J|A|C|S
COMMUNICATIONS

Published on Web 08/26/2003

Creation of a Productive, Highly Enantioselective Nitrilase through Gene Site Saturation Mutagenesis (GSSM)

Grace DeSantis,* Kelvin Wong, Bob Farwell, Kelly Chatman, Zoulin Zhu, Geoff Tomlinson, Hongjun Huang, Xuqiu Tan, Lisa Bibbs, Pei Chen, Keith Kretz, and Mark J. Burk*

Diversa Corporation, 4955 Directors Place, San Diego, California 92121

Received April 22, 2003; E-mail: gdesantis@diversa.com; mburk@diversa.com

Despite the successful development of biocatalysts for a variety of important transformations, enzymes historically have suffered from numerous shortcomings that have limited their utility in chemical, agrochemical, and pharmaceutical applications.[1,2] Attributes such as low specific activity, inadequate substrate scope, low or undesired enantiomer specificity, intolerance to organic solvents, and low volumetric productivity are challenges often encountered when implementing biocatalytic processes. By screening genomic libraries prepared from environmental samples collected around the globe, we recently discovered over 200 unique nitrilases that allow mild and selective hydrolysis of organonitriles to carboxylic acid derivatives.[3] Several enzymes were shown to be effective for enantioselective hydrolysis of hydroxynitriles to produce α- and β-hydroxy acids on small reaction scale. Of particular commercial interest was the nitrilase-catalyzed desymmetrization of 3-hydroxyglutaryl nitrile (**1**) to afford (*R*)-4-cyano-3-hydroxybutyric acid (**2**), the ethyl ester of which is an intermediate to the cholesterol-lowering drug Lipitor[3] (Scheme 1).

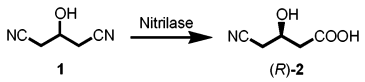
Unfortunately, efforts to develop a larger-scale process for economic production of this intermediate were plagued by lower enantiomeric excess (ee) as the substrate concentration was increased. We now report the application of a novel directed-evolution technique, the gene site saturation mutagenesis (GSSM)[4] method, to address this challenge. This led to a single amino acid nitrilase variant that has ideal properties for the synthesis of this important pharmaceutical intermediate, (*R*)-4-cyano-3-hydroxybutyric acid.

The most effective nitrilase identified in discovery efforts catalyzed the complete conversion of **1** to **2** in 24 h with a product ee of 95% at 100 mM substrate concentration.[3] However, attempts to operate this transformation under more practical conditions revealed a steady decrease in enantioselectivity with increasing concentration; at [**1**] = 0.5, 1, 2, and 3 M, enantiomeric excesses were 92.1, 90.7, 89.2, and 87.6%, respectively, determined at reaction completion.

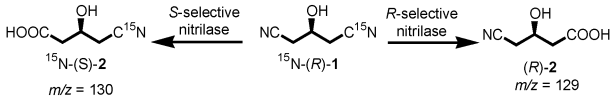
To develop a practical process, the overall objective was to identify a nitrilase that would catalyze the hydrolysis of **1** to (*R*)-**2** with high enantioselectivities at substrate concentrations up to 3 M. Enzymatic reactions rarely operate well at such high concentrations of organic substrate (33 wt %) due to substrate inhibition, product inhibition, enzyme instability, and lower activity and selectivity.[1,2] Our desire to develop a viable process for production of **2** which yields high enantiomeric excess at high volumetric productivities led us to employ our GSSM technology for directed evolution of a nitrilase enzyme.

The GSSM technology is a unique method for rapid laboratory evolution of proteins whereby each amino acid of a protein is replaced with each of the other 19 naturally occurring amino acids. This is accomplished at the genetic level through the use of degenerate primer sets, comprising either 32 or 64 codon variants,

Scheme 1. Synthesis (*R*)-4-Cyano-3-hydroxybutyric Acid



Scheme 2. High-Throughput Screen for Improved Selectivity



for each amino acid residue of the WT enzyme. Subsequent use of standard methods for DNA replication generates a library of genes possessing all codon variations required for saturation mutagenesis of the original gene.[4] Transformation and expression of this collection of genes in a host organism, such as *Escherichia coli*, furnishes a comprehensive library of single-site enzyme mutants.[5] The GSSM technique was employed to create a comprehensive library from our best wild-type nitrilase, which is 330 amino acids in length. The nitrilase GSSM library accesses every single amino acid variant and contained 10,528 genetic variants. It was screened with three times oversampling to ensure complete coverage of the library.[6] Hence, 31,584 clones were screened in an attempt to identify mutant enzymes with improved enantioselectivity.

Success in identification of the requisite phenotype in gene evolution experiments relies on the development of methods to rapidly evaluate libraries containing large numbers of mutants.[7] To efficiently probe our GSSM library of nitrilases for improved enantioselectivity, we developed a novel high-throughput screen based on mass spectrometric detection of isotopically differentiated products (Scheme 2).[8] When developing an assay it is imperative that the surrogate substrate be chemically similar to the actual substrate to ensure that any variants identified will perform the desired transformation effectively. Since substrate **1** is symmetric, a method was required to distinguish the two nitrile groups and detect the site of hydrolysis. Accordingly, we synthesized the ^{15}N-labeled analogue ^{15}N-(*R*)-**1**. Use of the chiral ^{15}N-(*R*)-**1** substrate is important since the labeled atom is removed from the product upon hydrolysis with a (*R*)-selective enzyme, whereas the label remains when hydrolyzed with a (*S*)-selective enzyme. The hydrolysis products differ by one mass unit and can be distinguished and quantified by mass spectrometry. In this assay the screening substrate and target substrate differ only by an isotope label, therefore ensuring that the identified nitrilase will perform equally well on the native unlabeled substrate. For the WT enzyme, it was shown that the observed ee is the same regardless of whether ^{15}N-(*R*)-**1** or ^{15}N-(*S*)-**1** is used.

Screening the nitrilase GSSM library by the method outlined in Scheme 2 was performed using 10 mM ^{15}N-(*R*)-**1** substrate in 100 mM phosphate buffer at 37 °C. Amino acid changes at 17 different residues led to enhanced enantioselectivity over the wild-type

10.1021/ja035742h CCC: $25.00

Table 1. GSSM Mutants with Enhanced Enantioselectivity

nitrilase	[HGN] 100 mM,[a] % ee	[HGN] 2.25 M,[b] % ee	time, h
WT	94.5 ± 0.1	87.8 ± 0.2	24
Ala55Gly	96.5 ± 0.4	nd	>160
Ile60Glu	96.5 ± 0.5	nd	>160
Ala190His	97.9 ± 0.1	98.1 ± 0.1	15
Ala190Ser	96.8 ± 0.2	95.5 ± 0.7	40
Ala190Thr	96.5 ± 0.2	96.6 ± 0.4	40
Asn111Ser	95.8 ± 0.5	96.1 ± 0.9	>160
Phe191Leu	97.9 ± 0.1	nd[c]	>160
Phe191Thr	97.9 ± 0.1	nd	>160
Phe191Met	97.9 ± 0.1	nd	>160
Phe191Val	97.9 ± 0.1	nd	>160
Met199Glu	97.9 ± 0.1	nd	>160
Met199Leu	97.9 ± 0.1	95.4 ± 0.1	160

[a] 100 mM reactions were performed with nitrilase expressed from *E. coli* in whole cell format and were complete within 36 h. [b] 2.25 M reactions were performed with nitrilase as lyophilized clarified cell lysate. All % ee data were determined at the indicated time and are reported as the average of three measurements with standard deviation of the mean. Time for reaction completion established by TLC. [c] nd = not determined.

enzyme. Secondary characterization of primary hits was performed by GC analysis using 100 mM of unlabeled substrate **1**. These studies showed that residues Ala190 and Phe191 were enantioselectivity "hot spots" with several mutants affording product **2** with higher ee (Table 1). Interestingly, in all identified up-mutations at position 190, alanine was replaced with a residue that has hydrogen-bonding potential.

Since the overall objective was to identify a nitrilase that would catalyze the hydrolysis of **1** to (*R*)-**2** with high enantioselectivity at >1 M substrate concentration, each of these up-mutant enzymes was evaluated at 2.25 M substrate concentration. Many of the variants did not perform well at this higher substrate loading, and reactions were not substantially complete after 6 days. However, for each of the serine, histidine, and threonine 190 variants, significant ee enhancement was observed. The Ala190His mutant is the most selective and most active of the GSSM mutants allowing complete conversion of **1** to (*R*)-**2** in 98% ee within 15 h. This is a dramatic improvement relative to WT, which yields (*R*)-**2** in only 88% ee after 24 h at 2.25 M [HGN].

Mutation of an Ala residue to a His necessarily requires a two-base change in the codon triplet -GCN to CAT/C. This result demonstrates one of the unique advantages offered by the GSSM approach for directed evolution of enzymes which permits unbiased access to all codon variations.[6] Random mutagenesis techniques based on mutagenic chemicals or error-prone PCR would be virtually precluded statistically from mutating two bases in the same codon. Moreover, techniques involving recombination of multiple genes also would not have been suitable for production of the Ala190His variant since the nitrilases available in public databases[9] do not have His in position 190.

With an ideal enzyme in hand, we sought to develop an efficient process for synthesis of the Lipitor intermediate (*R*)-**2** at 3 M substrate concentration. HGN (1 g, 9 mmol) was dissolved in phosphate buffer (2.03 mL), and A190H nitrilase (30 mg) was added. The reaction, was stirred at 20 °C for 15 h to furnish the desired acid (*R*)-**2** (1.1 g, 96%, 98.5% ee) with a 619 g L^{-1} d^{-1} volumetric productivity. This demonstrates that an efficient process achieving both high product enantiomeric excess and volumetric productivity may be accessed using the evolved nitrilase.

The benefits of the GSSM evolution strategy were demonstrated through identification and development of an improved nitrilase that provides practical access to a valuable intermediate for the drug Lipitor. Changing Ala to His provided a 10% increase in the enantiomeric excess at the commercially relevant 3 M substrate reaction concentration. Furthermore, error-prone PCR or gene recombination of this enzyme with known nitrilases would not have furnished the improved A190H variant enzyme. We currently are attempting to solve the structure of this variant as well as the wild-type enzyme to gain an understanding of the molecular interactions responsible the improvements.

Acknowledgment. We are grateful to Diversa Sequencing for sequence confirmation and J. Macomber, J. Verruto, and J. Poland for subcloning and library construction, A. Solbak and N. Cloar for colony picking, T. Todaro and M. Lafferty for robotics support, E. O'Donoghue, A. Flordeliza, and C. Cowden, for protein analysis support, M. Podar, M. Wall, and T. Richardson for bioinformatics support, and D. Robertson, B. Morgan, D. Weiner, J. Chaplin, and J. M Short are gratefully acknowledged for guidance.

Supporting Information Available: Materials and methods, including procedures for the preparation of all reagents, analytical details, library construction, and screening methods. (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Drauz, K., Waldmann, H., Roberts, S. R., Eds.; *Enzyme Catalysis in Organic Synthesis*, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2002. (b) Liese, A.; Seelbach, K.; Wandrey, C. *Industrial Biotransformations*; Wiley-VCH: Weinheim, Germany, 2000.
(2) (a) Schmidt, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt. B. *Nature* **2001**, *409*, 258. (b) Burk, M. J. *Adv. Synth. Catal.* **2001**, *343*, 499.
(3) DeSantis, G.; Zhu, Z.; Greenberg, W. A.; Wong, K.; Chaplin, J.; Hanson, S. R.; Farwell, B.; Nicholson, L. W.; Rand, C. L.; Weiner, D. P.; Robertson, D. E.; Burk, M. J. *J. Am. Chem. Soc.* **2002**, *124*, 9024.
(4) (a) Short, J. M. (Diversa Corporation). U.S. Patent 6,171,820, 2001. (b) Gray, 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*, 607.
(5) Saturation mutagenesis and unbiased access to all 32 (NNK) codons and 20 amino acids was confirmed by sequencing at nitrilase residues 40, 105, 190, 191, and 300.
(6) See Supporting Information for details about construction and screening of the GSSM library.
(7) (a) Wahler, D.; Reymond, J.-L. *Curr. Opin. Chem. Biol.* **2001**, *5*, 152. (b) Reetz, M. T. *Angew. Chem., Int. Ed.* **2001**, *40*, 284. (c) Zhao, H.; Arnold, F. *Curr. Opin. Struct. Biol.* **1997**, *7*, 480.
(8) (a) Reetz, M. T.; Becker, M. H.; Klein, H.-W.; Stockigt, D. *Angew. Chem., Int. Ed.* **1999**, *38*, 1758. (b) Guo, J.; Wu, J.; Siuzdak, G.; Finn, M. G. *Angew. Chem., Int. Ed.* **1999**, *38*, 1755.
(9) Pace, H C.; Brenner, C. *Genome Biol.* **2001**, *2* (1), 1.

JA035742H