Structural characterization of the cyclin-dependent protein kinase family

Jane A. Endicott1 and Martin E.M. Noble
Northern Institute for Cancer Research, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, U.K.

Abstract
Structural studies of members of the CDK (cyclin-dependent protein kinase) family have made a significant contribution to our understanding of the regulation of protein kinases. The structure of monomeric unphosphorylated CDK2 was the first of an inactive protein kinase to be determined and, since then, structures of other members of the CDK family, alone, in complex with regulatory proteins and in differing phosphorylation states, have enhanced our understanding of the molecular mechanisms regulating protein kinase activity. Recently, our knowledge of the structural biology of the CDK family has been extended by determination of structures for members of the transcriptional CDK and CDK-like kinase branches of the extended family. We include these recent structures in the present review and consider them in the light of current models for CDK activation and regulation.

Introduction
It was Louise Johnson who first drew an analogy between protein kinase families and human families, rewriting the opening lines of Leo Tolstoy’s Anna Karenina to state that “active protein kinases are all alike, every inactive kinase is inactive in its own way” [1]. As more protein kinase structures have been determined, however, it has become apparent that there are common structural features that recur in inactive kinase structures. Disruption of these features, which may be brought about by diverse mechanisms, ultimately leads to engagement within the kinase N-terminal lobe of a hydrophobic patch on the C-helix [that in the cell cycle CDKs (cyclin-dependent kinases) contains the conserved PSTAIRE (Pro-Ser-Thr-Ala-Ile-Arg-Glu) sequence] and rearrangement of the activation segment [defined as the sequence that lies between the conserved DFG (Asp-Phe-Gly) and APE (Ala-Pro-Glu) motifs] [2,3]. Together, these structural changes bring about activation by creating an active site that is competent to bind both ATP and protein substrates productively and to effect catalysis. CDK structures have been reviewed previously [4]. However, since 2010, a number of CDK structures have been determined that illuminate further the structural diversity of the extended CDK family. In the present review, we include these recent structures and consider them in the light of current models for CDK activation and regulation.

CDKs and the cell cycle
CDKs are implicated in many cellular regulatory processes, but were first characterized as regulators of the eukaryotic cell cycle [5]. In brief overview, the transition of cells through the early G1-phase of the cell cycle is co-ordinated by the activities of CDK4 and CDK6 complexes that are formed following the mitogen-dependent expression of D-type cyclins [6]. The key CDK4/6 substrate during G1-phase is pRB, the product of the retinoblastoma gene. pRB binds to members of the E2F transcription factor family within complexes that actively repress expression of genes required for cell cycle progression through G1-phase. A second important non-catalytic function of cyclin D-containing CDK complexes during early G1-phase is to act as a sink for members of the Cip/Kip CDK inhibitor family [7,8]. Cip/Kip proteins are potent inhibitors of CDK2 and only when they are sequestered can cyclin E, whose synthesis is promoted by E2F family members, activate CDK2 to bring about sustained pRB phosphorylation. These events also promote activation of E2F-responsive genes that ensure the timely licensing of DNA replication origins and subsequent successful completion of S-phase [9]. Later in the cell cycle, CDK1, predominantly in complex with cyclin B, directs cells into mitosis. In addition to association with an appropriate cognate cyclin, the cell cycle CDKs also require phosphorylation within the activation loop for full activity. In mammals, this phosphorylation event is catalysed by CDK7 [10].

CDKs that regulate transcription
RNA polymerase II transcribes protein-encoding eukaryotic genes and its activity is tightly regulated, in part by post-transcriptional processing by members of the CDK family. The human RNA polymerase II CTD (C-terminal domain) consists of 52 repeats (27 in yeast) of the sequence Y1S2P3T4S5P6S7 (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) and...
phosphorylation of these repeats couples transcription initiation, elongation and mRNA processing events. The roles of CDK8–cyclin C in transcription and its interaction with the core subunits of the Mediator complex are beginning to be clarified through a combination of genetic, biochemical and structural studies [11]. The Mediator complex appears to function at various stages throughout transcription and can be isolated in mutually exclusive complexes bound to either RNA polymerase II or to CDK8–cyclin C. Early events in transcription initiation also result from interplay between multiple complexes containing CDKs 7 and 9, the two best studied members of the transcriptional CDK family. During transcription initiation, Ser^3 is primarily phosphorylated by the general transcription factor TFIIH. This transcription factor consists of the trimeric CDK7–cyclin H–MAT1 CAK (CDK-activating kinase) complex and additional core complex subunits that also effect the essential function of TFIIH in nucleotide excision repair [12]. Following transcription initiation, promoter proximal pausing provides an opportunity for mRNA processing and further transcription factor recruitment. Release from this paused state requires the activity of CDK9–cyclin T1 [P-TEFb (positive transcription elongation factor b)] [13]. P-TEFb phosphorylates the CTD, primarily at Ser^2, and phosphorylates the negative elongation factors NELF (negative elongation factor) and DSIF [DRB (dichlorobenzimidazole riboside)-sensitivity-inducing factor]. These activities encourage promoter escape, recruit the splicing machinery and stimulate processive elongation. Increasingly, studies are implicating other members of the extended CDK family in the regulation of transcription. These studies are assigning functions to transcription to CDK11 bound to cyclin L [14] and CDKs 12 and 13 bound to cyclin K [15].

**Diverse functions of the extended CDK family**

The human kinome reveals that the CDK family forms a distinct branch of 26 proteins within the CMGC subfamily, of which 21 are classified as CDKs and five form a more distant group of CDKL (CDK-like) kinases [16,17]. Despite several decades of research on CDKs that control the cell cycle and transcription, many members of the family have not been studied extensively. Their distinctive sequence organization suggests that an understanding of their structure and function may challenge our current paradigms of CDK regulation. Recent determinations of the structures of the kinase domains of PCTAIRE-1 (CDK16) and CDKLs 1, 2, 3 and 5 have begun to elaborate the differences. Their close evolutionary relationship to kinases that play essential roles in diverse cellular processes and that are frequently associated with disease suggests that these kinases may offer novel targets for disease indications not historically associated with CDK dysfunction.

The three human PCTAIRE kinases, named originally for the amino acid sequences that encode their respective C-helices (i.e. Pro–Cys–Thr–Ala–Ile–Arg–Glu) and subsequently renamed CDKs 16, 17 and 18 [17], were isolated in a PCR-based screen to identify Cdc2-related proteins in two cancer cell lines [18]. PCTAIRE-1 (CDK16) was subsequently found to be expressed in neurons and spermatids and at high levels in transformed cell lines. Expression of PCTAIRE-3 (CDK18) is found in brain and testis and PCTAIRE-2 (CDK17) is limited to brain (reviewed in [19]). There is debate as to whether the PCTAIRE family has a role in cell cycle progression in normal cells, and/or whether its function is limited to differentiation of neurons or spermatids. Transfection of PCTAIRE-1 into neuroblastoma cell lines modulated neurite outgrowth, suggesting that it could function in the control of this process [20], and it has recently been shown to be essential for spermatogenesis [21]. PCTAIREs have also been shown to interact with complexes involved in protein transport through the early secretory pathway [22]. PCTAIRE-1 may have a role in modulating exocytosis as part of the process of transmitter release in neuronal cells where it has been shown to phosphorylate NEM (N-ethylmaleimide)-sensitive fusion protein [23]. In Caenorhabditis elegans, a PCTAIRE-1–cyclin Y complex has been found to act together with CDK5/p35 to target presynaptic components to the axon [24]. In vitro, PCTAIRE phosphorylates either histone H1 [25] or MBP (myelin basic protein) [20] and an optimized peptide substrate has been identified [26], although a bona fide substrate has not yet been identified. There are no PCTAIRE orthologues in C. elegans or S. pombe genomes, but PCTAIRE genes have been cloned from mice, Dicyostelium and Drosophila, suggesting that they encode a conserved function in higher eukaryotes.

More distant relatives of the CDKs are the CDKL kinases, CDKL1–CDKL5 [17]. Compared with other members of the family, very little is known about the expression and activity of these enzymes, although a common emerging theme is that they are selectively expressed in the brain and have been linked to behavioural and neurological disorders by genetic studies. CDKL2 function has been probed in mice where it is expressed in terminally differentiated neurons [27]. Expression patterns of CDKL3 in mice are also consistent with a role in brain function [28]. The most extensively studied member of the family is CDKL5 [29]. A recent large-scale clinical study has described the features of disease seen in patients with mutations in the CDKL5 gene [30]. Mutations in CDKL5 have been shown to be associated with a number of clinical conditions that are characterized by early-onset epilepsy and severe developmental delay. Taken together, studies of a number of neurocognitive disorders provide evidence for roles of CDKLs in regulating neuronal function.

**Structures of monomeric CDKs**

**The CDK2 paradigm**

Monomeric CDK2 shares with other active eukaryotic protein kinases a structure in which the catalytic site that binds ATP is sandwiched between N- and C-terminal lobes [31]. However, closer inspection reveals a more closed fold
in which conserved residues within the catalytic site are not correctly positioned, resulting in non-productive binding of ATP and the absence of a peptide-binding cleft. This inactive conformation arises mainly from the organization of two key elements of structure. These are (i) the C-helix (referred to as the PSTAIRE helix in CDK2) of the N-terminal lobe that forms the back of the active-site cleft, and (ii) the activation segment, which includes residues that lie between the conserved DFG and APE motifs (residues 145–147 and 170–172 respectively).

The start of the activation segment of inactive monomeric CDK2 adopts a ‘DFG-in’ conformation that is a characteristic of the active kinase fold (Figure 1A). In this position, the phenylalanine side chain points into the active site to complete a stack of hydrophobic amino acids that also includes CDK2 residues Leu66, Leu55, Phe146 and His125 that together compose the regulatory spine. These highly conserved amino acids first described in the structure of PKA (cAMP-dependent protein kinase) form a non-contiguous motif that traverses through both lobes of the protein kinase fold to ensure that events at the active site reverberate to more distant parts of the molecule [32,33].

CDK2 and Src are unusual among protein kinases in that, typically, inactive protein kinases display the alternative ‘DFG-out’ conformation (reviewed in [2,34]). In this conformation, the motif is flipped so that the phenylalanine side chain is within the ATP-binding pocket. This new location for the phenyl ring not only disrupts the regulatory spine, but also prevents the ‘catalytic spine’ from forming. This second spine of hydrophobic residues, again first defined following studies in PKA [33], is composed of CDK2 residues Ala31, Val18, Ile87, Leu153, Leu134, Ile135, Ile192 and Met196 (Figure 1A). When it is eventually properly assembled, it will also include the adenine ring of ATP correctly aligned for catalysis.

The residues immediately following DFG in the CDK2 activation loop adopt a short α-helical structure (αL12) and to accommodate it the C-helix is displaced out from the active site. As a result, the interaction between Glu51 (the E of the PSTAIRE sequence) and Lys33 (from β3) that is important for correct localization of the ATP triphosphate is not made. Towards the C-terminal end of the C-helix, the shift also results in considerable displacement of Leu195, one of the residues within the regulatory spine. Critically too, Arg50 that ultimately will form a ligand for phosphorylated Thr160 within the activation segment is also displaced.

A characteristic of protein kinase folds is their flexibility that allows them to respond to protein association and post-translational modification by structural change. A number of distantly related protein kinases have now been shown to adopt this ‘mixed’ inactive CDK2 conformation that shares features with both active (‘DFG-in’) and inactive (‘C-helix out’) protein kinase structures. It has been suggested that the inactive structure of CDK2 might represent an intermediate conformation in a pathway that would permit the switch from a ‘DFG-out’ to ‘DFG-in’ conformation (termed ‘DFG-flipping’) of the activation segment that is required for protein kinase activation [2]. Emerging from the αL12 helix, the activation segment then adopts an extended conformation that folds up in a β-hairpin turning, at Tyr159 and Thr160, over the glycine-rich loop before folding back towards the C-terminal lobe (Figure 1A). However, this part of CDK2 is quite flexible and in a number of structures that include ATP-competitive inhibitors cannot be unambiguously traced in the electron density maps.
The extended CDK family and CDKL protein kinases

The PCTAIRE protein kinases

The three PCTAIRE isoforms, unlike all other CDKs, have a long N-terminal extension (162, 189 and 64 residues for CDKs 16, 17 and 18 respectively) and a small C-terminal domain of approximately 40 amino acids. The catalytic domain is well conserved (CDK16 shows 52% and 57% identity with CDK2 and CDK5 respectively over the kinase domain) and contains all the amino acids that have been shown to be essential for catalytic activity. Within the N-terminal region adjacent to the kinase domain is a short region of sequence homology unique to the PCTAIREs called the PCTAIRE box. Several CDK16-interacting proteins have been identified in a recent screen [21]. In addition, earlier studies had characterized PKA, p11 (calpactin I light chain) [35], 14-3-3s [20,25] and CDK5-p25 [36] as interacting with PCTAIRE-1. In these studies, the N-terminal regulatory domain was essential for interactions and differences in sequence in this domain, particularly between CDKs 16, 17 and 18, imparted isofrom-specific protein-binding properties.

Human CDK16 has been shown to bind cyclin D3 [37], and human [21] and C. elegans [24] PCTAIRE-1 have been shown to bind to cyclin Y orthologues. Human cyclin Y is an unusual cyclin in that it is N-terminally myristoylated and associated with the plasma membrane. Formation of the CDK16-cyclin Y complex is dependent upon sequences N-terminal to the kinase fold suggesting that, in addition to binding PCTAIRE-1 regulators, the PCTAIRE box might also contribute to the CDK–cyclin interface [21]. Such a model is supported by an early study that showed that complex formation is inhibited by phosphorylation at Ser153, a known in vitro PKA site [20]. Taken together, these results suggest that the PCTAIRE box is an important motif in the formation of the CDK16-cyclin Y complex and for its subsequent activity.

The crystal structure of CDK16 (residues 205–473, PDB code 3MTL) has been determined at 2.4 Å (1 Å = 0.1 nm) resolution and shows that in the absence of its cyclin partner and PCTAIRE box the kinase domain adopts an inactive structure that is distinct from, but shares features with, the structure of monomeric CDK2 (Figure 1B). The CDK16 N-terminal fold is composed of a five-stranded β-sheet and, as seen in other protein kinases, the loop linking the end of β3 and the start of the C-helix is flexible. However, in a novel departure, the PCTAIRE sequence is not helical in the CDK16 crystal structure, but instead extends to form a β-strand. An inspection of the crystal lattice reveals that this structural re-arrangement leads to an unusual interdomain interaction in which the PCTAIRE sequence from one monomer aligns to form a fifth edge strand of the N-terminal β-sheet of an adjacent molecule (Figure 1C). Whether this β-strand is present in the authentic monomeric structure or more probably represents a crystallization artefact of a flexible sequence remains to be determined. The active site is disrupted further by the unusual conformation of the activation segment. The phenyl ring of Phe326 within the DFG motif points into the ATP phosphate-binding site (Figure 1B). Although the side chain of Lys194 (equivalent to CDK2 Lys35) swings towards the back of the active site and is able to make a salt bridge with Glu131 (equivalent to CDK2 Glu11), the disposition of these two residues within the active site does not resemble that seen in active kinase folds. Their location is incompatible with productive ATP binding, but does not prevent binding of the ATP-competitive inhibitor indirubin E804 ((2Z,3E)-2,3′-bi- indole-2,3′(1H,1′H)-dione 3-(O-[(3R)-3,4-dihydroxybutyl] oxime)). Beyond the DFG motif, the sequence that in CDK2 forms oL12 is unwound to form a short β-strand. The activation segment then heads out from the body of the fold and cannot be built between residues 312 and 323.

Overall, CDK2 and CDK16 superimpose rather well over their α-helical C-terminal domains (compare Figures 1A and 1B). As mentioned above, the PCTAIREs are characterized by additional sequences C-terminal to the kinase domain. In the CDK16 structure starting at residue 448, the next 20 residues diverge from CDK2 to form an extended loop that folds up towards the turn at the end of the CDK16 unravelled C-helix. The CDK16 sequence between residues 468 and 473 then returns to a position close to the CDK2 C-terminus. The final 23 CDK16 residues are not included in the model. It remains to be determined as to whether the CDK16 N- and/or C-terminal sequences are sufficient to order the C-helix in the full-length monomeric structure or whether cyclin binding is a prerequisite for stabilizing this region of the structure.

The CDKL protein kinases

CDKL1, CDKL2, CDKL3, CDKL4 and CDKL5 cluster as a separate branch of the kinase tree and share 35–40% sequence identity with CDK2 over the kinase domain. They also share sequence features with the MAPK (mitogen-activated protein kinase) and GSK3 (glycogen synthase kinase 3) families [16]. Among other sequence differences, they were originally distinguished from the CDKs by their variant PSTAIRE motifs within the C-helix namely KKIALRE (Lys-Lys-Ile-Ala-Leu-Arg-Glu) (CDKL1 and CDKL4), KKIAMRE (Lys-Lys-Ile-Ala-Met-Arg-Glu) (CDKL2), NKIAMRE (Asn-Lys-Ile-Ala-Met-Arg-Glu) (CDKL3) and KETTLRE (Lys-Glu-Thr-Thr-Leu-Arg-Glu) (CDKL5) [18]. However, all CDKL kinases contain residues equivalent to CDK2 Arg52 and Glu77 within this motif that are essential for catalysis. Unlike the cell cycle CDKs that encode little more than the kinase domain, the CDKLs are characterized by C-terminal extensions of various lengths. Structures of the kinase domains of CDKL1, CDKL2, CDKL3 and CDKL5, each bound to an ATP-competitive inhibitor, have been deposited recently in the PDB by the Structural Genomics Consortium (Figure 2).

Despite lacking extended sequences outside the catalytic domain, these structures provide interesting insights into CDKL kinase regulation and potential routes to their activation.
The structures of CDKL1 (residues 1–300, PDB code 4AGU), CDKL3 (residues 1–324, PDB code 3ZDU) and CDKL5 (residues 1–303, PDB code 4BGQ) all share structural features of the active kinase fold (Figures 2A, 2C and 2D respectively), especially in respect of the conformations of the DFG motif, the start of the activation segment, and the C-helix. In both the CDKL1 and CDKL3 structures, the salt bridge between a conserved lysine residue on β3 and a glutamate residue on αC (Lys33 and Glu51 in CDK2), normally characteristic of an active kinase conformation, is formed. In the CDKL5 structure, however, although the C-helix is swung in, the side chains of the equivalent residues (Lys42 and Glu60) are too far apart to form this salt bridge. The CDKL2 and CDKL3 structures share an additional extra C-terminal α-helix not present in CDK2 (compare Figures 1A, 2B and 2C).

Both the CDKs and MAPKs are activated by activation segment phosphorylation. This modification occurs on a residue equivalent to CDK2 Thr160 in the cell cycle CDKs and is catalysed in higher eukaryotes by CDK7 [6]. MAPKs are activated specifically by cognate MAPKKs (MAPK kinases) that catalyse dual phosphorylation of a conserved TXY (Thr-Xaa-Tyr) motif [38]. In the structures of CDK2–cyclin A phosphorylated on Thr160 [39] and ERK2 (extracellular-signal-regulated kinase 2) phosphorylated on Thr183 and Tyr185 [40], pThr160 and pThr183 are structurally equivalent and respectively act as a hub to organize a cluster of three arginine residues from the C-helix (Arg50/Arg68, CDK2/ERK2 numbering), catalytic loop (Arg126/Arg146) and activation segment (Arg150/Arg170). In the MAPK family, as illustrated by ERK2, phosphorylated Tyr185, although largely solvent-exposed, interacts with the side chains of Arg189 and

© The Authors Journal compilation © 2013 Biochemical Society
Arg192 and in so doing shapes the protein substrate-binding cleft at the P + 1 site. Despite the use of phosphomimetic mutations in structural studies of CDKL3, the available structures have not confirmed the formation of an equivalent constellation of positively charged residues around negatively charged activation segment in the CDKL3 family members.

In the two CDKL2 structures deposited, the kinase domain has been captured in an inactive conformation (PDB codes 4BBM and 4AAA). The structure of CDKL2 bound to the inhibitor TCS 2312 (PDB code 4BBM) includes the short αL2 helical element at the start of the activation segment and the C-helix is displaced, adopting a register and position reminiscent of that in monomeric CDK2 (compare Figures 1A and 2B).

As a result, the side chain of Arg52 is pointing out into solvent and that of Glu51 is sufficiently flexible as to not be included in the model. The activation segment is ordered into a long helix with Thr159 [in a TDY (Thr-Asp-Tyr) motif and equivalent to CDK2 Thr160] pointing out into solution. However, this part of the structure is clearly flexible, and is potentially ordered by inhibitor binding in the active site and/or lattice contacts given that it is disordered between residues 149 and 160 in the second CDKL2 structure bound to another ATP-competitive inhibitor (PDB code 4AAA; results not shown).

Outside the active-site cleft, both the MAPks and CDKs employ recruitment sites to enhance substrate selectivity [41]. MAPks engage their substrates in part through a docking site (‘D-site’) motif, a bipartite sequence composed of two or three tandem lysine-arginine residues followed by an hydrophobic-Xaa-hydrophobic motif, separated by one to six residues [42]. A comparison of the CDKL kinase structures with that of ERK2 bound to a D-site-containing peptide (PDB code 4H3Q [43]) suggests that the CDKL structure is not compatible with D-site motif binding (results not shown) and that this mechanism of substrate recruitment is unlikely to be conserved between the MAPK and CDKL kinase families.

Taken together, the conservation of potential ligands for both a phosphorylated threonine residue and a phosphorylated tyrosine residue within the CDKL activation segment and the observed sequence disorder in this region in the non-phosphorylated structures suggest that fully activated CDKL kinases are phosphorylated at one or both residues in the activation segment. However, the identity of a potential activating kinase has not been established.

Furthermore, given that the introduction of phosphomimetic residues did not result in an active structure, CDKL kinases might resemble CDKs more closely than MAPks in requiring a cyclin-like regulator to generate the fully active enzyme. Another alternative is that the ability to remodel the active site, that in the CDKs is dependent on the binding of the cyclin subunit, might be provided in the CDKL kinases in cis by the sequences that are not present in the current structures. Further work will dissect the mechanism of CDKL kinase activation, but the structural insights provided so far suggest that they will have novel features that might provide opportunities for selective inhibition if they are validated as targets in particular clinical settings.

Structures of CDK–cyclin complexes

A series of CDK2–cyclin A structures revealed the rearrangements that accompany CDK2 activation that has been applied to other CDK–cyclin cognate pairs as a general mechanism of CDK activation [39,44,45]. However, the generality of this model was challenged first by determination of the structure of a CDK5-p25 complex which adopts an active conformation in the absence of activation loop phosphorylation [46] and then subsequently by the determination of structures for CDK4–cyclin D which demonstrated that for this CDK–cyclin pair, cyclin binding and CDK phosphorylation did not lead to CDK4 adopting an active conformation [47,48]. These observations suggested that substrate binding makes a more significant contribution to the formation of a catalytically poised Michaelis complex than it does in other CDK–cyclin partnerships. The CDK2–cyclin A and CDK4–cyclin D complexes were also distinguished by the relative dispositions of their CDK and cyclin subunits. Residues from the CDK2 N- and C-terminal lobes and both the N- and C-terminal CBFs (cyclin box folds) of cyclin A contribute to a tight extended interface in this complex. In contrast, the interaction interface between CDK4 and cyclin D is much smaller and is predominantly mediated through residues in the CDK4 N-terminal lobe and N-terminal CBF of cyclin D. Since 2010, our knowledge of how structurally diverse the CDKs are has been extended by the determination of a number of structures for CDKs regulating transcription. These structures are beginning to reveal the features that distinguish them from CDKs regulating cell cycle progression.

The overall arrangement of CDK9–cyclin T (P-TEFb) is reminiscent of CDK4–cyclin D in that the elevated position of the cyclin T subunit results in a relatively small interface and a more open structure between the CDK9 and cyclin T C-terminal lobes [49] (Figure 3A). The transcriptional CDKs are also characterized by extended C-terminal sequences beyond the canonical protein kinase fold. It has been shown that this C-terminal tail determines the CDK9 kinetic pathway [50]. When present, CDK9 catalyses an ordered reaction in which ATP is the first substrate to be bound and ADP is the second product to be released. However, in its absence, CDK9 resembles CDK2 and both enzymes, encoding little more than the core kinase domain follow a random model for both substrate binding and product release. Given that this sequence feature is shared among the transcriptional CDKs, it has been proposed that the C-terminal tail might impose a shared kinetic mechanism that could also directly contribute to the pattern of CTD phosphorylation [50]. In vivo, the CDK9 C-terminal tail is (auto)phosphorylated at several positions and phosphorylation is required for P-TEFb binding to HIV TAR RNA [51]. It remains to be determined whether this phosphorylation also affects the recruitment of endogenous regulatory proteins and/or the P-TEFb kinetic mechanism.

The structure of CDK8–cyclin C bound to sorafenib (PDB code 3RGF [52]) resembles CDK9–cyclin T in that the two subunits again interact primarily through their N-terminal...
lobes. Two features of the CDK8–cyclin C structure create a unique interface between the two subunits at this interface (Figure 3B). First, the pose of CDK8 relative to cyclin C is intermediate between the extremes defined by CDK2–cyclin A and CDK9–cyclin T, and, secondly, the interface is extended and stabilized by an additional CDK8 N-terminal helix (αB). This helix has not been observed previously in CDK structures and interacts with helix α5 of the cyclin C N-terminal CBF, a significant and conserved element of the cyclin interface with its CDK partner. Collectively, these interactions stabilize the CDK8 αC helix in a swung-in active conformation.

Two structures of a CDK8–cyclin C apo structure have recently been deposited in which more of the activation segment is included in the model (PDB codes 4G6L and 4F7S). In these structures, the segment starts off with the DMG (Asp-Met-Gly) motif (starting at residue 173) in the ‘DFG-in’ conformation and is visible to Leu184. The segment is then built starting again at Pro246, but the APE motif (which includes Glu203) is not included. The structures suggest an unusual conformation of the activation segment and the immediate sequence following it, as the Cα–Cα distance between Leu184 and Pro246 is only 6.5 Å. The CDK8 structure is not phosphorylated in the activation segment and the CDK8 sequence does not contain a serine or threonine residue in the activation segment at a position comparable with Thr160 in CDK2. However, Arg65, Arg150 and Arg178 (equivalent to CDK2 Arg50, Arg126 and Arg152) are conserved and Arg65 and Arg150 are co-located, although Arg178 is considerably displaced from the position it adopts in other active phosphorylated CDK–cyclin structures. Intriguingly, CDK8 Ser182 in the CDK substrate consensus sequence SPXX (Ser-Pro-Xaa-Lys) is close to CDK2 pThr160 in an overlay of the phosphorylated CDK2–cyclin A structure. The answers to whether CDK8 activity is regulated by phosphorylation within the activation segment and whether the extended sequence mediates novel interactions await further characterization.

It is well documented that CDKs exploit a substrate recruitment site present on the N-terminal CBF of the cyclin subunit to assist in substrate selection. Both certain substrates, such as pRB, and members of the p21/p27 Cip/Kip CDK inhibitor family have been shown to bind to this site to enhance/inhibit CDK activity respectively. In both cases, the protein bound at the recruitment site engages the active site of the associated CDK molecule [53,54]. The structure of monomeric cyclin H (PDB codes 1JKW [55] and 1KXU [56]), the cognate partner of CDK7 was the first to suggest that this site is disabled in the transcriptional CDKs, where it is occluded by the C-terminal helix.

Recent determinations of two CDK9–cyclin T complexes bound to fragments of regulatory proteins raise the intriguing possibility that CDK9 at least might employ a similar mechanism to the cell cycle CDKs to recruit substrates and regulators, but using a novel surface on the cyclin T
C-terminal CBF [57,58] (Figures 3C and 3D). The fragments of HIV Tat (transactivator of transcription) (PDB code 3MIA [57]) (Figure 3C) and of AFF4 (PDB code 4MY [58]) (Figure 3D) that contact cyclin T are both relatively short and adopt α-helical conformations. The longer HIV Tat sequence extends across from the cyclin T subunit to bind to residues within the CDK9 activation segment (Figure 3C). In so doing, it occupies the expanded groove between the cyclin and CDK C-terminal lobes that is a feature of CDK9–cyclin T. AFF4 is an intrinsically disordered protein that helps to scaffold a subset of members of the super elongation complex-like family [59]. These protein assemblies contain P-TeFb and regulate the transcription elongation checkpoint control stage of transcription, mis-regulation of which has been associated with malignant cell phenotypes. The structures of AFF4 and HIV Tat bound to cyclin T show that the two P-TeFb regulators are bound at overlapping, but not perfectly analogous, sites on the cyclin T C-terminal CBF (Figures 4C and 4D). Additional structures containing intact endogenous CDK9 regulatory proteins will be required to determine to what extent the AFF4 and HIV Tat-bound structures delineate a universal binding site for P-TeFb regulators. It will be intriguing to discover whether this regulatory surface is a characteristic of CDKs that regulate transcription in the same way that the recruitment site is present in the cell cycle CDKs. In this regard, sequence and structural alignment of cyclin orthologues has identified a conserved groove specific to cyclin C that lies between the two CBFs [60]. Taken together, these studies suggest that the environment of the transcriptional CDKs within large assemblies may have led to the evolution of a number of conserved protein-binding sites to regulate and localize their activities.

Acknowledgements

We thank our colleagues for discussions and collaboration in studying the CDK–cyclin structure–function relationship.

Funding

Work on CDK complexes in our laboratory is supported by the Medical Research Council.

References
