Signal Transduction Proteins: Structural Basis of Control by Phosphorylation

L. N. JOHNSON, M. E. M. NOBLE, D. BARFORD, N. BROWN,
J. A. ENDICOTT, A. LAWRIE AND D. J. OWEN
Laboratory of Molecular Biophysics, University of Oxford,
South Parks Road, Oxford, OX1 3QU, U.K.

Introduction

It is an honour to deliver the Salimuzzaman Siddiqui Inaugural Lecture. Professor Salimuzzaman Siddiqui was a great scientist born 100 years ago in 1897. I never had the good fortune to meet him and I am indebted for my information to the excellent biographical memoir written by Professor Mohammed Akhtar for the UK Royal Society [1] (to which Professor Siddiqui has been elected in 1961). Professor Siddiqui was an outstanding natural product chemist who was able to isolate the active ingredients from medicinal plants and to develop knowledge of their chemical structure. He was also a visionary in the organisation of science in Pakistan who, as part of his responsibilities as Chairman of the Pakistan Council of Scientific and Industrial Research (1951-1966), founded five research institutes, including the Husein Ebrahim Jamal Institute in the University of Karachi. And he was a man of culture, well read in philosophy and in urdu and persian poetry and a gifted poet and artist. Towards the end of this lecture I shall describe the binding of a natural product to one of its target proteins and to demonstrate that, with current expertise in the field of protein crystallography, we can now build upon the basic science elucidated by the chemists and biochemists of earlier years to explain biological activity in terms of the molecular details of the interactions. It is a special significance for me to be in Karachi at this time and I am grateful to Professor Zaidi and to all friends for their hospitality and support.

Protein phosphorylation

The modification of proteins by phosphorylation and dephosphorylation reactions is the major mechanism for regulation of enzyme activity in the response of the cell to extracellular signals. Binding of hormones, such as adrenaline or insulin, or mitogens, such as growth factors, to their receptors on the outside of the cell membrane results in a series of recognition phenomena on the inside of the cell that triggers a cascade of protein phosphorylation/ dephosphorylation reactions. Protein phosphorylation is now recognised to effect diverse and important processes such as metabolism, gene transcription, membrane transport, cell growth, differentiation, motility, learning and memory. Phosphorylation provides a reversible process in which the forward and the back reactions are catalysed by different enzymes operating with different specificities so that reactions can be turned on or off in response to different stimuli. Phosphorylation on single or multiple sites can elicit a variety of molecular responses. It can result in enzyme activation or inhibition; it can alter the association/dissociation properties of protein-protein assemblies; or it can alter the surface recognition properties of a protein. The importance of phosphorylation has been emphasised from estimates, which indicate as many as 30% of intracellular proteins are phosphorylated and that 2% of eukaryotic genes encode protein kinases and 1% encode protein phosphatases.

Protein phosphorylation as an activatory mechanism was first observed with glycogen phosphorylase in the work of Fischer and Krebs in 1955 [2]. They established that the active form of glycogen phosphorylase (phosphorylase a), which was active in the absence of AMP, was phosphorylated at a site later shown to be Serine 14. In the same year, working with liver extracts, Sutherland and Wosilait showed that phosphorylase could be inactivated by a protein phosphatase [3] (Figure 1). For the next 13 years control of enzyme activity by phosphorylation was thought to be a peculiarity of glycogen metabolism. In 1968 Walsh *et al*, [4] reported the discovery of cyclic AMP dependent protein kinase, a kinase which had wider specificity and modified a number of important substrates both in the cytoplasm and in the nucleus. Cyclic AMP was found to activate cyclic AMP

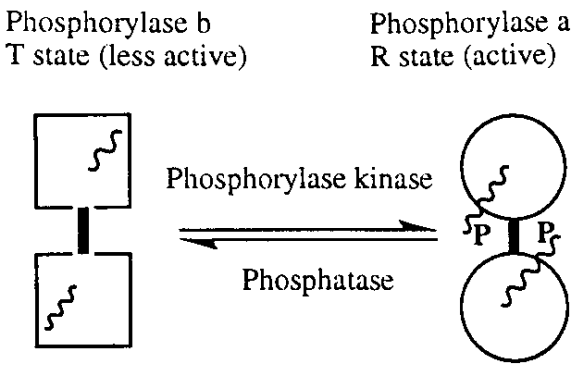


Fig. 1. The interconversion of glycogen phosphorylase b (less active; T state) and glycogen phosphorylase a (active; R state) through phosphorylation catalysed by phosphorylase kinase and dephosphorylation catalysed by protein phosphatase.

dependent protein kinase which in turn phosphorylated and activated phosphorylase kinase to phosphorylate and activate phosphorylase. Both cyclic AMP dependent protein kinase and phosphorylase kinase phosphorylate glycogen synthase and in this enzyme phosphorylation leads to inactivation (Figure 2). Today there are over 400 protein kinases that have been characterised by their amino acid sequences, forming one of the largest families of homologous proteins known in nature [5]. Figure 3 shows some of the intricate relationships between phosphorylation and dephosphorylation that take place in the muscle cell in response to hormones such as adrenaline and insulin. Adrenaline stimulates glycogen degradation in order to provide energy to sustain muscle contraction while insulin turns off glycogen degradation and stimulates glycogen synthesis through the action of protein phosphatases, which are themselves controlled by phosphorylation.

The response of a protein to phosphorylation is dictated by the special properties of the phosphate group that distinguish it from the naturally occurring amino acids (Figure 4). The phosphate group with 4 oxygen atoms can participate in extensive hydrogen bond interactions and these can link different parts of the polypeptide chain to create an ordered region. The phosphate group (approximate pK 6.7) is likely to be dianionic at physiological pH. The property of a double negative charge is a property that is not available to the naturally occurring amino acids and electrostatic effects are important in control by phosphorylation. Analysis of protein phosphate interactions in existing protein structures [6] has shown that the most common interaction is between the phosphate oxygens and the main chain nitrogens at the start of a helix where the most frequently found residue is glycine. In non-helix interactions, phosphate groups most commonly interact with

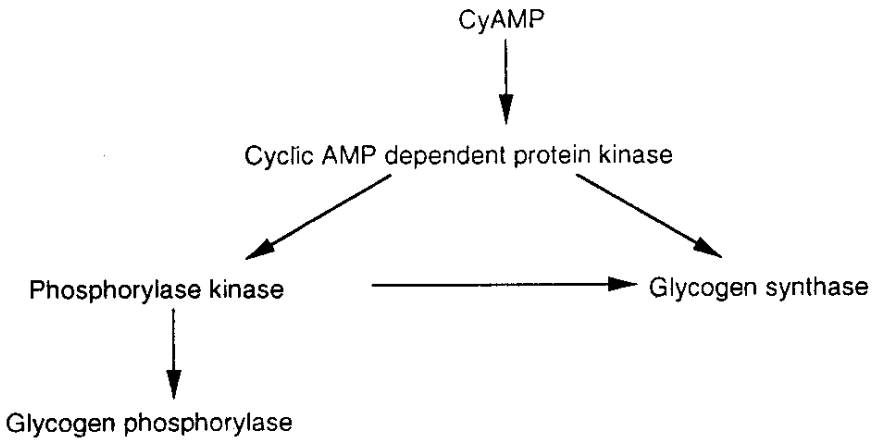


Fig. 2. The phosphorylation cascade initiated by cyclic AMP. Cyclic AMP dependent protein kinase phosphorylates phosphorylase kinase which activates glycogen phosphorylase and glycogen degradation. Cyclic AMP dependent protein kinase and phosphorylase kinase phosphorylate glycogen synthase which inactivates this enzyme and inhibits glycogen synthesis

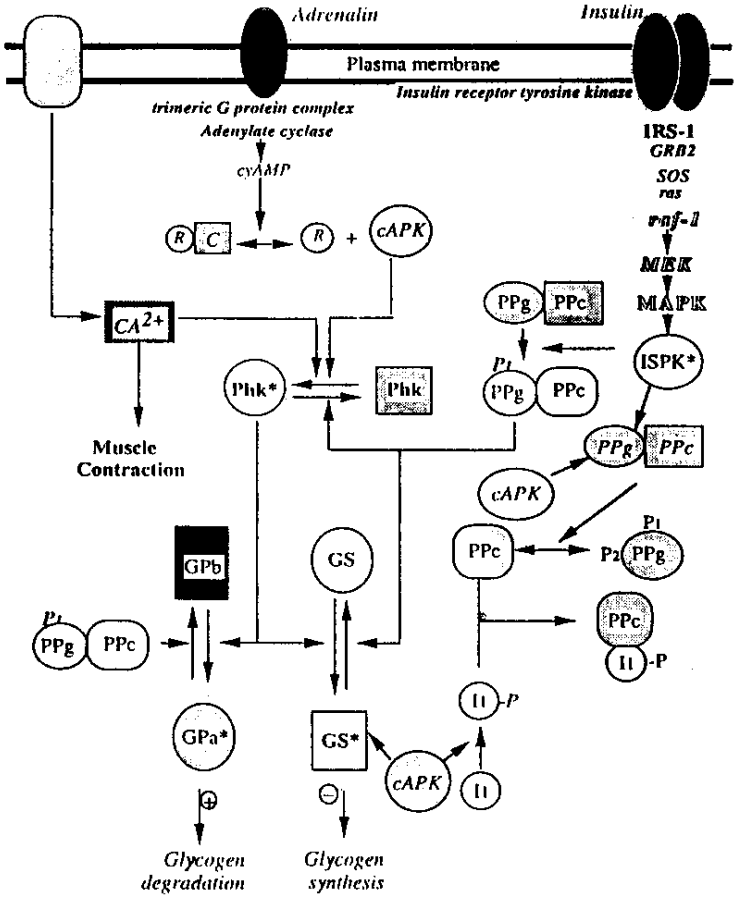
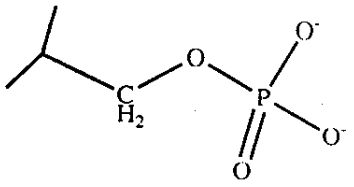


Fig. 3. Phosphorylation and dephosphorylation cascades that regulate glycogen metabolism in skeletal muscle. Inactive enzymes are represented by squares and active enzymes by circles or ellipses. Phosphorylated proteins are denoted by * or by the attachment of the label -P. Adrenaline binding to its receptor at the extracellular domain stimulates the intracellular synthesis of the second messenger cyclic AMP (cyAMP) through the interactions of the trimeric G protein complex which activates adenylate cyclase. cyAMP stimulates the cyclic AMP dependent protein kinase (cAPK) which in turn activates phosphorylase kinase (PhK). In addition PhK is activated by elevated Ca^{2+} released in response to neuronal signals. PhK activates glycogen phosphorylase (GPb to GPa) and both cAPK and PhK inactivate glycogen synthase. Glycogen degradation is turned off and glycogen synthesis is turned on in response to insulin. Binding of insulin to the extracellular domain of its receptor stimulates the intracellular insulin receptor tyrosine kinase which phosphorylates the insulin receptor substrate. In its phosphorylated state the insulin receptor substrate is recruited to the membrane through interactions with the GRB2 and SOS proteins to stimulate the ras protein which in turn activates the protein kinase Raf. Raf, MEK and MAPK form a protein kinase cascade where each in turn activates the next kinase on the pathway leading to the phosphorylation and activation of the insulin sensitive protein kinase (ISPK). ISPK phosphorylates the glycogen binding subunit (PPg) of the protein phosphatase at site 1 which leads to the activation of phosphatase activity by the catalytic subunit (PPc). The protein phosphatase then dephosphorylates PhK and GPa leading to inactivation of these enzymes and dephosphorylates GS leading to activation. The protein phosphatase can be inhibited by additional phosphorylation at site 2 on the glycogen binding subunit by cAPK which leads to dissociation of PPc. PPc is inhibited by the inhibitor protein I1 which inhibits when phosphorylated by cAPK.



1. Electrostatic - dianion
2. Polar - hydrogen bonds
3. Steric - bulky group

Fig. 4. Properties of a phosphorylated amino acid (Serine, Threonine or Tyrosine).

arginine residues. The guanidinium group is suited for interactions with phosphate by virtue of its planar structure and its ability to form multiple hydrogen bonds. Because of its resonance stabilisation, the guanidinium group is a poor proton donor (pK_a>12) and cannot function as a general acid catalyst in the hydrolysis of phosphorylated amino acids. Electrostatic interactions between arginine and phosphate groups provide tight binding sites that appear to play a dominant role in recognition and stabilisation of protein conformations [7].

Glycogen phosphorylase

Glycogen phosphorylase catalyses the first step in glycogen degradation: it promotes the attack of inorganic phosphate on the glycosidic bond at the non-reducing end of glycogen to yield glucose-1-phosphate. Glycogen phosphorylase is activated by phosphorylation on a single serine residue, Ser14, by the action of phosphorylase kinase, a highly specific enzyme whose only established physiological substrate is glycogen phosphorylase. The structural changes on conversion of inactive glycogen phosphorylase b (GPb) to active glycogen phosphorylase a (GPa) have been described [8-10] and reviewed [11,12]. In brief to a first approximation the enzyme can be viewed according to the Monod, Wyman, Changeux model in which the dimeric enzyme exists in 2 states, a T state which is less active and has low affinity for substrate and an active R state which has high affinity for substrate. The T and the R states are characterised by different tertiary conformations of the individual subunits and different quaternary structures that govern the relationships of the 2 subunits to each other. In the inactive T state of GPb the N terminal residues, 10 to 20, (residues 1 to 10 are not located in the crystal structure) are some of the least well defined regions of the structure and their positions were first established in the complex of the enzyme with glucose, a T state inhibitor. These residues make intrasubunit contacts (Figure 5a). On activation to GPa there is a conformational change involving changes in the dihedral angles of the main chain in the region of residues 22 and 23 so that the N terminal residues are directed through 120° and reach up to make intersubunit contacts (Figure 5a). The Ser-P docks between 2 arginine residues Arg69 from its own subunit and Arg43' from a symmetry related subunit (Figure 5b). Both these arginines shift to make contact with the Ser-P and these shifts together with other local conformational changes and a quaternary conformational change lead to a tightening of the subunit interface of the dimer. In GPb the C-terminal residues 836 to 842 are located in this intersubunit region and there is an intersubunit ion pair between Asp838 and His36'. On phosphorylation and localisation of the N-terminal tail at this site the C terminal region is displaced. Thus phosphorylation leads to ordering of the N-terminal tail and disorder of the C-terminal tail and concomitant other changes in structure.

The site of phosphorylation, Ser14, is located in a sequence of residues 10 to 16 (Arg-Lys-Gln-Ile-Ser*-Val-Arg) that contains 3 basic and no acidic residues. This characteristic of the N-terminal tail was noted in 1959 by Fischer *et al.*, [13], who speculated that it may have some importance in the control mechanism. There is a strong acidic patch on the protein surface caused by the proximity of a number of glutamic acid groups . The N-terminal tail is associated with this region. Ser14 is directed inwards and hydrogen bonds to Glu501. In order for these residues to become accessible to phosphorylase kinase the N-terminal region would need to be displaced from the surface and indeed those ligands that promote changes from the T state to the R state

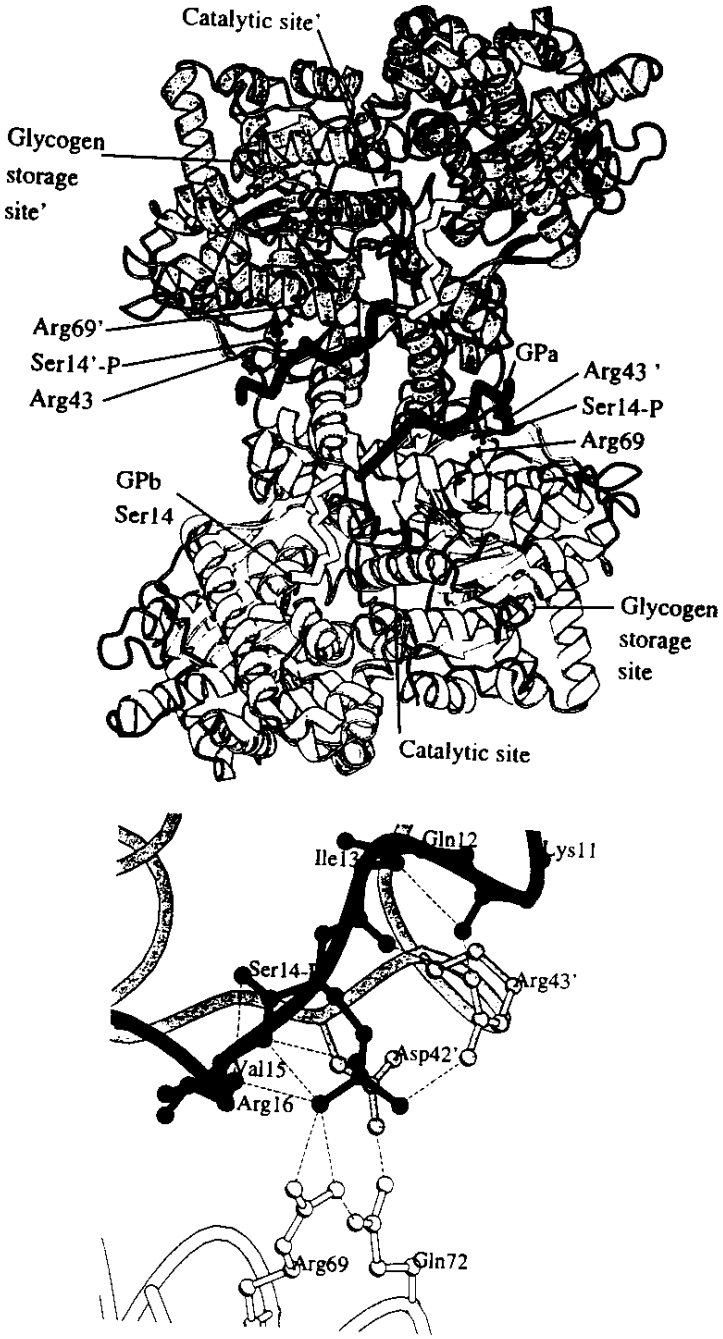


Fig. 5. Glycogen phosphorylase. a) A schematic diagram of the glycogen phosphorylase dimer viewed down the 2 fold axis of symmetry. The positions of the N-terminal 20 residues in the non-phosphorylated form (GPb) are shown as the thick white line. Their positions on phosphorylation of Ser14 are shown as the thick black line. The catalytic site is over 30 Å away from the site of phosphorylation. b) Details of the interaction of Ser14-P with 2 arginine residues at the subunit-subunit interface of GPa. Arg69, from the same subunit , contacts the phosphate and is also involved in hydrogen bonds to Gln72 which in turn hydrogen bonds to Asp42' from the other subunit. Arg43', from the other subunit, contacts the phosphate and also hydrogen bonds to main chain oxygen of Lys11 and Ile13. The Ser-phosphate group also hydrogen bonds to the main chain nitrogen of Val15 and Arg16.

also make phosphorylase a better substrate for its kinase [14] and references therein). This acidic environment is thus hospitable to the basic N-terminal peptide in the non-phosphorylated state but once phosphorylation takes place, there is electrostatic repulsion and the phosphorylated peptide must seek another location. As the structure clicks to the R state two arginines come together and provide suitable site for charge compensation of the phosphate. In the R state there are also changes at the subunit-subunit interface and changes in tertiary structure which lead to opening of the tunnel that leads to the catalytic site over 30 Å from the site of phosphorylation. At the catalytic site there is a change in the constellation of residues that leads to the creation of the phosphate substrate recognition site. The structural studies with glycogen phosphorylase have shown that phosphorylation at a single serine residue can cause pronounced conformational changes (the serine shifts 50 Å) and that this modification to an enzyme, that contains 842 amino acids per subunit, results in a change of organisation resulting in activation.

Phosphorylase kinase

Phosphorylase kinase represents one of the largest and most complex of the protein kinases. The enzyme exists as a hetero tetramer with stoichiometry $(\alpha\beta\gamma\delta)_4$ and a total molecular weight of approximately 1.3×10^6 Da (reviewed in [15]. The α and β subunits are regulatory and these subunits are the targets for phosphorylation by cyclic AMP dependent protein kinase (cAPK) (Figure 3). The δ subunit is an integral calmodulin subunit and provides sensitivity to calcium. Thus calcium released in response to nerve impulses both activates the actomyosin ATPase for muscle contraction and stimulates glycogen degradation to replenish ATP. The γ subunit is the catalytic subunit and is composed of an N-terminal kinase domain and a C-terminal regulatory domain.

We have cloned and expressed in *E. coli* the kinase domain of the γ subunit of rabbit muscle phosphorylase kinase (PhK). The expressed fragment represents a constitutively active protein kinase that requires no post-translational modification for activity. All the regulatory machinery for this particular kinase is located outside the kinase core domain. The binary structures of this kinase core of the catalytic γ subunit in complex with Mn^{2+}/AMPPNP (a non-hydrolysable analogue of ATP) and with the product Mg^{2+}/ADP have been solved [16] initially at 2.6 Å and at 3.0 Å resolution, respectively, and now at 2.1 Å for the former complex (Table 1).

Structure

The kinase core domain PhK, residues 1-298, represents a minimal kinase structure. The structure is composed of 2 lobes :an N-terminal lobe composed mostly of β sheet and a C-terminal lobe that is mostly α helix (Figure 6). The structure is similar to the kinase core of cAPK [17,18] and other kinases. Comparison of the relative lobe orientations between the binary PhK structure and the ternary complex of cAPK with inhibitor and ATP shows that PhK adopts a closed structure but one which is slightly more open than the ternary cAPK structure. The shift can be represented as a rotation of the N-terminal lobe with respect to the C-terminal lobe of about 5°.

Nucleotide binding

The nucleotide AMPPNP is associated mostly with the N-terminal domain but also makes important links to residues in the C-terminal domain. The polar contacts are shown in Figure 7. The adenine is docked into a non-polar pocket and the ribose makes 2 hydrogen bonds through its O2' and O3' hydroxyls to Glu110 and the main chain carbonyl oxygen of Glu153, respectively. The triphosphate moiety is stabilised by contacts to Lys48 and 2 metal ions. The Mn^{2+} in site 2 chelates the β and γ phosphates and the Mn^{2+} in site 1 chelates the α and γ phosphates. The metal ions make contact to residues Asp167 and Asn154 as shown in Figure 7. The major difference in the nucleotide contacts between PhK and cAPK is in the contacts to the glycine rich loop between strands β1 and β2. Because of the more open structure in PhK the loop is just too far away to contribute direct hydrogen bond interactions from the main chain NH groups. The sequence of residues 26-32 in PhK is Gly-Arg-Gly-Val-Ser-Ser-Val. The serine at position 31 replaces the third glycine that is normally found in the glycine rich loops of nucleotide binding sites, and the side chain of Ser31 may be responsible for

Table 1. Phosphorylase kinase: statistics of the data and refined structures.

Data		
Complex	Mn^{++}/AMPPNP	Mg^{++}/ADP
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell (Å)	a=47.6, b=67.4, c=110.8	a=47.9, b=69.1, c=112.9
X-ray source	Beamline 4 ESRF	Station 9.5 SRS
Detector	CCD Image intensifier	Mar Research Image Plate
Temperature	100K	Room temperature
Resolution	2.1 Å	3.0 Å
Number of Observations	280,000	21198
Number of Unique reflections	19,961	6964
R_{merge}[1]	0.051	0.096
Completeness	88.0%	87.8 %
Refinement		
R_{conv}[2]	0.21	0.172
R_{free}[3]	0.288	0.264
rms bonds[4] (Å)	0.01	0.013
rms angles[4] (°)	1.7	1.75

$$R_{merge} = \frac{\sum_j \sum_h \left| I_{h,j} - \bar{I}_h \right|}{\sum_j \sum_h I_{h,j}}$$

1 Where $I_{h,j}$ is the jth observation of reflection h.

$$R_{conv} = \frac{\sum_h \left\| Fobs_h \right| - \left| Fcalc_h \right\|}{\sum_h \left| Fobs_h \right|}$$

2 Where $Fobs_h$ and $Fcalc_h$ are the observed and calculated values respectively for structure factor h.

3 R_{free} is equivalent to R_{conv} for a randomly selected 5% subset of reflections not used in structure refinement.

4 As calculated by XPLOR.[37].

some of the displacement. In addition there may be further closure of the loop on the formation of the ternary complex. The results indicate an intricate arrangement of the protein that ensures the correct disposition of residues to locate the ATP substrate for catalysis. The correct location is also promoted by the "activation segment", a region responsible for control of protein kinases.

The activation segment of protein kinases

If cellular life is to function in an orderly manner, the switching on and off of protein kinases and phosphatases is as crucial for their function as their catalytic activity. Control mechanisms that have been recognised include: control by additional subunits or domains which may function in response to second messengers (e.g. cyAMP) or whose level of expression varies depending on the functional state of the cell (*e.g.* cyclin regulation of the cyclin dependent protein kinases); control by additional domains that target the kinase to different molecules or subcellular localisations (*e.g.* the SH2 and SH3 domains of the Src kinases); control by additional domains that inhibit the kinase activity by an autoregulatory process and whose inhibition may be relieved by Ca^{2+}; and control by phosphorylation and dephosphorylation by other kinases and by

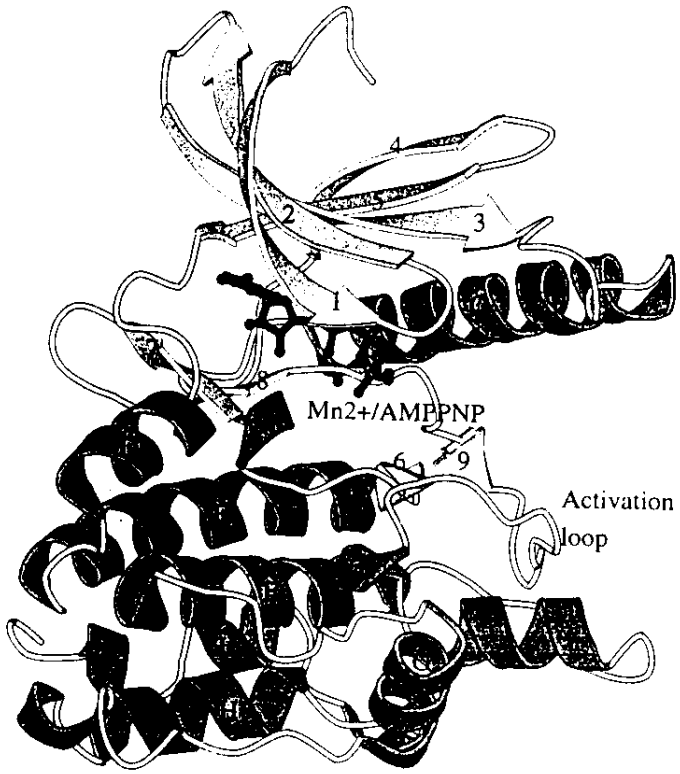


Fig. 6. A schematic diagram showing the fold of phosphorylase kinase catalytic domain. The catalytic site is located between the mostly β-sheet N-terminal domain and the α-helical C-terminal domain with Mn^{2+}/AMPPNP shown bound. The activation segment is labelled.

phosphatases. Phosphorylation of specific threonine, serine or tyrosine residues may occur at a number of sites. Some of these are located in the N-terminal or C-terminal portions of the polypeptide chain which lie outside the kinase domain. A key aspect of regulation recognised in recent years has been the observation that many protein kinases are phosphorylated on a residue(s) located in a particular segment in the centre of the kinase domain which is termed the "activation segment" [19]. The activation segment is defined as the region spanning conserved sequences "DFG" and "APE" and corresponds to residues 184 to 208 in cyclic AMP dependent protein kinase (cAPK) [20]. In cAPK the site of phosphorylation in the activation segment is Thr197

The conversion of an inactive kinase to an active kinase involves conformational changes in the protein that lead to the correct disposition of substrate binding and catalytic groups. The activation segment and the control of its conformation via phosphorylation plays a key role in these transformations. The crystal structure determination of cAPK showed the structural importance of Thr197 phosphorylation and demonstrated possible roles of phosphorylation in promotion of activation. The structure provided a definitive model to which other kinases could be related. Many important protein kinases are controlled by phosphorylation on their activation segments but some are not (Table 2). Comparison of the structures of cAPK and phosphorylase kinase allows us to provide some answers. All protein kinases contain an aspartate residue (Asp166 in cAPK) which has been implicated in the catalytic mechanism, most likely as a base which activates the incoming substrate seryl, threonyl or tyrosyl hydroxyl group. Most Ser/Thr and all Tyr protein kinases have an arginine immediately preceding this catalytic aspartate. Knowledge of the kinase structures so far suggests that these kinases, which we term "RD" kinases, require some ionic interactions of the arginine with a

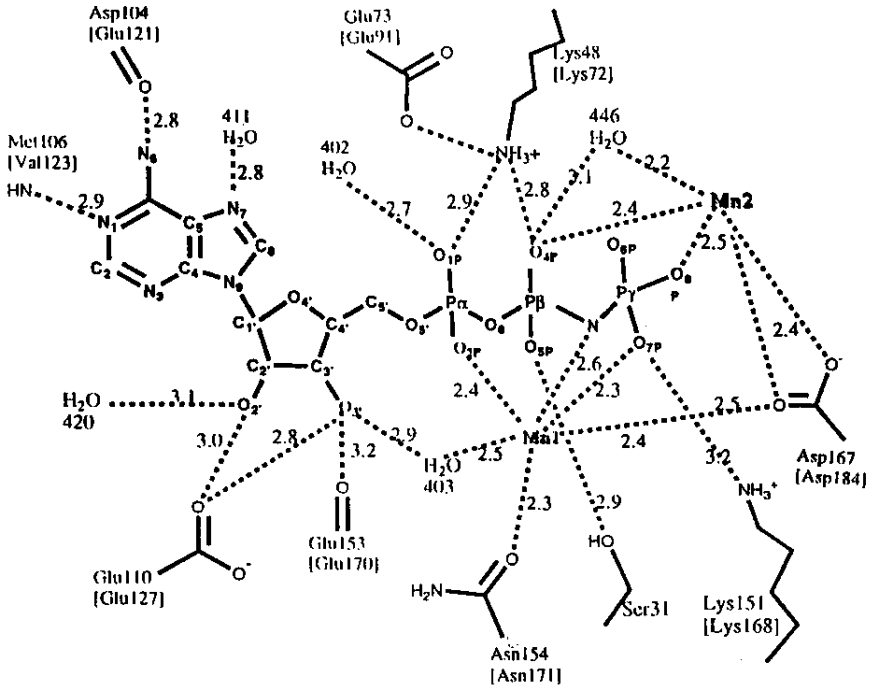


Fig. 7. Details of the polar interactions made by the non-hydrolysable ATP analogue, AMPPNP, at the catalytic site of phosphorylase kinase. The equivalent residues in cAPK are shown in square brackets. The adenine base makes 2 hydrogen bonds from its N1 and N6 atoms to main chain atoms close to the hinge region between the N and C terminal domains. The ribose contacts Glu110 from the hinge region and Glu153 main chain. The triphosphate is localised by interactions with the 2 manganese ions, labelled Mn1 and Mn2, and to Ser31, Lys48 and Lys151 and bridges the N and C terminal domains.

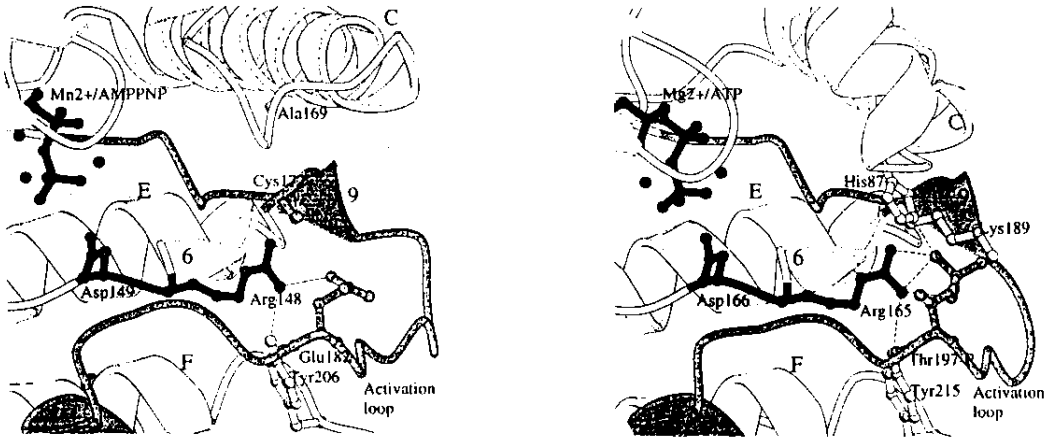


Fig. 8. Contacts from the activation segment to the catalytic site. a) In PhK a glutamate, Glu182, contacts Arg148 which precedes the catalytic aspartate, Asp149. , b) In cAPK the threonine-phosphate, Thr197-P, contacts Arg165 which precedes the catalytic aspartate, Asp166. The phosphate group also neutralises the charge on Lys189, from the start of the activation segment, and His87, from the N-terminal domain. In PhK the corresponding residues are Cys and Ala, respectively.

Table 2. Examples of protein kinases phosphorylated and not phosphorylated in the activation segment[5].

A. Kinases phosphorylated in the activation segment	B. Kinases not phosphorylated in the activation segment
"RD" kinases	*"RD" kinases*
Ser/Thr protein kinases	*Ser/Thr protein kinases*
Cyclic AMP dependent kinase (cAPK)[a]	Phosphorylase kinase (PhK)[a]
Cyclin dependent kinases ($p34^{cdc2}$, CDC2, $CDK2^{a,b}$, CDK7)	Casein kinase 1 (CK1)[a]
Microtubule associated kinase (MAPK/ERK-2)[b]	Calcium calmodulin dependent kinase II (CaMII)
Calcium calmodulin dependent kinase I (CaMKI)[b]	
Raf1 kinase	
Protein kinase C-βII (PKC)	*Tyrosine protein kinases*
MAP kinase kinase (Mek1)	EGF receptor tyrosine kinase (EGFRK)
Insulin stimulated protein kinase 1 (ISPK-1; rsk^{mo-2})	C-terminal Src kinase (Csk)
Glycogen synthase kinase 3 (GSK3)	
Tyrosine protein kinases	*Non-"RD" kinases*
Insulin receptor tyrosine kinase (IRK)[b]	Twitchin kinase[b]
FGF receptor tyrosine kinase[b]	Myosin light chain kinase (MLCK)
PDGF receptor tyrosine kinase (PDGFRK)	Wee1 kinase
c-Src tyrosine kinase[b] (and other kinases in the Src family e.g.Lck[a], Yes, Fyn, Fgr, Lyn, Mck, Blk)	

[a]: Kinases for which crystal structure is available in the active state.
[b]: Kinases for which crystal structure is available in the inactive state
"RD" kinases are defined as those kinases in which the conserved catalytic aspartate is preceded by an arginine residue

phosphate or with a carboxylate group. Kinases that are known to be activated by phosphorylation on the activation segment have an "RD" sequence at the catalytic aspartate (*e.g.* all those listed under section A of Table 2). The structural examples suggest that these kinases require charge neutralisation of a cluster of basic residues, including the arginine of the RD sequence, by the phospho-amino acid.

As shown in Figure 8b, in cAPK the phospho-threonine Thr197-P contacts His 87 from the amino terminal lobe, Arg165 which precedes the catalytic base (Asp166), Lys189 and Thr195. The phosphate is placed to compensate the cluster of positively charged residues. In phosphorylase kinase the equivalent residue to Thr197 is a glutamate, Glu182 (Figure 8a). Glu182 side chain is turned in to form an ionic link with Arg148, the arginine which precedes the catalytic base. In PhK the residues equivalent to His87 and Lys189 are Ala69 and Cys172, respectively. Thus in both kinase structures the activation segments contribute a stabilising ionic interaction with the conserved arginine preceding the catalytic aspartate; this interaction is mediated by Thr197-P in cAPK and by Glu182 in PhK. The sequence changes indicate that in PhK there is no cluster of basic residues that need a dianionic group for stabilisation and the role of an ionic interaction with the arginine can be accomplished by a glutamate. The structures of kinases in the inactive state show very different conformations of the activation segment, indicating that there are a variety of conformations accessible to different kinases in the inactive state.

A summary of our current understanding of the mechanism of protein kinases is given in Figure 9. The aspartate (Asp166 in cAPK) is presumed to act as a base to abstract a proton from the protein substrate hydroxyl group although the catalytic mechanism is not definitively established. The resulting alcoholate or phenolate ion is positioned to attack the γ phosphate of ATP. The proposals for the mechanism suggest several possible roles for the phosphothreonine in activation of cAPK and other kinases.

(i) The Thr-P stabilises the positively charged cluster of Arg165, Lys189 and His87 in a way that can only be promoted by a dianionic group

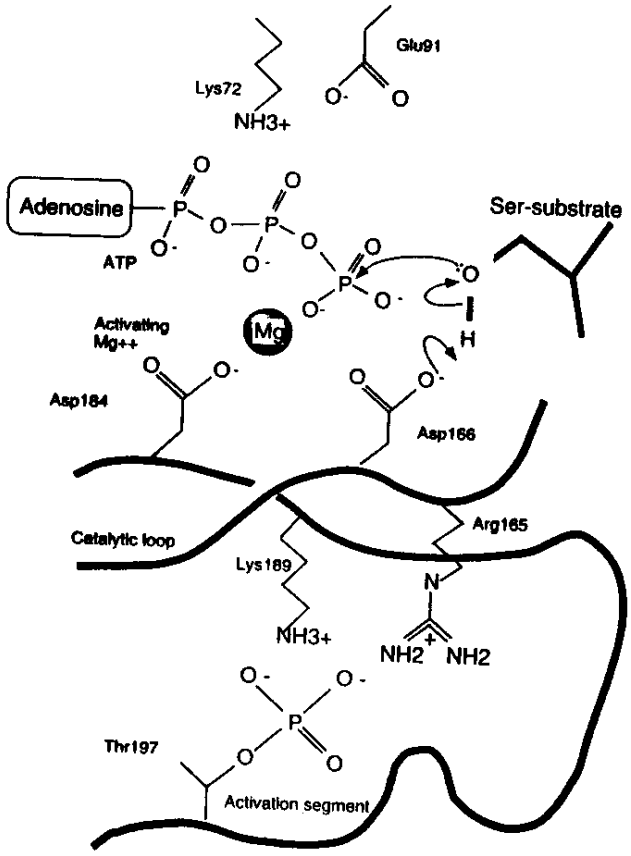


Fig. 9. A simplified representation of the catalytic site of cAPK showing ATP bound and a serine of a protein substrate. The diagram illustrates the connection between the activation segment and the catalytic loop via contacts through the Thr197-P. A possible mechanism is shown in which base-catalysed attack by Asp-166 on the substrate serine promotes the formation of an alchoholate ion, which can then attack the γ phosphate of ATP. (Diagram adapted from.[19]).

(ii) The interactions, especially those through Arg165, may help promote the correct orientation and electrostatic environment for the catalytic base Asp166

(iii) The location of the Thr197-P may determine the correct conformation for residues in the activation segment that are important for protein substrate recognition [18, 21].

(iv) Interactions between Thr-P and Lys189 may help assist the correct conformation for Asp184 which contacts the activatory Mg^{2+} and the interactions between residues 185 to 189 and the N-terminal lobe.

(v) The contact between His87 and Thr197-P may promote the correct domain/domain orientation that is also critical for ATP binding. However open conformations of phosphorylated cAPK are observed and hence phosphorylation is not sufficient for domain closure [22]. Thus this contact may not be important for the catalytic function in cAPK but it may be important for other functions, such as recognition of the regulatory subunit, and equivalent contacts may be important in other kinases.

Control of the cell cycle

Cell growth and division is a cornerstone of biology. The ordered processes by which a cell grows and divides into 2 daughter cells are summarised in

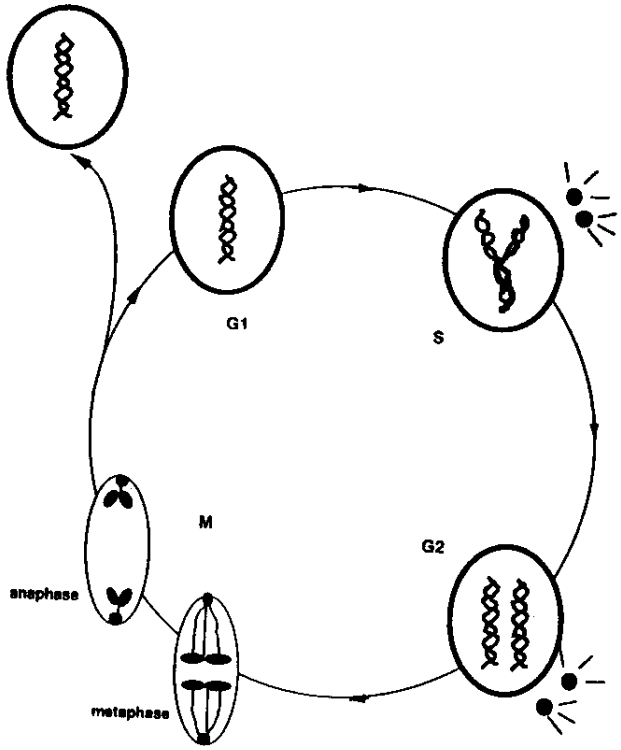


Fig. 10. Cell cycle events. In G1 the chromosomes are decondensed and there is a single spindle pole (centrosome). In S phase the DNA is replicated and the spindle poles are duplicated and move apart. In G2 the chromosomes are completely replicated. At M phase (mitosis) the nuclear envelope breaks down and the spindle poles organise an intranuclear spindle composed of microtubules; the chromosomes condense, attach to microtubules; at anaphase sister chromatids separate to the spindle poles and pull apart. At teleophase the chromosomes decondense, the intranuclear spindle is disassembled, the nuclear membrane reforms around the chromosomes, cytokinesis separates the 2 daughter cells, and cells re-enter G1 (modified from [38].

Figure 10. In the last few years there has been an explosion in knowledge of the checks and balances that ensure orderly process of these events such that the DNA is copied once and only once and that there is an even distribution of one copy of each chromosome to each daughter cell [23-27]. In the unified theory of cell cycle control it is recognised that the key events that modulate these processes are the phosphorylation events controlled by the cyclin dependent protein kinases (CDKs). These kinases are themselves controlled by phosphorylation in response to mitogens or growth factors and they are also controlled by the association with cyclin subunits and with inhibitor proteins. Cyclins are protein molecules whose concentrations vary during the cell cycle following cycles of expression and degradation. In eukaryotes a number of different CDKs have been recognised and different CDKs are important at different stages of the cell cycle (Figure 11). In particular CDK2, which becomes active at late G1 in response to binding cyclin E and during the start of S phase in response to binding cyclin A, has proved amenable to crystallographic studies.

The structure of the inactive form of CDK2 was solved in 1993 [28] and showed an inactive orientation of the activation segment and an incorrect disposition of the C helix, the so-called PSTAIRE helix because of this characteristic sequence motive in many CDKs (Figure 12). In 1995 the structure of the partially active form of CDK2 in association with cyclin A was published [29]. This

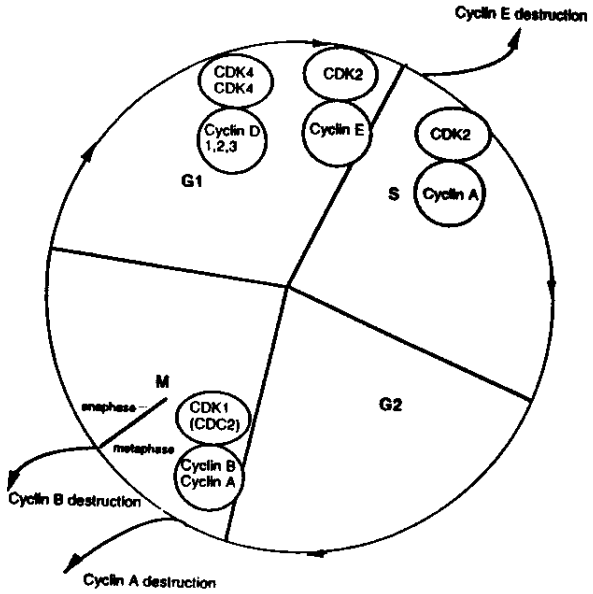


Fig. 11. Diagram showing the different CDKs and the different cyclins that become active during the cell cycle.

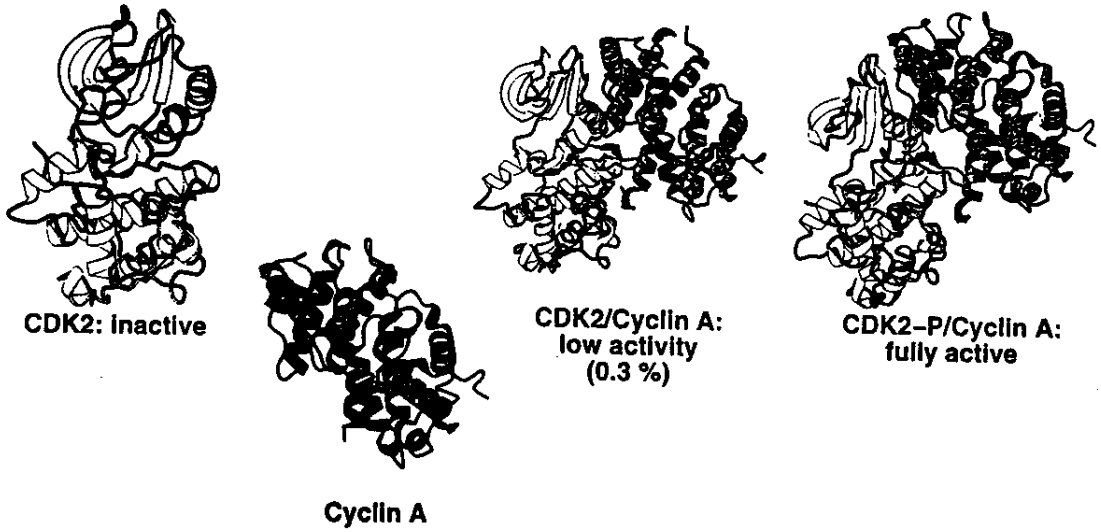


Fig. 12. Activation of CDK2 by association with cyclin A and by phosphorylation. The structure of inactive CDK2 (top left) has the activation segment (shown in black) in the wrong conformation and several other elements of the catalytic site are also wrongly disposed. The structure of cyclin A (unbound) is shown at lower left. There is no change in the conformation of cyclin A on complex formation with CDK2 (centre) but there is a substantial change in CDK2 which brings the activation segment into a conformation that is nearly correct. Full activation is only achieved with the phosphorylated CDK2 complexed with cyclin A (centre right). There is a further change in the activation segment which is important for protein substrate recognition.

work demonstrated that there is dramatic change in the conformation of the kinase principally involving a significant shift of the PSTAIRE helix and in the activation segment through interactions triggered by association with the cyclin. In the same year in Oxford we determined the crystal structure of cyclin A in its unbound state [30].The protein consists of a duplication of a 5 helical domain. There is only 17% identity in sequence between the 2 domains, yet despite this low homology the 2 domains adopt almost identical structures. Comparison between cyclin A in the bound and unbound states showed that there are no conformational changes (Figure 12). Evidently the cyclin acts as a rigid template against which the more pliable CDK2 docks. The CDK2/cyclin A complex showed that the catalytic site residues are in their correct relationship to promote catalysis and yet the complex only exhibits 0.3% of the activity displayed when CDK2 is phosphorylated at Thr160 in its activation segment. The final piece in the puzzle came with results in 1996 with the structure of the phosphorylated CDK2-cyclin A complex from Nicola Pavletich's laboratory [31]. On phosphorylation at Thr160, the activation segment undergoes further changes including shifts as great as 6Å in Thr160. The phosphate group is mostly buried and its charge is neutralised by 3 arginine residues, Arg126 adjacent to the catalytic aspartate, Arg50 from the PSTAIRE helix and Arg150 in the N-terminal part of the activation segment (equivalent to Lys189 in cAPK). Two of these arginines also hydrogen bond to main chain oxygen atoms on cyclin A. The phosphate acts as an organising centre with its influence extending beyond the immediate vicinity of the phosphorylation site, in a similar way to the organisational and long range effects of the Ser14-phosphate in glycogen phosphorylase. These changes allow the activation segment to adopt a conformation that is likely to be important for the recognition of the protein substrate, emphasising the importance of the correct juxtaposition of the 2 substrates, protein and ATP, before catalysis can take place. However the structure of a peptide bound CDK2 active complex has not yet been elucidated.

The binding of a potent inhibitor: CDK2-staurosporine complex

Uncontrolled cell proliferation is a characteristic of cancer, which kills one in six of us. Tumour cells most frequently have acquired damage to genes that regulate their cell cycles [32]. Because of the importance of the protein kinases in the control of cell growth and cell division, there has been keen interest in agents that inhibit these enzymes [33]. Staurosporine (Figure 13) is a microbial alkaloid which was characterised in 1986 as a potent inhibitor of protein kinase C [34] and later as a relatively non-specific but highly potent inhibitor of other protein kinases [33] including CDK1 (CDC2), a protein kinase which is 60% homologous to CDK2. IC_{50} values for many kinases are in the nanomolar range.The biological effects elicited by staurosporine are varied and complex, no doubt because of its ability to inhibit many different protein kinases lying on multiple signal transduction cascades [35]. It is widely used to induce cells into programmed cell death, although the precise molecular targets in this inhibition have not been identified. The biological effects of staurosporine indicate that it may have relevance in the treatment of diseases but that it would be valuable to have an agent that exhibited higher specificity and lower toxicity.

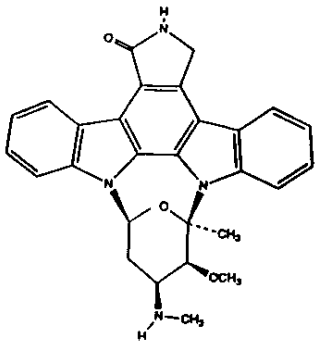


Fig. 13. The structure of staurosporine.

In recent work we have determined the structure of the CDK2-staurosporine complex to 2.1 Å resolution [36]. Staurosporine binds to the ATP recognition site in CDK2 between the 2 domains (Figure 14) and adopts a boat conformation for the tetrahydrophan ring, similar to the conformation seen in the small molecule structure of 4'-N-methyl staurosporine methiodide. Figure 15 shows a diagram of the hydrogen bonding pattern and non-polar contacts. The hydrogen bonds to the main chain atoms of Glu81 and Leu83 mimic the

Fig. 14. The structure of CDK2 shown as a ribbon diagram complexed with staurosporine bound at the ATP recognition site. The electron density is not definitive for a short stretch of chain following the third β strand and for part of the activation segment and hence there are 2 breaks in the chain.

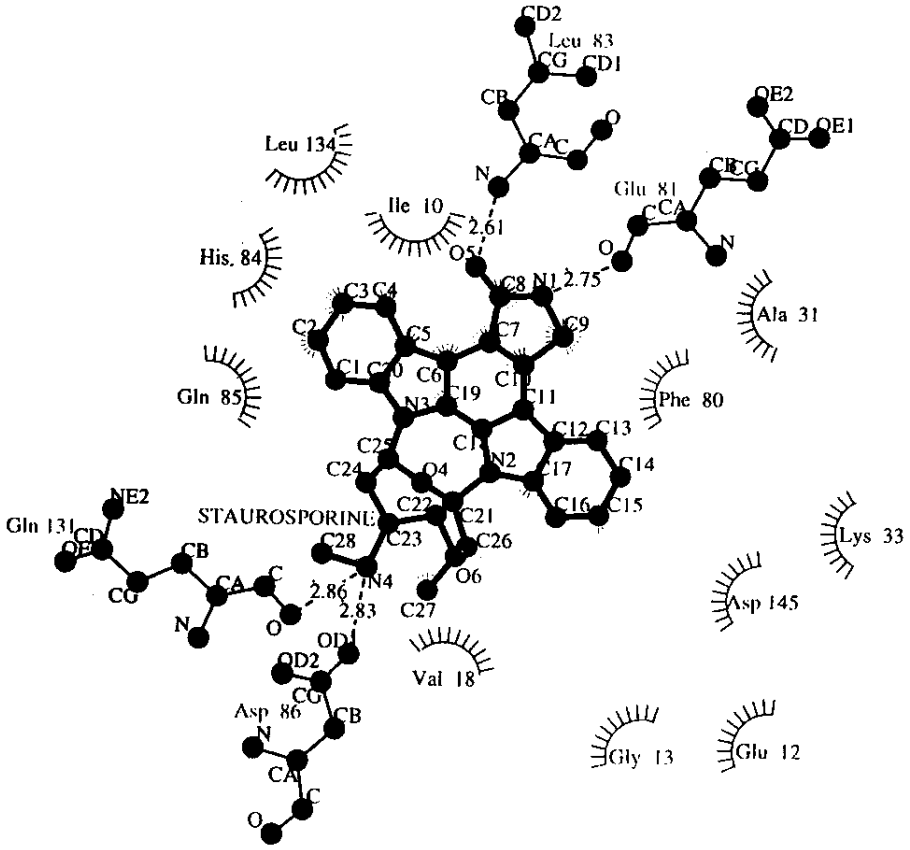


Fig.15. A diagram representing the hydrogen bonds and non-polar contacts made between staurosporine and CDK2. Hydrogen bonds are shown dashed. Van der Waals interactions to protein atoms are indicated by lines forming a fan shape and directed between the staurosporine atom and the residue of CDK2. Further details see text.

Fig. 16. Diagram showing the staurosporine molecule viewed edge on to the aromatic rings and bound in the ATP pocket of CDK2. The grey contours indicate the molecular surface of the protein.

ydrogen bonds made by the adenine base of ATP to CDK2 and similar interactions are seen in other kinases (e.g. Figure 7). The methylamino group interacts with a main chain oxygen and with an aspartate, Asp86, an interaction which exploits the charged nature of this group. Staurosporine analogues in which the methylamino group is replaced by a hydroxyl are generally weaker inhibitors of protein kinases. The major contribution to the binding appears to come from the non-polar interactions. Calculations show that the molecular surface buried in the CDK2-staurosporine complex is 657 $Å^2$, of which 544 $Å^2$ (82%) is non-polar. When ATP binds a total of 573$Å^2$ molecular surface is buried of which only 265 $Å^2$ (46%) is non-polar. The hydrophobicity of the pocket (Figure 16),which changes shape slightly to accommodate the staurosporine molecule, evidently contribute significantly to the binding. These interactions appear to be the major source of energy that produce a nanomolar inhibitor compared with the micromolar affinity exhibited with ATP.

Although staurosporine is a potent inhibitor of most protein kinases, it does display some specificity. For example the IC_{50} of staurosporine against CDC2-cyclin B is some 10^4-fold lower than against CDK4-cyclin D1. For CDK4 it is observed from sequence alignment (the structure of this kinase is not yet known) that a phenylalanine which contributes to the burial of the inhibitor is replaced by a histidine. It is possible that the polar histidine residue adopts a different conformation and makes less favourable contacts with staurosporine.

The determination of the CDK2-staurosporine crystal structure in large part explains the high potency for the protein kinase family. The extensive planar non-polar staurosporine surface is buried and the hydrogen bonding potential of the adenine recognition site is satisfied. The present structural results indicate where modifications to the staurosporine molecule may be made with the aim of providing an inhibitor with increased specificity for certain kinases. However even with the knowledge of

a structure, structure based drug design is not easy and there are several unforeseen (but, with the benefit of hindsight, explainable) phenomena that can occur in the development of new agents. As Salimuzzaman Siddiqui wrote "Scientific research has no easy gifts to offer. There is no finality in research and continued investigation in any field leads to improved products".

Acknowledgements

This work has been funded by the Medical Research Council and by the Oxford Centre for Molecular Sciences.

References

1. M. Akhtar, Biographical memoirs of Fellows of the Royal Society **42**: 399 (1996).
2. E.H. Fischer and E.G. Krebs, *J. Biol. Chem.* **216**, 121 (1955)
3. E.W. Sutherland, and W.D. Wosilait, *Nature*, **175**, 169 (1955).
4. D.A. Walsh, J.P. Perkins, and E.G. Krebs, *J. Biol. Chem.*, **243**, 3763 (1968).
5. G. Hardie, and S. Hanks, The protein kinase facts book (Academic Press, London, 1995).
6. R.R. Copley, and G.J. Barton, *J. Mol. Biol.*, **242**, 321 (1994).
7. L.N. Johnson, and M. O'Reilly, *Current Opinion in Structural Biology*, **6**, 762 (1996).
8. S.R. Sprang, *Nature*, **336**, 215 (1988).
9. D. Barford, and L.N. Johnson, *Nature*, **340**, 609 (1989).
10. D., Barford, S.-H Hu,. and L.N. Johnson, *J. Mol. Biol.*, **218**, 233 (1991).
11. L.N. Johnson, *FASEB J.* **6**, 2274 (1992).
12. L.N. Johnson, and D. Barford, *Annu. Rev. Biophys. Biomol. Struct.*, **22**, 199 (1993).
13. E.H. Fischer, D.J. Graves, E.B. Sydney-Crittenden, and E.G. Krebs, *J. Biol. Chem.*, **234**, 1698 (1959).
14. W.R. Harris, and D.J. Graves, *Arch. Biochem. Biophys.*, **276**, 102 (1990).
15. C.A. Pickett-Gies, and D.A. Walsh, in The Enzymes (eds. P.D. Boyer, and E.G. Krebs) 396 (Academic Press, Orlando, Florida, 1986).
16. D.J. Owen, M.E. Noble, E.F. Garman, A.C. Papageorgiou and L.N. Johnson, *Structure*, **3**, 467 (1995).
17. D.R. Knighton, *Science*, **253**, 407 (1991).
18. D. Bossmeyer, R.A. Engh, V. Kinzel, H. Ponstingl, and R. Huber, *EMBO J.*, **12**, 849 (1993).
19. L.N., Johnson, M.E.M. Noble, and D.J. Owen, *Cell*, **85**, 149 (1996).
20. S.S. Taylor, and E. Radzio-Andzelm, *Structure*, **2**, 345 (1994).
21. D.R. Knighton, *Science*, **253**, 414 (1991).
22. S., Cox, E. Radzio-Andzelm, and S.S. Taylor, *Curr. Opin. Struct. Biol.*, **4**, 893 (1994).
23. A. Murray, and T. Hunt, The cell cycle (W H Freeman, New York, 1993).
24. D.O. Morgan, *Nature*, **374**, 131 (1995).
25. C.J. Norbury, and P. Nurse, *Ann. Rev. Biochem.*, **61**, 441 (1992).
26. R.W. King, R.J. Deshaies, J.-M. Peters, and M.W. Kirschner, *Science*, **274**, 1652 (1996).
27. K. Nasmyth, *Science*, **274**, 1643 (1996).
28. H.L. De Bondt, *Nature*, **363**, 592 (1993).
29. P.D. Jeffrey, *Nature*, **376**, 313 (1995).
30. N.R. Brown, *Structure*, **3**, 1235 (1995).
31. A. Russo, P.D. Jeffrey, and N.P. Pavletich, *Nature Structural Biology*, **3**, 696 (1996).
32. C.J. Sher, *Science*, **274**, 1672 (1996).
33. L. Meijer, in Prog. in Cell Cycle Res. (ed. L., Meijor, S. Guidet, and H. Y. L. Tung) 1-13 (Plenum Press, New York, 1995).
34. T. Tamaoki, and e. al. *Biochem. Biophys. Res Commun.*, **135**, 397 (1986).
35. S. Omura, Y. Sasaki, Y. Iwai, and H. Takeshima, *J. Antibiot.*, **48**, 535 (1995).
36. A.M. Lawrie, *Nature Structural Biology*, Submitted (1997)
37. A.T. Brunger, X-PLOR: Version 3.1; a system for protein crystallography and NMR. (Yale University Press, New Haven, 1992).
38. L. Hartwell, in Cell Cycle Control (eds C. Hutchison & D. Glover) (IRL Press, Oxford, 1995).