Reconstitution of a Defunct Glycolytic Pathway via Recruitment of Ambiguous Sugar Kinases[†]

Brian G. Miller[‡,§] and Ronald T. Raines*[,‡,||]

Departments of Biochemistry and Chemistry, University of Wisconsin, Madison, Wisconsin 53706

Received April 5, 2005; Revised Manuscript Received June 7, 2005

ABSTRACT: During a recent investigation of the persistence of substrate ambiguity in contemporary enzymes, we identified three distinct ambiguous sugar kinases embedded within the modern *Escherichia coli* genome [Miller, B. G., and Raines, R. T. (2004) *Biochemistry 43*, 6387−6392]. These catalysts are the YajF, YcfX, and NanK polypeptides, all of which possess rudimentary glucokinase activities. Here, we report on the discovery of a fourth bacterial kinase with ambiguous substrate specificity. AlsK phosphorylates the glucose epimer, D-allose, with a k_{cat}/K_m value of 6.5×10^4 M^{-1} s^{-1}. AlsK also phosphorylates D-glucose, with a k_{cat}/K_m value that is 10^5-fold lower than the k_{cat}/K_m value displayed by native *E. coli* glucokinase. Overexpression of the *alsK* gene relieves the auxotrophy of a glucokinase-deficient bacterium, demonstrating that weak enzymatic activities derived from ambiguous catalysts can provide organisms with elaborated metabolic capacities. To explore how ambiguous catalysts are recruited to provide new functions, we placed the glucokinase-deficient bacterium under selection for growth at the expense of glucose. Under these conditions, the bacterium acquires a spontaneous mutation in the putative promoter region of the *yajF* gene, a locus previously shown to encode a sugar kinase with relaxed substrate specificity. The point mutation regenerates a consensus σ^{70} promoter sequence that leads to a 94-fold increase in the level of *yajF* expression. This increase provides sufficient glucokinase activity for reconstitution of the defunct glycolytic pathway of the bacterial auxotroph. Our current findings indicate that ambiguous enzymatic activities continue to play an important role in the evolution of new metabolic pathways, and provide insight into the molecular mechanisms that facilitate the recruitment of such catalysts during periods of natural selection.

Processes such as the acquisition of antibiotic resistance by pathogenic bacteria, the bioremediation of man-made pollutants by soil organisms, and the synthesis of unusual secondary metabolites by microbes and plants often require the evolution of protein catalysts capable of catalyzing novel chemical reactions. The mechanisms that facilitate the generation of enzymes with new or expanded catalytic repertoires, however, remain poorly understood. One possible source of new enzyme activities is promiscuous catalysts (*1*−*3*). Promiscuous catalysts are enzymes that possess the ability to catalyze multiple chemical reactions that often share a common intermediate or related mechanistic step (*4*−*7*). Biological catalysts that display significant flexibility in substrate recognition also represent a potential source of new enzyme activities. The capacity to catalyze the transformation of multiple, structurally related substrates is termed substrate ambiguity, and past investigators have suggested that such a feature may have been commonplace in primordial catalysts (*8*). Despite the fact that substrate ambiguity may continue to play a role in the functional diversification of proteins and metabolic pathways, conventional wisdom suggests that modern enzymes possess highly refined substrate specificities (*9*, *10*).

The potential of ambiguous catalysts to yield unique proteins led us to investigate the persistence of substrate ambiguity in contemporary enzymes. From within the genome of a common bacterium, we recently identified three distinct polypeptides that possess ambiguous substrate specificities (*11*). Notably, all three proteins belong to the same ROK[1] family of bacterial proteins (*12*), and include one previously identified enzyme, NanK, and two functionally uncharacterized polypeptides, YajF and YcfX. We found that all three enzymes could phosphorylate several structurally related metabolites, including the common carbohydrate

[†] B.G.M. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-1703-02). This work was supported, in part, by National Institutes of Health Grant GM44783. The Biophysics Instrumentation Facility was established with support from the University of Wisconsin-Madison, National Science Foundation Grant BIR-951257, and National Institutes of Health Grant S10 RR13790.

* To whom correspondence should be addressed: Department of Biochemistry, University of Wisconsin, 433 Babcock Dr., Madison, WI 53706-1544. Telephone: (608) 262-8588. Fax: (608) 262-3453. E-mail: raines@biochem.wisc.edu.

[‡] Department of Biochemistry.

[§] Current address: Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306.

[||] Department of Chemistry.

[1] Abbreviations: AlsK, allokinase; EDTA, ethylenediaminetetraacetic acid; Glk, glucokinase; IPTG, isopropyl β-D-1-thiogalactopyranoside; NADH, β-nicotinamide adenine dinucleotide, reduced form; NADP, β-nicotinamide adenine dinucleotide phosphate, oxidized form; NanK, *N*-acetyl-D-mannosamine kinase; OD, optical density; *pts*, phosphoenolpyruvate:carbohydrate phosphotransferase system; ROK, repressor, open reading frame, kinase; SDS, sodium dodecyl sulfate; TLC, thin-layer chromatography; Tris, tris(hydroxymethyl)aminomethane.

10.1021/bi0506268 CCC: $30.25 © 2005 American Chemical Society
Published on Web 07/22/2005

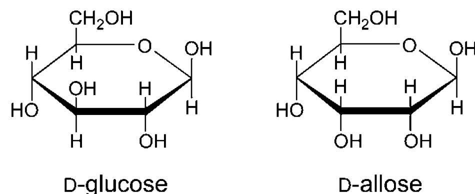

FIGURE 1: Structures of D-glucose and D-allose.

D-glucose. The glucokinase activity of each enzyme, however, was significantly lower than that of the endogenous bacterial glucokinase, with k_{cat}/K_m values ranging from 10^2 to 10^3 M^{-1} s^{-1} (*13*). Despite the fact that each possesses only rudimentary glucokinase activities, overproduction of YajF, YcfX, or NanK enabled the survival of a glucokinase-deficient bacterium on glucose minimal media. These results demonstrate that weak enzymatic activities derived from ambiguous catalysts can supply missing catalytic activities within living cells, and emphasize the importance of searching for such activities within the genomes of contemporary organisms.

In addition to the YajF, YcfX, and NanK polypeptides, *Escherichia coli* K-12 contains a fourth member of the ROK family of proteins with a putative sugar kinase activity, AlsK. On the basis of our previous findings, we decided to investigate the possibility that the AlsK polypeptide also manifests relaxed substrate specificity. Herein, we show that AlsK is capable of phosphorylating both D-allose and D-glucose (Figure 1). As with the *yajF*, *ycfX*, and *nanK* genes, overexpression of *alsK* from a powerful, extrachromosomal promoter enables growth of a glucokinase-deficient bacterium on glucose minimal medium. To examine how nature might recruit such ambiguous catalysts to provide new enzymatic activities, we place a glucokinase-deficient bacterium under selection for growth at the expense of glucose. We observe activation of the *yajF* gene, via acquisition of a single base pair change in the promoter region of this ambiguous kinase. This spontaneous mutation, which had been observed previously in a mutant strain of *E. coli* that is incapable of fructose utilization (*14*), permits recruitment of the ambiguous YajF enzyme and allows reconstitution of the defunct glycolytic pathway. These results support earlier speculation that ambiguous activities can provide selective advantages to their host cells under conditions of extreme selective pressure. Our findings also indicate that ambiguous enzymatic activities persist in modern organisms, where they could provide fertile substrata for the natural evolution of proteins with new catalytic activities (*1*−*4*, *8*, *15*).

EXPERIMENTAL PROCEDURES

Enzyme Production and Purification. Genomic DNA from *E. coli* strain MC4100 (*16*) was prepared according to standard procedures (*17*). Briefly, 1.5 mL of a saturated overnight culture of MC4100 was transferred to a sterile tube and subjected to centrifugation for 2 min at 10000*g*. The supernatant was removed, and the resulting cell pellet was resuspended in TE buffer (400 μL), which was 10 mM Tris-HCl buffer at pH 7.5, containing EDTA (1.0 mM). SDS and proteinase K were added to final concentrations of 1% (w/v) and 2 mg/mL, respectively, and the resulting reaction mixture was incubated at 37 °C for 1 h. The reaction mixture was then transferred to a Phase Lock Gel Light 2 mL tube (Eppendorf, Hamburg, Germany), and the nucleic acids were extracted by the addition of 0.5 mL of a phenol/chloroform/isoamyl alcohol mixture (25:24:1), according to the manufacturer's instructions. The extraction process was repeated three times, and the nucleic acids were precipitated by addition of ice-cold ethanol (1.0 mL) and 5 M NaCl (25 μL). DNA was isolated via centrifugation at 12000*g* for 10 min, and the isolated pellet was dried briefly and then resuspended in TE buffer (100 μL) containing ribonuclease A (250 μg/mL). The resulting reaction mixture was incubated for 30 min at 37 °C to promote RNA degradation. Genomic DNA was precipitated by the addition of an equal volume of 2-propanol, followed by the addition of 7.5 M ammonium acetate at pH 4.6 to a final concentration of 0.15 M. Genomic DNA was isolated by centrifugation at 12000*g* for 10 min.

The *alsK* gene was amplified with PfuTurbo DNA polymerase (Stratagene, La Jolla, CA) from genomic DNA using oligonucleotides that incorporated *Nco*I and *Eag*I recognition sites at the 5′ and 3′ ends of the gene, respectively. The amplified product was treated to eliminate polymerase (*18*) and purified with the Promega PCR cleanup kit. Amplified product was digested with *Nco*I and *Eag*I for 2 h at 37 °C and ligated into appropriately digested pBGM101, as described previously (*11*). The resulting plasmid (pBGM101-*alsK*) contains the *alsK* coding sequence downstream of the T7 promoter and in-frame with a C-terminal hexahistidine-encoding sequence. Transformation of glucokinase-deficient strain BM5340(DE3) with pBGM101-*alsK*, followed by plating on glucose minimal M9 agar, verified the ability of the *alsK* gene to complement the glucokinase auxotrophy.

To prevent contamination from endogenous glucokinases, AlsK was produced in glucokinase-deficient *E. coli* strain BM5340(DE3). Cultures (0.75 L) were inoculated to an initial OD of 0.005 at 600 nm and grown at 37 °C in Luria-Bertani broth supplemented with ampicillin (150 μg/mL). IPTG was added to a final concentration of 0.8 mM when the OD at 600 nm reached 1.1, and growth was continued for an additional 2 h at 37 °C. Bacteria were harvested by centrifugation at 10000*g* for 10 min and resuspended to 1 g of cells (wet weight) per 8 mL of buffer A [75 mM sodium phosphate buffer (pH 7.6) containing glucose (50 mM), imidazole (25 mM), glycerol (5%, w/v), and β-mercaptoethanol (5 mM)]. Cell extracts were prepared by being passed through a French pressure cell at 16 000 psi. Intact cells and cell debris were removed by centrifugation at 75000*g* for 70 min, and the resulting supernatant was loaded onto nickel−nitrilotriacetic acid (Ni−NTA) affinity resin (Qiagen, Valencia, CA) that had been equilibrated with buffer A. The column was washed with 8 column volumes of buffer A containing 40 mM imidazole, and AlsK was eluted in buffer A containing 250 mM imidazole. Purified AlsK was dialyzed overnight at 4 °C against 12 L of 50 mM Tris-HCl buffer (pH 7.6) containing $MgSO_4$ (5 mM), glycerol (5%, w/v), and dithiothreitol (0.5 mM). Protein concentrations were estimated from the absorbance at 280 nm using a molar extinction coefficient (34 900 M^{-1} cm^{-1}) determined by the method of Edelhoch (*19*). The molecular mass of purified AlsK was verified by MALDI mass spectrometry on an Applied Biosystems Voyager 6133 spectrometer at the University of Wisconsin-Madison Biophysics Instrumentation Facility.

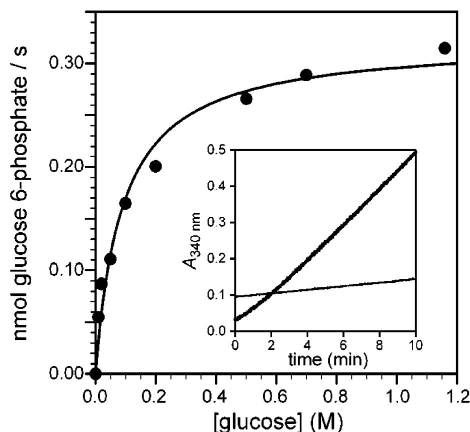

FIGURE 2: Rate of phosphorylation of D-glucose as catalyzed by AlsK. Data were obtained at 25 °C in 0.2−0.8 M Tris-HCl buffer at pH 7.6. The resulting kinetic parameters are listed in Table 1. The inset shows independent primary data for the glucokinase activity of AlsK at 0.40 M glucose (thick line) and background reduction of glucose by glucose-6-phosphate dehydrogenase (thin line).

Enzyme Assays. Glucokinase activity was measured spectrophotometrically at 340 nm by coupling the production of glucose 6-phosphate to the reduction of NADP via glucose-6-phosphate dehydrogenase. At the high concentrations of glucose needed to saturate AlsK, a small level of background NADP reduction was observed (Figure 2, inset). We tentatively assigned this activity to the ambiguous reduction of glucose by glucose-6-phosphate dehydrogenase, and the background rate of this reduction was subtracted from all kinetic measurements. Allokinase activity was determined by coupling production of Mg•ADP to oxidation of NADH via pyruvate kinase and lactate dehydrogenase, as previously described (*20*, *21*). Assays were conducted at 25 °C in 0.2−0.8 M Tris-HCl buffer at pH 7.6, and data were fitted to standard Michaelis−Menten kinetic equations for bisubstrate reactions. Kinetic constants reported in Table 1 are the average of values obtained from two independent preparations of the enzyme. The ability of allose to inhibit the glucokinase activity of AlsK was likewise assessed at three different glucose concentrations. The resulting apparent K_i value for this competing substrate was calculated from relative reaction velocities using standard kinetic equations.

Production of Allose 6-Phosphate by AlsK. To verify production of allose 6-phosphate, a 50 μL reaction mixture of 50 mM Tris-HCl buffer (pH 7.6) containing $MgCl_2$ (25 mM), D-allose (20 mM), ATP (20 mM), and purified AlsK (4 μM) was incubated for 20 min at 25 °C. An aliquot (5 μL) of this reaction mixture was spotted onto a normal phase Whatman Silica Gel TLC plate and developed in a solvent system comprised of butanol, ethanol, and H_2O (5:3:2). To visualize carbohydrates, plates were dipped in a solution of methanol containing concentrated H_2SO_4 (2%, v/v) and heated for 10 min at 120 °C. Unreacted allose and its phosphorylated product had R_f values of 0.5 and 0.1, respectively. Regions containing the phosphorylated product were scraped from untreated TLC plates, extracted with excess methanol, and evaporated to dryness. Negative ion electrospray mass spectrometry of the extracted compound verified the production of a product with a mass of 259 amu, which corresponds to that of allose 6-phosphate.

Isolation and Characterization of Revertants. A single colony of glucokinase-deficient BM5300 harboring plasmid pBGM101 (*11*) was inoculated into 7 mL of Luria-Bertani broth supplemented with ampicillin (150 μg/mL), and the culture was grown to saturation at 37 °C. A total of 1.8×10^9 colony-forming units of BM5300(pBGM101) was spread onto 20 M9 minimal plates supplemented with ampicillin (100 μg/mL), chloramphenicol (20 μg/mL), kanamycin (20 μg/mL), and glucose (0.1%, w/v). Following incubation for 9 days at 37 °C, approximately 750 small colonies were apparent on the glucose minimal plates. Eight clones were selected and restreaked onto glucose minimal plates to verify the Glk^+ phenotype. Four of these clones were chosen for further study, and each was inoculated into 5 mL of LB containing ampicillin (150 μg/mL), kanamycin (20 μg/mL), and chloramphenicol (20 μg/mL). Cultures were grown overnight at 37 °C, and cells were collected by centrifugation at 5000*g*. The supernatant was removed, and cell pellets were resuspended in 1.0 mL of lysis buffer, which was 100 mM Tris-HCl buffer (pH 7.6) containing $MgCl_2$ (10 mM), glycerol (10%, w/v), and dithiothreitol (1.0 mM). Crude cell extracts were prepared by sonication of resuspended cells at 80 W for 10 s, followed by centrifugation at 20000*g* for 20 min at 4 °C. Glucokinase activities from each cell lysate, including extracts prepared from unselected control BM5300-(pBGM101) cells, were measured spectrophotometrically, as described above. Genomic DNA was prepared from Glk^+ revertant cells using standard procedures (*16*), and the *yajF* gene, including ∼200 nucleotides upstream of the initiation codon, was amplified using oligonucleotides specific for this genomic locus. PCR products were purified with the Promega PCR cleanup kit, and sequencing was performed with the ABI Prism Big Dye Terminator Cycle Sequencing kit (Princeton Separations, Adelphia, NJ).

To gauge the level of transcriptional activation produced by the *yajF* promoter mutation, expression of the *E. coli glk* gene was placed under control of the wild-type and revertant *yajF* upstream sequence. A region encompassing 126 nucleotides upstream of the *yajF* initiation codon was PCR-amplified with oligonucleotides that incorporated *Xba*I and *Nco*I recognition sites at the 5′ and 3′ ends of the product, respectively. The amplified product was treated to eliminate polymerase (*18*) and purified with the Promega PCR cleanup kit. The amplified product was digested with *Xba*I and *Nco*I for 2 h at 37 °C and ligated into appropriately digested pBGM102, a derivative of pBGM101 harboring the *E. coli glk* gene (*11*). Plasmid pBGM129 contains the *glk* gene downstream of the wild-type *yajF* sequence, and pBGM130 contains the *glk* coding sequence downstream of the mutant *yajF* sequence. Three colonies each of BM5340(pBGM129) and BM5340(pBGM130) were inoculated into 25 mL of LB supplemented with ampicillin (150 μg/mL), and cultures were grown at 37 °C until the OD at 600 nm reached 1.0. Growth was quenched on ice, cells were isolated by centrifugation at 7000*g*, and cell pellets were resuspended in 1.0 mL of lysis buffer. Crude cell extracts were prepared via sonication, and glucokinase activity was determined spectrophotometrically, as described above.

RESULTS

Substrate Ambiguity of AlsK. In a previous study, we identified three different bacterial sugar kinases that possess

Table 1: Kinetic Parameters for the Phosphorylation of Carbohydrates by Glk and AlsK[a]

parameter	Glk[b]	AlsK	
	D-glucose	D-glucose	D-allose
k_{cat} (s^{-1})	410 ± 30	1.5 ± 0.3	17 ± 5
$K_{m,sugar}$ (M)	(7.6 ± 0.4) × 10^{-5}	0.10 ± 0.01	(2.6 ± 0.4) × 10^{-4}
$K_{m,ATP}$ (M)	(2.6 ± 0.2) × 10^{-4}	(6.9 ± 0.7) × 10^{-3}	(4.9 ± 0.7) × 10^{-4}
$k_{cat}/K_{m,sugar}$ (M^{-1} s^{-1})	(5.4 ± 0.7) × 10^{6}	15 ± 5	(6.5 ± 0.9) × 10^{4}

[a] Data were obtained at 25 °C in 0.2−0.8 M Tris-HCl buffer at pH 7.6. [b] From ref *11*.

sufficient substrate ambiguity to allow phosphorylation of glucose, in addition to phosphorylation of their physiological substrates (*11*). These kinases included the *E. coli* NanK, YajF, and YcfX polypeptides, each of which belongs to the ROK family of bacterial proteins (*12*). Interestingly, the only *E. coli* ROK family member with a predicted kinase activity that was not selected in our original complementation experiments was a polypeptide encoded by the *alsK* (*yjcT*) gene. A Clustal W alignment of the primary amino acid sequences of AlsK and *E. coli* Glk reveals a level of sequence identity of 17.5%. In addition, the sequence of the AlsK protein is 20% identical to that of NanK and 23% identical to those of both YajF and YcfX. On the basis of the sequence similarity between AlsK and the previously selected ROK family members, we decided to investigate the potential of AlsK to catalyze the phosphorylation of glucose. We also investigated whether overproduction of the *alsK* gene in the glucokinase-deficient bacteria, BM5340, allowed complementation of the auxotrophy of this strain.

The *E. coli* alsK gene encodes an enzyme with rudimentary glucokinase activity (Figure 2). To ensure that this activity did not arise from contaminating endogenous glucokinases, we produced and purified AlsK from glucokinase-deficient BM5340(DE3) cells as a C-terminal, hexahistidine-tagged polypeptide. Under saturating concentrations of glucose and ATP, AlsK catalyzes formation of glucose 6-phosphate with a turnover number, k_{cat}, of 1.5 s^{-1} (Table 1). The K_m value of AlsK for substrate glucose is 0.10 M, and the $k_{cat}/K_{m,sugar}$ value for the AlsK-catalyzed phosphorylation of glucose is 15 M^{-1} s^{-1}. This value is nearly 6 orders of magnitude lower than the k_{cat}/K_m value (5.4 × 10^6 M^{-1} s^{-1}) displayed by Glk, the endogenous bacterial glucokinase (*13*). Despite the disparity of these second-order rate constants, overexpression of the *alsK* gene from a powerful extrachromosomal promoter permits growth of a glucokinase-deficient bacterium on glucose minimal medium. BM5340(DE3) cells transformed with plasmid pBGM101-alsK, which harbors the *alsK* gene downstream of the T7 promoter, produce colonies on glucose-supplemented M9 minimal medium after incubation for just 2 days at 37 °C.

AlsK Is a Functional Allokinase. AlsK also catalyzes the efficient transfer of the γ-phosphoryl group of ATP to the glucose epimer, D-allose (Figure 3). Formation of allose 6-phosphate following incubation of AlsK with ATP and D-allose was confirmed by mass spectrometry of the isolated reaction product. The value of k_{cat} for the AlsK-catalyzed phosphorylation of allose is 17 s^{-1}, and the K_m value for substrate allose is 2.6 × 10^{-4} M (Table 1). Similarly, the apparent K_i value of allose acting as a competitive inhibitor of the glucokinase activity of AlsK is 3.3 × 10^{-4} M, suggesting that the same active site catalyzes the phosphorylation of both sugar substrates. The K_m value of ATP, measured during allose phosphorylation, is 4.9 × 10^{-4} M, whereas the K_m value of ATP, measured during glucose phosphorylation, is 6.9 × 10^{-3} M. A comparison of relative k_{cat}/K_m values for the phosphorylation of allose and glucose demonstrates that the affinity of the enzyme for the altered substrate in the transition state for allose phosphorylation exceeds its affinity for the altered substrate in the transition state for glucose phosphorylation by 5 kcal/mol.

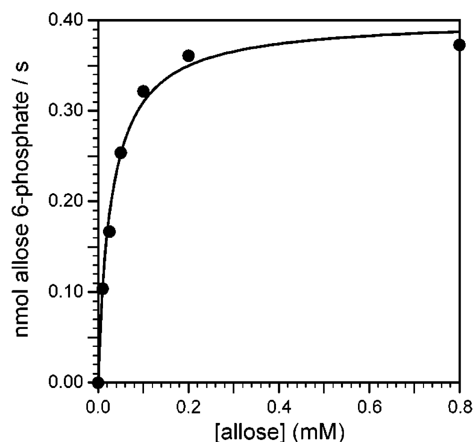

FIGURE 3: Rate of phosphorylation of D-allose as catalyzed by AlsK. Data were obtained at 25 °C in 0.2−0.8 M Tris-HCl buffer at pH 7.6. The resulting kinetic parameters are listed in Table 1.

Recruitment of the Ambiguous YajF Kinase via Spontaneous Mutation. E. coli strain BM5300 harbors chromosomal deletions in two glucokinase-encoding genetic loci, *glk* and *pts* (*11*, *22*, *23*). This organism cannot initiate glycolysis, and is unable to survive when glucose is supplied as a sole source of carbon and energy. Production of the glucose facilitator protein of *Zymomonas mobilis* from plasmid pBGM101 restores phosphorylation-independent glucose transport activity to BM5300 cells (*24*). When large numbers of glucokinase-deficient BM5300(pBGM101) cells were plated on glucose minimal medium, spontaneous glucose-metabolizing revertants appeared after 9 days of growth at 37 °C. Reversion of the Glk^- phenotype of BM5300-(pBGM101) occurred at a frequency of $\sim 10^{-7}$. Crude cell extracts from four different Glk^+ revertants showed a 7-fold increase in glucokinase activity over the background levels of activity present in extracts of BM5300(pBGM101) parent cells (Table 2). Thus, the capacity of BM5300 revertants to survive on glucose minimal medium is due to an increased level of glucokinase activity that likely results from a rare mutational event. Previous studies have identified three ambiguous sugar kinases that possess low levels of glucokinase activity (*11*). Increased levels of expression of any one of these three gene products could increase the level of glucokinase activity to an extent sufficient to restore the glycolytic pathway of the glucokinase-deficient host cell. To

Table 2: Relative Glucokinase Activities Present within Bacterial Revertants

cell line	relative glucokinase activity[a]
control BM5300(pBGM101)	1.0
revertant BM5301(pBGM101)	6.3
revertant BM5302(pBGM101)	4.6
revertant BM5303(pBGM101)	6.3
revertant BM5306(pBGM101)	10
BM5340(pBGM129)[b]	4.8
BM5340(pBGM130)[b]	450

[a] Normalized to the total protein concentration. [b] pBGM129 harbors the endogenous glucokinase gene (*glk*) under control of the wild-type *yajF* promoter, whereas pBGM130 contains the mutant *yajF* promoter sequence.

identify the genetic lesion responsible for the increased activity, we sequenced the chromosomal DNA of revertant cells in regions surrounding the *yajF*, *ycfX*, and *nanK* genes.

The genetic basis of the Glk$^+$ phenotype of BM5300-(pBGM101) revertants was mapped to a region in the 5′ untranslated region of the *yajF* gene, a locus previously shown to encode a sugar kinase with ambiguous substrate specificity (*11*). Revertant cells were found to possess a single transversion in the putative −35 promoter region of the *yajF* gene. The mutation involves the replacement of a cytidine nucleotide with an adenosine nucleotide, and creates a consensus σ^{70} promoter with the TTGACA sequence (Figure 4) (*25*, *26*). The failure to observe activation of any other ambiguous glucokinase in BM5300(pBGM101) revertant bacteria suggests that if such mutations are possible, they occur at a much lower frequency than the *yajF* promoter mutation. Because endogenous levels of YcfX, NanK, or other ambiguous kinases likely contribute to the background glucokinase activity observed in control BM5300(pBGM101) cells (Table 2), we decided to quantify the increased level of *yajF* expression afforded by the promoter mutation in the following manner. Expression of the *E. coli glk* gene was placed under control of either the wild-type *yajF* upstream sequence on the multicopy pBGM129 plasmid or the mutant *yajF* sequence on the multicopy pBGM130 plasmid. The glucokinase activities of cellular extracts from BM5340 cells harboring either pBGM129 or pBGM130 were determined spectrophotometrically and normalized to total protein content. As listed in Table 2, the mutation of C to A in the putative −35 promoter sequence of the *yajF* locus results in a 94-fold increase in the level of expression of the downstream gene product. This degree of *yajF* derepression provides sufficient glucokinase activity to enable growth of revertant BM5300(pBGM101) cells on glucose minimal media.

DISCUSSION

The glucokinase activity of AlsK described herein appears to originate from substrate ambiguity, as do the glucokinase activities of three other *E. coli* ROK family members, YcfX, YajF, and NanK (*11*). Flexibility in substrate recognition appears to be a common feature of the ROK scaffold, as indicated by the finding that members of this protein family include numerous sugar kinases and various carbohydrate responsive transcriptional repressors (*12*). Accordingly, this enzymic architecture could prove to be especially useful in the design of ligand receptors, or in the evolution of catalysts for the transfer of phosphoryl groups to various small molecules (*27*−*29*). The glucokinase activity of AlsK, with a k_{cat}/K_m value of 15 M^{-1} s^{-1}, is more than 10-fold lower than the glucokinase activities of the three previously selected enzymes. Nevertheless, overexpression of the *alsk* gene permits rapid growth of the glucokinase-deficient BM5340 bacterial strain. This observation suggests that BM5340 might provide a useful tool in the search for nonproteinogenic, glucose-phosphorylating catalysts.

The *E. coli* AlsK polypeptide also functions as an efficient allokinase. Initially annotated as a putative allokinase by Park and co-workers, the *alsK* gene represents the terminal open reading frame of an operon that is associated with allose metabolism by *E. coli* K-12 (*30*). The sequence similarity of AlsK with other sugar kinases in the ROK family provided evidence that *alsK* encodes an enzyme with allokinase activity. The *alsK* gene was found to be dispensable for allose utilization, however, and subsequent studies found that regulation of *alsK* expression was independent of the remaining components of the allose operon (*30*, *31*). Furthermore, expression of the *alsK* gene from a multicopy plasmid reportedly did not lead to an increased level of allokinase activity (*31*). Together, these findings led Hove-Jensen and co-workers to reject the functional assignment of allokinase activity to AlsK. To reconcile these apparent contradictions, we characterized the allokinase activity of highly purified, recombinant AlsK. Our results presented here demonstrate unequivocally that AlsK catalyzes the efficient phosphorylation of D-allose with ATP as the phosphoryl donor. Furthermore, the relatively low allose K_m value displayed by AlsK suggests that this glucose epimer is the true physiological substrate of AlsK.

Loss of repressive control over ambiguous gene products is a common first step in the acquisition of new metabolic capabilities by microorganisms (*32*). For example, mutations that permit the constitutive synthesis of ribitol dehydrogenase enable *Aerobacter aerogenes* to flourish on the unnatural sugars, xylitol and L-arabitol (*33*, *34*). In the studies presented here, recruitment of the ambiguous YajF sugar kinase requires mutational activation of the *yajF* locus, a gene characterized by a suboptimal promoter region and the rare GTG initiation codon (*14*, *35*). N-Terminal sequence analysis of purified YajF has shown that the GTG codon is the true translational start site, ruling out any possibility of post-translational processing of the YajF polypeptide (*11*). The presence of a weak promoter and a rare initiation codon suggests that *yajF* may be a cryptic gene whose expression requires the reconstitution of a consensus −35 promoter sequence. Cryptic genes are DNA sequences not normally expressed during the routine lifetime of an organism (*36*−*38*). Unlike pseudogenes, which have become nonfunctional as a result of genetic drift, cryptic genes can produce fully functional polypeptides when activated by a mutational event. The activation of cryptic genes during periods of natural selection has been observed in several bacteria, where their activities provide elaborated metabolic potentials or produce altered growth phenotypes (*36*, *38*). For example, the mutational correction of a natural frameshift in the *E. coli ilvG* gene results in expression of an α-acetohydroxyacid synthase isozyme that is insensitive to feedback inhibition by valine (*39*). Correction of the *ilvG* frameshift in revertant bacteria eliminates the valine-resistant growth phenotype that

yajF genomic locus

-35 -10 1 rbs Met Arg Ile

ttgAcaatccggttgtccgtctctacgctattgatattgaaaaaaataaggagagtacc *gtg cgt ata*

BM5300 (*glk*, Δ*pts*)
Glk+ revertants

FIGURE 4: Reversion of the glucokinase auxotrophy of strain BM5300 via spontaneous mutation of a cytidine nucleotide to an adenosine nucleotide (colored red) in the −35 promoter region of the *yajF* locus. This mutation regenerates a consensus σ^{70} promoter sequence, resulting in a 94-fold increase in the level of YajF production. The putative −35, −10, and +1 sites and the ribosome binding site (rbs) are indicated by lines and labels; the first three codons of the *yajF* gene, including the rare GTG initiation codon, are in italic type.

is characteristic of wild-type *E. coli* (*40*). Similarly, activation of the cryptic *bgl* operon enables certain microorganisms to survive on a variety of structurally diverse β-glucosides (*41*−*43*). The ability of *E. coli* to utilize citrate as a sole carbon and energy source requires a pair of mutations in the *citA* and *citB* genes, which activate a cryptic, ancestral citrate transport system (*44*).

The observed 94-fold increase in the level of *yajF* expression that accompanies a single base pair change in the −35 promoter region appears to be striking when compared with the more modest increases in the level of expression that often result from spontaneous alterations in promoter sequences. For example, several spontaneous mutations in the promoter region of the *E. coli* glucose-6-phosphate dehydrogenase coding sequence lead to increases in enzyme production of 3−18-fold (*45*, *46*). Similarly, spontaneous promoter mutations that activate a silent *gal* operon have been isolated in the bacterium *Streptococcus thermophilus* (*47*). These mutations, which belong to three distinct classes, produce a 5−7-fold increase in GalK activity in the uninduced state, and a 7−12-fold increase following induction with lactose or galactose. A survey of 54 naturally occurring *E. coli* promoter sequences revealed that a cytidine nucleotide is found in the fourth position of the −35 hexamer in approximately 20% of all promoter sequences (*48*). Indeed, cytidine is the second most frequently observed nucleotide at this position, behind only the consensus adenosine. Accordingly, the presence of a cytidine nucleotide in the fourth position of the hexameric −35 promoter region of the wild-type *yajF* gene does not appear to be sufficient to explain the cryptic nature of this locus. Interestingly, Gourse and co-workers found that mutation of the TTGTCA sequence of a wild-type rRNA promoter to the consensus TTGACA sequence resulted in a 24-fold increase in the level of expression of a downstream *lacZ* fusion operon (*49*). These findings suggest that the identity of single nucleotides within the σ^{70} −35 promoter region can have drastic effects upon the level of expression of downstream genes. The current observation of *yajF* activation via a "promoter-up" mutation echoes the work of Yanofsky and co-workers, who found that a single nucleotide in the promoter region of the *trp* operon of *Shigella dysenteriae* is responsible for the low levels of expression of tryptophan biosynthetic enzymes in this organism (*50*, *51*). Activation of the *trp* operon via spontaneous mutation of a thymidine nucleotide to a cytidine nucleotide in the −10 promoter region relieves the tryptophan auxotrophy of wild-type *S. dysenteriae*. Combined with these observations, our findings suggest that retention of suboptimal promoter sequences could be a common mechanism by which microorganisms suppress expression of rarely needed metabolic enzymes, but retain the ability to realize their rapid production during periods of intense selection.

It is worth noting that the precise *yajF* promoter mutation described herein has also been identified in an *E. coli* strain that is incapable of utilizing fructose via the endogenous phosphoenolpyruvate:carbohydrate phosphotransferase (PTS) system (*14*, *52*). Kornberg and co-workers found that acquisition of the same cytidine to adenosine transversion in the 5′ untranslated region of the *yajF* coding sequence provided PTS-deficient bacteria with sufficient fructokinase activity to permit survival on fructose minimal medium. Indeed, the fructokinase activity of YajF exceeds its glucokinase activity by nearly 2 orders of magnitude (*11*). Nature appears to use a similar adaptive mechanism to overcome two related but distinct selective pressures. Recruitment of the ambiguous YajF sugar kinase by mutational activation provides either fructokinase- or glucokinase-deficient bacteria with the ability to flourish on fructose or glucose, respectively. In both cases, the substrate ambiguity of YajF provides metabolic flexibility to the host organism.

Current estimates indicate that the modern *E. coli* genome harbors coding sequences for approximately 4300 gene products (*53*). Our results show that at least five different *E. coli* polypeptides are capable of performing the same chemical transformation, the phosphorylation of glucose. Thus, greater than 0.1% of all genomically encoded *E. coli* proteins possess sufficient glucokinase activity to enable survival of a glucokinase-deficient organism when their encoding genes are overexpressed at high levels. The extent to which such functional redundancy is widespread in modern metabolic pathways remains to be explored. It is possible that the sugar kinases that constitute the ROK superfamily are unique in their ability to catalyze each other's reactions at a low level. The possibility that substrate ambiguity is a ubiquitous feature of contemporary catalysts, however, raises

interesting possibilities for the identification of new enzyme activities from existing polypeptide scaffolds.

ACKNOWLEDGMENT

We are grateful to L. D. Lavis for assistance with characterization of allose 6-phosphate and S. M. Fuchs, J. Kalia, J. E. Lee, B. D. Smith, and Profs. Richard L. Gourse and W. W. Cleland for valuable discussions.

REFERENCES

1. O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities, *Chem. Biol. 6*, R91−R105.
2. James, L. C., and Tawfik, D. S. (2001) Catalytic and binding poly-reactivities shared by two unrelated proteins: The potential role of promiscuity in enzyme evolution, *Protein Sci. 10*, 2600−2607.
3. Copley, S. D. (2003) Enzymes with extra talents: Moonlighting functions and catalytic promiscuity, *Curr. Opin. Chem. Biol. 7*, 265−272.
4. Babbitt, P. C., and Gerlt, J. A. (1997) Understanding enzyme superfamilies: Chemistry as the fundamental determinant in the evolution of new catalytic activities, *J. Biol. Chem. 272*, 30591−30594.
5. Joerger, A. C., Mayer, A., and Fersht, A. R. (2003) Mimicking natural evolution in vitro: An *N*-acetylneuraminate lyase mutant with an increased dihydropicolinate synthase activity, *Proc. Natl. Acad. Sci. U.S.A. 100*, 5694−5699.
6. Yang, K., and Metcalf, W. M. (2004) A new activity for an old enzyme: *Escherichia coli* bacterial alkaline phosphatase is a phosphite-dependent hydrogenase, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7919−7924.
7. Gerlt, J. A., Babbitt, P. C., and Rayment, I. (2005) Divergent evolution in the enolase superfamily: The interplay of mechanism and specificity, *Arch. Biochem. Biophys. 433*, 59−70.
8. Jensen, R. A. (1976) Enzyme recruitment in evolution of new function, *Annu. Rev. Microbiol. 30*, 409−425.
9. Fischer, E. (1894) Einfluss der configuration auf die wirkung derenzyme, *Ber. Dtsch. Chem. Ges. 27*, 2985−2993.
10. Koshland, D. E., Jr. (1958) Application of a theory of enzyme specificity to protein synthesis, *Proc. Natl. Acad. Sci. U.S.A. 44*, 98−104.
11. Miller, B. G., and Raines, R. T. (2004) Identifying latent enzyme activities: Substrate ambiguity within modern bacterial sugar kinases, *Biochemistry 43*, 6387−6392.
12. Titgemeyer, F., Reizer, J., Reizer, A., and Saier, M. H. (1994) Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria, *Microbiology 140*, 2349−2354.
13. Meyer, D., Schneider-Fresenius, C., Horlacher, R., Peist, R., and Boos, W. (1997) Molecular characterization of glucokinase from *Escherichia coli* K-12, *J. Bacteriol. 179*, 1298−1306.
14. Sproul, A. A., Lambourne, L. T. M., Jean-Jacques, J., and Kornberg, H. L. (2001) Genetic control of manno(fructo)kinase activity in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A. 98*, 15257−15259.
15. Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions, *Nat. Genet. 37*, 73−76.
16. Casadaban, M. J. (1976) Transposition and fusion of the lac genes to selected promoters in *Escherichia coli* using bacteriophage λ and μ, *J. Mol. Biol. 104*, 541−555.
17. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1994) *Current Protocols in Molecular Biology*, John Wiley & Sons, Boston.
18. Matsumura, I., and Ellington, A. D. (2001) Mutagenic PCR of protein-encoding genes for *in vivo* evolution, *Methods Mol. Biol. 182*, 261−269.
19. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine residues in proteins, *Biochemistry 6*, 1948−1954.
20. Liu, F., Dong, Q., Meyers, A. M., and Fromm, H. J. (1991) Expression of human brain hexokinase in *Escherichia coli*: Purification and characterization of the expressed enzyme, *Biochem. Biophys. Res. Commun. 177*, 305−311.
21. Sabater, B., Sebastian, J., and Asensio, C. (1972) Identification and properties of an inducible mannokinase from *Streptomyces violaceoruber*, *Biochim. Biophys. Acta 284*, 406−414.
22. Curtis, S. J., and Epstein, W. (1975) Phosphorylation of D-glucose in *Escherichia coli* mutants defective in glucosephosphotransferase, mannosephosphotransferase, and glucokinase, *J. Bacteriol. 122*, 1189−1199.
23. Postma, P. W., Lengeler, J. W., and Jacobson, G. R. (1996) Phosphoenoylpyruvate:carbohydrate phosphotransferase systems in *Escherichia coli* and *Salmonella*, in *Cellular and Molecular Biology* (Neidhardt, F. C., Curtiss, R., and Lin, E. C., Eds.) American Society for Microbiology Press, Washington, DC.
24. Parker, C., Barnell, W. O., Snoepp, J. L., Ingram, L. O., and Conway, T. (1995) Characterization of the *Zymomonas mobilis* glucose facilitator gene product (glf) in recombinant *Escherichia coli*: Examination of transport mechanism, kinetics, and the role of glucose in glucose transport, *Mol. Microbiol. 15*, 795−802.
25. Harley, C. B., and Reynolds, R. P. (1987) Analysis of *E. coli* promoter sequences, *Nucleic Acids Res. 15*, 2343−2361.
26. Lisser, S., and Margalit, H. (1993) Compilation of *E. coli* mRNA promoter sequences, *Nucleic Acids Res. 21*, 1507−1516.
27. Westheimer, F. H. (1987) Why nature chose phosphates, *Science 1987*, 1173−1178.
28. Hoffmeister, D., Yang, J., Liu, L., and Thorson, J. S. (2003) Creation of the first anomeric D/L sugar kinase by means of directed evolution, *Proc. Natl. Acad. Sci. U.S.A. 100*, 13184−13189.
29. Mizanur, R. M., Zea, C. J., and Pohl, N. L. (2004) Unusually broad substrate tolerance of a heat-stable archaeal sugar nucleotidyltransferase for the synthesis of sugar nucleotides, *J. Am. Chem. Soc. 126*, 15993−15998.
30. Kim, C., Song, S., and Park, C. (1997) The D-allose operon of *Escherichia coli* K-12, *J. Bacteriol. 179*, 7631−7637.
31. Poulsen, T., Chang, Y.-Y., and Hove-Jensen, B. (1999) D-Allose catabolism of *Escherichia coli*: Involvement of *alsI* and regulation of *als* regulon expression by allose and ribose, *J. Bacteriol. 181*, 7126−7130.
32. Hegeman, G. D., and Rosenberg, S. L. (1970) The evolution of bacterial enzyme systems, *Annu. Rev. Microbiol. 24*, 429−462.
33. Lerner, S. A., Wu, T. T., and Lin, E. C. C. (1964) Evolution of a catabolic pathway in bacteria, *Science 146*, 1313−1315.
34. Mortlock, R. P., Fossitt, D. D., and Wood, W. A. (1965) A basis for the utilization of unnatural pentoses and pentitols by *Aerobacter aerogenes*, *Proc. Natl. Acad. Sci. U.S.A. 54*, 572−579.
35. Schneider, T. D., Stormo, G. D., Gold, L., and Ehrenfeucht, A. (1986) Information content of binding sites on nucleotide sequences, *J. Mol. Biol. 188*, 415−431.
36. Hall, B. G., Yokoyama, S., and Calhoun, D. H. (1983) Role of cryptic genes in microbial evolution, *Mol. Biol. Evol. 1*, 109−124.
37. Li, W.-H. (1984) Retention of cryptic genes in microbial populations, *Mol. Biol. Evol. 2*, 213−219.
38. Mukerji, M., and Mahadevan, S. (1997) Cryptic genes: Evolutionary puzzles, *J. Genet. 76*, 147−159.
39. Lawther, R., Calhoun, D. H., Adams, C. W., Hauser, C. A., Gray, J., and Hatfield, G. W. (1981) Molecular basis of valine resistance in *Escherichia coli* K-12, *Proc. Natl. Acad. Sci. U.S.A. 78*, 922−925.
40. Lawther, R. P., Calhoun, D. H., Gray, J., Adams, C. W., Hauser, C. A., and Hatfield, G. W. (1982) DNA sequence fine-structure analysis of *ilvG* ($IlvG^+$) mutations of *Escherichia coli* K-12, *J. Bacteriol. 149*, 294−298.
41. Schaeffler, S., and Mintzer, L. (1959) Acquisition of lactose fermenting properties by *Salmonellae*. I. Interrelationship between fermentation of cellobiose and lactose, *J. Bacteriol. 78*, 159−163.
42. Schaefler, S., and Schenkein, I. (1968) β-Glucoside permeases and phospho-β-glucosidases in *Aerobacter aerogenes*: Relationship with cryptic phospho-β-glucosidase in *Enterobacteriaceae*, *Proc. Natl. Acad. Sci. U.S.A. 59*, 285−292.
43. Reynolds, A., Felton, Y., and Wright, A. (1981) Insertion of DNA activates the cryptic *bgl* operon of *E.* coli K-12, *Nature 203*, 625−629.
44. Hall, B. G. (1982) Chromosomal mutation for citrate utilization in *Escherichia coli* K12, *J. Bacteriol. 151*, 269−273.
45. Fraenkel, D. G., and Banerjee, S. (1971) A mutation increasing the amount of constitutive enzyme in *Escherichia coli*, glucose 6-phosphate dehydrogenase, *J. Mol. Biol. 56*, 183−194.

46. Fraenkel, D. G., and Parola, A. (1972) "Up-promoter" mutations of glucose 6-phosphate dehydrogenase in *Escherichia coli*, *J. Mol. Biol. 71*, 107−111.
47. Vaughan, E. E., Van Den Bogaard, P. T. C., Catzeddu, P., Kuipers, O. P., and De Vos, W. M. (2001) Activation of silent *gal* genes in the *lac-gal* regulon of *Streptococcus thermophilus*, *J. Bacteriol. 183*, 1184−1194.
48. Siebenlist, U., Simpson, R. B., and Gilbert, W. (1980) *E. coli* RNA polymerase interacts homologously with two different promoters, *Cell 20*, 269−281.
49. Gaal, T., Barkei, J., Dickson, R. R., DeBoer, H. A., DeHaseth, P. L., Alavi, H., and Gourse, R. L. (1989) Saturation mutagenesis of an *Escherichia coli* rRNA promoter and initial characterization of promoter variants, *J. Bacteriol. 171*, 4852−4861.
50. Manson, M. D., and Yanofsky, C. (1976) Naturally occurring sites within the *Shigella dysenteriae* tryptophan operon severely limit tryptophan biosynthesis, *J. Bacteriol. 126*, 668−678.
51. Miozzari, G., and Yanofsky, C. (1978) Naturally occurring promoter down mutation: Nucleotide sequence of the *trp* promoter/operator/leader region of *Shigella dysenteriae* 16, *Proc. Natl. Acad. Sci. U.S.A. 75*, 5580−5584.
52. Kornberg, H. L., Lambourne, L. T. M., and Sproul, A. A. (2000) Facilitated diffusion of fructose via the phosphoenolpyruvate/glucose phosphotransferase system of *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A. 97*, 1808−1812.
53. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12, *Science 5331*, 1453−1474.

BI0506268