Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture

Christopher K. Savile *et al.*

Science **329**, 305 (2010);
DOI: 10.1126/science.1188934

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

counterintuitive. Based on the finding of a much stronger molecule-substrate interaction in the LT phase, we believe that the local atomic order at the interface, although different from molecule to molecule (no long-range order), is considerably higher in the LT (compared to the RT) phase, connected with a shorter distance between substrate and adsorbate atoms. Thus, heating leads to a reduction of the short-range order ($\Delta S > 0$), which is not completely compensated for by the appearance of long-range order ($\Delta S < 0$).

The reduction of the bonding strength is most probably accompanied by an increase of the vertical bond lengths, perhaps induced by the excitation of vertical molecular vibrations (*20*). Consequently, the lateral interaction between the molecules gains more weight and induces the lateral ordering (*16*).

At low temperature, the structural reorganization may be explained by different, energetically favorable adsorption sites of the NTCDA molecules if they are near the substrate, perhaps involving a distortion of the adsorbed molecule. If the molecules thereby differ slightly, such as in their rotational and/or translational alignment (the local bonding geometry), a statistic population of the respective sites would result in the observed loss of long-range order. The occurrence of both translational and rotational disorder upon cooling may occur, according to the existing data, and the order-disorder transition can thus indeed be a real melting process or a transition into a glassy state.

This scenario may bear some analogy to the description of inverse melting in metallic alloys (*4*). There, the phenomenon is possible because the (low-temperature) amorphous phase gains energy against the crystal because of increased chemical short-range order (*4*). The very pronounced anisotropy of the forces involved in the present strongly anisotropic (quasi-two-dimensional) case, where the ratio of the respective energy scales is extremely crucial, requires a different theoretical description.

References and Notes

1. N. Schupper, N. M. Shnerb, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **72**, 046107 (2005).
2. A. L. Greer, *Nature* **404**, 134 (2000).
3. P. M. Tedrow, D. M. Lee, *Phys. Rev.* **181**, 399 (1969).
4. W. Sinkler, C. Michaelsen, R. Bormann, D. Spilsbury, N. Cowlam, *Phys. Rev. B* **55**, 2874 (1997).
5. H. Y. Bai, C. Michaelsen, R. Bormann, *Phys. Rev. B* **56**, R11361 (1997).
6. M. R. Feeney, P. G. Debenedetti, F. H. Stillinger, *J. Chem. Phys.* **119**, 4582 (2003).
7. U. Stahl, D. Gador, A. Soukopp, R. Fink, E. Umbach, *Surf. Sci.* **414**, 423 (1998).
8. L. Kilian, thesis, Universität Würzburg, Würzburg, Germany (2002).
9. D. Gador *et al.*, *J. Electron Spectrosc. Relat. Phenom.* **101–103**, 523 (1999).
10. C. Stadler, S. Hansen, A. Schöll, C. Kumpf, E. Umbach, *N. J. Phys.* **9**, 50 (2007).
11. J. Stöhr, *NEXAFS Spectroscopy*, R. Gomer, Ed. (Spinger Series in Surface Science, Springer, Berlin, 1992), vol. 25.
12. A. Schöll *et al.*, *Phys. Rev. Lett.* **93**, 146406 (2004).
13. Y. Zou *et al.*, *Surf. Sci.* **600**, 1240 (2006).
14. A. Bendounan *et al.*, *Surf. Sci.* **601**, 4013 (2007).
15. A. Schöll, Y. Zou, T. Schmidt, R. Fink, E. Umbach, *J. Phys. Chem. B* **108**, 14741 (2004).
16. L. Kilian *et al.*, *Phys. Rev. Lett.* **100**, 136103 (2008).
17. F. S. Tautz, *Prog. Surf. Sci.* **82**, 479 (2007).
18. J. Ziroff, P. Gold, A. Bendounan, F. Forster, F. Reinert, *Surf. Sci.* **603**, 354 (2009).
19. N. Schupper, N. M. Shnerb, *Phys. Rev. Lett.* **93**, 037202 (2004).
20. O. Skibbe *et al.*, *J. Chem. Phys.* **131**, 024701 (2009).

21. Fruitful discussions with M. Sokolowski and Th. Schmidt are gratefully acknowledged. This work was funded by the German Bundesministerium für Bildung und Forschung (contracts 05 KS1WWA-5 and 05 KS7WE1) and by the Deutsche Forschungsgemeinschaft (contract DFG Re 1469/3-2). Experimental support by the BESSY (Berlin Electron Storage Ring Society for Synchrotron Radiation) staff during beam time is gratefully acknowledged.

Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5989/303/DC1
Methods
SOM Text
Figs. S1 and S2
References

4 March 2010; accepted 8 June 2010
10.1126/science.1189106

Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture

Christopher K. Savile,[1]* Jacob M. Janey,[2]* Emily C. Mundorff,[1] Jeffrey C. Moore,[2] Sarena Tam,[1] William R. Jarvis,[1] Jeffrey C. Colbeck,[1] Anke Krebber,[1] Fred J. Fleitz,[2] Jos Brands,[2] Paul N. Devine,[2] Gjalt W. Huisman,[1] Gregory J. Hughes[2]

Pharmaceutical synthesis can benefit greatly from the selectivity gains associated with enzymatic catalysis. Here, we report an efficient biocatalytic process to replace a recently implemented rhodium-catalyzed asymmetric enamine hydrogenation for the large-scale manufacture of the antidiabetic compound sitagliptin. Starting from an enzyme that had the catalytic machinery to perform the desired chemistry but lacked any activity toward the prositagliptin ketone, we applied a substrate walking, modeling, and mutation approach to create a transaminase with marginal activity for the synthesis of the chiral amine; this variant was then further engineered via directed evolution for practical application in a manufacturing setting. The resultant biocatalysts showed broad applicability toward the synthesis of chiral amines that previously were accessible only via resolution. This work underscores the maturation of biocatalysis to enable efficient, economical, and environmentally benign processes for the manufacture of pharmaceuticals.

Asymmetric hydrogenation technologies are being increasingly used for commercial-scale manufacture of fine chemicals and pharmaceuticals (*1*). Solutions that address remaining shortcomings of such technologies, including the need for high-pressure hydrogen, the use and subsequent removal of precious and toxic transition metals, the oftentimes lengthy process of ligand screening and synthesis, and the generally insufficient stereoselectivity (requiring further upgrading of the product), are still being sought. Even though biocatalysis is rarely constrained by these shortcomings, enzymes may suffer from other limitations, such as low turnover numbers, instability toward the demanding conditions of chemical processes, and postreaction processing issues.

The current synthesis of sitagliptin (*2*, *3*) involves asymmetric hydrogenation of an enamine at high pressure [250 pounds per square inch (psi)] using a rhodium-based chiral catalyst (Fig. 1A) (*4*). Within the manufacturing scheme, the chemistry suffers from inadequate stereoselectivity and a product stream contaminated with rhodium, necessitating additional purification steps at the expense of yield to upgrade both enantiomeric excess (e.e.) and chemical purity. By using a transaminase (*5*–*8*) scaffold and various protein engineering technologies, we have developed a catalyst and process that substantially improve the efficiency of sitagliptin manufacturing (Fig. 1B).

Transaminases have a limited substrate range, most accepting only substrates with a substituent no larger than a methyl group at the position adjacent to the ketone (Fig. 2A) (*9*–*14*). Not surprisingly, screening a variety of commercially available transaminases provided no enzyme with detectable activity for amination of the prositagliptin ketone (Fig. 2B). We therefore applied a combination of in silico design and directed evolution in an effort to confer such activity.

An (*R*)-selective transaminase [ATA-117 (*15*), a homolog of an enzyme from *Arthrobacter sp.*] (*16*) was previously used for *R*-specific transamination of methyl ketones and small cyclic ketones

[1]Codexis, Incorporated, 200 Penobscot Drive, Redwood City, CA 94063, USA. [2]Department of Process Research, Merck Research Laboratories, Merck and Company, Incorporated, Rahway, NJ 07065, USA.

*To whom correspondence should be addressed. E-mail: christopher.savile@codexis.com (C.K.S.); jacob_janey@merck.com (J.M.J.)

(*11*, *12*). To assess the feasibility of developing an enzyme for sitagliptin synthesis, we generated a structural homology model of ATA-117 (*17*) to develop hypotheses for initial library designs. Docking studies using this model suggested that the enzyme would be unable to bind prositagliptin ketone (Fig. 2B) because of steric interference in the small binding pocket and potentially undesired interactions in the large binding pocket (Fig. 2C). By using a substrate walking approach (*18*) with a truncated substrate (Fig. 2D), we first engineered the large binding pocket of the enzyme and then evolved that enzyme for activity toward prositagliptin ketone.

Consistent with the model, ATA-117 was poorly active on the truncated methyl ketone analog (Fig. 2D), giving 4% conversion at 2 g/l substrate loading (table S2). Site saturation libraries of residues lining the large pocket of the active site provided new variants with increased activity toward the methyl ketone analog. The best variant contained a $S^{223} \rightarrow P^{223}$ [S223P (*19*)] mutation and showed an 11-fold activity improvement (Fig. 3 and table S1). On the basis of this improved variant (ATA-117: S223P), we generated a small library of enzyme variants for potential activity on prositagliptin ketone. Analysis of the enzyme model suggested four residues that could potentially interact with the trifluorophenyl group [V69, F122, T283, and A284 (*19*)]. Each of these positions was individually subjected to saturation mutagenesis and also included in a combinatorial library that evaluated several residues at each position on the basis of structural considerations [$V^{69} \rightarrow G^{69}$ and $V^{69} \rightarrow A^{69}$ (V69GA), F122AVLIG, T283GAS, and A284GF; library size of 216 variants]. A variant containing four mutations, three in the small binding pocket and one in the large pocket, provided the first detectable transaminase activity on prositagliptin ketone (Fig. 3 and table S3). No detectable activity was identified in any of the variants from the single amino acid site saturation libraries. Initial activity was accomplished via an F122I, V, or L mutation in combination with V69G or A284G. Docking studies indicated that these mutations may relieve the steric interference in the small binding pocket (Fig. 2E). Enzyme loading of 10 g/l provided 0.7% conversion of 2 g/l of ketone over 24 hours, corresponding to an estimated turnover of 0.1 per day. Screening the same combinatorial library in the ATA-117 context without the S223P large binding pocket mutation did not provide any variant with detectible activity toward prositagliptin ketone. Having attained activity through computer-aided catalyst design, we started evolving an enzyme variant for a practical, large-scale process.

The variant with the highest activity toward prositagliptin ketone from round 1b was chosen as the parent for the second round of evolution, and all the beneficial mutations from both the small-pocket combinatorial library and the large-pocket saturation mutagenesis libraries were combined into a new library. Screening of this library resulted in a variant with 75-fold increased

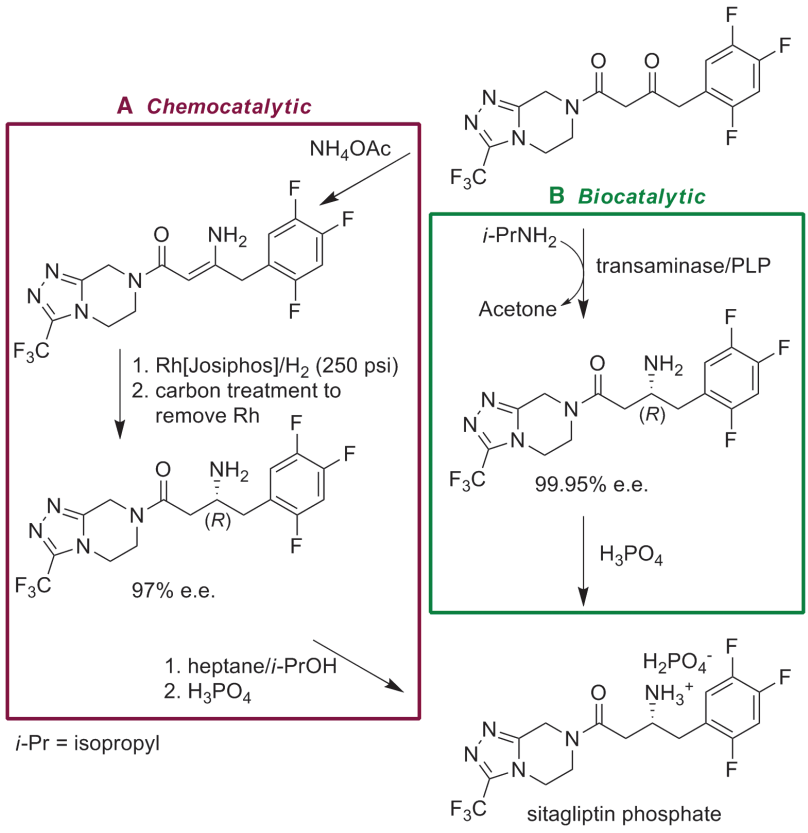


Fig. 1. (**A**) The current synthesis of sitagliptin involves enamine formation followed by asymmetric hydrogenation at high pressure (250 psi) using a rhodium-based chiral catalyst, providing sitagliptin in 97% e.e., with trace amounts of rhodium. Recrystallization to upgrade e.e. followed by phosphate salt formation provides sitagliptin phosphate. (**B**) Our biocatalytic route features direct amination of prositagliptin ketone to provide enantiopure sitagliptin, followed by phosphate salt formation to provide sitagliptin phosphate.

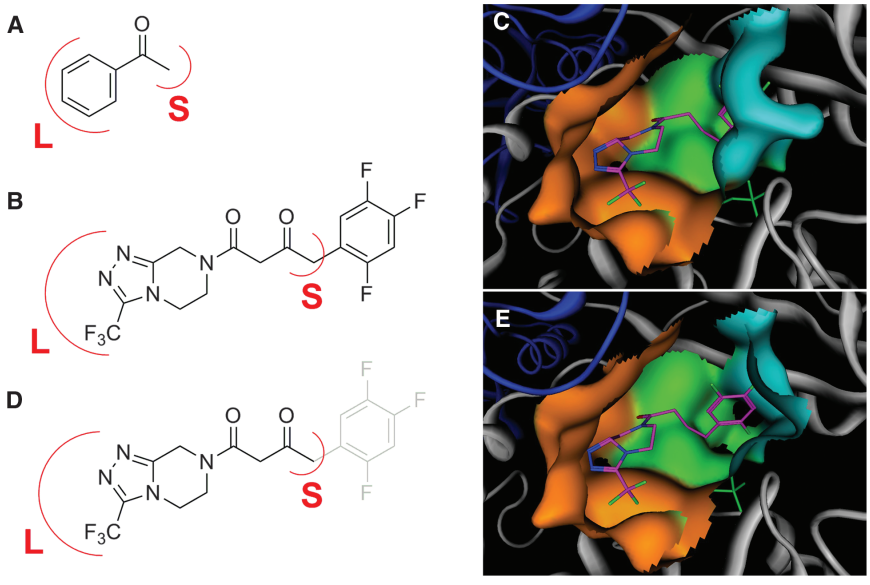
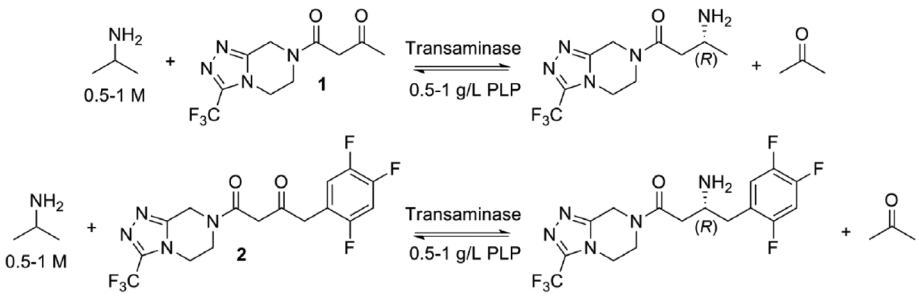


Fig. 2. Previous substrate range studies suggested that the active site of transaminase consists of large (L) and small (S, typically limited to substituents about the size of a methyl group) binding pockets as mapped on the structure of acetophenone (**A**). Accordingly, the structure of prositagliptin ketone (**B**) can be mapped on these binding pockets and docked into the active site of the homology model (**C**). A prositagliptin ketone analog (**D**) was designed to fit the large pocket for initial optimization of this part of the active site. After initial engineering of the large pocket, an enzyme variant was generated with activity on the desired substrate (**E**) by excavating the small pocket (gray/blue, transaminase homology model; orange, large binding pocket; turquoise, small binding pocket; green, PLP and catalytic residues).



Substrate	Added Mutations*	[Substrate] in g/l	Assay changes	Round identified	Improvement over parent†
1	ATA-117	2		–	N/A
1	G136Y	2		1a	6
1	S223P	2		1a	11
2	S223P	2		1a	Not active
2	Y26H;‡ V65A;‡ V69G; F122I; A284G	2		1b	first active
2	H62T; G136Y; E137I; V199I; A209L; T282S	2		2	75
2	S8P; H26Y; G69C; M94I; I137T; G215C	5	5% DMSO to 5% MeOH; RT to 30°C	3	9
2	L61Y; C69T; Y136F; T137E	10	0.5 to 1 M $iPrNH_2$; pH 7.5 to pH 8.5	4	4
2	D81G; I94L; I96L; T178S; L269P; P297S; S321P	40	5 to 10% MeOH; 30 to 45°C	5	1.4
2	Y60F; L94I; A169L; S178T; G217N; L273Y	100	10 to 20% MeOH	6	1.6
2	S124H	100	20% MeOH to 25% DMSO	7	1.1
2	I122M; H124N	100		8	1.1
2	Q329H	100		9	1.9
2	N124T; Y150S; V152C; H329Q	50	25 to 50% DMSO; 0.5% acetone	10	2
2	S126T	50		11	1.4

Fig. 3. Performance of evolvants in HTP screening over the course of evolution. *Mutations accumulated in each round of evolution; reference amino acid refers to parent from previous round of evolution. †Fold improvement over parent refers to the parent for that round of evolution. ‡Mutagenesis experiments showed that these mutations were not critical for activity toward prositagliptin ketone. The top variant from round 1b contained three mutations in the small pocket; however, two are sufficient for achieving detectable activity. Dash entry indicates starting enzyme; RT, room temperature.

activity compared with that of the second-round parent. This variant contained 12 mutations, 10 of which map to the binding pocket model (Fig. 3 and table S5).

Despite this substantial activity enhancement, this catalyst was not yet of practical utility. Reaction conditions in a chemical plant are much different from the conditions that an enzyme encounters in nature, making the development of practical biocatalytic processes a multidimensional challenge (*20*, *21*). The solubility of the prositagliptin ketone in water is low (<1 g/l), necessitating a substantial amount of organic cosolvent to prevent precipitation during reaction. After an evaluation of various cosolvents, methanol and dimethylsulfoxide (DMSO) were identified as optimal with regard to both enzyme performance and ketone solubility (table S7). Solubility can also be enhanced by operating at higher temperatures, prompting our selection of higher boiling DMSO as the preferred solvent. Because transaminase-catalyzed reactions are equilibrium controlled, product accumulation is favored by a large excess of the amine donor isopropylamine (*i*-$PrNH_2$) and/or removal of acetone coproduct. Thus, in order to meet the required performance for practical application, the biochemical characteristics of the transaminase needed to be improved to withstand the harsh conditions of at least 100 g/l (250 mM) prositagliptin ketone, 1 M *i*-$PrNH_2$, >25% DMSO, and a temperature of >40°C for a period of 24 hours, and to give product with >99.9% e.e.

Biocatalyst libraries were generated by using a variety of methods in subsequent iterative rounds of directed evolution (*22*–*25*). Screening of these libraries under processlike conditions provided variants that were increasingly tolerant to the desired process conditions. Specific reaction conditions could not be devised at project inception because no catalyst was available for evaluation of such conditions. Consequently, chemical process development and enzyme optimization were performed in parallel, with process conditions iteratively optimized as improved enzymes became available. Over the course of 11 rounds of evolution, high throughput (HTP) screening conditions were rendered more stringent with the rising activity and tolerance of the biocatalyst. Specifically, we increased the substrate concentration from 2 to 100 g/l, the *i*-$PrNH_2$ concentration from 0.5 to 1 M, the cosolvent from 5 to 50% DMSO, the pH from 7.5 to 8.5, and the temperature from 22° to 45°C, ultimately leading to a catalyst that met the required process targets (table S8). All active transaminase variants tested met the >99.9% e.e. target. Variants generated via this approach were simultaneously optimized for several performance criteria: activity; tolerance to DMSO, acetone, and *i*-$PrNH_2$; stability to the elevated temperature of the reaction medium; and expression in an *Escherichia coli* manufacturing host.

The sequence changes accumulated through this process gave rise to a compounded improvement of four orders of magnitude in activity from the initial prositagliptin transaminase (*26*). The final catalyst contained 27 mutations; of the 17 noncatalytically essential amino acid residues predicted to be interacting with the substrate, 10 were mutated in the final variant. The 27 mutations were obtained by screening various libraries that included diversity identified via structure-aided design of the small binding pocket (4 mutations), site saturation mutagenesis of the large binding pocket (5 mutations), site saturation mutagenesis of the small binding pocket (1 mutation), site saturation mutagenesis of positions outside the binding pockets (2 mutations), homology libraries (10 mutations), and random mutagenesis (5 mutations). The progression of mutagenesis is depicted on the homology model (Fig. 4). After the active site was engineered to accommodate the substrate in the initial rounds, further improvements to generate a productive catalyst came from substantial modification in the dimer interfacial region. The enzyme is presumably active only as a dimer (*27*); we can speculate that these mutations serve to strengthen dimer interactions in the face of increasingly destabilizing reaction conditions (*28*).

Under optimal conditions (*17*), the best variant converted 200 g/l prositagliptin ketone to sitagliptin of >99.95% e.e. (the undesired enantiomer was never detected) by using 6 g/l enzyme in 50% DMSO with a 92% assay yield at the end of reaction (table S11). In comparison with the rhodium-catalyzed process (Fig. 1A), the biocatalytic process provides sitagliptin with a 10 to 13% increase in overall yield, a 53% increase in

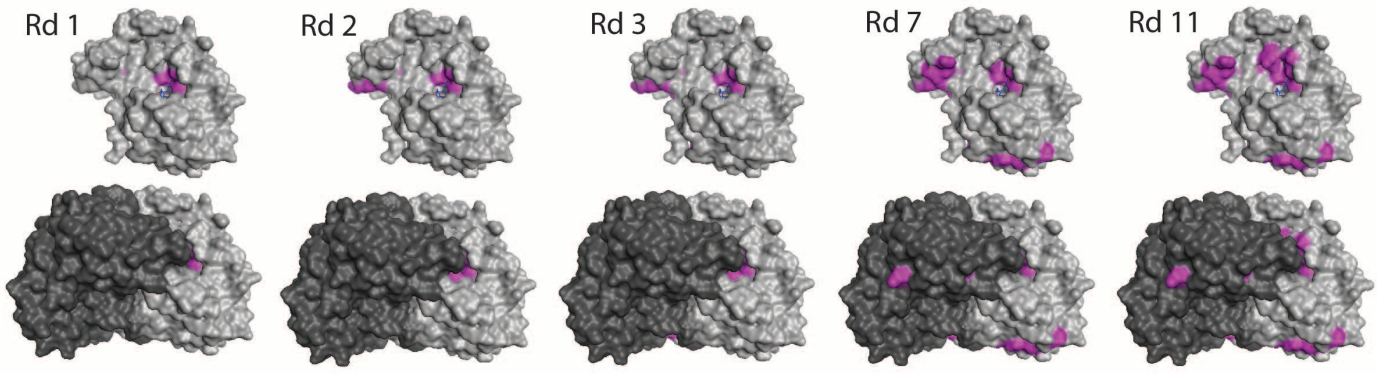
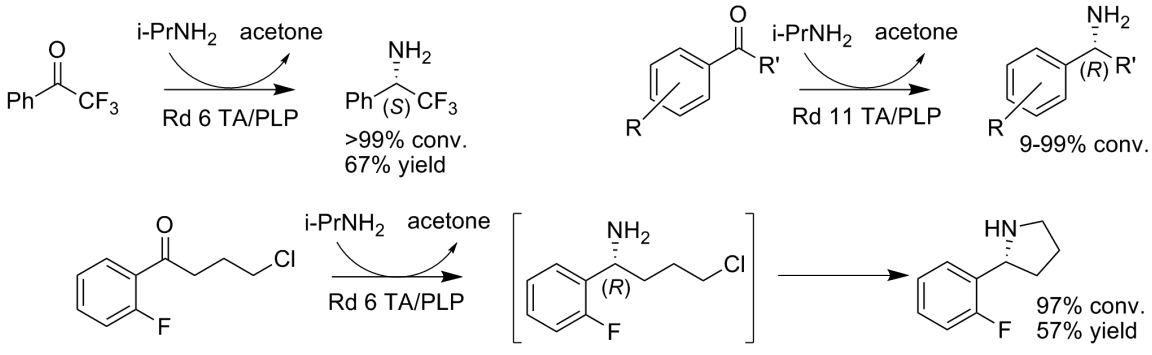


Fig. 4. Structural models of variants from sequential stages of the evolution program. Accumulated mutations are highlighted in purple on the homology model of the transaminase used in this study (size exclusion chromatography data show that the enzyme forms a dimer). The top row shows the mutations mapped on the monomer with the active site of the enzyme exposed and substrate bound. The bottom row shows the mutations mapped on the dimer. The active site, the subunit interface, and a few exposed surface areas can be identified as structural hot spots for mutations.

Fig. 5. Transaminase (TA)–catalyzed amination reactions that were heretofore not feasible either enzymatically or chemically but now proceed with the enzymes described in this work.



productivity (kg/l per day), a 19% reduction in total waste, the elimination of all heavy metals, and a reduction in total manufacturing cost; the enzymatic reaction is run in multipurpose vessels, avoiding the need for specialized high-pressure hydrogenation equipment.

The enzymes developed for sitagliptin synthesis have a broad substrate range and increased tolerance to high concentrations of *i*-PrNH$_2$ and organic solvent that enhances their practical utility. For instance, various trifluoromethyl-substituted amines as well as phenylethylamines with electron-rich substituents (Fig. 5 and fig. S6), which cannot be generated via traditional reductive amination, were prepared with near-perfect stereopurity. The engineered transaminases also enabled efficient access to chiral pyrrolidines. As such, these enzymes provide a general approach for the practical synthesis of chiral amines from prochiral ketones, which is of general interest in pharmaceutical manufacturing (*29*).

Although naturally occurring enzymes rarely offer ideal manufacturing catalysts, directed evolution provides an effective means to overcome this limitation. We have demonstrated that combining modeling with directed evolution offers a rapid means of creating an active enzyme that can operate under the demanding conditions required for the manufacture of pharmaceuticals. In addition, we have removed the intrinsic limitation that transaminases accept only a methyl substituent adjacent to the carbonyl functionality, thereby making substantial progress toward a general approach for the safe, efficient, environmentally friendly production of chiral primary amines. The manufacture of chiral alcohols is already often accomplished with biocatalysts (*30*–*32*); chiral amine synthesis via the enzymatic platform described here is expected to reach a similar level of broad utility and robustness. This development will serve as a model for the implementation of other biocatalytic manufacturing processes in which enzymes can be evolved to meet desired chiral process targets.

References and Notes

1. N. B. Johnson, I. C. Lennon, P. H. Moran, J. A. Ramsden, *Acc. Chem. Res.* **40**, 1291 (2007).
2. D. M. Kendall, R. M. Cuddihy, R. M. Bergenstal, *Eur. J. Intern. Med.* **20** (suppl.), S329 (2009).
3. D. Williams-Herman *et al.*, *BMC Endocr. Disord.* **8**, 14 (2008).
4. K. B. Hansen *et al.*, *J. Am. Chem. Soc.* **131**, 8798 (2009).
5. α-Transaminases are pyridoxal 5′-phosphate (PLP)-dependent enzymes widespread in nature for the synthesis of α-amino acids from the corresponding α-keto acids; comparatively few ω-transaminases, for the synthesis of chiral amines that are not adjacent to carboxylic acid functionalities, are known.
6. S. P. Crump, J. D. Rozzell, in *Biocatalytic Production of Amino Acids and Derivatives*, J. D. Rozzell, F. Wagner, Eds. (Wiley, New York, 1992), pp. 43–58.
7. J.-S. Shin, B.-G. Kim, *J. Org. Chem.* **67**, 2848 (2002).
8. B.-K. Cho *et al.*, *Biotechnol. Bioeng.* **99**, 275 (2008).
9. M. Höhne, S. Kühl, K. Robins, U. T. Bornscheuer, *ChemBioChem* **9**, 363 (2008).
10. M. D. Truppo, J. D. Rozzell, J. C. Moore, N. J. Turner, *Org. Biomol. Chem.* **7**, 395 (2009).
11. M. D. Truppo, N. J. Turner, J. D. Rozzell, *Chem. Commun.* **2009**, 2127 (2009).
12. D. Koszelewski, D. Clay, D. Rozzell, W. Kroutil, *Eur. J. Org. Chem.* **2009**, 2289 (2009).
13. D. Koszelewski *et al.*, *Angew. Chem. Int. Ed.* **47**, 9337 (2008).
14. D. Koszelewski, I. Lavandera, D. Clay, D. Rozzell, W. Kroutil, *Adv. Synth. Catal.* **350**, 2761 (2008).
15. The complete amino acid sequence of ATA-117 is as follows: MAFSADTSEIVYTHDTGLDYITYSDYELDPANPLAGGAA-WIEGAFVPPSEARISIFDQGYLHSDVTYTVFHVWNGNAFRLD-DHIERLFSNAESMRIIPPLTQDEVKEIALELVAKTELREAFVSVSITR-GYSSTPGERDITKHRPQVYMYAVPYQWIVPFDRIRDGVHAM-VAQSVRRTPRSSIDPQVKNFQWGDLIRAVQETHDRG-FEAPLLLDGDGLLAEGSGFNVVVIKDGVVRSPGRAALPGITRKTV-LEIAESLGHEAILADITLAELLDADEVLGCTTAGGVWPFVSVDGN-PISDGVPGPVTQSIIRRYWELNVESSSLLTPVQY.
16. A. Iwasaki, Y. Yamada, N. Kizaki, Y. Ikenaka, J. Hasegawa, *Appl. Microbiol. Biotechnol.* **69**, 499 (2006).
17. Materials and methods are detailed in supporting online material on *Science* Online.
18. Z. Chen, H. Zhao, *J. Mol. Biol.* **348**, 1273 (2005).
19. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
20. G. W. Huisman, J. J. Lalonde, in *Biocatalysis in the Pharmaceutical and Biotechnology Industries*, R. N. Patel, Ed. (CRC, Boca Raton, FL, 2006), pp. 717–742.
21. I. W. Davies, C. J. Welch, *Science* **325**, 701 (2009).
22. R. J. Fox *et al.*, *Nat. Biotechnol.* **25**, 338 (2007).
23. J. H. Zhang, G. Dawes, W. P. Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4504 (1997).
24. K. Stutzman-Engwall *et al.*, *Metab. Eng.* **7**, 27 (2005).
25. J. Liao *et al.*, *BMC Biotechnol.* **7**, 16 (2007).
26. See "quantitation of catalyst improvement," figs. S1 and S2, and tables S10 and S11 in supporting online material (SOM) on *Science* Online.

27. Size exclusion chromatography data show the enzyme forms a dimer (fig. S4).
28. V. G. H. Eijsink *et al.*, *J. Biotechnol.* **113**, 105 (2004).
29. D. J. C. Constable *et al.*, *Green Chem.* **9**, 411 (2007).
30. S. M. A. De Wildeman, T. Sonke, H. E. Schoemaker, O. May, *Acc. Chem. Res.* **40**, 1260 (2007).
31. J. C. Moore, D. J. Pollard, B. Kosjek, P. N. Devine, *Acc. Chem. Res.* **40**, 1412 (2007).
32. G. W. Huisman, J. Liang, A. Krebber, *Curr. Opin. Chem. Biol.* **14**, 122 (2010).
33. We thank C. Ng, D. Standish, M. Mayhew, D. Gray, C. Fletcher, F. Mazzini, P. Lattik, T. Brandon, and J. Postlethwaite for developing the transaminase production system; R. Fox for bioinformatics support; M. Jefferson, R. Trinidad, and M. Whitehorn for assay support; V. Mitchell for molecular biology support; J. Munger and O. Gooding for process development support; P. Fernandez, S. Grosser, and M. Mohan for pilot scale development; K. Morley and B. Grau for additional synthesis support; M. Foster and N. Wu for analytical support; and L. Moore, S. Lato, P. Seufer-Wasserthal, J. Grate, J. Liang, and J. Lalonde for helpful suggestions. Merck and Codexis have filed patent applications on the transaminases, the genes encoding them as well as their use. Materials can be provided under a materials transfer agreement; requests for enzymes should be directed to C.K.S. (christopher.savile@codexis.com), whereas requests for substrates, products, and process information should be directed to J.M.J. (jacob_janey@merck.com).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1188934/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S12
References

1 March 2010; accepted 19 May 2010
Published online 17 June 2010;
10.1126/science.1188934
Include this information when citing this paper.

Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction

Justin B. Siegel,[1,2]* Alexandre Zanghellini,[1,2]*† Helena M. Lovick,[3] Gert Kiss,[4] Abigail R. Lambert,[5] Jennifer L. St.Clair,[1] Jasmine L. Gallaher,[1] Donald Hilvert,[6] Michael H. Gelb,[3] Barry L. Stoddard,[5] Kendall N. Houk,[4] Forrest E. Michael,[3] David Baker[1,2,7]‡

The Diels-Alder reaction is a cornerstone in organic synthesis, forming two carbon-carbon bonds and up to four new stereogenic centers in one step. No naturally occurring enzymes have been shown to catalyze bimolecular Diels-Alder reactions. We describe the de novo computational design and experimental characterization of enzymes catalyzing a bimolecular Diels-Alder reaction with high stereoselectivity and substrate specificity. X-ray crystallography confirms that the structure matches the design for the most active of the enzymes, and binding site substitutions reprogram the substrate specificity. Designed stereoselective catalysts for carbon-carbon bond-forming reactions should be broadly useful in synthetic chemistry.

Intermolecular Diels-Alder reactions are important in organic synthesis (*1*–*3*), and enzyme Diels-Alder catalysts could be invaluable in increasing rates and stereoselectivity. No naturally occurring enzyme has been demonstrated (*4*) to catalyze an intermolecular Diels-Alder reaction (*1*, *2*), although catalytic antibodies have been generated for several Diels-Alder reactions (*3*, *4*). We have previously used the Rosetta computational design methodology to design novel enzymes (*5*, *6*) that catalyze bond-breaking reactions. However, bimolecular bond-forming reactions present a greater challenge, because both substrates must be bound in the proper relative orientation in order to accelerate the reaction and impart stereoselectivity. Also, previous successes with computational enzyme design have involved general acid-base catalysis and covalent catalysis, but the Diels-Alder reaction provides the opportunity to alter the reaction rate by modulation of molecular orbital energies (*7*). To investigate the feasibility of designing intermolecular Diels-Alder enzyme catalysts, we chose to focus on the well-studied model Diels-Alder reaction between 4-carboxybenzyl *trans*-1,3-butadiene-1-carbamate and *N,N*-dimethylacrylamide (Fig. 1, substrates 1 and 2, respectively) (*8*).

The first step in de novo enzyme design is to decide on a catalytic mechanism and an associated ideal active site. For normal-electron-demand Diels-Alder reactions, frontier molecular orbital theory dictates that the interaction of the highest occupied molecular orbital (HOMO) of the diene with the lowest unoccupied molecular orbital (LUMO) of the dienophile is the dominant interaction in the transition state (*7*). Narrowing the energy gap between the HOMO and LUMO will increase the rate of the Diels-Alder reaction. This can be accomplished by positioning a hydrogen bond acceptor to interact with the carbamate NH of the diene (thus raising the energy of the HOMO energy and stabilizing the positive charge accumulating in the transition state), and a hydrogen bond donor to interact with the carbonyl of the dienophile (lowering the LUMO energy and stabilizing the negative charge accumulating in the transition state) (*9*). Quantum mechanical (QM) calculations predict that these hydrogen bonds can stabilize the transition state by up to 4.7 kcal mol^{-1}. (fig. S1). In addition to electronic stabilization, binding of the two substrates in a relative orientation optimal for the reaction is expected to produce a large increase in rate through entropy reduction (*10*). Thus, a protein with a binding pocket (Fig. 1) that positions the two substrates in the proper relative orientation and has appropriately placed hydrogen bond donors and acceptors is expected to be an effective Diels-Alder catalyst.

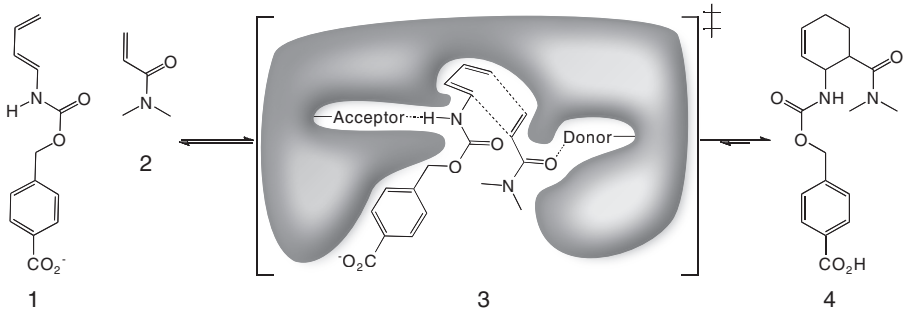


Fig. 1. The Diels-Alder reaction. Diene (1) and dienophile (2) undergo a pericyclic [4 + 2] cycloaddition (3) to form a chiral cyclohexene ring (4). Also shown in (3) is a schematic of the design target active site, with hydrogen bond acceptor and donor groups activating the diene and dienophile and a complementary binding pocket holding the two substrates in an orientation optimal for catalysis.

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [2]Biomolecular Structure and Design Program, University of Washington, Seattle, WA 98195, USA. [3]Department of Chemistry, University of Washington, Seattle, WA 98195, USA. [4]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. [5]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [6]Laboratory of Organic Chemistry, Eidgenössische Technische Hochschule (ETH) Zürich, 8093 Zürich, Switzerland. [7]Howard Hughes Medical Institute (HHMI), University of Washington, Seattle, WA 98195, USA.

*These authors contributed equally to this work.
†Present address: Arzeda Corporation, 2722 Eastlake Avenue East, Suite 150, Seattle, WA 98102, USA.
‡To whom correspondence should be addressed. E-mail: dabaker@u.washington.edu