

# Identifying Latent Enzyme Activities: Substrate Ambiguity within Modern Bacterial Sugar Kinases<sup>†</sup>

Brian G. Miller<sup>‡</sup> and Ronald T. Raines<sup>\*,‡,§</sup>

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

Received March 24, 2004; Revised Manuscript Received April 16, 2004

**ABSTRACT:** The ability of enzymes to catalyze the transformation of multiple, structurally related substrates could empower the natural evolution of new catalytic functions. The prevalence of such substrate ambiguity in modern catalysts, however, is largely unknown. To search for ambiguous sugar kinases, we generated a bacterium incapable of performing the first step of the glycolytic pathway, the phosphorylation of glucose. This organism cannot survive with glucose as its sole source of carbon. Within its genome, we find three DNA sequences that, when transcribed from a powerful extrachromosomal promoter, can complement the auxotrophy of the organism. These sequences contain the *nanK*, *yajF*, and *ycfX* genes. In vitro, the NanK, YajF, and YcfX proteins function as rudimentary glucokinases with ambiguous substrate specificities, displaying  $k_{cat}/K_m$  values for the phosphorylation of glucose that are  $10^4$ -fold lower than the  $k_{cat}/K_m$  value of endogenous bacterial glucokinase. Our findings suggest that modern genomes harbor a wealth of latent enzyme activities and that extant metabolic pathways are equivocal, in contrast to their usual depiction.

Primordial enzymes likely possessed broad substrate specificities, a feature that enabled their host cells to survive under a variety of nutritional conditions (1). As life evolved, gene duplications promoted the refinement of individual enzyme characteristics, resulting in enhanced specificities and higher catalytic efficiencies (2–5). The selective advantage conferred by ambiguous progenitor catalysts contrasts markedly with the dangers associated with the occurrence of substrate ambiguity in modern biological systems. Kinases, for example, must demonstrate extreme selectivity in their choice of phosphoryl acceptors, despite the similarity of their reactions and an abundance of potential cellular substrates. Inadvertent phosphorylation of an unintended target can lead to severe biological consequences, including death (6). How prevalent, then, is substrate ambiguity today?

The ability of enzymes to discriminate between a myriad of potential substrates of similar structure has long been appreciated as a defining characteristic of biological catalysts (7, 8). Indeed, distinguishing between related ground-state and transition-state species during the course of a chemical transformation provides the differential affinity upon which catalysis relies. Nevertheless, a few modern enzymes are known to possess sufficient substrate ambiguity to allow the acceptance of alternative substrates (9). The capacity to

catalyze the turnover of multiple substrates could power the evolution of novel catalytic functions and new metabolic pathways. Hence, a general method to identify ambiguous catalysts could prove beneficial by providing starting points for the directed evolution of enzymes with novel functions.

When faced with an unusable chemical resource, bacteria can mutate to allow the production of enzymes capable of catalyzing the conversion of the unusable substrate into a preferable energy source (10–13). For example, mutations that result in the constitutive expression of ribitol dehydrogenase enable *Aerobacter aerogenes* to flourish on the unnatural pentitols, xylitol, and L-arabitol (10, 11). In some instances, the constitutive expression of an ambiguous catalyst is accompanied by duplication of the encoding gene (14).

Here, we probe natural events in the evolution of enzymic catalysts. Specifically, we create a strain of *Escherichia coli* that lacks its endogenous glucokinase gene and hence cannot grow with glucose as its sole carbon source. From “over-expression” libraries that encompass the entire genome of this strain, we identify three distinct genes that enable its growth on glucose minimal medium. We purify the encoded proteins, determine their kinetic parameters for the phosphorylation of a variety of carbohydrates, and find that each is a rudimentary glucokinase of ambiguous substrate specificity.

## EXPERIMENTAL PROCEDURES

**Strain and Plasmid Construction.** The principal *E. coli* K-12 strains used in this study are described in Table 1. Strain TP2811, bearing a kanamycin-resistance cassette in place of the deleted *pts*<sup>1</sup> operon, was provided by Antoine Danchin (Institut Pasteur, Paris, France) (16). Strain DM1000, harboring a Tn10(Cam) insertion in the glucokinase gene

<sup>†</sup> B.G.M. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-1703-02). This work was supported by National Institutes of Health Grant GM44783. The Biophysics Instrumentation Facility is supported by 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—Madison, 433 Babcock Drive, Madison, WI 53706-1544. Telephone: (608) 262-8588. Fax: (608) 262-3453. E-mail: Raines@biochem.wisc.edu.

<sup>‡</sup> Department of Biochemistry.

<sup>§</sup> Department of Chemistry.

Table 1: *E. coli* Strains Used in This Paper

strain	relevant genotype	source
MC4100	F <sup>-</sup> <i>araD139</i> Δ( <i>argF-lac</i> )U169 <i>flbB5301</i> <i>deoC1 relA1 rbsR rpsL150 fruA25 thi-1</i>	ref 15
DM1000	MC4100 <i>glk15::Tn10</i> (Cam)	ref 17
TP2811	F <sup>-</sup> , <i>xyl</i> , <i>argH1</i> , Δ <i>lacX74</i> , <i>aroB</i> , <i>ilvA</i> , Δ( <i>ptsH</i> , <i>ptsI</i> , <i>crr</i> ), Kan <sup>R</sup>	ref 16
BM5300	DM1000 Δ( <i>ptsH</i> , <i>ptsI</i> , <i>crr</i> ), Kan <sup>R</sup>	this paper
BM5340	λDE3 lysogen of BM5300	this paper

was supplied by Winfried Boos (University of Konstanz, Germany) (17). Generalized transduction with P1kc phage of the Δ(*ptsH*, *ptsI*, *crr*), Kan<sup>R</sup> locus of strain TP2811 into strain DM1000, produced strain BM5300 (18). The glucokinase deficiency of BM5300 was verified by the production of white colonies on MacConkey agar (19) supplemented with glucose (1% w/v) and by the inability of BM5300 to produce colonies on K10 (20) minimal agar containing glucose (0.1% w/v), thiamine (1.0 mg/L), chloramphenicol (25 μg/mL), and kanamycin (40 μg/mL). Lysogenization of BM5300 with the λDE3 prophage yielded strain BM5340-DE3 and enabled production of recombinant proteins from T7-based expression plasmids (21).

Plasmid pTC323 containing the glucose facilitator gene (*glf*) of *Zymomonas mobilis* was obtained from Tyrell Conway (University of Oklahoma, Norman, Oklahoma) (22). The *glf* gene, under control of the *tac* promoter, was isolated from pTC323 by digestion with the *Bam*HI and *Hind*III restriction enzymes (New England Biolabs, Beverly, MA). The resulting 2.2-kb fragment was treated with T4 DNA polymerase (Promega, Madison, WI), purified by agarose gel electrophoresis, and ligated into the *Psh*AI site of pET22b(+) to yield plasmid pBGM101. The ability of pBGM101 to restore glucose uptake in strain TP2811 was verified by the production of dark red colonies on MacConkey agar supplemented with glucose (1% w/v). The internal *Eag*I and *Nco*I recognition sites of the *glf* coding sequence were replaced with silent mutations with the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA).

**Preparation of Genomic Libraries.** Genomic DNA from strain BM5300 was prepared according to standard procedures (23). Purified genomic DNA (2.2 μg) was digested for 16 h at 37 °C with an appropriate restriction enzyme (*Eag*I, *Nco*I, *Spe*I, or *Xba*I), and the resulting fragments were purified with the Promega PCR cleanup kit. The average size of these fragments varied between 8 and 120 kb, depending upon the restriction enzyme used in their generation. Genomic fragments were ligated into appropriately digested, dephosphorylated vector pBGM101 (3.4 μg) overnight at 14 °C with T4 DNA ligase (9 units). Ligation reactions were desalted by passage through two pre-equilibrated Autoseq G-50 Sephadex columns (Amersham Biosciences, Piscataway, NJ) and then transformed via electroporation into BM5340(DE3). After recovery for 1.5 h at

37 °C, transformed cells were pelleted at 2000g for 5 min, washed with 2 mL of M9 minimal medium, and plated onto M9 minimal agar containing glucose (0.1% w/v), MgCl<sub>2</sub> (2 mM), ampicillin (150 μg/mL), kanamycin (4 μg/mL), chloramphenicol (2.5 μg/mL), and IPTG (50 μM).<sup>2</sup> This procedure routinely produced genomic libraries of greater than 5 × 10<sup>5</sup> transformants. Colonies that grew within 10 days of plating were scored as positives and were selected for sequencing.

**Enzyme Production and Purification.** The sequencing of positive clones resulted in the identification of three independent open-reading frames (*nanK*, *ycfX*, and *yajF*) that share sequence identity with known sugar kinases (24). The three potential glucokinase genes were PCR-amplified with PfuTurbo DNA polymerase (Stratagene, La Jolla, CA) from the corresponding genomic fragment by using oligonucleotides that incorporated *Nco*I and *Eag*I recognition sites at the 5' and 3' ends of the gene, respectively. Purified PCR products were treated to eliminate polymerase (25) and purified with the Promega PCR cleanup kit. PCR products were digested with *Nco*I and *Eag*I for 2 h at 37 °C and ligated into appropriately digested vector pBGM101 as described above for the genomic library construction. The resulting plasmids (pBGM101-*nanK*, pBGM101-*ycfX*, and pBGM101-*yajF*) contained the *nanK*, *ycfX*, and *yajF* genes downstream of the T7 promoter and in-frame with a C-terminal hexahistidine-encoding sequence. Transformation of BM5340(DE3) with these plasmids, followed by plating on glucose minimal M9 agar, verified the ability of the isolated *nanK*, *ycfX*, and *yajF* genes to complement the glucokinase deficiency of this strain. Complementation of the glucokinase auxotrophy of BM5340(DE3) by *yajF* required the presence of an 87-base-pair sequence located upstream of the coding region of the gene, consistent with a role of this region in modulating the stability of the *yajF* transcript (26).

NanK, YcfX, and YajF were produced in *E. coli* strain BM5340(DE3) harboring pBGM101-*nanK*, pBGM101-*ycfX*, and pBGM101-*yajF*, respectively. Cultures (0.75 L) were inoculated to an initial OD of 0.005 at 600 nm and were grown at 37 °C in Luria-Bertani broth (18) supplemented with glucose (0.5% w/v), ampicillin (150 μg/mL), kanamycin (4.0 μg/mL), and chloramphenicol (2.5 μg/mL). For the production of NanK and YajF, IPTG was added to a final concentration of 0.8 mM when the OD reached 0.8 and the incubation was continued for 2.5 h at 37 °C. For the production of YcfX, IPTG was added to a final concentration of 50 μM when the OD reached 0.8 and the incubation was continued for 2.5 h at 25 °C. Bacteria were collected by centrifugation (10000g for 10 min) and resuspended to 1 g of cells (wet weight) per 8 mL of buffer A [75 mM sodium phosphate buffer at pH 7.6, containing glucose (50 mM), imidazole (25 mM), glycerol (5% w/v), and β-mercaptoethanol (5 mM)]. Cell extracts were prepared by passage through a French pressure cell at 16 000 psi. Intact cells and cell debris were removed by centrifugation at 75000g for 70 min, and the resulting supernatant was loaded onto a nickel-nitrilotriacetic acid (Ni-NTA) affinity column (Qiagen,

<sup>1</sup> Abbreviations: PTS, phosphoenolpyruvate:carbohydrate phosphotransferase system; ManZ, mannosephosphotransferase enzyme IIB; PtsG, glucosephosphotransferase enzyme II; Glf, glucose facilitator protein; ROK, repressor, open-reading frame, kinase; NanK, *N*-acetyl-D-mannosamine kinase; IPTG, isopropyl β-D-1-thiogalactopyranoside; Glk, glucokinase.

<sup>2</sup> An IPTG concentration of 50 μM was used because supplementation of BM5340(DE3) with ≥100 μM IPTG was found to retard its growth on solid medium.

Valencia, CA) that had been equilibrated with buffer A. The column was washed with eight column volumes of buffer A containing 40 mM imidazole, and the enzyme was eluted in buffer A containing 250 mM imidazole. Purified proteins were dialyzed overnight at 4 °C against 12 L of 50 mM Tris-HCl buffer at pH 7.6, containing MgSO<sub>4</sub> (5 mM) and glycerol (5% w/v).

The *E. coli* glucokinase gene was PCR-amplified with PfuTurbo DNA polymerase from MC4100 genomic DNA using oligonucleotides that incorporated *Nde*I and *Eag*I recognition sites at the 5' and 3' ends of the gene, respectively. The *glk* gene was ligated into appropriately digested pET22b(+), and the resulting plasmid was transformed into strain BL21(DE3) via electroporation. Glk was produced and purified as described above for NanK and YajF. All proteins were greater than 90% homogeneous as judged by SDS-PAGE analysis.

Protein concentration was estimated from the absorbance at 280 nm by using molar extinction coefficients determined with the method of Edelhoch (27). The molecular mass of each purified protein was verified by MALDI mass spectrometry on an Applied Biosystems Voyager System 6133 spectrometer at the University of Wisconsin—Madison Biophysics Instrumentation Facility.

**Enzyme Assays.** Kinase activity was determined by coupling the production of glucose 6-phosphate to the reduction of NADP via glucose 6-phosphate dehydrogenase or by coupling the production of MgADP to oxidation of NADH via pyruvate kinase and lactate dehydrogenase, as previously described (28, 29). Assays were conducted in 0.3–0.5 M Tris-HCl buffer at pH 7.6 and 25 °C, and data were fitted to standard Michaelis–Menten kinetic equations for bisubstrate reactions.

## RESULTS

We created an organism that requires the presence of a plasmid-borne glucokinase gene for survival. Previous genetic studies indicated that *E. coli* possesses three enzymes capable of phosphorylating glucose efficiently (30). Two of these enzymes, mannosephosphotransferase enzyme IIB (ManZ) and glucosephosphotransferase enzyme II (PtsG), belong to the phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) of bacteria. Carbohydrates transported by the PTS are phosphorylated concomitant with uptake, via a phosphoryl-group transfer cascade originating from phosphoenolpyruvate (31). In addition to ManZ and PtsG, *E. coli* also possesses a soluble ATP-dependent glucokinase encoded by the *glk* gene (17). We combined deletions in the *pts* operon and the *glk* gene to form strain BM5340(DE3) (Table 1). This bacterium fails to produce colonies on glucose minimal medium and is unable to metabolize glucose, as indicated by the production of white colonies on MacConkey agar.

The deletion in the *pts* operon of strain BM5340(DE3) prevents its transport of carbohydrates. To restore glucose transport activity, we supplied a gene encoding the glucose facilitator protein (Glf) of *Zymomonas mobilis* on an extra-chromosomal plasmid, pBGM101. Unlike the endogenous PTS, Glf transports glucose by facilitated diffusion (22). The action of Glf, in combination with transcription of an active glucokinase gene from the T7 polymerase promoter region,

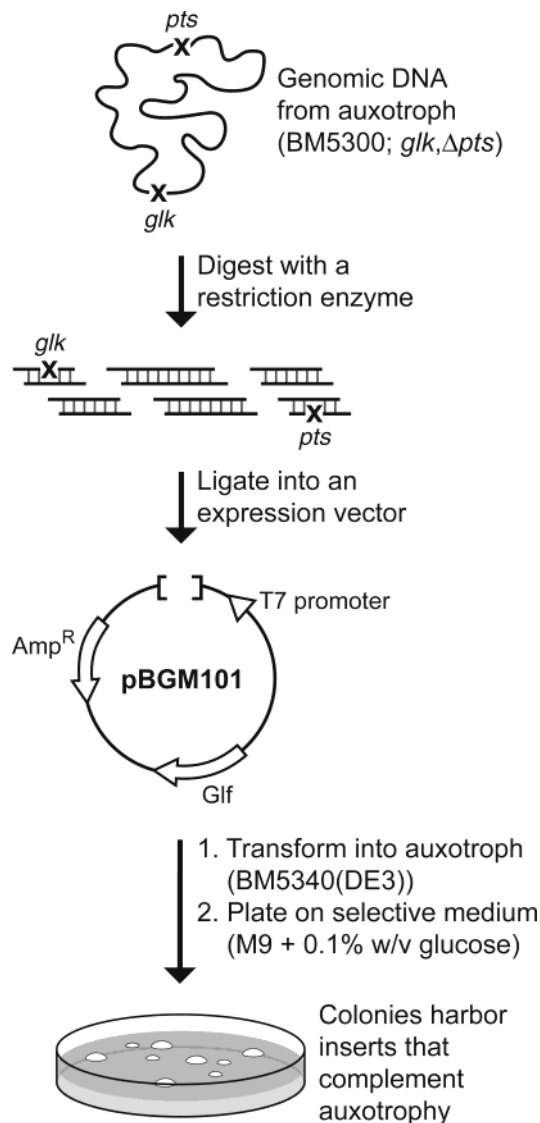


FIGURE 1: Identifying glucokinase activities embedded within the modern *E. coli* genome. Gene products encoding rudimentary glucokinases are selected from *glk*,  $\Delta$ *pts* genomic libraries via their ability to enable the survival of the auxotrophic BM5340(DE3) on glucose minimal medium, when overexpressed from the powerful T7 promoter.

enables BM5340(DE3) cells harboring pBGM101 to metabolize glucose and survive on glucose minimal medium.

To search for gene products that possess glucokinase activity, we constructed four different genomic libraries from the glucokinase-deficient strain BM5300 (Figure 1). These libraries were transformed into BM5340(DE3), overexpressed, and then selected by plating on M9 minimal medium containing glucose as the sole carbon source. After 3–8 days of growth at 37 °C, positive clones emerged at a rate of ca. 1 in 10<sup>4</sup> transformants. The doubling time of positive clones on solid medium was highly variable, a fact attributable to the size of the genomic fragment, the stability of the resulting transcript, and the activity of the rescuing gene product.

The sequencing of selected clones, followed by a comparison of the encoded genomic fragments with BLAST (32), resulted in our identification of three potential glucokinases. These clones arose from three different genes. None of these genes is identifiable from a BLAST search of the *E. coli* genome for amino acid sequences similar to that of *E. coli*

Table 2: Kinetic Parameters for the Phosphorylation of Carbohydrates by Glk, YcfX, NanK, and YajF<sup>a</sup>

parameter	Glk		YcfX		NanK		YajF	
	D-glucose	D-glucose	D-glucose	<i>N</i> -acetyl-D-mannosamine	D-glucose	D-fructose	D-glucose	D-fructose
$k_{\text{cat}}$ (s <sup>-1</sup> )	410	9.2	9.1	83	12	14		
$K_{\text{m,sugar}}$ (M)	$7.6 \times 10^{-5}$	$3.8 \times 10^{-3}$	$1.8 \times 10^{-2}$	$3.1 \times 10^{-4}$	$5.9 \times 10^{-2}$	$1.3 \times 10^{-3}$		
$K_{\text{m,ATP}}$ (M)	$2.6 \times 10^{-4}$	$3.4 \times 10^{-3}$	$2.9 \times 10^{-3}$	$5.1 \times 10^{-4}$	$1.2 \times 10^{-2}$	$1.8 \times 10^{-3}$		
$k_{\text{cat}}/K_{\text{m,sugar}}$ (M <sup>-1</sup> s <sup>-1</sup> )	$5.4 \times 10^6$	$2.4 \times 10^3$	$5.1 \times 10^2$	$2.7 \times 10^5$	$2.0 \times 10^2$	$1.1 \times 10^4$		

<sup>a</sup> Data were obtained at 25 °C in a 0.3–0.5 M Tris-HCl buffer at pH 7.6.

glucokinase, thereby validating the need for an experimental approach. The three selected gene products belong to the ROK (repressor, open-reading frame, kinase) protein family (24) and could have a ribonuclease H-like scaffold (33).

The selected genes were *nanK*, a gene postulated to encode a *N*-acetyl-D-mannosamine kinase (34); *yajF*, a gene previously shown to enable metabolism of fructose in a PTS-deficient strain of *E. coli* (26); and *ycfX*, a previously uncharacterized sequence with unknown function. To determine whether the production of each gene product is necessary and sufficient for complementation, we amplified each candidate gene from the corresponding genomic fragment and ligated the product into fresh pBGM101. Transformation of the resulting plasmids into BM5340(DE3), followed by growth on glucose minimal medium, demonstrated that production of NanK, YajF, or YcfX does indeed complement the glucokinase deficiency.

NanK, YajF, and YcfX have rudimentary glucokinase activity (Table 2). To ensure that this activity did not arise from contaminating endogenous glucokinases, we produced and purified NanK, YajF, and YcfX from glucokinase-deficient BM5340(DE3) cells as C-terminal, hexahistidine-tagged proteins. For all three enzymes, the values of  $K_{\text{m}}$  for ATP measured during glucose phosphorylation are similar, ranging from 2.9 mM for NanK to 12 mM for YajF. The affinities of each protein for substrate glucose are more variable, ranging from 3.8 mM for YcfX to 59 mM for NanK. Interestingly, NanK, YajF, and YcfX demonstrate nearly identical values of  $k_{\text{cat}}$  for glucose phosphorylation, perhaps indicative of a limitation of a common protein fold. The glucokinase catalytic efficiencies of NanK, YajF, and YcfX are  $5.1 \times 10^2$ ,  $2.0 \times 10^2$ , and  $2.4 \times 10^3$  M<sup>-1</sup> s<sup>-1</sup>, respectively. These values are four orders-of-magnitude lower than the value of  $k_{\text{cat}}/K_{\text{m}}$  ( $5.4 \times 10^6$  M<sup>-1</sup> s<sup>-1</sup>) for Glk, the known endogenous glucokinase.<sup>3</sup> In addition to their glucokinase activity, NanK and YajF phosphorylate *N*-acetyl-D-mannosamine and D-fructose, respectively (Figure 2 and Table 2), consistent with earlier hypotheses (26, 34).

## DISCUSSION

We have shown that an *E. coli* strain that is deficient in performing the first step of the glycolytic pathway cannot survive with glucose as its sole source of carbon. Surprisingly, this organism contains three other enzymes that can provide the missing catalytic activity, but only when these

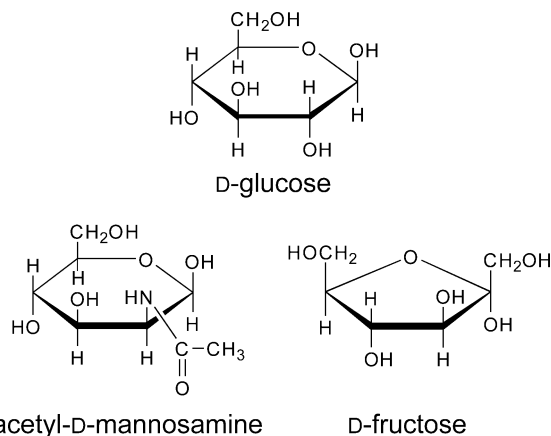


FIGURE 2: Structures of D-glucose, *N*-acetyl-D-mannosamine, and D-fructose.

enzymes are overproduced within the cell. These enzymes, NanK, YajF, and YcfX, have ambiguous substrate specificities.

Enzymatic catalysis is chemically demanding, requiring the precise alignment and coordinated action of numerous functional groups. For example, the catalytic base in the active site of the glycolytic enzyme triosephosphate isomerase is the carboxylate group of Glu165. Replacing Glu165 with an aspartate residue moves the base away from the substrate by only ca. 0.7 Å (35, 36) but causes a 10<sup>3</sup>-fold decrease in catalytic activity (37). Catalyzing the transfer of the  $\gamma$ -phosphoryl group of ATP to one oxygen of a particular carbohydrate presents similar challenges. Indeed, of the four *E. coli* enzymes that we have shown to possess glucokinase activity (Table 2), only YajF is capable of accepting D-fructose as a substrate and only NanK produces observable phosphorylation of *N*-acetyl-D-mannosamine, even in the presence of high concentrations of enzyme. We tested several potential substrates other than glucose for turnover by YcfX (including D-ribose, D-mannitol, D-sorbitol, inositol, and L-threonine) but detected no phosphoryl group transfer. Remarkably, NanK, YajF, and YcfX can all catalyze the phosphorylation of glucose.

NanK, YajF, and YcfX share limited sequence identity with *E. coli* glucokinase (Figure 3). Only 21 residues, or ca. 7% of the total sequence, are conserved in all four proteins. The sequences of NanK, YajF, and YcfX share a higher degree of identity with one another than they do with glucokinase, consistent with the divergence of NanK, YajF, and YcfX from a common ancestor. Transcription of the endogenous *nanK*, *yajF*, and *ycfX* genes is insufficient to complement the bacterial auxotrophy without their deliberate overexpression. Apparently, the glucokinase activity of each protein is not actively maintained by selective pressure. Thus, NanK, YajF, and YcfX could have diverged from *E. coli*

<sup>3</sup> The dynamic range of the selection system constructed herein spans more than four orders of magnitude, permitting the identification of glucokinases with  $k_{\text{cat}}/K_{\text{m}}$  values as low as  $5.1 \times 10^2$  M<sup>-1</sup> s<sup>-1</sup>. This threshold could be indicative of the absence of glucokinases with lower activity or an intrinsic limitation of our selection system (such as the instability or insolubility of glucokinases with lower activity).

<b>glk</b>	mtkyaalvgdv	<b>gg</b> t	narlalc	diasgeisqa	ktysgldyps	leavirvyle	ehkvekdkgc	iaiacpitgd	<b>70</b>
<b>yajF</b>	---mrigidl	<b>gg</b> t	ktevial	gdageqlyrh	rlptprddyr	qtietiatlv	dmaeagatgqr	gtvvgmgips	<b>67</b>
<b>ycfX</b>	---myygfdi	<b>gg</b> t	kialgvf	dsgrglqwek	rvptprdsyd	afldavcelv	aeadrgrfgck	gsvvgigipgm	<b>67</b>
<b>nanK</b>	--mttlaidi	<b>gg</b> t	klaaali	gadgqirdrr	elptpasqtp	ealrdalsal	vsplqahaqr	--vaiastgi	<b>66</b>
<b>glk</b>	wvamtntwa	fsiaemkkn-	--lgfshlei	indftavsm	ipmlkkehli	qfpgaepveg	kpiavygagt	<b>137</b>	
<b>yajF</b>	ispytgvvkn	anstwlngq-	--pfdkdlsa	rlqrevrlan	danclavsea	vdgaaagaqt	vfvaviigtgc	<b>134</b>	
<b>ycfX</b>	petedgtlya	anvpaasgk-	--plradlsa	rlldrdrvldn	dancfalsea	wddefqtqpl	vmgllilgtgv	<b>134</b>	
<b>nanK</b>	ir--dgsila	lnphnlggll	hfpvlvktleq	ltnlptiain	daqaaawaef	qaldgditdm	vfitvsttgv	<b>133</b>	
<b>glk</b>	glgvahlvhw	dkrwwslpge	gghvdfapns	eeeaiileil	r--aeighvs	aervlsggggl	vnlyraivka	<b>205</b>	
<b>yajF</b>	gagvafngra	higgngtage	wghnlpwmd	edelryreev	p-cycgkqgc	ietfilsqtgf	amdyrriisgh	<b>203</b>	
<b>ycfX</b>	ggglifngkp	itgksytige	fghmrlpva	ltmmgldfpl	rrcgcgqhgk	ienylsgngf	awlyqhyhqq	<b>204</b>	
<b>nanK</b>	gggvvsgckl	ltgpgglagh	ightladphg	-----	pvcgcgrtgc	veaiasgrgi	aaaaqgelag	<b>193</b>	
<b>glk</b>	dnrlpenlcp	kditeralad	sctdcrals	lfcvimgrfg	gnlalnlgtf	gg-vfiaggi	vprfleffka	<b>274</b>	
<b>yajF</b>	alkgseilrl	veesd---pv	aelalrnyel	rlakslahvv	nildpdivvl	gggmsnvdrl	yqtvqqlikq	<b>270</b>	
<b>ycfX</b>	plqapeiial	ydqgd---eq	arahveryld	llavclgnil	tivdpdlvvi	ggglnsfpai	ttqladrlpr	<b>271</b>	
<b>nanK</b>	-adaktiftr	agqgd---eq	aqqlihrsar	tlarliadik	attdcqcvv	ggsvglaeqy	lalvetylaq	<b>259</b>	
<b>glk</b>	sgfraafedk	grfkeyvhdi	pvylihndp	gllgsgahlr	qtlghil			<b>321</b>	
<b>yajF</b>	f-----	--vfggecet	pvrkakhgds	sgvrqaawlw	pqe----			<b>302</b>	
<b>ycfX</b>	h-----	--llpvarvp	rierarhgda	ggmrqaafllh	ltd----			<b>303</b>	
<b>nanK</b>	-----	---epaafhv	dllaahyrhd	agllgaalla	qgekl--			<b>291</b>	

FIGURE 3: Amino acid sequence alignment of *E. coli* Glk, NanK, YajF, and YcfX with ClustalW (38). Residues conserved in all four enzymes are on a black background; residues conserved in NanK, YajF, and YcfX are boxed.

glucokinase during ancient times, and their glucokinase activities may represent the nonessential retention of an ancestral function.

The identification of many distinct enzymes capable of catalyzing an identical chemical reaction provides new opportunities to illuminate the evolution of catalytic function. For example, within the ribonuclease H-like scaffold, the ability to bind glucose appears to have evolved independently in mammals, yeast, and bacteria (39). Moreover, glucokinase activity appears to have evolved independently on at least two different protein scaffolds by natural selection (40). Consequently, it would be of interest to determine whether refinement of glucokinase activity in NanK, YajF, and YcfX requires a common set of mutations that recapitulate site-specific amino acid identities present in *E. coli* glucokinase, or whether multiple solutions to this problem exist.

Few examples of natural catalytic promiscuity, the ability of an enzyme to catalyze multiple, chemically distinct reactions, are known (41–44). The experimental approach described herein provides a general strategy for the identification of promiscuous enzymes and affords a method to characterize these ambiguous activities systematically. Despite the presumed evolutionary advantages of catalytic promiscuity and substrate ambiguity (41), extant metabolic pathways are often portrayed as being unequivocal—in each biochemical transformation “the enzyme and the substrate must fit together like a lock and a key” (7). In contrast, our findings suggest that a wealth of ambiguous enzyme activities are embedded in the genomes of modern organisms and churn within living cells. In principle, these activities could originate from the partial retention of a previously optimized function, or they could result from the redundancy of chemical structures within the modern metabolic pool. Regardless of the source, such activities could provide convenient starting points for the natural or directed evolution of new enzymes with desirable functions.

## ACKNOWLEDGMENT

We are grateful to W. Boos, T. Conway, and A. Danchin for gifts of plasmids and bacterial strains and W. W. Cleland, S. Chandran, and B. D. Smith for valuable discussions.

## REFERENCES

- Jensen, R. A. (1976) Enzyme recruitment in evolution of new function, *Annu. Rev. Microbiol.* 30, 409–425.
- Lewis, E. B. (1951) Pseudoallelism and gene evolution, *Cold Spring Harbor Symp. Quant. Biol.* 16, 159–174.
- Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag, Berlin, Germany.
- Albery, W. J., and Knowles, J. R. (1976) Evolution of enzyme function and the development of catalytic efficiency, *Biochemistry* 15, 5631–5640.
- Burbaum, J. J., Raines, R. T., Albery, W. J., and Knowles, J. R. (1989) Evolutionary optimization of the catalytic effectiveness of an enzyme, *Biochemistry* 28, 9293–9305.
- Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signaling, *Nature* 411, 355–365.
- Fischer, E. (1894) Einfluss der configuration auf die wirkung derenzyme, *Ber. Dtsch. Chem. Ges.* 27, 2985–2993.
- 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.
- Hegeman, G. D., and Rosenberg, S. L. (1970) The evolution of bacterial enzyme systems, *Annu. Rev. Microbiol.* 24, 429–462.
- Lerner, S. A., Wu, T. T., and Lin, E. C. C. (1964) Evolution of a catabolic pathway in bacteria, *Science* 146, 1313–1315.
- 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.
- Campbell, J. H., Lengyel, J. A., and Langridge, J. (1973) Evolution of a second gene for  $\beta$ -galactosidase in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* 70, 1841–1845.
- Betz, J. L., Brown, P. R., Smyth, M. J., and Clarke, P. H. (1974) Evolution in action, *Nature* 247, 261–264.
- Rigby, P. W. J., Burleigh, B. D., Jr., and Hartley, B. S. (1974) Gene duplication in experimental enzyme evolution, *Nature* 251, 200–204.
- Casadaban, M. J. (1976) Transposition and fusion of the lac genes to selected promoters in *Escherichia coli* using bacteriophage  $\lambda$  and  $\mu$ , *J. Mol. Biol.* 104, 541–555.

16. Levy, S., Zeng, G.-Q., and Danchin, A. (1990) Cyclic AMP synthesis in *Escherichia coli* strains bearing known deletions in the pts phosphotransferase operon, *Gene* 86, 27–33.
17. 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.
18. Miller, J. H. (1972) *Experiments in Molecular Genetics*, pp 201–205, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. MacConkey, A. (1905) Lactose-fermenting bacteria in faeces, *J. Hyg.* 5, 333–379.
20. Epstein, W., and Davies, M. (1970) Potassium-dependent mutants of *Escherichia coli* K-12, *J. Bacteriol.* 101, 836–843.
21. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, *Methods Enzymol.* 185, 60–89.
22. Parker, C., Barnell, W. O., Snoep, 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.
23. 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, MA.
24. 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.
25. Matsumura, I., and Ellington, A. D. (2001) Mutagenic PCR of protein-encoding genes of in vivo evolution, *Methods Mol. Biol.* 182, 261–269.
26. 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.
27. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry* 6, 1948–1954.
28. Liu, F., Dong, Q., Myers, 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.
29. Sabater, B., Sebastian, J., and Asensio, C. (1972) Identification and properties of an inducible mannokinase from *Streptomyces violaceoruber*, *Biochim. Biophys. Acta* 284, 406–414.
30. 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.
31. Postma, P. W., Lengeler, J. W., and Jacobson, G. R. (1996) Phosphoenolpyruvate:carbohydrate phosphotransferase systems in *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology (Neidhardt, F. C., Curtiss, R., and Lin, E. C., Eds.), American Society for Microbiology Press, Washington, DC.
32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Biol.* 215, 403–410.
33. Cheek, S., Zhang, H., and Grishin, N. V. (2002) Sequence and structure classification of kinases, *J. Mol. Biol.* 320, 855–881.
34. Plumbridge, J., and Vimr, E. (1999) Convergent pathways for utilization of the amino sugars *N*-acetylglucosamine, *N*-acetylmannosamine, and *N*-acetylneuraminic acid by *Escherichia coli*, *J. Bacteriol.* 181, 47–54.
35. Joseph-McCarthy, D., Rost, L. E., Komives, E. A., and Petsko, G. A. (1994) Crystal structure of the mutant yeast triosephosphate isomerase in which the catalytic glutamic acid 165 is changed to aspartic acid, *Biochemistry* 33, 2824–2829.
36. Komives, E. A., Loughheed, J. C., Liu, K., Suglo, S., Zhang, Z., Petsko, G. A., and Ringe, D. (1995) The structural basis for pseudoreversion of the E165D lesion by the secondary S96P mutation in triosephosphate isomerase depends on the positions of the active site water molecules, *Biochemistry* 34, 13612–13621.
37. Raines, R. T., Sutton, E. L., Straus, D. R., Gilbert, W., and Knowles, J. R. (1986) Reaction energetics of a mutant triosephosphate isomerase in which the active-site glutamate has been changed to aspartate, *Biochemistry* 25, 7142–7154.
38. Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000) NPS@: Network protein analysis, *Trends Biochem. Sci.* 25, 147–150.
39. Bork, P., Sander, C., and Valencia, A. (1993) Convergent evolution of similar enzymatic function on different protein folds: The hexokinase, ribokinase, and galactokinase families of sugar kinases, *Protein Sci.* 2, 31–40.
40. Ito, S., Fushinobu, S., Yoshioka, I., Koga, S., Matsuzawa, H., and Wakagi, T. (2001) Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon, *Structure* 9, 205–214.
41. O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities, *Chem. Biol.* 6, R91–R105.
42. Joerger, A. C., Mayer, A., and Fersht, A. R. (2003) Mimicking natural evolution in vitro: An *N*-acetylneuraminidase lyase mutant with an increased dihydropicolinate synthase activity, *Proc. Natl. Acad. Sci. U.S.A.* 100, 5694–5699.
43. Copley, S. D. (2003) Enzymes with extra talents: Moonlighting functions and catalytic promiscuity, *Curr. Opin. Chem. Biol.* 7, 265–272.
44. Schmidt, D. M. Z., Mundorff, E. C., Dojka, M., Bermudez, E., Ness, J. E., Govindarajan, S., Babbitt, P. C., Minshull, J., and Gerlt, J. A. (2003) Evolutionary potential of ( $\beta/\alpha$ )<sub>8</sub>-barrels: Functional promiscuity produced by single substitutions in the enolase superfamily, *Biochemistry* 42, 8387–8393.

BI049424M