

Cellobiose-6-Phosphate Hydrolase (CelF) of *Escherichia coli*: Characterization and Assignment to the Unusual Family 4 of Glycosylhydrolases

JOHN THOMPSON,^{1*} SERGEI B. RUVINOV,² DARÓN I. FREEDBERG,³ AND BARRY G. HALL⁴

Microbial Biochemistry and Genetics Unit, Oral Infection and Immunity Branch, National Institute of Dental and Craniofacial Research,¹ and Laboratory of Biochemistry, National Heart, Lung, and Blood Institute,² National Institutes of Health, and Laboratory of Biophysics, Center for Biologics Evaluation and Research, Food and Drug Administration,³ Bethesda, Maryland 20892, and Biology Department, University of Rochester, Rochester, New York 14627-0211⁴

Received 21 July 1999/Accepted 22 September 1999

The gene *celF* of the cryptic *cel* operon of *Escherichia coli* has been cloned, and the encoded 6-phospho- β -glucosidase (cellobiose-6-phosphate [6P] hydrolase; CelF [EC 3.2.1.86]) has been expressed and purified in a catalytically active state. Among phospho- β -glucosidases, CelF exhibits unique requirements for a divalent metal ion and NAD^+ for activity and, by sequence alignment, is assigned to family 4 of the glycosylhydrolase superfamily. CelF hydrolyzed a variety of P- β -glucosides, including cellobiose-6P, salicin-6P, arbutin-6P, gentiobiose-6P, methyl- β -glucoside-6P, and the chromogenic analog, *p*-nitrophenyl- β -D-glucopyranoside-6P. In the absence of a metal ion and NAD^+ , purified CelF was rapidly and irreversibly inactivated. The functional roles of the cofactors have not been established, but NAD^+ appears not to be a reactant and there is no evidence for reduction of the nucleotide during substrate cleavage. In solution, native CelF exists as a homotetramer (M_w , ~200,000) composed of noncovalently linked subunits, and this oligomeric structure is maintained independently of the presence or absence of a metal ion. The molecular weight of the CelF monomer (M_r , ~50,000), estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, is in agreement with that calculated from the amino acid sequence of the polypeptide (450 residues; M_r = 50,512). Comparative sequence alignments provide tentative identification of the NAD^+ -binding domain (residues 7 to 40) and catalytically important glutamyl residues (Glu¹¹² and Glu³⁵⁶) of CelF.

The *cel* operon is one of three cryptic operons designated *bgl* (19, 28, 35), *asc* (11, 24), and *cel* (10, 16) that, when suitably activated (2, 31, 34), allow growth of *Escherichia coli* on β -glucosides. The *cel* operon maps to the 39.1- to 39.2-min region of the bacterial chromosome and contains six genes (*celABCDFG*) whose products promote the accumulation and dissimilation of cellobiose, salicin, and arbutin by the organism (10, 25). In sequential order, the genes *celABC* encode the phosphoenolpyruvate-dependent sugar-phosphotransferase components IIB^{cel}, IIC^{cel}, and IIA^{cel} (15, 25, 30) that catalyze the phosphorylation and simultaneous translocation of β -glucosides across the cytoplasmic membrane. (For a discussion of phosphotransferase functions and nomenclature, see references 21, 27, and 32). The next gene in the operon (*celD*) encodes a repressor, and *celF* encodes a phospho (P)- β -glucosidase (CelF) that hydrolyzes O- β -linked phosphorylated derivatives to generate glucose-6-phosphate (G6P) and appropriate aglycone (26). The product of *celG* has not been characterized. Activation of the *cel* operon can occur either via insertion of IS1, IS2, or IS5 into a region 72 to 180 bp upstream of the transcription start site or by base substitutions in *celD* such that the repressor is able to recognize cellobiose, arbutin, and salicin as inducers (26). In a recent report, Keyhani and Roseman suggest that the *cel* operon may also allow growth of *E. coli* on the chitin disaccharide *N,N'*-diacetylchitobiose without requirement for mutational activation (15).

Hall and coworkers (17, 26) obtained evidence for the CelF-catalyzed hydrolysis of P- β -glucosides by extracts prepared from cellobiose-grown cells of *E. coli*. However, despite repeated attempts, purification of CelF was never achieved because of the instability of the enzyme. That difficulty should be encountered in purification of CelF from *E. coli* was surprising, because the purification of two other P- β -glucosidases (BglA and BglB) from this organism had been relatively straightforward (28, 29, 46, 49). Additionally, hydrolysis of the chromogenic substrate *p*-nitrophenyl- β -D-glucopyranoside-6P (pNP β G6P) by BglA, BglB, and extracts containing CelF suggested catalytic similarity among the three P- β -glucosidases. Differences between CelF and the other P- β -glucosidases were evident at the molecular level, and comparative alignment of amino acid sequences revealed little homology between the sequence deduced for the polypeptide encoded by *celF* and those deduced for BglA and BglB. Furthermore, when classified by the amino acid sequence-based method of Henrissat and Bairoch (12, 13), BglA and BglB were included in family 1 whereas CelF was assigned to family 4 of the glycosylhydrolase superfamily (5, 8, 12, 40).

From microbial-genome-sequencing projects of the past decade, 16 bacterial glycosylhydrolases (some putative) can now be assigned to family 4 (see Fig. 2). However, it was only recently that the first two enzymes (both P- α -glucosidases) from this family were purified in active form. Both enzymes (MalH from *Fusobacterium mortiferum* [39] and GlvA from *Bacillus subtilis* [40]) are unstable, and both exhibit unique requirements for divalent metal ions (Mn^{2+} , Ni^{2+} , Co^{2+} , or Fe^{2+}) and nucleotide (NAD^+) for activity. Realizing that other members of family 4 might also require these (or addi-

* Corresponding author. Mailing address: National Institutes of Health, Bldg. 30, Room 528, Convent Dr. 4350, Bethesda, MD 20892-4350. Phone: (301) 496-4083. Fax: (301) 402-0396. E-mail: jthompson@dir.nidcr.nih.gov.

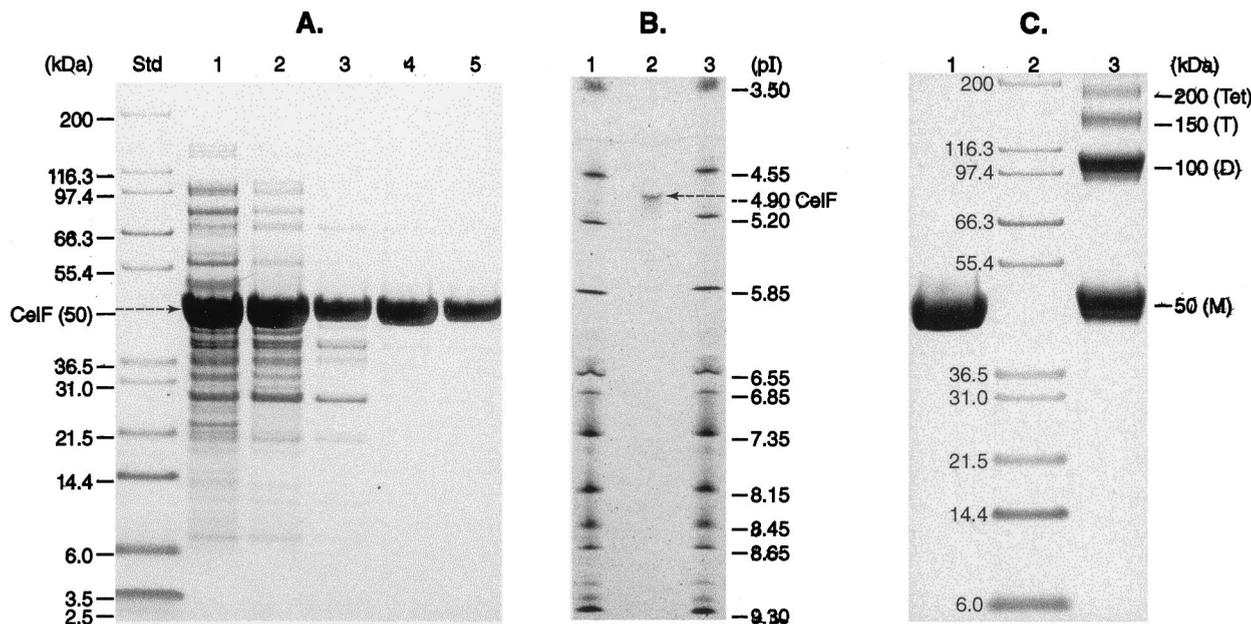


FIG. 1. Determination of the M_r , pI, and structural composition of purified CelF by analytical PAGE. (A) Samples from each of the five stages of enzyme purification were denatured, resolved by SDS-PAGE (4 to 12% acrylamide; Bis-Tris gel), and stained with Coomassie brilliant blue. Lane 1, high-speed supernatant; lane 2, DEAE Tris-Acryl M; lane 3, phenyl-Sepharose CL-4B; lane 4, Ultrogel Aca-44; and lane 5, purified CelF (M_r , ~50,000; ~8 μ g) from Blue Tris-Acryl. Also shown are Novex Mark 12 molecular weight markers (Std). (B) Determination of the pI of CelF (~4.9) by analytical electrofocusing. Approximately 1 μ g of enzyme was applied directly to the surface of the isoelectric focusing gel in lane 2 and protein standards (pI range, 3.5 to 9.3; Pharmacia) in lanes 1 and 3. (C) Cross-linking of the subunits of native CelF by dimethyl adipimidate (DMA). A mixture containing ~40 μ g of CelF and ~200 μ g of DMA was prepared in 20 μ l of 0.1 M triethanolamine buffer (pH 8). After 2 h of incubation at room temperature, 10 μ l of the sample was denatured and analyzed by SDS-PAGE (lane 3). M, D, T, and Tet, monomer, dimer, trimer, and tetrameric species, respectively. Denatured monomer (control; no DMA) and molecular weight markers are shown in lanes 1 and 2, respectively.

tional) cofactors, we renewed our attempts to purify CelF from *E. coli* in a catalytically active state. Success in this collaborative effort required (i) high expression of CelF in a strain of *E. coli* devoid of all other P- β -glucosidases, (ii) synthesis of both natural and artificial substrates of CelF, and (iii) inclusion of Mn^{2+} and NAD^+ in chromatography buffers during purification of the enzyme.

In this communication, we describe the purification, substrate specificity, and physicochemical properties of CelF from *E. coli*. By its intrinsic instability, unusual functional requirements (a divalent metal ion plus NAD^+), and oligomeric structure, CelF exhibits properties that have not previously been described for other β - or P- β -glycosylhydrolases. Inclusion of CelF in family 4 of the glycosylhydrolase superfamily raises questions concerning the predictability of enzyme specificity, and the catalytic mechanism, for proteins based solely upon sequence-based alignment.

MATERIALS AND METHODS

Materials. Isoelectric focusing standards, Ampholine PAG plates (pH 3.5 to 9.5), and phenyl-Sepharose CL-4B were purchased from Amersham Pharmacia Biotech. Ultrogel Aca-44 and Blue Tris-Acryl were from BioSeptra. Enzymes, nucleotides, nitrophenyl glycosides, and DEAE Tris-Acryl M were supplied by Sigma Chemical Co. High-purity sugars were obtained from Pfanstiehl Laboratories. Reagents for chemical syntheses, including trimethylphosphate, phosphorus oxychloride, and cyclohexylamine, were purchased from Aldrich.

***E. coli* K-12 strains.** Strain MK91 is $\Delta lacZ4680 celR1::IS2 \Delta(bgl-pho) rpsL ara-14 leuB6 \Delta lacZ4680 lacY trpE38 his-208 argG77 metA160 galK2(Oc)$ (16). Strain LP100 is $\Delta lacZ4680 celR1::IS2 \Delta celBCDF \Delta(bgl-pho) rpsL ara-14 leuB6 \Delta lacZ4680 lacY trpE38 his-208 argG77 metA160 galK2(Oc)$ (24). Strain PEP0 is dTn10_{cam} *ebgA5100 ebgR*⁺L532 *rpsL* $\Delta lacZ4680$ (Hall Laboratory collection). Strain FCY2007arb11 is IN(*rrnD-rmE*)1 *trpA46* $\Delta(bgl-pho)$ 201 $\Delta tna arbT$ dTn10_{kan}::*bglA1000* (Hall Laboratory collection). Strain PEP24 is *celR1::IS2* $\Delta celBCDF \Delta(bgl-pho) rpsL ara-14 leuB6 \Delta lacZ4680 lacY his-208 metA160 galK2(Oc)$ dTn10_{cam}::*ebgA5100 ebgR*⁺L532 dTn10_{kan}::*bglA1000*. Strain PEP24 was constructed as follows. LP100:dTn10_{cam}::*ebgA5100 ebgR*⁺L532 was trans-

duced by bacteriophage P1_{vir} from strain PEP0 into strain LP100 to create strain PEP21. A spontaneous Trp⁺ revertant of PEP21 was selected to create PEP22. A spontaneous Arg⁺ revertant of PEP22 was selected to create PEP23. dTn10_{kan}::*bglA1000* was transduced from strain FCY2007arb11 into strain PEP23 to create strain PEP24. Phenotypically, strain PEP24 exhibits no detectable P- β -glucosidase activity when assayed with pNP β G6P because the *cel* and *bglGFB* operons are deleted, *bglA* is disrupted by dTn10_{kan}, and the *asc* operon is cryptic (silent). Strain PEP14 is *celR1::IS2* $\Delta(bgl-pho) rpsL ara-14 leuB6 \Delta lacZ4680 lacY his-208 metA160 galK2(Oc)$ dTn10_{cam} *ebgA5100 ebgR*⁺L532 dTn10_{kan} *bglA* and was constructed as follows. dTn10_{cam}::*ebgA5100 ebgR*⁺L532 was transduced by bacteriophage P1_{vir} from strain PEP0 into strain MK91 to create strain PEP11. Trp⁺ was transduced from strain PEP23 into strain PEP11 to create strain PEP12. Arg⁺ was transduced from strain PEP23 into strain PEP12 to create strain PEP13. dTn10_{kan}::*bglA1000* was transduced from strain FCY2007arb11 into strain PEP13 to create strain PEP14. Phenotypically, strain PEP14 expresses the *cel* operon constitutively but expresses no other P- β -glucosidases because the *bglGFB* operon is deleted, *bglA* is disrupted by dTn10, and the *asc* operon is cryptic.

Construction of pUF4000. Genomic DNA was prepared from strain PEP14 by the cetyltrimethylammonium bromide method (3). The *celF* gene was amplified by the high-fidelity *Pfu* polymerase (Stratagene) according to the manufacturer's instructions with primers corresponding to bp 1816760 to 1816784 and bp 1815080 to 1815105 of the *E. coli* genome (4). The left primer also included an *Eco*RI site, while the right primer included a *Bam*HI site. The resulting 1.7-kb fragment was purified with a Qiagen PCR spin kit according to the manufacturer's instructions, digested with *Eco*RI and *Bam*HI, and ligated into similarly digested plasmid pSE380 (Invitrogen). The resulting construct, designated pUF4000, expresses *celF* under the control of the powerful pTAC promoter, which is controlled by the *lacI*-encoded Lac repressor. Because the plasmid also carries *lacI*, expression of *celF* is strongly repressed in the absence of isopropyl- β -D-thiogalactopyranoside (IPTG) and is fully induced in the presence of 1 mM IPTG.

The *celF* sequence that was originally reported (25) (GenBank accession no. M64438) included errors that resulted in reporting CelF as a 378-amino-acid protein. The complete sequence of *E. coli* K-12 reported by Blattner et al. (4) showed that CelF is 78 amino acids longer than originally reported (GenBank accession no. AE000268). We have confirmed that the nucleotide sequence of *celF* in strain PEP14 is identical to that reported by Blattner et al. To determine whether any mutations had occurred during amplification of *celF*, both strands of the insert in pUF4000 were sequenced on an Applied Biosystems model 377 automated DNA sequencer and found to correspond exactly to the sequence

reported by Blattner et al. (4). In pUF4000, translation proceeds from the strong ribosome binding site at bp 260 to 270 to generate an out-of-frame fusion with a C-terminal portion of the CelD protein that terminates at bp 365. It is likely that translation reinitiates to generate an in-frame C-terminal fragment of CelD that terminates at the normal stop codon of *celD* at bp 573. Translation of *celF* then initiates at bp 576 to yield an authentic CelF protein, as determined by sequencing the N-terminal 30 amino acids of the purified protein.

Growth of cells and preparation of cell extract. *E. coli* PEP24(pUF4000) was grown at 37°C in a Tris-HCl-phosphate-buffered Casamino Acids-salts medium (pH 7.2) containing ampicillin (200 µg/ml). When the culture had reached an optical density at 600 nm of ~0.1, IPTG was added to a final concentration of 1 mM and growth was continued until the optical density at 600 nm was ~1.0. The cells were harvested by centrifugation (13,000 × g for 10 min at 5°C) and washed by resuspension and centrifugation in 25 mM Tris-HCl buffer (pH 7.5) containing 1 mM dithiothreitol, 1 mM MnSO₄, and 50 µM NAD⁺ (designated TDMN buffer). The washed cells (55 g [wet weight]) were resuspended in 100 ml of TDMN buffer, and the organisms were disrupted at 0°C by two 1.5-min periods of sonic oscillation in a Branson sonifier operating at ~75% maximum power. The cell extract was clarified by ultracentrifugation (180,000 × g for 2 h at 5°C).

Purification of CelF. The enzyme was purified by conventional low-pressure chromatography in four stages: (i) DEAE Tris-Acryl M (anion exchange), (ii) phenyl-Sepharose CL-4B (hydrophobic interaction), (iii) Ultrogel AcA-44 (gel filtration), and (iv) Blue Tris-Acryl (affinity chromatography). All procedures were performed in the cold room.

Assay of CelF activity. During purification, the specific activity of the enzyme was determined in a discontinuous assay with chromogenic pNPβG6P as a substrate. The 2-ml reaction mixture (at 37°C) contained 50 mM Tris-HCl buffer (pH 7.5), 1 mM pNPβG6P, 1 mM MnSO₄, and 0.1 mM NAD⁺. After addition of the enzyme preparation, samples (0.25 ml) were removed at intervals (usually over a 2- to 3-min period) and immediately injected into 0.75 ml of 0.5 M Na₂CO₃ containing 0.1 M EDTA to stop the reaction. The A₄₀₀ was measured, and the amount of pNP was calculated from the molar extinction coefficient of the (yellow) *p*-nitrophenolate anion ($\epsilon = 18,300 \text{ M}^{-1} \text{ cm}^{-1}$). One unit of CelF activity is the amount of enzyme that catalyzes the formation of 1 µmol of pNP per min at 37°C. A continuous spectrophotometric assay was employed for substrate specificity studies and determination of kinetic parameters for CelF. In this assay, hydrolysis of *P*-β-glucoside (G6P formation) was monitored by the NAD⁺-plus-glucose-6-phosphate dehydrogenase (G6PDH) (EC 1.1.1.49) coupled reaction. The standard 500-µl assay mixture contained 50 mM Tris-HCl buffer (pH 7.5), 1 mM MnSO₄, 1 mM NAD⁺, a 0.5 to 4 mM concentration range of the desired *P*-β-glucoside, and 2 U of G6PDH (from *Leuconostoc mesenteroides*). Reactions were initiated by the addition of 5 µl of purified CelF (188 µg of protein), and the increase in the A₃₄₀ was recorded in a Beckman DU-70 spectrophotometer. A molar extinction coefficient ($\epsilon = 6,220 \text{ M}^{-1} \text{ cm}^{-1}$) for NADH was assumed for calculation of initial rates of substrate hydrolysis. Kinetic parameters were determined from Hofstee plots with an Enzyme Kinetics program (version 1.0c) (dogStar software). The products of cellobiose-6P cleavage were also determined by spectrophotometric assay with the further addition of ATP (5 mM) and 2 U of hexokinase (EC 2.7.1.1) for glucose estimation.

Preparation of *P*-β-glycosides. All *P*-β-glycosides were prepared in our laboratory. Chemical syntheses of pNPβG6P, pNPαG6P, pNPα-galactopyranoside-6P, pNPα-mannopyranoside-6P, and 4-methylumbelliferyl-β-glucopyranoside-6P (4MUβG6P) were initiated with the nonphosphorylated glycosides by the method of Wilson and Fox (46). Selective phosphorylation at the C-6 OH moiety of the glycopyranose was achieved by the use of a mixture of phosphorus oxychloride in trimethylphosphate containing small amounts of water. The phosphorylated derivatives were obtained as white, crystalline cyclohexylamine salts in 25 to 30% yield. Cellobiose-6P and other disaccharide and arylglycoside phosphates were prepared by enzymatic phosphorylation with the ATP-dependent β-glucoside kinase prepared from cellobiose-grown cells of *Klebsiella pneumoniae* (23). The *P*-β-glycosides were purified by a combination of Ba²⁺-ethanol precipitation, ion exchange, and paper chromatography.

Analytical procedures. Native gel electrophoresis and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were performed with the Novex XCell minicell system according to the manufacturer's instructions. Novex Tris-glycine (4 to 20%) gels were used for electrophoresis of CelF preparations under nonreducing conditions, and the Tris-glycine running buffer (pH 8.3) contained 0.1 mM NAD⁺ and 1 mM MnSO₄. Novex NuPage (4 to 12%) Bis-Tris gels, MES (2-[*N*-morpholino]ethanesulfonic acid)-SDS running buffer (pH 7.3), and Novex Mark 12 wide-range molecular weight protein standards were used for SDS-PAGE experiments. A Multiphor flat-bed electrophoresis unit (Pharmacia) and precast Ampholine PAG plates (pH range, 3.5 to 9.5) were used for electrofocusing experiments. Protein concentrations were routinely determined by the bicinchoninic acid protein assay reagent (Pierce). Microsequencing and electrospray mass spectrometry procedures have been described previously (40). Approximate *M_r* values for CelF were determined by fast protein liquid chromatography gel filtration with a Pharmacia Superose 6 HR 10/30 prepacked column and a Hewlett-Packard 1090 liquid chromatography system. The column was pre-equilibrated (flow rate, 0.4 ml/min; temperature, ~23°C) with 25 mM Tris-HCl buffer (pH 7.5) containing 0.1 M NaCl and calibrated with protein molecular weight standards. Protein samples (100 µg in 0.1 ml) were loaded onto the column, and elution was monitored at 280 nm.

Analytical ultracentrifugation. Experiments were performed in the Beckman Optima model XL-I analytical ultracentrifuge by previously described procedures (40). Data analysis was performed with the software package installed by Beckman, together with software provided by Allen P. Minton (National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health). The latter program can be downloaded from the World Wide Web (30a). The densities of dialysis buffers were determined at 20°C with an Anton Paar model DMA 58 densitometer. The partial specific volume for CelF ($V_{\text{bar}} = 0.729 \text{ ml/g}$) was calculated from the deduced amino acid composition of the protein together with the values of Zamyatin (50).

RESULTS AND DISCUSSION

Construction of *E. coli* PEP24. The activity of CelF is easily monitored by the formation of the yellow *p*-nitrophenolate ion generated by hydrolysis of the chromogenic substrate, pNPβG6P. Unfortunately, the constitutively expressed *BglA* of *E. coli* also hydrolyzes this compound, and CelF activity is virtually undetectable against the *BglA* background. To circumvent this problem, *E. coli* PEP24 was constructed, in which *bglA* is disrupted by mini-Tn10_{kan} insertion, the *asc* operon is cryptic, and both *bgl* and *cel* operons are deleted. Extracts of strain PEP24 contain no *P*-β-glucosidase activity.

Expression of CelF in *E. coli* PEP24(pUF4000). Cells of *E. coli* PEP24 when transformed with pUF4000 produced high levels of an IPTG-inducible protein whose estimated molecular weight (*M_r*, ~50,000) was that expected for the full-length polypeptide encoded by *celF*. The polypeptide was not present in a similarly prepared extract of the plasmid-free PEP24 parental strain. Native gel electrophoresis and subsequent *in situ* staining for CelF activity with 4MUβG6P as a substrate revealed a single zone of enzymatic activity (4MU fluorescence) only in the extract prepared from PEP24(pUF4000).

Purification and properties of CelF. Purification of CelF was achieved by conventional low-pressure chromatography, using 25 mM Tris-HCl buffer (pH 7.5) supplemented with 1 mM Mn²⁺ and 0.1 mM NAD⁺. The five-stage procedure yielded approximately 135 mg of purified enzyme from 55 g (wet weight) of cells, but disturbingly, the specific activity of the final preparation (~1.8 U/mg) was only twofold greater than that of the original high-speed supernatant (~0.9 U/mg). Although purified in active form, CelF is clearly (and progressively) inactivated during purification. Analysis of the purified enzyme by SDS-PAGE revealed a single polypeptide of ~50,000 *M_r* (Fig. 1A, lane 5). Analytical electrofocusing of CelF also revealed a single diffuse polypeptide (Fig. 1B, lane 2), but the experimentally determined pI of ~4.9 was considerably lower than the theoretical pI (6.34) calculated from the amino acid composition. The results of cross-linking experiments provided evidence of the oligomeric structure of CelF (Fig. 1C, lane 3), and a single high-molecular-weight tetrameric species (*M_r*, ~210,000) was detected by fast protein liquid chromatography gel filtration (data not shown). Microsequencing provided evidence of enzyme homogeneity, and (except for the terminal methionine residue) the unambiguous sequence of the first 30 amino acids from the N terminus agreed precisely with that expected from translation of *celF*: SQKLKVVITIGGSSYPLELLEGFRIKRYHEL. The *M_r* of 50,386 for CelF obtained by electrospray-mass spectrometry was less than the calculated molecular weight of the full-length polypeptide (50,512) by the mass equivalence of one methionine residue.

Assay of CelF activity *in vitro*. Addition of either the ultracentrifugal cell extract of *E. coli* PEP24(pUF4000) or purified CelF to assays containing both Mn²⁺ ion and NAD⁺ elicited immediate hydrolysis of pNPβG6P (ca. 1 to 2 µmol of pNP formed min⁻¹ mg⁻¹). Omission of the nucleotide markedly reduced the rate of enzyme activity, and in an Mn²⁺-free

TABLE 1. Metal ion specificity and NAD⁺ requirement for activity of CelF

Addition to assay ^a	Specific activity ^b
Control (no addition).....	0.017
Mn ²⁺	0.249
Mn ²⁺ + NAD ⁺	1.155
Ni ²⁺	0.042
Ni ²⁺ + NAD ⁺	0.880
Co ²⁺	0.197
Co ²⁺ + NAD ⁺	0.292
Mg ²⁺	ND ^c
Mg ²⁺ + NAD ⁺	0.065
Sr ²⁺	ND
Sr ²⁺ + NAD ⁺	0.061
Ca ²⁺	ND
Ca ²⁺ + NAD ⁺	0.048

^a The 2-ml assay mixture contained 50 mM Tris-HCl buffer (pH 7.5) and, when required, 1 mM appropriate divalent metal ion, 1 mM NAD⁺, and 0.38 mg of dialyzed CelF. After 5 min of preincubation at 37°C, the substrate (pNPβG6P) was added to a final concentration of 1 mM.

^b Expressed as micromoles of pNPβG6P hydrolyzed minute⁻¹ milligram of enzyme⁻¹.

^c ND, no detectable activity.

system pNPβG6P hydrolysis ceased within ~2 min. Preincubation in an assay lacking both a metal ion and nucleotide resulted in complete loss of CelF activity. Properties similar to those we report for CelF of *E. coli* have also been described for other family 4 glycosylhydrolases, including maltose 6-phosphate hydrolases from *F. mortiferum* (MalH [39]) and *B. subtilis* (GlvA [40]), respectively, α-galactosidase from *E. coli* (6), and α-glucosidase from *Thermotoga maritima* (17a). In the case of GlvA, loss of activity in the absence of a metal ion was in part attributable to dissociation of the Mn²⁺-stabilized active tetramer to a metal-free, inactive dimeric state (40). Accordingly, ultracentrifugal sedimentation velocity analysis was used to further characterize the apo and Mn²⁺ forms of CelF. In the presence of Mn²⁺ ion, a single monodisperse sedimenting boundary of solute was observed, but in solutions lacking Mn²⁺ ion a minor component comprising up to 13% of total solute (sedimentation coefficients [$s_{20,w}$], ~2 to 4 S) was also detected. This lower-molecular-weight fraction may represent a degradation product of the less conformationally stable apoenzyme. The major fraction of the enzyme existed as a single species of 8.7 ± 0.2 S for the metal-free form and 8.9 ± 0.2 S for the Mn²⁺-containing form. The tetrameric structure of CelF was confirmed by sedimentation equilibrium experiments conducted at three protein concentrations. A single component ($M_r = 202,000 \pm 5\%$) was found for both metal-free and Mn²⁺-associated forms of CelF. The rapid loss of

TABLE 2. Substrate specificity and kinetic parameters of purified CelF

P-β-glucoside	K_m (mM)	V_{max} (μmol min ⁻¹ mg ⁻¹)
4MUβG6P	0.30 ± 0.02	3.35 ± 0.08
Arbutin-6P ^a	0.35 ± 0.04	2.13 ± 0.08
pNPβG6P	0.44 ± 0.04	1.33 ± 0.05
Salicin-6P ^b	0.44 ± 0.09	1.38 ± 0.09
Cellobiose-6P	1.30 ± 0.09	1.74 ± 0.05
MethylβG6P	2.22 ± 0.21	0.56 ± 0.03
Gentiobiose-6P ^c	12.50 ± 1.26	1.11 ± 0.11

^a 4-hydroxyphenyl-β-D-glucopyranoside-6P.

^b O-hydroxy-methylphenyl-β-D-glucopyranoside-6P.

^c 6-O-β-D-glucopyranosyl-D-glucopyranoside-6P.

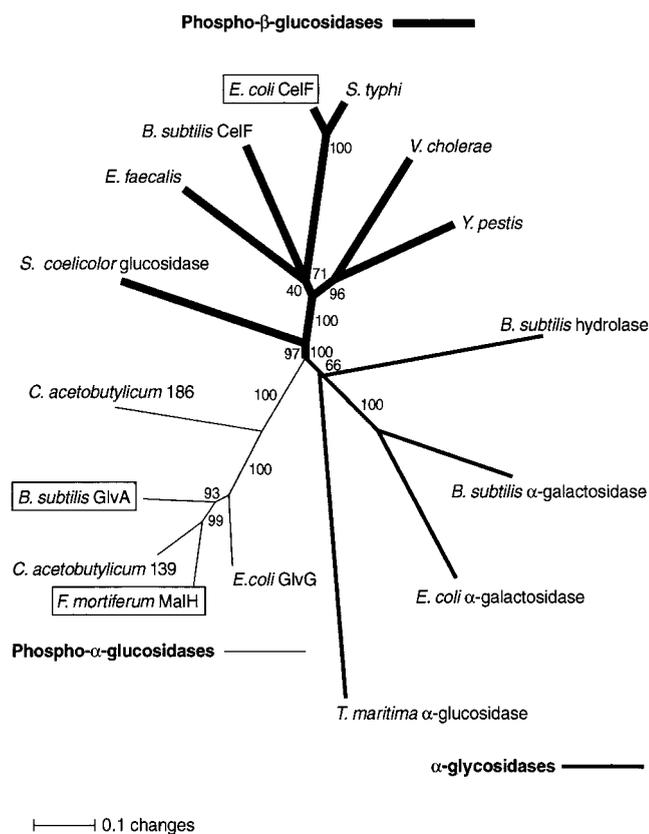


FIG. 2. Unrooted neighbor-joining phylogenetic tree of family 4 glycosylhydrolases. The 16 protein sequences were aligned by the program CLUSTAL X (41, 42). For pairwise alignments, a gap-opening penalty of 35 and a gap extension penalty of 0.75 were used. For multiple alignments, a gap-opening penalty of 15 and a gap extension penalty of 0.3 were used. The alignment was used to construct the neighbor-joining tree (33) as implemented by PAUP (Phylogenetic Analysis Using Parsimony) version 4.0b1a (37, 38). The numbers are bootstrap values, and branch lengths indicate the fraction of amino acid substitutions, as indicated on the scale. Proteins that have been purified to homogeneity and characterized are boxed. Entries without gene or protein names (*E. faecalis*, *V. cholerae*, etc.) are derived from sequences of unfinished genomes.

activity of CelF upon dilution in the Mn²⁺-free assay mixture cannot simply be attributed to tetramer dissociation but may result from conformational changes within the oligomeric structure (or active site) of the enzyme. In their studies of the instability of α-galactosidase from *E. coli*, Burstein and Kepes in 1971 proposed that dilution played a role in enzyme inactivation (6). In this context (see below), activity of a concentrated preparation of CelF was maintained after dialysis against Mn²⁺-free 25 mM Tris-HCl buffer (pH 7.5).

Metal ion and nucleotide specificity. In the standard assay, the activity of CelF was greatest within the temperature range of 34 to 38°C. CelF activity was not significantly affected by the buffer species, and within a pH range of 7.0 to 8.0, comparable rates of pNPβG6P hydrolysis were measured in BES {2-[bis(2-hydroxyethyl)amino]ethanesulfonic acid}-, TES [*N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid]-, HEPES-, or Tricine-buffered assays. To determine the specificity with respect to metal and nucleotide requirements, a concentrated preparation of CelF (35 mg ml⁻¹) was first dialyzed against 25 mM Tris-HCl buffer (pH 7.5). Enzyme activity was then determined in assays containing different metal ions in combination with NAD⁺ (Table 1). Mn²⁺, Co²⁺, and Ni²⁺ enhanced the activity of CelF, and for these ions, the presence of NAD⁺

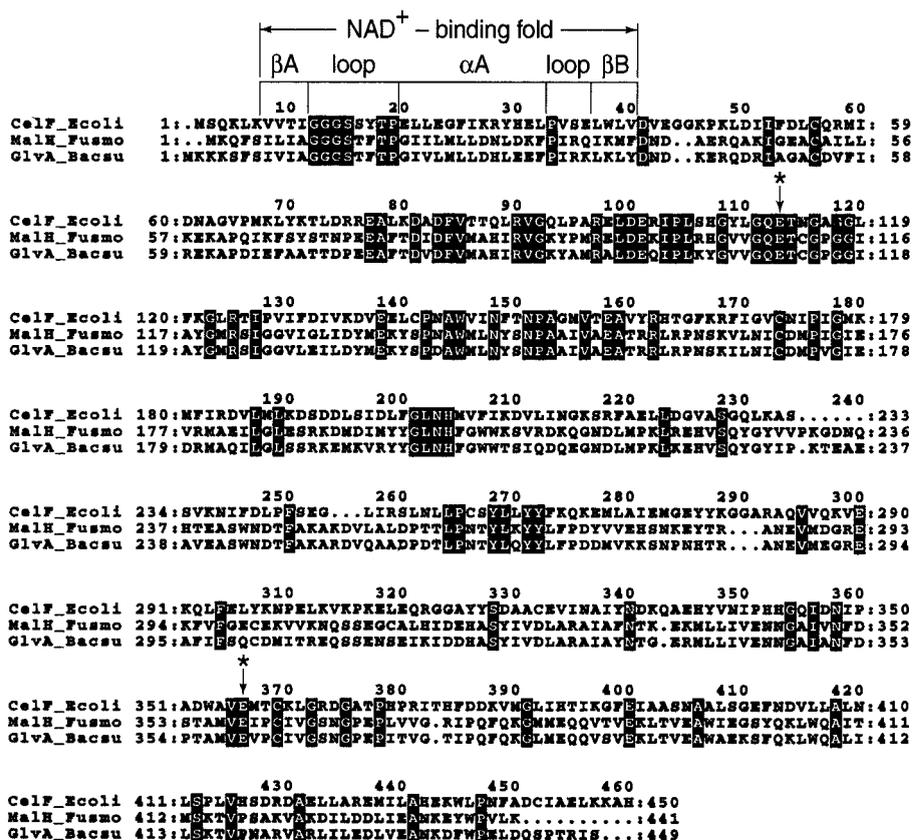


FIG. 3. Alignment of the amino acid sequences of the three enzymes from family 4 that have been purified and characterized: CelF (P- β -glucosidase from *E. coli* [this study]), MalH (P- α -glucosidase; *F. mortiferum* [5]), and GlvA (P- α -glucosidase; *B. subtilis* [40]). Identical residues are highlighted, and putative catalytic residues are marked by asterisks. The numbers to the left and right denote residue positions, and the numbers above the residues refer to alignment positions.

increased the rate of pNP β G6P hydrolysis by 2- to 20-fold. Other divalent metals, including Mg²⁺, Ca²⁺, Sr²⁺, and Zn²⁺, had little effect upon the activity of CelF either singly or in combination with the nucleotide. The concentrations of Me²⁺ and NAD⁺ required for optimum activity of CelF were determined to be 1 and 0.1 mM, respectively (data not shown). Of a variety of nucleotides tested, including NAD⁺, 3-acetylpyridine adenine dinucleotide, nicotinamide mononucleotide, NADP⁺, NADPH, NADPH (deamino), NAD⁺ (deamido), and NAD⁺ (bis dialdehyde), only NAD⁺ activated CelF.

Substrate specificity of CelF. Substrates hydrolyzed by CelF, and the relevant kinetic parameters, are presented in Table 2. In the presence of Mn²⁺ ion and NAD⁺, the enzyme hydrolyzed all P- β -glucosides tested. There was no detectable cleavage of the corresponding nonphosphorylated compounds (data not shown). Although exacting with respect to the presence of the glucopyranosyl-6P moiety and O- β glycosidic linkage, CelF enzyme is tolerant of wide variation in the structure and composition of the aglycone substituent of its substrates. The importance of the equatorial position of the C-4 OH group for pNP α G6P cleavage is evident from the fact that epimeric pNP β Gal6P (axial -OH at C-4) was not hydrolyzed by the enzyme. Spectrophotometric analyses showed that CelF catalyzed the hydrolysis of 0.05 μ mol of cellobiose-6P to yield 0.05 μ mol each of G6P and glucose. No evidence was obtained for reduction of NAD⁺ during cellobiose-6P cleavage.

Relationship of CelF to other family 4 members. The amino acid sequence of CelF (P17411) was used as a probe to search the nonredundant protein sequence database with the Gapped

BLAST programs (1). Nine homologs were identified: CelF of *B. subtilis* (P46320; a putative 6-phospho- β -glucosidase), a putative glucosidase of *Streptomyces coelicolor* (AL031107), MalH of *F. mortiferum* (O06901; 6-phospho- α -glucosidase), GlvA of *B. subtilis* (P54716; 6-phospho- α -glucosidase), GlvG of *E. coli* (P31450; truncated 6-phospho- α -glucosidase), a putative α -galactosidase of *B. subtilis* (O34645), MelA of *E. coli* (P06720; α -galactosidase), a putative hydrolase of *B. subtilis* (Z99107), and AglA of *T. maritima* (O33830; α -glucosidase). CelF was also used to search the National Center for Biotechnology Information database of the unpublished sequences of incomplete microbial genomes, and six additional protein homologs were identified: one from *Salmonella typhi* contig 679 (Sanger Centre), one from *Yersinia pestis* contig 792 (Sanger Centre), one from *Enterococcus faecalis* gef 6229 (The Institute for Genomic Research), one from *Vibrio cholerae* serotype O1 (The Institute for Genomic Research), and two from *Clostridium acetobutylicum* (contigs 139I and 186I [Genome Therapeutics Corp.]). The 16 protein sequences were first aligned by the program CLUSTAL X (41, 42), and the resulting alignment was used to construct a neighbor-joining (distance) tree (Fig. 2). Family 4 comprises three subgroups of glycosylhydrolases (all of bacterial origin), including (I) phospho- β -glucosidases, (II) phospho- α -glucosidases, and (III) α -glucosidases. At least one enzyme from each group has been purified in active form.

Stereospecificity of family 4 glycosylhydrolases. More than 60 sequence-based families of glycosylhydrolases have been established (13, 14), and this information is accessible from the

World Wide Web (36a). A fundamental property of sequence-based classification of glycosylhydrolases is the predictability of stereochemical specificity (α or β linkage in substrates) and molecular mechanism (inversion or retention) for all members of a particular family (7, 14, 20). Prior to the purification and characterization of CelF, only α -glycosidase activities had been described for members of family 4, including α -galactosidase (6, 18, 22), α -glucosidase (17a), and the two P- α -glucosidases (5, 40). With the addition of CelF, family 4 now contains both α - and β -glycoside-specific enzymes. In this context, family 4 is the exception among the glycosylhydrolase families, whose members invariably display either α or β specificity (see Tables 1 in references 14 and 20).

Catalytic mechanism(s) of family 4 glycosylhydrolases. Substrate cleavage by glycosylhydrolases occurs via one of two catalytic mechanisms that proceed with either inversion or retention of configuration at the anomeric center (7, 20, 36, 44, 47, 48). It is believed that the same mechanism is used by all members of a given family. The catalytic mechanism for members of family 4 has not yet been determined, but results obtained by site-directed mutagenesis of GlvA from *B. subtilis* identify Glu¹¹¹ and Glu³⁵⁹ as the probable proton donor and nucleophile or base, respectively (40). Comparative alignment of amino acid sequences (Fig. 3) shows that the two catalytic residues and the putative NAD⁺-binding domain of GlvA (P- α -glucosidase) are also positionally conserved in CelF (P- β -glucosidase) and MalH (P- α -glucosidase). Crystallographic data obtained from other glycosylhydrolases suggests that for "retaining" enzymes, the average distance between the carboxylate moieties of the two catalytic residues is 4.8 to 5.3 Å. For those enzymes in which hydrolysis proceeds with "inversion," the carboxylate groups are usually 9.0 to 9.5 Å apart (20, 45). We have recently reported the crystallization and preliminary X-ray analysis of GlvA (43), and crystallization trials for CelF are in progress (43a). Solution of the structures of GlvA and CelF will confirm (or refute) the active-site location for the two glutamyl residues, and their distance of separation may establish the catalytic mechanism(s) for the two P-glycosylhydrolases. It will be of interest to learn if CelF and GlvA use the same mechanism or whether family 4 of the glycosylhydrolase superfamily is exceptional in containing both inverting and retaining enzymes.

ACKNOWLEDGMENTS

We thank Saul Roseman, Nemat Keyhani, Edith Wolff, and Jack London for their interest and helpful discussions. We thank Nga Nguyen and Lewis Pannell for providing microsequence and mass spectrometry data, respectively. We acknowledge the assistance of Ann Ginsburg in analysis and interpretation of sedimentation data and thank Jack Folk for the synthesis of chromogenic and fluorogenic substrates.

REFERENCES

- 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.
- Amster-Choder, O., F. Houman, and A. Wright. 1989. Protein phosphorylation regulates transcription of the β -glucoside utilization operon in *E. coli*. *Cell* **58**:847–855.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1992. Short protocols in molecular biology, 2nd ed., John Wiley & Sons, Inc., New York, N.Y.
- 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.
- Bouma, C. L., J. Reizer, A. Reizer, S. A. Robrish, and J. Thompson. 1997. 6-Phospho- α -D-glucosidase from *Fusobacterium mortiferum*: cloning, expression, and assignment to family 4 of the glycosylhydrolases. *J. Bacteriol.* **179**:4129–4137.
- Burstein, C., and A. Kepes. 1971. The α -galactosidase from *Escherichia coli* K12. *Biochim. Biophys. Acta* **230**:52–63.
- Davies, G., and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**:853–859.
- El Hassouni, M., B. Henrissat, M. Chippaux, and F. Barras. 1992. Nucleotide sequences of the *arb* genes, which control β -glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli* *bgl* operon and evidence for a new β -glycosylhydrolase family including enzymes from eubacteria, archaeobacteria, and humans. *J. Bacteriol.* **174**:765–777.
- Fox, C. F., and G. Wilson. 1968. The role of a phosphoenolpyruvate-dependent kinase system in β -glucoside catabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **59**:988–995.
- Hall, B. G., P. W. Betts, and M. Krickler. 1986. Maintenance of the cellobiose utilization genes of *Escherichia coli* in a cryptic state. *Mol. Biol. Evol.* **3**:389–402.
- Hall, B. G., and L. Xu. 1992. Nucleotide sequence, function, activation, and evolution of the cryptic *asc* operon of *Escherichia coli* K12. *Mol. Biol. Evol.* **9**:688–706.
- Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**:309–316.
- Henrissat, B., and A. Bairoch. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**:781–788.
- Henrissat, B., and G. Davies. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**:637–644.
- Keyhani, N. O., and S. Roseman. 1997. Wild-type *Escherichia coli* grows on the chitin disaccharide, N,N'-diacetylchitobiose, by expressing the *cel* operon. *Proc. Natl. Acad. Sci. USA* **94**:14367–14371.
- Krickler, M., and B. G. Hall. 1984. Directed evolution of cellobiose utilization in *Escherichia coli* K12. *Mol. Biol. Evol.* **1**:171–182.
- Krickler, M., and B. G. Hall. 1987. Biochemical genetics of the cryptic gene system for cellobiose utilization in *Escherichia coli* K12. *Genetics* **115**:419–429.
- Lieli, W., et al. Personal communication.
- Liljeström, P. L., and P. Liljeström. 1987. Nucleotide sequence of the *melA* gene, coding for α -galactosidase in *Escherichia coli* K12. *Nucleic Acids Res.* **5**:2213–2220.
- Mahadevan, S., A. E. Reynolds, and A. Wright. 1987. Positive and negative regulation of the *bgl* operon in *Escherichia coli*. *J. Bacteriol.* **169**:2570–2578.
- McCarter, J. D., and S. G. Withers. 1994. Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**:885–892.
- Meadow, N. D., D. K. Fox, and S. Roseman. 1990. The bacterial phosphoenolpyruvate: glucose phosphotransferase system. *Annu. Rev. Biochem.* **59**:497–542.
- Nagao, Y., T. Nakada, M. Imoto, T. Shimamoto, S. Sakai, and T. Tsuchiya. 1988. Purification and analysis of the structure of α -galactosidase from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **151**:236–241.
- Palmer, R. E., and R. A. Anderson. 1972. Cellobiose metabolism in *Aerobacter aerogenes*. II. Phosphorylation of cellobiose with adenosine 5'-triphosphate by a β -glucoside kinase. *J. Biol. Chem.* **247**:3415–3419.
- Parker, L. L., and B. G. Hall. 1988. A fourth *Escherichia coli* gene system with the potential to evolve β -glucoside utilization. *Genetics* **119**:485–490.
- Parker, L. L., and B. G. Hall. 1990. Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. *Genetics* **124**:455–471.
- Parker, L. L., and B. G. Hall. 1990. Mechanisms of activation of the cryptic *cel* operon of *Escherichia coli* K12. *Genetics* **124**:473–482.
- Postma, P. W., J. W. Lengeler, and G. R. Jacobson. 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **57**:543–594.
- Prasad, I., and S. Schaeffer. 1974. Regulation of the β -glucoside system in *Escherichia coli* K12. *J. Bacteriol.* **120**:638–650.
- Prasad, I., B. Young, and S. Schaeffer. 1973. Genetic determination of the constitutive biosynthesis of phospho- β -glucosidase A in *Escherichia coli*. *J. Bacteriol.* **114**:909–915.
- Reizer, J., A. Reizer, and M. H. Saier, Jr. 1990. The cellobiose permease of *Escherichia coli* consists of three proteins and is homologous to the lactose permease of *Staphylococcus aureus*. *Res. Microbiol.* **141**:1061–1067.
- Reversible Associations in Structural and Molecular Biology RASNB Website. 6 November 1996. [Online.] <http://www.bbri.org/RASNB/rasmb.html>. [20 September 1999, last date accessed.]
- Reynolds, A. E., J. Felton, and A. Wright. 1981. Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K12. *Nature* **293**:625–629.
- Saier, M. H., Jr., and J. Reizer. 1992. Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. *J. Bacteriol.* **174**:1433–1438.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Schnetz, K., and B. Rak. 1988. Regulation of the *bgl* operon of *Escherichia coli* by transcriptional antitermination. *EMBO J.* **7**:3271–3277.

35. **Schnetz, K., C. Toloczyki, and B. Rak.** 1987. β -Glucoside (*bgl*) operon of *Escherichia coli* K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. *J. Bacteriol.* **169**:2579–2590.
36. **Sinnott, M. L.** 1990. Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.* **90**:1171–1202.
- 36a. **SWISS-PROT Protein Sequence Data Bank Website.** Release 38.0, July 1999. [Online.] <http://www.expasy.ch/cgi-bin/lists?glycosid.txt>. [20 September 1999, last date accessed.]
37. **Swofford, D. L.** 1993. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 3.1. Illinois Natural History Survey, Champaign, Ill.
38. **Swofford, D. L.** 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates, Sunderland, Mass.
39. **Thompson, J., C. R. Gentry-Weeks, N. Y. Nguyen, J. E. Folk, and S. A. Robrish.** 1995. Purification from *Fusobacterium mortiferum* ATCC 25557 of a 6-phosphoryl-*O*- α -D-glucopyranosyl:6-phosphoglucohydrolase that hydrolyzes maltose-6-phosphate and related phospho- α -D-glucosides. *J. Bacteriol.* **177**:2505–2512.
40. **Thompson, J., A. Pikis, S. B. Ruvinov, B. Henrissat, H. Yamamoto, and J. Sekiguchi.** 1998. The gene *glvA* of *Bacillus subtilis* 168 encodes a metal-requiring, NAD(H)-dependent 6-phospho- α -glucosidase: assignment to Family 4 of the glycosylhydrolase superfamily. *J. Biol. Chem.* **273**:27347–27356.
41. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
42. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
43. **Varrot, A., H. Yamamoto, J. Sekiguchi, J. Thompson, and G. J. Davies.** 1999. Crystallization and preliminary X-ray analysis of the 6-phospho- α -glucosidase from *Bacillus subtilis*. *Acta Crystallogr.* **D55**:1212–1214.
- 43a. **Varrot, A., et al.** Unpublished data.
44. **White, A., and D. R. Rose.** 1997. Mechanism of catalysis by retaining β -glycosyl hydrolases. *Curr. Opin. Struct. Biol.* **7**:645–651.
45. **Wiesmann, C., W. Hengstenberg, and G. E. Schulz.** 1997. Crystal structures and mechanism of 6-phospho- β -galactosidase from *Lactococcus lactis*. *J. Mol. Biol.* **269**:851–860.
46. **Wilson, G., and C. F. Fox.** 1974. The β -glucosidase system of *Escherichia coli*. IV. Purification and properties of phospho- β -glucosidases A and B. *J. Biol. Chem.* **249**:5586–5598.
47. **Withers, S. G., and R. Aebersold.** 1995. Approaches to labelling and identification of active site residues in glycosidases. *Protein Sci.* **4**:361–372.
48. **Withers, S. G., R. A. J. Warren, I. P. Street, K. Rupitz, J. B. Kempton, and R. Aebersold.** 1990. Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a “retaining” glycosidase. *J. Am. Chem. Soc.* **112**:5887–5889.
49. **Witt, E., R. Frank, and W. Hengstenberg.** 1993. 6-Phospho- β -galactosidases of Gram-positive and 6-phospho- β -glucosidase B of Gram-negative bacteria: comparison of structure and function by kinetic and immunological methods and mutagenesis of the *lacG* gene of *Staphylococcus aureus*. *Protein Eng.* **6**:913–920.
50. **Zamyatnin, A. A.** 1984. Amino acid, peptide and protein volume in solution. *Annu. Rev. Biophys. Bioeng.* **13**:145–165.