Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of *Methanocaldococcus jannaschii* nucleoside kinase: an archaeal member of the ribokinase family

Linda Arnfors,[a] Thomas Hansen,[b] Peter Schönheit,[b] Rudolf Ladenstein[a] and Winfried Meining[a]*

[a]Karolinska Institutet, Department of Biosciences at Novum, Center for Structural Biochemistry, Sweden, and [b]Christian-Albrechts-Universität Kiel, Institut für Allgemeine Mikrobiologie, Germany

Correspondence e-mail: winfried.meining@biosci.ki.se

Received 5 April 2006
Accepted 27 June 2006

PDB References: MjNK, 2c4e, r2c4esf; MjNK complex, 2c49, r2c49sf.

Nucleoside kinase from the hyperthermophilic archaeon *Methanocaldococcus jannaschii* (MjNK) is a member of the ribokinase family. In the presence of ATP and Mg^{2+}, MjNK is able to catalyze the phosphorylation of a variety of nucleosides, including inosine, cytidine, guanosine and adenosine. Here, the crystal structure of MjNK, the first structure of an archaeal representative of the ribokinase family, is presented. The structure was solved using the multiple-wavelength anomalous dispersion technique. Three-dimensional structures of the unliganded enzyme and a complex of MjNK, an ATP analogue and adenosine were determined to 1.7 and 1.9 Å resolution, respectively. Each subunit comprises an α/β-domain and a smaller lid domain and has an overall fold characteristic of the ribokinase superfamily. MjNK shares highest structural similarity to the ribokinases from *Escherichia coli* and *Thermotoga maritima*. Similar to ribokinase and other superfamily members, the lid domain of MjNK undergoes a significant conformational change upon substrate binding. In the crystal structure of the MjNK complex, subunit *A* adopts a closed conformation and subunit *B* an open conformation. In subunit *A* all substrates and Mg^{2+} were observed, whereas in subunit *B* only the ATP analogue could be clearly identified in the electron density. The structures of MjNK and *E. coli* ribokinase (EcRK) were compared with respect to putative determinants of thermal stability. Relative to EcRK, MjNK shows an increased charged and a decreased hydrophobic accessible surface area, as well as a higher fraction of charged residues, ionic networks and large aromatic clusters, characteristics that are frequently observed in enzymes from hyperthermophiles.

1. Introduction

Hyperthermophiles are mainly, but not exclusively, found in the kingdom of archaea. Enzymes from hyperthermophiles are often homologous to their mesophilic counterparts and their reaction mechanisms are generally identical. However, enzymes from hyperthermophiles can withstand higher temperatures and in many cases even need high temperatures to maintain their enzymatic activity. There is no single explanation of high thermotolerance; however, several parameters, such as a statistical prevalence of ionic interactions at the protein surface, increased formation of ionic networks and reduction of hydrophobic accessible surface area (Karshikoff & Ladenstein, 2001; Spassov *et al.*, 1995; Yip *et al.*, 1995), are found more frequently in hyperthermophiles.

Methanocaldococcus jannaschii nucleoside kinase (MjNK) is a hyperthermostable enzyme capable of phosphorylating a wide range of nucleosides in the presence of ATP and

Table 1
The ribokinase superfamily.

Family	Enzyme	Organism	Lid	Quaternary structure
PfkB family of carbohydrate kinases (ribokinase family, PF00294)	Ribokinase	*E. coli* (Sigrell *et al.*, 1998)	+	Dimer
		Thermotoga maritima (Joint Center for Structural Genomics, unpublished work)	+	Dimer
	Adenosine kinase	Human (Mathews *et al.*, 1998)	+	Monomer
		Toxoplasma gondii (Schumacher *et al.*, 2000)	+	Monomer
	2-keto-3-deoxygluconate (KDG) kinase	*Thermus thermophilus* (Ohshima *et al.*, 2004)	+	Hexamer
		Thermotoga maritima (Joint Center for Structural Genomics, unpublished work)	+	Dimer
	PdxY protein	*E. coli* (Safo *et al.*, 2004)	−	Dimer
	Brain pyridoxal kinase	Sheep (Li *et al.*, 2002)	−	Dimer
	Aminoimidazole riboside kinase	*Salmonella enterica* (Zhang *et al.*, 2004)	+	Dimer
	PfkB carbohydrate kinase	*Thermotoga maritima* (Joint Center for Structural Genomics, unpublished work)	−	Monomer
	Possible 1-phosphofructokinase	*Thermotoga maritima* (Joint Center for Structural Genomics, unpublished work)	+	Dimer
	Phosphomethylpyrimidine kinase	*Thermus thermophilus* (Bagautdinov *et al.*, unpublished work)	−	Dimer
	Tagatose-6-phosphate kinase	*Enterococcus faecalis* (Gorman & Shapiro, unpublished work)	+	Dimer
ADP-specific phosphofructokinase/ glucokinase family (PF04587)	ADP-dependent glucokinases	*Thermococcus litoralis* (Ito *et al.*, 2001)	+	Monomer
		Pyrococcus furiosus (Ito *et al.*, 2003)	+	Dimer
		Pyrococcus horikoshii (Tsuge *et al.*, 2002)	+	Monomer
	ADP-dependent phosphofructokinase	*Pyrococcus horikoshii* (Wong & Jia, unpublished work)	+	Monomer
Hydroxyethylthiazole kinase family (PF02110)	4-Methyl-5-β-hydroxyethylthiazole kinase	*Bacillus subtilis* (Campobasso *et al.*, 2000)	−	Trimer
	Hydroxyethylthiazole kinase protein	*Pyrococcus horikoshii* (Jeyakanthan & Tahirov, unpublished work)	−	Trimer
Phosphomethylpyrimidinekinase family (PF08543)	4-Amino-5-hydroxymethyl-2-methyl-pyrimidine phosphate kinase	*Salmonella typhimurium* (Cheng *et al.*, 2002)	−	Dimer
Carbohydrate kinase family (PF01256)	Carbohydrate kinase	*Bacillus subtilis* hypothetical protein in SIGY-CYDD intergenic region (Zhang *et al.*, 2002)	−	Tetramer

magnesium (Fig. 1). MjNK shows the highest catalytic activity for inosine, guanosine and cytidine, but also phosphorylates adenosine, thymidine, uridine and xanthosine to some extent (Hansen *et al.*, 2006). MjNK is a member of the phosphofructokinase B (PfkB) family of carbohydrate kinases, also known as the ribokinase family (Pfam PF00294; Bork *et al.*, 1993; Wu *et al.*, 1991). This diverse protein family is present in all three domains of life: archaea, bacteria and eukarya. The first crystal structures determined within the ribokinase family were *Escherichia coli* ribokinase (EcRK; Sigrell *et al.*, 1998) and human adenosine kinase (hAK; Mathews *et al.*, 1998). The structures of the ribokinase family members share a similar overall fold (SCOP 53613; Andreeva *et al.*, 2004; Murzin *et al.*, 1995). In addition to the structures of EcRK and hAK, the structures of various other members of the ribokinase superfamily have been determined (see Table 1).

Enzymes belonging to the ribokinase superfamily are phosphotransferases with a methyl hydroxyl group as the phosphate-group acceptor. However, *Salmonella typhimurium* 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase (StHMPPK; Cheng *et al.*, 2002) catalyzes both the phosphorylation of a hydroxymethyl pyrimidine (HMP) and the phosphorylation of the phosphomethyl group of HMPP. Members of the ribokinase superfamily share an α/β-fold core consisting of a central eight-stranded β-sheet and eight flanking α-helices. The active site is positioned in a shallow groove along one edge of the β-sheet. The phosphate-acceptor hydroxyl group and γ-phosphate are close neighbours in the centre of the groove and the substrate and adenosine part of ATP are located at opposite ends of the groove. In addition to the α/β-domain, many members of the ribokinase superfamily possess a smaller lid domain. Most of the enzymes comprising the lid domain undergo a conformational change upon substrate binding. The conformational change results in a movement of the lid domain that brings it closer to the α/β-domain and thereby covers the active site and shields the substrate from the solvent. The lid domain has been suggested to be a morphological marker of evolution within the ribokinase superfamily (Cheng *et al.*, 2002; Zhang *et al.*, 2004). Even though the members of the ribokinase superfamily share a similar monomeric overall fold, the enzymes show significant variation in their quaternary structure (see Table 1).

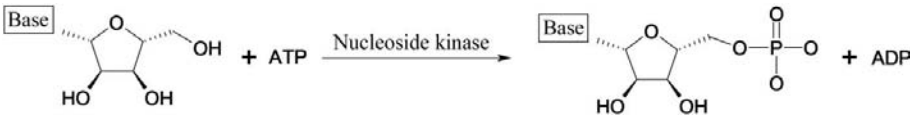

Figure 1
The reaction catalyzed by nucleoside kinase.

Here, we report the first structure of an archaeal member of the ribokinase family. The structure of MjNK has been determined for the apo enzyme as well as the enzyme in complex with the ATP analogue phosphoaminophosphonic acid-adenylate ester (AMP-

PNP) and the substrate adenosine. In order to identify possible reasons for differences in thermostability between nucleoside kinase and ribokinase, we have also compared certain structural features of MjNK and EcRK.

2. Materials and methods

2.1. Expression and purification

Expression and purification of the native enzyme have been described elsewhere (Arnfors *et al.*, 2005).

E. coli BL21 (DE3) pLys S cells were transformed with a pET-17b plasmid, which contained the *mjnk* gene under the control of a T5 promoter and a *lac* operator (Hansen & Schönheit, unpublished work). Transformed cells were grown at 310 K in 400 ml M9 minimal medium supplemented with 20 m*M* glucose as the sole carbon source in 2 l shaking flasks. For selection, 100 µg ml^{-1} carbenicillin and 34 µg ml^{-1} chloramphenicol were used. When the cells reached an optical density at 600 nm of 0.3, the following seven solid amino acids were added to the final concentrations given in parentheses: L-lysine (100 µg ml^{-1}), L-phenylalanine (100 µg ml^{-1}), L-threonine (100 µg ml^{-1}), L-isoleucine (50 µg ml^{-1}), L-leucine (50 µg ml^{-1}), L-valine (50 µg ml^{-1}), L-selenomethione (50 µg ml^{-1}).

Following 15 min of further growth, protein expression was induced by the addition of 0.8 m*M* IPTG. After further 6 h of growth ($OD_{600} \simeq 2.0$–2.6), the cells were harvested by centrifugation at 277 K and washed in a buffer containing 50 m*M* Tris–HCl pH 7.0 and 50 m*M* NaCl. The resulting recombinant selonomethionine derivate of MjNK was purified as recently described for recombinant MjNK (Arnfors *et al.*, 2005), except that 5 m*M* DTT was added to all protein solutions and buffers used during the purification.

The His-tagged ribokinase from *E. coli* was overexpressed and purified using a slightly modified version of the procedure described by Andersson & Mowbray (2002). The cells (wet weight 3 g l^{-1}) were stored in 10 ml 50 m*M* Tris–HCl pH 7.8, 300 m*M* NaCl at 253 K. Thawed cells were treated with one Complete EDTA-free tablet (Roche Diagnostics). Cells were lysed using a French press and the supernatant after centrifugation was added to 2 ml Talon metal-affinity resin (Clontech). The purification of the ribokinase was performed at room temperature using the batch method presented by Clontech. The enzyme was dialysed into 50 m*M* Tris–HCl pH 7.8, 150 m*M* NaCl and stored at 277 K.

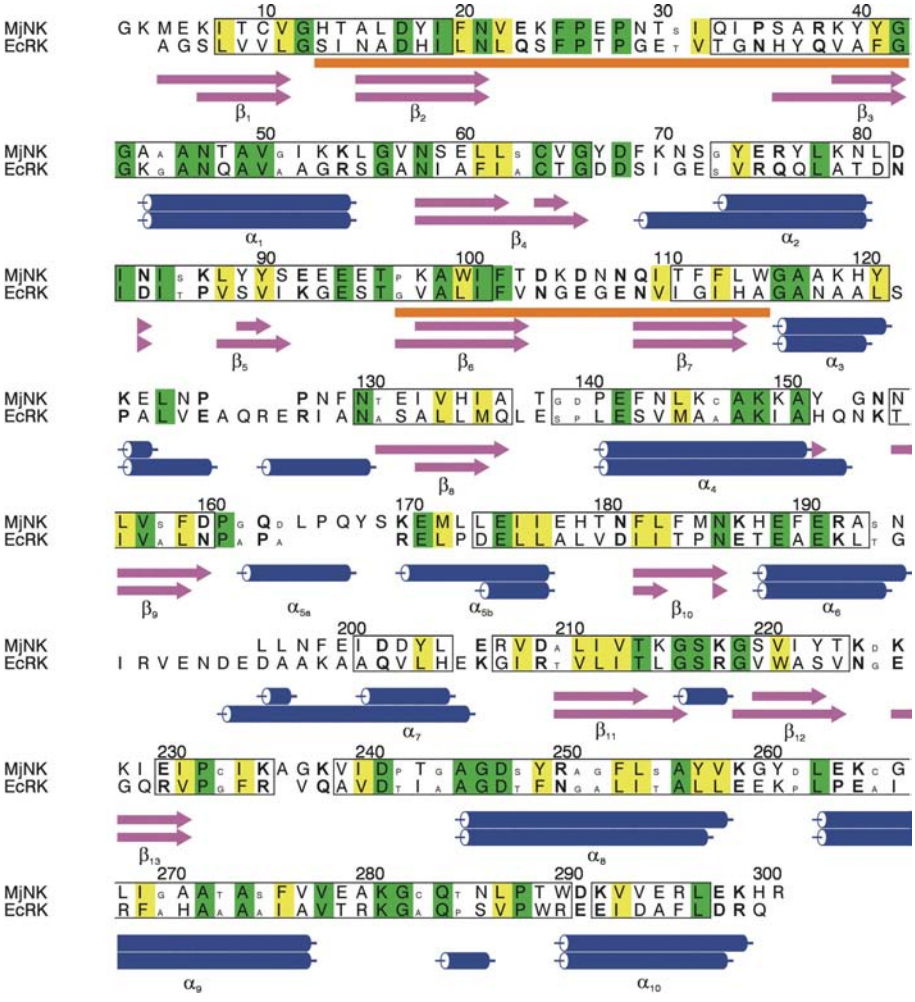

Figure 2
Structure-based sequence alignment of monomer *A* of the MjNK complex with *E. coli* ribokinase (EcRK; PDB code 1rka). Helices and strands are shown as cylinders and arrows, respectively. Residues within boxes represent structurally equivalent regions, conserved residues are coloured green, conserved residues with hydrophobic character are coloured yellow, conserved residues with polar or charged character are shown in bold letters and residues showing a conservation of small size are shown in a smaller font size. The sequences are numbered according to MjNK.

2.2. Differential scanning calorimetry

Differential scanning calorimetry (DSC) was performed on a Microcal MCS calorimeter controlled by the *MCS OBSERVER* program (Microcal). Prior to DSC experiments, the sample buffer was exchanged to 50 m*M* sodium phosphate pH 7.0, 150 m*M* NaCl by dialysis and the ribokinase concentration was adjusted to 1 mg ml^{-1} using an extinction coefficient of 12 660 M^{-1} cm^{-1} (Andersson, personal communication). Both buffer and sample were degassed for 30 min before loading into the calorimeter. To stabilize the baseline, six prescans were carried out before the sample scan. Scans were run at 1 K min^{-1} from 293 to 363 K.

2.3. Molecular-mass determination

The molecular masses of the native enzyme and the selenomethionine

(SeMet) derivative were analyzed on a Quadrupole-orthogonal injection time-of-flight mass spectrometer (QTOF1, Waters/Micromass plc, Manchester, England) equipped with a standard Z-spray source. The enzymes were dialyzed against autoclaved water using a microprep 10 kDa cutoff concentration tube. Prior to injection into the mass spectrometer, the enzyme solutions were diluted in 50% acetonitrile containing 0.2% formic acid to a final protein concentration of 16 pmol μl^{-1}. The samples were sprayed from gold-plated borosilicate capillaries (Protana AS, Odense, Denmark) using a capillary voltage of +1000 V. The mass scale of the instrument was calibrated using caesium iodine clusters over the m/z range 130–2000. The data were analysed using the *MassLynx* v.3.5 software (supplied by the manufacturer), which includes the Maxent deconvolution algorithm.

2.4. Crystallization

The crystallization procedure of the apo enzyme (MjNK-apo) and the enzyme in complex with the ATP analogue AMP-PNP and magnesium has been described elsewhere (Arnfors *et al.*, 2005).

The SeMet derivative was dialyzed against 20 m*M* Tris–HCl pH 8.0 containing 0.02% azide and concentrated to a protein concentration of 15 mg ml^{-1}. The crystallization condition for the SeMet derivative was optimized from the MjNK-apo crystallization condition (10 m*M* PEG 3350, 0.15 *M* magnesium acetate) using the sitting- and hanging-drop vapour-diffusion methods. In order to keep selenium in reduced form, 10 m*M* DTT was added to the protein solution 2 h prior to the crystallization setup. A volume of 2–3 µl mother liquor was added to an equal volume of the protein solution. All crystallization drops contained 0.02% azide. The drops were allowed to equilibrate against 500 µl mother liquor at a temperature of 293 K. Crystals were obtained from a solution containing 15%(*w*/*v*) PEG 3350, 0.15 *M* magnesium acetate and 0.1 M Tris–HCl pH 7.5. Microcrystals appeared after 1 d and after 4 d the crystals had reached their full size of approximately 0.2 × 0.2 × 0.2 mm. The cryoprotectant solution used for data collection had the same composition as the mother liquor with an additional 20% ethylene glycol.

2.5. Data collection

Prior to data collection, the derivative crystal was soaked in the cryoprotectant solution for 1 min and flash-frozen in liquid nitrogen. Data collection was performed at the tunable beamline BW7A at the EMBL Outstation, DESY (Hamburg, Germany) at three different wavelengths. A complete MAD (multiple-wavelength anomalous dispersion) data set was collected around the selenium edge from a single crystal using a MAR CCD detector. At the remote wavelength, both high- and low-resolution data sets were collected. Data were indexed and scaled using the *HKL* programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

2.6. Structure determination and refinement

The complete data sets collected at the peak, inflection and remote wavelengths were used for MAD phasing to 2.2 Å as well as for initial refinement of the SeMet-derivative structure. All data were sorted and reindexed using *XPREP* (Bruker). Searches for selenium positions were performed with *SOLVE* (Terwilliger & Berendzen, 1999) using the remote wavelength data as reference. Phase combination with Hendrickson–Lattman coefficients was performed after each of the following refinement steps. The three selenium sites found were used for density modification with local pattern matching with *RESOLVE* (Terwilliger, 2000). An $F_o - F_c$ electron-density map was calculated using *FFT* (Read & Schierbeek, 1988) from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). Automated model building with *RESOLVE* (Terwilliger, 2003) was performed to obtain an initial starting model. The amino-acid segments from *RESOLVE* were evaluated in *O* (Jones *et al.*, 1991) and manually organized by comparison of the $F_o - F_c$ map and the amino-acid sequence. Simulated annealing in *CNS* (Brünger *et al.*, 1998) was performed using torsion angles and a starting temperature of 5000 K that was decreased in steps of 25 K. Further refinement was performed in alternating steps of refinement in reciprocal space with *REFMAC*5 (Murshudov, 1997) and manual model building and refinement using the program *O* (Jones *et al.*, 1991). The refinement was stopped when R_{cryst} and R_{free} were 24.0 and 27.6%, respectively.

Automated molecular-replacement searches with the native data sets were performed with *AMoRe* within the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994). The refined structure of the SeMet derivative was used as a search model with the 1.7 Å apoenzyme data. The cross-rotation function was calculated in the resolution range 10–4 Å with an integration-sphere radius of 21 Å. The translation function was automatically calculated from the 20 best rotation-search results. The resulting model of the apoenzyme was manually rebuilt in *O* (Jones *et al.*, 1991) and refined in *REFMAC*5 (Murshudov, 1997) during several cycles. When the R_{free} was below 30%, the *REFMAC*5 option TLS refinement was introduced, where one subunit was treated as one TLS group.

To determine phases of the 1.9 Å complex data *via* molecular replacement, the dimer related by crystallographic symmetry of the apo enzyme structure containing amino acids 8–18, 39–100 and 112–295 in each subunit was used as a search model. The automated cross-rotation and translation searches with *AMoRe* were run with a resolution range of 10–4 Å and an integration sphere radius of 32 Å. The resulting model was treated by rigid-body refinement in *REFMAC*5. Each of the three segments of the search model described above was treated as one rigid body. The phases and figure of merit from the rigid-body refinement were used for restrained refinement in *REFMAC*5 of the model from the rigid-body refinement. The model was built in *O* and further refined using *REFMAC*5. When the R_{free} dropped to below 30%, TLS refinement was introduced with TLS groups corresponding to the domains of MjNK. The adenosine and AMP-PNP molecules were incorporated into the model at a later stage of the refinement.

Table 2
Data-collection parameters and statistics for the Se-MAD data.

Values in parentheses are for the lowest resolution shell.

	Peak	Inflection	Remote
Wavelength (Å)	0.9779	0.9784	0.9070
Crystal-to-detector distance (mm)	155	155	155/300
Oscillation range (°)	0.47	0.47	0.47/2.35
Resolution range (Å)	30–2.05 (2.07–2.05)	30–2.05 (2.07–2.05)	30–2.0 (2.02–2.00)
Unique reflections	46720	46785	50653
Measured reflections	541700	661708	901733
Completeness (%)	98.8 (95.9)	99.4 (99.7)	99.7 (99.4)
$R_{\rm merge}$ (%)	5.6 (48.7)	6.3 (50.6)	7.7 (47.5)
$\langle I\rangle/\langle\sigma(I)\rangle$	18.2 (3.2)	19.3 (3.3)	18.9 (2.0)

Table 3
Refinement statistics of the apoenzyme and the MjNK complex.

	MjNK apo form	MjNK–adenosine–AMP-PNP
PDB code	2c4e	2c49
Refinement		
Non-H protein atoms	2366	4739
Non-H substrate atoms	0	100 (2 adenosine, 2 AMP-PNP)
Non-H ion atoms	7 (MO6)	26 (2 MO6, 2 MO5)
Missing residues	*A*1, *A*301–302	*A*1–2, *A*302, *B*1–*B*2, *B*301–302
Water molecules	107	140
Resolution limit (Å)	1.70	1.92
$R_{\rm cryst}$† (%)	23.5	24.5
$R_{\rm free}$‡ (%)	25.3	28.1
Ramachandran plot		
Most favourable regions (%)	93.2	91.6
Allowed regions (%)	6.8	8.4
Disallowed regions (%)	0.0	0.0
R.m.s. standard deviation		
Bond lengths (Å)	0.010	0.016
Bond angles (°)	1.247	1.700
Average *B* factor (Å)	24.0	21.4

† $R_{\rm cryst} = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$. ‡ $R_{\rm free}$ is the cross-validation *R* factor computed for the test set of 5% of the unique reflections.

Water molecules were added either with the program *ARP/wARP* v.5.0 (Lamzin, 1993; Perrakis, 1997) from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) or manually in *O* (Jones *et al.*, 1991). Only water molecules which had a clear density above 1σ in the $2F_{\rm o} - F_{\rm c}$ map and above 3σ in the $F_{\rm o} - F_{\rm c}$ map and a temperature factor below 60 Å^2 were kept. The substrate molecules and magnesium ions were identified from the $F_{\rm o} - F_{\rm c}$ density maps.

The stereochemical quality of the models was checked with the programs *SFCHECK* (Vaguine *et al.*, 1999) and *PROCHECK* (Laskowski *et al.*, 1993) as implemented in *CCP*4 (Collaborative Computational Project, Number 4, 1994). All structural alignments were performed and root-mean-square differences were determined with the program *LSQMAN* (Kleywegt & Jones, 1994). In case residues of one subunit were aligned with residues of another subunit, the 'Brute_force' command was used to find the best assignment of residues to each other. In case residues from a dimer were aligned with residues from another dimer, the corresponding residues were explicitly assigned to each other as shown in Fig. 2 and a first alignment was calculated with the 'Explicit' command. In either case the alignment was improved using the 'Improve' command. Alignments of α/β-domains were restricted to residues 3–12, 42–96 and 116–300 (MjNK) and 4–11, 41–95 and 115–308 (EcRK). Lid alignments were restricted to residues 13–41 and 97–115 (MjNK) and 12–40 and 96–114 (EcRK). The amino-acid sequence alignment shown in Fig. 2 was generated with the programs *STAMP* (Russell & Barton, 1992) and *ALSCRIPT* (Barton, 1993) and is based on secondary-structure assignments obtained from the program *DSSP* (Kabsch & Sander, 1983). All molecular-graphics figures were prepared with the program *PyMOL* (DeLano, 2002).

2.7. Surface calculations and determination of noncovalent interactions

Ion pairs and ionic networks were analyzed using the programs *CONTACT* (Collaborative Computational Project, Number 4, 1994) and *IONSTAT* (Winfried Meining, unpublished program) with a maximum distance of 4 or 6 Å. Accessible surface areas were calculated in *CNS* (Brünger *et al.*, 1998). The water probe radius was 1.4 Å and the accuracy of the numerical integration was set to 0.12. Net charges were calculated by summarizing the total number of positive (Arg, Lys, His) and negative (Asp, Glu) charges, assuming 50% of the histidine residues as charged, which is consistent with the p$K_{\rm a}$ value of this amino acid at neutral pH. Hydrogen bonds were calculated with *HBPLUS* v.3.15 (McDonald & Thornton, 1994) using the default parameters: the maximum distance for D—A was 3.9 Å and for H—A was 2.5 Å and the minimum angles for D—H—A, D—A—AA and H—A—AA were 90°. Ion pairs counted as hydrogen bonds by the program were excluded. Aromatic–aromatic interactions were defined by a distance of less than 7 Å between the aromatic ring centroids (Burley & Petsko, 1985).

3. Results and discussion

3.1. Expression, purification and characterization

Recombinant MjNK was purified and appeared to be homogeneous as judged by SDS–PAGE. Approximately 40 mg of pure ribokinase was produced from 3 g of wet cells. Differential scanning calorimetry experiments revealed an apparent melting temperature ($T_{\rm m}$) of 329 K for *E. coli* ribokinase in 50 m*M* sodium phosphate pH 7.0, 150 m*M* NaCl. The molecular weight was determined to be 33 937 ± 1 Da for the native protein and 34 079 ± 1 Da for the SeMet protein. The difference of 142 ± 1 Da corresponds to the incorporation of three selenomethionines. N- and C-terminal sequencing indicated that MjNK was lacking Met1 for both the native and derivative protein. From these data, it can be concluded that the purified protein only contained three methionines instead of four and thus the SeMet derivative was fully incorporated.

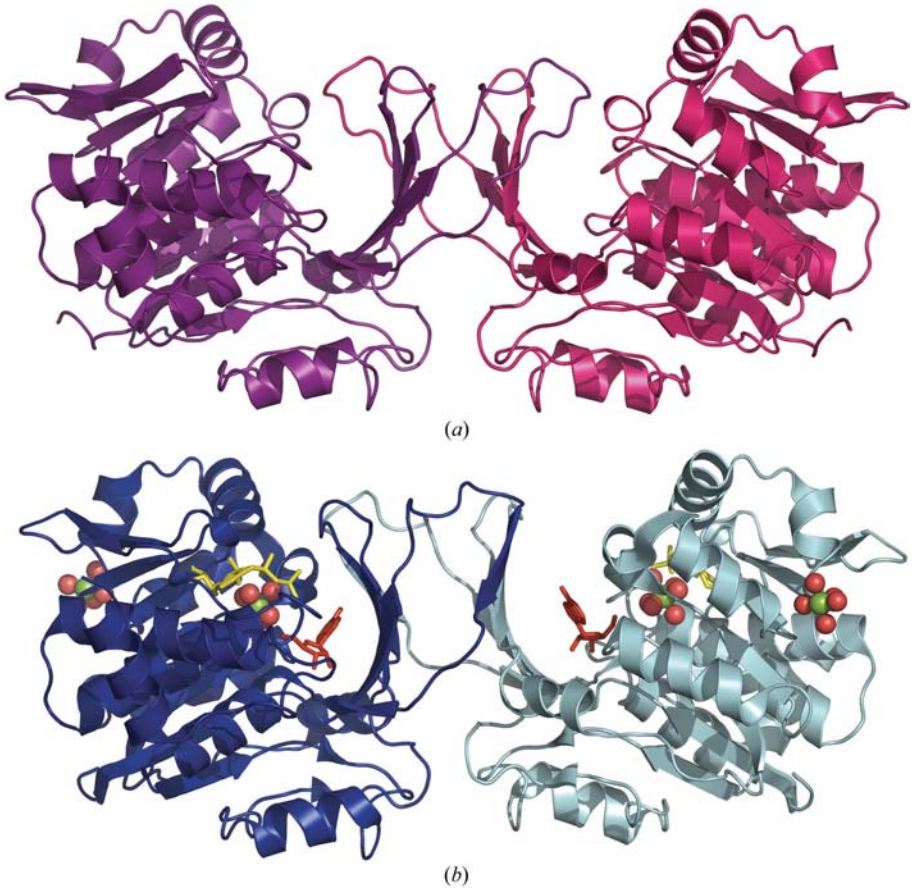

Figure 3
Ribbon representation of MjNK. (*a*) Crystallographic dimer of the apo form (subunit *A*, purple; subunit *A*′, pink) and (*b*) in complex with adenosine (red) and AMP-PNP (yellow). Subunit *A* (dark blue) is in the closed conformation with the lid domain closer to the α/β-domain and contains the strongly bound adenosine, whereas subunit *B* (light blue) is in the open conformation and contains an adenosine that is bound with low occupancy. Mg atoms and surrounding waters are shown as green and red spheres, respectively.

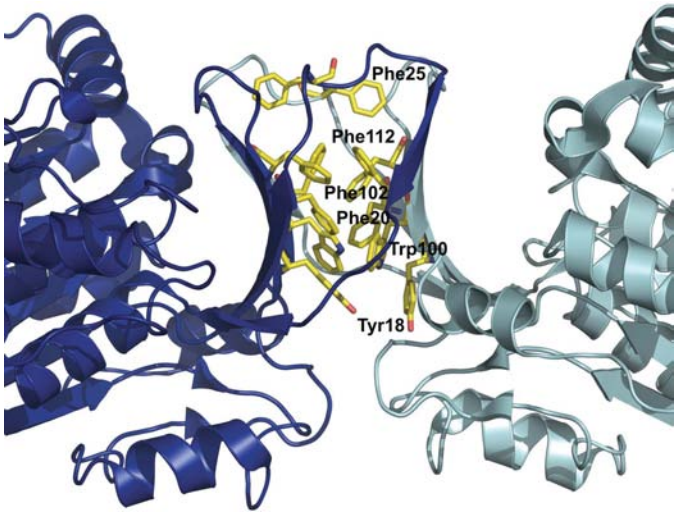

Figure 4
A 12-member aromatic cluster in the dimer interface of nucleoside kinase. The aromatic cluster is formed by interactions between six aromatic residues from each subunit.

3.2. Crystallization and data collection

The crystallization and data collection of the native enzyme have been described elsewhere (Arnfors *et al.*, 2005). The selenomethionine derivative of nucleoside kinase crystallized in space group $P2_12_12$, with unit-cell parameters a = 64.4, b = 146.7, c = 41.1 Å, $\alpha = \beta = \gamma = 90^\circ$. The data-collection statistics of the SeMet data are shown in Table 2. The space group and the unit-cell parameters were identical to those of the apoenzyme. A calculated Matthews coefficient (V_M; Matthews, 1968) of 2.86 Å^3 Da^{-1} for four monomers in the unit cell of the selenomethionine derivative indicated that the asymmetric unit contains one subunit. The corresponding solvent content is 56.4%.

3.3. Structure determination and refinement

Structure-determination attempts with both isomorphous replacement and molecular replacement using apoenzyme data as well as data for the complex were unsuccessful. *E. coli* ribokinase, *Toxoplasma gondii* adenosine kinase and human adenosine kinase were the only three-dimensional structures available when the molecular-replacement trials were performed and were used as search models in molecular replacement. The MjNK structure was eventually determined by the multiple-wavelength anomalous dispersion (MAD) technique using a selenomethionine derivative. The first electron-density map from *SOLVE* (Terwilliger & Berendzen, 1999) showed clear protein density. The electron density was interpretable after density modification with local pattern matching in *RESOLVE* (Terwilliger, 2000). The figure of merit (FOM) improved from 0.34 to 0.60 after density modification. The first initial model contained 75% of the amino acids, of which 22% were alanines or glycines in polyalanine segments. After manual localization and organization of the segments, 85% of all residues could be assigned and built into electron density. Only 3% of the residues that were built into the model had an undefined side chain. The remaining 15% of the amino acids that could not be built in belonged to the N- and C-termini, the lid domain and residues 234–247. Initial refinement with simulated annealing was successful and the model showed a good fit to the density map. After this stage the R_{cryst} and R_{free} were 33.4 and 34.9%, respectively. After several cycles of manual model building in *O* (Jones *et al.*, 1991) and refinement

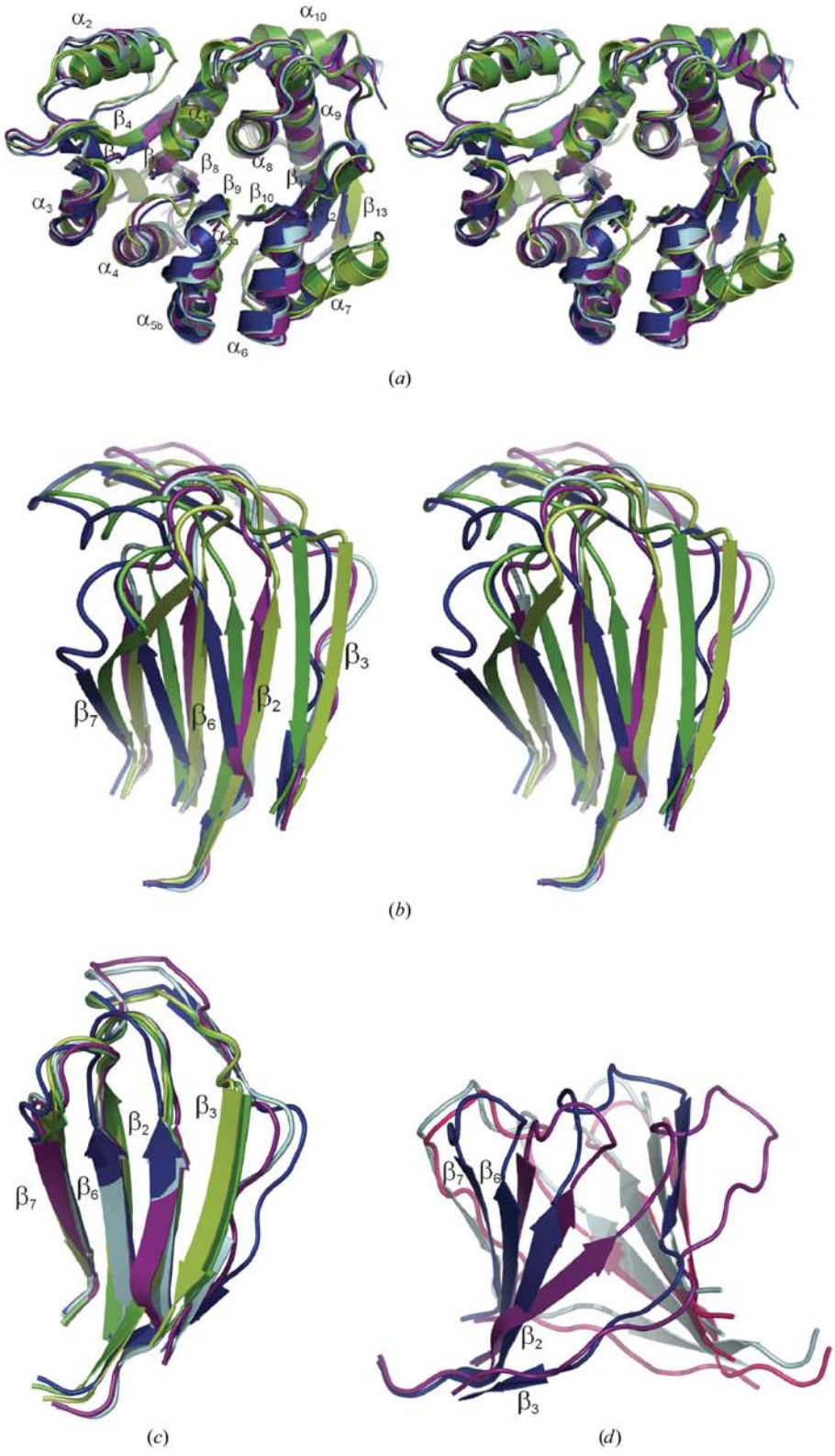

Figure 5
Secondary-structure alignments of MjNK apo form (purple), subunit *A* (dark blue) and *B* (light blue) of the MjNK complex, the EcRK apo form (yellow) and subunit *A* of the EcRK complex (green). In (*a*)–(*c*) subunit *B* of the EcRK complex has been omitted. (*a*) Alignment of α/β-domains (only the α/β-domains are shown). (*b*) Lid domains after alignment of the α/β-domains. (*c*) Lid domains after alignment of lid domains. (*d*) Dimers of lid domains of the MjNK apo form (purple and red) and the complex form (dark blue and light blue) after alignment of α/β-domains of both subunits *A* and *B*. The alignments were prepared according to the procedure described in §2.

with *REFMAC*5 (Murshudov, 1997), $R_{\rm cryst}$ and $R_{\rm free}$ dropped to 24.0 and 27.6%, respectively.

Molecular replacement using the apoenzyme data and the SeMet-derivative structure as a search model yielded one clear solution with an *R* factor of 30.1% and a correlation factor of 77.0%. We have previously described the collection of two data sets of protein–ligand complexes, one of which (MjNK-A) was obtained from crystals that grew in the presence of AMP-PNP and the other of which (MjNK-AF) was collected after soaking MjNK-A crystals with fructose-6-phosphate (F-6-P; Arnfors *et al.*, 2005). F-6-P was chosen as ligand as it can be used as a substrate by many members of the ribokinase family. No density for F-6-P could be identified. Instead, clear density for an adenosine molecule, probably a hydrolysis product of AMP-PNP, was observed in the substrate-binding region of the active site. Thus, the complex crystal forms obtained before and after soaking with F-6-P were identical. In the further refinement only the data obtained from MjNK-AF were processed. A clear solution with an *R* factor of 46.5% and a correlation factor of 50.1% was found for the complex data when the dimer related by crystallographic symmetry of the apoenzyme was used as search model. After rigid-body refinement in *REFMAC*5 (Murshudov, 1997) the $R_{\rm cryst}$ and $R_{\rm free}$ were 42.4 and 43.1%, respectively. Restrained refinement in *REFMAC*5 (Murshudov, 1997) using the phase and FOM from the rigid-body refinement reduced $R_{\rm cryst}$ and $R_{\rm free}$ to 32.5 and 36.6%, respectively. Refinement statistics are presented in Table 3.

3.4. Description of the final model

Two crystal structures of *M. jannaschii* nucleoside kinase were determined: unliganded MjNK (MjNK-apo) and MjNK in complex with adenosine, the non-hydrolyzable ATP-analogue AMP-PNP and magnesium (MjNK complex). Both crystals belong to orthorhombic space groups and in both crystal modifications the enzyme forms a homodimer (subunits *A* and *B*).

Unliganded MjNK crystallized in space group $P2_12_12$. In this form the asymmetric unit contains a single subunit and the dimer axis coincides with the crystallographic *a* axis. The MjNK complex crystals belong to space group $P2_12_12_1$ and the asymmetric unit contains a dimer related by a twofold non-crystallographic axis which is parallel to the crystallographic *b* axis.

The structure of MjNK was determined to 2.2 Å resolution by the multiple-wavelength anomalous dispersion method using a selenomethionine derivative. The apoenzyme model was refined at 1.7 Å and contains residues 2–300 (residues 104–111 have zero occupancy), one magnesium ion linked to six water molecules (MO6) and 107 water molecules. The crystallographic *R* factor is 23.5% and $R_{\rm free}$ is 25.3%. The model of the MjNK complex was refined at 1.9 Å resolution and contains 597 residues (subunit *A*, 3–301; subunit *B*, 3–300), two adenosine molecules (the adenosine in subunit *B* has low occupancy), two AMP-PNP molecules and four magnesium ions, of which two are coordinated to six water molecules (MO6; MO6 in subunit *B* has low occupancy) and the other two to five water molecules (MO5) and 140 water molecules. The crystallographic *R* factor is 24.5% and $R_{\rm free}$ is 28.1%. The unusally high $R_{\rm free}$ of the complex structure might be caused by the conformational variability of the lid domain, in particular of the *B* subunit.

3.5. Overall structures

Nucleoside kinase exists as a homodimer in both crystal structures (Fig. 3). The overall structure of the MjNK subunit can be divided into two parts: a central α/β-fold and a protruding lid domain. The α/β-domain contains residues 1–12, 42–96 and 116–302 and the lid domain contains residues 13–41 and 97–115. Each subunit contains nine α-helices, two 3_{10}-helices and 14 β-strands. The central β-sheet in each subunit contains eight strands with topology $\beta4\uparrow\beta1\uparrow\beta8\uparrow$-$\beta9\uparrow\beta10\uparrow\beta11\uparrow\beta12\downarrow\beta13\uparrow$. The β-sheet is flanked by five α-helices ($\alpha3$, $\alpha4$, $\alpha5$, $\alpha6$, $\alpha7$) on one side and by four α-helices ($\alpha1$, $\alpha2$, $\alpha8$, $\alpha9$) on the opposite side. α-Helix $\alpha10$ is in contact with helix $\alpha9$. With the exception of $\alpha10$, all α-helices are approximately antiparallel to the strands of the central β-sheet. The protruding lid domain is made up by four strands with topology $\beta3\downarrow\beta2\uparrow\beta6\uparrow\beta7\downarrow$ and two loops connecting strand $\beta2$ to $\beta3$ and $\beta6$ to $\beta7$. The lid folds back over the substrate-binding site and thus shields the substrate from the solvent.

3.6. Dimer interface

Gel-filtration chromatography indicates that nucleoside kinase is a dimer in solution (Hansen *et al.*, unpublished results). In the crystal forms, MjNK is an elongated dimer with approximate dimensions of 42 × 47 × 85 Å. The dimeric interface results mainly from hydrophobic interactions between the lid domains of each subunit. The aromatic residues Tyr18, Phe20, Phe25, Trp100, Phe102 and Phe112 from each subunit form a large 12-member cluster of aromatic–aromatic interactions (Fig. 4). There are 23 hydrogen bonds in total and one ion-pair interaction between the two subunits in the enzyme complex. Strand $\beta7$ in one subunit is involved in hydrogen bonding with loop $\beta2\beta3$ such that strands $\beta2A$, $\beta6A$, $\beta7A$ and loop $\beta2B\beta3B$ fold back on subunit *A* and the opposite for subunit *B*. As with EcRK, the MjNK subunit assumes either an open or a closed conformation (Fig. 3). In the apoenzyme, the subunits are in the open conformation. In the MjNK complex, the two subunits show different conformations. Subunit *A* exists in the closed conformation and also contains a tightly bound adenosine. Subunit *B* assumes the open conformation (Figs. 5*a* and 5*b*). Weak density is observed for the adenosine in subunit *B*; however, most atoms were not covered by density at a $\sigma = 1$ level. The nucleoside is thus only partially bound, which might be a consequence of the open conformation of subunit *B*. Upon lid closure, the α/β-domain remains in the same conformation (Fig. 5*a*) and so does the shorter β-sheet protrusion in the lid domain (Fig. 5*c*). However, residues from Phe25 and to the end of the longer β-sheet protrusion ($\beta3$, loop $\beta2\beta3$, $\beta2$) are shifted away from $\beta2$ compared with the open-form conformer and down towards the α/β-domain and the substrate-binding site (Fig. 5*c*). This results in an elongation of strand $\beta3$ with two residues in the open subunit. A superposition of the C^{α} atoms of residues 3–12, 42–96 and 116–300 in the α/β-domain of subunit *A* and subunit *B* in the MjNK complex reveals a root-mean-square deviation of 0.49 Å. A superposition of the C^{α} atoms of residues 98–114 in the lid domain gives a root-mean-square deviation of 0.36 Å. For comparison, a superposition of the whole lid domains (residues 13–41 and 97–115) gives an r.m.s. deviation of 1.35 Å.

3.7. The active site

There are two active sites per MjNK dimer. Each is located along one edge of the central β-sheet and is partially capped by the lid. In the active site of subunit *A* (the closed conformer) density was observed for adenosine, MO6 and AMP-PNP (Figs. 6 and 7). No density was observed for the γ-phosphate of AMP-PNP in subunit *A*. In the active site of subunit *B* (the open conformer) density was observed for AMP-PNP, but was only partially visible for adenosine and MO6. The nucleoside and magnesium are probably bound with lower occupancy in the open-form subunit, thus leading to the weak density.

The density for adenosine was strong and clear in subunit *A*, but weak in subunit *B*, which indicates that adenosine is bound with lower occupancy in the open conformer. The ribose ring of adenosine is in the energetically favourable C2′- or C3′-*endo* conformation. The lid domain covers the adenosine molecule. The O atoms of the ribose ring of the nucleoside in subunit *A* forms hydrogen bonds with side-chain atoms of Asp17, Asn47, Gln163 and Asp247 and the main-chain N atom of Gly43. The O5′ oxygen also hydrogen bonds with water O2 of the magnesium ion. In the apoenzyme, a water molecule is positioned at the same place as the ribityl O5′ oxygen. The adenine base makes hydrogen bonds from N atom N1 to Asp164, from N3 *via* water Z67 to the side chain of Thr138 and

the main-chain N of Ala44, from N6 to the side chain of Gln33 in subunit *B* and from N7 to the side chain of Thr111. The pyrimidine ring of adenine forms aromatic π–π stacking interactions with the benzene ring of Phe113.

The ATP analogue AMP-PNP is located in a shallow groove in the α/β-domain next to adenosine and the magnesium ion (MO6) but that is not capped by the lid. There are only few interactions between AMP-PNP and the enzyme. No hydrogen bonding occurs between the adenine base and MjNK. The O2′ atom of the ribose ring forms hydrogen bonds *via* water Z53 to Ser220 and *via* water Z62 to Ile234 N and Ala271 O. O3′ makes hydrogen bonds with the main-chain O of Gly219. One of the O atoms of the α-phosphate moiety forms hydrogen bonds with the side chain of Thr214 and another O atom of the β-phosphate forms hydrogen bonds with Asn186 OD1 and water O6 of the MO6. In the apo enzyme, the position of the O2′ atom of AMP-PNP is occupied by a water molecule.

3.8. Magnesium-binding sites

There are two magnesium-binding sites per subunit: one magnesium ion (MO6) is coordinated octahedrally by six water molecules and is located between adenosine and AMP-PNP in the active site and one magnesium ion (MO5) is coordinated by five water molecules and is positioned in the interface between the dimers in the crystal. Clear density was observed for MO5 in both subunits and MO6 in subunit *A*; however, the density was weak for MO6 in subunit *B*. The water molecules of MO6 in the active site of subunit *A* form hydrogen bonds with the side chains of Asp160, Glu189 and Asp247, one O atom of the β-phosphate in AMP-PNP and the O5′ atom of the substrate adenosine. The density in the active site of subunit *B* suggests that one water molecule in MO6 might also form hydrogen bonds to a γ-phosphoryl O atom. This fits well to previously described mechanisms for kinases, where the electrophilic magnesium ion located in the active site simultaneously coordinates both the β- and γ-phosphates and thereby decreases their negative charge, which in turn enhances the leaving of ADP.

3.9. Conserved residues and reaction mechanism

The location of the substrate-binding site as well as certain conserved residues and regions are important common features of members of the ribokinase superfamily. The position of adenosine, ATP-analogue and MO6 in MjNK are the same as in human adenosine kinase. There is a slight tilt of the adenine base of the adenosine molecule in MjNK compared with that in hAK, which can be explained by the π–π stacking interaction of the respective phenylalanine residues that are positioned on opposite faces with respect to the adenine base. In MjNK, this phenylalanine is represented by Phe113 and in hAK by Phe170. They are not at corresponding positions (Ala136 in hAK is at the same position as Phe113 in MjNK) in the structures and belong to different segments of the α/β-domain. In addition, when comparing the active sites, the ribose ring of adenosine in MjNK is at the same position as ribose in ribokinase.

All conserved residues present in members of the ribokinase superfamily are also found in MjNK. The aspartate and asparagine involved in hydrogen bonding with the 2′- and 3′-hydroxyl groups in the ribose ring are represented in MjNK by Asp17 and Asn47. One of the fingerprint regions of the ribokinase family is the glycine-glycine dipeptide sequence (Bork *et al.*, 1993; Wu *et al.*, 1991). In MjNK, this conserved GG region is represented by residues 42 and 43. Gly42 is the last residue in strand β3 and the GG dipeptide is part of the hinge region between the lid domain and the α/β-domain. It is believed that the conformational shift of the GG motif upon substrate binding plays an important role in forming the closed conformation of the enzymes (Schumacher *et al.*, 2000). The N*XX*E sequence, the side chains of the asparagine and glutamate of which are involved in ATP binding, corresponds to residues NLHE 186–189 in MjNK. Asn186 forms hydrogen bonds with one O atom of the β-phosphate group and Glu189 hydrogen bonds with two water molecules of the magnesium ion MO6 in subunit *A*. The most highly conserved region present among all members of the ribokinase superfamily is the anion hole. It is made up by the amide N atoms of the amino-acid sequence GAGD (AAGD in EcRK; Sigrell *et al.*, 1998, 1999) and is represented by residues 244–247 in MjNK. The anion hole assists in neutralizing accumulated negative charge during the phosphate-group transfer. The aspartate residue in the anion hole forms hydrogen bonds with the phosphate acceptor and is important for catalysis. In *Bacillus subtilis* 4-methyl-5-β-hydroxyethylthiazole kinase (BsThiK; Campobasso *et al.*, 2000) and StHMPPK (Cheng *et al.*, 2002), the aspartate is replaced by cysteine. Interestingly, the enzymatic activity increases tenfold when the cysteine is mutated to an aspartate in BsThiK (Campobasso *et al.*, 2000). The nonconserved arginine or lysine that stabilizes the transition state during phosphate transfer and in some members forms hydrogen bonds with ribose is most likely to be Arg250 in MjNK. The side chain of Arg250 is positioned below the magnesium ion when adenosine is above to the left and the phosphate part of AMP-PNP above to the right with respect to the magnesium ion (like a Y formation). Moreover, the side-chain N atoms of arginine are located next to and in line with the main-chain N atoms of residues 244–246 in the anion hole, which leads to an extension of the anion hole and an increase in the positive charges that stabilize the phosphate transfer. The distance between Arg250 NZ and adenosine O5′ is 8 Å in subunit *A*. In ribokinase, the positively charged residue is represented by the positively charged Lys43 and in hAK it is assumed to be Arg132.

The reaction mechanism of MjNK is believed to be similar to the proposed reaction mechanism of EcRK (Sigrell *et al.*, 1998). The first step is deprotonation of the 5′-hydroxyl group of the ribose ring in the substrate by Asp247. Subsequently, the negatively charged 5′-hydroxyl group makes a nucleophilic attack on the γ-phosphate of ATP through a transition state where the γ-phosphate binds pentacovalently to the five O atoms. The magnesium ion is expected to be involved in

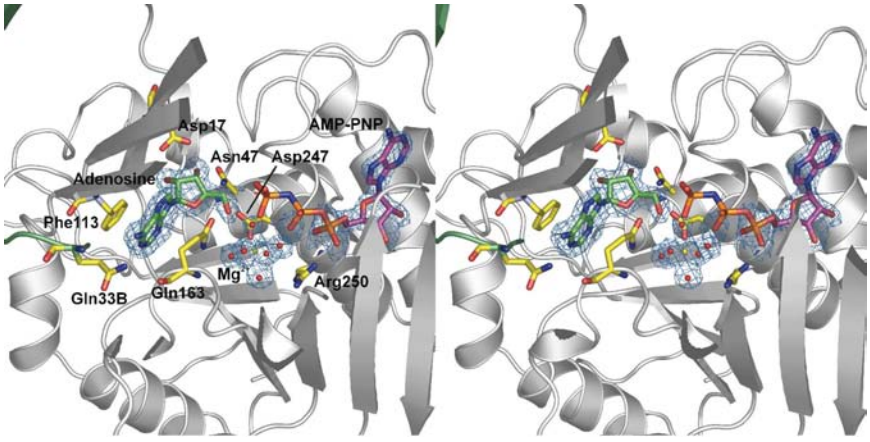

Figure 6
The active site in subunit *A* of nucleoside kinase. The $2F_o - F_c$ electron density is contoured at 1σ around adenosine, AMP-PNP and the magnesium ion. Subunit *A* backbone, grey; subunit *B* backbone, pale green; residues and substrates according to figure. For clarity, not all residues involved in substrate binding are shown.

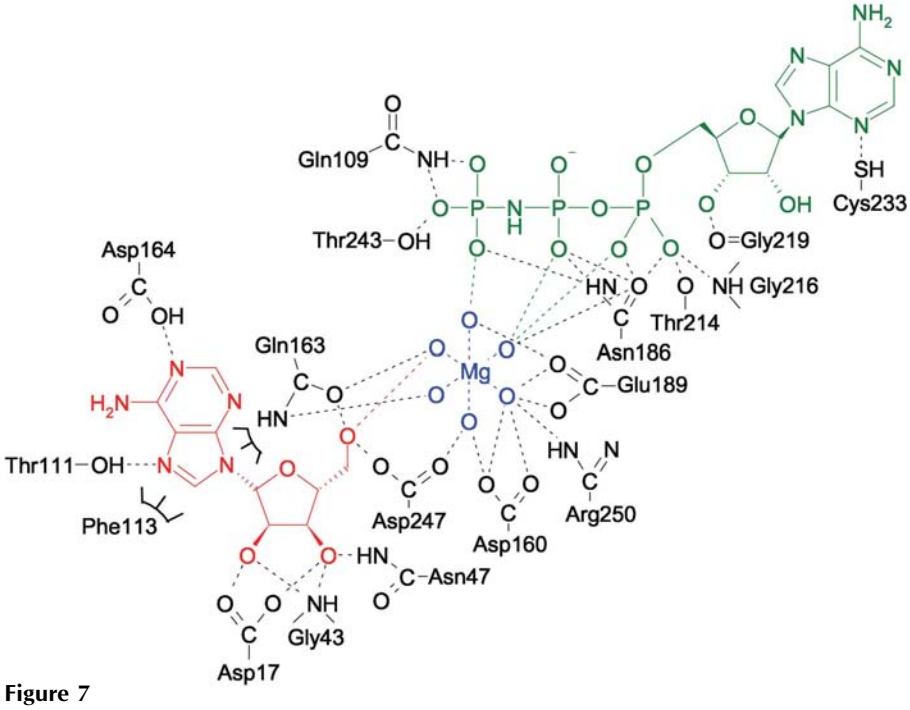

Figure 7
Schematic overview of contacts between adenosine (red), AMP-PNP (green), the solvated magnesium ion (blue) and the protein environment (black). Dashed lines indicate atomic distances between non-C atoms of less than 3.5 Å. The imidazole ring of adenosine is sandwiched between the aromatic ring of Phe113 and the amid group of Gln163.

orienting the γ-phosphate as well as in stabilizing the phosphate transfer. The transition state is stabilized by the main-chain N atoms of the anion hole and the positively charged Arg250. Finally, the monophosphorylated nucleotide and ADP are formed.

3.10. Structural comparisons and correlation with determinants of thermostability

A search for proteins which are structurally similar to subunit *A* of MjNK with the program *DALI* (Holm & Sander, 1993) revealed a *Z* score higher than 12 for 13 members of the ribokinase superfamily. Seven of these proteins, *E. coli* ribokinase (EcRKl PDB code 1rkd), *Thermatoga maritima* ribokinase (TmRK; PDB code 1vm7), *Salmonella enterica* serovar *typhimurium* aminoimidazole riboside kinase (PDB code 1tz6), *Thermus thermophilus* 2-keto-3-deoxygluconate kinase (PDB code 1v19), human adenosine kinase (hAK; PDB code 1bx4), *Enterococcus faecalis* tagatose-6-phosphate kinase (EfTPK; PDB code 2awd) and *Thermotoga maritima* phosphofructokinase (PDB code 1o14), score higher than 24. The sequence identities of these proteins are in the range of 18–22%. The closest relatives are EcRK (*Z* score 36) and TmRK (*Z* score 35). All these structural homologues belong to the ribokinase family of carbohydrate kinases.

A sequence alignment of MjNK and EcRK, which is based on a structural alignment, is shown in Fig. 2. MjNK and EcRK are dimers and contain the α/β-domain as well as a protruding lid domain forming a β-clasp. The overall fold and the secondary-structure elements of MjNK and EcRK are highly similar (Figs. 5*a* and 5c). Structural differences between these two enzymes include variations of the extent and orientation of secondary-structure elements and the orientation of the lid domain with respect to the α/β-domain. A structural alignment of the α/β-domain of subunit *A* of the MjNK complex with the α/β-domain of the EcRK complex results in an r.m.s.d. of 1.33 Å (207 C^{α} atoms). A comparison of the corresponding lid domains results in an r.m.s.d. of 1.26 Å (45 C^{α} atoms). Whereas the conformations of the α/β-domain and the lid domain of the MjNK complex are similar to their respective counterparts in the EcRK complex, the relative orientation of the lid domain against the α/β-domain is quite different (Fig. 5*b*). The differences of the lid orientation between the apo form and the complex form are much more pronounced in MjNK than in EcRK (Fig. 5*b*). The aligned lid domains of EcRK assume an intermediate orientation between the two extrema of the closed and opened form of the MjNK complex. The transformation between the lid domains of the opened and closed form of the MjNK complex can be described by a rotation by 26° followed by a translation by 32 Å, whereas the transformation of the lid of the *B* subunit

Table 4
Characteristics of nucleoside kinase and *E. coli* ribokinase.

	Nucleoside kinase	*E. coli* ribokinase
PDB code	2c49	1rk2
Apparent melting temperature (K)	363	329
Optimal growth temperature of organism (K)	358	310
No. of amino-acid residues per subunit	302	309
Hydrophobic/polar/charged residues† (%)	39.1/33.4/27.5	48.9/29.8/21.3
R.m.s.d. of 263 C^{α} atoms (comparison of subunit *A*) (Å)	—	1.48
Secondary-structure contents (%)	58.9	61.2
α-Helix/β-sheet/3_{10}-helix content (%)	29.8/23.5/5.6	35.0/23.6/2.6
Accessible surface area of dimer (Å^2)	25,639	24,704
Buried intersubunit surface (% of subunit)	4.9	4.3
Distribution of hydrophobic/polar/charged residues at accessible surface‡ (%)	17.7/30.3/52.0	25.8/30.9/43.3
Distribution of hydrophobic/polar/charged residues at interface‡ (%)	57.5/34.2/8.3	60.4/27.8/11.8
No. of ion pairs‡	29/69	24/43
No. of ion pairs per residue‡	0.048/0.11	0.039/0.070
Charged residues forming ion pairs‡ (%)	38/60	36/61
No. of residues forming two ion pairs‡	7/30	3/10
No. of residues forming three ion pairs‡	0/4	0/3
Aromatic residues (%)	13.2	5.5
Phe/Tyr/Trp/His content (%)	5.0/5.3/1.0/2.0	2.6/0.3/0.6/1.9
Aromatic residues involved in aromatic–aromatic interactions§ (%)	87.5	47.1
Phe/Tyr/Trp/His involved in aromatic–aromatic interactions§ (%)	93.3/87.5/100.0/66.7	50.0/100.0/0.0/50.0

† Hydrophobic residues are A, V, L, I, W, F, P, M; polar residues are G, S, T, Y, N, Q, C; charged residues are R, K, H, D, E. ‡ 4.0/6.0 Å cutoff radius. § 7.0 Å distance between aromatic ring centres.

to the lid of the *A* subunit of the EcRK complex corresponds to a rotation by only 13° followed by a translation by 13 Å. Strand β3 is shorter in the closed conformation of subunit *A* (dark blue in Fig. 5*c*) and the loop connecting strand β2 with strand β3 is shifted away from β2 and down towards the active site. The lid domain of EcRK contains four β-strands ($\beta3\beta2\beta6\beta7$) and two loops connecting β2 with β3 and β6 with β7. In MjNK, strand β3 is much shorter than in EcRK and a four-stranded β-sheet is only observed in the hinge region of the lid domain. Moreover, strand β3 varies in length depending on whether the lid domain is open or closed. The β-strands form a β-sheet that together with the β-sheet from the other subunit make up the dimeric interface (Fig. 5*d*). Instead of the longer β-strand, MjNK possesses an ordered loop that forms dimer interactions with β7 of the other subunit (Figs. 2 and 5*d*). In the MjNK complex part of the loop connecting β2 and β3 participates in the β-sheet of the other subunit, whereas in the apo form this loop does not contribute to the adjacent β-sheet. It seems that the asymmetrical lid arrangement in the MjNK complex gains additional stability owing to the extension of the β-sheets by a fifth β-strand. As was the case in the MjNK dimer, the dimeric interface of EcRK is stabilized by interactions between a set of aromatic residues. This group of residues comprises a six-membered aromatic cluster (His17, Tyr36 and His113 from each subunit) and one aromatic pair (Phe101 from each subunit).

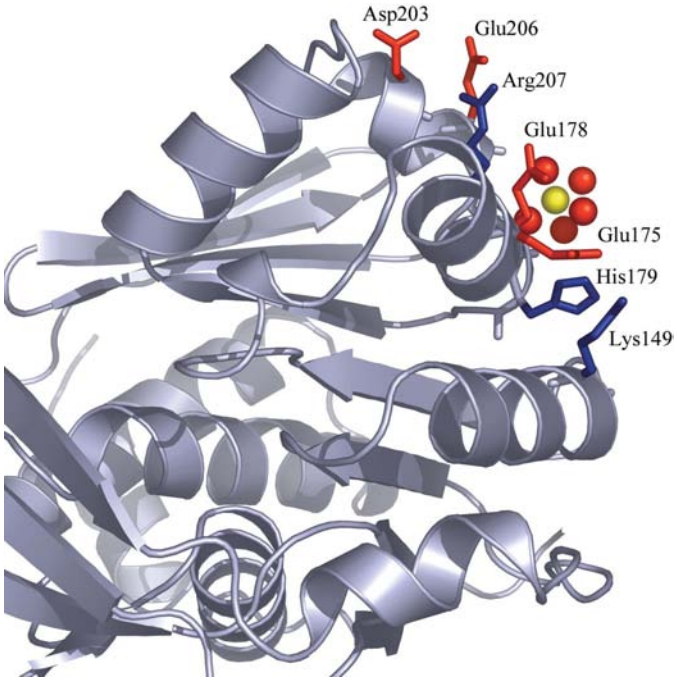

Figure 8
The seven-member ionic network in subunit *A* of the nucleoside kinase binary complex. The network is close to the magnesium ion (yellow sphere) coordinated by five water molecules (red spheres) and is located at the interface between crystallographically related dimers.

Many hyperthermostable proteins have shorter loop regions than their mesophilic counterparts. In MjNK, helix α6 is prolonged, but the loop connecting it to α7 is much shorter than the corresponding loop in EcRK. Furthermore, the loop connecting β9 and α5 in EcRK is stabilized by a small α-helix (α5a) in MjNK (Fig. 5*a*).

Detailed comparisons of members of the ribokinase superfamily have been reported by Campobasso *et al.* (2000) and Cheng *et al.* (2002).

Since MjNK and EcRK show a similar overall fold and dimerize in the same way, it is worthwhile to compare them regarding thermostability. The apparent melting temperature (T_m) is 363 K for MjNK at pH 7.0 (Hansen *et al.*, unpublished results) and 329 K for EcRK at pH 7.0 (this study). We will compare structural determinants between EcRK and hyperthermostable MjNK. The enzyme parts of the MjNK complex and the EcRK dimer (subunits *A* and *B* in 1rk2) have been compared and data from the comparison can be found in Table 4.

3.11. Accessible surface area

Compared with their mesophilic counterparts, proteins from hyperthermophiles show an increase in the accessible surface area contributed by polar and charged residues, but a

decrease in the accessible surface area contributed by hydrophobic residues (Karshikoff & Ladenstein, 2001; Spassov *et al.*, 1995; Yip *et al.*, 1995, 1998). Compared with EcRK, MjNK has an increased charged accessible surface and a decreased hydrophobic accessible surface (Table 4). The accessible surfaces contributed by polar residues for the two enzymes are similar.

The dimer interface of MjNK buries 1255 Å^2 of the approximately 14 075 Å^2 accessible surface of the subunit, leading to a solvent-accessible surface area of 25 639 Å^2 for the dimer. For comparison, the dimer of EcRK buries 1062 Å^2 of the approximately 13 414 Å^2 accessible surface of the subunit, giving a solvent-accessible area of 24 704 Å^2 for the dimer. In other words, 4.9% of the subunit of MjNK was buried in the dimer interface compared with 4.3% for the subunit of EcRK. The distribution of hydrophobic, polar and charged residues at the interface of MjNK and EcRK is shown in Table 4.

3.12. Charged residues and ionic interactions

MjNK has an increase in the number of polar and charged residues and a lower fraction of hydrophobic residues per dimer compared with the EcRK dimer. The numbers of hydrophobic, polar and charged residues for the MjNK dimer are 236 (39.1%), 202 (33.4%) and 166 (27.5%). The respective numbers for the EcRK dimer are 302 (48.9%), 184 (29.8%) and 132 (21.3%). A cutoff distance of 6 Å revealed the most obvious differences between the ionic interactions in MjNK and EcRK, although differences could also be observed with cutoffs of 4 and 5 Å.

The MjNK dimer contains 18 two-member, ten three-member, three four-member, three five-member and one seven-member ionic networks. For comparison, the EcRK dimer contains 23 two-member, two three-member, two four-member and two five-member networks, *i.e.* a much smaller number of multiple-member networks. The most remarkable difference between the two enzymes is a four-member intersubunit network in MjNK, while there are no intersubunit networks observed in EcRK. This network consists of Lys39*A*, Asp68*B*, Glu95*B* and Lys98*B* and is a direct effect of the different subunit conformations. Upon lid closure, Lys39 takes on another position and conformation which enable it to contribute to an intersubunit network. The three residues belonging to the same subunit in the intersubunit network form a corresponding three-member network comprising Asp68*A*, Glu95*A* and Lys98*A* in subunit *A*.

An additional difference in ionic interactions between the subunits is observed to arise from different side-chain conformations of Glu93. In subunit *A*, Glu94*A*, Lys119*A*, His120*A* and Glu123*A* make up a four-member network, which is extended to a five-member network by Glu93*B* in subunit *B*.

Another interesting interaction is the seven-member network consisting of Lys149*A*, Glu175*A*, His179*A*, Glu178*A*, Arg207*A*, Asp203*A* and Glu206*A*, which is located at the surface near the magnesium ion MO5 and the crystallographically related dimer (Fig. 8). As a consequence of smaller conformational changes of the side chains of His179*B* and Glu178*B*, the equivalent residues in subunit *B* are divided into one three-member network comprising Lys149*B*, Glu175*B* and His179*B* and one four-member network consisting of Glu178*B*, Arg207*B*, Asp203*B* and Glu206*B*.

A five-member network located around MO6 in the active site is present in both subunits and includes His135, Asp160, Asp247, Arg250 and Glu189. Three of these residues, Asp160, Asp247 and Glu189, are involved in hydrogen bonding to the MO6 waters and the highly conserved and catalytically important Asp247 is also bound to the substrate. In subunit *A* of EcRK there is a corresponding four-member network and in subunit *B* a three-member network. The conserved Asp255 (corresponding to Asp247 in MjNK), Lys43 and Glu143 contribute, as does Glu190 (corresponding to Glu189 in MjNK) in subunit *A*. It is suggested that Lys43 in EcRK stabilizes the transition state during phosphorylation of the substrate. The ionic network in MjNK will, however, possibly contribute to the stabilization of the magnesium ion binding. The two five-member networks in EcRK originate from the interaction between Glu267, Lys269, Arg298, Glu273 and Asp302 in both subunits.

It has been observed that most hyperthermostable enzymes not only contain a high content of ion-pair interactions, but also a minimization of repulsive and unfavourable contacts (Spassov *et al.*, 1995). At a cutoff of 4 Å the number of contacts between equally charged residues is seven in MjNK and four in EcRK. For comparison, a cutoff of 3.5 Å results in two unfavourable contacts in both enzymes.

Many hyperthermostable proteins are also characterized by an electrostatic optimization of the charged residues and a minimization of excess charges (Spassov *et al.*, 1995), which leads to a net charge close to zero. Nucleoside kinase has a net charge of zero, compared with −9 for ribokinase.

3.13. Aromatic interactions

Aromatic interactions are an important contribution to the structural stability of proteins (Burley & Petsko, 1985), but are not as frequently observed in hyperthermostable enzymes as, for example, an increased number of ion-pair interactions. MjNK has a much higher number of aromatic residues than EcRK. The number of aromatic residues per subunit for MjNK and EcRK, respectively, are 15 and eight Phe, 16 and one Tyr, three and two Trp and six and six His. In addition, the positively charged residues Lys, Arg and His can form cation–π interactions with aromatic ring systems and have been proposed to be an important factor in protein stabilization (Burley & Petsko, 1985; Dougherty, 1996). Besides the six-member aromatic cluster (His17, Tyr36 and His113 from each subunit) and the aromatic pair (Phe101 from each subunit) in the dimer interface of EcRK, there are only two additional aromatic interactions observed in EcRK and these involve the four-member cluster of Phe243, Phe257, His279 and Phe304 in both subunits. MjNK, on the other hand, contains many aromatic interactions and especially aromatic clusters. As has

already been mentioned, there is a 12-member aromatic cluster in the dimer interface resulting from Tyr18, Phe20, Phe25, Trp100, Phe102 and Phe112 from both subunits (Fig. 4). In addition, each subunit comprises two six-member aromatic clusters (His13, Phe113, Thr115, Tyr121, Phe129, Phe143 and His135, Phe182, Phe184, Tyr249, Phe253, Tyr257), one five-member aromatic cluster (Tyr41, Phe69, Tyr74, Tyr77, Tyr90), one four-member aromatic cluster (Phe190, Phe199, Tyr204, Tyr223) and four aromatic pairs (Tyr89 and His120, Tyr152 and His179, Phe159 and Tyr168, Phe261 and Trp290). Out of 40 aromatic residues, there are only five that are not involved in aromatic interactions. These are His188, His300, Phe276, Tyr40 and Tyr67. However, cation–π interactions occur between Tyr40 and Arg38 and Phe276 and Arg296. The vast number of aromatic interactions certainly plays an important role in the contribution to the thermal stability of MjNK.

We thank the EMBL Outstation at DESY (Hamburg, Germany) for provision of synchrotron-radiation facilities. LA thanks Alexander Popov at the EMBL Outstation at DESY (Hamburg, Germany) for outstanding help during MAD data collection. Many thanks to Professor Sherry Mowbray at Swedish University of Agricultural Sciences and Dr Evalena Andersson at Uppsala University for kindly providing the ribokinase plasmid as well as the expression and purification protocols.

References

Andersson, C. E. & Mowbray, S. L. (2002). *J. Mol. Biol.* **315**, 409–419.

Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J., Chothia, C. & Murzin, A. G. (2004). *Nucleic Acids Res.* **32**, D226–D229.

Arnfors, L., Hansen, T., Meining, W., Schönheit, P. & Ladenstein, R. (2005). *Acta Cryst.* F**61**, 591–594.

Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.

Bork, P., Sander, C. & Valencia, A. (1993). *Protein Sci.* **2**, 31–40.

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., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Burley, S. K. & Petsko, G. A. (1985). *Science*, **229**, 23–28.

Campobasso, N., Mathews, I. I., Begley, T. P. & Ealick, S. E. (2000). *Biochemistry*, **39**, 7868–7877.

Cheng, G., Bennett, E. M., Begley, T. P. & Ealick, S. E. (2002). *Structure*, **10**, 225–235.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

DeLano, W. L. (2002). *The PyMOL Molecular Visualization System.* http://www.pymol.org.

Dougherty, D. A. (1996). *Science*, **271**, 163–168.

Hansen, T., Arnfors, L., Ladenstein, R. & Schönheit, P. (2006). In the press.

Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.

Ito, S., Fushinobu, S., Jeong, J. J., Yoshioka, I., Koga, S., Shoun, H. & Wakagi, T. (2003). *J. Mol. Biol.* **331**, 871–883.

Ito, S., Fushinobu, S., Yoshioka, I., Koga, S., Matsuzawa, H. & Wakagi, T. (2001). *Structure*, **9**, 205–214.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Karshikoff, A. & Ladenstein, R. (2001). *Trends Biochem. Sci.* **26**, 550–556.

Kleywegt, G. J. & Jones, T. A. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 9–14.

Lamzin, V. S. (1993). *Acta Cryst.* D**49**, 129–147.

Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.

Li, M. H., Kwok, F., Chang, W. R., Lau, C. K., Zhang, J. P., Lo, S. C., Jiang, T. & Liang, D.-C. (2002). *J. Biol. Chem.* **277**, 46385–46390.

McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.

Mathews, I. I., Erion, M. D. & Ealick, S. E. (1998). *Biochemistry*, **37**, 15607–15620.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Murshudov, G. N. (1997). *Acta Cryst.* D**53**, 240–255.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.

Ohshima, N., Inagaki, E., Yasuike, K., Takio, K. & Tahirov, T. H. (2004). *J. Mol. Biol.* **340**, 477–489.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perrakis, A. (1997). *Acta Cryst.* D**53**, 448–455.

Read, R. J. & Schierbeek, A. J. (1988). *J. Appl. Cryst.* **21**, 490–495.

Russell, R. B. & Barton, G. J. (1992). *Proteins*, **14**, 309–323.

Safo, M. K., Musayev, F. N., Hunt, S., di Salvo, M. L., Scarsdale, N. & Schirch, V. (2004). *J. Bacteriol.* **186**, 8074–8082.

Schumacher, M. A., Scott, D. M., Mathews, I. I., Ealick, S. E., Roos, D. S., Ullman, B. & Brennan, R. G. (2000). *J. Mol. Biol.* **298**, 875–893.

Sigrell, J. A., Cameron, A. D., Jones, T. A. & Mowbray, S. L. (1998). *Structure*, **6**, 183–193.

Sigrell, J. A., Cameron, A. D. & Mowbray, S. L. (1999). *J. Mol. Biol.* **290**, 1009–1018.

Spassov, V. Z., Karshikoff, A. D. & Ladenstein, R. (1995). *Protein Sci.* **4**, 1516–1527.

Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Tsuge, H., Sakuraba, H., Kobe, T., Kujime, A., Katunuma, N. & Ohshima, T. (2002). *Protein Sci.* **11**, 2456–2463.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.

Wu, L.-F., Reizer, A., Reizer, J., Cai, B., Tomich, J. M. & Saier, M. H. Jr (1991). *J. Bacteriol.* **173**, 3117–3127.

Yip, K. S., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R. & Consalvi, V. (1995). *Structure*, **3**, 1147–1158.

Yip, K. S., Britton, K. L., Stillman, T. J., Lebbink, J., de Vos, W. M., Robb, F. T., Vetriani, C., Maeder, D. & Rice, D. W. (1998). *Eur. J. Biochem.* **255**, 336–346.

Zhang, R. G., Grembecka, J., Vinokour, E., Collart, F., Dementieva, I., Minor, W. & Joachimiak, A. (2002). *J. Struct. Biol.* **139**, 161–170.

Zhang, Y., Dougherty, M., Downs, D. M. & Ealick, S. E. (2004). *Structure*, **12**, 1809–1821.