PNAS www.pnas.org

Supplementary Information for

The cryo-electron microscopy structure of the human CDK-activating kinase

Basil J. Greber, Juan M. Perez-Bertoldi, Kif Lim, Anthony T. Iavarone, Daniel B. Toso, Eva Nogales

Correspondence to: Eva Nogales, Basil J. Greber Email: ENogales@lbl.gov, basil.greber@icr.ac.uk

This PDF file includes:

Figures S1 to S9 Tables S1 and S2 SI References



Figure S1. CAK expression, purification, and specimen preparation. (A) Schematic of baculovirus-based expression constructs used for expression of full-length CAK and CAK-MAT1 Δ 219 in insect cells. (B) Elution profile from Superdex 200 10/300 GL column showing a monodisperse peak for full-length CAK. (C) SDS-PAGE gel of purified CAK. The bands for cyclin H and MAT1 overlap due to their similar molecular weights. (D) Representative electron micrograph showing CAK particles, acquired on a 200 kV-Talos Arctica electron microscope using a K3 direct electron detector.



Figure S2. Structure determination of the CAK-ATPyS complex. (**A**) Data processing scheme. Due to the large number of operations, only selected intermediate results are shown. The strategy was analogous to the processing of the CAK-THZ1 dataset (Fig. S8), for which detailed depictions of classifications and intermediate maps are shown. Inset: Projection orientation distribution of the final refinement. CAK is preferentially oriented but shows a continuous series of views along one rotation axis, which is sufficient to obtain complete sampling of Fourier space and reconstruction of an isotropic map (see C). (**B**) Fourier shell curves between independent cryo-EM half-maps (black) and between the refined coordinate model and the cryo-EM map (red). FSC thresholds for resolution estimation according to (1). (**C**) Validation of the cryo-EM map of the human CAK using the 3D FSC server (2) confirms that the map contains structural information at better than approx. 3.3-Å resolution in all directions (sphericity > 0.96). (**D**) Density for ATPyS in the active site of CDK7. Additional densities near the nucleotide (indicated by asterisks) may correspond to additional metal ions or unassigned ordered solvent molecules. (**E**) Section of the cryo-EM map with the fitted model.



Figure S3. Comparison to other CDK-cyclin structures. (A) CDK7, cyclin H, and MAT1 colored by residue from N-terminus (blue) to C-terminus (red). (B) Superposition of the CAK structure with the cell-cycle controlling CDK2-cyclin A complex (PDB ID 1FIN) (3). Cyclin H (light brown) in the CAK is rotated to the back slightly and shows a widened gap between the CDK and the cyclin compared to CDK2-cyclin A (yellow). MAT1 would not fit between the CDK and the cyclin in the cell-cycle CDK-like configuration. (C) Superposition of the CAK with the structure of the transcription-controlling complex pTEF-B (CDK9-cyclin T1; cyclin T1 in blue) (PDB ID 3BLG) (4). Cyclin T1 is rotated away from CDK9 even further than cyclin H is from CDK7. (D) Cyclin A (yellow) in the CDK2-cyclin A complex (PDB ID 1FIN) (3) has an α -helix that extends further from the location of the N-terminus of cyclin H (light brown). The C-terminal α -helix of MAT1 is localized similarly to this cyclin A helix and also interacts with CDK7.



Figure S4. Conservation of MAT1. (A-C) Residues that are conserved between human MAT1 and yeast Tfb3 are shown in violet (darker shade = identical; lighter shade = similar). (**D**) Sequence alignment of MAT1 and Tfb3, extracted from the multiple sequence alignment shown in Fig. S5. Residues are color-coded as in A-C.

2000 2000 1 10 20 30 40 P51948 Homo sapiens	
1 10 20 30 40 P51948 Homo sapiensMDDQCPRCKTKWRNBSLKLMVN.VCGHTLCESCVDLLDVRGAGNOP. P51949 Mus musculus	
P51948 Homo sapiens	
	. EC
	. EC
EIBVPO GallusMDEQGCPRCKTTKYRNPSLKLMVN.VCGHTLCESCVELLEVRCAGNCH.	. EC
A0A151 Alligator mississippiensisMDDQGCPRCKTTKYRNPSLKLMVN.VCGHTLCESCVELLEVRGAGNCQ.	. EC
P51951 Xenopus laevisMDDQGCPRCKTTKYRNPSLKLMVN.VCGHTLCESCVELLEVRCSGSCQ.	. EC
B8A5G8 Danio rerioMDDQGCPRCKTTKYRNPSLKLMVN.VCGHTLCESCVDMLFVRGSGNCV.	. Q <mark>C</mark>
Q7KPG8 Drosophila melanogasterMDDQACPRCKITKYRNPSLKLMVN.VCGHTLCESCVDLLPLKGSGACP.	. E C
Q7SHN1 Neurospora crassa MARPLDPLGFEDEMCPVCKSRKYLNPDIVFVFNPECYHSMCLNCANRLFNDGPNQCPH.	AG <mark>C</mark>
094684 Schizosaccharomyces pombe . MDDEGARKVDEKCELCQADRYLNENMKLLINPECYHKMCESCVDRIETTCPAQCPTI	PGC
<i>Q03290 Saccharomyces cerevisiae</i> MLMDEYEENKDM <mark>OD</mark> I <u>CKT</u> DR <u>Y</u> LS <u>P</u> DVKFLVNPECY <u>H</u> RI <u>CESC</u> VDRIBSLCPAQOPYI	KGC

		B2	α.3			α4
		>	ففففففف	LLLLL		000000000000000000000000000000000000000
	50	60	70	80	90	100
Homo sapiens	GTPLRKSN	FRVQLFE	DPTVDKEVE I	IRKKVLKIYN	KREEDEP.S	LREYNDFLEEVEEIVFNL
Mus musculus	GTPLRKSN	FRVQLFE	DPTVDK <mark>EV</mark> E]	IRKKVLKIY <mark>N</mark>	KREEDFP.S	LREYNDFLEEVEEIVFNL
Gallus gallus	DTPLRKSN	FRVQLFD	DPAVDK <mark>EV</mark> E 1	IRKKVLKIYN	KREDD <mark>F</mark> P.S	LSEYNDFLEEIEEIVFNL
Alligator mississippiensis	DTPLRKSN	FRVQLFE	DPTIDK <mark>EV</mark> DI	IRKKVLKIY <mark>N</mark>	KREDD <mark>F</mark> P.S	LREYNDFLEEVEEIVFNL
Xenopus laevis	DTPLRKSN	FKVQLFE	DPTIDK <mark>EV</mark> EI	IRKKILKIY <mark>N</mark>	KREEDFP.S	LREYNDFLEEIEEIVLNL
Danio rerio	DTPLRKSN	FRVQLFE	DPAIDK <mark>EV</mark> EI	IRKKVLKIY <mark>N</mark>	KREFDFS.S	LTEYNDYLEQVEDIVFNL
Drosophila melanogaster	MVP LRNN	FRVQLFE	DPMVEK <mark>EV</mark> DI	I <mark>R</mark> RRI LRDY <mark>N</mark>	KREEDFA.S	LAEYNDYLEEIEDIVYNL
Neurospora crassa	N K T <mark>L R </mark> K G	FRSAFFG	DLAVEREVDI	IRRRVAAVFN	QVEDDFE.T	LQDYNNYLQMVEDLTFEL
Schizosaccharomyces pombe	NKI <mark>LR</mark> KAK	FREQTFE	DAQIEREVDV	/ <mark>R</mark> K R I S R I F <mark>N</mark>	KGQQEFD.S	LQAYNDYLEEVEILTFNL
Saccharomyces cerevisiae	DKI <mark>LR</mark> KNK	FKTQIFD	DVEVEKEVDI	IRKRVFNVFN	KTIDDENGD	LVEYNKYLEEVEDIIYKL

			α.5				α6
	ll	leele.	. 2000	ووووووو	2000000000	0000000000000	200000000000000000000000000000000000000
	110		120	130	140	150	160
Homo sapiens	TNN	/DLDNT.	. K K K M I	EIYQKENKI	VIQKNKLKLT	REQEELEEALEV	/ERQENEQR <mark>R</mark> LFIQKE
Mus musculus	TNN	VDLENT.	. KKKMI	EIYQKENKI) V I Q K N K L K L T	REQEELEEALEV	/ERQEHEQRRLFIQKE
Gallus gallus	TNN	VDLENT.	. KRKMI	ELYQKDNKE	ZVIQKNKIKLT	REQEELEEALEV	/ E R Q E N E Q <mark>R R</mark> L L E Q K E
Alligator mississippiensis	TNN T	DLENT.	. KKKMI	EMYQKDNKI) V I Q K N K I K L T	REQEELEEALEV	/ E R Q E N E Q R R L L I Q K E
Xenopus laevis	TNN	VDLDNT.	. RRKII	DMYQKENKI) T <mark>I</mark> QR <mark>N</mark> KIKMT	REQEELEEALEN	4 E K H E N E Q R R L H L Q K E
Danio rerio	ANN	MDVEMT.	. KQKMI	EQYQRDNKI) V <mark>I</mark> Q R <mark>N</mark> K A K L T	REQEELEELLLG	EQQDSELRR LETLQE
Drosophila melanogaster	CNN	IDIIET.	.NKRII	EAYKRDNRE	ZVIQRNKTRVG	RDEYALEEMLEI	L <mark>EK</mark> VQEEA <mark>RR</mark> KELEEL
Neurospora crassa	VNG:	I <mark>d</mark> err <mark>r</mark> q	AEAQL	Q A W E A E H R A	D I E R N K K A G R	EADEISRVRLAA	A <mark>ER</mark> DAVRQ <mark>RR</mark> IEAIKE
Schizosaccharomyces pombe	IYK	IDVEET.	. EEKVI	KQYEKQNRI) S <mark>I</mark> AA <mark>N</mark> SARAA	AEARILAQNEII	LKRQKQEA <mark>R</mark> EAAIRE
Saccharomyces cerevisiae	DHG	IDVAKT.	. EEKLI	RTYEELNKO	LIMNNLERSR	TEIESFEOROKH	EKEMKLKK <mark>R</mark> LLEROI
=							

				α7		
	170	180	190	2000	2200	220
Homo sapiens	EQLQQIL	K R K N <mark>K Q</mark> A F L I	ELESS	DLPVALLLAQHKDR	STQLEMQLE <mark>K</mark> PH	KPVKPVTF
Mus musculus	EELQQAL	KRKN <mark>KQ</mark> AFLI	ELESS	DLPVALLLAQHKDR:	STQLEMQLE <mark>K</mark> PH	RSMKPVTF
Gallus gallus	EQLQQMM	K R K N <mark>K Q</mark> A L <mark>L</mark> I	ELESS	SLPASLLLAQHKDR:	S T Q L E M Q L E <mark>K</mark> P H	KPVKPVTF
Alligator mississippiensis	EQMQQII	KRKN <mark>KQ</mark> ALLI	ELESS	HLPASLLLAQHKDR:	STQLEMQLE <mark>K</mark> PH	KPVKPVTF
Xenopus laevis	EQFQQMM	KRKN <mark>KQ</mark> EL <mark>L</mark> I	Q <mark>L</mark> ETS	HLPASILLAQHKGK:	S V Q A E M Q V E <mark>K</mark> P H	RSFKTDTF
Danio rerio	EQRQLQA	KRKN <mark>KQ</mark> ALL	ELENS	KLPAAVLLAQHKDR.	A A H L E T Q I E <mark>K</mark> Q H	KQNVKPTNIF
Drosophila melanogaster	ENEHKKK	K A R D <mark>K Q</mark> A L I E	E <mark>L</mark> MYS(GKDAAQIVTEFAEK.	AEK.QREEE <mark>K</mark> QI	LPPPKPANEF
Neurospora crassa	AEAEKRE	RVRSREMELI	N <mark>L</mark> AKGTTAM:	TAEPATKVQLKR!	RGQVNRVAE <mark>T</mark> AS	SNPAIATTGM
Schizosaccharomyces pombe	HQKEKER	REQVEQQIIF	'D <mark>l</mark> atso	GKDPNKIIQLS	DSLKKQQI	ENIASSV
Saccharomyces cerevisiae	EEERMN	K E W T <mark>K K</mark> E I V M	IR <mark>L</mark> STT	CQDINETIEGVK	NTVKLKKS	SSARRKLEEL

	230	240	250	260	η1 η2 2.20 200	280
	239	249	239	200	2/9	209
Homo sapiens	STGIKMGQHIS.	.LAPIHK.LEE	ALYEYQPLQI	ETY <mark>GP</mark> HVPELEI	MLG.RLG <mark>YL</mark> NH	VRAASP
Mus musculus	STGIKMGQQIS.	.LAPIQK.LEE	ALYEYQPLQI	ETCGPQVPEQE:	LLG.RLG <mark>YL</mark> NH	VRAASP
Gallus gallus	STGIKMGQHIS.	.LAPIQK.LEE	TLYEYQPLQV	ETYGPPVPELE:	SLG.RLG <mark>YL</mark> NH	VRAASP
Alligator mississippiensis	STGIKMGQHIS.	.LAPIQK.LEE	ALYEYQPLQV	EMFGPQVPDPE	LLG.RLG <mark>YL</mark> NH	VRAASP
Xenopus laevis	STGIKKGHHIA.	.SVPVTK.IEE	ALYQYQPIHI	ETY <mark>GP</mark> QVPHIEI	MLG.RQG <mark>YL</mark> NH	VRAAAP
Danio rerio	ST <mark>GI</mark> MMGQTVS.	.LASVSR.VEE	VLYVYQPLYI	DTY <mark>GP</mark> PVPELD(QLG.RKG <mark>FL</mark> NH	VRAASL
Drosophila melanogaster	STGIKFGQTADF	SLLPVPKSEEG	PLFVYEPLVP	F S E <mark>G P</mark> A M P P T N I	EIV.SRG <mark>YI</mark> AH	IRAETP
Neurospora crassa	GTGASEVERLSI	RGLKEKPKAPA	PQGPYDPFGG	MDF <mark>AP</mark> SRYKLH(GGL.SHPLMEK	Y.RLDR
Schizosaccharomyces pombe	S.NISRSSSILI	SDVQQVAE	DTTPFSPLAG	EKD <mark>GS</mark> KYFSYSI	K N T Y Q D L <mark>Y L</mark> E K	V.SHEP
Saccharomyces cerevisiae	N. RVLKNNPYFN	ISNVNVQÑSRLK	DAVPFTPFNG	DRE <mark>AH</mark> PRFTLK(G S V Y N D P <mark>F I</mark> K D	L.EHRK

	n3		cr.8					
	0000	0000	0000000					
		290	300					
Homo sapiens	QDLAGGY	TSSLAC	HRALODAF	GLFWQPS				
Mus musculus	QDLAGGY	TSSLAC	HRALQDAFS	5 <mark>GL</mark> FWQPR				
Gallus gallus	QDLAGGY	TSSLAC	HRALQDAFS	5 <mark>GL</mark> FWFPS	3			
Alligator mississippiensis	QDLAGGY	TSSLAC	HRALQDAFS	5 <mark>GL</mark> FWHPS	3			
Xenopus laevis	QDLAGGY	VSSLAC	HRALQDAFS	5 <mark>GL</mark> FWQTH				
Danio rerio	QDQAGGY	TSGLAC	HRAIQDAFS	G <mark>GL</mark> FSL				
Drosophila melanogaster	QENAGGE	TSALAC	ERALQEAL	2 <mark>GL</mark> YYTAT	TGVVSGI			
Neurospora crassa	QHVAGGY	SFDDFA	ARALYEAF	A <mark>g l</mark> gvfve	DEKEVGS	GIGLSQET	SMAAGIAA	VVDADKMDL
Schizosaccharomyces pombe	GR.KCGF	RIEDCH	LRCLYEAFS	5 <mark>G</mark> IDYDLE	SLKKLEV	/AS		
Saccharomyces cerevisiae	EFIASGE	NTNYAY	ERVLTEAFI	4 <mark>GL</mark> GCVIS	SEEL			

Figure S5. Multiple sequence alignment of MAT1 from different animal and fungal species. Species names and accession numbers are indicated. The alignment was computed with Clustal Omega (5) and visualized using the ESPRIPT web server (6).



Figure S6. CDK7 conformation and CAK activity. (A) Extended conformation of the T-loop (teal) in our structure of the CAK. (B) The T-loop in the structure of isolated CDK7 (PDB ID 1UA2) is folded across the active site and would preclude substrate binding (7). (C, D) T-loop conformation in the well-studied CDK2 system for comparison (8, 9). In the active CDK2-cyclin A complex (C) (PDB ID 1JST), the T-loop is extended towards cyclin A (yellow). In inactive, isolated CDK2 (D), the T-loop (teal) is folded across the active site (PDB ID 1HCK). (E) Activity assay for the CAK complex. Phosphorylated (YSPTSPS)₃KKKK-biotin peptide was detected using a phospho-Ser5-specific antibody (see Methods). CDK7 was detected as a loading control. (F) Mass-spectrometric analysis of the activity assay reaction after 2 hours. Peptides without phosphorylation (expected mass of (YSPTSPS)₃KKKK-biotin = 2915 Da) and with one and two added phosphate groups (2996 Da and 3075 Da) could be detected.



Figure S7. Mass spectrometric analysis of THZ1 binding to CAK. CAK-MAT1 Δ 219 and CAK-MAT1 Δ 219-THZ1 complexes were analyzed using intact electrospray ionization mass spectrometry. Spectra for CDK7 from untreated CAK are plotted in orange and spectra for CDK7 from THZ1-treated CAK are plotted in blue. Cyclin H and MAT1 Δ 219 did not show any difference between the treated and untreated samples, confirming that THZ1 did not unspecifically modify cysteine residues under our assay conditions. A mass difference of 566 Da corresponds to the mass of THZ1 (chemical structure shown on the left). Mass differences of approx. 80 Da likely correspond to phosphorylations on CDK7.



Figure S8. Structure determination of CAK-THZ1. (**A**) Data processing scheme. All maps shown are un-sharpened RELION outputs. (**B**) FSC curves (half-map FSC black, model vs. map FSC red). (**C**) Local resolution of the cryo-EM reconstruction. (**D**) Analysis of the cryo-EM map using the 3D FSC server (2) indicates isotropic resolution (sphericity = 0.9).



Figure S9. Comparison to CDK12-cyclin K-THZ531 complex. (A) Superposition of CDK12cyclin K-THZ531 (PDB ID: 5ACB) (10) onto our CAK-THZ1 structure. (B) The structure of the nucleotide-binding pocket is highly conserved between CDK7 and CDK12, explaining the ability of structurally similar inhibitors to target the two kinases and the ability of THZ1 to bind to both of them. THZ1 is shown in purple, THZ531 in pink. Side chains near the ligands are shown in teal (CDK7) and light cyan (CDK12) with C_{α} atoms in grey. (C) The aromatic ring systems of THZ1 and THZ531 in the nucleotide-binding pockets of CDK7 and CDK12 are tightly bound in a very similar conformation, but the substituent protruding from the pocket can assume variable conformations, depending on the position of the covalently reacting cysteines, as evidenced by the two conformations of THZ531 observed in the crystal structure (10) and the weaker density of the acrylamide arm of THZ1 in our structure (Fig. 4B).

Dataset	CAK wild type	CAK-MAT1∆219-THZ1
Microscope	Talos Arctica	Talos Arctica
Stage type	Autoloader	Autoloader
Voltage (kV)	200	200
Detector	Gatan K3	Gatan K3
Acquisition mode	Super-resolution	Super-resolution
Physical pixel size (Å)	0.686	0.686
Defocus range (µm)	0.3-2.5	0.5-2.5
Electron exposure (e^{-}/A^2)	69	69
Reconstruction	EMD-22123	EMD-22131
Software	RELION 3.1	RELION 3.1
Particles picked	8,982,448	4,884,787
Particles final	136,859	31,198
Extraction box size (pixels)	384 x 384 x 384	384 x 384 x 384
Rescaled box size (pixels)	216 x 216 x 216	192 x 192 x 192
Final pixel size (Å)	1.220	1.372
Accuracy rotations (°)	1.00	1.13
Accuracy translations (Å)	0.35	0.43
Map resolution (Å)	2.8	3.3
Map resolution range	2.73-3.4	3.2-4.4
Map sharpening B-factor (Å ²)	-45	-40
Coordinate refinement		
Software	PHENIX	PHENIX
Refinement algorithm	REAL SPACE	REAL SPACE
Clipped box size (pixels)	150	134
Resolution cutoff (Å)	2.8	3.3
FSC _{model-vs-map} =0.5 (Å)	3.0	3.5
Model	PDB-6XBZ	PDB-6XD3
Number of residues	646	645
Protein	644	644
Ligand (ATPγS, THZ1, Mg ²⁺)	2	1
B-factors overall	49.7	61.3
Protein	49.7	61.3
Ligand (THZ1)	-	71.0
Ligand (ATPγS-Mg ²⁺)	81.1	-
R.M.S. deviations		
Bond lengths (Å)	0.008	0.006
Bond angles (°)	0.819	0.757
Validation		
Molprobity score	2.0	2.3
Molprobity clashscore	12.9	14.0
Rotamer outliers (%)	0.0	0.2
C _β deviations (%)	0	0
Ramachandran plot		
Favored (%)	94.8	93.8
Allowed (%)	5.2	6.2
Outliers (%)	0.0	0.0

 Table S1. Cryo-EM data collection, 3D reconstruction, and refinement statistics.

Protein	Chain	Length (aa)	Residues built	Ligand	Comments
		C	AK		
MAT1	Н	309	244-308		
Cyclin H	I	323	1-38, 42-284		
CDK7	J	346	10-45, 51-312	ATPγS-Mg²+, residues 400, 401	S164- phosphate
		CAK	K-THZ1		
MAT1	Н	309	244-308		Residues 1-219 deleted
Cyclin H	Ι	323	1-38, 42-284		
CDK7	J	346	10-45, 51-312	THZ1, residue 401	S164- phosphate

Table S2. Components of the refined atomic models.

SI References

- 1. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745 (2003).
- 2. Y. Z. Tan *et al.*, Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Meth* **14**, 793-796 (2017).
- 3. P. D. Jeffrey *et al.*, Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376**, 313-320 (1995).
- 4. S. Baumli *et al.*, The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO J* 27, 1907-1918 (2008).
- 5. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539 (2011).
- 6. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324 (2014).
- 7. G. Lolli, E. D. Lowe, N. R. Brown, L. N. Johnson, The crystal structure of human CDK7 and its protein recognition properties. *Structure* **12**, 2067-2079 (2004).
- 8. A. A. Russo, P. D. Jeffrey, N. P. Pavletich, Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol* **3**, 696-700 (1996).
- 9. U. Schulze-Gahmen, H. L. De Bondt, S. H. Kim, High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design. *J Med Chem* **39**, 4540-4546 (1996).
- 10. T. Zhang *et al.*, Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat Chem Biol* **12**, 876-884 (2016).