Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 11 (2003) 43–52

BIOORGANIC & MEDICINAL CHEMISTRY

Structure-Based Mutagenesis Approaches Toward Expanding the Substrate Specificity of D-2-Deoxyribose-5-phosphate Aldolase

Grace DeSantis,[a,†] Junjie Liu,[a] David P. Clark,[b] Andreas Heine,[c] Ian A. Wilson[c] and Chi-Huey Wong[a,*]

[a]*Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA*
[b]*Department of Microbiology, University of Illinois, Urbana, IL 61801, USA*
[c]*Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA*

Received 15 August 2002; accepted 15 August 2002

Abstract—2-Deoxyribose-5-phosphate aldolase (DERA, EC 4.1.2.4) catalyzes the reversible aldol reaction between acetaldehyde and D-glyceraldehyde-3-phosphate to generate D-2-deoxyribose-5-phosphate. It is unique among the aldolases as it catalyzes the reversible asymmetric aldol addition reaction of two aldehydes. In order to expand the substrate scope and stereoselectivity of DERA, structure-based substrate design as well as site-specific mutation has been investigated. Using the 1.05 Å crystal structure of DERA in complex with its natural substrate as a guide, five site-directed mutants were designed in order to improve its activity with the unnatural nonphosphorylated substrate, D-2-deoxyribose. Of these, the S238D variant exhibited a 2.5-fold improvement over the wild-type enzyme in the retroaldol reaction of 2-deoxyribose. Interestingly, this S238D mutant enzyme was shown to accept 3-azidopropinaldehyde as a substrate in a sequential asymmetric aldol reaction to form a deoxy-azidoethyl pyranose, which is a precursor to the corresponding lactone and the cholesterol-lowering agent Lipitor™. This azidoaldehyde is not a substrate for the wild-type enzyme. Another structure-based design of new nonphosphorylated substrates was focused on the aldol reaction with inversion in enantioselectivity using the wild type or the S238D variant as the catalyst and 2-methyl-substituted aldehydes as substrates. An example was demonstrated in the asymmetric synthesis of a deoxypyranose as a new effective synthon for the total synthesis of epothilones. In addition, to facilitate the discovery of new enzymatic reactions, the engineered *E. coli* strain *SELECT* (*Δace, adhC, DE3*) was developed to be used in the future for selection of DERA variants with novel nonphosphorylated acceptor specificity.

Introduction

Enzymes are now widely exploited as catalysts in asymmetric organic synthesis, due to their exquisite chemo-, regio- and stereo-specificity.[1–4] The aldolases are a particularly useful class of enzymes because these enzymes catalyze a C–C bond formation with high stereoselective control at the newly formed stereogenic centers.[5] More than 20 aldolase structures have been reported to date and most contain a common $\alpha_8\beta_8$ barrel structural motif.[6]

2-Deoxyribose-5-phosphate aldolase (DERA, EC 4.1.2.4) is unique amongst the aldolases since it catalyzes the reversible condensation of two aldehydes (Fig. 1).[6] Its application in organic synthesis, especially the synthesis of the antitumor agent Epothilone A, has been established.[5] However, the specificity of wild-type DERA is somewhat limited, in particular it has a strong preference for phosphorylated substrates.[4] Expanding the range of unnatural substrates that aldolases will accommodate as well as overcoming their instability and high cost is crucial to further increasing the scope of their synthetic application. One of the primary objectives of this study is to expand the specificity of DERA beyond its natural substrate D-2-deoxyribose-5-phosphate (DRP) and improve its activity with nonphosphorylated substrates.

Numerous methods to alter enzyme properties now exist. These include, for example, solvent or substrate engineering, enzyme adsorption and covalent chemical modifications of enzymes.[7] More recently, site-directed mutagenesis and random mutagenesis approaches to

*Corresponding author. Tel.: +1-619-784-2487; fax: +1-619-784-2409; e-mail: wong@scripps.edu
†Current address: Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA.

0968-0896/03/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0968-0896(02)00429-7

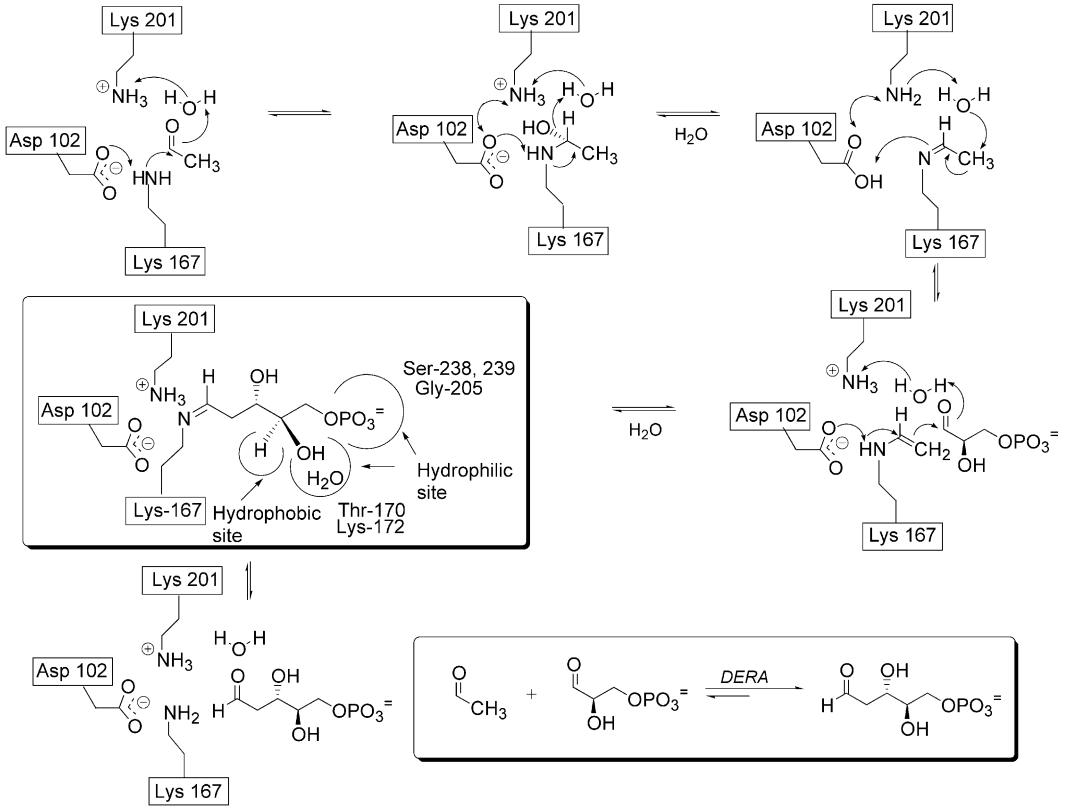

Figure 1. Mechanism of DERA catalyzed aldol reaction between the natural donor acetaldehyde and acceptor D-glyceraldehyde-3-phosphate to generate D-2-deoxyribose-5-phosphate. The reaction proceeds through a Schiff base intermediate. The proton shuffling system composed of Asp102 and Lys201, a key water molecule to mediate proton transfer,[6] another key water involved in enantioselective acceptor binding and the residues interacting with the acceptor substrate side chains are shown.

alter enzyme specificity have been exploited.[8,9] The former often requires a detailed understanding of the enzyme's catalytic mechanism, substrate specificity determinants and tertiary structure. By contrast, random mutagenesis approaches do not require prior understanding of specificity determinants nor knowledge of the structure. Numerous robust methods to generate gene libraries now exist.[8–13] The limitation of this approach is the lack of high-throughput methods to identify the desired phenotype. With 20^X variants possible for an x-amino acid protein, this search becomes an impossible task. General approaches that may be used to identify the desired enzyme activity or property are: in vitro screening for activity,[14] in vitro screening for binding[15] and in vivo selection for activity.[16] The respective shortcoming of each is low throughput in the absence of automation, difficulty of linking binding to catalysis and difficulty in implementation for unnatural activity. Therefore, development of general high throughput methods to screen for the desired enzyme activity is critical for the advancement of organic synthesis using enzymes as catalysts.

In the current investigation, the X-ray structure of DERA and its proposed catalytic mechanism (Fig. 1) are used as a guide to design new nonphosphorylated substrates for the enzymatic reaction with inverted enantioselectivety and to alter the enzyme with mutagenesis to improve the turnover of the retro-aldol reaction of the nonphosphorylated unnatural substrate D-2-deoxyribose (DR).[6] Since the active site of DERA as well as most aldolases is a typical α/β barrel (Fig. 2), which has been shown to be a common scaffold (about 10% of known proteins have this fold) useful for alteration of the catalytic activity of other enzymes by directed evolution,[16–18] it is thought to be a good model for development of novel DERA catalysts with expanded substrate specificity.

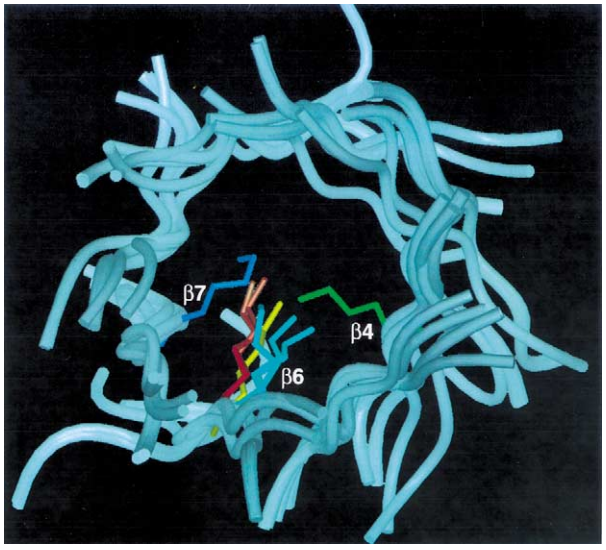

Figure 2. Overlap of eight known aldolase αβ barrels showing the Lys residue for Schiff base formation. DERA Lys167 is depicted in yellow.

Results and Discussion

Rational design of substrates and active site mutants for asymmetric aldol reactions

With the recently determined 1.05 Å three-dimensional structure of *E. coli* DERA in a carbinolamine covalent complex with bound DRP (Fig. 3),[6] five mutants were designed in hopes of improving activity for the unnatural substrate DR. The phosphate binding pocket is comprised of residues Gly171, Lys172, Gly204, Gly205, Val206, Arg207, Gly236, Ser238 and Ser239. However, only the side-chain of Ser238 forms a direct hydrogen bonding contact with the phosphate moiety of DRP.

The utility of a simple approach for changing substrate specificity by altering the electrostatic environment in an enzyme active site to one which is complementary to the electrostatic nature of the unnatural substrate has been demonstrated.[19] Thus, by inspection of the enzyme active site, two basic residues were targeted for mutagenesis to acidic residues. The K172E and R207E mutants were therefore prepared. In addition, three neutral side chains in the phosphate binding pocket were replaced with acidic ones, generating G205E, S238D and S239E mutants. The goal of these designed mutations was to change the substrate specificity of WT-DERA from a preference for the negatively charged DRP to the nonphosphorylated, neutral DR substrate.

It was anticipated that an expanded substrate specificity of DERA in the retro-aldol direction would parallel an expanded substrate specificity in the aldol direction. Accordingly, these mutants are characterized in the retro-aldol direction.[6]

For each of the five mutants, the activity with the natural substrate, DRP (Table 1) is substantially decreased as expected due to electrostatic repulsion between the introduced negatively charged residue and the negatively charged phosphate moiety of DRP. In all cases, especially for R207E, the specificity for the unnatural substrate is improved as shown by the increase in the ratio of specificity constants for DR compared to DRP $\{k_{cat}/K_M(DR)\}/\{k_{cat}/K_M\ (DRP)\}$ of the mutants versus WT. Clearly this residue is critical to DRP transition state binding as evidenced by the data and is in agreement with the conserved nature of this residue for the nine closest homologues of *E. coli* DERA.[6] However, for the shorter DR substrate, residue 207 may not be in sufficient proximity to effect a substantial change since, for this mutant, DR specificity is virtually unchanged compared to WT. Two of the designed DERA mutants exhibited higher than WT activity with DR as the substrate. Of these, the S238D variant is the most active, with a 2.5-fold improvement in k_{cat}/K_M compared to WT-DERA. S239E exhibits a 1.3-fold improvement in k_{cat}/K_M compared to WT. For both S239E and S238D, k_{cat}/K_M for the natural phosphorylated substrate is

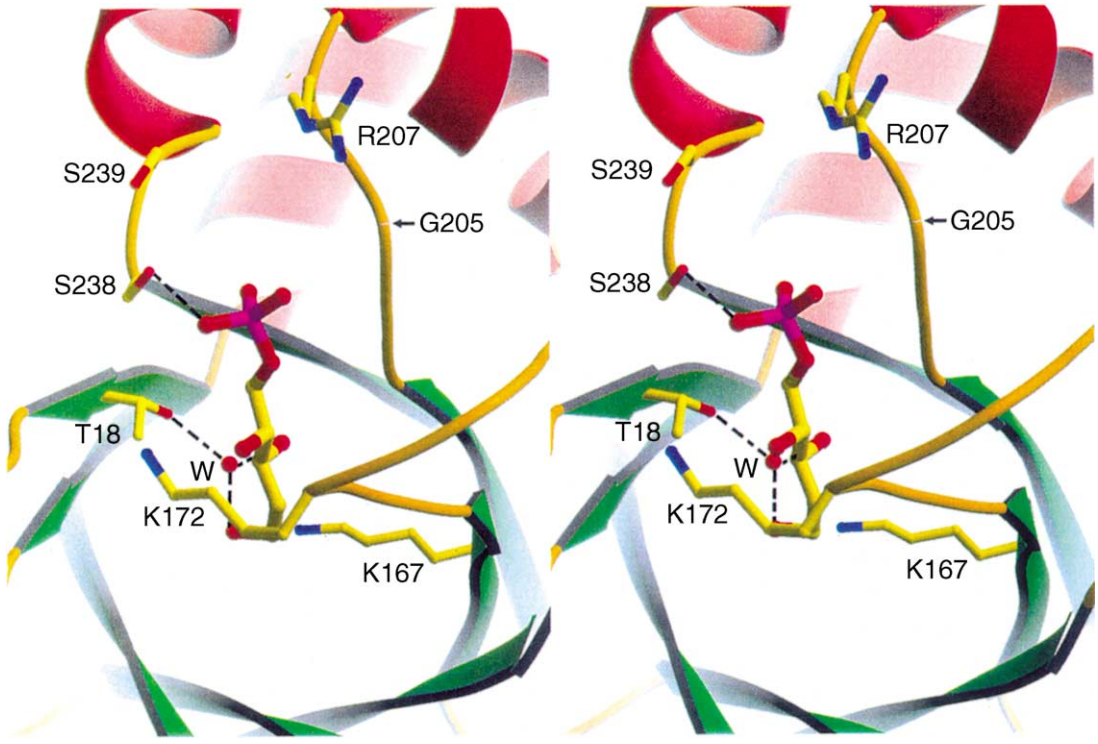

Figure 3. Stereoview of the active site of the DERA carbinolamine complex. The central beta barrel is depicted by the β-strand colored in green. Upon reaction of the substrate D-2-deoxyribose-5-phosphate with the reactive lysine, Lys167, the carbinolamine is formed. Thr18 is hydrogen bonded via a water molecule to the substrate. Also Ser238 forms a hydrogen bond with the phosphate group of the substrate. Residues shown in the figure, which are mutated in the present study are K172E, G205E, R207E, S238D and S239E. A water molecule is shown to have H-bonding interaction with K172, T18 and the C4–OH group of the substrate.

substantially decreased as would be expected due to electrostatic repulsion. Interestingly, in the WT structure only the side chain of S238 is in direct contact with the substrate and it seems that its proximity permits a degree of modulation of substrate specificity even for the smaller DR substrate. The G205E mutation yields a protein that is virtually inactive both with respect to DR and DRP. This residue is strictly conserved in the nine homologues of DERA and its mutation may effect a structural perturbation. The K172E mutation results in a 5-fold decrease in k_{cat}/K_M with the DR substrate.

In order to establish whether the improvement in the DERA catalyzed retro-aldol reaction is synthetically useful, we evaluate the efficiency of the DERA mutants compared to WT to catalyze the aldol reaction between acetaldehyde and (±)-glyceraldehyde. In the aldol direction, the relative activity of the DERA variants as evaluated both by a spectraphotometric coupled-assay of substrate consumption and by thin layer chromatographic analysis of product formation is: S238D ~ S239W > WT > R207E > K172E > G205E. The aldol reaction activity thus parallels the kinetic retro-aldol activity data and validates this approach. Therefore, two improved variants of DERA which catalyze both the aldol and retro-aldol reaction of a nonphosphorylated substrate have been developed.

Molecular modeling (Fig. 4) shows that the terminal hydroxyl group of the product is able to form a 2.9–3.2 Å hydrogen bond to Asp238-CO_2. This may explain the increased activity of the S238D mutant toward the nonphosphoylated substrate. Furthermore, optimization of the Asp side chain conformation (rotamer) results in the gain of a hydrogen bond (2.5 Å) to a water molecule in the active site. This water molecule forms a second hydrogen bond of 2.9 Å to the Nζ of Lys172. The first product complex (Fig. 4A) shows formation of a hydrogen bond (2.7 Å) between the hydroxyl group at the (*R*)-configured or D-configured C4 position with this water molecule. The hydrogen bond is absent in the second complex (Fig. 4B) with the (*S*)-configured C4 position and may explain the observed preference for the product formation shown in Fig. 4A.

The role of two water molecules in catalysis and enantioselection

In addition to D-glyceraldehyde, DERA and the S238D variant accept other 2-substituted 3-hydroxypropinaldehydes and inversion of enantioselectivity has been observed when 2-methyl- or 2-methoxy-3-hydroxypropinaldehyde is used as the substrate (Fig. 5).[5] In both cases, the L-enantiomer is the preferred substrate for the wild type, but facial selectivity remains unchanged (Fig. 5A).[5] This methyl-derived product has been used in the total synthesis of epothilones.[5] In addition, the S238D mutant accepts the L-2-methyl derivative as a better substrate with a 5-fold improvement in k_{cat}/K_M compared to that of the wild type. These results are consistent with structure-based molecular modeling. As described above and in Figure 4, the water molecule interacting with the 2-hydroxy group of D-glyceraldehyde (corresponding to the 4-hydroxy group of the product) plays a key role in determining the enantioselectivity of DERA catalysis. The corresponding D-2-methyl derivative is not a substrate as the methyl group would be in close contact with the water molecule and the carbonyl oxygen of Thr170. On the other hand, binding of the 2-methyl group of the L-enantiomer to the enzyme is energetically more favorable (Fig. 4C), with the methyl group pointing to a more hydrophobic environment in van der Waals contact with Cα of Gly171 (3.5 Å), Cβ of Ala203 (3.9 Å) and Cα of Gly204 (3.6 Å). Both mechanistic and modeling studies thus reveal the important roles of the two water molecules in DERA catalysis: one is acting as acid and base in catalysis and the other is involved in the enantioselective binding of the acceptor substrate, as shown in Figure 1.

A new sequential aldol reaction catalyzed by S238D variant

While the S238D variant is in general better than the wild-type DERA to accept nonphosphorylated sub-

Table 1. Effect of rationally designed DERA phosphate binding pocket mutations

DERA mutant	Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} mM^{-1})	$\frac{[k_{cat}/K_M(DR)]}{[k_{cat}/K_M(DRP)]}$	$\frac{[k_{cat}/K_M(mutant)]}{[k_{cat}/K_M(WT)]}$
WT	DRP	68±1	0.64±0.01	106±2	1.9×10^{-5}	1
	DR	0.107±0.005	57±7	0.0020±0.0003		1
K172E	DRP	—[a]	—	0.0013	0.27	1.2×10^{-5}
	DR	0.022±0.02	63±22	0.0003±0.0001		0.18
G205E	DRP	—	—	1.3×10^{-6}	3.1×10^{-3}	1.2×10^{-8}
	DR	—	—	4.0×10^{-9}		2.0×10^{-6}
R207E	DRP	—	—	0.0009±0.0001	2.2	8.0×10^{-6}
	DR	0.064±0.001	33±3	0.0019±0.0002		0.95
S238D	DRP	0.58±0.05	61±11	0.01±0.001	0.13	9.0×10^{-5}
	DR	0.21±0.01	39±6	0.005±0.0009		2.5
S239E	DRP	41±1	4.3±0.3	9.5±0.7	2.7×10^{-4}	0.09
	DR	0.175±0.007	67±8	0.0026±0.0004		1.3

[a]No data.

strates as acceptors, it also catalyzes a novel sequential aldol reaction using 3-azidopropinaldehyde as the first acceptor and two molecules of acetaldehyde as donor to form an azidoethyl pyranose, a key intermediate useful for the synthesis of the cholesterol lowering agent Lipitor™ (Fig. 5B). The azidoaldehyde is, however, not a substrate for the wild-type enzyme.

Design and implementation of an in vivo selection system

While the 2.5-fold improvement in activity reported here is encouraging, considering that most mutations lead to decreases in activity, further enhancements are desirable. Though increasing the substrate scope of aldolases has previously been established by random

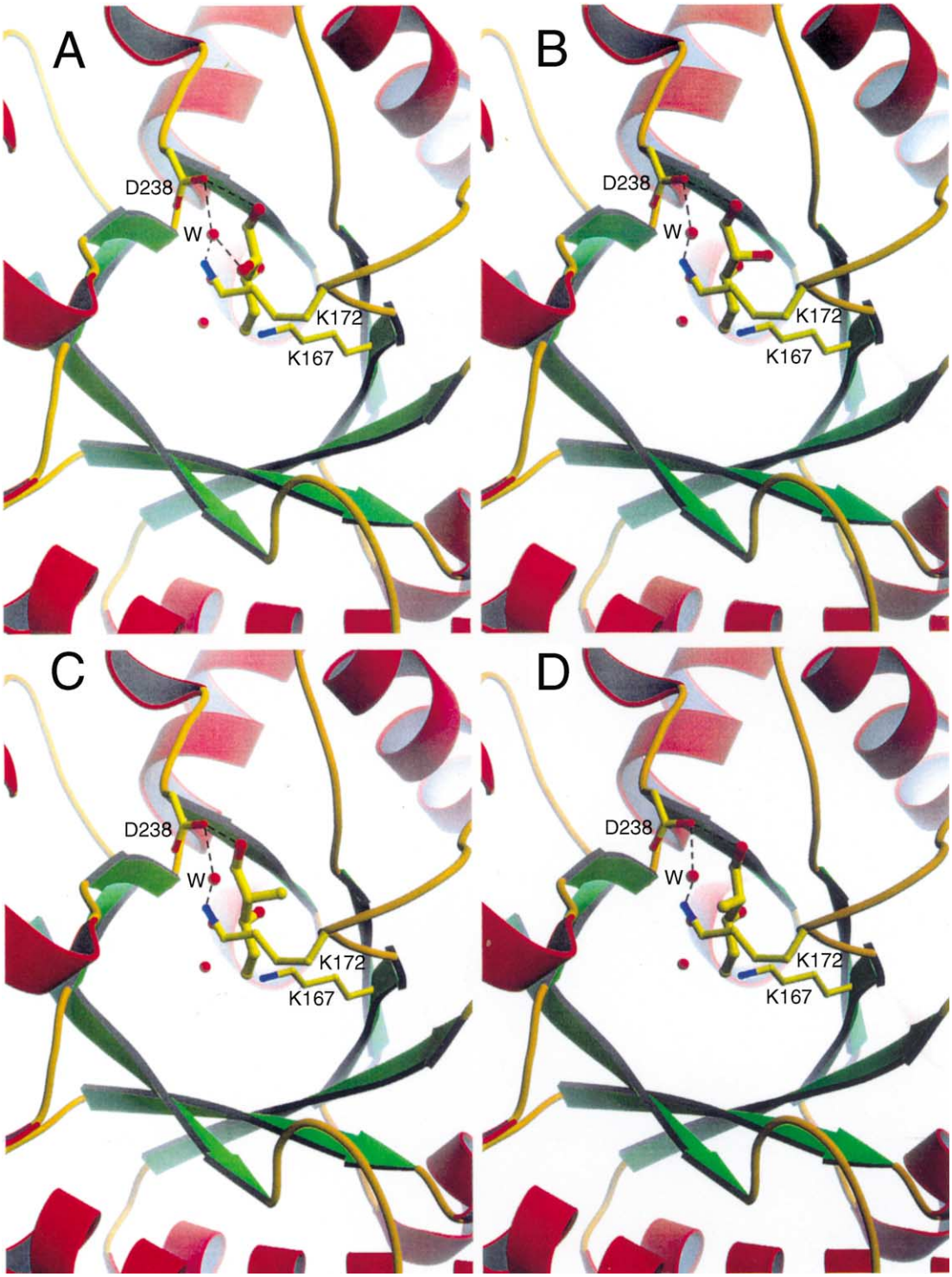

Figure 4. DERA product modeling based on the Schiff base complex structure (PDB code 1JCJ).[6] The enzyme model contains the S238D mutation. Four product molecules were modeled into the active site and the energy of the complex minimized. The enzyme is able to catalyze the aldol reaction with D-glyceraldehyde and D-2-methyl-3-hydroxy propanol resulting in the products displayed in A and C, respectively (Figs 1 and 5). The structures shown in B and D resulting from the corresponding enantiomeric acceptors are not substrates for DERA.

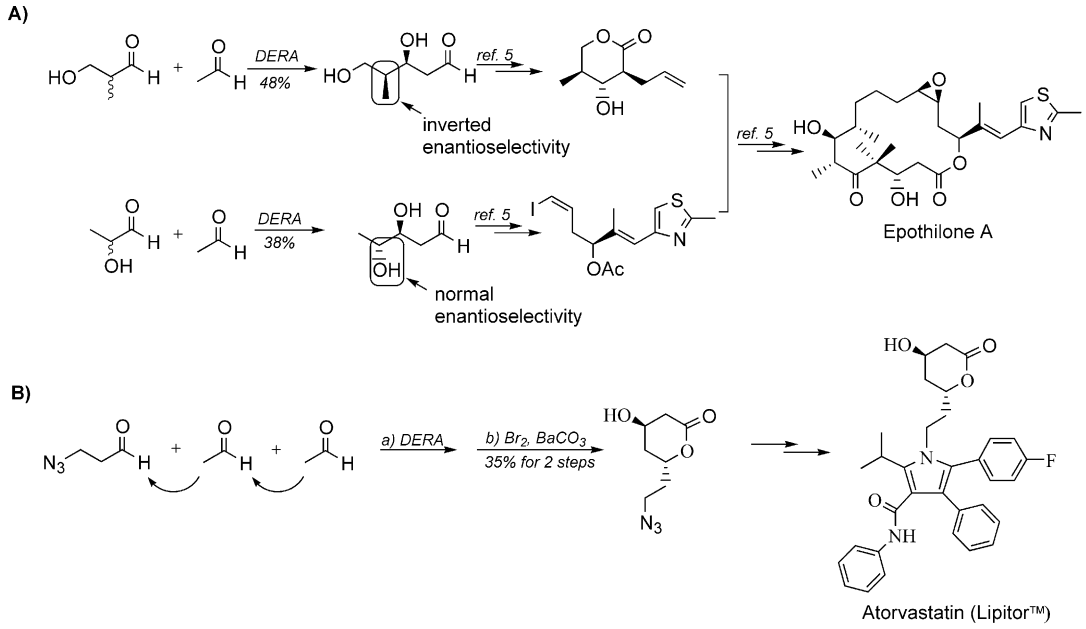

Figure 5. DERA catalyzed synthesis of designed substrates.

mutagenesis,[18] throughput limitations have allowed only a small percentage of the gene to be characterized. Thus, in order to rapidly evaluate the activity of a significant population of variants, a higher throughput activity-based screening methodology is essential. In preparation for a directed evolution program to identify DERA variants with expanded substrate scope, an in vivo selection system suitable for high-throughput analysis was therefore developed.

Having established the validity of screening for improved retro-aldol activity as indicative of the synthetic potential of the DERA, the retro-aldol direction was chosen for the development of a selection system. A cell that utilizes acetaldehyde as its sole carbon source or is dependent on acetaldehyde for growth was desired to aid selection of DERA variants with improved activity for DR or alternative unnatural substrates. *SELECT* (*Δace, adhC, DE3*), an *E. coli* strain that requires acetaldehyde for growth was engineered. Two features of *SELECT* are key. Firstly, the absence of a viable pyruvate dehydrogenase (*aceF*) affects an acetate auxotroph when grown in glucose as the sole carbon source. Secondly, the constitutive overproduction of an aerotolerant version of *adhE,* which has both alcohol dehydrogenase and acetaldehyde dehydrogenase activities, affects conversion of acetaldehyde to acetyl-CoA thus overcoming the acetate auxotroph (Fig. 6).

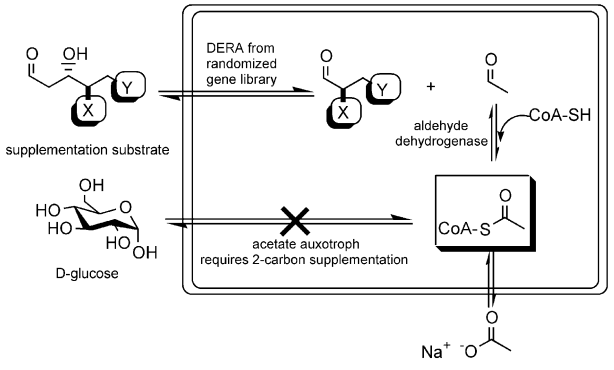

Figure 6. The metabolically engineered *SELECT* (*Δace, adhC, DE3*) *E. coli* strain requires 2-carbon supplementation for cell growth. In the absence of acetate supplementation, only those cells transformed with a DERA gene variant can catalyze the retro-aldol reaction of the supplementation substrate and survive. Thus, any enzyme selectivity that releases acetaldehyde or acetate may be selected.

E. coli SELECT grows well in medium supplemented with either acetate or an acetaldehyde source and exhibits the desired phenotype (Fig. 7). *SELECT* was transformed with the DERA expressing plasmid,

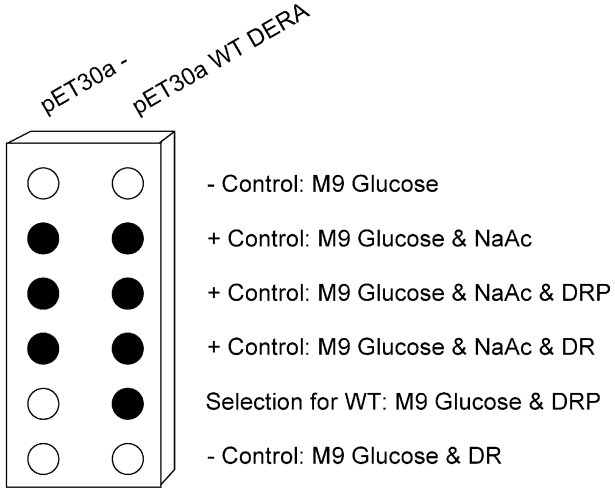

Figure 7. Proof of concept for the selection protocol using *SELECT*. *SELECT* cells require a 2-carbon source for growth. Neither DR nor DRP is toxic to the cells. The cells transformed with the empty vector (−) do not grow in the presence of DRP but the WT DERA expressing cells do. Neither transformant grows when supplemented only with DR for which DERA has a 53,000-fold poorer specificity constant compared to DPR. Both tranformants grow when the medium is supplemented with sodium acetate (NaAc).

pET30a WT DERA, and growth conditions were then optimized for the expression of soluble active DERA enzyme. Selection conditions were further optimized using DERA's natural substrate DRP as a supplementation substrate and Figure 6 illustrates that viable selection conditions are achieved. In the absence of 2-carbon supplementation, neither *SELECT* cells transformed with a plasmid which expresses WT DERA nor those transformed with a nonexpressing plasmid (−) grow. By contrast, both grow in the presence of sodium acetate supplementation. That both also grow in the presence of either sodium acetate together with DRP, or sodium acetate together with DR, demonstrates that neither of these supplementation substrates nor their metabolic products are toxic to the cells. Proof of principle for this selection system arises from the fact that only *E. coli SELECT* cells transformed with plasmid that expresses WT DERA grow when DRP is used as the supplementation substrate. Furthermore, that the endogenous genomic *E. coli* DERA is not expressed at a sufficiently high level to affect the use of DRP as a 2-carbon source by virtue of its metabolism to acetaldehyde and D-glyceraldehyde-3-phosphate was established. Since WT DERA cannot accept the unnatural substrate DR efficiently, neither of the SELECT cells transformed with nonexpressing plasmid (pET30a−) nor those transformed with DERA expressing plasmid (pET30a WT DERA) grow when supplemented with DR. A novel activity-based selection system is thus established and can be used to select for a DERA variant which can catalyze the retro-aldol reaction of DR and other nonphosphorylated substrates. Work is in progress to identify novel DERA variants for this purpose.

Several examples demonstrating the power of in vivo selection based methods for identifying mutant enzymes which reverse the phenotype of a bacterial strain deficient in an enzyme with the desired activity have been reported.[12,16,20–24] However, in most examples, such systems have been utilized to identify mutations which transform the activity of a natural enzyme into another natural enzyme to overcome auxotroph.[16,21] In addition, several examples for which selection has been used to identify variants with native activity for an inactivated enzyme have also been demonstrated.[20] To date, the reported examples of in vivo selection that have identified unnatural enzyme specificity or activity involve gene products which confer antibiotic resistance.[12] However, more recently, an innovative growth selection based assay method for the identification of an error-prone T7 polymerase,[23] and identification of a four-base codon *t*RNA[24] were developed using an antibiotic resistance selection. Each of these elegant examples demonstrates the potential power a selection or complementation approach can have in identifying variants with improved or altered activity. Thus, the in vivo activity based selection system which utilizes the engineered *E. coli* strain *SELECT* to identify DERA variants with expanded substrate scope described here is one of the first examples of a selection method able to identify an enzyme with unnatural and synthetically useful substrate specificity in an ultra-high throughput manner.

Conclusion

Using the high-resolution X-ray structure of DERA and its catalytic mechanism, we have demonstrated that both the acceptor substrate and the enzyme can be changed to alter the efficiency and specificity of the enzymatic aldol reaction, including inversion of enantioselectivity using nonphosphorylated substrates and wild-type or S238D variants and new substrate specificity using the S238D variant. The S238D variant showed a 2.5-fold improvement in DERA activity with the unnatural substrate DR. It accepts 3-azidopropinaldehyde as a new substrate in a sequential aldol reaction to form a novel azidopyranose, while the wild-type enzyme is inactive toward this azidoaldehyde. To further improve the efficiency for identification of DERA variants to catalyze novel aldol reactions with nonphosphorylated substrates, we have developed a selection system which will be used to expand the acceptor specificity and stereoselectivity of this type of aldol reaction.

Experimental

Nucleic acid manipulations were done according to standard procedures.[25] TAQ DNA polymerase was from Stratagene. The Quiagen QIAprep Spin Miniprep Kit was utilized for plasmid preparation. PCR products were purified by electrophoresis on a 1% agarose gel and then extracted using the QIAEXII Agarose Gel Extraction Kit. Restriction endonucleases and T4 ligase were from New England Biolabs. Electrocompetent *E. coli* BL21 (DE3) cells, pET30 LIC and pET30a plasmids, and His-bind metal chelation resin were from Novagen. Oligonucleotide primers were prepared by Operon Technologies (San Diego, CA). DNA sequencing was performed at the Protein and Nucleic Acid Core Facility at The Scripps Research Institute on a ABI50 automated sequencer. UV kinetic assays were performed on a Cary 3 Bio UV–Vis spectrophotometer. Curve fitting was done by the non-linear least squares method using KaleidaGraph (Abelbeck Software). All reagents were purchased at highest commercial quality and used without further purification unless otherwise stated. Silica gel 60 (230–240 mesh) from Merck was used in chromatography. High resolution mass spectra (HRMS) were recorded on IONSPEC-FTMS spectrometer (MALDI) with DHB as matrix. ^{1}H NMR and ^{13}C NMR spectra were performed on a Bruker AMX-500 instrument. IR spectra were recorded on a Perkin–Elmer 1600 series FT-IR spectrometer. Optical rotations were recorded on a Perkin-Elmer 241 polarimeter.

Cloning of WT DERA

The *E. coli* D-2-Deoxyribose-5-phosphate aldolase (DERA, EC 4.1.2.4) gene was PCR amplified from plasmid pVH17 (ATCC86963),[26] using the forward primer 5′-ACCGATGACGACGACAAGGCCATGGCTATGACTGATCTGAAAG and the reverse primer 5′ - TGGTTGAGGAGAAGCCAAGCTTAGTAGCTGCTGGCGCT and subcloned into the pET30 LIC vector (Novagen).[6] *E. coli* D-2-Deoxyribose-5-phosphate

aldolase (DERA, EC 4.1.2.4) gene was PCR amplified from the above construct, WT DERA pET30 LIC using the forward primer 5′-ACCGATGACGACGACAAGGCCATGGCTATGACTGATCTGAAAG and the reverse primer 5′-TGGTTGAGGAGAAGCCAAGCTTAGCTGCTGGCGCT and then subcloned into the pET30a vector (Novagen) using the NcoI and HindIII restriction sites.

Site-directed mutagenesis

The following cloning primers were used: 5′-GACGACGACAAGATGCATATG, (forward) 5′-GAGGAGAAGCCCGGTTTAGTA (reverse). A 810-bp fragment was obtained by PCR using 20 μM of each of the dNTPs, 10 pM oligonucleotide primers, 10 ng template and 5 U *Taq* polymerase (Stratagene) in 100 μL DNA polymerase buffer. Mutagenesis primers used for double-sided overlap extension PCR were: G207E, 5′-GCGGGCGGCGTG*GAA*ACTGCGGAAGAT (forward), 5′-ACTTTCCGCAGT*TTC*CACGCCGCCCGC (reverse). S239E, 5′-TTTGGCGCTTCC*GAA*CTGCTCGCAAGC (forward), 5′-GCTTGCCAGCAG*TTC*GGAAGCGCCAAA (reverse). S238D, 5′-CGCTTTGGCGCT*GAC*AGCCTGCTGGCA. K172E, TCTACCGGT*GAA*GTGGCTGTG (forward), CACAGCCAC*TTC*ACCGGTAGA (reverse). G205E, 5′-AAACCGGCGGGCGAAGTGCGTACTGCG (forward), 5′-CGC AGTACGCACTTCGCCCGCCGGTTT (reverse). The mixture was thermocycled for 1 cycle at of 94 °C for 5 min, then 30 cycles of {94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s} and then one cycle of 72 °C for 10 min.

Protein expression and purification of DERA mutants

Plasmids were transformed into electrocompetent BL21 (DE3) and subjected to 1 h outgrowth at 37 °C in 1 mL SOC medium. These transformants (10–200 μL) were plated on LB_{kan} plates and incubated at 37 °C overnight. A starter culture was prepared by picking individual colony to inoculate a 100 mL Luria–Bertani (LB) starter culture containing 10 μg/mL kanamycin (kan) grown at 37 °C, 220 rpm overnight. The starter culture was used to inoculate 1L LB_{kan}. Protein expression was induced at $OD_{600} = 0.6$- 0.8 by the addition of isopropyl-β-D-thiogalactoside (IPTG) to a final concentration of 0.5 mM. Cells were harvested 6 h after induction, by centrifugation at 4 °C, 8000 rpm for 10 min and were stored at −78 °C. The cell pellet was resuspended in 25 mL of 100 mM phosphate, 200 mM sodium chloride pH 7.5 chilled on ice. The cells were lysed by passing through a French press (SLM Instruments, Urbana, IL) compressed to 1500 psi and then released to ambient pressure, three times. Cell debris was pelleted by centrifugation at 4 °C, 14,000 rpm for 1 h. The supernatant was filtered through a 0.2 μm cellular acetate membrane filter (Corning), and was loaded onto a Ni^{2+}-NTA-agarose column with a bed volume of 2.5 mL pre-equilibrated with 100 mM phosphate, 200 mM sodium chloride, 5 mM imidazole, 5 mM β-mercaptoethanol pH 7.5 buffer. The column was washed with 40 mL of 100 mM phosphate, 200 mM sodium chloride, 20 mM imidazole, 5 mM β-mercaptoethanol, pH 7.5 buffer. Bound enzyme was then eluted with 20 mL of 100 mM phosphate, 200 mM sodium chloride, 20 mM imidazole, 5 mM β-mercaptoethanol pH 7.5 buffer, and was dialyzed against 50 mM triethanolamine hydrochloride pH 7.5 buffer at 4 °C. Eluted enzymes were analyzed by SDS-PAGE and were found to be >95% pure in all cases. Enzyme solutions were aliquoted and frozen in liquid nitrogen and stored at −78 °C prior to use. Enzyme concentrations were determined by the Bradford procedure (Bio-Rad) using bovine serum albumin as a calibration standard.

DERA cleavage (retroaldol) assay

Enzyme activity was monitored by the standard coupled assay using α-Glycerophosphate Dehydrogenase, (α-GPD, EC 1.1.18) and Triosephosphate Isomerase (TPI, EC 5.3.1.1). Enzyme activity was assayed in the retro-aldol, decomposition direction with 0.01–4 mM D-2-deoxyribose-5-phosphate (DRP) or 5 to 200 mM D-2-deoxyribose in 50 mM triethanolamine hydrochloride pH 7.5 buffer using a GPD/TPI (1.6 U/mL Sigma G-1881) coupled enzyme system at 25 °C in the presence of 0.3 mM NADH by observing the rate of decrease of NADH concentration as monitored at 340 nm, $\varepsilon = 6220\ M^{-1}\ cm^{-1}$.

DERA addition (aldol) assay

DERA enzyme activity was assayed in the aldol synthesis direction by determining the concentration of acetaldehyde remaining by a coupled endpoint assay with yeast alcohol dehydrogenase (YADH, EC 1.1.1.1). 200 mM acetaldehyde, which had been freshly distilled under anerobic conditions, 200 mM (±)-glyceraldehyde and 0.2 mg/mL DERA in 50 mM triethanolamine, pH 7.5 buffer which had been deoxygenated with N_2, were incubated under an N_2 atmosphere at 22 °C. At various time points, 50 μL aliquots were withdrawn and quenched into 15 μL of 60% perchloric acid. After a 5 min, incubation on ice, 890 μL 1 M triethanolamine, pH 7.5 buffer and 45 μL 4 N NaOH were added to neutralize the solution. 20 μL this solution was then assayed for remaining acetaldehyde. The amount of acetaldehyde remaining was equated to moles NADH consumed, as determined in triethanolamine pH 7.5 buffer containing 0.3 mM NADH, 20 μL the above quenched reaction aliquot and 0.05 mg/mL YADH. DR product formation was also confirmed by silica gel TLC with ethylacetate running solvent and *p*-anisaldehyde developing stain. R_f: glyceraldehyde = 0.04 (stains brown) R_f: 2-deoxyribose = 0.1 (stains blue).

Construction of *E. coli SELECT* strain

First, DC81 was transduced with P1 grown on JC1552 (aceF+ leu-) and transductants able to grow without acetate were selected in the presence of leucine. DC119 was one such aceF+ transductant, which also received the leu mutation from JC1552 and hence required leucine. Next, DC119 was transduced with P1 grown on DC34 (ΔaceEF leu+) and transductants able to grow without leucine were selected on minimal medium E

containing succinate (0.4%) plus acetate (0.2%) as carbon source. Transductants were screened for those unable to grow on succinate alone, that is, those receiving succinate (0.4%) plus acetate (0.2%) as the carbon source. Transductants were screened for those unable to grow on succinate alone, that is, those receiving the Δ(aroP-aceEF)15 deletion and therefore requiring exogenous acetate. DC489 was one such transductant. *E. coli* strain *SELECT* was then prepared by generating the λDE3 lysogen of DC489 using the Novagen λDE3 lysogenization kit (69734-3) according to manufacturer's directions. *E. coli* strains DC81, DC34, and JC1552 were used for construction of *SELECT*.[27]

Development of liquid selection conditions

Plasmids were transformed into electrocompetent *SELECT* cells and subjected to 1 h outgrowth at 37 °C in 1 mL SOC medium supplemented with 0.1% sodium acetate. The cells were then collected by centrifugation at 4 °C, 3000 rpm for 10 min. The supernatant was discarded and the pellet gently re-suspended M9 0.2% glucose. This was repeated twice. The cells were then diluted to $OD_{600} = 0.001$ in M9 0.2% glucose, 0.01 mM IPTG, 10 μg/mL kanamycin. The appropriation supplementation substrate (sodium acetate, D-2-deoxyribose-5-phosphate or D-2-deoxyribose) was then added at 0.1% w/w concentration. After an appropriate selection time at 37 °C, typically 24–72 h, the cells were harvested by centrifugation and their amplified plasmids isolated.

Molecular modeling

The DERA enzyme S238D mutation was generated using the program O[1] and the side chain placed in a common rotamer position. The product molecules displayed in Figure 5A–D were generated using the Builder module of InsightII (2000) (Accelrys Inc.) and energy minimized. They were manually placed in the enzyme active site based on the existing Schiff base crystal structure (PDB code 1JCJ).[6] Hydrogen atoms on the protein residues and on relevant water oxygen atoms were added using the Biopolymer module. For energy minimization, the CVFF force field was used. All minimizations were carried out with the Discover module using a distant dependent dielectric constant. In the first round of minimization all non-hydrogen atoms were constrained to fixed positions, and steepest descent and conjugate gradient energy minimizations were performed for 100 iterations each. Thereafter, constraints for the product molecule and Asp238 were released and the minimization procedure was repeated.[19]

Sequential asymmetric aldol reaction

To a mixture containing 3-azidopropinaldehyde (600 mg, 6.0 mmol) was added a buffer solution (36 mL, pH = 7.5), which contained mutant S238D DERA (about 200 U based on the assay using DRP as substrate). The resulting solution was stirred in the dark for 6 days under argon. The reaction was quenched with 2 volumes of acetone. The mixture was then stirred at 0 °C for 1 h and centrifuged to remove the precipitated enzyme. The aqueous phase was concentrated *in* vacuo, and the residue was passed through a short silica column eluted with EtOAc. The elutant was concentrated and afforded the crude product (560 mg, 3.0 mmol).

To a mixture of the lactol above (560 mg, 3.0 mmol) and $BaCO_3$ (0.8 g, 4.0 mmol) in H_2O (20 mL) at 0 °C was added slowly freshly opened Br_2 (180 μL, 3.4 mmol). The mixture was stirred in the dark overnight. After filtration, water was removed in vacuo. Purification of the residue by flash chromatography (silica, 1:1 hexane/EtOAc) afforded the product (391 mg, 35% for 2 steps). $[\alpha]_D = 72.0°$ (c = 1.0, $CHCl_3$); IR (film): 3421.1, 2928.0, 2102.8, 1718.2, 1254.8, 1072.2 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 4.85 (m, 1H), 4.40 (m, 1H), 3.54 (dd, $J = 5.8$, 7.3 Hz, 2H), 2.76 (br. s, 1H), 2.67 (m, 2H), 2.00 (br. d, $J = 14.3$ Hz, 1H), 1.95 (m, 1H), 1.87 (m, 1H), 1.77 (m, 1H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.30, 72.86, 62.37, 47.06, 38.45, 35.72, 34.73; HRMS *m/e* calcd for (M^+) $C_7H_{11}N_3O_3$: 185.0800; found: 208.0693 (M + Na).

Acknowledgements

This work was funded by the NIH GM44154 grant to CHW. A Natural Sciences and Engineering Research Council of Canada postdoctoral fellowship to GD is gratefully acknowledged. The authors would thank Drs. Sun Fong, Pam Sears, John Lutz and Ms. Flora Huang for helpful discussions and Dr. M. A. Elslinger for computational assistance and help with the docking calculations.

References and Notes

1. Wong, C.-H.; Whitesides, G. M. *Enzymes in Synthetic Organic Chemistry*; Pergamon: Oxford, 1994.
2. Faber, K.; Patel, R. *Curr. Opin. Biotechnol.* **2000**, *11* (6), 517.
3. BioCatalysis, Accelrys Inc. http://www.accelrys.com/chem_db/biocat.html.
4. Machajewski, T. D.; Wong, C.-H. *Angew. Chem. Int. Ed.* **2000**, *39*, 1352.
5. Liu, J.; Wong, C.-H. *Angew. Chem. Int. Ed.* **2002**, *41*, 1404.
6. Heine, A.; DeSantis, G; Lutz, J. G.; Mitchell, M.; Wong, C.-H.; Wilson, I. A. *Science* **2001**, *294*, 369.
7. DeSantis, G.; Jones, J. B. *Curr. Opin. Biotechnol* **1999**, *10* (4), 324.
8. Chen, R. *Biotechnol* **2001**, *19*, 13.
9. Zaccolo, M.; Gherardi, E. *J. Mol. Biol.* **1999**, *285*, 775.
10. Cadwell, R. C.; Joyce, G. F. *PCR Methods Appl.* **1992**, *2*, 28.
11. Christians, F. C.; Loeb, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6124.
12. Stemmer, W. P. C. *Nature* **1994**, *370*, 389.
13. Grey, K. A.; Richardson, T. H.; Kretz, K.; Short, J. M.; Bartnek, F.; Knowles, R.; Kan, L.; Swanson, P. E.; Robertson, D. E. *Adv. Synth. Catal.* **2001**, *343*, 618.
14. Moore, J. C.; Arnold, F. H. *Nat. Biotechnol* **1996**, *14*, 458.
15. Forrer, P.; Jung, S.; Pluckthun, A. *Curr. Opin. Struct. Biol.* **1999**, *9*, 514.

16. Altamirano, M. M.; Blackburn, J. M.; Aguayo, C.; Fersht, A. R. *Nature* **2000**, *403*, 617.
17. Wymer, N.; Buchanan, L. V.; Herderson, D.; Mehta, N.; Botting, C. H.; Pocivavsek, L.; Fierke, C. A.; Toone, E. J.; Naismith, J. H. *Structure* **2000**, *9*, 1.
18. Fong, S.; Machajewski, T. D.; Mak, C. C.; Wong, C.-H. *Chem. Biol.* **2000**, *7*, 873.
19. DeSantis, G.; Shang, X.; Jones, J. B. *Biochemistry* **1999**, *38*, 13391.
20. MacBeath, G.; Kast, P.; Hilvert, D. *Science* **1998**, *279*, 1958.
21. Yano, T.; Oue, S.; Kagamiyama, H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5511.
22. Black, M. E.; Newcomb, T. G.; Wilson, H.-M. P.; Loeb, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3525.
23. Brakmann, S.; Grzeszik, S. *Chembiochem. Eur. J. Chem. Biol.* **2001**, *2*, 212.
24. Magliery, T. J.; Anderson, J. C.; Schultz, P. G. *J. Mol. Biol.* **2001**, *307*, 755.
25. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Laboratory Press: Plainview, NY, 1989.
26. Valentin-Hansen, P.; Boetius, F.; Hammer-Jespersen, K.; Svendsen, I. *Eur. J. Biochem.* **1982**, *125*, 561.
27. Clark, C. P.; Cronan, J. E., Jr. *J. Bacteriol.* **1980**, *141*, 177.