].

The previously reported engineering of TAms have reversed their enantioselectivity or modulated their substrate specificity without altering their transaminase subclass. However, the transition between subclasses, such as from an ωTAm to an α-TAm, has not been previously explored through directed evolution experiments.

Here we have aligned a diverse range of transaminase structures sharing the same type-I fold, to highlight significant structural differences between the active sites of αTAms and ωTAms, exemplified by SPAT and CV2025 in Figure 1. A phylogeny analysis of these structures has further delineated an evolutionary pathway between the substrate preferences of α- and ωTAms, that shifts from α-amino acids, via amines that are progressively more distal to the carboxylate moiety, and finally to amines with no carboxylate moiety.

Structural differences between α- and ωTAms were found in regions containing loops that form parts of the active site, and include their shortening, lengthening, and remodeling, and also the appearance and disappearance of short beta strands and alpha helices. Given the correlation between this extensive restructuring and the evolution of their substrate preferences, engineering the interconversion of their substrate specificities to improve αTAm type activity within an ωTAm enzyme, might be expected to be very challenging.

Generally, αTAms are not able to utilize non-amino acid amines, whereas ωTAms have evolved to accept them [15]. The ωTAms typically discriminate against amino-acids other than alanine or glutamate, formed respectively by aminating their preferred acceptor substrates pyruvate and □-ketoglutarate. Thus, engineering an αTAm into an ωTAm is potentially a significant challenge given that it must introduce acceptance of a non-carboxylated amine substrate. By comparison, engineering an ωTAm to include more αTAm-like activity, while retaining or excluding activity on non-carboxylated amines, should be a more achievable task. We therefore set out to test this notion using single-site saturation mutagenesis of CV2025 active-site residues, and demonstrated significant improvements in SPAT-like L-serine acceptance without also engineering the full SPAT-like active-site backbone topology. Comparison of our variants to sequences along the natural evolutionary pathway between the CV2025 and SPAT, has identified key similarities at the single residue level, even under a background of significant backbone structure shifts during the natural evolution. This has significant implications for our understanding of the natural evolution of enzyme specificities, as well as for improving the available approaches for guiding enzyme engineering using sequence comparisons and phylogenetic analyses.

Results and Discussion

Structural alignment and comparison of transaminases

The crystal structures of ω-transaminase from CV2025 and 28 other diversely related transaminases, including the serine:pyruvate transaminase from *Sulfolobus solfataricus* (SPAT), were aligned using the DaliLite server [30] to compare their active-site structure and topology. The CV2025 and SPAT structures were also aligned independently using the PLP cofactors in Pymol [31], to focus on key structural differences in one active-site forming region (Figure 1). While the two sequences share only 11% identity, their overall secondary and tertiary structure topologies are very similar and are clearly evolutionarily related. Monomers from each of the two structures aligned for 309 out of 376 SPAT residues to an RMSD of 3.4 Å in DaliLite, with most of the deviation originating from regions in their active sites.

Overall the SPAT sequence is 75 residues shorter than for CV2025 and there are significant deviations in four particular regions (shown in Figure 1 as regions A, B, C and D), around the active-site entrances of the two enzymes. In region A, the first 55 N-terminal residues of CV2025 are missing in SPAT, and the following 10-12 loop-residues present within both active sites are also significantly remodeled as a result. The missing 55 residues include two short helices within the first 28 amino acids, followed by a three-stranded beta sheet on one side of the active-site entrance as shown in Figure 1. However, the same region is

occupied by part of an extended loop and a new 13-residue helix in SPAT from the other monomer. This significantly extends the 11-residue loop (denoted region B) found in CV2025 (D314-H325). Based on a BLAST-P search [32] the first 55 residues of CV2025 are not present in 85% of the 226 closest homologues to SPAT.

A relatively smaller structural difference in the active site is in the length of the loop (denoted region C) leading from helix 4 (annotation from Humble et al. 2012 [23]) at the surface, directly down into the bottom of the active site, and back out again to connect to helix 5, also at the surface, thus forming a significant section of the central entrance channel. CV2025 accommodates a 13-residue loop (L81-H93), leading to position F88 located deep within the central active-site channel, compared to an 8-residue loop in SPAT. The space created by the shorter loop length in SPAT is also filled by part of the extended loop of region B described above. Interestingly, this loop in CV2025 was found to become displaced upon binding to the inhibitor gabaculine [25], and the authors suggested that the loop flexibility may account for differences in substrate specificity between CV2025 and a closely related *Pseudomonas* ωTAm.

The active-site of CV2025 contains a loop (G152-T157), followed by a single helical turn (I158-L163), a 3-residue turn, and finally a second helical turn (K167-G173) which extends furthest into the active-site entrance. Here we denote this as region D. In SPAT the two helices are replaced by a single longer 13-residue helix at the location of the more buried of the two CV2025 helices, thus providing a more accessible active-site entrance along with the deletion of at least ten interconnecting loop residues found in CV2025.

Other differences reflect the isolation of SPAT from a thermophilic organism, where thermophilic enzymes often contain shorter and less flexible loops at the protein surface [24]. For example, the C-terminal end of a 19-residue helix present in CV2025 (Y134-P141) is missing one helical turn and three subsequent loop residues in SPAT. The CV2025 helix thus protrudes further out into solvent at the enzyme surface than in SPAT, while the orientation of the helix relative to the PLP cofactor is significantly altered. Another surface loop containing a single helical turn, plus the first N-terminal turn of a 20-residue helix in CV2025 (residues Y189-F200) are also missing in SPAT.

Phylogenetic analysis of transaminase structures

The structural alignments of 29 transaminases were used to improve their sequence alignment and phylogenetic tree construction, and also provided a subsequent analysis of the evolution and structural differences in the active-site regions A, B, C and D as shown in Figure 2. The evolutionary tree shows a clade delineation into α-transaminases (Pfam subgroup I) and ω-transaminases (Pfam subgroup III) as expected [18], with the exceptions of the glutamine:2-deoxy-scyllo-inosose aminotransferase (PDB ID: 2C81) which groups with the αTAms, and the stereo-inverting D-phenylglycine aminotransferase (PDB ID: 2CY8) which groups with the ωTAms. The overall grouping is consistent with the changes in sequence identity and RMSD relative to CV2025. As the sequence identity to CV2025 increases there is a clear shift from αTAms (subgroup I) which prefer α-amino acids, to ωTAms which aminate at ketones positioned further from, or without, a carboxylate moiety (subgroup III). However there is also a progression of the substrate preferences (as annotated in the PDB) from beta, through gamma, delta, epsilon and zeta-amines, and ultimately on to non-carboxylated amines. While this range of substrates accepted within the ωTAms has been previously reported [18], the evolutionary progression above has not been previously resolved.

Figure 2 also shows that the RMSD for structures aligned to that of CV2025 shifts smoothly from below 3 Å in ωTAms to above 3 Å in αTAms, concurrent with a gradual sequence identity drop to below 13% to that of CV2025. Visual comparison of all structure alignments allowed us to identify clusters of structure types that were similar in regions A, B, and D. These clusters were found to map directly onto each clade within the phylogenetic tree as shown in Figure 2, demonstrating a clear link between the remodeling of each active site region and the substrate specificity of each enzyme subgroup. By contrast, the length of the loop denoted region C does not cluster with transaminase type, but rather varies considerably across all transaminases, with a difference of up to 10 residues from the shortest to the longest, and no clear trend.

With such significant differences between the active sites of ωTAms and αTAms, we set out to determine whether SPAT-like activity could still be engineered into the CV2025 scaffold using only single active-site mutations, and without remodeling of regions A, B, C and D towards a SPAT-like active-site topology. The grouping of glutamine:2-deoxy-scyllo-inosose aminotransferase (PDB ID: 2C81) with the αTAms indicates that the αTAm active site architecture can accommodate ωTAm-type activity. However, ability to engineer an interconversion between the two enzyme classes in the opposite direction, and with only single mutations, is not yet known.

Selection of target residues

Two computational structure modeling approaches were combined to select CV2025 ωTAm active-site residues with the potential to influence specificity through direct interaction with substrates. The active-site cavities were first mapped using CAVER [33] on the holo-enzyme in which pyridoxal-phosphate (PLP) forms a Schiff's base to K288. This mapped the large and small substrate-binding pockets in addition to the central channel leading out of the active site, and defined the region of structure in which to confine substrate docking by AutoDock [34]. Six substrates known to be either well (pyruvate, alanine, methylbenzylamine, acetophenone), or poorly (serine and hydroxypyruvate) accepted by CV2025 [5] were docked into the active-site, and the major clusters of final poses were used to identify all nearby residues with the potential to contribute to their binding, and thus provided targets for saturation mutagenesis. Details of the contacts made with each substrate and residue are provided in the supplementary information (Table S1). The selected residues are shown in Figure 3 along with the final docked positions for each substrate.

A total of 11 residues were selected for site-saturation mutagenesis (P21, F22, L59, W60, F88, Y153, A231, G319, F320, T321, R416), that were predicted to interact with at least one of the substrates. Three additional residues were selected due to their close proximity to the PLP cofactor (Y322), or their potential to create large cavities at the active-site entrance upon mutation (F89, Y168). All but one of these residues are highly conserved among 100 sequences with at least 40% sequence identity to CV2025, as summarised in Table 1. Four target residues, L59, Y153, G319 and T321 are completely conserved (H(x)=0). Only Y168 is highly variable (H(x)>2.0). The amino acid found at each equivalent position in SPAT is also shown in Table 1. These are different to any of the residues found in ωTAms, with four exceptions. F88 in CV2025 is F in SPAT and also in 90% of related ωTAms. F89 in CV2025 is S in SPAT and also in 10% of related ωTAms. A231 in CV2025 is S in SPAT and also in 3% of ωTAms. F320 in CV2025 is F in SPAT and also in 38% of ωTAms. The active-site residues in CV2025 therefore appear to be highly specialised toward ωTAm activity, and have little similarity to those found in SPAT.

Library generation and high-throughput screening

Saturation mutagenesis with the NNK codon was targeted to each of the fourteen residues independently, and 76 colonies randomly picked into 96-well plates, along with controls, as the basis for each residue library. Enzyme variants were expressed in *E. coli* BL21(DE3):pQR801 using IPTG induction, as described in Materials and Methods. DNA sequencing from random wells demonstrated that a wide diversity of mutations had been created in each library. A high-throughput colorimetric screen was used to assess the total enzyme activity from the conversion of serine and pyruvate to hydroxypyruvate and alanine after 3 hours, for each variant prepared as a crude lysate. The hydroxypyruvate generated was detected by reaction with WST-1 tetrazolium salt, giving rise to a blue coloration due to reduction of the tetrazolium and oxidation of the hydroxypyruvate. Complete sequencing of the Y153X library revealed a bias of 30% wild-type colonies, but a random distribution of between 0 and 7 occurrences for all other mutations, with all but five amino-acids (I,D,H,Q,T) represented. The expected number of distinct variants from the remaining 70% of 76 colonies, calculated using GLUE-IT [35], is 17. Sequencing of the 13 highest performing (non wild-type) W60X colonies revealed 6 of the 19 possible substitutions.

The maximum activities obtained within each library are shown in Figure 4. Transaminase expression levels in clarified lysates were also determined by SDS-PAGE with densitometry and found to vary only slightly in all of those measured, from 0.25-0.41 mg/ml (±0.1 mg/ml). Positions Y153, W60, F88 and T321 gave mutants with up to 4-fold, 4-fold, 3-fold and 2-fold improvements, respectively, in total activities towards 10 mM serine, 10 mM pyruvate at pH 7.0. Position L59 gave a modest 1.4-fold improvement, while all other residues led to no improvement in activity above that of wild type, even though a range of mutants were always identified within the libraries by random DNA sequencing. It is notable that five of the seven most highly conserved (H(x)≤0.4) residues among ωTAms could be mutated to give variants with improved SPAT-like activity, whereas all of the lesser conserved residues yielded no better than the activity of wild-type CV2025. This is consistent with our previous analysis of saturation mutagenesis of enzymes for improved activity or substrate specificity, and other examples [36-38], where the most effective mutations were found at highly conserved residues in direct contact with substrates or cofactors. It is anticipated that such residues are those that most strongly define enzyme substrate specificity, and are therefore precisely those that must be altered to allow the improved acceptance of new substrates.

For ωTAm, all of the mutants with at least 2-fold improved total activity towards serine were obtained at residues with H(x) values of 0.4 or below. If this had been used as a cut-off criterion for choosing target residues, the same mutants could have been obtained by creating and screening only half of the libraries. This echoes our previous work on transketolase where an H(x) value of 0.4 or below would have found all mutants that altered substrate specificity, and decreased the target residues from twenty to only nine [37, 39]. However, using this cut-off value of H(x) alone can come with some case-specific knock-on effects. For example, saturation mutagenesis of *T. litoralis* aminoacylase residues with H(x) values below 0.3, did give the greatest substrate specificity changes, but they also resulted in a partial decrease in the overall catalytic efficiency [38].

The distribution of mutations tolerated at the top-three sites provides further insight into the functional importance of each residue for maintaining the substrate specificity of ωTAm, and also the overall catalytic efficiency. The averaged activities for all sequenced mutants within the Y153X, W60X and F88X libraries are shown in Figure 5. Strikingly, serine:pyruvate activity was highly tolerant to many different types of mutation at these three sites.

Position Y153, yielded ten mutants with between 2-fold and 4-fold improvements over the wild-type activity. Two more (Y153E and Y153K) retained 50% activity, while only two (Y153R and Y153P), completely abolished activity. This is in stark contrast to the 100%

conservation of this residue in the ωTAms related to CV2025, although Phe is often observed at this position in the more distantly related ωTAms. The equivalent residue in SPAT is also Phe, and the Y153F mutation in CV2025 gave 3.5-fold improved SPAT activity. However Y153M, Y153N and Y153S each gave 4-fold improvements, indicating that simple mutation of CV2025 towards the residues present in SPAT, while a successful strategy, will not necessarily provide the best possible activities. During the review of this paper, the equivalent mutations to Y153F and Y153M in *V. fluvialis* ωTAm have been identified independently, and were found to decrease the specific activity towards methylbenzylamine, yet increase it towards (R)-phenylglycinol [40].

The hydroxyl group of Y153 hydrogen bonds via water to the phosphate moiety of PLP [41], positioning Y153 across the active site and directly above the PLP cofactor, potentially acting as a gateway residue for substrate access. Mutations towards charged residues (E, K, R) are all detrimental to activity, implying an electrostatic destabilisation of PLP binding into a functional position. Humble et al previously observed that Y153K decreased the enzyme activity, and attributed this to the loss of the edge-on interaction of the Y153 phenyl ring with the pyridine ring of PLP, as well as loss of the hydrogen bond via water to the phosphate moiety of PLP. However, mutations to almost any other residue than charged variants, and notably Y153F, remove the hydrogen bond interaction yet lead to increased activity towards serine. The magnitude of this increase does not appear to be related to the size of the new residue side-chain, indicating that steric hindrance is no longer a major factor for serine access to the PLP cofactor once the hydrogen bond is removed. Proline, which abolishes activity completely, presumably is structurally destabilising to the active site.

Position W60 was also very tolerant to substitution, with six mutations (W60Q, W60S, W60H, W60T, W60N and W60P), giving between 2-fold and 3.5-fold improvements in activity. The equivalent residue in SPAT is G9, yet mutation to Gly was not observed. Close relatives to CV2025 often contain a Tyr (2%) or Phe (10%) at this position, but none of these were observed under serine:pyruvate screening conditions. More distantly related ωTAms identified in the structure alignments, such as ornithine-AT (PDB ID: 1OAT), 4-aminobutyrate-AT (PDB ID: 3OKS and 1SFF), acetylornithine-AT (PDB ID: 2ORD), L-lysine-epsilon-AT (PDB ID: 2CJG), beta-phenylalanine-AT (PDB ID: 4AO9), and D-phenylglycine-AT (PDB ID: 2CY8), contain Ser, Ala, Thr or Gly at equivalent sites. These enzymes are all evolutionarily mid-way between SPAT and CV2025 (Figure 2) in both sequence and structure, and they also aminate ketones that are distal to a carboxylate moiety (with the exception of D-phenylglycine-AT). However, they are clearly clustered evolutionarily with the ωTAms. The identification also of W60S and W60T mutations with greater SPAT-like activity, and the presence of Ser and Thr in natural ωTAms which accept amino-acids with more distal amines, suggests the possible role of the W60 mutations in controlling the tolerance of a carboxylate moiety in the amine acceptor substrate. Interestingly, an equivalent mutation (W57G) in *Vibrio fluvialis* ωTAm, which has a preference for aromatic amines, was found to improve the acceptance of aliphatic amines [29]. Furthermore, the W60C mutation in CV2025 has been previously shown to improve k_{cat}/K_m for 4'-substituted acetophenones and (*S*)-MBA [12]. This alternatively suggests that our mutations of W60 to smaller residues may also be providing generally increased steric access and activities, including serine.

Residue F88 was similarly very tolerant to substitution, such that six mutants (F88R, F88H, F88Q, F88S, F88T, F88I), gave between 3-fold and 2-fold improvements in activity. Again, the only natural variation at this position within the close relatives of CV2025 was Tyr (10%), which was not observed when screening for SPAT-like activity. Within more distantly related ωTAms, the equivalent residue can be Met, Leu, Gly, Arg, Ala, Asn, Gln, or Ser, and three of these were found in the top four F88 mutations (F88R, F88Q and F88S). At this position, Arg, Gln and Ser are found respectively in ornithine-AT (PDB ID: 1OAT and 1GBN), 4-aminobutyrate-AT (PDB ID: 1SFF but not 3OKS), and L-lysine-epsilon-AT (PDB ID: 2CJG).

As previously with the W60 mutations, the phylogenetic analysis of natural TAms infers that F88R, F88Q and F88S mutations might better accommodate the carboxylate moiety within the amine acceptor substrate. Hence, these same mutations could provide improved affinity to serine through the same mechanism. The selection of positively charged (Arg) or polar hydrogen bonding (Gln, Ser) residues at F88 is consistent with an interaction with a carboxylate. However, in SPAT and in many other αTAms the F88-containing structure (Region C) is distinctly different due to a widely varying loop length as described above, and so no fully equivalent residue to F88 exists. Instead, SPAT contains a shorter loop in region C, leading to a partially equivalent residue F28 in region C of SPAT (Figure 1B). An extended loop in region B of SPAT also partially replaces F88 in CV2025 with SPAT residue Y240. The net result is a small cavity in SPAT that is filled by the aromatic ring of F88 in CV2025.

The binding of gabaculine to the holo-CV2025 enzyme in a co-crystal structure significantly alters the region B loop (residues 81-93) resulting in F88 pointing out of the active site [25]. It is therefore also plausible that some F88 mutations could have disrupted the loop in the holo-CV2025 form, improving access to the active site by pre-forming this putative substrate binding conformation.

Other residues also displayed interesting behaviour in terms of tolerance to mutation. For example, L59 gave a 1.4-fold improvement with L59M, but no other mutants were identified at this residue with greater than wild-type activity. The equivalent residue in SPAT is V8, and other ωTAm structures contain Trp, Ile, or Tyr at the equivalent site, but not Met. Residue T321 tolerated mutations to Ala and Ser with no loss in activity, and mutations to Gly and Lys with 50% of the activity retained. However, T321R abolished activity completely. The equivalent residue in SPAT is A242. Similarly, F89 tolerated mutation to Leu without loss in activity, whereas substitutions to Glu, Pro and Ser each resulted in 15-20% retention of activity. The equivalent residue in SPAT is S30. A Glu at the equivalent position is also found in diaminopelargonic acid aminotransferase, an ωTAm distantly related to CV2025. Residue F22 retained wild-type levels of activity from mutations to Trp and Met. There is no equivalent residue in SPAT. Y322 also tolerated substitutions to Cys, Leu, Thr, Ser and Phe with between 10% and 35% retention of activity. The equivalent residue in SPAT is T243. F320 gave mutations to Cys and Ser with 30% and 50% retained activity. The equivalent residue in SPAT is F241. A231 was mutated to Thr with 70% activity retained. Threonine is found in 9% of ωTAms closely related to CV2025. The equivalent residue in SPAT is S139, which is also similar to threonine.

By contrast, only residue R416 appeared to be intolerant to mutation. Fourteen colonies in total for R416X were sequenced, including all of those that were active. All active colonies were wild-type transformants, and all other randomly identified mutants R416Q, R416L and R416S resulted in completely abolished activity. The equivalent residue in SPAT is A326. It is possible that the two naturally observed ωTAm variants of His (8%) and Thr (1%), and the SPAT equivalent of Ala at this position were simply not generated in the library, or otherwise that they are not active towards serine:pyruvate.

To summarise, nine libraries were generated at positions for which there is a clear equivalent residue in SPAT (excluding CV2025:F88). The amino-acid found at each site in SPAT was identified in five libraries: one (Y153F) with improved SPAT activity, and four (F89, F320, T321, Y322) where they tolerated the mutation with no improvement in SPAT-like activity above that of wild-type CV2025. One other residue (A231) tolerated a threonine mutation, which is homologous to the serine found in SPAT. This demonstrates that mutation towards the equivalent residues in SPAT can be successful for gaining SPAT-like substrate specificity, even though the active-site structure in regions A-D around the mutated residues, differ significantly between the two wild-type enzymes. Therefore, the CV2025 structural form of regions A-D does not preclude SPAT activity.

Three libraries (W60X, F88X, Y153X) yielded improved SPAT-like activity, from mutations that also occurred at equivalent positions in the ωTAms found on the evolutionary path between CV2025 and SPAT. Another four (F89X, A231X, F320X, T321X) libraries tolerated such mutations with 15-100% retained activity. This finding demonstrates that a more SPAT-like preference for serine can already emerge through single active-site mutations found in variants only part way along the evolutionary path to SPAT. Improved activity with serine involves a shift from acceptance of non-carboxylated amines to a carboxylated amine, which reflects a similar shift along the natural evolutionary path from non-carboxylated amines in CV2025, through (β/γ/δ/ε/ζ)-amines (relative to the carboxylate moieties) in other ωTAm variants, and finally to α-amino acids in SPAT and other αTAms.

Kinetic analysis of purified Y153 mutants

Two mutants, Y153M and Y153S, which performed best in the high-throughput screen, were chosen for detailed kinetic analysis along with the wild-type enzyme, to compare the acceptance of serine to that of two substrates *rac*-aminoindane and (*S*)-MBA preferred by wild-type CV2025. Neither of the two mutations were observed in SPAT, or any of the TAm structures, or from the BLASTP sequences closely related to CV2025. Therefore, it was interesting to understand them further in terms of their kinetic effects. The mutant enzymes were purified and characterised at a range of serine, *rac*-aminoindane or (*S*)-MBA concentrations and a constant pyruvate concentration. Pyruvate concentration was fixed at 10 mM to avoid substrate inhibition observed at higher concentrations.

The initial velocity enhancements, expected in the high-throughput screen with serine, can be calculated from the kinetic parameters at 10 mM substrates shown in Table 2. This gives 21.7-fold and 5.3-fold improvements for Y153M and Y153S respectively, which are higher than the 4-fold activity improvements observed by high-throughput screening. This difference simply results from the non-linear relationship between initial velocities, and the degree of conversion at a fixed time-point used for high-throughput screening. It does not result from any differences in expression level for the clarified lysates used in the screening. Therefore the improvements shown in Figures 4 and 5 give the correct relative order of mutant activities, and the same conclusions discussed above, but the enhancements in activity relative to wild-type should actually stretch to up to 22-fold for the fastest (Y153M).

The k_{cat} of wild-type CV2025 enzyme towards aminoindane and MBA were each 16-fold higher than that for serine (0.5 s^{-1}). Meanwhile, the K_m for serine (260 mM) was 17-fold and 9-fold higher respectively than those for aminoindane and MBA. These results are consistent with the specific activities previously reported for CV2025 [5]. The Y153M mutation had little effect on k_{cat} for the three substrates. However, the K_m decreased 1.5-fold for aminoindane and 2.5-fold for MBA. By contrast, the Y153M mutation led to a 57-fold improvement in K_m for L-serine. The overall improvement in k_{cat}/K_m for Y153M relative to wild type was 67(±12)-fold for serine, but only 1.5-fold for aminoindane and 2.5-fold for MBA, indicating a strong shift in substrate preference towards serine relative to the two non-carboxylated amines.

The Y153S mutation led to a small (1.2-fold) decrease in k_{cat} for serine. This loss was more than countered by a 20-fold decrease in K_m for serine. By contrast, k_{cat} decreased 160-fold and 7-fold for aminoindane and MBA respectively, but K_m decreased by only 34-fold and 3-fold respectively. The net result was again a significant shift in substrate preference towards serine relative to the two amines, such that k_{cat}/K_m improved 16(±3)-fold for serine, but decreased 4.5-fold and 2-fold for aminoindane and MBA respectively.

The effect of the Y153M and Y153S mutations occurs mainly via K_m, showing a clear impact on substrate affinity by the mutations. The CV2025 transaminase active-site thus shows a high plasticity in its ability to significantly alter substrate specificity with just single mutations. The two mutants Y153S and Y153M both remain as □-transaminases with k_{cat}/K_m values for aminoindane and MBA at 4 to 6.5-fold higher, than for serine. However, this has closed the gap considerably compared to the wild-type CV2025 for which k_{cat}/K_m for aminoindane and MBA was 290-fold and 147-fold higher than for serine. Thus the Y153S and Y153M mutations have largely removed the discrimination against serine observed in the wild-type CV2025, while also retaining catalytic efficiency, particularly in Y153M. Only Y153S shows any impact on catalytic efficiency, with a minor decrease for serine but more significant decreases for aminoindane and MBA.

The specific activity of SPAT at 5 mM serine and 5 mM pyruvate is 0.015 μmol mg^{-1} min^{-1} [24], which is 50% higher than that calculated for CV2025 (0.01 μmol mg^{-1} min^{-1}) from the kinetic parameters under similar conditions. By this measure, CV2025 is already a reasonable acceptor of serine relative to SPAT. The specific activities of the two new CV2025 mutants Y153S and Y153M are similarly calculated from the kinetic parameters at 5 mM substrates to be 0.13 μmol mg^{-1} min^{-1} and 0.36 μmol mg^{-1} min^{-1}, respectively. Clearly then, the mutant CV2025 enzymes both have significantly higher SPAT activity than the SPAT from *S. solfataricus*.

Conclusion

Analysis of the natural evolutionary pathway between SPAT and CV2025 has revealed a gradual progression from the acceptance of α-amino acids, through amines that are increasingly distal to the carboxylate moiety, and finally to a preference for non-carboxylated amine substrates in CV2025. This pathway correlates with significant structural shifts in regions A-D within the enzyme active-site. Surprisingly, up to a 67-fold improvement in k_{cat}/K_m could be elicited towards SPAT activity from a single mutation, while 3-4 fold increases in SPAT activity could also be obtained from a wide range of different substitutions at any one of three key residues. The k_{cat}/K_m values for the favoured ωTAm substrates, aminoindane and MBA, were also improved in Y153M but decreased in Y153S. This showed how saturation mutagenesis of single residues is sufficient to significantly modulate the substrate specificity of the ωTAm CV2025 to include that of SPAT from the αTAm subclass. However, the extensive active-site structure and sequence differences, found to correlate to subclass specificities, may yet still be required to engineer these new mutants for the complete discrimination against non-amino acid amines. Our results also indicate that Y153, W60 and F88 in CV2025 are all important positions for defining ωTAm substrate specificity. Y153 appears to be a key gateway residue which controls access to the PLP cofactor, and which can be opened up to allow serine access through almost any mutation except charged variants and proline. Some of the most highly conserved active-site positions in CV2025, including Y153, W60, F88 and T321, were highly tolerant to many alternative mutations when screening for SPAT activity, demonstrating the high plasticity of the transaminase active site, and also suggesting that ω-TAm substrate specificity is partly driven by conserving the exclusion of α-amino acids.

The mutants found were often similar to those found at equivalent positions in sequence, or at least in structural space, along the natural evolutionary pathway between the CV2025 and SPAT. This is consistent with the important role and conservation of particular active-site residues spatially within the active site. It also suggests that the extensive structural remodeling, observed as distinct transaminase sub-classes have evolved, are not critical for conserving one particular substrate preference. It is thus tempting to conclude that single-site saturation mutagenesis provided a more efficient route to SPAT activity than by

mimicking the natural evolutionary pathway. However, the larger structural differences in active-site regions A-D between SPAT and CV2025 may conserve additional aspects such as the wider profile of their substrate preferences, feedback regulation/inhibition behaviours, or even fine-tune their relative catalytic proficiencies (k_{cat}) to levels which optimise flux through their respective metabolic pathways.

Finally, our engineered CV2025 ωTAm variant with improved activity towards serine, also generates hydroxypyruvate (HPA) as a byproduct. This has potential use in reactions coupled after transketolase (TK) which also accepts HPA, similar to a synthetic scheme demonstrated for the synthesis of 6-deoxy-L-sorbose [42]. This could now be applied within our *de novo* TK-TAm pathways [4, 8] to remove the need for HPA addition.

Materials and methods

Except where stated, all reagents were obtained from Sigma–Aldrich (Gillingham,UK) and used without further purification.

Structural alignment of transaminases

A similarity search within the protein databank webserver [43] was used to identify 28 TAms that are structurally related to that of *Chromobacterium violaceum* ω-Transaminase (CV2025 ωTAm) in the holo form (PDB ID: 4A6U) [23]. The structures included the serine:pyruvate transaminase from *Sulfolobus solfataricus* (SPAT) (PDB ID: 3ZRP) [24], along with a diverse range from high to very low sequence identity. From over 200 initial structures, those with unknown or non-TAm functions were removed. Near-identical structures and their mutants were also removed. The 28 structures remaining were aligned to the CV2025 chain A monomer using the DaliLite server [30], to compare their active-site structure and topology, and to obtain RMSD data. RMSD data reported are for alignments of the monomeric structures alone, and typically aligned 300-451 residues. These structure alignments were also used to generate sequence alignments, which were then refined manually using the visualised structures as guidance, particularly in active site regions, to obtain accurate sequence identity values as well as a reliable table of amino acid substitutions at key active-site residues. This refinement was particularly necessary for structure pairs with higher RMSD values.

Phylogenetic analysis of the transaminase structure-sequence alignment

The 29 aligned sequences from the TAm structures were used to obtain a Maximum Likelihood tree phylogeny with the PHYLIP package program ProML [44]. Neighbour-joining (NJ) tree phylogenies were also obtained using Protdist via BioEdit [45] for comparison. Identical results were obtained with each method, and the NJ tree was retained.

Multiple sequence alignments for CV2025 and SPAT

Multiple protein sequence alignments were obtained with compositional matrix adjusted identities of >40% using sequence searches from NCBI-BLASTP [32] and both CV2025 TAm and SPAT as the query sequences, and multiple sequence alignment with ClustalX and the BLOSUM62 alignment matrix [46]. The initial BLASTP searches were set to yield 1000 and 500 sequences related to CV2025 and SPAT, respectively. Only one sequence from any single species was accepted, taking the one with the lowest sequence identity to maximise diversity, giving 100 and 226 final sequences related to CV2025 and SPAT, respectively, for the multiple sequence alignment. Each alignment was manually improved by eye in BioEdit.

In-silico active site mapping

Cavities within the CV2025 TAm structure and the immediately surrounding residues were mapped using CAVER [33] and default parameters, with the crystal structure of the enzyme in the holo form with PLP forming a Schiff's base to K288 (PDB ID: 4AH3, provided in advance by personal communication with Prof Jennifer Littlechild) [25]. The same sites were reconfirmed using another published structure (PDB ID: 4A6U) [23]. Surrounding residues were then identified by visual inspection of the structure overlaid with the highlighted cavities in Pymol [31]. Active-site residues that could potentially contact substrates were identified by independently docking MBA, acetophenone, serine, pyruvate, alanine and hydroxypyruvate into the CV2025 TAm structure, using AutoDock software version 4.0 [34]. Substrate docking models were obtained in the active site, using a cubic grid centred at (-24.896 -17.433 -48.359) with sides of 80 Å to cover the active-site cavities identified above. Defaults were used for docking each substrate except for the following: the maximum number of energy evaluations was increased to 1 million, the number of genetic algorithm (GA) runs was increased from 10 to 200, and the grid spacing used was 0.375 Å. Cluster analysis by AutoDock on each final conformation obtained from the 200 GA runs grouped those with an RMSD of less than 0.5 Å. Clusters were then ranked in order of increasing energy. Visual analysis of docked conformations was carried out in Pymol. All amino acids found to interact with at least one substrate were selected for saturation mutagenesis.

Sequence conservation and entropy

Protein sequence conservation at each site in CV2025 TAm was determined from the multiple protein sequence alignment of 100 sequences with compositional matrix adjusted identities of >40%. Excess sequence beyond the N and C-terminal ends of the CV2025 query sequence, or gaps within CV2025 sequence, were deleted. Sequence entropy, H(x), was calculated at each site using BioEdit, where the maximum entropy for 21 possible amino acids (including stop codon) is 3.04 and zero represents a fully conserved residue.

Saturation mutagenesis and screening of ω-Transaminase mutants

CV2025 ω-Transaminase was expressed throughout upon IPTG induction, from plasmid pQR801 [5] transformed into *E. coli* BL21(DE3) (Stratagene, Netherlands). Saturation site-directed mutagenesis was performed using Quikchange kit (Stratagene, Netherlands) according to the manufacturers instructions, and mutated plasmids retransformed into chemically competent *E. coli* BL21(DE3) cells. Saturation mutagenesis primers used, along with their reverse complement sequences, were as follows:

P21 CCATCACCTGCAT**NNK**TTCACCGATACC
F22 CCG**NNK**ACCGATACCGCATCGCTGAACC
L59 CGACGGCATGGCCGGA**NNK**TGGTGCGTGAACGTCG
W60 GGCATGGCCGGACTG**NNK**TGCGTGAACGTCGG
F88 CCGTTCTACAACACC**NNK**TTCAAGACCACC
F89 CTACAACACCTTC**NNK**AAGACCACCCATCC
Y153 GCTGGAACGGC**NNK**CACGGCTCC
Y168 GCATGAAG**NNK**ATGCACGAGCAGG
A231 GCGAACCCATCCAGGGC**NNK**GGCGGCGTGATCGTCC
G319 CGACTTCAACCAC**NNK**TTCACCTACTCC
F320 CGACTTCAACCACGGC**NNK**ACCTACTCC
T321 CGACTTCAACCACGGCTTC**NNK**TACTCCGGCCACCCGGT
Y322 CGACTTCAACCACGGCTTCACC**NNK**TCCGGCCACCCGGT
R416 CCTGATCATG**NNK**GCATGCGGCGACC

For each library 76 colonies were picked and inoculated into wells of a 96 deep-square well (DSW) plate containing 900 µl of terrific broth (TB) with 100 µg/ml kanamycin. Each 96-well plate contained twelve negative control wells of media only, and eight wells with *E. coli* BL21(DE3):pQR801 expressing wild-type CV2025 ω-Transaminase, of which four each were used as reaction positive and negative controls respectively. Microplates were covered with an inverted DSW plate, sealed, and incubated at 37 °C, 300 rpm, 50mm orbital throw, 85% humidity for 16 hours to stationary phase in an incubated shaker (INFORS HT Multitron, Bottmingen Switzerland). Liquid handling was performed with a Genesis robot (Tecan, Reading, UK) to minimise errors. From each well, 100 µl culture was inoculated into fresh 900 µl terrific broth with 100 µg/ml kanamycin, grown to OD_{600} 2.0 under the same conditions, then induced with 0.89 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG), and the temperature decreased to 32 °C for 5 hours. Cells were harvested by plate centrifugation at 4300 rpm (3102g) for 15 minutes, the medium discarded and cells resuspended in 1 ml 0.1 M potassium phosphate, pH 7.0, 0.2 mM pyridoxal phosphate (PLP), by shaking for 10 min at 37 °C, 300 rpm, 50 mm orbital throw, 85% humidity. Centrifugation and resuspension was repeated once, then the plates frozen and stored at -80 °C for up to 6 months. At least six randomly picked colonies from each residue library were sequenced to assess the diversity of the libraries. In addition, the top performing colonies from each library were sequenced, even if the activity did not improve above wild-type. Finally, the Y153X library was sequenced completely.

Reaction plates were created by transferring 30 µl of each thawed cell suspension into standard 96-well plates containing 225 µl lysis buffer (0.2 mg/ml lysozyme, 0.2 mM PLP, 0.1 M potassium phosphate pH 7.0, 10% bug buster (Life Technologies, Paisley, UK). Cells were resuspended for 10 minutes in a Thermomixer Comfort microshaker (Eppendorf, Hauppauge, NY) at 21 °C, 1000 rpm, 3 mm orbital throw, then 30 minutes at 37 °C, 300 rpm, 50 mm orbital throw, 85% humidity. Lysed cells were centrifuged for 60 minutes at 4300 rpm (3102 g) at 4 °C then 55 µl of the clarified lysate transferred into standard 96-well plates, stored at 4 °C, and used for reaction screens within 6 hours. Expression levels within clarified lysates were assessed by SDS-PAGE and densitometry. Reaction plates were equilibrated to 37 °C, and enzyme assays initiated by adding 145 µl 0.1 M potassium phosphate, pH 7.0, 0.2 mM PLP, and 13.8 mM substrates to give 0.2 mM PLP, 10 mM serine and 10 mM of pyruvate. Reaction plates were sealed and shaken at 37 °C, for 3 hours at 50 rpm, then 10 µl transferred to another plate containing 190 µl of 0.125 mM WST-1 tetrazolium salt dissolved in 0.11 M NaOH, shaken at 50 rpm for 10 minutes and the absorbance at 595 nm measured in a platereader. Reaction with the hydroxypyruvate product leads to an increase in blue colour.

Enzyme mutant purification and detailed kinetic analysis

Wild-type enzyme and two promising mutants Y153M and Y153S were overexpressed and purified using Sepharose fast flow Ni-NTA resin (GE Healthcare, Little Chalfont, UK). Eluted samples were buffer exchanged using PD-10 desalting columns (GE Healthcare, Little Chalfont, UK) into 0.1 M potassium phosphate, pH 7.0, 0.2 mM PLP. Serine/pyruvate reactions were performed in triplicate with 10 mM pyruvate, and 5-200 mM L-serine, using 0.3 mg ml^{-1} wild-type enzyme or 0.03 mg ml^{-1} mutant enzymes at 37 °C. The hydroxypyruvate product was determined by sampling at regular intervals with an Aminex HPX-87H ion exchange HPLC column (300 mm x 7.8 mm; Bio-Rad, Hemel Hempstead, UK) as described previously [5]. Reactions with (*S*)-MBA or aminoindane were performed with 10 mM pyruvate, and 0.15-50 mM (*S*)-MBA or racemic aminoindane, using 0.4 mg ml^{-1} enzyme at 37 °C. Formation of acetophenone from MBA or 1-indanone from aminoindane was followed by RP-HPLC using an ACE 5 C18 reverse phase column (150 mm × 4.6 mm, 5 µm particle size; Hichrom Limited, Berkshire, UK). A gradient was run from 15% acetonitrile / 85% 0.1% (v/v) trifluoroacetic acid (TFA) to 72% acetonitrile / 28% TFA over

8 min, followed by a re-equilibration step for 2 min (oven temperature 30 °C, flow rate 1 mL/min). UV detection was carried out at 250 nm. No substrate inhibition was apparent. Kinetic parameters were obtained by fitting to the standard Michaelis-Menten equation by linear regression.

Acknowledgments

DD was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) (grant ref BB/G016593/1). The crystal structure coordinates of the CV2025 ω-transaminase were kindly provided by Prof Jennifer Littlechild, Exeter University, UK.

Author contributions

D. Deszcz designed, created and screened the enzyme libraries, purified and characterised selected variants, and performed the structural modeling. P. Affaticati and A. Gegel helped screening of the enzyme libraries, N. Ladkau carried out enzyme kinetic and expression analyses, J.M. Ward and H. C. Hailes helped plan, oversee and supervise the enzyme characterisations. P.A. Dalby conceived the project and carried out the structural and sequence alignments. All authors wrote the manuscript.

References

1. Midelfort KS, Kumar R, Han S, Karmilowicz MJ, McConnell K, Gehlhaar DK, Mistry A, Chang JS, Anderson M, Villalobos A, Minshull J, Govindarajan S & Wong JW (2013) Redesigning and characterizing the substrate specificity and activity of *Vibrio fluvialis* aminotransferase for the synthesis of imagabalin. *Prot Eng Des Sel* **26**, 25–33.

2. Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, Devine PN, Huisman GW & Hughes, GJ (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309.

3. Sehl T, Hailes HC, Ward JM, Menyes U, Pohl M & Rother D (2014) Efficient 2-step biocatalytic strategies for the synthesis of all nor(pseudo)ephedrine isomers. *Green Chem* **16**, 3341, doi:10.1039/c4gc00100a

4. Ingram CU, Bommer M, Dalby PA, Ward JM, Hailes HC & Lye GJ (2007) One-pot synthesis of amino-alcohols using a de-novo transketolase and beta-alanine:pyruvate transaminase pathway in *Escherichia coli*. *Biotechnol Bioeng* **96**, 559–569.

5. Kaulmann U, Smithies K, Smith MEB, Hailes HC & Ward JM (2007) Substrate spectrum of ω-transaminase from *Chromobacterium violaceum* DSM30191 and its potential for biocatalysis. *Enz Microb Technol* **41**, 628–637.

6. Smithies K, Smith MEB, Kaulmann U, Galman JL, Ward JM & Hailes HC (2009) Stereoselectivity of an ω-transaminase-mediated amination of 1,3-dihydroxy-1-phenylpropane-2-one. *Tet: Asymm* **20**, 570–574.

7. Smith MEB, Chen BH, Hibbert EG, Kaulmann U, Smithies K, Galman JL, Baganz F, Dalby PA, Hailes HC, Lye GJ, Ward JM, Woodley JM & Micheletti M (2010) A multidisciplinary approach toward the rapid and preparative-scale biocatalytic synthesis of chiral amino alcohols: A concise transketolase-/ω-transaminase-mediated synthesis of (2S, 3S)-2-aminopentane-1,3-diol. *Org Proc Res Dev* **14**, 99–107.

8. Rios-Solis L, Halim M, Cázares A, Morris P, Ward JM, Hailes HC, Dalby PA, Baganz F & Lye GJ (2011) A toolbox approach for the rapid evaluation of multi-step enzymatic syntheses comprising a “mix and match” *E. coli* expression system with microscale experimentation. *Biocatal Biotrans* **29**, 192–203.

9. Chen BH, Sayar A, Kaulmann U, Dalby PA, Ward JM & Woodley JM (2006) Reaction modeling and simulation to assess the integrated use of transketolase and ω-transaminase for the synthesis of an aminotriol. *Biocatal Biotrans* **24**, 449–457.

10. Truppo MD, Rozzell JD, Moore JC & Turner NJ (2009) Rapid screening and scale-up of transaminase catalysed reactions. *Org Biomol Chem* **7**, 395–398.

11. Truppo MD, Rozzell JD & Turner NJ (2010) Efficient production of enantiomerically pure chiral amines at concentrations of 50 g / L using transaminases. *Org Proc Res Devel* **14**, 234–237.

12. Cassimjee KE, Branneby C, Abedi V, Wells A & Berglund P (2010) Transaminations with isopropyl amine: equilibrium displacement with yeast alcohol dehydrogenase coupled to in situ cofactor regeneration. *Chem Comm* **46**, 5569–5571.

13. Halim M, Rios-Solis L, Micheletti M, Ward JM & Lye GJ (2014) Microscale methods to rapidly evaluate bioprocess options for increasing bioconversion yields: application to the ω-transaminase synthesis of chiral amines. *Bioproc Biosys Eng* **37**, 931–941.

14. Grishin NV, Phillips MA & Goldsmith EJ (1995) Modeling of the spatial structure of eukaryotic ornithine decarboxylases. *Prot Science* **4**, 1291–304.

15. Höhne M, Schätzle S, Jochens H, Robins K & Bornscheuer UT (2010) Rational assignment of key motifs for function guides in silico enzyme identification. *Nature Chem Biol* **6**, 807–813.

16. Sayer C, Martinez-Torres RJ, Richter N, Isupov M, Hailes HC, Littlechild J & Ward JM (2014) The substrate specificity, enantioselectivity and structure of the (R)-selective amine : pyruvate transaminase from *Nectria haematococca FEBS J* **281**, 2240–2253.

17. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H-R, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL & Bateman A (2008) The Pfam protein families database. *Nucl Acids Res* **36**, 281–288.

18. Mehta PK, Hale TI & Christen P (1993) Aminotransferases: demonstration of homology and division into evolutionary subgroups. *Eur J Biochem* **214**, 549–561.

19. Burnett G & Walsh C (1979) Stereospecificity of enzymatic transamination of gamma-aminobutyrate. *Chem Comm* **19**, 826–828.

20. Yonaha K, Suzuki K, Minei H & Toyama S (1983) Distribution of omega-amino acid - pyruvate transaminase and aminobutyrate - alpha-ketoglutarate transaminase in microorganisms. *Agric Biol Chem* **47**, 2257–2265.

21. Samsonova NN, Smirnov SV, Altman IB & Ptitsyn LR (2003) Molecular cloning and characterization of *Escherichia coli* K12 *ygj*G gene. *BMC Microbiol* **3**, 1–10.

22. Ward JM & Wohlgemuth R (2010) High-yield biocatalytic amination reactions in organic synthesis. *Curr Org Chem* **14**, 1914-1927.

23. Humble MS, Cassimjee KE, Håkansson M, Kimbung YR, Walse B, Abedi V, Federsel HJ, Berglund P & Logan, DT (2012) Crystal structures of the *Chromobacterium violaceum* ω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP. *FEBS J* **279**, 779–792.

24. Sayer C, Bommer M, Isupov M, Ward J & Littlechild J (2012) Crystal structure and substrate specificity of the thermophilic serine:pyruvate aminotransferase from *Sulfolobus solfataricus*. *Acta Crystallog. Sect D* **68**, 763–772.

25. Sayer C, Isupov M, Westlake A & Littlechild J (2013) Structural studies of *Pseudomonas* and *Chromobacterium* ω-aminotransferases provide insights into their differing substrate specificity *Acta Crystallog. Sect D* **69**, 564–576.

26. Martin AR, DiSanto R, Plotnikov I, Kamat S, Shonnard D & Pannuri S (2007) Improved activity and thermostability of (S)-aminotransferase by error-prone polymerase chain reaction for the production of a chiral amine. *Biochem Eng J* **37**, 246–255.

27. Bum-Yeol H, Ko S-H, Park H-Y, Seo J-H, Lee B-S & Kim B-G (2008) Identification of ω-aminotransferase from *Caulobacter crescentus* and site-directed mutagenesis to broaden substrate specificity. *J Microbiol Biotechnol* **18**, 48–54.

28. Svedendahl M, Branneby C, Lindberg L & Berglund P (2010) Reversed enantiopreference of an ω-transaminase by a single-point mutation. *ChemCatChem* **2**, 976–980.

29. Cho B, Park H, Seo J, Kim J, Kang T, Lee B & Kim B (2008) Redesigning the substrate specificity of ω-aminotransferase for the kinetic resolution of aliphatic chiral amines. *Biotechnol Bioeng* **99**, 275–284.

30. Holm L & Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**, 566–567.

31. DeLano WL (2002) The PyMOL molecular graphics system. Available at: http://www.pymol.org.

32. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* **25**, 3389–3402.

33. Kozlikova B, Sebestova E, Sustr V, Brezovsky J, Strnad O, Daniel L, Bednar D, Pavelka A, Manak M, Bezedka M, Benes P, Kotry M, Gora A, Damborsky J & Sochor J (2014) CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* **30**, 2684-2685.

34. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS & Olson AJ (2009) AutoDock4 and AutoDockTools4 : Automated docking with selective receptor flexibility. *J Comp Chem* **30**, 2785–2791.

35. Firth AE & Patrick WM (2005) Statistics of protein library construction. *Bioinformatics* **21**, 3314–3315.

36. Paramesvaran J, Hibbert EG, Russell AJ & Dalby PA (2009) Distributions of enzyme residues yielding mutants with improved substrate specificities from two different directed evolution strategies. *Prot Eng Des Sel* **22**, 401–411.

37. Hibbert EG, Senussi T, Costelloe SJ, Lei W, Smith MEB, Ward JM, Hailes HC & Dalby PA (2007) Directed evolution of transketolase activity on non-phosphorylated substrates. *J Biotechnol* **131**, 425–432.

38. Parker BM, Taylor IN, Woodley JM, Ward JM & Dalby PA (2011) Directed evolution of a thermostable L-aminoacylase biocatalyst. *J Biotechnol* **155**, 396–405.

39. Hibbert EG, Senussi T, Smith MEB, Costelloe SJ, Ward JM, Hailes HC & Dalby PA (2008) Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. *J Biotechnol* 134, 240–5.

40. Nobili A, Steffen-Munsberg F, Kohls H, Trentin I, Schulzke C, Höhne M & Bornscheuer UT (2014) Engineering the active site of the amine-transaminase from Vibrio fluvialis for the asymmetric synthesis of aryl-alkyl amines and amino alcohols. *ChemCatChem*, in press.

41. Humble MS, Cassimjee KE, Abedi V, Federsel HJ & Berglund P (2012) Key amino acid residues for reversed or improved enantiospecificity of an ω-transaminase. *ChemCatChem* **4**, 1167–1172.

42. Hecquet L, Bolte J & Demuynck C (1996) Enzymatic synthesis of “natural-labeled” 6-deoxy-L-sorbose precursor of an important food flavor. *Tetrahedron* **52**, 8223–8232.

43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE (2000) The Protein Data Bank. *Nucl Acids Res* **28**, 235–242.

44. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package), version 35c. Department of Genetics, University of Washington, Seattle

45. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Series* **41**, 95–98.

46. Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res* **22**, 4673–4680.

Supplementary Information

		Substrate						
ChainA	**ChainB**	**HPA**	**MBA**	**Acetophenone**	**Serine**	**Pyr1**	**Pyr2**	**Alanine**
P21			HF					
F22			HF	HF				
L59		HF	HF	HF	HF			
W60				HF			PiP	
	F88		HF					
Y153			HB/HF	HF/N	HF/N			PiP
Y168						PiP		
A231		HF	HF	HF	HF			
	G319		HF	HF/N				
	F320		HF					
	T321	HB/N	HF/N	HF	HF/N			
R416		N		HF	HB	N		N

Table S1.

Interactions of computationally docked substrates with active-site residues.

Interactions denoted are: (N) with amino acid backbone nitrogen; (HF) hydrophobic; (HB) hydrogen bond; (PiP) polar-pi interaction. All substrates docked into a single cluster with the lowest binding energy, with the exception of pyruvate which clustered into two major groups denoted Pyr1 and Pyr2.

Tables

Residue		H(x)	Amino-acid frequency (%)													Residue
			A	F	G	H	L	N	P	R	S	T	W	Y	~	in SPAT [a]
P	21	0.7	6	0	1	0	0	0	81	0	10	0	0	0	2	-
F	22	1.1	0	64	1	0	1	0	0	0	9	0	0	20	2	-
L	59	0.0	0	0	0	0	100	0	0	0	0	0	0	0	0	V8
W	60	0.4	0	2	0	0	0	0	0	0	0	0	88	10	0	G9
F	88	0.3	0	90	0	0	0	0	0	0	0	0	0	10	0	F28
F	89	1.1	5	69	2	0	0	1	0	0	10	6	0	0	0	S30
Y	153	0.0	0	0	0	0	0	0	0	0	0	0	0	100	0	F88
Y	168	2.1	13	10	27	3	0	13	3	0	6	0	1	16	0	-
A	231	0.4	88	0	0	0	0	0	0	0	3	9	0	0	0	S139
G	319	0.0	0	0	100	0	0	0	0	0	0	0	0	0	0	-
F	320	1.4	0	38	3	0	6	0	0	0	0	0	10	39	0	F241
T	321	0.0	0	0	0	0	0	0	0	0	0	100	0	0	0	A242
Y	322	0.6	12	1	3	0	0	0	0	0	0	0	1	83	0	T243
R	416	0.9	0	0	0	8	0	0	0	69	0	1	0	0	21	A326

Table 1.

Amino-acid sequence entropy, H(x), and positional frequency at targeted CV2025 active-site residues.

Values derived from 100 ωTAm sequences with at least 40% sequence identity to CV2025. Only amino acids with the highest positional frequency are shown. Residues in monomer A are underlined, the rest are in monomer B. [a] "-" no equivalent residue in SPAT. Residue numbering for SPAT is from its UniProt accession number Q97VM5.

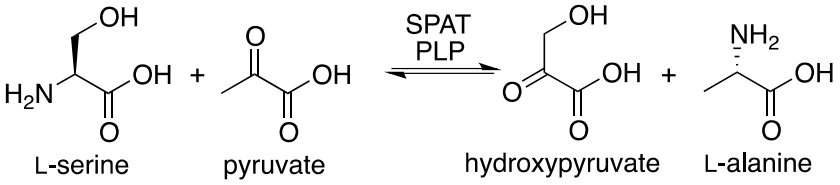
	Y153S	Y153M	WT
Serine			
K_m **(mM)**	13 (1)	4.6 (0.4)	260 (40)
k_{cat} **(s^{-1})**	0.40 (0.01)	0.59 (0.01)	0.51 (0.05)
k_{cat}/K_m **(M^{-1} s^{-1})**	30 (3)	130 (10)	1.9 (0.3)
rac-Aminoindane			
K_m **(mM)**	0.43 (0.04)	10 (2.5)	15 (1)
k_{cat} **(s^{-1})**	0.05 (0.001)	8.5 (0.8)	8 (1)
k_{cat}/K_m **(M^{-1} s^{-1})**	120 (13)	850 (200)	550 (90)
(*S*)-MBA			
K_m **(mM)**	8.7 (0.3)	11 (3)	28 (5)
k_{cat} **(s^{-1})**	1.2 (0.1)	7.5 (0.5)	8 (1.7)
k_{cat}/K_m **(M^{-1} s^{-1})**	140 (10)	660 (200)	280 (80)

Table 2.

Kinetic parameters for wild-type CV2025 transaminase and two mutants, Y153S and Y153M.

K_m is for the amine donor indicated in each case. Standard deviations are given in parentheses. The aminoindane used was racemic.

Schema



Scheme 1

Figures

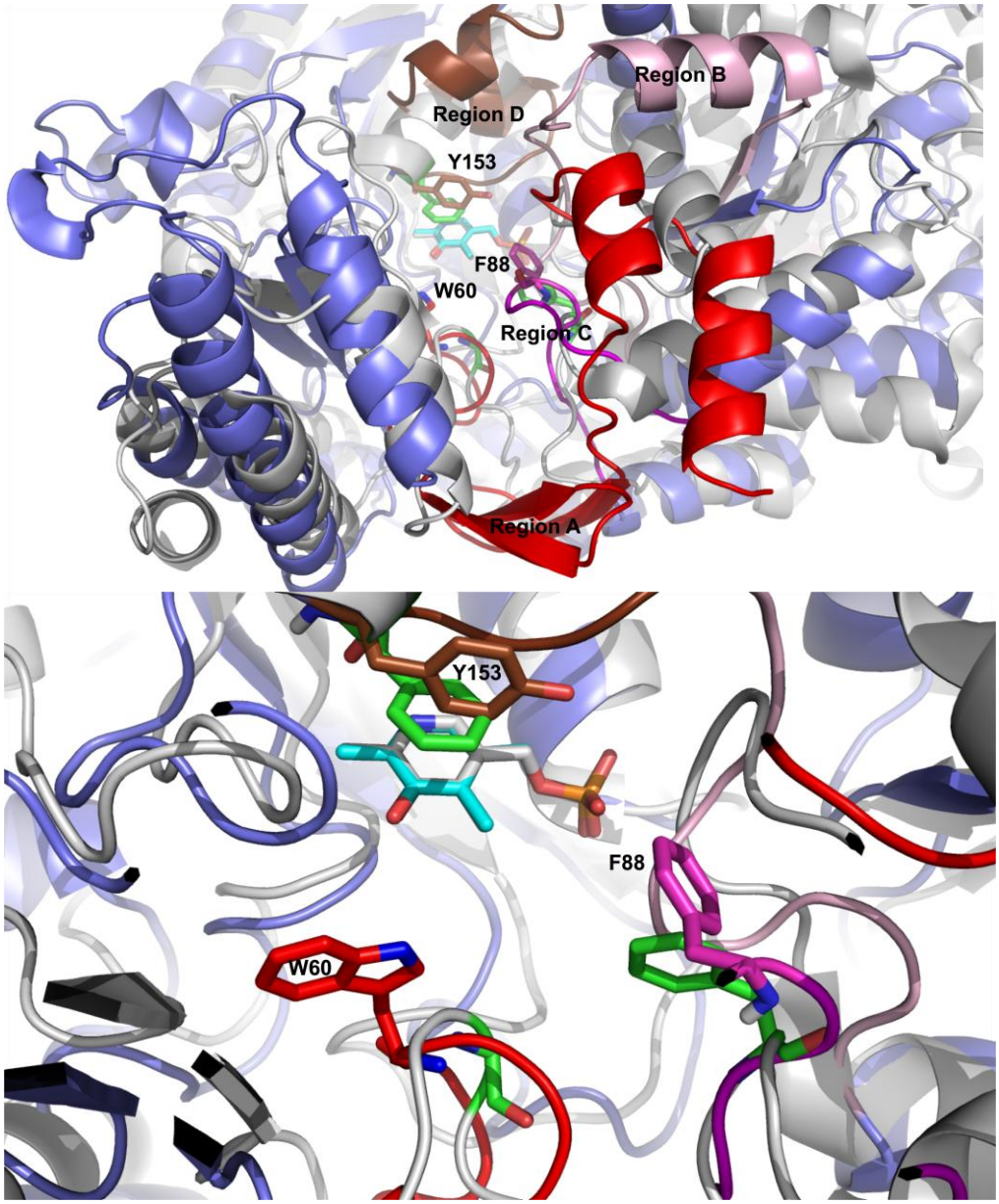


Figure 1.

Structurally aligned active-sites of *CV2025* ω-transaminase (blue) and the *Sulfolobus solfataricus* Serine:Pyruvate aminotransferase (SPAT) (grey).

CV2025 ω-transaminase (PDB ID: 4A6U) (Humble *et al*. 2012) and SPAT (PDB ID: 3ZRP) (Sayer et al 2012) share 11% sequence identity, and aligned overall with DaliLite (Holm and Park 2000) to an RMSD of 3.4 Å. Image A: Alignment of just one active-site, weighted on the PLP cofactor, is shown with active-site regions of CV2025 deviated significantly to SPAT, in red (region A), pink (region B), magenta (region C), and brown (region D). PLP is shown as cyan sticks. Image B: close up of W60, F88 and Y153 in CV2025 shown in sticks, with G9, F28 and F88 in SPAT shown in green sticks. W60, F88 and Y153 (CV2025) align to G9, F28 and F88 (SPAT), respectively. Images generated in Pymol (Delano 2002).

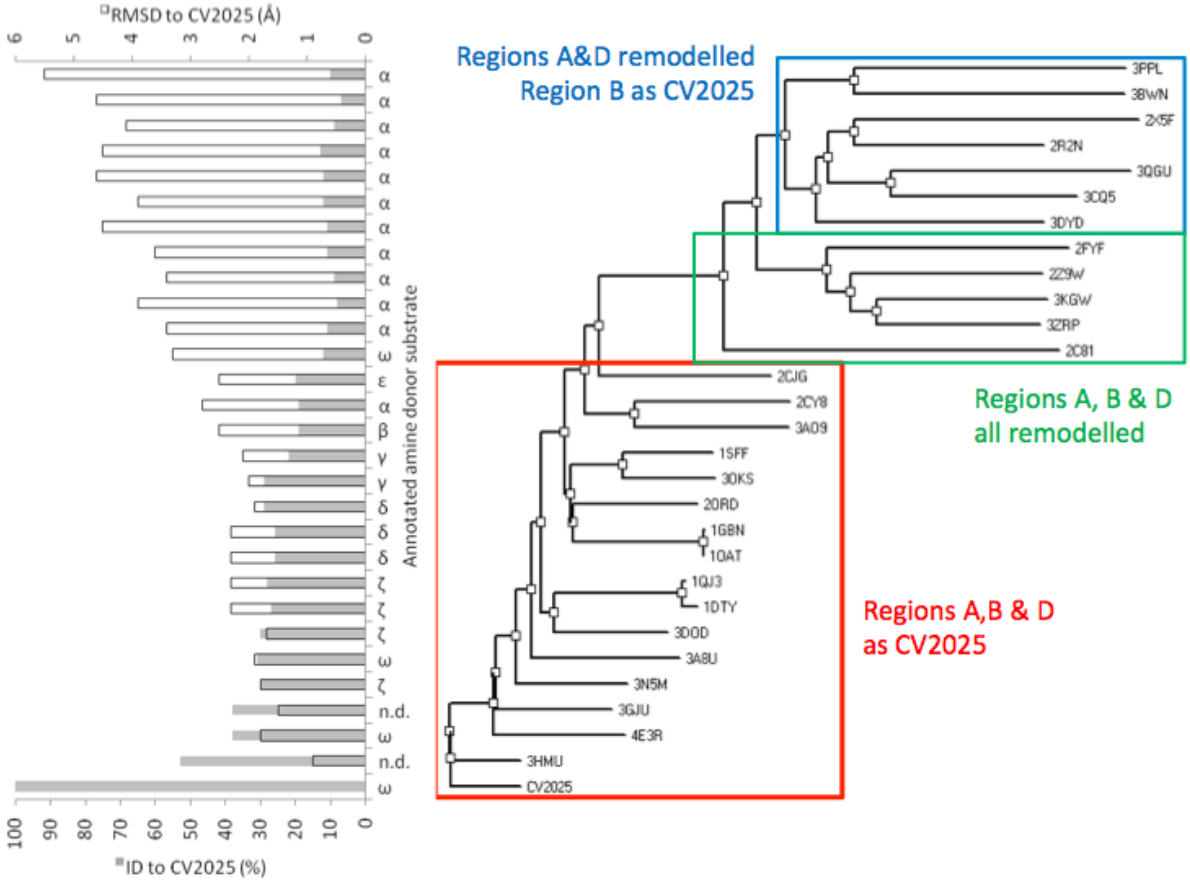


Figure 2.

Phylogenetic analysis of transaminase structures.

The sequences from 29 transaminase structures, refined by structure alignment, were used to construct a Neighbour-Joining tree. PDB identifiers are shown on the tree. The tree shown is aligned to the graph on the left which indicates the sequence identity (ID) and RMSD to CV2025 in each case. Also shown is the position (α-ω, n.d. = not determined) of the amine relative to the carboxylate moiety of the primary substrates associated with each structure, as annotated in the PDB and related literature. Transaminases cluster primarily into αTAm and ωTAm clades. The αTAm clade can be split into two sub-clades in which structural region B is either similar to that in CV2025 or remodeled. SPAT is denoted as 3ZRP.

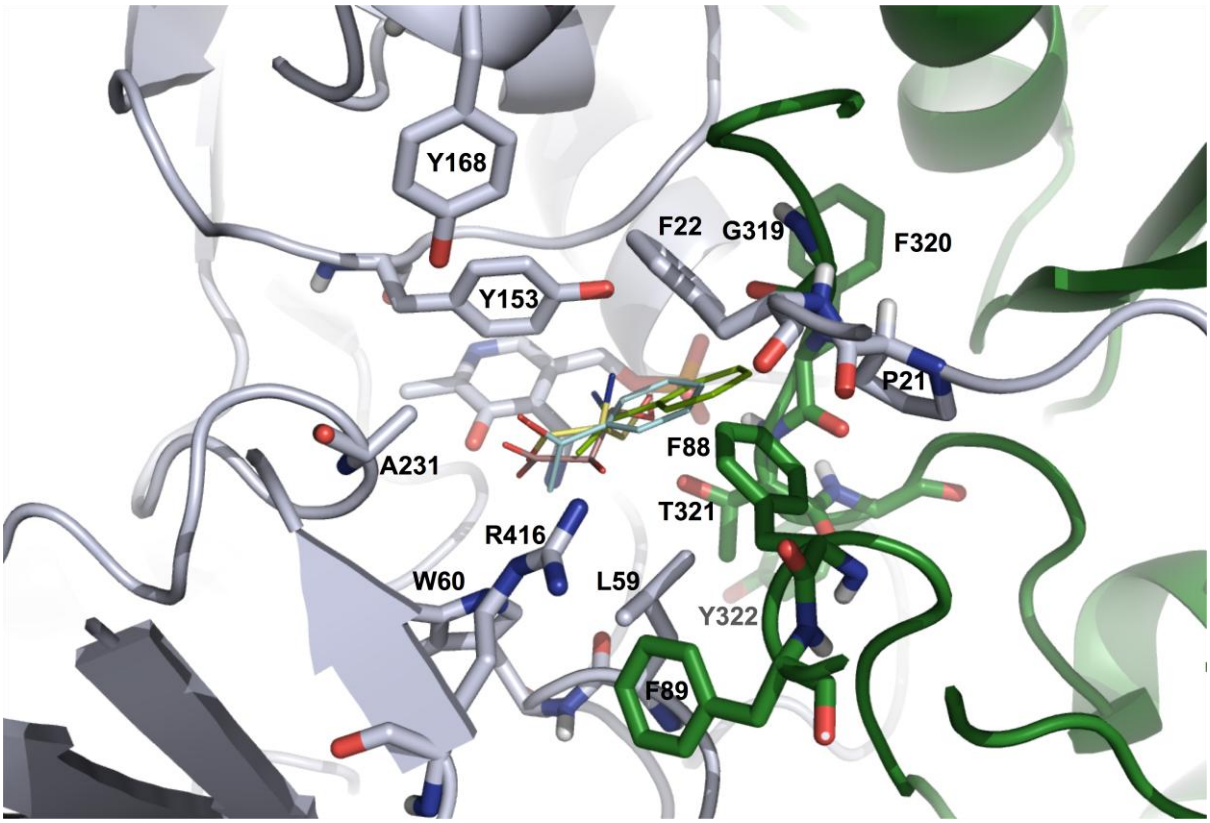


Figure 3.
Residues targeted for independent site-directed saturation mutagenesis.
Each monomer forming the active site is coloured separately (green and grey). Target residues, and the pyridoxamine cofactor are highlighted in CPK sticks. Four of the substrates docked with AUTODOCK into holo-CV2025 (PDB ID: 4A6U), are also show in thick lines: MBA (green); acetophenone (light blue); serine (yellow); HPA (purple). Image generated in Pymol (Delano 2002).

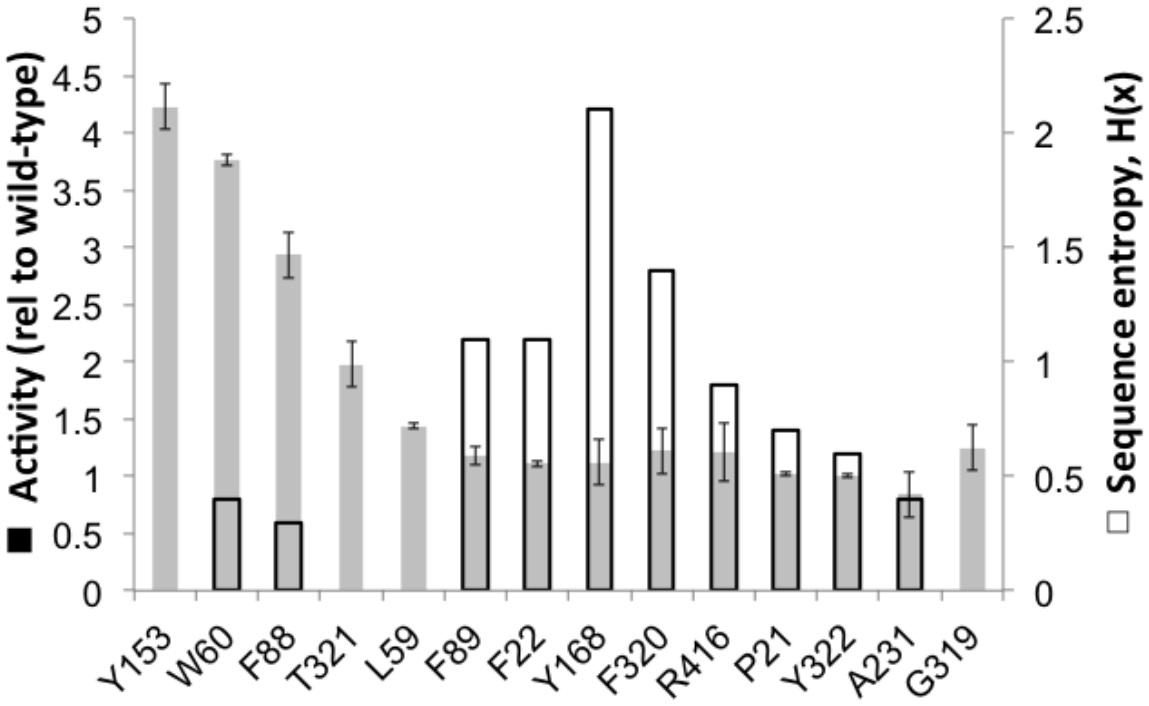
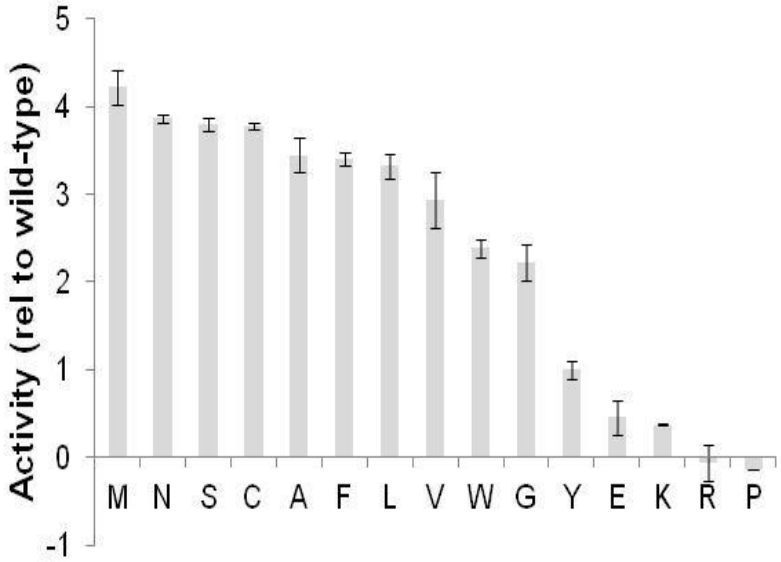


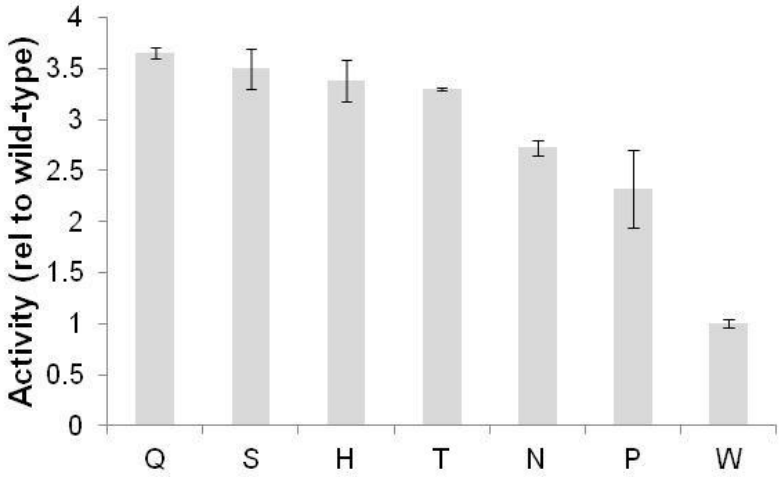
Figure 4.

Comparison of maximum activity improvements with the natural sequence entropies at active-site residues subjected to saturation mutagenesis.

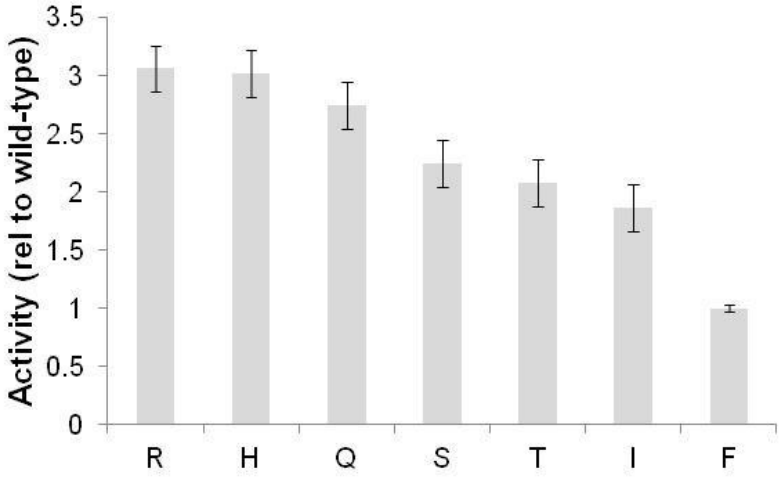
Maximum total activities relative to wild-type controls. Positional natural sequence entropies from an alignment of 100 ωTAm sequences with at least 40% identity to CV2025. The first seven sites are ordered by decreasing maximum activity, where mutants with activities higher than wild-type were identified. For the remainder (Y168 onwards), maximum activities were from wild-type transformants as determined by DNA sequencing, and so these were plotted simply in order of H(x). Error bars are standard deviations of the mean, except for uniquely sequenced mutants, for which a conservative standard error of 0.2 was assumed. Activities were measured in crude lysates towards 10 mM serine and 10 mM pyruvate, 0.2 mM PLP, in 0.1 M potassium phosphate, pH 7.0 at 37 °C.


Activity (rel to wild-type)
5
4
3
2
1
0
-1
M N S C A F L V W G Y E K R P

A


Activity (rel to wild-type)
4
3.5
3
2.5
2
1.5
1
0.5
0
Q S H T N P W

B


Activity (rel to wild-type)
3.5
3
2.5
2
1.5
1
0.5
0
R H Q S T I F

C

Figure 5.

Distribution of total activities for A) Y153X, B) W60X and C) F88X mutants, relative to wild-type.
Error bars are standard deviations of the mean, except for uniquely sequenced mutants, for which a conservative standard error of 0.2 was assumed. Activities were measured in crude lysates towards 10 mM serine and 10 mM pyruvate, 0.2 mM PLP, in 0.1 M potassium phosphate, pH 7.0 at 37 °C.