Creation of a Pair of Stereochemically Complementary Biocatalysts

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · JANUARY 2007

Impact Factor: 12.11 · DOI: 10.1021/ja065233q · Source: PubMed

5 AUTHORS, INCLUDING:

Gavin J Williams
North Carolina State University
29 PUBLICATIONS **730** CITATIONS

SEE PROFILE

J|A|C|S
ARTICLES
Published on Web 11/21/2006

Creation of a Pair of Stereochemically Complementary Biocatalysts

Gavin J. Williams, Thomas Woodhall, Lorna M. Farnsworth, Adam Nelson,* and Alan Berry*

Contribution from the Astbury Centre for Structural Molecular Biology and School of Chemistry, University of Leeds, Leeds, LS2 9JT, United Kingdom

Received July 21, 2006; E-mail: a.s.nelson@leeds.ac.uk; a.berry@leeds.ac.uk

Abstract: *N*-Acetylneuraminic acid lyase (NAL) exhibits poor facial selectivity during carbon−carbon formation, and as such, its utility as a catalyst for use in synthetic chemistry is limited. For example, the NAL-catalyzed condensation between pyruvate and (2*R*,3*S*)-2,3-dihydroxy-4-oxo-*N,N*-dipropylbutyramide yields ca. 3:1 mixtures of diastereomeric products under either kinetic or thermodynamic control. Engineering the stereochemical course of NAL-catalyzed reactions could remove this limitation. We used directed evolution to create a pair of stereochemically complementary variant NALs for the synthesis of sialic acid mimetics. The E192N variant, a highly efficient catalyst for aldol reactions of (2*R*,3*S*)-2,3-dihydroxy-4-oxo-*N,N*-dialkylbutyramides, was chosen as a starting point. Initially, error-prone PCR identified residues in the active site of NAL that contributed to the stereochemical control of an aldolase-catalyzed reaction. Subsequently, an intense structure-guided program of saturation and site-directed mutagenesis was used to identify a complementary pair of variants, E192N/T167G and E192N/T167V/S208V, which were ∼50-fold selective toward the cleavage of the alternative 4*S*- and 4*R*-configured condensation products, respectively. It was shown that wild-type NAL could not be used for the highly stereoselective synthesis of a 6-dipropylamide sialic acid mimetic because the 4*S*-configured product was only ∼3-fold kinetically favored and only ∼3-fold thermodynamically favored over the alternative 4*R*-configured product. However, the complementary 4*R*- and 4*S*-selective variants allowed the highly (>98:<2) diastereoselective synthesis of both 4*S*- and 4*R*-configured products under kinetic control from the same starting materials. Conversion of an essentially nonselective aldolase into a pair of complementary biocatalysts will be of enormous interest to synthetic chemists. Furthermore, since residues identified as critical for stereoselectivity are conserved among members of the NAL superfamily, the approach might be extended to the evolution of other useful biocatalysts for the stereoselective synthesis of biologically active molecules.

Introduction

Enzymatic biocatalysis has tremendous potential in the synthesis of pharmaceuticals, agrochemicals, and fine chemicals.[1−5] However, although enzymes are highly efficient catalysts that exhibit remarkable control over the reaction catalyzed, its substrate selectivity, and its stereochemical course, industrial biocatalysis was projected to account for less than 15% of the global worldwide revenue ($9.5 billion) for chiral products by 2005.[6] This is due, in part, to the instability of enzymes, especially in organic solvents, and problems of substrate and product inhibition. Recent advances in directed evolution and high-throughput screening for new enzymes have helped to overcome a number of these drawbacks, and engineered enzymes are now finding use in a range of important transformations.[7−9]

The most useful catalysts for synthetic chemistry are those which may be exploited in the preparation of a wide range of products, but here, the exquisite selectivity of enzymes can pose a problem: enzymes generally have a limited substrate repertoire and are often available for only one of the possible stereoisomeric products. By contrast, the synthetic chemist may require a stereoisomer that is not the product of an available enzyme or more than one stereoisomeric product. Directed evolution has again been used to tackle these problems by altering the stereochemistry of different enzymes.[9] Two distinct approaches have been used. In the first approach (Figure 1A) the enzyme is engineered or evolved to accept a stereoisomer of the natural substrate.[10−14] The directed evolution of more effective enzymes

(1) Liese, A.; Seelbach, K.; Wandrey, C. *Industrial Biotransformations*; Wiley-VCH: Weinheim, Germany, 2000.
(2) Schmidt-Dannert, C.; Arnold, F. H. *Trends Biotechnol.* **1999**, *17*, 135−136.
(3) Silvestri, M.; DeSantis, G.; Mitchell, M.; Wong, C. H. *Top. Stereochem.* **2003**, *23*, 267−342.
(4) Straathof, A. J. J.; Panke, S.; Schmid, A. *Curr. Opin. Biotechnol.* **2002**, *13*, 548−556.
(5) Thayer, A. M. *Chem. Eng. News*, Aug 14, 2006, pp 15−25.
(6) Rouhi, A. M. *Chem. Eng. News*, Jun 14, 2004, pp 47−62.
(7) DeSantis, G.; Zhu, Z.; Greenberg, W. A.; Wong, K.; Chaplin, J.; Hanson, S. R.; Farwell, B.; Nicholson, L. W. R.; Weiner, C. L. *J. Am. Chem. Soc.* **2002**, *124*, 9024−9025.
(8) Greenberg, W. A.; Varvak, A.; Hanson, S. R.; Wong, K.; Huang, H.; Chen, P.; Burk, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5788−5793.
(9) Reetz, M. *Adv. Catal.* **2006**, *49*, 1−69.
(10) Liebeton, K.; Zonta, A.; Schimossek, K.; Nardini, M.; Lang, D.; Dijkstra, B. W.; Reetz, M. T.; Jaeger, K. E. *Chem. Biol.* **2000**, *7*, 709−718.

10.1021/ja065233q CCC: $33.50

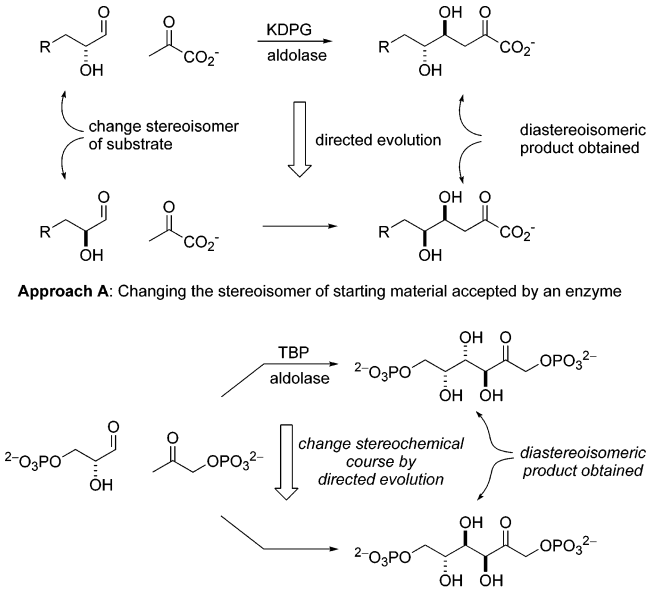

Figure 1. Directed evolution of enzyme stereochemistry. Two approaches are possible for the directed evolution of enzymes to produce the opposite stereochemistry of products compared to the wild-type enzyme. These approaches are illustrated by specific examples. Approach A: An enzyme (e.g., KDPG-aldolase) is evolved to select the enantiomeric substrate to obtain the diastereoisomeric product.[14] Approach B: The relative configuration of the product is altered by modifying the stereochemical course of the reaction catalyzed. This is illustrated by the evolution of the mechanism of tagatose-1,6-bisphosphate aldolase to catalyze *si*-face attack of dihydroxyacetone phosphate on glyceraldehyde 3-phosphate rather than the *re*-face attack catalyzed by the wild-type enzyme. This results in production of the diastereoisomer fructose 1,6-bisphosphate rather than the natural tagatose 1,6-bisphosphate.[19]

for the kinetic resolution of racemic substrates is based on a similar premise: by enhancing the discrimination between enantiomeric starting materials, more highly enantiomerically enriched products are obtained.[7,15] A fundamentally different approach (Figure 1B) involves modification of the stereochemical *course* of enzyme-catalyzed bond formation. Here, the starting materials accepted by the engineered enzymes are the same as those used by the wild-type enzyme; however, *by controlling the configuration of new stereogenic centers formed during the reaction*, stereoisomeric products have been obtained with enzymes which catalyze P−O bond formation,[16] reduction of carbonyl groups,[17] or addition of water to a *p*-quinone methide intermediate.[18] We further demonstrated that the stereochemical course of enzyme-catalyzed carbon−carbon bond formation may be modified by directed evolution. We engineered an aldolase to synthesize a (3*S*,4*S*)-configured product, fructose 1,6-bisphosphate, instead of a (3*S*,4*R*)-configured product, tagatose 1,6-bisphosphate.[19] This change must perturb the orientation of the substrates such that the normal attack of the donor, dihydroxyacetone phosphate, on the *re* face of the acceptor aldehyde, glyceraldehyde 3-phosphate, is converted into attack on the opposite, *si* face, leading to a stereoisomeric product.

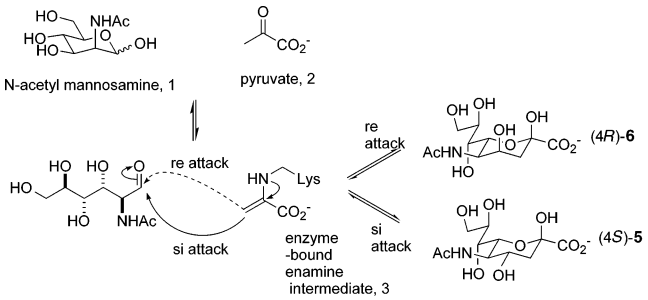

Figure 2. Reaction catalyzed by NAL. Wild-type NAL catalyzes the aldol condensation of *N*-acetylmannosamine (1) with pyruvate (2) to generate *N*-acetylneuraminic acid (4). The new carbon−carbon bond is formed between C3 and C4 of the product, and the configuration at C4 depends on the stereochemical course of the reaction. Pyruvate is bound via a Schiff base linkage to a lysine residue, and this facilitates proton abstraction from the methyl group to generate the enzyme bound enamine (3). Attack of this enamine on the *si* face of the aldehyde carbonyl group (solid line; above the plane of the paper) yields the 4*S*-configured product, **5**, whereas attack on the *re* face of the aldehyde (dotted line; behind the plane of the paper) yields the 4*R*-configured product, **6**. The facial stereoselectivity of the wild-type NAL reaction is poor.[36]

The aldol reaction is one of the cornerstones of synthetic organic chemistry, leading to formation of a new carbon−carbon bond and up to two new stereogenic centers. The enzymes that catalyze this general reaction type—the aldolases—have therefore found growing use as synthetic tools.[4,20−23] In addition, direct asymmetric aldol condensations have been developed which exploit catalytic antibodies,[24−26] organocatalysts,[27,28] and bimetallic zinc catalysts.[29] One particularly useful aldolase is the enzyme *N*-acetylneuraminic acid lyase (NAL, otherwise known as *N*-acetylneuraminic aldolase or sialic acid aldolase), which catalyzes the reversible aldol condensation (Figure 2) of *N*-acetylmannosamine (ManNAc, **1**) with pyruvate (**2**) to produce the sialic acid, *N*-acetylneuraminic acid (**4**). Sialic acids are essential components of complex carbohydrates which play pivotal roles as recognition signals in a variety of biological processes.[30,31] These events are particularly well exemplified in host−pathogen and host−parasite interactions where the oligosaccharide is often required for invasion, infectivity, and survival of the invading organism in the host (for a review of

(11) May, O.; Nguyen, P. T.; Arnold, F. H. *Nat. Biotechnol.* **2000**, *18*, 317−320.
(12) Reetz, M. T.; Torre, C.; Eipper, A.; Lohmer, R.; Hermes, M.; Brunner, B.; Maichele, A.; Bocola, M.; Arand, M.; Cronin, A.; Genzel, Y.; Archelas, A.; Furstoss, R. *Org. Lett.* **2004**, *6*, 177−180.
(13) Horsman, G. P.; Liu, A. M. F.; Henke, E.; Bornscheuer, U. T.; Kazlauskas, R. J. *Chem. Eur. J.* **2003**, *9*, 1933−1939.
(14) Fong, S.; Machajewski, T. D.; Mak, C. C.; Wong, C. *Chem. Biol.* **2000**, *7*, 873−883.
(15) Alexeeva, M.; Enright, A.; Dawson, M. J.; Mahmoudian, M.; Turner, N. J. *Angew. Chem., Int. Ed.* **2002**, *41*, 3177.
(16) Jiang, R. T.; Dahnke, T.; Tsai, M. D. *J. Am. Chem. Soc.* **1991**, *113*, 5485−5486.
(17) Nakajima, K.; Kato, H.; Oda, J.; Yamada, Y.; Hashimoto, T. *J Biol. Chem.* **1999**, *274*, 16563−16568.
(18) van Den Heuvel, R. H.; Fraaije, M. W.; Ferrer, M.; Mattevi, A.; van Berkel, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9455−9460.
(19) Williams, G. J.; Domann, S.; Nelson, A.; Berry, A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3143−3148.
(20) Wong, C.-H.; Whitesides, G. M. *Enzymes in Synthetic Organic Chemistry*; Pergamon: Oxford, 1994.
(21) Gijsen, H. J. M.; Qiao, L.; Fitz, W.; Wong, C. H. *Chem. Rev.* **1996**, *96*, 443−473.
(22) Breuer, M.; Hauer, B. *Curr. Opin. Biotechnol.* **2003**, *14*, 570−576.
(23) Machajewski, T. D.; Wong, C. H. *Angew. Chem., Int. Ed.* **2000**, *39*, 1352−1374.
(24) Wagner, J.; Lerner, R. A.; Barbas, C. F. *Science* **1995**, *270*, 1797−1800.
(25) Barbas, C. F.; Heine, A.; Zhong, G. F.; Hoffmann, T.; Gramatikova, S.; Bjornestedt, R.; List, B.; Anderson, J.; Stura, E. A.; Wilson, I. A.; Lerner, R. A. *Science* **1997**, *278*, 2085−2092.
(26) Zhong, G.; Lerner, R. A.; Barbas, C. F. *Angew. Chem., Int. Ed.* **1999**, *38*, 3738−3741.
(27) Nothrup, A. B.; MacMillan, D. W. C. *J. Am. Chem. Soc.* **2002**, *124*, 6798−6799.
(28) Bogevig, A.; Kumaragurubaran, N.; Jorgensen, K. A. *Chem. Commun.* **2002**, 620−621.
(29) Trost, B. M.; Ito, H. *J. Am. Chem. Soc.* **2000**, *122*, 12003−12004.
(30) *Biology of Sialic acids*; Plenum Press: New York and London, 1995.
(31) McGuire, E. J. In *The Biology of Sialic Acids*; Rosenberg, A., Schengrund, C. L., Eds.; Plenum Publishing Corp.: New York, 1976.

the roles of sialic acids see ref 30). Sialic acid mimetics can, therefore, be important chemotherapeutic agents,[32] for example, the anti-influenza drugs Oseltamivir, Relenza, and Tamiflu.[33] NAL has been used in a number of syntheses of such analogues,[23] and structure-guided protein engineering has also been used to tailor the enzyme for the parallel synthesis of sialic acid mimetics.[34,35]

Despite its usefulness in the synthesis of complex carbohydrates, wild-type NAL[36] as well as other members of the NAL-superfamily[37] show poor facial stereoselectivity during carbon−carbon bond formation. During the condensation of ManNAc **1** with pyruvate **2**, one stereogenic center is created at C4 of the product, and two diastereomeric products—with 4*R* and 4*S* configurations—are possible (Figure 2). The *R* configuration is produced via attack of an intermediate enamine complex (**3** in Figure 2) between pyruvate and a lysine residue (Lys-165 in *Escherichia coli*)[38] onto the *re* face of the aldehyde carbonyl of ManNAc, whereas the *S* configuration is produced by attack of the *same* intermediate onto the opposite face of the *same* aldehyde substrate. The poor facial stereoselectivity of the enzyme means that the stereochemical outcome of reactions catalyzed by wild-type NAL is generally thermodynamically controlled, depending only on the relative thermodynamic stability of the two possible products.[21,23,36,39,40] In principle, synthesis of the less thermodynamically favored product is also possible if the route to it is kinetically favored, i.e., if the free energy of the transition state leading to it is lower than that leading to the thermodynamically favored product. Under these conditions, it is possible to isolate the thermodynamically less-favored product well before thermodynamic equilibrium has been reached. However, where the desired product is neither kinetically nor thermodynamically favored, it is impossible to prepare using wild-type NAL. To allow the use of NAL to access either stereoisomer of sialic acid mimetics, we set ourselves the challenging goal of engineering the relatively nondiscriminatory wild-type NAL into a pair of stereochemically complementary aldolases which catalyze the selective formation of either possible diastereoisomeric aldol product from the *same* substrates under kinetic control.

Materials and Methods

Materials. The restriction enzymes *Eco*RI and *Hin*dIII, T4 DNA ligase, *Pfu* DNA polymerase, Wizard Miniprep DNA purification kits, and DNA markers were from Promega, Southampton, U.K. Lactate dehydrogenase was from Boehringer Mannheim Gmbh (Germany). Nicotinamide adenine dinucleotide (reduced form) [NADH] lithium salt, *N*-acetylneuraminic acid (Type IV), and pyruvate were from Sigma Chemicals Ltd. (Poole, U.K.). The QIAquick DNA extraction kit was from QIAgen. Chelating Sepharose Fast Flow resin was supplied by Pharmacia (Milton Keynes, U.K.). Dialysis tubing was from Medicell International, Ltd. All other chemicals were of analytical grade.

Bacterial Strains and Plasmids. *E. coli* EP-Max 10B competent cells {Δ(*mrr-hsd*RMS-mrcBC) Φ80d *lac*ZΔM15 Δ*lac*X74 *deoR recA1 araD139* Δ(*ara, leu*)7697 *gal*U *gal*K *rps*L *nup*G λ^-} were from Bio-Rad, U.K. The expression vector pK*nan*A-His_6 (expressing the wild-type NAL) and pK*nan*A-His_6/E192N (expressing the E192N mutant NAL) were as previously described.[34] pKK223-3 was obtained from Pharmacia.

Error-Prone PCR. Error-prone PCR was performed using the Stratagene GeneMorph I kit as described by the manufacturer. Briefly, 1 ng of the plasmid pK*nan*A-His_6/E192N was used as the template with the oligos, pKK-FOR (5′-GGA TAA CAA TTT CAC ACA GG-3′) and pKK-REV (5′-GGC TGA AAA TCT TCT TCT-3′). Mutant genes were cloned into pKK223-3 and ∼2000 library members screened for activity with the *R* and *S* conformations of the dipropylamide, **5** and **6** (0.2 mM).

Saturation Mutagenesis at Positions Ala-10, Thr-48, Thr-167, and Ser-208. Saturation mutagenesis at positions Ala-10, Thr-48, Thr-167, and Ser-208 was performed using the Megaprimer PCR method[41] as previously described.[34] The oligonucleotides used were A10X, 5′-CGT GGC GTA ATG NNK GCA CTC CTG ACT-3′; T48X, 5′-GTG GGT GGT TCG NNK GGC GAG GCC TTT GTA-3′; T167X, 5′-GCG CTG AAA CAG NNK TCT GGC GAT CTC-3′; and S208X, 5′-GGT GGT ATC GGC NNK ACC TAC AAC ATC-3′. Full-length PCR products were purified from agarose gels and subcloned into pKK223-3 as described previously.[34]

Library Screening, Protein Expression, and Purification. Individual colonies from the library were picked from agar plates, transferred to 96-well microtitreplates, and screened for activity with the 4*R* and 4*S* diastereomers of the dipropylamide (**5** and **6**; Figure 3) by coupling the production of pyruvate to the oxidation of NADH using lactate dehydrogenase.[34] No correction was made for levels of protein expression or bacterial cell growth during the initial library screening. Active library members were selected, and the variants of NAL were expressed and purified as previously described.[34]

Enzyme Assay. Kinetic parameters for cleavage of sialic acid or analogues were determined for the pure enzymes by a standard coupled assay with lactate dehydrogenase and NADH. The assay was performed in a 1 mL cuvette at 25 °C containing 50 mM Tris-HCl (pH 7.5), 0.2 mM NADH, 0.5U LDH, and a suitable aliquot of aldolase (1−100 μg). The reaction was initiated by addition of varying concentrations of substrate. The decrease in absorbance at 340 nm was recorded as the measure of enzyme activity on a Uvikon 930 spectrophotometer. One unit of aldolase activity is defined as the amount of enzyme which catalyzes the oxidation of 1 μmol of NADH/min in this system using the molar extinction coefficient of NADH as 6220 M^{-1} cm^{-1}. Kinetic parameters were estimated by nonlinear regression analysis.[42]

Time Course of the Reaction Followed by 500 MHz ^{1}H NMR Spectroscopy. Oxygen (10 min, 0.6 psi), ozone (3 min, 0.6 psi), and oxygen (5 min, 0.6 psi), generated using a Welsbach generator, were bubbled sequentially through a stirred solution of γ,δ-unsaturated amide[43] (22 mg, 0.1 mmol) in methanol (1 mL) at −78 °C. Dimethyl sulfide (0.15 mL) was added, and the mixture was allowed to warm to room temperature, stirred for 2.5 h, and evaporated under reduced pressure. The residue was dissolved in a deuterated aqueous sodium phosphate buffer (20 mM, pD7.4), sodium pyruvate (55 mg, 0.5 mmol) was added, the pD was adjusted to 7.4 with aqueous ammonia (1 M),

(32) Koketsu, M.; Nitoda, T.; Sugino, H.; Juneja, L. R.; Kim, M.; Yamamoto, T.; Abe, N.; Kajimoto, T.; Wong, C. H. *J. Med. Chem.* **1997**, *40*, 3332−3335.
(33) Babu, Y. S.; Chand, P.; Bantia, S.; Kotian, P.; Dehghani, A.; El-Kattan, Y.; Lin, T.-H.; Hutchinson, T. L.; Elliott, A. J.; Parker, C. D.; Ananth, S. L.; Horn, L. L.; Laver, G. W.; Montgomery, J. A. *J. Med. Chem.* **2000**, *43*, 3482−3486.
(34) Williams, G. J.; Woodhall, T.; Nelson, A.; Berry, A. *PEDS* **2005**, *18*, 239−246.
(35) Woodhall, T.; Williams, G. J.; Berry, A.; Nelson, A. *Angew. Chem., Int. Ed.* **2005**, *44*, 2109−2112.
(36) Smith, B. J.; Lawrence, M. C.; Barbosa, J. A. R. G. *J. Org. Chem.* **1999**, *64*, 945−949.
(37) Theodossis, A.; Walden, H.; Westwick, E. J.; Connaris, H.; Lamble, H. J.; Hough, D. W.; Danson, M. J.; Taylor, G. L. *J. Biol. Chem.* **2004**, *279*, 43886−43892.
(38) Barbosa, J. A. R. G.; Smith, B. J.; DeGori, R.; Ooi, H. C.; Marcuccio, S. M.; Campi, E. M.; Jackson, W. R.; Brossmer, R.; Sommer, M.; Lawrence, M. C. *J. Mol. Biol.* **2000**, *303*, 405−421.
(39) Fitz, W.; Schwark, J. R.; Wong, C. H. *J. Org. Chem.* **1995**, *60*, 3663−3670.
(40) Lin, C.-H.; Sugai, T.; Halcomb, R. L.; Ichikawa, Y.; Wong, C.-H. *J. Am. Chem. Soc.* **1992**, *114*, 10138−10145.
(41) Reikofski, Y.; Tao, B. Y. *Biotechnol. Adv.* **1992**, *10*, 535−547.
(42) Cleland, W. W. *Methods Enzymol.* **1979**, *63*, 103−138.
(43) Woodhall, T.; Williams, G. J.; Berry, A.; Nelson, A. *Org. Biomol. Chem* **2005**, *3*, 1795−1800.

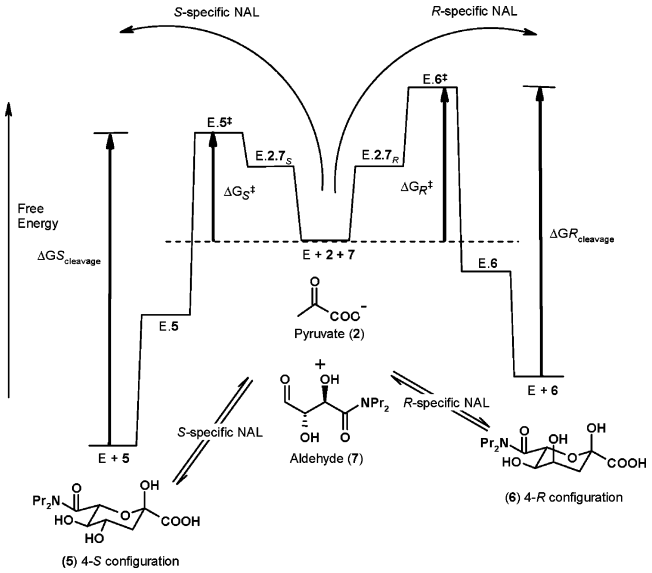

Figure 3. Theoretical reaction profiles of aldol condensation reactions from pyruvate (**2**) and aldehyde (**7**) to the alternative (4*S*- and 4*R*-configured) dipropylamides (**5** and **6**, respectively) are shown (center outward) as well as the retro-aldol cleavage reactions used in the screening assays (outside toward the center) under conditions where [substrate] $< K_m$. The thermodynamic products of the aldol condensation reactions (E + **2** + **7** → E + **5** and E + **2** + **7** → E + **6**) are controlled by the relative energies of (E + **5**) and (E + **6**). The kinetic consequences of the E + **2** + **7** → E + **5** and E + **2** + **7** → E + **6** reactions are determined by the activation energies $\Delta G_S^{\ddagger}$ and $\Delta G_R^{\ddagger}$, respectively. A measure of the heights of the E·**5**‡ and E·**6**‡ transition states can be gained from $\Delta G_{S_{cleavage}}{}^{\ddagger}$ and $\Delta G_{R_{cleavage}}{}^{\ddagger}$, which in turn are related to k_{cat}/K_m for the cleavage of the *S*-isomer **5** and k_{cat}/K_m for the *R*-isomer **6**, respectively. Note that the changes in Gibbs free energies schematically shown are for the concentrations of substrates used in the experiments ([substrate] $< K_m$) and not for standard states of 1 M.

the volume was adjusted to 700 μL, and the sample was transferred to an NMR tube, and an initial 500 MHz ^{1}H NMR spectrum was obtained. The enzyme (3 or 4 mg of the *S*- or *R*-selective enzymes, respectively) was added, and 500 MHz ^{1}H NMR spectra were acquired at 5 min intervals. The mole fractions of the components were determined by integration of the spectra using X-WinNMR [4-H of the major pyranose form of the *S*-configured product **5**, 3.97−3.91 ppm; 4-H of the major pyranose form and 5-H of a furanose form of the *R*-configured product **6**, 4.28−4.20 ppm; 4-H of the hydrated form of the aldehyde **7**: 5.21−5.18 ppm],[43] accounting for the mixtures of the components present [**5**, 79:8:7:6 mixture of two furanose and two pyranose forms; **6**, 44:32:16:8 mixture of two pyranose and two furanose forms],[43] using the total integral for the terminal methyl groups of the NPr_2 substituent as an internal standard. Time courses of the normalized data were plotted using SigmaPlot.

General Procedure for the Enzymatic Synthesis of the Sialic Acid Mimetics 5 and 6. Oxygen (10 min, 0.6 psi), ozone (3 min, 0.6 psi), and oxygen (5 min, 0.6 psi), generated using a Welsbach generator, were bubbled sequentially through a stirred solution of γ,δ-unsaturated amide[43] (88 mg, 0.4 mmol) in methanol (1 mL) at −78 °C. Dimethyl sulfide (0.15 mL) was added, and the mixture was allowed to warm to room temperature, stirred for 3 h, and evaporated under reduced pressure. The residue was dissolved in an aqueous sodium phosphate buffer (20 mM, pD7.4), sodium pyruvate (94 mg, 0.85 mmol) was added, the pH was adjusted to 7.4 with aqueous ammonia (1 M), and the variant enzyme was added (400 μL, 22 mg mL^{-1} of the *R*-selective mutant in phosphate buffer; 700 μL, 31 mg mL^{-1} of the *S*-selective enzyme in phosphate buffer). Conversion of the reaction was determined by analysis of 100 μL aliquots of the reaction mixture by 500 MHz ^{1}H NMR spectroscopy. After 5 h, the reaction mixture was acidified with formic acid and evaporated under reduced pressure, and the crude product was purified by ion-exchange chromatography (Dowex 1 × 8-100, formate form, 2 g, gradient elution: 0 → 1.0 M formic acid), evaporated under reduced pressure, purified by stirring in water with 2 mL of Bakers' yeast in water, the pH adjusted to 5.5 with aqueous ammonium (1 M), and dialyzed to give either the sialic mimetic **5** (64 mg, 66%) or **6** (68 mg, 70%) as colorless foams, spectroscopically identical to those previously reported.[43] The yields of products take into account removal of aliquots for determination of the reaction conversion by NMR spectroscopy.

Results

In previous work, we created a variant of NAL, E192N, which has broad substrate specificity toward a range of mimetics, such as **5** and **6** (Figure 3), in which the glycerol moiety of the natural substrate, *N*-acetylneuraminic acid,[34,35] has been replaced with a dialkylaminocarbonyl group. This enzyme has a higher catalytic activity toward **5** and **6** than the wild-type enzyme has for its natural substrate. The E192N variant was thus chosen as an ideal starting point for evolving/engineering a pair of complementary enzymes for the stereocontrolled *synthesis* of the 4*R*- or 4*S*-configured dipropylamides, **5** and **6**. Since the NAL reaction is fully reversible, assays and activity screens were performed in the cleavage (retro-aldol) direction with either **5** or **6** separately as substrate to yield pyruvate **2** and aldehyde **7**, and the pyruvate produced was detected by its reduction with NADH using lactate dehydrogenase. By the principle of microscopic reversibility, stereoselective enzymes in the cleavage assay will also show increased stereoselectivity in the desired, condensation direction. Synthetic chemical routes were therefore developed to the two diastereoisomeric screening substrates,[43] **5** and **6**, and a measure of the stereochemical preference of NAL and the evolved variants for the cleavage of these substrates was made by comparison of the initial rates of reaction and their resulting kinetic parameters in separate assays using each substrate.

The kinetic parameters of wild-type NAL and the E192N variant with the dipropylamides **5** and **6** are given in Table 1. As judged by the k_{cat}/K_m values for the two diastereoisomers, the wild-type enzyme functions with the 4*S*-configured dipropylamide **5** ~3-fold better than the 4*R*-configured substrate **6**, mainly due to a ~3-fold higher k_{cat} with the *S* isomer. There is no significant difference in the K_m values for the two diastereoisomers, implying that the enzyme is able to bind each stereoisomeric substrate with similar affinity, in agreement with previous data for the binding of the natural substrates to the wild-type enzyme.[36] The E192N variant shows marginally relaxed stereoselectivity compared with wild-type NAL, preferring the 4*S* isomer (**5**) only ~2-fold better than the 4*R* isomer (**6**) (Figure 4).

Evolution Strategy and Selection of Variants. The subtleties of enzymic reactions make rational engineering of the stereochemical course of bond formation extremely challenging, even when high-resolution X-ray crystal structures of the enzyme in the presence of substrates or substrate analogues are available. Thus, although a range of crystal structures of NAL have been solved in the presence of substrate analogues,[38] we used a combination of error-prone PCR (to determine which residues in NAL are important for stereochemical discrimination) and saturation mutagenesis (to discover the 'best' amino acid substitutions) to create a pair of stereochemically complementary enzymes. At each stage in the directed evolution variants were selected on the basis of their stereoselectivity, measured by the

Table 1. Steady-State Kinetic Parameters of the Wild-Type and Variant Enzymes

	dipropylamide (**5**) (4*S*-configuration)			dipropylamide (**6**) (4*R*-configuration)			
enzyme	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m (min^{-1}•mM^{-1})	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m (min^{-1}•mM^{-1})	(k_{cat}/K_m)(*R*-**6**)/ (k_{cat}/K_m)(*S*-**5**)
wild-type	230 ± 6	12 ± 1	19	73 ± 4	11 ± 2	6.6	0.35
E192N	450 ± 14	0.8 ± 0.07	563	130 ± 3	0.4 ± 0.04	325	0.58
19D3	320 ± 10	0.7 ± 0.09	457	3 ± 0.1	0.06 ± 0.01	50	0.11
15C7	51 ± 3	2.7 ± 0.4	18.9	16 ± 0.4	0.7 ± 0.05	22.9	1.21
E192N/T48A	510 ± 35	2.3 ± 0.4	222	5 ± 0.1	0.03 ± 0.002	166	0.75
E192N/T167V	12 ± 0.8	2.3 ± 0.4	5	18 ± 0.7	0.83 ± 0.09	22	4.4
E192N/T167G	5 ± 0.1	0.5 ± 0.03	10	NM[a]	NM[a]	<0.2[a]	<0.02[a]
E192N/A10V	300 ± 14	2.8 ± 0.3	107	62 ± 2.3	0.54 ± 0.07	115	1.1
E192N/S208G	58 ± 1.2	0.5 ± 0.04	116	23 ± 0.5	0.11 ± 0.01	209	1.8
E192N/S208V	7.6 ± 0.4	2.6 ± 0.3	2.9	4.4 ± 0.1	0.55 ± 0.05	8	2.75
E192N/A10V/T48A	170 ± 3	0.74 ± 0.05	230	3.1 ± 0.1	0.15 ± 0.02	21	0.09
E192N/T167V/S208 V	0.1 ± 0.003	8.5 ± 0.4	0.012	1.1 ± 0.04	1.9 ± 0.2	0.58	48

[a] No activity greater than the background rate of 5 nmol of NADH consumed/min could be detected even with 4 mg of protein.

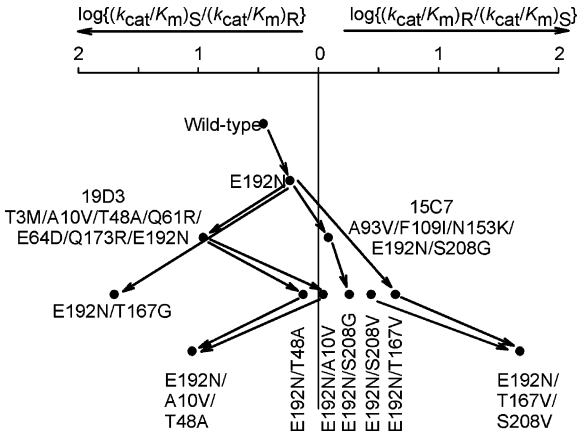

Figure 4. Schematic of the divergent creation of two stereocomplementary NAL enzymes. The preferred stereochemistries of reaction of wild-type and variant NAL enzymes are shown as the logarithm of the ratio of catalytic efficiencies (k_{cat}/K_m) toward the 4-*S*- and 4-*R*-configured substrates **5** and **6**, respectively. A value of zero thus represents an enzyme which is nondiscriminatory between the two diastereoisomers.

ratio of the k_{cat}/K_m values for the two diastereoisomeric screening substrates **6** and **5** (i.e., $(k_{cat}/K_m)_{(6)}/(k_{cat}/K_m)_{(5)}$). *R*-Selective enzymes therefore have stereoselectivity ratios greater than 1, whereas *S*-selective enzymes have ratios less than 1. The 'parental' E192N enzyme has a stereoselectivity ratio of 0.58, reflecting its low *S* selectivity.

Error-Prone PCR. An error-prone PCR library was constructed using the gene encoding the E192N enzyme as template, and the library was screened separately with the 4*R*- and 4*S*-configured dipropylamides (**5** and **6**). The screens were performed at a substrate concentration (0.2 mM) lower than the K_m values for the parent E192N protein (Table 1) since under these conditions, the activation energies, $\Delta G_{S^{\ddagger}_{cleavage}}$ and $\Delta G_{R_{cleavage}}{}^{\ddagger}$, are for the E + **5** → E•**5**‡ and E + **6** → E•**6**‡ reactions[44] (Figure 3). Importantly, in the desired synthetic reactions aldehyde **7** + pyruvate → **5** and aldehyde **7** + pyruvate → **6** the relative ability to form the two alternative products under kinetic control depends only on the relative heights of these two transition states (E•**5**‡ and E•**6**‡). Comparison of the screening substrate cleavage rates under conditions where most variants are not saturated with either substrate (**5** or **6**) thus allows stereoselectivity in the aldol condensation to be estimated. We previously validated this approach in directed evolution of enzyme reaction stereochemistry for TBP-aldolase.[19]

The error-prone PCR library (~2500 clones) was expressed in *E. coli*, and crude cell extracts were screened, uncorrected for the expression level of the enzyme, at 0.2 mM substrate concentration. The results indicated that most library members displayed approximately the same stereochemical control as the 'parental' E192N variant, even though the level of activity with both isomers was often significantly reduced (data not shown). However, about 1% of library members showed significant differences in activity between the two diastereoisomers. Two variants, designated 15C7 and 19D3, were selected as displaying greater stereoselectivity (15C7 approximately 2-fold more *R*-selective and 19D3 approximately 6-fold more *S*-selective) than the E192N 'parental' variant. DNA sequencing of these clones revealed that both enzymes contained multiple amino acid substitutions. 15C7 contained four changes in addition to the E192N substitution: A93V, F109I, N153K, and S208G. In contrast, the 19D3 variant contained an additional six substitutions: T3M, A10V, T48A, Q61R, E64D, and Q173R.

The two selected variants, 19D3 and 15C7, were purified to homogeneity, and the steady-state kinetic parameters with both diastereoisomers of the dipropylamides (**5** and **6**) were determined (Table 1). 19D3 had a stereoselectivity ratio of 0.11 and therefore discriminates more strongly in favor of the *S*-configured isomer **5** than the E192N 'parental' variant (Figure 4), mainly due to a 43-fold decrease in k_{cat} for the *R*-isomer **6**. By contrast, variant 15C7 has a stereoselectivity ratio of 1.21: a slight, but significant, preference for the *R*-configured substrate (Figure 4) which stems mainly from a slightly larger increase in the K_m value for the *S*-isomer **5** (~3-fold) than for the *R*-isomer **6** (~2-fold) (Table 1).

Since synergy between mutated residues is rare in novel enzymes identified by directed evolution,[45,46] we reasoned that the changes in stereochemical control might be brought about by a subset of the mutations present in 19D3 and 15C7 and that some of the other amino acid substitutions could be detrimental to activity or stability. Examination of the crystal structures of the *E. coli* and *Haemophilus influenzae* NAL enzymes[38,47] revealed that, of the substituted residues found in

(44) Fersht, A. R. *Structure and Mechanism in Protein Science: A guide to catalysis and protein folding*; W. H. Freeman and Co.: New York, 1999.

(45) Zaccolo, M.; Gherardi, E. *J. Mol. Biol.* **1999**, *285*, 775−783.
(46) Matsumura, I.; Ellington, A. D. *J. Mol. Biol.* **2001**, *305*, 331−339.

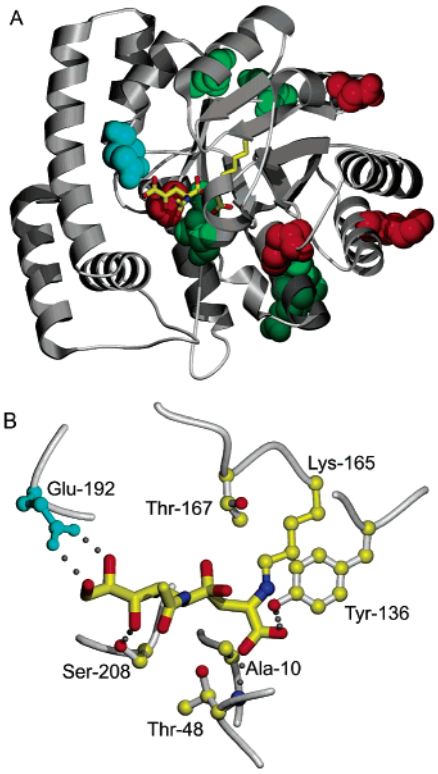

Figure 5. Location of amino acid substitutions in NAL. The structure of the *H. influenzae* NAL in complex with the substrate analogue 4-oxosialic acid[38] is shown, but the residues are numbered according to the *E. coli* protein. 4-Oxosialic acid forms a Schiff base to the enzyme via Lys-165. The wild-type residue Glu-192 (cyan) is replaced by Asn in the 'parental' E192N and in all the variants selected. This substitution results in an enzyme that is highly active with the dipropylamide screening substrates used.[34,35] (A) The locations of residues Ala-93, Phe-109, Asn-153, and Ser-208 substituted in variant 15C7 are shown in red, and the location of residues Thr-3, Ala-10, Thr-48, Gln61, Glu-64, and Gln173 substituted in variant 19D3 are shown in green. (B) The location of amino acid residues (ball-and-stick representation) targeted for saturation mutagenesis: Ala-10, Thr-48, and Ser-208 substituted in 15C7 and 19D3 and identified as lying in the enzyme active site along with residue Thr-167 which is also located near the C4 position of the substrate (plain bond representation).

15C7 and 19D3, only Ala-10 and Thr-48 (in 19D3) and Ser-208 (in 15C7) (*E. coli* numbering) made direct contact with the substrate (Figure 5A). We reasoned that these residues would be most likely to influence the reaction stereochemistry, and we therefore used saturation mutagenesis of Ala-10, Thr-48, and Ser-208 separately to determine whether substitutions in these positions alone are sufficient to explain the changes found in 15C7 and 19D3 and find the best substitutions for stereochemical discrimination. In addition, we noted that Thr-157 in the related enzyme 2-keto-3-deoxygluconate aldolase (KDGA) (equivalent to Thr-167 in NAL) makes hydrogen bonds to the epimeric C4-OH of the product[37] (Figure 5B), and we also chose to carry out saturation mutagenesis of Thr-167 to determine whether it might play a role in determining the stereochemistry of the reaction.

Saturation Mutagenesis. Saturation mutagenesis libraries (A10X, T48X, T167X, and S208X) were constructed, and crude cell extracts from ~300 colonies from each library were screened for activity with both the 4*S*- and 4*R*-configured isomers (**5** and **6**). The results, uncorrected for enzyme expression and bacterial cell growth levels, are shown in Figure 6. In each case, a significant proportion of the library members were inactive with both substrates, showing that many amino acid substitutions are not tolerated for activity, as might be expected for active site residues. For each library, many of the variants showed equal changes in activity with both the *R*- and *S*-configured substrates, i.e., altered activity but no change in stereochemical control. It thus appears that no single substitution at position Ala-10 produces an enzyme with altered stereochemical course (Figure 6), suggesting that the A10V substitution found in 19D3 is not sufficient to explain its improved *S* selectivity. The switch in selectivity for 19D3 must therefore stem from one or more of the other amino acid substitutions or from A10V acting in concert with one or more of the other changes. The S208V variant was isolated from relatively inactive variants in the S208X library and had a slight preference for the *R*-configured substrate. A number of particularly interesting variants were identified in the T48X and T167X libraries (populations **A**−**E**, Figure 6), suggesting that Thr-48 and Thr-167 may both play important roles in controlling the stereochemistry of the reaction. It is particularly interesting to note that some of the subpopulations of variants in the T167X library appear to have opposite stereochemical preferences (for example, population **C** compared with **D** or **E**), suggesting a major role for Thr-167 in stereochemical discrimination. DNA sequencing of four selected clones revealed that population **A** in the T48X library (Figure 6) contained the substitution T48P, whereas those variants with higher activity (**B**; Figure 6) contained the changes T48A or T48V. Similarly, representative clones from subpopulations **D** and **E** from the T167X library contained T167G or T167A substitutions, respectively. Representatives of population **C**, the only group of variants which appeared to demonstrate a small but significant preference for the *R*-configured screening substrate, contained the mutation T167V.

Detailed kinetic characterization of a number of purified enzymes was carried out in order to determine accurately the importance of each of the identified amino acid substitutions in the discrimination between the *R*- and *S*-configured screening substrates. As examples of *S*-selective enzymes, T48A variants were chosen over T48P and T48V variants since the former showed consistently higher activities. Similarly, since T167G variants (**D** in Figure 6) showed excellent stereochemical discrimination, having little or no measurable *R* activity, this substitution was chosen in preference to T167A (**E** in Figure 6) despite the fact that the latter generally showed higher overall activity. The T167V (population **C**) and S208V variants were selected as variants showing *R*-isomer selectivity. The E192N/T48A, E192N/T167G, E192N/T167V, and E192N/S208V variants were therefore purified and characterized kinetically (Table 1 and Figure 4).

Construction of an *S*-Selective Enzyme. The purified E192N/T48A variant shows a very similar selectivity (stereoselectivity ratio 0.75) for the stereochemical course of the reaction as the E192N single mutant (stereoselectivitiy ratio 0.58). This result, which appears to contradict the high *S* selectivity found in the screening results shown in Figure 6 (population **B**), is explained by the offset in K_m values for **5** and **6** with the E192N/T48A variant (Table 1). Coupled with

(47) Izard, T.; Lawrence, M. C.; Malby, R. L.; Lilley, G. G.; Colman, P. M. *Structure* **1994**, *2*, 361−369.

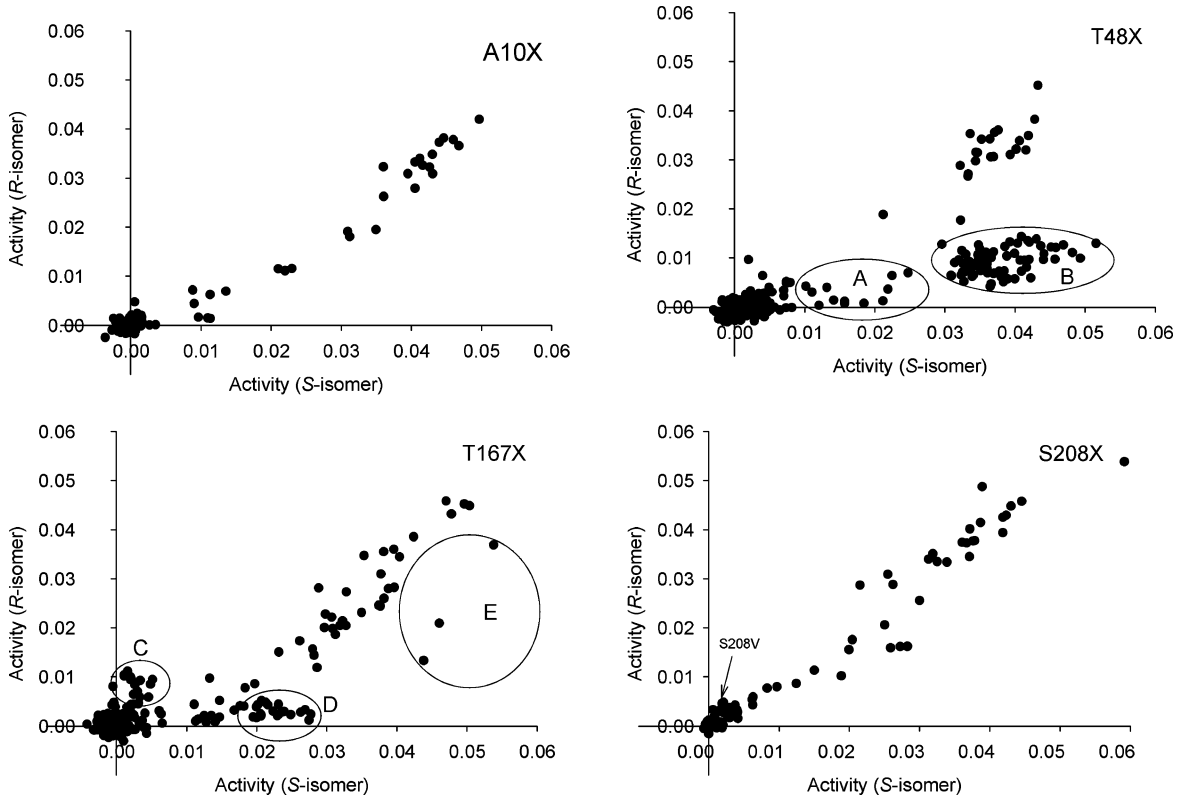

Figure 6. Activity screens for stereoselective variants of E192N from saturation mutagenesis libraries. Identical copies of 96-well microtitreplates containing crude cell lysates of saturation mutagenesis libraries A10X, T48X, T167X, and S208X were screened for activity with the 4*R*- or 4*S*-diastereomeric dipropylamides (**5** and **6**; Figure 3) by coupling the production of pyruvate to the oxidation of NADH using lactate dehydrogenase.[34] Activities for both substrates are plotted as rate of change of absorbance at 340 nm per minute and uncorrected for bacterial cell growth or level of NAL expression: (A) A10X; (B) T48X; (C) T167X; (D) S208X.

the differences in the V_{max} values for the two substrates (Table 1), the expected rate of cleavage of **5** should be ~6-fold greater than that of **6**, entirely consistent with the apparent ~4-fold discrimination displayed by population **B** in the screen (Figure 6). Thus, although substrate concentrations may be chosen under which E192N/T48A will cleave **5** more rapidly than **6**, the kinetic parameters point to this variant being rather nondiscriminatory (Figure 4).

Our saturation mutagenesis data suggest that neither A10V nor T48A are individually responsible for the improved *S* selectivity of 19D3. To further investigate the role of the A10V substitution, we prepared E192N/A10V and E192N/A10V/T48A variants to explore whether their combination in 19D3 might account for the increased stereochemical discrimination observed. The E192N/A10V double mutant was found to have a stereoselectivity ratio of 1.1, resulting from a ~3-fold lower catalytic efficiency (k_{cat}/K_m) for the *R*-configured substrate (**6**) than the E192N single mutant and a ~6-fold lower k_{cat}/K_m with **5**. A10V in the scaffold of E192N wipes out the slight stereochemical discrimination shown by the 'parental' E192N enzyme and results in the least stereoselective enzyme generated (Table 1 and Figure 4). However, combination of T48A with A10V results in an *S*-selective enzyme with a stereoselectivity ratio of 0.09, brought about mainly by a 55-fold higher k_{cat} for **5** than **6**. Thus, combination of T48A with A10V has a synergistic effect, removing the unexpected fall in K_m for the *R*-configured substrate seen in T48A (Table 1) and resulting in a remarkably *S*-selective enzyme. The stereoselectivity ratio (0.09) of this variant closely mirrors that of 19D3 (0.11), and combination of A10V and T48 in 19D3 explains its selection from the error-prone PCR library.

Screening of the saturation libraries (Figure 6) also identified the T167G substitution (population **D**) as leading to an *S*-selective catalyst. In these initial screens, members of this population showed lower activity with the *S* isomer than the T48A clones (population **B**), but most importantly for our selection criteria, T167G variants appeared to have little or no activity with the *R*-configured substrate, even when as much as 4 mg of enzyme was used in the assay. Determination of the kinetic parameters of this enzyme revealed that this simple amino acid substitution results in a dramatic effect on the stereochemistry of the reaction, such that the stereoselectivity ratio is less than 0.02 (Table 1). The E192N/T167G variant is at least 5.5 times more *S* selective than 19D3 and at least 30 times more *S* selective than the E192N 'parental' variant.

Construction of an *R*-Selective Enzyme. The only active site substitution found in the initial *R*-selective variant 15C7 was S208G. Measurement of the kinetic parameters of an E192N/S208G variant constructed by site-directed mutagenesis (Table 1 and Figure 4) showed that a slight change in stereoselectivity (stereoselectivity ratio 1.8) is brought about by the S208G substitution. This preference for the *R*-configured substrate **6** over the *S*-configuration **5** overturns a 1.7-fold preference for the *S*-epimer for the parental E192N variant, and this probably accounts for the selection of 15C7 from the error-prone PCR library. Furthermore, detailed kinetic evaluation of the purified E192N/S208V enzyme identified from the S208X saturation library showed that the S208V substitution produces

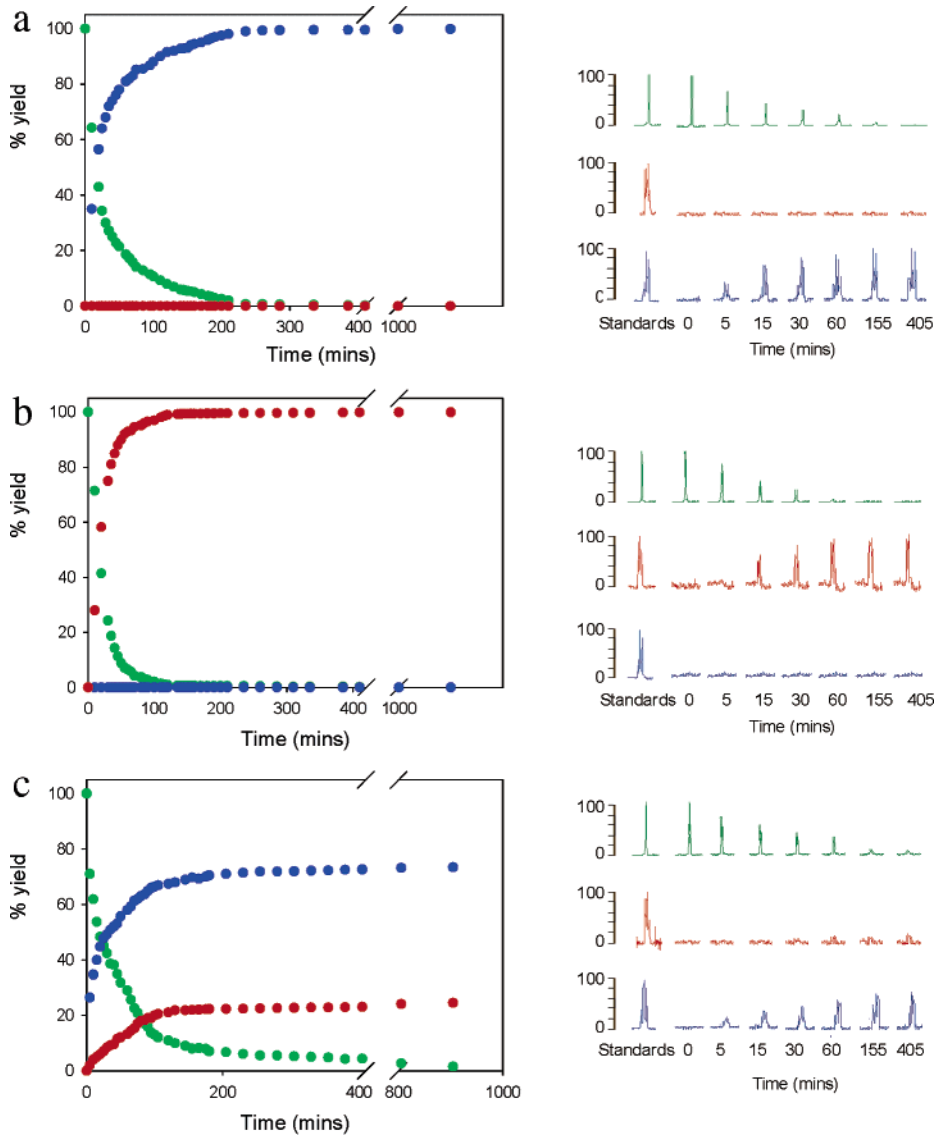

Figure 7. Time course of the stereocontrolled synthesis of diastereomeric sialic acid analogues using the evolved enzymes. The enzymic syntheses of **5** and **6** from the aldehyde **7** and pyruvate **2** (5 mol equiv), catalyzed by the 'parental' E192N enzyme and evolved *S*-selective (E192N/T167G; 3 mg) or *R*-selective (E192/T167V/S208V; 4 mg) aldolases were followed by acquisition of 500 MHz ^{1}H NMR spectra at 5 min intervals. The mole fractions of **5**, **6**, and **7** (left) were determined by integration of diagnostic resonances, accounting for the mixtures of components present: the *S*-configured product **5** (blue), the *R*-configured product **6** (red), and the aldehyde **7** (green). The diagnostic resonances are shown in the right panel and have been scaled such that their heights reflect the proportion of each compound (**5**, **6**, or **7**) determined by integration: 4-H of the major pyranose form of the *S*-configured product **5** (3.97−3.91 ppm), 4-H of the major pyranose form and 5-H of one of the furanose forms of the *R*-configured product **6** (4.28−4.20 ppm), and 4-H of the hydrated form of the aldehyde **7** (5.21−5.18 ppm). (a) Time course with the evolved *S*-selective enzyme, E192N/T167G. (b) Time course with the evolved *R*-selective enzyme, E192/T167V/S208V. (c) Time course with the 'parental' E192N enzyme.

an enzyme with an improved stereoselectivity ratio toward the *R*-configured isomer of 2.75, 1.5-fold better than that of the E192N/S208G variant (Table 1).

The most *R*-selective enzyme found during screening of the saturation mutagenesis libraries was the E192N/T167V variant (population **C** in Figure 6). The kinetics of this purified enzyme showed that it has a stereoselectivity ratio of 4.4, brought about by a slightly higher k_{cat}, as well as a lower (∼3-fold) K_m for the *R*-epimer than for the *S*-epimer (Table 1) (Figure 4). In an attempt to improve the stereoselectivity for the *R*-epimer, we combined the mutations present in the two best *R*-selective enzymes (T167V and S208V) by site-directed mutagenesis to create the E192N/T167V/S208V triple mutant. Characterization of this purified variant showed a dramatic change in stereoselectivity compared with the E192N variant. The E192N/T167V/S208V enzyme has a stereoselectivity ratio of 48, representing a complete reversal of the stereoselectivity of the poorly *S*-stereoselective wild-type and E192N enzymes (Table 1).

The evolutionary strategy adopted (Figure 4) has therefore successfully produced a pair of enzymes with opposite stereoselectivity (E192N/T167G, stereoselectivity ratio < 0.02; and E192N/T167V/S208V, stereoselectivity ratio 48) from an enzyme (E192N) that displayed almost no stereochemical preference for the reaction stereochemistry of aldol cleavage. Both enzymes discriminate in favor of their preferred substrate by a factor of around 50, making them suitable for the clean, stereocontrolled synthesis of sialic acid mimetics.

Stereochemically Controlled Synthesis of Sialic Acid Mimetics. The premise of our strategy was that stereoselective enzymes in the retro-aldol cleavage direction would be valuable

catalysts in synthetically useful and stereoselective aldol condensation reactions. We therefore used 500 MHz ^{1}H NMR spectroscopy to follow the reaction of the aldehyde (**7**) (generated by ozonolysis of the corresponding γ,δ-unsaturated amide[35]) with pyruvate (**2**) in the presence of either of the evolved *R*- or *S*-selective enzymes (Figure 7). The disappearance of the aldehyde (**7**) was judged by the decrease in signal intensity at 5.21−5.18 ppm, while synthesis of the *R*- and *S*-configured products was gauged from the intensity of signals at 4.28−4.20 and 3.97−3.91 ppm, respectively. The peak intensities were normalized to reflect the relative amounts of furanose and pyranose forms of the products present (see Materials and Methods). Unlike the E192N 'parental' variant, the time courses for the reactions catalyzed by the evolved enzymes (Figure 7) show that each enzyme was highly stereoselective (>95:<5 stereoselectivity by NMR) for the synthesis of the appropriate diastereoisomer of the sialic acid mimetics **5** and **6** in high yield (>97% yield by NMR) and within a reasonable reaction time (2 h). The reaction mixtures were purified to give products **5** and **6** as >98:<2 mixtures of diastereoisomers in 66% and 70% yield, respectively, from the corresponding γ,δ-unsaturated amide.

Discussion

Creation of a pair of complementary *R*- and *S*-stereoselective variant enzymes for the synthesis of the sialic acid mimetics from the essentially *non*stereoselective E192N NAL was accomplished (Figure 4) using a combination of structure-guided mutagenesis and directed evolution. Specifically, error-prone PCR was used to identify residues responsible for controlling the facial selectivity of attack of the intermediate enamine (**3**) on the substrate aldehyde (Figure 2) and hence for determining the stereochemical course of the reaction. The most important residues for stereocontrol were Ala-10, Ser-208, and two crucial threonine residues, Thr-48 and Thr-167. Both Ala-10 and Thr-48 lie close to the carboxylate of pyruvate in the structure of the enzyme with Thr-48 forming part of the primary group of conserved residues in the NAL superfamily.[38] The pyruvate carboxylate in turn makes a hydrogen bond with a tyrosine residue, Tyr-136, which is believed to mediate proton transfer between the substrate carboxylate and the incoming aldehyde carbonyl group in a substrate-assisted catalysis model of the enzyme reaction mechanism.[36,38] A recent protein simulation study of the *E. coli* NAL in complex with the enamine and aldehyde substrates concluded that Ser-208 would form a hydrogen bond to the C4-OH only in the configuration leading to the *R*-configured product.[36] It is therefore likely that the substitutions of Ala-10 and Thr-48 described here interfere with this hydrogen-bonding network to control the stereochemical sense of carbon−carbon bond formation.

Thr-167 had not previously been identified as important in determining the stereochemistry of the enzyme reaction. Interestingly, however, substitution of Thr-167 with glycine produced the most *S*-selective enzyme, whereas substitution with valine contributed critically in the construction of the most *R*-selective enzyme. Thr-167 lies 4.7 Å from O4 of the substrate, and it is interesting to speculate whether steric clashes with the substituted valine, compared with the increased space in the glycine variant, account for the changes in selectivity. Substitution of a very limited set of amino acids in the active site of NAL suggests a modular arrangement in which some residues

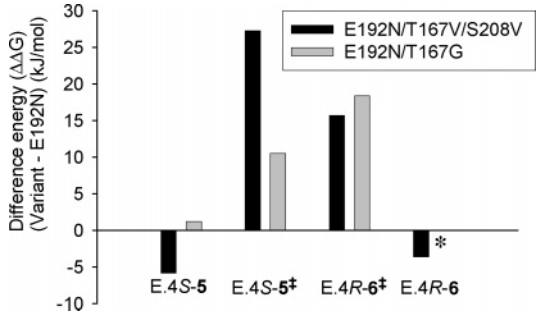

Figure 8. Difference energy diagram for the reactions catalyzed by E192N and the two stereoselective NAL enzymes. The free energies of the binding of the two diastereoisomers (**5** and **6**) to the E192N, E192N/T167G, and E192N/T167V/S208V variants were calculated from the K_m values measured for these enzymes, assuming that K_m provides an indication of the binding affinity between the enzyme and substrate. The activation energies of k_{cat}/K_m were also calculated to provide the heights of the transition state barriers E·**5**‡ and E·**6**‡ (see Figure 3), and the difference energy ($\Delta\Delta G$) was calculated (energy of variant minus energy of E192N 'parent'). The relevant intermediates are E·**5**, the enzyme·4*S*-dipropylamide complex; E·**5**‡, the transition state for formation of the 4*S*-dipropylamide from pyruvate and aldehyde; E·**6**‡, the transition state for formation of the 4*R*-dipropylamide from pyruvate and aldehyde; and E·**6**, the enzyme·4*R*-dipropylamide complex. The asterisk indicates that no difference energy could be calculated for the E·**6** complex with the E192N/T167G variant because we were unable to estimate K_m for this enzyme and substrate, and the height of the E·**6**‡ transition state for the same variant is a lowest estimate based on the background rate of change of absorbance in the screening assay.

(particularly E192) are mainly responsible for discriminating between substrates and other residues (Ala-10, Ser-208, Thr-48, and Thr-167) play a major role in determining the stereochemistry of the reaction.

The progression of the evolution of the two new activities was followed by measuring the reaction free energy profiles for the cleavage of the diastereomeric screening substrates **5** and **6** relative, for convenience, to the uncomplexed enzymes.[48] A difference energy diagram (energy of the variant minus energy of the E192N parent) (Figure 8) was then used to understand the thermodynamics of the changes in enzyme reaction stereoselectivity for the condensation of the aldehyde **7** with pyruvate **2**. Variants E192N/T167G and E192/T167V/S208V, which were selected for changes in stereochemical course of the reaction, show the largest $\Delta\Delta G$ values for the transition states E·**5**‡ and E·**6**‡, while only small changes are observed for the binding free energies of E·**5** and E·**6**. As expected, therefore, the changes in reaction stereochemistry are brought about by the relative changes in the free energies of the transition-state barriers. However, none of the variants selected showed any decreases in the free energy of these barriers, even for the favored reaction pathway. When Thr-167 is substituted by glycine, $\Delta G_R^{\ddagger}$ is increased by at least 18.3 kJ mol^{-1} (since no reaction could be detected with the *R*-configured screening substrate **6**, only a lower limit can be placed on this energy barrier), whereas $\Delta G_S^{\ddagger}$ increases by only 10.5 kJ mol^{-1} and the resulting enzyme is highly *S*-selective. The *R*-selective enzyme E192N/T167V/S208G also shows increases in both $\Delta G_S^{\ddagger}$ and $\Delta G_R^{\ddagger}$; however, the smaller increase in $\Delta G_R^{\ddagger}$ results in a highly *R*-selective enzyme.

The directed evolutionary approach yielded active and useful biocatalysts for the stereochemically controlled synthesis of either C4-diastereoisomer of mimetics of sialic acid. The

(48) Fersht, A. R.; Matouschek, A.; Serrano, L. *J. Mol. Biol.* **1992**, *224*, 771−782.

approach is an interesting alternative to the use of aldolase catalytic antibodies, a few of which have catalytic proficiencies which compare well with those of enzymes.[26] Antibodies which catalyze aldol reactions with alternative stereochemical courses have been identified and exploited in synthetically useful aldol and retro-aldol (kinetic resolution) reactions.

The lower activity of the selected variants, relative to the 'parental' E192N variant, is perhaps a consequence of the evolutionary strategy adopted. The selection pressure applied allowed the creation of enzymes with excellent stereochemical discrimination but not necessarily high activity. The result fits well with the paradigm of directed evolution "You get what you screen for".[49] It is also possible that a similar course was followed in the natural evolution of stereoselective enzymes where the creation of a new enzyme selective for a *required* diastereoisomer would give cells containing the protein a selective advantage, even if the enzyme was of low activity. After this initial appearance of the enzyme, further selection pressure *in vivo*, for example, to increase enzyme stability, to generate high levels of activity, and for appropriate metabolic control, would then tailor the enzyme into the highly active, stereoselective enzymes found today. Our directed evolution experiments on NAL may have mimicked the first of these possible stages in natural evolution of highly stereoselective enzymes.

Acknowledgment. This work was supported by the Biotechnology and Biological Sciences Research Council, the Engineering and Physical Sciences Research Council, and the Wellcome Trust and is a contribution from the Astbury Centre for Structural Molecular Biology.

JA065233Q

(49) You, L.; Arnold, F. H. *Protein Eng* **1996**, *9*, 77−83.