Structure, Vol. 12, 1809–1821, October, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.str.2004.07.020

Crystal Structure of an Aminoimidazole Riboside Kinase from *Salmonella enterica*: Implications for the Evolution of the Ribokinase Superfamily

Yan Zhang,[1] Michael Dougherty,[3] Diana M. Downs,[3] and Steven E. Ealick[2,*]

[1]Department of Molecular Biology and Genetics
[2]Department of Chemistry and Chemical Biology
Cornell University
Ithaca, New York 14853
[3]Department of Bacteriology
University of Wisconsin-Madison
Madison, Wisconsin 53706

Summary

The crystal structures of a *Salmonella enterica* aminoimidazole riboside (AIRs) kinase, its complex with the substrate AIRs, and its complex with AIRs and an ATP analog were determined at 2.6 Å, 2.9 Å, and 2.7 Å, respectively. The product of the *Salmonella*-specific gene *stm4066*, AIRs kinase, is a homodimer with one active site per monomer. The core structure, consisting of an eight-stranded β sheet flanked by eight α helices, indicates that AIRs kinase is a member of the ribokinase superfamily. Unlike ribokinase and adenosine kinase in this superfamily, AIRs kinase does not show significant conformational changes upon substrate binding. The active site is covered by a lid formed by residues 16–28 and 86–100. A comparison of the structure of AIRs kinase with other ribokinase superfamily members suggests that the active site lid and conformational changes that occur upon substrate binding may be advanced features in the evolution of the ribokinase superfamily.

Introduction

In bacteria, 5-aminoimidazole ribotide (AIR) serves as a branch point metabolite in the biosynthetic pathways of purines and thiamin (Newell and Tucker, 1968). AIR is then converted to carboxyaminoimidazole ribotide, the next intermediate in purine biosynthesis, or 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMPP), the pyrimidine moiety of thiamin. Biosynthesis of AIR requires five enzymatic steps and cellular levels of AIR control the flux through the purine-thiamin branch point (Allen et al., 2002). Although AIR cannot be efficiently taken up by most bacteria, it can be salvaged by the phosphorylation of aminoimidazole riboside (AIRs), which is easily taken up by the cells (Newell and Tucker, 1968). Genetic studies in *Salmonella enterica* showed that overexpression of gene *stm4066* facilitated AIRs salvage. Subsequent studies demonstrated the *stm4066* gene product had aminoimidazole riboside kinase activity (Dougherty and Downs, 2003).

The product of the *stm4066* gene (AIRs kinase) contains 319 amino acid residues and in vitro studies show this enzyme is capable of phosphorylating AIRs with a k_{cat} of 0.08 s^{-1}. Neither the purine nucleosides adenosine, inosine, or guanosine, nor the pyrimidine nucleosides cytosine, thymidine, or uridine serve as substrates for AIRs kinase (Dougherty and Downs, 2003). The AIRs analog 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranoside (AICARs), a precursor of a purine pathway intermediate 5-aminoimidazole-4-carboxamide ribotide also is not phosphorylated by AIRs kinase (Dougherty and Downs, 2003).

Amino acid sequence alignments using BLAST (Altschul et al., 1990) indicate that AIRs kinase is homologous to the enzymes in the ribokinase family (Bork et al., 1993; Wu et al., 1991). This family includes a number of prokaryotic enzymes: ribokinase, fructokinases, a minor 6-phosphofructokinase, 1-phosphofructokinase, and 6-phosphofructokinase (Blatch et al., 1990; Daldal, 1984; Hope et al., 1986; Orchard and Kornberg, 1990; van Rooijen et al., 1991). Eukaryotic members of the family were later identified, and include yeast ribokinase (Thierry et al., 1990) and human adenosine kinase (Spychala et al., 1996). The first crystal structures solved in this family were *E. coli* ribokinase (RK) (Sigrell et al., 1998) and human adenosine kinase (AK) (Mathews et al., 1998). Based on the structural similarity, some enzymes lacking significant primary sequence homology were also identified as members of the ribokinse family and include 4-methyl-5-β-hydroxy-ethylthiazole (THZ) kinase (Campobasso et al., 2000), ADP-dependent glucokinase (GK) (Ito et al., 2001), and pyridoxal kinase (PLK) (Li et al., 2002). By generalization of the original definition of the ribokinase family, this class of enzymes can be defined as the ribokinase superfamily (Andreeva et al., 2004; Murzin et al., 1995). All of these enzymes are phosphotransferases that have an alcohol group as an acceptor (E.C.2.7.1.−). However, 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase (HMPP kinase) (Cheng et al., 2002) catalyzes two phosphorylation reactions: one to a hydroxymethyl group of hydroxymethyl pyrimidine (HMP) and the second to the phosphomethyl group of HMPP.

The common structural feature for the enzymes in this superfamily is a central eight-stranded β sheet that is flanked by eight structurally conserved α helices, five on one side and three on the other. The active site is located in a shallow groove along one edge of the β sheet, with the phosphate acceptor hydroxyl group and γ-phosphate of ATP close together in the middle of the groove, and substrate and ATP binding at the ends. Some of the family members, such as RK (Sigrell et al., 1998, 1999), AK (Mathews et al., 1998; Schumacher et al., 2000), and GK (Ito et al., 2001), have a lid structure that covers the active site upon substrate binding. However, others, such as THZ kinase (Campobasso et al., 2000) and HMPP kinase (Cheng et al., 2002), lack this structural feature. This active site lid structure has been suggested to be a morphological marker for evolution within the superfamily (Cheng et al., 2002).

Though AIRs is a substrate for AIRs kinase, it is not considered to be the physiologically relevant substrate.

*Correspondence: see3@cornell.edu

The lack of significant levels of AIRs in the natural environment of *S. enterica*, together with the instability of AIRs suggest other substrates exist (Dougherty and Downs, 2003). The operon containing *stm4066* in *S. enterica* is not present in the genome of the closely related species *E. coli* (Dougherty and Downs, 2003). This suggests that a sugar(s) specific to the natural environment of *S. enterica* could be the physiological substrate for AIRs kinase. The structure of AIRs kinase and the knowledge of the chemical environment of the substrate binding site discussed herein will facilitate the identification of candidate substrates for in vivo and in vitro analysis.

Here we report the structure of AIRs kinase from *S. enterica* and compare its structure to other members of the ribokinase superfamily. This comparison supports the hypothesis that the active site lid is a morphological marker for evolution within this superfamily. It also suggests that the conformational changes that occur upon substrate binding may be advanced features in the evolution of the superfamily. The identification of key residues at the active site suggests chemical properties that must be present in candidate substrates.

Results

Description of the Final Models

Three crystal structures of *Salmonella enterica* AIRs kinase were determined: unliganded AIRs kinase, AIRs kinase in complex with AIRs (AIRs kinase/AIRs), and AIRs kinase in complex with AIRs and adenosine-5′-(β, γ-methylene)triphosphate (AIRs kinase/AIRs/AMP-PCP). All of the crystals belong to the orthorhombic space group $P2_12_12_1$, and the asymmetric unit contains two monomers (designated A and B), which are related by 2-fold noncrystallographic symmetry. The structure of free AIRs kinase was determined to 2.6 Å resolution by the single wavelength anomalous diffraction method using selenomethionine (SeMet)-labeled protein. The AIRs kinase homodimer model contains 593 residues (monomer A: 5–90, 98–308, monomer B: 5–90, 99–291, 294–310), four potassium ions, and 55 water molecules. The crystallographic R factor is 22.2% and R_{free} is 29.3%. The model of the complex AIRs kinase/AIRs, refined at 2.9 Å resolution, includes 597 residues (monomer A: 5–90, 98–310, monomer B: 5–90, 99–310), one AIRs molecule, four potassium ions, and 42 water molecules. The crystallographic R factor is 20.5% and R_{free} is 27.4%. The model of the complex AIRs kinase/AIRs/AMP-PCP, refined at 2.7 Å, includes 594 residues (monomer A: 5–90, 98–308, monomer B: 5–90, 99–309), one AIRs molecule, two AMP-PCP molecules, two magnesium ions, four potassium ions, and 75 water molecules. The crystallographic R factor is 21.6% and R_{free} is 28.5%.

Overall Structure

AIRs kinase is a homodimer that shows some asymmetry in all three structures. Monomer A consists of eight α helices, seven 3_{10} helices, and thirteen β strands (Figure 1B). In monomer B, helix 11, which is a 3_{10} helix in monomer A, is a short α helix. The overall AIRs kinase fold is an αβα three-layer sandwich (Figures 1A and 1C). The central β sheet contains nine strands with a topology of β5↑β4↑β1↑β8↑β9↑β10↑β11↑β12↓β13↑. All of the strands are parallel except for strand 12, which is antiparallel to its neighboring strands. Five α helices (α1, α2, α13, α14, and α15) flank one side of the β sheet, and the remaining three α helices (α6, α8, and α9) and all seven 3_{10} helices lie on the opposite side. Among these eight α helices, all except α14 are approximately antiparallel to the strands of the central β sheet. A four-stranded antiparallel β sheet consisting of β2, β3, β6, and β7 forms a lid that folds back over the substrate binding site. The homodimer forms through lid-to-lid interactions.

A comparison of monomers A and B shows a root-mean-square deviation (rmsd) of 0.7 Å for 274 of 299 pairs of ordered C_α atoms. The major differences between the two monomers occur in the lid regions. Residues 14–27 (β2, β3, and the loop connecting them) show an rmsd of 5.3 Å and residues 86–90 and 100–106 (β6, β7 and the following loop) show an rmsd of 3.1 Å. Most of the loop from 91–97 is disordered in both monomers. Although the lid regions are not directly involved in crystal contacts, it is possible that the differences between the two monomers are the indirect result of crystal packing.

The overall structures of unliganded AIRs kinase and liganded AIRs kinase are similar. No significant conformational changes are observed in the complexes with respect to the unliganded form; however, each complex displays the same asymmetry as the unliganded AIRs kinase. The rmsd for ordered C_α atoms between free AIRs kinase and the AIRs kinase/AIRs complex, and free AIRs kinase and the AIRs kinase/AIRs/AMP-PCP complex are 0.4 Å and 0.3 Å, respectively.

Dimer Interface

AIRs kinase is an elongated dimer with approximate dimensions 52 Å × 46 Å × 95 Å. The subunit interface in the dimer is formed by lids (β3β2β6β7) from each monomer. Strands β3A and β3B are hydrogen bonded such that the two four-stranded β sheets form a highly twisted eight-stranded β sheet. The severe twisting causes the eight-stranded β sheet to fold back on itself, resulting in van der Waals interactions across the noncrystallographic 2-fold axis between β6, β7, residues 57–58 and residues 101–104. The total surface area buried at the interface is about 1514 Å, which is at the low end of the range observed for a dimer (Janin, 1997). Nevertheless, gel filtration chromatography and nondenaturing polyacrylamide gel electrophoresis results also suggest that AIRs kinase is a dimer in solution (data not shown).

AIRs Binding Site

Each AIRs kinase monomer has one active site, located along one edge of the central β sheet, with the AIRs binding site near the lid and the ATP binding site at the other end. In the electron density map, AIRs was observed only in monomer A, whereas AMP-PCP was observed in both monomers (Figure 2). The AIRs molecule is almost completely buried in a pocket formed by the loop following β1 (12–14), the loop preceding α1 (30–32), α7 and the preceding loop, and the N terminus

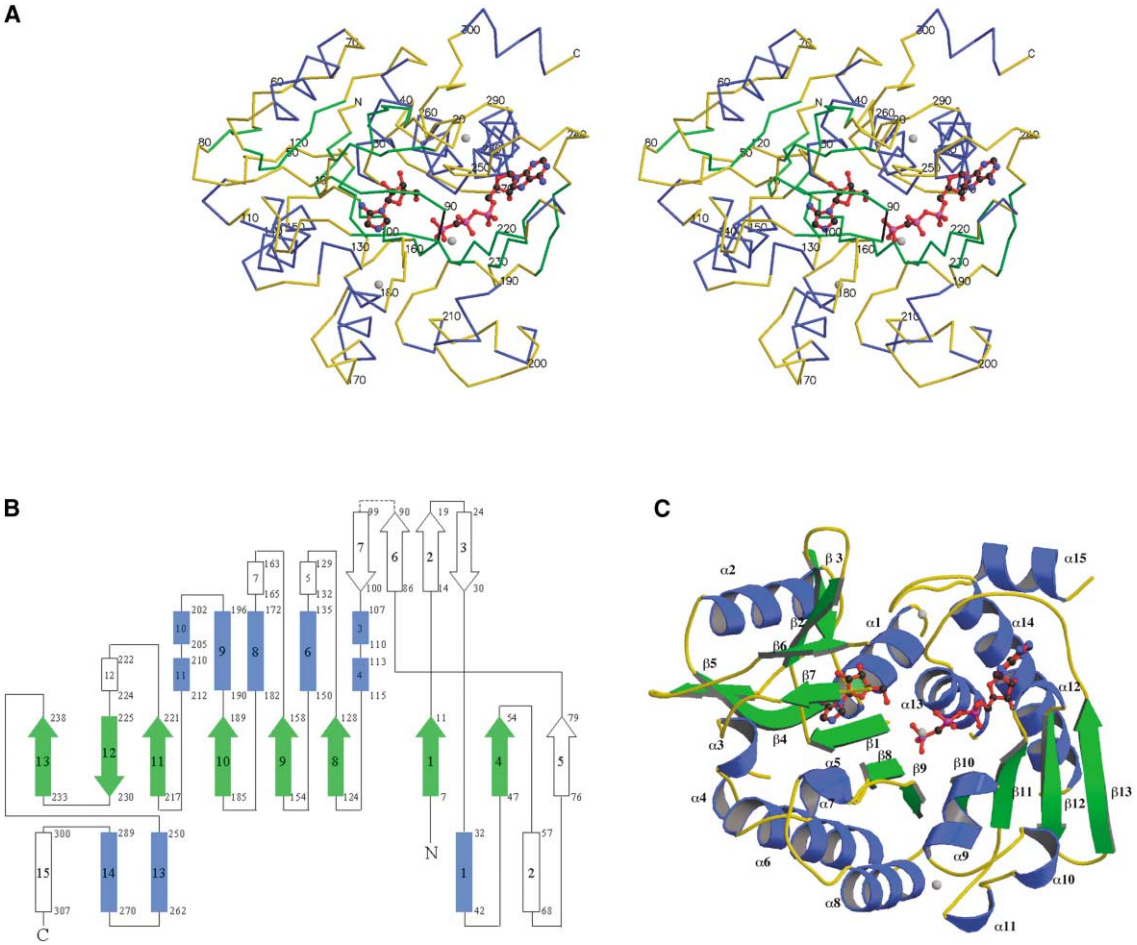

Figure 1. Structure of AIRs Kinase

(A) Stereoview of a C_α trace colored by secondary structural element (blue for α helices, green for β strands, and yellow for loops) and with every tenth residue labeled with a sequence number. The substrate AIRs, ATP analog AMP-PCP, and ions are shown in ball-and-stick representation. A break in the backbone is connected with a dashed line.

(B) Topology diagram. The α helices are shown as rectangles and the β strands are shown as arrows. Helices 3, 4, 5, 7, 10, 11, and 12 are 3_{10} helices. The dotted line depicts the disordered loop. Each secondary structural element is labeled in its center with its designator and with its beginning and ending sequence number. The elements common to all members of the ribokinase family are in blue for α helices and in green for β strands.

(C) A ribbon diagram representing the overall fold (α helices, blue; β strands, green). The substrate AIRs, the ATP analog AMP-PCP, and ions are shown in ball-and-stick representation.

(A) and (C) were prepared using MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

of α13. Strands β2, β3, β6, and β7, which form the lid, cover the AIRs molecule. The AIRs molecule assumes a standard *anti* conformation with a glycosidic torsion angle of −137°. An uncommon ribose sugar pucker of O4′-*endo* provides the best fit to the unbiased difference electron density. One face of the ribose ring stacks against the aromatic ring of Phe99 and the other against a highly conserved Gly-Gly dipeptide. The ribose ring forms hydrogen bonds between the O2′, O3′ and O5′ oxygen atoms with the side chains of residues Asp16, Asn35, Asp252, and Arg162, and the main chain N of residue Gly31 (Figure 2B). Additional interactions between AIRs and the active site come from the imidazole ring. The N3 atom of AIRs forms a hydrogen bond to the hydroxyl group of Tyr101 and N6 donates a hydrogen bond to O_δ of Asp12. Both of these residues are in the lid region.

ATP Analog Binding Site

The ATP analog AMP-PCP is positioned in a shallow groove, mostly exposed to the solvent (Figure 2). The adenine ring is stacked between the main chain atoms in 3_{10} helix 12 connecting strands β11 and β12 and the side chain of Val288 in helix α14. The loop connecting strand β13 to helix α13 helps to further position the base ring. There are no hydrogen bonds between the adenine ring and the enzyme. The sugar ring of AMP-PCP is in the common C3′-*endo* conformation and the glycosidic torsion angle is in an *anti* conformation with a value of −179°. The O2′ atom of the ribose moiety forms a bifurcated hydrogen bond with the carbonyl oxygen and the O_δ atoms of Asn281, while the O3′ atom forms a bifurcated hydrogen bond with the O_δ atoms of Asn281 and the carbonyl oxygen atom of Gly225 (Figure 2B).

The α-phosphate of AMP-PCP is about 6 Å from the

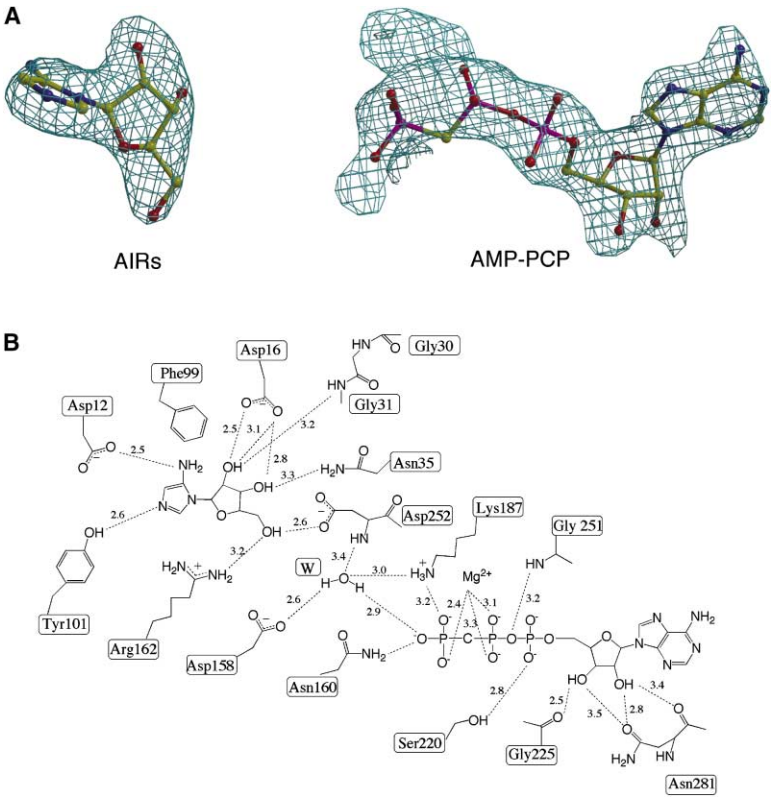

Figure 2. The Active Site of AIRs Kinase

(A) Electron density of AIRs and AMP-PCP. The $F_o - F_c$ omit map was calculated to 2.7 Å after one cycle of refinement without AIRs and AMP-PCP from the model (bluegreen mesh, contoured at 3σ). AIRs and AMP-PCP are drawn in ball-and-stick representations. Prepared using BOBSCRIPT (Esnouf, 1997).

(B) A schematic drawing of the active site. Key hydrogen bonds are indicated by dashed lines with the corresponding donor-acceptor distance labeled.

C terminus of β11. One of the α-phosphate oxygen atoms is hydrogen bonded to the hydroxyl group of Ser220. The oxygen atom connecting the α-phosphate and β-phosphate is hydrogen bonded to the amide group of Gly251. Two of the oxygen atoms of the γ-phosphate, N_ζ of Lys187, N_δ of Asn160, the main chain N of Asp252, O_δ of Asp158, and a water molecule form a hydrogen bond network. One Mg^{2+} ion is in the ATP analog binding site, located between the β- and γ-phosphates and interacts with the oxygen atoms of both.

Potassium Ion Binding Sites

Two tightly bound potassium ions were built into each monomer based on the presence of 0.15 M K_2SO_4 in the crystallization solution, the coordination geometry for each site, and the refined *B* factors. Ribokinase is also suggested to bind monovalent cations, such as K^+ (Andersson and Mowbray, 2002). The first K^+ binding site is adjacent to the active site and involves the two loops connecting β13 and α13, and α14 and α15. The coordination geometry is best described as octahedral with the carbonyl oxygen atoms of Asp246, Thr248, Ala290, Ala287, Gly292, and the O_δ of Asp246 as vertices. The second K^+ is located near the surface of the protein. It interacts primarily with carbonyl oxygen atoms in the main chain of loop 180–184 and loop 212–214. Two water molecules are also within coordination distance. The coordination geometry of the second K^+ binding site is also octahedral. The average observed distances between potassium ions and oxygen atoms are 2.8 Å for the first ion and 2.9 Å for the second ion.

Discussion

Comparison with Other Protein Structures

A structural comparison using DALI (Holm and Sander, 1993) revealed a similarity with Z scores higher than 12 between AIRs kinase and five members of the ribokinase superfamily. The ribokinase family was first identified as a family of carbohydrate kinases (Bork et al., 1993; Wu et al., 1991), and subsequently some small molecule kinases that phosphorylate a hydroxymethyl (Cheng et al., 2002; Li et al., 2002) or hydroxyethyl (Campobasso et al., 2000) group of noncarbohydrates were reported to share the same fold as the members of the ribokinase family (Andreeva et al., 2004; Miller et al., 1994). Eight structures of proteins with known function in this superfamily have been reported to date: RK, human AK, *T. gondii* AK, THZ kinase, GK, HMPP kinase, PLK, and KDG kinase (Table 1). Two structures (1O14 and 1KYH) deposited in the Protein Data Bank have the core structure of the ribokinase superfamily but have uncharacterized functions. The members of the ribokinase superfamily share a common αβα core structure (Figure 3). The central β sheet of the core has eight strands with an order of β2β1β3β4β5β6β7β8, in which all but strand β7 are parallel. Eight structurally conserved α helices flank the central sheet, five on one side and three on the other. Each of the known structures has a few additional secondary structural elements flanking the conserved core.

In the case of AIRs kinase, the central β sheet is composed of strands β4, β1, β8, β9, β10, β11, β12, and β13. There are two extra β strands inserted between

Table 1. Members of the Ribokinase Superfamily Listed in the Order of Proposed Evolution Pathway

Enzyme	Abbreviation	PDB ID	Residues/ Monomer	Monomers/ Molecule[c]	Conformational Change[d]
4-methyl-5-β-hydroxy-ethylthiazole kinase	THZ kinase	1EKK	272	3	none
Unknown function enzyme	—	1KYH	276	4	unknown
4-amino-5-hydroxymethy-2-methylpyrimidine phosphate kinase	HMPP kinase	1JXH	288	2	local
ribokinase	RK	1RKD	309	2	global
pyridoxal kinase	PLK	1LHP	312	2	local
Unknown function enzyme	—	1O14	331	2	unknown
adenosine kinase[a]	hAK	1BX4	345	1	global
2-keto-3-deoxygluconate kinase	KDG kinase	1J5V	351	2	unknown
adenosine kinase[b]	TgAK	1DGY	363	1	global
ADP-dependent glucokinase	GK	1GC5	467	1	global

[a] Human adenosine kinase.
[b] *T. gondii* adenosine kinase.
[c] The number of monomers per biological molecule or in the crystal structure if the biological function is not known.
[d] Conformational changes upon substrate binding: global, domain movement when substrates are bound; local, loops movement when substrates are bound; none, no significant conformational changes observed; and unknown, data not available.

strands β1 and β4 and three extra β strands between β4 and β8. Helices α1, α13, and α14, corresponding to the structurally conserved three helices, flank on one side of the central β sheet. Helices α3, α4, α6, α8, α9, α10, and α11 match the five conserved helices on the other side, of which two pairs of 3_{10} α helices (α3 and α4, and α10 and α11) are equivalent to the two α helices in the other members of the ribokinase superfamily (Figure 1B).

A total of 107 C_α positions of AIRs kinase, corresponding to the structurally conserved core of the ribokinase superfamily, were manually selected for superposition with structurally equivalent residues in the other ten members using the computer program ProFit (Martin, 1998) (Figure 3). The comparison showed that AIRs kinase is most similar to substrate-bound RK, human AK, and *T. gondii* AK with an rmsd of 1.5 Å, whereas the rmsd for AIRs kinase with other members is 1.9–2.9 Å. Among the ribokinase superfamily, human AK and RK (excluding the two functionally equivalent AKs) are the most similar pair, with an rmsd of 1.0 Å, whereas the most dissimilar pairs are the proteins of unknown function, 1O14 and 1KYH, with an rmsd of 3.1 Å.

Quaternary Structures of the Ribokinase Superfamily

Although the monomer fold of the ribokinase superfamily is well conserved, there are differences in the quater-

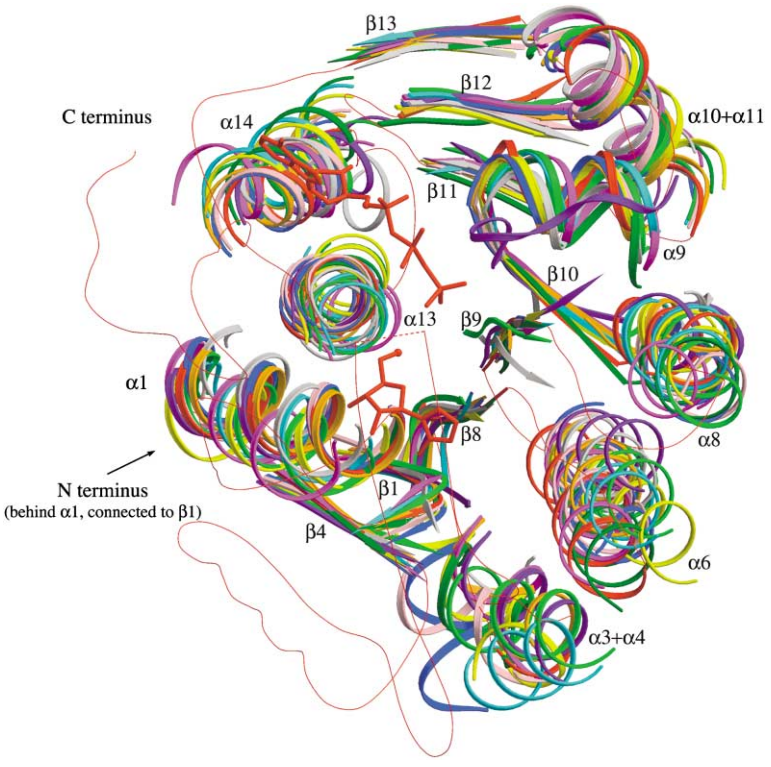

Figure 3. Superposition of the Members of the Ribokinase Superfamily Based on 107 Structurally Conserved C_α Positions

Conserved secondary structural elements from all superfamily members are shown using thick lines and labeled using AIRs kinase notation. Nonconserved secondary structual elements for AIRs kinase are shown by a thin red line, with backbone break connected by a dashed line. Colors: AIRs kinase, red; RK, orange; human AK, blue; *T. gondii* AK, pink; GK, yellow; HMPP kinase, green; THZ kinase, magenta; pyridoxal kinase, cyan; KDG kinase, olive; 1O14, purple; and 1KYH gray.

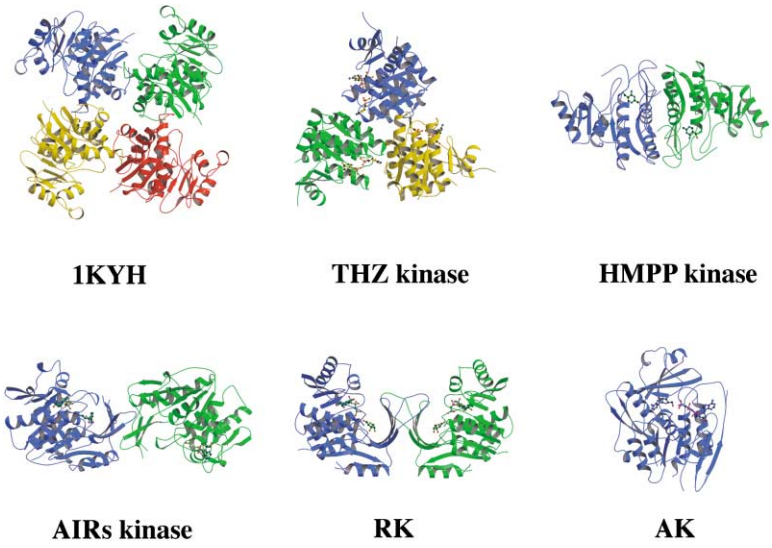

Figure 4. Quaternary Structures of the Ribokinase Superfamily Members

Ribbon diagrams representing six classes of quaternary structures of the ribokinase superfamily members. The dimers for HMPP kinase and PLK are similar, the dimers for KDG kinase and AIRs kinase are similar and the dimers for RK and 1O14 are similar. GDK is a monomer similar to AK. The substrates and ATP or ATP analog are shown in ball-and-stick representation if available. The figure was prepared using MOLSCRIPT (Kraulis, 1991).

nary structures for the active forms of the protein in this superfamily (Figure 4): 1KYH is tetrameric; THZ kinase is trimeric (Campobasso et al., 2000); RK, AIRs kinase, KDG kinase, 1O14, HMPP kinase, and PLK are dimeric (Cheng et al., 2002; Kwok et al., 1987; Sigrell et al., 1997); and human and *T. gondii* AKs, and GK are monomeric (Darling et al., 1999; Dougherty and Downs, 2003; Koga et al., 2000; Mathews et al., 1998). 1KYH forms a tetramer with 4-fold symmetry through the interactions along the edges of the central β sheet. The four predicted active sites are at the interfaces of adjacent momomers. The THZ kinase trimer contains three identical active sites (Campobasso et al., 2000) located at the interface between two subunits within a trimer and covered by the adjacent subunit. The dimerization patterns for HMPP kinase (Cheng et al., 2002) and PLK (Li et al., 2002) are similar, although they are distinctly different from those of RK (Sigrell et al., 1998) and AIRs kinase. Both HMPP kinase and PLK have one independent active site per monomer and the dimer interface involves a large area at one side of the central β sheet with extensive hydrogen bonding and van der Waals interactions. It can be described as side-to-side dimerization.

The active site flaps in both PLK and HMPP kinase are not located at the dimer interface. In contrast, the RK dimer interface involves an interlocking β clasp at the one edge of the central β sheet, which can be described as edge-to-edge dimerization. The β clasp forms between two four-stranded β sheets from each subunit and covers the active site as a lid, which closes upon substrate binding (Sigrell et al., 1999). Similarly to RK, the active-site lids of AIRs kinase also contribute hydrogen bonding and van der Waals interactions to the interface of two monomers in the crystal structure. When one monomer of AIRs kinase is superimposed on one monomer of RK, the 2-fold axis for the AIRs kinase dimer is almost perpendicular to that of the RK dimer. Therefore, the four-stranded β sheet from AIRs kinase cannot form a β clasp. Instead, strands β3 from two adjacent monomers lie side by side to form a highly twisted antiparallel β sheet. Interestingly, KDG kinase forms a dimer in the same way as AIRs kinase, while 1O14 forms a dimer similar to that of RK. Both human and *T. gondii* adenosine kinase are monomeric, with a small domain consisting of a five-stranded β sheet flanked by two α helices, which acts as a lid and covers the active site upon substrate binding (Mathews et al., 1998; Schumacher et al., 2000). In the case of ADP-dependent glucokinase, the active site lid is composed of a five-stranded β sheet and four α helices (Ito et al., 2001).

Substrate Binding Sites

For all the members in the ribokinase superfamily, the substrate binding site is surrounded by the N-terminal portions of α1 and α13 and the C-terminal portions of β1 and β8 using AIRs kinase notation (Figure 3), which are well conserved compared to helices α4, α6, and α8. Superposition of the six available structures with the bound substrates shows that the general geometry of the binding site is similar (Figure 5A). The sugar rings of AIRs for AIRs kinase, ribose for RK, and adenosine for human and *T. gondii* AK superimpose closely, and the hydroxymethyl groups of the carbohydrates are oriented in the same way. The five-membered ring of THZ in the case of THZ kinase and the six-membered ring of HMP in the case of HMPP kinase are bound differently from the carbohydrates. The ring of THZ is aligned with the average planes of the sugars, however the ring of HMP is nearly perpendicular to the sugar rings. The hydroxyethyl group in THZ and the hydroxymethyl group in HMP, nevertheless, are positioned in almost identical locations as the other phosphate acceptors.

Only one of the active sites was occupied by AIRs in the crystal structures of both the AIRs kinase/AIRs complex and the AIRs kinase/AIRs/AMP-PCP complex. Comparison of the two monomers in the asymmetric unit indicates a significant difference (rmsd 3.1–5.1 Å) in the lid region that covers the substrate binding site. The lid is more open for monomer B, which does not have the substrate bound, than for monomer A. The aromatic ring of Phe99 in β7 from monomer A helps to close the substrate binding site through hydrophobic interactions with the ribose ring, but it is disordered in

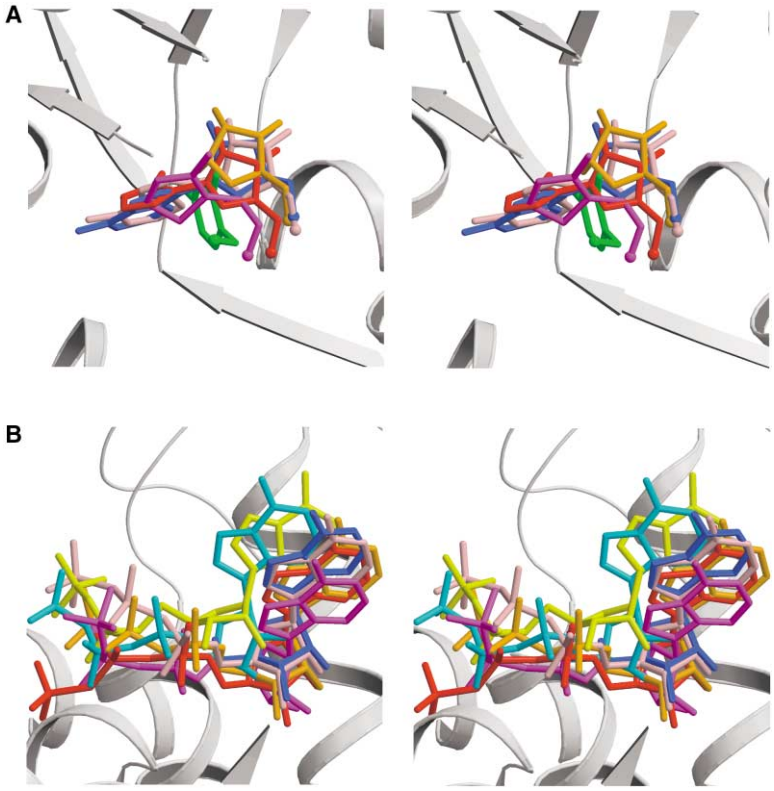

Figure 5. Stereoview of the Active Sites of Ribokinase Superfamily Members

AIRs kinase backbone, grayscale; ligands from AIRs kinase, red; RK, orange; human AK, blue; *T. gondii* AK, pink; GK, yellow; HMPP kinase, green; THZ kinase, magenta; and pyridoxal kinase, cyan.

(A) Superposition of substrates. The hydroxyl group that accepts phosphate is indicated as a large sphere, while the rest of the substrate is represented by bonds only. Substrates are shown for AIRs kinase (AIRs), RK (ribose), human AK (adenosine), *T. gondii* AK (adenosine), HMPP kinase (HMP), and THZ kinase (Thz).

(B) Superposition of ligands at the ATP or ADP binding sites. Ligands are ATP from THZ kinase and pyridoxal kinase, AMP-PCP from AIRs kinase, ribokinase, and *T. gondii* AK, ADP from glucokinase, and adenosine from human AK.

monomer B. The difference in substrate binding may be the result of differences in crystal packing. However, the possibility of the presence of an asymmetric dimer in solution cannot be ruled out.

Prediction of Substrate Specificities for Ribokinase Superfamily Members

AIRs is deeply buried in the active site with the ribose ring anchored by several hydrogen bonds. The O2′ and O3′ ribose hydroxyl groups hydrogen bond with the side chain of Asp16 (Figure 2B). The O2′ ribose hydroxyl group also hydrogen bonds with the backbone amide nitrogen of Gly31 and the O3′ hydroxyl makes a hydrogen bond with the N_δ of Asn35. Asp16 and Asn35 are well conserved among sugar kinases, including RK, both human and *T. gondii* AK, GK, KDG kinase, and protein 1O14; however, they are not conserved in HMPP kinase, THZ kinase, and PLK (Figure 6). Besides the conserved residues, the side chain of the nonconserved residue Arg162 forms hydrogen bonds with O5′ in AIRs. In RK, the side chain of Lys 43 contributes to similar hydrogen bonds but comes from the other side of the ribose ring and in KDG kinase Arg168 exits in the corresponding position.

One of the fingerprint regions of the ribokinase family is the glycine-glycine dipeptide (Gly30Gly31 in AIRs kinase) sequence (Bork et al., 1993; Wu et al., 1991). The GG dipeptide was found to undergo a conformational switch upon substrate binding, and is believed to play an important role in forming the closed conformation of the enzymes and subsequent substrate sequestration (Schumacher et al., 2000). All eleven members of the ribokinase superfamily were examined, and GG is present in all except THZ kinase, PLK, and protein 1KYH (Figure 6). In addition to lacking the GG motif, THZ kinase and PLK do not have a lid or flap structure that closes upon substrate binding.

	16	30	35	246	254
AIRsK	D	GG	N	DTTGAGD	AF
RK	D	GG	N	DTIAAGD	TF
H_AK	D	GG	N	DTNGAGD	AF
Tg_AK	D	GG	N	DTNGAGD	AF
GK	D	GG	I	STVGIGD	TI
HMPPK	-	GG	A	NTHGTGC	TL
THZK	-	VV	F	KVTGAGC	LL
PLK	-	VG	A	VFVGTGD	LF
KDGK	-	GG	N	DRVGAGD	SF
1O14	D	GG	N	HLLGAGD	AY
1KYH	-	GA	L	AKGGTGD	TL

Figure 6. Conserved Residues among the Members of the Ribokinase Family

The conserved residues involved in hydrogen bonding with the 2′ and 3′ hydroxyl groups in the ribose ring are shown in magenta; the conserved GG residues in green; the conserved aspartate or cysteine serving as general base for nucleophilic attack in red; and the residues that form the anion hole in the blue frame. The other residues are in black. The absence of a conserved residue at the corresponding position is represented by a dash. The residues are numbered according to AIRs kinase notation. Abbreviations: AIRsK, AIRs kinase; RK, ribokinase; H_AK, human adenosine kinase; Tg_AK, *T. gondii* adenosine kinase; GK, glucokinase; HMPPK, HMPP kinase; THZK, THZ kinase; PLK, pyridoxal kinase; KDGK, 2-keto-3-deoxylgluconate kinase; 1O14, protein with PDB ID 1O14; and 1KYH, protein with PDB ID 1KYH.

The phosphate acceptor 5′-hydroxyl group of AIRs extends away from the sugar ring, making a hydrogen bond with the side chain of Asp252. This aspartic acid residue near the phosphate acceptor is one of the most highly conserved residues among the ribokinase superfamily of proteins (Figure 6), and is believed to act as a general base to abstract the proton from the 5′-hydroxyl group (Mathews et al., 1998; Schumacher et al., 2000; Sigrell et al., 1998). The only exceptions are HMPP kinase and THZ kinase, which have cysteine as a catalytic base for the activation of the phosphate acceptor in a nucleophilic attack (Campobasso et al., 2000; Cheng et al., 2002). Interestingly, mutation of cysteine to aspartate at the active site in THZ kinase increases the enzymatic activity 9-fold.

Analysis of the hydrogen bonding scheme of the conserved residues to the ribose ring suggests that the unknown function of protein 1O14 may involve the phosphorylation of a ribosyl hydroxyl group, similar to AIRs kinase, RK, and AK, while the protein 1KYH may have a noncarbohydrate substrate similar to HMPP or THZ. The presence of a well-defined lid structure in the protein 1O14 and the absence of such a structure in the protein 1KYH are also consistent with that prediction. The structural comparison of these two proteins with other members of the ribokinase superfamily shows that protein 1O14 is more similar to human AK (rmsd 2.3 Å) and *T. gondii* AK (rmsd 2.4 Å) than to THZ kinase (rmsd 3.0 Å), while protein 1KYH is more similar to HMPP kinase (rmsd 2.3 Å) and THZ kinase (rmsd 2.4 Å) than to AIRs kinase or GK (rmsd 2.9 Å), which is also consistent with the prediction.

ATP Binding Site

The ATP binding site is clearly defined by the binding of the ATP analog, AMP-PCP. It is located in a shallow groove that stretches along the C-terminal edge of the central β sheet with the strands β9, β10, β11, β12, and β13 at one side, and α13 and α14 at the other side, a region structurally highly conserved in the ribokinase superfamily. The ATP binding sites from seven members of the ribokinase superfamily, for which the structures in the presence of adenosine, ADP, ATP, or ATP analog are available, are superimposed in Figure 5B. For AIRs kinase, RK, human AK, and *T. gondii* AK, the adenine base, ribose, α-phosphate, and β-phosphate are positioned similarly. However, the γ-phosphate of AIRs kinase points away from those of the other enzymes. The distance between the 5′-hydroxyl oxygen atom of AIRs and the γ-phosphate group of AMP-PCP is about 5.5 Å, which is comparable to RK (Andersson and Mowbray, 2002).

For PLK, the adenosine in ATP is pushed away by the side chain of Met223 inside the ATP binding site, while the phosphates are located similarly to those of RK and AK (Schumacher et al., 2000; Sigrell et al., 1998). The adenosine of ATP in PLK moves toward the central β sheet, creating space for a 3_{10} helix between strand β8 and helix α9, which corresponds to a long loop connecting strand β13 and helix α13 in AIRs kinase (Li et al., 2002). In the case of THZ kinase, the ATP is bound in the presence of product monophosphate, which displaces the γ-phosphate from the active site (Campobasso et al., 2000). In the case of the only ADP-dependent enzyme in this superfamily, GK, two large hydrophobic residues, Tyr357 and Ile453, which form the base of ribose binding pocket, force the ADP to move toward the direction of the adenine base and the β-phosphate (Ito et al., 2001). Consequently, the α- and β-phosphates of ADP overlap with the β- and γ-phosphates of ATP, respectively, for the other family members. Therefore, the main catalytic apparatus remains and it is the binding pattern of the adenosine moiety that changes in the ADP-dependent kinase.

The Anion Hole

Residues 249–252 (GAGD) in AIRs kinase form the predicted kinase anion hole (Schumacher et al., 2000; Sigrell et al., 1999). This region is the most highly conserved region among the entire ribokinase superfamily of proteins and is present in all the current members (Figure 6). The anion hole helps to neutralize accumulated negative charge during the phosphate group transfer. The last residue in the conserved anion hole sequence is the aspartate or cysteine that is postulated to deprotonate the alcohol during the catalysis. Adjacent to the anion hole in AIRs kinase, a potassium ion was found to be bound to the two loops connecting β13 and α13, and α14 and α15, namely, the carbonyl oxygen atoms of Asp246, Thr248, Ala290, Ala287, Gly292, and the O_δ of Asp246. RK was reported to be activated by potassium, and the structure of RK in the presence of cesium was solved (Andersson and Mowbray, 2002). The cesium ion was bound immediately adjacent to the anion hole and was proposed to induce the formation of the anion hole and promote allosteric activation of the enzyme. When the structures of AIRs kinase and RK are superimposed, the potassium binding sites overlap within experimental error. The presence of both an anion hole and a potassium ion in all three structures of AIRs kinase, AIRs kinase/AIRs complex, and AIRs kinase/AIRs/AMP-PCP complex suggests that potassium could play the same role in AIRs kinase.

Conformational Changes upon Substrate and ATP Binding

Many members in the ribokinase superfamily undergo global conformational changes when the substrates and ATP are bound (Schumacher et al., 2000; Sigrell et al., 1999). The binding of substrate to the enzyme induces the "closed" conformation of the enzyme followed by substrate sequestration. Subsequent formation of the anion hole, induced by the binding of ATP, completes the structural requirements for the catalysis (Schumacher et al., 2000; Sigrell et al., 1999). The structure of AIRs kinase/AIRs/AMP-PCP, however, did not show any significant conformational changes compared with that of unliganded AIRs kinase. Furthermore, the AIRs kinase structure is more similar to the "closed" form of RK than to the "open" form. In contrast to HMPP kinase, the loop connecting β6 and β7 in the active-site lid remains disordered even when the substrate was bound (Cheng et al., 2002). Interestingly, the lids of AIRs kinase, which are different in monomers A and B, do not change in the presence of substrate and ATP. The presence of a preformed anion hole in unliganded AIRs kinase and

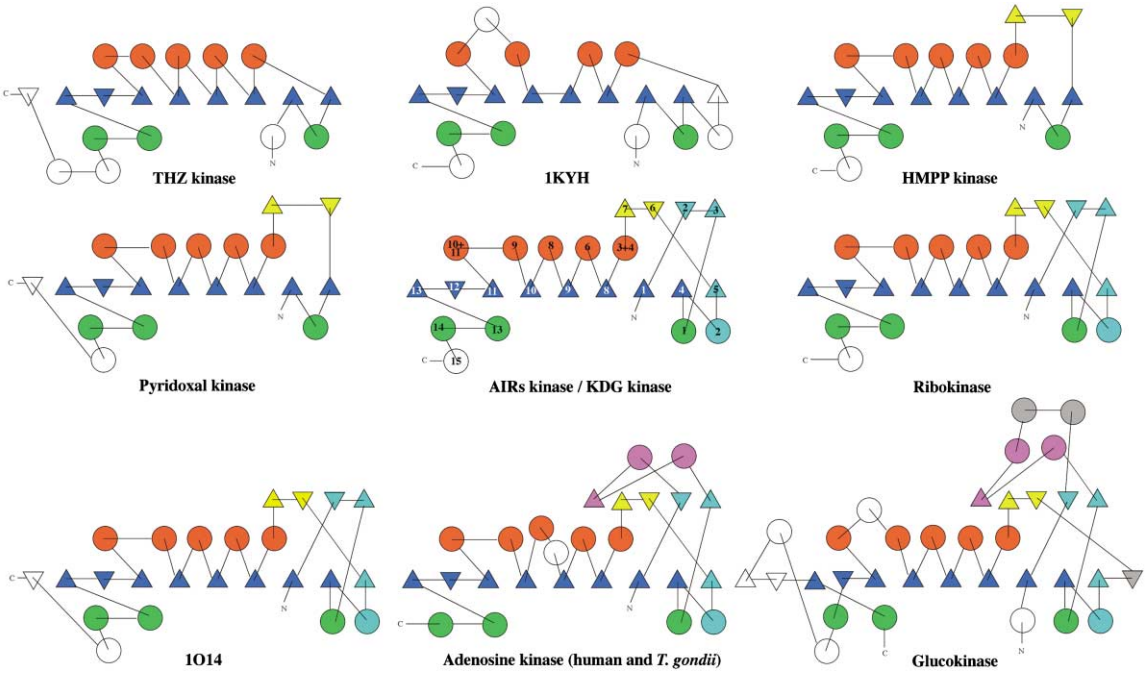

Figure 7. Topology Diagrams of Ribokinase Superfamily

α helices are shown as circles and β strands as triangles. Triangles pointing in the same direction represent parallel strands and triangles pointing in opposite directions represent antiparallel strands. The secondary structural elements conserved in all ribokinase superfamily members are shown in the same color: central β sheet, blue; α helices on each side: red and green. The secondary structural elements in yellow represent an insertion relative to THZ kinase. The secondary structural elements in cyan represent an insertion relative to HMPP kinase. The secondary structural elements in magenta represent an insertion relative to ribokinase. The secondary structural elements in gray represent an insertion relative to adenosine kinase. The unique secondary structural elements for each enzyme are shown in white. Each of the two pairs of 3_{10} helices in AIRs kinase (α3 and α4, α10 and α11) is shown as an α helix because they correspond to the topologically conserved α helices. Other 3_{10} helices and β turns are not shown. The secondary structural elements for AIRs kinase are labeled. The figures were prepared with TOPS (Westhead et al., 1999).

AIRs kinase/AIRs complex may be the result of potassium binding, which suggests that the local conformational changes resulting in the formation of an anion hole are independent of substrate binding.

Implications for Evolution of the Ribokinase Superfamily

Comparison of ten structures of proteins in the ribokinase superfamily suggests a common ancestor and an evolutionary pathway illustrated in Figure 7. Several factors serve as indicators of change along the evolutionary pathway, including quaternary structure, size of the monomer, active site lid elaboration, and conformational changes upon substrate binding. In general, as the family members progress from simple to more complex, the active form of the protein tends to become monomeric, the number of amino acids in the monomer increases, the active site lid becomes structurally more complex,

Table 2. Summary of Data Collection and Processing Statistics

	SeMet Data1	SeMet Data2	AIRs Kinase/AIRs	AIRs kinase/ AIRs/AMP-PCP
Wavelength (Å)	0.9791	0.9792	0.9796	0.9760
Resolution (Å)	3.1	2.6	2.9	2.7
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a (Å)	136.27	137.43	135.85	137.01
b (Å)	54.06	53.99	54.10	54.10
c (Å)	89.81	89.49	89.80	89.90
No. of reflections	95,186	96,181	50,206	88,221
No. of unique reflections	12,297	18,143	14,276	18,879
Redundancy[a]	7.7 (7.8)	5.3 (4.6)	3.5 (3.3)	4.7 (4.8)
Completeness (%)[a]	97.0 (95.3)	85.4 (69.3)	91.0 (75.0)	99.1 (99.8)
R_{sym}(%)[a,b]	7.6 (23.4)	6.9 (28.9)	4.9 (28.9)	5.9 (29.5)
I/σ[a]	21.5 (9.1)	20.5 (4.1)	23.1 (4.2)	23.6 (4.5)

[a] Values for the outer resolution shell are given in parentheses.

[b] $R_{sym} = \Sigma\Sigma_i|I_i - \langle I\rangle|/\Sigma\langle I\rangle$, where $\langle I\rangle$ is the mean intensity of N reflections with intensities I_i and common indices *h*, *k*, and *l*.

Table 3. Summary of Refinement Statistics

	AIRs Kinase	AIRs Kinase/AIRs	AIRs Kinase/ AIRs/AMP-PCP
Resolution (Å)	50–2.6	50–2.9	50–2.7
Total no. of non-hydrogen atoms	4432	4488	4629
No. of protein atoms	4373	4427	4471
No. of ligand atoms	4	19	83
No. of water oxygen atoms	55	42	75
No. of reflections in refinement	31,754	22,326	16,286
No. of reflections in test set	3,165	2,415	1,780
R factor[a] (%)	22.2	20.5	21.6
R_{free}[b] (%)	29.3	27.4	28.5
Rms deviation from ideal geometry			
Bonds (Å)	0.008	0.007	0.007
Angles (°)	1.4	1.4	1.5
Ramachandran plot			
Most favored region (%)	87.7	82.1	89.5
Additionally allowed region (%)	11.1	16.7	10.5
Generously allowed region (%)	1.2	1.2	0
Disallowed region (%)	0	0	0
Average B factors			
Main chain	68.9	54.3	49.2
Side chain	69.7	55.0	54.7
AIRs		75.3	61.8
AMP-PCP			67.7
Metal ions	76.5	59.9	53.0
Water	63.1	54.2	52.6

[a] R factor $= \Sigma_{hkl}||F_{obs}| - k|F_{cal}||/\Sigma_{hkl}|F_{obs}|$, where F_{obs} and F_{cal} are the observed and calculated structure factors, respectively.
[b] For R_{free}, the sum is extended over a subset of reflections (10%) excluded from all stages of refinement.

and the protein utilizes a substrate-induced conformational change.

The simplest fold, which contains all the necessary elements for catalytic activity and may represent the most ancient fold, is the THZ kinase monomer. THZ kinase is trimeric with one active site and one anion hole per subunit. It has a less efficient conserved cysteine as the base for deprotonating the hydroxyl group. Because THZ kinase lacks a lid or flap structure to cover the substrate binding site, the active site is shielded by extensive contacts with the adjacent monomer. THZ kinase does not undergo any conformational changes upon substrate or ATP binding (Campobasso et al., 2000).

Similar to THZ kinase, HMPP kinase has a conserved cysteine residue at the active site, and it also has a conserved aspartate close to the hydroxyl group of HMP; however, neither residue is close enough to the hydroxyl group to serve as the general base. HMPP kinase forms a dimer with a large contact surface. It still lacks a well-defined lid but utilizes a flap formed by two short β strands and a flexible loop. Upon substrate binding, no significant conformational changes occur, although the disordered loop in the flap becomes ordered (Cheng et al., 2002).

PLK forms a dimer in the crystal in exactly the same way as HMPP kinase does. However, both the dimer and the monomer are biologically active (Li et al., 2002). PLK utilizes a conserved aspartate residue as the base to deprotonate the phosphate acceptor. The two short β strands in the flap in HMPP kinase still exist in PLK, though they do not act as a lid. The lack of a well defined lid in PLK is also supported by the tolerance of enzymatic activity for variations in substrates (McCormick et al., 1961; McCormick and Snell, 1959, 1961). The loop in the flap of PLK is longer than for HMPP kinase and occupies the pyridoxal binding site. When ATP binds, the loop swings away to a position partially covering the ATP binding site to prevent unproductive hydrolysis of ATP.

AIRs kinase is an interesting enzyme because it shows both ancient and advanced features in the proposed evolution pathway. In AIRs kinase, the lid is fully formed by extending the two short β strands of HMPP kinase and inserting two additional β strands to form a four-stranded β sheet. This flap forms the dimer interface in the crystal structure. The current crystal structures do not show any significant conformational changes upon substrate and ATP binding. KDG kinase has almost exactly the same monomer topology and dimerization pattern as AIRs kinase.

The monomer of RK is very similar to AIRs kinase and KDG kinase; however, two four-stranded β sheets from the adjacent RK monomers form a β clasp structure resulting in a biological dimer. Furthermore, these two β sheets interchange a pair of strands, such that each ATP binding site is contributed by both β sheets (Sigrell et al., 1998). Ribose binding induces lid closing, while ATP binding causes only local conformational changes that are necessary for the phosphoryltransfer reaction (Sigrell et al., 1999).

Adenosine kinase employs the same catalytic mechanism as other superfamily members (Schumacher et al., 2000); however, the monomer is fully independent, with the lid as a small domain formed by the insertion of an additional β strand and two amphipathic α helices. The new β strand replaces the interchanged strand from the neighboring subunit in RK, and the two α helices

stabilize the five-stranded β sheet. In ADP-dependent glucokinase, the lid becomes more structurally complex with respect to the adenosine kinase by the extension of two β strands and the insertion of two additional α helices (Ito et al., 2001).

The two kinases with unknown functions are also placed into the possible evolution pathway based on their conserved active site residues, monomer topology, and quaternary structure. 1KYH has a simple fold similar to THZ kinase, with one conserved α helix replaced by a loop. It forms a tetramer in the crystal structure and does not possess a flap or lid structure. It represents an early structure in the evolution pathway of the ribokinase superfamily. 1O14 is very similar to RK, with a well evolved β clasp lid structure. It is expected to undergo conformational changes after the substrate is bound and is a relatively advanced enzyme in the ribokinase superfamily.

Experimental Procedures

Protein Expression and Purification

The stm4066 gene was cloned into a pET14b vector between NdeI and BamHI sites, which carries a sequence encoding an N-terminal polyhistidine tag. The plasmid was transformed into *E. coli* strain B834(DE3)pLysS (Novagen). The transformed cells were grown in 50 ml of LB containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol at 37°C until the OD_{600} reached 0.8. This culture was inoculated into 2 liters of fresh LB containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol. The resulting culture was grown at 37°C until the OD_{600} reached 0.6. Expression was induced by addition of isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.5 mM. The cells were harvested after 5 hr by centrifugation at 3000 × g for 20 min at 4°C and stored at −80°C.

All purification steps were carried out at 4°C. Cells were resuspended in 40 ml buffer A (50 mM sodium phosphate and 300 mM sodium chloride [pH 8.0]). The cells were lysed by one freezing-thawing cycle and two cycles of 3 × 10 s pulse sonication with a 40 s pause between each burst on ice. The crude extract was centrifuged at 12,000 × g for 25 min, and the supernatant was mixed for 1 hr with TALON IMAC resin (BD) preequilibrated with buffer A. The resin was poured into a column and then washed with buffer B (50 mM sodium phosphate, 300 mM sodium chloride, and 10 mM imidazole [pH 8.0]). AIRs kinase was eluted from the resin with buffer C (50 mM sodium phosphate, 300 mM sodium chloride, and 300 mM imidazole [pH 8.0]) and buffer exchanged into buffer D (10 mM Tris and 1 mM dithiothreitol (DTT) [pH 8.0]) using an Econo-Pac 10DG column (Bio-Rad). The purity of AIRs kinase was determined by Coomassie-stained SDS-PAGE and found to be greater than 95% (data not shown). The purified protein was then concentrated to 10 mg/mL using a 10 kDa cutoff concentrator (Amicon) and stored at −80°C. Protein concentrations were determined by the Bradford method (Bradford, 1976) using bovine serum albumin as the standard.

For the production of selenomethionine (SeMet)-incorporated protein, the cell growth conditions were similar to those described above. The growth medium contained M9 salts supplemented with 40 μg/l of all L-amino acids except L-methionine, which was replaced by L-SeMet, 0.4% (w/v) glucose, 2 mM $MgSO_4$, 25 μg/L $FeSO_4 \cdot 7H_2O$, 1 mM $CaCl_2$, 1% BME vitamin solution (GIBCO-BRL), 100 μg/l ampicillin, and 34 μg/l chloramphenicol. The purification procedure for the SeMet protein was the same as that for the native protein.

Crystallization

All crystallization experiments were carried out at 22°C using the hanging drop vapor diffusion method. Crystallization conditions were screened with native protein using Crystal Screen Kits I and II (Hampton Research). One microliter of 10 mg/ml protein was mixed with an equal amount of reservoir solution and equilibrated against 1 ml reservoir solution. The reservoir solution for the optimized conditions contained 14%–16% polyethylene glycol 3350 (PEG 3350), 0.15 M K_2SO_4, 4% tertiary-butanol or 4% isopropanol, 2 mM DTT, and 100 mM MES (pH 6.1). Crystals appeared within 1 day and grew to their maximum size (0.5 mm × 0.4 mm × 0.05 mm) in about 5 days. SeMet AIRs kinase was crystallized under the same conditions.

The crystals of the SeMet AIRs kinase/AIRs complex were obtained using solutions of SeMet AIRs kinase at 8 mg/ml and 7 mM AIRs with 5 mM $MgSO_4$. The crystals of the native AIRs kinase/AIRs/AMP-PCP complex were obtained using solutions of native AIRs kinase at 8 mg/ml, 7 mM AIRs, 7 mM AMP-PCP, and 5 mM $MgSO_4$. The reservoir solutions and crystallization conditions for the complex were the same as those for SeMet AIRs kinase except 5% glycerol was added to the reservoir.

All the crystals belong to space group $P2_12_12_1$ with approximate unit cell dimensions of a = 137.0 Å, b = 54.0 Å, and c = 89.9 Å. The crystals contain two monomers per asymmetric unit, with a solvent content of 48%.

X-Ray Data Collection and Processing

All data were collected at cryogenic temperatures. For cryoprotection, the crystals of SeMet AIRs kinase were gently transferred to a stabilization solution consisting of the mother liquor with 15% glycerol and then frozen directly in the nitrogen cold stream. The crystals of complex were taken out of the mother liquor and frozen directly in the nitrogen cold stream.

Monochromatic X-ray intensity data were measured for the unliganded AIRs kinase and the SeMet AIRs kinase/AIRs at the NE-CAT 8-BM beam line at the Advanced Photon Source using a Quantum-315 CCD detector (Area Detector Systems Corp.). The first data set was collected to 3.1 Å resolution at the wavelength corresponding to the peak of the selenium spectrum over a range of 110° using 30 s for each 1.0° oscillation at a crystal to detector distance of 400 mm. The second data set was collected to 2.6 Å resolution over a range of 160° using 30 s for each 0.5° oscillation at a crystal to detector distance of 375 mm. Bijovet pairs were measured after each 20° wedge using inverse beam geometry for both data sets. The data for the SeMet AIRs kinase/AIRs complex were collected over a range of 100° using 60 s for each 1.0° oscillation at a crystal to detector distance of 370 mm.

A data set of native AIRs kinase/AIRs/AMP-PCP was collected at the A1 station of the Cornell High Energy Synchrotron Source using a Quantum-210 CCD detector (Area Detector Systems Corp.) over a range of 130° using 30 s for each 1.0° oscillation at a crystal to detector distance of 219 mm.

The HKL2000 suite (Otwinowski and Minor, 1997) of programs was used for integration and scaling of all data sets. Details of the data collection and processing are given in Table 2.

Structure Determination

The initial Se atom positions were determined using shake-and-bake direct methods (Miller et al., 1993) as implemented in SnB (Weeks and Miller, 1999). The 3.1 Å data set was used to calculate normalized anomalous differences (ΔE) using the DREAR (Blessing and Smith, 1999) suite of programs. Twenty-five trials produced solutions that identified 10 out of 14 expected Se atom positions. These sites showed a noncrystallographic 2-fold symmetry. The 10 Se sites and scaled 2.6 Å data set were employed for phasing using SOLVE (Terwilliger and Berendzen, 1999). The phases were further improved by density modification using RESOLVE (Terwilliger, 2000). The figure of merit (FOM) improved from 0.33 to 0.55 after density modification. The resultant electron density map was interpretable.

Model Building and Refinement

All model building was performed using the computer graphics program O (Jones et al., 1991). The initial model was built with a polyalanine model. The map was subsequently further improved by density modification using a protein mask created from the structure of *E. coli* ribokinase and an NCS constraint using the RAVE program package (Kleywegt and Jones, 1996). The chain was traced for residues 5–88 and 98–307, and the second monomer was generated

using NCS. The side chains were then included for both monomers. The model refinements were carried out using the program CNS (Brünger et al., 1998) with the 2.6 Å resolution unliganded AIRs kinase data set. An NCS restraint was applied. The refinement procedure involved successive rounds of simulated annealing refinement, energy minimization, temperature factor refinement, and model rebuilding. As the model refinements proceeded, two strong peaks for each monomer in the electron density map were revealed and modeled as potassium. Water molecules were included in subsequent rounds of refinement.

The structure of the AIRs kinase/AIRs complex was determined by refining the unliganded AIRs kinase model against the data for the complex. AIRs was defined in an F_o–F_c electron density map contoured at 3σ and manually fitted. Bond distance and angle restraints for AIRs were generated using the program PRODRG (van Aalter et al., 1996). A similar strategy was employed for the AIRs kinase/AIRs/AMP-PCP complex. The electron density for AIRs and AMP-PCP was remarkably clear. One magnesium ion, two potassium ions, and water molecules for each monomer were added to the model in subsequent rounds of refinement. The final refinement statistics are given in Table 3.

Acknowledgments

We thank the NE-CAT beamline 8-BM of the Advanced Photon Source and the Cornell High Energy Synchrotron Source for providing beam time for the study. We thank Drs. Angela V. Toms and Ethan C. Settembre for assistance with data collection and Leslie Kinsland for assistance with the preparation of this manuscript. This work was supported by NIH grant RR-15301 and GM47296 (D.D.). M.D. was supported by a Biotechnology Traineeship from the N.I.H. (T32 GM08349) and a Louis and Elsa Thomsen Wisconsin Distinguished Fellowship Award.

Received: May 27, 2004
Revised: July 13, 2004
Accepted: July 13, 2004
Published: October 5, 2004

References

Allen, S., Zilles, J.L., and Downs, D.M. (2002). Metabolic flux in both the purine mononucleotide and histidine biosynthetic pathways can influence synthesis of the hydroxymethyl pyrimidine moiety of thiamine in *Salmonella enterica*. J. Bacteriol. *184*, 6130–6137.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Andersson, C.E., and Mowbray, S.L. (2002). Activation of ribokinase by monovalent cations. J. Mol. Biol. *315*, 409–419.

Andreeva, A., Howorth, D., Brenner, S., Hubbard, T., Chothia, C., and Murzin, A. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res. *32*, D226–D229.

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

Blessing, R.H., and Smith, G.D. (1999). Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations. J. Appl. Crystallogr. *32*, 664–670.

Bork, P., Sander, C., and Valencia, A. (1993). Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. Protein Sci. *2*, 31–40.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Campobasso, N., Mathews, I.I., Begley, T.P., and Ealick, S.E. (2000). Crystal structure of 4-methyl-5-b-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 Å resolution. Biochemistry *39*, 7868–7877.

Cheng, G., Bennett, E.M., Begley, T.P., and Ealick, S.E. (2002). Crystal structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase from *Salmonella typhimurium* at 2.3 Å resolution. Structure *10*, 225–235.

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

Darling, J.A., Sullivan, W.J.J., Carter, D., Ullman, B., and Roos, D.S. (1999). Recombinant expression, purification, and characterization of *Toxoplasma gondii* adenosine kinase. Mol. Biochem. Parasitol. *103*, 15–23.

Dougherty, M., and Downs, D.M. (2003). The *stm4066* Gene product of *Salmonella enterica* serovar Typhimurium has aminoimidazole riboside (AIRs) kinase activity and allows AIRs to satisfy the thiamine requirement of *pur* mutant strains. J. Bacteriol. *185*, 332–339.

Esnouf, R. (1997). An extensively modified version of Molscript which includes greatly enhanced colouring capabilities. J. Mol. Graph. *15*, 132–134.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrixes. J. Mol. Biol. *233*, 123–138.

Hope, J.N., Bell, A.W., Hermodson, M.A., and Groarke, J.M. (1986). Ribokinase from *Escherichia coli* K12. Nucleotide sequence and overexpression of the rbsK gene and purification of ribokinase. J. Biol. Chem. *261*, 7663–7668.

Ito, S., Fushinobu, S., Yoshioka, I., Koga, S., Matsuzawa, H., and Wakagi, T. (2001). Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. Structure *9*, 205–214.

Janin, J. (1997). Specific versus non-specific contacts in protein crystals. Nat. Struct. Biol. *4*, 973–974.

Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for the building of protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kleywegt, G.J., and Jones, T.A. (1996). xdlMAPMAN and xdlDATAMAN-programs for reformatting, analysis, and manipulation of biomacromolecular electron-density maps and reflection datasets. Acta Crystallogr. D Biol. Crystallogr. *52*, 826–828.

Koga, S., Yoshioka, I., Sakuraba, H., Takahashi, M., Sakasegawa, S., Shimizu, S., and Ohshima, T. (2000). Biochemical characterization, cloning, and sequencing of ADP-dependent (AMP-forming) glucokinase from two hyperthermophilic archaea, *Pyrococcus furiosus* and *Thermococcus litoralis*. J. Biochem. (Tokyo) *128*, 1079–1085.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Kwok, F., Scholz, G., and Churchich, J. (1987). Brain pyridoxal kinase dissociation of the dimeric structure and catalytic activity of the monomeric species. Eur. J. Biochem. *168*, 577–583.

Li, M.-H., Kwok, F., Chang, W.-R., Lau, C.-K., Zhang, J.-P., Lo, S.C.L., Jiang, T., and Liang, D.-C. (2002). Crystal structure of brain pyridoxal kinase, a novel member of the ribokinase superfamily. J. Biol. Chem. *277*, 46385–46390.

Martin, A.C. (1998). ProFit (London, SciTech Software).

Mathews, I.I., Erion, M.D., and Ealick, S.E. (1998). Structure of human adenosine kinase at 1.5 Å resolution. Biochemistry *37*, 15607–15620.

McCormick, D.B., and Snell, E.E. (1959). Pyridoxal kinase of human brain and its inhibition by hydrazine derivatives. Proc. Natl. Acad. Sci. USA *45*, 1371–1379.

McCormick, D.B., and Snell, E.E. (1961). Pyridoxal phosphokinases. II. Effects of inhibitors. J. Biol. Chem. *236*, 2085–2088.

McCormick, D.B., Gregory, M.E., and Snell, E.E. (1961). Pyridoxal phosphokinases. I. assay, distribution, purification, and properties. J. Biol. Chem. *236*, 2076–2084.

Merritt, E.A., and Bacon, D.J. (1997). Raster3D. Photorealistic Molecular Graphics. Methods Enzymol. *277*, 505–524.

Miller, R., DeTitta, G.T., Jones, R., Langs, D.A., Weeks, C.M., and Hauptman, H.A. (1993). On the application of the minimal principle to solve unknown structures. Science *259*, 1430–1433.

Miller, R., Gallo, S.M., Khalak, H.G., and Weeks, C.M. (1994). SnB: crystal structure determination via shake-and-bake. J. Appl. Crystallogr. *27*, 613–621.

Murzin, A., Brenner, S., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Newell, P.C., and Tucker, R.G. (1968). Biosynthesis of the pyrimidine moiety of thiamine. A new route of pyrimidine biosynthesis involving purine intermediates. Biochem. J. *106*, 279–287.

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

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Schumacher, M.A., Scott, D.M., Mathews, I.I., Ealick, S.E., Roos, D.S., Ullman, B., and Brennan, R.G. (2000). Crystal structures of *Toxoplasma gondii* adenosine kinase reveal a novel catalytic mechanism and prodrug binding. J. Mol. Biol. *298*, 875–893.

Sigrell, J.A., Cameron, A.D., Jones, T.A., and Mowbray, S.L. (1997). Purification, characterization, and crystallization of *Escherichia coli* ribokinase. Protein Sci. *6*, 2474–2476.

Sigrell, J.A., Cameron, A.D., Jones, T.A., and Mowbray, S.L. (1998). Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 Å resolution: insights into a new family of kinase structures. Structure *6*, 183–193.

Sigrell, J.A., Cameron, A.D., and Mowbray, S.L. (1999). Induced fit on sugar binding activates ribokinase. J. Mol. Biol. *290*, 1009–1018.

Spychala, J., Datta, N.S., Takabayashi, K., Datta, M., Fox, I.H., Gribbin, T., and Mitchell, B.S. (1996). Cloning of human adenosine kinase cDNA: sequence similarity to microbial ribokinases and fructokinases. Proc. Natl. Acad. Sci. USA *93*, 1232–1237.

Terwilliger, T.C. (2000). Maximum likelihood density modification. Acta Crystallogr. D Biol. Crystallogr. *56*, 965–972.

Terwilliger, T.C., and Berendzen, J. (1999). Automated structure solution for MIR and MAD (http://www.solve.lanl.gov). Acta Crystallogr. D *55*, 849–861.

Thierry, A., Fairhead, C., and Dujon, B. (1990). 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.

van Aalter, D.M., Bywater, R., Findlay, J.B., Hendlich, M., Hooft, R.W., and Vriend, G. (1996). PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. J. Comput. Aided Mol. Des. *10*, 255–262.

van Rooijen, R.J., van Schalkwijk, S., and 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.

Weeks, C.M., and Miller, R. (1999). The design and implementation of SnB v2.0. J. Appl. Crystallogr. *32*, 120–124.

Westhead, D.R., Slidel, T.W., Flores, T.P., and Thornton, J.M. (1999). Protein structural topology: automated analysis and diagrammatic representation. Protein Sci. *8*, 897–904.

Wu, L.F., Reizer, A., Reizer, J., Cai, B., Tomich, J.M., and 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 *Escherichia coli*. J. Bacteriol. *173*, 3117–3127.

Accession Numbers

The Protein Data Bank codes for AIRs kinase are as follows: 1TYY for unliganded AIRs kinase, 1TZ3 for the AIRs kinase/AIRs complex, and 1TZ6 for the AIRs kinase/AIRs/AMP-PCP complex.