the FEBS Journal

Substrate scope and selectivity in offspring to an enzyme subjected to directed evolution

Cecilia Blikstad[1], Käthe M. Dahlström[2], Tiina A. Salminen[2] and Mikael Widersten[1]

1 Department of Chemistry – BMC, Uppsala University, Sweden
2 Structural Bioinformatics Laboratory, Åbo Akademi University, Turku, Finland

Keywords
directed evolution; kinetic characterization; oxidoreductase; structure modeling; vicinal diol

Correspondence
M. Widersten, Department of Chemistry – BMC, Uppsala University, Box 576, SE-751 23 Uppsala, Sweden
Fax: +46 18 558431
Tel: +46 18 471 4992
E-mail: mikael.widersten@kemi.uu.se

(Received 20 December 2013, revised 12 March 2014, accepted 24 March 2014)

doi:10.1111/febs.12791

We have analyzed the effects of mutations inserted during directed evolution of a specialized enzyme, *Escherichia coli* *S*-1,2-propanediol oxidoreductase (FucO). The kinetic properties of evolved variants have been determined and the observed differences have been rationalized by modeling the tertiary structures of isolated variants and the wild-type enzyme. The native substrate, *S*-1,2-propanediol, as well as phenylacetaldehyde and 2*S*-3-phenylpropane-1,2-diol, which are new substrates accepted by isolated variants, were docked into the active sites. The study provides a comprehensive picture of how acquired catalytic properties have arisen via an intermediate generalist enzyme, which had acquired a single mutation (L259V) in the active site. Further mutagenesis of this generalist resulted in a new specialist catalyst. We have also been able to relate the native enzyme activities to the evolved ones and linked the differences to individual amino acid residues important for activity and selectivity. F254 plays a dual role in the enzyme function. First, mutation of F254 into an isoleucine weakens the interactions with the coenzyme thereby increasing its dissociation rate from the active site and resulting in a four-fold increase in turnover number with *S*-1,2-propanediol. Second, F254 is directly involved in binding of aryl-substituted substrates via π–π interactions. On the other hand, N151 is critical in determining the substrate scope since the side chain amide group stabilizes binding of 1,2-substituted diols and is apparently necessary for enzymatic activity with these substrates. Moreover, the side chain of N151 introduces steric hindrance, which prevents high activity with phenylacetaldehyde. Additionally, the hydroxyl group of T149 is required to maintain the catalytically important hydrogen bonding network.

Introduction

Propanediol oxidoreductase (FucO), a homodimeric class III alcohol dehydrogenase [1] from *Escherichia coli*, is a component of the catabolic pathway of fucose and rhamnose where it catalyzes the reduction of *S*-lactaldehyde into *S*-1,2-propanediol (*S*-**2** in Fig. 1) [2–4]. FucO is highly specialized to accept aliphatic primary alcohols of low molecular weight as substrates and displays very low or no activity with aryl-substituted or other bulky substrates [5]. FucO strongly prefers 2-*S*-enantiomers and is strictly regioselective towards primary alcohols. Due to the high degree of stereoselectivity the enzyme is interesting as a potential biocatalyst for production of α-hydroxyaldehydes and other chiral aldehydes and alcohols.

We had earlier concluded from a survey of the substrate range of FucO [5] that unbranched aliphatic

Abbreviations
Adh3, *Oenococcus oeni* alcohol dehydrogenase; FucO, *S*-1,2-propanediol oxidoreductase.

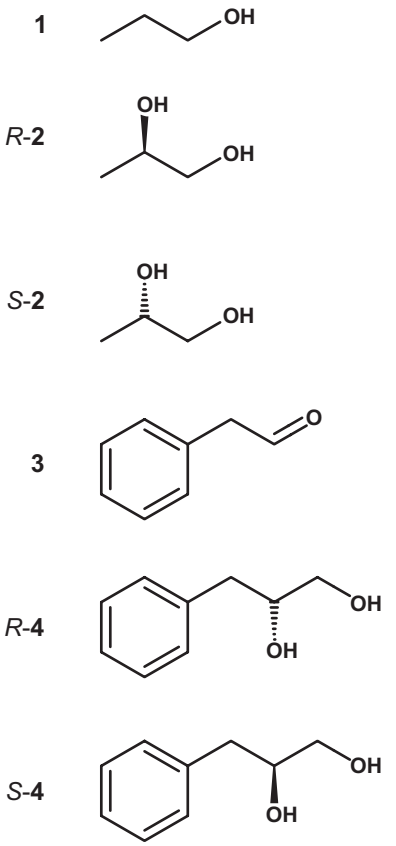

Fig. 1. Compounds tested as substrates in this study. Alcohols **1** and *S*-**2** are efficiently oxidized by the wild-type FucO enzyme whereas aldehyde **3** is a very poor substrate and diols **4** are not transformed at all.

mono-, di- and tri-substituted alcohols up to chain lengths of four carbons are efficiently oxidized if k_{cat} is applied as a measure of enzyme efficiency. k_{cat}/K_M, however, decreases drastically if substrate chain length is changed from three carbons to two or four carbons, illustrating the evolutionary adaptation of this enzyme to act on **2** or the corresponding lactaldehyde. 2*S*-3-phenylpropane-1,2-diol (**4**) is the 3-phenyl-substituted derivative of *S*-**2** but is not transformed at detectable levels by this enzyme. The limited substrate scope of FucO can be linked to the narrow entrance to the active site [6], whose structural design effectively prevents productive binding of larger substrates.

In order to expand the range of accepted substrates to include also aryl-substituted compounds, we recently subjected FucO to iterative saturation mutagenesis driven [7,8] directed evolution. Mutation sites were chosen using the CASTing method [9] and the constructed variant libraries maintained proteins in which residues lining the active-site entrance had been randomized (Fig. 2) [10]. Variants displaying novel

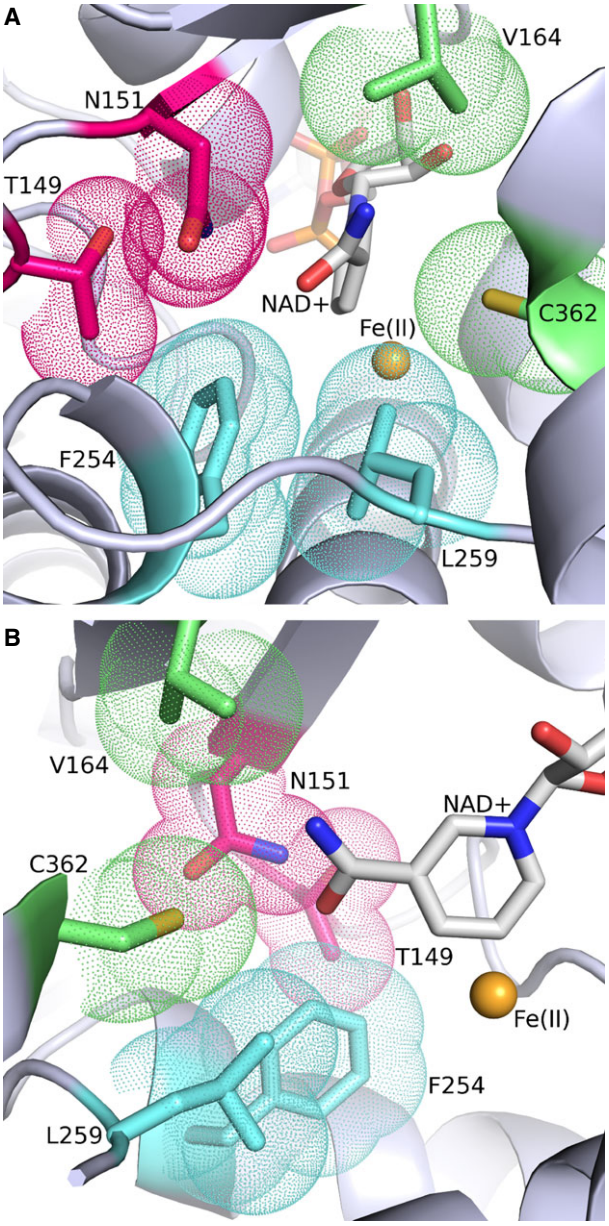

Fig. 2. Mutated residues lining the substrate entry of the FucO active site. Focused libraries containing two residues were constructed: library A (red), T149 and N151; library C (green), V164; library D (cyan) F254 and L259; library E (green) V164 and C362. (A) and (B) display different views rotated 90° around the *y*-axis. Image created with PYMOL 1.5 (http://www.pymol.org) from the atomic coordinates in 1RRM [6].

substrate profiles were kinetically characterized and the most active variant from each generation was used as the template for another round of mutagenesis. Due to the total absence of activity of the wild-type enzyme with diol **4**, it was a challenging aim to isolate new FucO variants that had acquired activity with these

aryl-substituted substrates. However, the wild-type enzyme exhibits poor, albeit measurable, activity with the benzyl-substituted aldehyde phenylacetaldehyde (3) suggesting that FucO might be malleable to accept also aryl-substituted diols as substrates. Aldehyde **3** was chosen as a selective substrate during the initial round of library screening because of its structural relatedness to the aldehyde product resulting from oxidation of **4**. In subsequent rounds, **4** was also included as a selective substrate. The engineering process was successful and enzyme variants active with *S*-**4** and **3** were isolated.

The aim with the present work was to complement the previous study, which was performed on a relatively small set of isolated FucO variants and with a limited set of substrates [10]. In this study, we present a more complete structure–function analysis of these variants together with other variants obtained during the screening process that displayed altered activity profiles. The range of tested alcohol substrates has also been increased to provide a more comprehensive investigation of the substrate scope of these *in vitro* evolved enzymes and to relate their substrate selectivities to their structures. Three-dimensional models of the most interesting variants have been produced to complement the kinetic data and substrates have been docked in to the active site.

Results and discussion

In general, highly specialized enzymes such as FucO are considered less evolvable compared with generalist enzymes displaying activity with a broad range of substrates [11]. Specialists often appear to acquire new catalytic activities by evolving via a more promiscuous generalist enzyme from which, again, more specialized variants can be obtained [12,13]. This is exemplified by variant D93 (L259V), which is a mediocre catalyst but displays activity with almost any substrate presented to it in this study (Fig. 3, Table 1). Addition of the N151G mutation into D93 results in a new specialist enzyme, variant DA1472 (N151G, L259V). This variant is a highly efficient catalyst for the reduction of **3**, but has lost activity with virtually all the other substrates. The other evolutionary branch derived from the D93 variant developed activity with **4**. The DE variants (DE461: V164C, L259V, C362G; DE1028: V164M, L259V, C362Y; and DE1130: V164I, L259V) display higher activity with **4** but still exhibit many features of the parent D93 generalist being all-round in their substrate selectivities and relatively inefficient as catalysts. They do, however, represent important intermediates towards new specialists, since they display improved activity with the **4** enantiomers.

Structure–activity relationships of the variant FucO enzymes

In an enzyme like FucO, the chemical mechanism which is primarily dependent on the chemistry of the cofactor(s) and to a lesser extent on the properties of amino acids contributing to catalysis, effects on catalytic efficiencies as a consequence of mutagenesis may be attributed to changes in the abilities to form productive enzyme–substrate complexes. In general, the overall modeling and docking results show that the size of the active-site cavity is a major limiting factor for FucO to be able to use aryl-substituted molecules as substrates. Additionally, a few active-site amino acids contribute to stabilizing the complex through hydrogen bonds and π–π stacking interactions. The basis for the mutagenesis and selection strategy, aiming to produce variants that had acquired activity with the putative substrate *S*-**4**, was to increase the space in the entrance to the active site. This strategy proved to be effective since the *S*-**2** substrate preference of the wild-type enzyme was totally changed in the most active variants. Variant DA1472 (N151G, L259V) displayed a ratio of $(k_{cat}/K_M)^{\mathbf{3}}/(k_{cat}/K_M)^{S\text{-}\mathbf{2}}$ that was shifted seven orders of magnitude and all tested variants displayed different degrees of decreased catalytic efficiencies with *S*-**2** (Tables 1 and S1).

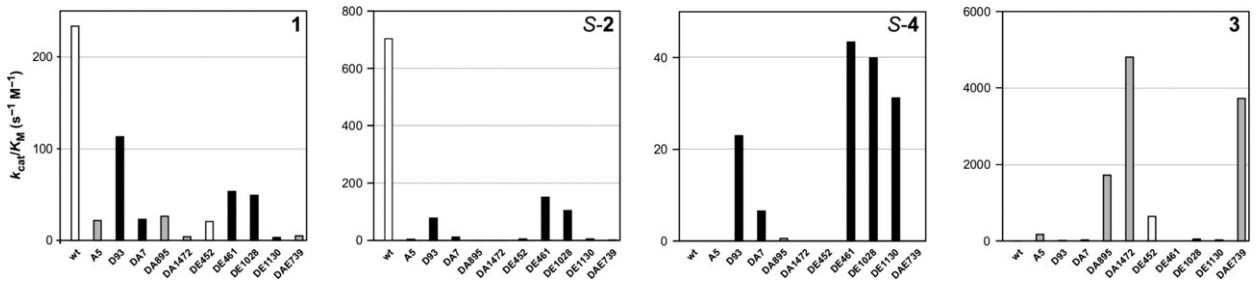

Fig. 3. Enzyme activities expressed as k_{cat}/K_M ($s^{-1}\cdot M^{-1}$) of FucO variants: wild-type (white), N151G-containing variants (grey), L259V-containing variants (black) and DE452 (light grey). See Tables 1 and S1 for additional kinetic data and sequences of these mutants.

Table 1. Steady state kinetic parameters of a selection of variant FucO enzymes. –, not measured; **1**, 1-propanol; **2**, 1,2-propanediol; **3**, phenylacetaldehyde; **4**, 3-phenyl-1,2-propanediol

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}\cdot M^{-1}$)	Enzyme	Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}\cdot M^{-1}$)
Wild-type	**1**	2.8 ± 0.05	12 ± 0.6	240 ± 8	DA7	**1**	0.75 ± 0.04[c]	33 ± 10[c]	23 ± 7[c]
	R-**2**	0.16 ± 0.01	74 ± 7	2.2 ± 0.1	(T149S,	*R*-**2**	n.s.	n.s.	0.36 ± 0.2[c]
	S-**2**	3.8 ± 0.04	5.4 ± 0.1	710 ± 10	L259V)	*S*-**2**	0.64 ± 0.04	55 ± 8	11 ± 2
	3	0.023 ± 0.01	23 ± 10	0.99 ± 0.6		**3**	0.74 ± 0.06	30 ± 3	24 ± 3
	R-**4**	n.s.[b]	n.s.	< 0.1[a]		*R*-**4**	0.0074 ± 0.001	10 ± 2	0.69 ± 0.2
	S-**4**	n.s.[b]	n.s.	< 0.1		*S*-**4**	0.13 ± 0.007	19 ± 2	6.6 ± 0.3
A5	**1**	1.3 ± 0.07	59 ± 8	22 ± 3	DA895	**1**	0.28 ± 0.02[c]	11 ± 1[c]	26 ± 2[c]
(N151G)	*R*-**2**	n.s.[b]	n.s.	0.22 ± 0.1[c]	(T149S,	*R*-**2**	0.045 ± 0.006[c]	150 ± 90[c]	0.30 ± 0.2[c]
	S-**2**	1.3 ± 0.04	330 ± 20	3.9 ± 0.3	N151G,	*S*-**2**	0.070 ± 0.03[c]	500 ± 400[c]	0.14 ± 0.07[c]
	3	2.1 ± 0.1	12 ± 0.7	170 ± 10	L259V)	**3**	3.8 ± 0.2	2.2 ± 0.4	1700 ± 300
	R-**4**	n.s.[b]	n.s.	< 0.1		*R*-**4**	n.s.[b]	n.s.	< 0.1
	S-**4**	n.s.[b]	n.s.	< 0.1		*S*-**4**	n.s.[b]	n.s.	0.53 ± 0.2
D93	**1**	3.1 ± 0.1	27 ± 2	110 ± 4	DE1130	**1**	0.98 ± 0.2[c]	310 ± 100[c]	3.1 ± 1[c]
(L259V)	*R*-**2**	0.61 ± 0.03	280 ± 20	2.2 ± 0.2	(V164I,	*R*-**2**	0.30 ± 0.04[c]	100 ± 80[c]	3.0 ± 0.4[c]
	S-**2**	1.9 ± 0.1	25 ± 2	77 ± 40	L259V)	*S*-**2**	1.3 ± 0.04	270 ± 20	5.0 ± 0.4
	3	0.62 ± 0.1	48 ± 10	13 ± 3		**3**	0.90 ± 0.1	38 ± 10	24 ± 1
	R-**4**	n.s.[b]	n.s.	0.50 ± 0.1		*R*-**4**	0.014 ± 0.003	35 ± 10	0.40 ± 0.05
	S-**4**	0.15 ± 0.004	6.6 ± 0.4	23 ± 2		*S*-**4**	0.29 ± 0.01	9.3 ± 0.5	31 ± 1
DA1472	**1**	0.44 ± 0.05[c]	110 ± 40[c]	4.0 ± 2[c]	DE461	**1**	2.0 ± 0.1	37 ± 4	53 ± 6
(N151G,	*R*-**2**	n.s.	n.s.	0.62 ± 0.06[c]	(V164C,	*R*-**2**	0.61 ± 0.03	360 ± 30	1.7 ± 0.1
L259V)	*S*-**2**	n.s.	n.s.	0.23 ± 0.1[c]	L259V,	*S*-**2**	3.5 ± 0.08	23 ± 1	150 ± 7
	3	21 ± 1	4.4 ± 0.7	4800 ± 800	C362G)	**3**	n.s.	n.s.	25 ± 1
	R-**4**	n.s.[b]	n.s.	< 0.1		*R*-**4**	0.050 ± 0.005	26 ± 4	1.9 ± 0.1
	S-**4**	n.s.[b]	n.s.	< 0.1		*S*-**4**	0.32 ± 0.01	7.3 ± 0.4	43 ± 2
C191 (V164I)	**1**	4.3 ± 0.2	75 ± 6	57 ± 3	D47	**1**	4.6 ± 0.2	49 ± 5	94 ± 6
	R-**2**	0.56 ± 0.02	210 ± 10	2.7 ± 0.1	(F254I)	*R*-**2**	0.65 ± 0.02	320 ± 20	2.1 ± 0.04
	S-**2**	7.2 ± 0.2	22 ± 2	320 ± 20		*S*-**2**	14 ± 0.5	50 ± 4	270 ± 20
	3	0.070 ± 0.001	4.9 ± 2	1.4 ± 0.4		**3**	0.026 ± 0.003	24 ± 5	1.1 ± 0.1
	R-**4**	–	–	–		*R*-**4**	–	–	–
	S-**4**	–	–	–		*S*-**4**	–	–	–

[a] From estimated level of detection of enzyme activity. [b] n.s., unable to reach substrate saturation within the applicable concentration range. [c] Negative cooperativity. K_M values represent $K_{0.5}$, half-saturating substrate concentration, and k_{cat}/K_M represents $k_{cat}/K_{0.5}$.

The mutated residues were chosen solely based on their anticipated influence on substrate binding and access to the active-site cavity. Analysis of their conservation in related proteins, however, suggests that these residues may indeed also be involved in determining the substrate scope of the enzymes (Table 2). Some of these proteins have been characterized for their functional properties and certain distinctions can be made. The relatively similar alcohol dehydrogenase (Adh3) from *Oenococcus oeni*, which shares four of the six residues that were mutated in FucO, displays some differences in substrate preferences [14]. Adh3 catalyzes the oxidation of ethanol with 20-fold higher efficiency compared with FucO, while displaying an equal activity in the acetaldehyde reduction. The same enzyme, however, displays a 100-fold lower activity with **2** compared with FucO. Hence, it appears that Adh3 is primarily adjusted to efficiently oxidize mono-substituted short-chained alcohols while FucO preferentially facilitates oxidation of likewise short-chained 1,2-diols. Although other parts of the protein structures also contribute to substrate selectivity [19], it is anticipated that the structural differences in the active sites cause the functional differences between the proteins listed in Table 2.

Of the variants that were produced and screened for catalytic activity, the following displayed particularly interesting functional properties: A5 (N151G), D47 (F254I), D93 (L259V), DA7 (T149S, L259V), DE1130 (V164I, L259V), DA895 (T149S, N151G, L259V) and DE461 (V164C, L259V, C362G). To search for structural determinants for the acquired activities, models of the tertiary structures of these variants were generated and the binding modes of the transformed substrates were predicted by docking studies into the active sites.

Table 2. Structure-guided sequence alignment of FucO-related proteins.

Protein	PDB code	Residue^{position}						Sequence identity (%)
E. coli FucO	2BL4 [6]	T^{149}	N^{151}	V^{164}	F^{254}	L^{259}	C^{362}	(100)
O. oeni Adh3	4FR2 [14]	T^{151}	H^{153}	V^{166}	F^{258}	L^{263}	N^{366}	39
G. thermoglucosidasius Adhe	3ZDR [15]	T^{155}	F^{157}	P^{170}	F^{262}	L^{267}	C^{386}	34
T. maritima ADH	1O2D [16]	T^{145}	Y^{147}	G^{159}	I^{247}	T^{252}	H^{336}	30
R. eutropha putative	3JZD	T^{128}	V^{130}	T^{141}	L^{233}	M^{238}	–[a]	29
S. meliloti probable glycerol DH	3UHJ	S^{147}	I^{148}	A^{162}	F^{269}	C^{274}	I^{359}	29
C. glutamicum ADH	3IV7	T^{127}	V^{129}	T^{140}	F^{232}	S^{237}	–	27
E. coli unknown function	1OJ7 [17]	N^{147}	G^{149}	A^{162}	–	–	–	26
T. maritima ADH	1VLJ	N^{148}	N^{150}	G^{163}	–	–	–	26
T. maritima glycerol DH	1KQ3 [18]	S^{123}	L^{125}	Y^{138}	F^{243}	L^{248}	E^{332}	23
S. denitrificans ADH	3RF7	S^{149}	T^{151}	G^{162}	I^{249}	V^{254}	L^{331}	22

[a] Corresponding residue missing in structure due to differences in local folding.

Removal of N151 improves activity with the benzyl-substituted aldehyde but abolishes activity with diols

An N151G mutation was expected to result in an enlargement of the active-site entrance that would allow relatively unrestricted entry of bulkier aryl-substituted substrates. The strategy was clearly successful in that the first-generation variant A5 (N151G) indeed exhibited improved activity with **3**, and all other variants that had also acquired the N151G mutation, DA1472 (N151G, L259V) and DA895 (T149S, N151G, L259V), displayed from reasonable to very high activities with **3** (Table 1). Based on our previous docking studies of DA1472, we proposed that the acquired activity resulted from removal of the sterically interfering side chain of N151 [10]. In the present study, modeling of the parent variant A5 (N151G) supports this hypothesis. In these N151G-containing variants, **3** is able to form favorable interactions with F254, which is expected to stabilize the ternary complex. In variants that have retained N151, clashes between the side chain and **3** prevent the π–π stacking interactions with F254 and most probably they are therefore unable to utilize **3** as a substrate. Hence, **3** is proposed to be bound productively only if the side chain of N151 is removed.

All the N151G variants, however, lost much of the wild-type activity with the native *S*-**2** diol substrate (Fig. 3, Table 1) and were devoid of activity with both **4** enantiomers. The relative activities of these variants with the unselected-for substrate 1-propanol (**1**) followed the same trend but to a lesser degree (Fig. 3). Therefore, it seems that the overall activity with 2-hydroxy-substituted substrates was more severely affected in these variants compared with the activity with the mono-substituted **1**. The native alcohol substrate, *S*-**2**, mainly interacts with the wild-type FucO through hydrogen bonds to NAD^+ and to the amide side chain of N151 (Fig. 4A). N151 hydrogen bonds to both hydroxyl groups of the bound *S*-**4** diol, thereby enforcing the interactions between the enzyme and the substrate molecule (Fig. 4B–D). It is thus plausible that *S*-1,2-diols require the asparagine side chain amide for adequately stable interactions with the enzyme. Based on the results from our previous modeling and docking studies, the N151G mutation reduces the hydrogen bond interactions with the diol substrates [10]. Hence, the loss of the N151 mediated hydrogen bonds destabilizes the enzyme–substrate complexes in the case of *S*-**2** and *S*-**4**. Even though the N151G mutation, as expected, enabled activity with benzyl-substituted aldehyde **3**, the results with **2** and **4** illustrated that a more refined fine-tuning of the active-site structure is required for FucO to accept bulkier diol substrates.

L259V mutation installs activity with aryl-substituted diols

The fact that variant D93 (L259V) catalyzes the oxidation of *S*-**4** with reasonable activity illustrates that quite subtle structural changes play a role in tailoring the substrate scope. The L259V mutation adds 25 $Å^3$ more volume to the active site [20] (corresponding to a methylene group) and pulls the isopropyl group closer to the peptide main chain. The volume difference between **2** and **4** resulting from the insertion of the phenyl substituent is approximately 100 $Å^3$. This discrepancy in the amount of the added volume of the substrate and the increased space in the active-site cavity suggests that the binding geometry is the key to

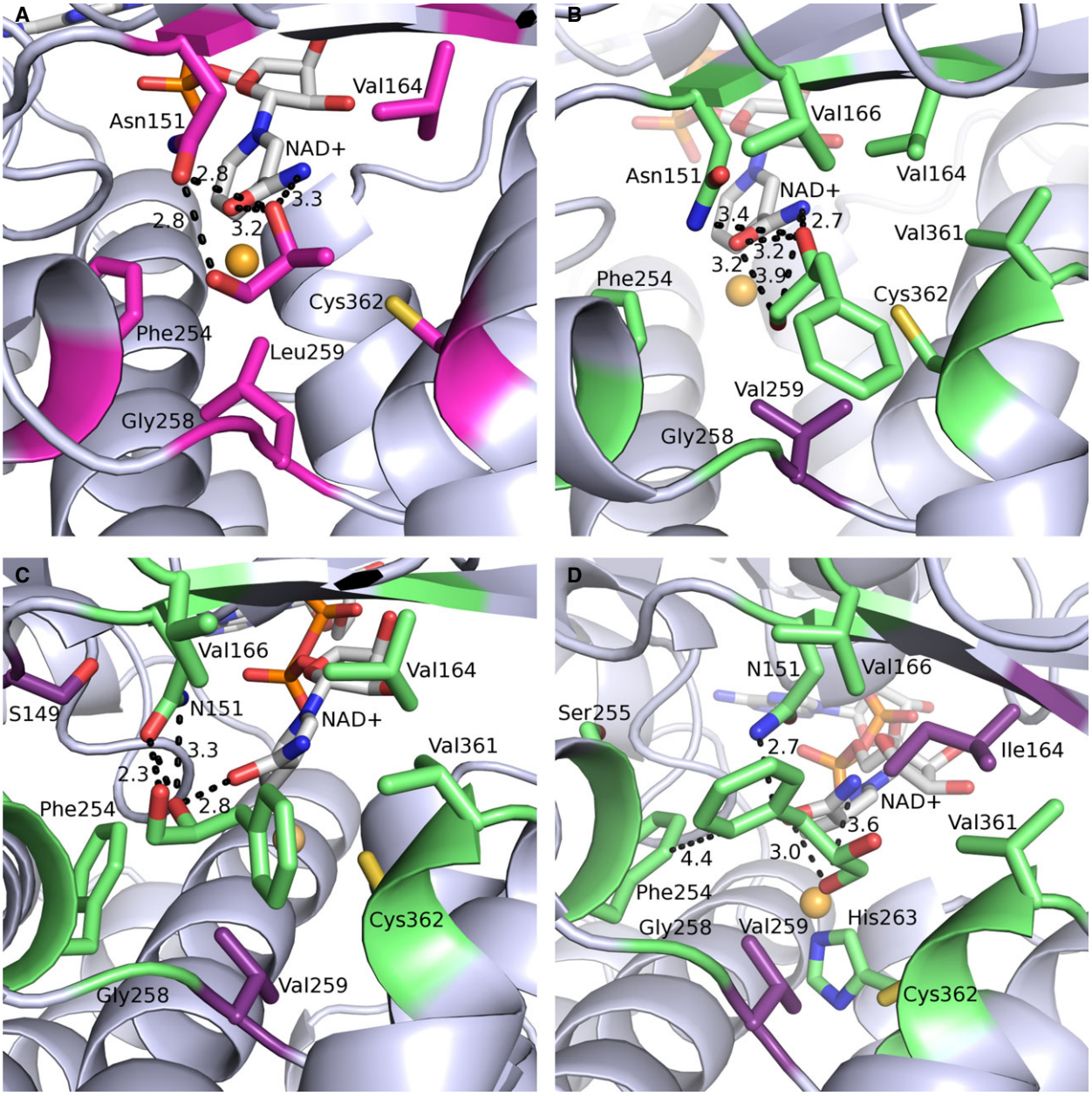

Fig. 4. (A) *S*-**2** (pink) hydrogen bonds to both NAD^+ (grey) and N151 when docked into the wild-type FucO enzyme. Residues within 4 Å of the substrate are shown as pink sticks. *S*-**4** (green) hydrogen bonds through both hydroxyl groups to NAD^+ and N151 when docked to the D93 (L259V) (B), DA7 (T149S, L259V) (C) and DE1130 (V164I, L259V) (D) variants. In the DE1130 mutant *S*-**4** also π–π stacks with F254. In (B), (C) and (D), the residues within 4 Å of the substrate molecule are shown as green sticks, while the mutated residues are shown as purple sticks.

formation of productive enzyme–substrate complexes with bulky diol *S*-**4**. However, modeling of the interactions between *S*-**4** and D93 (L259V) and DE1130 (V164I, L259V) (Fig. 4B,D) suggests that the enzyme activities are facilitated by the larger active-site cavities caused by the mutations. Although the added volume is relatively small it allows for alternative, productive binding modes in the active sites. As already mentioned, the activity also requires the presence of N151. The present results are in accordance with previous modeling on the related DE461 (V164C, L259V, C362G) variant [10].

Introduction of a V153I mutation to DE1130 abolishes activity with phenyl-substituted diols

Interestingly, the DE452 (V153I, V164I, L259V) variant, which is a V153I variant of DE1130 (V164I, L259V), is also able to use **3** as substrate, although it lacks the otherwise required N151G replacement. DE452, however, is unable to use *S*-**4** as substrate whereas DE1130 is active with the same substrate. Therefore, the V153I substitution, in this structural context, installs activity with **3** but abolishes activity with *S*-**4**. Based on the modeling studies, the introduction of a bulkier isoleucine at position 153 in DE452 may limit the possible side chain conformations of I164, which, as a result, points directly into the active-site entrance, thereby limiting its *S*-**4** access. On the other hand, in DE1130 where V153 is retained, I164 can adopt a different conformation, which allows the bulky *S*-**4** substrate to enter into the active site.

V164I and F254I mutations increase native activity with *S*-1,2-propanediol

The V164I (C191) and F254I (D47) variants stood out in the analysis since both of them exhibited increased turnover numbers with *S*-**2**. These enzymes showed two- and four-fold increases in k_{cat}, respectively. This is noteworthy because the turnover numbers of the wild-type enzyme have never exceeded 4 s^{-1} when its activity has been tested with a range of alcohol substrates. This suggests a rate limiting step uncoupled from the substrate structure, which is in accordance with our previous result that NADH release is rate limiting for the FucO-catalyzed oxidation reaction [5]. The fact that the V164I and F254I variants display higher turnover numbers indicates that the process of cofactor release has been affected and the dissociation rate increased. The predicted structural effects of these mutations are consistent with the observed kinetics. Both V164 and F254 are in near van der Waals distance (< 4 Å) from the nicotinamide ring of the bound NAD^+ (PDB code 2BL4 [6]) and thus both the V164I and F254I replacements are expected to affect the cofactor binding directly.

The docking poses of *S*-**2** in the wild-type and the F254I variant are similar to each other. In the wild-type enzyme, F254 interacts with the amide oxygen (3.0 Å) and the pyridine ring (4.9 Å) of NAD^+. In the F254I variant enzyme, however, I254 is not able to interact with NAD^+ at all. Therefore, it seems likely that the loss of interaction between the enzyme and the nucleotide in this variant makes NAD^+ less tightly bound to the mutated enzyme and hence the rate determining dissociation may become faster. *S*-**4** has a similar binding profile to *S*-**2**, but its phenyl ring forms an additional π–π stacking interaction with F254. The F254I mutation abolishes the possibility of forming π–π stacking interactions, which explains why the F254I variant lacks activity with *S*-**4**.

T149 has a role in FucO catalysis

One of the mutated residues, T149, which is located at the active-site entry, distant from the cofactor and the catalytic iron (Fig. 1), seems to be important for catalysis. Structure-based sequence alignment of FucO-related proteins reveals that T149 is highly conserved in the most similar enzymes and its hydrogen bonding capacity is retained in the other related proteins by either serine or asparagine substitutions at the corresponding positions (Table 2). The apparent importance of a hydroxyl-containing amino acid at this position of the active site is also reflected in the variants isolated in the directed evolution. In the construction of library A (Fig. 2), T149 was allowed to become substituted only by alanine or serine [10]. The only isolated variants that had preserved appreciable activity with *S*-**2** were those where T149 had been retained or was conservatively replaced by serine. Two variants carrying a T149A substitution (variants A210, T149A, N151W and A203, T149A, N151S) were both severely impaired in their abilities to catalyze the oxidation of *S*-**2**, with k_{cat}/K_M values three orders of magnitude lower than with the wild-type (Table S1). In the second-generation library DA, in which full randomization of T149 was allowed, only variants that had retained T149 or contained the T149S replacement were identified to be active with *S*-**2**.

The effects of the T149 substitutions in the activities with the aryl-substituted substrates emphasize the importance of T149. Variant DA7 (T149S, L259V) is 3.5-fold less active with *S*-**4** compared with D93 (L259V), primarily as a result of an elevated $K_M^{S\text{-}4}$. When the related variants DA1472 (N151G, L259V) and DA895 (T149S, N151G, L259V) are compared, the reduction reaction with aldehyde **3** does not provide the same result since the primary effect is a 5.5-fold reduction in k_{cat}[3]. Hence, the importance of T149 is also reflected in this comparison but the effects are distinct, possibly due to differences in rate-limiting steps in the oxidation and reduction reactions.

Furthermore, in the modeled structures of wild-type FucO and variants with retained N151, a hydrogen bond is formed between N151 and the hydroxyl group of T149. The distance between the atoms is 3.2 Å in the wild-type enzyme and in the set of 10 T149S

variants it ranges from 2.7 to 5.0 Å. Additionally, the β-methyl group in the T149 side chain interacts with the phenyl ring of F254 restricting the rotational freedom of T149. When the β-methyl group is lost by the T149S mutation, the side chain packing is affected and the unbranched serine residue becomes more flexible. The increased rotational freedom could destabilize the deduced hydrogen bonding between S149 and N151, which in turn would render N151 more flexible and might compromise its hydrogen bonding capacity to the substrate and cofactor, hence destabilizing protein–substrate interactions and affecting the catalyzed reduction rate. Our results show that the hydroxyl group in position 149 is required for effective catalytic reaction and even its spatial position is important for catalysis. Crystal structures of T149S mutants would be needed to give more insight into the speculated issue.

Thermodynamic cycle analysis of mutagenesis impact

Some of the isolated variants can be viewed as mutants of each other. An example is the A5/DA1472 pair, where the latter can be regarded as an L259V variant of A5 (or an N151G variant of D93). Comparison of the catalytic efficiencies (k_{cat}/K_M) of these related variants allows direct comparisons of the transition state (TS) stabilization and the effects caused by the inserted mutations [21]. The results are summarized in Fig. 5.

As already mentioned, variants active with **3** have lost the catalytic ability to oxidize the native *S*-**2** substrate to the opposite extent. The comparison of $\Delta\Delta G$ values of variants A5 (N151G), D93 (L259V) and the double mutant A5/D93 (DA1472) reveals that the N151G mutation causes the most severe change for the catalytic reaction with *S*-**2** with an increase of 13 kJ·mol^{-1} in $\Delta\Delta G$ (Fig. 5A), which reflects how much this substitution weakens the free energy of the TS stabilization. The decrease in $(k_{cat}/K_M)^{S\text{-}\mathbf{2}}$ in the A5 variant, compared with the wild-type enzyme, is primarily manifested as a 60-fold increase in $K_M{}^{S\text{-}\mathbf{2}}$. Assuming that K_M (to an extent) is a reflection of the dissociation constant of *S*-**2** from the ternary complex (in the ground state), a higher value of $K_M{}^{S\text{-}\mathbf{2}}$ suggests poorer affinity between the enzyme and the ternary complex. Therefore the observed decrease in activity indicates that the binding interactions, which are lost in the ground state and cause an increase in K_M, also affect TS stabilization unfavorably. The value roughly equals one lost hydrogen bond, which is not unreasonable considering the structural effects of the N151G substitution. The effects on TS stabilization resulting from the L259V mutation are less drastic while the double mutant DA1472 displays, within experimental error, the sum of the effects caused by the respective single substitutions. The additive effect of mutations provides another type of information: namely, since the effects of the structural alterations act independently of each other, the concomitant double substitutions do not provide any synergistic effects.

The same analysis with aldehyde **3** provides the energetic gains in the stabilization of TS in the reduction reaction (Fig. 5B). Since the same variants are compared, the mutations, as one would expect, also act additively with **3**. There is a clear qualitative difference, however, compared with the effects of mutations on the *S*-**2** oxidation. Based on the observed relatively small impact on $K_M{}^{\mathbf{3}}$ (Table 1), the gain in TS stabilization with **3** appears to be realized only in the TS but not in the ground state.

Analyzing the impact of the T149S substitution by thermodynamic cycles reveals that in the catalyzed oxidation of the not-selected-for substrate **1**, the negative impact of the N151G mutation on activity is essentially rescued by the addition of T149S (Fig. 5C). Hence, the T149S substitution, going from variant DA1472 (N151G, L259V) to DA895 (T149S, N151G, L259V), changes the catalytic mechanism in the oxidation of **1**. The favorable effect of the T149S mutation is manifested in a 10-fold lowering of $K_M{}^{\mathbf{1}}$ (Table 1).

In a variant with retained N151, such as D93 (L259V), the T149S substitution primarily results in a lower turnover number. The T149S variant of D93, DA7, displays similar $K_M{}^{\mathbf{1}}$ but a four-fold lower k_{cat}. Consequently, when N151 is replaced by a glycine, the side chain of S149 may contribute to stabilization of the ternary complex in the oxidation of the mono-substituted alcohol **1**. A decrease in turnover numbers was also observed in other isolated T149S variants but to a lesser degree. For the α-hydroxy-substituted diols **2** and **4**, the effects are different and the T149S substitution does not cure the negative effects caused by the N151G mutation (Fig. 5D), which is in accordance with the modeling results suggesting stabilizing hydrogen bonds between N151 and the α-hydroxyl of the diols.

Negative cooperativity in mutants catalyzing reactions with non-selected-for substrates

Deviations from Michaelis–Menten kinetics are relatively frequent in oligomeric enzymes. Observed cooperativity in substrate binding is a general phenomenon and fulfills physiological functions in regulating enzyme activities. In the case of negative cooperativity,

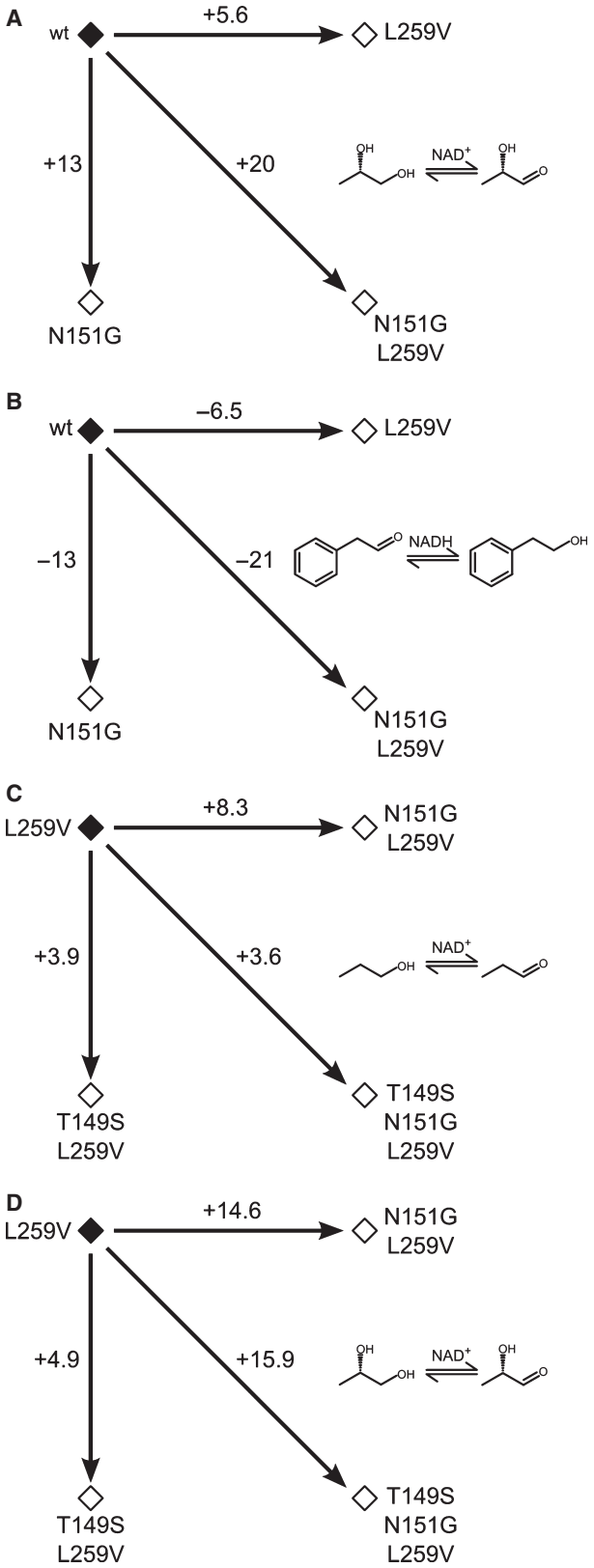

Fig. 5. Thermodynamic cycles of the reduction of *S*-**2** (A) and the oxidation of **3** (B) catalyzed by either the single mutants A5 (N151G) or D93 (L259V) or the double mutant DA1472. (C) Analysis of the impact of the T149S mutation on the oxidation of **1** and (D) the oxidation of *S*-**2**. The $\Delta\Delta G$ values given (in kJ·mol^{-1}) have been calculated from $\Delta\Delta G = -RT \ln[(k_{cat}/K_M)^{mutant}/(k_{cat}/K_M)^{reference}]$ [21]. It is clear that the inserted mutations are beneficial for the reduction of **3** but, to a similar extent, unfavorable for catalysis of the oxidation of *S*-**2**. The additivity of the calculated $\Delta\Delta G$ values in cases (A) and (B) also suggests that these mutations act independently of each other and without synergistic effects.

the (effector) ligand binding causes the enzyme to be inhibited by some mechanism (e.g. conformational change or non-productive, inhibitory binding) [22–24]. The presence of cooperativity is easily revealed by the Hill coefficient [25] that deviates from unity, with values < 1 for negative cooperativity.

During the kinetic analysis of the different FucO variants that were isolated in the evolution process, several variants displayed the hallmarks of negative cooperativity, with Hill coefficients of less than unity. The absolute activities for these enzymes were in all cases low and the negative cooperativity behavior was only observed with the three-carbon alcohols **1**, *R*-**2** and *S*-**2** (Table 1). *R*-**2** is clearly the most potent homotropic inducer of negative cooperativity as the majority of all active mutants exhibited this feature with this substrate. The underlying mechanism for this behavior is as yet unknown but a plausible cause could be that the active-site mutations have made alternative non-productive binding modes possible for these substrates. These binding modes may result in half-of-the-sites reactivity and thus cause the observed kinetic behavior.

Conclusion

The present analysis of a population of enzyme variants, all isolated after selection for activities towards 'unnatural' substrates, provides insight into the structural underpinnings of molecular evolution for new functions. Iterative directed evolution allows one to follow the breadcrumbs that end up in the emergence of new catalytic functions. In general, the starting protein, in this case the highly specialized *S*-lactaldehyde reducing enzyme, appears to be altered from a specialist into a more promiscuous generalist. Regarding activity with phenylacetaldehyde the FucO enzyme went through the generalist D93 (L259V) to the highly specialized DA1472 enzyme. Variants DE1028 and DE1130 have gained activity with *S*-3-phenylpropane-1,2-diol (*S*-**4**)

and simultaneously lost activity with 1-propanol (**1**) and 1,2-propanediol (**2**). By additional mutagenesis, these variants thus have the potential to become specialized toward the target substrate *S*-**4**. Even though the amino acid change in the L259V mutation can be regarded as conservative, the extra space created seems to be a crucial factor for variants that are able to use **4** as a substrate. Furthermore, residue N151, which based on docking studies stabilizes the binding of diols *S*-**4** and *S*-**2** to the enzyme, clearly has an important role since all N151G variants were unable to use these diols as substrates. The π–π stacking property of F254 has an additional impact on *S*-**4** binding as the F254I variant displays no activity with this diol. In most variants, both hydroxyl groups of *S*-**4** are predicted to form hydrogen bonds with NAD^+, which further stabilizes the enzyme–substrate interactions. Mutations such as F254I and V164I result in decreased interactions between the enzyme and the nucleotide and thereby give rise to increased k_{cat} values for *S*-**2**. This finding opens up new possibilities for enhancing the activity for other variants by engineering the nucleotide binding site for increased dissociation rate of the cofactor. We shall continue the development of new catalytic activities in future directed evolution efforts to fulfill the goal of generating an efficient biocatalyst. Additionally, we also aim to gather more experimental data describing how substrate selectivity is evolved in the FucO enzyme variants.

Experimental procedures

Chemical and reagents

Commercially available chemicals were purchased from Sigma-Aldrich (St. Louise, MO, USA) at the highest purity available. Chiral purity of *S*- and *R*-1,2-propanediol (**2**) was ≥ 96%. *S*- and *R*-(2,3- epoxypropyl)benzene at > 98% chiral purity were purchased from TCI Europe N.V. (Zwijndrecht, Belgium). Chromatographic resins were purchased from GE Healthcare (Uppsala, Sweden). The *S*- and *R*-enantiomers of **4** were synthesized by *Solanum tuberosum* epoxide hydrolase catalyzed ring opening of the corresponding epoxide enantiomers according to the protocol previously described [10].

Selection of enzyme variants

Enzyme variants analyzed in this study emerged from earlier work in which FucO was subjected to directed evolution in order to enhance its activity with aryl-substituted substrates [10]. In addition to the variants displaying improved activity with the target substrates, variants assigned as hits from the screen but, due to their apparently low specific activities with these substrates when measured on purified proteins, were discarded from further evolution were also included into this structure–activity relationship study.

Larger scale expression and purification of FucO variants

C-terminally 5-His-tagged FucO mutants were expressed in *E. coli* according to the protocol previously described [5] and purified by Ni^{2+}-affinity chromatography as described previously [10].

Kinetic characterization of FucO variant enzymes

Catalyzed initial rates were measured for **1**, *S*- and *R*-**2**, **3**, *S*- and *R*-**4** in the presence of 0.2 mM of NAD^+ or NADH (oxidation or reduction, respectively) and the steady state parameters k_{cat}, K_M and k_{cat}/K_M were extracted. Reactions were followed spectrophotometrically, at 30 °C, by monitoring the change in NADH concentration ($\Delta\varepsilon$ = 6.22 $mM^{-1}\cdot s^{-1}$) using a Molecular Device SpectraMAX 190 plate reader. In the case of *S*- and *R*-**4** a Shimadzu UV-1700 spectrophotometer was used. Oxidation reactions were performed in 0.1 M glycine pH 10.0 and reduction reactions in 0.1 M sodium phosphate, pH 7.0. Measurements were performed in duplicate. Kinetic parameters were extracted by fitting the Michaelis–Menten equation to the data by nonlinear regression using programs MMFIT and RFFIT in the SIMFIT package (http://www.simfit.man.ac.uk/). Measurements were performed essentially as described in [5] with the exception that each substrate concentration was measured in duplicate.

Structure-guided sequence alignment

The primary amino acid sequence of FucO was used as the query in a BLAST [26] search of the protein structure data bank (PDB). Ten hits of varying degrees of sequence identity with the query sequence were selected and subjected to a CLUSTAL W2 [27] alignment. The resulting alignment based only on primary structure similarities was corrected following superimposition of each protein tertiary structure onto the FucO structure. See Table 2 for protein identities and degrees of similarity with FucO.

Structural modeling and docking

The three-dimensional models of FucO variants A5 (N151G), D93 (L259V), D47 (F254I), DA1472 (N151G, L259V), DA7 (T149S, L259V), DE452 (V153I, V164I, L259V) and DE1130 (V164I, L259V) were created based on the crystal structure of FucO in complex with NAD^+ (PDB code 2BL4 [6]). MODELLER [28] was used to build a set of 10 models for each variant with NAD^+ and Fe^{2+} in

the active site. The model with the lowest value of the MODELLER objective function was used for docking studies.

For ligand docking, *S*-**2** and **3** were taken from the crystal structure of *Klebsiella oxytoca* diol dehydratase (PDB code 1UC4 [29]) and *E. coli* amine oxidase (PDB code 1D6U [30]), respectively. The ligand from the crystal structure complex between *Scytalidium lignicola* Scytalidopepsin B (PDB code 2IFR [31]) and (2*R*, 3*S*)-3-amino-3-phenylpropane-1,2-diol was edited to *S*-**4** with MAESTRO MOLECULAR MODELING INTERFACE (Version 9.3; Schrödinger Inc.).

All protein models and substrates were prepared for docking in DISCOVERY STUDIO (Version 3.5; Accelrys Inc.), and active-site cavities were detected automatically. O7N and H23/H25 of NAD^+ were defined as hydrogen bond donor and acceptor, respectively. *S*-**2** O4 and O5 were constrained to hydrogen bond acceptors, while H12 and H13 were handled as donors. In *S*-**4**, O1 and O5 were defined as hydrogen bond acceptors, and H22 and H23 as donors. O9 in **3** was constrained to hydrogen bond acceptor. GOLD via DISCOVERY STUDIO (Version 3.5; Accelrys Inc.) was used for docking the three ligands to the wild-type FucO structure and all the variants. The docking poses were analyzed and scored with the SCORE LIGANDS function in DISCOVERY STUDIO (Version 3.5; Accelrys Inc.), and the best pose was the one with highest PLP2 score. PYMOL (Version 1.5; Schrödinger LLC) was used to prepare pictures of the complexes.

Acknowledgements

Professor M.S. Johnson and Biocenter Finland (bioinformatics, structural biology and translational activities) are acknowledged for the excellent computational infrastructure at the Structural Bioinformatics Laboratory, Åbo Akademi University, and the National Doctoral Programme in Informational and Structural Biology is also acknowledged. This work was funded by the Swedish Research Council (MW, grant 621-2011-6055), Sigrid Juselius Foundation (TAS, KMD), Tor, Joe and Pentti Borg Foundation (TAS, KMD) and Åbo Akademi Graduate School (KMD).

Author contribution

CB performed the experimental experiments. TAS and KMD performed the structural modeling and docking experiments. All authors contributed to planning and writing of the manuscript.

References

1 Reid MF & Fewson CA (1994) Molecular characterization of microbial alcohol dehydrogenases. *Crit Rev Microbiol* **20**, 13–56.

2 Boronat A & Aguilar J (1979) Rhamnose-induced propanediol oxidoreductase in *Escherichia coli*: purification, properties, and comparison with the fucose-induced enzyme. *J Biol Chem* **140**, 320–326.

3 Baldomá L & Aguilar J (1988) Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: aerobic–anaerobic regulation of L-lactaldehyde dissimilation. *J Bacteriol* **170**, 416–421.

4 Conway T & Ingram LO (1989) Similarity of *Escherichia coli* propanediol oxidoreductase (fucO product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. *J Bacteriol* **171**, 3754–3759.

5 Blikstad C & Widersten M (2010) Functional characterization of a stereospecific diol dehydrogenase, FucO, from *Escherichia coli*: substrate specificity, pH dependence, kinetic isotope effects and influence of solvent viscosity. *J Mol Catal B Enzym* **66**, 148–155.

6 Montella C, Bellsolell L, Pérez-Luque R, Badía J, Baldoma L, Coll M & Aguilar J (2005) Crystal structure of an iron-dependent group III dehydrogenase that interconverts L-lactaldehyde and L-1,2-propanediol in *Escherichia coli*. *J Bacteriol* **187**, 4957–4966.

7 Reetz MT & Carballeira JD (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc* **2**, 891–903.

8 Reetz MT, Prasad S, Carballeira JD, Gunulya Y & Bocola M (2010) Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J Am Chem Soc* **132**, 9144–9152.

9 Reetz MT, Wang L-W & Bocola M (2006) Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew Chem Int Ed* **45**, 1236–1241.

10 Blikstad C, Dahlström KM, Salminen TA & Widersten M (2013) Stereoselective oxidation of aryl-substituted vicinal diols into chiral α-hydroxy aldehydes by re-engineered propanediol oxidoreductase. *ACS Catal* **3**, 3016–3025.

11 Tracewell CA & Arnold FH (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol* **13**, 3–9.

12 Matsumura I & Ellington AD (2001) *In vitro* evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J Mol Biol* **305**, 331–339.

13 Rockha-Shmuel L & Tawfik DS (2012) Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. *Nucleic Acids Res* **40**, 11627–11637.

14 Elleuche S, Fodor K, Klippel B, von der Heyde A, Wilmanns M & Antranikian G (2013) Structural and biochemical characterisation of a NAD^+-dependent alcohol dehydrogenase from *Oenococcus oeni* as a new

model molecule for industrial biotechnology applications. *Appl Microbiol Biotechnol* **97**, 8963–8975.
15 Extance J, Crennell SJ, Eley K, Cripps R, Hough DW & Danson MJ (2013) Structure of a bifunctional alcohol dehydrogenase involved in bioethanol generation in *Geobacillus thermoglucosidasius*. *Acta Crystallogr D Biol Crystallogr* **69**, 2104–2115.
16 Schwarzenbacher R, von Delft F, Canaves JM, Brinen LS, Dai X, Deacon AM, Elsliger MA, Eshaghi S, Floyd R, Godzik A *et al.* (2004) Crystal structure of an iron-containing 1,3-propanediol dehydrogenase (TM0920) from *Thermotoga maritima* at 1.3 Å resolution. *Proteins* **54**, 174–177.
17 Sulzenbacher G, Alvarez K, Van Den Heuvel RH, Versluis C, Spinelli S, Campanacci V, Valencia C, Cambillau C, Eklund H & Tegoni M (2004) Crystal structure of *E. coli* alcohol dehydrogenase YqhD: evidence of a covalently modified NADP coenzyme. *J Mol Biol* **342**, 489–502.
18 Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T *et al.* (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc Natl Acad Sci USA* **99**, 11664–11669.
19 Zheng H, Wang X, Yomano LP, Geddes RD, Shanmugam KT & Ingram LO (2013) Improving *Escherichia coli* FucO for furfural tolerance by saturation mutagenesis of individual amino acid positions. *Appl Environ Microbiol* **79**, 3202–3208.
20 Counterman AE & Clemmer DE (1999) Volumes of individual amino acid residues in gas-phase peptide ions. *J Am Chem Soc* **121**, 4031–4039.
21 Fersht A (1999) Thermodynamic Cycles. In *Structure and Mechanism in Protein Science*, pp. 129−131. W.H. Freeman, New York, NY.
22 Conway A & Koshland DE Jr (1968) Negative cooperativity in enzyme action. The binding of diphosphopyridine nucleotide to glyceraldehyde 3-phosphate dehydrogenase. *Biochemistry* **7**, 4011–4023.
23 Levitzki A & Koshland DE Jr (1989) Negative cooperativity in regulatory enzymes. *Proc Natl Acad Sci USA* **62**, 1121–1128.
24 Koshland DE Jr (1996) The structural basis of negative cooperativity: receptors and enzymes. *Curr Opin Struct Biol* **6**, 757–761.
25 Hill AV (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol* **40** (Suppl), iv–vii.
26 Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
27 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.
28 Sali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779–815.
29 Shibata N, Nakanishi Y, Fukuoka M, Yamanishi M, Yasuoka N & Torava T (2003) Structural rationalization for the lack of stereospecificity in coenzyme B12-dependent diol dehydratase. *J Biol Chem* **278**, 22717–22725.
30 Wilmot CM, Hajdu J, McPherson MJ, Knowles PF & Phillips SE (1999) Visualization of dioxygen bound to copper during enzyme catalysis. *Science* **286**, 1724–1728.
31 Pillai B, Cherney MM, Hiraga K, Takada K, Oda K & James MN (2007) Crystal structure of scytalidoglutamic peptidase with its first potent inhibitor provides insights into substrate specificity and catalysis. *J Mol Biol* **365**, 343–361.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Table S1. Steady state kinetic parameters of a selection of variant FucO enzymes.