Divergent Evolution of Function in the ROK Sugar Kinase Superfamily: Role of Enzyme Loops in Substrate Specificity

Mioara Larion,[‡] Lauren B. Moore,[‡] Steven M. Thompson,[§] and Brian G. Miller*[,‡]

Department of Chemistry and Biochemistry and School of Computational Science, The Florida State University, Tallahassee, Florida 32306

Received May 14, 2007; Revised Manuscript Received August 16, 2007

ABSTRACT: The D-allose and *N*-acetyl-D-mannosamine kinases of *Escherichia coli* K-12 are divergent members of the functionally diverse ROK (repressor, open reading frame, kinase) superfamily. Previous work in our laboratory has demonstrated that AlsK and NanK possess weak phosphoryl transfer activity toward the alternate substrate D-glucose. To gain insight into the evolutionary mechanisms that fuel the specialization of individual enzyme function, experimental laboratory evolution was conducted to improve the glucokinase activities of AlsK and NanK. Error-prone PCR was combined with in vivo functional selection in a glucokinase-deficient bacterium to identify four independent single nucleotide substitutions in the *alsK* and *nanK* genes that improve the glucokinase activity of each enzyme. The most advantageous substitutions, L84P in NanK and A73G in AlsK, enhance the k_{cat}/K_m values for phosphoryl transfer to glucose by 12-fold and 60-fold, respectively. Both substitutions co-localize to a variable loop region located between the fourth β-sheet and the second α-helix of the ROK scaffold. A multiple sequence alignment of diverse ROK family members reveals that the A73G substitution in AlsK recapitulates a conserved glycine residue present in many ROK proteins, including some transcriptional repressors. Steady-state kinetic analyses of the selected AlsK and NanK variants demonstrate that their native activities toward D-allose and *N*-acetyl-D-mannosamine are largely unaffected by the glucokinase-enhancing substitutions. Substrate specificity profiling reveals that the A73G AlsK and L84P NanK variants display systematic improvements in the k_{cat}/K_m values for a variety of nonnative carbohydrates. This finding is consistent with an evolutionary process that includes the formation of intermediates possessing relaxed substrate specificities during the initial steps of enzyme functional divergence.

Modern experimental techniques in molecular biology and microbial genetics have made possible the rapid evolution of protein catalysts within the laboratory in a manner that closely resembles the process of natural divergent evolution (*1*−*4*). These methods offer a unique opportunity to identify underlying principles that govern which features of a progenitor catalyst are targeted for alteration during the specialization of enzyme function. A detailed investigation of laboratory-based experimental enzyme evolution promises to reveal common mechanisms for the generation of new catalytic activities during sequential rounds of mutational drift and natural selection. In addition, the ability to understand how single amino acid substitutions within a given enzyme scaffold stimulate changes in individual kinetic constants should provide unique insight into the linkage between a protein's three-dimensional structure and its biological function.

If common structural and/or kinetic features are targeted by evolution during the specialization of function, one might expect these to be most easily identified in a superfamily of protein catalysts that share a common ancestor and possess an overlapping alternate function (*5*, *6*). Recently, our laboratory has discovered a series of divergent bacterial sugar kinases that belong to the ROK[1] superfamily (Pfam 00480), which fits this description (*7*, *8*). The ROK superfamily represents a functionally diverse group of prokaryotic proteins that includes carbohydrate-responsive transcriptional repressors and sugar kinases (*9*). ROK transcriptional repressors are signified by a conserved N-terminal helix−turn−helix motif that affords DNA binding, whereas ROK kinases contain an N-terminal ATP-binding motif denoted by the sequence DXGXT (*10*). Presently, the Pfam database contains more than 1600 ROK members, many of which possess unknown physiological functions. Within the last 3 years, the crystal structures of several ROK polypeptides have been determined. Included among these are unliganded structures of *Escherichia coli N*-acetyl-D-mannosamine kinase (PDB

* To whom correspondence should be addressed: 213 Dittmer Laboratory of Chemistry, Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306-4390. Tel (850) 645-6570; fax (850) 644-8281; e-mail: miller@chem.fsu.edu.

[‡] Department of Chemistry and Biochemistry.

[§] School of Computational Science.

[1] Abbreviations: AlsK, allokinase; Glk, glucokinase; Mak, manno-(fructo)kinase; NagK, *N*-acetyl-D-glucosamine kinase; NanK, *N*-acetyl-D-mannosamine kinase; PgmK, inorganic polyphosphate/ATP glucomannokinase; OD, optical density; PCR, polymerase chain reaction; ROK, repressor, open reading frame, kinase, G6PDH, glucose-6-phosphate dehydrogenase, Mlc, making large colonies protein; YdhR, putative fructokinase; IPTG, isopropyl thio-β-D-galactoside; $NAD(P)^+$, nicotinamide adenine dinucleotide (phosphate). oxidized form; NAD(P)H, nicotinamide adenine dinucleotide (phosphate), reduced form; DTT, dithiothreitol.

10.1021/bi700924d CCC: $37.00
Published on Web 11/03/2007

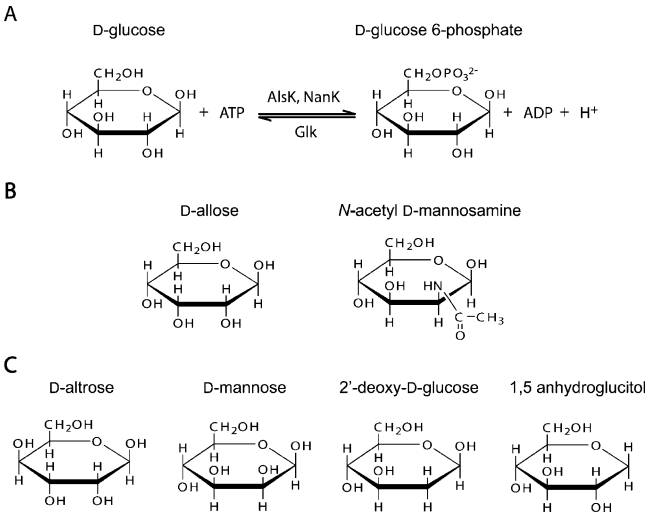

FIGURE 1: (A) Phosphorylation of glucose, as catalyzed by AlsK, NanK and Glk. (B) Structures of the AlsK and NanK native substrates. (C) Structures of alternate AlsK and NanK substrates.

2AA4) and *Salmonella typhimurium N*-acetyl-D-glucosamine kinase (PDB 2AP1), as well as the structure of the *E. coli* transcriptional repressor Mlc (PDB 1Z6R) (*11*−*13*). Notably, the unique inorganic polyphosphate/ATP-glucomannokinase from *Arthrobacter* sp. strain KM is the only ROK kinase whose structure has been determined with a carbohydrate bound at the active site (PDB 1WOQ) (*14*). This enzyme lacks the CXCGXXGC sequence motif present in the founding members of the ROK family first described by Saier and colleagues, but it retains the core ROK family signature pattern (*15*). The crystal structure of *E. coli* glucokinase has also been recently determined in the presence of glucose; however, this enzyme is distantly related to other ROK proteins and it is not currently classified as a member of this superfamily (*16*).

The *E. coli* K-12 genome harbors the coding sequences for four ROK sugar kinases: allokinase (AlsK) (*17*, *18*), *N*-acetyl-D-mannosamine kinase (NanK) (*19*, *20*), *N*-acetyl-D-glucosamine kinase (NagK) (*21*), and manno(fructo)kinase (Mak) (*22*). These four enzymes are derived from a common ancestor and each possesses a low level of phosphoryl transfer activity toward the alternate substrate, D-glucose (Figure 1). The latent glucokinase activities of these enzymes, which range from ~100 to 1000 M^{-1} s^{-1}, are several orders of magnitude lower than the catalytic efficiency displayed by native *E. coli* glucokinase (*7*, *8*). In contrast, the k_{cat}/K_m value of each enzyme acting upon its native substrate exceeds 10^5 M^{-1} s^{-1}, as expected for a physiologically relevant activity. How differences in substrate specificity have arisen within the ROK superfamily of bacterial sugar kinases, given the high level of structural similarity between the native carbohydrate substrates of each member (Figure 1), is unknown.

The evolutionary origin of enzyme specificity is of particular interest to our laboratory. Most contemporary enzymes possess the ability to recognize subtle structural differences in individual substrate molecules, often at the single-atom level, and to transform these differences into large variations in catalytic efficiency. Indeed, substrate selectivity is a distinguishing characteristic of protein catalysts, one that is largely unrealized in the world of synthetic small-molecule catalysts (*23*). Specificity is defined by the ratio of second-order rate constants, k_{cat}/K_m, that govern the transformations of two competing substrates. Several models have been put forth to explain the structural, thermodynamic, and kinetic basis for enzyme specificity, often amid much discussion (*24*, *25*). One particularly controversial issue is the extent to which enzyme conformational changes participate in substrate discrimination (*26*−*28*). A consensus has yet to be reached on this issue (*29*−*31*); however studies aimed at uncovering the evolutionary origins of substrate discrimination in protein catalysts may provide additional insight.

The sugar kinases encompassed by the ROK superfamily offer a unique opportunity to explore the evolutionary origins of enzyme specificity. Here, we probe evolutionary events that occur during the optimization of glucokinase activity in two members of the ROK superfamily, Alsk and NanK, which share only 21% sequence identity with one another. Following random mutagenesis and in vivo functional selection, we identify two structurally overlapping mutational "hot spots" in the sugar kinase scaffold. Steady-state kinetic analyses of the selected variants demonstrate that the native activities of AlsK and NanK are largely unaffected by the glucokinase-enhancing substitutions. Furthermore, the variants have acquired an increased ability to phosphorylate a variety of nonnatural carbohydrate substrates as a result of the evolutionary process. These findings provide insight into the determinants of substrate specificity in the ROK superfamily and support previous reports suggesting that alterations in enzyme−substrate selectivity proceed through nonspecialized intermediates (*32*, *33*).

MATERIALS AND METHODS

Random Mutagenesis of alsK and nanK. Error-prone PCR of the native *alsK* and *nanK* genes were carried out with the Gene Morph II random mutagenesis kit (Stratagene) and the following oligonucleotide primers: AlsK forward, 5′-CTT TAA GAA GGA GAT ATA CCA TG-3′; AlsK reverse, 5′-GGT GCT CGA GTG CGG CCG-3′; NanK forward, 5′-GGC GGT GCC ATG GCC ACA CTG GCG ATT GAT ATC G-3′; and NanK reverse, 5′-CTG TAC TTC ACC CAT CCG GCC GAT TTT TCT CCC-3′. Template plasmid (pBGM101-*alsK* or pBGM101-*nanK*; 1 μg) (*7*, *8*) was used in each reaction, and 25 amplification cycles were performed to achieve an average mutation rate of 1 base pair change per gene. Following amplification, the reaction mixture was treated with *Dpn*I (20 units) for 1 h at 37 °C to digest parental DNA and the products were purified with the Promega PCR cleanup kit. The resulting mutagenized product was digested with *Eag*I (5 units) and *Nco*I (5 units) for 2 h at 37 °C, purified again with the PCR cleanup kit, and added to a ligation reaction containing appropriately digested pBGM101.1 vector (1 μg) and T_4 DNA ligase (40 units). The ligation reaction was incubated overnight at 17 °C, and then the reaction was quenched by heating at 70 °C for 10 min. The ligation products were precipitated via the addition of sodium acetate (0.3 M) and ethanol (70%) with incubation at −20 °C for 1 h. The reaction mixture was centrifuged at 16000*g* for 10 min to collect the DNA pellet, which was subsequently washed with 200 μL of ice-cold ethanol (70%). Following a second centrifugation step at 16000*g*, the pellet

was isolated and resuspended in 15 μL of 37 °C nuclease-free water.

To evaluate the quality and mutational frequency of the *alsK* and *nanK* libraries, a small aliquot of each library was transformed into electrocompetent BM5340(DE3) cells via electroporation. Transformed cells were plated on Luria−Bertani plates containing ampicillin (150 μg/mL), chloramphenicol (25 μg/ mL), and kanamycin (40 μg/mL), and single colonies were isolated following overnight incubation at 37 °C. *Nco*I and *Eag*I double digests of miniprep DNA prepared from individual clones demonstrated that more than 80% of library members possessed an insert. A total of 18 clones were selected from rich Luria−Bertani plates and the plasmid DNA from each was sequenced to reveal a mutation frequency of 0.9 ± 0.6 bps per gene. Serial dilutions of transformed BM5340(DE3) cells were plated on antibiotic-supplemented Luria−Bertani plates to yield estimated library sizes of 7.3×10^6 transformants for the *alsK* library and 8.6×10^6 transformants for the *nanK* library. These library sizes are sufficiently large to ensure that all possible single base pair changes are sampled in each library.

Genetic Selection Experiments. To select for amino acid substitutions that improved the glucokinase activity of AlsK and NanK, error-prone PCR libraries were transformed into electrocompetent BM5340(DE3) cells. Following recovery at 37 °C for 1.5 h in SOB medium (20 g tryptone, 5 g yeast extract, 0.5 g NaCl, and 0.19 g KCl per liter of water) supplemented with glucose (20 mM) and $MgCl_2$ (20 mM), transformed cells were pelleted by centrifugation at 3000*g*, washed twice with 1 mL of M9 minimal medium, and plated on M9 minimal plates containing glucose (0.005% w/v), ampicillin (150 μg/mL), chloramphenicol (25 μg/mL), and kanamycin (40 μg/mL). Selection plates were also supplemented with IPTG at a final concentration of either 50 μM for NanK selection experiments or 10 μM for AlsK selection experiments to induce protein production. The total fraction of complementing clones was 0.128% for the *nanK* library and 0.064% for the *alsK* library. Approximately 1% of the total complementing clones from each library afforded colonies within 2 days of growth at 37 °C on glucose-supplemented M9 minimal plates, and these subpopulations were selected for further study. Control experiments verified that expression of the wild-type *nanK* and *alsK* genes does not produce colonies within this time frame under the selective growth conditions described above, and the wild-type genes were never recovered in the library selection experiments. Mini-prep DNA prepared from individual colonies was retransformed into electrocompetent BM5340-(DE3) cells and the selection was repeated as described above to verify the growth phenotype of primary selectants. The fastest growing library members from this second round of selection were submitted for sequencing.

Enzyme Expression and Purification. Enzymes were produced in the glucokinase-deficient *E. coli* strain BM5340-(DE3) to avoid contamination with endogenous glucokinases. Cultures were inoculated to an initial OD_{600nm} of 0.01 and cells were grown at 37 °C in Luria−Bertani medium containing ampicillin (150 μg/mL), chloramphenicol (25 μg/ mL), and kanamycin (40 μg/mL) until the OD_{600nm} reached 0.9. IPTG was added to a final concentration of 1 mM to induce gene expression, and cultures were grown for an additional 2−3 h at 37 °C. Bacteria were harvested by centrifugation at 7000*g* for 15 min, and the resulting cell pellet (3 g) was resuspended in 12 mL of buffer A, containing sodium phosphate (75 mM, pH 7.6), glycerol (5% w/v), and imidazole (25 mM). Cell extracts were prepared by passage through a model 110L microfluidizer (Microfluidics Corp), and the resulting cell lysate was cleared by centrifugation at 15000*g* for 30 min at 4 °C. The supernatant was syringe-tip-filtered (0.2 μm) and then loaded onto a 5 mL HisTrap FF crude column that had been previously equilibrated with buffer A. Following loading, the column was washed with 10 additional column volumes of buffer A containing 40 mM imidazole. Purified enzymes were eluted in buffer A containing 250 mM imidazole and were dialyzed overnight against two sequential 2 L aliquots of Tris-HCl buffer (25 mM, pH 7.6) containing $MgSO_4$ (5 mM), glycerol (5% w/v), and DTT (0.5 mM). Protein concentrations were estimated from absorbance readings at 280 nm by use of previously determined molar extinction coefficients of 15 000 cm^{-1} for NanK and 34 900 cm^{-1} for AlsK (*7*, *8*). The purity of all enzymes was judged to be ≥95% on the basis of polyacrylamide gel analysis.

Enzyme Assays. The glucokinase activity of individual enzymes was measured by coupling the production of glucose 6-phosphate to the reduction of $NADP^+$ via glucose-6-phosphate dehydrogenase (*34*). To determine apparent K_m values for glucose, reaction mixtures contained Tris (0.2 M, pH 7.6), $NADP^+$ (0.5 mM), DTT (1 mM), ATP (7.5 mM), $MgCl_2$ (8.5 mM), and G6PDH (7.5 units). Enzyme activity with allose, altrose, 2-deoxyglucose, mannose, and *N*-acetyl-D-mannosamine as substrates was measured by coupling the production of ADP to the oxidation of NADH via the combined action of pyruvate kinase (15 units) and lactate dehydrogenase (25 units) (*35*). Assays contained Tris (0.2 M, pH 7.6), NADH (0.5 mM), DTT (1 mM), phosphoenolpyruvate (PEP) (5 mM), ATP (5−25 mM), $MgCl_2$ (6−26 mM), and KCl (5 mM). To determine the apparent K_m values for ATP, the same experimental conditions were employed at saturating concentrations of carbohydrate. Assays were conducted at 25 °C and were initiated by addition of the enzyme under investigation. Data were fitted to standard Michaelis−Menten kinetic equations, with each data point representing the average of at least two individual rate determinations. Assays of wild-type and variant NanK acting upon 1,5-anhydroglucitol were conducted by the lactate dehydrogenase/pyruvate kinase linked assay at saturating ATP concentrations (25 mM) and subsaturating concentrations of carbohydrate. The k_{cat}/K_m value was determined from the slope of the linear plot of the rate of NAD production as a function of 1,5-anhydroglucitol concentration.

Multiple Sequence Alignment and Phylogenetic Analysis. An initial ClustalW (*36*) alignment of *E. coli* AlsK and NanK was used to produce a HMM profile (*37*) with HMMerBuild. This HMM profile was then used to search for homologous sequences in UniProt Release 9.6 with HMMerSearch. A natural break in *Expectation values* was apparent within the results between 1×10^{-76} and 3×10^{-50}. Therefore, in spite of the presence of hundreds of homologous sequences, all sequences with *E values* higher than 1×10^{-76} were eliminated from the data set. HMMerAlign was then used to align the remaining data set to the initial HMM. Nearly redundant sequences were then eliminated and a new HMM

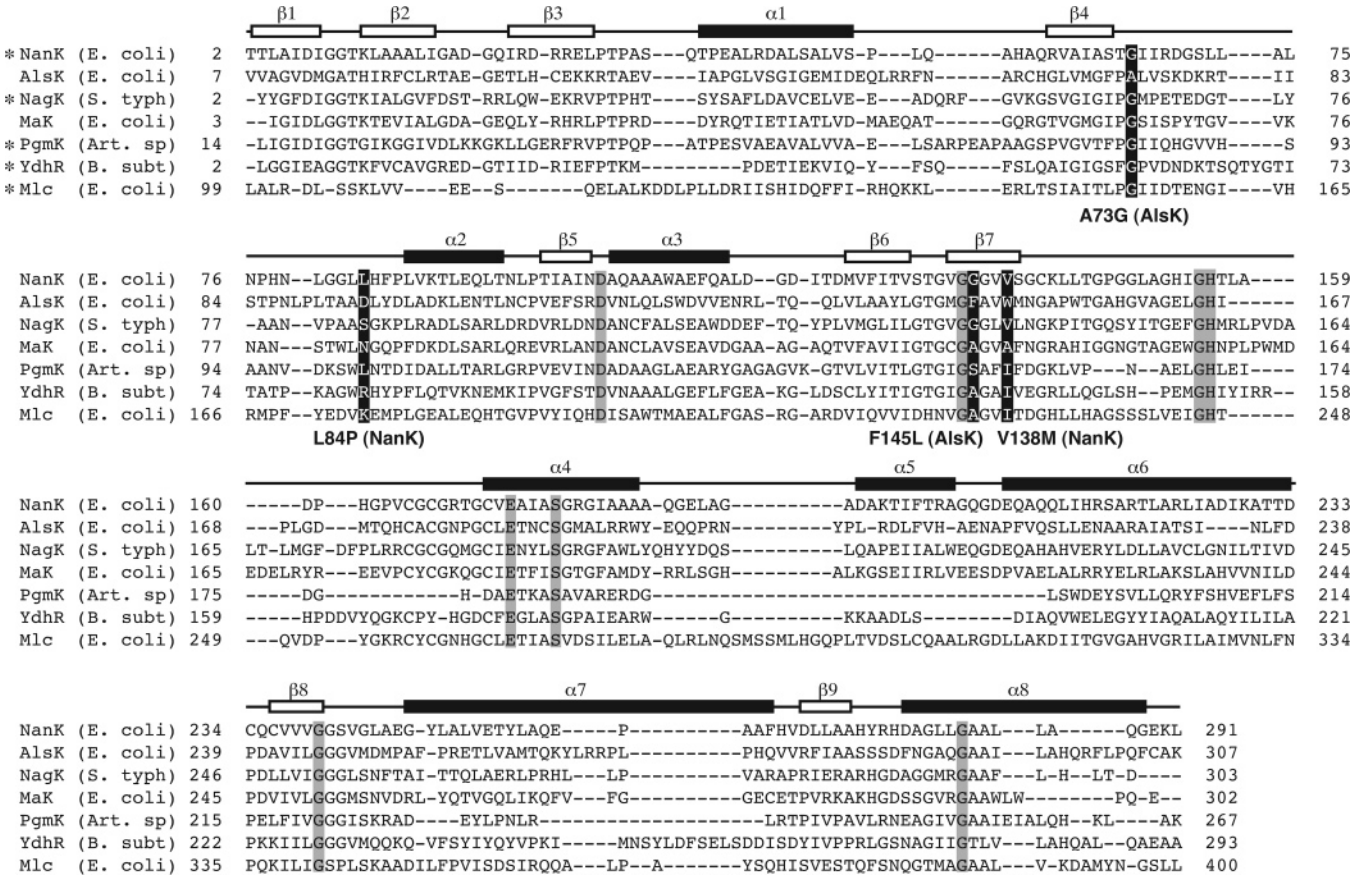

FIGURE 2: Multiple sequence alignment of ROK superfamily members. Locations of the amino acid substitutions observed in this study are shown on a black background and their identities are depicted in boldface type. The secondary structural elements of the ROK scaffold are represented above the sequence alignment and conserved residues are shown on a gray background. Asterisks denote ROK family members whose structures are available in the Protein Data Bank. A description of the construction of this alignment is provided in the Materials and Methods section.

profile was built. This new HMM was then used to merge MLC_Ecoli (P50456 *E value* 1.5×10^{-2} in initial search), YDHR_Bacsu (O05510 *E value* 2.9×10^{-6} in initial search), PGMK_Arts (Q7WT42 *E value* 6.9×10^{-4} in initial search), MAK_Ecoli (P23917 *E value* 5.7×10^{-7} in initial search), NAGK_Ecoli (P75959 *E value* 7.5×10^{-9} in initial search), and NAGK_Salty (Q8ZPZ9 *E value* 5.1×10^{-8} in initial search) into the alignment. Finally, this alignment was sent to the Advanced T-Coffee Expresso (*38*) server to be refined by structural consistency methodology, and then minor manual adjustments were made with the GCG SeqLab editor. The alignment was next masked to eliminate homoplastic columns of amino acids with less than 10% similarity and exported to FastA format, which was then converted to PHYLIP format with ReadSeq (*39*). The final data matrix consisted of 25 taxa by 311 characters.

ProML (version 3.67) (*40*) was used to infer the maximum likelihood (ML) tree for the data set by use of a JTT protein model (*41*) with a four discrete rate γ correction model and an α value of 1.00. The optimal model was determined with ProtTest (*42*). The heuristic search strategy consisted of 10 random additions with global tree rearrangements for each replicate. SeqBoot (version 3.67) (*40*) was then used to produce 100 pseudoreplicate data sets, which were then passed back to ProML for bootstrapped maximum likelihood analysis. An identical evolutionary model and search strategy was employed. The resulting 100 boostrap ML trees were then run through Consense (version 3.67) (*40*) to produce the majority-rule consensus tree for the bootstrap analysis. The single maximum likelihood tree, available in Supporting Information, had a negative log likelihood value of 9148.29.

All sequence analysis [with the exception of the format converter ReadSeq (*39*)] was performed within SeqLab (*43*) from the GCG (*44*) package. Phylogenetic analyses were done with the PHYLIP package (*40*).

RESULTS

Co-localization of Selected AlsK and NanK Variants. To investigate the accessible evolutionary pathway(s) for improving the glucokinase activities of *E. coli* AlsK and NanK, error-prone PCR was conducted on the wild-type *alsK* and *nanK* genetic templates. In these directed evolution experiments, a low rate of mutation was chosen in an attempt to mimic the single nucleotide changes that likely power the evolutionary refinement of weak promiscuous catalytic activities in natural biological systems. Single amino acid substitutions that increased the glucokinase activity of AlsK and NanK were selected by the ability of their encoded genes to efficiently complement the glucokinase auxotrophy of BM5340(DE3). Two single amino acid substitutions in NanK, L84P and V138M, produced the fastest growth rate of BM5340(DE3) cells on glucose minimal medium. Similarly, two single amino acid variants of AlsK, A73G and F145L, were most effective at complementing the glucokinase deficiency of BM5340(DE3).

Table 1: Kinetic Parameters for Phosphorylation of Native and Ambiguous Substrates by Wild-Type NanK and Selected Variants[a]

parameter	wt NanK	L84P NanK loop region	V138M NanK β-sheet region
		N-acetyl-D-mannosamine[b]	
$k_{cat,\ sugar}$ (s^{-1})	56 ± 1.0	18 ± 0.4	20 ± 0.5
$K_{m,\ sugar}$ (M)	$(3.6 \pm 0.3) \times 10^{-4}$	$(2.3 \pm 0.3) \times 10^{-4}$	$(1.6 \pm 0.1) \times 10^{-4}$
$K_{m,\ ATP}$ (M)	$(2.6 \pm 0.1) \times 10^{-4}$	$(8.4 \pm 0.6) \times 10^{-5}$	$(1.1 \pm 0.1) \times 10^{-4}$
$k_{cat}/K_{m,\ sugar}$ ($M^{-1}\ s^{-1}$)	$(1.5 \pm 0.1) \times 10^{5}$	$(7.8 \pm 0.7) \times 10^{4}$	$(1.2 \pm 0.1) \times 10^{5}$
		D-glucose[c]	
$k_{cat,\ sugar}$ (s^{-1})	22 ± 0.5	84 ± 0.2	60 ± 0.3
$K_{m,\ sugar}$ (M)	$(2.0 \pm 0.1) \times 10^{-2}$	$(6.4 \pm 0.1) \times 10^{-3}$	$(8.6 \pm 0.2) \times 10^{-3}$
$K_{m,\ ATP}$ (M)	$(1.2 \pm 0.1) \times 10^{-3}$	$(9.2 \pm 1.0) \times 10^{-4}$	$(6.3 \pm 1.0) \times 10^{-4}$
$k_{cat}/K_{m,\ sugar}$ ($M^{-1}\ s^{-1}$)	$(1.1 \pm 0.1) \times 10^{3}$	$(1.3 \pm 0.1) \times 10^{4}$	$(7.0 \pm 0.1) \times 10^{3}$

[a] Kinetic parameters were determined at 25 °C, in 0.2 M Tris, pH 7.6. [b] Determined by a pyruvate kinase/lactate dehydrogenase coupled assay. [c] Determined by a glucose-6-phosphate dehydrogenase coupled assay (see Materials and Methods).

Table 2: Kinetic Parameters for Phosphorylation of Native and Ambiguous Substrates by Wild-Type AlsK and Selected Variants[a]

parameter	wt AlsK	A73G AlsK putative loop region	F145L AlsK putative β-sheet region
		D-allose[b]	
$k_{cat,\ sugar}$ (s^{-1})	47 ± 0.7	40 ± 0.5	49 ± 0.8
$K_{m,\ sugar}$ (M)	$(1.9 \pm 0.1) \times 10^{-4}$	$(2.0 \pm 0.1) \times 10^{-4}$	$(1.5 \pm 0.1) \times 10^{-4}$
$K_{m,\ ATP}$ (M)	$(2.7 \pm 0.2) \times 10^{-4}$	$(6.0 \pm 0.5) \times 10^{-5}$	$(1.4 \pm 0.1) \times 10^{-4}$
$k_{cat}/K_{m,\ sugar}$ ($M^{-1}\ s^{-1}$)	$(2.5 \pm 0.1) \times 10^{5}$	$(2.0 \pm 0.1) \times 10^{5}$	$(3.2 \pm 0.2) \times 10^{5}$
		D-glucose[c]	
$k_{cat,\ sugar}$ (s^{-1})	10 ± 0.2	13 ± 0.5	28 ± 0.5
$K_{m,\ sugar}$ (M)	$(2.9 \pm 0.2) \times 10^{-2}$	$(6.5 \pm 0.9) \times 10^{-4}$	$(7.2 \pm 0.5) \times 10^{-3}$
$K_{m,\ ATP}$ (M)	$(5.4 \pm 0.5) \times 10^{-3}$	$(4.5 \pm 0.2) \times 10^{-4}$	$(9.6 \pm 0.1) \times 10^{-4}$
$k_{cat}/K_{m,\ sugar}$ ($M^{-1}\ s^{-1}$)	$(3.4 \pm 0.3) \times 10^{2}$	$(2.1 \pm 0.3) \times 10^{4}$	$(3.9 \pm 0.3) \times 10^{3}$

[a] Kinetic parameters were determined at 25 °C, in 0.2 M Tris, pH 7.6. [b] Determined by a pyruvate kinase/lactate dehydrogenase coupled assay. [c] Determined by a glucose-6-phosphate dehydrogenase coupled assay (see Materials and Methods).

Table 3: Specificity Profiles of Wild-Type and A73G AlsK[a]

substrate	k_{cat} (s^{-1})	K_m (M)	k_{cat}/K_m ($M^{-1}\ s^{-1}$)	*x*-fold change in k_{cat}/K_m
		wild-type AlsK		
D-allose	47 ± 0.7	$(1.9 \pm 0.1) \times 10^{-4}$	$(2.5 \pm 0.1) \times 10^{5}$	
D-glucose	10 ± 0.2	$(2.9 \pm 0.2) \times 10^{-2}$	$(3.4 \pm 0.3) \times 10^{2}$	
D-altrose	34 ± 2.2	$(2.1 \pm 0.3) \times 10^{-1}$	$(1.6 \pm 0.2) \times 10^{2}$	
2′-deoxy-D-glucose	1.5 ± 0.1	$(3.8 \pm 0.4) \times 10^{-1}$	4.1 ± 0.5	
D-mannose	0.5 ± 0.1	$(3.9 \pm 0.7) \times 10^{-1}$	1.2 ± 0.2	
		A73G AlsK		
D-allose	40 ± 0.5	$(2.0 \pm 0.1) \times 10^{-4}$	$(2.0 \pm 0.1) \times 10^{5}$	0.8
D-glucose	13 ± 0.5	$(6.5 \pm 0.9) \times 10^{-4}$	$(2.1 \pm 0.3) \times 10^{4}$	62
D-altrose	94 ± 2.1	$(4.8 \pm 0.3) \times 10^{-2}$	$(2.0 \pm 0.1) \times 10^{3}$	12.5
2′-deoxy-D-glucose	19 ± 0.5	$(5.1 \pm 0.4) \times 10^{-2}$	$(3.7 \pm 0.3) \times 10^{2}$	90
D-mannose	9.5 ± 0.4	$(2.1 \pm 0.2) \times 10^{-1}$	$(4.5 \pm 0.4) \times 10^{1}$	37.5

[a] Kinetic parameters were determined at 25 °C, in 0.2 M Tris, pH 7.6, by a glucose-6-phosphate dehydrogenase coupled assay for glucose or a pyruvate kinase/lactate dehydrogenase coupled assay for all other sugars (see Materials and Methods).

A comparison of the primary amino acid sequences of AlsK and NanK reveals that the sites of substitution co-localize to similar positions within the sugar kinase scaffold. A sequence alignment of individual ROK family members, shown in Figure 2, indicates that Leu-84 of NanK is located near Ala-73 of AlsK. Similarly, Val-138 of NanK nearly overlaps with the position of Phe-145 in the AlsK primary sequence. When the conserved secondary structural elements of the ROK scaffold were mapped onto the sequence alignment, the glucokinase-enhancing substitutions correspond to the same secondary regions. In particular, Leu-84 of NanK and Ala-73 of AlsK reside in a variable loop region that is located between the fourth β-sheet and the second α-helix of the ROK scaffold. Conversely, Val-138 of NanK and Phe-145 of AlsK both reside in the seventh β-sheet. The pairwise co-localization of these selected substitutions suggests the existence of two "hot spots" within the polypeptide scaffold, which are targeted by evolution during the optimization of glucokinase activity.

Activity of Selected Variants toward Glucose and Their Native Substrates. To assess the enhancements in glucokinase activity afforded by individual AlsK and NanK variants, kinetic assays were performed on highly purified preparations of each protein. Steady-state kinetic analysis of the most active NanK variant, L84P, demonstrates that this substitution increases the second-order rate constant for glucose phosphorylation, $k_{cat}/K_{m\ glucose}$, by 12-fold, from 1.1×10^3 to 1.3×10^4 $M^{-1}\ s^{-1}$ (Table 1). This increase is achieved by a 3-fold decrease in the K_m value for glucose and a 4-fold increase in the k_{cat} value. The L84P NanK variant also displays a modest 2-fold decrease in the catalytic efficiency toward its native substrate, *N*-acetyl-D-mannosamine. V138M, the second NanK variant selected from the error-prone PCR library, displays a 6-fold increase in the k_{cat}/K_m value for

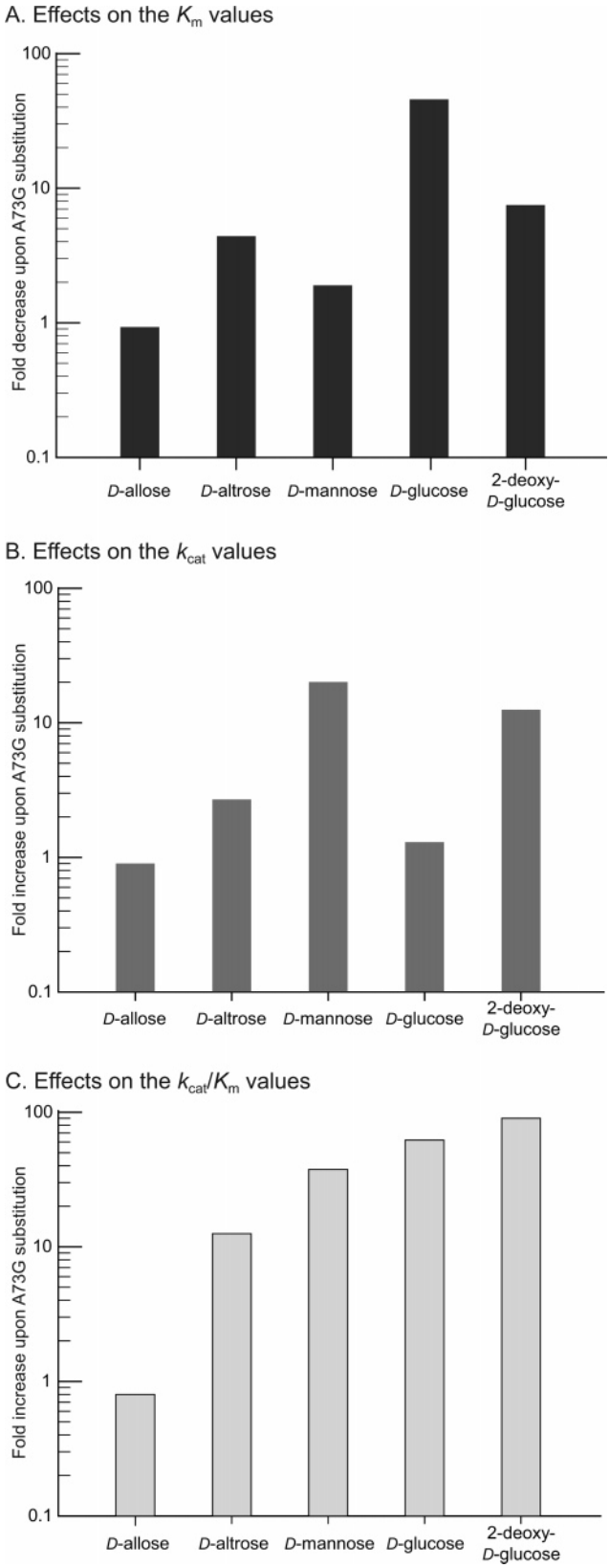

FIGURE 3: Effects of the A73G AlsK loop substitution upon the kinetic parameters for a variety of carbohydrate substrates. *x*-Fold improvements were calculated as follows: (A) $(K_{m,wt-AlsK})/(K_{m,A73GAlsK})$; (B) $(k_{cat,A73GAlsK})/(k_{cat,wt-AlsK})$; (C) $(k_{cat}/K_{m,A73GAlsK})/(k_{cat}/K_{m,wt-AlsK})$.

glucose, while the k_{cat}/K_m value for *N*-acetyl-D-mannosamine remains unchanged. To investigate the consequences of combining the L84P and V138M substitutions in a single polypeptide, the doubly mutated *nanK* coding sequence was constructed and the resulting gene product was purified. The k_{cat}/K_m value for glucose phosphorylation by the L84P/V138M variant increased 5-fold compared to wild-type NanK, from 1.1×10^3 M^{-1} s^{-1} to 5.3×10^3 M^{-1} s^{-1}. Thus, the effects of the individual L84P and V138M substitutions were nonadditive.

Steady-state kinetic analysis of the most active AlsK variant, A73G, reveals a 60-fold increase in the k_{cat}/K_m value for glucose phosphorylation (Table 2). The effect of this mutation is exclusively due to an improved K_m value for substrate glucose, which decreases from 29 mM to 650 μM. The second AlsK variant selected from our randomized library is F145L, an enzyme that displays a 10-fold increase in glucokinase activity compared with the k_{cat}/K_m value of the wild-type enzyme. Similar to the selected NanK variants, the A73G and F145L substitutions in AlsK had no significant effect upon the second-order rate constant for phosphoryl transfer when the native carbohydrate, D-allose, was the substrate.

Substrate Specificity Profiles of AlsK and NanK Loop Variants. To understand the basis for the 60-fold increase in glucokinase activity provided by the AlsK loop substitution, A73G, several carbohydrates were tested as potential substrates. Altrose, 2′-deoxyglucose, and mannose were found to be substrates for phosphoryl transfer by both wild-type and A73G AlsK. Together with the native substrate, D-allose, this collection provides a group of carbohydrates that are transformed by the wild-type enzyme with a range of catalytic efficiencies that span more than 5 orders of magnitude. As shown in Table 3, substitution of Ala-73 with Gly in AlsK enhances the k_{cat}/K_m for all nonnative substrates tested. This includes a 13-fold improvement in activity toward altrose, a 38-fold increase in activity toward mannose, and a 90-fold increase in the k_{cat}/K_m value for 2′-deoxyglucose. These improvements are achieved, in part, via 2-, 4-, and 8-fold decreases in the K_m values for mannose, altrose, and 2′-deoxyglucose, respectively. Similarly, the A73G loop substitution increases the k_{cat} value for altrose, 2′-deoxyglucose, and mannose by 3-, 13-, and 20-fold, respectively. These observations stand in stark contrast to the unchanged value of k_{cat} for glucose phosphorylation resulting from the A73G substitution. A summary of the effects of the A73G substitution upon individual catalytic constants for all substrates examined in this study is provided in Figure 3.

The effect of the L84P loop substitution upon the specificity profile of native NanK was also measured for two poor substrates, D-mannose and 1,5-anhydroglucitol. It was impossible to achieve saturating concentrations of 1,5-anhydroglucitol for either wild-type or L84P NanK; thus, only a value for k_{cat}/K_m is reported for this substrate. Similar to the effects of the AlsK loop substitution, the L84P substitution enhanced the catalytic efficiency of NanK toward both 1,5-anhydroglucitol and D-mannose by 6- and 10-fold, respectively. In the case of mannose, this increase stems from a modest decrease in the K_m value and a slight increase in the k_{cat} value (Table 4).

DISCUSSION

Previous work in our laboratory has identified a divergent superfamily of bacterial sugar kinases that possess ambiguous

Table 4: Specificity Profile of Wild-Type and L84P NanK[a]

substrate	k_{cat} (s^{-1})	K_m (M)	k_{cat}/K_m (M^{-1} s^{-1})	*x*-fold change in k_{cat}/K_m
		wild-type NanK		
N-acetyl-D-mannosamine	56 ± 1.0	$(3.6 \pm 0.3) \times 10^{-4}$	$(1.5 \pm 0.1) \times 10^{5}$	
D-glucose	22 ± 0.5	$(2.0 \pm 0.1) \times 10^{-2}$	$(1.1 \pm 0.1) \times 10^{3}$	
D-mannose	25 ± 0.2	$(8.4 \pm 0.2) \times 10^{-2}$	$(3.0 \pm 0.1) \times 10^{2}$	
1,5-anhydroglucitol	nd[b]	nd	1.7 ± 0.1	
		L84P NanK		
N-acetyl-D-mannosamine	18 ± 0.4	$(2.3 \pm 0.3) \times 10^{-4}$	$(7.8 \pm 0.7) \times 10^{4}$	0.5
D-glucose	84 ± 0.2	$(6.4 \pm 0.1) \times 10^{-3}$	$(1.3 \pm 0.6) \times 10^{4}$	12
D-mannose	62 ± 0.7	$(2.3 \pm 0.2) \times 10^{-2}$	$(2.8 \pm 0.2) \times 10^{3}$	9.3
1,5-anhydroglucitol	nd	nd	$(1.0 \pm 0.1) \times 10^{1}$	5.9

[a] Kinetic parameters were determined at 25 °C, in 0.2 M Tris, pH 7.6, by a glucose-6-phosphate dehydrogenase coupled assay for glucose or a pyruvate kinase/lactate dehydrogenase coupled assay for all other sugars (see Materials and Methods). [b] Values were not determined due to the high concentrations of substrate required for saturation.

substrate specificities toward the alternate substrate D-glucose (*7*, *8*). This family includes the AlsK and NanK polypeptides of *Escherichia coli* K-12, which display k_{cat}/K_m values for glucose phosphorylation of 3.4×10^2 M^{-1} s^{-1} and 1.1×10^3 M^{-1} s^{-1}, respectively. The identification of an overlapping alternate function in these enzymes, in combination with the availability of a sensitive genetic selection system linking glucokinase activity to bacterial survival, provided a unique opportunity to trace the evolutionary pathway(s) leading to the specialization of individual enzyme function within the ROK superfamily. In particular, we wished to investigate whether common structural and/or kinetic features of the sugar kinase scaffold could be targeted for alteration during the evolutionary refinement of glucokinase activity in this superfamily. Here, we show that substitutions at two similar sites within the AlsK and NanK scaffold can improve catalytic activity toward an alternate substrate. This result is similar to the findings of a previous investigation in the enolase superfamily, which demonstrated that single amino acid replacements at overlapping active-site locations in the L-Ala-D/L-Glu epimerase from *Escherichia coli* and the muconate lactonizing enzyme II from *Pseudomonas* sp. P51 enhance the alternate *o*-succinylbenzoate synthase activity of these enzymes (*5*). Similarly, substitutions at structurally superimposable active-site residues in two divergent aminohydrolyase family members were found to increase the promiscuous esterase activity of each enzyme (*6*). Taken together, these examples strongly suggest that mutational events that target a common structural region of a polypeptide scaffold often power the initial steps of the divergent evolution of alternate function within a superfamily of related protein catalysts.

Two distinct categories of enhancing substitutions were selected from our random libraries. The first class of variants includes F145L of AlsK and V138M of NanK, both of which produced modest increases in glucokinase catalytic efficiency. Phe-145 and Val-138 reside on a β-sheet whose amino terminus includes two highly conserved glycine residues that form close interactions with the C6 position of glucose in the crystal structure of the inorganic polyphosphate/ATP glucomannokinase from *Arthrobacter* sp. strain KM (PgmK) (*16*). In fact, Phe-145 of AlsK is directly adjacent to one of these glycine residues. Replacement of this bulky residue with a smaller leucine side chain could potentially perturb the positioning of the sugar substrate within the AlsK active site. Although the Val-138 side chain is further removed from the NanK active site, subtle perturbations at this location could also impact positioning of carbohydrates within the active site via the collective reorientation of intervening residues.

The second category of substitutions, which produced larger improvements in glucokinase activity, occurred at positions occupied by Ala-73 of AlsK and Leu-84 of NanK. These two residues are located in a variable region of the ROK scaffold located between the fourth β-sheet and the second α-helix. To better understand the nature of the glucokinase-enhancing substitutions, we compared the amino acid sequences of NanK and AlsK in this region with a diverse range of other ROK family members (Figure 2). There was no detectable conservation in other ROK proteins of the position analogous to Leu-84 in NanK. Several gaps appear in the multiple sequence alignment in the region surrounding Leu-84, consistent with significant alterations occurring within this loop during the divergent evolution of function in this superfamily. The structure of the unliganded *E. coli* NanK is available (*11*); however, the absence of a structure that reveals how the carbohydrate substrate binds in the active site makes inferences regarding the impact of the L84P replacement upon NanK specificity difficult. In contrast to the variability in the sequence alignment near Leu-84, Figure 2 reveals that Ala-73 of AlsK is a highly conserved glycine residue in numerous other superfamily members. Indeed, the conservation of glycine at this position persists across the entire phylogeny of the ROK superfamily and is not limited to sugar kinases, as many transcriptional repressors also retain this residue. Unfortunately, the structure of AlsK is unknown. However, the structure of another ROK sugar kinase, the inorganic polyphosphate/ATP glucomannokinase (PgmK) from *Arthrobacter* sp. strain KM, was recently determined in the presence of substrate glucose (*14*). The active site of this enzyme reveals that the Cα atom of the conserved glycine residue (Gly-84 in PgmK) is located within 3.3 Å of the O3 atom of bound glucose (Figure 4). Notably, the structure of D-allose differs from that of D-glucose at the 3′-OH position (Figure 1). On the basis of the available sequence and structural data, we postulate that removal of the Ala-73 side chain in AlsK provides sufficient space near the 3′-OH of D-glucose to allow this allose epimer to bind in a position that is conducive to efficient phosphoryl transfer.

To understand the consequences of the evolutionary process more generally, the L84P NanK and A73G AlsK enzymes were tested for activity with a variety of nonnative sugar substrates. Interestingly, the A73G and L84P replace-

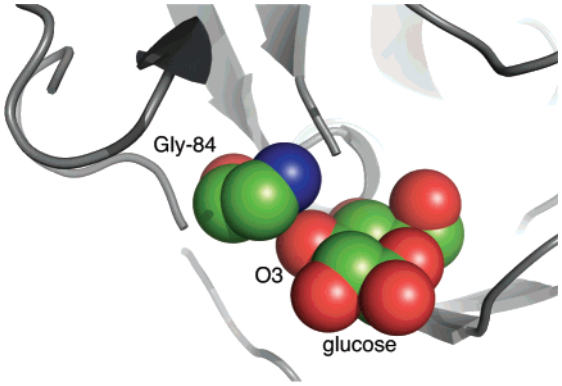

FIGURE 4: Active site of inorganic polyphosphate/ATP-glucomannokinase from *Arthrobacter* sp. strain KM in complex with D-glucose, revealing the position of the highly conserved glycine residue (Gly-84) with respect to the O3 position of the bound substrate. Images were created with the program PyMOL (Delano Scientific) and PDB entry 1WOQ (*14*).

ments enhanced the catalytic efficiencies for all substrates examined. This observation is consistent with the results of other experimental evolution studies that support the formation of nonspecific protein catalysts during the early stages of functional divergence (*32*, *33*). In the present study, the effects of A73G and L84P substitutions upon the steady-state catalytic parameters of individual carbohydrates depended upon the nature of the substrate. For example, the A73G substitution increased the second-order rate constants for the promiscuous phosphorylation of D-altrose, D-mannose, and 2′-deoxyglucose, in part by enhancing the k_{cat} value of each substrate. By contrast, the value of k_{cat}/K_m for glucose was enhanced solely via an altered K_m value. This observation suggests that different intrinsic steps may limit the transformation rates of distinct carbohydrate substrates by each enzyme. For one carbohydrate, 2-deoxyglucose, the A73G substitution produced a greater increase in this substrate's k_{cat}/K_m value (90-fold) than for D-glucose (60-fold). This observation was unexpected since the evolved AlsK was isolated under glucose-selective growth conditions, and to our knowledge the in vivo selection conditions imposed in this study provided no impetus to enhance activity toward 2′-deoxyglucose. Nevertheless, this finding demonstrates how the mutational expansion of enzymatic function via the formation of nonspecific intermediates can serve to amplify additional promiscuous activities, which could prove advantageous to a host organism.

SUPPORTING INFORMATION AVAILABLE

List of the protein sequences used to construct the multiple sequence alignment and the maximum likelihood phylogenetic analysis of these sequences. This material is available free of charge via the Internet at http://pubs.acs.org

REFERENCES

1. Taylor, S. V., Kast, P., and Hilvert, D. (2001) Investigating and engineering enzymes by genetic selection, *Angew. Chem., Int. Ed. 40*, 3310−3335.
2. Farinas, E. T., Butler, T., and Arnold, F. H. (2001) Directed enzyme evolution, *Curr. Opin. Biotechnol. 12*, 545−551.
3. Lutz, S., and Patrick, W. M. (2004) Novel methods for directed evolution of enzymes: Quality, not quantity, *Curr. Opin. Biotechnol. 15*, 291−297.
4. Kaur, J., and Sharma, R. (2006) Directed evolution: An approach to engineer enzymes, *Crit. Rev. Biotechnol. 26*, 165−199.
5. Vick, J. E., Schmidt, D. M. Z., and Gerlt, J. A. (2005) Evolutionary potential of $(\alpha/\beta)_8$-Barrels: In vitro enhancement of a "new" reaction in the enolase superfamily, *Biochemistry 44*, 11722−11729.
6. Roodveldt, C., and Tawfik, D. S. (2005) Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily, *Biochemistry 44*, 12728−12736.
7. Miller, B. G., and Raines, R. T. (2004) Identifying latent enzyme activities: Substrate ambiguity within modern bacterial sugar kinases, *Biochemistry 43*, 6387−6392.
8. Miller, B. G., and Raines, R. T. (2005) Reconstitution of a defunct glycolytic pathway via recruitment of ambiguous sugar kinases, *Biochemistry 44*, 10776−10783.
9. 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.
10. Holmes, K. C., Sander, C., and Valencia, A. (1993) A new ATP-binding fold in actin, hexokinase and Hsc70, *Trends Cell Biol. 3*, 53−59.
11. Patskovsky, Y., and Almo, S., (2005) Crystal structure of *Escherichia coli* putative *N*-acetyl-D-mannosamine kinase, New York Structural Genomics Consortium (to be published).
12. Brunzelle, J. S., Minasov, G., Shuvalova, L., Collart, F. R., and Anderson, W. F. (2005) Crystal structure of the putative regulatory protein, Midwest Center for Structural Genomics (to be published).
13. Schiefner, A., Gerber, K., Seitz, S., Welte, W., Diederichs, K., and Boos, W. (2005) The crystal structure of Mlc, a global regulator of sugar metabolism in *Escherichia coli*, *J. Biol. Chem. 280*, 29073−29079.
14. Mukai, T., Kawai, S., Mori, S., Mikami, B., and Murata, K. (2004) Crystal structure of bacterial inorganic polyphosphate/ATP-glucomannokinase, *J. Biol. Chem. 279*, 50591−50600.
15. Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999) The PROSITE database, its status in 1999, *Nucleic Acids Res. 27*, 215−219.
16. Lunin, V. V., Li, Y., Schrag, J. D., Iannuzzi, P., Cygler, M., and Matte, A. (2004) Crystal structure of *Escherichia coli* ATP-dependent glucokinase and its complex with glucose, *J. Bacteriol. 186*, 6915−6927.
17. Kim, C., Song, S., and Park, C. (1997) The D-allose operon of *Escherichia coli* K-12, *J. Bacteriol. 179*, 7631−7637.
18. 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.
19. Plumbridge, J., and Vimr, E. (1999) Convergent pathways for utilization of the amino sugars *N*-acetylglucosamine, *N*-acetylmannosamine, and *N*-acetylneuramic Acid by *Escherichia coli*, *J. Bacteriol. 181*, 47−54.
20. Ringenberg, M. A., Steenberg, S. M., and Vimr, E. R. (2003) The first committed step in the biosynthesis of sialic acid by *Escherichia coli* K1 does not involve a phosphorylated *N*-acetylmannosamine intermediate, *Mol. Microbiol. 50*, 961−975.
21. Uehara, T., and Park, J. T. (2004) The *N*-acetyl-D-glucosamine kinase of *Escherichia coli* and its role in murein recycling, *J. Bacteriol. 186*, 7273−7279.
22. Sproul, A. A., Lambourne, L. T. M., Jean-Jacques, D. 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.
23. D'Souza, V. T., and Bender, M. L. (1987) Miniature organic models of enzymes, *Acc. Chem. Res. 20*, 146−152.
24. Fischer, E. (1894) Einfluss der configuration auf die wirkung derenzyme, *Ber. Dtsch. Chem. Ges. 27*, 2985−2993.
25. 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.
26. Fersht, A. (1974) Catalysis, binding and enzyme−substrate complementarity, *Proc. R. Soc. London, Ser. B 187*, 397−407.
27. Herschlag, D. (1988) The role of induced fit and conformational changes of enzymes in specificity and catalysis, *Bioorg. Chem. 16*, 62−96.
28. Post, C. B., and Ray, W. J. (1995) Reexamination of induced fit as a determinant of substrate specificity in enzymatic reactions, *Biochemistry 34*, 15881−15885.
29. Hedstrom, L. (1996) Trypsin: A case study in the structural determinants of enzyme specificity, *Biol. Chem. 377*, 465−470.

30. Pasternak, A., White, A., Jeffery, J. C., Medina, N., Cahoon, M., Ringe, D., and Hedstrom, L. (2001) The energetic cost of induced fit catalysis: Crystal structures of trypsinogen mutants with enhanced activity and inhibitor affinity, *Protein Sci. 10*, 1331−1342.
31. Tsai, Y.-C., and Johnson, A. K. (2006) A new paradigm for DNA polymerase specificity, *Biochemistry 45*, 9675−9687.
32. Matsumura, I., and Ellington, A. D. (2001) *In vitro* evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates, *J. Mol. Biol. 305*, 331−339.
33. Aharoni, A., Gaidukov, L., Khersonsky, O., McQ, G. S., Roodveldt, C., and Tawfik, D. S. (2005) The evolvability of promiscuous protein functions, *Nat. Genet. 37*, 73−76.
34. Slein, M. W. (1963) D-Glucose determinations with hexokinase and glucose 6-phosphate dehydrogenase, in *Methods of Enzymatic Analysis* (Bergmeyer, H. U., Ed.) Academic Press, San Diego, CA, p 117.
35. Easterby, J. S. (1973) Coupled enzyme assays: A general expression for the transient, *Biochim. Biophys. Acta. 293*, 552−558.
36. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673−4680.
37. Eddy, S. R. (1998) Profile hidden Markov models, *Bioinformatics 14*, 755−763.
38. Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., and Notredame, C. (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee, *Nucleic Acids Res. 34*, 604−608.
39. Gilbert, D. G. (1990−2006) ReadSeq(dDistributed by the author), Biology Department, Indiana University, Bloomington, IN. See http://iubio.bio.indiana.edu/soft/molbio/readseq/.
40. Felsenstein, J. (1980−2007) PHYLIP (Phylogeny Inference Package) version 3.6 (distributed by the author), Department of Genome Sciences, University of Washington, Seattle, WA. Available at http://evolution.genetics.washington.edu/phylip.html.
41. Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences, *Comput. Appl. Biosci. 8*, 275−282.
42. Abascal, F., Zardoya, R., and Posada, D. (2005) ProtTest: Selection of best-fit models of protein evolution, *Bioinformatics 21*, 2104−2105.
43. Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W., and Gillevet, P. M. (1994) The Genetic Data Environment, an expandable GUI for multiple sequence analysis, *Comput. Appl. Biosci. 10*, 671−675.
44. Genetics Computer Group. (Copyright 1982−2007) *Program Manual for the Wisconsin Package*, version 11, Accelrys, Inc., San Diego, CA.

BI700924D