Supporting Information

Elucidating dynamics of adenylate kinase from enzyme opening to ligand release

Kwangho Nam,[1,*] Abdul Raafik Arattu Thodika,[1] Christin Grundström,[2] Uwe H. Sauer,[2] Magnus Wolf-Watz[2]

[1]Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, USA
[2]Department of Chemistry, Umeå University, Umeå SE-90187, Sweden

Decomposition of the SMCV free energy profiles.

In this study, the free energy profile was determined using the thermodynamic integration as outlined in eq. (1). Then, the contribution of each collective variable (CV) to the total free energy values can be determined through the formula:

$$\Delta F(\alpha) = F(\alpha) - F(0) = \sum_i^N \Delta F_i(\alpha) \quad \text{(S1)}$$

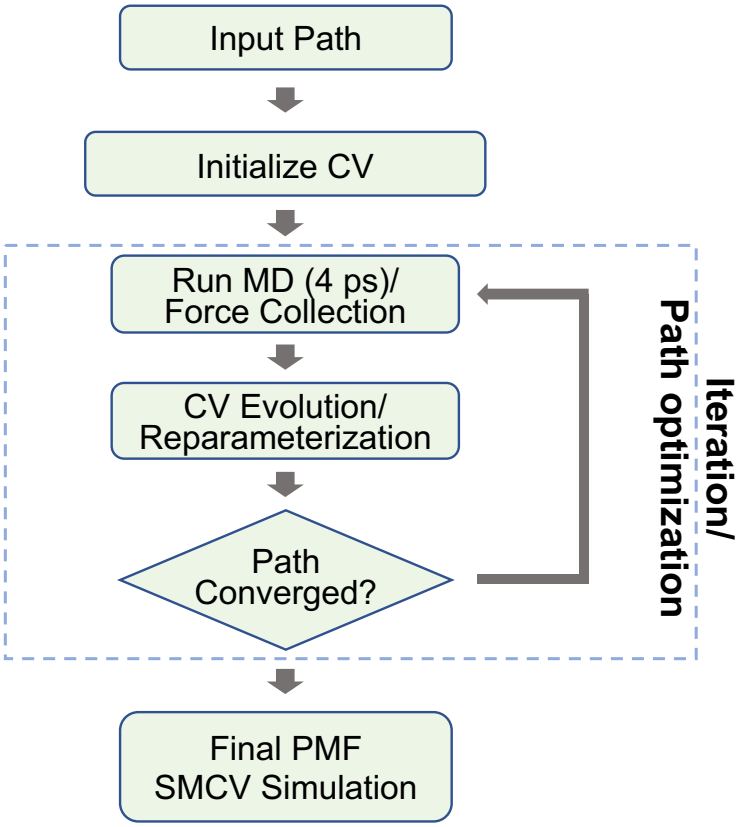
where N is the total number of CVs, and $\Delta F_i(\alpha)$ represents the free energy contribution of i-th CV. The $\Delta F_i(\alpha)$ can be computed by projecting the free energy gradient of the i-th CV onto the path, followed by the thermodynamic integration:

$$\begin{aligned} \Delta F_i(\alpha) &= \int_0^{\alpha} \frac{dF_i}{dc_i'} \cdot \frac{dc_i'}{d\alpha_i'} d\alpha' \\ &= \sum_{i=2}^{N} \frac{1}{2} \left(\frac{dF_i(\alpha_i)}{dc_i} + \frac{dF_i(\alpha_{i-i})}{dc_i} \right) (c_i(\alpha_i) - c_i(\alpha_{i-i})) \end{aligned} \quad \text{(S2)}$$

In the above equation, the second equation refers to the trapezoidal rule integration, where c_i refers to the i-th CV at the specified α_i value. Here, we assume the arc length of the path along α between 0 and 1 is 1.0, thus a normalized arc length, and the string images are evenly distributed along the path.

Table S1. Crystallographic data collection and refinement statistics.

	D158A AK variant with Ap5A (PDB ID: 8Q2B)
Data collection	Beam line ID23-2 ESRF, France
Wavelength (Å)	0.87293
Space group (Nr.)	P 1 21 1 ($P2_1$) (4)
Unit cell dimensions	
a, b, c (Å]	57.28, 77.97, 59.91
α, β, γ (°)	90.00, 93.61, 90.00
Molecules per asymmetric unit	2
Resolution range (outer shell) (Å)	47.45 – 1.76 (1.82 – 1.76)
Accepted reflections	361085 (36562)
Unique reflections	52176 (5114)
Completeness (%]	100.0 (99.9)
Multiplicity	6.9 (7.1)
Mean ($I/\sigma I$)	11.2 (1.82)
R_{merge} (I) (%)	10.9 (109.7) (all I+ and I-)
R_{meas} (I) (%)	11.8 (118.3) (all I+ & I-)
R_{pim} (I) (%)	4.5 (44.0) (all I+ & I-)
$CC_{1/2}$ (%)	99.8 (62.7)
Wilson B-factor	22.1
Refinement	
Resolution range (outer shell) (Å)	42.69 - 1.76 (1.79 – 1.76)
Reflections used for refinement	52150 (2592)
Reflections used for R_{free}	2662 (144)
R_{work} (%)	17.4 (26.3)
R_{free} (%)	20.9 (31.1)
Overall map CC (Fc, 2mFo-DFc) (%)	89.4
R. M. S. deviations	
Bond length (Å)	0.008
Bond angles (°)	0.99
Ramanchandran plot statistics	
Residues in favored region (%)	99.53
Residues in allowed region (%)	0.47
Residues in outlier region (%)	0.00
All-atom clash score	5.26
No. of non-hydrogen atoms	4004
Protein (non-hydrogen)	3338
Ap5A	114
MPO	13
SO_4	5
Solvent (H_2O) (non-hydrogen)	534
Average B-factors ($Å^2$):	29.0
Protein chains	26.6
Water molecules	36.1



Scheme S1. Workflow of the SMCV simulation. Starting from the initial path obtained from the unbiased opening simulation of AK, followed by the initialization of the CV values, the iterative path optimizations were carried out. First, a short (4ps) MD simulation with restraints on the CV positions was performed, during which the forces on each CV were evaluated. Then, the CV positions were evolved based on the averaged forces on them, followed by reparameterization of the CV positions to distribute them with equal spacing along the path. The updated CV positions were checked for convergence. Once the path was converged, the PMF simulations were run without updating the path.

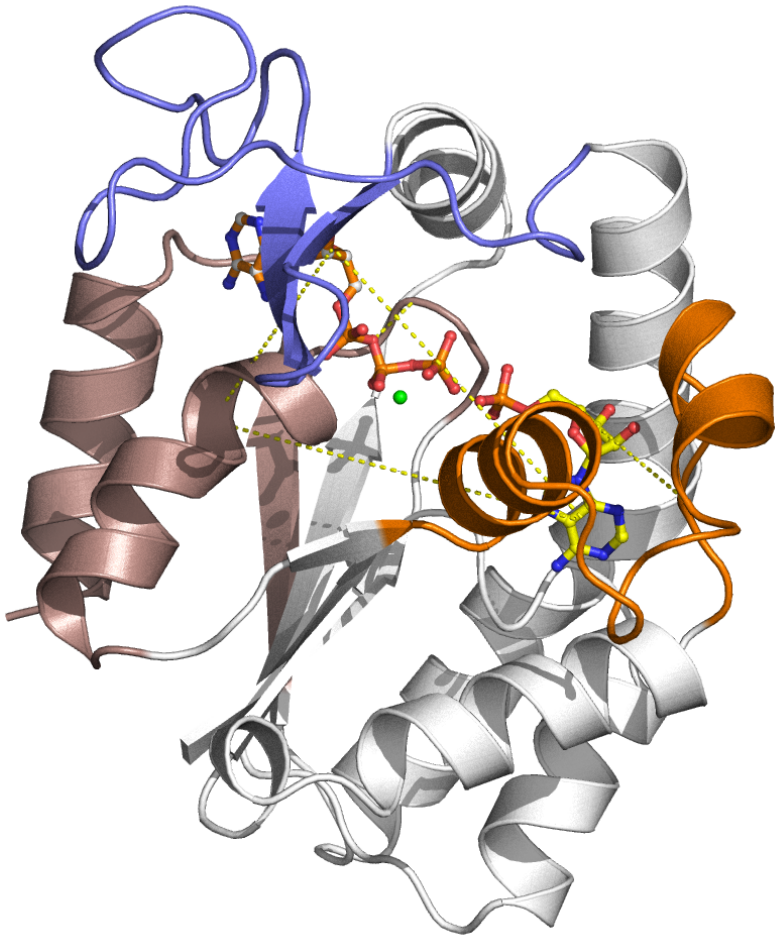


Figure S1. Definition of collective variables (CVs) for the close-to-open conformational change and ligand release. The residues used to define the center of mass (COM) of the ATP lid, AMP lid and CORE subdomains are shown in blue, orange and dark salmon, respectively, with their distances denoted by dashed yellow lines. The two dashed lines from ATP (orange ball and stick) and AMP (yellow ball and stick) illustrate the distances from the COM of ATP to COM of the backbone heavy atoms of AK residues 7 – 15 and 119 – 123 and from the COM of AMP to COM of the backbone heavy atoms of AK residues 30 – 35 and 84 – 88.

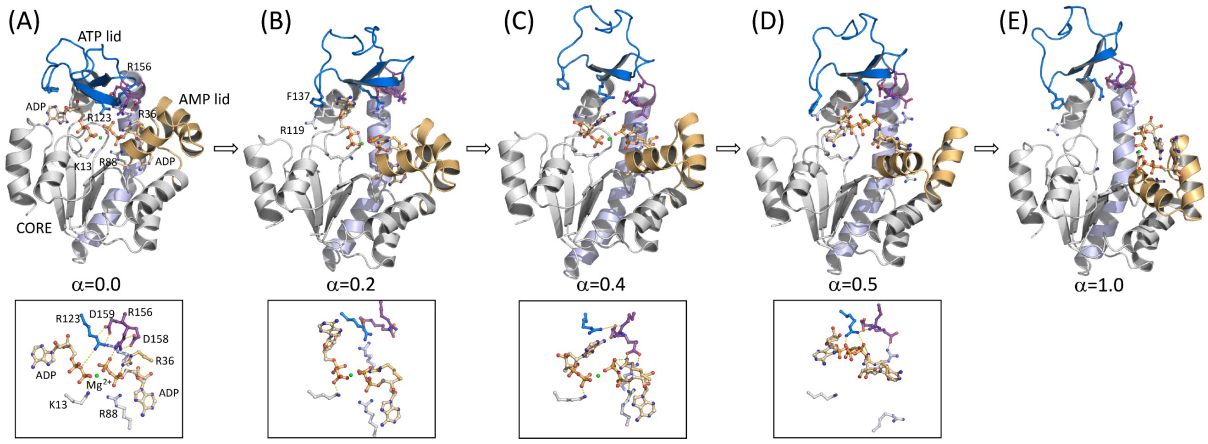


Figure S2. Representative snapshots of the enzyme and products (PS) along the SMCV pathway determined in **Figure 1A**: (**A**) α=0.0, (**B**) α=0.2, (**C**) α=0.4, (**D**) α=0.5 and (**E**) α=1.0. For (**A**) – (**D**), the close-up view of the products (2ADPs and Mg^{2+}) with residues that interact with them are shown in the bottom panel. For the reactant state (RS) system, the same snapshots are shown in **Figure 3**. In the figure, the ATP lid is shown in blue cartoon, the AMP lid in gold cartoon, the CORE subdomain in white cartoon, the α7 helix in light blue cartoon and the catalytic loop composed of residues 156 – 159 in purple cartoon, respectively. In each figure, key enzyme residues and products are shown in the ball-and-stick model. In the bottom panel, key interactions are indicated with dashed line.

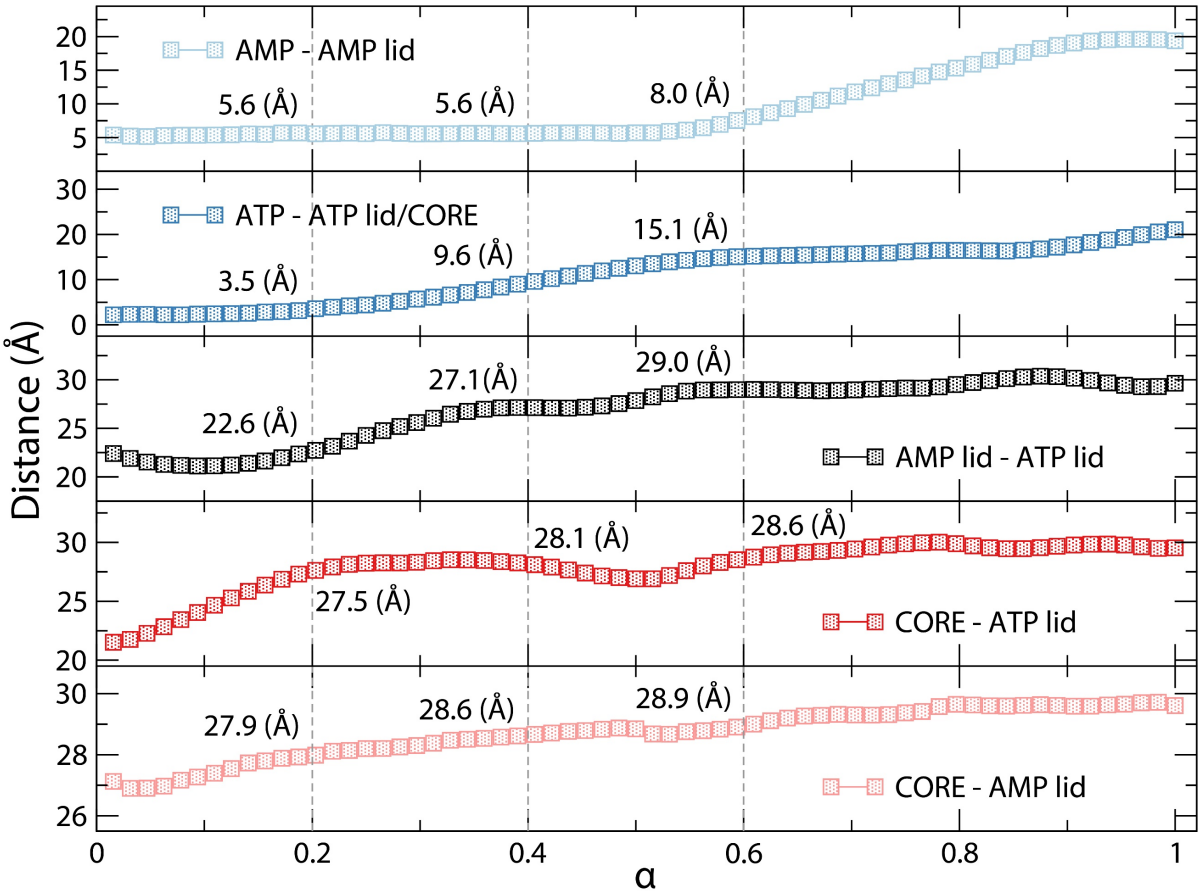


Figure S3. Changes of collective variables along the RS SMCV free energy profile provided in **Figure 2A**. At each indicated α value, the CV values are also shown.

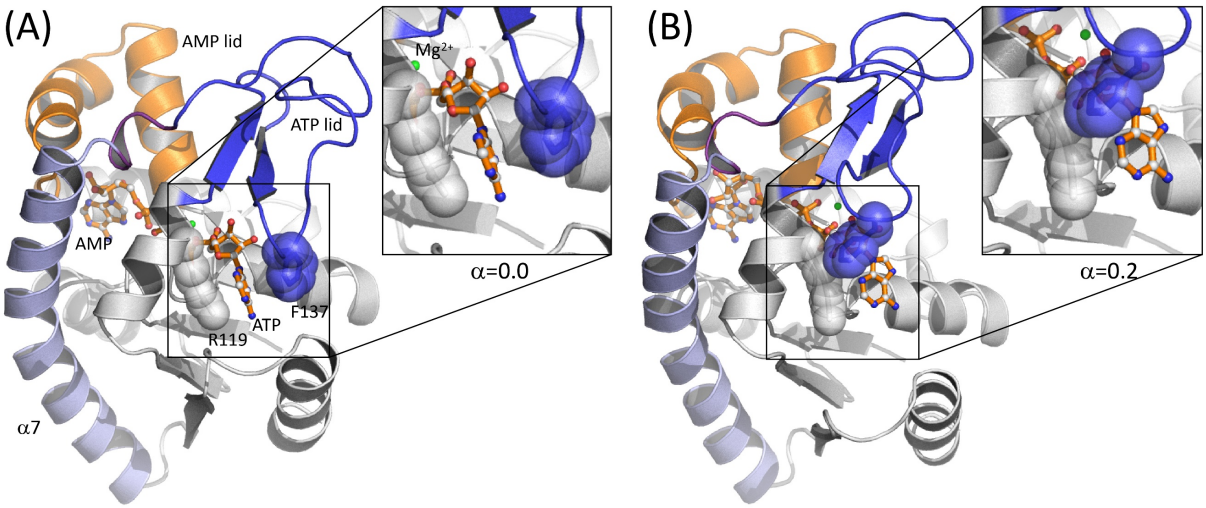


Figure S4. R119-ATP base-F137 stack at the base binding pocket of the enzyme at two different states: (**A**) α=0.0 and (**B**) α=0.2. The figure highlights the breakage of the stacking interaction at α=0.2, where the ATP base/ribose groups move out from their binding pocket allowing the upward movement of the ATP lid.

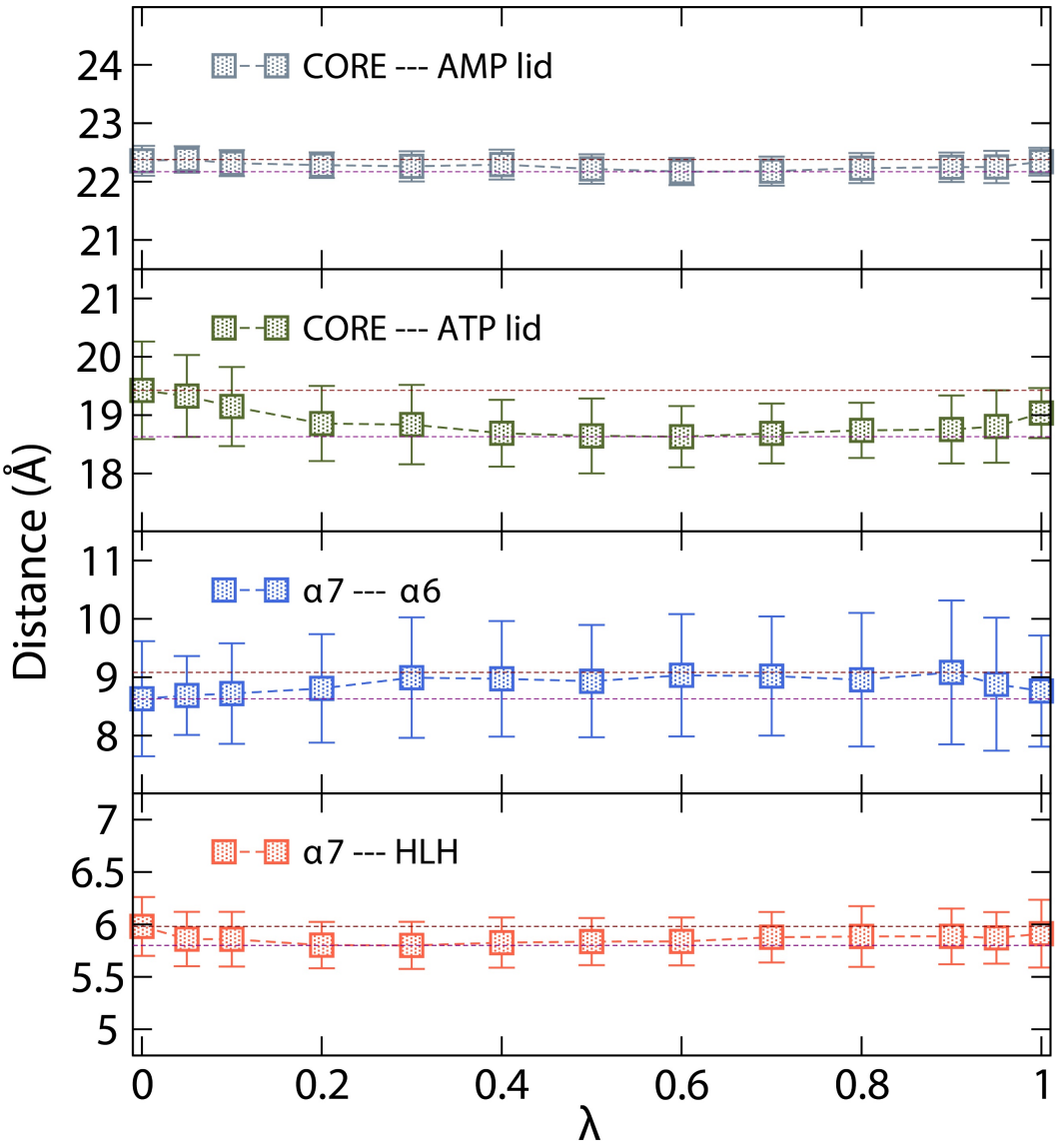


Figure S5. Changes of distances during the RS-to-PS alchemical transformation simulations between λ = 0 and 1.

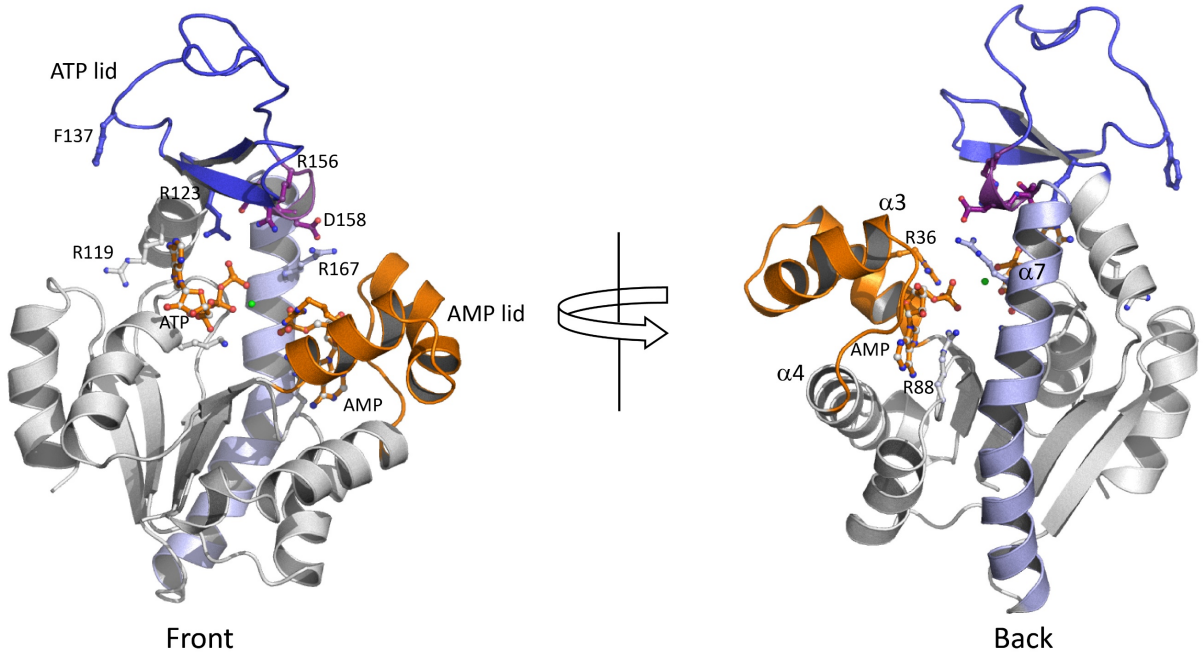


Figure S6. Representative snapshot of the reactant state (RS) system opening at α=0.3, where the ATP lid opens and the AMP lid remains closed. The ATP's base and ribose groups are completely moved out from their binding pockets while AMP remains bound in the AMP lid.

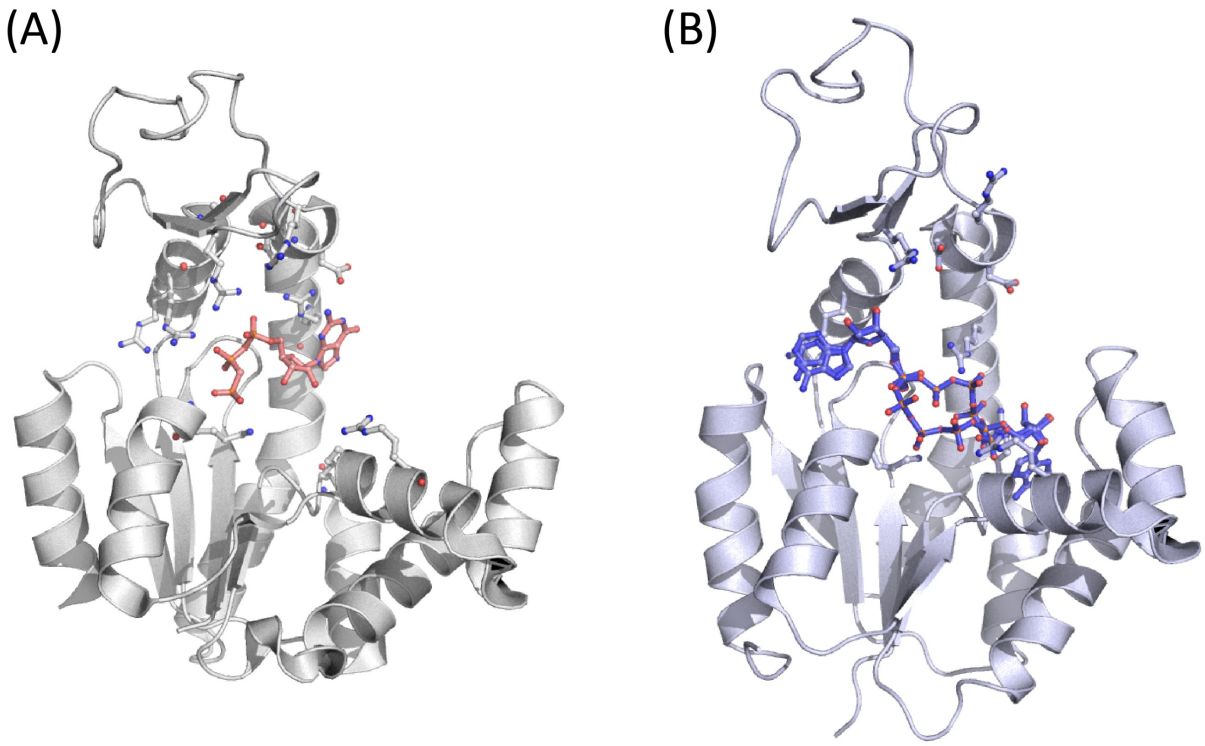


Figure S7. X-ray structures representing the non-specific contact complex of AK: (**A**) PDB ID: 6F7U for *E. coli* AK and (**B**) PDB ID: 6ZJE for human adenylate kinase 3 (hAK3). In (**A**), the bound ligand is GTP, thus non-specific substrate of AK and in (**B**) the bound ligand is Ap5A (in two orientations), while the natural substrates for hAK3 is GTP and AMP, thus non-specific inhibitor.