

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Oliver, MR;Horne, CR;Shrestha, S;Keown, JR;Liang, LY;Young, SN;Sandow, JJ;Webb, AI;Goldstone, DC;Lucet, IS;Kannan, N;Metcalf, P;Murphy, JM

Title:

Granulovirus PK-1 kinase activity relies on a side-to-side dimerization mode centered on the regulatory  $\alpha$ C helix

Date:

2021-12-01

Citation:

Oliver, M. R., Horne, C. R., Shrestha, S., Keown, J. R., Liang, L. Y., Young, S. N., Sandow, J. J., Webb, A. I., Goldstone, D. C., Lucet, I. S., Kannan, N., Metcalf, P. & Murphy, J. M. (2021). Granulovirus PK-1 kinase activity relies on a side-to-side dimerization mode centered on the regulatory  $\alpha$ C helix. *Nature Communications*, 12 (1), <https://doi.org/10.1038/s41467-021-21191-7>.









Persistent Link:

<https://hdl.handle.net/11343/273019>

License:

CC BY

# Granulovirus PK-1 kinase activity relies on a side-to-side dimerization mode centered on the regulatory $\alpha$ C helix

Michael R. Oliver<sup>1,6,8</sup>, Christopher R. Horne <sup>2,3,8</sup>, Safal Shrestha<sup>4</sup>, Jeremy R. Keown <sup>1,7</sup>, Lung-Yu Liang<sup>2,3</sup>, Samuel N. Young<sup>2</sup>, Jarrod J. Sandow <sup>2,3</sup>, Andrew I. Webb <sup>2,3</sup>, David C. Goldstone <sup>1</sup>, Isabelle S. Lucet <sup>2,3</sup>, Natarajan Kannan<sup>4,5</sup>, Peter Metcalf <sup>1</sup>✉ & James M. Murphy <sup>2,3</sup>✉

The life cycle of *Baculoviridae* family insect viruses depends on the viral protein kinase, PK-1, to phosphorylate the regulatory protein, p6.9, to induce baculoviral genome release. Here, we report the crystal structure of *Cydia pomonella* granulovirus PK-1, which, owing to its likely ancestral origin among host cell AGC kinases, exhibits a eukaryotic protein kinase fold. PK-1 occurs as a rigid dimer, where an antiparallel arrangement of the  $\alpha$ C helices at the dimer core stabilizes PK-1 in a closed, active conformation. Dimerization is facilitated by C-lobe:C-lobe and N-lobe:N-lobe interactions between protomers, including the domain-swapping of an N-terminal helix that crowns a contiguous  $\beta$ -sheet formed by the two N-lobes. PK-1 retains a dimeric conformation in solution, which is crucial for catalytic activity. Our studies raise the prospect that parallel, side-to-side dimeric arrangements that lock kinase domains in a catalytically-active conformation could function more broadly as a regulatory mechanism among eukaryotic protein kinases.

<sup>1</sup>School of Biological Sciences, University of Auckland, Auckland, New Zealand. <sup>2</sup>Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia. <sup>3</sup>Department of Medical Biology, University of Melbourne, Parkville, VIC, Australia. <sup>4</sup>Institute of Bioinformatics, University of Georgia, Athens, GA, USA. <sup>5</sup>Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA. <sup>6</sup>Present address: School of Biological Sciences, University of Edinburgh, Edinburgh, UK. <sup>7</sup>Present address: Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>8</sup>These authors contributed equally: Michael R. Oliver, Christopher R. Horne. ✉email: [peter.metcalf@auckland.ac.nz](mailto:peter.metcalf@auckland.ac.nz); [jamesm@wehi.edu.au](mailto:jamesm@wehi.edu.au)

**B**aculoviridae are a widespread family of large DNA viruses that specifically infect the orders of *Lepidoptera*, *Hymenoptera* and *Diptera* holometabolic insects<sup>1,2</sup>. Following primary infection and localization to the host cell nucleus, the initiation of baculoviral gene transcription, DNA replication and nucleocapsid assembly occurs within an electron-dense structure called the virogenic stroma<sup>2,3</sup>. A key modulator of this process is a highly-conserved, small, positively-charged DNA-binding protein named p6.9, which is expressed by all members of the *Baculoviridae* family<sup>4,5</sup>. p6.9 shares structural similarity to the protamines that are synthesised late in spermiogenesis to replace histones and condense the spermatid genome into a genetically inactive state<sup>6,7</sup>. These similarities include an arginine-rich sequence that confers the ability to bind DNA, neutralize the negative charge of the phosphate backbone and condense DNA into a highly compact chromatin-like structure<sup>8</sup>. In baculoviruses, this p6.9-mediated condensation is required for the encapsulation of the viral genome inside the nucleocapsid core<sup>9</sup>. However, much like protamine, the p6.9–DNA interaction can be attenuated by post-translational phosphorylation, leading to the release of baculoviral DNA from the nucleocapsid, triggering a cascade of early to very-late gene transcription<sup>10–13</sup>. Therefore, both the encapsulation and release of the baculoviral genome is regulated by the phosphorylation status of p6.9, thus supporting a pivotal role for post-translational modifications in the baculovirus life cycle.

Phosphorylation of p6.9 is mediated by protein kinase-1 (PK-1), a protein kinase expressed within the nucleocapsid of lepidopteran-infecting baculoviruses<sup>14–17</sup>. Deletion of PK-1 compromises encapsulation of the baculoviral genome in nucleocapsid assembly<sup>16</sup>, and mutation of 7 PK-1 substrate Ser/Thr residues in p6.9 markedly reduced the hyper-expression of very-late genes<sup>17</sup>, consistent with PK-1 serving a crucial role in the regulation of viral proliferation. PK-1 primarily phosphorylates the N-terminal region of p6.9 and has been reported to promote phosphorylation of arginine residues present within this region<sup>11,17</sup>, although the precise molecular mechanism is currently unclear.

Here, we present the crystal structure of *Cydia pomonella* granulovirus PK-1: a serine/threonine protein kinase that, despite being encoded by the baculoviral genome, shares the fold of a eukaryotic protein kinase. Phylogenetically, PK-1 appears to have originated from its host's genome, with the greatest similarities to the AGC kinase domain of ribosomal protein S6 kinase A5 (RPS6KA5) homologues in arthropods. The acquisition of an atypical N-terminal helix, which is peculiar to PK-1 orthologs in granuloviruses and some nucleopolyhedroviruses, facilitates constitutive dimerization via domain swapping between protomers, and is augmented by a contiguous  $\beta$ -sheet between the  $\beta$ 1– $\beta$ 5 strands of each protomer, and interactions between the  $\alpha$ C helices via an antiparallel arrangement. Biochemical studies support the idea that this dimerization mode locks the PK-1 dimer into a catalytically-active conformation and that dimerization is crucial to its catalytic activity. PK-1 harbours ATP-binding and catalytic residues typical of classical eukaryotic protein kinases within its active site, although in contrast to conventional protein kinases, the locked active conformation reduces the dependence on the N-lobe  $\beta$ 3-strand Lys and the Gly-rich loop for ATP-binding. The unusual parallel, side-to-side arrangement of kinase domains within the dimer, characterized by the N-lobe:N-lobe and C-lobe:C-lobe interactions revealed by our crystal structure, positions each active site facing in opposing directions and primed for substrate engagement. These findings suggest that stabilizing a catalytically-active conformation via a parallel side-to-side dimerization mode could represent a mechanism used by other protein kinases to regulate their catalytic activities.

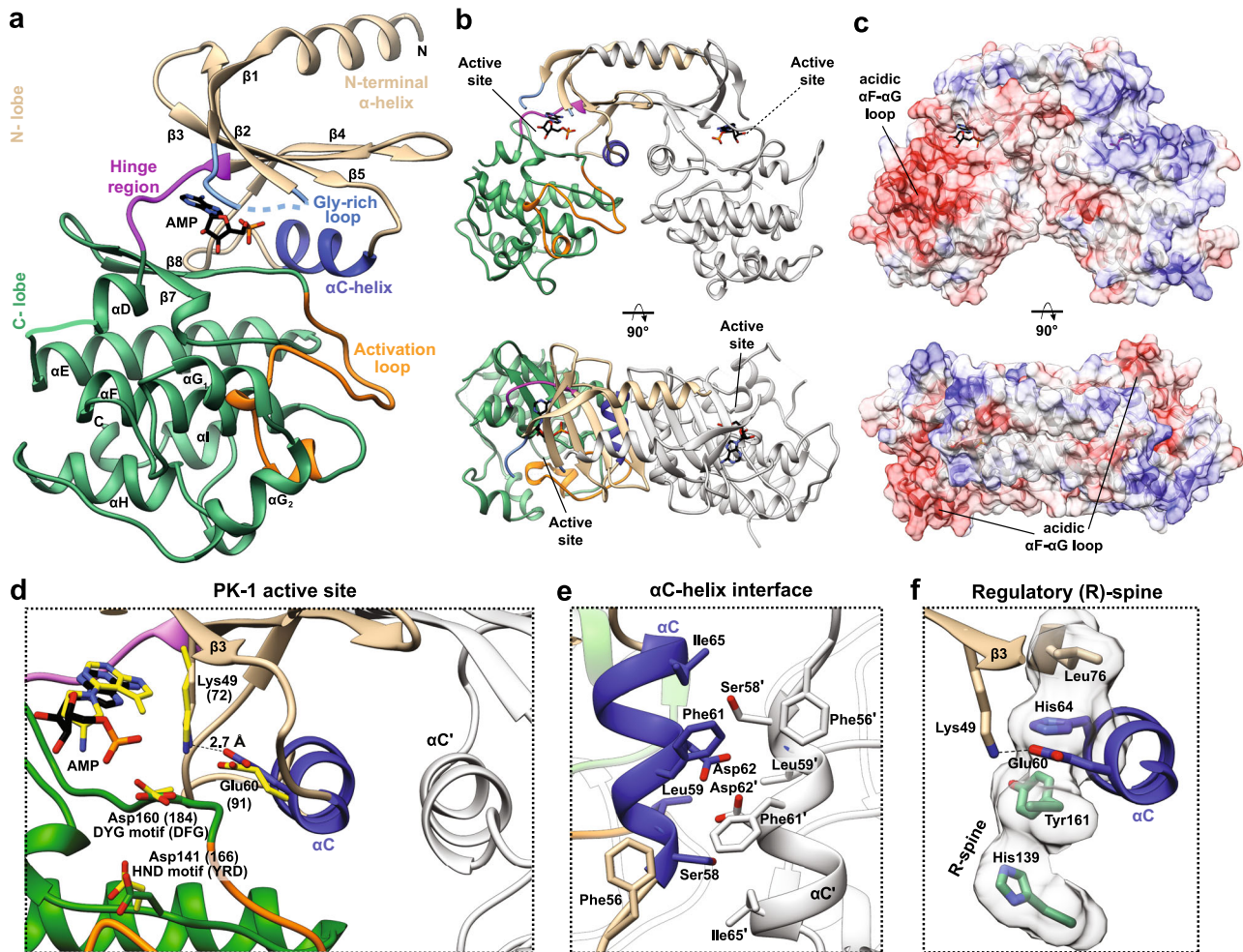
## Results

### *C. pomonella* GV PK-1 adopts an active kinase conformation.

To gain atomic level insight into the PK-1 structure and an understanding of the regulation of its catalytic activity, we solved the crystal structure of PK-1 from *C. pomonella* granulovirus (CpGV) in complex with AMP at 2 Å resolution (Fig. 1; coding sequence in Supplementary Table 1; crystallographic statistics in Supplementary Table 2). PK-1 exhibited a bi-lobal architecture, comprising a smaller N-terminal lobe (Fig. 1a, tan) and a large C-terminal lobe (Fig. 1a, green) that are connected by a short hinge region (Fig. 1a, purple), which resembles the classical eukaryotic protein kinase fold<sup>18</sup>. The N-lobe contains five  $\beta$ -strands ( $\beta$ 1– $\beta$ 5) that form an antiparallel  $\beta$ -sheet, the regulatory  $\alpha$ C-helix (Fig. 1a, dark blue) and an unusual additional  $\alpha$ -helix at the very N-terminus that sits atop the N-lobe. The C-lobe is typical of protein kinases, and mostly comprises  $\alpha$ -helices ( $\alpha$ D– $\alpha$ I), in addition to the activation loop (Fig. 1a, orange) and two additional  $\beta$ -strands ( $\beta$ 7– $\beta$ 8).

The crystal structure contained two PK-1 protomers in the asymmetric unit, which closely associate to form a parallel, side-to-side dimeric assembly (Fig. 1b, Supplementary Fig. 2a). Analysis using the PDBePISA server<sup>19</sup> confirmed that the dimer interface was extensive, with a buried surface area of  $\sim$ 1652 Å<sup>2</sup> and  $\Delta$ G of  $-20.6$  kcal/mol, consistent with stable association. This dimerization interface is formed by four key regions, which include: an atypical N-terminal  $\alpha$ -helix that is domain-swapped between protomers (Supplementary Fig. 2b); a contiguous  $\beta$ -sheet formed by the N-lobe  $\beta$ 1– $\beta$ 5 strands of each protomer (Supplementary Fig. 2c); a hydrophobic core stabilized by aromatic stacking interactions between the N-lobe  $\beta$ -sheet and the  $\alpha$ C-helix, primarily through Phe61 and Phe79 (Supplementary Fig. 2c); and a network of van der Waals interactions and hydrogen bonds between the activation loop and the C-terminus of the  $\alpha$ E-helix from opposing protomers (Supplementary Fig. 2d). As a consequence of these interactions, the  $\alpha$ C helices from each protomer are positioned side-by-side in an antiparallel arrangement (Fig. 1b) and the active sites of each protomer face in opposite directions. Mapping the electrostatic potential to the molecular surface of PK-1 revealed a cluster of negative charge within the  $\alpha$ F– $\alpha$ G loop in the C-lobe, near the active site (Fig. 1c).

The active site of PK-1 grossly resembles that of a canonical eukaryotic protein kinase, and the core catalytic residues<sup>20</sup>—the N-lobe  $\beta$ 3-strand Lys, the catalytic loop Asp and activation loop Asp—are all present and conserved among PK-1 orthologs (Fig. 1d). Accordingly, co-crystallized AMP was bound in the active site in a conformation analogous to that commonly observed for ATP in eukaryotic protein kinases (Fig. 1d). The closed conformation of the PK-1 kinase domain resembles the structures of active conformations of eukaryotic protein kinases<sup>21,22</sup>, which appears to be stabilized allosterically by the  $\alpha$ C helix of the dimeric partner PK-1 protomer (Fig. 1d, e). The  $\alpha$ C helix from one PK-1 protomer associates with that of its dimer partner in an antiparallel fashion (Fig. 1e) anchored by hydrophobic interactions between Phe61 in the  $\alpha$ C helix of one protomer and Phe56 of the dimer partner  $\alpha$ C helix. This conformation is augmented by hydrophobic interactions of Phe61 ( $\alpha$ C helix) with Phe79 and Phe81 ( $\beta$ 4 strand, N-lobe), the latter of which also forms hydrophobic contacts with Ile65 ( $\alpha$ C helix) (Supplementary Fig. 2c). In addition, hydrogen bonds between the Asp62 pairs and with Ser58 of the dimer partner contribute to the dimer interface between  $\alpha$ C helices (Fig. 1e). In addition to the ATP-binding  $\beta$ 3-strand residue, Lys49, forming a salt bridge with Glu60 of the  $\alpha$ C-helix (the counterpart of the Lys72:Glu91 salt bridge in protein kinase A), we observed an intact regulatory (R) spine comprising His64, Leu76, His139 and Tyr161 (Fig. 1f). Although unusual in protein kinase structures,



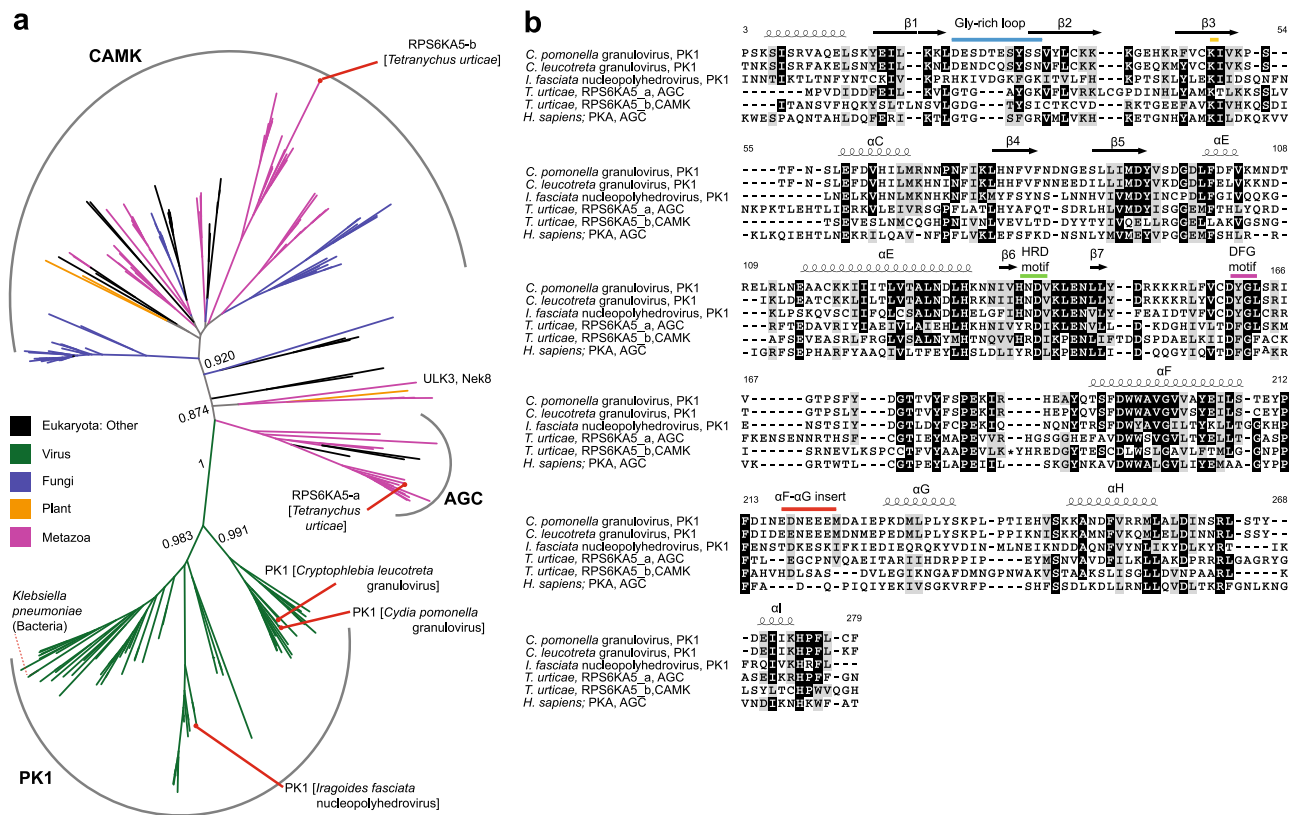
**Fig. 1** Overall structure of CpGV PK-1. **a** The PK-1 catalytic core has two lobes—an N-terminal lobe (tan) and a C-terminal lobe (green), connected via a short hinge region (purple). The N-terminal lobe contains an extended  $\alpha$ -helix at the N-terminus, a bound AMP nucleotide (black), the Gly-rich loop (light blue, disordered) and the  $\alpha$ C-helix (dark blue). The C-terminal lobe is largely  $\alpha$ -helical and contains the activation loop (orange). **b** Cartoon representation of the side-to-side dimeric assembly formed by PK-1 (coloured by **a**, grey). The active site is located within a cleft formed between the bi-lobed architecture, which face in opposite directions between each protomer. **c** The electrostatic potential mapped onto the molecular surface of PK-1 shows a cluster of negative charge on the front face (red) at the  $\alpha$ F- $\alpha$ G loop. **d** The active site residues of PK-1 (shown as sticks) grossly resemble those of the canonical eukaryotic protein kinase A (PKA; PDB ID: 3MVJ)<sup>75</sup> (yellow sticks, residue number and catalytic motifs in parentheses). Conventional protein kinases contain an ATP-binding VAIK motif in the  $\beta$ 3-strand (present as FVCK in PK-1, YAMK in PKA), a  $Mg^{2+}$  binding DFG motif at the start of the activation loop (DYG in PK-1, DFG in PKA) and the catalytic HRD motif in the catalytic loop (HND in PK-1, YRD in PKA). AMP is bound in a conformation analogous to that of the nucleotide in PKA (yellow). **e** In PK-1, an antiparallel arrangement between dimer-related  $\alpha$ C helices, mediated principally by hydrophobic interactions, allosterically locks PK-1 in an activated conformation. **f** The residues that form an intact regulatory (R)-spine are shown as sticks and by surface representation. As a result, the salt bridge between the  $\beta$ 3-strand Lys (Lys49) and  $\alpha$ C-helix Glu (Glu60) is preserved.

the entire activation loop was resolved in our PK-1 structure. The activation loop adopted an extended conformation, which appears to have been stabilized by interactions with its dimeric partner, including residues C-terminal to the  $\alpha$ E-helix located at the dimeric interface (Supplementary Fig. 2d, e).

#### PK-1 sequences diverge from eukaryotic protein kinases.

Although encoded by a granulovirus genome, our structure revealed PK-1 to possess a eukaryotic protein kinase fold. This led us to propose the origin of the kinase might lie in acquisition from the host. We performed Basic Local Alignment Search Tool (BLAST)<sup>23</sup> searches to identify sequences closely related to CpGV PK-1, followed by phylogenetic analysis to determine the relationships between PK-1 and other eukaryotic protein kinases. We identified PK-1 homologues in granuloviruses (98–100% coverage, and 44–100% identity) and nucleopolyhedroviruses (61–99% coverage and 33–46% sequence identity). In several granulovirus

sequences and a subset of nucleopolyhedrovirus sequences, sequence homology extended beyond the kinase domain to include the N-terminal extension that formed a domain-swapped helix in the CpGV PK-1 structure. However, in most distantly related viral and eukaryotic protein kinase sequences, sequence similarity with PK-1 is limited to the catalytic domain, suggesting that the N-terminal helix is specific to the baculovirus clade, and likely acquired after incorporation into the viral genome. Remarkably, viral PK-1 sequences were found to be most closely related to metazoan homologues of RPS6KA5 kinases in the AGC family (bootstrap value of 0.874), and to a lesser extent to metazoan homologues of RPS6KA5 kinases belonging to  $Ca^{2+}$ /calmodulin protein kinase (CAMK) family (Fig. 2a). Because the greatest sequence similarity was to those from arthropods, such as *Tetranychus urticae* (Red spider mite), we propose that PK-1 was acquired from the genome of the insect cell host. Surprisingly, we identified an ortholog (WP\_143445091.1) in



**Fig. 2** Phylogenetic analysis of baculoviral PK-1. **a** Unrooted tree of PK-1 and closely related sequences as identified through a BLAST search. Types of organisms are differentiated by colour; major bootstrap confidence levels are shown and each prominent kinase family are highlighted. **b** Multiple sequence alignment of CpGV PK-1 with other representative sequences on the phylogenetic tree in (a). The alignment was visualized using ESPrnt 3.076 with CpGV PK-1 crystal structure as input to define the secondary structures. The black background highlights identical residues, while the grey background indicates similar residues. Residues are labelled with respect to CpGV PK-1. The Gly-rich loop (blue),  $\beta$ -strand Lys (orange), canonical HRD (green) and DFG (pink) motif, and acidic  $\alpha$ F- $\alpha$ G insert loop (red) are highlighted. Secondary structure of CpGV PK-1 is shown above the alignment. Accession numbers for the sequence analysis are as follows: *C. leucotreta* PK-1: NP\_891851.1; *I. fasciata* PK-1: ACJ04629.1; *T. urticae* AGC, Ca<sup>2+</sup>/calmodulin-dependent protein kinase (CAMK): XP\_015782813.1; *H. sapiens* PKA: NP\_002721.1. \* denotes an insert sequence in RPS6KA5.

*Klebsiella pneumoniae*, a bacterium, raising the prospect of horizontal gene transfer from baculoviruses to prokaryotes or convergent evolution following acquisition from an insect host.

Next, we sought to identify evolutionary constraints distinguishing viral PK-1 and related CpGV sequences from other eukaryotic protein kinase sequences (ePKs). To quantify the constraints, we performed a Bayesian statistical analysis on the sequence alignment containing PK-1 and ePK sequences. The analysis revealed constraints in several key regions, including the catalytic loop, the activation loop, and the  $\alpha$ F-helix (Supplementary Fig. 3a, b). Specifically, two different clusters of interactions were formed by the PK-1 specific residue constraints in the N- and C-termini of the activation loop (Supplementary Fig. 3b, Insets). At the N-terminus of the activation loop, Y161 (the equivalent of the Phe in the DFG motif) forms van der Waals contacts with H64 and F73. Hydrophobic interactions are extended by van der Waals contacts between M67, N70 and F73 (Supplementary Fig. 3b, Top Inset). At the C-terminal end of the activation loop, a salt bridge between D174 in the activation loop and K184 immediately following the end of the activation loop is further stabilized by additional hydrogen bonds from N140 in the catalytic loop, with K184 further stabilized by van der Waal contacts with F180 and F194 (Supplementary Fig. 3b, Bottom Inset). Together, these PK-1-specific variations appear to further stabilize the activation loop in an extended conformation.

The CpGV PK-1 active site contains the  $\beta$ 3-strand Lys, and Asp residues in the catalytic (HRD motif; HND in PK-1) and

activation (DFG motif; DYG in PK-1) loops, which are typical motifs associated with protein kinase activity, as mentioned above. We note that despite the substitution of Phe in the canonical DFG motif, the Tyr present still contributes to the assembly of an intact R-spine in our CpGV crystal structure (Fig. 1d). Two structural elements, however, are of interest owing to their divergence relative to conventional protein kinases: a loss of Gly residues from the Gly-rich loop; and an acidic insert in the loop connecting the  $\alpha$ F and  $\alpha$ G helices (Fig. 2b). Located between the  $\beta$ 1- and  $\beta$ 2-strands of the N-lobe, the Gly-rich loop is a highly conserved motif among kinase proteins<sup>24–26</sup>. The component G $\delta$ GX $\phi$ G (G, Gly;  $\delta$ , hydrophilic; X, any amino acid;  $\phi$ , large hydrophobic amino acid) consensus motif facilitates loop mobility to shield the ATP substrate from bulk solvent and position the ATP  $\gamma$ -phosphate for catalysis<sup>27–29</sup>. Unusually, the corresponding sequence in CpGV PK-1, DESdteSYS (where dte is an insert sequence), and those of other granulovirus PK-1 orthologs (Fig. 2b; Supplementary Fig. 3), lack all three conserved glycine residues, while only one of the glycine residues is lost in nucleopolyhedrovirus homologues (Fig. 2b; Supplementary Fig. 3). This divergence in the granulovirus PK-1 Gly-rich loop is reminiscent of that observed among small molecule kinases, where glycine residues are commonly lost from this loop. In contrast to typical protein kinases, the loss of Gly-rich loop conformational flexibility is believed to confer an evolutionary advantage upon small molecule kinases by promoting efficient, constitutive substrate

phosphorylation<sup>30,31</sup>, raising the prospect of a similar function in viral protein kinases.

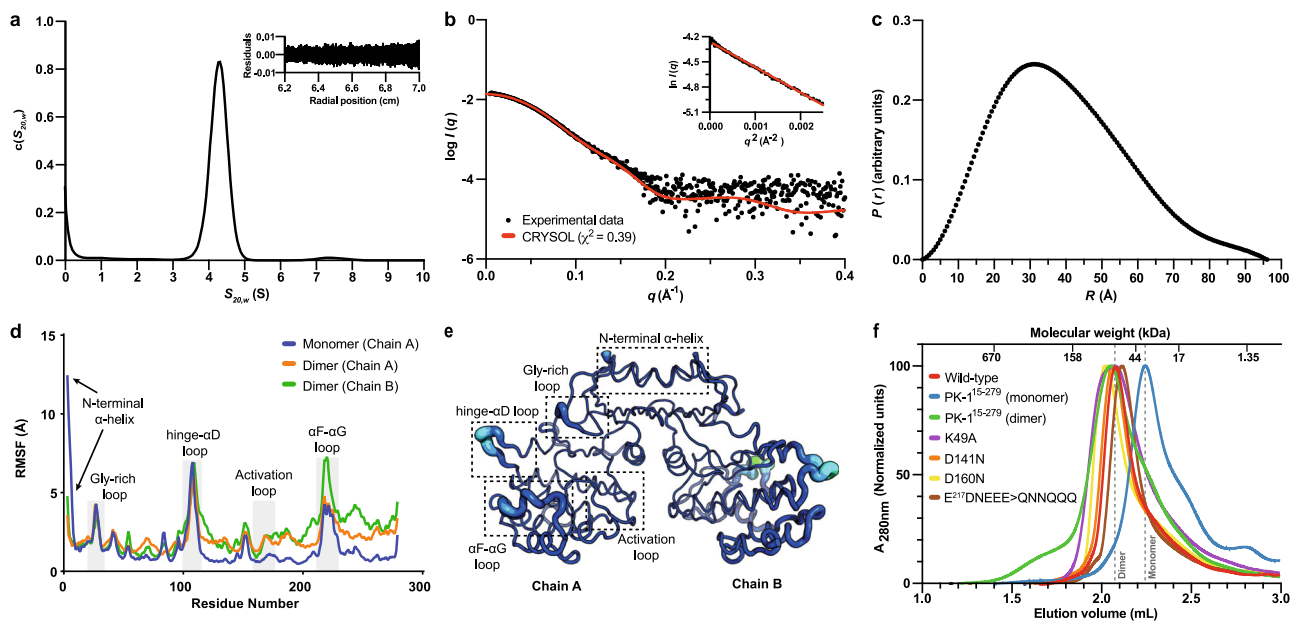
Another atypical feature of PK-1 is an insertion of acidic residues in the  $\alpha$ F- $\alpha$ G loop in the C-lobe (Fig. 2b; Supplementary Fig. 3), which are surface-exposed (Fig. 1c). This loop is known to contribute to substrate recognition in other kinases<sup>32,33</sup>, suggesting that the sequence, E<sup>217</sup>DNEEE<sup>222</sup>, might contribute to recognition of highly-basic substrates, such as p6.9. Consistent with such a conserved function, an acidic sequence in this loop is conserved among granulovirus PK-1 orthologs, but not within nucleopolyhedrovirus orthologs (Fig. 2b). Furthermore, in the crystal structure, the acidic insert lines a putative substrate binding pocket, presumably for selectively binding residues in the p6.9 substrate.

**PK-1 adopts a rigid dimeric conformation in solution.** Because PK-1 occurred as an intertwined dimer within our crystal structure, we sought to evaluate its oligomeric state in solution by performing sedimentation velocity analytical ultracentrifugation (AUC) experiments at a concentration of 0.9 mg mL<sup>-1</sup> (27  $\mu$ M). When the sedimentation data were fitted with a continuous sedimentation coefficient [c(s)] distribution model, a single species was observed with a sedimentation coefficient of 4.4S and a frictional ratio ( $f/f_0$ ) of 1.29, suggesting that PK-1 is asymmetric (Fig. 3a; statistics shown in Supplementary Table 3). Furthermore, the measured molar mass for this single species was 66.8 kDa, consistent with a calculated dimer mass of 66 kDa, confirming that PK-1 exists as a dimer in solution.

We then proceeded to characterise the arrangement of the dimer in solution using small-angle X-ray scattering (SAXS)

experiments with an inline size exclusion chromatography (SEC) setup (statistics shown in Supplementary Table 4). As illustrated in the inset accompanying the experimental scattering profile (Fig. 3b), the Guinier analysis produced a linear plot, consistent with a single monodisperse species in solution and an absence of substantive aggregation or interparticle interference. The theoretical scattering profile, which was generated from the CpGV PK-1 crystal structure was consistent with the experimental data ( $\chi^2$  value of 0.39 using CRY SOL; Fig. 3b), which supports that the arrangement of protomers is biologically relevant in solution and not just a feature within the crystal. The maximum particle dimension ( $D_{\max}$ ) of PK-1 was determined to be 96 Å from the  $P(r)$  analysis (Fig. 3c), which is in agreement with the envelope, based on the crystal structure, calculated using CRY SOL (92 Å). The modest difference is likely associated with flexibility of the protomers in solution, an observation supported by the  $f/f_0$  in our AUC experiments. Measured on a relative scale using SAXS-MoW<sup>234</sup>, the molecular weight of PK-1 from the SAXS experiment was 74.3 kDa, consistent with the calculated dimeric mass (66 kDa).

To further explore the molecular basis of this dimeric interface and assess overall flexibility, we subjected the apo form of the CpGV PK-1 dimer to 1 microsecond ( $\mu$ s) long unbiased molecular dynamics (MD) simulations. During the simulation, the two chains in the dimeric form moved away from each other initially but remained bound throughout the simulation (Supplementary Movie 1). The movement away from each other is seen as an initial increase in the radius of gyration ( $R_g$ ) followed by fluctuations around a mean  $R_g$  of 2.8 nm (Supplementary Fig. 4). Interestingly, three regions distal to the dimer interface were dynamic, as indicated by their relatively higher root mean square



**Fig. 3 PK-1 adopts a rigid dimer assembly in solution.** **a** Sedimentation velocity data at a protein concentration of 0.9 mg mL<sup>-1</sup> were fitted to a continuous sedimentation coefficient [c(s)] distribution model. The residuals and best fit of the data are shown as an inset. **b** SAXS scatter profile for PK-1. The scattering data is best fit (red line,  $\chi^2 = 0.39$ ) to the theoretical scatter of the PK-1 crystal structure. The Guinier plot is linear (inset), showing the sample is free from any measurable amounts of aggregation or inter-particle interference. **c** The pairwise distribution plot, calculated from the scattering data, estimate the maximum interparticle dimension ( $D_{\max}$ ) is 96 Å. **d** Plot showing the root mean square fluctuations (RMSF) of monomeric PK-1 (blue) and chain A and B of dimeric PK-1 (orange and green, respectively) in MD simulations. Residue fluctuations, monitored for the C $\alpha$  atoms of the protein over the entire trajectory, are shown. The activation loop and most dynamic regions of PK-1 are highlighted. **e** Modelled from the MD simulation, the PK-1 dimer is represented as a putty cartoon, highlighting the atomic fluctuations of PK-1. These fluctuations are also graded by colour from lowest (dark blue) to highest (light blue) RMSF value. As per the RMSF plot in (d), the activation loop and most dynamic regions of PK-1 are shown. **f** Analytical SEC of wild-type PK-1 and mutant constructs. The elution volumes of molecular weight standards are shown above the chromatograms. Absorbance at 280 nm ( $A_{280\text{nm}}$ ) is normalized across all samples.

fluctuations (RMSF) values (Fig. 3d, orange and green lines). These regions included the Gly-rich loop, hinge- $\alpha$ D loop and  $\alpha$ F- $\alpha$ G loop (Fig. 3d, e). When the apo monomeric form of PK-1 was subjected to an approximately 2.5  $\mu$ s long unbiased MD simulation, the same regions were also dynamic in addition to the N-terminal  $\alpha$ -helix, which displayed higher RMSF values compared to the dimer (Fig. 3d, Supplementary Movie 2). Notably, in both simulations, the activation loop was relatively stable as indicated by low RMSF values. On mapping the PK-1 specific constraints to the monomeric MD trajectory, we find that the two clusters of interactions formed by PK-1 specific constraint residues at the N- and C-termini of the activation loop are stable (Supplementary Movie 2). This observation suggests that specific evolutionary changes within PK-1 might contribute to the stabilization of the activation loop in an extended conformation. Our observations further support the view that the activation loop – an element typically flexible in protein kinases – is rigidified in PK-1, presumably to facilitate constitutive substrate recognition and phosphorylation.

Based on this notion and the dimer observed in our crystal structure, we sought to investigate if a monomeric form of PK-1 could be obtained by generating a truncated PK-1 construct (PK-1<sup>15–279</sup>) in which the atypical domain-swapped N-terminal helix is absent. While full-length wild-type PK-1 eluted as a single peak from SEC during purification, PK-1<sup>15–279</sup> eluted as two distinct peaks corresponding to dimeric (~66 kDa) and monomeric (~33 kDa) species (Fig. 3f), when analysed individually against a molecular weight standard, indicating this truncation of the N-terminal helix alters the oligomeric state. By comparison, substitutions of putative catalytic residues (K49A, D141N and D160N) and the acidic sequence in the  $\alpha$ F- $\alpha$ G loop (E<sup>217</sup>DNEEE>QNNQQQ) did not impact the oligomeric state, because these proteins eluted solely as dimeric species using analytical SEC. Taken together, our AUC, SAXS and SEC data, and molecular dynamics simulations demonstrate that full-length PK-1 exists as a rigid dimer in solution, which can be disrupted by truncation of the atypical, domain-swapped N-terminal helix.

**Biochemical characterization of PK-1.** We next sought to define the role of dimerization in regulating PK-1 catalytic activity by examining the activities of wild-type PK-1 relative to monomeric and dimeric PK-1<sup>15–279</sup>, acidic  $\alpha$ F- $\alpha$ G loop and active site mutant PK-1 proteins (Fig. 4a) in an ADP-Glo assay in the presence of a p6.9 substrate peptide. While wild-type PK-1 exhibited robust catalytic activity (Fig. 4b, red), the activity of the monomeric form of PK-1<sup>15–279</sup> was severely compromised (Fig. 4b, blue), with activity of ~20% of that of full-length wild-type PK-1. These data clearly implicate dimerization as serving a crucial function in promoting PK-1 catalytic activity. The dimeric form of the PK-1<sup>15–279</sup> truncation mutant retained ~70% of full-length wild-type PK-1 activity (Fig. 4b, green), suggesting that the active conformation dictated by the dimer, rather than dimerization per se, is crucial for optimal p6.9 phosphorylation.

PK-1 contains a conserved, atypical acidic sequence in  $\alpha$ F- $\alpha$ G loop (E<sup>217</sup>DNEEE) (Fig. 4a) that could potentially contribute to recognition of highly basic substrates, such as p6.9. We substituted this sequence to eliminate the negative charge (E<sup>217</sup>DNEEE>QNNQQQ) and examined the impact on catalysis using the ADP-Glo assay. Surprisingly, this mutant exhibited comparable catalytic activity to full-length wild-type PK-1 (Fig. 4b, brown), indicating that in an in vitro context in which substrate is in excess, this sequence does not measurably enhance substrate recognition.

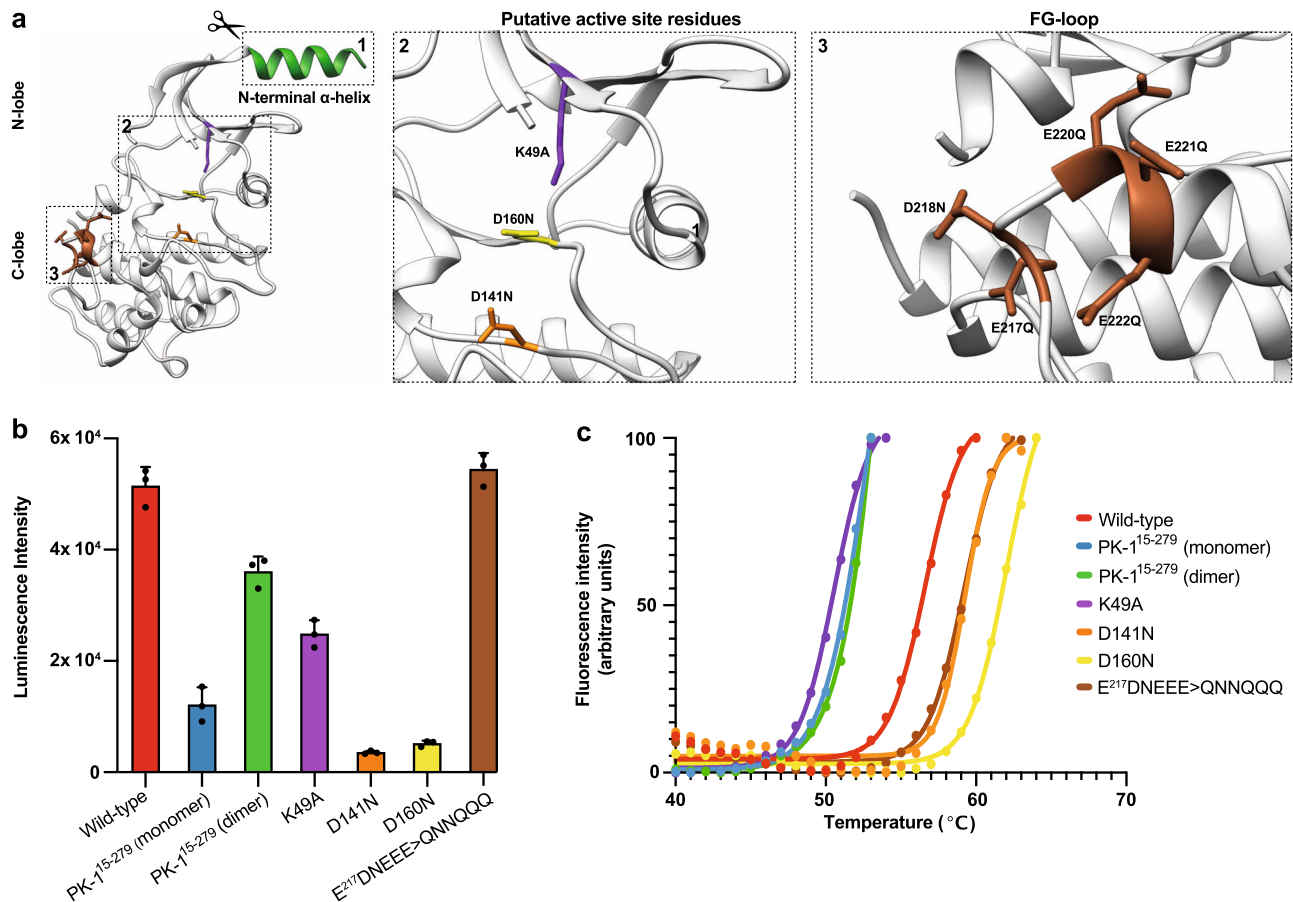
Because PK-1 adopts the classical eukaryotic protein kinase fold, it was of interest to establish whether the residues

corresponding to core catalytic residues in conventional protein kinases serve comparable functions in PK-1. Accordingly, we compared the catalytic activities of K49A, D141N and D160N full-length PK-1 in ADP-Glo assays (Fig. 4a). As expected, the activities of the D141N and D160N PK-1 mutants were ablated (Fig. 4b, orange and yellow, respectively), illustrating that PK-1 catalytic activity relies on residues that are the counterparts of the catalytic residues within a conventional protein kinase. Unexpectedly, however, K49A PK-1 exhibited ~50% of wild-type PK-1 activity (Fig. 4b, purple). These data indicate that K49 in PK-1, which typically functions in positioning ATP via interaction with the  $\alpha$ - and  $\beta$ -phosphates in conventional eukaryotic protein kinases, is dispensable for PK-1 catalytic activity. This raises the possibility that the active form of PK-1, stabilized by dimerization, and the Gly-rich loop sequence that is divergent from those of conventional protein kinases (Fig. 2a), obviates the reliance on K49 for ATP-binding and positioning for catalysis. To ensure the differences in activity we observed for each PK-1 construct were not attributable to compromised protein folding, we evaluated each variant's thermal stability by differential scanning fluorimetry. Wild-type and mutant PK-1 proteins exhibited melting temperatures ( $T_m$ ) ranging from 50 to 61 °C (Fig. 4c), which are values typical of folded kinase domains<sup>35</sup>. Collectively, these data demonstrate that PK-1 is catalytically-active and that the assembly and conformation of PK-1 dimers is critical for optimal activity.

## Discussion

Here, we report the structure of a baculoviral kinase, PK-1, which adopts a fold typical of eukaryotic protein kinases and harbours active site residues typical of conventional eukaryotic kinases. Our sequence analyses indicate a eukaryotic ancestral origin of PK-1, where the kinase was acquired via co-opting features from multiple eukaryotic kinases, in particular RPS6KA5 from the insect host over its evolutionary trajectory. In contrast to most eukaryotic protein kinases, PK-1 occurs as a constitutive dimer. Dimerization via an atypical domain swapped N-terminal helix and the formation of a contiguous  $\beta$ -sheet between the two protomer's N-lobes positions the two protomers in a parallel side-to-side arrangement with an antiparallel arrangement of the two  $\alpha$ C helices at the dimer core, such that the active sites of the two monomers face in opposite directions. The parallel side-to-side dimer adopted by PK-1 is unusual, because prior structural studies have reported only other modes of  $\alpha$ C-helix-focused allosteric regulation to date (Fig. 5). Various modes of protein kinase allosteric regulation by protein ligands, such as cyclin binding to cyclin-dependent protein kinases (CDKs)<sup>36</sup> or the binding of ligands to the PDK-1 interacting fragment (PIF) pocket of PDK1<sup>37</sup>, and kinases/pseudokinases via homo- and heterodimerization have been reported. In common with the parallel side-to-side dimerization mode that we report here for PK-1, each of the previously-reported binding modes influence the position of the central regulatory element, the  $\alpha$ C helix, to dictate whether the kinase/pseudokinase adopts a closed conformation synonymous with the active form of a conventional protein kinase (Fig. 5)<sup>38–42</sup>.

Our data lead us to propose an unusual mechanism, whereby the dimerization of PK-1 via an atypical dimer interface facilitates constitutive phosphorylation of its substrates. This dimerization mode stabilizes elements typically flexible in kinases, the Gly-rich and activation loops, and by locking the kinase into a rigid, active conformation marked by an intact R-spine and  $\beta$ 3-strand K49: $\alpha$ C helix E60 salt bridge, predisposes PK-1 to binding ATP in a productive orientation for catalysis. The  $\beta$ 3-strand residue, K49, which is known to mediate ATP  $\alpha$ - and  $\beta$ -phosphate binding in

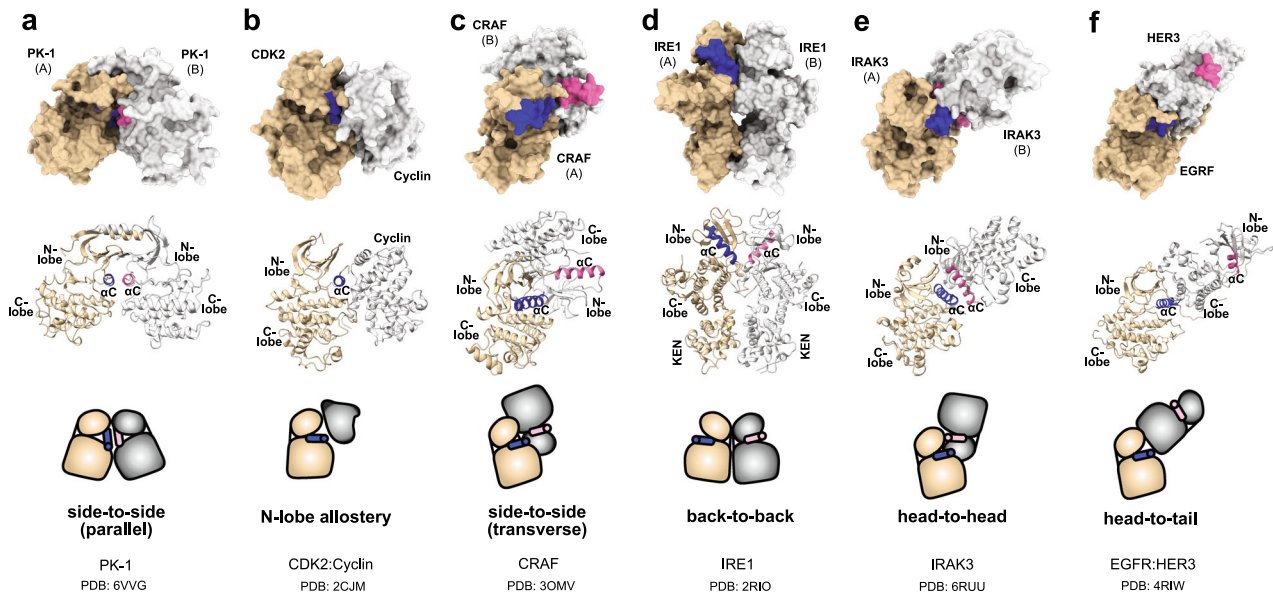


**Fig. 4 Biochemical characterisation of PK-1.** **a** Residues mutated in the present study are shown as sticks on the PK-1 crystal structure solved in this study (Fig. 1), including the truncated N-terminal  $\alpha$ -helix (green ribbon) to generate PK-1<sup>15-279</sup> (box 1) and zoomed views of the putative active site residues (box 2) and the patch of acidic residues (brown sticks) within the  $\alpha$ F- $\alpha$ G loop (box 3) shown on the right. **b** In vitro kinase activities of the PK-1 wild-type and mutant constructs. Individual data points are plotted; the bar and error bars shown represent mean  $\pm$  SD of three independent ADP-Glo assays. **c** Thermal melt curves of the PK-1 wild-type and mutant constructs, performed using differential scanning fluorimetry confirms that these variants are folded. Data represent mean of two independent experiments. Data are plotted throughout for wild-type PK-1 in red, and mutants are color-coded: truncated PK-1 monomer in blue and dimer in green; K49A, purple; D141N, orange; D160N, yellow; and E<sup>217</sup>DNEEE>QNNQQQ, brown.

conventional protein kinases and pseudokinases<sup>35,43–46</sup>, and serve an indispensable role in catalysis in conventional kinases<sup>47,48</sup>, could be substituted to Ala and retain ~50% of wild-type activity. This contrasts with Asn substitution of the canonical catalytic Asp (D160) and Mg<sup>2+</sup>-binding Asp (D141), which lead to almost complete ablation of catalytic activity. In addition to the dispensability of K49 for catalytic activity, the Gly-rich loop (D<sup>24</sup>ESdteSYS) is divergent relative to those in conventional protein kinases (G $\delta$ GX $\phi$ G). We propose that the structural rigidification in PK-1 dimers that arises from the parallel side-to-side dimerization mode supersedes the necessity of Gly-rich loop and activation loop flexibility and the  $\beta$ 3-strand Lys for ATP-binding and catalysis. Although this dimerization mode has not been reported for other kinases or pseudokinases<sup>49</sup>, our findings raise the prospect more broadly that inter-kinase/pseudokinase interactions may serve functions to rigidify the core active site elements responsible for ATP-binding and positioning for catalysis.

PK-1 contains an acidic insertion within the  $\alpha$ F- $\alpha$ G loop (E<sup>217</sup>DNEEE), which based on our bioinformatic analysis, is present in *Baculoviridae* PK-1 sequences. Considering PK-1's function as a kinase that targets Arg-rich substrates, we hypothesized that this sequence could serve a function in recruitment of polybasic substrates. However, in our hands, mutation of this sequence to neutralize charge did not impact catalytic activity relative to wild-type protein. As a result, our data suggest that, at

least under conditions where substrate is in excess, this element, which is highly mobile in molecular dynamics simulations, does not markedly influence substrate recognition. One possibility is that the acidic sequence of the atypical Gly-rich loop on the opposing lip of the active site may compensate for loss of the acidic  $\alpha$ F- $\alpha$ G loop sequence to facilitate recognition of basic substrates. Functionally, recognition of the Arg-rich substrate, p6.9, by PK-1 occurs in nuclear occlusions where the substrate is highly concentrated and, accordingly, would not be expected to rely on the complementary charges of p6.9 and PK-1's acidic  $\alpha$ F- $\alpha$ G loop for phosphorylation. Instead, it appears that the role of PK-1 phosphorylation is to reduce the net positive charge of p6.9 to disrupt DNA interaction so transcription of viral very late genes can proceed. In support of this idea, 7 serines within p6.9 were identified as crucial PK-1 substrates, where their mutation to Ala prevented progression of very late gene transcription<sup>17</sup>. Deletion of PK-1 from *Autographa californica* multiple nucleopolyhedrovirus compromised p6.9 phosphorylation in Sf9 insect cells, revealing an important role for PK-1 in promoting hyperphosphorylation of Ser and Thr residues and apparently priming p6.9 for Arg phosphorylation. Because some Arg and Ser/Thr phosphorylation was observed within p6.9 in the absence of PK-1<sup>17</sup>, this suggests some redundancy among the p6.9-targeting kinases encoded within baculoviral genomes. Furthermore, these data indicate that some viral kinases catalyse Arg phosphorylation



**Fig. 5 Comparison of known modes of kinase dimerization and allostery.** **a** Parallel side-to-side homodimer PK-1 (PDB 6VVZ; solved in this study); **b** allosteric regulation of CDK2 via N-lobe binding to Cyclin (PDB 2CJM)<sup>77</sup>; **c** transverse side-to-side homodimer CRAF (PDB 30MV)<sup>38</sup>; **d** back-to-back homodimer IRE1 (PDB 2RIO)<sup>41</sup>; **e** head-to-head homodimer IRAK3 (PDB 6RUU)<sup>40</sup>; **f** head-to-tail heterodimer EGFR:HER3 (PDB 4RIW)<sup>42</sup>. Each allosteric or dimerization mode is illustrated by molecular surface (top panel), cartoon ribbon (middle panel) and schematic representation (lower panel). Protomer A and B are coloured tan and grey, respectively. The  $\alpha$ -helix ( $\alpha$ C) from protomer A and B are highlighted in dark blue and pink, respectively. The N-lobe (N) and C-lobe (C) are shown, and the KEN domain that contributes to the IRE1 dimer interface is indicated in **d**.

that is primed by Ser/Thr phosphosites within p6.9. Thus, whether the Gly-rich loop and  $\alpha$ F- $\alpha$ G loop acidic sites work in concert to contribute to Arg-rich peptide substrate recognition remains of outstanding interest.

The dimerization mode observed for PK-1 in our structural and biophysical studies is unusual among protein kinases. The combination of allostery, where each protomer locks its dimer-opposed partner into a rigid conformation primed for catalysis, and the parallel dimerization mode with one subunit facing in each direction, distinguish PK-1 from other reported modes of kinase dimerization. Such a dimerization mode would allow processive phosphorylation of repetitive substrates, like protamines and p6.9 in the case of PK-1, via two active sites poised in a catalytically-active conformation by virtue of the locked, side-to-side dimerization. Knowledge of the extent to which side-to-side dimerization is used more broadly in nature among viral and eukaryotic protein kinases to regulate catalytic activity awaits future detailed studies.

## Methods

**Protein expression and purification.** The gene coding for full-length *C. pomonella* GV PK-1 (isolate Mexico/1963; Uniprot Q91F41) was codon-optimized and synthesized for *E. coli* expression (Genewiz; Supplementary Table 1) and sub-cloned into the expression vector pPROEX Htb as an in-frame fusion with a TEV protease-cleavable N-terminal hexahistidine tag. The D141N and D160N mutations were introduced into the wild-type template using oligonucleotide-directed overlap PCR (Supplementary Table 1), with remaining mutant sequences synthesized by GeneWiz. All insert sequences were verified by Sanger sequencing (AGRF, Australia). Wild-type and mutant PK-1 constructs were expressed in *E. coli* BL21-CodonPlus-RIL (Agilent) or C41 (DE3) cells cultured in Super Broth supplemented with ampicillin ( $100 \mu\text{g mL}^{-1}$ ) at  $37^\circ\text{C}$  with shaking at 220 rpm to an  $\text{OD}_{600}$  of  $\sim 0.6$ – $0.8$ . Protein expression was induced by the addition of isopropyl  $\beta$ -D-1-thiogalactopyranoside ( $250 \mu\text{M}$ ) and the temperature was lowered to  $18^\circ\text{C}$  for incubation overnight. Cell pellets were resuspended in lysis buffer (20 mM HEPES, pH 7.5, 200 mM NaCl, 5% glycerol, 0.5 mM TCEP), supplemented with Complete protease cocktail inhibitor (Roche), and lysed by sonication. Cell debris and insoluble material was pelleted via centrifugation at  $45,000 \times g$  and the lysate was incubated with pre-equilibrated Ni-NTA agarose (HisTag, Roche) at  $4^\circ\text{C}$  for 1 h with gentle agitation. Ni-NTA beads were then pelleted via centrifugation and washed thoroughly with wash buffer (20 mM HEPES, pH 7.5, 200 mM NaCl, 5 mM imidazole, 5% glycerol, 0.5 mM TCEP). Bound protein was eluted from the beads

using elution buffer (20 mM HEPES, pH 7.5, 200 mM NaCl, 250 mM imidazole, 5% glycerol, 0.5 mM TCEP), filtered through a  $0.45\text{-}\mu\text{m}$  filter, mixed with  $300 \mu\text{g}$  of recombinant His<sub>6</sub>-TEV and dialysed overnight in size exclusion buffer (20 mM HEPES, pH 7.5, 200 mM NaCl, 5% glycerol) at  $4^\circ\text{C}$ . Following protease cleavage, the dialysate was incubated with Ni-NTA agarose (Roche) pre-equilibrated in the same buffer at  $4^\circ\text{C}$  for 1 h to remove TEV and uncleaved protein. Following incubation, the sample was centrifuged to eliminate particulates, the supernatant  $0.45\text{-}\mu\text{m}$  filtered, concentrated via centrifugal ultrafiltration (30 kDa molecular weight cut-off; Millipore) and loaded onto a HiLoad 16/600 Superdex 200 prep grade size exclusion column (Cytiva) equilibrated in size exclusion buffer. For AUC experiments, glycerol was omitted from the buffer. The final purity of the eluted protein was assessed by SDS-PAGE (Supplementary Fig. 1). Protein that was not immediately used in experiments was aliquoted, flash-frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$ .

**Crystallization, data collection and structure determination.** For crystallization trials, the gene encoding full-length *C. pomonella* GV PK-1 was amplified from genomic DNA (using primer sequences reported in Supplementary Table 1) and subcloned into the Gateway expression vector, pDEST566. The resulting protein, bearing a TEV protease-cleavable N-terminal hexahistidine and maltose-binding protein tag, was expressed in *E. coli* Rosetta (DE3) cells cultured in Luria Broth containing ampicillin ( $100 \mu\text{g mL}^{-1}$ ) at  $37^\circ\text{C}$  with shaking at 220 rpm to an  $\text{OD}_{600}$  of  $\sim 0.6$ – $0.8$ . Protein expression was induced by the addition of isopropyl  $\beta$ -D-1-thiogalactopyranoside ( $100 \mu\text{M}$ ) and the temperature was lowered to  $18^\circ\text{C}$  for incubation overnight. Purification was performed essentially as described above, but Ni<sup>2+</sup>-chromatography was performed using a 5 mL HisTrap FF NiNTA column (Cytiva). Protein eluted from a HiLoad 16/600 Superdex 200 size exclusion column (Cytiva) in 50 mM HEPES, pH 8.0, 150 mM NaCl, 2 mM TCEP was concentrated to  $5 \text{ mg mL}^{-1}$  ( $152 \mu\text{M}$ ) and subjected to sitting-drop vapour diffusion experiments at  $18^\circ\text{C}$  using the JCSG+ and PACT premier (Molecular Dimensions) commercial crystallization screens. A volume of  $0.15 \mu\text{L}$  protein solution and  $0.15 \mu\text{L}$  reservoir solution was equilibrated over  $50 \mu\text{L}$  of reservoir solution using an Oryx 4 Protein Crystallisation Robot (Douglas Instruments LTD, Berkshire, UK). Spontaneously nucleating PK-1 crystals were of insufficient size for data collection, so a fine screen was developed from initial hits. Because initial hits did not diffract beyond  $6 \text{ \AA}$ , we supplemented the protein solution with 10 mM  $\text{MnCl}_2$  and 1 mM adenosine monophosphate (AMP) and obtained suitable PK-1 crystals in 300 mM  $\text{NH}_4\text{Cl}$  and 16.5% (w/v) PEG 3350 at  $18^\circ\text{C}$ . For data collection, the crystals were cryoprotected in this condition supplemented with 20% (w/v) glycerol and then flash-frozen. Data were processed using XDS<sup>50</sup> and AIMLESS from the CCP4 program suite<sup>51</sup> to  $2.01 \text{ \AA}$  in the space group  $P 2_1 2_1 2_1$ . The structure of PK-1 was determined by molecular replacement in PHASER<sup>52</sup> using the human  $\text{Ca}^{2+}$ /Calmodulin-dependent protein kinase II gamma (CaMK2G, PDB accession: 2V7O)<sup>53</sup>. The resulting model was improved by iterative rounds of refinement using PHENIX<sup>54</sup> and manual building in COOT<sup>55</sup>, which included the

addition of AMP and water molecules. Due to insufficient electron density, residues 1, 2, 26–29 and 217–222 in chain A, and residues 1, 2, 27–31, 105–111 and 279 in chain B were not modelled. Final model validation was performed using MOLPROBITY<sup>56</sup>. The dimer interface was analysed using PDBePISA<sup>57</sup>. All structural graphics were prepared using UCSF Chimera<sup>58</sup> and PyMOL. All data collection and refinement statistics are summarized in Supplementary Table 2.

**Sequence and phylogenetic analysis.** NCBI BLAST was performed using full length CpGV PK-1 sequence as the query searched against the non-redundant protein sequences (NR) database. The kinase domains of the sequences were aligned using MAPGAPS<sup>59</sup> and a curated profile built using structural alignments of eukaryotic protein kinases (ePKs) and the CpGV PK-1 crystal structure (PDB: 6VVG) and purged to 90% identity. A maximum-likelihood tree was constructed using FastTree 2.1.11 (ref. 60) and visualized using Interactive Tree Of Life (ITOL)<sup>61</sup>. Evolutionary constraints imposed on PK-1 sequences were quantified using Contrast Hierarchical Alignment Interaction Network (CHAIN)<sup>62</sup> analysis, which quantifies the extent to which aligned residues in the foreground alignment (PK-1 sequence alignment) diverge from the corresponding position in the background alignment (alignment of representative ePK sequences).

**Analytical ultracentrifugation.** Sedimentation velocity experiments were performed in a Beckman Coulter Optima analytical ultracentrifuge using double sector epon-charcoal centerpieces fitted with sapphire windows in an An-50 Ti eight-hole rotor at 20 °C. Data were obtained at 50,000 rpm using the absorbance optical system at 280 nm, measuring protein at 0.9 mg mL<sup>-1</sup> (27 µM) against a buffer reference (20 mM HEPES, pH 7.5, 200 mM NaCl). Sedimentation data were fitted to a continuous size distribution [c(s)] model and converted to a standardized sedimentation coefficient ( $s_{20,w}$ ) distribution using SEDFIT<sup>63</sup>. The buffer density, buffer viscosity and an estimate of the partial specific volume of the protein sample, based on the amino acid sequence were calculated using SEDNTERP. All data collection and analysis statistics are summarized in Supplementary Table 3.

**Small-angle X-ray scattering.** SAXS data were collected at the Australian Synchrotron SAXS/WAXS beamline using an inline co-flow size-exclusion chromatography setup to minimize sample dilution and maximize signal-to-noise<sup>64</sup>. Protein at 5 mg mL<sup>-1</sup> (152 µM) was injected (80 µL) onto an inline Superdex 200 5/150 Increase GL column (Cytiva), equilibrated with size exclusion buffer (20 mM HEPES, pH 7.5, 200 mM NaCl, 5% glycerol) at 12 °C, at a flow rate of 0.2 mL min<sup>-1</sup> using an established protocol<sup>65,66</sup>. 2D intensity plots were radially averaged, normalized against sample transmission, and background-subtracted using the Scatterbrain software package (Australian Synchrotron). The ATSAS software package was used to perform the Guinier analysis (PrimusQT)<sup>67</sup>, to calculate the pairwise distribution function  $P(r)$  and the maximum interparticle dimension ( $D_{max}$ ), and to evaluate the solution scattering against the crystal structure solved in this study (CRY SOL)<sup>68</sup>. The molecular mass of each sample was estimated using the SAXS-MoW2 package<sup>69</sup>. All data collection and processing statistics are summarized in Supplementary Table 4.

**Analytical size exclusion chromatography.** Analytical size exclusion chromatography runs were performed in size exclusion buffer (20 mM HEPES, pH 7.5, 200 mM NaCl, 5% glycerol) on a Superdex 200 5/150 Increase GL column (Cytiva). Purified protein was injected (100 µL) onto the column at concentrations varied from 1 to 5 mg mL<sup>-1</sup> (30–152 µM). Gel filtration standard (BioRad, CA) was used to define the molecular weight marker elution volumes (1.35–670 kDa).

**Molecular dynamic simulations.** Unbiased full atom MD simulation of apo CpGV PK1 in the monomeric and dimeric forms were performed using GROMACS 2018.1. Hetero atoms and water molecules were removed from the co-crystal structure of CpGV PK-1 dimer bound to AMP molecules (PDB ID: 6VVG). The main and sidechain atoms for residues 27–29 in chain A and residues 27–79, 105–110 in chain B were modelled using RosettaLoop<sup>70</sup> using the cyclic coordinate descent (CCD) protocol. Chain A was used for the monomer simulation whereas both chains were used for the dimer simulation. The hydrogen atoms were converted to virtual sites to remove the fastest vibrational freedom. The protein was parameterized with Amber99SB-ILDN and solvated with TIP3P waters modelled in a dodecahedron box. The charges were neutralized by addition of sodium and chloride ions. The neighbour list for nonbonded interactions was defined using Verlet cutoff<sup>71</sup>. Particle Mesh Ewald (PME) was used to calculate long-range interactions. Energy minimization was performed with steepest-descent algorithm followed by conjugate descent with  $F_{max}$  less than 500 kJmol<sup>-1</sup>nm<sup>-1</sup>. The canonical ensemble was carried out by heating the system from 0 K to 310 K, using velocity rescaling for 100 ps<sup>72</sup>. The isothermal-isobaric ensemble ( $P = 1$  bar,  $T = 310$  K) was carried out using the Berendsen barostat for 100 ps<sup>73</sup>. The unrestrained MD productions were collected using a time step of 5 fs. The trajectories were processed and analysed using the built-in GROMACS tools. Visualization was done using PyMOL 2.3.2.

**In vitro kinase activity assay.** The in vitro kinase activity assay was performed with an ADP-Glo kinase assay kit (Promega), following standard procedures. Briefly, a 25 µL kinase reaction contained 0.1 mg mL<sup>-1</sup> PK-1, 0.5 mg mL<sup>-1</sup> p6.9 peptide (with the sequence MVRRRR SRSPNR RRSYRS RSRSRs RSRSRs RSRSRs PYRS, synthesized to >95% purity by Mimotopes Pty Ltd), and 1 mM ATP in kinase reaction buffer (20 mM HEPES, pH 7.5, 20 mM MgCl<sub>2</sub>, 2 mM MnCl<sub>2</sub>, 1 mM DTT, and 0.1 mg mL<sup>-1</sup> BSA). Each reaction was performed at 25 °C for 1 h and terminated by the addition of the ADP-Glo reagent. After incubation for 40 min with the ADP-Glo reagent, the Kinase Detection reagent was added and after an incubation time of 1 h, the luminescence was detected with a microplate reader (CLARIOstar, BMG LabTech).

**Thermal shift assays.** Thermal shift assays were performed as described previously<sup>44,74</sup> using a Corbett Real Time PCR machine with proteins diluted in 20 mM HEPES, pH 7.5, 200 mM NaCl to 10 µg in a total reaction volume of 25 µL. SYPRO Orange (Molecular Probes, CA) was used as a probe with fluorescence detected at 530 nm. Two independent assays were performed for wild-type and mutant PK-1 constructs; averaged data are shown for each in Fig. 4.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Data are available from the corresponding author upon reasonable request. The atomic coordinates for PK-1 have been deposited in the Protein Data Bank with the accession code 6VVG. Source data are provided with this paper.

Received: 24 November 2020; Accepted: 19 January 2021;

Published online: 12 February 2021

## References

- Martignoni M. E., Iwai P. J. *A Catalog of Viral Diseases of Insects, Mites, and Ticks* (Department of Agriculture, Forest Service, Pacific Northwest Research Station, 1986).
- Rohrmann G. F. *Baculovirus Molecular Biology*, 4th edn. (National Center for Biotechnology Information (US), Bethesda (MD), 2019).
- Fraser, M. J. Ultrastructural observations of virion maturation in *Autographa californica* nuclear polyhedrosis virus infected *Spodoptera frugiperda* cell cultures. *J. Ultrastruct. Mol. Struct. Res.* **95**, 189–195 (1986).
- Tweeten, K. A., Bulla, L. A. & Consigli, R. A. Structural polypeptides of the granulosis virus of *Plodia interpunctella*. *J. Virol.* **33**, 877–886 (1980).
- van Oers, M. M. & Vlak, J. M. Baculovirus genomics. *Curr. Drug Targets* **8**, 1051–1068 (2007).
- Balhorn, R. The protamine family of sperm nuclear proteins. *Genome Biol.* **8**, 227 (2007).
- Lewis, J. D. & Ausio, J. Protamine-like proteins: evidence for a novel chromatin structure. *Biochem Cell Biol.* **80**, 353–361 (2002).
- Brewer, L. R., Corzett, M. & Balhorn, R. Protamine-induced condensation and decondensation of the same DNA molecule. *Science* **286**, 120–123 (1999).
- Wang, R. et al. Proteomics of the *Autographa californica* nucleopolyhedrovirus budded virions. *J. Virol.* **84**, 7233–7242 (2010).
- Wilson, M. E. & Consigli, R. A. Functions of a protein kinase activity associated with purified capsids of the granulosis virus infecting *Plodia interpunctella*. *Virology* **143**, 526–535 (1985).
- Liu, X. et al. Distribution and phosphorylation of the basic protein P6.9 of *Autographa californica* nucleopolyhedrovirus. *J. Virol.* **86**, 12217–12227 (2012).
- Kelly, D. C. & Lescott, T. Baculovirus replication: phosphorylation of polypeptides synthesized in *Trichoplusia ni* nuclear polyhedrosis virus-infected cells. *J. Gen. Virol.* **65**, 1183–1191 (1984).
- Kool, M., Ahrens, C. H., Vlak, J. M. & Rohrmann, G. F. Replication of baculovirus DNA. *J. Gen. Virol.* **76**, 2103–2118 (1995).
- Luque, T., Finch, R., Crook, N., O'Reilly, D. R. & Winstanley, D. The complete sequence of the *Cydia pomonella* granulovirus genome. *J. Gen. Virol.* **82**, 2531–2547 (2001).
- Reilly, L. M. & Guarino, L. A. The pk-1 gene of *Autographa californica* multinucleocapsid nuclear polyhedrosis virus encodes a protein kinase. *J. Gen. Virol.* **75**, 2999–3006 (1994).
- Liang, C. et al. *Autographa californica* multiple nucleopolyhedrovirus PK-1 is essential for nucleocapsid assembly. *Virology* **443**, 349–357 (2013).
- Li, A. et al. Posttranslational modifications of baculovirus protamine-like protein P6.9 and the significance of its hyperphosphorylation for viral very late gene hyperexpression. *J. Virol.* **89**, 7646–7659 (2015).

18. Taylor, S. S. & Radzio-Andzelm, E. Three protein kinase structures define a common motif. *Structure* **2**, 345–355 (1994).
19. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
20. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
21. Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl Acad. Sci. USA* **103**, 17783–17788 (2006).
22. Modi, V., Dunbrack, R. L. Jr. & Structurally-Validated, A. Multiple sequence alignment of 497 human protein kinase domains. *Sci. Rep.* **9**, 19790 (2019).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
24. Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011).
25. Cowan-Jacob, S. W. Structural biology of protein tyrosine kinases. *Cell Mol. Life Sci.* **63**, 2608–2625 (2006).
26. Nolen, B., Taylor, S. & Ghosh, G. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* **15**, 661–675 (2004).
27. Zheng, J. et al. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* **32**, 2154–2161 (1993).
28. Fabbro, D., Cowan-Jacob, S. W. & Moebitz, H. Ten things you should know about protein kinases: IUPHAR Review 14. *Br. J. Pharm.* **172**, 2675–2700 (2015).
29. Taylor, S. S., Keshwani, M. M., Steichen, J. M. & Kornev, A. P. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2517–2528 (2012).
30. Oruganty, K. & Kannan, N. Design principles underpinning the regulatory diversity of protein kinases. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2529–2539 (2012).
31. Oruganty, K., Talevich, E. E., Neuwald, A. F. & Kannan, N. Identification and classification of small molecule kinases: insights into substrate recognition and specificity. *BMC Evol. Biol.* **16**, 7 (2016).
32. Chu, W.-T., Chu, X. & Wang, J. Binding mechanism and dynamic conformational change of C subunit of PKA with different pathways. *Proc. Natl Acad. Sci. USA* **114**, E7959–E7968 (2017).
33. Brown, S. H. J., Wu, J., Kim, C., Alberto, K. & Taylor, S. S. Novel isoform-specific interfaces revealed by PKA RIIBeta holoenzyme structures. *J. Mol. Biol.* **393**, 1070–1082 (2009).
34. Piiadov V., Ara'jo E. A. D., Neto M. O., Craievich A. & Polikarpov I. SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Science* **28**, 454–463 (2019).
35. Murphy, J. M. et al. A robust methodology to subclassify pseudokinases based on their nucleotide binding properties. *Biochemical J.* **457**, 323–334 (2014).
36. Wood D. J. & Endicott J. A. Structural insights into the functional diversity of the CDK-cyclin family. *Open Biol.* **8**, 180112 (2018).
37. Biondi, R. M. et al. Identification of a pocket in the PDK1 kinase domain that interacts with PIF and the C-terminal residues of PKA. *EMBO J.* **19**, 979–988 (2000).
38. Hatzivassiliou, G. et al. RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature* **464**, 431–435 (2010).
39. Rajakulendran, T., Sahmi, M., Lefrancois, M., Sicheri, F. & Therrien, M. A dimerization-dependent mechanism drives RAF catalytic activation. *Nature* **461**, 542–545 (2009).
40. Lange S. M., Nelen M. I., Cohen P. & Kulathu Y. Dimeric structure of the pseudokinase IRAK3 suggests an allosteric mechanism for negative regulation. *Structure* <https://doi.org/10.1016/j.str.2020.11.004> (2020).
41. Lee, K. P. et al. Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing. *Cell* **132**, 89–100 (2008).
42. Littlefield, P. et al. Structural analysis of the EGFR/HER3 heterodimer reveals the molecular basis for activating HER3 mutations. *Sci. Signal* **7**, ra114 (2014).
43. Jacobsen, A. V. & Murphy, J. M. The secret life of kinases: insights into non-catalytic signalling functions from pseudokinases. *Biochem Soc. Trans.* **45**, 665–681 (2017).
44. Murphy, J. M. et al. The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. *Immunity* **39**, 443–453 (2013).
45. Murphy, J. M. et al. Insights into the evolution of divergent nucleotide-binding mechanisms among pseudokinases revealed by crystal structures of human and mouse MLKL. *Biochemical J.* **457**, 369–377 (2014).
46. Scheeff, E. D., Eswaran, J., Bunkoczi, G., Knapp, S. & Manning, G. Structure of the pseudokinase VRK3 reveals a degraded catalytic site, a highly conserved kinase fold, and a putative regulatory binding site. *Structure* **17**, 128–138 (2009).
47. Carrera, A. C., Alexandrov, K. & Roberts, T. M. The conserved lysine of the catalytic domain of protein kinases is actively involved in the phosphotransfer reaction and not required for anchoring ATP. *Proc. Natl Acad. Sci. USA* **90**, 442–446 (1993).
48. Iyer, G. H., Moore, M. J. & Taylor, S. S. Consequences of lysine 72 mutation on the phosphorylation and activation state of cAMP-dependent kinase. *J. Biol. Chem.* **280**, 8800–8807 (2005).
49. Lavoie, H., Li, J. J., Thevakumaran, N., Therrien, M. & Sicheri, F. Dimerization-induced allostery in protein kinase regulation. *Trends Biochem. Sci.* **39**, 475–486 (2014).
50. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
51. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D.* **50**, 760–763 (1994).
52. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
53. Rellos, P. et al. Structure of the CaMKIIdelta/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS Biol.* **8**, e1000426 (2010).
54. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352–367 (2012).
55. Emsley, P. & Cowtan, K. Coot: model building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
56. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D.* **66**, 12–21 (2010).
57. Krissinel E. & Henrick K. in *Computational Life Sciences, Proceedings* (eds Berthold M. R., Glen R., Diederichs K., Kohlbacher O. & Fischer I.) (2005).
58. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
59. Neuwald, A. F. Rapid detection, classification and accurate alignment of up to a million or more related protein sequences. *Bioinformatics* **25**, 1869–1875 (2009).
60. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
61. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
62. Neuwald, A. F. The CHAIN program: forging evolutionary links to underlying mechanisms. *Trends Biochem. Sci.* **32**, 487–493 (2007).
63. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
64. Kirby, N. et al. Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta Crystallogr. D. Struct. Biol.* **72**, 1254–1266 (2016).
65. Chen K., et al. Crystal structure of the hinge domain of Smchd1 reveals its dimerization mode and nucleic acid-binding residues. *Sci. Signal* **13**, eaaz5599 (2020).
66. Trevelyan S. J., et al. Structure-based mechanism of preferential complex formation by apoptosis signal-regulating kinases. *Sci. Signal* **13**, eaay6318 (2020).
67. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
68. Svergun, D., Barberato, C. & Koch, M. H. CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
69. Fischer, H., Neto, M. D., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109 (2010).
70. Wang, C., Bradley, P. & Baker, D. Protein-protein docking with backbone flexibility. *J. Mol. Biol.* **373**, 503–519 (2007).
71. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
72. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
73. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. V., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
74. Murphy, J. M. et al. A robust methodology to subclassify pseudokinases based on their nucleotide-binding properties. *Biochem. J.* **457**, 323–334 (2014).
75. Freeman-Cook, K. D. et al. Design of selective, ATP-competitive inhibitors of Akt. *J. Med. Chem.* **53**, 4615–4622 (2010).
76. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
77. Welburn, J. P. et al. How tyrosine 15 phosphorylation inhibits the activity of cyclin-dependent kinase 2-cyclin A. *J. Biol. Chem.* **282**, 3173–3181 (2007).

## Acknowledgements

We thank staff at the Australian Synchrotron MX2 and SAXS/WAXS beamlines for their assistance in data collection and the New Zealand Synchrotron Group for enabling access; Yee-Foong Mok (Bio21 Institute, University of Melbourne) for assistance with analytical ultracentrifugation experiments. J.M.M. gratefully acknowledges National Health and Medical Research Council of Australia support (Investigator grant 1172929 and IRISS grant 9000653) and the Victorian Government Operational Infrastructure Support Scheme. Funding for N.K. from the National Institutes of Health (R01GM114409) is acknowledged. P.M. gratefully acknowledges support of the New Zealand Royal Society Marsden (grant UOA221).

## Author contributions

M.R.O. and C.R.H. designed and performed experiments, analysed data and co-wrote the paper with J.M.M.; J.R.K. and D.C.G. contributed to protein crystallization, X-ray data collection and structure determination; S.S. and N.K. performed and analysed molecular dynamic simulations, and conducted phylogenetic analyses; L.Y.L., S.N.Y., J.J.S., A.I.W. and I.S.L. performed experiments and analysed data; P.M. and J.M.M. supervised the project and contributed to experimental design and data analysis. All authors commented on the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21191-7>.

**Correspondence** and requests for materials should be addressed to P.M. or J.M.M.

**Peer review information** *Nature Communications* thanks Takayoshi Kinoshita, Kemal Yelekçi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021