The crystal structure of cyclin A

NR Brown[1†], MEM Noble[1†], JA Endicott[1*†], EF Garman[1], S Wakatsuki[2], E Mitchell[2], B Rasmussen[2], T Hunt[3] and LN Johnson[1]

[1]Laboratory of Molecular Biophysics and Oxford Centre for Molecular Sciences, Rex Richards Building, South Parks Road, Oxford, OX1 3QU, UK, [2]ESRF, BP 220, F-38043 Grenoble Cedex, France and [3]Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Herts, EN6 3LD, UK

Background: Eukaryotic cell cycle progression is regulated by cyclin dependent protein kinases (CDKs) whose activity is regulated by association with cyclins and by reversible phosphorylation. Cyclins also determine the subcellular location and substrate specificity of CDKs. Cyclins exhibit diverse sequences but all share homology over a region of approximately 100 amino acids, termed the cyclin box. From the determination of the structure of cyclin A, together with results from biochemical and genetic analyses, we can identify which parts of the cyclin molecule may contribute to cyclin A structure and function.

Results: We have solved the crystal structure, at 2.0 Å resolution, of an active recombinant fragment of bovine cyclin A, cyclin A-3, corresponding to residues 171–432 of human cyclin A. The cyclin box has an α-helical fold comprising five α helices. This fold is repeated in the C-terminal region, although this region shares negligible sequence similarity with the cyclin box.

Conclusions: Analysis of residues that are conserved throughout the A, B, and E cyclins identifies two exposed clusters of residues, one of which has recently been shown to be involved in the association with human CDK2. The second cluster may identify another site of cyclin A–protein interaction. Comparison of the structure of the unbound cyclin with the structure of cyclin A complexed with CDK2 reveals that cyclin A does not undergo any significant conformational changes on complex formation. Threading analysis shows that the cyclin-box fold is consistent with the sequences of the transcription factor TFIIB and other functionally related proteins. The structural results indicate a role for the cyclin-box fold both as a template for the cyclin family and as a generalised adaptor molecule in the regulation of transcription.

Structure 15 November 1995, **3**:1235–1247

Key words: cell cycle, cyclin, protein kinase, regulation, X-ray structure

Introduction

Progress through the eukaryotic cell cycle depends on the appropriate activation of members of the cyclin dependent protein kinase (CDK) family (reviewed in [1,2]). Members of the CDK family have homologous sequences which include the subset of residues that are found in all serine/threonine protein kinases and that are essential for catalytic activity [3]. CDKs are activated by binding positive regulatory molecules called cyclins, that were originally named because of their cyclical accumulation and destruction, which occurs concomitantly with rapid rounds of division in cleaving embryos of marine invertebrates and amphibians [4,5]. The subsequent isolation of multiple cyclins has shown that they are diverse in sequence, sharing homology only over approximately a 100-amino-acid region, now called the cyclin box [6,7]. In higher eukaryotes, cyclins have so far been classified into eight families (A to H) on the basis of sequence similarity and their timing of appearance during the cell cycle (reviewed in [2]).

Cyclin A is required for progression through S phase [8–10] and for entry into mitosis [9,11,12]. This dual function probably stems from the ability of cyclin A to form stable, active complexes with either CDC2 or CDK2, unlike the B-type cyclins, which associate exclusively with CDC2, or cyclin E, which forms complexes only with CDK2. Cyclin A is first detected in late G1/early S phase [10,13–15], and its expression steadily rises throughout S and G2. When bound to the kinase, cyclin A is abruptly degraded, just before the metaphase→anaphase transition, by ubiquitin-mediated proteolysis (reviewed in [16]). The binding of cyclin A to CDC2 (sometimes referred to as CDK1) or CDK2 is a prerequisite for kinase activity and promotes phosphorylation on Thr161 or Thr160 (for CDC2 or CDK2, respectively), by the CDK-activating kinase (CAK) [17–21].

The role of cyclin A is controversial, partly because it can form an active protein kinase with either CDK2 or CDC2. At the start of S phase, cyclin A is detectable in a stable complex containing CDK2, the transcription factor E2F, and p107 (a protein related to the tumour suppressor gene product, pRB). This complex can bind to DNA and has histone-kinase activity [22–25]. E2F-binding sites are found upstream of a number of genes required for S phase onset (reviewed in [26]). Binding of cyclin A–CDK2 to E2F may be an important regulatory interaction linking progression through the cell cycle with appropriate control of S-phase-specific gene transcription. The N-terminal sequences of E2F -1, -2 and -3 contain a binding sequence for cyclin A, and through association with E2F-1, cyclin A–CDK2 has been shown to phosphorylate DP-1, the partner of E2F, resulting in decreased DP-1 DNA binding and thus decreased gene transcription *in vitro* [27]. Cyclin A also associates directly with p107,

*Corresponding author. †NRB, MEMN and JAE made equal contributions to this work.

 ISSN 0969-2126

binding to the p107 linker sequence between the two consensus pocket sequences (the 'A pocket' and the 'B pocket') which are essential for E2F binding [28,29].

Whereas association of E2F with cyclin A–CDK2 appears to be important in the regulation of E2F activities at the beginning of S phase, certain D and E cyclins have been implicated in releasing transcriptionally active E2F in G1, from complexes containing E2F and pRB (reviewed in [2]). Hyperphosphorylation of pRB leads to the release of E2F, and increased expression of certain genes (reviewed in [30]). Ectopic expression of cyclin A gives rise to pRB phosphorylation [31] but *in vivo*, pRB hyperphosphorylation occurs at a time in the cell cycle when cyclin A expression is low [10,13–15].

Cyclin A–CDC2 complexes appear to be required later in the cell cycle for entry into M phase and are functionally distinct from cyclin B–CDC2 [8,9,11,13,32,33]. The peak of kinase activity associated with cyclin A occurs before that of cyclin B–CDC2 [13–15], and both cyclins A and B have to be degraded for cells to exit mitosis and enter G1 [34].

Cyclins bind to and activate specific CDK partners. They also mediate the substrate specificity and subcellular localisation of CDKs [35,36]. In order to determine how these specific binding interactions are accommodated in cyclin molecules we have determined the crystal structure of an active fragment of bovine cyclin A, termed cyclin A-3. Our results are considered in the light of the recently published structure of the cyclin A–CDK2 complex [37].

Results

Construction of an active bovine cyclin A fragment

Mutational analysis of cyclin A has shown that a short peptide sequence of nine amino acids, the destruction box, situated near the N terminus is required for ubiquitin-mediated degradation (reviewed in [16]). Cyclin A from which this sequence has been deleted retains CDC2- and CDK2-binding activity, is constitutively active, blocks M-phase exit when expressed *in vivo*, and is not destroyed [38]. Further N-terminal deletions, that were carried out on the basis of secondary structure prediction and the results of limited *in vitro* proteolysis, have shown that cyclin A sequences sufficient for CDK binding and activation of associated kinase activity reside within a proteolytically stable C-terminal fragment ([38]; J Adamczewski, personal communication). The construct cyclin A-3 encodes the bovine cyclin A sequence (EMBL accession no. X68321) beginning at the residue equivalent to Val171 in human cyclin A and has a hexahistidine tag following Leu432 at the C-terminus (Fig. 1). (Throughout this manuscript residues are numbered according to their equivalent in human cyclin A [39]). Cyclin A-3–CDC2 complexes, generated by the addition of bacterially expressed cyclin A-3 to interphase *Xenopus* egg extract (which has low levels of endogenous cyclins), display high histone H1 kinase activity ([38] and J Gannon, personal communication [data not shown]). In this assay system, cyclin A-3 can fully activate the histone-kinase activity of endogenous CDC2. Because the construct lacks the N-terminal destruction box it does not undergo ubiquitin-mediated proteolysis. As described in the Materials and methods section, this construct formed crystals that diffracted well enough to allow an unambiguous structure determination.

The structure of cyclin A-3

The feature of the cyclin A-3 structure that is immediately striking is its high α-helical content. The structure contains twelve helices of more than one turn and up to 18 amino acids in length (Fig. 1), arranged in two lobes linked by two short extended polypeptide regions (Fig. 2). Helices 1–5 encode a compact N-terminal fold. The N-terminal helix, together with helices 1′–5′ and the C-terminal helix, form a larger C-terminal lobe. The hinge between the two lobes is formed by residues Lys194–Gly198, and the extended sequence, Ala307–Pro309, between helices 5 and 1′. Of the 249 residues (Ile182–Leu430) for which a clear conformation can be deduced from the electron density, 156 (63%) are classified as helical by Kabsch and Sander analysis [40]. The first eleven residues and the hexahistidine tag are poorly defined in the electron density, suggesting that they may be flexible or disordered.

After this structure determination was completed, the structure of a human cyclin A–CDK2 complex was reported [37]. The complexed cyclin was prepared by proteolytic digestion and corresponds to residues 173–432. The uncomplexed cyclin (this work) was prepared from recombinant bovine cyclin A corresponding to residues 171–432 in the human sequence. Within this region there are eight amino-acid differences between bovine and human cyclin A. An important finding from a comparison of the structures of free and bound cyclin A is that cyclin A undergoes no significant structural change on association with CDK2. The rms (root mean square) difference is 0.41 Å for 252 equivalent Cα atoms.

The cyclin box encodes a compact α-helical domain

Consideration of the intramolecular interactions that stabilise the cyclin A-3 structure suggests that the cyclin box runs from Tyr199–Leu306. This defines the limits of the cyclin box more securely than has been possible by sequence analysis [6,7,41,42]. The cyclin box itself constitutes a compact domain consisting of five α helices (Figs 1,2). The first helix in the fold, helix 1, is the most highly conserved in the cyclin family. Beginning at Asn208, it occupies a unique position, contributing surface residues, contact residues to the C-terminal domain and residues to the cyclin core. The next two helices, 2 and 3, which are largely buried, form the core of the fold. Residues in helix 2 also contribute to the interdomain interface. Helix 3 is predominantly hydrophobic and interacts extensively, through conserved hydrophobic packing interactions, with amino acids in helical

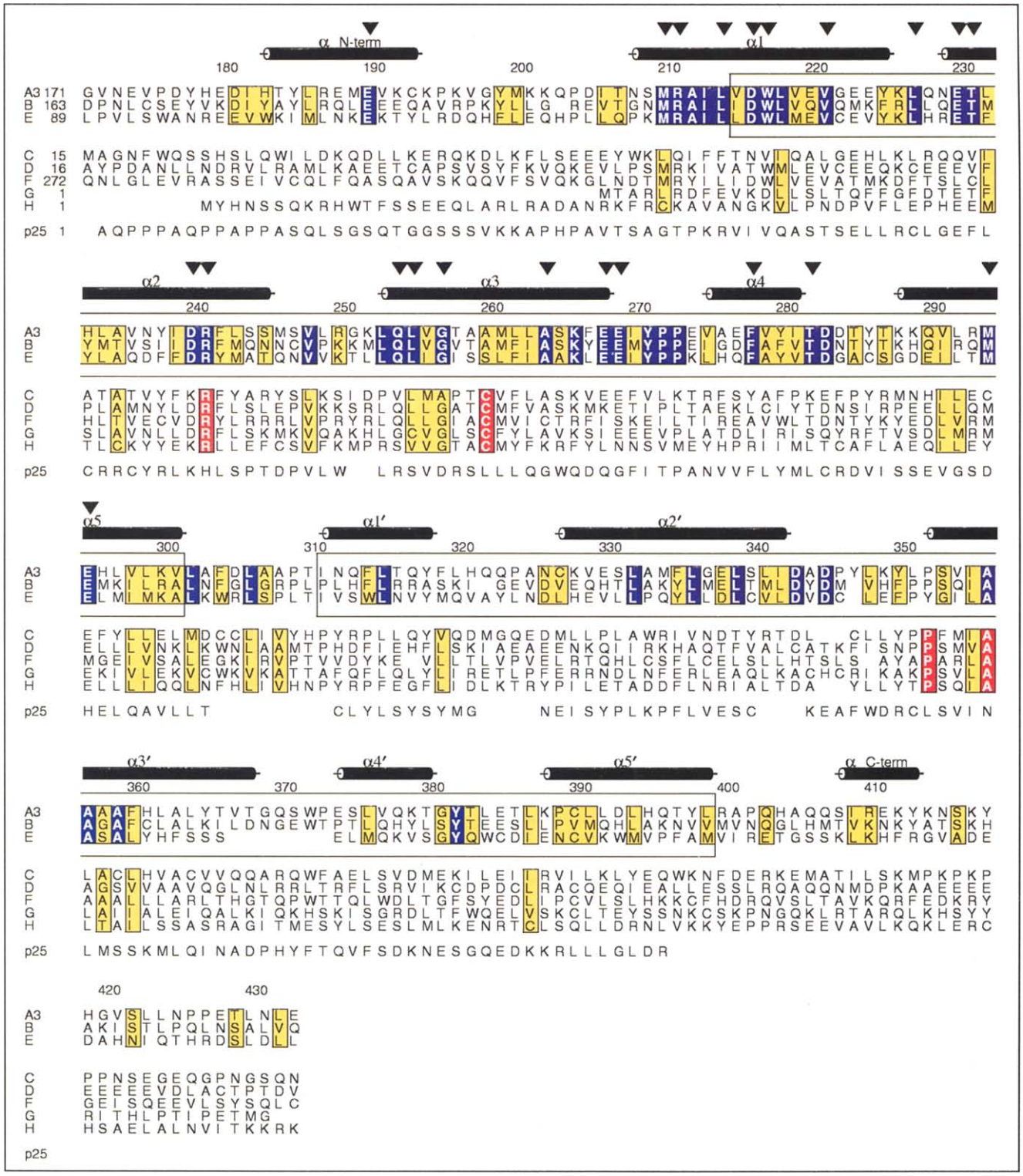

Fig. 1. Multiple sequence alignment of bovine A-3, representative cyclin B–H sequences, and p25, to illustrate sequence diversity between the different cyclin classes. The results of conservation analysis are shown for the group of A-3, B and E cyclins (cyclins interacting with CDC2 or CDK2), and for the group of G, F, D, H and C cyclins. Identities within groups are boxed and coloured blue (for the A, B and E cyclin group) or red (for the C, D, F, G and H cyclin group). Conserved positions are boxed and coloured yellow. The subset of residues that are identical in multiple A, B and E cyclins are shown by black triangles. Experimentally determined helices are represented above the alignment as cylinders. Numbers to the left of the sequences indicate the first residues displayed in the alignment. For the A, B and E sequences, structurally equivalent regions in the two domains are boxed. Sequences shown are A-3 (EMBL accession number X68321), Human B [6], human E [79], rat G [62], human F [80], human H [21], human C and D [78] and human p35 [50]. The figure was generated using ALSCRIPT [76], using an alignment created by MaxHom [77], manually adjusted to minimise deletions within the secondary structural elements.

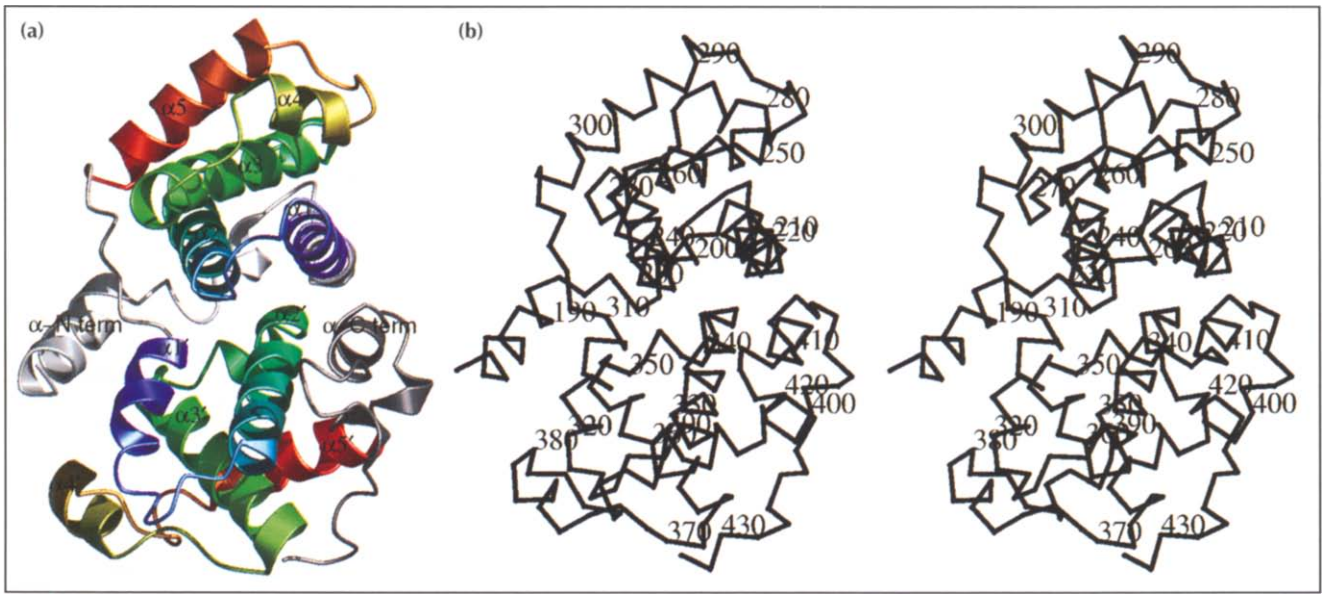

Fig. 2. The fold of bovine cyclin A-3. **(a)** Cartoon representation of the structure of cyclin A-3. Secondary structure defined by DSSP is shown by different representations for helices (spiral ribbons) or random coil stretches of protein. The two domains sharing a common fold (residues 44 (214) to 133 (303) and 141 (311) to 229 (399)) are each colour ramped from blue to red between N- and C-terminal residues of the equivalent region, so that equivalent parts of the structure are approximately the same colour. Regions not equivalent between the two domains are coloured grey (residues 11 (181) to 43 (213), 134 (304) to 140 (310) and 230 (400) to 262 (432). Helices of more than four residues in length are numbered 1–5, and 1′–5′ for each of the cyclin folds. The two helices outside these domains are labelled α-N term and α-C term. The view is close to the rough pseudo-twofold axis which relates helices 1–5 with helices 1′–5′. **(b)** Stereo Cα trace of cyclin A-3. Every tenth residue is labelled. The view is the same as in (a).

regions 1, 2, 4, and 5. The helices cross the molecule so that the surface loops linking the N-terminal helix to helix 1, and helix 2 to helix 3, pack together on one side of the molecule, and those joining helices 1 to 2 and 3 to 4 are adjacent and on the other side of the structure (Fig. 2). A search for known protein structures related to the cyclin fold was performed using the program STAMP [43]. This identified only a limited similarity between cyclin A-3 and members of the α-helical globin family (data not shown).

Sequence conservation within the cyclin fold

In the cyclin fold 22 residues are identical in multiple examples of A, B and E cyclins. These residues (indicated by inverted triangles in Fig. 1 and coloured yellow in Fig. 3) are (with one exception) in the N-terminal cyclin-box fold. Some play a structural role, and others are exposed and form two notable surface clusters. One cluster is part of the CDK2–binding site of cyclin A [37] (Fig. 4) and the second may also identify a site of cyclin–protein interaction.

Arg211 and Asp240 form a striking buried, conserved ionic pair in a hydrophobic environment created by the surrounding residues Tyr199, Ala212, Leu214, Val215, Val236, Ile239, Leu243, Leu341, Ala344, and Leu348 (Fig. 5). These residues have conserved hydrophobic character in many cyclins (Fig. 1), illustrating the importance of this region to structural stability. The Arg211 guanidinium group also makes interdomain contact, through hydrogen bonds, to the main-chain carbonyls of Ser340 and Leu341 at the C-terminal end of helix 2′, indicating a possible role for these residues in domain–domain stabilisation. Mutational analysis supports the structural evidence that these two residues are essential for cyclin A activity. Alteration of Arg211 to alanine or lysine, or alteration of Asp240 to alanine or glutamate results in loss of the ability of cyclin A to bind to and activate CDC2 [38,44]. In human cyclin B, conservative mutation of these residues to lysine–aspartate or arginine–glutamate pairs resulted in proteins that could still bind CDC2, which became phosphorylated on Tyr15, but the complexes containing these mutant cyclins could not be activated. A double mutation of the arginine residue and the adjacent alanine to a serine–serine pair gave a completely inactive cyclin B [45].

Arg241 is partially exposed (Fig. 3b) and is involved in interactions with the main-chain carbonyl oxygen of Asp305 which is part of the surface loop that links helices 5 and 1′. Glu295 is also partially exposed and structurally important (Fig. 3b). This residue interacts with Lys266 in a hydrophobic surface environment created by Leu262 and Leu263. A number of other conserved amino acids (Thr231, Leu255, Gly257, Ala264, Phe278, Thr282, and Met294) are buried. The tight hydrophobic packing around sequence loci 257 and 264 necessitates that the residues occupying these positions should be small.

The cyclins A, B and E (i.e. those cyclins that bind to CDC2 or CDK2 or both) have a marked area of negative surface charge generated by the folding together of Glu230 at the end of helix 2, with Glu268 and Glu269 at the C terminus of helix 3 (Fig. 3b). Above this negative

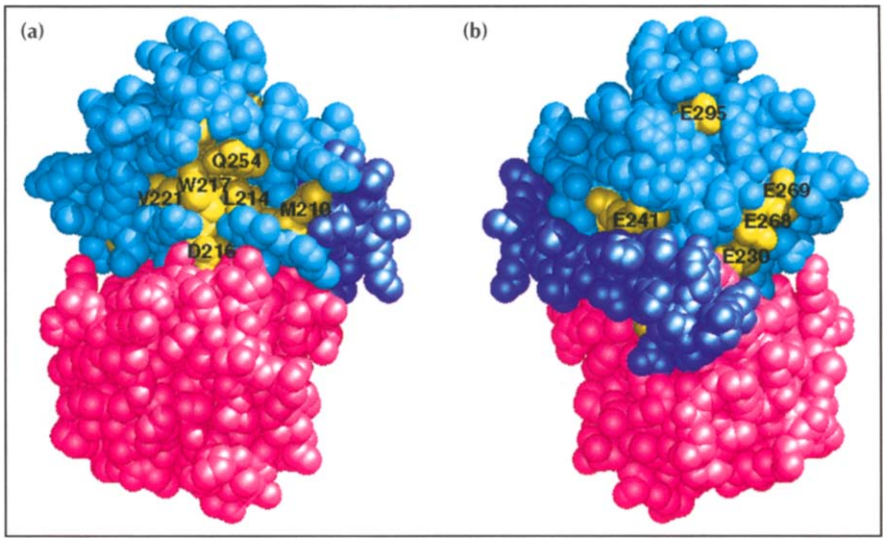

Fig. 3. Colour coded CPK models of cyclin A-3. Residues preceding the cyclin box (1–37, (171–207)) are coloured blue, residues of the N-terminal cyclin box domain (38–137, (208–307)) are coloured cyan, and residues of the C-terminal cyclin box domain (138–262 (308–432)) are coloured magenta. Residues conserved throughout cyclins A, B, and E, indicated by an inverted triangle in Figure 1, are coloured yellow. Residues conserved by this criterion which present a significant surface feature are identified by their residue type (single letter code) and sequence number (relative to the cyclin A-3 sequence). The views in part **(a)** and **(b)** are related by 180° rotation around a vertical axis.

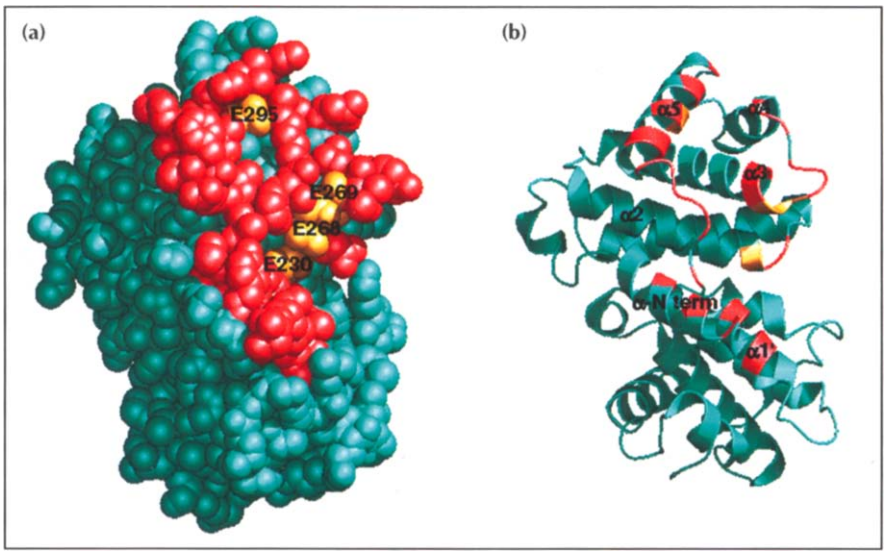

Fig. 4. Location of cyclin A residues in contact with CDK2. **(a)** A CPK model. **(b)** A ribbon diagram. Contact residues identified by Jeffrey *et al.* in the cyclin A–CDK2 complex structure [37] are coloured red. Residues that were also identified by sequence conservation analysis are labelled and coloured gold in (a). Helical regions that contribute contact residues are labelled in (b). The view in this figure is identical to that of Figure 3b.

surface, a conserved leucine residue, Leu227, is also exposed (Fig. 1). The cyclin A–CDK2 structure shows that these three glutamate residues contribute to the extended surface of the cyclin A–CDK2 interaction [37] (Fig. 4). Mutation of Glu230 to glutamine had shown that, at least in cyclin B, this residue is important in promoting dephosphorylation and activation of CDC2 [45]. Interestingly, in the D-cyclin family, members of which bind specifically to CDK4 and CDK6, Glu268 is replaced by a residue that is positively charged.

On the opposite side of the cyclin box from the CDK2-binding site, there is a second notable cluster of conserved residues that may identify another site of cyclin A–protein interaction. Trp217 and Gln254 are adjacent and form the core of a conserved surface on the cyclin A molecule that is shared with the B and E cyclins (Fig. 3a). Trp217 sits in a cluster of four hydrophobic residues (Val221, Leu218, Leu214, and Ile213). Other amino acids in helix 1 (Val215, Asp216, Val219, and Glu223) pack with complementary sequences in the C-terminal domain. Adjacent to Trp217, the A cyclins have a cluster of negatively charged, solvent-accessible residues at the end of helix 1 (Glu220, Glu223 and Glu224), whereas the B cyclins share neutral or positively charged residues in this region (Fig. 1). Two residues in helix 1, conserved in most cyclins, complete the site: Met210 and Ile213 (Figs. 1,3a).

The C-terminal domain fold contains a cyclin-box fold

The C-terminal domain starts with residue Thr310, at the start of helix 1′. Residues 310–432 show very little sequence conservation between members of the cyclin family and alignments are difficult. The surprising result from the determination of the cyclin A-3 structure is that, although in this region the cyclin sequences share little sequence similarity with each other or with the cyclin box consensus sequence, the fold of this domain is strikingly similar to that of the cyclin box (Fig. 2); for 70 Cα positions the rms difference is 1.7 Å. Residues 215–303 in the N-terminal domain superimpose on residues 311–399. Helices 1–5 of the cyclin-box fold are equivalent, respectively, to helices 1′–5′. Residues in the C-terminal cyclin-box fold that correspond to those that

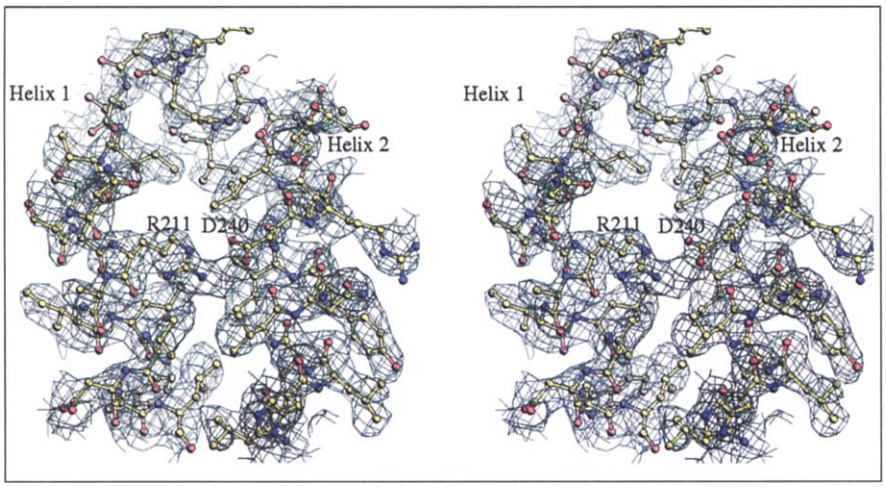

Fig. 5. Refined model and experimental electron density around the two conserved residues Arg211 and Asp240. Electron density is from the DM-modified MIRAS map at 2.4 Å resolution, and is contoured at the rms value of the electron density of the map. Arg211 and Asp240 form a buried salt bridge connecting helices 1 and 2. Arg211 is further involved in interactions, across the domain–domain interface, with carbonyl oxygens at the end of helix 2′.

interact with CDK2 in the N-terminal fold, are partly shielded by residues close to the C terminus. In addition, the C-terminal fold lacks an equivalent to the N-terminal helix. These structural differences may explain why the C-terminal domain does not have a CDK-binding site.

Within the cyclin-A family there is strong sequence conservation in the C-terminal domain, but there is less conservation between cyclins A and B (one exception being conservation of Phe314), and almost no detectable sequence similarity between A-type and E-type cyclins. There is a cluster of conserved cyclin A residues in the C-terminal domain at structurally equivalent positions to the conserved amino acids Trp217, Gln254 and Tyr286 in the cyclin box. Ser353, in the proline–serine motif at the beginning of helix 3′, is equivalent to Gln254; Phe314, in helix 1′, is the C-terminal analogue of Trp217; and Tyr382, in the loop between helices 4′ and 5′, is equivalent to Tyr286. The N-terminal helix interacts closely with residues in the C-terminal domain. His183, Leu186, Arg187 and Glu190 screen Phe314 from the solvent. Glu190 is partly buried and interacts with side-chain and amino groups of Ser353. Met189 and Cys193 shield Pro309 from the solvent.

N-terminal deletions of residues up to the equivalent of residue 176 in the human sequence have no effect on the function of *Xenopus* cyclin, but deletion of residues from 183 onwards (the start of the N-terminal helix) results in loss of activity [38]. The structure of cyclin A–CDK2 confirms that the N-terminal helix contributes residues required for CDK2 binding [37]. Deletions of the N-terminal helix would expose conserved hydrophobic residues in the C-terminal fold and so its loss could affect the structural integrity of this region.

Phe314 is the first residue in the motif Phe–Leu–Arg–Arg–X–Ser–Lys (where X is any amino acid), which is a characteristic of the cyclin B sequences [38]. It is intriguing that mutational and sequence analysis that defined regions characteristic of the A- and B-type cyclins should have identified regions that are equivalent in the cyclin-box and C-terminal folds, respectively [38,41]. The residues in the cyclin A-3 structure that are equivalent to the two arginine and lysine residues in the cyclin B Phe–Leu–Arg–Arg–X–Ser–Lys motif are on the surface.

Helices 2′ and 3′, like their counterparts (helices 2 and 3) in the cyclin box, are hydrophobic and pack into the core of the fold and again, this is a tightly packed hydrophobic region. An exception, however, is the buried glutamate residue Glu338 in helix 2′. This amino acid forms an ion pair with Lys412 in the C-terminal helix. Both of these residues are conserved between the A and B cyclins. The most notable sequence conservation in the C-terminal domain is the cluster of small residues in the middle of helix 3′ (Fig. 1).

Stabilisation of the domain interface results from extensive hydrophobic packing between residues Leu341 and Ile342 at the end of helix 2′, Leu232 and Val236 from helix 2, and Ala212, Val215 and Val219 in helix 1. The helix 2′–helix 3′ loop is the site of a 44-amino-acid insertion in WHI1, a G1 (CLN) *S. cerevisiae* cyclin [46]. These residues would lie below the surface loop between the N-terminal helix and helix 1, away from the conserved surface regions identified by comparative structure and sequence analysis.

Residues Leu332, Phe335 and Leu339 in helix 2′ are conserved in the A and B cyclins and pack against hydrophobic residues in the C-terminal helix. Cyclin A or B constructs with short deletions at the C terminus cannot bind CDK2 [38,41,42,47]. The cyclin A structure shows that deletions of the region C-terminal to Ser422 would expose Leu332 to solvent. Further deletions into the C-terminal helix would increasingly expose the hydrophobic face of helix 2′ and ultimately affect residues interacting with helix 1. Like deletions of the N-terminal helix, exposure of buried hydrophobic residues could affect the folding of this domain.

Sequences compatible with a cyclin-box fold are present in other regulatory proteins

The two domains of the cyclin A-3 structure illustrate that proteins with few identities (13%) can adopt similar

folds. The technique of threading has been shown to be a sensitive tool for detecting whether an amino-acid sequence is compatible with a given tertiary structure [48]. The implementation of the threading algorithm in the program THREADER [48] produces a pseudo-energy term from the optimal alignment of an amino-acid sequence of unknown structure onto a three-dimensional (3D) template of known structure. The pseudo-energy term comprises the sum of residue-to-residue distance-dependent pair potentials resulting from the optimal threading of sequence onto template. By comparing a candidate sequence with a library of potential folds, the technique can identify which of the potential folds is the most likely to be adopted by the candidate sequence.

Threading was used, in this case, to test hypotheses from the literature that the cyclin fold is common to a diverse family of proteins, extending beyond the previously recognised cyclin family [49]. For this analysis, candidate sequences were threaded against a database of 254 folds, representing diverse tertiary structures and three additional templates corresponding to the entire structure of cyclin A-3; the N-terminal region (residues 182–309); and the C-terminal region (residues 310–430). (Only results for the entire cyclin structure are shown.) The candidate sequences used corresponded to p25, an active fragment of p35 (the positive regulatory subunit of CDK5 [50,51]) various portions of human p107 [52], and transcription factor IIB (TFIIB) [53].

The results of this analysis are summarised in Table 1. This table presents both the ranking and the filtered 'Z score' of the intact cyclin template, and of the highest-scoring non-cyclin template, for each of the candidate sequences tested. A ranking of 1 for a given template indicates that THREADER has identified that template as the most likely tertiary structure for the candidate sequence. The filtered Z score is a measure of how many fold lower (negative) or higher (positive) the energy of a particular sequence-template threading is, compared with the average energy observed for that sequence threaded against all of the templates. Thus, a large negative number indicates a low energy, and so a high compatibility of the candidate sequence with a particular template. According to Jones *et al.* [54], a Z score of –3.5 or lower is indicative of a positive result to the analysis.

The most striking result is that the B-pocket sequence of p107, TFIIB, and p25 all have the cyclin fold ranked 1 (Table 1, rows 3, 5, and 6). Thus, threading lends strong support to the hypothesis that each of these sequences contains a cyclin-box-like motif. In the case of p107 and TFIIB, this hypothesis was advanced by Gibson *et al.* [49] on the basis of sequence comparison results using a profile-based approach. They identified the B-pocket sequence of p107 as containing a single copy of the repeated unit present in a cyclin molecule, which they termed the 'TR' (for TFIIB repeat). We tested this conclusion by using the portion of p107 C-terminal to the identified TR as a candidate sequence in the threading protocol (Table 1, row 4). For this candidate sequence, however, the cyclin templates are uniformly ranked low down the list of potential tertiary folds, indicating that a second cyclin-box fold is unlikely in this region.

Table 1. Threading of sequences onto the cyclin structure.

	Sequence	Cyclin A3 Rank†	Cyclin A3 Score‡	Non-cyclin* CODE§	Non-cyclin* Rank†	Non-cyclin* Score‡
1	Phosphorylase kinase (γ-subunit; 20–298)	15	–1.8	E1APM	1	–3.0
2	p107 (A pocket; 252–451)	12	–1.6	A1COL	1	–3.5
3	p107 (B pocket; 648–816)	1	–4.2	1EZM	4	–2.5
4	p107 (B pocket to end; 817–935)	73	1.0	A1GDH	1	–2.0
5	p25 (99–307 of p35)	1	–3.6	1OFV	3	–2.8
6	TFIIB (1–316)	1	–3.8	A2HHM	2	–3.5
7	UDG (Cyclin box to C terminus; 193–326)	250	–1.4	1ECA	1	–3.1

*Highest ranking template not derived from the cyclin A-3 structure. †Position in list of potential folds ordered by filtered Z score. ‡Filtered Z score=(pseudo energy–mean pseudo energy)/(standard deviation of pseudo energy), where sequence-template matches involving only small fractions of the candidate sequence are excluded from the calculations. §PDB code of the highest ranking non-cyclin template: 1APM, cAMP dependent protein kinase; 1COL, colicin A; 1EZM, elastase; 1GDH, D-glycerate dehydrogenase; 1OFV, flavodoxin; 2HHM, inositol monophosphatase; 1ECA, erythrocruorin.

An alternative possibility, that a second copy of the cyclin-box fold might constitute the A-pocket sequence of p107, was tested by using the p107 A-pocket sequence as a candidate sequence in the threading protocol. Although none of the cyclin templates are ranked 1 (Table 1, row 2), a template corresponding to the first cyclin-box fold is ranked fourth, with an energy not far from that of the preferred template (data not shown). The most preferred template, with PDB code 1COL, is the all-helical protein colicin A. We cannot rule out the possibility that the fold of the A-pocket sequence of p107 is at least related to the cyclin-box fold on the basis of this result.

A protein (here termed UDG), produced from a human genomic DNA library, and that is believed to have uracil DNA glycosylase activity, has been shown to bear extensive homology to members of the cyclin family [55]. In this case, the similarity is also limited to the cyclin-box region, raising the question of whether this protein contains a single copy of the cyclin-box fold, or whether, like cyclin A, it contains a tandem repeat of this structural motif. This was tested by threading the sequence of UDG that is downstream from the sequence-related cyclin-box region (Table 1, row 7). Again, none of the

cyclin templates were ranked 1, but the cyclin-box folds 1 and 2 are both ranked highly (fifth and sixth, respectively, data not shown); thus it cannot be ruled out that the candidate sequence contains a cyclin-box-like structural motif. The occurrence of cyclin-box folds in some of these proteins is summarised in Figure 6.

The catalytic core domain of phosphorylase kinase was included as a negative control. This protein, which is of similar length to cyclin A-3, exhibits a classic kinase fold composed of a β sheet and an α-helical domain. Threading showed that this sequence is not compatible with the cyclin fold (Table 1, row 1) but is compatible with that of the cyclic AMP dependent protein kinase structure (PDB code 1APM), with a filtered Z score of –3.0. This result gives further confidence in the ability of the threading algorithm to distinguish correct folds.

Discussion

The cyclin box

Cyclins are a sequence-diverse family of proteins that are essential for CDK activity. Sequence alignments have previously shown that cyclins share a conserved region of approximately 100 amino acids in the centre of the molecule, and these comparisons, together with the results of site-directed mutagenesis, supported the idea that this region is essential for CDK binding and activation. The determination of the crystal structure of cyclin A-3 now shows that the cyclin box sequence encodes a compact five-helical domain. The cyclin-box fold is highly structured and there are no extended loops. Notably, the diverse cyclin sequences are very similar in length over the cyclin box, with the exceptions of the CLN-encoded cyclins. The cyclin A structure, however, shows that the insertions present in the Clns could easily be accommodated in a surface loop.

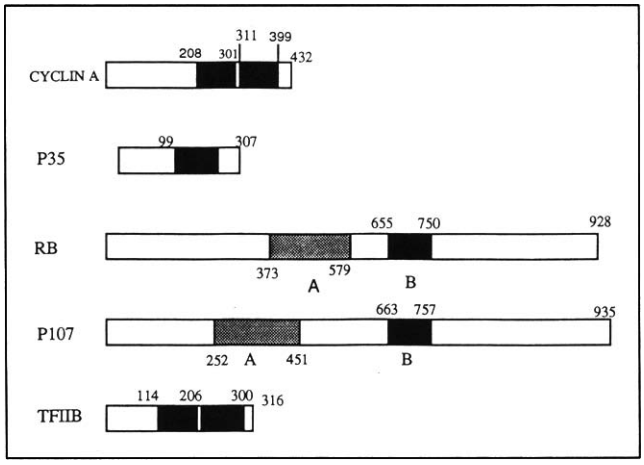

Fig. 6. A schematic representation of the occurrence of cyclin-box folds. Cyclin-box folds are denoted in black. In cyclin A (this work) the upper numbers refer to the full length cyclin A molecule. In pRB and p107 the numbering of the A-pocket sequence is according to [52], and the numbering in the B-pocket sequence corresponds to that of the sequences with similarities to TFIIB and cyclin, identified by [49]. For TFIIB the numbering of the domains is according to [49].

Comparison with cyclin A–CDK2

A comparison of the structure of cyclin A-3 with that of cyclin A–CDK2 [37] reveals no significant structural changes between the free and bound forms of cyclin A-3. In striking contrast, CDK2 undergoes a significant conformational response on complex formation that results in activation of the enzyme. The surface of cyclin A-3 thus appears to provide a rigid template against which the CDK responds to produce a complementary surface. These comparisons show a rigid regulatory molecule and a pliable enzyme structure.

The surface residues of cyclin A-3 identified in the complex as being important for complex formation include the exposed surface patch of Glu268 and Glu269 and the residues Met294 and Glu295. These residues were also identified as conserved surface residues from sequence conservation of the A, B and E cyclins in this analysis of the free cyclin structure. The cyclin A–CDK2 interface, however, also contains other cyclin A residues that are not completely conserved in these other cyclins, suggesting that amino-acid differences can be tolerated within this binding region. The highly conserved patch identified in this work, that includes Trp217 and Gln254, is not involved in the interface with CDK2 and it may be that this region plays a role in mediating contacts to other protein partners. The recently identified CDK inhibitors p21 and/or p27 are attractive candidates for those partners (reviewed in [2]). One model that can be envisaged is that those molecules that directly regulate kinase activity bind to the highly conserved N-terminal cyclin-box fold, whereas molecules that effect localisation or regulate the complex, such as p107, bind to the C-terminal cyclin-box fold.

Although there is little (13%) sequence identity between the two domains of the cyclin A-3 structure, these sequences adopt similar folds. This result parallels the observation that, even within the cyclin box, diverse cyclins do not share identity at residues that appear important for the structural integrity of cyclin A. A number of protein classes are found that share similar 3D structures but no sequence identity (reviewed in [56]). The tandem duplication of the cyclin-box structure in cyclin A was predicted by extensive cyclin-sequence analysis [49]. The sequence similarity that suggested the tandem duplication in the cyclin structure centred on the conserved sequence of alanine and small hydrophobic residues that are now shown to comprise helices 3 and 3′, respectively. The structural importance of these residues is emphasised by their conservation in diverse cyclin-box fold sequences.

Implications for other cyclins

Sequence alignment suggests that the structures of cyclins G and H are compatible with a tandemly duplicated cyclin-box fold. Cyclin H, together with CDK7,

constitutes CAK, the CDK-activating kinase [21,57]. CAK was re-isolated as an integral component of transcription factor IIH (TFIIH) which phosphorylates the C-terminal domain of the largest subunit of RNA polymerase II [58–61]. Alignment of cyclin H with cyclin A shows that cyclin H would contain little more than the duplicated cyclin-box fold (Fig. 1). Intriguingly, the naturally occurring first residue of cyclin H corresponds to the first residue of cyclin A-3, the sequence of which has been identified by mutational analysis to be the minimum required for CDK binding and catalytic activity [38]. The cyclin H sequence is of sufficient length to encode a helix equivalent to the N-terminal helix. This helix, lying outside the cyclin-box fold but in intimate contact with residues in the C-terminal domain, may be a conserved feature of CDK binding. The cyclin A–CDK2 structure shows that this helix contributes residues to the CDK2-binding site [37]. A counter example is cyclin G which, by sequence alignment, starts at helix 1 and lacks the N-terminal helix. A CDK partner for cyclin G has, however, not been isolated and its function is still unknown [62].

Threading analysis strongly suggests that p25, an active proteolytic fragment of p35, [50,51] a CDK5 partner and activator with essentially no sequence similarity to cyclins, shares structural similarity to the cyclins. p25 results from an N-terminal deletion of 98 residues from the p35 sequence. Alignment from the threading against cyclin A-3 shows that the sequence is not long enough to encode two complete cyclin-box folds. This result suggests that, at least for the CDK5–p25 complex, a second cyclin-box fold is not required for the positive regulatory function of p25.

Structural similarity with TFIIB and other functionally related molecules

A wider role for the cyclin fold as a more generalised adaptor motif is supported by the results of threading analysis, which show that sequences compatible with the cyclin-box fold are to be found in a wide range of regulatory molecules, most notably the transcription factor TFIIB and the B pocket of the tumour suppressor gene products pRB and the related protein, p107 (Fig. 6). The B-pocket sequence in pRB and p107 is part of a bipartite binding domain required for their association with the transcription factor E2F/DP [63]. The results presented in this paper are consistent with at least one of the regions associated with E2F binding having a compact fold. Still to be determined are the relative contributions of the B-pocket sequence (compatible with a cyclin-box fold) and the A-pocket sequence to the binding of E2F/DP to the pocket-protein family. Sequence similarity between the pRB A-pocket sequence and TFIID, and between the pRB B-pocket sequence and TFIIB, has been reported previously [64]. This result provides additional supporting evidence for a model in which the A and B pocket sequences encode independent (but possibly cooperative) structures. Intriguingly, the cyclin family, the pocket-protein family, and UDG, have all been functionally linked to E2F activity. In the case of the association of E2F-1 with UDG, the suggestion was made that UDG and E2F-1 may interact biochemically [65], although the authors stated that no evidence was available to support the contention. The fact that a sequence compatible with the cyclin-box fold is important for E2F binding in pRB supports the contention that UDG may directly bind E2F.

The discovery that molecules controlling progression through the cell cycle are also closely involved in regulating gene transcription has been one of the most exciting recent developments in these two fields. The determination and analysis of the structure of cyclin A strongly suggests that a number of molecules important to both processes share the cyclin-box fold as a structural motif. Within the cyclin family, the concentration of conserved residues in the N-terminal cyclin box suggested that this region of these molecules is responsible for their shared function of CDK binding. The unique properties of the different cyclins may be associated with the other domains of the proteins of this family. The characterisation of protein binding to the cyclin-box fold will be a productive and illuminating task.

Biological implications

Cyclins are a diverse family of proteins that have only limited sequence similarity but share the common function of activating members of the cyclin-dependent kinase (CDK) family. The activity of cyclin–CDK complexes can be further regulated by both phosphorylation and association with additional proteins. Seven CDK and eight cyclin classes have been identified in higher eukaryotes. The CDKs share sequence similarity and are structurally homologous to other protein kinases. The cyclins, however, share only limited homology within the cyclin-box sequence. CDKs undergo large changes in conformation on binding cyclins but the crystal structure of unbound cyclin now shows that cyclin conformation does not change significantly. Instead, it provides a rigid surface of complementary architecture to the active CDK.

The structure of cyclin A-3, an active recombinant fragment of bovine cyclin A, shows that the cyclin box encodes a novel protein fold comprising a five-helix bundle that is tandemly repeated in the C-terminal domain. Sequence comparison between the A, B and E cyclin classes, which bind CDK2 and /or CDC2, shows that there are two surface clusters of conserved residues in the cyclin-box fold. One cluster is part of the extended surface of cyclin A–CDK2 interaction. The second is on the other side of the molecule and may identify an additional site of cyclin–protein interaction. The cyclin A structure provides a framework for sequence analysis aimed at

determining the origins of the similarities and differences between the cyclin classes in the binding of CDK and other proteins.

Threading analysis has been used to test the extent to which the cyclin-box motif occurs in a range of regulatory proteins, including the transcription factor TFIIB and the B-pocket sequence of the pocket proteins. The results of this analysis strongly suggest that the cyclin-box fold is present in the extended cyclin family and, as previously proposed, in a number of molecules that regulate transcription. We propose that the cyclin box encodes a structurally stable protein fold that, through protein–protein association, can modulate the activity of a number of molecules involved in the regulation of the cell cycle and transcription.

Material and methods

Expression and purification of cyclin A-3

A cDNA fragment encoding the bovine cyclin A sequence (equivalent to that of human cyclin A from Val171 to the C terminus) was amplified by PCR and cloned into the bacterial expression vector pET21d (Novogen, Madison, WI). PCR amplification introduced a methionine residue at the N terminus and changed the next residue in the sequence from serine to glycine. Subsequent amino-acid sequencing showed that the N-terminal residue of cyclin A-3 is a glycine. This construct also modified the bovine sequence by introducing six consecutive histidine residues following the C-terminal valine. Protein expression was induced in *E. coli* B834/LYS-S cells in log-phase by addition of IPTG to a concentration of 0.1 mM, followed by incubation at 37°C for 30 min, then at 30°C for 3 h. The *E. coli* cell pellet was harvested by centrifugation and resuspended in N buffer (300 mM NaCl, 0.01% monothioglycerol, 0.01% NaN_3, 50 mM Tris-HCl, pH 7.5) containing added protease inhibitors, (0.1 mM PMSF, 0.7 $\mu g\ ml^{-1}$ pepstatin, 0.5 $\mu g\ ml^{-1}$ leupeptin, 0.1 mM benzamidine), 20 $\mu g\ ml^{-1}$ DNAase, 20 $\mu g\ ml^{-1}$ RNAse, 0.1% Triton X-100 and 10 mM $MgCl_2$. Following cell lysis by freeze/thaw, the lysates were incubated on ice for 30 min, then centrifuged at 10000 g for 20 min at 4°C. The cleared lysate was immediately loaded onto a Ni-NTA column (QIAGEN) equilibrated in N buffer and eluted with a linear gradient of imidazole (0.0–0.5 M) in N buffer. Fractions containing cyclin A-3 were identified by SDS-PAGE and pooled, and $MgCl_2$ was added to a final concentration of 0.1 M. $MgCl_2$ addition appears to be essential at this stage of the preparation to prevent further protein aggregation. The volume of the pool was reduced to approximately 5 ml (Centriprep, Amicon, Beverly, MA) before gel filtration on Superdex75 (HR26/60, Pharmacia, Uppsala, Sweden) equilibrated in 0.1M $MgCl_2$, 0.01% monothioglycerol, 0.01% NaN_3, 50 mM Tris-HCl, pH 8.5. Cyclin A-3 eluted in two peaks, the first corresponding to oligomeric aggregates and the second consistent with a monomer population. This latter peak was pooled and concentrated to 5–10 $mg\ ml^{-1}$ (Centricon, Amicon) for crystallisation trials. Approximately 10 mg of cyclin A-3 of sufficient purity for crystallisation trials was produced from 1 L of bacterial cell culture. Dynamic light scattering measurements (M801 Molecular Size Detector, Protein Solutions Inc., Charlottesville, VA) confirmed the monodispersity of the cyclin A-3 preparation. Cyclin A-3 appears to be stable when stored under these conditions at 4°C for several weeks.

Cyclin A-3 crystallisation

Initial crystallisation conditions were identified following screening using sparse matrix sampling conditions (Crystal Screen, Hampton Research, Riverside, CA). Cyclin A-3 crystals were grown using the sitting drop vapour diffusion technique at 20°C. Typically 1–2 μl of cyclin A-3, at 5–10 $mg\ ml^{-1}$ in 0.1 M $MgCl_2$, 0.01% monothioglycerol, 0.01% NaN_3, 50 mM Tris-HCl, pH 8.5, were mixed with an equal volume of 25–26% methoxyPEG 5K (Fluka, Buchs, Switzerland) containing 0.2 M $LiSO_4$ and 0.1 M Tris-HCl, pH 7.0–7.5 and allowed to equilibrate with 1 ml of the same buffer. Crystals appeared within one week, and had a tendency to be plate-like (typically 100–200 μm by 50–75 μm and rarely thicker than 10 μm), with clustered growth. Extensive screening around these conditions failed to eliminate these problems.

Data collection and processing

Cyclin A-3 crystallised in the orthorhombic spacegroup $P2_12_12_1$ and with unit-cell dimensions (at 100 K) a=52.7 Å, b=60.9 Å, c=83.5 Å. There is one molecule of cyclin A-3 (262 amino acids; 32417 Da) per asymmetric unit, giving a V_m=2.18 $Å^3\ Da^{-1}$, which indicates a solvent content of approximately 30%. Because the crystals were typically only 5–10 μm in the smallest dimension, data could only be collected at Synchrotron Radiation Sources. All data sets were collected at 100 K, and crystals were soaked in cryoprotectant (30% methoxyPEG, 0.2 M $LiSO_4$, 15% glycerol, 0.1 M Tris, pH7–7.5) 5 min before mounting, using the method of Teng [66]. An initial native data set was collected at beamline 9.6 of the Daresbury SRS at a wavelength of 0.87 Å, with a crystal-to-detector distance of 300 mm on a 30 cm Mar Research imaging plate detector, giving a resolution of 1.9 Å at the edge. Data were collected with individual oscillations of 1.5° and exposures of approximately 2 min. Subsequent screening of heavy atom compounds was carried out at both the ESRF, BL4 and at Daresbury stations 7.2 and 9.6. BL4 was operating at a wavelength of 0.993 Å, equipped with a Mar30 imaging plate set at a distance of 200 mm, giving a resolution of 2.4 Å at the edge of the plate. Typically images were collected over 2° oscillations, with 10 s exposures. Further derivative data and an improved native data set were collected at Daresbury station 7.2 operating at a wavelength of 1.488 Å. Data were collected on a Mar18 imaging plate detector set to give either 2.0 Å or 2.4 Å resolution at the edge, with oscillations of 1.5° and typical exposures of 180 s. The images were integrated with the DENZO package [67], and reflections were subsequently scaled and merged using SCALEPACK [67]. Structure factors were derived from intensities by the program TRUNCATE from the CCP4 package [68]. Statistics of the data sets used in the structure determination are given in Table 2.

Structure solution and refinement

A single mercury site was identified in the λ=0.993 Å ETMS data set by inspection of the Harker sections of an anomalous Patterson. Phases from this data set were used in difference Fouriers to identify sites in the EMP derivative and in the ETMS derivative collected at λ=1.488 Å. All three derivatives shared one common site, with the EMP derivative having three subsidiary sites, and the λ=1.488 Å EMTS data set having one subsidiary site. All other potential derivatives were found to contain either no interpretable heavy atom sites, or only the major site. Heavy atom refinement and phasing were carried out to the limit of available data for each derivative (typically 2.4 Å) using MLPHARE [69]. Statistics of the phasing are summarised in Table 2. Only one of the two possible

enantiomeric heavy atom structures gave an MIR map with a defined solvent boundary. Phases from this procedure were improved through density modification by the program DM [70], taking advantage of the solvent flattening, histogram matching, and Sayre's equation options in a resolution-based extension scheme, beginning at 4.5 Å and extending to 2.4 Å. The resulting phases were used to calculate a map at 2.4 Å resolution which was skeletonized by the BONES algorithm [71]. α-helical stretches could clearly be seen in the skeleton, allowing an essentially complete model to be built directly, with the exception of the N and C termini, and the connection between helices 1′ and 2′. The map was of sufficient quality to allow unambiguous assignment of the sequence on the basis of features in the electron density.

Refinement of the structure at 2.0 Å resolution was pursued by alternate cycles of Simulated Annealing refinement using X-PLOR [72] and manual refitting using O [71]. In the X-PLOR refinement, a 2σ cutoff was applied to the data. The C terminus could be completed in the first $2F_o-F_c$ map up to residue Leu432, corresponding to the last genuine residue of bovine cyclin A. When the conventional R factor reached 25%, waters were included in the model with the help of the program ARP [73]. Statistics of the refined structure are given in Table 2.

Structure analysis

After refinement, the structure was analysed for geometric quality using the PROCHECK package [74]. This awarded the structure a 'G-value' more than two standard deviations better than the mean for structures of the same nominal resolution.

The equivalence of the N and C terminal domains was identified by the program STAMP [43], which aligned 66 Cα positions of the N and C-terminal cyclin domains with a rms agreement of 1.8 Å. This alignment was improved using the program O [71]. A STAMP comparison of the cyclin fold with a data base of other known structures revealed some similarity with members of the globin superfamily. For this comparison, however, only some 40 Cα positions could be equivalenced, corresponding to four of the helices of the cyclin fold.

Threading analyses

Comparison of other potential cyclin-fold sequences with the cyclin structure was performed with the program THREADER [48]. This program performs an optimal alignment of a candidate sequence against a known structure, with a scoring table based on distance and residue-type dependent pair potentials. Candidate sequences were threaded separately onto the intact cyclin fold, and the two cyclin-box domains that constitute the cyclin fold (results not shown). To achieve a measure of significance of the scores, the candidate sequences were also threaded against a data base of 254 different folds provided with the program. Default parameters were used except for a 'd-value' of 200, and an 'x-value' of 1. These two parameters govern the depth of search used in sequence alignment, and the gap-extension penalty, respectively. A d-value of 200 increases the program's chances of establishing the correct alignment between candidate sequence and the trial structure, while an x-value of 1 discourages long insertions or deletions in the sequence–structure alignment. Candidate sequences threaded included the p25 construct, which is a fragment of p35 (residues 99–307), domain A of human p107 (residues 252–451), domain B of human p107 (residues 648–816), the C-terminal portion of human p107 (residues 817–935), intact human transcription factor IIB (TFIIB, residues 1–316), and the C-terminal portion of human uracil deglycosylase starting just beyond the previously identified cyclin box motif of this protein (residues 193–326). As an additional control, the sequence of a truncated form of the γ-subunit of phosphorylase kinase, which is known not to have a cyclin fold [75], was threaded against the database of known folds and the cyclin fold. Results of these analyses are summarised in Table 1.

Table 2. Statistics of the data sets used and of the refined structure.

	Native (λ=1.488 Å)	EMP (λ=0.99 Å)	EMTS (λ=0.99 Å)	EMTS-2 (λ=1.488 Å)
Resolution (Å)	2.0	2.4	2.4	2.5
Completeness (%)	91.4*	89.1	80.3	96.6
R_{merge} †(%)	6.8	14.5	6.2	6.3
R_{iso} ‡(%)		25.7	19.1	26.2
Phasing power§		1.13	1.46	1.47
R_{cullis} #(%)		78.0	64.0	64.0
Number of sites		4	1	2
Refined structure				
R_{conv} **(%)	21.6			
R_{free} ††(%)	29.9			
G_{value} ‡‡	0.41			
Rmsd from ideal values				
Bonds (Å)	0.01			
Angles (°)	1.41			
Number of atoms	2204 (2030 protein atoms constituting residues 11–262, and 174 water molecules)			

*Redundancy of the native data set is 4.1.

†$R_{merge}=\sum_h\sum_j|I_{h,j}-I_h|/\sum_h\sum_j|I_{h,j}|$

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

‡$R_{iso}=\sum_h||Fph_h|-|Fp_h||/\sum_h|Fp_h|$

where Fph_h and Fp_h are the measured structure factor amplitudes for reflection h in the derivative and native data sets, respectively.

§Phasing power=$|Fh_{calc}|/E$

where Fh_{calc} is the calculated structure factor amplitude of reflection h due to scattering by the heavy atoms, and E is the residual lack of closure error.

#$R_{cullis}=\sum_h||Fph_h\pm Fp_h|-|Fhcalc_h||/\sum_h|Fph_h\pm Fp_h|$

where Fph_h and Fp_h are the measured structure factors for centric reflection h in the derivative and native data sets, respectively, and $Fhcalc_h$ is the calculated structure factor amplitude of reflection h due to scattering by the heavy atoms.

**$R_{conv}=\sum_h||Fobs_h|-|Fcalc_h||/\sum_h|Fobs_h|$,

where $Fobs_h$ and $Fcalc_h$ are the observed and calculated structure factor amplitudes, respectively, for reflection h.

††R_{free} is equivalent to R_{conv} for a five percent subset of reflections not used in the refinement. ‡‡G_{value} is the overall measure of structure quality from the PROCHECK package [74]. Higher numbers are indicative of a better structure. The mean G value for protein structures determined at 2.0 Å resolution is approximately –0.4.

The coordinates have been submitted to the Brookhaven Protein Data Bank.

Acknowledgements: The authors would like to thank R Bryan and R Esnouf for computing facilities, I Taylor for technical assistance, A Willis (Oxford) and D Pappin (ICRF) for amino-acid sequencing and analysis, G Barton for his assistance in running STAMP, J Adamczewski for characterisation of the proteolysis patterns of cyclin A, J Gannon for histone-kinase assays, and S Lee for figures. The staff at Daresbury SRS (beamlines 9.6 and 7.2) and at the ESRF provided excellent facilities during data collection. P Nurse and members of his laboratory are thanked for their support

throughout the work on this cell-cycle project. We also thank N Pavletich, P Jeffrey and A Russo for exchange of CDK2–cyclin A coordinates. This work is supported by the Medical Research Council (UK), the Zeneca/DTI/BBSRC Link grant to the Oxford Centre for Molecular Sciences and the Imperial Cancer Research Fund.

Note added in proof

The crystal structure determination of the transcription factor TFIIB in complex with the TATA box binding protein and DNA (Nikolov, D.B. *et al.*, *Nature* **377**:119–128), and the NMR determination of free TFIIB (Bagby, S. *et al.*, *Cell* **82**:857–867) have shown that the transcription factor TFIIB does contain a repeat cyclin-box fold in agreement with the result presented here.

References

1. Norbury, C. & Nurse, P. (1992). Animal cell cycles and their control. *Ann. Rev. Biochem.* **61**, 441–470.
2. Hunter, T. & Pines, J. (1994). Cyclins and cancer II: Cyclin D and CDK inhibitors come of age. *Cell* **79**, 573–582.
3. Meyerson, M., *et al.*, & Tsai, L.-H. (1992). A family of human cdc2-related protein kinases. *EMBO J.* **11**, 2909–2917.
4. Evans, T., Rosenthal, E.T., Youngblom, J., Distel, D. & Hunt, T. (1983). Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* **33**, 389–396.
5. Murray, A.W. & Kirschner, M. (1989). Cyclin synthesis drives the early embryonic cell cycle. *Nature* **339**, 275–280.
6. Pines, J. & Hunter, T. (1989). Isolation of a human cyclin cDNA: Evidence for cyclin mRNA and protein regulation in the cell cycle and for interaction with $p34^{cdc2}$. *Cell* **58**, 833–846.
7. Minshull, J., *et al.*, & Hunt, T. (1989). The role of cyclin synthesis, modification and destruction in the control of cell division. *J. Cell Sci.* **12**, 77–97.
8. Girard, F., Strausfeld, U., Fernadez, A. & Lamb, N.J.C. (1991). Cyclin A is required for the onset of DNA replication in mammalian fibroblasts. *Cell* **67**, 1169–1179.
9. Pagano, M., Pepperkok, R., Verde, F., Ansorge, W. & Draetta, G. (1992). Cyclin A is required at two points in the human cell cycle. *EMBO J.* **11**, 961–971.
10. Zindy, F., *et al.*, & Bréchot, C. (1992). Cyclin A is required in S phase in normal epithelial cells. *Biochem. Biophys. Res. Commun.* **182**, 1144–1154.
11. Walker, D.H. & Maller, J.L. (1991). Role for cyclin A in the dependence of mitosis on completion of DNA replication. *Nature* **354**, 314–317.
12. Lehner, C.F. & O'Farrell, P.H. (1989). Expression and function of Drosophila cyclin A during embryonic cell cycle progression. *Cell* **56**, 957–968.
13. Minshull, J., Golsteyn, R., Hill, C.S. & Hunt, T. (1990). The A- and B-type cyclin associated cdc2 kinases in Xenopus turn on and off at different times in the cell cycle. *EMBO J.* **9**, 2865–2875.
14. Pines, J. & Hunter, T. (1990). Human cyclin A is adenovirus E1A-associated protein p60 and behaves differently from cyclin B. *Nature* **346**, 760–763.
15. Rosenblatt, J., Gu, Y. & Morgan, D.O. (1992). Human cyclin-dependent kinase 2 is activated during the S and G2 phases of the cell cycle and associates with cyclin A. *Proc. Natl. Acad. Sci. USA* **89**, 2824–2828.
16. Murray, A. (1995). Cyclin ubiquitination: The destructive end of mitosis. *Cell* **81**, 149–152.
17. Desai, D., Wessling, H.C., Fisher, R.P. & Morgan, D.O. (1995). Effects of phosphorylation by CAK on cyclin binding by CDC2 and CDK2. *Mol. Cell. Biol.* **15**, 345–350.
18. Fesquet, D., *et al.*, & Cavadore, J.-C. (1993). The *MO15* gene encodes the catalytic subunit of a protein kinase that activates cdc2 and other cyclin-dependent kinases (CDKs) through phosphorylation of Thr161 and its homologues. *EMBO J.* **12**, 3111–3121.
19. Poon, R.Y.C., Yamashita, K., Adamczewski, J.P., Hunt, T. & Shuttleworth, J. (1993). The cdc2-related protein $p40^{MO15}$ is the catalytic subunit of a protein kinase that can activate $p33^{cdk2}$ and $p34^{cdc2}$. *EMBO J.* **12**, 3123–3132.
20. Solomon, M.J., Wade-Harper, J. & Shuttleworth, J. (1993). CAK, the $p34^{cdc2}$ activating kinase, contains a protein identical or closely related to $p40^{MO15}$. *EMBO J.* **12**, 3133–3142.
21. Mäkelä, T.P., Tassan, J.-P., Nigg, E.A., Frutiger, S., Hughes, G.J. & Weinberg, R.A. (1994). A cyclin associated with the CDK-activating kinase MO15. *Nature* **371**, 254–257.
22. Shirodkar, S., Ewen, N., DeCaprio, J.A., Morgan, J., Livingston, D.M. & Chittenden, T. (1992). The transcription factor E2F interacts with the retinoblastoma product and a p107-cyclin A complex in a cell cycle-regulated manner. *Cell* **68**, 157–166.
23. Devoto, S.H., Mudryj, M., Pines, J., Hunter, T. & Nevins, J.R. (1992). A cyclin A-protein kinase complex possesses sequence-specific DNA binding activity: $p33^{cdk2}$ is a component of the E2F-cyclin A complex. *Cell* **68**, 167–176.
24. Smith, E.J. & Nevins, J.R. (1995). The RB-related p107 protein can suppress E2F function independently of binding to cyclin A/CDK2. *Mol. Cell. Biol.* **15**, 338–344.
25. Zhu, L., Enders, G., Lees, J.A., Beijersbergen, R.L., Bernards, R. & Harlow, E. (1995). The pRB-related protein p107 contains two growth suppression domains: Independent interactions with E2F and cyclin/cdk complexes. *EMBO J.* **14**, 1904–1913.
26. Nevins, J. R. (1992). E2F: A link between the Rb tumor suppressor protein and viral oncoproteins. *Science* **258**, 424–429.
27. Krek, W., *et al.*, & Livingston, D.M. (1994). Negative regulation of the growth-promoting transcription factor E2F-1 by a stably bound cyclin A-dependent protein kinase. *Cell* **78**, 161–172.
28. Faha, B., Ewen, M.E., Tsai, L.-H., Livingston, D.M. & Harlow, E. (1992). Interaction between cyclin A and Adenovirus E1A-associated p107 protein. *Science* **255**, 87–90.
29. Ewen, M.E., Faha, B., Harlow, E. & Livingston, D.M. (1992). Interaction of p107 with cyclin A independent of complex formation with viral oncoproteins. *Science* **255**, 85–87.
30. Hollingsworth, R.E. Jr, Chen, P.-L. & Lee, W.-H. (1993). Integration of cell cycle control with transcriptional regulation by the retinoblastoma protein. *Curr. Op. Cell Biol.* **5**, 194–200.
31. Hinds, P.W., Mittnacht, S., Dulic, V., Arnold, A., Reed, S.I. & Weinberg, R.A. (1992). Regulation of the retinoblastoma protein functions by ectopic expression of human cyclins. *Cell* **70**, 993–1006.
32. Whitfield, W.G.F., Gonzalez, C., Maldonado-Codina, G. & Glover, D.M. (1990). The A-and B-type cyclins of Drosophila are accumulated and destroyed in temporally distinct events that define separable phases of the G2-M transition. *EMBO J.* **9**, 2563–2572.
33. Knoblich, J.A. & Lehner, C.F. (1993). Synergistic action of Drosophila cyclins A and B during the G2-M transition. *EMBO J.* **12**, 65–74.
34. Murray, A.W., Solomon, M.J. & Kirschner, M.W. (1989). The role of cyclin synthesis and degradation in the control of maturation promoting factor activity. *Nature* **339**, 280–286.
35. Peeper, D.S., *et al.*, & Piwnica-Worms, H. (1993). A- and B-type cyclins differentially modulate substrate specificity of cyclin-cdk complexes. *EMBO J.* **12**, 1947–1954.
36. Pines, J. & Hunter, T. (1994). The differential localization of human cyclins A and B is due to a cytoplasmic retention signal in cyclin B. *EMBO J.* **13**, 3772–3781.
37. Jeffrey, P. D., *et al.*, & Pavletich, N. (1995). Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376**, 313–320.
38. Kobayashi, H., Stewart, E., Poon, R., Adamczewski, J.P., Gannon, J. & Hunt, T. (1992). Identification of the domains in cyclin-A required for binding to, and activation of, $p34^{cdc2}$ and $p32^{cdk2}$ protein kinase subunits. *Mol. Biol. Cell.* **3**, 1279–1294.
39. Wang, J., Chenivesse, X., Henglein, B. & Bréchot, C. (1990). Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma. *Nature* **343**, 555–557.
40. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* **22**, 2577-2637.
41. Maridor, G., Gallant, P., Golsteyn, R. & Nigg, E.A. (1993). Nuclear localization of vertebrate cyclin A correlates with its ability to form complexes with cdk catalytic subunits. *J. Cell Sci.* **106**, 535–544.
42. Lees, E.M. & Harlow, E. (1993). Sequences within the conserved cyclin box of human cyclin-A are sufficient for binding to and activation of cdc2 kinase. *Mol. Cell. Biol.* **13**, 1194–1201.
43. Russell, R.B. & Barton, G.J. (1992). Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins* **14**, 309–323.
44. Stewart, E., Kobayashi, H., Harrison, D. & Hunt, T. (1994). Destruction of Xenopus cyclins A and B2, but not B1, requires binding to $p34^{cdc2}$. *EMBO J.* **13**, 584–594.
45. Zheng, X.-F. & Ruderman, J.V. (1993). Functional analysis of the P box, a domain in cyclin B required for the activation of Cdc25. *Cell* **75**, 155–164.
46. Nash, R., Tokiwa, G., Anand, S., Erickson, K. & Futcher, A.B. (1988). The *WHI1+* gene of *Saccharomyces cerevisiae* tethers cell division to cell size and is a cyclin homolog. *EMBO J.* **7**, 4335–4346.

47. Gallant, P. & Nigg, E.A. (1994). Identification of a novel vertebrate cyclin: cyclin B3 shares properties with both A- and B-type cyclins. *EMBO J.* **13**, 595–605.
48. Jones, D.T., Taylor, W.R. & Thornton, J.M. (1992). A new approach to protein fold recognition. *Nature* **358**, 86–89.
49. Gibson, T.J., Thompson, J.D., Blocker, A. & Kouzarides, T. (1994). Evidence for a protein domain superfamily shared by the cyclins, TFIIB, and RB/p107. *Nucl. Acids Res.* **22**, 946–952.
50. Tsai, L.H., Delalle, I., Caviness, J.V.S., Chae, T. & Harlow, E. (1994). p35 is a neural-specific regulatory subunit of cyclin-dependent kinase 5. *Nature* **371**, 419–423.
51. Lew, J., *et al.*, & Wang, J.H. (1994). A brain-specific activator of cyclin-dependent kinase 5. *Nature* **371**, 423–426.
52. Ewen, M.E., Xing, Y.G., Lawrence, J.B. & Livingston, D.M. (1991). Molecular cloning, chromosomal mapping, and expression of the cDNA for p107, a retinoblastoma gene product-related protein. *Cell* **66**, 1155–1164.
53. Ha, I., Lane, W.S. & Reinberg, D. (1991). Cloning of a human gene encoding the general transcription initiation factor IIB. *Nature* **352**, 689–695.
54. Jones, D.T. (1995). Threader-optimal sequence threading program. University College London.
55. Muller, S.J. & Caradonna, S. (1993). Cell cycle regulation of a human cyclin-like gene encoding uracil-DNA glycosylase. *J. Biol. Chem.* **268**, 1310–1319.
56. Orengo, C.A., Jones, D.T. & Thornton, J.M. (1994). Protein superfamilies and domain superfolds. *Nature* **372**, 631–634.
57. Fisher, R.P. & Morgan, D.O. (1994). A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell* **78**, 713–724.
58. Roy, R., *et al.*, & Egly, J.-M. (1994). The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell* **79**, 1093–1101.
59. Feaver, W.J., Svejstrup, J.Q., Henry, N.L. & Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell* **79**, 1103–1109.
60. Serizawa, H., Mäkelä, T.P., Conaway, J.W., Conaway, R.C., Weinberg, R.A. & Young, R.A. (1995). Association of Cdk-activating kinase subunits with transcription factor TFIIH. *Nature* **374**, 280–282.
61. Shiekhattar, R., *et al.*, & Reinberg, D. (1995). CDK-activating kinase complex is a component of human transcription factor TFIIH. *Nature* **374**, 283–287.
62. Tamura, K., *et al.*, & Okayama, H. (1993). Cyclin G: a new mammalian cyclin with homology to fission yeast Cig1. *Oncogene* **8**, 2113–2118.
63. Hu, Q., Dyson, N. & Harlow, E. (1990). The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations. *EMBO J.* **9**, 1147–1155.
64. Hagemeier, C., Bannister, A.J., Cook, A. & Kouzarides, T. (1993). The activation domain of transcription factor PU.1 binds the retinoblastoma (RB) protein and the transcription factor TFIID in vitro: RB shows sequence similarity to TFIID and TFIIB. *Proc. Natl. Acad. Sci. USA* **90**, 1580–1584.
65. Walsh, M.J., Shue, G., Spidoni, K. & Kapoor, A. (1995). E2F-1 and a cyclin-like DNA repair enzyme, uracil-DNA glycosylase, provide evidence for an autoregulatory mechanism for transcription. *J. Biol. Chem.* **270**, 5289–5298.
66. Teng, T.-Y. (1990). Mounting of crystals for Macromolecular Crystallography in a free-standing thin film. *J. Appl. Cryst.* **23**, 387–391.
67. Otwinowski, Z. (1993). Data collection and processing. In Proceedings of the CCP4 study weekend, DL/SCI/R34 (Sawyer, L., Isaacs, N. & Bailey, S. eds), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
68. Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
69. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering. Proceedings of the CCP4 study weekend,* DL/SCI/R32 (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80–85, SERC Daresbury Laboratory, Warrington, UK.
70. Cowtan, K. (1994). "dm": an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on protein crystallography* **31**, 34–38.
71. Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
72. Brünger, A.T. (1992). *X-PLOR Manual. version 3.1* Yale University, New Haven, CT.
73. Lamzin, V.S. & Wilson, K.S. (1993). Automated refinement of protein models. *Acta Cryst. D* **49**, 129–147.
74. Morris, A.L., MacArthur, M.W., Hutchinson, E.G., & Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* **12**, 354–364.
75. Owen, D.J., Noble, M.E.M., Garman, E.F., Papageorgiou, A.C. & Johnson, L.N. (1995). Two structures of the catalytic domain of phosphorylase kinase: an active protein kinase complexed with substrate analogue and product. *Structure* **3**, 467–482.
76. Barton, G.J. (1993). ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40.
77. Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* **9**, 56–68.
78. Lew, D.J., Dulic, V. & Reed, S.I. (1991). Isolation of three novel human cyclins by rescue of G1 cyclin (Cln) function in yeast. *Cell* **66**, 1197–1206.
79. Koff, A., *et al.*, & Roberts, J.M. (1991). Human cyclin E, a new cyclin that interacts with two members of the CDC2 gene family. *Cell* **66**, 1217–1228.
80. Bai, C., Richman, R. & Elledge, S.J. (1994). Human cyclin F. *EMBO J.* **13**, 6087–6098.

Received: **11 Aug 1995**; revisions requested: **30 Aug 1995**; revisions received: **7 Sept 1995**. Accepted: **11 Sept 1995**.