# Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates

**William A. Greenberg\*, Alexander Varvak, Sarah R. Hanson, Kelvin Wong, Hongjun Huang, Pei Chen, and Mark J. Burk\***

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

A process is reported for efficient, enantioselective production of key intermediates for the common chiral side chain of statin-type cholesterol-lowering drugs such as Lipitor (atorvastatin) and Crestor (rosuvastatin). The process features a one-pot tandem aldol reaction catalyzed by a deoxyribose-5-phosphate aldolase (DERA) to form a 6-carbon intermediate with installation of two stereogenic centers from 2-carbon starting materials. An improvement of almost 400-fold in volumetric productivity relative to the published enzymatic reaction conditions has been achieved, resulting in a commercially attractive process that has been run on up to a 100-g scale in a single batch at a rate of 30.6 g/liter per h. Catalyst load has been improved by 10-fold as well, from 20 to 2.0 wt % DERA. These improvements were achieved by a combination of discovery from environmental DNA of DERAs with improved activity and reaction optimization to overcome substrate inhibition. The two stereogenic centers are set by DERA with enantiomeric excess at >99.9% and diastereomeric excess at 96.6%. In addition, downstream chemical steps have been developed to convert the enzymatic product efficiently to versatile intermediates applicable to preparation of atorvastatin and rosuvastatin.

**B**iocatalysis has received attention recently as a tool of enormous potential for the synthesis of pharmaceutical, industrial, and agricultural chemicals and intermediates (1–5). To capitalize further on this potential, enzymes must demonstrate not only superior performance characteristics such as high enantioselectivity and low catalyst loading but also cost-effectiveness and scalability. A successful industrial biocatalytic process requires enzymes that can be produced cheaply in kilogram quantities and are stable and active at very high substrate and product concentrations, frequently in the presence of organic solvents. Although many remarkable biocatalytic transformations have been reported on the milligram-to-gram scale, especially in light of recent advances in directed evolution and high-throughput screening for novel enzymes (6–8), the number of processes that have been applied on an industrial scale remains limited (5, 9). This is especially true for biocatalytic processes involving formation of carbon–carbon bonds.

The aldol reaction is an extremely useful transformation that allows the formation of new carbon–carbon bonds and the introduction of up to two new stereocenters into the product. Controlling the stereochemistry of aldol reactions can be challenging and has been an area of intense research. Although several elegant asymmetric methods have been developed (10–12), there remains a need for additional methods, and in recent years enzyme-catalyzed approaches using aldolases have received attention as alternatives to chemical methodologies (13).

Several years ago, Wong and coworkers (14, 15) reported unprecedented one-pot tandem aldol reactions catalyzed by a deoxyribose-5-phosphate aldolase (DERA), in which 2 eq of acetaldehyde were added in sequence to 2-carbon aldehyde acceptors to afford six-membered lactol derivatives (Fig. 1). Because the DERA-catalyzed reaction is an equilibrium process, the intermediate 4-carbon adduct (Fig. 1) is reversibly formed

under the reaction conditions. The second condensation between this intermediate and a second equivalent of acetaldehyde drives the equilibrium favorably because of the stability of the cyclized lactol form of the product. The authors noted the structural similarity of the enzymatic products to the lactone moiety (and its interchangeable 3,5-dihydroxy acid form) of the cholesterol-lowering 3-hydroxy3-methylglutaryl (HMG)-CoA reductase inhibitors mevastatin and lovastatin. Today, HMG-CoA reductase inhibitors (collectively known as statins) have worldwide sales of approximately $20 billion, led by Pfizer's Lipitor (atorvastatin), the world's top-selling drug. Atorvastatin and AstraZeneca's newly approved Crestor (rosuvastatin) (Fig. 2) are completely synthetic molecules, unlike the earlier generation of statins that were fungal metabolites or semisynthetic derivatives thereof. The synthetic statins share the chiral 3,5-dihydroxy acid side chain also found in the natural products, which is essential for activity and represents the greatest challenge for preparation of these drugs.

Although the published DERA-catalyzed reaction offered the potential to greatly simplify the process for synthesizing the chiral side chain of statins, several issues would limit its practicability for large-scale production. First, the catalyst load was very high, ≈200 mg of DERA per gram of isolated product (R═Cl), or 20% (wt/wt). Use of such a high concentration of enzyme would make the process prohibitively expensive, in addition to making isolation of product from the reaction mixture difficult. Second, the reaction time was on the order of several days. Third, the concentration of the limiting reagent, chloroacetaldehyde, was only 100 mM. Taken together, the volumetric productivity of the published process was 2 g/liter per day. Hence, in this process, the value of building an advanced statin intermediate, with establishment of both stereogenic centers from achiral 2-carbon starting materials in a single reaction, would be offset on process scale by a very large reaction volume, long reaction time, high catalyst concentration, and laborious product isolation. In addition to these points, the enantioselectivity and diastereoselectivity of the DERA-catalyzed reaction was unknown.

We addressed these limitations by (*i*) discovering an improved DERA by screening genomic libraries prepared from environmental DNA and (*ii*) developing a fed-batch reaction process to overcome significant substrate inhibition. In addition, an inexpensive oxidation method has been developed to convert the enzymatic lactol product into (3*R*,5*S*)-6-chloro-2,4,6-trideoxy-*erythro*-hexonolactone (**1**). The crude lactone was determined to have an enantiomeric excess (ee) of >99.9% and diastereomeric excess (de) of 96.6%. After crystallization, the lactone had an ee

---
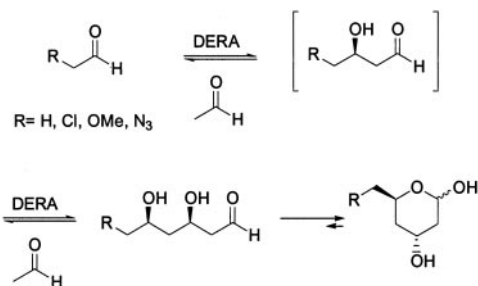
**Fig. 1.** DERA-catalyzed tandem aldol reaction.



>99.9 % ee
99.8% de

**Fig. 3.** Synthesis of lactone (**1**). Reagents and conditions: a, DERA, $H_2O$, room temperature, 3 h; b, NaOCl, HOAc, $H_2O$, room temperature, 3 h.

of >99.9% and de of 99.8%. Compound **1** was functionalized efficiently to provide side-chain intermediates for synthesis of a variety of statins.

## Materials and Methods

**General.** All reagents were purchased and used without additional purification. Reactions were monitored by analytical TLC on silica gel 60 F254 plates and visualized by staining with acidic ceric ammonium molybdate or potassium permanganate. ^{1}H and ^{13}C NMR spectra were obtained on a Bruker (Billerica, MA) 500-MHz UltraShield spectrometer. Fed-batch reactions were performed on a Dasgip (Juelich, Germany) Fedbatch-pro system. GC was performed on an Agilent 6890 gas chromatograph.

**Improved DERA.** The improved DERA was discovered by high-throughput screening of environmental DNA libraries, as described in *Results and Discussion*. The protein sequence of this aldolase is MNIAKMIDHTLLKPEATEQQIVQLCTEAK-QYGFAAVCVNPTWVKTAARELSGTDVRVCTVIGF-PLGATTPETKAFETTNAIENGAREVDMVINIGALK-SGQDELVERDIRAVVEAAAGRALVKVIVETALLTD-EEKVRACQLAVKAGADYVKTSTGFSGGGATVEDV-ALMRKTVGDRAGVKASGGVRDWKTAEAMINAGA-TRIGTSSGVAIVTGGTGRADY.

**Synthesis.** The synthetic procedures described below are illustrated in Figs. 3–6.

**(3R,5S)-6-Chloro-2,4,6-trideoxy-erythro-hexose (2).** Lyophilized crude DERA lysate (1.3 g of powder, 220 mg of pure DERA) was dissolved in 70 ml of water. An aqueous solution containing 1.5 M chloroacetaldehyde and 3.1 M acetaldehyde was fed into the stirred enzyme solution at a rate of 16.66 ml/h for 3 h, resulting in a final volume of 120 ml. Acetone (200 ml) was added to
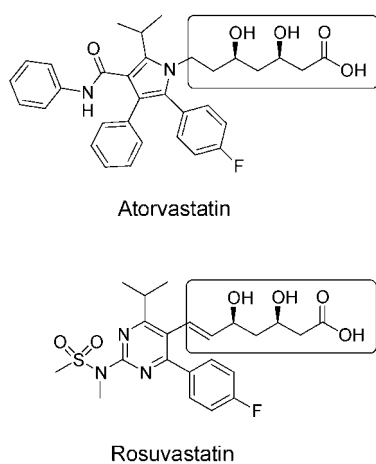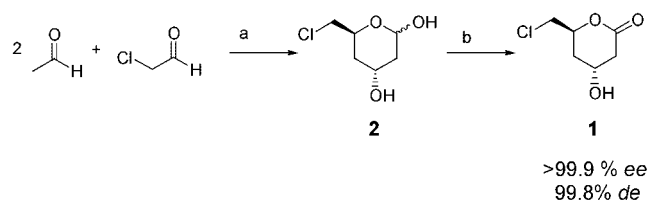
precipitate protein, and the mixture was filtered through Celite. The filtrate was concentrated under reduced pressure to a volume of 50 ml and extracted three times with 75 ml of ethyl acetate. The organic fraction was dried over anhydrous sodium sulfate and concentrated under reduced pressure to provide crude lactol (11.12 g) as a light-yellow oil. The crude product was carried on to the oxidation step without additional purification. **(3R,5S)-6-Chloro-2,4,6-trideoxy-erythro-hexonolactone (1).** Crude lactol (**2**) (75.0 g, ≈0.45 mol) was dissolved in glacial acetic acid (600 ml). Aqueous sodium hypochlorite (265 ml of a 13% solution) was added dropwise to the solution over 3 h. The reaction mixture was concentrated under reduced pressure to a volume of 200 ml, neutralized by addition of solid sodium bicarbonate, and extracted twice with 500 ml of ethyl acetate. The organic fraction was dried over sodium sulfate, concentrated to dryness under reduced pressure, and dissolved in 350 ml of chloroform. The solution was cooled slowly to −20°C, and after 12 h, crystalline product was collected, washed with cold chloroform, and dried under vacuum. Lactone **1** was obtained as white crystals (44.0 g, 0.267 mol, 45% over two steps): ^{1}H NMR (500 MHz) $CDCl_3$ δ 5.01 (1H, m), 4.47 (1H, m), 3.80 (1H, m), 3.68 (1H, m), 2.69 (2H, d, $J$ = 3.6 Hz), 2.09 (1H, m), 1.97 (1H, m). ^{13}C NMR (125 MHz) $CDCl_3$ δ 170.15, 74.81, 62.50, 46.55, 38.54, 32.77.

**(4R,6R)-[6-Cyanomethyl-2,2-dimethyl(1,3)dioxan-4-yl]acetic acid methyl ester (3).** Lactone **1** (10.0 g, 60.7 mmol) was dissolved in 200 ml of dimethylformamide (DMF). Sodium cyanide (8.93 g, 182 mmol) was dissolved in 10 ml of water. The aqueous cyanide solution was added to the DMF solution of chlorolactone. The mixture was stirred at 40°C for 16 h. At the 4-h time point, NMR analysis indicated the presence of chloro diol carboxylate, an epoxide intermediate, and cyano diol carboxylate. After 16 h, the nitrile was the major product. The crude product mixture was partially concentrated under reduced pressure to remove the water (to a final volume of ≈100 ml). It then was acidified carefully to pH ≈ 3 with concentrated sulfuric acid (≈3 ml) and purged for 1 h with nitrogen gas. Dimethoxypropane (21.2 g, 203.3 mmol) was added, and the mixture was stirred at room temperature for 2 h. The solution then was titrated with tri-methylsilyldiazomethane (2 M solution in hexane) until bubbling
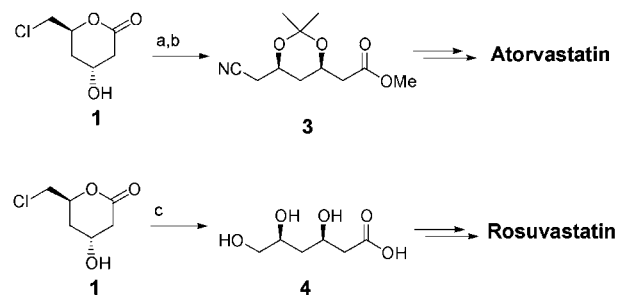


Atorvastatin

Rosuvastatin

**Fig. 2.** Structures of Lipitor (atorvastatin) and Crestor (rosuvastatin).



**Fig. 4.** Reagents and conditions: a, NaCN, DMF, 5% $H_2O$, 40°C, 16 h; b, dimethoxypropane, DMF, catalytic $H_2SO_4$, then trimethylsilyldiazomethane; c, NaOH, $H_2O$, 40°C, 16 h.
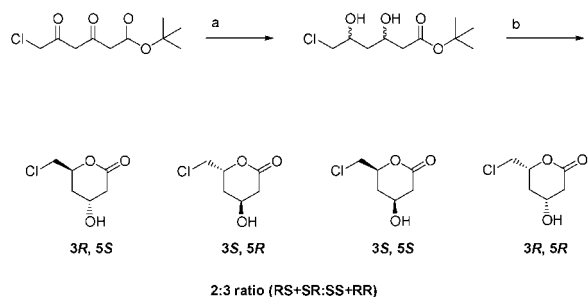
**Fig. 5.** Preparation of chiral standards for analysis of **1**. Reagents and conditions: a, NaBH$_4$, MeOH; b, 15% trifluoroacetic acid, CH$_2$Cl$_2$, 1% triethylsilane.

was no longer observed and the final pH was ≈7. Excess trimethylsilyldiazomethane was quenched with 0.1 ml of acetic acid. Water (200 ml) was added, and the product was extracted with 200 ml of ethyl acetate. The organic layer was back-extracted with 100 ml of water and then dried over sodium sulfate and concentrated under reduced pressure. The residue was purified on silica, eluting with 4:1 hexane/ethyl acetate, yielding 6.52 g of light-yellow oil (48% over 3 steps from chlorolactone **1**). Epoxide intermediate: ^{1}H NMR (sodium salt) (500 MHz) [2]H$_2$O δ 3.87 (1H, m), 2.88 (1H, m), 2.58 (1H, app. t, $J$ = 4.2 Hz), 2.34 (1H, dd, $J_1$ = 3.1 Hz, $J_2$ = 2.8 Hz), 2.09 (2H, d, $J$ = 6.7 Hz), 1.47 (1H, dt, $J_1$ = 9.5 Hz, $J_2$ = 5.0 Hz), 1.39 (1H, m). Compound **3**: ^{1}H NMR (500 MHz) CDCl$_3$ δ 4.35 (1H, m), 4.17 (1H, m), 3.70 (3H, s), 2.59 (1H, dd, $J_1$ = 15.8 Hz, $J_2$ = 6.9 Hz), 2.51 (2H, m), 2.43 (1H, dd, $J_1$ = 15.8 Hz, $J_2$ = 6.1 Hz), 1.78 (1H, dt, $J_1$ = 12.6 Hz, $J_2$ = 2.5 Hz), 1.48 (3H, s), 1.39 (3H, s), 1.33 (1H, dd, $J_1$ = 18.3 Hz, $J_2$ = 11.6 Hz); ^{13}C NMR (125 MHz) CDCl$_3$ δ 171.22, 116.92, 99.76, 65.56, 65.16, 51.97, 41.04, 35.58, 29.90, 25.12, 19.79. **(3R,5S)-3,5,6-Trihydroxyhexanoic acid (4).** Lactone 1 (330 mg, 2 mmol) was dissolved in 5 ml of water. Sodium hydroxide (176 mg, 4.4 mmol) was added, and the mixture was stirred at 40°C for 16 h. The solvent was removed under reduced pressure to provide a white solid, which was dissolved in [2]H$_2$O for NMR analysis: ^{1}H NMR (sodium salt) (500 MHz) [2]H$_2$O δ 4.08 (1H, m), 3.79 (1H, m), 3.50 (1H, m), 3.41 (1H, m), 2.31 (2H, m), 1.52 (1H, m), 1.48 (1H, m); ^{13}C NMR (125 MHz) [2]H$_2$O δ 179.93, 70.41, 67.92, 67.19, 45.80, 41.52.
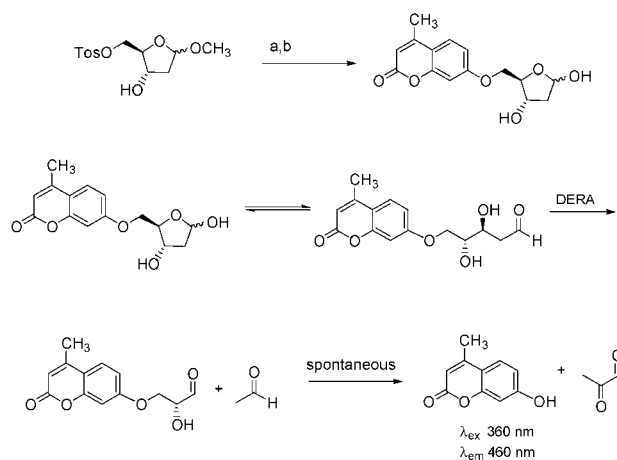


**Fig. 6.** Synthesis and retro-aldol reaction of fluorogenic DERA substrate. a, 4-methylumbelliferone, K$_2$CO$_3$, DMF, 75°C, 16 h; b, H$_2$O/20% CH$_3$CN, Dowex 50WX8-100, 2 days.

**Chiral Analysis.** Enantio- and diastereoselectivity of crude and recrystallized lactone **1** were determined by chiral GC using β-Dex 225 + β-Dex 120 columns in series. A representative GC trace is shown in Fig. 7. A mixture of all four possible diastereomers was prepared as illustrated in Fig. 5 and described below.

6-Chloro-3,5-dioxohexanoic acid tert-butyl ester (16) (2.0 g, 8.5 mmol) was dissolved in 40 ml of methanol at 0°C. Sodium borohydride (750 mg, 20 mmol) was added, and the mixture was stirred for 30 min at 0°C. The solvent was removed under reduced pressure, and the residue was extracted between 75 ml of ethyl acetate and 50 ml of water. The organic layer was dried over anhydrous sodium sulfate and concentrated under reduced pressure. The crude diol product (300 mg) was dissolved in 5 ml of 15% trifluoroacetic acid in dichloromethane with 1% triethylsilane. The mixture was stirred at room temperature for 1.5 h and then concentrated under reduced pressure. The residue was extracted between 30 ml of ethyl acetate and 30 ml of concentrated aqueous sodium bicarbonate. The organic layer was dried over sodium sulfate, concentrated, and purified by silica gel chromatography, eluting with 3:1 ethyl acetate/hexane, with partial separation of diastereomeric pairs of lactones. Fraction A consisted of 52 mg of a 1:1 mixture, and fraction B consisted of 27 mg of a 1:3 mixture in favor of the 3R,5R/3S,5S pair. Spectral data for the 3R,5S/3S,5R racemic mixture: ^{1}H NMR (500 MHz) CDCl$_3$ δ 5.01 (1H, m), 4.47 (1H, m), 3.80 (1H, m), 3.68 (1H, m), 2.69 (2H, d, $J$ = 3.6 Hz), 2.09 (1H, m), 1.97 (1H, m); ^{13}C NMR (125 MHz) CDCl$_3$ δ 170.15, 74.81, 62.50, 46.55, 38.54, 32.77. Spectral data for the 3S,5S/3R,5R racemic mixture: ^{1}H NMR (500 MHz) CDCl$_3$ δ 4.50 (1H, m), 4.32 (1H, m), 3.72 (2H, dd, $J_1$ = 5.1 Hz, $J_2$ = 1.1 Hz), 2.92 (1H, m), 2.53 (1H, m), 2.39 (1H, m), 1.83 (1H, m); ^{13}C NMR (125 MHz) CDCl$_3$ δ 170.01, 76.24, 63.47, 45.81, 39.49, 34.95.

**Fluorogenic DERA Substrate.** The substrate was prepared from known methyl 5-toluenesulfonyl-2-deoxyriboside (17). The toluenesulfonate (12.75 g, 42.2 mmol) was dissolved in 75 ml of DMF. To the solution was added potassium carbonate (11.68 g, 84.5 mmol) and 4-methylumbelliferone (9.29 g, 52.8 mmol). The mixture was stirred at 75°C for 16 h. Water (300 ml) was added, and the mixture was extracted twice with 200 ml of ethyl acetate. The organic layer was back-extracted with 100 ml of 0.1 M aqueous sodium hydroxide, dried over sodium sulfate, and concentrated. The crude product (9.05 g) was dissolved in 25 ml of acetonitrile and 100 ml of water. Dowex 50WX8-100 (2.5 g) was added, and the mixture was stirred at room temperature. After 1.5 h, the mixture was exposed to reduced pressure to remove methanol and then returned to atmospheric pressure. After 2 days, the mixture was filtered, concentrated under reduced pressure, and purified by silica gel chromatography, eluting with a gradient from 100% ethyl acetate to 10% acetone/ethyl acetate. The product was obtained as a white foam (5.31 g, 62%) as a 1:1 mixture of anomers: ^{1}H NMR (500 MHz) DMSO-d$_6$ δ 7.70 (1H, m), 6.97 (2H, m), 6.32 (1H, m, 1α), 6.20 (1H, m, 1β), 6.19 (1H, m), 5.40 (1H, br), 5.17 (1H, br), 4.0–4.3 (4H, m), 2.38 (3H, s), 1.65–2.0 (2H, m); ^{13}C NMR (125 MHz) CDCl$_3$ δ 161.63, 160.12, 154.68, 153.35, 126.43, 113.18, 112.38, 111.17, 101.28, 97.86, 97.05, 82.93, 80.89, 70.93, 70.69, 70.56, 68.94, 42.11, 18.09.

## Results and Discussion

**Discovery of an Improved DERA.** To improve the commercial viability of the process, we sought to discover and develop a DERA that was improved over the enzyme from *Escherichia coli* in two specific parameters: tolerance to high substrate concentrations and reduced catalyst load. Thus, we desired an enzyme that would produce more product per liter while at the same time
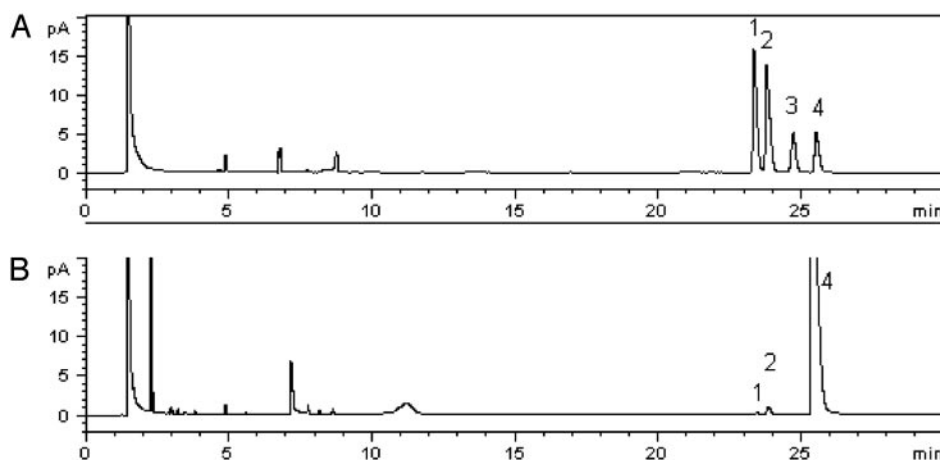
Greenberg *et al.*

**Fig. 7.** GC conditions: columns, $\beta$-Dex 225 (30 m $\times$ 0.25 mm $\times$ 0.25 $\mu$m) plus $\beta$-Dex 120 (30 m $\times$ 0.25 mm $\times$ 0.25 $\mu$m); oven temperature, 180°C; carrier gas He flow rate, 1.0 ml/min. (*A*) Chromatogram of lactone **1** standard with an approximate ratio of (3*R*,5*S*)/(3*S*,5*R*)/(3*R*,5*R*)/(3*S*,5*S*) = 1:1:3:3. (*B*) Chromatogram of actual sample from the chemoenzymatic process with an analysis result of >99.94% ee and 96.60% de using the indicated GC method. Peak 1, (3*R*,5*R*) or (3*S*,5*S*) stereoisomer; peak 2, (3*R*,5*R*) or (3*S*,5*S*) stereoisomer; peak 3, (3*S*,5*R*) stereoisomer; peak 4, (3*R*,5*S*) stereoisomer.

using less enzyme than would be required with *E. coli* DERA. In an effort to access the most diverse range of enzymes that occur in nature, we create large genomic libraries by extracting DNA directly from environmental samples collected around the world from a variety of habitats (18). These environmental genomic libraries are screened through a variety of high-throughput methods including both fluorogenic activity-based screens and sequence-based screens (19). Our fluorescent DERA assay, based on technology developed by Reymond and coworkers (20), uses a fluorogenic substrate analog that undergoes a retro-aldol reaction in the presence of a DERA (Fig. 6). The resultant fragment spontaneously eliminates to produce fluorescent 4-methylumbelliferone. Using these methods, we discovered >15 DERAs that exhibited activity in the title reaction. All these enzymes were overexpressed and compared directly to the *E. coli* DERA in the commercially relevant reaction (Fig. 1). One enzyme was found to be superior, and results for this DERA are described in the following section. Because it was discovered from large, mixed genomic libraries created from environmental DNA, the source organism for the improved DERA is unknown. The protein sequence has <30% identity to *E. coli* DERA and

has closest homology to putative DERAs from Gram-positive organisms.

**Aldolase Reaction.** As noted previously, the reported volumetric productivity levels observed for the reaction between acetaldehyde and chloroacetaldehyde using *E. coli* DERA were exceedingly low (2 g/liter per day). To address this point, we further investigated the cause for limited volumetric productivity as well as the high catalyst load requirement in this transformation. Both the *E. coli* enzyme and our DERA were examined for inhibition, and no effect was observed when acetaldehyde concentration was increased when holding chloroacetaldehyde concentration constant. By contrast, when attempts were made to increase the chloroacetaldehyde concentration, the reaction was halted completely. Inhibition by chloroacetaldehyde, a potent electrophile, was overcome by process improvements that entailed slow feeding of the substrates over 3 h, at a constant ratio of 2:1 acetaldehyde/chloroacetaldehyde, into the reaction at a rate such that they were consumed as fast as they were added. Thus, chloroacetaldehyde did not reach inhibitory concentrations, and the reaction proceeded to completion. The improved
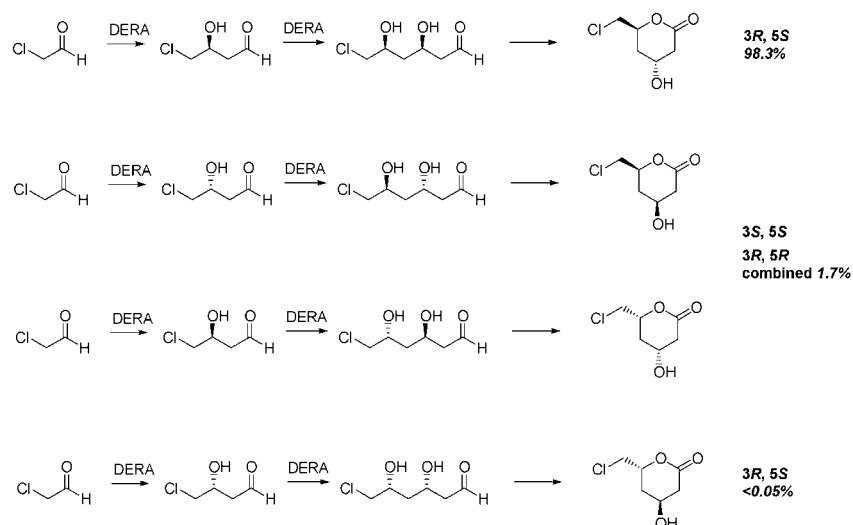


**Fig. 8.** Enantio- and diastereoselectivity of the DERA-catalyzed process.

DERA was expressed on a scale of hundreds of grams, and directly compared (with respect to catalyst load and volumetric productivity) with the *E. coli* enzyme under the same substrate feeding conditions.

We examined lactol productivity levels that could be achieved under the improved process conditions. The *E. coli* DERA performed significantly better, producing a maximum of 76 g/liter (456 mM) product over 3 h, with a minimum catalyst load of 4.8% (wt/wt). When lower concentrations of *E. coli* DERA were used, much lower product yields were achieved. Even greater performance was observed with our improved DERA, which reached 93 g/liter (558 mM) product over 3 h, with a catalyst load of only 2.0% (wt/wt). By discovering a DERA with improved activity and substrate tolerance and converting from a batch reaction to a fed-batch process, we were able to reach ≈7-fold higher product concentrations (>0.5 M) while at the same time using less enzyme (<2 g DERA per liter, or ≈75 μM), compared with the published process. Tolerance to even higher aldehyde concentrations may be engineered into the aldolase by directed evolution. Thus, higher concentrations of product could be achieved in the same time while expending less than half as much enzyme, rendering the process significantly more cost-effective.

**Oxidation to Lactone (1).** The published procedure for oxidation of the enzymatic lactol product to the lactone entailed reaction with bromine and barium carbonate (14, 21). Although this process was effective on a laboratory scale, the expense and toxicity of bromine would prohibit its use in a large-scale process. We identified a very inexpensive method for the selective oxidation to lactone **1** by using aqueous sodium hypochlorite and acetic acid. Lactone **1** is crystalline, providing a readily scalable method for purification by recrystallization. This process yielded a product of exceptional enantiomeric and diastereomeric purity, as described in the following section.

**Enantio- and Diastereoselectivity.** Because the enzymatic product **2** exists as an equilibrium mixture of α and β anomers, chiral analysis of the enzymatic reaction was performed after oxidation to compound **1** to simplify the process. Because two stereogenic centers are set in a single pot, four isomers (two enantiomeric pairs of diastereomers) can be formed. Relative to the desired (3R,5S) product, improper setting of the first, or second, center leads to the diastereomeric (3R,5R) and (3S,5S) isomers, respectively. Improper setting of both centers leads to the enantiomeric (3S,5R) isomer, as illustrated in Fig. 8. Because both centers are set in the same reaction, exquisite enantioselectivity is observed. Each of the two individual centers is set with 98–99% ee, thus the likelihood that the undesired stereochemistry is formed at both centers on the same molecule is vanishingly small. For instance, if the first center is set improperly, it is most likely that the second center will be set properly, leading to the diastereomeric (3S,5S) product. The two diastereomeric products were observed as a combined 1.7% of the total crude lactone concentration, and the enantiomeric (3S,5R) isomer was observed in <0.05%, giving a de of 96.6% and ee of >99.9%. The diastereomeric impurities were separated effectively from the desired product by recrystallization, providing **1** with an ee of >99.9% and de of 99.8%. Diastereoselectivity may be improved further by performing directed evolution on the aldolase.

**Conversion of 1 to Statin Intermediates.** Although the DERA-catalyzed process economically provided highly pure **1** in only two steps, it remained to convert this compound to desired statin intermediates by converting the chloride to the relevant func-
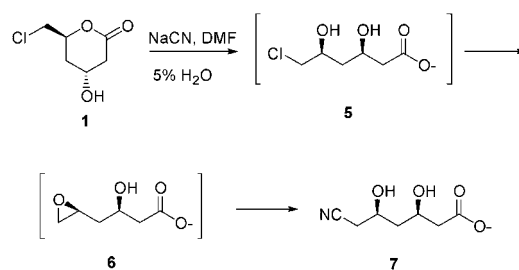


**Fig. 9.** Cyanide displacement through an epoxide intermediate.

tionality (e.g., a nitrile for atorvastatin and hydroxyl group for rosuvastatin). This process entailed opening the lactone ring of **1** with appropriate nucleophiles and introducing suitable protecting groups. For example, triol acid **4** was formed cleanly and quantitatively after treatment of lactone **1** with aqueous NaOH at 45°C over 16 h (Fig. 4). Furthermore, when **1** was stirred at 45°C in wet DMF with 3 eq of NaCN (Fig. 9), the lactone was hydrolyzed within minutes to the chloro diol intermediate **5**. This intermediate was consumed over several hours, and epoxide intermediate **6** concomitantly increased in concentration, as revealed by $^1$H and $^{13}$C NMR spectroscopy. The epoxide subsequently was opened regioselectively by cyanide, with complete conversion in 16 h. At the 4-h time point, the chloro diol, epoxide, and cyano diol were present in approximately equimolar quantities. It should be noted that epoxide **6** also was observed as an intermediate during the basic hydrolysis of lactone **1** to afford the triol acid **4**.

Protection of the cyano diol **7** in a two-step, single-pot process provided intermediate **3** in 48% overall yield from **1**. It should be noted that these are preliminary results, and there is potential for additional optimization of the chemical steps that follow the DERA-catalyzed process. Likewise, it should be possible to further improve the enzyme's specific activity and tolerance of high aldehyde concentrations in the DERA process by directed evolution (8). The complete process to the protected, fully elaborated, enantiomerically pure atorvastatin side chain intermediate **3** entailed only five steps from chloroacetaldehyde and acetaldehyde.

## Conclusions

By a combination of discovery of superior DERAs from environmental DNA libraries and process improvements to overcome substrate inhibition, we developed a highly efficient, cost-effective, scalable, and enantioselective biocatalytic process. The aldolase-catalyzed step proceeded with an ee of >99.9% and de of 96.6%. After oxidation, crystalline lactone **1** was obtained with an ee of >99.9% and de of 99.8%. The volumetric productivity of the enzymatic process, which has been performed on a 100-g scale, was improved to 30.6 g/liter per h (from 0.08 g/liter per h for a published process). Downstream chemical transformations have been developed to convert the enzymatic product to key statin intermediates. The complete process, integrating a key biocatalytic step forming two carbon–carbon bonds stereoselectively with subsequent chemical transformations, represents a general, practical, and economical route to statins.

1. Drauz, K., Waldmann, H. & Roberts, S. R., eds. (2002) *Enzyme Catalysis in Organic Synthesis* (Wiley–VCH, Weinheim, Germany), 2nd Ed.

2. Liese, A., Seelbach, K. & Wandrey, C. (2000) *Industrial Biotransformations* (Wiley–VCH, Weinheim, Germany).

3. Burk, M. J. (2001) *Adv. Synth. Catal.* **343,** 499–500.
4. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M. & Witholt, B. (2001) *Nature* **409,** 258–268.
5. Schoemaker, H. E., Mink, D. & Wubbolts, M. G. (2003) *Science* **299,** 1694–1697.
6. Reetz, M. (2001) *Angew. Chem. Int. Ed. Engl.* **40,** 284–310.
7. 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., *et al.* (2002) *J. Am. Chem. Soc.* **124,** 9024–9025.
8. DeSantis, G., Wong, K., Farwell, B., Chatman, K., Zhu, Z., Tomlinson, G., Huang, H., Tan, X., Bibbs, L., Chen, P., *et al.* (2003) *J. Am. Chem. Soc.* **125,** 11476–11477.
9. Straathof, A., Panke, S. & Schmid, A. (2002) *Curr. Opin. Biotechnol.* **13,** 548–556.
10. Trost, B. M. & Ito, H. (2000) *J. Am. Chem. Soc.* **122,** 12003–12004.
11. Johnson, J. S. & Evans, D. A. (2000) *Acc. Chem. Res.* **33,** 325–335.
12. Machajewski, T. D., Wong, C.-H. & Lerner, R. A. (2000) *Angew. Chem. Int. Ed. Engl.* **39,** 1352–1374.
13. Silvestri, M., DeSantis, G., Mitchell, M. & Wong, C.-H. (2003) *Top. Stereochem.* **23,** 267–342.
14. Gijsen, H. J. M. & Wong, C.-H. (1994) *J. Am. Chem. Soc.* **116,** 8422–8423.
15. Wong, C.-H., Garcia-Junceda, E., Chen, L., Blanco, O., Gijsen, H. J. M. & Steensma, D. H. (1995) *J. Am. Chem. Soc.* **117,** 3333–3339.
16. Wolberg, M., Hummel, W. & Muller, M. (2001) *Chem. Eur. J.* **7,** 4562–4571.
17. Tulshian, D. B., Fundes, A. F. & Czarniecki, M. (1992) *Bioorg. Med. Chem. Lett.* **2,** 515–518.
18. Short, J. M. (1997) *Nat. Biotechnol.* **15,** 1322–1323.
19. Robertson, D. E., Mathur, E. J, Swanson, R. V., Marrs, B. L. & Short, J. M. (1996) *SIM News* **43,** 3.
20. Carlon, R. P., Jourdain, N. & Reymond, J.-L. (2000) *Chem. Eur. J.* **6,** 4154–4162.
21. Morrison, I. M. & Perry, M. B. (1966) *Can. J. Biochem.* **44,** 1115–1126.