Supporting Information

A Fluorescent Probe Identifies Active Site Ligands of Inositol Pentakisphosphate 2-Kinase

Hayley Whitfield[†], Megan Gilmartin[†], Kendall Baker[†], Andrew M Riley[‡], Himali Y Godage[+], Barry VL Potter[‡ +], Andrew M Hemmings[†], Charles A Brearley[†*]

Protein purification and crystallisation methods…… **2**

Structure determination and refinement……..... **2**

Molecular docking calculations…….. **2**

Figure S1. Binding of 2-FAM-IP5 to AtIP5 2-K fitted to a 1-site binding model with Hill slope……. **4**

Figure S2. Predicted lowest energy binding poses of 2-FAM-IP5…… **5**

Figure S3. Double difference Fourier electron density maps for ligands bound to AtIP5 2-K……... **6**

Figure S4. Single difference ligand omit electron density maps……... **7**

Figure S5. AtIP5 2-K ligand interactions…….. **8**

Table S1. X-ray Data Collection and Refinement Statistics……. **9**

Table S2. Pairwise comparison of conformations of PDB entries…….. **12**

Table S3. Definition of AtIP5 2-K residues involved in binding substrate, nucleotide or inositide……. **12**

Table S4. Pairwise comparison of conformations……... **12**

Table S5. Ligand Validation Statistics…… **13**

Protein Purification and Crystallization

AtIP5 2-K was prepared according to the method of [1]. Crystals were grown using the sitting drop method at 16 °C, equilibrated against a reservoir containing 50 µl precipitant (18 % (w/v) PEG 3350, 0.1 M bis-tris propane pH 6.5 and 2 mM $MgCl_2$). Protein at a concentration of 5 or 7.5 mg/mL was incubated with 2 mM ligand and 2mM ADP prior to setting the crystallisation drops. The pre-incubated protein was subsequently mixed in a 1:1 ratio with the precipitant. Initially yielding only small crystals, microseeding was employed to provide crystals for X-ray data collection.

Structure Determination and Refinement

Single crystals were harvested into either 18 % (w/v) PEG 3350, 0.1 M bis-tris propane pH 6.5, 2 mM $MgCl_2$ with 25 % (v/v) ethylene glycol or into 35 % (w/v) PEG 3350, 0.1 M bis-tris propane pH 6.5, 2 mM $MgCl_2$ and flash frozen in liquid nitrogen. X-ray diffraction data was collected at the Diamond Light Source (Oxford) on beamlines I04-1 and I04. Data was indexed and processed using the automatic xia2 pipeline [2]. Reprocessing was carried out as necessary using Aimless [3] as part of the CCP4 programme suite [4]. Molecular replacement was performed using AutoMR in the Phenix suite [5] and structures refined using Phenix Refine [6]. Coot [7] was used to perform manual adjustment to models between rounds of automatic refinement using Phenix. Jligand [8] or eLBOW [9] were used to construct geometry restraint definitions for ligands. Simulated annealing ligand omit maps were calculated using Phenix to confirm the positioning of ligands [10]. X-ray data collection and refinement statistics can be found in Supporting Information Table 1. Difference Fourier and simulated annealing omit maps for all ligands are available as Supporting Information Tables 3 and 4, respectively. Ligand contacts to active site residues were rendered with LigPlot+ [11] for all complexes and shown in Supporting Information Figure 5. Molecular structures were rendered in PyMOL [12].

Structure factor amplitudes and refined atomic coordinates for the complexes with *myo*-Ins(1,3,4,5,6)P_5, *myo*-InsP_6, *neo*-InsP_6, D-*chiro*-InsP_6 and purpurogallin have been deposited in the RCSB Protein Data Bank with accession codes 6FL3, 6FJK, 6GFH, 6GFH and 6FL8, respectively.

Molecular docking calculations

Molecular docking experiments using torsionally-flexible 2-FAM-InsP_5 as ligand and the crystal structures of the three known conformers of AtIP5 2-K: open (PDB 4AXC), half-closed (PDB 4AXE) and closed (PDB 2XAM) as receptor were carried out using AutoDock Vina [13]. As the purified enzyme samples used in binding assays were treated to remove cofactor, the coordinates for ADP were removed from the receptor models prior to docking simulations. A D-2 axial model for InsP_5 (D-2 axial and five equatorial phosphates) was used as representative of the presumed predominant conformation of the ligand at the pH used in our assays. Atomic coordinates for the ligand were by merging ligand entries for InsP_6 and FAM taken from the PDB. The structures of ligand and receptor were formatted with AutoDockTools [14]. The chemical bonds involving the phosphate and FAM groups of 2-FAM-InsP_5 were defined as rotatable giving a total of

12 degrees of torsional freedom. Fixed (i.e. inflexible) enzyme models were used. A search space of 34 x 24 x 24 $Å^3$, centred on and encompassing the various enzyme active sites, was used. Productive binding modes were assigned as those poses found with binding energies within 0.5 $kcal.mol^{-1}$ of the global energy minimum pose for each docking calculation.

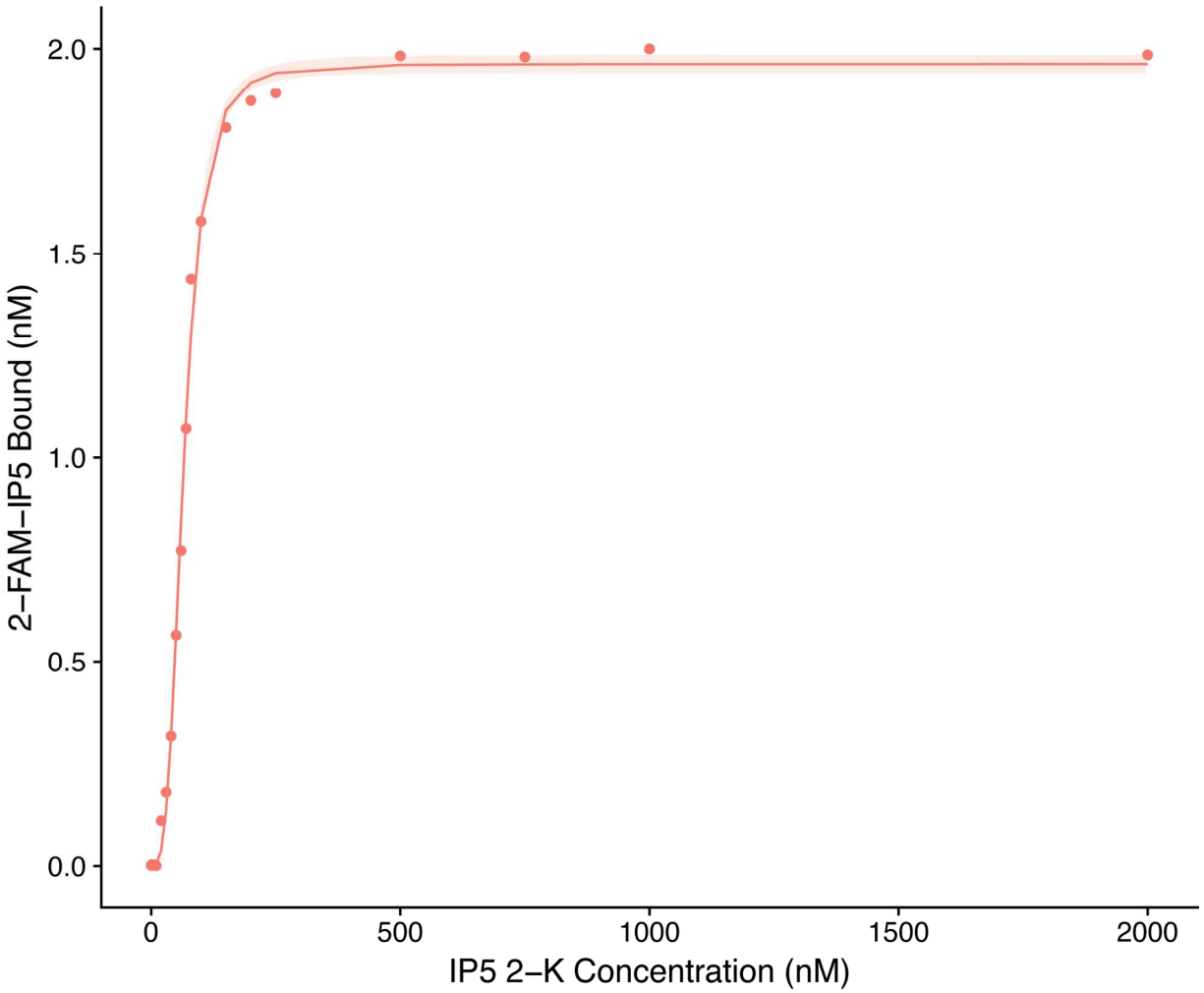

Figure S1. Binding of 2-FAM-IP_5 to AtIP5 2-K fitted to a 1-site binding model with Hill slope.

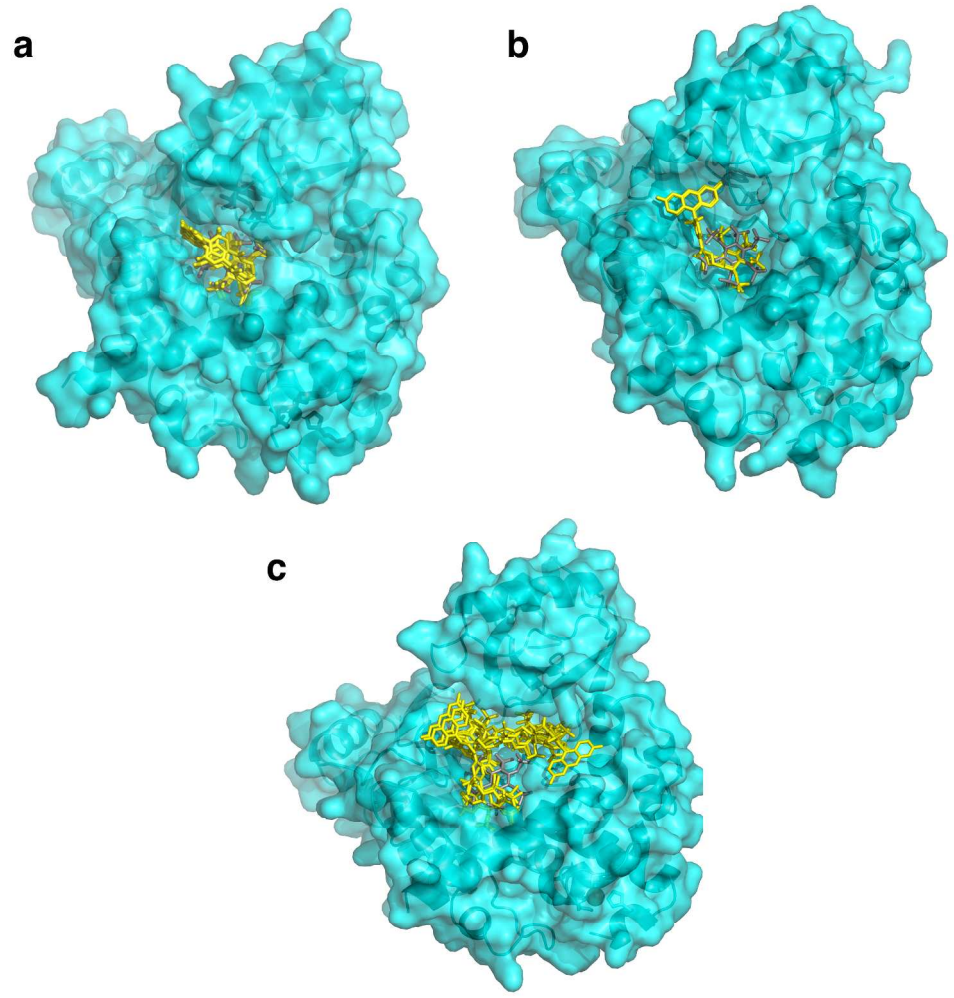

Figure S2. Predicted lowest energy binding poses of 2-FAM-IP_5 (yellow) to (a) the closed form (PDB 2XAM), (b) the half-closed form (PDB 4AXE) and (C) the open (apo) form (PDB 4AXC) of AtIPK5 2-K. Myo-$InsP_6$ is shown (grey).

Figure S3. Double difference Fourier electron density maps for ligands. $2m|F_o| - D|F_c|$ *electron densities (contoured at 1σ) for (a) myo-InsP*$_6$*, (b) myo-Ins(1,3,4,5,6)P*$_5$*, (c) neo-Ins(1,3,4,5,6)P*$_5$*, (d) D-chiro-InsP*$_6$ *and (e) purpurogallin ligands bound to AtIP5 2-K.*

Figure S4. $m|F_o| - D|F_c|$ *single difference ligand omit electron density maps for (a) myo-InsP_6 (3σ); (b) myo-Ins(1,3,4,5,6)P_5 (3σ); (c) neo-Ins(1,3,4,5,6)P_5 (3σ); (d) D-chiro-InsP_6 (2.5σ) and (e)purpurogallin (2.5σ) ligands of AtIP5 2-K. Contour levels are indicated in brackets.*

Figure S5. AtIP5 2-K ligand interactions. (a) myo-InsP_6:ADP; (b) myo-Ins(1,3,4,5,6)P_5:ATP; (c) neo-Ins(1,3,4,5,6)P_5:ADP; (d) D-chiro-InsP_6:ADP and (e) purpurogallin:ADP. Plots generated

1 **SI Tables**

2 **Table S1 X-ray Data Collection and Refinement Statistics**

PDB ID	6JFK	6FL3	6GFH	6GFG	6FL8
Ligands	*myo*-IP6:ADP	*myo*-IP5:ADP	*neo*-IP5:ATP	D-*chiro*-IP6:ADP	Purpurogallin: ADP
DATA COLLECTION					
Wavelength	0.9173	0.9173	0.9173	0.9795	0.9282
Space group	P 1	P 1	P 1	P 1	P 1
Unit cell a,b,c (Å) α,β,χ (˚)	59.8,60.3,84.0 88.5,89.2,63.4	59.0,59.0,83.4 88.0,89.6,63.7	59.6,60.6,84.5 87.8,88.4,63.0	59.6,60.8,82.4 89.5,88.3, 62.5	60.0,61.5,83.4 89.6,88.0,62.1
Resolution (□):	83.95-2.03 (2.08-2.03)	52.87-2.36 (2.42-2.36)	29.34 - 2.65 (2.78 - 2.65)	29.02 - 3.00 (3.18 - 3.00)	54.29-2.10 (2.15-2.10)
Number of unique reflections:	61224 (4591)	37982 (2806)	27766 (3658)	19896 (3147)	55908 (4199)
Completeness (%):	90 (90)	92 (94)	91 (90)	97 (94)	91 (92)
Multiplicity:	1.9 (2.0)	2.1 (2.2)	2.1 (2.1)	2.3 (2.1)	2.3 (2.3)
Rmerge	0.068 (0.219)	0.074 (0.408)	0.169 (0.705)	0.149 (0.569)	0.079 (0.675)
Rpim	0.070 (0.413)	0.069 (0.360)	0.110 (0.625)	0.131 (0.501)	0.066 (0.565)

Rmeas	0.099 (0.584)	0.103 (0.535)	0.169 (0.945)	0.199 (0.761)	0.104 (0.884)
<I/sigma(I)>	6.9 (3.5)	8.2 (1.9)	3.5 (1.2)	7.1 (1.8)	6.3 (1.4)
Wilson B factor ($Å^2$)	28.57	32.59	47.33	55.76	38.53
REFINEMENT					
Resolution range	83.95-2.03 (2.10-2.03)	52.83-2.36 (2.44-2.36)	29.34 - 2.65 (2.74 - 2.65)	29.02 - 3.00 (3.14 - 3.00)	54.29-2.10 (2.18-2.10)
Completeness (%)	90 (90)	92 (93)	91 (89)	97 (93)	91 (92)
Reflections used in refinement	61202 (6182)	37978 (3861)	27753 (2623)	19872 (1896)	55881 (5647)
Reflections used for R-free	3100 (319)	1898 (188)	1425 (97)	1062 (100)	2765 (297)
R-work (%)	17.8 (23.3)	15.5 (21.2)	19.9 (25.3)	19.7 (28.0)	18.2 (28.5)
R-free (%)	22.5 (29.2)	21.8 (28.2)	26.8 (29.1)	26.5 (38.3)	23.5 (34.3)
Number of non-hydrogen atoms	7266	7291	6890	6838	7108
Protein residues	840	845	830	840	845
RMS(bonds)	0.007	0.007	0.009	0.010	0.008
RMS(angles)	0.83	0.93	1.08	1.15	0.88
Ramachandran favoured (%)	97	96	89	85	96
Ramachandran outliers (%)	0.6	0.6	2.0	3.2	0.7

Rotamer outliers (%)	4	5.1	9.6	11	4.6 3
Average B-factor	40.56	38.42	52.50	52.44	51.00

Table S2 Pairwise comparison of conformations of PDB entries discussed in the text

	RMSD with number of residues in parentheses compared to *myo*-IP_6 ADP (6JFK)		
	'lid' residues 6-42 (N-lobe)+103-148 (B4 and N-II)	63-102 (N-I) +161-435 (C- lobe)	6-435
myo-IP_5 ADP (6FL3)	0.21 (83)	0.48 (299)	0.46 (400)
neo-IP_5 ATP (6GFH)	0.36 (83)	0.53 (298)	0.56 (399)
D-*chiro*-IP_6 ADP (6GFG)	0.50 (83)	0.42 (298)	0.48 (391)
Purpurogallin ADP (6FL8)	0.48 (83)	0.43 (299)	0.60 (397)
closed (PDB ID 4AQK)	0.35(83)	0.48 (296)	0.77 (398)
half-closed (PDB ID 4AXE)	1.14 (83)	0.86 (296)	1.29 (398)
4AXE vs 4AQK	1.11 (83)	0.78 (296)	2.81 (398)

Table S3 Definition of AtIP5 2-K residues involved in binding substrate, nucleotide or inositide

Binding site	Number residues	Residues
Substrate	28	R16, G19, G20, A21, N22, V24, V38, R40, R45, R130, L146, H149, E166, K168, K170, R192, H196, K200, N238, R241, D368, M372, I406, D407, S409, K411, R415, Y419
Nucleotide	18	R16, G19, G20, A21, N22, V24, V38, R40, L146, H149, E166, K168*, R241, M372, D368*, I406, D407, S409
Inositide	12	R45, R130, K168*, K170, R192, H196, K200, N238, D368*, K411, R415, Y419

'Substrate' residues cover both 'inositide' and 'nucleotide' binding sites.
*Residues that appear in both 'inositide' and 'nucleotide' sites.

Table S4 Pairwise comparison of conformations of PDB entries listed vs the *myo*-IP_6 ADP-liganded closed conformation of PDB 6FJK

	RMSD of listed PDB entries vs *myo*-IP_6 ADP-liganded protein (PDB 6JFK) monomer A		
	active site (28 residues)	Inositide site (12 residues)	Nucleotide site (18 residues)
myo-IP_5 ADP (6FL3)	0.26	0.23	0.16
neo-IP_5 ATP (6GFH)	0.43	0.41	0.31
D-*chiro*-IP_6 ADP (6GFG)	0.36	0.42	0.25
Purpurogallin ADP (6FL8)	0.50	0.52	0.35

Table S5 Ligand Validation Statistics

Structure	Ligand	Resolution/Å	Molecule	Occupancy	RSCC[1]	RSR[1]	B-factor[2]
6FJK	*myo*-IP6	2.02	A	0.97	0.98	0.12	27 (22)
			B	0.98	0.97	0.13	29 (30)
6FL3	*myo*-IP5	2.36	A	1.00	0.98	0.13	28 (26)
			B	1.00	0.99	0.12	26 (27)
6GFH	*neo*-IP5	2.65	A	0.91	0.95	0.14	54 (41)
			B	1.00	0.97	0.12	48 (38)
6GFG	*D-chiro*-IP6	3.00	A	0.86	0.91	0.20	70 (52)
			B	0.90	0.88	0.20	78 (49)
6FL8	Purpurogallin	2.10	A	0.90	0.75	0.21	57 (45)
			B	0.90	0.62	0.35	58 (45)

[1] Real Space Correlation Coefficient (RSCC) and Real Space R-factor (RSR) calculated using the wwPDB validation server at https://validate.wwpdb.org.
[2]The average B-factor calculated for non-hydrogen ligand atoms is followed in brackets by the average B-factor of protein atoms forming direct interactions with that ligand.

1. Gonzalez, B., Banos-Sanz, J. I., Villate, M., Brearley, C. A., and Sanz-Aparicio, J. (2010) Inositol 1,3,4,5,6-pentakisphosphate 2-kinase is a distant IPK member with a singular inositide binding site for axial 2-OH recognition, *Proc Natl Acad Sci U S A 107*, 9608-9613.
2. Winter, G. (2010) XIA2: an expert system for macromolecular crystallography data reduction., *J Appl Crystallogr 43*, 186-190.
3. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics, *Acta Crystallogr D Biol Crystallogr 67*, 282-292.
4. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr D Biol Crystallogr 67*, 235-242.
5. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.
6. Afonine, P. V., Grosse-Kunstleve, R. W., and Adams, P. D. (2005) A robust bulk-solvent correction and anisotropic scaling procedure, *Acta Crystallogr D Biol Crystallogr 61*, 850-855.
7. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.
8. Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A., and Murshudov, G. N. (2012) JLigand: a graphical tool for the CCP4 template-restraint library, *Acta Crystallogr D Biol Crystallogr 68*, 431-440.
9. Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Crystallogr D Biol Crystallogr 65*, 1074-1080.

10. Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., and Adams, P. D. (2017) Polder maps: improving OMIT maps by excluding bulk solvent, *Acta Crystallogr D Struct Biol 73*, 148-157.
11. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery, *J Chem Inf Model 51*, 2778-2786.
12. Schrodinger, LLC. (2015) The PyMOL Molecular Graphics System, Version 1.8.
13. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J Comput Chem 31*, 455-461.
14. Morris, G. M., Huey, R., and Olson, A. J. (2008) Using AutoDock for ligand-receptor docking, *Curr Protoc Bioinformatics Chapter 8*, Unit 8 14.