Protein Engineering, Design & Selection, 2015, vol. 28 no. 8, pp. 251–258
doi: 10.1093/protein/gzv028
Advance Access Publication Date: 1 June 2015
Original Article

OXFORD

Original Article

Substitutions at the cofactor phosphate-binding site of a clostridial alcohol dehydrogenase lead to unexpected changes in substrate specificity

Danielle J. Maddock, Wayne M. Patrick, and Monica L. Gerth*

Department of Biochemistry, University of Otago, Dunedin 9010, New Zealand

*To whom correspondence should be addressed. E-mail: monica.gerth@otago.ac.nz
Edited by David Ollis

Received 7 March 2015; Revised 23 April 2015; Accepted 5 May 2015

Abstract

Changing the cofactor specificity of an enzyme from nicotinamide adenine dinucleotide 2′-phosphate (NADPH) to the more abundant NADH is a common strategy for increasing overall enzyme efficiency in microbial metabolic engineering. The aim of this study was to switch the cofactor specificity of the primary–secondary alcohol dehydrogenase from *Clostridium autoethanogenum*, a bacterium with considerable promise for the bio-manufacturing of fuels and other petrochemicals, from strictly NADPH-dependent to NADH-dependent. We used insights from a homology model to build a site-saturation library focussed on residue S199, the position deemed most likely to disrupt binding of the 2′-phosphate of NADPH. Although the CaADH(S199X) library did not yield any NADH-dependent enzymes, it did reveal that substitutions at the cofactor phosphate-binding site can cause unanticipated changes in the substrate specificity of the enzyme. Using consensus-guided site-directed mutagenesis, we were able to create an enzyme that was stringently NADH-dependent, albeit with a concomitant reduction in activity. This study highlights the role that distal residues play in substrate specificity and the complexity of enzyme–cofactor interactions.

Key words: alcohol dehydrogenase, *Clostridium autoethanogenum*, cofactor specificity, enzyme engineering, mutagenesis

Introduction

The use of microorganisms for the production of fuels and chemicals offers an environmentally friendly alternative to their extraction from finite global supplies of crude oil. *Clostridium autoethanogenum* has emerged as a promising platform for such microbial bio-manufacturing. This Gram-positive bacterium possesses the Wood–Ljungdahl pathway (Ragsdale and Pierce, 2008), enabling it to grow with CO and/or CO_2 plus H_2 as its sole carbon and energy sources (Abrini *et al.*, 1994). *Clostridium autoethanogenum* is also highly tolerant of feedstock contaminants (Tracy *et al.*, 2012); this allows the use of industrial waste gas streams, such as those produced by steel mills, as cheap and abundant raw materials for product synthesis (Köpke *et al.*, 2011).

The only chemicals that are synthesised natively by *C.autoethanogenum* at commercially appreciable levels are acetate and ethanol (Köpke *et al.*, 2011). In other acetogenic clostridial species, such as *Clostridium acetobutylicum* (Siemerink *et al.*, 2011) and *Clostridium ljungdahlii* (Köpke *et al.*, 2010), the range of potentially valuable chemicals produced has been extended through the introduction of heterologous enzymes. An alternative strategy is to use genetic engineering to modify native enzymes. This technique can be advantageous, as native enzymes are already optimised for expression and activity in the organism of interest.

In this study, we have focussed on engineering the nicotinamide adenine dinucleotide 2′-phosphate (NADPH)-dependent primary–secondary alcohol dehydrogenase from *C.autoethanogenum* (CaADH). In the previous work, we showed that CaADH was important for catalysing two pathway-ending reactions in the metabolic network of *C.autoethanogenum*: the reduction of acetaldehyde to ethanol; and the reduction of the hydroxy-ketone, *R*-acetoin, to 2*R*,3*R*-butanediol (Köpke *et al.*, 2014). We also showed that CaADH was highly active

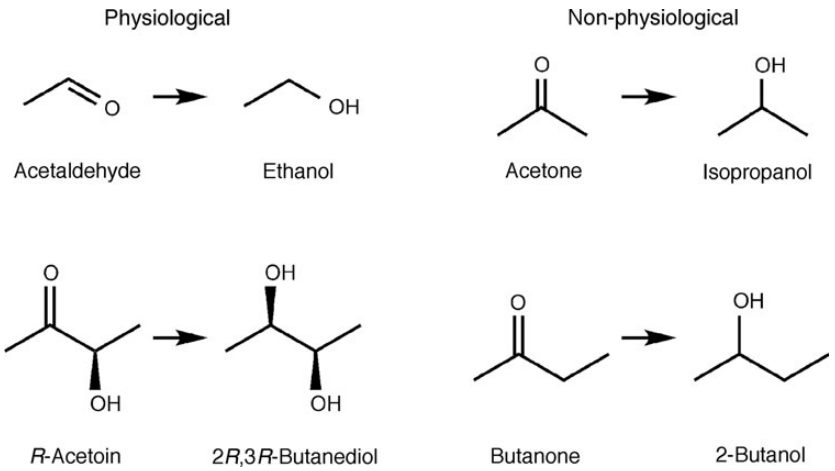

Fig. 1 Physiological and non-physiological reactions of CaADH.

towards a range of aldehyde and ketone substrates, from 2-carbon acetaldehyde, to 3-carbon acetone (with which it was most active), to the 4-carbon substrates *R*-acetoin and butanone (Fig. 1).

Our overall goal is to engineer highly active, highly specific CaADH variants, in order to direct carbon flux in gas-fermenting *C. autoethanogenum* towards desired products with maximal yields. As a first step, the aim of this study was to engineer CaADH to be NADH-dependent. NADH is the favoured cofactor as it generally present at higher cellular concentrations (Bennett *et al.*, 2009). Although this switch has been attempted in a number of different enzymes (see, e.g. Scrutton *et al.*, 1990; Chen *et al.*, 1995; Rane and Calvo, 1997; Lerchner *et al.*, 2013; Pick *et al.*, 2014), it is often accompanied by a loss in catalytic efficiency. However, a study in which engineering a ketol-acid reductoisomerase (KARI) to be NADH-dependent caused a 3-fold increase in titre yield, despite decreased activity *in vitro*, indicates that *in vitro* activity is not always an accurate representation of *in vivo* activity (Bastian *et al.*, 2011). Furthermore, additional engineering can also be employed to restore losses in catalytic efficiency.

In this study, we attempted to switch the cofactor specificity of CaADH using rational design and site-saturation mutagenesis, before ultimately succeeding via a consensus-guided mutagenesis approach. While the site-saturation mutagenesis was targeted at cofactor-binding residues located ~20 Å from the substrate-binding site, variants from the library showed unexpected changes in substrate specificities. Thus, this study highlights the surprising complexity of seemingly straightforward enzyme engineering problems.

Materials and methods

Materials

Restriction enzymes and the Gibson Assembly kit were from New England BioLabs (Ipswich, MA, USA). Taq polymerase was from KAPA Biosystems (Wilmington, MA, USA). Phusion polymerase was from ThermoFisher Scientific (Waltham, MA, USA). Oligonucleotides were from Integrated DNA Technologies (Coralville, IA, USA). Bugbuster reagent and reduced NADH were from Merck Millipore (Billerica, MA, USA). Phenazine methosulfate (PMS) was from J. T. Baker Chemical Co. (Centre Valley, PA, USA). L-Arabinose and ampicillin were from Gold Biotechnology (St. Louis, MO, USA). Dithiothreitol (DTT) was from Melford Laboratories (Ipswich, UK). Acetone, butanone, acetoin, protease inhibitor cocktail and reduced NADPH were from Sigma Chemical Co. (St. Louis, MO, USA). 4-Nitroblue tetrazolium (NBT) chloride was from Boehringer Mannheim (Stuttgart, Germany). The QuikChange II Site-Directed Mutagenesis Kit, including *Escherichia coli* strain XL1-Blue, was from Agilent (Santa Clara, CA, USA). Talon metal affinity resin was from ClonTech (Mountain View, CA, USA).

Bacterial strains

Escherichia coli strain MC1061 was from the *E.coli* Genetic Stock Center (New Haven, CT, USA). *Escherichia coli* strain LMG194 was from Invitrogen (Carlsbad, CA, USA).

Cloning and library construction

The construction of pBAD(KpnI)-CaADH, for the arabinose-induced expression of hexahistidine (His_6) tagged CaADH, was described previously (Köpke *et al.*, 2014). The first variant tested in this study was CaADH (G198D/S199G/R200G/Y218F), abbreviated hereafter as CaADH(DGGF). A synthetic gBlock gene fragment covering 392 bp of the CaADH gene and containing the desired mutations was purchased from Integrated DNA Technologies (Coralville, IA, USA). The vector, pBAD(KpnI)-CaADH, was amplified with primers CaADH_gBlock_for and CaADH_gBlock_rev (Table I), enabling the gBlock to be cloned using Gibson assembly (Gibson *et al.*, 2009).

The CaADH(S199X) site-saturation mutagenesis library was constructed using overlap-extension PCR with Phusion polymerase. Two primary products were generated by amplifying pBAD(KpnI)-CaADH with the primer pairs CaADH_KpnI_for and CaADH_S199X_rev, and CaADH_HindIII_rev and CaADH_S199X_rev (Table I). These primary products were then used in a second PCR to generate full-length product using the flanking primers CaADH_KpnI_for and CaADH_HindIII_rev. This assembled product was subcloned back into pBAD(KpnI)-CaADH after vector and insert had both been digested to completion with restriction enzymes KpnI-HF and HindIII-HF. The ligated plasmids were used to transform *E.coli* MC1061 by electroporation. Of the resulting transformants, 192 colonies were picked into two 96-well microplates (Costar, Corning, NY, USA), in which each well contained 100 µl LB medium supplemented with ampicillin (100 µg ml^{-1}). After 18 h incubation at 28°C, a 50 µl aliquot of glycerol (50%, w/v) was added to each well, and the plates were stored at −80°C. Diversity was confirmed by sequencing 10 of the CaADH genes from the library.

Table I. Oligonucleotides used in this study

Oligonucleotide	Sequence (5′ → 3′)[a]
CaADH_gBlock_for	CAACTACCATGGAAGCGGTGATACTTTACCAATACCTCG
CaADH_gBlock_rev	CCAGTAGTCATCATGTCTGTCATCATAACTGCACTTTCTAAAGG
CaADH_KpnI_for	CAGGTACCGAGAACCTGTATTTCCAAGGAAAAGGTTTTGCAATGTTAGGTATTAAC
CaADH_HindIII_rev	TCTAGAAGCTTAGAATGTAACTACTGATTTAATTAAATCTTTTGG
CaADH_S199X_for	CGGTGTTGGANNKAGACCTGTTTG
CaADH_S199X_rev	CAAACAGGTCTMNNTCCAACACCG
CaADH_DVEA_for	GTGCCACGGACATCGTCAATGCCAAAAATGGCGAATTGTTG
CaADH_DVEA_rev	CAACAATGTCGCCATTTTTGGCATTGACGATGTCCGTGGCAC
CaADH_Y218A_for	ATGGAGCAACTGATATTGTAAATGCTAAAAATGGTGATATAGTTGAAC
CaADH_Y218A_rev	GTTCAACTATATCACCATTTTTAGCATTTACAATATCAGTTGCTCCAT

[a]Restriction sites are underlined; mutagenic sites are shown in bold.

Expression of CaADH(S199X) library variants

Cells from each well of the two microplates containing the CaADH (S199X) library were used to inoculate fresh 96-well plates containing 150 µl LB supplemented with ampicillin (100 µg ml^{-1}). The plates were incubated at 37°C with shaking (500 rpm) in a SelectMix 56 Vortexing Incubator (Select Bioproducts, Edison, NJ, USA) for 16–18 h. Aliquots of the saturated cultures (50 µl) were then used to inoculate deep-well microplates (Labcon), which contained 1 ml LB supplemented with ampicillin (100 µg ml^{-1}) and zinc acetate (50 µM) in each well. The plates were incubated at 37°C with shaking (1200 rpm) for 3 h. Expression of each CaADH(S199X) library variant was induced by adding L-arabinose to each well, to a final concentration of 0.2% (w/v). Induction was for 5 h at 28°C before cells were harvested by centrifugation (3200 *g*, 15 min). Supernatants were removed and the cells were lysed by re-suspension in 50 µl BugBuster reagent followed by incubation at room temperature, with shaking (500 rpm), for 20 min. The insoluble fractions were pelleted by centrifugation (3200 *g*, 15 min) and aliquots of the soluble supernatants were used for activity assays.

Microplate-based activity screening

The activities of the CaADH(S199X) variants were assessed using a discontinuous colourimetric assay based on reduction of NBT to formazan dye in the presence of NAD(P)H and with PMS as a catalyst. Formation of a purple product indicated a lack of enzymatic activity, as an active CaADH would oxidise all NAD(P)H to NAD(P)$^+$, precluding the subsequent reduction of NBT (Glieder and Meinhold, 2003). To test for a change in cofactor specificity, 50 µl of the enzyme-containing soluble cell lysate from each well was added to 150 µl activity buffer (50 mM Tris pH 7.5 with 0.2 mM NADPH or NADH), containing 20 mM acetone as the substrate. Enzyme-catalysed reduction of the acetone was allowed to proceed for 30 min at room temperature. At this point, 50 µl of colour development solution [0.01% (w/v) NBT and 0.0006% (w/v) PMS] was added. Colour development was allowed to develop for 1 min, and determined qualitatively by observation. To test substrate specificity, the protocol was repeated with butanone or acetoin (each at 20 mM) as the substrate. Three independent biological replicates were performed for each combination of cofactor and substrate. The genes encoding variants that showed reproducible changes in specificity were sequenced.

Expression and purification of CaADH and selected variants

Recombinant cell cultures (500 ml) were grown at 37°C with shaking (200 rpm) until mid-log phase ($OD_{600} \approx 0.6$), and then protein expression was induced by adding L-arabinose to a final concentration of 0.2%. At this point, the cultures were moved to 28°C and incubated with shaking for a further 16 h. Cells were harvested by centrifugation (5000 *g*, 10 min) and the pellets were stored frozen at −20°C. On the day of purification, cell pellets were thawed and re-suspended in 10 ml lysis buffer (50 mM Tris–HCl, 100 mM KCl, pH 7.5) with added lysozyme (final concentration of 0.5 mg ml^{-1}), Benzonase nuclease (50 U) and protease inhibitor cocktail (100 µl). After incubation at room temperature for 20 min, cells were lysed by sonication. The insoluble cell debris was pelleted by centrifugation (25 000 *g*, 20 min), and the clarified lysate was mixed with 200–500 µl Talon metal affinity resin. This mixture was incubated for 1 h on a rotating platform at 4°C to allow the binding of His_6-tagged proteins to the resin. The resin was washed twice with 5 bed volumes of lysis buffer, and then transferred to a gravity flow column. The resin was washed twice more with 5 bed volumes of lysis buffer, and then once with 5 bed volumes of lysis buffer supplemented with 5 mM imidazole, and once with 5 bed volumes of lysis buffer plus 10 mM imidazole. The His_6-tagged protein was eluted using lysis buffer containing 150 mM imidazole, and 500 µl fractions were collected. Amicon Ultra centrifugal filter units (10 kDa cut-off; Merck Millipore) were used to exchange each protein into storage buffer (50 mM potassium phosphate, 200 mM NaCl, pH 7.0). Protein concentration was quantified by measuring the absorbance at A_{280} [ε_{280} = 33 920 M^{-1} cm^{-1} for CaADH and all variants except CaADH(S199W), for which ε_{280} = 39 420 M^{-1} cm^{-1}; calculated using ProtParam on the ExPASy server (Gasteiger *et al.*, 2003)]. Aliquots of each protein were stored at −80°C, and only exposed to a single freeze–thaw cycle. Controls showed that this treatment did not affect activity.

Activity assays

The activities of CaADH and variants were measured using spectrophotometric assays. A Cary 100 UV–visible spectrophotometer with a Peltier temperature controller was used to measure the decrease in absorbance at A_{340}, which is indicative of the oxidation of NAD(P)H to NAD(P)$^+$ (NAD(P)H ε_{340} = 6220 M^{-1} cm^{-1}). All assays were carried out at 25°C. To determine the kinetic parameters with fixed cofactor and varying substrate concentrations, the standard assay mixture contained 50 mM Tris–HCl, pH 7.5, 1 mM DTT and 0.2 mM of the appropriate cofactor. The range of substrate concentrations assayed was approximately equivalent to zero to five times the estimated K_M value, except for acetoin, for which the maximum concentration assayed was 250 mM. After pre-equilibration for 1 min, assays were initiated by the addition of enzyme. Wild-type CaADH was added to a final concentration of 5 nM, and the CaADH(S199X) variants

were present at 10 nM. Initial reaction rates were measured with at least five different concentrations of each substrate. All conditions were tested in triplicate and corrected for background. To determine Michaelis constants (K_M) for the cofactors, assays were repeated as described earlier, but with a fixed acetone concentration of 50 mM. The range of cofactor concentrations assayed was approximately equivalent to zero to five times the estimated K_M value for the cofactor. Kinetic parameters (k_{cat} and K_M) were determined by fitting the data to the Michaelis–Menten equation using non-linear regression analysis in Prism (GraphPad, La Jolla, CA, USA). All values are reported as means ± standard errors.

In silico characterisation

The RosettaBackrub flexible backbone algorithm (Smith and Kortemme, 2008) was used to determine the potential effects of mutations at amino acid position 199 on the structure of CaADH. It was used to generate ensembles of 10 most probable structures for each of CaADH, CaADH (S199A) and CaADH(S199R), using the CaADH homology model as the starting template. The backrub motion was applied at a radius of 30 Å to ensure both the cofactor-binding and the substrate-binding site cavities were included in the simulations. To quantify the differences between the ensembles, the active site cavity of each generated structure was measured using the Pocket Volume Measurer (POVME) algorithm (Durrant *et al.*, 2014). To use this script, the user must enter the co-ordinates and radii for a number 'inclusion' and 'exclusion' zones. These zones determine the approximate position of the cavity to be measured. A script was written to define a large inclusion zone that covered an area larger than the putative active site of CaADH, without including any other protein cavities. This script was then applied to all of the RosettaBackrub ensembles generated for CaADH, CaADH(S199A) and CaADH(S199R).

Consensus mutagenesis

In order to identify consensus sequence motifs associated with either NADPH dependence or NADH dependence, we used FATCAT (Ye and Godzik, 2003) to perform a structure-based alignment of CaADH with four NADPH-dependent dehydrogenases and seven NADH-dependent dehydrogenases. The PDB accession codes of the NADPH-dependent enzymes were 1KEV, 3TQH, 2H6E and 3QWB; the accession codes of the NADH-dependent enzymes were 1JVB, 1F8F, 2DFV, 1H2B, 3QE3, 2JHF and 1KOL.

Based on this bioinformatics analysis, we began by purchasing a synthetic version of the CaADH gene from DNA 2.0 (Menlo Park, CA, USA), which had been codon-optimised for expression in *E.coli* and which encoded three amino acid substitutions (G198D/S199V/P201E). This gene was subcloned into pBAD(KpnI), to yield the expression vector pBAD(KpnI)-CaADH(DVE). In turn, pBAD(KpnI)-CaADH(DVE) was used as the template for constructing a quadruple mutant, CaADH (G198D/S199 V/P201E/Y218A)—abbreviated CaADH(DVEA)—by QuikChange mutagenesis. The primers used to introduce the Y218A mutation were CaADH_DVEA_for and CaADH_DVEA_rev (Table I). The final variant, CaADH(Y218A), was also constructed using QuikChange mutagenesis, with the template pBAD(KpnI)-CaADH and the primers CaADH_Y218A_for and CaADH_Y218A_rev (Table I).

Arabinose-induced expression and purification of CaADH (Y218A) were carried out as described earlier. The expression vectors for CaADH(DVE) and CaADH(DVEA) were used to transform electrocompetent *E.coli* LMG194 cells that also contained plasmid pGro7 (Takara Bio, Japan), for co-expression of the GroES-GroEL chaperone. For purification of these two proteins, the lysis buffer was supplemented with 5 mM ATP and 5 mM $MgCl_2$ (to facilitate removal of the chaperones). Activity assays with acetone were carried out as described earlier, except CaADH(DVEA) was present at a concentration of 300 nM. For determining the K_M of CaADH(DVEA) for NADH, the cofactor concentration was varied while acetone concentration was fixed at 75 mM. The maximum NADH concentration assayed was 700 μM; higher concentrations exceeded the measurable range of the spectrophotometer.

Results

Construction and expression of CaADH(DGGF)

While the structure of CaADH is yet to be determined, X-ray crystallography has been used to solve the structure of the homologous NADPH-dependent alcohol dehydrogenase from *Clostridium beijerinckii* (Korkhin *et al.*, 1998). The *C.beijerinckii* ADH (CbADH) shares 86% amino acid identity and 94% similarity with CaADH. Further, Korkhin *et al.* used their structure (PDB entry 1KEV) to predict four amino acid substitutions required to effect a switch from using NADPH to NADH in CbADH: G198D, S199G, R200G and Y218F. Given our goal of switching cofactor specificity, and the high level of sequence conservation, we began by introducing these four mutations into CaADH. However, the resulting variant, CaADH(DGGF), could only be expressed in insoluble inclusion bodies. No soluble protein was recovered under any of the conditions tested, thus no further characterisation of this variant was undertaken.

Construction of a CaADH(S199X) site-saturation library

We used the Phyre2 protein structure prediction server (Kelley and Sternberg, 2009) to build a homology model of CaADH (Fig. 2A). The server modelled CaADH using the CbADH structure, with 100% confidence and with 99% coverage. A high confidence indicates that the two proteins are homologous, whereas the per cent coverage indicates how much of the sequence is covered when generating the model. Similar to the CbADH structure, CaADH is predicted to be composed of two domains, the catalytic domain and the cofactor-binding domain. The cofactor-binding domain is characterised by the presence of a Rossmann fold, typical of nicotinamide-binding proteins (Rossmann *et al.*, 1974). To model the NADPH molecule and zinc ion into the active site, the CbADH NADPH- and zinc-bound structure (Korkhin *et al.*, 1998) was overlaid with the predicted CaADH structure. When aligned, these two structures had a root mean square deviation of 0.426 Å over 294 C_α atoms. The only difference between the native cofactor (NADPH) and the desired target (NADH) is the 2′-phosphate moiety of NADPH. Using our model, we identified serine 199 as a key residue that was predicted to interact with this phosphate, via a hydrogen bonding interaction (Fig. 2B).

With the aim of disfavouring NADPH binding, while retaining activity with NADH, we created a site-saturation library focussed on position 199. Primers with the degenerate NNK codon, where N represents any nucleotide base and K represents either guanine or thymine, were used to randomise S199. A total of 192 independent colonies were picked for activity screening, giving a 97% chance that all 20 amino acids would be sampled at the mutated position according to the GLUE-IT algorithm (Firth and Patrick, 2008).

High-throughput protein expression and screening

Crude lysates were used to test enzyme activity in a 96-well plate format. Initially, each plate was tested with an assay mixture containing

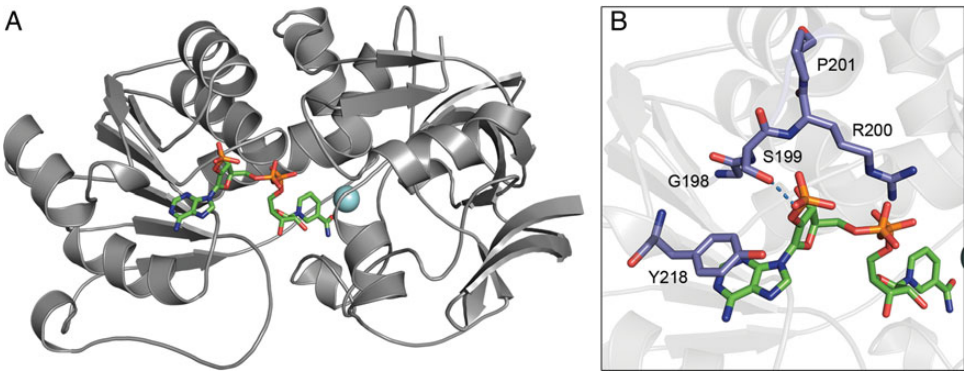

Fig. 2 (**A**) Homology model of a CaADH monomer. NADPH is shown as sticks with a green backbone, and Zn^{2+} is depicted as a light blue sphere. NADPH and Zn^{2+} were modelled into the structure by aligning it with the NADPH- and Zn^{2+}-bound CbADH. (**B**) The 2′-phosphate-binding site of CaADH showing the positions of all residues mutated in this study as sticks. The blue dashed line indicated the most likely position of a hydrogen bond between S199 and one of the oxygen atoms of the phosphate moiety.

Table II. Kinetic parameters for CaADH and variants identified from the CaADH(S199X) library screen

Variant	Acetone			Butanone			Acetoin		
	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} M^{-1})	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} M^{-1})	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} M^{-1})
CaADH	55 ± 3	1.0 ± 0.1	5.5×10^4	37 ± 2	1.8 ± 0.2	2.1×10^4	150 ± 8	98 ± 10	1.6×10^3
CaADH(S199A)	60 ± 1	0.9 ± 0.1	6.7×10^4	46 ± 2	0.8 ± 0.1	6.1×10^4	230 ± 10	120 ± 20	1.9×10^3
CaADH(S199C)	36 ± 1	1.2 ± 0.1	3.0×10^4	23 ± 1	1.5 ± 0.2	1.5×10^4	41 ± 4	45 ± 10	0.9×10^3
CaADH(S199G)	62 ± 1	1.6 ± 0.1	3.9×10^4	40 ± 1	2.4 ± 0.2	1.7×10^4	110 ± 10	230 ± 40	4.8×10^2
CaADH(S199R)	89 ± 5	1.1 ± 0.2	8.1×10^4	50 ± 2	1.9 ± 0.3	2.6×10^4	95 ± 4	34 ± 5	2.8×10^3
CaADH(S199W)	48 ± 3	2.0 ± 0.3	2.4×10^4	28 ± 1	3.6 ± 0.5	7.8×10^3	66 ± 7	190 ± 30	3.5×10^2

acetone and either NADPH or NADH. None of the variants from the CaADH(S199X) library showed observable activity with NADH as the cofactor. However, many of the library variants did show activity with acetone and NADPH. As we had previously demonstrated that CaADH has a wide substrate range (Köpke *et al.*, 2014), we decided to rescreen the library with NADPH and the additional substrates, butanone and acetoin (Fig. 1). When comparing the plates with different substrates, we observed that some variants showed a change in substrate specificity relative to CaADH. Given that our homology model (Fig. 2) showed a distance of 21 Å between Ser199 the catalytic zinc ion, this finding was unexpected. We identified five mutants that appeared to have differences in substrate specificities: CaADH (S199A), CaADH(S199C), CaADH(S199G), CaADH(S199R) and CaADH(S199W).

Steady-state kinetics of CaADH(S199X) variants

The five variants identified from the 96-well plate-based screen were solubly over-expressed and purified using immobilised metal affinity chromatography. The yields of CaADH(S199A), CaADH(S199G) and CaADH(S199C) were typically ~5 mg of purified protein per litre of culture, whereas CaADH(S199R) and CaADH(S199W) had lower yields of ~0.6 mg per litre of culture.

The activities of each purified variant were measured with NADPH and either acetone, butanone or acetoin (Table II). CaADH(S199C), CaADH(S199G) and CaADH(S199W) all showed lower catalytic efficiencies than CaADH (Fig. 3), with no significant changes in substrate preference. In contrast, CaADH(S199A) and CaADH(S199R) had higher catalytic efficiencies with all three substrates. Both of

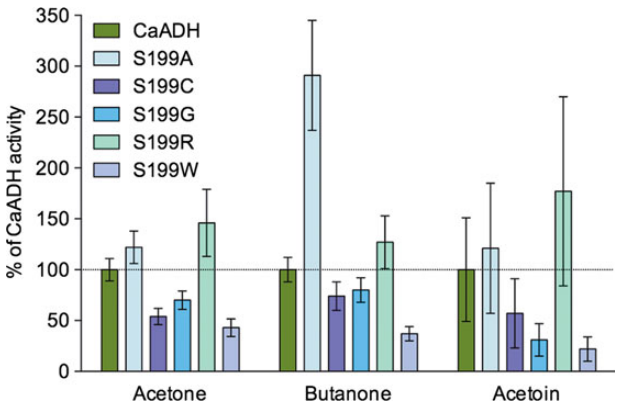

Fig. 3 The activities of selected S199X variants with acetone, butanone and acetoin relative to CaADH. Percentage ratios are calculated using the k_{cat}/K_M values shown in Table II, which are calculated from triplicates.

these variants also displayed changes in their substrate preferences. CaADH(S199A) displayed the most dramatic change: a 3-fold increase in activity with butanone relative to CaADH, and only minor increases in activity with acetone and acetoin (Fig. 3). These changes make CaADH(S199A) almost as active with butanone as it is with acetone. The primary effect is a 55% decrease in the K_M for butanone relative to CaADH, although this is also accompanied by a slight increase in k_{cat} (Table II). On the other hand, CaADH(S199R) showed increases in activity of 47 and 75% with acetone and acetoin, respectively, while activity with butanone remained comparable with wild-type CaADH (Fig. 3). The increase in activity with acetone was due

Table III. Kinetic parameters of selected CaADH variants for NADPH and NADH

Variant	NADPH			NADH		
	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} M^{-1})	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} M^{-1})
CaADH	68 ± 3	1.7 ± 0.3	3.9×10^7		ND	
CaADH(S199A)	50 ± 5	5.6 ± 1.3	8.9×10^6		ND	
CaADH(S199R)	32 ± 2	36 ± 6	8.9×10^5		ND	
CaADH(DVEA)		ND		12 ± 3	1000 ± 400	1.2×10^4

ND, not detected.

to a 60% increase in k_{cat} (Table II). In contrast, it was a 3-fold decrease in K_M that led to an overall improvement in activity towards acetoin. Each of the changes in activity for CaADH(S199A) and CaADH (S199R) was found to be statistically significant using an unpaired *t*-test.

As the original goal of our site-saturation mutagenesis was to disfavour NADPH binding, we were interested to determine the effects of the mutations on the K_M for NADPH. Our steady-state kinetic analysis showed that CaADH had a K_M of 1.7 μM for its cofactor (Table III). This is substantially lower than the NADPH K_M reported for the homologous *C.beijerinckii* enzyme, CbADH ($K_M^{NADPH} \approx 20$ μM; Ismaiel *et al.*, 1993). Predictably, K_M values for NADPH were changed in both CaADH(S199A) and CaADH(S199R). CaADH(S199A) showed a 3-fold increase in K_M relative to the wild-type enzyme. CaADH (S199R) showed a more substantial increase, with a K_M^{NADPH} of 36 μM.

These results showed that mutations distant from the substrate-binding site of CaADH could nevertheless affect the substrate specificity of the enzyme. The enzyme was also highly tolerant of substitutions at position S199. Amino acid substitutions that yielded the most soluble and/or active variants differed markedly in terms of side-chain charge, hydrophobicity and size (from glycine to tryptophan), with no apparent trend in tolerated residues at that site. Mutagenesis at this site also, unsurprisingly, had an effect on the ability of the enzyme to bind NADPH. Overall, the site-saturation mutagenesis and subsequence analyses of kinetics highlighted that localised changes in protein sequence and structure could have varied effects on function.

Modelling structural changes in CaADH variants

Next, we used the point mutation function of the RosettaBackrub webserver (Smith and Kortemme, 2008) to gain insights into how the S199A and S199R mutations may have affected the overall structure of the CaADH active site. This algorithm gives an ensemble of the most probable protein structures in an attempt to capture the conformational changes that proteins undergo when in solution. We generated ensembles of 10 structures each for CaADH, CaADH(S199A) and CaADH (S199R). To quantify global changes in each substrate-binding site, we measured their volumes in the structures from each ensemble, using POVME 2.0 (Durrant *et al.*, 2014). Although their mean cavity volumes were similar (145 ± 19 $Å^3$ for CaADH, 145 ± 36 $Å^3$ for CaADH (S199A) and 150 ± 22 $Å^3$ for CaADH(S199R); mean ± standard deviation), there were differences in the predicted ranges of cavity sizes sampled by the three variants. The analysis suggested that CaADH has the most rigid substrate-binding cavity (Fig. 4). CaADH(S199A) is predicted to sample the most variable range of cavity conformations, while CaADH(S199R) is intermediate (Fig. 4). It is tempting to speculate that these apparent increases in conformational diversity may explain why CaADH(S199A) and CaADH(S199R) are particularly active with the large, 4-carbon substrates butanone and acetoin, respectively.

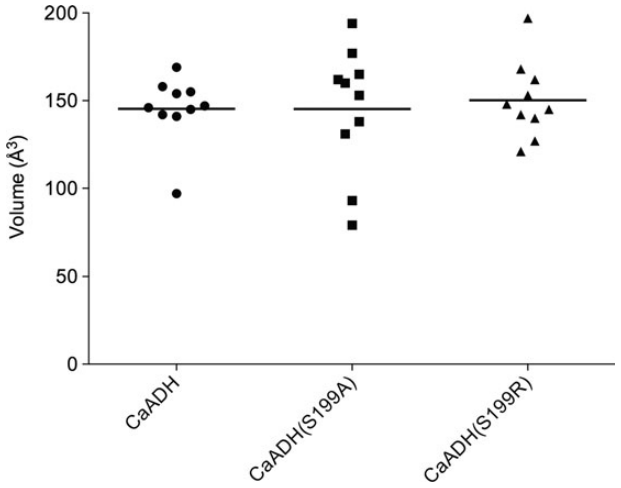

Fig. 4 The range of volumes of the predicted active site cavities of CaADH, CaADH(S199A) and CaADH(S199R). The horizontal line indicates the mean cavity volume.

Consensus-guided mutagenesis to switch cofactor specificity

While substitution of S199 led to unexpected changes in substrate specificity, we still had not achieved our goal of engineering an NADH-dependent CaADH variant. Therefore, we performed a structure-based alignment of four NADPH-dependent and seven NADH-dependent dehydrogenases. Similar to the previous analysis of CbADH (Korkhin *et al.*, 1998), this identified residues G198, S199, P201 and Y218 (but not R200) as important for cofactor specificity in CaADH (Fig. 2B). In the NADPH-dependent enzymes, position 198 was small and uncharged, while it was exclusively acidic in NADH-dependent enzymes. Position 199 was found to be hydrophobic (V/I/L) in NADH-dependent enzymes, while the other two positions were more variable but most commonly glutamate in position 201, and alanine, proline or serine in position 218.

Compiling these data, we began by synthesising and testing the triple mutant CaADH(DVE), containing the G198D/S199V/P201E substitutions. The His_6-tagged protein could be expressed solubly when the GroES-GroEL chaperone was co-expressed. It remained highly active with NADPH as the cofactor and acetone as the substrate ($k_{cat} = 39 \pm 1$ s^{-1}; $K_M^{acetone} = 0.6 \pm 0.1$ mM). Promisingly, CaADH (DVE) showed trace activity for the reduction of acetone in the presence of NADH. However, this was barely above background and unable to be quantified. Therefore, we introduced the additional Y218A mutation to generate CaADH(G198D/S199V/P201E/Y218A), abbreviated CaADH(DVEA). When co-expressed with GroES-GroEL, yields of purified CaADH(DVEA) were typically 1–3 mg/l of culture.

CaADH(DVEA) had no detectable activity with NADPH as the cofactor. On the other hand, the quadruple mutant had improved activity with NADH compared with CaADH(DVE), to a level that allowed kinetic parameters to be determined for the reduction of acetone. Overall, the catalytic efficiency of the enzyme was $k_{cat}/K_M = 1.6 \times 10^2\ s^{-1}\ M^{-1}$, which was ~350-fold lower than the efficiency of wild-type CaADH with NADPH and acetone. This was due to both a decreased turnover number ($k_{cat} = 2.3 \pm 0.1\ s^{-1}$ compared with $55 \pm 3\ s^{-1}$ for CaADH) and an increased Michaelis constant ($K_M^{acetone} = 14 \pm 1$ mM compared with 1.0 ± 0.1 for CaADH). At 1.0 ± 0.4 mM, the K_M for NADH was also almost 600-fold greater than the CaADH K_M for NADPH (1.7 μM; Table III). The large error associated with the measurement of the CaADH(DVEA) K_M^{NADH} was indicative of the fact that we were unable to saturate the enzyme with cofactor, without exceeding the absorbance range of the spectrophotometer. Finally, we introduced the Y218A mutation into CaADH, to control for the possibility that it was the sole mutation required for the cofactor switch. CaADH(Y218A) was solubly expressed without the need to co-express GroES-GroEL, but it was inactive with NADH.

Discussion

In this study, we used a variety of mutagenesis approaches in our attempts to change the cofactor specificity of CaADH. Ultimately, we achieved a complete switch in cofactor specificity with a structure-guided consensus mutagenesis approach. Without an experimentally determined structure of the CaADH(DVEA) variant, it is difficult to be certain why it is now completely specific for NADH. However, we can draw some conclusions based on the characteristics of the residues in the native enzyme, and those that have replaced them. The substitution of G198 with aspartate is an intuitive change; the introduction of a negatively charged side chain at this position (Fig. 2B) should disfavour the binding of the negatively charged 2′-phosphate of NADPH. The switch from serine to valine at position 199 was a less obvious substitution, and one that was not identified in isolation from the CaADH(S199X) library. Modelling suggested a hydrogen bond between the side-chain hydroxyl of S199 and one of the oxygens of the NADPH 2′-phosphate (Fig 2B); this would be absent in CaADH (DVEA). The side chain of valine is also much more hydrophobic than that of serine (Monera *et al.*, 1995). It is therefore likely to favour an orientation towards the new NADH cofactor and the hydrophobic core of the protein, rather than allowing the loop on which it is located to reorient towards bulk solvent. Similar to G198D, the P201E substitution presumably increased the overall negative charge of the cofactor-binding pocket, while also decreasing its size (thus favouring NADH over NADPH). However, these three mutations were not sufficient to effect a switch in cofactor usage: while they imparted trace activity with NADH, variant CaADH(DVE) was still highly active with NADPH.

In the highly homologous CbADH structure, Y218 plays a critical role in binding NADPH. A comparison of the apo- and cofactor-bound structures revealed a ~120° rotation of the Y218 side chain upon binding, enabling it to form stacking interactions with the adenine moiety of NADPH as well as a hydrogen bond from the side-chain hydroxyl to one of the 2′-phosphate oxygens (Korkhin *et al.*, 1998). While the Y218A substitution alone was insufficient to induce a cofactor switch, the additive effects of all four mutations in the CaADH(DVEA) variant resulted in an enzyme that was strictly NADH-dependent. On the other hand, Y218 was mutated to phenylalanine in the insoluble variant, CaADH(DGGF). Together, these results show that the stacking interaction from Y218 is not essential for NAD(P)H binding, and that wholesale restructuring of the 2′-phosphate-binding pocket is required to switch cofactor dependence.

Our results add to the body of literature to suggest that there are no 'golden rules', as yet, for the design challenge of interconverting cofactor preference between NADPH and NADH. One early attempt to define these rules focussed on a conserved fingerprint sequence that was identified on the loop between the first β strand and α-helix of the Rossmann fold in a number of dinucleotide binding enzymes (Wierenga *et al.*, 1985). Alignments suggested that this sequence was GxGxxG in NADH-dependent enzymes, but GxGxxA in NADPH-binding enzymes (Hanukoglu and Gutfinger, 1989). However, an experimental test of this hypothesis with NADPH-dependent glutathione reductase showed that the alanine-to-glycine substitution only decreased NADPH-dependent activity by 13%, albeit while imparting a 6-fold increase in activity with NADH. Six further mutations were required to create a variant with higher activity with NADH than NADPH (Scrutton *et al.*, 1990), and the inevitable expansion of sequence databases quickly showed that the GxGxxG/A fingerprint was not a reliable predictor of cofactor preference (Baker *et al.*, 1992). Indeed, CaADH is one such exception, containing the GxGxxG motif. On the other hand, we found that it was critical to introduce the G198D substitution in order to switch the cofactor preference of CaADH. In the cofactor-binding Rossmann fold, G198 is the final residue in the second β strand of the βαβαβ motif. It has been observed that NADH-binding proteins typically possess an acidic residue at this position (Wierenga *et al.*, 1985), and our results suggest it is an important target with which to begin a mutagenesis campaign.

For one particular family of oxidoreductases, the KARIs, a generic set of rules to switch the cofactor specificity from NADPH to NADH has been proposed (Brinkmann-Chen *et al.*, 2013). In this study, the authors observed that the loop between the second α helix and β strand is the determinant of cofactor specificity in these enzymes. They showed that this loop is typically one of three lengths (6, 7 or 12 residues), and allowing for this they were able to identify amino acid commonalities found between all loop lengths that were consistent with cofactor preference. Based on this, the authors propose a number of amino acid substitutions to impart a cofactor switch. Interestingly, three of the four substitutions needed for the cofactor switch in CaADH (G198D, S199V and P201E) are located on the corresponding loop, indicating that this loop is important for cofactor specificity in a number of Rossmann-fold containing enzymes. Having said that, the residues that Brinkmann-Chen *et al.* found to be conserved in KARIs are not conserved in CaADH, nor are the lengths of the β2αB loop conserved within the ADHs we used for our consensus work, indicating that the strategy developed in the study is not applicable to ADHs. Nevertheless, this study indicates that focussing on a specific family of enzymes, rather than an entire class or all those with a shared fold, is a more successful strategy for developing rules to switch cofactors.

A recently described algorithm, Cofactory (Geertz-Hansen *et al.*, 2014), uses primary sequence information to predict the cofactor-binding specificities of Rossmann folds. When we retrospectively entered the CaADH sequence into this webserver, it could not predict whether the enzyme bound NADH or NADPH. However, entering the sequence for CaADH(DVEA) led to the correct prediction that the enzyme would be NADH-dependent. It would be interesting to extrapolate from this observation and directly test the utility of Cofactory for guiding future design work.

The most unexpected finding of this work was that mutagenesis at position S199 alone altered substrate specificity, rather than cofactor preference. The large distance between S199 and the substrate-binding site (~20 Å; Fig. 2A) implied that structural changes at the point of

mutation were being translated through the backbone to distal regions of the protein. Our analysis of a high-confidence homology model using a combination of RosettaBackrub and POVME suggested that the S199A and S199R mutations each allowed the CaADH active site to adopt a greater range of conformations, which varied significantly in total volume (Fig. 4). We hypothesised that these differences in conformational diversity might explain why CaADH (S199A) and CaADH(S199R) showed increased activity with 4-carbon substrates. However, it is less clear why the S199R mutation has improved the K_M for acetoin (but not butanone), while S199A has had the opposite effect (improving the K_M for butanone; Table II). Ligand-bound structures will be necessary to elucidate the long-range structural effects of the S199A and S199R mutations; this work is under way in our laboratory.

In their highly influential survey of the enzyme engineering literature, Morley and Kazlauskas found that mutations between 5 and 10 Å away from the catalytic site were most effective at changing enantioselectivity and substrate specificity, whereas both close and distal mutations (between 5 and 32 Å) were effective at changing catalytic activity and thermal stability (Morley and Kazlauskas, 2005). This study has provided a counter-example, and emphasised the roles that distal mutations can play in altering substrate specificity.

Funding

This work was supported by a Smart Ideas grant from the New Zealand Ministry of Business, Innovation and Employment. D.J.M is supported by the University of Otago's Fanny Evans Postgraduate Scholarship for Women. W.M.P. is the grateful recipient of a Rutherford Discovery Fellowship. Funding to pay the Open Access publication charges for this article was provided by a Smart Ideas grant from the New Zealand Ministry of Business, Innovation and Employment.

References

Abrini,J., Naveau,H. and Nyns,E.J. (1994) *Arch. Microbiol.*, **161**, 345–351.
Baker,P.J., Britton,K.L., Rice,D.W., Rob,A. and Stillman,T.J. (1992) *J. Mol. Biol.*, **228**, 662–671.
Bastian,S., Liu,X., Meyerowitz,J.T., Snow,C.D., Chen,M.M.Y. and Arnold,F. H. (2011) *Metab. Eng.*, **13**, 345–352.
Bennett,B.D., Kimball,E.H., Gao,M., Osterhout,R., Van Dien,S.J. and Rabinowitz,J.D. (2009) *Nat. Chem. Biol.*, **5**, 593–599.
Brinkmann-Chen,S., Flock,T., Cahn,J.K.B., Snow,C.D., Brustad,E.M., McIntosh,J.A., Meinhold,P., Zhang,L. and Arnold,F.H. (2013) *Proc. Natl Acad. Sci. U.S.A.*, **110**, 10946–10951.
Chen,R.D., Greer,A. and Dean,A.M. (1995) *Proc. Natl Acad. Sci. U.S.A.*, **92**, 11666–11670.
Durrant,J.D., Votapka,L., Sorensen,J. and Amaro,R.E. (2014) *J. Chem. Theory Comput.*, **10**, 5047–5056.
Firth,A.E. and Patrick,W.M. (2008) *Nucleic Acids Res.*, **36**, W281–W285.
Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D. and Bairoch,A. (2003) *Nucleic Acids Res.*, **31**, 3784–3788.
Geertz-Hansen,H.M., Blom,N., Feist,A.M., Brunak,S. and Petersen,T.N. (2014) *Proteins*, **82**, 1819–1828.
Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison,C.A. and Smith, H.O. (2009) *Nat. Methods*, **6**, 343–345.
Glieder,A. and Meinhold,P. (2003) *Methods Mol. Biol.*, **230**, 157–170.
Hanukoglu,I. and Gutfinger,T. (1989) *Eur. J. Biochem.*, **180**, 479–484.
Ismaiel,A.A., Zhu,C.X., Colby,G.D. and Chen,J.S. (1993) *J. Bacteriol.*, **175**, 5097–5105.
Kelley,L.A. and Sternberg,M.J.E. (2009) *Nat. Protoc.*, **4**, 363–371.
Köpke,M., Held,C., Hujer,S., *et al.* (2010) *Proc. Natl Acad. Sci. U.S.A.*, **107**, 13087–13092.
Köpke,M., Mihalcea,C., Liew,F.M., Tizard,J.H., Ali,M.S., Conolly,J.J., Al-Sinawi,B. and Simpson,S.D. (2011) *Appl. Environ. Microbiol.*, **77**, 5467–5475.
Köpke,M., Gerth,M.L., Maddock,D.J., Mueller,A.P., Liew,F., Simpson,S.D. and Patrick,W.M. (2014) *Appl. Environ. Microbiol.*, **80**, 3394–3403.
Korkhin,Y., Kalb,A.J., Peretz,M., Bogin,O., Burstein,Y. and Frolow,F. (1998) *J. Mol. Biol.*, **278**, 967–981.
Lerchner,A., Jarasch,A., Meining,W., Schiefner,A. and Skerra,A. (2013) *Biotechnol. Bioeng.*, **110**, 2803–2814.
Monera,O.D., Sereda,T.J., Zhou,N.E., Kay,C.M. and Hodges,R.S. (1995) *J. Pept. Sci.*, **1**, 319–329.
Morley,K.L. and Kazlauskas,R.J. (2005) *Trends Biotechnol.*, **23**, 231–237.
Pick,A., Ott,W., Howe,T., Schmid,J. and Sieber,V. (2014) *J. Biotechnol.*, **189**, 157–165.
Ragsdale,S.W. and Pierce,E. (2008) *Biochim. Biophys. Acta*, **1784**, 1873–1898.
Rane,M.J. and Calvo,K.C. (1997) *Arch. Biochem. Biophys.*, **338**, 83–89.
Rossmann,M.G., Moras,D. and Olsen,K.W. (1974) *Nature*, **250**, 194–199.
Scrutton,N.S., Berry,A. and Perham,R.N. (1990) *Nature*, **343**, 38–43.
Siemerink,M.A.J., Kuit,W., Contreras,A.M.L., Eggink,G., van der Oost,J. and Kengen,S.W.M. (2011) *Appl. Environ. Microbiol.*, **77**, 2582–2588.
Smith,C.A. and Kortemme,T. (2008) *J. Mol. Biol.*, **380**, 742–756.
Tracy,B.P., Jones,S.W., Fast,A.G., Indurthi,D.C. and Papoutsakis,E.T. (2012) *Curr. Opin. Biotechnol.*, **23**, 364–381.
Wierenga,R.K., De Maeyer,M.C.H. and Hol,W.G.J. (1985) *Biochemistry*, **24**, 1346–1357.
Ye,Y.Z. and Godzik,A. (2003) *Bioinformatics*, **19**, II246–II255.