Specificity evolution of the ADP-dependent sugar kinase family – *in silico* studies of the glucokinase/phosphofructokinase bifunctional enzyme from *Methanocaldococcus jannaschii*

Felipe Merino and Victoria Guixé

Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

Keywords
ADP-dependent kinase family; glucokinase/phosphofructokinase bifunctional enzyme; homology modeling; protein–ligand docking; specificity determinants

Correspondence
V. Guixé, Laboratorio de Bioquímica y Biología Molecular, Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile
Fax: +56 2 2712983
Tel: +56 2 9787335
E-mail: vguixe@uchile.cl

In several archaea of the *Euryarchaeota*, the glycolytic flux proceeds through a modified version of the Embden–Meyerhof pathway, where the phosphofructokinase and glucokinase enzymes use ADP as the phosphoryl donor. These enzymes are homologous to each other. In the hyperthermophilic methanogenic archaeon *Methanocaldococcus jannaschii*, it has been possible to identify only one homolog for these enzymes, which shows both ADP-dependent glucokinase and phosphofructokinase activity. This enzyme has been proposed as an ancestral form in this family. In this work we studied the evolution of this protein family using the Bayesian method of phylogenetic inference and real value evolutionary trace in order to test the ancestral character of the bifunctional enzyme. Additionally, to search for specificity determinants of these two functions, we have modeled the bifunctional protein and its interactions with both sugar substrates using protein–ligand docking and restricted molecular dynamics. The results show that the evolutionary story of this family is complex. The root of the family is located inside the glucokinase group, showing that the bifunctional enzyme is not an ancestral form, but could be a transitional form from glucokinase to phosphofructokinase, due to its basal location within the phosphofructokinase group. The evolutionary trace and the molecular modeling experiments showed that the specificity for fructose 6-phosphate is mainly related to the stabilization of a negative charge in the phosphate group, whereas the specificity for glucose is related to the presence of some histidines instead of glutamines/asparagines and to the interaction of this ligand with a glutamic acid residue corresponding to Glu82 in the bifunctional enzyme.

Several archaea of the *Euryarchaeota* present a uniquely modified Embden–Meyerhof pathway that involves only four of the classical enzymes present in the canonical pathway [1]. One of the most striking features of this modified glycolysis is the phosphorylation of glucose and fructose 6-phosphate by ADP as the phosphoryl donor, instead of ATP. These ADP-dependent kinases are homologous to each other and they show no sequence similarity to any of the hitherto known ATP-dependent enzymes. It was therefore proposed that they were part of a new family of kinases [2]. The presence of these ADP-dependent enzymes has been reported in several members of the *Thermococcales* [2–5], and also, on the basis of kinetic and

Abbreviations
ADP-GK, ADP-dependent glucokinase; ADP-GK/PFK, ADP-dependent glucokinase/phosphofructokinase; ADP-PFK, ADP-dependent phosphofructokinase; rvET, real value evolutionary trace.

genomic data, Verhees *et al*. proposed the operation of this pathway in some methanogenic archaea of the *Methanococcales* and *Methanosarcinales*. [6]. Recently, Ronimus and Morgan [7] reported a close homolog for these enzymes in higher eukaryotic genomes that has significant ADP-dependent glucokinase activity.

To date, the structures of the ADP-dependent glucokinases (ADP-GKs) from *Thermococcus litoralis* [8], *Pyrococcus furiosus* [9], and *Pyrococcus horikoshii* [10], and the ADP-dependent phosphofructokinase (ADP-PFK) from *P. horikoshii*, have been solved (Protein Data Bank: 1U2X). Surprisingly, despite the low sequence identity of these enzymes with the hitherto known ATP-dependent kinases, they can be classified, from a structural point of view, as members of the ribokinase superfamily [8]. The ribokinase-like fold is basically composed of an eight-stranded β-sheet surrounded by eight α-helices, three on one side and five on the other. This family was first proposed by Bork *et al*. [11] on the basis of sequence alignments only, as the first structure for a member of this family was published 5 years later [12]. Members of this family comprise ATP-dependent kinases of fructose 6-phosphate, fructose 1-phosphate, tagatose 6-phosphate, fructose, ribose, and nucleosides. Now, with more structural information available, it has been possible to recognize that the ribokinase superfamily also contains enzymes that can transfer the γ-phosphate of ATP to some vitamins involved in B6 synthesis, such as pyridoxal kinase [13] and ADP-dependent kinases. Thus, this superfamily can be subdivided into three major groups: the ATP-dependent sugar kinases described earlier by Bork; the ATP-dependent vitamin kinases; and the ADP-dependent sugar kinases. The main structural difference between the vitamin kinase enzymes and the other two groups is that the former present only the core αβα ribokinase-like fold (large domain), whereas the enzymes belonging to the other groups have, in addition, a small domain composed always of a β-sheet and sometimes of some α-helical insertions. This domain acts as a lid in the active site, and has been proposed as a good phylogenetic marker for the evolution of this superfamily [14]. Although the structures of several members of this superfamily are available, there are few reports that address, from a structural perspective, the evolution of the group [14].

As a result of the biological importance of the identity of the phosphoryl donor, some attention has been given to the structural determinants for the ATP/ADP specificity [8]. On the other hand, to the best of our knowledge, there are no studies addressing, from a structural point of view, the issue of the sugar specificity. Probably, this is mainly because there is no structure of an ADP-PFK cocrystallized with fructose 6-phosphate available to date.

The hyperthermophilic methanogenic archeon *Methanocaldococcus jannaschii* has just one ADP-dependent enzyme, which has been kinetically characterized by two different groups [6,15]. Interestingly, this enzyme has significant ADP-dependent kinase activity with glucose and fructose 6-phosphate. Inhibition studies showed that probably both sugars can bind to the same site in the enzyme [15]. On the basis of this feature, it was proposed that this enzyme is an ancestral form from which the two separated specificities originated through gene duplication. However, although several dendrograms have been published for the ADP-dependent family, there are no phylogenetic studies available.

In this work, we analyze the evolution of the ADP-dependent family of kinases using the Bayesian method of phylogenetic inference and real value evolutionary trace (rvET) [16] to test the ancestral character of the bifunctional enzyme and to search for specificity determinants of function. Also, we have modeled the ADP-dependent glucokinase/phosphofructokinase (ADP-GK/PFK) enzyme and its interaction with glucose and fructose 6-phosphate. To the best of our knowledge, this is the first time that a structural model of protein bound to fructose 6-phosphate has been available for a member of the ribokinase superfamily.

Results and Discussion

Evolutionary analysis of the ADP-dependent kinase family

To study the evolution of the archaeal members of the ADP-dependent family, a phylogenetic tree was constructed using the Bayesian method of phylogenetic inference with the eukaryotic members as the outgroup. The archaeal group presents ADP-GKs and ADP-PFKs, whereas the eukaryotic group presents only ADP-GKs. For this reason, it seems reasonable to assume that the divergence between the archaeal and eukaryotic groups occurred prior to the gene duplication event in the archaea. On the basis of this hypothesis, the eukaryotic group was chosen as the outgroup.

The inferred tree was very robust with regard to the improvements in the alignment used in terms of topology and posterior probability (not shown). Also, it was very similar to the dendrogram proposed by Ronimus and Morgan [7] using the maximum parsimony and neighbor joining methods. It shows basically five groups: the ADP-GKs from eukaryotic sources (out-

group sequences), the ADP-GKs from the *Thermococcales*, the ADP-GKs from the *Methanosarcinales*, the ADP-PFKs from the *Thermococcales* and *Methanococcales*, and the ADP-PFKs from the *Methanosarcinales* (Fig. 1). Surprisingly, the root of the cladogram was not between the ADP-PFKs and the ADP-GKs. Instead, it was located between the ADP-GKs from the *Thermococcales* and *Methanosarcinales*. This means that there was not a bifunctional origin as was proposed early by Sakuraba *et al.* [15]. The two ADP-PFK groups appear in the same clade as the *Thermococcales* ADP-GKs. Unfortunately, the only characterized gene from a methanogenic source is the ADP-GK/PFK from *M. jannaschii*. On the basis of the evolutionary trace and phylogenetic results (see below), it is possible that the ADP-PFKs from the *Methanosarcinales* can also use glucose as a phosphoryl acceptor. This could be important, especially for the operation of the Embden–Meyerhof pathway in *Methanococcoides burtonii*, as the ADP-GK gene in this organism has a large C-terminal deletion and this protein is probably not functional. Interestingly, the rest of the sequence shows the same conservation pattern of the other glucokinases, suggesting that either the event is recent or that this protein is still performing its function in some way.

To test the robustness of the root's place, a new rooting was performed using other members of the ribokinase superfamily as the outgroup. Again, the root of the group appears in the same place with a high posterior probability (not shown).

To explain the topology of the obtained phylogenetic tree, more than a gene duplication event is necessary, because the ADP-GKs from the *Thermococcales* appear to be more similar to the ADP-PFKs than to other ADP-GKs. One possible explanation could be that a gene duplication event occurred after the divergence between the *Thermococcales* and *Methanosarcinales*, and later, a lateral gene transfer event added the ADP-PFK activity to the methanogenic group. A similar scenario for the generation of paralogous proteins involving lateral gene transfer has been proposed by Gogarten *et al.* [17]. To test this hypothesis, the archaeal genes included in the former analysis were grouped according to their relative synonymous codon usage [18]. The clustering of the genes shows basically two groups (Fig. 2). One of them is formed by the genes from methanogenic archaea, and the other is formed by the genes from the hyperthermophilic archaea. This shows that the codon usage of these genes is in good agreement with the phylogeny of the archaeal group, and does not support the horizontal gene transfer proposal. However, this methodology fails when the lateral movement has occurred early in evolution, because codon usage is masked through time due to the accumulation of several mutations. This means that it is possible that the event that we are searching for is too ancient to be captured by this methodology. However, it has been shown that horizontal gene transfer is continuously modifying the prokaryotic genomes and that the rate of transfer of housekeeping genes is

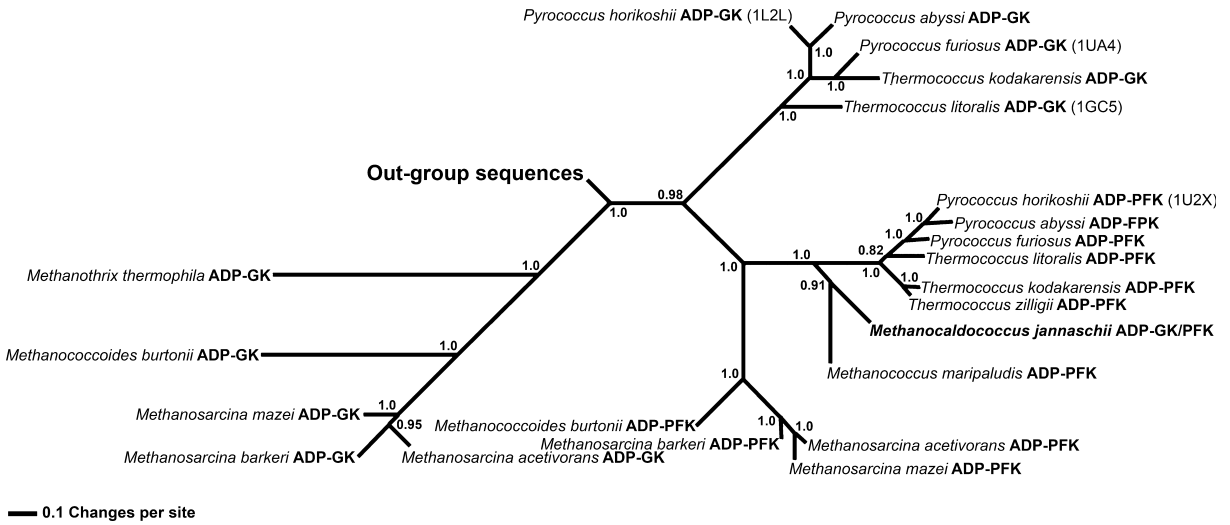

Fig. 1. Phylogenetic tree of the ADP-dependent sugar kinase family. The eukaryotic group was used as the outgroup. The posterior probability of each split is shown in the nodes. The distance to the outgroup sequences does not represent the distance in the real tree, because they are only used in the figure to point out the place where the root of the archaeal part is located. The position of the bifunctional enzyme is highlighted by bold letters.

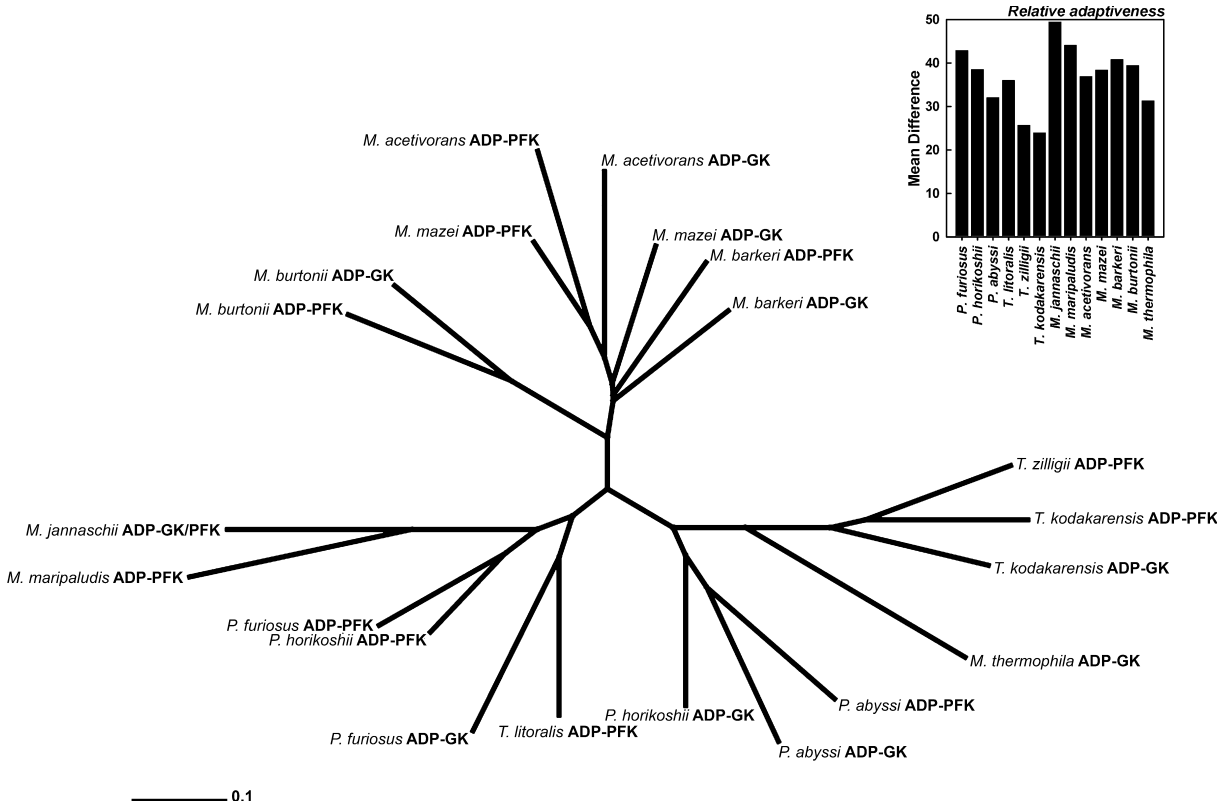

Fig. 2. Codon usage comparison for the ADP-dependent genes from archaeal sources. Dendrogram grouping the genes according to their average difference in relative synonymous codon usage. Inset: comparison between the codon usage of the ADP-GK from *M. thermophila* and the genomes of the archaea included in the relative synonymous codon usage analysis in terms of relative adaptiveness.

significant [19]. The only exception in the clustering seen in Fig. 2 is the ADP-GK from *Methanothrix thermophila*, which appears inside the hyperthermophilic archaeal group. Then, the codon usage of this gene was compared to the codon usage of the whole genomes of the archaea used in the phylogenetic inference in terms of relative adaptiveness [20]. This analysis showed that, in fact, the codon usage of this ADP-GK is more similar to the codon usage of the *Thermococcus* genus than to its own genome (Fig. 2, inset). *M. thermophila* is a thermophilic archeon that can grow at temperatures between 35 °C and 75 °C, whereas the archaea from the *Thermococcales* can grow between 65 °C and 100 °C. This small temperature overlap suggests that *M. thermophila* can live in the same habitat as some of the *Thermococcales* (at least in terms of growth temperature), giving the possibility of lateral gene transfer. Although this does not explain the topology of the phylogenetic tree, it provides evidence to confirm that lateral gene transfer occurred within this family. However, when the same procedure is used to compare the ADP-GKs and ADP-PFKs from the *Methanosarcinales* with their genomes, it is not possible to see any significant difference (not shown).

Molecular modeling of the bifunctional enzyme

To study in detail the structural determinants of the sugar specificity within the ADP-dependent kinase family, the structure of the ADP-GK/PFK enzyme from *M. jannaschii* was modeled using homology modeling. The active site of these enzymes is located in a cleft between the two domains. It has been previously reported that binding of the phosphoryl acceptor can change the relative orientation between the large and small domains in several kinases of the ribokinase superfamily [10,21,22]. Specifically, the two domains approach each other when the ligand is bound, changing the enzyme from an open to a closed conformation. For this reason, the bifunctional enzyme was modeled in two different conformations.

Table 1 shows the protein quality scores for the models used in the further analysis. In general, the models show above 92% of their residues within the core residues of the Ramachandran space and a

Table 1. Quality measurements of the chosen protein models in the open and closed conformation. The VERIFY 3D values presented are the sum of the scores for all the residues

	PROSA 2003	VERIFY 3D	PROCHECK				PROQ	
	Z-score		Core (%)	Allowed (%)	Generally allowed (%)	Disallowed (%)	LGSCORE	MAXSUB
Open conformation	−12.85	195.09	92.3	6.6	0.9	0.2	5.716	0.534
Closed conformation	−12.59	201.53	92	6.8	0.7	0.5	6.724	0.644
Templates								
1U2X	−13.99	213.9	93.1	6.9	0	0	6.956	0.642
1UA4	−14.32	220.67	93.3	6.2	0.5	0	7.562	0.776

very low percentage of them within the disallowed regions. The prosa2003 combined *Z*-score [23] is very close to that expected for a protein of 462 residues. Also, PROQ [24] shows that all the models are classified as good models and the VERIFY3D scores [25] of the chosen models are always above 0.1. The protein quality scores for the 30 models constructed are available in supplementary Table S1.

The models for the open and closed conformations are available in the Protein Model DataBase (http://mi.caspur.it/PMDB/) with the accession numbers PM0075152 and PM0075151 respectively.

To study in more detail the conformational change induced by the phosphoryl acceptor, models of the two conformations were compared using DYNDOM [26]. The angle of closure between the domains is about 47°. The analysis showed that 13 (Val27, Asp28, Ala29, Glu102, Glu103, Arg104, Lys170, Ile171, Asn172, Arg173, Ala201, Ser202, and Arg203) of the 462 residues in the protein act as a hinge, whereas the others belong to one of the two domains (Fig. 3). These residues belong to the coils between the domains and to a small part of the β-sheet of the small domain. Of these 13 residues, only Val27, Asp28, Ala29 and Arg104 are well conserved among the ADP-dependent kinase family. Of these, Asp28 is the only one that seems to be involved in sugar binding for both specificities (see below). Lys170 and Arg203 are well conserved within the ADP-PFK family, and seem to be related to ligand specificity (see below). Asn172 is an interesting residue; it also seems to be related to sugar discrimination, because it forms part of a conserved asparagine/histidine position in ADP-PFK and ADP-GK respectively. Glu102 and Glu103 are located in a region where there is at least a glutamic acid and a hydrophobic residue or another glutamic acid in all the thermophilic enzymes. In the mesophilic enzymes, there are always two hydrophobic residues. It has been shown that for a thermophilic/mesophilic adenylate kinase pair, the rate of opening of the nucleotide-binding lids is the limiting step for the observed k_{cat} [27]. As a result of the rigidity of the thermophilic enzyme, it was not able to perform catalysis at low temperatures. A similar mechanism could be operating in the ADP-dependent family, and those glutamic acids could be related to the rigidity of the hinge.

When the results of the rvET analysis [16] are mapped in the structures of the enzymes, two obvious

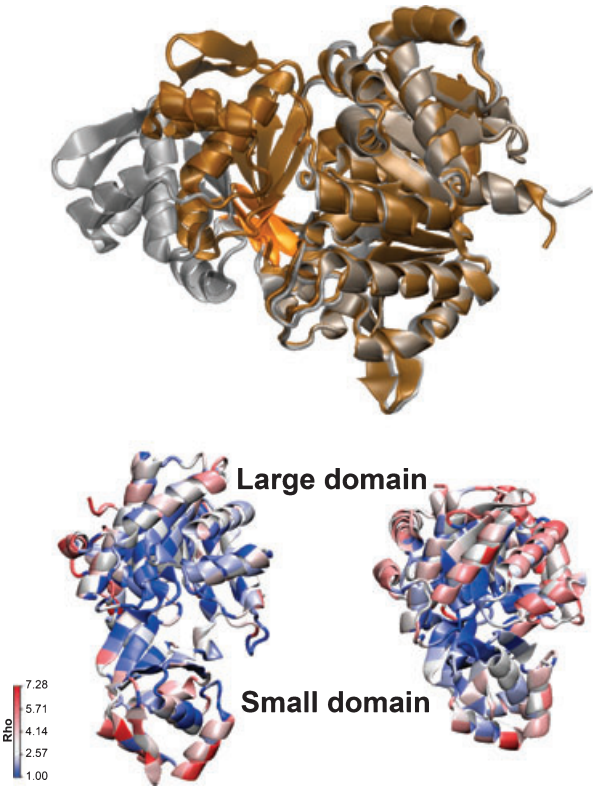

Fig. 3. Comparison between closed and open conformations of the bifunctional enzyme. Upper: The residues acting as a hinge between both domains are shown in orange. Lower: crystal structure of the ADP-PFK from *P. horikoshii* (left) and the ADP-GK from *P. furiosus* (right). Each residue is colored according to its rho value from the rvET analysis for the corresponding specificity.

patterns emerge (Fig. 3). First, it is clear that the ADP-PFK group has more conserved residues than the ADP-GK group. This seems reasonable, as the former group is evolutionarily newer. Second, a group of highly conserved residues in a cleft between the two domains, which forms the active site of these enzymes, can be seen. It is interesting that when the enzymes are in the closed conformation, the residues form a very dense cluster, which is not the case for the open conformation. From this point of view, it seems reasonable to assume that the closed conformation of these enzymes is necessary for the proper orientation of the ligands to produce catalysis as observed for other members of the superfamily [9,21,22].

Protein–ligand interaction modeling

The ADP-GK/PFK from *M. jannaschii* shows significant activity with glucose and fructose 6-phosphate *in vitro*. However, the relevance of this enzyme in the phosphorylation of these two sugars *in vivo* depends on the intracellular concentration of the metabolites as well as the specificity of the enzyme for them. It has been previously shown that this enzyme has specificity constants (k_{cat}/K_m) of 7.5×10^5 s^{-1} M^{-1} for fructose 6-phosphate and 1.2×10^4 s^{-1} M^{-1} for glucose [15]. For two competing substrates such as these, this means that in an equimolar mixture of both sugars, the rate at which fructose 6-phosphate is phosphorylated will be approximately 60 times higher than the rate at which glucose is phosphorylated at any substrate concentration [28]. This shows that, in fact, the enzyme from *M. jannaschii* is an unspecific phosphofructokinase, as can be also deduced by its position in the ADP-PFK branch of the phylogenetic tree (Fig. 1). Beyond this fact, the high K_m value seen for glucose could be a consequence of a high intracellular concentration of this metabolite. If this is true the ADP-GK/PFK could be performing both functions *in vivo*. It has been previously suggested by Verhees *et al.* that the modified Embeden–Meryerhof pathway could be operative in this archeon, on the basis of the presence of the ADP-GK/PFK enzyme (characterized in that publication just as ADP-PFK) and other characteristic enzymes of the pathway, such as glyceraldehyde-3-phosphate ferredoxin oxidoreductase [6]. The occurrence of the dual function *in vivo* would then be in good agreement with that hypothesis. Interestingly, it has been shown that this enzyme can use acetyl phosphate as phosphoryl donor, with a k_{cat} value similar to that obtained with ADP, but with a higher K_m value [6], demonstrating that there is promiscuity in both binding sites.

Another member of the ribokinase superfamily from *M. jannaschii* that is very similar to the ATP-dependent phosphofructokinase group has been characterized [29]. This enzyme was reported as a nucleoside kinase and, although it has a broad substrate specificity, it cannot use fructose 6-phosphate as a phosphoryl acceptor.

In order to obtain docking results comparable with the kinetic parameters measured for the bifunctional enzyme [15], the protonation state of the ionizable residues in the protein was predicted in the H++ server [30] at the pH value used in the kinetic characterization. It has been observed for the members of the ADP-dependent kinase family that the optimum pH value is lower for the phosphofructokinase activity than for the glucokinase activity [2–4]. To address this point, the protonation state of the ionizable residues in the bifunctional enzyme was also predicted using a pH value of 7.8, which is close the pH optimum for the glucokinase activity. There are several residues with significantly shifted pK_a values, as could be predicted for a protein with several ionizable residues. Among them, four lysines change their protonation state in the pH range studied (6.5 or 7.8). Lys9, Lys170, Lys178 and Lys238 were protonated at pH 6.5, whereas they were uncharged at pH 7.8. Of these four residues, Lys170 is the most interesting in relation to the sugar specificity problem. In the rvET experiments, this residue appears to be important for ligand discrimination, and also, in *Escherichia coli* Pfk-2, a member of the ribokinase superfamily, a structurally equivalent residue is crucial for fructose 6-phosphate binding (R. Cabrera and V. Guixé, unpublished results). This observation could explain the lower pH optimum for the phosphofructokinase than the glucokinase activity in the family.

Beyond the different way that ligands bind to the active site, one would expect that the catalytic mechanism would be conserved among members of a protein superfamily. In the ribokinase superfamily, it has been shown that an aspartic acid acts as the catalytic base in the phosphoryl transfer reaction, and that site-directed mutagenesis of this residue causes a dramatic loss of activity [9,31,32]. Additionally, a hydrogen bond between this residue and the OH phosphoryl acceptor in the ligand has always been seen by X-ray crystallography in members of the superfamily [9,21,22,33,34]. In the bifunctional enzyme, this aspartic acid corresponds to Asp442. In order for a docking conformation to be considered 'correct', there must be a contact between this residue and the phosphoryl acceptor OH group in the sugar.

The docking experiments described below showed that, as was mentioned before, the open conformation

of the enzyme was not appropriate for catalysis. When the docking experiment was performed with glucose as the ligand, 11 clusters of solutions were encountered using a 0.5 Å clustering cut-off. The most populated one contains 18 conformations with dissociation constants estimated at 298 K around 25 mM. Asp442 makes a hydrogen bond with the C2 hydroxyl group, whereas the C6 hydroxyl group makes two hydrogen bonds with the Asn26 and Asp28 side chains. The backbone of Gly107 also makes a hydrogen bond with the endocyclic oxygen. This suggests that the conformation is not appropriate for catalysis, as the hydroxyl acceptor is far away from the catalytic base.

When the ligand was fructose 6-phosphate, there were 37 clusters using the same cut-off as above, showing that, in fact, it was not possible to find a low-energy binding site. However, all the lower-energy clusters show substrate conformations that are bound to the protein mainly through ionic interactions with the phosphate group of the sugar and that are far away from the catalytic Asp442.

When glucose was docked to the closed conformation of the bifunctional enzyme, only three clusters were encountered. The first of them showed 48 conformations, and they are very similar to the conformation seen for glucose in the ADP-GK from *P. furiosus* [9]. The lower estimated free energy of binding in this cluster was -2.51 kcal·mol^{-1}, which corresponds to a K_d of 14.36 mM at 298 K, i.e. about one order of magnitude higher than the observed K_m of 1.6 mM.

In the case of the docking between the closed conformation and fructose 6-phosphate, 20 clusters were encountered. The first cluster showed only three conformations, with a very broad range of estimated free energy of binding. The second cluster contained nine conformations. The ring of the sugar in this cluster binds in a very similar position to that of glucose. Additionally, the phosphoryl acceptor OH group makes a hydrogen bond with the lateral chain of Asp442. For these reasons, this cluster was used for further analysis. The lower estimated free energy of binding in this cluster was -5.79 kcal·mol^{-1}, which corresponds to a K_d of 56 μM at 298 K, what is about five times higher than the observed K_m of 10 μM. Although there are differences between the predicted and observed binding free energies, the higher affinity of the enzyme for fructose 6-phosphate than for glucose is consistent with the kinetic data. Interestingly, when the same experiments are performed with the partial charges of the ligands derived using the Gasteiger method [35] the results are structurally very similar, but the predicted dissociation constants are 1.20 mM and 2.9 μM for glucose and fructose 6-phosphate respectively, which are in good agreement with the experimental data. However, the 'quantum mechanics-derived' charges are theoretically more appropriate.

Searching for contacts between the sugars and the protein, we inspected a 5 Å radius. Glucose makes contact with Asn24, Asn26, Asp28, Glu82, Gly107, Gln108, Ile111, Asn172, Ile174, Arg197, Ile199, Gln235, Val438, Gly439, and Asp442 (Fig. 4). Fructose 6-phosphate has the same contacts, with the exception of Glu82 and the inclusion of Arg203 and Lys170. The last residue is beyond the 5 Å cut-off, but it appears to make a weak ionic pair with the phosphate group of the ligand (Fig. 4).

Among the several contacts between the proteins and the sugars, only five residues seem to be important for the sugar specificity, on the basis of the rvET analysis (Fig. 4). This is a very striking result on the basis of the differences between a pyranose and a furanose phosphate, and it suggests high plasticity of the sugar-binding site within this protein family.

The Glu82 side chain makes a hydrogen bond with the OH group of C2 in glucose, but makes no contact with fructose 6-phosphate. The rvET rho values for this position in the alignment were 1 for the glucokinase group, 1.89 for the phosphofructokinase group, and 2.44 for the whole family. This residue has been proposed to be important for the discrimination of glucose specificity over fructose 6-phosphate [9,15]. Interestingly, this residue is present in all the methanogenic ADP-PFKs, suggesting the possibility of these enzymes binding glucose. Asn172 and Gln235 interact with the phosphate group when the ligand is fructose 6-phosphate. When the ligand is glucose, Asn172 interacts with the C3 hydroxyl group, whereas Gln235 makes a hydrogen bond with the side chain of Glu82. The equivalent residue to Gln235 in the ADP-GK from *P. furiosus* makes contact with the C1 hydroxyl group of glucose through a water molecule [9]. This interaction is not seen in our simulation, due to the absence of water. Asn172 and Gln235 are strictly conserved as asparagine and glutamine in the phosphofructokinase group. On the other hand, the Asn172 position is occupied by a highly conserved histidine, whereas the Gln235 position is occupied by glutamine or histidine in the glucokinase group. The explanation for this preference is not clear.

Lys170 and Arg203 interact with the phosphate group of fructose 6-phosphate through a weak ionic pair and a strong salt bridge respectively, as judged by their distance. Within the phosphofructokinase group, these residues are strictly conserved, whereas in the

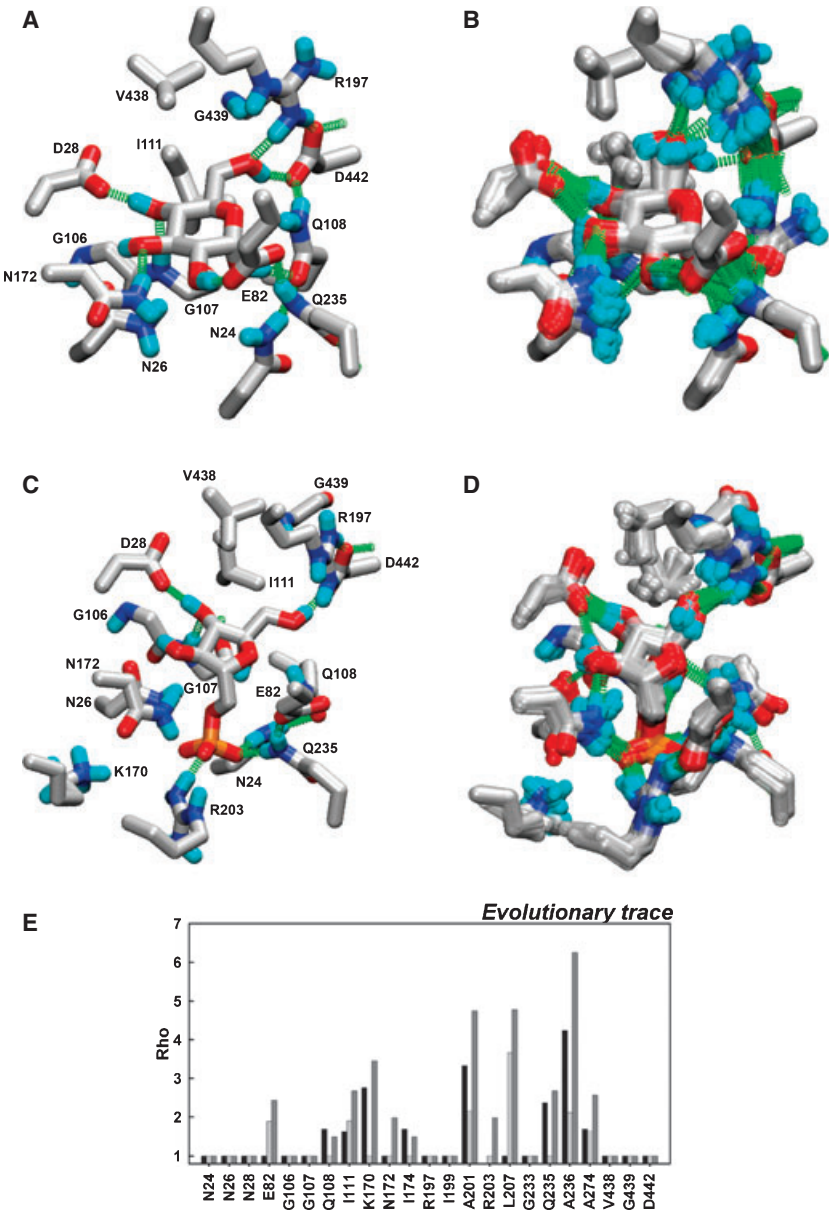

Fig. 4. Protein–ligand interaction modeling results. (A) α-D-glucose docked in the closed conformation of the bifunctional enzyme after the energy minimization procedure. (B) Set of representative frames of the α-D-glucose molecular dynamics simulation. (C) β-D-Fructose 6-phosphate docked in the closed conformation of the bifunctional enzyme after the energy minimization procedure. (D) Set of representative frames of the β-D-fructose 6-phosphate molecular dynamics simulation. (E) Results of the rvET analysis for all the residues within 5 Å from the ligands. The results for the glucokinase specificity are shown in black, those for the phosphofructokinase specificity in light gray, and those for the whole family in dark gray. Lys170 is beyond the 5 Å cut-off, but it seems to be important for phosphate binding. Residues 201, 207, 236 and 274 are not in contact with the ligand, but they are included in the 5 Å radius. Rho values for Val438, Gly439 and Asp442 were calculated excluding the ADP-GK from *Methanococcoides burtonii*, as the gene encoding this has a C-terminal deletion and the protein is probably not functional.

glucokinase group, the Lys170 position has an rvET rho value of 2.76 and the Arg203 position is a gap. In the glucokinase group, the two positions that follow Arg203 in the alignment are negatively charged residues that may repel the negative charge of the phosphate group in fructose 6-phosphate and thus prevent the binding of this ligand.

To test the stability of the contacts seen in the docking experiments and in order to refine the interaction with ligands, especially for fructose 6-phosphate, short

restricted molecular dynamics simulations were performed with the final docking results. Figure 4 shows a representative set of frames for each simulation. The interactions were conserved in the 0.5 ns of simulation, showing that the predicted contacts are stable. However, the hydroxyl acceptor in both ligands turned towards the sugar ring in the simulations far from the catalytic base, to interact with Arg197 and Gly439. This interaction occurred through the whole simulation for fructose 6-phosphate, and was intermittent with the interaction with Asp442 in the case of glucose. Arg197 and Gly439 have been seen to interact with glucose through a pair of water molecules in the crystal structure of *P. furiosus* [9]. The absence of these water molecules within the active site in the molecular dynamics simulations could explain the observed behavior. Although the position of the internal water for glucose could be predicted by copying the ones seen in the crystal structure of the ADP-GK from *P. furiosus*, there is no way to predict the internal waters for fructose 6-phosphate. Because in the closed conformation the active site is very deep in the protein structure, it would require an extremely long dynamic simulation to see the waters from the salvation sphere and to place them correctly within the sugar binding site.

In conclusion, we have shown that the root of the ADP-dependent kinase family is within the glucokinase group, which refutes the hypothesis of the ancestral character of the bifunctional enzyme. The results also suggest at least one horizontal gene transfer event in this family. Finally, it seems that the principal determinant for the fructose 6-phosphate specificity is the presence of residues with the capability to stabilize the negative charge of the phosphate group, whereas the presence of the Glu82 side chain and the preference for histidine over asparagine/glutamine are key elements for glucose specificity in the active site.

Experimental procedures

Sequence and structural alignments, evolutionary trace, codon usage, and phylogenetic analysis

The structure of the ADP-GKs from *P. horikoshii*, *P. furiosus* and *T. litoralis* (Protein Data Bank: 1L2L, 1UA4, and 1GC5, respectively) and the structure of the ADP-PFK from *P. horikoshii* (Protein Data Bank: 1U2X) were structurally aligned using the CE-MC program [36]. Additionally, each pdb file was split off into small and large domains. A new structural alignment was performed for the separated domains using the DEEPVIEW program [37]. Using these two structural alignments, a final sequence alignment was constructed. All the archaeal sequences for ADP-dependent kinases available in the GenBank nonredundant database were aligned in CLUSTALX [38] using the structural alignment previously constructed as a profile, and finally, this alignment was corrected manually. Then, the sequences of the eukaryotic ADP-GKs were aligned in CLUSTALX with the former alignment. This final alignment was again corrected manually, using the secondary structure prediction performed by the JPRED server [39] (available at http://www.compbio.dundee.ac.uk/~www-jpred) for the *Mus musculus* ADP-GK as a guide.

To study possible specificity determinants within this family, an rvET [16] analysis was performed for the archaeal sequences, including both specificities and the two specificities separately.

A phylogenetic tree for the archaeal ADP-dependent family of kinases was constructed using the Bayesian method of phylogenetic inference implemented in MRBAYES 3.1 [40,41]. The eukaryotic sequences were used as the outgroup. To test the possibility of horizontal gene transfer in this family, the genes were grouped according to their relative synonymous codon usage with the GCUA 1.2 software [18]. Additionally, the ADP-GK gene from *M. thermophila* was compared with all the archaeal genomes included in the phylogenetic analysis, using the graphical codon usage analysis software [20]. Also, ADP-PFKs and ADP-GKs from the *Methanosarcinales* were compared to their own genomes using the same procedure.

Molecular modeling of the bifunctional enzyme

As it has been shown that binding of the phosphoryl-acceptor ligand can change the angle between the two domains in members of the ribokinase superfamily, the bifunctional enzyme was modeled in two different conformations. The open conformation was modeled using 1U2X as template. This enzyme has 40% sequence identity with the ADP-GK/PFK. There is a unique structure in the closed conformation (1UA4) that shares only 25% sequence identity with the ADP-GK/PFK. Because this level of sequence similarity produces models that are not appropriate for the studies performed in this work, 1U2X and 1UA4 were used as templates, as follows. 1U2X was split off into its large and small domains and then aligned with 1UA4. Later, the closed enzyme and also the two fragments were used as templates for the closed conformation.

Fifteen models were constructed for each conformation with MODELLER 8 [42]. The quality of the models was evaluated using PROCKECK [43], PROSA2003 [22], VERIFY3D [25], and PROQ [24].

Partial charge derivation and docking experiments

To test the results of the rvET analysis, α-D-glucose and β-D-fructose 6-phosphate were docked to the best model of the enzyme in each conformation. The atomic partial

charges of each molecule were derived with the RED III program [44], using the rigid body reorientation algorithm and the RESP-A1 model for the charge fitting. The quantum mechanics software used was PC-GAMESS 7.1 (A. A. Granovsk, http://classic.chem.msu.su/gran/gamess/index.html). The derivation for fructose 6-phosphate was performed using the multimolecule approach with β-D-fructose and methyl-phosphate as fragments.

The K_m measurements performed by Sakuraba *et al*. [15] were performed at pH 6.5 for fructose 6-phosphate and glucose. To take this effect into account, the protonation state of the ionizable residues was predicted at pH 6.5 in each conformation using the web server H++ (available at http://biophysics.cs.vt.edu/H++) [30]. After this procedure, partial charges were added with the CHARMM force field [45].

The docking experiments were performed with AUTODOCK 4 [46], with the protonation states in accord with the pH values used in the kinetic experiments performed by Sakuraba *et al*. [15]. Fifty separated runs of the Lamarkian genetic algorithm using 2 500 000 energy evaluations and the other parameters as default were performed for both ligands in the open conformation. The same procedure was used for the closed conformation, except that when the ligand was fructose 6-phosphate, the number of energy evaluations was raised to 25 000 000.

All of the side chains were rigid in the experiments with the open conformation and in the experiment with the closed conformation and glucose. After the modeling procedure, the side chain of Lys170 pointed far from the binding site in the closed conformation, probably due to charge constraints in the absence of the phosphate group of fructose 6-phosphate. For this reason, this side chain was free to move when the ligand was fructose 6-phosphate in this conformation.

In the two docking experiments that involved glucose, only the bond between C5 and C6 was free to rotate, whereas when the ligand was fructose 6-phosphate, the bonds between C1 and C2, C5 and C6, C6 and O6, and O6 and the phosphorous atom were free to rotate.

Restricted molecular dynamics experiments

To test whether the interactions predicted by the docking experiments for both sugars in the closed conformation were stable, restricted molecular dynamics simulations were performed. The system was solvated with the smallest possible sphere of water, and all the simulations were performed with NAMD 2.6, using the spherical boundary condition and the SHAKE constraints [47]. The systems were minimized with 1500 iterations of the conjugated gradient algorithm. Later, the system was equilibrated at 300 K for 50 ps, and then the simulation was extended for 0.5 ns using a time step of 2 fs. All of the residues with at least one atom inside a 7 Å radius of the ligand were free to move, whereas the others were fixed in the equilibration and simulation steps. All of the atoms were free in the minimization procedure. Both simulations were performed using the CHARMM force field [45,48,49]. The parameters for glucose were available in CHARMM. For fructose 6-phosphate, we used the parameters for the sugar phosphate available for the nucleic acids [48,49]. The same partial charges used in the docking experiments were used here.

All the structural figures were drawn, and the trajectory analyses were performed, using the VMD software [50].

Acknowledgements

We acknowledge Dr Ricardo Cabrera for critical reading of the manuscript. This work was supported by Grant 1070111 from the Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) Chile.

References

1 Verhees CH, Kengen SW, Tuininga JE, Schut GJ, Adams MW, De Vos WM & Van Der Oost J (2003) The unique features of glycolytic pathways in Archaea. *Biochem J* **375**, 231–246.

2 Tuininga JE, Verhees CH, van der Oost J, Kengen SW, Stams AJ & de Vos WM (1999) Molecular and biochemical characterization of the ADP-dependent phosphofructokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Biol Chem* **274**, 21023–21028.

3 Ronimus RS, Koning J & Morgan HW (1999) Purification and characterization of an ADP-dependent phosphofructokinase from *Thermococcus zilligii*. *Extremophiles* **3**, 121–129.

4 Koga S, Yoshioka I, Sakuraba H, Takahashi M, Sakasegawa S, Shimizu S & 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* **128**, 1079–1085.

5 Jeong JJ, Fushinobu S, Ito S, Shoun H & Wakagi T (2003) Archaeal ADP-dependent phosphofructokinase: expression, purification, crystallization and preliminary crystallographic analysis. *Acta Crystallogr D Biol Crystallogr* **59**, 1327–1329.

6 Verhees CH, Tuininga JE, Kengen SW, Stams AJ, van der Oost J & de Vos WM (2001) ADP-dependent phosphofructokinases in mesophilic and thermophilic methanogenic archaea. *J Bacteriol* **183**, 7145–7153.

7 Ronimus RS & Morgan HW (2004) Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase. *Biochem Biophys Res Commun* **315**, 652–658.

8 Ito S, Fushinobu S, Yoshioka I, Koga S, Matsuzawa H & Wakagi T (2001) Structural basis for the ADP-speci-

ficity of a novel glucokinase from a hyperthermophilic archaeon. *Structure* **9**, 205–214.
9 Ito S, Fushinobu S, Jeong JJ, Yoshioka I, Koga S, Shoun H & Wakagi T (2003) Crystal structure of an ADP-dependent glucokinase from *Pyrococcus furiosus*: implications for a sugar-induced conformational change in ADP-dependent kinase. *J Mol Biol* **331**, 871–883.
10 Tsuge H, Sakuraba H, Kobe T, Kujime A, Katunuma N & Ohshima T (2002) Crystal structure of the ADP-dependent glucokinase from *Pyrococcus horikoshii* at 2.0-A resolution: a large conformational change in ADP-dependent glucokinase. *Protein Sci* **11**, 2456–2463.
11 Bork P, Sander C & 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.
12 Sigrell JA, Cameron AD, Jones TA & Mowbray SL (1998) Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 A resolution: insights into a new family of kinase structures. *Structure* **6**, 183–193.
13 Li MH, Kwok F, Chang WR, Lau CK, Zhang JP, Lo SC, Jiang T & Liang DC (2002) Crystal structure of brain pyridoxal kinase, a novel member of the ribokinase superfamily. *J Biol Chem* **277**, 46385–46390.
14 Zhang Y, Dougherty M, Downs DM & Ealick SE (2004) Crystal structure of an aminoimidazole riboside kinase from *Salmonella enterica*: implications for the evolution of the ribokinase superfamily. *Structure* **12**, 1809–1821.
15 Sakuraba H, Yoshioka I, Koga S, Takahashi M, Kitahama Y, Satomura T, Kawakami R & Ohshima T (2002) ADP-dependent glucokinase/phosphofructokinase, a novel bifunctional enzyme from the hyperthermophilic archaeon *Methanococcus jannaschii*. *J Biol Chem* **277**, 12495–12498.
16 Mihalek I, Res I & Lichtarge O (2004) A family of evolution–entropy hybrid methods for ranking protein residues by importance. *J Mol Biol* **336**, 1265–1282.
17 Gogarten JP, Doolittle WF & Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**, 2226–2238.
18 McInerney JO (1998) GCUA: general codon usage analysis. *Bioinformatics* **14**, 372–373.
19 Jain R, Rivera MC & Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* **96**, 3801–3806.
20 Fuhrmann M, Hausherr A, Ferbitz L, Schodl T, Heitzer M & Hegemann P (2004) Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Mol Biol* **55**, 869–881.
21 Sigrell JA, Cameron AD & Mowbray SL (1999) Induced fit on sugar binding activates ribokinase. *J Mol Biol* **290**, 1009–1018.
22 Schumacher MA, Scott DM, Mathews II, Ealick SE, Roos DS, Ullman B & Brennan RG (2000) Crystal structures of *Toxoplasma gondii* adenosine kinase reveal a novel catalytic mechanism and prodrug binding. *J Mol Biol* **298**, 875–893.
23 Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**, 355–362.
24 Wallner B & Elofsson A (2003) Can correct protein models be identified? *Protein Sci* **12**, 1073–1086.
25 Luthy R, Bowie JU & Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83–85.
26 Hayward S & Lee RA (2002) Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* **21**, 181–183.
27 Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ & Kern D (2004) Linkage between dynamics and catalysis in a thermophilic–mesophilic enzyme pair. *Nat Struct Mol Biol* **11**, 945–949.
28 Cornish-Bowden A (1995) *Fundamentals of Enzyme Kinetics*, pp 105–108. Portand Press, London.
29 Hansen T, Arnfors L, Ladenstein R & Schonheit P (2007) The phosphofructokinase-B (MJ0406) from *Methanocaldococcus jannaschii* represents a nucleoside kinase with a broad substrate specificity. *Extremophiles* **11**, 105–114.
30 Gordon JC, Myers JB, Folta T, Shoja V, Heath LS & Onufriev A (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **33**, 368–371.
31 Maj MC, Singh B & Gupta RS (2000) Structure–activity studies on mammalian adenosine kinase. *Biochem Biophys Res Commun* **275**, 386–393.
32 Campobasso N, Mathews II, Begley TP & Ealick SE (2000) Crystal structure of 4-methyl-5-beta-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 A resolution. *Biochemistry* **39**, 7868–7877.
33 Ohshima N, Inagaki E, Yasuike K, Takio K & Tahirov TH (2004) Structure of *Thermus thermophilus* 2-keto-3-deoxygluconate kinase: evidence for recognition of an open chain substrate. *J Mol Biol* **340**, 477–489.
34 Safo MK, Musayev FN, Hunt S, di Salvo ML, Scarsdale N & Schirch V (2004) Crystal structure of the PdxY protein from *Escherichia coli*. *J Bacteriol* **186**, 8074–8082.
35 Gasteiger J & Marsili M (1980) Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron* **36**, 3219–3228.
36 Guda C, Scheeff ED, Bourne PE & Shindyalov IN (2001) A new algorithm for the alignment of multiple protein structures using Monte Carlo optimization. *Pac Symp Biocomput* **7**, 275–286.

37 Guex N & Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723.
38 Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG (1997) The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
39 Cuff JA, Clamp ME, Siddiqui AS, Finlay M & Barton GJ (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892–893.
40 Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.
41 Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
42 Sali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779–815.
43 Laskowski RA, MacArthur MW, Moss DS & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.
44 Pigache A, Cieplak P & Dupradeau FY (2004) *Automatic and highly reproducible RESP and ESP charge derivation: application to the development of programs RED and X RED*. 227th ACS National Meeting, Anaheim, CA, USA (http://q4md-forcefieldtools.org/RED).
45 MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S *et al*. (1998) All-Atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586–3616.
46 Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK & Olson AJ (1998) Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* **19**, 1639–1662.
47 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L & Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802.
48 Foloppe N & MacKerell AD (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem* **21**, 86–104.
49 MacKerell AD & Banavali NK (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J Comput Chem* **21**, 105–120.
50 Humphrey W, Dalke A & Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**, 33–38.

Supplementary material

The following supplementary material is available online:

Table S1. Quality measurements of the protein models in the open and closed conformations.

This material is available as part of the online article from http://www.blackwell-synergy.com

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.