The Eleventh Datta Lecture

The structural basis for substrate recognition and control by protein kinases

Louise N. Johnson*, Edward D. Lowe, Martin E.M. Noble, David J. Owen[1]

Laboratory of Molecular Biophysics and Oxford Centre for Molecular Sciences, University of Oxford, Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK

Received 14 May 1998

Abstract Protein kinases catalyse phospho transfer reactions from ATP to serine, threonine or tyrosine residues in target substrates and provide key mechanisms for control of cellular signalling processes. The crystal structures of 12 protein kinases are now known. These include structures of kinases in the active state in ternary complexes with ATP (or analogues) and inhibitor or peptide substrates (e.g. cyclic AMP dependent protein kinase, phosphorylase kinase and insulin receptor tyrosine kinase); kinases in both active and inactive states (e.g. CDK2/cyclin A, insulin receptor tyrosine kinase and MAPK); kinases in the active state (e.g. casein kinase 1, Lck); and kinases in inactive states (e.g. twitchin kinase, calcium calmodulin kinase 1, FGF receptor kinase, c-Src and Hck). This paper summarises the detailed information obtained with active phosphorylase kinase ternary complex and reviews the results with reference to other kinase structures for insights into mechanisms for substrate recognition and control.

Key words: Phosphorylase kinase; Protein kinase; Control mechanism; Protein phosphorylation; Protein crystallography

1. Introduction

Reversible protein phosphorylation on serine, threonine or tyrosine residues is an essential mechanism for regulation of numerous cellular signalling pathways and metabolic functions. Control by phosphorylation is achieved by the protein kinases and protein phosphatases, that catalyse the opposing phosphorylation and dephosphorylation reactions, respectively. Both protein phosphorylation and dephosphorylation were first discovered in 1955 with glycogen phosphorylase (GP) [1,2]. For the next 13 years, protein phosphorylation on serine residues was considered to be a peculiarity of glycogen metabolism, a mechanism for controlling the activities of phosphorylase, phosphorylase kinase (PhK) and glycogen synthase. The discovery of cyclic AMP dependent protein kinase (cAPK) in 1968, with its broad substrate specificity and capability for both serine and threonine phosphorylation, and the discovery of tyrosine phosphorylation by the product of the Rous sarcoma virus *src* gene in 1980, opened the way for the identification of a host of other protein kinases involved in the regulation of cell growth, cell division, cell motility, metabolism, membrane transport, gene expression, learning and memory. While some kinases are highly specific and will only recognise a few target molecules in vivo, others have a broader specificity and once activated, phosphorylate multiple targets. Regulation of protein kinases is achieved through a variety of mechanisms that include phosphorylation and control by additional regulatory domains or subunits. In some instances target substrates are restricted by the sub-cellular localisation of the kinase.

The eukaryotic protein kinases represent one of the largest protein super-families. In the recently sequenced budding yeast genome there are 113 conventional protein kinase genes corresponding to about 2% of the total genome [3] and the estimate that there may be as many as 1000 protein kinases in the mammalian genome [4] still appears valid. In the yeast genome, there are there are several dual specificity kinases but no true tyrosine protein kinases, suggesting that tyrosine kinases evolved in response to the need for cell-cell signalling within multicellular organisms. Similarities in sequence indicate that all protein kinases share a common core of about 270 amino acids [5]. The achievement of different specificity and different mechanisms for control within a common molecular framework of the protein kinases provides a fascinating problem in structural biology. Recent results from protein crystallography have begun to supply the answers. The crystal structures of 12 protein kinases (Table 1) have been determined in both active and inactive conformations. They show that the kinases have a common fold, as anticipated. The structures of active kinases display a similar conformation in key regions involved in ATP and protein substrate recognition but the structures of inactive kinases show considerable differences. This review describes current structural information on the protein kinases, with special reference to the structure of active phosphorylase kinase.

2. Phosphorylase kinase

Phosphorylase kinase (PhK) is a key enzyme involved in the control of glycogen degradation. The enzyme integrates extracellular signals, that arise from hormone receptor interactions and from neuronal impulses mediated through calcium,

*Corresponding author. Fax: (44) (1865) 510454.
E-mail: louise@biop.ox.ac.uk

[1]Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

The previous Datta Lectures were given by: F. Melchers (1st, 1986), N. Sharon (2nd, FEBS Letters 217 (1987) 145–157); B.G. Malmström (3rd, FEBS Letters 250 (1989) 9–21; J.C. Skou (4th, FEBS Letters 268 (1990) 314–324); B.A. Lynch and D.E. Koshland, Jr. (5th, FEBS Letters 307 (1992) 3–9); A.R. Fersht (6th, FEBS Letters 325 (1993) 5–16); E. Sackmann (7th, FEBS Letters 346 (1994) 3–16); P. De Camilli (8th, FEBS Letters 369 (1995) 3–12); C. Weissmann (9th, FEBS Letters 389 (1996) 3–11); Ph. Cohen et al. (10th, FEBS Letters 410 (1997) 3–10).

PII: S0014-5793(98)00606-1

Table 1
X-ray crystal structures of protein kinases

Protein kinase	Inactive state	Active state
a: Serine/threonine kinases		
cAPK	–	Ternary complex with AMPPNP and inhibitor peptide[a] [6,7]
CDK2	Apo CDK2 and binary complex with ATP [8,9]	Partially active CDK2/cyclin A complex [10] Active phospho-CDK2/cyclin A complex with ATP [11]
MAPK	Apo ERK2 and binary complex with ATP [12]; p38 MAPK[13,14]	Active phospho MAPK [15]
PhK	–	Binary complex with ADP and AMPPNP; ternary complex with AMPPNP and peptide substrate[16,17]
CK1	–	*S. pombe* CK1 with ATP [18]; Truncated mammalian CK1 [19]
Twitchin kinase	Auto-inhibited form [20,21]	–
CaM kinase 1	Auto-inhibited form [22]	–
b: Tyrosine kinases		
IRK	Apo inactive IRK [23]	Active phospho-IRK with AMPPNP and peptide substrate [24]
FGFR1K	Apo FGFR1K and binary complex with AMPPCP [25]	–
Lck	–	Active phospho-Lck [26]
c-Src	Apo Src with regulatory SH2 and SH3 domains and auto-inhibited by phosphorylation [27,28]	–
Hck	Apo Hck with regulatory SH2 and SH3 domains and auto-inhibited by phosphorylation and binary complex with ATP [29]	–

Kinases: cAPK: cyclic AMP dependent protein kinase; CDK2: cyclin dependent protein kinase 2; MAPK: mitogen activated protein kinase; ERK2: extracellular regulated kinase 2; PhK: phosphorylase kinase; CK1: casein kinase 1; CaM kinase 1: calcium/calmodulin dependent protein kinase 1; IRK: insulin receptor kinase; FGFR1K: fibroblast growth factor receptor 1 kinase; Lck: lymphoid cell kinase; c-Src: cellular kinase named after oncogene of Rous sarcoma virus; Hck: haematopoietic cell kinase. ATP analogues: AMPPNP: adenylyl imidophosphate; AMPPCP: adenylyl diphosphonate.
[a]Structures of other cAPK complexes are summarised in [30].

with those that arise from intracellular events, and provides a tightly controlled kinase activity which activates glycogen phosphorylase (reviewed in [31]). Phosphorylase kinase is one of the largest of the protein kinases and is composed of four types of subunit, with stoichiometry $(\alpha\beta\gamma\delta)_4$ and a total molecular weight of 1.3×10^6 Da. The α and β subunits are regulatory and are the targets for control by phosphorylation by cAPK. The δ subunit is essentially identical to calmodulin and confers calcium sensitivity. The 386 residue γ subunit is the catalytic subunit which comprises an N-terminal catalytic region and a C-terminal calmodulin binding region. The fragment of the γ subunit comprising residues 1–298 has been expressed and crystallised and the crystal structure solved [16] to give a structure of a fully active protein kinase that is constitutively active without the requirement for activation by phosphorylation or other effectors.

In muscle and liver, phosphorylase kinase catalyses the phosphorylation of inactive phosphorylase b (GPb) to active phosphorylase a (GPa) through the phosphorylation of a single serine residue, serine 14. In vivo phosphorylase is the only recognised substrate of phosphorylase kinase, but in vitro the enzyme will catalyse phosphorylation of a number of other proteins and peptide substrates with decreased efficiency compared with the natural substrate. Analysis of sequences surrounding the sites of phosphorylation indicate a consensus sequence of R/KXXS*V/IZ where S* is the serine phosphorylated and X and Z may be any amino acid but activity is increased if Z is an arginine as in phosphorylase. Peptide substrates exhibit K_m values (about 1 mM) that are higher than the K_m observed for phosphorylase (values range from 44 to 270 μM for the holoenzyme; and from 9 to 27 μM for the truncated γ subunit) suggesting that the phosphorylase molecule may make additional contributions to recognition and affinity in regions distinct from the immediate surroundings of the serine phosphorylated [32]. Indeed only one of the 29 serines in phosphorylase is phosphorylated, although some other serines are surrounded by a consensus sequence motif.

3. Protein kinase core structure

The core of all eukaryotic protein kinases adopt a common fold, first shown with the structure of cAPK [33] and later with other protein kinase structures [34]. Surprisingly this fold is also adopted by an aminoglycoside kinase, an enzyme from pathogenic drug-resistant bacteria of the Enterococci and Staphylococci genera that has less than 6% identity to the eukaryotic protein kinases [35]. The fold is illustrated for PhK [16,17] in Fig. 1. The N-terminal lobe (residues 18–105 in PhK numbering) is comprised of five strands of antiparallel β sheet (β1–β5) and one α-helix, the C helix (shown white in Fig. 1). The larger C-terminal lobe (residues 110–292) is comprised of a four helix bundle (α-helices D, E, F and H) flanked by two additional helices (G and I) and two short antiparallel β sheets (β7 with β8 and β6 with β9) that occur between helices E and F (shown salmon in Fig. 1). The two lobes are connected by a hinge region, residues 105–110 (shown magenta). The ATP binding site is situated at the interface of the 2 lobes while the peptide substrate (green in Fig. 1) is located mostly in the C-terminal lobe. The loop between strands β1 and β2 contains the glycine loop that is important for localising the phosphates of ATP. The catalytic loop (residues 145–154), which includes the strand β6 (gold in Fig. 1), contains the conserved aspartate, Asp^{149} (Asp^{166} in cAPK), that is presumed to be a catalytic residue, and Asn^{154} (Asn^{171} in cAPK), a metal binding residue for ATP binding. A key aspect of regulation that has received considerable attention, because of its function in signal cascade processes, is phosphorylation on a residue (or residues) located in the activation segment at the centre of the kinase domain [36]. The activation segment (residues 167–191 in PhK following β8 and including β9) (shown cyan in Fig. 1) is defined as the region

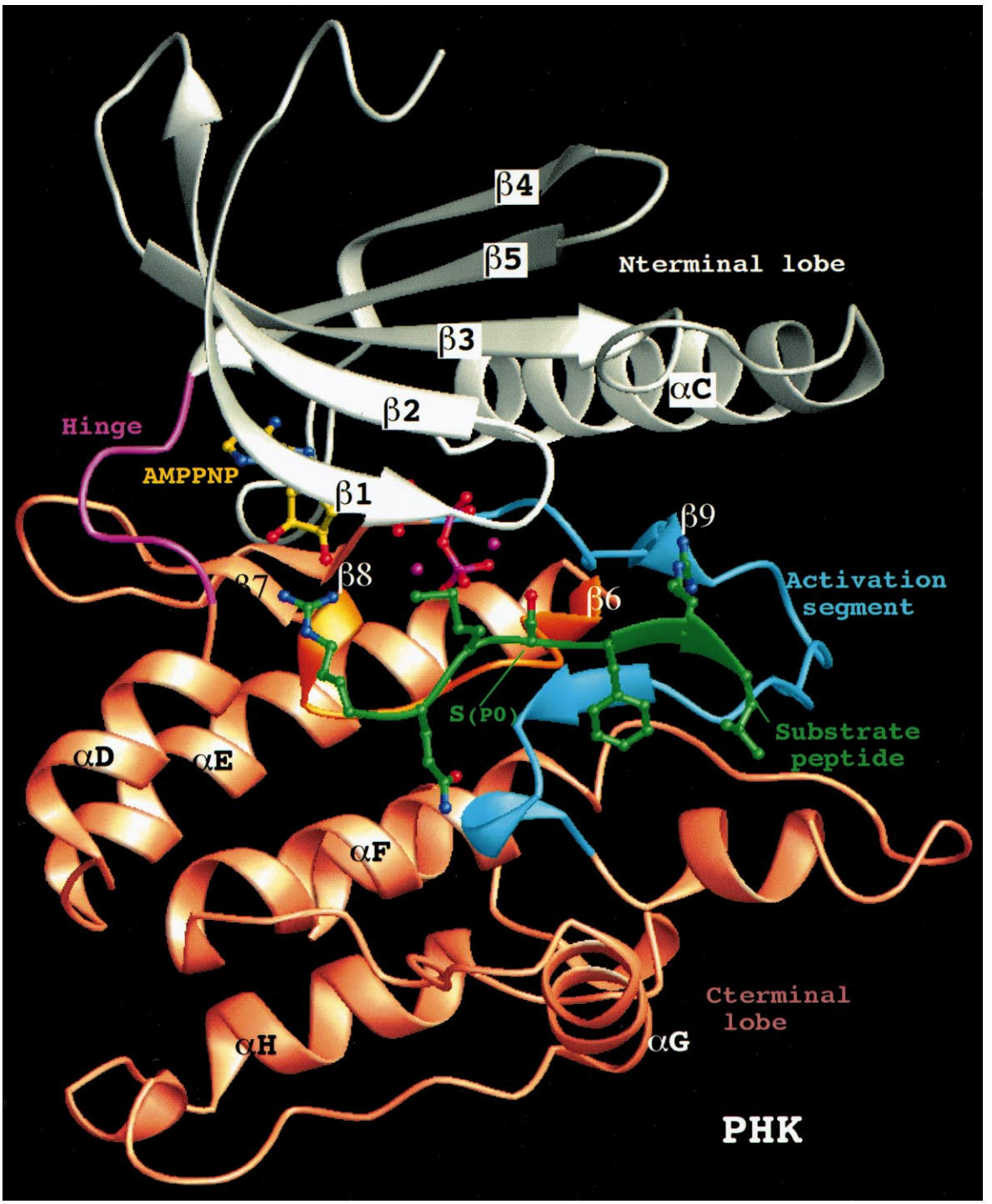


Fig. 1. Schematic diagram of the fold of phosphorylase kinase (PhK), illustrating a typical protein kinase fold. The N-terminal lobe is shown in white, the hinge region in magenta and the C-terminal lobe in salmon. The catalytic loop is shown in gold and the activation segment in cyan. α-Helices are labelled from αC to αI (αI is obscured by αF). β-Strands are labelled β1–β9. The substrate peptide is shown in green with all residues displayed and the serine to be phosphorylated marked by S(P0). The short antiparallel β-sheet between the peptide substrate and the activation segment is indicated by arrows. AMPPNP and two manganese atoms (magenta) are shown with all atoms displayed.

spanning the conserved sequences DFG (residues 167–169 from sub-domain VII) and APE (residues 191–193 from sub-domain VIII) (using the single letter amino acid code and the sub-domain nomenclature of [5]). Residue Asp[167] (residue Asp[184] in cAPK; the D of the DFG motif) is a highly conserved residue in all protein kinases and is involved in binding an ATP chelating metal. The length of the activation segment varies from 19 residues in cAPK, 21 residues in PhK and up to 32 residues in other kinases. The central part of this region shows little sequence conservation and yet the segment plays a crucial role in the activation of different kinases. PhK is active without the requirement for phosphorylation of residues in the activation segment. The residue that is phosphorylated in other kinases (e.g. Thr[197] in cAPK, Thr[160]

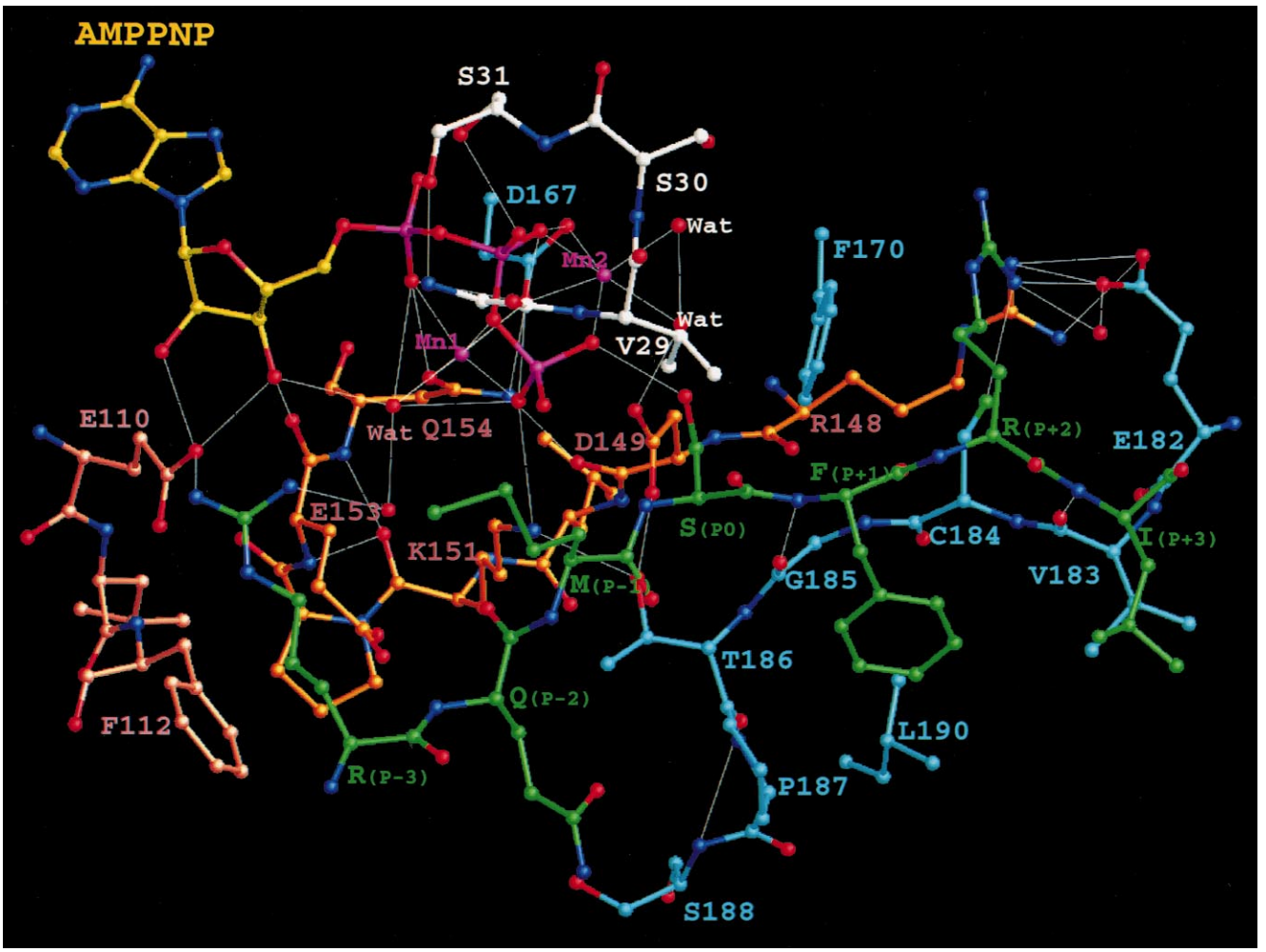


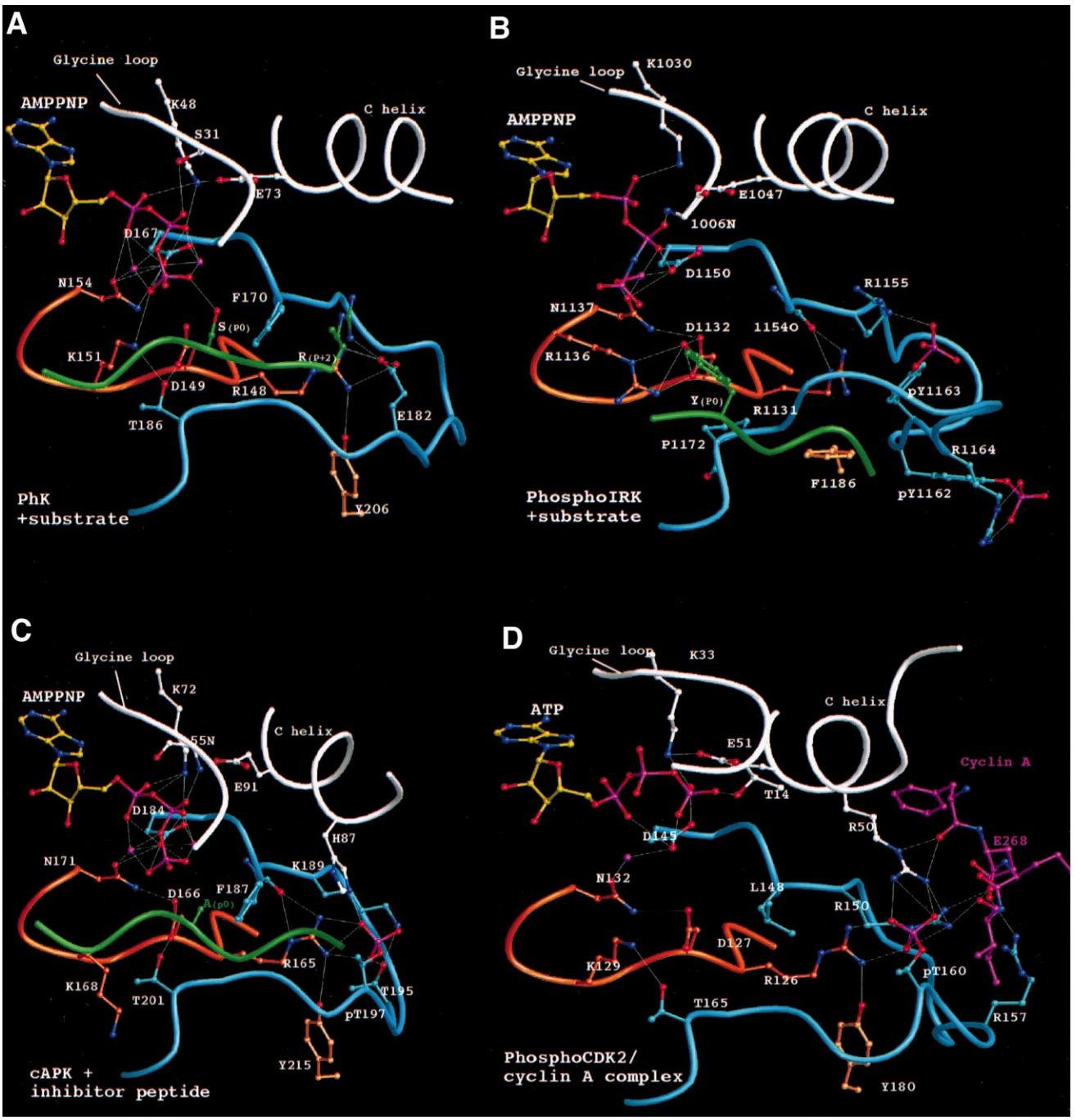
Fig. 2. Details of AMPPNP and peptide substrate binding to PhK. The single letter amino acid code is used to label residues. Polar contacts less than 3.3 Å are shown as thin white lines. Waters are labeled Wat and shown as red spheres. The view is similar and the colour scheme is identical to that in Fig. 1. The contact from Lys^{48} to the α and β phosphates has been omitted for clarity. For further details see text.

→

Fig. 3. Details for four active protein kinases (PhK, IRK, cAPK and CDK2/cyclin A) showing the interdependence of the ATP binding regions, the substrate peptide interactions and the activation segment. The kinase structures were superimposed using the C-terminal lobes as a reference and are displayed in identical views similar to that of Fig. 1 and with the same colour scheme (residues from the N-terminal lobe are in white, from the catalytic loop in gold, from the activation segment in cyan and from the peptide inhibitor or substrate in green). Water molecules are not shown. A: PhK/AMPPNP/peptide substrate ternary complex showing the key interactions of the triphosphate groups with residues from the N-terminal lobe (including the interaction of the lysine, Lys^{48}, with the glutamate, Glu^{73}, from the C helix) and the C-terminal lobe. The activation segment is stabilised by an interaction between Glu^{182} and Arg^{148}, the arginine that precedes the catalytic aspartate, Asp^{149}, that also interacts with Tyr^{206}. The peptide substrate and the interactions of the serine S(P0), that is to be phosphorylated, with the γ phosphate of AMPPNP and the interactions of Arg(P+2) with Glu^{182} are shown. B: IRK/AMPPNP/peptide substrate ternary complex showing key interactions as above. The triphosphate groups are in a slightly different positions to PhK. The activation segment is stabilised by the interactions of $pTyr^{1163}$ with Arg^{1155}. $pTyr^{1163}$ is more distant from the arginine preceding the catalytic base, Arg^{1131}, than the equivalent residue, Glu^{182}, in PhK. The second phosphotyrosine $pTyr^{1162}$ contacts Arg^{1164} and is exposed, as is $pTyr^{1158}$ (not shown). The non-polar residue (Met^{1153}), which interposes between the catalytic base and the preceding arginine in all protein kinases (e.g. Phe^{170} in PhK), is not shown for clarity. The activation segment holds the substrate peptide further away from the kinase compared with PhK and allows the longer tyrosine residue, (Y(P0)), to reach the catalytic aspartate, Asp^{1132}. C: cAPK/AMPPNP/PKI ternary complex. The activation segment is stabilised by interactions between $pThr^{197}$ and His^{87}, Arg^{189} and Arg^{165}, the last residue being the arginine that precedes the catalytic base, Asp^{166}. The position of His^{87} is similar to the position of Arg(P+2) of the substrate peptide in PhK. The inhibitor peptide has an alanine (AP(0)) in place of serine at the phosphorylation site. In both cAPK and PhK the catalytic aspartate interacts with a threonine (Thr^{186} in PhK and Thr^{201} in cAPK), an interaction that is not possible with IRK where this residue is a proline. D: Phospho-CDK2/cyclin/ATP complex without substrate peptide. The cyclin atoms, which interact with the activation segment, are shown in magenta. The phosphothreonine, $pThr^{160}$, interacts with Arg^{50}, Arg^{126} and Arg^{150}. The interactions of Arg^{50} and Glu^{51} (the R and E from the C-helix PSTAIRE motif) are key features of the activation process. Both Arg^{50} and Arg^{150} make hydrogen bonds to groups in cyclin A. Structures of activated CDK2 with substrate peptides have not yet been determined and may promote further changes for the ATP. For further details see text.

in CDK2 and Thr[183] in MAPK) is a glutamate in PhK, Glu[182].

Open and closed lobe structures have been identified for cAPK from solution scattering and X-ray crystal structures [37,38]. Numerous crystallographic studies on binary and ternary complexes of cAPK, summarised by Narayana et al. [30], provide evidence for flexibility between lobes in response to substrate binding via an induced fit mechanism. cAPK is permanently activated by autophosphorylation on Thr[197] and these open and closed structures represent fluctuations of an active kinase structure in which there is little conformational change within the lobes. In superposition studies, the C-terminal lobe is taken as the reference lobe and the open or closed state is characterised by the relative disposition of the N-terminal lobe. The ternary complex of cAPK with protein kinase inhibitor (PKI) and AMPPNP represents the most closed conformation. A binary complex with a modified PKI represents the most open cAPK structure (by about 11°). To a first approximation the lobe shift can be visualised as a rotation close to the hinge point between the two lobes, although the movement is more complex and may contain shifts of the glycine loop with respect to the β-sheet of the N-terminal lobe and translations. Computational energy calculations suggest that the binding of the inhibitor peptide and ATP are energetically additive but that it is the binding of the inhibitor peptide that drives the final closure of the two lobes [39]. Studies on other kinases in the active conformation show similar variations of lobe orientations. Thus constitutively ac-



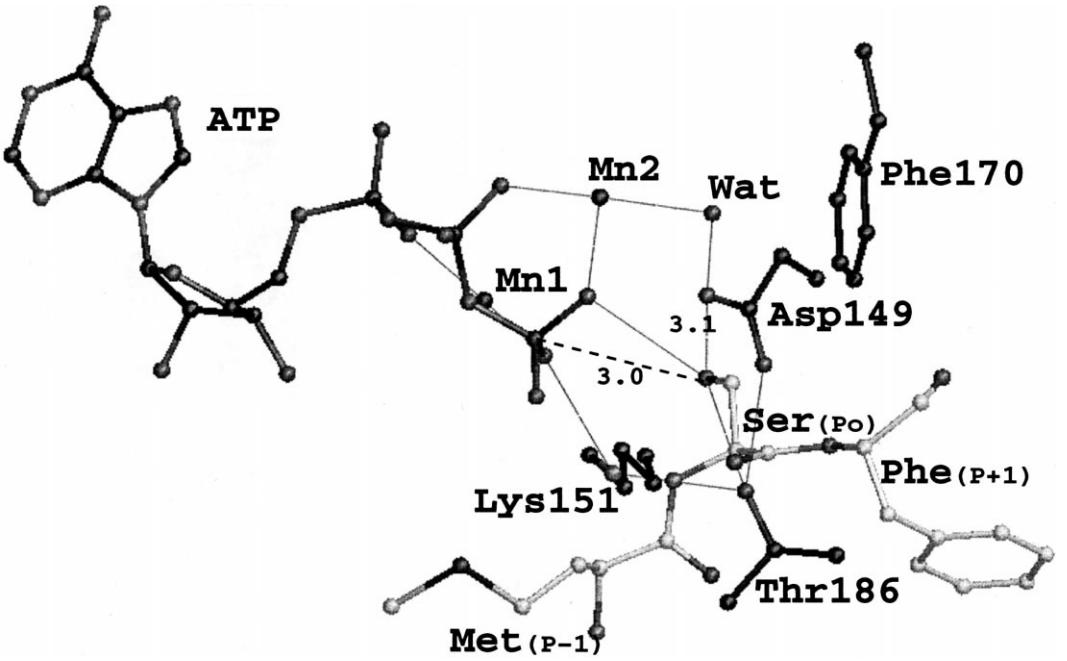


Fig. 4. Proposed geometry of catalytic groups and substrates for phospho transfer from ATP to serine substrate in PhK. The substrate serine has been rotated about the Cα-Cβ bond from the gauche$^-$ position seen in the crystal structure (e.g. as shown in Figs. 2 and 3A) to the gauche$^+$ position where it is now within hydrogen bonding distance of the proposed catalytic residue, Asp^{149}. The rotation also bring the OG atom of the serine more directly in line to promote attack on the γ phosphate of ATP. The aspartate, Asp^{149}, is in a buried environment shielded by Phe^{170} but its hydrogen bonding capacity is fully satisfied by hydrogen bonds from OD2 to Thr^{186} to Lys^{151} and from OD1 through water to Mn2 that chelates the β and γ phosphates. Such an environment could result in a pK_a for the aspartate similar to that observed for exposed aspartates (perhaps slightly lower because of the proximity of the positive charges or perhaps elevated by the non-polar environment). The arrangement is consistent with a mechanism in which the aspartate acts to hydrogen bond with or to abstract a proton from the OH group of the serine and the alcoholate ion promotes nucleophilic attack on the γ phosphate of ATP. The negative charges developed on the γ and β phosphates during phospho transfer are stabilised by the metal ion contacts and other interactions. Although plausible, the mechanism remains to be definitively established and the details of whether the process is associative or dissociative elucidated.

tive PhK in the binary complex with nucleotide is more open (5°) and PhK in the ternary complex is only slightly more open (2°) than the ternary cAPK complex [17]. Active CK1 from *Schizosaccharomyces pombe* in binary complex with ATP [18] is closed similar to cAPK but mammalian apo-CK1δ (56% identical to *S. pombe* CK1) is even more closed than ternary cAPK [19]. Activated apo-phospho-MAPK [15] and activated apo-phospho-Lck [26] exhibit structures that are 11–12° more open than ternary cAPK complex, consistent with the notion that substrates are required to induce the final stages of closure.

The kinases in their inactive conformations exhibit a wide range of lobe orientations, both open and closed. Inactive FGFR1K [25], IRK [40] and MAPK [12] exhibit very open structures (20°, 26° and 17°, respectively) compared with the ternary cAPK structure. On activation of IRK by triple phosphorylation on the activation segment and formation of the active ternary phospho-IRK complex with AMPPNP and peptide substrate, the lobes close by 21° [24], although the structure is still more open than in the ternary complex of cAPK. The two auto-inhibited kinases CaMK1 and twitchin kinase, each have a C-terminal extension that carries either a calmodulin binding site or S100 binding site, respectively. The C-terminal extension occupies the substrate peptide binding groove in a pseudo-substrate binding mode. In CaMK1 [22] the short tail region wraps around causing severe distortions of the N-terminal lobe while in twitchin kinase [20,21] the longer tail occupies part of the ATP binding site. Both structures have open conformations (18° for CaMK1 and 30° for twitchin kinase). The inactive conformation of CDK2 [8] exhibits a closed structure in which the activation segment partially blocks the peptide substrate binding site and there is a significant displacement of the C-helix in the N-terminal lobe compared with its conformation in the active kinase structures. On activation by cyclin A binding and by phosphorylation on Thr^{160} there are profound changes in the conformation of CDK2 that result in a more open conformation (by 14°). In addition there are shifts (of about 7 Å) and a rotation in the C-helix and a shift of 12–21 Å in residues of the activation segment [11]. The structures of the two kinases from the Src family, c-Src [27,28] and Hck [29], represent the down regulated state in which a phospho-tyrosine in the C-terminal tail is bound to the SH2 recognition site. Their conformations are more similar to the closed conformation of cAPK but there is also a displacement of the C-helix, as in the inactive conformation of CDK2.

Williams et al. [28] have shown that, if the lobe orientation is defined solely with respect to the core structure of the two lobes (the antiparallel β sheet for the N-terminal lobe and the α-helix core for the C-terminal lobe), the path from open to closed is not a random process but one that follows a defined route for the kinase crystal structures analysed. Three regions act as hinge points, the start and end of the loop from α-C to β4 and the loop between the lobes. The pliability between the lobes is exploited in the active kinase state for substrate recognition via an induced fit mechanism that also has significance for product release. In the conformations of the inactive kinase, it is not only lobe rotation that is exploited but other

alterations, such as those in the C helix and activation segment, that ensure that the kinase is inactive and cannot be simply activated by equilibrium fluctuations of the lobes. In inactive IRK the interactions of the DFG conserved sequence with the glycine rich loop keep the lobes open while in MAPK interactions from the DFG region with residues from the αC helix keep the lobes open. The correct disposition of the two lobes is a key feature of the activation process that may be triggered for example by cyclin binding to CDK2, or phosphorylation on the activation segment in MAPK and IRK. The structural results show that significant movement of other elements, are required to bring these kinases into their correct conformation.

4. ATP recognition

In PhK, the ATP binding site has been identified in studies with ADP and the non-hydrolysable ATP analogue, adenylyl imidophosphate, AMPPNP. An identical binding mode for AMPPNP in the binary complex and in the ternary complex with peptide substrate [17] was observed that is representative of nucleotide binding to the other kinases in their active conformations. AMPPNP binds in a crevice making contacts through the adenine base to main chain atoms close to the hinge region between the two lobes (Asp^{104} carbonyl oxygen) and (Met^{106} amide nitrogen) (Fig. 1). There are a number of hydrophobic interactions with the base from residues Leu (β1), Val (β2) and Ala (β3), Ile (β5), Phe (hinge) and Met from the C-terminal lobe. The ribose hydrogen bonds to residues in the C-terminal lobe (Glu^{110} side chain and Glu^{153} main chain carbonyl oxygen) (Fig. 2). The location of the triphosphate moiety is crucial for protein kinase catalysis and in PhK, as in other active protein kinases, is promoted by four major interactions that involve residues from both lobes (Fig. 2). (i) Two metal ions are involved. In crystallographic work with PhK, studies on ADP-Mg^{2+} and AMPPNP-Mn^{2+} complexes show that the Mg^{2+} and Mn^{2+} ions occupy almost identical sites. One metal site (Mn1 in the complex of PhK with AMPPNP) is coordinated by the terminal oxygens from the α and γ phosphates of AMPPNP, the bridging nitrogen between the β and γ phosphates, the amide oxygen of the side chain of the conserved asparagine, Asn^{154} (Asn^{171} in cAPK), from β7, and a carboxyl oxygen from the conserved aspartate, Asp^{167} (Asp^{184} in cAPK; the D of the DFG sequence at the start of the activation segment), and a water molecule. The second site, Mn2, is coordinated by the terminal oxygens of the β and γ phosphates, two carboxyl oxygens of Asp^{167} and two water molecules. (ii) Lys^{48} (Lys^{72} in cAPK) from β3 of the N-terminal lobe contacts terminal oxygens from the α and β phosphates. This conserved lysine is localised by a contact to the conserved glutamate, Glu^{73} (Glu^{91} in cAPK) from the C helix. Thus the C helix plays an indirect but important role in localisation of the phosphates. (This contact is not shown in Fig. 2, for clarity, but is illustrated in Fig. 3A.) (iii) The glycine loop from the β1 to β2 connection is not involved in PhK in direct contacts through main chain amides to the phosphates, as it is in cAPK. The third glycine of the 'glycine loop' is a serine residue in PhK. The serine side chain hydrogen bonds to the β phosphate (Fig. 2) and the glycine loop is slightly more open than in cAPK. (iv) Lys^{151} (Lys^{168} in cAPK) contacts the γ phosphate.

The positions of AMPPNP seen in active PhK and those of ATP or AMPPNP seen in active cAPK are closely similar and provide support for the notion that this constellation represents the ATP bound poised for catalysis (Fig. 3A,C). In the structure of the ternary complex of activated phospho-IRK with AMPPNP and peptide substrate, there are differences in the kinase contacts to the phosphate groups [24] (Fig. 3B). The lysine from β3 (Lys^{1030}) is hydrogen bonded to just the α phosphate and the contact of the lysine to the glutamate on the C-helix is too long for a hydrogen bond. Although the metal ions and their chelating aspartate and asparagine residues occupy similar positions to those in PhK and cAPK, the positions of the phosphates differ in their torsion angles at the link from the ribose to the α phosphate so that the co-ordination of the metals by the phosphates is different. Metal 1 is coordinated by the β and γ phosphates and metal 2 by just the γ phosphate. It remains to be seen if these differences are characteristic of differences between tyrosine kinases and serine/threonine kinases. In the activated phospho-CDK2/cyclin A complex with ATP [11] the lysine/glutamate pair is made but there are differences in the conformation of the phosphates compared to those seen in the ternary complexes of cAPK and PhK (Fig. 3D). The differences result in a contact of the lysine (Lys^{33}) to the γ phosphate and there is only one metal site that is closest to the α phosphate oxygen. It may be that the presence of protein substrate is required to help orient the ATP.

In binary complexes of activated kinases with nucleotide, the interactions of the adenine and ribose components of ATP, ATP analogues or ADP are similar to those described above but there are slight differences in the phosphate positions. In PhK binary complex with ADP, derived from hydrolysis of ATP, there is a change in phosphate positions such that Mg1 links the α and β phosphates and Mg2 binds the β phosphate and there is a closure of the glycine loop [16]. These changes result in more extensive contacts to the nucleotide diphosphate than the nucleotide triphosphate complex which may explain why release of product ADP appears to be the rate limiting step in kinase reactions.

The inactive kinase conformations show considerable differences in their ATP binding sites. The results indicate that a key part of the activation process involves the correct binding of the triphosphate moiety of ATP. In the inactive IRK, twitchin kinase and CaMK1 structures the nucleotide recognition site is blocked by residues from the activation segment, from the C-terminal tail or from N-terminal lobe residues, respectively. However ATP binding has been observed to both the closed conformation of inactive CDK2 [8] and Hck [29] and the open conformations of inactive MAPK [12] and FGFR1K [25]. The interactions to the adenine are similar to those that are seen in ternary complexes with active kinases, indicating that the specific hydrogen bonds made by the adenine to main chain residues are sufficient to direct specificity. These inactive binary nucleotide-kinase complexes are characterised by the failure to localise the phosphate groups or by the phosphate groups held in incorrect positions for catalysis. In CDK2, Hck and MAPK the lysine from β3 and the aspartate from the DFG at the start of the activation segment are in the wrong positions to contact the phosphates or metal ions, indicating the importance of domain orientation, C-helix localisation through the glutamate/lysine pair and activation segment interactions for the promotion of the correct recog-

nition site. In FGFR1K the lysine to glutamate interaction of the C-helix is made but the very open structure results in loss of contacts to the ribose, the γ phosphate is disordered and there are no metals bound.

The C-helix plays a key role in correct positioning of the ATP phosphates, although it is not involved in direct phosphate interactions. In inactive CDK2 and inactive down regulated Src and Hck, the C-helix is swung outwards and rotated by about 90° relative to its position in the active kinase structures. Thus the ion pair between the glutamate from the C-helix and the lysine, that chelates the α and β phosphates, is not made. In Src and Hck, interactions at the C-terminal end of the C-helix with the linker region between the SH2 domain and the kinase domain contribute to the inactive conformation [27–29]. In CDK2, interactions at the N-terminal end of the C-helix on binding the activatory cyclin A result in a rotation and shift of the C-helix to its active conformation [10]. The previously exposed glutamate is driven inwards towards the phosphate binding lysine and the non-polar residues and the arginine of the PSTAIRE motif of CDK2's C-helix interact with the cyclin A.

5. Peptide substrate binding

We found that attempts to co-crystallise PhK with peptide substrates corresponding to residues 9–18 from the natural substrate phosphorylase were not successful, possibly because PhK exhibits low affinity for such substrates ($K_m \approx 1$ mM). In a study of protein kinase specificity, Songyang et al. [41] identified an optimal peptide substrate (sequence KRMMS*FFLF) from analyses with an oriented degenerate peptide library. On the basis of this result and from knowledge of the catalytic site of PhK, we designed a modified peptide (RQMS*FRL) that corresponded more closely to the natural substrate, fitted the space available in the crystal, and was soluble at the high concentrations required for crystallisation. Kinetic analysis showed for the peptide RQMS*FRL, $k_{cat} = 400$ min^{-1} and $K_m = 0.4$ mM compared to values for the natural peptide KQIS*VRG, $k_{cat} = 70$ min^{-1} and $K_m = 1.8$ mM [17].

The crystal structure of the ternary complex of PhK with AMPPNP and the peptide substrate in the presence of manganese [17] showed that the peptide bound to the active site groove associated with the C-terminal lobe. PhK recognition of the peptide is achieved through marked complementarity of shape, hydropathy and electrostatic potential that provide an explanation for the specificity of PhK and the higher affinity of the modified peptide (Fig. 2). Following convention, the phosphorylatable serine is identified as sub-site P0; residues N-terminal to this are indicated as P−1, P−2 etc. and those C-terminal from the serine as P+1, P+2, etc. The most extensive non-polar interactions occur between Phe(P+1) and a hydrophobic pocket created by the activation segment residues Val^{183}, Pro^{187} and Leu^{190}. These contacts explain the preference for a non-polar residue at the +1 sub-site. Leu(P+3) also contacts a non-polar pocket but one of the residues is generated from a crystallographic two-fold related molecule and this contact may or may not be relevant in vivo. Arg(P−3) forms an ion pair with Glu^{110}, a residue close to the hinge region that also interacts with the O2′ and O3′ ribose oxygens of the nucleotide substrate and this interaction explains the specificity of PhK for basic residues in the −3 position. Arg(P+2) has some conformational mobility in the crystal and in its best defined position makes contact with Glu^{182} from the activation segment. Glu^{182} plays an analogous role to the phosphorylated residue in other protein kinases that are controlled by phosphorylation on the activation segment, as discussed later. The most striking aspect of the polar interactions between the peptide substrate and PhK is a short stretch of antiparallel β-sheet formed by Phe(P+1) and Leu(P+3) from the peptide and residues Gly^{185} and Val^{183} from the activation segment of PhK. This interaction is dependent upon a defined conformation of the activation segment and implicates this region in a critical role in substrate recognition.

The binding of the peptide substrate to PhK closely resembles the binding of PKI to cAPK [6,7] in the conformation and interactions of residues in the positions P−3 to P+1. In particular the interactions and conformation of Arg(P−3) are identical in the two kinases. However there is a significant difference in the conformation of residues in the P+2 and P+3 positions. In cAPK, the site for interaction of the (P+2) residue with the equivalent phospho-threonine 197 is occupied by a histidine (His^{87}) from the N-terminal lobe in cAPK. Consequently the antiparallel sheet interaction of the inhibitor peptide with the activation segment is not seen in cAPK (Fig. 3A,C).

Comparison with the structure of the active phospho-IRK in complex with AMPPNP and peptide substrate leads to insights into specificity determinants of serine and tyrosine kinases. In IRK, six residues (GDYMNM) from a putative insulin receptor phosphorylation site have been localised in subsites P−2 to P+3. On activation of IRK by triple phosphorylation at residues Tyr^{1158}, Tyr^{1162} and Tyr^{1163} there is a dramatic shift in the activation segment (shifts of up to 30 Å) that displaces the segment from the position in which it blocks the catalytic site to a conformation in which it creates a substrate recognition site, similar to that observed for active cAPK, PhK, active phospho-CDK2/cyclin A and the active phospho-Lck conformations [24]. In the peptide substrate complex, residues at the P+1 and P+3 positions form a short anti-parallel β-sheet with Leu^{1171} and Gly^{1169} from the activation segment of IRK [24], making similar interactions to those observed for the PhK peptide complex. The activation segments of PhK and IRK superimpose approximately for residues 167–181 (1150–1164 in IRK). Thereafter the activation segment of IRK, which is three residues longer, follows a different path form that of PhK (Fig. 3A,B). The two chains superimpose again at residue 189 (1175 in IRK). Residues 1171–1172 in IRK are held further away from the catalytic site than the corresponding residues in PhK thus allowing the larger tyrosine residue to be accommodated. A tyrosine could not be accommodate in the PhK conformation. A threonine is conserved (Thr^{186} in PhK) in all serine/threonine protein kinases and this residue appears to play a role in determining the specificity for serine. Thr^{186} is hydrogen bonded to the catalytic residue, Asp^{149} (discussed later), and the lysine, Lys^{151}, that binds to the γ-phosphate of AMPPNP (Fig. 3A). In IRK, and all tyrosine kinases, the equivalent residue is a proline, Pro^{1172}. A proline at this site provides a non-polar surface for the tyrosine substrate and appears to influence the conformation of the activation segment creating a more open site for tyrosine at this critical position for substrate binding.

6. The role of the activation segment and phosphorylation in protein kinases

The results with PhK, IRK and cAPK show that the activation segment plays a crucial role in substrate recognition and in promoting the correct environment for the catalytic residues. All protein kinases contain a conserved aspartate (Asp^{149} in PhK) that has been implicated in the catalytic mechanism. In many, but not all, protein kinases the catalytic aspartate is preceded by an arginine residue. In PhK, this arginine, Arg^{148}, makes an ion pair with Glu^{182} from the activation segment and Tyr^{206} (Fig. 3A). These and other interactions serve to create the correct orientation of the activation segment to recognise substrate and to promote the correct environment for Asp^{149}, which is shielded from the arginine by a non-polar residue, Phe^{170}. The activation segment plays a critical control in control of protein kinase activity by phosphorylation [36]. In cAPK, which is activated by autophosphorylation of Thr^{197}, the phospho-threonine forms an ion pair with the arginine (Arg^{165}) that precedes that catalytic aspartate (Asp^{166}), and makes additional contacts to Lys^{189}, from the start of the activation segment and to His^{87} from the N-terminal lobe α-C helix (Fig. 3C). The cluster of three basic residues (His^{87}, Arg^{165} and Lys^{189} in cAPK) demand neutralisation by a dianionic phosphate group in order to maintain their relative orientations. By contrast, in PhK there is only one residue, Arg^{148}, that requires compensation and this can be accomplished by the glutamate, Glu^{182}.

The importance of the interactions from the phospho-amino acid in the activation segment for activation is emphasised from a comparison of kinases whose structures are known in their inactive and active phosphorylated conformations. In inactive IRK, as described above, the activation segment blocks the ATP binding site. The large changes observed on the triple phosphorylation at Tyr^{1158}, Tyr^{1162} and Tyr^{1163} lead to the creation of the ATP binding site and the peptide substrate binding site [24]. P-Tyr^{1163} contacts Arg^{1155} from the start of the activation segment (similar to Lys^{189} in cAPK) but makes a rather long contact to Arg^{1131}, the arginine that precedes the catalytic base (Fig. 3B). In inactive CDK2 the activation segment makes contact with the N-terminal lobe [8]. On binding cyclin A (which results in 0.3% of maximal activity), significant changes take place that bring the C-helix to its correct conformation, the domains open, and the activation segment is reoriented so that a glutamate, Glu^{162}, contacts the arginine preceding the catalytic base [10]. On phosphorylation of Thr^{160} in the CDK2/cyclin A complex a fully active kinase is obtained in which pThr^{160} is directed in towards the arginine cluster composed of Arg^{50} from the N-terminal lobe C-helix, Arg^{126} (preceding the catalytic base) and Arg^{150} from the start of the activation segment [11] (Fig. 3D). This readjustment of the activation segment results in a 300-fold increase in catalytic activity observed between CDK2/cyclin A complex and phospho-CDK2/cyclin A complex and is significant for creation of the peptide substrate recognition site. In MAPK, phosphorylation on Thr^{183} and Tyr^{185} leads to a significant rearrangement of the activation segment and lobe closure [15]. The phospho threonine, pThr^{183}, is directed in to contact three arginines equivalent to those in CDK2: Arg^{68} from C-helix, Arg^{146} preceding the catalytic base and Arg^{170} in the activation segment. The pTyr^{185} is directed outwards and contacts further arginine residues that form an anion binding site in non-phosphorylated MAPK. The changes in Tyr^{185} create a putative proline recognition site for the P+1 residue of substrate peptide, similar to that observed for activated CDK2, and thus providing a possible explanation of the specificity of both kinases for a proline following the site of phosphorylation. In active Lck, the phosphotyrosine residue pTyr^{394} interacts with two arginines, Arg^{363} (preceding the catalytic base) and Arg^{387} from the activation segment but comparison with the active IRK ternary complex suggests the C-terminal end of the activation segment may not be properly positioned for peptide substrate binding.

7. Comparison of the conformation of the bound substrate peptide with that in the native GP structures

The structures of both the inactive non-phosphorylated form of phosphorylase, GPb, and the active phosphorylated form, GPa, are known [42,43]. We can ask which of the conformations, that of the substrate or that of the product, does the phosphorylase kinase recognise? The answer is that the conformation of the peptide bound to the catalytic site of PhK resembles neither the conformation seen in GPb or GPa. The bound peptide exhibits an extended conformation. In GPb, the N-terminal residues around the site of phosphorylation (Ser^{14}) adopt an irregular conformation in which residues 8–12 and 14–16 are α-helical [44]. In GPa the conformation around Ser^{14} is extended but the conformations of residues 15–17 are α-helical and the Ser^{14} phosphate group hydrogen bonds to the main chain nitrogens of residues 15 and 16. Thus the kinase tailors the conformation of the peptide to adopt a conformation that is characteristic of neither the natural substrate nor the natural product. The N-terminal residues in GPb exhibit high mobility in the structure and it appears that they can readily modify their conformation to fit the kinase. Rapid release of the phosphorylated protein or peptide product is a characteristic of most protein kinase reactions. The fact that the observed conformation adopted in the phosphorylated protein is different from the conformation of the bound product may suggest conformational energetic reasons, in addition to proximity of charged groups, provide an incentive for the product to leave.

It is noteworthy that the activation segments around the sites of phosphorylation in CDK2, MAPK, CaMK1, IRK, FGFRK, Src and Hck are all mobile in the non-phosphorylated state. Mobility in the region around the site of phosphorylation may be an important factor that allows protein kinase substrate recognition and accommodation of the bound substrate in an extended conformation.

8. The catalytic mechanism

In cAPK, stereochemical arguments indicate that the γ-phosphate of ATP is transferred to the hydroxyl of the serine by direct, nucleophilic displacement [45]. Enzymatic phosphorylation proceeds with rapid binding of substrates, and the chemical step is fast (500 s^{-1}) relative to the rate-determining dissociation of the product ADP [46,47]. Metal ions are required for the formation of productive enzyme-substrate complex and the fast transfer of the phosphate group. They serve both to stabilise and orient the γ-phosphate in the correct orientation for catalysis and to promote and stabilise the

developing negative charges during catalysis. The effects of pH on phosphorylation of peptide substrates by cAPK led to proposals for a general base catalysed reaction in which the base accepts a proton from the seryl hydroxyl concurrent with nucleophilic attack of the alcoholate ion on the γ phosphate of ATP [48]. The role of the catalytic base has been probed in more detail in recent experiments [49]. Because of the similarities in structures between cAPK and PhK (33% sequence similarity) and the conservation of key residues, much of the inferences made about the catalytic proposals with cAPK may be also applied to PhK.

In PhK, Asp^{149} has been identified as the catalytic residue by its proximity to substrate in the catalytic site (Figs. 2 and 3A). The peptide substrate serine OG atom is 3.6 Å from the γ-phosphorus atom of AMPPNP and is hydrogen bonded to one of the phosphate oxygens which in turn contacts the metal, Mn2. The Asp^{149} OD2 atom is roughly equidistant from both the substrate Ser OG (4.3 Å) and the γ-phosphate oxygen (4.1 Å). These distances are larger than might be anticipated for a direct participation of Asp^{149} in catalysis. Although the position of AMPPNP is identical to the position observed in cAPK for ATP, it has been argued [17] that, since the inactive ATP analogue (AMPPNP) has a poor leaving group and may differ in ionisation properties, the position of the serine may differ slightly from the position adopted in an active ternary complex with ATP. Indeed if the conformation of the peptide substrate serine is altered by a single rotation about the Cα-Cβ bond from a gauche$^-$ to a gauche$^+$ conformation, then the OG atom of the serine is within 3.1 Å of the Asp^{149} OD2 atom and 3.0 Å from the γ-phosphorus atom of the AMPPNP, consistent with a direct participation of the aspartate in catalysis (Fig. 4). A key role for the aspartate is supported by recent site directed mutagenesis experiments in which the mutant Asp^{148}Ala exhibited a decrease in k_{cat}/K_m of 10^4 compared with the native enzyme (V. Skamnaki, D.J. Owen et al., manuscript in preparation).

9. Concluding remarks

The crystal structures of the active kinases in ternary complexes with ATP analogues and substrate or inhibitor peptides have provided considerable insights into mechanisms of activation, substrate recognition and catalysis. However, there is more to be learnt, especially from the proline (P+1) specific kinases such as MAPK and CDK2 when these structures with substrate become available. The studies with CDK2 with cyclin and modification by phosphorylation have provided a detailed description of the activation processes. It will be interesting to see if these mechanisms are also exploited in other CDK/cyclin pairs and if so what is the origin of the specificity for activation of the different cyclins for their cognate CDKs. For other kinases the shift in focus is towards regulation by different domains, such as calmodulin and calcium for PhK or the complete complex of regulatory and catalytic subunits of cAPK, to test pseudo-substrate recognition proposals. The structure of the active conformation of an intact Src kinase should test proposals for the roles of the SH2 and SH3 domains and phosphorylation in regulation. Undoubtedly there will be more riches to come that demonstrate how a common fold can be utilised in many different ways to generate specificity and control mechanism that are essential for all living organisms.

References

[1] Fischer, E.H. and Krebs, E.G. (1955) J. Biol. Chem. 216, 121–132.
[2] Sutherland, E.W. and Wosilait, W.D. (1955) Nature 175, 169–170.
[3] Hunter, T. and Plowman, G.D. (1997) Trends Biochem. Sci. 22, 18–22.
[4] Hunter, T. (1987) Cell 50, 823–825.
[5] Hanks, S.K. and Hunter, T. (1995) FASEB J. 9, 576–596.
[6] Zheng, J., Trafny, E.A., Knighton, D.R., Xuong, N., Taylor, S.S., Eyck, L.F.T. and Sowadski, J.M. (1993) Acta Crystallogr. D 49, 362–365.
[7] Bossmeyer, D., Engh, R.A., Kinzel, V., Ponstingl, H. and Huber, R. (1993) EMBO J. 12, 849–859.
[8] De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O. and Kim, S.-H. (1993) Nature 363, 592–602.
[9] Schulze-Gahmen, U., de Bondt, H.L. and Kim, S.H. (1996) J. Med. Chem. 39, 4540–4546.
[10] Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J. and Pavletich, N.P. (1995) Nature 376, 313–320.
[11] Russo, A., Jeffrey, P.D. and Pavletich, N.P. (1996) Nature Struct. Biol. 3, 696–700.
[12] Zhang, F., Strand, A., Robbins, D., Cobb, M.H. and Goldsmith, E.J. (1994) Nature 367, 704–711.
[13] Wilson, K.P., Fitzgibbon, M.J., Caron, P.R., Griffith, J.P., Chen, W., McCaffrey, P.G., Chambers, S.P. and Su, M.S. (1996) J. Biol. Chem. 271, 27696–27700.
[14] Wang, Z., Harkins, P.C., Ulevitch, R.J., Han, J., Cobb, M.H. and Goldsmith, E.J. (1997) Proc. Natl. Acad. Sci. USA 94, 2327–2332.
[15] Canagarajah, B.J., Khokhlatchev, A., Cobb, M.H. and Goldsmith, E.J. (1997) Cell 90, 859–869.
[16] Owen, D.J., Noble, M.E., Garman, E.F., Papageorgiou, A.C. and Johnson, L.N. (1995) Structure 3, 467–482.
[17] Lowe, E.D., Noble, M.E.M., Skamnaki, V.T., Oikonomakos, N.G., Owen, D.J. and Johnson, L.N. (1997) EMBO J. 16, 6646–6658.
[18] Xu, R.-M., Carmel, G., Sweet, R.M., Kuret, J. and Cheng, X. (1995) EMBO J. 14, 1015–1023.
[19] Longenecker, K.L., Roach, P. and Hurley, T.D. (1995) J. Mol. Biol. 257, 618–631.
[20] Hu, S.H., Parker, M.W., Lei, J.Y., Wilce, M.C.J., Benian, G.M. and Kemp, B.E. (1994) Nature 369, 581–584.
[21] Kobe, B., Heierhorst, J., Feil, S.C., Parekr, M.W., Benian, G.M., Weiss, K.R. and Kemp, B.E. (1996) EMBO J. 15, 6810–6821.
[22] Goldberg, J., Nairn, A.C. and Kuriyan, J. (1996) Cell 84, 875–887.
[23] Hubbard, S.R., Wei, L., Ellis, L. and Hendrickson, W.A. (1994) Nature 372, 746–754.
[24] Hubbard, S.R. (1997) EMBO J. 16, 5572–5581.
[25] Mohammadi, M., Schlessinger, J. and Hubbard, S. (1996) Cell 86, 577–587.
[26] Yamaguchi, H. and Hendrickson, W.A. (1996) Nature 384, 484–489.
[27] Xu, W., Harrison, S.C. and Eck, M.J. (1997) Nature 385, 595–602.
[28] Williams, J.C., Weijland, A., Gonfloni, S., Thompson, A., Courtneidge, S.A., Superti-Furga, G. and Wierenga, R.K. (1997) J. Mol. Biol. 274, 757–775.
[29] Sicheri, F., Moarefi, I. and Kuriyan, J. (1997) Nature 385, 602–609.
[30] Narayana, N., Cox, S., Xuong, N., Ten Eyck, L.F. and Taylor, S.S. (1997) Structure 5, 921–935.
[31] Pickett-Gies, C.A. and Walsh, D.A. (1986) in: The Enzymes (Boyer, P.D. and Krebs, E.G., Eds.), Vol. 17, pp. 396–459, Academic Press, Orlando, FL.
[32] Graves, D.J. (1983) Methods Enzymol. 99, 268–278.
[33] Knighton, D.R., Zheng, J., Ten Eyck, L.F., Ashford, V.A., Xuong, N.-H., Taylor, S.S. and Sowadski, J.M. (1991) Science 253, 407–413.
[34] Taylor, S.S. and Radzio-Andzelm, E. (1994) Structure 2, 345–355.
[35] Hon, W.-C., McKay, G.A., Thompson, P.R., Sweet, R.M.,

Yang, D.S.C., Wright, G.D. and Berghuis, A.M. (1997) Cell 89, 887–895.
[36] Johnson, L.N., Noble, M.E.M. and Owen, D.J. (1996) Cell 85, 149–158.
[37] Olah, G.A., Mitchell, R.D., Sosnick, T.R., Walsh, D.A. and Trewhella, J. (1993) Biochemistry 32, 3649–3657.
[38] Cox, S., Radzio-Andzelm, E. and Taylor, S.S. (1994) Curr. Opin. Struct. Biol. 4, 893–901.
[39] Helms, V. and McCammon, J.A. (1997) Protein Sci. 6, 2336–2343.
[40] Hubbard, S.R., Wei, L., Ellis, L. and Hendrickson, W.A. (1994) Nature 372, 746–754.
[41] Songyang, Z. et al. (1996) Mol. Cell. Biol. 16, 6486–6493.
[42] Acharya, K.R., Stuart, D.I., Varvill, K.M. and Johnson, L.N. (1991), Glycogen Phosphorylase pp. 123, World Scientific, Singapore.
[43] Barford, D., Hu, S.-H. and Johnson, L.N. (1991) J. Mol. Biol. 218, 233–260.
[44] Gregoriou, M. et al. (1998) Protein Sci. (in press).
[45] Ho, M.F., Bramson, H.N., Hansen, D.E., Knowles, J.R. and Kaiser, E.T. (1988) J. Am. Chem. Soc. 110, 2680–2681.
[46] Adams, J.A. and Taylor, S.S. (1992) Biochemistry 31, 8516–8522.
[47] Grant, B.D. and Adams, J.A. (1996) Biochemistry 35, 2022–2029.
[48] Yoon, M.-Y. and Cook, P.F. (1987) Biochemistry 26, 4118–4125.
[49] Zhou, J. and Adams, J.A. (1997) Biochemistry 36, 2977–2984.