Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Free and ATP-bound structures of Ap_4A hydrolase from *Aquifex aeolicus* V5

Jeyaraman Jeyakanthan,[a] Shankar Prasad Kanaujia,[b] Yuya Nishida,[c] Noriko Nakagawa,[d,e] Surendran Praveen,[b] Akeo Shinkai,[d] Seiki Kuramitsu,[d,e] Shigeyuki Yokoyama[f,g] and Kanagaraj Sekar[b]*

[a]Life Science Group, National Synchrotron Radiation Research Center, 101 Hsin-Ann Road, Hsinchu Science Park, Hsinchu 30076, Taiwan, [b]Bioinformatics Centre (Centre of Excellence in Structural Biology and Bio-computing), Indian Institute of Science, Bangalore 560 012, India, [c]Graduate School of Frontier Biosciences, Osaka University, Suita, Osaka 560-0043, Japan, [d]RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, [e]Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan, [f]RIKEN Systems and Structural Biology Center, Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, and [g]Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Correspondence e-mail: sekar@physics.iisc.ernet.in

Received 23 July 2009
Accepted 7 November 2009

PDB References: asymmetric diadenosine tetraphosphate (Ap_4A) hydrolase, 3i7u; ATP-bound, 3i7v.

Asymmetric diadenosine tetraphosphate (Ap_4A) hydrolases degrade the metabolite Ap_4A back into ATP and AMP. The three-dimensional crystal structure of Ap_4A hydrolase (16 kDa) from *Aquifex aeolicus* has been determined in free and ATP-bound forms at 1.8 and 1.95 Å resolution, respectively. The overall three-dimensional crystal structure of the enzyme shows an $\alpha\beta\alpha$-sandwich architecture with a characteristic loop adjacent to the catalytic site of the protein molecule. The ATP molecule is bound in the primary active site and the adenine moiety of the nucleotide binds in a ring-stacking arrangement equivalent to that observed in the X-ray structure of Ap_4A hydrolase from *Caenorhabditis elegans*. Binding of ATP in the active site induces local conformational changes which may have important implications in the mechanism of substrate recognition in this class of enzymes. Furthermore, two invariant water molecules have been identified and their possible structural and/or functional roles are discussed. In addition, modelling of the substrate molecule at the primary active site of the enzyme suggests a possible path for entry and/or exit of the substrate and/or product molecule.

1. Introduction

Diadenosine 5′,5‴-P^1,P^4-tetraphosphate (Ap_4A), which was first discovered by Zamecnik *et al.* (1966), is synthesized by many enzymes (Brevet *et al.*, 1982; Goerlich *et al.*, 1982; Hilderman & Ortwerth, 1987; Sillero & Sillero, 2000). The ubiquitous nucleotide Ap_4A is found at intracellular concentrations in the micromolar range (Garrison & Barnes, 1992) in both prokaryotes and eukaryotes. The Ap_4A molecule has been proposed to play roles in signalling by ATP-sensitive K^+ channels (Jovanovic *et al.*, 1997; Martin *et al.*, 1998), initiation of apoptosis (Vartanian *et al.*, 1999), transcriptional activity (Lee *et al.*, 2004; Lee & Razin, 2005), cell division and stress response (Kisselev *et al.*, 1998) and activation of gene expression (Pinto *et al.*, 1986). In addition, the Ap_4A molecule is considered to be an alarm signal as it plays roles in cell division (Nishimura *et al.*, 1997; Nishimura, 1998), heat shock and oxidative stress (Lee *et al.*, 1983). Thus, both the synthesis and the degradation of Ap_4A must be tightly controlled.

Nudix hydrolases, a superfamily of Mg^{2+}-requiring enzymes, catalyse the hydrolysis of nucleoside diphosphates linked to any other moiety *X*. Enzymes of this family are recognized by a highly conserved loop–helix–loop structural motif (known as the 'Nudix motif'), which functions as a versatile metal-binding and catalytic site (Mildvan *et al.*, 2005). This motif contains 23 amino acids, $GX_5EX_7REUXEEXGU$, where *X* is

any residue and *U* is a bulky hydrophobic residue, preferably Leu, Ile or Val (Bessman *et al.*, 1996; Xu *et al.*, 2004). Depending upon their substrate specificity, Nudix hydrolases have been classified into many classes, of which the most studied are asymmetric Ap_4A hydrolases, MutT pyrophosphohydrolases, ADP-ribose pyrophosphatases and GDP-mannose mannosyl hydrolases. The asymmetric Ap_4A hydrolase (EC 3.6.1.17) catalyses the hydrolysis of Ap_4A to ATP and AMP in the presence of a divalent metal ion (Hohn *et al.*, 1982). Bacterial Ap_4A hydrolases, which are associated with pathogenic invasion, have been suggested to be similar to those of plants (Ismail *et al.*, 2003). The mechanism of Ap_4A hydrolases from plants (Guranowski *et al.*, 1994) and bacteria (Conyers *et al.*, 2000) has been studied previously.

Studies of Ap_4A hydrolases from a number of different species have been reported. The X-ray crystal structure of Ap_4A hydrolase from *Caenorhabditis elegans* (*Ce*Ap_4A hydrolase; Bailey *et al.*, 2002) revealed that the enzyme has the $\alpha\beta\alpha$-sandwich architecture of a Nudix fold and that the adenine ring of the product molecule stacks between two highly conserved aromatic residues, most often tyrosines as observed in the NMR structures of free and complexed forms of a 17 kDa human asymmetric Ap_4A hydrolase (*Hs*Ap_4A hydrolase; Swarbrick *et al.*, 2005). The NMR structure of the Ap_4A hydrolase from narrow-leafed blue *Lupinus angustifolius* (*La*Ap_4A hydrolase; Swarbrick *et al.*, 2000) shows the insertion of a helix, named the 'external helix', compared with those of *Ce*Ap_4A and *Hs*Ap_4A hydrolases and is observed to show movement upon substrate binding (Fletcher *et al.*, 2002). However, the binding site of animal-type Ap_4A hydrolase was shown to differ from those of the plant/pathogenic bacterial class of enzymes. In addition, these studies suggest that the reorientation of important side chains that play a major role in substrate binding as found in *La*Ap_4A hydrolase may not be required in animal-type Ap_4A hydrolases. To further understand the mechanism of substrate binding and hydrolysis, three-dimensional crystal structures of Ap_4A hydrolase from *Aquifex aeolicus* (*Aa*Ap_4A) have been determined in free and ATP-bound forms. The binding of the product molecule (ATP) in the active site of the enzyme confirms the functional role of the enzyme.

2. Materials and methods

2.1. Cloning, expression and purification

AaAp_4A hydrolase was amplified by the polymerase chain reaction (PCR) from *A. aeolicus* VF5 genomic DNA and cloned into the expression plasmid pET-21a (Novagen). *Escherichia coli* BL21-CodonPlus (DE3)-RIL cells were transformed by the expression plasmid. The transformants were cultured at 310 K overnight in a medium containing 1.0% polypeptone, 0.5% yeast extract, 0.5% NaCl and 100 µg ml^{-1} ampicillin pH 7.0. The cells were lysed by sonication in 20 m*M* Tris–HCl pH 8.0 containing 500 m*M* NaCl, 5 m*M* β-mercaptoethanol. The lysate was incubated at 363 K for 10 min and centrifuged at 15 000 rev min^{-1} for 30 min at 277 K. After buffer exchange with 20 m*M* Tris–HCl pH 8.0, the protein sample was loaded onto a Toyopearl SuperQ-650M column (Tosoh) pre-equilibrated with 20 m*M* Tris–HCl pH 8.0. The protein bound to the column and was eluted with a linear gradient from 0 to 1000 m*M* NaCl. The protein-containing fraction was subjected to buffer exchange with 20 m*M* Tris–HCl pH 8.0 and was loaded onto a Resource Q column (GE Healthcare Biosciences) pre-equilibrated with the same buffer. The protein bound to the column and was eluted with a linear gradient from 0 to 400 m*M* NaCl. The protein was subjected to buffer exchange with 10 m*M* potassium phosphate buffer pH 7.0 and loaded onto a Bio-Scale CHT2-I column (Bio-Rad) pre-equilibrated with the same buffer. The protein bound to the column and was eluted with a linear gradient from 10 to 500 m*M* potassium phosphate pH 7.0. The protein was subjected to gel filtration on a HiLoad 16/60 Superdex 75pg column (GE Healthcare Biosciences) pre-equilibrated with 20 m*M* Tris–HCl, 150 m*M* NaCl pH 8.0. Fractions were collected and the purified protein was concentrated using Vivaspin 5K molecular-weight cutoff filters (Sartorius). The final preparation was dissolved to a protein concentration of 9.0 mg ml^{-1} in 20 m*M* Tris–HCl containing 150 m*M* NaCl and 1 m*M* DTT (pH 8.0).

2.2. Enzyme assay

Substrate hydrolysis was measured using a previously described method with some modifications (Iwai *et al.*, 2004). In brief, reaction mixtures (300 µl) containing 50 m*M* Tris–HCl, 100 m*M* KCl, 5 m*M* $MgCl_2$, 0–400 µ*M* Ap_nA (Ap_3A, Ap_4A, Ap_5A and Ap_6A) and 10 n*M* enzyme at pH 7.6 were incubated at 298 K. The reaction was stopped by adding 60 µl 500 m*M* EDTA and the protein was removed by ultrafiltration using a membrane filter (Vivaspin VS0112, molecular-weight cutoff 5000, Vivascience). The filtrate was applied onto a reversed-phase column (Capcell Pak C18 MG, 4.6 × 75 mm, Shiseido Co.) equilibrated with 50 m*M* Tris–HCl, 5 m*M* tetra-*n*-butyl ammonium phosphate and 10% methanol pH 7.6. Elution was performed using a linear gradient from 10 to 50% methanol. The substrates and products were monitored at 260 nm and their identification was based on their retention times. The concentrations of the above were calculated by integration of their respective peak areas. The initial velocity was calculated from the product concentration and was plotted against the substrate concentration. The data were fitted to the Michaelis–Menten equation and kinetic constants were calculated using the software *IGOR Pro* v.3.14 (Wave Metrics).

2.3. Crystallization and data collection

Initial *Aa*Ap_4A hydrolase crystallization trials were performed using the sitting-drop vapour-diffusion method at 293 K by mixing 1 µl protein solution (9.0 mg ml^{-1}) with 1 µl reservoir solution. Crystals of *Aa*Ap_4A hydrolase (free form) were obtained by the hanging-drop vapour-diffusion method at 293 K by mixing 1 µl protein solution with an equal volume of reservoir solution comprising 0.1 *M* Tris–HCl pH 7.6,

29%(*w*/*v*) PEG 3350 and 0.75 *M* NaCl. Crystals of the protein–ligand complex were obtained with the same drop ratio of protein solution (1 µl incubated with 10 m*M* Ap_4A for 10 min) and reservoir solution [1 µl 0.1 *M* Tris–HCl pH 8.8, 25%(*w*/*v*) PEG 3350 and 0.2 *M* NaCl]. Crystals of the free and ATP-bound forms were soaked in precipitant solution containing 20%(*v*/*v*) ethylene glycol and 20%(*v*/*v*) ethylene glycol and glycerol mixture, respectively, before flash-freezing and exposing the crystals to X-rays at 100 K. X-ray diffraction data were collected on RIKEN Structural Genomics Beamline II (BL26B2) at SPring-8 (Hyogo, Japan) using SPACE (SPring-8 Precise Automatic Cryo-sample Exchanger; Ueno *et al.*, 2006). The crystals were mounted on the goniometer using a sample-handling robot of SPACE and the crystal-to-detector distance was maintained at 150 mm. A total of 270° data were collected with a 1° oscillation angle and an exposure time of 8 s per degree of oscillation. The collected data were processed with the *HKL* suite (Otwinowski & Minor, 1997).

2.4. Structure solution, refinement and validation

The ligand-free crystal structure was solved using the molecular-replacement program *MOLREP* (Vagin & Teplyakov, 1997; Read, 2001). The three-dimensional atomic coordinates of the Nudix protein Ndx1 (Ap_6A hydrolase) from *Thermus thermophilus* HB8 (PDB code 1vcd; T. Iwai, N. Nakagawa, S. Kuramitsu & R. Masui, unpublished work) were used as the search model as it had a higher sequence identity (40%) than Ap_4A hydrolases from plants and animals. Calculation of Matthews coefficients suggested the presence of four monomers in the asymmetric unit (V_M = 2.23 Å^3 Da^{-1}; Matthews, 1968). A total of 5% of the reflections were set aside for the calculation of R_{free} (Brünger, 1992). After 50 cycles of rigid-body refinement followed by 50 cycles of positional refinement using the program *CNS* (Brünger *et al.*, 1998), R_{work} and R_{free} fell to 28% and 31%, respectively, in the resolution range 50.0–1.8 Å. Subsequently, the residues were refitted using difference electron-density maps and the model was subjected to simulated annealing by heating the system to 3000 K and slowly cooling to 100 K in 10 K steps. At this stage, R_{work} and R_{free} fell to 25% and 28%, respectively. Strong peaks (of greater than 10σ in the $F_o - F_c$ map) were observed for chloride ions and these were added in the subsequent refinements. Water molecules were located and added from difference electron-density maps with the criteria of peak heights greater than 2.8σ and hydrogen-bonding distances of 3.5 Å or less to polar atoms of the protein molecule or other water molecules. The final model was refined to a crystallographic R_{value} of 20.3% (R_{free} = 23.9%). The final refined model contained 4452 protein atoms, 578 water O atoms, 11 chloride ions, 11 ethylene glycol (EDO) molecules, four polyethylene glycol (PEG) molecules, two triethylene glycol (PGE) molecules, one tetraethylene glycol (PG4) and one tris(hydroxymethyl)aminomethane (Tris) molecule.

For the ATP-bound form, preliminary calculations (V_M = 2.23 Å^3 Da^{-1}) suggested the presence of two monomers in the asymmetric unit. An initial model was obtained by molecular replacement (*MOLREP*), using the atomic coordinates of the ligand-free *Aa*Ap_4A hydrolase as a model. A similar approach to that described above was used to refine the ATP-bound form of the enzyme. Clear electron density (of up to 10σ in the $F_o - F_c$ map) for the phosphate groups of the ATP molecule was observed in the active site of the enzyme. In addition, weak electron density (of up to 4.8σ in the $F_o - F_c$ map) was observed near the N-terminal region of the protein molecule. However, water molecules were first located in the model in order to improve the electron density for the ligand molecules. Subsequently, the Ap_4A (with 0.5 occupancy) and ATP molecules were added to the model using the improved electron-density maps. The refined final model (R_{work} = 19.4% and R_{free} = 25.1%) contained 2234 protein atoms, 345 water

Table 1
X-ray data and refinement statistics for the free and ATP-bound forms of *Aa*Ap_4A hydrolase.

Values in parentheses are for the high-resolution shell.

	3i7u (native)	3i7v (complex)
Space group	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a = 85.59, b = 38.05, c = 87.59, β = 93.1	a = 35.30, b = 59.23, c = 67.42, β = 92.6
Temperature (K)	100	100
Matthews coefficient (Å^3 Da^{-1})	2.23	2.20
Solvent content (%)	44.8	44.0
Resolution range (Å)	50–1.8 (1.86–1.8)	50–1.95 (2.02–1.95)
Observed reflections	270072	107834
Unique reflections	52495 (5086)	20328 (1996)
Completeness (%)	99.5 (97.9)	99.8 (99.8)
R_{merge}† (%)	4.6 (25.2)	5.8 (24.9)
$I/\sigma(I)$	30.6 (3.8)	27.8 (3.7)
R_{work} (%)	20.3	19.4
R_{free} (%)	23.9	25.1
Protein model		
No. of monomers per ASU	4	2
Protein atoms	4452	2234
Water molecules	558	345
Product (ATP) molecules	—	2
Substrate (Ap_4A) molecules	—	1
Chloride ions	11	—
Ethylene glycol (EDO)	11	7
Polyethylene glycol (PEG)	4	—
Triethylene glycol (PGE)	2	—
Tetraethylene glycol (PG4)	1	—
Tris molecules	1	—
Glycerol (GOL) molecules	—	2
Deviations from ideal geometry		
Bond lengths (Å)	0.005	0.008
Bond angles (°)	1.2	1.7
Dihedral angles (°)	22.8	22.3
Improper angles (°)	0.7	1.3
Average B factors (Å^2)		
Protein atoms	25.6	24.1
Water molecules	38.3	39.2
ATP molecules	—	23.8
Substrate (Ap_4A) molecule	—	32.7
Chloride ions	26.1	—
Other molecules‡	45.9	47.7
Ramachandran plot (%)		
Most favoured	90.5	90.0
Additionally allowed	9.5	9.6
Generously allowed	0	0.4

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum_{hkl}$ is the sum over all reflections and $\sum_i$ is the sum over i measurements of reflection hkl. ‡ Ethylene glycol (EDO), polyethylene glycol (PEG), triethylene glycol (PGE), tetraethylene glycol (PG4), Tris and glycerol (GOL) molecules.

atoms, two ATP molecules, one Ap_4A molecule, seven ethylene glycol (EDO) molecules and two glycerol (GOL) molecules. The topology parameters for the Ap_4A and ATP molecules were generated using the *HIC-UP* web server (Kleywegt, 2007).

To summarize, the program *CNS* (Brünger *et al.*, 1998) was used to refine both forms of the enzyme. The molecular-modelling program *Coot* (Emsley & Cowtan, 2004) was used for model fitting and minor adjustments. All atoms were refined with unit occupancies except for those of Ap_4A (occupancy 0.5). Simulated-annealing OMIT maps were calculated using *CNS* and were used to correct or check the final protein model. The program *PROCHECK* (Laskowski *et al.*, 1993) was used to validate and check the quality of the final refined model. The details of refinement of both forms of the structure are given in Table 1. The atomic coordinates and the structure factors for the free (PDB code 3i7u) and ATP-bound forms (PDB code 3i7v) have been deposited in the Protein Data Bank (Berman *et al.*, 2000). The structural superposition server 3*dSS* (Sumathi *et al.*, 2006) was used to identify invariant water molecules and for structural superposition. The webserver *PDB Goodies* (Hussain *et al.*, 2002) was also used at various stages of the refinement. The figures were generated using the program *PyMOL* (DeLano Scientific LLC; http://www.pymol.org). Protein–ligand hydrogen bonds were calculated using the program *HBPLUS* (McDonald & Thornton, 1994). A donor–hydrogen acceptor angle greater than or equal to 90° and a donor–acceptor distance less than or equal to 3.5 Å were used as the criteria for delineating hydrogen bonds. The secondary-structure elements of the protein were assigned using the program *DSSP* (Kabsch & Sander, 1983). The structure-based sequence alignment was generated using the program *MUSTANG* (Konagurthu *et al.*, 2006). For NMR structures, representative structures (structures closest to the centroid of the largest cluster) were obtained using the OLDERADO database (Kelley *et al.*, 1996).

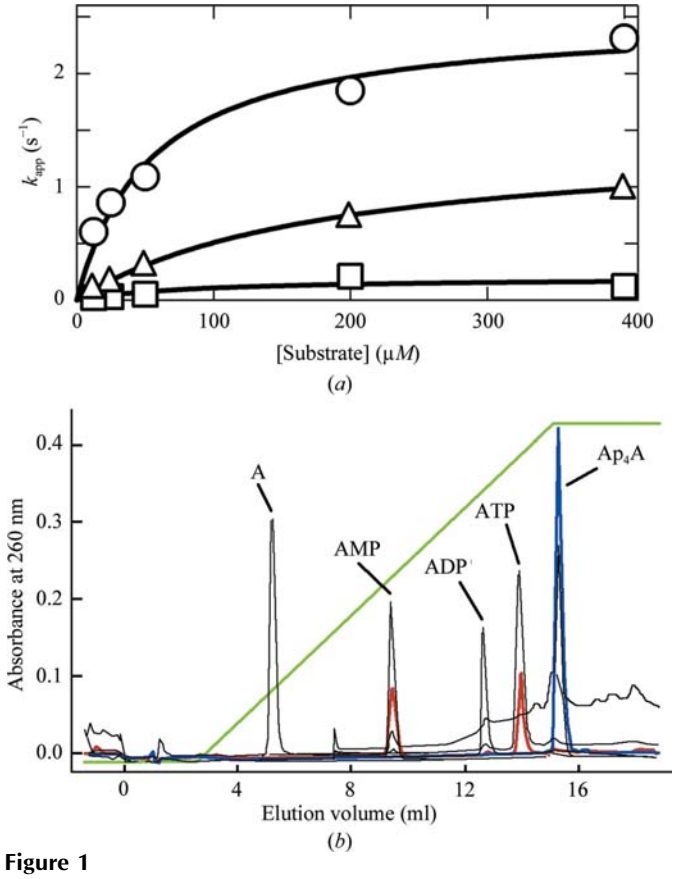

Figure 1
(*a*) The hydrolysis of Ap_nA measured at 298 K. Circles, Ap_4A; triangles, Ap_5A; squares, Ap_6A. (*b*) Elution profiles for Ap_4A, ATP, ADP and AMP. Green, concentration of methanol in the elution buffer; red, elution profile for the reaction mixture containing 1 μ*M* Ap_4A and 1 μ*M* *Aa*Ap_4A hydrolase; blue, elution profile for a control mixture containing 1 μ*M* Ap_4A without enzyme; black, elution profile for standards including ANP and Ap_4A.

Table 2
Kinetic constants for *Aa*Ap_4A hydrolase activity.

Substrate	Product	k_{cat} (s^{-1})	K_m (μ*M*)	k_{cat}/K_m (M^{-1} s^{-1})
Ap_4A	ATP, AMP	2.5	54	4.6×10^4
Ap_5A	ATP, ADP	1.4	183	7.6×10^3
Ap_6A	ATP	0.2	89	2.2×10^3

2.5. Molecular-dynamics simulations

Molecular-dynamics (MD) simulations were performed using the package *GROMACS* v.3.3.3 (Berendsen *et al.*, 1995; Lindahl *et al.*, 2001) running on parallel processors with the AMBER03 all-atom force field (Duan *et al.*, 2003; Sorin & Pande, 2005). During molecular simulations, the crystallographic water molecules were removed from the protein models. A cubic box of size 6.6 nm^3 was generated using the *editconf* module of *GROMACS*. The protein models were solvated with the SPC (simple point charge) water model using the *genbox* program available in the *GROMACS* suite. H atoms were added to the ligand molecule using the *MolProbity* webserver (Davis *et al.*, 2007). The topology and parameter files for the ligand molecule were generated using *AMBER* suite (Case *et al.*, 2006). Furthermore, the charges for the ligand atoms were computed using the *ab initio* program *Gaussian* (Frisch *et al.*, 2004). Energy minimization was performed using the conjugate-gradient method for 200 ps with a maximum force-field cutoff of 1 kJ mol^{-1} nm^{-1}. Chloride ions were used to neutralize the overall charge of the system. Simulations utilized the NPT ensembles with isotropic pressure coupling (τ_p = 0.5 ps) to 10^5 Pa and temperature coupling (τ_t = 0.1 ps) to 300 K. Parrinello–Rahman and Nose–Hoover coupling protocols were used for pressure and temperature, respectively. Long-range electrostatics were computed using the Particle Mesh Ewald (PME) method (Darden *et al.*, 1993) and Lennard–Jones energies were cut off at 1.2 nm. Bond lengths were constrained with the *LINCS* algorithm (Hess *et al.*, 1997). MD simulations were performed for a time period of 10 ns. Analyses were performed with tools available in the *GROMACS* suite. The average structures used for comparison and analyses were calculated using ensembles generated every 2 ps between 1 and 10 ns.

3. Results and discussion

3.1. Enzyme assay

The *Aa*Ap$_4$A hydrolase activity results show that it can catalyse the degradation of Ap$_4$A, Ap$_5$A and Ap$_6$A (Figs. 1*a* and 1*b*) but not Ap$_3$A. The substrates yielded the following products: Ap$_4$A gave ATP and AMP, Ap$_5$A gave ATP and ADP and Ap$_6$A gave two ATP molecules (Table 2). The reactions with all these substrates produced ATP as a common product. The catalytic constant (k_{cat}) and the catalytic efficiency (k_{cat}/K_m; Fersht, 1999) were both in the order Ap$_4$A > Ap$_5$A > Ap$_6$A (Table 2) and these data indicate that the enzyme is an ATP-producing Ap$_4$A hydrolase.

3.2. Overall structure

The crystal asymmetric units of the free and ATP-bound forms of *Aa*Ap$_4$A hydrolase contained four and two monomers, respectively. The monomers of both forms are structurally similar, with average inter-chain root-mean-square deviations (r.m.s.d.s) of 0.9 and 0.4 Å, respectively. The overall tertiary structure of *Aa*Ap$_4$A hydrolase has approximate dimensions of 40 × 36 × 44 Å and shows the common $\alpha\beta\alpha$ Nudix fold. This consists of three α-helices (α_1–α_3), two β-sheets (made up of seven β-strands β_1–β_7) and nine loops (L1–L9) (Fig. 2). Three loops (L2, L3 and L4) are located near the active-site cleft. The central part of the molecule, which contains a twisted four-stranded β-sheet (β_1, β_4, β_5 and β_6), is sandwiched between two antiparallel α-helices (α_2 and α_3) on one side and the α_1 α-helix (oriented perpendicular to the α_2 and α_3 helices) on the other. The angle formed by helices α_2 and α_3 is ∼162°. Furthermore, the angles between α_1 and α_2 and between α_1 and α_3 are approximately 97° and −100°, respectively. All three α-helices of Ap$_4$A hydrolase are highly conserved in related enzymes. The Nudix-family sequence motif (GX_5EX_7RE*UX*EE*X*G*U*) is located mainly in the loop–helix–loop (L4–α_1–L5) region (residues 32–54) as found previously in other homologous structures. However, the substrate-binding region mainly contains residues from the β_1, β_4, β_6 and β_7 strands. This shows that the Nudix-family sequence motif is highly conserved and is important for maintaining the overall tertiary structure as well as for the catalytic activity of the enzyme.

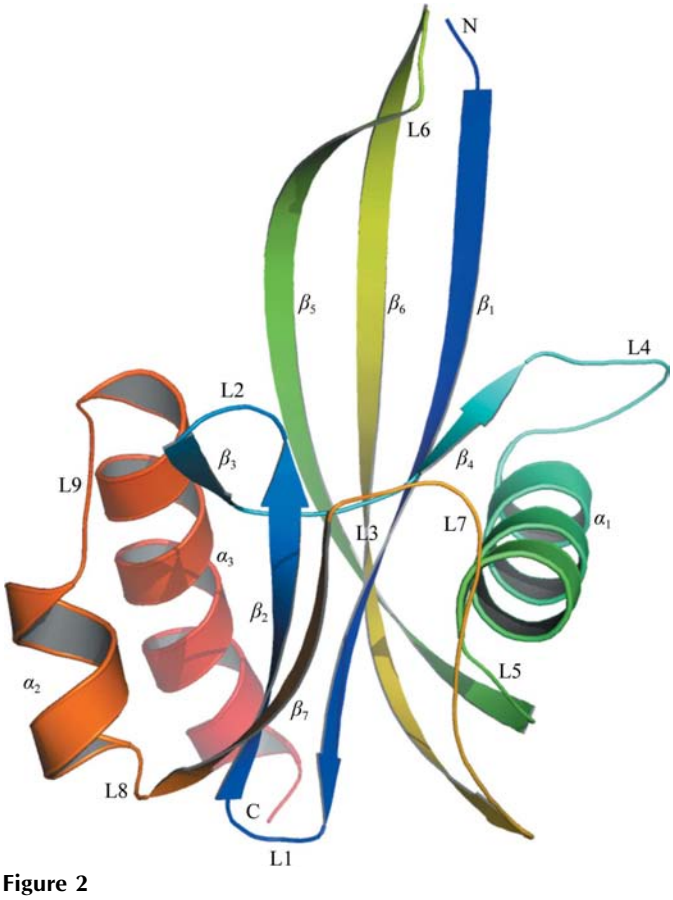

Figure 2
The overall three-dimensional tertiary structure of *Aa*Ap$_4$A hydrolase is shown together with the secondary structures.

3.3. ATP binding

As mentioned previously (§2.3), the substrate molecule Ap$_4$A was cocrystallized with the enzyme. A careful examination of difference electron-density maps revealed several features in the substrate-binding site. However, electron density for the whole Ap$_4$A molecule was not observed in the active site. Instead, electron density (of up to 10σ in the $F_o - F_c$ map for the phosphate groups) for the product molecule ATP was clearly observed (Fig. 3), suggesting cleavage of the substrate molecule during crystallization of the enzyme. The phosphate groups are named based on the previous work of Guranowski *et al.* (1994). The phosphate group closest to the adenine moiety was named the P^1 phosphate and the phosphate farthest from the adenine ring was named the P^3 phosphate.

The adenine moiety of the ATP molecule sits in a groove formed by the central β-sheet and loop L2 and is oriented towards helices α_2 and α_3, while the phosphate groups point towards the catalytic helix α_1. The glycosidic bond of the adenosine group is in the *anti* conformation and the ribose sugar adopts the C2′-*endo* and C3′-*exo* conformation as observed in previous structures of Ap$_4$A hydrolase from other species. The adenine moiety of the product molecule ATP is sandwiched between two highly conserved aromatic residues (Tyr67 and Tyr115), which are approximately parallel to each other and are located at equal distances (∼3.6 Å) from the adenine moiety. The N atoms N1 and N3 of the adenine ring are directly hydrogen bonded to two water molecules (HOH216 and HOH248). These water molecules are also hydrogen bonded to the backbone N atom of Gly117 (located near the active-site pocket) and the backbone carbonyl O atom of Trp68, respectively. In addition, an EDO molecule is found to interact with the ATP molecule in contact with the adenine ring (Fig. 4).

The conserved residue Tyr115 and three water molecules (HOH249, #HOH334 and #HOH335, where symmetry-related water molecules are prefixed with #) are found to coordinate to the sugar O atoms of the ATP molecule. The adenine moiety stacking residue Tyr67 together with the conserved

residues Lys31 and Lys78 and the EDO molecule are hydrogen bonded to the P^1 phosphate of the product molecule ATP. A water molecule (HOH251) anchors the P^2 phosphate of the ATP molecule (Fig. 4). This water molecule is further hydrogen bonded to the phenolic hydroxyl of Tyr69, which is replaced by alanine in the *Ce*Ap$_4$A and *Hs*Ap$_4$A hydrolases (Fig. 5). Furthermore, the residue Ser6 and two water molecules (HOH253 and HOH513) are hydrogen bonded to the phosphoryl O atoms of the P^3 phosphate of the ATP molecule (Fig. 4). However, Ser6 is not conserved and is replaced by asparagine and alanine residues in plant (*La*Ap$_4$A) and animal (*Ce*Ap$_4$A and *Hs*Ap$_4$A) Ap$_4$A hydrolases, respectively (Fig. 5).

3.4. Substrate (Ap$_4$A) binding at a noncatalytic site

In the ATP-bound structure, weak electron density (of up to 4.8σ in the $F_o - F_c$ map) was observed at a site between monomers. An Ap$_4$A molecule was fitted to this density with partial occupancy (0.5). This substrate molecule (Ap$_4$A) is anchored by residues at the monomer–monomer interface. The α- and β-adenine moieties interact with residues Met1*B*, #Trp96*B* and Lys3*A*, respectively. Residues Lys2*B*, Phe5*A*, Phe5*B*, Ile34*A*, Ile34*B* and Pro40*A* further stabilize the substrate molecule either by making hydrogen bonds or by contributing to the hydrophobic environment (Fig. 6). Furthermore, residues Lys3*A*, Lys3*B* and #Asn25*A*, together with several water molecules, coordinate the sugar and phosphate groups (Fig. 6). However, none of these residues involved in the stabilization of the Ap$_4$A molecule at this intermolecular site are conserved, suggesting nonspecific binding of the substrate molecule.

3.5. Structural changes in the active site

The tertiary structures of the free and ATP-bound forms are very similar (with a C^α-atom r.m.s.d. of 1.0 Å). However, several regions (1–3, 22–25, 69–76, 89–91, 96–99 and 105–116) deviate with an average r.m.s.d. of 1.6 Å. The large deviation in the N-terminal residues may possibly arise from binding of the substrate (Ap$_4$A) molecule, although the terminal residues of protein molecules are generally known to be flexible. The regions 22–25, 69–75, 89–91 and 96–99 belong to loops L2, L6 and L7, while the region 105–116 belongs to a highly conserved helix (α_2). It is noteworthy that one of the residues (Tyr115) is involved in the stacking of the adenine moiety of the ATP molecule in the active site. It is observed that although the conformations of the residue Tyr115 are similar in both forms of the enzyme (Fig. 7), the α_2-helix has moved by 1.48 Å towards the active site compared with the native structure.

Conformational changes about the χ^4 angle are observed for the active-site residues Lys31 and Lys78 (Fig. 7). The

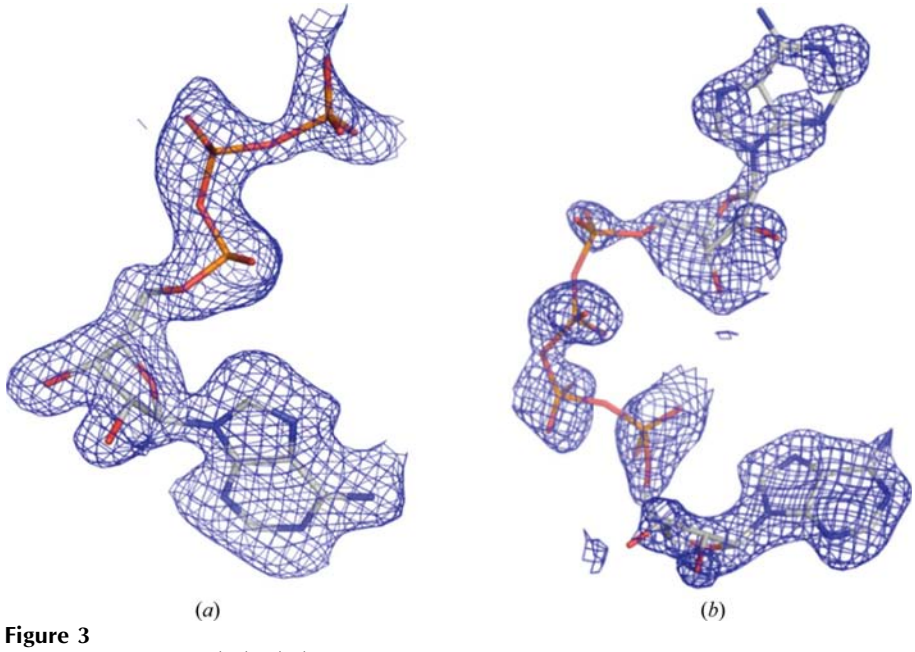

Figure 3
Simulated-annealing ($2|F_o| - |F_c|$) OMIT map contoured at 0.8σ for (*a*) the product (ATP) and (*b*) the substrate (Ap$_4$A) molecules observed in the asymmetric unit of the ATP-bound form of *Aa*Ap$_4$A hydrolase.

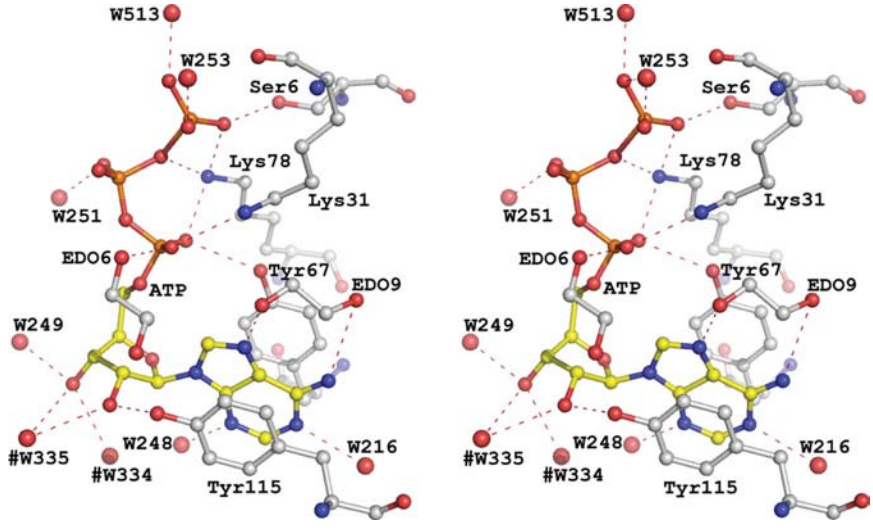

Figure 4
Hydrogen-bonding interactions of the product molecule (ATP) with the protein molecule. Protein atoms are coloured by atom type (red, oxygen; blue, nitrogen; white, carbon) and the ATP molecule is shown in yellow. The interacting atoms of the ATP molecule are also labelled for clarity. Symmetry-related water molecules are prefixed with #.

residue Asn33, which is replaced by a histidine in the *Hs*Ap$_4$A and *Ce*Ap$_4$A hydrolases, also moves away upon substrate binding (Fig. 7). This may be attributed to the conformational change in the residues Glu4, Tyr69 and Lys79 in the vicinity of Asn33. The negatively charged residues Glu35, Glu51 and Glu97 in the substrate-binding site have also rearranged themselves upon substrate binding. In addition, residue Ser24 has also reoriented its O$^{\gamma}$ atom upon substrate binding, causing the removal of a water molecule (HOH255). However, Ser6, which is hydrogen bonded to the ATP molecule, does not show any conformational change. This suggests a possible role of Ser6 in anchoring the substrate molecule in the active site during the hydrolytic process.

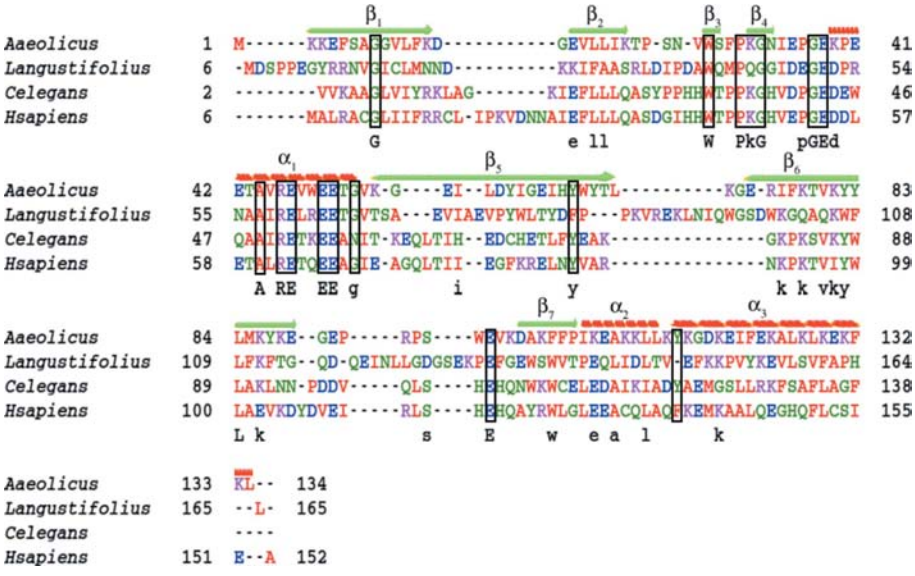

Figure 5
Structure-based sequence alignment of *Aa*Ap$_4$A (PDB code 3i7u; present study), *La*Ap$_4$A (PDB code 1jkn; Fletcher *et al.*, 2002), *Ce*Ap$_4$A (PDB code 1ktg; Bailey *et al.*, 2002) and *Hs*Ap$_4$A (PDB code 1xsc; Swarbrick *et al.*, 2005) hydrolases. The secondary-structural elements are shown for *Aa*Ap$_4$A hydrolase. The important residues located in the active site are outlined. The residues are coloured according to their chemical nature (red, A, F, I, L, M, P, V, W; blue, D, E; magenta, K, R; green, C, G, H, N, Q, S, T, Y). The 'markup row' below each sequence-alignment stretch shows completely conserved residues (upper case) and semi-conserved residues (lower case).

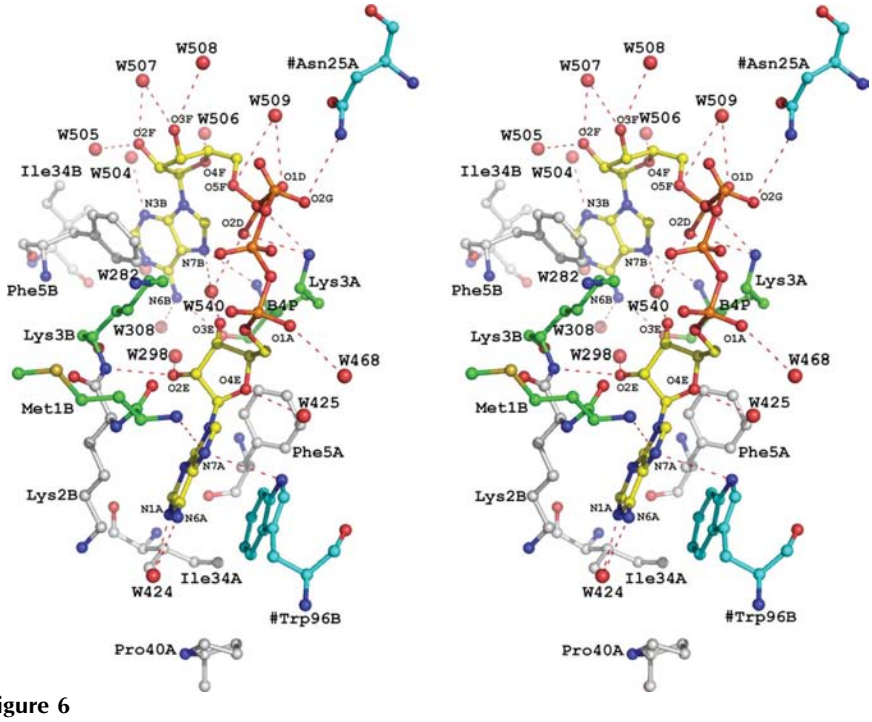

Figure 6
Hydrogen-bond interactions of the substrate molecule (Ap$_4$A) at the intermolecular site. C atoms of the amino acids from the protein molecule are coloured white, whereas those of the Ap$_4$A molecule are coloured yellow. The interacting atoms of the Ap$_4$A molecule are also labelled for clarity. The symmetry-related residues and water molecules are prefixed with #. The C atoms of the symmetry-related residues interacting with the substrate molecule are coloured cyan, whereas those of the adjacent chain residues are coloured green.

3.6. The nucleophilic water molecule

It is well known that a water molecule acts as a nucleophile during substrate hydrolysis by Ap$_4$A hydrolases (Dixon & Lowe, 1989; McLennan, Prescott *et al.*, 1989; McLennan, Taylor *et al.*, 1989; Guranowski *et al.*, 1994). Also, as stated in the literature, the residues equivalent to Glu51 (Glu56, Glu64 and Glu67 in *Ce*Ap$_4$A, *La*Ap$_4$A and *Hs*Ap$_4$A hydrolases, respectively) act as a base holding the nucleophilic water molecule (Swarbrick *et al.*, 2000, 2005; Maksel *et al.*, 2001; Abdelghany *et al.*, 2003). However, another study (Mildvan *et al.*, 2005) favoured the residue equivalent to Glu47 as the general base (Glu52 in *Ce*Ap$_4$A hydrolase). The invariant water molecule in the present structure was identified by pairwise superposition of all nine subunits (three from the *Ce*Ap$_4$A hydrolase and six from the present study). It was observed that two water molecules (HOH235 and HOH236 in PDB entry 3i7u) were invariant in all the subunits. HOH236 (average *B* factor = 34.7 Å^2) is hydrogen bonded to Glu51 O$^{\epsilon1}$, but is located a little away from the ATP molecule in the ATP-bound structure. HOH235 (average *B* factor = 27.6 Å^2) is hydrogen bonded to Glu51 O$^{\epsilon2}$ and the backbone O atom of Lys31. It is further hydrogen bonded to the O3G atom of the ATP molecule. However, no invariant water molecule is found near Glu47. To further explore the structural and/or functional role of these invariant water molecules, a 10 ns molecular-dynamics

simulation was carried out with the substrate molecule (Ap_4A) modelled in the active site. The Ap_4A and ATP molecules observed in the crystal structure as well as the Ap_4A molecule modelled in the active site before and after the MD simulations are shown in Fig. 8. The average structure computed from the ensemble of MD simulations revealed that a water molecule is hydrogen bonded to the P^4 phosphate of the Ap_4A molecule at a distance of 2.48 Å. The residence frequency calculated using MD ensembles for a water molecule around the P^4 phosphate atom of the modelled substrate molecule is 90%. Furthermore, the hydrogen-bond frequency of Glu51 with any water molecule was calculated to be 100% during the MD simulations. Apart from locating the invariant water molecule and verifying it with the MD simulations, we also wished to study the conformational changes of the Ap_4A molecule during MD simulations. The average structure of the Ap_4A hydrolase computed from the ensembles generated during the MD calculations shows that the P^4 phosphate of the Ap_4A molecule was oriented towards one of the invariant water molecules (HOH235). Also, the β-adenine ring of the Ap_4A molecule in the average structure is seen to be orientated towards the N-terminal region of the protein molecule. These observations may suggest the possible path taken by the substrate/product molecule to enter/exit the active site of the enzyme.

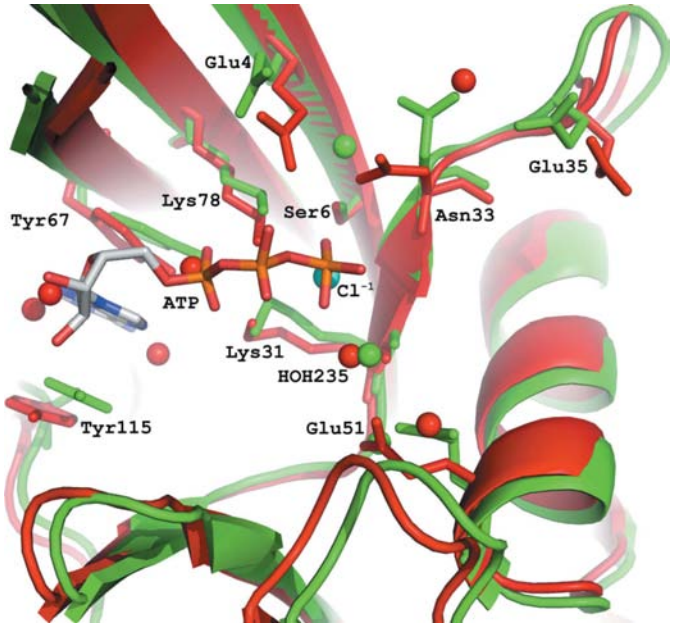

Figure 7
Changes arising from substrate binding in the active site of the enzyme. The free and ATP-bound forms of Ap_4A hydrolase are shown in red and green colours, respectively. One of the invariant water molecules (235) is labelled for clarity. The chloride ion observed in the ATP-free crystal structure is coloured cyan.

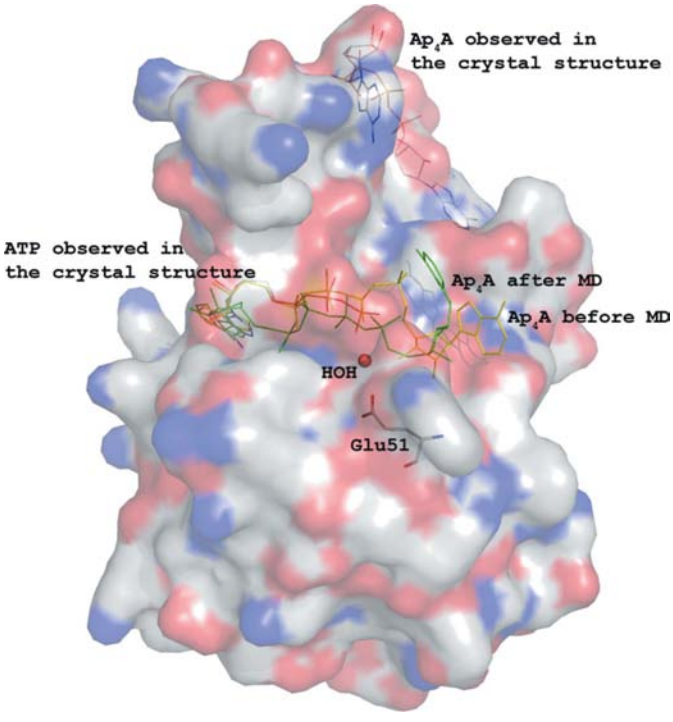

Figure 8
Surface diagram of *Aa*Ap_4A hydrolase with bound substrate (Ap_4A) and product (ATP) molecules. The modelled Ap_4A molecule in the active site of Ap_4A hydrolase before and after MD simulations is coloured yellow and green, respectively. The base residue (Glu51) and the invariant water molecule are also shown.

3.7. Comparison with other related enzymes

A structure-based sequence alignment (Fig. 5) of *Aa*Ap_4A and plant-type (*La*Ap_4A) and animal-type (*Ce*Ap_4A and *Hs*Ap_4A) Ap_4A hydrolases showed sequence identities of 15 and 23%, respectively. Structural superposition of all three Ap_4A hydrolases showed average r.m.s.d.s of 1.7 and 1.8 Å and of 1.5 and 1.4 Å for the free and ligand-bound forms, respectively. However, there are some significant differences: an insertion of 12 residues at the N-terminus and six residues in loop L7 and an additional α-helix are observed in *La*Ap_4A hydrolase. In *Hs*Ap_4A hydrolase there is an insertion of a short helix (α_2) in addition to the three catalytically and structurally important helices. There is also an insertion of six residues at the N-terminus and ten residues corresponding to loop L1 of *Aa*Ap_4A hydrolase. A comparison of *Aa*Ap_4A hydrolase with Ap_6A hydrolase from *T. thermophilus* (*Tt*Ap_6A) shows that the two enzymes have similar three-dimensional structures, with an r.m.s.d. of 1.0 Å and an approximate sequence identity of 40%. In addition, the results obtained from the structural search with the Ap_6A hydrolase using the *DALI* server suggest that *Tt*Ap_6A hydrolase may be an orthologue of *Aa*Ap_4A hydrolase.

JJ thanks beamlines BL12B2, BL26B1 and BL26B2 of SPring-8 for excellent facilities and assistance. KS and SPK thank the Bioinformatics Centre and the Interactive Graphics-Based Molecular Modelling Facility. We also thank Kayoko Matsumoto, Miwa Ohmori, Yayoi Fujimoto, Toshi Arima and Seiki Baba for sample preparation. The authors thank the anonymous referees for their useful suggestions. This work was supported by the RIKEN Structural Genomic/ Proteomics Initiative (RSGI), the National Project on Protein Structural

and Functional Analyses and Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Abdelghany, H. M., Bailey, S., Blackburn, G. M., Rafferty, J. B. & McLennan, A. G. (2003). *J. Biol. Chem.* **278**, 4435–4439.

Bailey, S., Sedelnikova, S. E., Blackburn, G. M., Abdelghany, H. M., Baker, P. J., McLennan, A. G. & Rafferty, J. B. (2002). *Structure*, **10**, 589–600.

Berendsen, H. J. C., van der Spoel, D. A. & van Drunen, R. (1995). *Comput. Phys. Commun.* **91**, 43–56.

Berman, M. H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bessman, M. J., Frick, D. N. & O'Handley, S. F. (1996). *J. Biol. Chem.* **271**, 25059–25062.

Brevet, A., Plateau, P., Cirakoglu, B., Pailliez, J. P. & Blanquet, S. (1982). *J. Biol. Chem.* **257**, 14613–14615.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

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.

Case, D. A. *et al.* (2006). *AMBER* 9. University of California, San Francisco, USA.

Conyers, G. B., Wu, G., Bessman, M. J. & Mildvan, A. S. (2000). *Biochemistry*, **39**, 2347–2454.

Darden, T., York, D. & Pedersen, L. (1993). *J. Chem. Phys.* **98**, 10089–10092.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. III, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids. Res.* **35**, W375–W383.

Dixon, R. M. & Lowe, G. (1989). *J. Biol. Chem.* **264**, 2069–2074.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. & Kollman, P. (2003). *J. Comput. Chem.* **24**, 1999–2012.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fersht, A. (1999). *Structure and Mechanism in Protein Science.* New York: W. H. Freeman & Co.

Fletcher, J. I., Swarbrick, J. D., Maksel, D., Gayler, K. R. & Gooley, P. R. (2002). *Structure*, **10**, 205–213.

Frisch, M. J. *et al.* (2004). *Gaussian.* Gaussian Inc., Wallingford, Connecticut, USA.

Garrison, P. N. & Barnes, L. D. (1992). *Ap_4A and Other Dinucleoside Polyphosphates*, edited by A. G. McLennan, pp. 29–61. Boca Raton: CRC Press.

Goerlich, O., Foeckler, R. & Holler, E. (1982). *Eur. J. Biochem.* **126**, 135–142.

Guranowski, A., Brown, P., Ashton, P. A. & Blackburn, G. M. (1994). *Biochemistry*, **33**, 235–240.

Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997). *J. Comput. Chem.* **18**, 1463–1472.

Hilderman, R. H. & Ortwerth, B. J. (1987). *Biochemistry*, **26**, 1586–1591.

Hohn, M., Albert, W. & Grummt, F. (1982). *J. Biol. Chem.* **257**, 3003–3006.

Hussain, A. S. Z., Shanthi, V., Sheik, S. S., Jeyakanthan, J., Selvarani, P. & Sekar, K. (2002). *Acta Cryst.* D**58**, 1385–1386.

Ismail, T. M., Hart, C. A. & McLennan, A. G. (2003). *J. Biol. Chem.* **278**, 32602–32607.

Iwai, T., Kuramitsu, S. & Masui, R. (2004). *J. Biol. Chem.* **279**, 21732–21739.

Jovanovic, A., Alekseev, A. E. & Terzica, A. (1997). *Biochem. Pharmacol.* **54**, 219–225.

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

Kelley, L. A., Gardner, S. P. & Sutclife, M. J. (1996). *Protein Eng.* **9**, 1063–1065.

Kisselev, L. L., Justesen, J., Wolfson, A. D. & Frolova, L. Y. (1998). *FEBS Lett.* **427**, 157–163.

Kleywegt, G. J. (2007). *Acta Cryst.* D**63**, 94–100.

Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Leskh, A. M. (2006). *Proteins*, **64**, 559–574.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lee, P. C., Bochner, B. R. & Ames, B. N. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 7496–7500.

Lee, Y. N., Nechushtan, H., Figov, N. & Razin, E. (2004). *Immunity*, **20**, 145–151.

Lee, Y. N. & Razin, E. (2005). *Mol. Cell. Biol.* **20**, 8904–8912.

Lindahl, E., Hess, B. & van der Spoel, D. (2001). *J. Mol. Model.* **7**, 306–317.

Maksel, D., Gooley, P. R., Swarbrick, J. D., Guranowski, A., Gange, C., Blackburn, G. M. & Gayler, K. R. (2001). *Biochem. J.* **357**, 399–405.

Martin, F., Pintor, J., Rovira, J. M., Ripoll, C., Miras-Portugal, M. T. & Soria, B. (1998). *FASEB J.* **12**, 1499–1506.

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

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

McLennan, A. G., Prescott, M. & Evershed, R. P. (1989). *Biomed. Environ. Mass Spectrom.* **18**, 450–452.

McLennan, A. G., Taylor, G. E., Prescott, M. & Blackburn, G. M. (1989). *Biochemistry*, **28**, 3868–3875.

Mildvan, A. S., Xia, Z., Azurmendi, H. F., Sawaswat, V., Legler, P. M., Massiah, M. A., Gabelli, S. B., Bianchet, M. A., Kang, L. W. & Amzel, L. M. (2005). *Arch. Biochem. Biophys.* **433**, 129–143.

Nishimura, A. (1998). *Trends Biochem. Sci.* **23**, 157–159.

Nishimura, A., Moriya, S., Ukai, H., Nagai, K., Wachi, M. & Yamada, Y. (1997). *Genes Cells*, **2**, 401–413.

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

Pinto, R. M., Canales, J., Sillero, M. A. G. & Sillero, A. (1986). *Biochem. Biophys.* **138**, 261–267.

Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.

Sillero, A. & Sillero, M. A. (2000). *Pharmacol. Ther.* **87**, 91–102.

Sorin, E. J. & Pande, V. S. (2005). *Biophys. J.* **88**, 2472–2493.

Sumathi, K., Ananthalakshmi, P., Roshan, M. N. A. M. & Sekar, K. (2006). *Nucleic Acids Res.* **34**, W128–W134.

Swarbrick, J. D., Bashtannyk, T., Maksel, D., Zhang, X. R., Blackburn, G. M., Gayler, K. R. & Gooley, P. R. (2000). *J. Mol. Biol.* **302**, 1165–1177.

Swarbrick, J. D., Buyya, S., Gunawardana, D., Gayler, K. R., McLennan, A. G. & Gooley, P. R. (2005). *J. Biol. Chem.* **280**, 8471–8481.

Ueno, G., Kanda, H., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. (2006). *J. Struct. Funct. Genomics*, **7**, 15–22.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Vartanian, A., Alexandrov, I., Prudowski, I., McLennan, A. & Kisselev, L. (1999). *FEBS Lett.* **456**, 175–180.

Xu, W., Dunn, C. A., Jones, C. R., D'Souza, G. & Bessman, M. J. (2004). *J. Biol. Chem.* **279**, 24861–24865.

Zamecnik, P. C., Stephenson, M. L., Janeway, C. M. & Randerath, K. (1966). *Biochem. Biophys. Res. Commun.* **24**, 91–97.