

## DNA from Uncultured Organisms as a Source of 2,5-Diketo-D-Gluconic Acid Reductases

WILLIAM H. ESCHENFELDT,<sup>1</sup> LUCY STOLS,<sup>2</sup> HELGA ROSENBAUM,<sup>1</sup> ZUBIN S. KHAMBATTA,<sup>1</sup>  
ELSIE QUAITE-RANDALL,<sup>1</sup> SHAN WU,<sup>3</sup> DONNA C. KILGORE,<sup>4</sup>  
JONATHAN D. TRENT,<sup>1</sup>† AND MARK I. DONNELLY<sup>2\*</sup>

*Environmental Research Division<sup>2</sup> and Biological Sciences Division,<sup>1</sup> Argonne National Laboratory,  
Argonne, Illinois 60439; Genencor International, Inc., Palo Alto, California 94304<sup>3</sup>;  
and Eastman Chemical Company, Kingsport, Tennessee 37662<sup>4</sup>*

Received 26 February 2001/Accepted 27 June 2001

**Total DNA of a population of uncultured organisms was extracted from soil samples, and by using PCR methods, the genes encoding two different 2,5-diketo-D-gluconic acid reductases (DKGRs) were recovered. Degenerate PCR primers based on published sequence information gave internal gene fragments homologous to known DKGRs. Nested primers specific for the internal fragments were combined with random primers to amplify flanking gene fragments from the environmental DNA, and two hypothetical full-length genes were predicted from the combined sequences. Based on these predictions, specific primers were used to amplify the two complete genes in single PCRs. These genes were cloned and expressed in *Escherichia coli*. The purified gene products catalyzed the reduction of 2,5-diketo-D-gluconic acid to 2-keto-L-gulonic acid. Compared to previously described DKGRs isolated from *Corynebacterium* spp., these environmental reductases possessed some valuable properties. Both exhibited greater than 20-fold-higher  $k_{cat}/K_m$  values than those previously determined, primarily as a result of better binding of substrate. The  $K_m$  values for the two new reductases were 57 and 67  $\mu\text{M}$ , versus 2 and 13 mM for the *Corynebacterium* enzymes. Both environmental DKGRs accepted NADH as well as NADPH as a cosubstrate; other DKGRs and most related aldo-keto reductases use only NADPH. In addition, one of the new reductases was more thermostable than known DKGRs.**

Interest in 2,5-diketo-D-gluconic acid reductases (DKGRs) is related to a new biotechnological process for the production of vitamin C (ascorbic acid) from glucose. Conversion of glucose to ascorbic acid is a complicated process that involves selective epimerization, oxidation, and lactonization reactions. The natural biosynthetic pathways are long and incorporate many energy-consuming reactions (8, 21, 38). The present commercial process for ascorbic acid production (the Reichstein process) couples a single biological step—the microbial oxidation of sorbitol to sorbose—with a subsequent, multistep, chemical conversion of blocked derivatives of sorbose to ascorbic acid (7, 26). An alternative commercial process that has been proposed (2, 10, 34) consists of biological conversion of glucose to 2-keto-L-gulonic acid, which is then lactonized chemically to ascorbic acid (Fig. 1). The biological metabolism involved is simpler than that of natural biosynthetic routes and requires less metabolic energy (less ATP and NADPH). In this process, glucose is first converted to 2,5-diketo-D-gluconic acid by endogenous oxidases of a suitable bacterial strain, with molecular oxygen as the ultimate electron acceptor. 2,5-Diketo-D-gluconic acid is then reduced enzymatically to 2-keto-L-gulonic acid by a heterologous DKGR expressed in the production strain. The NADPH required for the reaction is generated by the metabolism of the host strain. Finally, chem-

ical lactonization of 2-keto-L-gulonic acid generates ascorbic acid.

To date, only two DKGRs have been extensively characterized, both isolated from a species of *Corynebacterium* (20, 23, 35). These enzymes, designated DKGR A and B, reduce 2,5-diketo-D-gluconic acid rather inefficiently, exhibiting  $K_m$  values for substrate greater than 1 mM and catalytic efficiencies ( $k_{cat}/K_m$ ) less than 20  $\text{mM}^{-1} \text{s}^{-1}$ . The enzymes are members of the aldo-keto reductase superfamily (14, 30). Like almost all other aldo-keto reductases, the known DKGRs are specific for NADPH (14, 30). Recently, additional aldo-keto reductases that can convert 2,5-diketo-D-gluconic acid to 2-keto-L-gulonic acid have been isolated from *Escherichia coli* through a search of the genome sequence for homologues (40, 41). However, these enzymes also catalyze the reaction relatively inefficiently. The *Corynebacterium* DKGRs also lack stability (23, 35). Consequently, DKGRs with superior enzymatic activity and enhanced stability were sought to facilitate the development of the alternative production processes for ascorbic acid.

Rather than search through culture collections for a known microbe with a superior DKGR, we explored the vast reservoir of uncultured microorganisms. Molecular analyses of microbial communities have revealed that less than 1% of the microbes present are known from culture collections. These undiscovered microbes certainly contain many homologues of known enzymes, some of which may have valuable properties. Present methods for accessing these enzymes rely on construction of libraries of DNA fragments of uncultured organisms and screening of clones for the desired enzymes (12, 33). Here we describe an alternative approach that relies solely on PCR

\* Corresponding author. Mailing address: Argonne National Laboratory, Bldg. 202/Rm. BE111, 9700 South Cass Ave., Argonne, IL 60439. Phone: (630) 252-7432. Fax: (630) 252-7709. E-mail: donnelly@anl.gov.

† Present address: NASA Ames Research Center, Moffett Field, CA 94035.

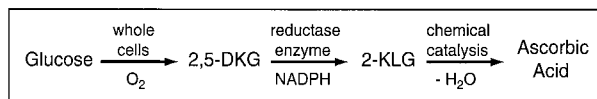


FIG. 1. Alternative process for conversion of glucose to ascorbic acid. An alternative process for conversion of glucose to ascorbic acid effected by whole cells, a recombinant enzyme, and a chemical catalyst. Membrane-bound oxidases convert glucose to 2,5-diketo-D-gluconic acid (2,5-DKG), with oxygen as the electron acceptor. A heterologous 2,5-diketo-D-gluconic acid reductase, in practice obtained from another organism, reduces 2,5-diketo-D-gluconic acid to 2-keto-L-gulonic acid (2-KLG). Known reductases are NADPH dependent. Finally, chemically catalyzed lactonization of 2-keto-L-gulonic acid generates ascorbic acid.

to obtain full-length, functional genes from environmental DNA.

#### MATERIALS AND METHODS

**Extraction and purification of environmental DNA.** Environmental samples were collected in the vicinity of Argonne National Laboratory, Argonne, Illinois. Pond water was collected in plastic carboys, and the suspended matter was concentrated either by flowthrough centrifugation (model A5-16; Sharples) or for small volumes by filtration through 0.22- $\mu$ m-pore-size nitrocellulose filters. The DNA was extracted from the concentrates using the Puregene whole blood and tissue DNA extraction kit (Genra Systems). Soil samples from a deciduous forest and a residential garden were taken from 3 to 6 cm below the surface and extracted essentially as described by Selenska and Klingmüller (31). Two grams (wet weight) of soil were suspended in 4 ml of extraction buffer (120 mM  $\text{Na}_2\text{HPO}_4$ , pH 8.0, containing 1% sodium dodecyl sulfate) in a 50-ml conical tube, shaken at 200 rpm for 1 h at 70°C, and then centrifuged at 3,000  $\times$  g for 5 min at room temperature. The DNA-containing supernatant was collected, and the soil pellet was extracted twice more with 2 ml of extraction buffer. The combined supernatants were centrifuged at 20,000  $\times$  g for 10 min at room temperature to remove residual particles.

Humic substances were removed from soil extracts by size exclusion and ion-exchange chromatography. A mixture of 1.4 ml of the soil DNA extract and 150  $\mu$ l of glycerol was applied to a 1.0-by-20 cm Sepharose CL-4B column (Pharmacia) equilibrated in 10 mM Tris-HCl, pH 7.5, containing 1 mM EDTA and 0.1 M NaCl. The void-volume fractions were pooled, and the DNA was precipitated with ethanol, dissolved in 10 mM Tris-HCl, pH 8.0, containing 0.75 M NaCl, was applied to a Tip 500G column (Qiagen), and was eluted according to the manufacturer's instructions. The final, isopropanol-precipitated DNA was dissolved in 500  $\mu$ l of 10 mM Tris-HCl, pH 8.0, and the DNA concentration was determined by absorbance at 260 nm. DNA samples were stored at -20°C.

**Amplification of complete DKGR genes.** Homologues of known DKGR genes were amplified from environmental DNA samples in three steps. In the first step, internal fragments were generated using degenerate primers based on known DKGR gene sequences (Table 1). Following sequencing of the internal fragments, their flanking regions were amplified in a series of nested PCRs in which one primer was specific for the internal fragment and the other was a random primer that targeted the unknown flanking DNA. For subsequent amplifications, the product of the previous PCR was used as the template and a different, internally nested primer was paired with the same random primer. Once distinct PCR products were generated in good yield (three or four rounds of PCR were required), the products were isolated and sequenced, allowing prediction of the sequence of the full-length gene (Fig. 2). The entire gene was then amplified in a single PCR with primers specific for the flanking regions of the assembled, putative gene using the original environmental DNA as the template.

The primers used are listed in Table 1. The degenerate primers DU1 and DL1 were designed by comparing the amino acid sequences of the two known DKGRs from *Corynebacterium* spp. (GenBank accession nos. M12799 [2] and M21193 [10]) and the related morphine dehydrogenase from *Pseudomonas putida* (GenBank accession no. M94775 [39]). The sequences were aligned by the Clustal method, and degenerate primers were designed targeting regions of identity or strong similarity for at least 7 amino acids. For the second stage of the process, nested primers were designed to amplify DNA flanking two of the internal fragments generated in the first stage, I-14 and I-28 (Table 2). On the basis of comparison of the sequences of the internal fragments, clone-specific primers for

TABLE 1. PCR primer sequences

| Oligo-nucleotide <sup>a</sup> | Sequence                              |
|-------------------------------|---------------------------------------|
| DU1 .....                     | GGCTACCGNCWSMTCGACAC                  |
| DL1 .....                     | GGGTGSAGCTCGAYCTGGTT                  |
| 14nU1 .....                   | CTATGACAATGAGGCAGAGGTC                |
| 14nU2 .....                   | CGCGCGCGAGGAAGTTTTTGTGACA             |
| 14nU3 .....                   | CCGTGCCCCGAAGCAAGACAA                 |
| 14nL1 .....                   | GCTGCAAGAGCTTCGAGATC                  |
| 14nL2 .....                   | AGTTCGAGACTCCGATGCCCTTAAC             |
| 14nL3 .....                   | CGAATGCGTGCCAAGTCTCAA                 |
| 14nL4 .....                   | GACCTCTGCCTCATTGTTCATAG               |
| 28nU1 .....                   | TTATGACAACGAGCCGAGGTT                 |
| 28nU2 .....                   | GCCATTCAAGAGTCGGTTCGACA               |
| 28nU3 .....                   | CGAAACCGGATTGGTGAATCA                 |
| 28nL1 .....                   | ACAACATTCGCAGCCCAAGAA                 |
| 28nL2 .....                   | GAAAGTTTGAGACCAGTGTAT                 |
| 28nL3 .....                   | ACCGATTGATTTACCAATCCG                 |
| 28nL4 .....                   | TTGCCTCGTGGTATCCGTGGCG                |
| 14fU .....                    | GCCGTTTTGCTGTACCTA                    |
| 14fL .....                    | TTTCTTCGTCCAGGGGAGTTTG                |
| 28fU .....                    | TCGCCCCGTGGAGCCAAAAC                  |
| 28fL .....                    | TCGCGCTCTGAATCGTTCTG                  |
| 14expU .....                  | GAGAACAATTGTATGAGCGCAGAACAGCCT        |
| 14expL .....                  | TCTTCTAAGCTTCACTAATTCATATCGTCAGGATT   |
| 28expU .....                  | GAGAACAATTGTATGGCATCGCCGCTGGTT        |
| 28expL .....                  | TCTTCTAAGCTTCACTAATTCATATGTCGTCTGGGTT |

<sup>a</sup> Primers designated D were degenerate primers used to generate internal fragments of the genes. U or L refers to the upper or lower DNA strand. Those with the letter "n" are the nested primers used in generating the flanking regions of fragments 14 and 28. Nested primers ending with the number 1, 2, 3, or 4 were used in the first, second, third, or fourth round, respectively. The set with the letter "f" amplified the full-length genes from environmental DNA, and that with the letters "exp" introduced restriction sites to allow introduction of the genes into an expression vector (Materials and Methods). The locations of the primers in the assembled gene and flanking regions are illustrated in Fig. 2.

I-14 and I-28 (designated 14n and 28n, respectively) were chosen from areas that had the least sequence homology with the other internal fragments. To amplify flanking DNA, a single specific primer was paired with a random restriction-site PCR (RSPCR) primer (29). These primers are of the general structure N<sub>10</sub>GAATTC, where the first 10 positions are completely degenerate and the final 6 specify a restriction site, *Eco*RI in the example. *Nco*I, *Pvu*II, *Xho*I, *Bgl*I, and *Hind*III primers were also used. These primers are not listed in Table 1. For the final stage, amplification of the full-length gene, specific primers (designated 14f and 28f) identical to the flanking regions outside the assembled, putative genes were selected based on comparison of the two assembled sequences. Sequence comparisons were performed using the program MegAlign of the software suite LaserGene (DNASStar). All primers were analyzed for hairpin and duplex formation, melting temperature, and free energy of association with the program Oligo5 (National Biosciences, Inc.), and all primers were synthesized by the HHMI Biopolymer Laboratory and W. M. Keck Foundation Biotechnology Research Laboratory, Yale University.

Optimal conditions for PCR were determined for each stage. For the initial reactions, the degenerate primers were tested using plasmid ptrp1-35a as the template, which contains the *Corynebacterium* DKGR A gene (2). Before amplifying the flanking regions from the environmental DNA, the specificity of the nested primers was confirmed using the I-14 and I-28 fragments as templates; a primer was considered specific if it generated the expected band only with its specific template and an opposing primer that also targeted that template. Unless stated otherwise, PCR mixtures (50  $\mu$ l) contained Mg-free buffer (Promega), a 200  $\mu$ M concentration of each of the four deoxynucleoside triphosphates, 2.5 mM  $\text{MgCl}_2$ , a 2  $\mu$ M concentration of each of the primers, 1.5 U of *Taq* polymerase (Promega), and 25 to 100 ng of environmental DNA as the template. Reactions were initiated at 94°C (1 min), followed by 30 to 40 cycles of 94°C (30 s), 58°C (45 s), and 72°C (1 min), and ended with incubation at 72°C for at least 15 min. For amplification of flanking regions, an initial reaction used one RSPCR primer paired with one of the outermost specific primers, for example,

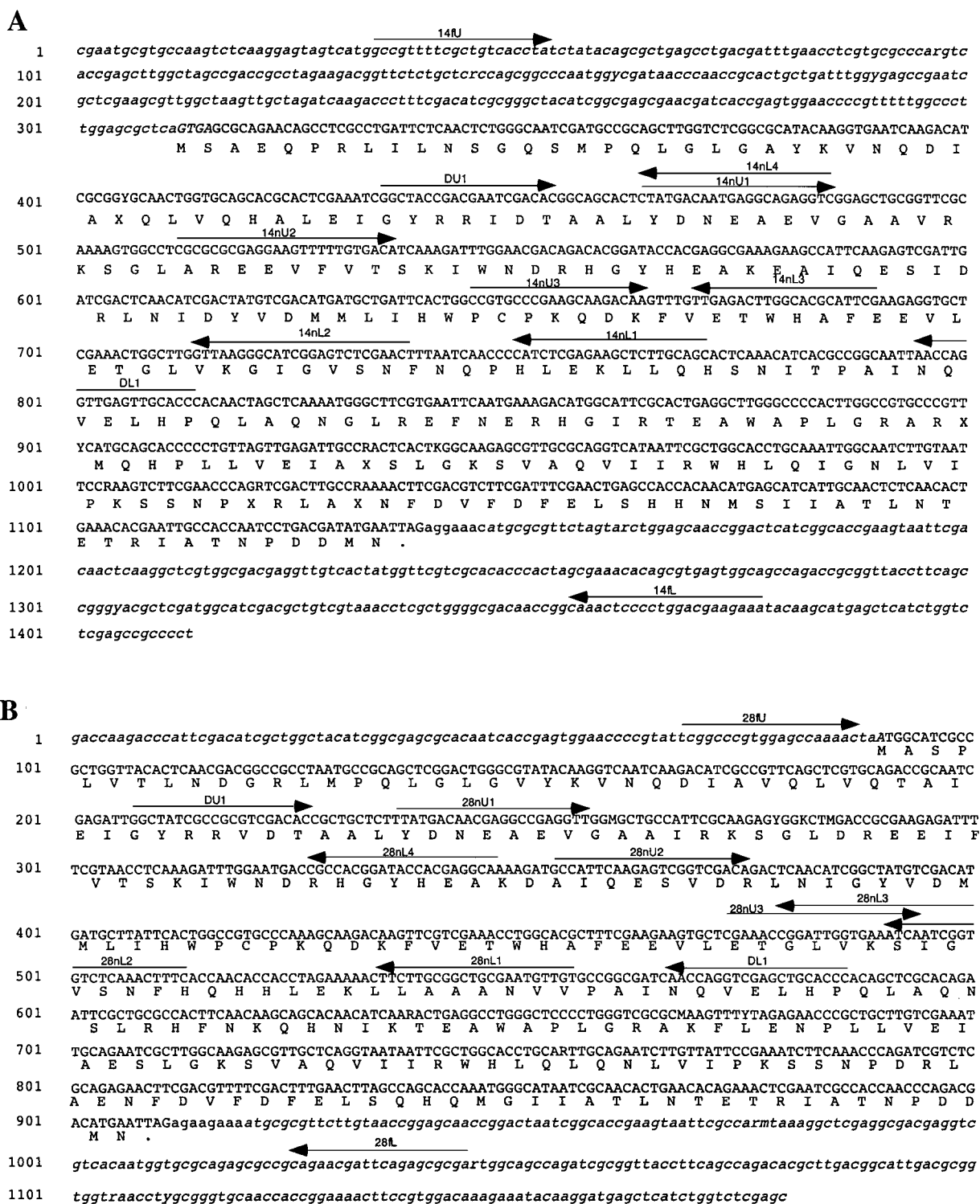


FIG. 2. Assembled putative DKGR genes and flanking regions. The putative DKGR genes, I-14 (A) and I-28 (B), and their flanking sequences were assembled from the three fragments generated by PCR of environmental DNA. The coding sequence is shown in uppercase letters, and below it the predicted amino acid sequence of the DKGR homologue is shown. Flanking sequence is shown in lowercase letters, and flanking ORFs are italicized. The arrows indicate primers used in the PCR amplifications (Table 1). In each case, the regions of overlap span from the degenerate primers DU1 and DL1 to the nearest opposing nested primer.

TABLE 2. Cloned internal gene fragments

| Clone  | Source             | Size (bp) | % Identity <sup>a</sup> |         |
|--------|--------------------|-----------|-------------------------|---------|
|        |                    |           | to DKGR A               | to I-14 |
| I-14   | Residential garden | 340       | 45                      | 100     |
| I-28   | Residential garden | 340       | 44                      | 76      |
| II-4   | Deciduous forest   | 331       | 35                      | 33      |
| III-6  | Woodland pond      | 337       | 32                      | 26      |
| III-19 | Woodland pond      | 370       | 30                      | 25      |
| III-24 | Woodland pond      | 331       | 37                      | 36      |

<sup>a</sup> The percent identity values determined for the nucleotide sequences of the cloned fragments exclude those bases derived from the PCR primers.

14nU1 or 14nL1 in the case of I-14 (Table 1). In these reactions, the primer concentration was 20  $\mu$ M and the annealing temperature was 50°C. For subsequent rounds, the appropriate next inner nested primer was paired with the same RSPCR primer, and 1  $\mu$ l of the PCR from the previous round was used as the template under the same conditions. For amplification of the full-length gene, the standard conditions were modified only in the use of 1.5 mM MgCl<sub>2</sub>. In all cases, products were analyzed by electrophoresis in a 1% agarose gel (28). PCRs were performed using either a GeneAmp PCR system 9600 (Perkin-Elmer) or a Robocycler Gradient 96 thermal cycler (Stratagene).

**Cloning of internal gene fragments and DKGR genes.** The internal gene fragments generated in the initial amplifications were purified by electrophoresis in 1.0% agarose gels in Tris-borate-EDTA buffer (28). The bands of interest were excised, purified with a QiaQuick gel purification kit (Qiagen), and ligated into pBluescript SK+ (Stratagene) that had been treated with *EcoRV* and *Taq* polymerase plus dTTP to introduce a single unpaired T residue at its 3' ends (3). PCR products were ligated into the vector with T4 DNA ligase (Promega) and were transformed into *E. coli* DH5 $\alpha$  (MaxEfficiency; GIBCO/BRL). White transformants that arose on Luria-Bertani agar plates containing ampicillin, isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) (U.S. Biochemicals), and 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-Gal) (28) were analyzed for the presence of inserts of the expected size by PCR with primers specific for the T3 and T7 promoter regions of the vector.

The DKGR homologues were cloned into the expression vector pJF118EH (9) by PCR with the directly amplified full-length genes as the template. Because both genes had an internal *EcoRI* site, the forward expression primers (14expU and 28expU) (Table 1) added a *MunI* restriction site immediately upstream of the initiation codon. For I-14, the forward primer also changed the GTG initiation codon to ATG. The reverse primers for both clones (14expL and 28expL) added a second, in-frame termination codon immediately adjacent to the existing termination codon, along with a *HindIII* restriction site. Standard PCR conditions were used. The products were purified, digested with *MunI* and *HindIII*, and cloned into pJF118EH that had been digested with *EcoRI* and *HindIII* by standard techniques (28) to give plasmids pI-14jf and pI-28jf.

**Analysis of PCR products and genes.** The cloned internal fragments, flanking-region PCR products, full-length PCR products, and cloned DKGR genes were sequenced using the ABI Prism Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer Applied Biosystems) in a Perkin-Elmer GeneAmp PCR system 9600 thermal cycler. All component concentrations and the incubation and cycling conditions followed the manufacturer's instructions. Samples were separated on a 6% acrylamide gel containing 8 M urea in an Applied Biosystems 373A DNA sequencer (Perkin-Elmer Applied Biosystems). Sequences were analyzed with the Seqman program (DNASar). The flanking regions amplified from environmental DNA were not cloned. Rather, their PCR products were sequenced directly using the clone-specific primers.

**Expression and purification of DKGRs.** Cultures of *E. coli* DH5 $\alpha$  containing pI-14jf and of *E. coli* JM109 containing pI-28jf were grown at 37 and 30°C, respectively, in 500 ml of Luria broth (28) in 2-liter, notched Erlenmeyer flasks shaken at 250 rpm. IPTG was added to 1 mM at an optical density at 600 nm of 0.5, and cells were harvested after 4-h (37°C experiment) or overnight growth (30°C) and were then washed once with Tris-EDTA buffer. Cells were resuspended in approximately 2 volumes of 10 mM Tris-HCl (pH 7.5) containing 1 mM EDTA, 0.5 mM dithiothreitol, and 0.001% phenylmethylsulfonyl fluoride and were then lysed by passing the suspension twice through a French press. Cell debris and membranes were removed by centrifugation at 950  $\times$  g, followed by ultracentrifugation at 435,000  $\times$  g in a Beckman TL-100 ultracentrifuge. Both reductases were purified by affinity chromatography (Matrix Red A; Amicon)

followed by ion exchange on a MonoQ column (Pharmacia). All buffers contained 0.5 mM dithiothreitol.

A 2- by 8-cm column of Active Red Matrix equilibrated with 10 mM Tris-HCl, pH 7.2, containing 0.5 mM EDTA was loaded and eluted using a fast protein liquid chromatography system (Pharmacia Biotech). For DKGR C, approximately 5 ml of ultracentrifuged extract was loaded onto the column at a flow rate of 0.5 ml/min. The column was washed with 40 ml of equilibration buffer at a flow rate of 2 ml/min and was then eluted stepwise, first with 40 ml of equilibration buffer containing 1.5 M NaCl, then with buffer containing 2.5 M NaCl. Enzymatic activity eluting in the 2.5 M NaCl wash.

For DKGR D, this procedure was modified as follows. After loading the enzyme, the column was washed with equilibration buffer, as described above. The enzyme was then eluted with a 100-ml linear gradient from 0 to 1.5 M NaCl in equilibration buffer. The enzyme eluted at a NaCl concentration of approximately 0.6 M. In both purifications, the fractions containing DKGR activity were pooled and dialyzed against buffer lacking salt.

The pooled, dialyzed fractions were loaded onto a MonoQ HR 10/10 column by using a Superloop. DKGR C was eluted with a linear gradient (2.5%/min) of 0 to 1.0 M NaCl in 0.1 M Tris-HCl buffer, pH 7.5. Purification of DKGR D required two MonoQ steps, performed at pH 7.5 and 8.0. The enzyme was eluted from the first column with a linear gradient (1%/min) of 0 to 1.0 M NaCl in 0.1 M Tris-HCl buffer, pH 7.5. Fractions containing reductase activity were pooled, dialyzed overnight against 0.1 M Tris-HCl, pH 8.0, and loaded onto the MonoQ column equilibrated with the same buffer. The enzyme was eluted with a linear gradient (1.25%/min) of 0 to 1.0 M NaCl in 0.1 M Tris, pH 8.0, containing 0.5 mM dithiothreitol. In each case, the enzyme was eluted as a sharp peak ( $A_{280}$ ) in the final gradient. Purity was evaluated by denaturing gel electrophoresis (16).

**Enzymatic analyses.** Standard assays were performed in the direction of reduction of 2,5-diketo-D-gluconic acid at 30°C in 1.0 ml of 100 mM Tris-HCl buffer, pH 7.2, containing 0.1 mM NADPH and 1 mM 2,5-diketo-D-gluconic acid (provided by Genencor International). The decrease in absorbance due to the oxidation of NADPH was measured with a Shimadzu UV 160U spectrophotometer. One unit of enzyme is defined as the amount of enzyme that catalyzes the oxidation of 1  $\mu$ mol of NADPH per min. Alternative substrates, obtained from Sigma, were evaluated at 1 mM. For determination of the pH optima, solutions of 100 mM bis-Tris and bis-Tris propane were prepared in increments of 0.5 pH unit from pH 5.5 to 9.0. The enzymes were assayed at each pH level to determine optimal activity.

Kinetic parameters were evaluated in duplicate and were calculated by a least-squares fit of the data to the hyperbola with the curve-fitting algorithm of DeltaGraph (DeltaPoint, Monterey, Calif.) or Prism (GraphPad Software, San Diego, Calif.). Cosubstrates were present at the concentrations described for the standard assay (above). For the determination of the  $K_m$  for NADPH and NADH, assays were performed with a Varian Cary 1G spectrophotometer.

For determination of the parameters for NADH-dependent activity, higher concentrations of both cofactor and substrate were required. Consequently, the initial absorbance at 340 nm was above the linear range of the spectrophotometer, and we measured the change in absorbance at 385 nm. We determined that the extinction coefficient of NADH was 7.74-fold lower at 385 nm than at 340 nm, and the rate data were adjusted accordingly.

Protein concentrations were assayed by the method of Bradford (5) using the protocol and reagent from Bio-Rad Laboratories with bovine serum albumin as the standard.

**Characterization of reaction product.** To a 1-ml solution that contained 10  $\mu$ mol of highly pure 2,5-diketo-D-gluconic acid in 65 mM bis-Tris buffer at pH 7.0, 5  $\mu$ mol of NADPH and approximately 40 U of purified reductase were added. The progress of the reaction was monitored by determination of the concentration of NADPH that remained. When the reaction mixture reached an optical density at 340 nm of less than 2.0, an additional 5  $\mu$ mol of NADPH and 40 U of purified reductase were added and the incubation continued. The product was then evaporated to dryness under reduced pressure and derivatized as follows. Approximately 0.3 mg of dried material was mixed with 0.40 ml of pyridine plus 0.1 ml of solution containing 60 mg of hydroxylamine hydrochloride in 3 ml of pyridine. The sample was heated for 30 min at 75°C in a sealed vial, and then 1 ml of *N,O*-bis-(trimethylsilyl)trifluoroacetylacetamide (BSTFA; Pierce) containing 1% trichloromethylsilane was added slowly while the vial was agitated by hand. The resulting solution was heated for 30 min at 75°C to induce silylation. After cooling, 1- $\mu$ l samples were injected directly into a Finnigan TSO-700 gas chromatograph-mass spectrometer (GC/MS) equipped with an electron impact detector and a 0.25- $\mu$ m-thickness film DB-5 capillary column (0.23-mm inside diameter, 30-m length). The column temperature was ramped from 50 to 290°C at 15°C/min and was then held at 290°C for 14 min. Other parameters were as follows: injector, 280°C; transfer line, 280°C; splitless injec-

tion; valve A open after 0.5 min; multiplier, 1,200 V; and filament, 300 mA. The mass/charge ratio ( $m/z$ ) range of 35 to 650 was scanned in 1-s intervals. The trimethylsilyl oxime derivatives of 2-keto-L-gulonic acid (*syn* and *anti*) were eluted as a doublet at approximately 13 min.

**Thermal inactivation of DKGRs.** The thermal stability of each reductase was evaluated at low protein concentration (0.085 mg/ml) in 100 mM bis-Tris buffer, pH 7.0. The half-life at 45°C was determined by incubating 30- $\mu$ l aliquots of purified enzyme in thin-walled PCR tubes at 4°C. The temperature was shifted rapidly to 45°C by means of a Robocycler Gradient 96 thermal cycler (Stratagene) and was held at 45°C for 0.5, 5, 10, 20, 30, or 60 min before the sample was returned to 4°C. Each tube was assayed later by the standard procedure. The rate constant for loss of activity was determined by fit to the equation for exponential decay using Prism (GraphPad). The midpoint temperature of inactivation of DKGR D was determined by incubating the enzyme for 10 min over a range of temperatures defined by the Robocycler. The Robocycler was programmed to move samples from 4°C to a gradient of defined temperatures ranging from 30 to 52°C in 2°C increments. After 10 min the samples were returned to 4°C. All samples were assayed in duplicate.

## RESULTS

**Amplification of internal gene fragments.** In a control PCR using a plasmid bearing the *Corynebacterium* spp. DKGR A gene as the template, the degenerate primers DU1 and DL1 (Table 1) generated a well-defined band of the expected 380 bp. When various environmental DNA extracts were used as the template, PCR gave broad bands between 350 and 400 bp in size. These bands were excised from the gel and cloned, resulting in a total of six independently derived, putative gene fragment clones that were studied further (Table 2). Sequencing revealed that all six clones were different from one another. A basic local sequence alignment tool, BLASTX (1), search of the GenBank database indicated that all six possessed significant homology with known members of the aldo-keto reductase superfamily. Each was most highly homologous to a gene of unknown function: fragments I-14, I-28, and II-4 to the *yvsB* gene of *Bacillus subtilis*; fragment III-6 to the *E. coli yafB* gene; fragment III-19 to *ytbE* of *B. subtilis*; and fragment III-24 to protein CAA22355 of *Streptomyces coelicolor*. Two of the fragments, I-14 and I-28, were very similar to each other (76% nucleotide sequence identity, excluding the primer sequences). Of the six, these two were the most homologous to the *Corynebacterium* DKGR A gene (GenBank accession no. M12799 [2]), possessing 45 and 44% nucleotide sequence identity, respectively.

**PCR amplification of upstream and downstream regions.** The 5' and 3' flanking sequences for clones pI-14 and pI-28 were obtained by RSPCR (29). Nested, clone-specific primers (Table 1) were paired with several different RSPCR primers. The initial amplification, using environmental DNA as the template, generated a diffuse smear of products with a few faintly discernible bands. Subsequent rounds of PCR used the product of the previous reaction as the template, the same RSPCR primer, and a nested primer. With each round, increasingly discrete products were generated. After three or four rounds, discrete products were formed in good yield. For both the pI-14 and pI-28 clones, an approximately 500-bp fragment of 5' flanking sequence was generated using the *Bgl*I RSPCR primer and nested primers 14nL4 and 28nL4, respectively, and an approximately 800-bp fragment of 3' flanking sequence was generated using the *Xho*I RSPCR primer and nested primers 14nU3 and 28nU3, respectively.

Sequencing of the proximal regions of the flanking PCR

|     |   |         |
|-----|---|---------|
| 1   | M S A E Q P R L L I N S G Q S M P Q L G L G A Y K V N Q D I T V Q L V Q H A L E | pI-14jf |
|     | M A S . L V T . . D . R L . . . . . V . . . . . A . . . . . T . I . .           | pI-28jf |
| 41  | I G Y R R I D T A A L Y D N E A E V G A A V R K S G L A R E E V F V T S K I W N | pI-14jf |
| 39  | . . . . . V . . . . . L . . . . . L . . . . . L . . . . . E . . . . .           | pI-28jf |
| 81  | D R H G Y H E A K E A I Q E S J D R L N I D Y V D M M L I H W P C P K Q D K F V | pI-14jf |
| 79  | . . . . . V . . . . . G . . . . . H . H . . . . . A A A T V V . .               | pI-28jf |
| 121 | E T W H A F E E V L E T G L V K G I G V S N F N Q P H L E K L L Q H S N I T P A | pI-14jf |
| 119 | . . . . . V . . . . . S . . . . . H F . K Q . N . K . . . . . K . L E N         | pI-28jf |
| 161 | I N Q V E L H P Q L A Q N G L R E L N E R H G I R T E A W A P L G R A R F M Q H | pI-14jf |
| 159 | . . . . . V . . . . . S . . . . . H F . K Q . N . K . . . . . K . L E N         | pI-28jf |
| 201 | P L L I E I A E S L G K S V A Q V I I R W H L Q I G N L V I P K S S N P D R L A | pI-14jf |
| 199 | . . . . . V . . . . . S . . . . . H F . K Q . N . K . . . . . K . L E N         | pI-28jf |
| 241 | E N F D V F D F E L S H H N M S I I A T L N T E T R I A T N P D D M N           | pI-14jf |
| 239 | . . . . . Q . Q . G . . . . . T E T R I A T N P D D M N                         | pI-28jf |

FIG. 3. Amino acid sequences of expressed DKGRs. The complete amino acid sequences of the final, expressed genes, obtained from direct amplification of environmental DNA, are given here. Only the residues of DKGR D (pI-28jf) that differ from those of DKGR C (pI-14jf) are shown. The numbering of DKGR D matches that of DKGR A, whose structure has been determined.

products confirmed that their sequences overlapped those of the original clones. Putative nucleotide sequences for the complete genes and flanking regions of I-14 and I-28 were constructed from the overlapping fragments (Fig. 2) (GenBank accession nos. AF385140 and AF385141, respectively). Each putative gene encodes a protein of approximately 32 kDa with significant homology to the *Corynebacterium* DKGR A. The genes are predicted to start for I-14 at the GTG codon at position 312 of the assembled fragments and for I-28 at the ATG codon at position 94. The difference in the positions of the putative start codons results from a poorer yield of unambiguous sequence data for the upstream flanking region of I-28, not from any difference in the organization of the genes.

The putative genes are flanked on both sides by partial open reading frames (ORFs) (Fig. 2). The upstream ORF begins outside the amplified fragments and covers 104 amino acids in the I-14 assembly but only 29 in the I-28 assembly because of the smaller amount of reliable sequence data for that fragment. In both cases, the stop codon of this ORF overlaps the start codon of the putative DKGR gene. The downstream ORF starts just beyond the termination codon of both genes and extends beyond the range of the amplified DNA. BLASTP searches indicated homology of the upstream ORF to the protein ACC74333 (4) of *E. coli* and of the downstream ORF to protein CAB51274 (25) of *S. coelicolor* (downstream ORF). Both proteins are hypothetical, with no assigned function.

**Direct amplification and expression of complete genes.** To establish that the assembled I-14 and I-28 genes are truly present in the environment and are not chimeras of multiple homologous genes, the full-length genes were amplified in a single PCR using the original environmental DNA as the template and primers specific for the flanking regions of the assembled fragments (Table 1; Fig. 2). In each case, a single band of the expected size, 1.34 and 0.98 kb, respectively, was obtained in good yield. Sequencing of these PCR products confirmed their identity as the assembled I-14 and I-28 genes. To allow expression of the genes, restriction sites were introduced adjacent to the start and termination codons by PCR. The primers (Table 1) also changed the GTG start codon of I-14 to ATG and added a second termination codon to each gene. These PCR products were cloned into pJF118EH (9) to give the expression vectors pI-14jf and pI-28jf, respectively.

Sequencing of these two clones revealed an overall amino acid sequence identity of 82.5% for the two proteins (Fig. 3).

TABLE 3. Purification of 2,5-diketo-L-gulonic acid reductases<sup>a</sup>

| Form | Step               | Amt (U) | Sp act (U/mg) | Purification (fold) | Recovery (%) |
|------|--------------------|---------|---------------|---------------------|--------------|
| C    | Extract            | 97,500  | 2.5           | 1                   | 100          |
|      | Matrix Red agarose | 59,900  | 20.4          | 8.0                 | 61.4         |
|      | MonoQ              | 56,920  | 60.2          | 23.7                | 58.4         |
| D    | Extract            | 116,400 | 0.36          | 1                   | 100          |
|      | Matrix Red agarose | 68,928  | 2.1           | 5.8                 | 59.2         |
|      | MonoQ I            | 62,230  | 16.9          | 46.6                | 53.5         |
|      | MonoQ II           | 49,788  | 20.5          | 56.9                | 42.8         |

<sup>a</sup> Enzymes were purified as described in Materials and Methods.

Gene I-14 (GenBank accession no. AF385142) encodes a 275-amino-acid protein of 31,203 Da. Gene I-28 (GenBank accession no. AF385143) encodes a 273-amino-acid protein of 30,881 Da. The DNA and protein sequences of each directly amplified gene differed slightly from those predicted by the assembled gene fragments (Fig. 2); I-14 differed by 4% from the original predicted sequence and clone pI-28 by 1%. Such differences are very likely due to the large number of PCR cycles used to generate the original clones, which could allow errors to accumulate in the assembled sequences. Errors may also be present in the expressed, cloned genes; each was amplified in two rounds of PCR. However, given the high activity and specificity of the expressed enzymes (see below), it is unlikely that errors, if present, impaired the enzymes significantly. A BLASTP search of the GenBank database indicated that both final sequences most closely resemble a putative aldo-keto reductase gene from *S. coelicolor*, CAA22355 (25), with sequence identities of 47 and 48%, respectively, for the I-14 and I-28 genes. Both sequences also are 41 and 42% identical, respectively, to DKGR A of *Corynebacterium* spp.

Expression of the I-14 and I-28 genes generated proteins of an approximate subunit molecular mass of 31 kDa. Both proteins reduced 2,5-diketo-D-gluconic acid rapidly and were designated DKGR C (product of I-14) and DKGR D (product of I-28) to conform to the earlier designation of the *Corynebacterium* enzymes. Extracts of control cells containing only the parent vector, pJF118EH, lacked DKGR activity.

**Purification of reductases.** The two DKGRs were purified by affinity chromatography with Matrix Red agarose followed by ion-exchange chromatography on a MonoQ column (Table 3). DKGR C, which was more highly overexpressed, was purified to homogeneity in two steps. This reductase bound tightly to the Matrix Red agarose and was eluted at relatively high purity with 2.5 M NaCl, dialyzed to remove salt, and then purified to homogeneity on the MonoQ column. The enzyme was eluted as a sharp, symmetrical, well-isolated peak at approximately 0.4 M NaCl.

DKGR D, which was less highly overexpressed and bound less tightly to the Matrix Red agarose and MonoQ resins, was more difficult to purify. This protein was first eluted from the Matrix Red agarose column with a gradient of NaCl rather than by stepwise elution and was then purified to near-homogeneity in two MonoQ steps, the second of which was performed at a different pH with a shallow gradient of salt concentration. The resulting protein was free of major contaminants and was estimated by densitometry to be more than 97% pure. The recovery

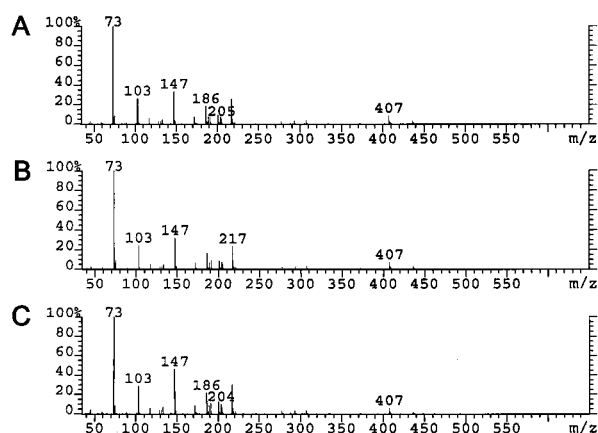


FIG. 4. Mass spectra of product of DKGRs and of authentic 2-KLG. Mass spectra of the trimethylsilyl oxime derivative of the product of the reduction of 2,5-diketo-D-gluconic acid by DKGR C (A) and DKGR D (B) compared to a library spectrum of the derivative of authentic 2-keto-L-gulonic acid (C).

of both reductases was approximately 50%, with the major loss of activity occurring during affinity chromatography (Table 3). The greater degree of purification of DKGR D reflects its expression being poorer than that of DKGR C.

Purified DKGR C and DKGR D had apparent native molecular masses of 31 and 30 kDa, respectively, consistent with the predicted molecular weights, 31,203 and 30,881.

**Determination of product of DKGRs.** Reduction of 2,5-diketo-D-gluconic acid can generate any of four products depending on the carbonyl reduced and the stereochemistry of the resulting alcohol. We established the product of both DKGRs to be 2-keto-L-gulonic acid by high-pressure liquid chromatography and GC/MS. A 10 mM solution of 2,5-diketo-D-gluconic acid was prepared and converted to product by each purified reductase as described in Materials and Methods. High-pressure liquid chromatography analysis revealed that all of the 2,5-diketo-D-gluconic acid had been converted to a compound that coeluted with authentic 2-keto-L-gulonic acid. The reaction mixture was subsequently derivatized to give the trimethylsilyl oximes of the product and was analyzed by GC/MS. The product of both reactions was eluted as a doublet (*syn* and *anti* oximes) at 13.1 min, as was the derivative of authentic 2-keto-L-gulonic acid (data not shown). The mass spectra of the three derivatives were identical (Fig. 4). All other components present in the chromatogram were identified as derivatives of buffer components or components arising from the derivatization reagents (data not shown). No other product attributable to reduction of 2,5-diketo-D-gluconic acid was observed.

**Kinetic analyses of environmental reductases.** The purified reductases were evaluated for their ability to reduce ketosugars other than 2,5-diketo-D-gluconic acid. No reduction of 2-keto-D-gluconic acid, 5-keto-D-gluconic acid, 2-keto-L-gulonic acid, xylulose, xylulose-5-phosphate, or D-fructose was observed. Kinetic parameters of the enzymes with 2,5-diketo-D-gluconic acid as the substrate were determined at 30°C (Table 4). The  $K_m$  values determined for 2,5-diketo-D-gluconic

TABLE 4. Kinetic parameters of purified DKGRs<sup>b</sup>

| Form | Parameters                                      |   |  |                                      |
|------|---|---|--|--------------------------------------|
|      | $K_m$ for DKG <sup>a</sup><br>( $\mu\text{M}$ ) | $k_{\text{cat}}$<br>( $\text{s}^{-1}$ ) | $k_{\text{cat}}/K_m$<br>( $\text{mM}^{-1} \text{s}^{-1}$ ) | $K_m$ for NADPH<br>( $\mu\text{M}$ ) |
| A    | 13,500  | 4                                       | 0.3  | 13                                   |
| B    | 2,000   | 39                                      | 19   | 10                                   |
| C    | 57  | 31                                      | 550  | 3.0                                  |
| D    | 67  | 27                                      | 400  | 2.7                                  |

<sup>a</sup> DKG, 2,5-diketo-D-gluconic acid.

<sup>b</sup> Values for forms A and B are from Sonoyama and Kobayashi (35). For forms C and D, kinetic parameters for 2,5-diketo-D-gluconic acid were determined in the presence of 0.15 mM NADPH, and values for NADPH were determined in the presence of 1 mM 2,5-diketo-D-gluconic acid.

acid were 57 and 67  $\mu\text{M}$  for DKGR C and D, respectively. These values are much lower than the values of 2 mM and 13 mM reported previously for the *Corynebacterium* reductases (35). The observed  $k_{\text{cat}}$  values for both environmental reductases were close to that of the more active *Corynebacterium* enzyme, DKGR B (Table 4). As a result, the  $k_{\text{cat}}/K_m$  values were much higher for the new enzymes. The new DKGRs had catalytic efficiencies more than 1,000 times higher than those of the *Corynebacterium* DKGR A enzyme and 20 times higher than those of DKGR B.

The pH profiles of both reductases revealed a preference for acidic pH, but good activity was observed at all pH values below 7.5 (data not shown). Both enzymes demonstrated maximum activity at pH 6.0, with slightly less activity at pH 5.5. This trend was observed for all buffers evaluated, but activity was highly dependent on the buffer used. Amine buffers such as Tris and bis-Tris gave the best activity. In phosphate and pyrophosphate buffers, both enzymes were approximately one-third as active at pH 6.0. Sulfonate buffers such as morpholineethanesulfonic acid (MES) and HEPES gave intermediate activities. The preference of DKGR D for acidic pH was slightly more pronounced.

**NADH-dependent activity.** With rare exceptions, aldo-keto reductases are specific for NADPH as a cosubstrate, including the *Corynebacterium* DKGRs (24, 37). Purified DKGR C and D both catalyzed the NADH-dependent reduction of 2,5-diketo-D-gluconic acid (Table 5). The reduction was much less efficient with NADH as the cosubstrate, however. The  $K_m$  for NADH was nearly 3 orders of magnitude higher than for NADPH, and the apparent  $k_{\text{cat}}$  was much lower with NADH as the cosubstrate. These combined effects result in catalytic efficiencies ( $k_{\text{cat}}/K_m$ ) 600- and 130-fold lower than those mea-

TABLE 5. Comparison of kinetic parameters with NADH as cofactor<sup>b</sup>

| Form | Parameters                          |   |   |   |                                |
|------|-------------------------------------|---|---|---|--------------------------------|
|      | $K_m$ for NADH<br>( $\mu\text{M}$ ) | $k_{\text{cat}}$<br>( $\text{s}^{-1}$ ) | $k_{\text{cat}}/K_m$ for NADH<br>( $\text{mM}^{-1} \text{s}^{-1}$ ) | $K_m$ for DKG <sup>a</sup><br>( $\mu\text{M}$ ) | $K_m$ for $\text{P}_i$<br>(mM) |
| C    | 1,800                               | 1.6                                     | 0.9   | 2,260   | 16.8                           |
| D    | 3,900                               | 12.2                                    | 3.1   | 1,150   | 10.6                           |

<sup>a</sup> DKG, 2,5-diketo-D-gluconic acid.

<sup>b</sup> Values for 2,5-diketo-D-gluconic acid were determined in the presence of 0.15 mM NADH, and values for NADH were determined in the presence of 1 mM 2,5-diketo-D-gluconic acid.

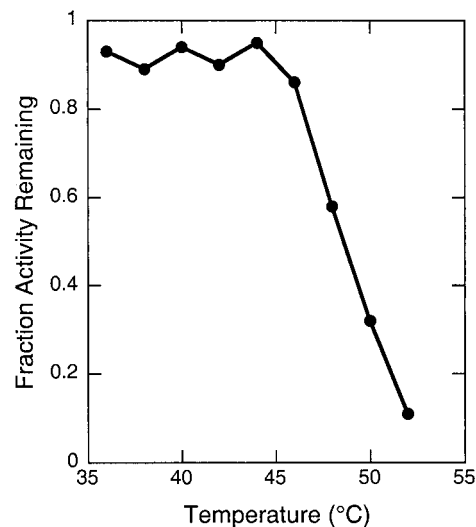


FIG. 5. Thermal inactivation of DKGR D. Residual activity of DKGR D was determined after 10 min of incubation at various temperatures established in a Robocycler (Stratagene). Samples were chilled to 4°C and were then assayed under standard conditions (30°C). The protein concentration was 0.08  $\mu\text{g}/\text{ml}$ .

sured with NADPH for DKGR C and DKGR D, respectively. Replacement of NADPH by NADH also affected the apparent  $K_m$  for 2,5-diketo-D-gluconic acid, increasing it 40- and 17-fold, respectively (Tables 4 and 5). The NADH-dependent activity was enhanced by inclusion of inorganic phosphate in the reaction buffer (Table 5). The stimulation was saturable, indicating that the phenomenon was due to binding of inorganic phosphate to the enzyme.

**Thermal stability of DKGRs.** The *Corynebacterium* reductases are thermally labile and lose activity rapidly at temperatures above 40°C (23, 35). To determine the thermal stability of the environmental reductases, we incubated each at 44°C for various periods of time using a robotic PCR thermal cycler (Robocycler Gradient 96; Stratagene) to establish temperatures rapidly and precisely (see Materials and Methods). Under these conditions, DKGR C was quite labile, losing over half its activity at the earliest time point, 0.5 min. In contrast, DKGR D was relatively stable under these conditions, with a half-life of 53.4 min. The thermal inactivation temperature of DKGR D was determined by incubating the enzyme for 10 min over a temperature gradient established by the Robocycler (Fig. 5). The enzyme retained nearly complete activity up to 45°C, after which the activity declined rapidly. The temperature under which half the activity was lost under these conditions was estimated to be 47°C.

## DISCUSSION

Because of their high specificity and selectivity, enzymes are potentially ideal catalysts for industrial processes. However, under the conditions required by many processes, known enzymes are often deficient in some way. The discovery of enzymes with better reaction rates, pH optima, thermal stability, solvent tolerance, substrate specificity, or other properties may be critical to the successful commercialization of new industrial

processes. A proposed source of better enzymes is the vast reservoir of microbes that have not as yet been cultured successfully. Useful enzymes have been recovered from uncultured microbes by screening environmental DNA expression libraries for the desired enzymatic activities (12, 33). As an alternative approach, internal fragments of a variety of genes have been amplified directly from environmental DNA by PCR using degenerate primers to conserved sequences (6, 11, 13, 18, 19, 22, 27, 31, 32).

In the present examples we extended the latter approach by applying PCR-based gene walking techniques to internal fragments. By a series of PCRs using nested primers specific for the internal fragment paired with a random primer, the less conserved, flanking regions of the genes were obtained. Putative full-length genes were predicted from the amplified fragments, and complete, functional genes encoding the targeted enzymatic activity were obtained from environmental DNA in single PCRs. The approach was successful using a convenient method of DNA extraction that has been shown recently to yield less DNA per gram of soil than alternative methods (36). As such, the environmental DNA that we used most likely represents a subsample of the total DNA in the habitats sampled, but our success shows that it still contained sufficient diversity to yield valuable genes by this approach.

The targeted enzyme, DKGR, catalyzes a critical step in a proposed biotechnological process for the production of ascorbic acid, and the goal of our research was to discover new variants with better properties. Because DKGR is a member of the aldo-keto reductase superfamily of enzymes, it is possible that genes isolated based on homology to known examples could be other aldo-keto reductases, not true DKGRs. We found that both purified enzymes catalyzed the reduction of 2,5-diketo-D-gluconic acid to 2-keto-L-gulonic acid. The reaction was highly specific; several other structurally related ketosugars were not reduced. The enzymes also possessed three properties of potential value in vitamin C production—higher catalytic efficiency, use of NADH as the cosubstrate, and in the case of DKGR D, higher thermostability.

Both DKGR C and DKGR D reduced 2,5-diketo-D-gluconic acid more efficiently than the *Corynebacterium* enzymes (35), exhibiting >20-fold-higher  $k_{cat}/K_m$  values (Table 4). The increased efficiency was primarily due to much higher affinity for 2,5-diketo-D-gluconic acid. The environmental reductases possessed  $K_m$  values of 57 and 67  $\mu$ M, whereas those for the *Corynebacterium* enzymes were 2 and 13 mM. Two recently described aldo-keto reductases encoded by the *E. coli* genes *yafB* and *yqhE* also reduce 2,5-diketo-D-gluconic acid to 2-keto-L-gulonic acid (41), but these enzymes also are inefficient catalysts with millimolar  $K_m$  values (L. Stols, unpublished observations).

The environmental DKGRs also accepted NADH as the cosubstrate, a rare characteristic among aldo-keto reductases (14, 17, 24). The *Corynebacterium* reductases are specific for NADPH (35). For a commercial process, such as the production of ascorbic acid, use of NADH in place of NADPH could result in improved performance. NADH is more abundant physiologically than NADPH and is the reduced cofactor most commonly generated from metabolism of glucose. For possible in vitro processes, NADH is more stable and much less expensive. However, the NADH-dependent activity of the new

DKGRs is much less than the NADPH-dependent activity (Tables 4 and 5), and improved variants of these DKGRs would be required for efficient use of NADH. The crystal structures of aldo-keto reductases (14) and of the *Corynebacterium* DKGR A (15) indicate that two highly conserved, positively charged residues interact with the 2'-phosphate of NADPH. Based on sequence alignments, those residues are present in the environmental reductases as well. Apparently other differences allow these enzymes to accept NADH.

A common limitation of enzymes is their sensitivity to high temperatures. The previously known DKGRs are sensitive to thermal denaturation, exhibiting midpoint denaturation temperatures of less than 40°C (23). The two new environmental reductases differ considerably in their thermal stability in spite of being 82.5% percent identical in amino acid sequence. DKGR C is more labile than the *Corynebacterium* enzymes, but DKGR D is the most stable DKGR characterized to date. However, with an inactivation temperature of only 47°C (Fig. 5), the enzyme is still relatively labile. Modeling and crystallographic studies of these enzymes should provide insights into the determinants of their thermal stability, cofactor specificity, and high affinity for substrate, providing a basis for construction of mutants or chimeras optimized for commercial use.

The new DKGRs provide potential catalysts for industry but also offer a glimpse into the possible metabolic diversity of uncultured organisms. Considering the much higher efficiencies of the new enzymes, they may represent the first true examples of enzymes whose physiological function is reduction of 2,5-diketo-D-gluconic acid to 2-keto-L-gulonic acid. This reaction, however, is not known to occur in any established metabolic pathway. ORFs flanking the DKGR genes suggest that they are part of an operon, possibly encoding sequential metabolic reactions. BLAST searches found these ORFs and the DKGRs to be most homologous to hypothetical proteins of *S. coelicolor* or *E. coli*, but inspection of those genomes showed that none of these homologues was located in a similar operon. The true physiological role of DKGRs remains to be determined. Regardless, the results support the hypothesis that enzymes from uncultured organisms can provide new, alternative catalysts with desirable properties. The success of this PCR-based approach provides a valuable complement to the established, library-based methods for extracting those catalysts. Future studies targeting different enzymes and using diverse methods of DNA preparation will define the true scope and utility of the method.

#### ACKNOWLEDGMENTS

Genencor International provided 2,5-diketo-D-gluconic acid, 2-keto-L-gulonic acid, and *Corynebacterium* spp. DKGR A. We thank Ed St. Martin, Barbara Swanson, and Steve Perri for helpful discussions.

This work was supported by the U.S. Department of Commerce's Advanced Technology Program through a grant to Genencor International, Inc., Eastman Chemical Company, Inc., Electrosynthesis Company, Inc., Microgenomics, Inc., and Argonne National Laboratory and by the Assistant Secretary for Energy Efficiency and Renewable Energy, U.S. Department of Energy, under contract W-31-109-Eng-38. W.H.E. and L.S. contributed equally to the results.

#### REFERENCES

1. Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.



2. Anderson, S., C. Marks, R. Lazarus, J. Miller, K. Stafford, J. Seymour, D. Light, W. Rastetter, and D. Estell. 1985. Production of 2-keto-L-gulonate, an intermediate in L-ascorbate synthesis, by a genetically modified *Erwinia herbicola*. *Science* **230**:144–149.
3. Ausubel, F. M., K. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1988. Current protocols in molecular biology. John Wiley and Sons, Inc., New York, N.Y.
4. Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1474.
5. Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248–254.
6. Bruce, K. 1997. Analysis of *mer* gene subclasses within bacterial communities in soils and sediments resolved by fluorescent-PCR-restriction fragment length polymorphism profiling. *Appl. Environ. Microbiol.* **63**:4914–4919.
7. Crawford, T. C. 1982. Synthesis of L-ascorbic acid, p. 1–36. In P. A. Seib and B. M. Tolbert (ed.), *Ascorbic acid: chemistry, metabolism and uses*. American Chemical Society, Washington, D.C.
8. Davey, M. W., C. Gilot, G. Persiau, J. Ostergaard, Y. Han, G. C. Bauw, and M. C. Van Montagu. 1999. Ascorbate biosynthesis in *Arabidopsis* cell suspension culture. *Plant Physiol.* **121**:535–543.
9. Fürste, J. P., W. Pansegrau, R. Frank, H. Blöcker, P. Scholz, M. Bagdasarjan, and E. Lanka. 1986. Molecular cloning of the plasmid RP4 primase region in a multi-host-range *tacP* expression vector. *Gene* **48**:119–131.
10. Grindley, J. F., M. A. Payton, H. van de Pol, and K. G. Hardy. 1988. Conversion of glucose to 2-keto-L-gulonate, an intermediate in L-ascorbate synthesis, by a recombinant strain of *Erwinia citreus*. *Appl. Environ. Microbiol.* **54**:1770–1775.
11. Hales, B. A., C. Edwards, D. A. Ritchie, G. Hall, R. W. Pickup, and J. R. Saunders. 1996. Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR amplification and sequence analysis. *Appl. Environ. Microbiol.* **62**:668–675.
12. Henne, A., R. Daniel, R. A. Schmitz, and G. Gottschalk. 1999. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.* **65**:3901–3907.
13. Herrick, J. B., E. L. Madsen, C. A. Batt, and W. C. Ghiorse. 1993. Polymerase chain reaction amplification of naphthalene-catabolic and 16S rRNA gene sequences from indigenous sediment bacteria. *Appl. Environ. Microbiol.* **59**:687–694.
14. Jez, J. M., M. J. Bennett, B. P. Schlegel, M. Lewis, and T. M. Penning. 1997. Comparative anatomy of the aldo-keto reductase superfamily. *Biochem. J.* **326**:625–636.
15. Khurana, S., D. B. Powers, S. Anderson, and M. Blaber. 1998. Crystal structure of 2,5-diketo-D-gluconic acid reductase A complexed with NADPH at 2.1-Å resolution. *Proc. Natl. Acad. Sci. USA* **95**:6768–6773.
16. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.
17. Ma, H., K. Ratnam, and T. M. Penning. 2000. Mutation of nicotinamide pocket residues in rat liver 3 alpha-hydroxysteroid dehydrogenase reveals different modes of cofactor binding. *Biochemistry* **39**:102–109.
18. McDonald, I. R., E. M. Kenna, and J. C. Murrell. 1995. Detection of methanotrophic bacteria in environmental samples with the PCR. *Appl. Environ. Microbiol.* **61**:116–121.
19. Miguez, C. B., D. Bourque, J. A. Sealy, C. W. Greer, and D. Groleau. 1997. Detection and isolation of methanotrophic bacteria possessing soluble methane monooxygenase (sMMO) genes using the polymerase chain reaction (PCR). *Microb. Ecol.* **33**:21–31.
20. Miller, J. V., D. A. Estell, and R. A. Lazarus. 1987. Purification and characterization of 2,5-diketo-D-gluconate reductase from *Corynebacterium* sp. *J. Biol. Chem.* **262**:9016–9020.
21. Nishikimi, M., and K. Yagi. 1996. Biochemistry and molecular biology of ascorbic acid biosynthesis. *Subcell. Biochem.* **25**:17–39.
22. Okuta, A., K. Ohnishi, and S. Harayama. 1998. PCR isolation of catechol 2,3-dioxygenase gene fragments from environmental samples and their assembly into functional genes. *Gene* **212**:221–228.
23. Powers, D. B., and S. Anderson. August 1998. Mutants of 2,5-diketo-D-gluconic acid (2,5-DKG) reductase A. U.S. patent 5,795,761.
24. Ratnam, K., H. Ma, and T. M. Penning. 1999. The arginine 276 anchor for NADP(H) dictates fluorescence kinetic transients in 3 alpha-hydroxysteroid dehydrogenase, a representative aldo-keto reductase. *Biochemistry* **38**:7856–7864.
25. Redenbach, M., H. M. Kieser, D. Denapaita, A. Eichner, J. Cullum, H. Kinashi, and D. A. Hopwood. 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol. Microbiol.* **21**:77–96.
26. Reichstein, T., and A. Grüssner. 1934. Eine ergiebige Synthese der 1-Ascorbinsäure (C-vitamin). *Helv. Chim. Acta* **17**:311–328.
27. Rosado, A. S., G. F. Duarte, L. Seldin, and J. D. Van Elsas. 1998. Genetic diversity of *nifH* gene sequences in *Paenibacillus azotofixans* strains and soil samples analyzed by denaturing gradient gel electrophoresis of PCR-amplified gene fragments. *Appl. Environ. Microbiol.* **64**:2770–2779.
28. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
29. Sarkar, G., R. T. Turner, and M. E. Bolander. 1993. Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *PCR Methods Appl.* **2**:318–322.
30. Seery, L. T., P. V. Nestor, and G. A. FitzGerald. 1998. Molecular evolution of the aldo-keto reductase gene superfamily. *J. Mol. Evol.* **46**:139–146.
31. Selenska, S., and W. Klingmüller. 1991. DNA recovery and direct detection of Tn5 sequences from soil. *Lett. Appl. Microbiol.* **13**:21–24.
32. Seow, K.-T., G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C. R. Hutchinson, and J. Davies. 1997. A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J. Bacteriol.* **179**:7360–7368.
33. Short, J. M. September 1999. Protein activity screening of clones having DNA from uncultivated microorganisms. U.S. patent 5,958,672.
34. Sonoyama, T., B. Kageyama, and T. Honjo. November 1975. Process for producing 2-keto-L-gulonic acid. U.S. patent 3,922,194.
35. Sonoyama, T., and K. Kobayashi. 1987. Purification and properties of two 2,5-diketo-D-gluconate reductases from a mutant strain derived from *Corynebacterium* sp. *J. Ferment. Technol.* **65**:311–317.
36. Tien, C. C., C. C. Chao, and W. L. Chao. 1999. Methods for DNA extraction from various soils: a comparison. *J. Appl. Microbiol.* **86**:937–943.
37. Todaka, T., S. Yamano, and S. Toki. 2000. Purification and characterization of NAD-dependent morphine 6-dehydrogenase from hamster liver cytosol, a new member of the aldo-keto reductase superfamily. *Arch. Biochem. Biophys.* **374**:189–197.
38. Wheeler, G. L., M. A. Jones, and N. Smirnov. 1998. The biosynthetic pathway of vitamin C in higher plants. *Nature* **393**:365–369.
39. Willey, D. L., D. A. Caswell, C. R. Lowe, and N. C. Bruce. 1993. Nucleotide sequence and over-expression of morphine dehydrogenase, a plasmid-encoded gene from *Pseudomonas putida* M10. *Biochem. J.* **290**:539–544.
40. Yum, D.-Y., B.-Y. Lee, D.-H. Hamm, and J.-G. Pan. 1998. The *viaE* gene, located at 80.1 minutes on the *Escherichia coli* chromosome, encodes a 2-ketoaldonate reductase. *J. Bacteriol.* **180**:5984–5988.
41. Yum, D.-Y., B.-Y. Lee, and J.-G. Pan. 1999. Identification of the *yqhE* and *yafB* genes encoding two 2,5-diketo-D-gluconate reductases in *Escherichia coli*. *Appl. Environ. Microbiol.* **65**:3341–3346.