Protein Science (1993), *2*, 31–40. Cambridge University Press. Printed in the USA.

Convergent evolution of similar enzymatic function on different protein folds: The hexokinase, ribokinase, and galactokinase families of sugar kinases

PEER BORK, CHRIS SANDER, AND ALFONSO VALENCIA
European Molecular Biology Laboratory, Heidelberg, Germany

(RECEIVED August 26, 1992; REVISED MANUSCRIPT RECEIVED September 28, 1992)

Abstract

Kinases that catalyze phosphorylation of sugars, called here sugar kinases, can be divided into at least three distinct nonhomologous families. The first is the hexokinase family, which contains many prokaryotic and eukaryotic sugar kinases with diverse specificities, including a new member, rhamnokinase from *Salmonella typhimurium*. The three-dimensional structure of hexokinase is known and can be used to build models of functionally important regions of other kinases in this family. The second is the ribokinase family, of unknown three-dimensional structure, and comprises pro- and eukaryotic ribokinases, bacterial fructokinases, the minor 6-phosphofructokinase 2 from *Escherichia coli*, 6-phosphotagatokinase, 1-phosphofructokinase, and, possibly, inosine–guanosine kinase. The third family, also of unknown three-dimensional structure, contains several bacterial and yeast galactokinases and eukaryotic mevalonate and phosphomevalonate kinases and may have a substrate binding region in common with homoserine kinases.

Each of the three families of sugar kinases appears to have a distinct three-dimensional fold, since conserved sequence patterns are strikingly different for the three families. Yet each catalyzes chemically equivalent reactions on similar or identical substrates. The enzymatic function of sugar phosphorylation appears to have evolved independently on the three distinct structural frameworks, by convergent evolution. In addition, evolutionary trees reveal that (1) fructokinase specificity has evolved independently in both the hexokinase and ribokinase families and (2) glucose specificity has evolved independently in different branches of the hexokinase family. These are examples of independent Darwinian adaptation of a structure to the same substrate at different evolutionary times. The flexible combination of active sites and three-dimensional folds observed in nature can be exploited by protein engineers in designing and optimizing enzymatic function.

Keywords: enzymatic function; galactokinase; hexokinase; ribokinase; substrate specificity evolution; sugar kinases

Sugars are used by cells as a source of energy or carbon. The first step in sugar metabolism after transport into the cell is phosphorylation catalyzed by specific sugar kinases. All metabolic sugars are phosphorylated to trap them inside the cell and to prepare them for further chemical reactions, either catabolic or anabolic. The main catabolic pathway is glycolysis, with glucose as the key sugar. The other key sugar is ribose, essential in nucleotide biosynthesis. Other abundant sugars are fructose and galactose, whereas ribulose, xylulose, or fucose and others are rare and can only be processed by specialized microorganisms.

Advances in sequencing technology have led to a rapidly increasing number of available primary structures. In combination with recent biochemical data, this sequence information is used here to obtain some insight into the evolution of this important class of enzymes and into functionally and structurally important sequence regions. A detailed understanding of sugar kinases is particularly important in the context of human hereditary diseases, glycerol kinase deficiency (Seltzer et al., 1989), and non-insulin-dependent diabetes mellitus (Hattersley et al., 1992; Stoffel et al., 1992; Vionnet et al., 1992).

Reprint requests to: Protein Design Group, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany.

Results

The hexokinase family

Hexokinase has an ATPase domain with the same basic fold and active site as that of actin and the hsp70 family

of heat shock proteins (Bennet & Steitz, 1980; Flaherty et al., 1990, 1991; Kabsch et al., 1990; Holmes et al., 1993). Probable additional members of the family can be identified by sequence pattern searches. These include various sugar kinases as well as the prokaryotic cell cycle proteins Mreb, Ftsa, and Stba (Bork et al., 1992). The ligand ATP is located between two main domains, and its phosphate tail is bound by two beta-hairpins, one from either domain (hexokinase differs slightly from actin and hsc70 in one of the phosphate-binding beta-hairpin loops). The adenosine base is bound by a nearby loop, and two contacting helices appear to form a hinge region between the two main domains (see kinemages).

Diverse specificities

The hexokinase family (Table 1) includes eukaryotic hexokinases and glucokinases as well as prokaryotic gluco-, ribulo-, glucono-, xylulo-, glycero-, fructo-, rhamno-, and

Table 1. *The sugar kinase families*

Sequence[a]	E.C.	Pathway[b]	Sequence code[c]
Hexokinase-like		Mainly catabolic	
Hexo-	2.7.1.1	Glycolysis	HXKA_YEAST, HXKB_YEAST, HXKA_MOUSE, HXKB_HUMAN, HXKB_RAT, HXK2/RAT, HXK3/RAT, HXST/HUMAN, HXKH/HUMAN, HXKP/MOUSE, HXKC/BOVIN
Gluco-	2.7.1.2	Glycolysis	HXKG_ZYMMO, HXKG/STRC, HXKG/HUMAN, HXKH_RAT[d], HXKP_RAT, HXKG_YEAST
Fructo-	2.7.1.4	Glycolysis	FRK/ZYMMO
Ribulo-	2.7.1.16	Pentose cycle	KIRI_ECOLI, KIRI_SALTY
Glucono-	2.7.1.12	Pentose cycle	GNTK_BACSU
Xylulo-	2.7.1.17	Pentose cycle	XYLK_ECOLI, XYLK_LACPE, XYLK_ACTMI, XYLK/STRU, XYLK/ARTBA, XYLK/THERM
Glycero-	2.7.1.30	Glycolysis	GLPK_BACSU, GLPK_ECOLI
Fuco-	2.7.1.52	Pentose cycle	FUCK_ECOLI
Rhamnulo-	2.7.1.5		RHAK/SALTY
Ribokinase-like		Several roles	
Fructo-	2.7.1.4	Glycolysis, sucrose	SCRK_VIBAL, SCRK_SALTY, SCRK/KLEPN
Ribo-	2.7.1.15	Pentose cycle, nucleotide metabolism	RBSK_ECOLI, RBSK/YEAST, RBSK/PACDE
6-Phosphofructo-Minor enzyme[e]	2.7.1.11	Glycolysis?	K6P2_ECOLI
6-Phosphotagato-		Glycolysis?	K6PT/STAAU (LACC_STAAU), K6PT/STRMU, K6PT/LACLA (LACC_LACLA)
1-Phosphofructo-	2.7.1.56	Glycolysis	K1PF_ECOLI, K1PF_RHOCA, K1PF_XANCP
Inosine-guanosine-	2.7.1.73	Purine metabolism	INGK_ECOLI
Galactokinase-like		Mainly anabolic	
Galacto-	2.7.1.6	Lactose cycle	GAL1_ECOLI, GAL1_KLULA, GAL1_SACCA, GAL1_SALTY, GAL1_STRLI, GAL1/LACHE
Mevalonate-	2.7.1.36	Cholesterol synthesis	KIME_RAT, KIME_YEAST
P-mevalonate-	2.7.4.2	Cholesterol synthesis	ERG8_YEAST
Homoserine-[f]	2.7.1.39	Threonine synthesis	KHSE_BACSU, KHSE_BRELA, KHSE_CORGL, KHSE_ECOLI, KHDE_FREDI, KHSE_YEAST

[a] For the names of the sugar kinases, append the word kinase after the hyphen.

[b] Metabolic pathways in which the enzymes are active, with a question mark where the pathway is very uncertain.

[c] Sequence codes are from SWISSPROT where available and have an "_" between enzyme and species abbreviation (Bairoch & Boeckmann, 1992). Other sequences were taken from the literature (see Materials and methods for references), with names chosen by us and "/" as the separator between enzyme name and species. Detailed information for each SWISSPROT sequence including references appears on the Diskette Appendix. The codes for the species are: ACTMI, *Actinoplanes missouriensis*; ARTBA, *Artherobacter*; BACSU, *Bacillus subtilis*; BRELA, *Brevibacterium lactofermentum*; CORGL, *Corynebacterium glutamicum*; ECOLI, *Escherichia coli*; FREDI, *Fremyella diplosiphon*; KLEPN, *Klebsiella aerogenes*; KLULA, *Kluyveromyces lactis*; LACHE, *Lactobacillus helveticus*; LACLA, *Lactococcus lactis*; LACPE, *Lactobacillus pentosus*; RHOCA, *Rhodobacter capsulatus*; SACCA, *Saccharomyces carlsbergensis*; SALTY, *Salmonella typhimurium*; STAAU, *Staphylococcus aureus*; STRC, *Streptomyces coelicolor*; STRLI, *Streptomyces lividans*; STRMU, *Streptococcus mutans*; STRU, *Streptomyces rubiginosus*; THERM, *Thermus thermophilus*; VIBAL, *Vibrio alginolyticus*; XANCP, *Xanthomonas campestris*; ZYMMO, *Zymomonas mobilis*.

[d] The glucokinase sequence Hxkh_Rat differs from the second isoenzyme, Hxkp_Rat, only in the first 15 amino acids due to differential splicing of the gene. Hxkp_Rat is therefore omitted from the tree (Fig. 2).

[e] The minor isoenzyme 2 of 6-phosphofructokinase from *E. coli* has only about 10% in vitro activity compared to the major, structurally unrelated, counterpart (Daldal, 1984).

[f] Homoserine kinase shares only one conserved region with the other enzymes of this family (Fig. 5; for interpretation see text).

fucokinases. In spite of their different substrate specificities, most of the prokaryotic sugar kinases have clear overall similarities to each other (Fig. 2, right), detectable by standard homology search programs like Fasta (Pearson & Lipman, 1988). Two members difficult to detect are ribulokinase, with only about 20% identical residues relative to other members of this branch, and rhamnokinase, which was identified here by a combination of alignment and pattern searches. The member closest to rhamnokinase is fucokinase (25% amino acid identity within 409 residues; Fasta score of 266; Figs. 1, 2).

Three-dimensional models

Based on common sequence patterns in the structural core region (Bork et al., 1992; Fig. 1), a plausible three-dimensional model can be built of the core of any protein in this family, based on the structure of yeast hexokinase. Using such a model, structural and functional information can be transferred from one member of the family to the others, even when sequence similarity is low. This can be extremely useful for the identification of residues that participate in active sites or in the interpretation of mutant phenotypes. For example, Stoffel et al. (1992) used a model of human glucokinase to suggest that the possible role two point mutants, in residues Thr-228 and Gly-261, in early-onset non-insulin-dependent diabetes mellitus is to disrupt ATP and glucose binding.

Evolutionary tree

An evolutionary tree (Fig. 2) of the hexokinase family has several interesting features. The major division is into eukaryotic and prokaryotic enzymes. Within the prokaryotic subfamily there is a very wide spread of specificities to different sugars (Fig. 2, right and middle). The eukaryotic enzymes of this family (Fig. 2, left) are hexokinases (type I, II, and III isoenzymes) and glucokinases (type IV hexokinases). The mammalian hexokinases have about twice the molecular weight of all other sugar kinases of this family and contain two similar domains, probably as a result of gene duplication (Easterby, 1971; Nishi et al., 1988; Schwab & Wilson, 1989). Within one species, the mammalian glucokinases differ mainly at their N-terminal 15 amino acids as a result of differential splicing.

Nt and Ct domains of mammalian hexokinases—Functional or regulatory?

The eukaryotic members form four separate clusters (Fig. 2, left): the C-terminal domains of mammalian hexokinases, their N-terminal domains, mammalian glucokinases, and yeast sequences; each of these, except for the N-terminal domain of rat hexokinase type III, has all the residues identified, by analogy, to be involved in ATP binding (Bork et al., 1992). However, biochemical experiments indicate that the N-terminal domain is regulatory and the C-terminal domain catalytic (White & Wilson, 1989). Unfortunately, the separation in the tree of N-terminal and C-terminal domains of mammalian hexokinases, although indicative of a divergence of function, throws no further light on the possible regulatory role of the N-terminal domain.

```
HXKB_YEAST    81 DFLAIDLGGTNLRVVLVKLGG -46- SEPIPLGFTFSFPASQNKIN - 37- VVALINDTTGTLVASYYTD - 1- ETKMGVIFGTGVNGAYYD
HXKG_YEAST    80 VLLAADLGGTNFRICSVNLHG -53- AKPMKLGFTFSYPVDQTSLN - 38- VVALTNDTVGTYLSHCYTS -12- EPVIGCIFGTGTNGCYME
HXKH_RAT      73 DFLSLDLGGTNFRVMLVKVGG -47- HKKLPLGFTFSFPVRHEDLD - 38- VVAMVNDTVATMISCYYED - 1- QCEVGMIVGTGCNACYME
HXKA_MOUSE1   79 DFIALDLGGSSFRILRVQVNH -45- DKKLPVGFTFSFPCRQSKID - 38- IVAVVNDTVGTMMTCGYDD - 1- QCEVGLIIGTGTNACYME
HXKA_MOUSE2 527 DFLALDLGGTNFRVLLVKIRG -45- GPRMPLGFTFSFPCKQTSLD - 38- VVAVVNDTVGTMMTCAYEE - 1- SCEIGLIVGTGSNACYME
HXKG_ZYMMO    2 EIVAIDIGGTHARFSIAEVSS -57- NPWVLRPATLNEKLD..IID -  0- THVLINDFGAVAHAVAHMD -17- GVITILGPGTGLGVAHLL
Kiri_Salty    3 IAIGLDFGSDSVRALAVDCAT -54- ANVVGIGVDSTGSTP.APID -171- ISGGAFDCHMGAVGAGA.Q - 0- PNTLVKVIGTSTCDILIA
Glpk_Ecoli    5 YIVALDQGTTSSRAVVMDHDA -46- DQIAAIGITNQRETT.IVWE -149- ISGIAGDQQAALFGQLCVK - 0- EGMAKNTYGTGCFMLMNT
Xylk_Ecoli    1 MYIGIDLGTSGVKVILLNEQG -43- QDVKALGIAGQMHGA.TLLD -142- VVAGGGDNAAGAVGVGMVD - 0- ANQAMLSLGTSGVYFAVS
Gntk_Bacsu    4 YMLGIDIGTTSTKAVLFSENG -43- KRISFISFSSAMHSV.IAID -145- FVIGASDGVLSNLGVNAIK - 0- KGEIAVTIGTSGAIRTII
Fuck_Ecoli   15 VILVLDCGATNVRAIAVNRQG -45- CHIRGIAVTTFGVDG.ALVD -145- VISAGHDTQFALFGAGA.E - 0- QNEPVLSSGTWEILMVRS
Rhak/Salty    5 HCVAVDLGASSGRVMLARYDS -45- ILIDSIGIDTWGVDY.VLLD -140- VAVASHDTASAVIASPLAN - 0- KHSAYLSSGTWSLMGFES
                      phosphate 1                 sugar binding?              connect 1              phosphate 2

HXKB_YEAST      -144- IGARAARLSVCGI - 7- GYKTGHIAADGSVYNRYPGFKEKAANALKDIY -13- IVPAEDGSGAGAAVIAALAQ
HXKG_YEAST      -152- ISRRSAYLAAVPL -13- YHGEVEIGCDGSVVEYYPGFRSMLRHALALSP -11- LKIAKDGSGVGAALCALVA*
HXKH_RAT        -138- VSTRAAHMCSAGL -13- DVMRITVGVDGSVYKLHPSFKERFHASVRRLT - 5- FIESEEGSGRGAALVSAVAC
HXKA_MOUSE1     -138- VSFRSANLVAATL -13- PRLRTTVGVDGSLYKMHPQYSRRFHKTLRRLV - 5- FLLSESGSGKGAAMVTAVAY
HXKA_MOUSE2     -138- VSKRAAQLCGAGM -13- DHLNVTVGVDGTLYKLHPHFSRIMHQTVKELS - 5- FLLSEDGSGKGAALITAVGV
HXKG_ZYMMO       - 34- RFRRVSIERIISG - 4- NIYEALAAIEGVPFSLLDDIKLWQMALEGKDN -21  LAQGRTSVVIGGGVGLRIAS
Kiri_Salty      -131- FGARAIQECFTDQ - 0- GIAVNNVMALGGIARKNQVIMQVCCDVLNRPL - 1- IVASDQCCALGAAIFAAVAA
Glpk_Ecoli       - 89- HIIRATLESIAYQ -12- GIRLHALRVDGGAVA.NNFLMQFQSDILGTRV - 1- RPEVREVTALGAAYLAGLAV
Xylk_Ecoli       - 87- ELARAVLEGVGYA -11- GIKPQSVTLIGGGAR.SEYWRQMLADISGQQL - 2- RTGGDVGPALGAARLAQIAA
Gntk_Bacsu       - 97- HMIRAALEGVIYN -12- DGPVTRIQATGGFAR.SEVWRQMMSDIFESEV - 1- VPESYESSCLGACILGLYAT
Fuck_Ecoli       - 83- HFYRAALEGLTAQ -12- HFKASELLLVGGGSR.NTLWNQIKANMLDIPV - 1- VLDDAETTVAGAALFGWYGV
Rhak/Salty       - 88- ELARCIFDSLALL -12- GEKFTQLHIVGGGCQ.NSLLNQLCADACGIRV - 0- MAGPVEASTLGNIGIQLMTL
                         helix                      adenosine                      connect 2
```

Fig. 1. Hexokinase family: aligned conserved regions of a representative subset of family members. Numbers between the blocks indicate the length of intervening segments. Protein names indicate the name of the protein followed by a species identifier, taken from SWISSPROT when available (e.g., Hxkb_Yeast) or chosen by us (e.g., Rhak/Salty). A list of all the enzymes used is in Table 1. Bottom line: functional and structural identification of the conserved regions, in analogy to the known 3D structure of yeast hexokinase: phosphate 1 and 2 and adenosine are involved in ATP binding, connect 1 and connect 2 in putative hinge motion (Bork et al., 1992); sugar binding corresponds to a possible glucose-binding site in hexokinase (Arora et al., 1990); and the helix is an additional conserved region common to the hexokinase-like sugar kinases.

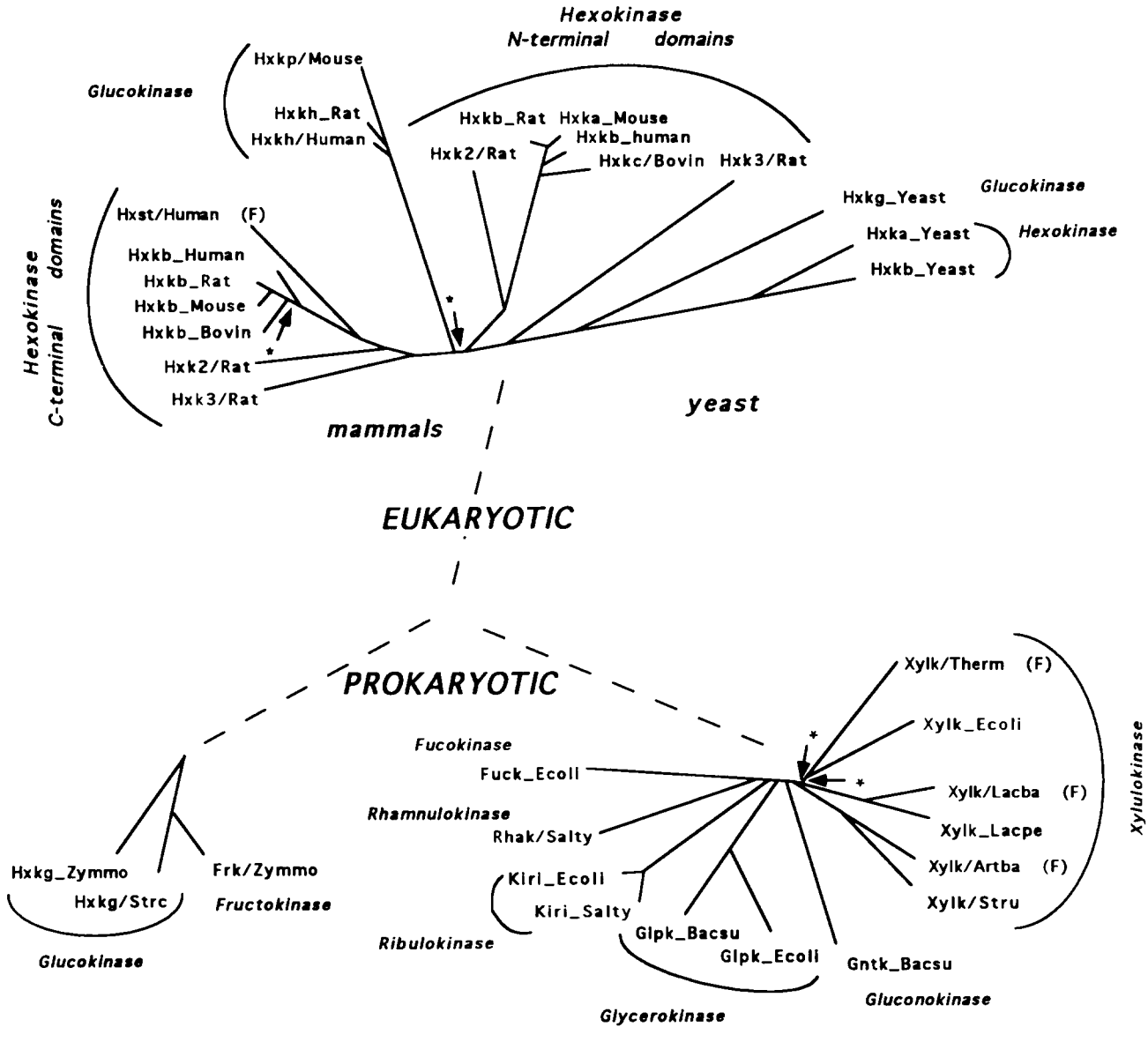

Fig. 2. Hexokinase family: evolutionary tree of all currently available family members. The distances from the root of the tree are undetermined (dotted lines) as the similarity between the three subfamilies is limited to the seven conserved boxes. The branch lengths are roughly proportional to the amount of sequence difference, as calculated using the program Clustal V (Higgins et al., 1992) with full-length sequences as input. An asterisk marks lack of confidence in the precise branching order (confidence less than 95% in the bootstrapping option of Clustal V). (F) indicates sequence fragments. The N-terminal and C-terminal domains of mammalian hexokinases, a probable result of gene duplication, are entered separately (left side of figure). All proteins in the tree probably have a fold similar to that of yeast hexokinase and actins as well as of the ATPase domains of hsp70 heat shock proteins and some prokaryotic cell cycle proteins (Bork et al., 1992). The evolutionary distances (sums of branch lengths) between the xylulokinases of different species are as large as those between the xylulokinases and enzymes of other substrate specificities (right) indicating that substrate specificity is only one of several evolutionary constraints governing sequence variation. The presence of glucokinases in three separate branches (two in the left subfamily; one in the center one) is suggestive of multiple independent evolution of this specificity. Note also that fructokinase activity (center of figure) appears to have evolved independently in this structural family—fructokinases with a different fold are in Figure 4. Notation of protein names as in Table 1.

Independent evolution of glucose specificity

An interesting feature of the tree (Fig. 2) is the fact that glucokinases appear in three clusters in separate branches: (1) mammalian glucokinases form one cluster (Fig. 2, far left), (2) yeast glucokinase appears to cluster with yeast hexokinases rather than with mammalian glucokinases (Fig. 2, left), and (3) bacterial glucokinases from *Zymomonas mobilis* and *Streptomyces coelicolor* are grouped with *Zymomonas* fructokinase (Fig. 2, middle). A divergent evolutionary relation between these

apparently disconnected clusters is unlikely. Rather, evolutionary convergence to glucose specificity independently in mammals, yeast, and bacteria probably must be invoked. It is unlikely, however, that identical active site constellations are responsible for glucose specificity in the three cases. Convergence may be limited to similarity in the ability to bind glucose specifically and may not extend to the precise way in which this is achieved. Overall, the tree of hexokinase-related sugar kinases present a remarkable diversity of specificities (Table 1), all achieved by fine tuning of specific regions in the same basic three-dimensional framework.

The ribokinase family

The second structurally distinct family is composed of prokaryotic sequences related to ribokinase (Table 1), including several fructokinases, the minor 6-phosphofructokinase from *Escherichia coli*, 1-phosphofructokinases, and 6-phosphotagatokinases, with some similarities already described (Daldal, 1983, 1984; Blatch et al., 1990; Orchard & Kornberg, 1990; van Rooijen et al., 1991) and reviewed by Wu et al. (1991). Here we have attempted to assemble a complete set of the currently known members of this family (Fig. 3) that is also called the pfkb family in the Prosite collection of sequence patterns (Bairoch, 1992). The first eukaryotic member of the family was recently identified in yeast (Thierry et al., 1992). Multiple alignment, identification of conserved regions, and a subsequent database search with corresponding property patterns (Bork & Grunwald, 1990) allowed us to recognize an additional member of this family, inosine-guanosine kinase from *E. coli* (Table 1; Fig. 3), which also is a sugar kinase in that it phosphorylates the ribose of inosine or guanosine.

Divergence of specificity before species divergence

The evolutionary tree (Fig. 4) based on the multiple sequence alignment (Fig. 3) suggests that in this family divergence of specificity occurred before species divergence: fructokinases branched off from ribokinases (Fig. 4, middle and right) before the divergence of yeast and *E. coli* (Fig. 4, right). That substrate specificity is the major organizing principle of the tree is also evident from the fact that phosphosugar kinases form a major cluster (Fig. 4, left), with early divergence of the approximately equidistant subfamilies 1-phosphofructokinases (K1pf), minor 6-phosphofructokinases (K6p2), and 6-phosphotagatokinases (K6pt). Apparently, the changes required to change specificity from the 1 to the 6 hydroxyl position of phosphofructose are about as difficult as the change in specificity from phosphofructose to phosphotagatose.

The galactokinase family

The third structurally distinct family (Fig. 5) is characterized by three highly conserved boxes (Adams et al., 1988). The sequence segments between these boxes are highly variable in length and amino acid composition. We constructed consensus patterns for the three conserved regions and clearly identified rat mevalonate kinase (Tanaka et al., 1990), its yeast equivalent (RAR1; Kearsey & Edwards, 1987; Oulmouden & Karst, 1991) and phosphomevalonate kinase (EC 2.7.4.2) as members of this family. The regions most conserved between the mevalonate kinases are exactly the conserved boxes of the galactokinases (Schafer et al., 1992). The second box is particularly strongly conserved. A further search with a sequence pattern for only the second box reveals a closely related sequence box in homoserine kinases (EC 2.7.1.39). This box is the most conserved of five conserved regions in homoserine kinases (Mannhaupt et al., 1990). None of the other four match any of the remaining patterns of galactokinase. These facts suggest a common function for the single conserved segment of galactose, mevalonate, phosphomevalonate, and homoserine kinases (e.g., an ATP binding site, as proposed by Tsai & Robinson [1991]).

Jumping into another pathway?

Mevalonate kinase, phosphomevalonate kinase, and homoserine kinase are examples of enzymes homologous to sugar kinases but located in different metabolic pathways. This homology raises an interesting evolutionary

```
K6pt/Lacla   35 AGGKGLNVT...RVLYESGDKVTATGFLGGK -152- GIEWIVVSLGRNGAFAK -11- DIPVVNPVGSGDSTVAGIASAL
K6pt/Strmu   35 PGGKGLNVT...RVLSQLGDDVLASGLLGGK -152- GIDWIIVSLGSQGAFAK -11- KIAVVNPVGSGDSTVAGITSAL
K6pt/Staau   35 AGGKGLNVT...RVLAQVGEPVLASGFIGGE -152- GIEWIIVSLGAQGAFAK -11- TISVLNPVGSGDSTVAGITSAI
K1pf_Xancp   38 AGGKGINVA...ACLADWGSQVAALGVLGVG -154- GIQLVVISMGTEGALFV -11- RLAQGSSVGAGDAMVAGLAAAL
K1pf_Rhoca   38 AGGKGVNVA...SFLAHVGHGVAVTGLLGAE -151- GVGLVAVSMGAGGAVLV -11- ATPIASTVGAGDAMVAGLIHAA
K1pf_Ecoli   38 AAGKGINVA...KVLKDLGIDVTVGGFLGKD -149- GIAHVVISLGAEGALVW -11- SVDVVSTVGAGDSMVGGLIYGL
K6p2_Ecoli   37 PGG.GINVA...RAIAHLGGSATAIFPAGGA -150- KAKRVVVSLGPQGALGV -11- ALKSQSTVGAGDRLVGAMTLKL
Scrk/Klepn   27 PGGAPANVA...VGVARLGGDSGFIGRVGDD -155- QPTLLLVTQGKAGVQAA -11- PVVAVDTTGAGDAFVAGLLAGL
Scrk_Salty   27 PGGAPANVA...VGVARLGGNSGFIGAVGGD -155- QPELLLVTRGKAGVLAA -11- PVASVDTTGAGDAFVAGLLASL
Scrk_Vibal   26 PGGAPANVA...VAIARLSGKSAFFGRVGDD -154- NIALVLVTQGAKGVWRV -11- VVSPIDTTGAGDAFVGGLLACL
Rbsk_Ecoli   40 FGGKGANQA...VAAGRSGANIAFIACTGDD -147- GIRTVLITLGSRGVWAS -11- RVQAVDTIAAGDTFNGALITAL
Rbsk_Yeast   36 AGGKGLNQA...AAIGKLKNPSSRYSVRMIG -176- KRGIVVMTLGSRGVLFC -14- NVSVVDTTGAGDTFLGGLVTQL
Ingk_Ecoli   92 AGGTIGNTMHNYSVLADDRSVLLGVMCSNIE -156- WVDLVLCTAGPIGLYMA -55- PEKIMNTNGAGDGALAALLHDI
```

Fig. 3. Ribokinase family: aligned conserved regions. Each of the three sequence "boxes" alone is sufficient to identify the members of the family detected so far. Substrate specificities and protein names are as in Table 1.

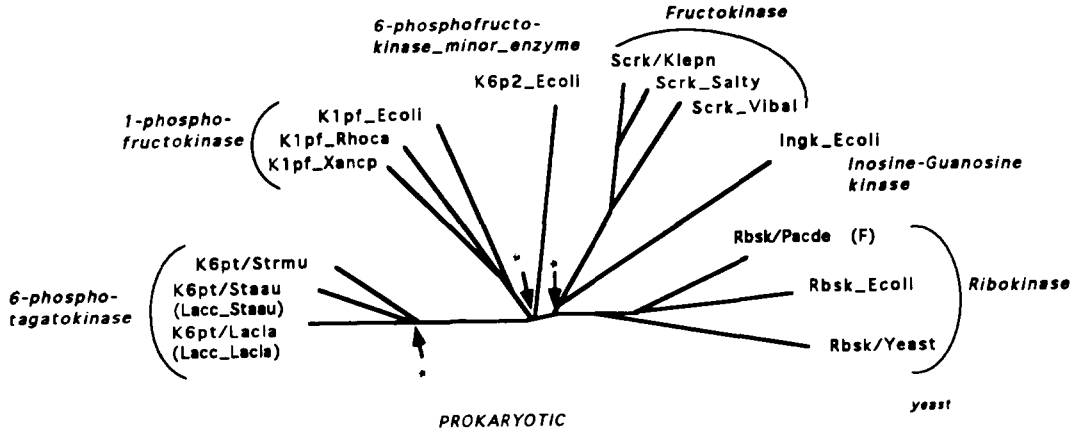

Fig. 4. Ribokinase family: evolutionary tree. Calculated with the program Clustal V using full-length sequences. Fructokinases in this family are structurally distinct from those in the hexokinase family (Fig. 2), a result of independent (convergent) evolution. Notation of protein names as in Table 1.

question. Homoserine kinase catalyzes an important step in the biosynthesis of the amino acid threonine; all other enzymes of this pathway are pyridoxal phosphate-dependent, homologous to one another, and have apparently evolved via several gene duplication events (Bork & Rohde, 1990). So how does homoserine kinase fit in? Perhaps the energy-consuming ATP-dependent homoserine kinase has been inserted into the pathway during evolution to overcome an energy barrier in threonine synthesis. Another puzzling observation is that mevalonate kinase is apparently less similar in sequence to phosphomevalonate kinase (related substrates in subsequent steps of the cholesterol synthesis pathway) than it is to galactokinase (sugar substrate in a completely different pathway).

The galactokinase family, like the other two major families of sugar kinases, has a wide diversity of substrate specificity, all accommodated within the same structural framework, in spite of the fact that the chemistry of the substrates for galactokinase and mevalonate kinase is quite different (phosphorylation of the first versus the last hydroxyl group).

Conclusions

Three distinct families of sugar kinases

We have compared more than 60 sequences of sugar kinases with 20 different sugar-binding activities. Three major families can be distinguished, here called the hexokinase, ribokinase, and galactokinase families, corresponding to three biologically abundant sugars. There are strong reasons to believe that these three sugar kinase families have different three-dimensional structures: (1) there is no significant sequence similarity between the three families; (2) quality, length, and distribution of the conserved boxes within each family are different; and (3) secondary structure predictions result in very different orders and frequencies of alpha-helices and beta-strands (data not shown).

```
GAL1_ECOLI galactose      19 THTIQAPGRVNLIGEHTDYNDGFVLPCAID - 62- GVDMVISGNVPQGAGLSSSASLEVAVGTV -195- KGGVRMTGGGFGGCIVALIPEEL
GAL1_SACCA     "          44 DFVARSPGRVNLIGEHIDYCDFSVLPLAID - 78- GLQVFCEGDIPTGSGLSSSAAFICAVALA -277- SYGSRLTGAGWGGCTVHLVPGGP
GAL1_KLULA     "          35 FFITRSPGRVNLIGEHIDYCQFHV.PMASE - 75- GMEIYVKGDIPSGGGLSSSAAFICAVSLA -275- SFGSRLTGAGWGGCTVHLCSTDT
GAL1/LACHE     "          18 KDVFFSPGRINVIGEHTDYNGGHVFPAPIS - 66- GFNLYIEANLPSGSGLSSSAAIEMLMGII -192- VLGARMIGGGFGGSAIAIVKKSE
GAL1_STRLI     "          22 RGCGRRAGRENLIGEHTDYNDGFVMPSPCR - 60- GADVHLASTVPSGAGLSSSAALEVRPLAM -195- GPRRRMTGGGFGGSAIVLVEAAA
GAL3_YEAST gal3 protein   39 ILSLGLLGRVNLIGEHIDYCDFSVLPLAID - 78- GAQIFCQSDIPTGGGLSSAFTCAGRLATI *
KIME_RAT   mevalonate      5 VLLVSAPGKVILHGEHAVVHGKVALAVALN - 93- SLDIMVWSELPPGAGLGSSAAYSVCVAAA -169- GLHSKLTGAGGGGCGITLLKPGL
KIME_YEAST     "           4 PFLTSAPGKVIIFGEHSAVYNKPAVAASVS - 97- NIKFSLKSTLPIGAGLGSSASISVSLALA -184- IGSTKLTGAGGGGCSLTLLRRDI
ERG8_YEAST P-mevalonate    4 LRAFSAPGKALLAGGYVLDTKYEAFVVGLS -108- SFHSHRIEEVP.KTGLGSSAGLVTVLTTA -322- VLTCLIPGAGGYDAIAVLTKQDV
KHSE_BACSU homoserine                                          80 PVHVKVWSDIPLARGLGSSAAAIVAAIEL
KHSE_BRELA     "                                               85 GLRVVCHNNIPQSRGLGSSAAAAVAGVAA
KHSE_CORGL     "                                               85 GLRVVCHNNIPQSRGLGSSAAAAVAGVAA
KHSE_ECOLI     "                                               81 PVAMTLEKNMPIGSGLGSSACSVVAALMA
KHSE_FREDI     "                                               82 SVKIEIDLGVPLARGLGSSATAIVGGLVA
KHSE_YEAST     "                                               92 GTKVHVSNPIPLGRGLGSSGAAVVAGVIL
```

Fig. 5. Galactokinase family: aligned conserved regions in enzymes structurally related to galactokinases. The pattern describing the central box is the only one in common with homoserine kinase and is the most stringent and discriminating pattern (the random background of unrelated enzymes is separated from the true positive sequences by more than eight mismatches, see Materials and methods). Homoserine kinase fits the pattern with only one mismatch and therefore is proposed to belong to the family. For interpretation see text. Substrate specificities and protein names are as in Table 1.

Why different structural families?

If sugar kinase activity with a wide range of specificities can be constructed on different structural frameworks, then why do we see at least three distinct structural families? Are there subtle ways in which one particular structure performs the function of sugar phosphorylation differently? To answer these questions definitively we need much more biochemical information about each of the enzymes: details of the catalytic mechanism, rate constants, quantitative specificities, and regulatory dependencies. Lacking such information, we can only state some possible reasons. Perhaps at a fairly early stage in evolution, the three classes reflected different basic metabolic roles of different basic sugars, such as glucose as a source of energy or ribose as a central raw material in the pentose phosphate cycle and in nucleic acid biosynthesis. Alternatively, different folds are perhaps needed for different types of metabolic regulation. There is another possibility—nature may simply have reinvented kinase activity on different occasions as a result of historical accident, exploiting different protein folds available at that time in that organism when one basic fold would have done just as well. Whatever happened early in evolution, the three frameworks have, by now, adapted to very diverse specificities in different organisms.

Convergent evolution: Independent evolution of similar specificity

In addition to the surprisingly broad distribution of specificities, the evolutionary trees of the hexokinase and ribokinase families provide multiple evidence of convergent evolution in the sense that similar specificity evolved independently in different structural families (examples: fructokinases) or in different branches of the same structural family (example: glucokinase). Two specific examples follow.

Fructokinases

Flexible adaptation of different structural frameworks to the ability of phosphorylating a particular sugar is evident for fructokinases. The *Zymomonas* fructokinase (Frk/Zymmo; Fig. 2, middle; Zembrzuski et al., 1992) belongs to the hexokinase family, has all of the characteristic sequence patterns of this family, and can be significantly aligned with other members of this family, e.g., it has 26% amino acid identity over 300 residues with *Streptomyces* glucokinase. The other fructokinases are members of the ribokinase family (Fig. 3; Fig. 4, middle), i.e., they are built onto a different three-dimensional structure.

Glucokinases

Also, within one structural family we observe independent adaptation to a similar specificity. For example, the ability to specifically select a glucokinase substrate has arisen at least twice in the hexokinase family, apparently independently. Yeast and mammalian glucokinases (Hxkg_Yeast, Hxkh_Rat) are in the left branch of Figure 2, grouped with hexokinases, whereas the bacterial glucokinases Hxkg_Zymmo, are in a separate branch (Fig. 2, middle), grouped with fructokinases. Apparently independent evolution of similar specificity (substrate: adrenaline) has also been observed in different branches of the evolutionary tree of G-protein-coupled receptors (Doolittle, 1991). We use the term convergent evolution to describe the independent evolution of the same specificity, the same or different structural scaffolds, as well as the independent evolution of similar catalytic function on different structural frameworks—without implying that details of active site geometry are necessarily similar.

Gene duplication, not exon shuffling

What are the mechanisms for generating so much diversity so easily and occurring so many times? We cannot entirely exclude some kind of domain shuffling mechanism, but at least for the hexokinase family, where we know the three-dimensional structure, there is no obvious compact structural module that carries sugar specificity, nor have we so far seen any sequence evidence for domain shuffling. We do have some evidence that genomic organization (data not shown) appears to be uniform in one family and different between the different structural families. In several cases, a set of enzymes belonging to the same metabolic pathway are coded on one operon. In these cases the entire operon appears to have been copied as a unit before specificity diverged or converged—a very simple and the most plausible mechanism.

Convergent evolution: Other examples

Evolution of similar enzymatic function on different structural frameworks is not an entirely uncommon event (Fig. 6A). A classical example is that of serine proteases: the Ser-His-Asp triad is present in an almost identical three-dimensional constellation in the distinctly different structural frames of trypsin and subtilisin (and their relatives) (Wright et al., 1969). Two more recent examples: there are at least two structurally distinct classes of tRNA synthases (Nagel & Doolittle, 1991) and two different structural classes of superoxide dismutases (Smith & Doolittle, 1992). On several occasions nature appears to have reinvented similar enzymes and even similar specificity using distinctly different three-dimensional folds.

Flexible adaptation of protein structures to diverse functions

These examples of convergent evolution support the idea that active sites can be adapted quite easily, whereas three-

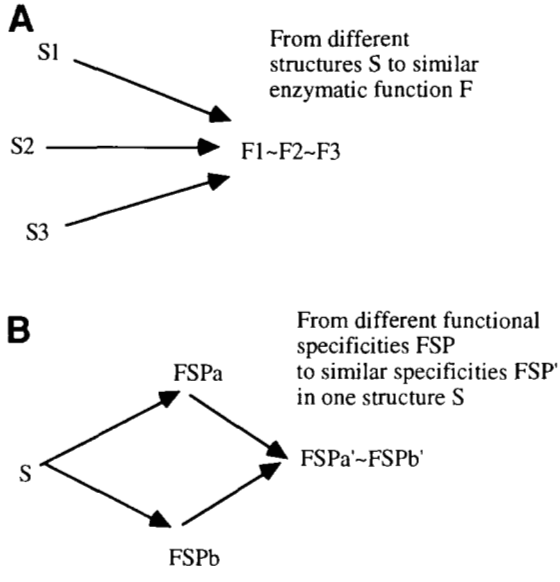

Fig. 6. Two kinds of convergent evolution. **A:** Independent evolution of functionally equivalent active sites in each of three different three-dimensional frameworks. **B:** Independent evolution of the same substrate specificity in different branches of one structural family.

dimensional folds are extremely persistent in evolution. Not only does this appear to be a basic principle of the natural evolution of enzymes, but it is of practical value in protein engineering. In designing and optimizing enzymatic function, the protein engineer can, following nature, rather freely combine different protein structural scaffolds with a variety of specific active site architectures.

In summary, sugar kinases can be grouped by their sequence similarity (and probably by their three-dimensional structures) into three distinct families. Convergent evolution has resulted in functionally equivalent active sites in each of the three different frameworks (Fig. 6A). Within the structurally related families the substrate specificity can be modified without affecting the basic protein fold (Fig. 6B).

Materials and methods

Sequence data

Sequences were taken from the SWISSPROT protein sequence database (release 21; Bairoch & Boeckmann, 1992) or from the literature. All the identifiers are given in Table 1. Sequences from sources other than SWISSPROT are: rat hexokinases type II (Hxk2/Rat; Thelen & Wilson, 1991) and type III (Hxk3/Rat; Schwab & Wilson, 1991), human liver glucokinases (Hxkh/Human; Tanizawa et al., 1991), mouse pituitary glucokinase (Hxkp/Mouse; Hughes et al., 1991), a fragment of a putative isoenzyme of human brain hexokinase (Hxst/Human; Adams et al., 1992), liver cattle hexokinase (Hxkc/Bovin; EMBL accession number M65140; partly published in Griffin et al., 1989), glucokinase from *Streptomyces coelicolor* (Hxkg/Strc; EMBL accession number X65932; S. Angell, E. Schwarz, & M.J. Bibb, unpubl.), rhamnokinase from *Salmonella typhimurium* (Rhak/Salty; Nishitani & Wilcox, 1991), xylose kinases (Xylk/Stru from *Streptomyces rubiginosus*, EMBL accession number M73789, H.C. Wong et al., unpubl.; Xylk/Artba from *Artherobacter*, Loviny-Anderton et al., 1991; Xylk/Lacba from *Lactobacillus brevis*, Bor et al., 1992; Xylk/Therm from *Thermus thermophilus*, Dekker et al., 1991), yeast ribokinase (Rbsk/Yeast; Thierry et al., 1992), fructokinases from *Klebsiella pneumoniae* (Scrk/Klepn; Aulkmeyer et al., 1991) and *Zymomonas mobilis* (Frk/Zymmo; Zembrzuski et al., 1992), 6-phosphotagatokinase from *Streptococcus mutans* (K6pt/Strmu; EMBL accession number M80797; Rosey & Steward, 1992), and galactokinase from *Lactobacillus helveticus* (Gal1/Lache; Mollet & Pilloud, 1991).

Property pattern searches in sequence databases

In order to identify additional members of each family, property pattern searches (Bork & Grunwald, 1990) were carried out as follows: The most conserved regions were extracted from a multiple sequence alignment and to each position of these regions a set of steric and physicochemical amino acid properties were assigned. The amino acid property pattern was then translated to a consensus pattern of preferred amino acids, providing a generalization of the original observed amino acid frequencies at each position, and the database of known sequences was screened for similarities. To identify remotely related proteins, deviations from the consensus pattern (mismatches) were allowed. A mismatch is a deviating amino acid property relative to the pattern. In this way a single amino acid difference can be translated to more than one mismatch depending on its physicochemical relation with the other amino acids present in this position. The overall mismatch score is calculated as the sum over all positions of the pattern. If the mismatch score for a particular sequence is well below the random background of scores of apparently nonrelated proteins, this is interpreted as a sequence homology, and the sequence is included in the family for the next iteration of multiple sequence alignment, pattern definition, and database search. Depending on the permissiveness of the cutoffs used, this procedure either explodes by including very many unrelated proteins, not well separated from background, or it converges to a situation where the scores for all family members are well separated from the background. Convergence after a few iterations, with minor adjustments of the multiple sequence alignment, is taken as a strong indication of significance. We use this empirical rule, as a mathematically sound statistical significance estimate has yet to be derived for pattern searches based on multiple sequence alignment insertions and deletions.

Other sequence alignment tools

As an additional, independent check, all sequences included in a family were individually compared with the protein sequence database using the program Fasta (Pearson & Lipman, 1988). Multiple sequence alignments were carried out using MaxHom (Sander & Schneider, 1991) or Clustal V (Higgins et al., 1992). Tree calculations were performed by the neighbor-joining method of (Saitou & Nei, 1987) using Kimura's correction for multiple hits (Kimura, 1983) as implemented in Clustal V (Higgins et al., 1992). Accordingly, insertions/deletions were not considered in calculation of sequence distances. The reliability of the sequence distances represented by the length of tree branches was estimated with a bootstrapping method included in the Clustal V package. The bootstrapping method assesses the statistical reliability by generating a large set (5,000 in this case) of independent trees, each of them representing a different random selection of the sequence positions from the multiple sequence alignment. Thus, the fewer the positions of the multiple sequence alignment that support a grouping the lower is the reliability of the corresponding branch.

Note added in proof

After acceptance of our manuscript, the sequence of a human galactokinase with similarity to the respective family described here was published (Lee et al., 1992). This sequence contains the consensus patterns and underlines the widespread occurrence of this family of sugar kinases.

Acknowledgments

We thank D. Higgins for modifying Clustal V and the EMBL protein design group for help and discussions. We are grateful to A. Bairoch, M. Saraste, D. Higgins, and L. Serrano for critical reading of the manuscript; to J.E. Wilson for extensive discussion and critical comments; and to E. Melendez for valuable advice.

References

Adams, C.W., Fornwald, J.A., Schmidt, F.J., Rosenberg, M., & Brawner, M.E. (1988). Gene organization and structure of the *Streptomyces lividans* gal operon. *J. Bacteriol. 170*, 203–212.

Adams, M.D., Dubnich, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, S.C., & Venter, C. (1992). Sequence identification of 2375 human brain genes. *Nature 355*, 632–634.

Arora, K.K., Fanciulli, M., & Pedersen, P.L. (1990). Glucose phosphorylation in tumor cells. *J. Biol. Chem. 265*, 6481–6488.

Aulkmeyer, P., Ebner, R., Heilmann, G., Jahreis, K., Schmid, K., Wrieden, S., & Lengeler, J.W. (1991). Molecular analysis of two fructokinases involved in sucrose metabolism of enteric bacteria. *Mol. Microbiol. 5*, 2913–2922.

Bairoch, A. (1992). PROSITE: A dictionary of sites and patterns in proteins. *Nucleic Acids Res. 20*, 2013–2018.

Bairoch, A. & Boeckmann, B. (1992). The Swiss-Prot protein sequence data bank. *Nucleic Acids Res. 20*, 2019–2022.

Bennet, W.S. & Steitz, T.A. (1980). Structure of a complex between yeast hexokinase A and glucose. *J. Mol. Biol. 140*, 183–230.

Blatch, G.L., Scholle, R.R., & Woods, D.R. (1990). Nucleotide sequence and analysis of the *Vibrio alginolyticus* sucrose uptake-encoding region. *Gene 95*, 17–23.

Bor, Y.C., Oraes, L., Lee, S.P., Crosby, W.L., Sinsken, A.J., & Butt, C.A. (1992). Cloning and sequencing the *Lactobacillus brevis* gene encoding xylose isomerase. *Gene 114*, 127–131.

Bork, P. & Grunwald, C. (1990). Recognition of different nucleotide-binding sites in primary structures using a property pattern approach. *Eur. J. Biochem. 191*, 347–358.

Bork, P. & Rohde, K. (1990). Sequence similarities between tryptophan synthase b subunit and other pyridoxal-phosphate-dependent enzymes. *Biochem. Biophys. Res. Commun. 171*, 1316–1325.

Bork, P., Sander, C., & Valencia, A. (1992). An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsc70 heat shock proteins. *Proc. Natl. Acad. Sci. USA 89*, 7290–7294.

Daldal, F. (1983). Molecular cloning of the gene for phosphofructokinase-2 of *E. coli* and the nature of a mutation, pfkB1, causing a high level of the enzyme. *J. Mol. Biol. 167*, 285–305.

Daldal, F. (1984). Nucleotide sequence of gene pfkB encoding the minor phosphofructokinase of *E. coli* K-12. *Gene 28*, 337–342.

Dekker, K.A., Yamagoto, H., Sakaguchi, K., & Udaka, S. (1991). Xylose (glucose) isomerase gene from *Thermus thermophilus*: Cloning, sequencing, and comparison with other thermostable xylose isomerases. *J. Bacteriol. 173*, 3078–3083.

Doolittle, R.F. (1991). New perspectives on evolution provided by protein sequences. In *New Perspectives on Evolution* (Leonard, W., Ed.), pp. 165–173. Wiley-Liss, New York.

Easterby, J.S. (1971). The polypeptide chain molecular weight of a mammalian hexokinase. *FEBS Lett. 18*, 23–26.

Flaherty, K.M., Deluca-Flaherty, C., & McKay, D.B. (1990). Three-dimensional structure of the ATPase fragment of a 70k heat shock cognate protein. *Nature 346*, 623–628.

Flaherty, K.M., McKay, D.B., Kabsch, W., & Holmes, K.C. (1991). Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kd heat shock cognate protein. *Proc. Natl. Acad. Sci. USA 88*, 5041–5045.

Griffin, L.D., MacGregor, G.R., Muzny, D.M., Harten, J., Cook, R.G., & McCabe, E.R. (1989). Synthesis and characterization of a bovine hexokinase I (HK1) cDNA probe by mixed oligonucleotide primed amplification of cDNA using high complexity primer mixtures. *Biochem. Metab. Biol. 41*, 125–131.

Hattersley, A.T., Turner, R.C., Permutt, M.A., Patel, P., Tanizawa, Y., Chiu, K.C., O'Rahilly, S., Watkins, P.J., & Wainscoat, J.S. (1992). Linkage of type 2 diabetes to the glucokinase gene. *Lancet 339*, 1307–1310.

Higgins, D.G., Bleasby, A.J., & Fuchs, R. (1992). CLUSTAL V: Improved software for multiple sequence alignment. *CABIOS 8*, 189–191.

Holmes, K.C., Sander, C., & Valencia, A. (1993). A new ATP-binding fold for actin, hexokinase and Hsc70. *Trends Cell Biol.*, in press.

Hughes, S.D., Quaade, C., Milburn, J.L., Cassidy, L., & Newgard, C.B. (1991). Expression of normal and novel glucokinase mRNAs in anterior pituitary islet cells. *J. Biol. Chem. 266*, 4521–4530.

Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F., & Holmes, K.C. (1990). Atomic structure of the actin-DNase I complex. *Nature 347*, 37–44.

Kearsey, S.E. & Edwards, J. (1987). Mutations that increase the mitotic stability of minichromosomes in yeast: Characterization of RAR1. *Mol. Gen. Genet. 210*, 509–517.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, England.

Lee, R.T., Peterson, C.L., Calman, A.F., Herskowitz, I., & O'Donnell, J.J. (1992). Cloning of a human galactokinase gene (GK2) on chromosome 15 by complementation in yeast. *Proc. Natl. Acad. Sci. USA 89*, 10887–10891.

Loviny-Anderton, T., Shaw, P.C., Shin, M.K., & Hartley, B.S. (1991). D-Xylose (D-glucose) isomerase from *Artherobacter* strain N.R.R.L. B3728. *Biochem. J. 277*, 263–271.

Mannhaupt, G., Pohlenz, H.-D., Seefluth, A.K., Pilz, U., & Feldmann, H. (1990). Yeast homoserine kinase. *Eur. J. Biochem. 191*, 115–122.

Mollet, B. & Pilloud, N. (1991). Galactose utilization in *Lactobacillus helveticus*: Isolation and characterization of the galactokinase (galK)

and galactose-1-phosphate uridyl transferase (galT) genes. *J. Bacteriol. 173*, 4464–4473.

Nagel, G.M. & Doolittle, R.F. (1991). Evolution and relatedness in two aminoacyl-tRNA synthetase families. *Proc. Natl. Acad. Sci. USA 88*, 8121–8125.

Nishi, S., Seino, S., & Bell, G.I. (1988). Human hexokinase: Sequences of amino- and carboxyl-terminal halves are homologous. *Biochem. Biophys. Res. Commun. 157*, 937–943.

Nishitani, J. & Wilcox, G. (1991). Cloning and characterisation of the L-rhamnose regulon in *Salmonella typhimurium* LT2. *Gene 105*, 37–42.

Orchard, L.M. & Kornberg, H.L. (1990). Sequence similarities between the gene specifying 1-phosphofructokinase (fruK), genes specifying other kinases in *E. coli* K12, and lacC of *Staphylococcus aureus*. *Proc. R. Soc. Lond. B 242*, 87–90.

Oulmouden, A. & Karst, F. (1991). Nucleotide sequence of the ERG12 gene of *Saccharomyces cerevisiae* encoding mevalonate kinase. *Curr. Genet.* 19, 9–14.

Pearson, W.R. & Lipman, D.J. (1988). Improved tools for biological sequence comparisons. *Proc. Natl. Acad. Sci. USA 85*, 2444–2448.

Rosey, E.L. & Steward, G.C. (1992). Nucleotide and deduced amino acid sequence of the lacR, lacABCD, and lacFE genes encoding the repressor, tagatose 6-phosphate gencluster and sugar-specific PTS components of the lactose operon of *Streptococcus mutans*. *J. Bacteriol. 174*, 6159–6170.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4*, 406–425.

Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins Struct. Funct. Genet. 9*, 56–68.

Schafer, B.L., Bishop., R.W., Kratunis, V.J., Kalinovski, S.S., Mosley, S.T., Gibson, K.M., & Tanaka, R.D. (1992). Molecular cloning of human mevalonate kinase and identification of a missense mutation in the genetic disease mevalonic aciduria. *J. Biol. Chem. 267*, 13229–13238.

Schwab, D.A. & Wilson, J.E. (1989). Complete amino acid sequence of rat brain hexokinase, deduced from the cloned cDNA, and proposed structure of a mammalian hexokinase. *Proc. Natl. Acad. Sci. USA 86*, 2563–2567.

Schwab, D.A. & Wilson, J.E. (1991). Complete amino acid sequence of the type III isozyme of rat hexokinase, deducted from the cloned cDNA. *Arch. Biochem. Biophys. 285*, 365–370.

Seltzer, W.K., Angelini, C., Dhariwal, G., Ringel, S.P., & McCabe, E.R. (1989). Muscle glycerol kinase in Duchenne dystrophy and glycerol kinase deficiency. *Muscle & Nerve 12*, 307–313.

Smith, M.W. & Doolittle, R.F. (1992). A comparison of evolutionary rates of the two major kinds of superoxide dismutase. *J. Mol. Evol. 34*, 175–184.

Stoffel, M., et al. (1992). Human glucokinase gene: Isolation, characterization, and identification of two missense mutations linked to early-onset non-insulin-dependent (type 2) diabetes mellitus. *Proc. Natl. Acad. Sci. USA 89*, 7698–7702.

Tanaka, R.D., Lee, L.Y., Schäfer, B.L., Kratunis, V.J., Mohler, W.A., Robinson, G.W., & Mosley, S.T. (1990). Molecular cloning of mevalonate kinase and regulation of its mRNA levels in rat liver. *Proc. Natl. Acad. Sci. USA 87*, 2872–2876.

Tanizawa, Y., Koranyi, L.I., Welling, C.M., & Permutt, M.A. (1991). Human liver glucokinase gene: Cloning and sequence determination of two alternatively spliced cDNAs. *Proc. Natl. Acad. Sci. USA 88*, 7294–7297.

Thelen, A.P. & Wilson, J.E. (1991). Complete amino acid sequence of the type II isozyme of rat hexokinase, deducted from the cloned cDNA: Comparison with a hexokinase from Novikoff ascites tumor. *Arch. Biochem. Biophys. 286*, 645–651.

Thierry, A., Faihead, C., & Dujon, B. (1992). The complete sequence of the 8.2 kb segment left of MAT on chromosome III reveals five ORFs, including a gene for a yeast ribokinase. *Yeast 6*, 521–534.

Tsai, Y.H. & Robinson, G.W. (1991). Cloning and characterization of ERG8, an essential gene of *Saccharomyces cerevisiae* that encodes phosphomevalonate kinase. *Mol. Cell. Biol. 11*, 620–631.

Van Rooijen, R.J., Van Schalkwijk, S., & De Vos, W.M. (1991). Molecular cloning, characterization, and nucleotide sequence of the tagatose 6-phosphate pathway gene cluster of the lactose operon of *Lactococcus lactis*. *J. Biol. Chem. 266*, 7176–7181.

Vionnet, N., et al. (1992). Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. *Nature 356*, 721–722.

White, T.K. & Wilson, J.E. (1989). Isolation and characterization of the discrete N- and C-terminal halves of rat brain hexokinase: Retention of full catalytic activity in the isolated C-terminal half. *Arch. Biochem. Biophys. 174*, 375–393.

Wright, C.S., Alden, R.A., & Kraut, J. (1969). Structure of subtilisin BPN′ at 2.5 Å resolution. *Nature 221*, 235–242.

Wu, L.-F., Reizer, A., Reizer, J., Cai, B., Tomich, J.M., & Saier, M.H., Jr. (1991). Nucleotide sequence of the *Rhodobacter capsulatus* fruK gene, which encodes fructose-1-phosphate kinase: Evidence for a kinase superfamily including both phosphofructokinases of *E. coli*. *J. Bacteriol. 173*, 3117–3127.

Zembrzuski, B., Chilco, P., Liu, X.-L., Liu, J., Conway, T., & Scopes, R. (1992). Cloning, sequencing, and expression of the *Zymomonas mobilis* fructokinase gene and structural comparison of the enzyme with other hexose kinases. *J. Bacteriol. 174*, 3455–3460.