Extending enzyme molecular recognition with an expanded amino acid alphabet

Claire L. Windle, Katie J. Simmons, James R. Ault, Chi H. Trinh, Adam Nelson, Arwen R. Pearson, and Alan Berry

Natural enzymes are constructed from the 20 proteogenic amino acids, which may then require posttranslational modification or the recruitment of coenzymes or metal ions to achieve catalytic function. Here, we demonstrate that expansion of the alphabet of amino acids can also enable the properties of enzymes to be extended. A chemical mutagenesis strategy allowed a wide range of noncanonical amino acids to be systematically incorporated throughout an active site to alter enzymic substrate specificity. Specifically, 13 different noncanonical side chains were incorporated at 12 different positions within the active site of N-acetylenuraminic acid lyase (NAL), and the resulting chemically modified enzymes were screened for activity with a range of aldehyde substrates. A modified enzyme containing a 2,3-dihydroxypropyl cysteine at position 190 was identified that had significantly increased activity for the aldol reaction of erythrose with pyruvate compared with the wild-type enzyme. Kinetic investigation of a saturation library of the canonical amino acids at the same position showed that this increased activity was not achievable with any of the 20 proteogenic amino acids. Structural and modeling studies revealed that the unique shape and functionality of the noncanonical side chain enabled the active site to be remodeled to enable more efficient stabilization of the transition state of the reaction. The ability to exploit an expanded amino acid alphabet can thus heighten the ambitions of protein engineers wishing to develop enzymes with new catalytic properties.

Enzymes are phenomenally powerful catalysts that increase reaction rates by up to 10^{18}-fold (1, 2), and a new era of enzyme applications has been opened by the advancement of protein engineering and directed evolution to provide new, or improved, enzymes for industrial biocatalysis. Enzymes are attractive catalysts because they are highly selective, carrying out regio-, chemo-, and stereoselective reactions that are challenging for conventional chemistry. Moreover, enzymes are efficient catalysts, function under mild conditions with relatively nontoxic reagents, and enable the production of relatively pure products, minimizing waste generation. In recent years, there has been much success in engineering enzymes for desired reactions (3–5) using methods such as rational protein engineering (6–8), directed evolution (9–11), and, most recently, computational enzyme design (12–15).

Enzymes found in Nature achieve catalysis using active sites generally composed of only 20 canonical amino acids, which are encoded at the genetic level, plus the rarer selenocysteine and 1-pyrrolysine. However, many enzymes also rely on one or more of 27 small organic cofactor molecules and/or 13 metal ions for their function (16). In addition, in some cases, Nature has exploited noncanonical amino acids (Ncas) in catalysis to extend its catalytic repertoire: for example, the quinones TPQ, LTQ, TTQ, and CTQ, respectively, in adenosylmethionine decarboxylases (17); formyl glycine residues in hydrogenase, and quinohemoprotein amine dehydrogenase (17); respectively, in amine oxidase, lysyl oxidase, methylamine dehydrogenase (18); and Alan Berry (19); and 4-methylideneimidazole-5-one (MIO) in N-acetylneuraminic acid lyase (NAL), and the resulting chemically modified enzymes were screened for activity with a range of aldehyde substrates. A modified enzyme containing a 2,3-dihydroxypropyl cysteine at position 190 was identified that had significantly increased activity for the aldol reaction of erythrose with pyruvate compared with the wild-type enzyme. Kinetic investigation of a saturation library of the canonical amino acids at the same position showed that this increased activity was not achievable with any of the 20 proteogenic amino acids. Structural and modeling studies revealed that the unique shape and functionality of the noncanonical side chain enabled the active site to be remodeled to enable more efficient stabilization of the transition state of the reaction. The ability to exploit an expanded amino acid alphabet can thus heighten the ambitions of protein engineers wishing to develop enzymes with new catalytic properties.

Significance

The remarkable power of enzymes as catalysts is derived from the precise spatial positioning of amino acids as a result of a polypeptide folding into its native, active fold. Protein engineers have a wide arsenal of tools available to alter the properties of enzymes but, until recently, have been limited to replacement of amino acids with one of the other naturally occurring proteogenic amino acids. Here we describe a protein engineering approach to introduce a noncanonical amino acid that results in altered substrate specificity of an aldolase to produce a novel activity that cannot be achieved by simple substitution with any of the canonical amino acids.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID code 5LKY).

1To whom correspondence may be addressed. Email: a.berry@leeds.ac.uk or a.s.nelson@leeds.ac.uk.

2Present address: Hamburg Centre for Ultrafast Imaging, Universität Hamburg, Hamburg 22761, Germany.

3Present address: Institute for Nanostructure and Solid State Physics, Universität Hamburg, Hamburg 22761, Germany.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616816114/-/DCSupplemental.
chemical modification procedure to introduce the Nca relies on the
were selected as positions for Nca incorporation (Fig. 2). The
throughout the active site on the reaction catalyzed.
(2-aminoethyl cysteine) in place of the catalytic Lys-165 retains
and we have shown that an enzyme bearing the Nca
specificity and stereochemistry of this enzyme is highly malleable
pyruvate to form
NAL catalyses the aldol reaction of
Windle et al. PNAS
Fig. 2.
Amino acid positions modified, modifying thiols, and screening substrates. Cysteine residues were introduced individually into
Production of Modified Enzymes.
Results
Production of Modified Enzymes. N-acetylneuraminic acid lyase (NAL) catalyses the aldol reaction of N-acetyl-d-mannosamine and
pyruvate to form N-acetylneuraminic acid (Fig. L4). The substrate
specificity and stereochemistry of this enzyme is highly malleable
through protein engineering and directed evolution (34, 35),
and we have shown that an enzyme bearing the Nca γ-thialysine
(2-aminoethyl cysteine) in place of the catalytic Lys-165 retains
activity (36). Here we have explored the effect of introducing Ncas
throughout the active site on the reaction catalyzed.
Twelve residues in the active site of Staphylococcus aureus NAL
were selected as positions for Nca incorporation (Fig. 2). The
chemical modification procedure to introduce the Nca relies on the
conversion of a cysteine into dehydroalanine by a bis-alkylation/
elimination sequence (27, 36), followed by Michael addition of a thiol
to introduce the new Nca side chain. The S. aureus NAL naturally
contains no cysteines, making it an ideal target protein for chemical
introduction of Ncas (36). The residues chosen for modification were
picked based on their proximity to the aldehyde substrate binding
pocket in the previously solved structure of an Escherichia coli NAL
variant in complex with 4-epi-N-acetylneuraminic acid (PDB ID code
4BWL) (37) (Fig. 3). The equivalent positions in the S. aureus NAL
based on structural alignment were then mutated into cysteine
residues by site-directed mutagenesis, and expressed and purified
as previously described (36). The Ncas were then incorporated by
first converting the cysteine to dehydroalanine (Dha) using 2,5-
dibromohexan-1,6-diamide and then Michael addition with 13 differ-
thiols resulting in installation of 13 different Ncas at each of the
12 positions (27) (Fig. 2). The thiols chosen encompass a wide
variety of noncanonical side chains, some mimicking elongated
versions of canonical amino acids and some incorporating func-
tionalities that are either absent or uncommon in the canonical
amino acids. Incorporation was carried out on a small (2.5 mg) scale
under denaturing conditions to ensure the cysteine residue was ac-
cessible for modification before the protein was refolded. This
procedure was previously shown to generate active, modified,
refolded enzyme (36). The conversions of Cys→Dha and Dha→Nca
were both monitored by ESI mass spectrometry (Figs. S1 and S2).

Screening and Activity. NAL is highly specific for its ketone donor
(pyruvate) while accepting a range of aldehyde acceptors
(38, 39). However, the wild-type enzyme has a strong preference
for longer aldehydes (38–40): C4 aldehydes are generally poorly
accepted (38) and C3 aldehydes are not substrates (39). The
differently modified enzymes at each of the positions in the
polypeptide chain were screened in 96-well plates for altered
activity and specificity in condensing pyruvate with a range of
aldehyde substrates with different lengths (C4 to C6), stereo-
chemistry, and functional groups (Fig. 2). Enzyme activity was
assessed using a variation of the established thiobarbituric acid
(TBA) assay (41). Aldol reaction of pyruvate with the aldehydes
in Table 1 generates 2-oxo-4,5-dihydroxy (or 2-oxo-4-acetamido-
5-hydroxy) carboxylic acids. Subsequent cleavage of these prod-
ucts with periodate generates 1,3-dicarboxyls that react with
TBA to generate an intense pink chromophore that can be
quantified at 550 nm. Absorbance at 550 nm for each well was
measured after 16 h condensation. The results were analyzed by
comparing activity of the wild-type (unmodified) NAL with that of
the Nca-containing enzyme with the same aldehyde substrate. An
electric analysis (Fig. 4) shows the result of screening NAL

<table>
<thead>
<tr>
<th>Cysteine variants</th>
<th>Thiols</th>
</tr>
</thead>
<tbody>
<tr>
<td>L247C I251C</td>
<td></td>
</tr>
<tr>
<td>L142C F190C</td>
<td></td>
</tr>
<tr>
<td>Y252C G207C</td>
<td></td>
</tr>
<tr>
<td>E192C I139C</td>
<td></td>
</tr>
<tr>
<td>F172C N170C</td>
<td></td>
</tr>
<tr>
<td>T209C S208C</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Amino acid positions modified, modifying thiols, and screening substrates. Cysteine residues were introduced individually into S. aureus NAL by site-
directed mutagenesis at each position shown. After conversion into dehydroalanine, each position was modified separately with the 13 thiols shown. The
resulting enzymes were screened using the thiobarbituric acid assay for activity in condensing pyruvate with the aldehydes shown.
enzymes bearing Ncas at position 190 against a range of aldehydes. Whereas most modifications either generated enzymes with lower activity than the wild-type (Fig. 4, blue) or with the same level of activity (white), three modified enzymes had activities for the condensation of pyruvate with erythrose considerably greater than that of wild-type enzyme (purple). The improved enzymes had 2,3-dihydroxypropyl cysteine (Hpc), 4-hydroxybutyl cysteine (Hbc), or 2-hydroxypropyl cysteine (Hpc), with the 2,3-dihydroxypropyl cysteine having the greatest activity. Interestingly, all of the active enzymes have hydroxylated Nca side chains.

To confirm that Phe190Dpc NAL catalyzes the aldol reaction of pyruvate with erythrose to generate 3-deoxy-2-heptulosonic acid (DHA) (PDB ID code 12108643), a large-scale reaction was incubated for 48 h with Phe190Dpc NAL, and the resulting diastereomeric mixture of product (DHA) was purified by anion exchange chromatography. The 500 MHz 1H, COSY, and TOCSY NMR spectra (Fig. S3) were in agreement with reported data (42) for the expected product (DHA) and showed that a ~75:25 mixture of C4-configured products had been formed (Table S1).

Table 1. Steady-state kinetic parameters for the aldol condensation of erythrose and pyruvate and the aldol condensation of ManNac and pyruvate for both wild-type and F190Dpc NAL

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>k_{cat, app}</th>
<th>mM</th>
<th>k_{cat, app}/K_m, min^{-1}/mM^{-1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>F190Dpc</td>
<td>Erythrose</td>
<td>7.6 ± 0.56</td>
<td>4.4 ± 0.81</td>
<td>1.7</td>
</tr>
<tr>
<td>F190Dpc</td>
<td>ManNac</td>
<td>0.22 ± 0.01</td>
<td>2.5 ± 0.43</td>
<td>0.09</td>
</tr>
<tr>
<td>Wild type</td>
<td>Erythrose</td>
<td>0.5 ± 0.05</td>
<td>3.0 ± 0.9</td>
<td>0.17</td>
</tr>
<tr>
<td>Wild type</td>
<td>ManNac</td>
<td>0.8 ± 0.05</td>
<td>3.1 ± 0.7</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Initial rates of enzyme reactions were measured in duplicate using the thioharbitoric acid assay at a fixed concentration of pyruvate of 80 mM while varying the concentration of aldheyde (0.8–15 mM). Data were fitted to the Michaelis–Menten equation to estimate the apparent kinetic parameters. Parameter values ± SE of the fit are shown.
modeling of the reaction mechanism of the N substrate that, as expected, the DHA, being shorter than the natural both the wild-type and Phe190Dpc NALs. The models revealed out on the Schiff bases formed between the product DHA and the structural basis for the change of specificity. b o u n d, a n d s o w e t u r n e d t o m o l e c u l a r m o d e l i n g t o r y t o e l u c i d a t e the modified Phe190Dpc NAL with the reaction product DHA fortunately, we were unable to determine the crystal structure of help to stabilize the transition state of the reaction. Un-
suggested that Thr167 binds to the aldehyde oxygen atom and
same interactions with Glu-192 of the enzyme (37) (Figs. S7 and
A550 between the reaction with the modified enzyme and the wild-type enzyme
assayed using the TBA assay and the absorbance read at 550 nm. The difference in
of enzyme (0.2 mg; 5.8 nmol) with aldehyde (8 mM) and pyruvate (80 mM) were
are blue with the intensity proportional to the activity difference. Values greater
than zero (modified enzyme is more active than wild type) are similarly colored in pink. Modified enzymes are ranked by order of activity with pyruvate and erythrose.
configuration. It is highly unlikely that the Michael addition of racemic 2,3-dihydroxy propanethiol to the unfolded protein was stereocontrolled; rather, we hypothesize, as previously (36), that only one stereoisomer of the modified protein can refold correctly and is recovered after gel filtration. The Dpc190 side chain is positioned between Glu192 and Asp141 and points into the active site where it extends further (average 0.8 ± 0.1 Å from all four subunits) into the active site than the corresponding phenylalanine in the wild-type structure. The noncanonical side chain exhibits a high degree of flexibility in the active site, adopting different conformations in all four subunits (Fig. S5). In subunit A, the Dpc side chain was modeled in two different conformations that resulted from a rotation around the Cα–Cβ bond. In both subunits A and B, the hydroxyl groups of the Dpc side chain make hydrogen bond interactions to the side chains of Glu192 and Asp141 (Fig. 6B). Another notable difference between wild-type and Phe190Dpc NAL was at the region 167–169. In the Phe190Dpc structure, the removal of the bulky hydrophobic phenylalanine allows the backbone chain containing residues Thr167, Ala168, and Pro169 to move closer to the cata-
ytic Lys165 in the active site (Fig. S6). This results in the Cα of Thr167 in the Phe190Dpc structure being displaced on average 1.0 ± 0.1 Å and adopting a different rotamer compared with the wild-type structure (Fig. 6A and B and Fig. S6). Previous QM/MM modeling of the reaction mechanism of the E. coli NAL (37) has suggested that Thr167 binds to the aldehyde oxygen atom and helps to stabilize the transition state of the reaction. Unfortunately, we were unable to determine the crystal structure of the modified Phe190Dpc NAL with the reaction product DHA bound, and so we turned to molecular modeling to try to elucidate the structural basis for the change of specificity.
Computational energy minimization experiments were carried out on the Schiff bases formed between the product DHA and both the wild-type and Phe190Dpc NALs. The models revealed that, as expected, the DHA, being shorter than the natural substrate N-acetylneuraminic acid (Neu5Ac), does not make the same interactions with Glu-192 of the enzyme (37) (Figs. S7 and S8). Because of our finding that both 4R and 4S DHA were formed by Phe190Dpc NAL, both stereoisomers were modeled into the active sites. The minimizations revealed that C1–C3 of either C4 diastereoisomer of DHA were bound in the same way in both the wild-type and Phe190Dpc models. However, there were significantly different interactions at the other end (C4–C7) of the product. Tyr252 lies slightly further from the product in both 4(R) and 4(S) DHA-Phe190Dpc NAL models, meaning that it can no longer hydrogen bond with the product C5 hydroxyl (∼4.7 Å compared with 3.1 Å in the wild-type model; Fig. 6C and D). In addition, Thr167 lies ∼1.6 Å closer to the C4 hydroxyl of the product in the DHA-Phe190Dpc model than in the wild-type model: here, the difference in the preferred rotamer of Thr167 may enable better stabilization of the enzyme transition state. Dpc does not interact directly with DHA in the Phe190Dpc model; it forms an entirely different network of interactions between the carboxylate side chains of Asp141 and Glu192 in the wild-type model, which brings the Dpc 3.0 Å closer to the

![Fig. 4. Screening data for enzymes modified at position 190. Overnight reactions of enzyme (0.2 mg; 5.8 nmol) with aldehyde (8 mM) and pyruvate (80 mM) were assayed using the TBA assay and the absorbance read at 550 nm. The difference in A550 between the reaction with the modified enzyme and the wild-type enzyme were calculated. Values below zero (modified enzyme is less active than wild type) are blue with the intensity proportional to the activity difference. Values greater than zero (modified enzyme is more active than wild type) are similarly colored in pink. Modified enzymes are ranked by order of activity with pyruvate and erythrose.](image-url)

![Fig. 5. Comparison of activities for the saturation library at position 190 against F190Dpc variant. Activity of NAL variants bearing the 20 canonical amino acids at position 190 was measured using the TBA assay and compared with NAL bearing the Nca Dpc at position 190. Initial reaction rates were determined using the TBA assay under conditions where the time course of product formation is linear (Fig. S4). Steady-state kinetic parameters were determined at a fixed pyruvate concentra-
tion of 80 mM and erythrose concentrations between 0.8 and 15 mM. Each reaction contained 18 μg (0.52 nmol) of enzyme and were carried out in duplicate and data were fitted to the Michaelis-Menten equation. The fitted value ± SE of the fit is shown. (A) Comparison of kcat values; (B) Steady-state parameters.](image-url)
product than the phenylalanine in the wild-type model (Fig. S7). Thus, molecular modeling suggests that the newly introduced Dpc side chain at residue 190 alters the substrate specificity not by direct interaction with DHA but rather by altering the conformation and interaction network in the active site, resulting in better transition state stabilization. This observation is entirely consistent with the kinetic data (Table 1) where the change in substrate specificity measured by $k_{cat}/K_{m(app)}$ is almost entirely due to alterations in $k_{cat(app)}$ rather than in $K_{m(app)}$.

Discussion

Aldolases are an important class of biocatalysts that are finding increasing uses in the synthesis of complex compounds (33); they are particularly useful in that they can create two stereochemical centers increasing uses in the synthesis of complex compounds (33); they are particularly useful in that they can create two stereochemical centers. The recent advances in computational (12, 13) and (semi)rational methods (6) for enzyme redesign, coupled with the development of recent powerful chemical mutagenesis methods (51) and the enormous variety of Ncas side chains and chemistries, will open the way to engineer enzymes with catalytic functions that are not found in Nature.

Materials and Methods

Further details may be found in SI Materials and Methods.

Chemical Modification Using Various Thiols. The conversion of the cysteine residues (introduced by site-directed mutagenesis) to dehydrololine was carried out as previously described (36) using 2,5-dibromohexan-1,6-diamide synthesized (27). ESI-MS was used to check for complete conversion into the Nca-containing protein. Small-scale (2.5 mg) modifications were used for the screening process. For detailed characterization, large-scale modifications (up to 50 mg) were performed. Modified enzymes were refolded by first dialyzing into sodium phosphate buffer (50 mM, pH 7.4) containing urea (6 M) to remove excess modification reagents, followed by dialysis into buffer without urea (36). Large-scale protein modifications were purified using size-exclusion chromatography performed using an ÄKTA Prime purification system (GE Healthcare Life Sciences) with a Superdex 200 column. Protein (8 mg/mL, 5 mL) was injected onto the column, which was run in Tris-HCl buffer (50 mM, pH 7.4) at 2 mL/min.

Screening Using the TBA Assay. Modified enzymes were screened for aldol reaction activity with pyruvate and a variety of aldehydes. Reactions in 96 deep-well plates contained pyruvate (100 μL, final concentration 80 mM), aldehyde (100 μL, final concentration 8 mM), and modified enzyme [50 μL of 1.0 mg/mL (1.5 nmol)]. Reactions were incubated for 16 h at room temperature, and activity was assessed by TBA assay. Each reaction was oxidized by the addition of 11 μL of sodium periodate (0.2 M in 9 M H$_3$PO$_4$) and incubated for 20 min at room temperature. Oxidation was terminated by the addition of 45 μL of sodium arsenite (10% w/v) in 0.5 M Na$_2$SO$_4$ and 0.05 M H$_2$SO$_4$, and reactions were agitated until all brown discoloration had dissipated. A total of 135 μL of TBA (0.6% in 0.5 M Na$_2$SO$_4$) was added to each reaction and heated to 70 °C for 30 min. The 100-μL samples of each reaction were transferred to a flat-bottomed 96-well plate, and the absorbance was read at 550 nm.

Synthesis of DHA. Erthrose (500 mg, 4.2 mmol) and sodium pyruvate (2.29 g, 21 mmol) were dissolved in sodium phosphate buffer (50 mM, pH 7.4, 10 mL), and Phe190Dpc (0.8 mg) was added. The reaction was incubated at room temperature for a minimum of 48 h before purification by ion exchange chromatography on AG1 × 8 resin (HCO$_3$-; 100-200 mesh). Product was eluted using a 0-0.4 M ammonium bicitarone linear gradient (52). Fractions containing product were identified using the TBA assay on a 20-μL sample and were pooled, freeze-dried, and redissolved in D$_2$O before analysis by 500 MHz $^1$H, COSY, and TOCSY NMR spectroscopy.

Kinetic Assays. Kinetic parameters were determined using the TBA assay. The 100-μL reactions contained erthrose (0.8–15 mM) and pyruvate (80 mM). Reactions were initiated by the addition of 25 μL of enzyme (0.7 mg/mL, 0.5 nmol) and then incubated at room temperature for 1.5 h. A 100-μL sample was taken from each reaction and stopped by the addition of 10 μL trichloroacetic acid (12% w/v). Precipitated proteins were removed by centrifugation and samples were analyzed. Under these conditions the rate of formation of product was linear over 90 min (Fig. 34) and the initial rate of the reaction was determined from the A$_{550 nm}$ against a standard curve of N-acetylenameric acid. Reaction rates were analyzed in duplicate and kinetic parameters were estimated by fitting to the Michaelis–Menten equation using nonlinear regression analysis. Parameter values ± SE of the fit are reported throughout.

Fig. 6. Structures and models of wild-type and modified NALs. Structures of (A) wild-type (PDB ID code 4AH7) (34) and (B) Phe190Dpc (PDB ID code 5LKY) NAL enzymes in complex with pyruvate were structurally aligned. B shows the network of interactions that the Nca side chain Dpc forms with Asp141 and Glu192. Energy-minimized models of product Schiff base structures [illustrated with 4(S)-DHA] were generated using Maestro and are displayed from the same view. (C) Wild-type NAL. (D) Phe190Dpc enzyme complex.
lsine of subunit B of the wild-type or Phe190Dpc NAL. Energy minimization of the products formed was carried out using Macromodel with the normal and the inverted stereoselectivity. J. Org. Chem. 70(12):3663–3670.


