Identifying Latent Enzyme Activities: Substrate Ambiguity within Modern Bacterial Sugar Kinases†

Brian G. Miller‡ and Ronald T. Raines*§¶

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 allow 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' operon, was provided by Antoine Danchin (Institut Pasteur, Paris, France) (16). Strain DM1000, harboring a TnJ0(Cam) insertion in the glucokinase gene...
was supplied by Winfried Boos (University of Konstanz, Germany) (17). Generalized transduction with P1kc phage of the Δ(pshH, psl, crr), KanR 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 ßDE3 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 BamHI and HindIII 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 PshAI 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 EagI and NcoI recognition sites of the glf coding sequence were replaced with silent mutations with the QuikChange site-directed mutagensis 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 (EagI, NcoI, SphI, or XbaI), 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 preequilibrated 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), MgCl2 (2 mM), ampicillin (150 µg/mL), kanamycin (4 µg/mL), chloramphenicol (2.5 µg/mL), and IPTG (50 µM). This procedure routinely produced genomic libraries of greater than 5 x 10^4 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 NcoI and EagI 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 NcoI and EagI 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 (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 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-nitriolactiic acid (Ni-NTA) affinity column (Qiagen, 1 Abbreviations: PTS, phosphoenolpyruvate:carbohydrate phosphotransferase system; ManZ, mannosephosphotransferase enzyme IIb; PstG, 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; Glik, glucokinase.

2 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 MgSO₄ (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 NdeI and EagI 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, 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 10⁴ 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...
nearly identical values of variable, ranging from 3.8 mM for YcfX to 59 mM for **affinities of each protein for substrate glucose are more ranging from 2.9 mM for NanK to 12 mM for YajF. The ATP measured during glucose phosphorylation are similar, mannosamine and D-fructose, respectively (Figure 2 and YcfX are 5.1**

**DISCUSSION**

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 BMS340(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 BMS340(DE3) cells as C-terminal, hexahistidine-tagged proteins. For all three enzymes, the values of $K_m$ 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 $k_{cat}$ for glucose phosphorylation, perhaps indicative of a limitation of a common protein fold. The glucokinase catalytic efficiencies of NanK, YajF, and YcfX are $5.1 \times 10^9$, $2.0 \times 10^7$, and $2.4 \times 10^7 \text{M}^{-1} \text{s}^{-1}$, respectively. These values are four orders-of-magnitude lower than the value of $k_{cat}/K_m$ ($5.4 \times 10^6 \text{M}^{-1} \text{s}^{-1}$) for Glk, the known endogenous glucokinase. 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).

**Table 2: Kinetic Parameters for the Phosphorylation of Carbohydrates by Glk, YcfX, NanK, and YajF**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Glk</th>
<th>YcfX</th>
<th>NanK</th>
<th>YajF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>410</td>
<td>9.2</td>
<td>9.1</td>
<td>83</td>
</tr>
<tr>
<td>$K_m$ (M)</td>
<td>$7.6 \times 10^{-5}$</td>
<td>$3.8 \times 10^{-3}$</td>
<td>$1.8 \times 10^{-2}$</td>
<td>$3.1 \times 10^{-4}$</td>
</tr>
<tr>
<td>$K_m$ (M)</td>
<td>$2.6 \times 10^{-4}$</td>
<td>$3.4 \times 10^{-3}$</td>
<td>$2.9 \times 10^{-3}$</td>
<td>$5.1 \times 10^{-4}$</td>
</tr>
<tr>
<td>$k_{cat}/K_m$ (M$^{-1}$ s$^{-1}$)</td>
<td>$5.4 \times 10^9$</td>
<td>$2.4 \times 10^9$</td>
<td>$5.1 \times 10^9$</td>
<td>$2.7 \times 10^5$</td>
</tr>
</tbody>
</table>

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

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$^3$-fold decrease in catalytic activity (37). Catalyzing the transfer of the γ-phosphoryl group of ATP to one oxygen of a particular carbohydrate presents similar challenges. Indeed, 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.

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

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

The sequences of 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.
glk  mtkyalvscg ggtarlalc diaesigesqa ktyesglgyps leavirvyle ehkvkvdgce iaiaagipie 70
yajF  ---mrigidh ggtvevial gqageqlyrh rlihptcvey drqyietytv dmartergq gtvvpmgoq 67
ycFX  ---myggfog sfqgaalgv dfgrglgewk arptpcyvdy afdladvcev yeadagfogc gavviegifs 67
nanK  --mtiaalig ggtlaali gadtqgdr qelepmsqtp ealdalsr nmvlhagq - vialastq 66

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.

glk  ggwavhvh hvkwrvalpgq ghvdpans eeeaiieli r-aeigvav gcvrlgc gvnlyraivka 205
yajF  gggavfngra hqgntae wgbnpqmd ederyrevve p-ecvgck dezcgc qmysdrq amdyrylqgh 203
yceFX  gccilfngkp itkscytyog fgbrmlvpva tmmqdlpfpl rrcgqrcgc gcvyldgc awlyghyhgq 204
nanK  cvdrvsgckl ltpgoalq egltlphyg ppgmgtdgq valniaegq gaaagqelag 193

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


BI049424M