BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Structure of a double-domain phosphagen kinase reveals an asymmetric arrangement of the tandem domains

Zhiming Wang,[a,b] Zhu Qiao,[a] Sheng Ye[a]* and Rongguang Zhang[a,c]*

[a]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China, [b]University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China, and [c]National Center for Protein Science Shanghai, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201203, People's Republic of China. *Correspondence e-mail: yesheng@moon.ibp.ac.cn, rzhang@ibp.ac.cn

Received 1 October 2014
Accepted 20 January 2015

Edited by Q. Hao, University of Hong Kong

Keywords: double-domain arginine kinase.

PDB references: double-domain arginine kinase, native, 4rf6; complex with L-arginine, 4rf7; complex with ADP, 4rf8; complex with L-arginine and ATPγS, 4rf9

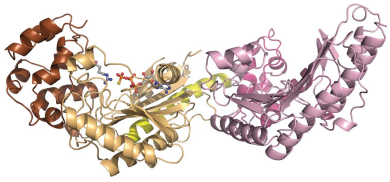
Supporting information: this article has supporting information at journals.iucr.org/d

Tandem duplications and fusions of single genes have led to magnificent expansions in the divergence of protein structures and functions over evolutionary timescales. One of the possible results is polydomain enzymes with interdomain cooperativities, few examples of which have been structurally characterized at the full-length level to explore their innate synergistic mechanisms. This work reports the crystal structures of a double-domain phosphagen kinase in both apo and ligand-bound states, revealing a novel asymmetric L-shaped arrangement of the two domains. Unexpectedly, the interdomain connections are not based on a flexible hinge linker but on a rigid secondary-structure element: a long α-helix that tethers the tandem domains in relatively fixed positions. Besides the connective helix, the two domains also contact each other directly and form an interdomain interface in which hydrogen bonds and hydrophobic interactions further stabilize the L-shaped domain arrangement. Molecular-dynamics simulations show that the interface is generally stable, suggesting that the asymmetric domain arrangement crystallographically observed in the present study is not a conformational state simply restrained by crystal-packing forces. It is possible that the asymmetrically arranged tandem domains could provide a structural basis for further studies of the interdomain synergy.

1. Introduction

Gene duplication plays a major role in the evolutionary process of the emergence of new genes and usually leads to two kinds of results: a gene family with homologous members or a single gene with tandem elements that are similar to each other (Lynch & Conery, 2000). For enzymes, the former is common but the latter is rare. It is apparent that if an enzyme with polydomains functions as efficiently as the simple sum of the isolated domains, there is no reason for this enzyme to survive natural selection. For the polydomain enzymes that are retained, interdomain synergism could be one of the evolutionary advantages that they possess, which often enhances the catalytic activities of each single domain according to the results of biochemical studies (Tada & Suzuki, 2010). However, few examples of polydomain enzymes have been structurally characterized at the full-length level to explore their innate synergistic mechanisms.

Polydomain phosphagen kinase (PK) is a representative polydomain enzyme with two or three highly conserved



domains. PK catalyzes the reversible transfer of a γ-phosphoryl group of ATP to a reactive guanidinium group, yielding ADP and a phosphorylated guanidino compound (known as a phosphagen) that functions as an agent for the temporal and spatial buffering of ATP levels in cell. Therefore, PKs play a remarkably important role, majorly as an energy homeostasis controller in tissues demanding high and variable rates of energy turnover, such as muscle fibres, neuronal tissues, spermatozoa, retinas and gills (Ellington, 2001). To date, eight phosphagens have been identified with their own respective phosphagen kinases, among which arginine kinases (AKs) and creatine kinases (CKs) are the most extensively studied subfamilies. CK is the sole PK in vertebrates, while AK is widely distributed throughout bacteria, protozoa, invertebrates and lower chordates (Andrews *et al.*, 2008; Ellington, 2001).

Almost all PKs are single-domain enzymes comprising only one PK domain, with the exception of several unusual polydomain phosphagen kinases that have been identified in recent decades (Tombes & Shapiro, 1985; Suzuki *et al.*, 1997). Bioinformatic analysis of polydomain PK genes suggested that they resulted from tandem gene duplications and subsequent divergence (Compaan & Ellington, 2003). Most of the polydomain PKs found to date are double-domain AKs (ddAKs), the first of which to be identified was ddAK from the sea anemone *Anthopleura japonicus*, which was isolated and sequenced in 1997 (Suzuki *et al.*, 1997). Its domain 1 (D1) and domain 2 (D2) share relatively high identity (77%), which is consistent with their genesis by duplication. A series of biochemical studies of isolated D1 and D2 showed that they have lower catalytic activities compared with intact ddAK (Tada *et al.*, 2008; Tada & Suzuki, 2010). Intriguingly, with the exception of ddAK from *A. japonicus*, the other ddAKs have a D2-dependent D1, meaning that the isolated D1 has little or even zero activity (Suzuki *et al.*, 2003; Uda *et al.*, 2008) and thus suggesting an asymmetric arrangement of the tandem domains.

The catalytic mechanism of single-domain PK has been thoroughly investigated during the past fifty years, showing a random-order and rapid-equilibrium kinetic mechanism (Griffiths *et al.*, 1957; Morrison & James, 1965). Although most AKs function in a monomeric form, CKs act as homodimers or heterodimers in the cytosol or as homo-octamers in mitochondria (tetramerized dimers) and have been shown both biochemically and structurally to possess negative cooperativity between protomers (Hornemann *et al.*, 2000). Further structural studies of a dimeric single-domain AK (dsdAK) revealed the involvement of a reciprocating mechanism, *i.e.* the twofold-symmetric dimers of PK catalyze in an asymmetric way with only one protomer working alternatively at any time (Wu *et al.*, 2010). To fulfill this negative cooperativity, Wu and coworkers also observed that a symmetric dimerization interface with sufficient hydrogen-bond connections is critically necessary. However, the mechanism of negative cooperativity in dimeric PKs has remained controversial to date; Lim and coworkers reported that there was no obvious structural support for negative cooperativity in glycocyamine kinase, which is a heterodimeric PK (Lim *et al.*, 2010). Since previous biochemical evidence has shown that there is cooperativity to some extent between the two domains of ddAKs (Tada *et al.*, 2008; Tada & Suzuki, 2010), the immediate question here is whether or not a double-domain PK with two tandem domains can adopt a symmetric structural arrangement as in dimeric PKs.

In the present work, we determined the crystal structures of a ddAK from *A. japonicas* in both apo and ligand-bound states, which to our best knowledge are the first reported structures of a polydomain PK. The structures were then compared with that of dsdAK to uncover the differences in domain arrangements and interdomain interactions. We believe that these findings could shed light on our understanding of polydomain enzymes.

2. Materials and methods

2.1. Protein preparation

The cDNA encoding ddAK from *A. japonicus* was a kind gift from Professor Tomohiko Suzuki, Kochi University, Japan. Wild-type ddAK was cloned into pET-28a expression vector (Novagen, USA). Subsequent DNA sequencing revealed the presence of three missense mutations, I52T, G83A and E658Q, compared with the sequence from GenBank (AB008014.1 and BAA22888.1), which might be the result of specimen heterogeneity or PCR misincorporation, according to Suzuki and coworkers (Tada *et al.*, 2008). Triple site-directed mutations were generated to acquire the correct ddAK sequence, which was further confirmed by DNA sequencing. In order to obtain target proteins with short N-terminal tags in an inexpensive way, wild-type ddAK was cloned into a homemade expression vector with a PreScission protease cleavage site introduced in pET-28a just before its multiple cloning site, dubbed pET-28p.

Escherichia coli strain BL21 (DE3) cells (Novagen, USA) were transformed with pET-28p-ddAK and then cultured in LB medium to express native ddAK protein after the addition of 0.2 m*M* IPTG. After centrifugation, the collected cells were resuspended and lysed in PBS buffer using a high-pressure homogenizer (JNBIO, China). Target proteins were isolated from the clarified lysate *via* immobilized-nickel affinity chromatography followed by overnight treatment with PreScission protease at 4°C to remove the fusion tag. The cleavage mixture was then loaded onto a Resource S column (GE Healthcare, USA) for cation-exchange chromatography and a Superdex 200 column (GE Healthcare, USA) for subsequent size-exclusion chromatography in a solution consisting of 20 m*M* Tris–HCl pH 8.0, 5 m*M* $MgCl_2$, 50 m*M* NaCl. Purified target proteins were flash-frozen in liquid nitrogen and then stored at −80°C for future use.

2.2. ThermoFluor assays

ThermoFluor assays were conducted using 1 mg ml^{-1} ddAK protein in test conditions (different buffer solutions or ligands; see Supplementary Tables S1 and S2) supplemented with 1:200 diluted SYPRO Orange dye (Invitrogen, USA).

The solution used to store the protein was employed as a control, since ddAK tends to aggregate in pure water. The fluorescence changes in the wells of a 96-well plate were monitored simultaneously with a real-time PCR machine (CFX96, Bio-Rad, USA) in FRET mode, where the fluorescence intensity was measured with excitation and emission wavelengths of 490 and 575 nm, respectively. The temperature gradient was set in the range 20–95°C with a increment of 0.5°C per 15 s. The data were analyzed with *DSF Analysis* (an Excel-based tool) and the Boltzmann model was used to plot the melting curves of ddAK to obtain the midpoint of the thermal unfolding value for ddAK using the curve-fitting software *XLfit* 5 (ID Business Solutions Ltd). All measurements were repeated three times to avoid random errors (Boivin *et al.*, 2013).

2.3. Crystallization

Crystallization screening using the hanging-drop vapour-diffusion method was performed with ddAK at 16°C in 96-well plates set up automatically by a Mosquito robot (TTP Labtech, England). Hanging drops consisting of 0.2 µl protein solution (10 mg ml^{-1}) and 0.2 µl reservoir solution were equilibrated against 100 µl reservoir solution. Screening of 400 commercially available crystallization conditions gave several hits, which were further optimized manually in 2 µl drops against 400 µl reservoir solution. Diffraction-quality crystals of ddAK were usually obtained 2 d after setup in a solution consisting of 100 m*M* bis-tris pH 6.5, 20% PEG 5000. Native ddAK was then co-crystallized with various ligands or their combinations, including the substrate arginine, the product ADP and the substrate analogue ATPγS. Both ddAK–ADP and ddAK–arginine complex crystals grew in a condition consisting of 200 m*M* potassium acetate pH 7.8, 20% PEG 3350; the latter crystals were further soaked in different ligand solutions, eventually yielding ddAK–arginine–ATPγS complex crystals. All diffraction-quality crystals were soaked in cryoprotectant consisting of reservoir solution plus 30%(*v*/*v*) glycerol and flash-cooled in liquid nitrogen prior to data collection.

2.4. Data collection, phasing and model refinement

Crystallographic diffraction data sets for crystals of ddAK and its complexes (ddAK–ADP and ddAK–arginine–ATPγS) were collected on the BL-19U and BL-17U beamlines of Shanghai Synchrotron Radiation Facility, People's Republic of China, respectively. The data set for the ddAK–arginine complex crystal was collected using the in-house X-ray source (MicroMax-007 generator with VariMax HR optics, Rigaku, Japan) at the Institute of Biophysics, Chinese Academy of Sciences.

The phasing problem was solved for the apo ddAK structure by the molecular-replacement (MR) method with *phenix.phaser* (McCoy *et al.*, 2007) using the structure of dsdAK from the sea cucumber *Stichopus japonicus* (PDB entry 3ju5; Wu *et al.*, 2010) as a search model. The final model was manually adjusted using *Coot* (Emsley & Cowtan, 2004) and refined with *phenix.refine* (Adams *et al.*, 2010; Afonine *et al.*, 2012). The refined model of ddAK was then used as a search model in the MR method to solve the structures of the ddAK complexes. Figures showing protein structures were prepared using *PyMOL* (http://www.pymol.org).

2.5. Molecular-dynamics simulations

Since the ddAK structure in the apo state has the highest resolution (1.95 Å) among the four data sets, it was selected to provide the atomic coordinates for molecular-dynamics (MD) simulations. After solvation in a rectangular box of ionized

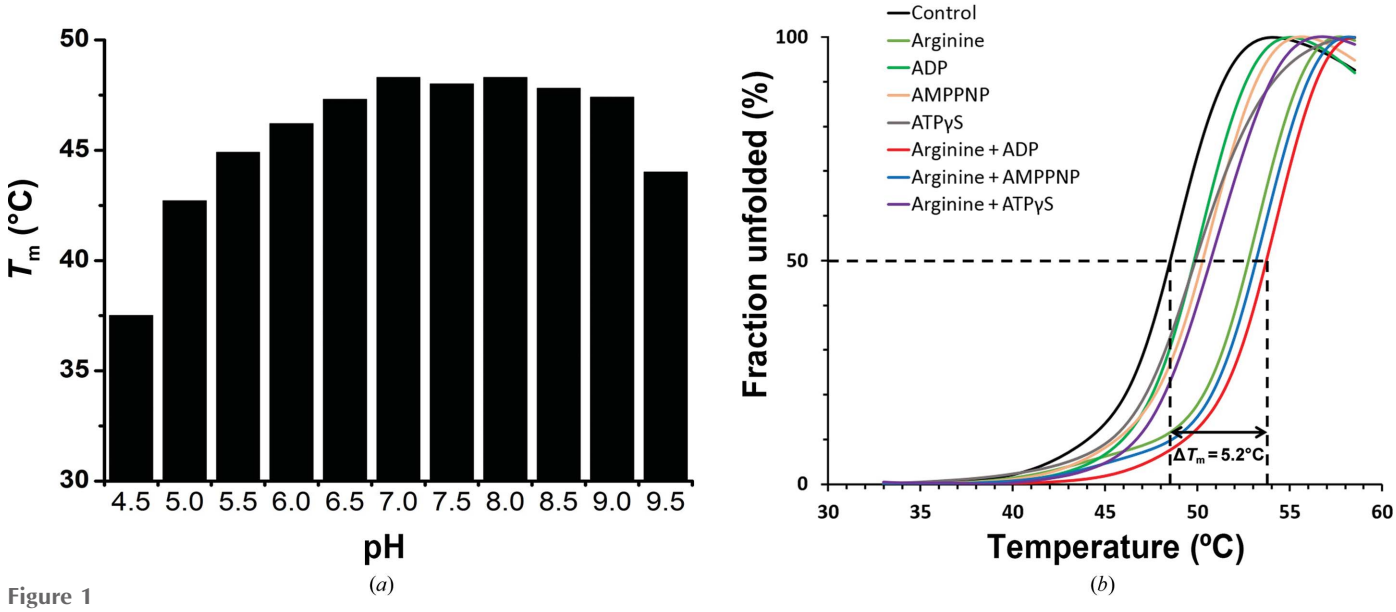


Figure 1
Thermal shift assays of ddAK in buffers with different pH values or with ligands. (*a*) The critical T_m of ddAK in buffers with different pH values. (*b*) Representative melting curves from different combinations of ligands. The results obtained in the absence of ligands are indicated as a black solid line, while the results in the presence of different ligands or their combinations are indicated as different coloured solid lines and are labelled accordingly. The condition containing arginine and ADP was shown to be the most stable, leading to an increase in the T_m of 5.2°C.

water containing Na^+ and Cl^- ions at a concentration of 0.1 *M*, the ddAK system included roughly 70 600 atoms in a box of dimensions 131 × 70 × 83 Å.

All MD simulations were performed using *NAMD* 2.9 (Phillips *et al.*, 2005) on the ScGrid of the Supercomputing Environment of the Chinese Academy of Sciences and were analyzed with *VMD* 1.9 (Humphrey *et al.*, 1996). The CHARMM22 force field with CMAP corrections was employed (MacKerell *et al.*, 1998, 2001) and the TIP3P model was used for water molecules (Price & Brooks, 2004). The van der Waals interaction cutoff distances were set at 12 Å and long-range electrostatic forces were computed using the particle mesh Ewald (PME) summation method with a grid size of 1 Å (Darden *et al.*, 1993). In every simulation run, the ddAK system was equilibrated in the NPT ensemble (T = 300 K, P = 101 325 Pa) for 5 ns. At least four runs were carried out for each simulation process to avoid the influence of randomly chosen initial velocities.

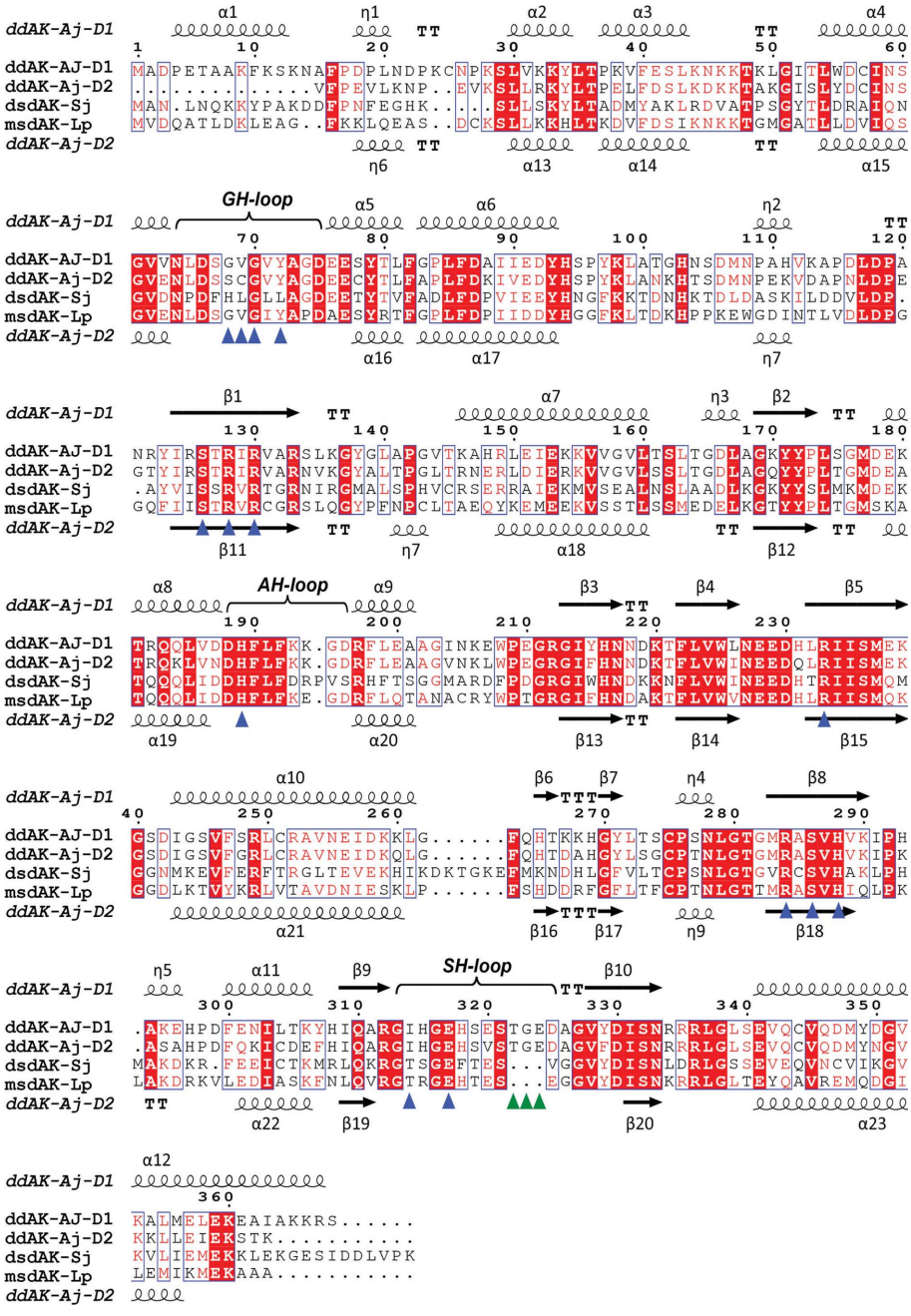


Figure 2
Structure-based sequence alignment of AKs. Sequences of structurally characterized representative AKs are aligned and numbered accordingly. Conserved residues are coloured red and similar residues are coloured blue. Secondary-structure elements of ddAK D1 are shown and labelled above the sequences and those of D2 are labelled below the sequences. The GH-, AH- and SH-loops are indicated. The three extra residues in the SH-loop are indicated by green triangles; the residues involved in ligand binding are indicated by blue triangles. From top to bottom, the sequences are *A. japonicas* ddAK D1 (1–368), *A. japonicas* ddAK D2 (369–715), *S. japonicas* dsdAK (PDB entry 3ju6; Wu *et al.*, 2010) and *Limulus polyphemus* msdAK (PDB entry 1m80; Niu *et al.*, 2011).

3. Results

3.1. Buffer and additive screening

As used as a routine approach prior to crystallization trials in a large number of structural studies, the thermostabilities of recombinantly expressed ddAK in different conditions were assessed using ThermoFluor assays to seek an optimized buffer system for crystallization experiments. For a pH range from 4.5 to 9.5 with different buffering chemicals, the melting temperature (T_m) of ddAK reached its highest value (~48°C) at a pH between 7.0 and 8.5, which is about 10°C higher than in a buffer at pH 4.5 (Fig. 1*a*). Consistent with this result, Suzuki and coworkers showed that the optimum pH for the forward reaction of ddAK extracted from *A. japonicus* body-wall muscle is in the range from 7.9 to 9.1 (Suzuki *et al.*, 1997). Accordingly, Tris buffer at pH 8.0 was used to prepare ddAK samples for crystallization and further analysis.

Subsequently, ThermoFluor was also employed, as was its primary purpose when developed, to examine the thermostabilities of ddAK in the presence of different ligands as well as Mg^{2+} ions. There are two substrates in the reaction catalyzed by ddAK, arginine and ATP, the latter of which was substituted with its chemically inert analogues AMPPNP and ATPγS in our ThermoFluor assays owing to autohydrolysis of ATP. Surprisingly, the results showed that arginine is a notable factor in increasing the thermostability of ddAK. Compared

with ddAK in apo, ADP-bound, AMPPNP-bound and ATPγS-bound states, adding free L-arginine to the solutions dramatically increased the T_m by about 4.6°C (Fig. 1*b*), indicating an important role of the substrate arginine in inducing the conformational change of ddAK. Meanwhile, the most effective combination of ligands is arginine and ADP, which increased the T_m by about 5.2°C.

3.2. Overall structure

As a representative member of the polydomain PKs, ddAK consists of two 40 kDa kinase domains, the N-terminal D1 (1–368) and the C-terminal D2 (369–715), which share a relatively high amino-acid sequence identity (77%; Fig. 2). These two domains have an identity of 48–53% to other AKs. Thus, as expected, both D1 and D2 share the overall architecture of PK family members, composed of a small N-terminal subdomain (NsD) and a large C-terminal subdomain (CsD), the counterparts of which in single-domain PKs are usually called the N-terminal and C-terminal domains, respectively (Fig. 3*a*). Consistent with the high sequence identity between D1 and D2, their structural similarity is quite high according to the topology schematic shown in Fig. 3(*d*) and superposition resulted in an r.m.s.d. of 0.7 Å (over 326 C^{α} atoms; Fig. 3*b*). In

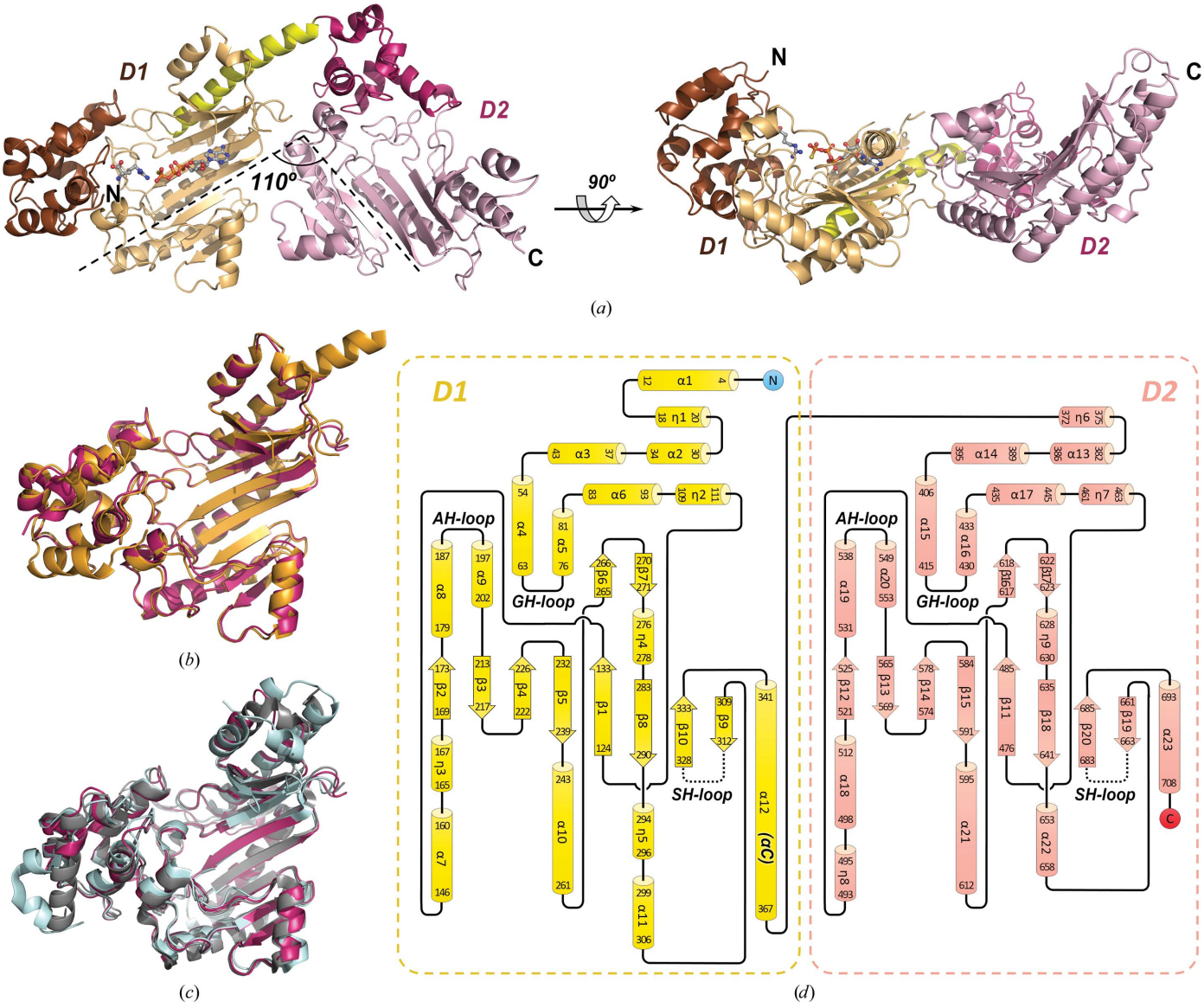


Figure 3
Overall structure of the ddAK–arginine–ATPγS complex. (*a*) Cartoon representation of ddAK in complex with ATPγS and arginine. D1 is coloured in gold, D2 in pink and the αC helix in yellow. In each domain, the NsD is shaded darker than the CsD. D2 can be roughly obtained by a ∼110° rotation of D1, which is shown by the crossing dashed lines. Both the substrate arginine and the substrate analogue ATPγS were found in the active site of D1. Ligands are depicted in ball-and-stick representation (carbon, white; nitrogen, blue; oxygen, red; sulfur, yellow; phosphorus, orange). (*b*) Superposition of the structures of ddAK D1 (gold) and D2 (pink). (*c*) Superposition of the structures of ddAK D2 (pink), msdAK (grey) and dsdAK (cyan) in the substrate-free state. (*d*) Overall topology schematic of ddAK. Secondary-structure elements comprising D1 and D2 are framed with dashed boxes and coloured using the same scheme as in (*a*). Helix αC and the GH-, AH- and SH-loops are indicated; the SH-loops are missing in the complex structure and are shown as dashed lines.

Table 1
Diffraction data-collection and structure-refinement statistics.

Values in parentheses are for the highest resolution shell.

	ddAK	ddAK–arginine	ddAK–ADP	ddAK–arginine–ATPγS
PDB code	4rf6	4rf7	4rf8	4rf9
Data collection				
Resolution limits (Å)	47.7–1.95 (1.99–1.95)	47.9–2.10 (2.14–2.10)	47.5–2.17 (2.19–2.17)	48.0–2.35 (2.43–2.40)
Wavelength (Å)	1.03317	1.54178	0.97915	0.97915
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters				
a (Å)	77.4	78.5	79.5	78.3
b (Å)	58.9	59.3	59.3	59.5
c (Å)	162.1	162.3	163.7	162.1
$\alpha = \gamma$ (°)	90.0	90.0	90.0	90.0
β (°)	91.0	91.3	90.8	91.4
Measured reflections	640982	355818	292008	186134
Multiplicity	6.3 (3.9)	4.1 (4.0)	3.7 (3.7)	3.2 (3.1)
Mean $I/\sigma(I)$	6.3 (2.2)	10.1 (1.9)	10.7 (2.3)	11.4 (2.7)
Completeness (%)	95.7 (65.9)	100.0 (100.0)	99.8 (100.0)	96.9 (98.6)
$R_{\rm merge}$†	0.096 (0.562)	0.072 (0.794)	0.091 (0.681)	0.096 (0.710)
Refinement statistics				
Resolution range (Å)	47.7–1.95	47.9–2.10	47.5–2.17	48.0–2.35
Reflections				
Working set	88394	83363	75670	57120
Test set	4687	4399	3994	3036
$R_{\rm cryst}$‡ (%)	17.5	18.2	19.2	19.9
$R_{\rm free}$§ (%)	21.2	21.8	23.9	24.8
R.m.s.d.				
Bonds (Å)	0.005	0.004	0.003	0.003
Angles (°)	0.936	0.802	0.833	0.702
Average B factor (Å^2)				
Overall	29.8	43.2	51.0	48.9
Protein	28.4	42.9	50.7	49.0
Waters	40.2	46.3	56.5	48.0
Modelled residues	1394	1392	1379	1376
Missing residues	42	44	57	60
No. of atoms				
Total	12619	12184	11708	11638
Protein	11081	11008	10921	10884
Ligands	—	24	54	43
Solvent	1538	1152	733	692
Ramachandran plot				
Favoured regions (%)	97.4	97.0	96.8	96.8
Outliers (%)	0.4	0.4	0.7	0.6

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the integrated intensity of a given reflection. ‡ $R_{\rm cryst} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ denote observed and calculated structure factors, respectively. § $R_{\rm free}$ is equivalent to $R_{\rm cryst}$ but was calculated using a randomly chosen 10% of the reflections as the test set, which were excluded from the refinement process.

addition, superposition of ddAK D2, dsdAK and monomeric single-domain AK (msdAK) shows that the overall fold of AKs is highly conserved (Fig. 3*c*).

In contrast to the dimeric PKs with a pseudo-twofold symmetric arrangement, ddAK has its two domains joined in an L-shape, meaning that D2 can be roughly obtained by a ~110° rotation of D1 (Fig. 3*a*). The linker between D1 and D2 is a relatively rigid connection based on a connecting α-helix (αC; yellow in Fig. 3*a*) fused to the very last α-helix of D1 and the very first helix of D2. Additionally, the two domains build several indirect connections at the interface between the CsD of D1 and both the NsD and the CsD of D2, which will be discussed later in detail.

Prior to the present work, D1 and D2 were assigned, according to non-structure-based sequence alignments, as residues 1–364 and 365–715, respectively (Suzuki *et al.*, 1997). However, our structure of ddAK apparently shows a unified αC helix between D1 and D2, which, as a secondary-structure element, is not supposed to be broken in domain assignment. Given that the long αC helix is mainly packed against the bottom of D1, we propose defining D1 from the N-terminus of ddAK to the C-terminal end of αC, *i.e.* D1 and D2 are structurally assigned as residues 1–368 and 369–715, respectively.

Despite the high resolution of the ddAK structures that we have determined, some loop regions could not be traced owing to a lack of clear electron density. The loop region around residues 667–676 is missing in all of the ddAK chains in our four structures. Meanwhile, the loop region around residues 315–323 can only be observed in chain *B* of the apo ddAK and ddAK–arginine structures. These two loop regions are closely related to the fixation of bound substrates. Further details of the data-collection and refinement statistics are summarized in Table 1. The coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 4rf6 (apo ddAK), 4rf7 (ddAK–arginine), 4rf8 (ddAK–ADP) and 4rf9 (ddAK–arginine–ATPγS).

3.3. Active sites

To further reveal the allostery induced by ligand binding, ddAK was co-crystallized in the presence of Mg^{2+} with various ligands or their combinations, including the substrate arginine, the product ADP, the substrate analogue AMPPNP or ATPγS and a combination of arginine, ADP and NO_3^-, as in the so-called transition-state analogue complex (TSAC) that is commonly employed in structural studies of PKs. However, only crystals of ddAK–arginine and ddAK–ADP complexes could be grown, which were further soaked in different ligand solutions, eventually yielding crystals of the ddAK–arginine–ATPγS complex. All three ddAK complex crystals belonged to the same space group $P2_1$, with similar unit-cell parameters as the apo ddAK structure.

With two similar PK domains, ddAK possesses two active sites in D1 and D2, which, however, behave differently in ligand binding according to our complex structures. In both the ddAK–arginine and ddAK–ADP complexes only the active site of D1 harbours ligands; that of D2 is left empty. In the ddAK–arginine–ATPγS complex the situation is more

complicated. It is D1 that binds the substrate arginine and the substrate analogue ATPγS (Fig. 4*a*), which both show clear electron densities. However, the active site of D2 is not completely vacant but is occupied by some unrecognizable small molecules with a strong electron density around the arginine residues responsible for the binding of ATP phosphates (Supplementary Fig. S1). A possible interpretation of the bulk of the density could be the negatively charged phosphoryl groups of ATPγS captured and fixed by Arg585, though there could be other possibilities.

The asymmetry of ligand binding in ddAK is consistent with the asymmetry of its domain arrangement. In fact, both D1 and D2 of ddAK are catalytically functional, as has been demonstrated by previous biochemical studies (Tada & Suzuki, 2010). Moreover, according to a structure-based sequence alignment, the residues involved in ligand binding (indicated with blue triangles in Fig. 2) are highly conserved, except for the residues related to arginine binding, which, however, mostly contribute their main-chain atoms to binding the substrate arginine. Thus, the vacant active centre of D2 in the complex structures indicates an innate asymmetry of ddAK instead of a plausible dysfunction of D2.

At the active site of D1 in the ddAK–arginine–ATPγS complex, the substrate arginine is captured by the side chains of Tyr72 and Glu229 and the main-chain N atoms of Gly68, Val69 and Gly70, all *via* salt-bridge/hydrogen-bond interactions (Fig. 4*b*). Among these residues, only Glu229 belongs to the CsD and binds the guanidinium group of the substrate arginine, while its amino-acid group is fixed by the other residues which are all from the NsD. This could explain why the substrate arginine is so important in stabilizing ddAK: its binding pulls the NsD and the CsD together. For the binding of ATPγS, Arg128, Arg130, Arg233 and Arg284 constitute a positively charged region to attract the α-phosphoryl, β-phosphoryl and γ-thiophosphoryl groups of ATPγS, while Ser126 and Ser286 tether the nucleoside group *via* hydrogen bonds, as well as His288 *via* hydrophobic packing (Fig. 4*b*). Unexpectedly, the observed distances between the guanidino N atom of the substrate arginine and the two O^{γ} atoms of the γ-thiophosphoryl group are 2.5 and 2.6 Å, respectively, indicating the formation of hydrogen bonds.

In the ddAK–ADP complex structure, D1 exhibits a similar mode to bind the sole bound ligand, the product ADP, where the only difference is that the side chain of Arg233 protrudes

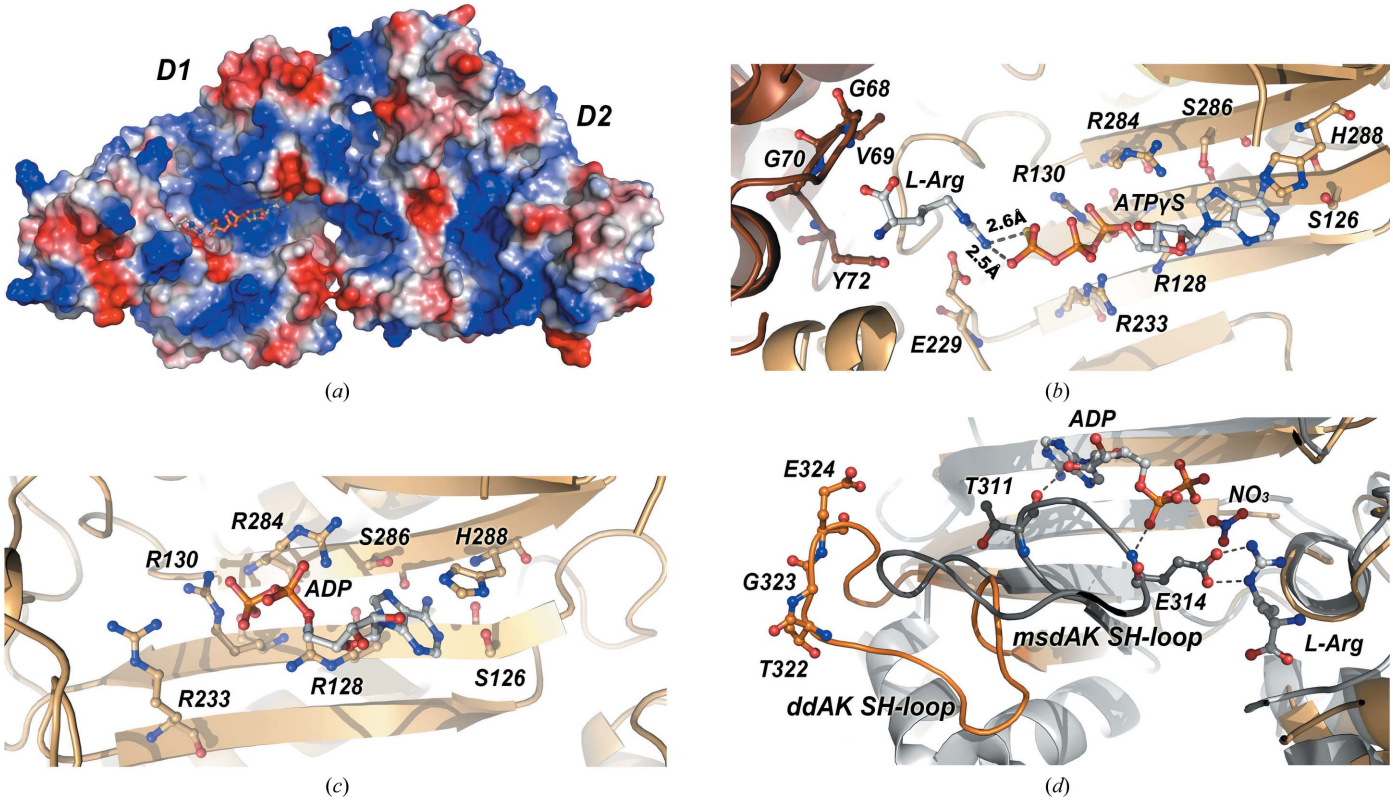


Figure 4
Active sites of ddAK. (*a*) Electrostatic potential surface calculated using *PyMOL* shown in the same view as in Fig. 3(*a*). Red indicates areas of negative potential and blue indicates areas of positive potential. The substrate analogue ATPγS and the substrate arginine bound to the active site of D1 are depicted in ball-and-stick representation coloured as in Fig. 3(*a*). (*b*) Cartoon representation of the active site of ddAK D1 in the arginine/ATPγS-bound complex. Both the ligands and the residues involved in ligand binding are depicted in ball-and-stick representation coloured as in Fig. 3(*a*) and are labelled accordingly. Hydrogen bonds between the substrate arginine and the substrate analogue ATPγS are depicted as black dashed lines. (*c*) Cartoon representation of the active site of ddAK D1 in the ADP-bound complex. Both ADP and the residues involved in ligand binding are depicted in ball-and-stick representation coloured as in Fig. 3(*a*) and are labelled accordingly. (*d*) SH-loops in a superposition of apo ddAK and msdAK TSAC. The colour scheme is the same as in Fig. 3(*c*). The SH-loops are highlighted with darker colours. Residues involved in ligand binding in msdAK and the extra Thr-Gly-Glu are depicted in ball-and-stick representation. Hydrogen bonds are depicted as black dashed lines.

in an opposite direction away from ADP, making no interaction with it (Fig. 4*c*). In the TSAC structure of msdAK, which also contains a bound ADP, Arg229 (the counterpart of Arg233 in ddAK) formed hydrogen bonds to both the β-phosphoryl group and the nitrate ion, suggesting its critical role in transfer of the γ-phosphoryl group (Yousef *et al.*, 2003; Zhou *et al.*, 2000). A possible reason that may cause this difference is that although Mg^{2+} was present in the co-crystallization conditions, we could not find any electron density around ADP that could be interpreted as Mg^{2+}.

It is worth noting that no apparent conformational changes are induced by ligand binding in our ddAK complex structures, according to the unexpectedly low r.m.s.d. values among the four ddAK structures (0.2–0.6 Å for D1 and 0.3–0.4 Å for D2). Actually, somewhat similar situations have been observed in previous crystallographic studies of single-domain PKs (Wu *et al.*, 2010; Bush *et al.*, 2011; López-Zavala *et al.*, 2013), indicating that a TSAC of PK is needed to reach the fully closed state, which we failed to approach. For the ddAK–arginine–ATPγS complex crystals obtained by soaking, the packing forces could be the restraints that prevent further allostery that might assist ligand binding to the other domain. Nevertheless, in the ddAK–arginine and ddAK–ADP complex crystals obtained by co-crystallization, no such kind of restraint will be in effect. Therefore, the ligand-binding asymmetry between D1 and D2 indicates that substrates of ddAK could have a higher affinity for D1 than for D2.

3.4. The SH-, AH- and GH-loops

As enzymes, PKs have two conformational states, the open state and the closed state, the latter of which was structurally revealed by the classical TSAC structure. As shown in the TSAC structure of msdAK (Zhou *et al.*, 1998), two loop regions play critical roles in the conformational change induced by ligand binding: loop 310–322 and loop 185–194. When msdAK binds its substrates, these two loops move towards each other to close the active site, such that the ATP substrate inside will not suffer from a potential wasteful hydrolytic loss of a γ-phosphoryl group (Yousef *et al.*, 2003; Zhou *et al.*, 2000). In the TSAC structure of msdAK, loop 310–322 is involved in the binding of both ADP (*via* the main-chain atoms of Glu314 and Thr311; Fig. 4*d*) and arginine (*via* the side chain of Glu314; Fig. 4*d*) and therefore will be referred to

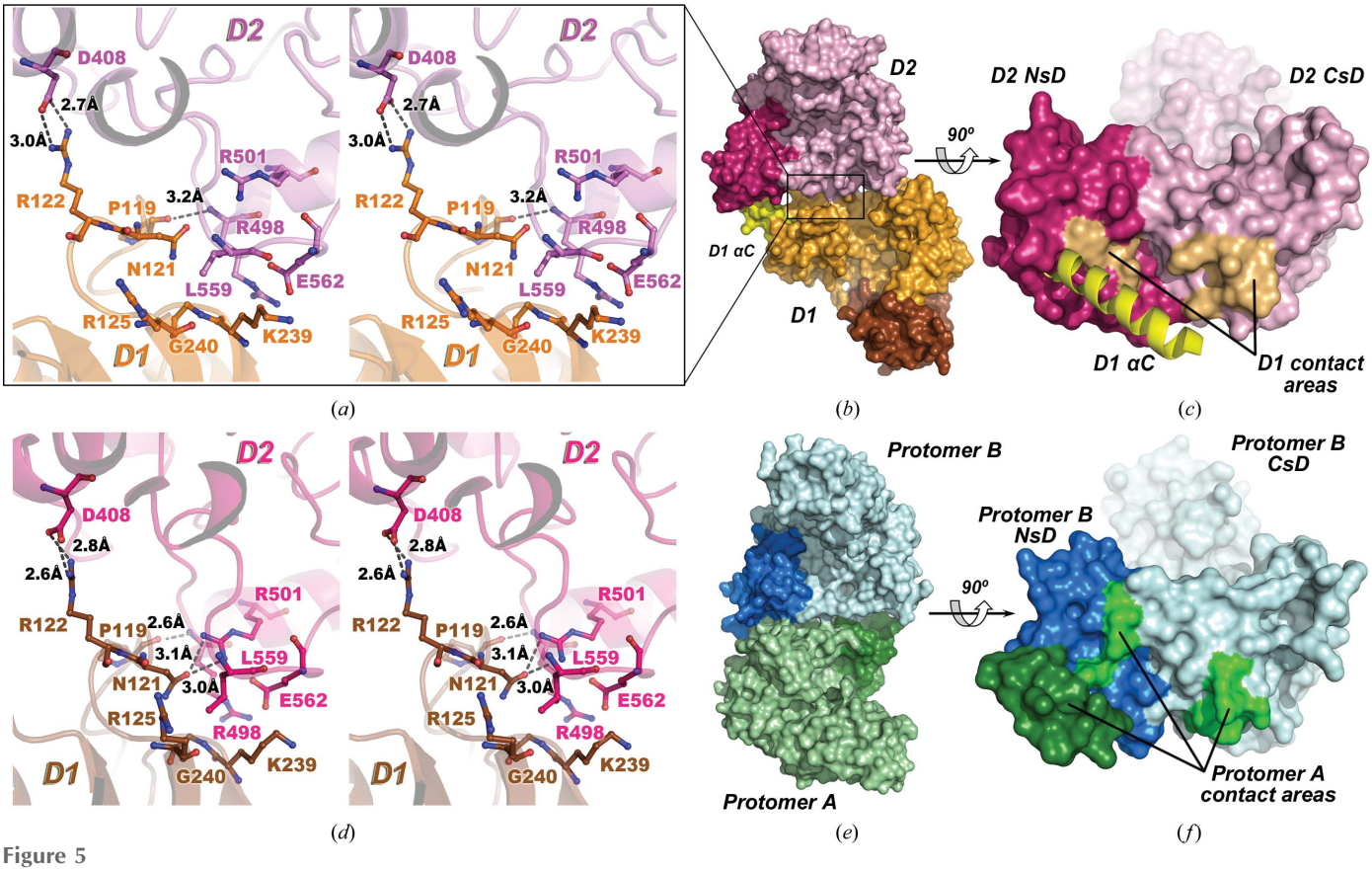


Figure 5
Comparison of the interdomain interface in ddAK and the interprotomer interface in dsdAK. (*a*) Wall-eyed stereo representation of the interface between ddAK D1 and D2 in the crystal structure. Residues involved in interdomain interactions are depicted in ball-and-stick representation coloured as in Fig. 3(*a*) and are labelled accordingly. Hydrogen bonds between D1 and D2 are depicted as black dashed lines. (*b*) Solvent-accessible surface of ddAK coloured as in Fig. 3(*a*). (*c*) Solvent-accessible surface of ddAK D2 coloured as in Fig. 3(*a*). The D1 contact areas are coloured gold. Helix αC is depicted in cartoon representation and is coloured yellow. (*d*) Wall-eyed stereo representation of the interface between ddAK D1 and D2 in an equilibrated MD simulation model. The colour and representation scheme are the same as in (*a*). (*e*) Solvent-accessible surface of dsdAK. Promoter *A* is coloured green and protomer *B* is coloured blue. In each promoter, the NsD is shaded darker than the CsD. (*f*) Solvent-accessible surface of protomer *B* of dsdAK. The protomer *A* contact areas are coloured light and dark green.

as the substrate-holding loop (SH-loop). Loop 185–194 only contributes its His185 to further fix ADP and will be named the ATP/ADP-holding loop (AH-loop). In addition, loop 61–72 that interacts with the amino-acid group of the substrate arginine will be referred to as the guanidinium-holding loop (GH-loop), which is also often called the 'specificity loop' owing to its role in guanidino substrate selectivity (Yousef *et al.*, 2002; Azzi *et al.*, 2004).

Considering that the residues involved in substrate binding are highly conserved between ddAK and msdAK (Fig. 2), it is reasonable to believe that the counterparts of the SH-, AH- and GH-loops in ddAK (residues 313–328, 189–198 and 65–76 in D1 and residues 665–680, 541–550 and 417–428 in D2, respectively) should adopt the same roles. Actually, our structures of both the ddAK–arginine and ddAK–arginine–ATPγS complexes indeed show that the GH-loop of ddAK D1 is deeply involved in the binding of the substrate arginine in the same way as in msdAK. However, in all three complex structures that we have determined the AH-loop does not contact any ligands, while the SH-loop is untraceable. Apparently, in the complex structures the ligand-bound D1 is not in a fully closed state as in the TSAC, leaving the SH-loop unfixed and flexible, consistent with the previously reported high flexibility of this loop based on NMR studies of single-domain PKs (Rivière *et al.*, 2012; Niu *et al.*, 2011; Davulcu *et al.*, 2011).

Another possible cause of its high flexibility could be that the SH-loops of both ddAK D1 and D2 are three residues (322–324 in D1 and 674–676 in D2) longer than that of msdAK (Fig. 1*a*, indicated by blue triangles). In spite of the lack of the SH-loop in our complex structures, the apo structure of ddAK shows the conformation of a continuous complete SH-loop, in which the three extra residues, Thr-Gly-Glu, are located on the protein surface as a small bulk, excluding their chance of interaction with the substrate (Fig. 4*d*). To understand whether the Thr-Gly-Glu bulk is involved in the catalysis by ddAK requires further enzymological studies.

3.5. Domain–domain interaction

Since the domains in ddAK adopt an asymmetric arrangement, it is reasonable that the interaction interfaces of the two domains are totally different. Owing to the L-shape of ddAK, the N-terminus of D1 is located at the opposite side of the D1–D2 interface, resulting in only one region of D1 CsD contributing to forming the interface, involving Pro119, Asn121, Arg122, Arg125, Lys239 and Gly240 (Figs. 5*a* and 5*b*). On the other side, two regions of D2 are involved: a single residue Asp408 of the D2 NsD and a region of the D2 CsD consisting of Arg498, Arg501, Leu559 and Glu562 (Figs. 5*a* and 5*b*). The area of the entire interdomain interface is approximately 450–500 Å^2, which is merely a third of the interprotomer interface area in dsdAK (1454 Å^2; Wu *et al.*, 2010) and only about 3% of the area of one PK domain.

Based on the crystal structures of ddAK in different states, there are three or four hydrogen bonds between D1 and D2, as well as hydrophobic interactions (Fig. 5*a*). The connection closest to the αC helix is the two parallel hydrogen bonds between the side chains of Arg122 and Asp408, which further strengthens the linker between the D1 and D2 NsDs as a supplement to αC. The third hydrogen bond is formed between the main-chain atoms of Pro119 and Arg498, next to a small patch of hydrophobic regions formed between Arg125, Gly240 and Lys239 from D1 and Arg498, Leu559 and Glu562 from D2. Although Leu559 is the only hydrophobic residue among those residues, the other large

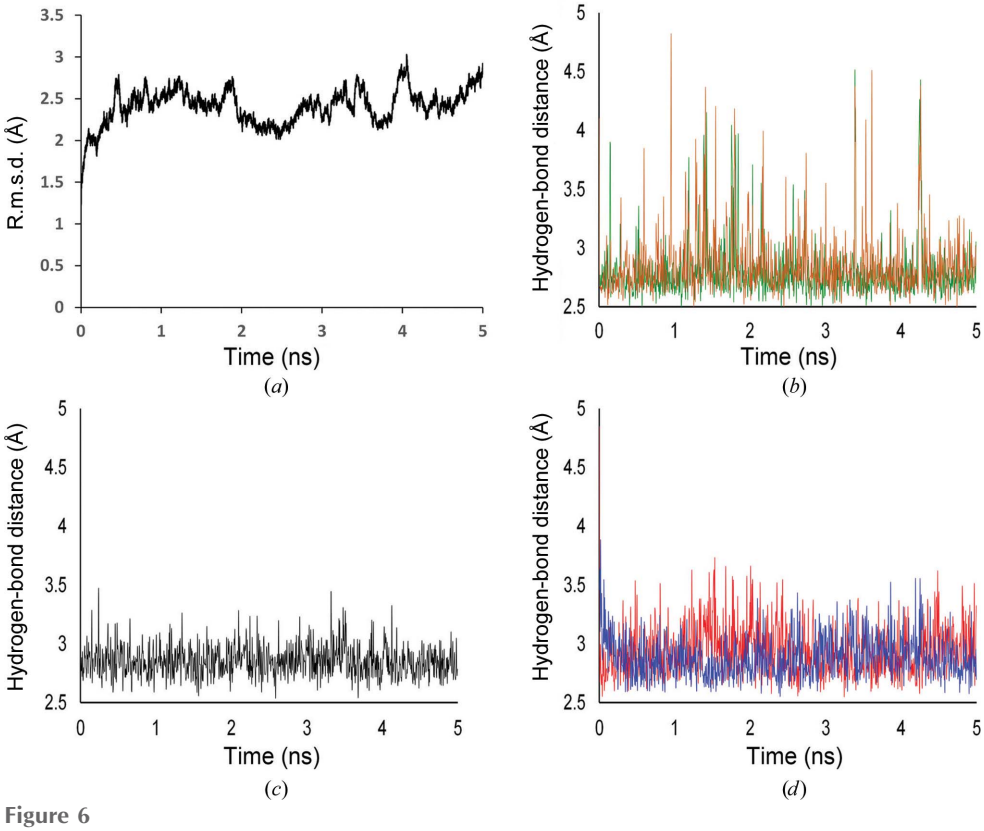


Figure 6
MD simulations and mutagenesis studies of ddAK. (*a*) The overall r.m.s.d. between ddAK models in an equilibrium simulation run and the crystal structure. The r.m.s.d. increases by about 2.5 Å after 0.5 ns equilibration and is maintained around this value during the 5 ns simulation run. (*b*) Hydrogen-bond distances between the side chains of Arg122 and Asp408. The curves representing the two hydrogen bonds are coloured orange and green, respectively. (*c*) Hydrogen-bond distances between the main-chain atoms of Pro119 and Arg498. (*d*) Hydrogen-bond distances between the side chains of Asn121 and Arg501. The curves representing the two hydrogen bonds are coloured red and blue, respectively.

residues contribute the alkyl portion of their side chains to the hydrophobic interaction. The fourth hydrogen bond bears clear instability, since it can be only found between Lys239 and Trp560 in a couple of chains in all of the ddAK structures that we have solved.

The crystal structures of ddAK clearly confirmed that interdomain interactions exist between D1 and D2 besides the connecting αC helix; however, considering that the interdomain interface is quite small compared with that of a dsdAK dimer, there is still the possibility that the interactions are dynamic and maintained merely by crystal-packing forces. In order to test whether or not the D1–D2 interactions are artificial, conventional MD simulations were performed on ddAK. In each 5 ns equilibrium simulation run, the overall r.m.s.d. increased at first and eventually stabilized at a value of around 2.5 Å (Fig. 6*a*). The interdomain hydrogen bonds observed in the crystal structures of ddAK were also monitored during the simulations. The two hydrogen bonds between the side chains of Arg122 and Asp408 are maintained during most of the period of simulation (Fig. 6*b*), but are not as stable as the hydrogen bond between the main-chain atoms of Pro119 and Arg498, which almost remains in the appropriate range for hydrogen-bond distances during the entire equilibrium trajectory of 5 ns (Fig. 6*c*). As a matter of fact, our MD simulation runs unexpectedly revealed another hydrogen-bond connection between D1 and D2, which is formed *via* the side chains of Asn121 and Arg501 (Fig. 5*d*) after about 0.1 ns of equilibrium and is even more stable than those between Arg122 and Asp408 (Fig. 6*d*). The results of MD simulation suggested that the interdomain connections of ddAK are not a result of crystal packing but reflect a natural feature of this polydomain PK.

We also compared the interdomain interface of ddAK with the interprotomer interface of dsdAK. The interface on the surface of D2 includes three D1 contact areas: (i) the N-terminal end of α18 and an adjacent loop region connecting α20 and β13, all in the CsD, (ii) the N-terminal end of α15 in the NsD and (iii) the NsD region connecting the αC helix (Fig. 5*c*). On the surface of a dsdAK protomer, the interprotomer interface also contains three contact areas, of which the first two (bright green in Fig. 5*f*) are the counterparts of (i) and (ii) described above and the last one (dark green in Fig. 5*f*) is the protein N-terminus that protrudes into the adjacent protomer to form tight binding (Figs. 5*e* and 5*f*). Generally, the interaction interface on D2 of ddAK closely resembles that of a dsdAK protomer, although the domain arrangements are quite different between a ddAK monomer and a dsdAK dimer.

4. Discussion

Dimeric enzymes usually adopt twofold-symmetric arrangements of their two assembled protomers, leading to an identical interface on both protomers. For polydomain enzymes, the domain arrangements are usually asymmetric, with a back-to-face configuration. In some cases, the two domains tightly fold together to form an interdomain active centre, as in amino-acid racemases, which, however, might not be quite appropriate to define as exact tandem-domain enzymes since their N- and C-termini are both located in one domain (Liu, Iwata, Kita *et al.*, 2002; Liu, Iwata, Yohda *et al.*, 2002). In other cases, such as in hexokinase I, the tandem domains are away from each other, with the only connection *via* a long linker, such as an α-helix, thus excluding the chance of interdomain allosteric propagation (Aleshin *et al.*, 1998). Another well known polydomain enzyme similar to polydomain PKs is the triple-domain luciferase from marine creatures (Liu *et al.*, 2004; Schultz *et al.*, 2005), the interdomain synergism and full-length structure of which have not yet been studied.

In the present study, the crystal structures that we have determined show that ddAK also adopts an asymmetric L-shaped domain arrangement with a long α-helix connecting the two domains, as well as other interactions between them. Further MD simulations revealed that although the interdomain interface is relatively small, the domain arrangement of ddAK is fairly stable, basically excluding the possibility that the arrangement is a conformational state restrained by crystal-packing forces. In future studies of ddAK, we will try to employ the small-angle X-ray scattering method to study its solution conformation, which will provide further evidence.

Owing to the asymmetry of the domain arrangement, ddAK possesses an asymmetric interdomain interface. Via the interface, CsD of D1 contacts both the NsD and the CsD of D2, but not *vice versa*. Surprisingly, the pattern of the interface on D2 closely resembles that of a protomer in dsdAK dimers, indicating that the allosteric effect propagated from D1 to D2 could mimic that which occurs in dsdAK dimers to some extent. It is well known that a large number of oligomeric PKs have interprotomer synergy, which, according to our assumptions, could also exist between the tandem domains of ddAK and be based on the interdomain interactions. This hypothesis needs to be critically assessed by future enzymological studies.

After the ddAKs had been identified, Suzuki and coworkers speculated that there are interdomain interactions between the tandem domains which will lead to functional cooperativity. However, owing to the lack of structural information prior to the present work, it is nearly impossible to examine the interdomain connections by mutagenesis studies. The only exception is a bold attempt to insert six Lys residues between Lys366 and Arg367, which dramatically reduced the activity of ddAK to 17% of that of the wild type (Tada & Suzuki, 2010), thus confirming the speculation about interdomain synergy, although the detailed mechanism remained unclear. With the atomic coordinates of full-length ddAK reported in the present work, we are able to identify the exact interdomain interface, which provides a structural basis for further mutagenesis studies to investigate the exact relationship between the asymmetrically arranged tandem domains and the previously reported interdomain synergy.

Acknowledgements

The gene for ddAK from *A. japonicas* was kindly provided by Professor Tomohiko Suzuki, Kochi University, Japan.

Supercomputing time was freely provided by the ScGrid of the Supercomputing Environment of the Chinese Academy of Sciences. We are very grateful to the staff of the Structural Biology Core Facility (Institute of Biophysics, Chinese Academy of Sciences) for their technical assistance, especially Ms Ya Wang, Mr Yi Han and Mr Jianhui Li. We are also grateful to Associate Professor Wenming Qin of National Center for Protein Science Shanghai for his assistance in synchrotron data collection. This work was supported by the '973 Program' of the Ministry of Science and Technology of China (Grant Nos. 2011CB911101 and 2012CB910304) and the National Natural Science Foundation of China (Project Nos. 31271491 and 31470792).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Aleshin, A. E., Zeng, C. B., Bourenkov, G. P., Bartunik, H. D., Fromm, H. J. & Honzatko, R. B. (1998). *Structure*, **6**, 39–50.
Andrews, L. D., Graham, J., Snider, M. J. & Fraga, D. (2008). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **150**, 312–319.
Azzi, A., Clark, S. A., Ellington, W. R. & Chapman, M. S. (2004). *Protein Sci.* **13**, 575–585.
Boivin, S., Kozak, S. & Meijers, R. (2013). *Protein Expr. Purif.* **91**, 192–206.
Bush, D. J., Kirillova, O., Clark, S. A., Davulcu, O., Fabiola, F., Xie, Q., Somasundaram, T., Ellington, W. R. & Chapman, M. S. (2011). *J. Biol. Chem.* **286**, 9338–9350.
Compaan, D. M. & Ellington, W. R. (2003). *J. Exp. Biol.* **206**, 1545–1556.
Darden, T., York, D. & Pedersen, L. (1993). *J. Chem. Phys.* **98**, 10089–10092.
Davulcu, O., Skalicky, J. J. & Chapman, M. S. (2011). *Biochemistry*, **50**, 4011–4018.
Ellington, W. R. (2001). *Annu. Rev. Physiol.* **63**, 289–325.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Griffiths, D. E., Morrison, J. F. & Ennor, A. H. (1957). *Biochem. J.* **65**, 153–162.
Hornemann, T., Rutishauser, D. & Wallimann, T. (2000). *Biochim. Biophys. Acta*, **1480**, 365–373.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Lim, K., Pullalarevu, S., Surabian, K. T., Howard, A., Suzuki, T., Moult, J. & Herzberg, O. (2010). *Biochemistry*, **49**, 2031–2041.
Liu, L. J., Iwata, K., Kita, A., Kawarabayasi, Y., Yohda, M. & Miki, K. (2002). *J. Mol. Biol.* **319**, 479–489.
Liu, L. J., Iwata, K., Yohda, M. & Miki, K. (2002). *FEBS Lett.* **528**, 114–118.
Liu, L. Y., Wilson, T. & Hastings, J. W. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 16555–16560.
López-Zavala, A. A., García-Orozco, K. D., Carrasco-Miranda, J. S., Sugich-Miranda, R., Velázquez-Contreras, E. F., Criscitiello, M. F., Brieba, L. G., Rudiño-Piñera, E. & Sotelo-Mundo, R. R. (2013). *J. Bioenerg. Biomembr.* **45**, 511–518.
Lynch, M. & Conery, J. S. (2000). *Science*, **290**, 1151–1155.
MacKerell, A. D., Banavali, N. & Foloppe, N. (2001). *Biopolymers*, **56**, 257–265.
MacKerell, A. D. *et al.* (1998). *J. Phys. Chem. B*, **102**, 3586–3616.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Morrison, J. F. & James, E. (1965). *Biochem. J.* **97**, 37–52.
Niu, X. G., Bruschweiler-Li, L., Davulcu, O., Skalicky, J. J., Brüschweiler, R. & Chapman, M. S. (2011). *J. Mol. Biol.* **405**, 479–496.
Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. (2005). *J. Comput. Chem.* **26**, 1781–1802.
Price, D. J. & Brooks, C. L. (2004). *J. Chem. Phys.* **121**, 10096–10103.
Rivière, G., Hologne, M., Marcillat, O. & Lancelin, J.-M. (2012). *FEBS J.* **279**, 2863–2875.
Schultz, L. W., Liu, L. Y., Cegielski, M. & Hastings, J. W. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 1378–1383.
Suzuki, T., Kawasaki, Y. & Furukohri, T. (1997). *Biochem. J.* **328**, 301–306.
Suzuki, T., Tomoyuki, T. & Uda, K. (2003). *FEBS Lett.* **533**, 95–98.
Tada, H., Nishimura, Y. & Suzuki, T. (2008). *Int. J. Biol. Macromol.* **42**, 46–51.
Tada, H. & Suzuki, T. (2010). *Int. J. Biol. Macromol.* **47**, 250–254.
Tombes, R. M. & Shapiro, B. M. (1985). *Cell*, **41**, 325–334.
Uda, K., Yamamoto, K., Iwasaki, N., Iwai, M., Fujikura, K., Ellington, W. R. & Suzuki, T. (2008). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **151**, 176–182.
Wu, X. A., Ye, S., Guo, S. Y., Yan, W. P., Bartlam, M. & Rao, Z. H. (2010). *FASEB J.* **24**, 242–252.
Yousef, M. S., Clark, S. A., Pruett, P. K., Somasundaram, T., Ellington, W. R. & Chapman, M. (2003). *Protein Sci.* **12**, 103–111.
Yousef, M. S., Fabiola, F., Gattis, J. L., Somasundaram, T. & Chapman, M. S. (2002). *Acta Cryst.* D**58**, 2009–2017.
Zhou, G. F., Ellington, W. R. & Chapman, M. S. (2000). *Biophys. J.* **78**, 1541–1550.
Zhou, G. F., Somasundaram, T., Blanc, E., Parthasarathy, G., Ellington, W. R. & Chapman, M. S. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 8449–8454.