The EMBO Journal Vol.16 No.5 pp.958–967, 1997

The structure of cyclin H: common mode of kinase activation and specific features

Gregers Andersen, Didier Busso, Arnaud Poterszman, Jae Ryoung Hwang, Jean-Marie Wurtz, Raymond Ripp, Jean-Claude Thierry, Jean-Marc Egly and Dino Moras

Institut de Génétique et de Biologie Moléculaire et Cellulaire, 1 rue Laurent Fries, BP 163, 67404 Illkirch Cédex, CU de Strasbourg, France

G.Andersen and D.Busso should be considered as joint first authors

The crystal structure of human cyclin H refined at 2.6 Å resolution is compared with that of cyclin A. The core of the molecule consists of two repeats containing five helices each and forming the canonical cyclin fold also observed in TFIIB. One hundred and thirty-two out of the 217 Cα atoms from the cyclin fold can be superposed with a root-mean-square difference of 1.8 Å. The structural homology is even higher for the residues at the interface with the kinase, which is of functional significance, as shown by our observation that cyclin H binds to cyclin-dependent kinase 2 (cdk2) and that cyclin A is able to activate cdk7 in the presence of MAT1. Based on this superposition, a new signature sequence for cyclins was found. The specificity of the cyclin H molecule is provided mainly by two long helices which extend the cyclin fold at its N- and C-termini and pack together against the first repeat on the side opposite to the kinase. Deletion mutants show that the terminal helices are required for a functionally active cyclin H.

Keywords: CAK/cdk/crystal structure/cyclin H/TFIIH

Introduction

Cyclin-dependent kinases (cdks) are members of the Ser/Thr protein kinase family. They were first identified as components that regulate the passage through the cell cycle. The activity of these enzymes is controlled by transient association with their cyclin regulatory subunits, binding of inhibitory proteins and/or phosphorylation and dephosphorylation reactions (Morgan, 1995; Nigg, 1996). Such transient associations lead to the phosphorylation of specific protein substrates involved in the overall cdk cascade, DNA replication or transcription. Thus, different cyclin–kinase pairs such as cyclin A–cdk2, cyclin B–cdk1, cyclin D–cdk6, cyclin D–cdk4 or cyclin E–cdk2 have been isolated from various cell extracts. This association is followed by phosphorylation of the kinase at the threonine located between amino acids 160 and 170, by a distinct kinase activity called cyclin-activating kinase (CAK) (Fisher and Morgan, 1994). The crystal structures of cyclin A alone, complexed with cdk2 and in complex with both cdk2 and p27, an inhibitory molecule, have been determined (Brown *et al.*, 1995; Jeffrey *et al.*, 1995; Russo *et al.*, 1996a,b). These results show that cyclin A keeps the same conformation in its free state as it does in complexes with various partners, while the kinase adapts to fit the molecular interface.

The CAK kinase (named cdk7 in human, MO15 in frog oocytes, Kin 28 in *Saccharomyces cerevisiae*; see Figure 1B) recently has been purified and has interesting properties. It is found associated with cyclin H and a third subunit MAT1. This ternary complex is found either in a free state (for a review, see Morgan, 1995), or as part of a multisubunit complex TFIIH (for a review, see Svejstrup *et al.*, 1996). First identified as a basal transcription factor absolutely required for transcription of protein-coding genes, TFIIH was also shown to play a major role in DNA repair (Schaeffer *et al.*, 1993). Mutations in some of the TFIIH subunits give rise to various genetic diseases such as xeroderma pigmentosum, Cockayne syndrome and trichotiodystrophy (for a review, see Cleaver and Kraemer, 1996).

Three of the nine TFIIH subunits have been shown to possess enzymatic activities. One of them is the kinase cdk7 whereas the two others are XPB and XPD, two ATP-dependent helicases with opposite polarity. Only cdk7 was shown to be regulated by cyclin H, whose structure and function is the subject of the present study.

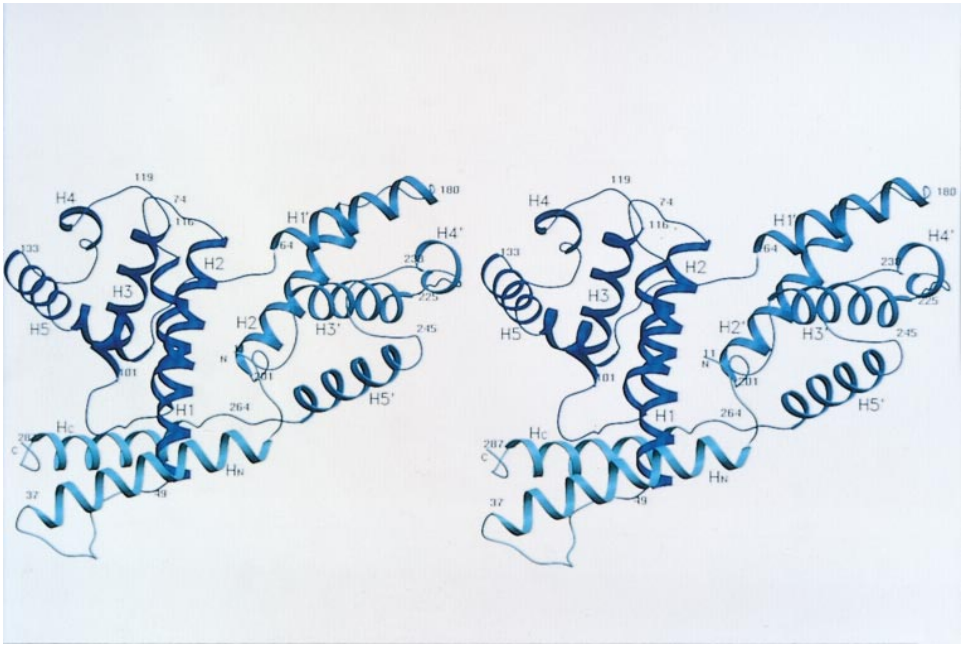
Human cyclin H is a polypeptide of 323 amino acids, with a molecular weight of 38 kDa. When cyclin H is associated with cdk7, the kinase is able to phosphorylate either cdk2 or the C-terminal domain of the largest subunit of RNA polymerase II (ctd), in addition to TATA box-binding protein (TBP) or TFIIE; the activity is enhanced by the third subunit, MAT1, of the kinase complex which has some affinity for cdk7. The levels of cyclin H and MAT1 subunits, as well as the resulting kinase activity, do not vary over the eukaryotic cell cycle (Fisher and Morgan, 1994; Fisher *et al.*, 1995; Adamczewski *et al.*, 1996).

In order to understand the structure–function of TFIIH, a key multisubunit complex involved in several mechanisms of cell life, in transcription (Drapkin *et al.*, 1994; Feaver *et al.*, 1994; Roy *et al.*, 1994) as well as in cell cycle regulation (for review, see Morgan, 1995; Nigg, 1995), we have undertaken the structural study of the various subunits. We present here our results on human cyclin H, the regulatory subunit of the CAK complex.

Results

Overall description

As described briefly in Andersen *et al.* (1996), human cyclin H is an elongated molecule with two characteristic α-helical domains (each containing five helices), referred



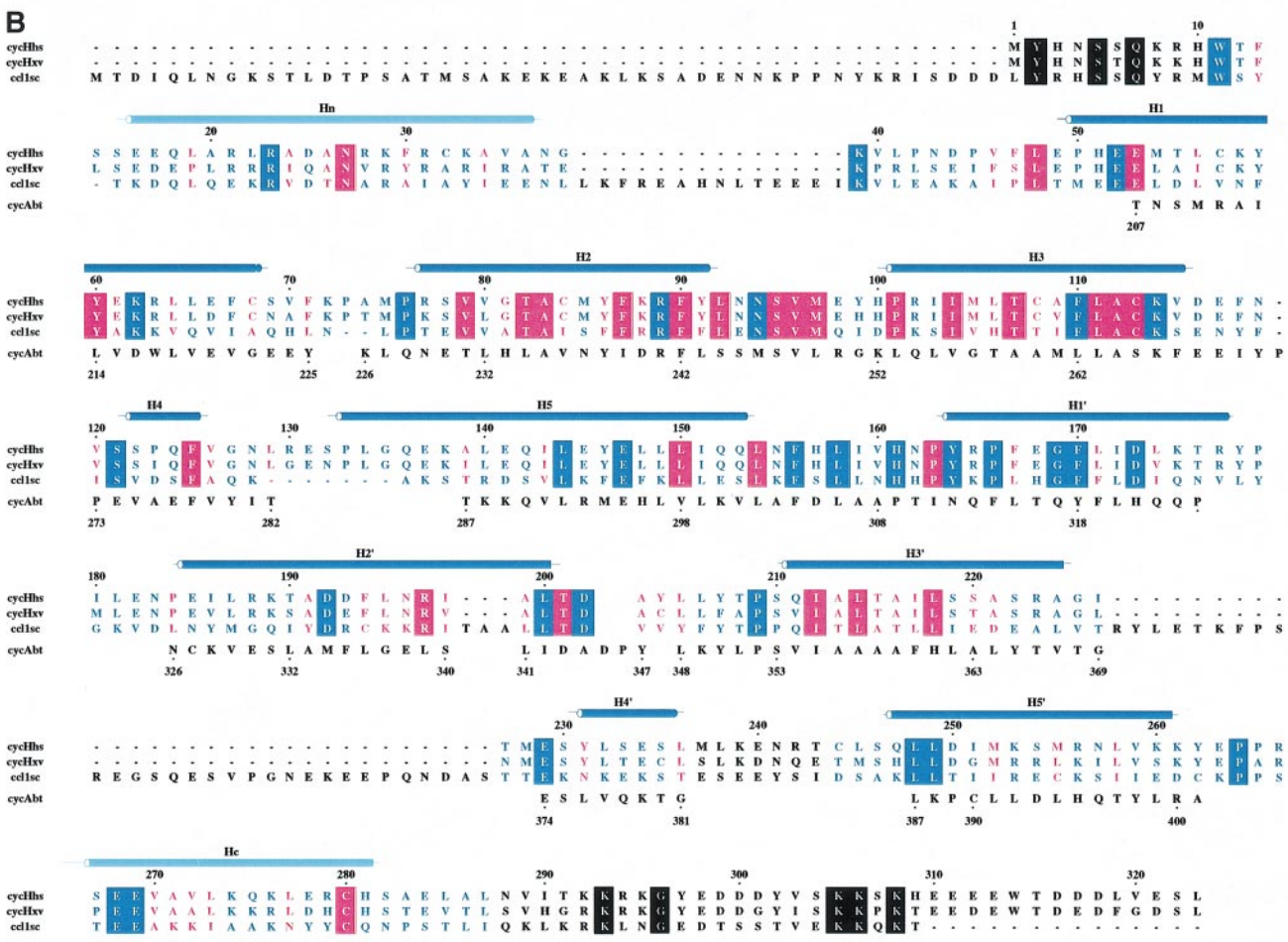
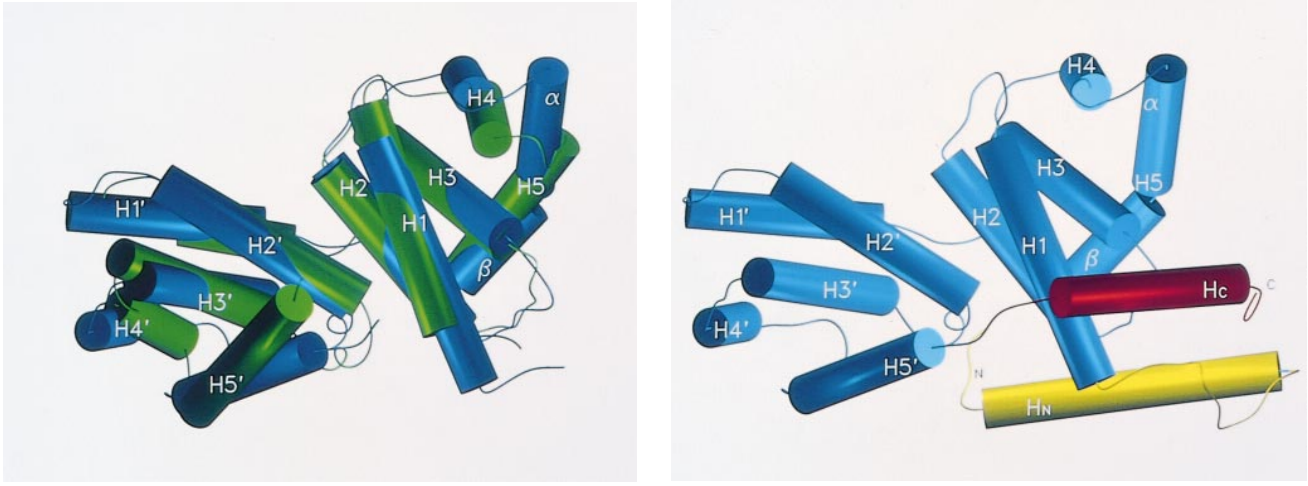


Fig. 1. (**A**) Stereoview showing the main chain trace of cyclin H (residues 11–287) drawn with SETOR (Evans, 1993). Repeat 1 (H1–H5) is drawn in magenta, Repeat 2 (H1′–H5′) in blue. Helices H_N and H_C, which form a layer at the end of the first helical repeat, are coloured in cyan. (**B**) Sequence alignment of human (cycHhs), *Xenopus laevis* (cycHxv) and yeast (ccl1sc) cyclin H. Strictly conserved residues are shaded. H1–H5 are the helices from the first repeat of the cyclin box and H1′–H5′ are helices from the second repeat. H_N and H_C designate the N- and C-terminal helices, respectively. Buried residues (accessibility <10 Å^2) are drawn in magenta, other residues in blue. Residues which could not be traced are in black. A structural alignment of bovine cyclin A (cycAbt) is indicated below the cyclin H sequences for residues where such a comparison is adequate.



long helices located at the N- and C-termini of the molecule (helices H_N and H_C respectively) and both helices pack against Repeat 1.

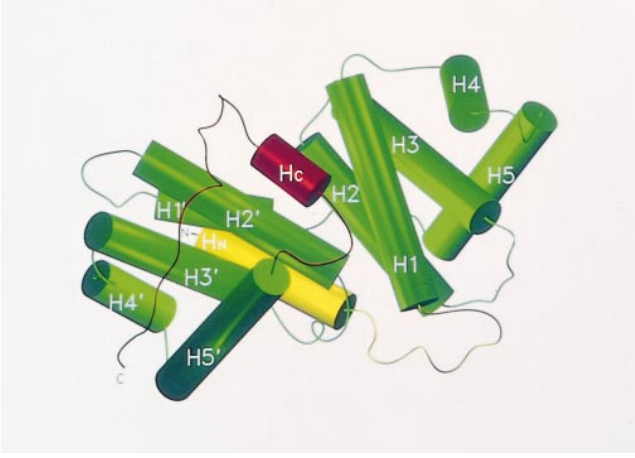


Fig. 2. (**A**) Superposition of cyclin H (blue) and cyclin A (green). The two molecules were first superposed using H1, H2 and H3. The superposition was then extended to residues 48–265 in cyclin H and residues 200–405 in cyclin A. (B and C) Cyclin H (**B**) and cyclin A (**C**) alone are represented in the same orientation, with helices HN and HC in yellow and red, respectively.

to as Repeat 1 and Repeat 2, which adopt the so-called cyclin-fold topology (Figure 1A) previously observed in cyclin A and TFIIB (Jeffrey *et al.*, 1995; Nikolov *et al.*, 1995). Repeat 1 corresponds to the cyclin box, initially identified by sequence analysis, and consists of five α-helices, numbered H1–H5, highly conserved among the various species (Figure 1A and B). The core helix H3 is flanked on one side by helices H1 and H2, which form an interface with the analogous helices of Repeat 2, and by H4 and H5 on the other. Repeat 2 can be derived from Repeat 1 by a rotation of 160° around the pseudo 2-fold axis of the molecule. The two repeats are connected by a linker peptide stabilized via water-mediated interactions. A third domain, characteristic of cyclin H, consists of two

Orientation of the two repeats in distantly related cyclins

To analyse further the spatial organization of cyclin H, we compared the mutual orientation of the two repeats in cyclin H and cyclin A (Figure 2A). The two cyclin folds (217 and 205 Cα atoms in cyclin H and A respectively) were matched, such that 132 Cα atoms from helices H1, H2, H3 and H5 (Repeat 1), from the linker peptide between H5 and H1′ and from helices H1′, H2′ and H3′ (Repeat 2) superpose with a root-mean-square difference (r.m.s.d.) of 1.8 Å, a value characteristic of the similarity between the two structures (Flores *et al.*, 1993). Knowing that cyclins A and H are representative members of two phylogenetically distant cyclin types, both the cyclin fold and the relative positioning of the two repeats are likely to be conserved in all cyclins. This result could not be predicted by sequence analysis alone, because cyclins exhibit a strong homology only in Repeat 1 but differ markedly in the rest of their sequences (Figure 3). Note that in TFIIB, a general transcription factor with a cyclin fold, the two repeats are oriented differently (Nikolov *et al.*, 1995).

In both cyclin H and cyclin A the interface between Repeat 1 and Repeat 2 harbours a buried surface of a similar size (428 $Å^2$ and 395 $Å^2$ respectively). The interactions are strengthened by the N- and C-terminal ends which contribute equally (549 $Å^2$ and 422 $Å^2$ respectively). The area of the resulting interface is comparable with that observed between subunits of similar size within oligomeric proteins (Miller *et al.*, 1987). The relative positioning of the two repeats is the result of a dense network of both hydrophobic interactions and salt bridges involving water molecules, which restrict the movements of the two lobes. The contact between the two repeats involves H1, H2, H1′, H2′ and the linker peptide (Figure 1A), highly conserved in size among the various cyclins and the various species. In cyclin H, a cluster of

A

```
H1
pH-C              l  ϕ    l  ls  l
cycH       53 EMTLCKYYEKRLLEFCSVF      71
motif                    l  l
cycA      207 TNSMRAILVDWLVEVGEEY     225
pA-BDE           MR llh Wl pV

H2
pH-C            h   slsTA  ϕl RFlh
cycH       74 AMPRSVVGTACMYFKRFYLNNSVM  97
motif           h   s   s  hl  l
cycA      226 KLQNETLHLAVNYIDRFLSSMSVL 249
pA-BDE          ph  -sl hs  llDRϕhp

H3
pH-C          ph  lh TshlLAsKs-p
cycH      100 HPRIIMLTCAFLACKVDEF     118
motif             sshlhss+  ##
cycA      252 KLQLVGTAAMLLASKFEEI     270
pA-BDE          lQLlG sslllAsKlp-
                        *  **  *

H5,1,H1'
pH-C          ppl phEl  lp hp  Ll  pPYpslll    h p
cycH   139 ALEQILEYELLLIQQLNFHLIVHNPYRPFEGFLIDLKT  176
motif                  hh  h
cycA   287 TKKQVLRMEHLVLKVLAFDLAAPTINQFLTQYFLHQQP  324
pA-BDE        pl   E  ll  l h l       ϕh  h
             **    *   *   *****

H2'
pH-C                ph pp
cycH    184 PEILRKTADDFLNRIALT    201

cycA    326 NCKVESLAMFLGELSLID    343
pA-BDE                  l   s  -

H3'
pH-C          P  Islsl
cycH    209 PSQIALTAILSSASRA     224

cycA    352 PSVIAGAAFHLALYTV     367
pA-BDE        s  h ss        l
```

B

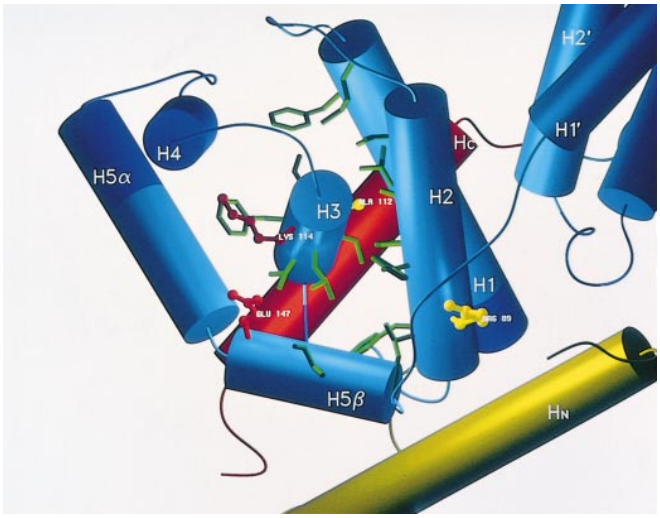


Fig. 3. (**A**) Structural alignment of cyclin H (cycH) and cyclin A (cycA) based on an overall superposition of the two molecules (132 Cα atoms, r.m.s.d. = 1.8 Å). The sequence of cyclin H (cycH) together with a pattern of conserved (upper case letters) or semi-conserved residues [ϕ (F,Y,W), h (A,C,V,I,L,M,F,Y,W), l (V,I,L,M,F,YW), s (G,A,S,C,T,V), + (K,R), – (E,D), p (S,T,D,E,H,K,R)] derived from an alignment of representative members of cyclin H and cyclin C is presented for the superposable regions (pH-C). The same procedure was applied for the A-BDE subgroup (cycA, pA-BDE). Residues from the linker region between H1 and H5′ are in italics. Residues which interact with the kinase are indicated by an asterisk (*). Glu117 and Phe118 from cyclin H and the corresponding Glu269 and Ile270 from cyclin A do not superpose, but they are important for kinase recognition (Jeffrey *et al.*, 1995) and are labelled with (#). Between the cyclin H and A sequences, the consensus motif of conserved residues deduced from an alignment of 100 cyclin sequences is represented. (**B**) Location of conserved or semi-conserved residues in cyclins shown on the 3D structure of cyclin H. Side chains of the functional Lys114 and Glu147 are in red, side chains of the structural Arg89 and Ala112 in yellow. The remaining side chains, drawn in green, belong to the motif extracted from the global sequence alignment.

charged residues is formed by Arg77 (H2), Glu168 (H1′), Arg188 (H1′) and Asp192 (H2′). A similar cluster is also observed in cyclin A; it involves Asn229 (H2), Glu338 (H2′) and Lys412 (H_C). A highly conserved amino acid, Arg89 (Arg241 in cyclin A), plays a key role in the stabilization of the relative orientation of the two repeats (Figure 3A). It anchors H2 of the first repeat to the linker peptide by interacting with the main chain carbonyl atom of His157 (Figure 3B). This interaction is likely to be conserved within the cyclin family (see also below).

H_N and H_C are required for a functionally active cyclin H

The main differences between the two cyclins lie outside the cyclin box. The respective orientations of H_N and H_C are different from those observed in cyclin A (compare Figure 2B and C). In the latter, H_N and H_C are an integral part of the interface between the two repeats whereas in cyclin H only the N-terminal end of H_N and the connecting loop H5′–H_C are involved in an equivalent contact (Figure 2). Amino acids located between the C-terminus of H_N and H1 as well as the N-terminus of H_C clip H1, while packing onto Repeat 1. The compact structure of cyclin H can be correlated to mild digestion experiments which show a proteolytically stable fragment of 31 kDa. This fragment starts at residue 12 according to microsequence analysis (data not shown). Its molecular weight indicates that a segment of ~35 residues was removed from the C-terminal end of the molecule. Note that these residues are poorly defined in the electron density map, indicating a probable flexibility in this region in the absence of a partner molecule (Figure 1B).

In order to analyse precisely the role of the N- and C-terminal regions of cyclin H, we generated truncated proteins containing evolutionarily conserved N- or C-terminal motifs and tested them for cdk7 binding and ctd kinase activity (Figure 4A). A soluble extract of Sf9 cells overexpressing an active cdk7 (Fisher and Morgan, 1994; Adamczewski *et al.*, 1996) was incubated with an excess of each truncated cyclin H recombinant over-expressed in *Escherichia coli*. The incubation mixture was immunoprecipitated with antibodies against cdk7 (Ab-cdk7), previously cross-linked to protein G–Sepharose. Ab-cdk7 monoclonal antibodies were directed against an epitope located at the C-terminal end of cdk7, which is on the side opposite to the cyclin-binding site. After extensive washing of the immunoadsorbent with 100 mM KCl, the remaining proteins were analysed by Western blotting (Figure 4B) and assayed for ctd kinase activity using ctd4, a synthetic peptide that mimics the C-terminal domain of the largest subunit of RNA polymerase II, as substrate (Figure 4C).

The deletion mutant CTΔ290, where the C-terminus was removed (Figure 4A), binds and activates cdk7 (Figure 4B, lane 9, and C, lane 5). When the deletion encompasses HC, the truncated cyclin H (CTΔ257) loses its regulatory function although it still binds to cdk7 (Figure 4C, lane 6, and B, lane 11). The same phenomenon is observed when H_N (NTΔ35) is deleted. In this case, although the mutant binds cdk7, the resulting complex is not able to

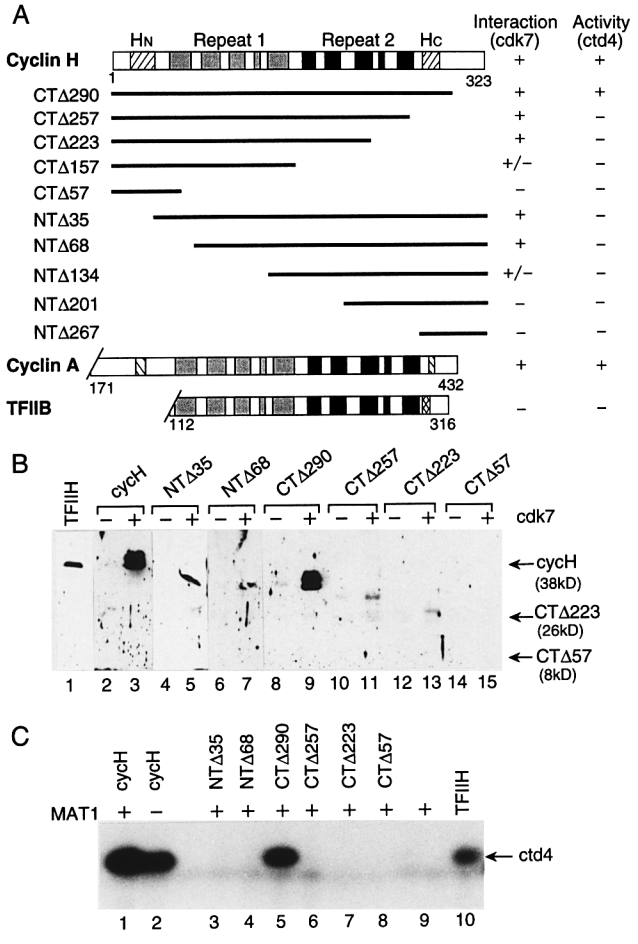


Fig. 4. Activity of cyclin H deletion mutants. (**A**) Schematic drawing of C- and N-terminal deletion mutants of cyclin H, cyclin A and TFIIB, expressed in and purified from *E.coli*, which were employed for interaction and in cdk7 kinase activity assay. The positions of Repeat 1 and Repeat 2 containing five helices each (light and dark grey shaded respectively) as well as HN and HC representing the N- and C-terminal helices located outside the cyclin fold are indicated. (+) or (–) indicates the ability of the deletion mutants to interact with cdk7 and to phosphorylate ctd4 substrate in the presence of cdk7 and MAT1. (**B**) Wild-type and deletion mutants of cyclin H were pre-incubated in the absence (–) or in the presence (+) of cdk7, before being immunoprecipitated with Ab-cdk7 antibody. Immunoprecipitated fractions were analysed by Western blotting using polyclonal antibodies directed against cyclin H (lanes 2–15) and for the ctd kinase activity (C) in the presence or absence of MAT1 as indicated at the top of the panel. In (B), 5 μl of TFIIH is used as control (lane 1); MAT1 is present in each fraction blotted (lanes 2–15). The molecular weight of cyclin H mutants is indicated at the right of the panel. (**C**) The immunoprecipitated fractions were tested in a kinase assay using ctd4 as a substrate (lanes 1–10).

phosphorylate the ctd4 substrate. As expected, further deletions of cyclin H abolish the cdk7-binding property (see Figure 4B). These results demonstrate the absolute requirement for the N- and C-terminal regions of human cyclin H for the ctd kinase activity.

Sequence conservation within the cyclin family

Cyclin H and cyclin A are representative members of the cyclin family which contains two evolutionary divergent supergoups (Bazan, 1996). To analyse the sequence and structure–function relationship, we focused initially on cyclin types H-C and A-BDE, as the structures of cyclin H and cyclin A have been determined. Using the three-dimensional (3D) superposition of cyclins H and A, sequences from representative members of the different cyclin types were aligned. The analysis was restricted to regions of conserved 3D structures. Thus, only residues from helices H1, H2, H3 and H5, from the linker peptide (between H5 and H1′) and from helices H1′, H2′ and H3′ were taken into consideration. The alignment was then extended to a representative set of 100 cyclin sequences for residues encompassing Repeat 1 (see Materials and methods and Figure 3).

Cyclins possess a pseudo 2-fold axis which relates the two helical repeats (the first and the second repeat of cyclin H can be superposed over 56 Cα atoms with an r.m.s.d. of 2.2 Å). The degree of amino acid conservation is significantly higher in Repeat 1 than in Repeat 2 (Figure 3A). Only 19 out of the 66 conserved residues which form the pattern common to H-C cyclins (pH-C Figure 3A) are found in Repeat 2. When the second repeat of cyclins H and A are superposed, helices H1′, H2′ and H3′ match quite well, but not helices H4′ and H5′ (Figure 2A). Due to the topological position of the C-terminal residues, helices H2′ and H3′ are in a different environment. In cyclin A, H2′ and H3′ are partially buried by C-terminal residues located between H_C and the protein C-terminus (Figure 2C). On the contrary, in cyclin H, as C-terminal residues pack against the first repeat, the corresponding residues are exposed to the solvent (Figure 2B). Functionally, the two repeats are not equivalent (see below).

Most of the conserved residues play a structural role. Two of them, Arg89 and Ala112 of cyclin H, are invariant among H-C and A-BDE cyclins (Figure 3A and B). Ala112 in the central helix H3 makes hydrophobic interactions with conserved residues from surrounding helices (Figure 3). Two other residues, Lys114 and Glu147 in cyclin H, are exposed and are invariant among H-C and A-BDE cyclins (Figure 3A and B). In cyclin A, they form a salt bridge and are involved in kinase recognition [see the analogous Lys266 and Glu295 (Jeffrey *et al.*, 1995)]. In cyclin F, the lysine is replaced by an arginine and the glutamic acid by a conserved methionine. The positive charge at position 114 of cyclin H is therefore conserved within the cyclin family.

Using the pattern of conserved residues (Figure 3A, motif), a database search (translated EMBL database V47.0) retrieves 275 sequences, of which 238 are cyclins or annotated as cyclins. Due to the degeneracy of the motif, the remaining 37 sequences found are either suspected false positives or may contain a pseudo cyclin fold (see Materials and methods). Interestingly, among the 37 sequences, 10 TFIIB members, well known to share a common fold with cyclins (Nikolov *et al.*, 1995), are detected. Indeed, in TFIIB, as also observed in cyclin F, the conserved Lys114 is replaced by an arginine and Glu147 by a hydrophobic residue. Note also that the positively charged Lys114 in human cyclin H as well as Arg169 in TFIIB, located in H3, are involved in protein–protein interactions in their respective cyclin–cdk and TFIIB–TBP complexes. The glutamic acid or the hydrophobic residue at position 147 might be important in stabilizing the structure either by hydrophobic contacts or through a salt bridge.

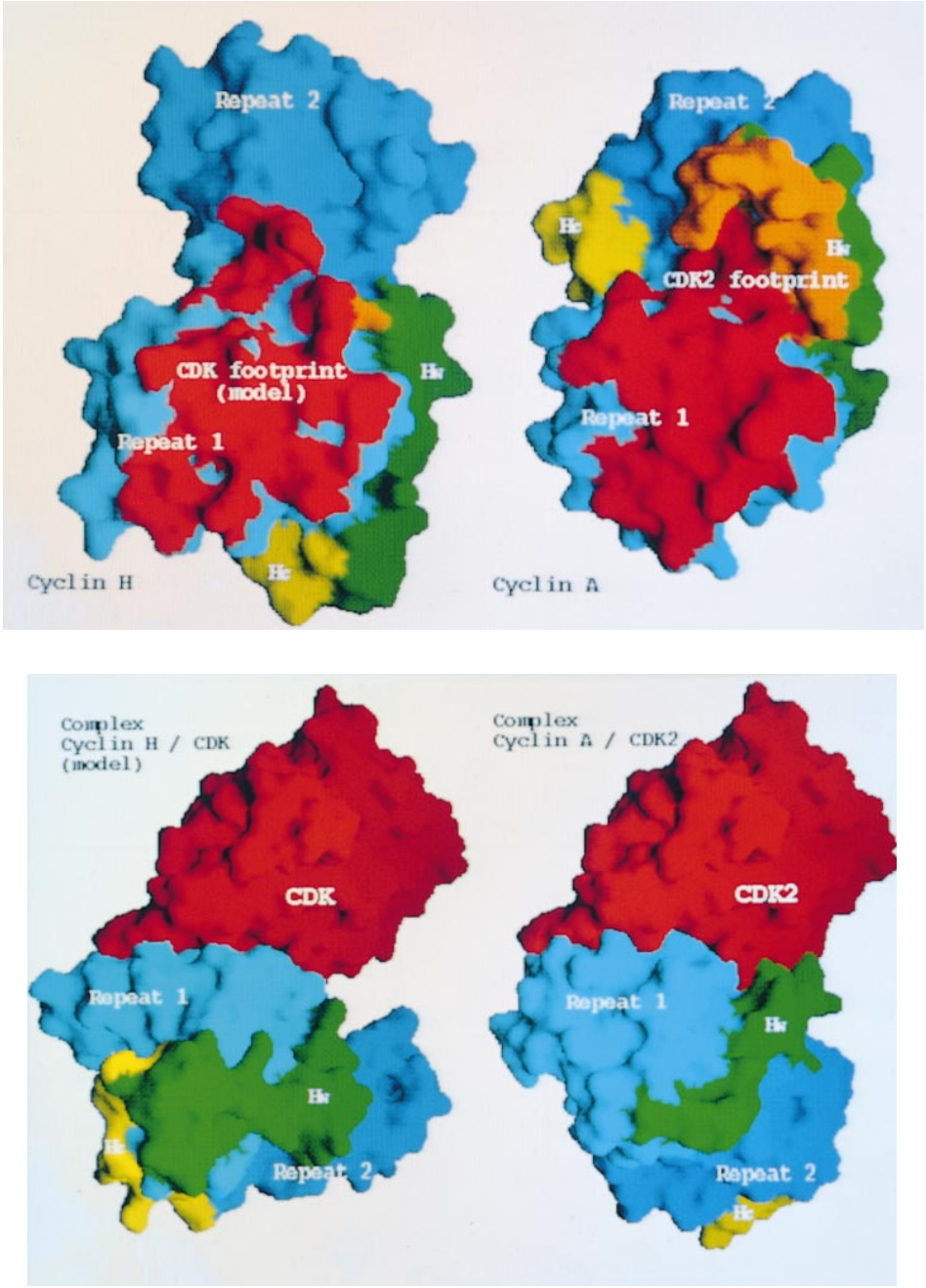


Fig. 5. Upper part: molecular surface of human cyclin H (left) and cyclin A (right) drawn in the same orientation. The first and the second repeat are coloured in light and dark blue, respectively. The N- and C-terminal extensions from both cyclins are coloured in green and yellow, respectively. Surface buried upon complex formation with the kinase is shown in red (Repeats 1 and 2) and orange (HN). Lower part: molecular surfaces of human cyclins H and A complexed with cdks (in red).

Cdk recognition

The strong structural conservation of the core domain between cyclin H (this work) and cyclin A, either alone or complexed with cdk2 (Brown *et al.*, 1995; Jeffrey *et al.*, 1995) suggests that cyclins behave as rigid templates for kinase recognition. As cdks show high sequence conservation (67% similarity and 47% identity between cdk2 and cdk7), a model of the cyclin H–cdk7 complex was assembled on the basis of the crystal structure of cyclin A complexed with cdk2. To optimize contacts between cyclin H and the kinase, helices H2, H3 and H5 (C-terminal end) from cyclin H and from cyclin A (complexed with cdk2) were superposed (Figure 5, lower panel). In the cyclin A–cdk2 complex, kinase recognition involves Repeat 1 (interface 1248 Å^2 red surface; Figure 5, upper panel) as well as the H_N helix (interface 645 Å^2 orange surface Figure 5, upper panel). Residues which would play a role analogous to that of H_N in cyclin A could not be identified in cyclin H. The comparison of cyclin H with cyclin A in its complex form shows a large conservation of the residues located at the interface (Figures 5 and 6). In this case, cdk contacts would involve at least Repeat 1. Cα atoms from residues in contact with the kinase in the cyclin A–cdk2 complex superpose to the corresponding residues of cyclin H with an r.m.s.d. of 1.3 Å; the cyclin A–cdk2 complex is thus a reliable model

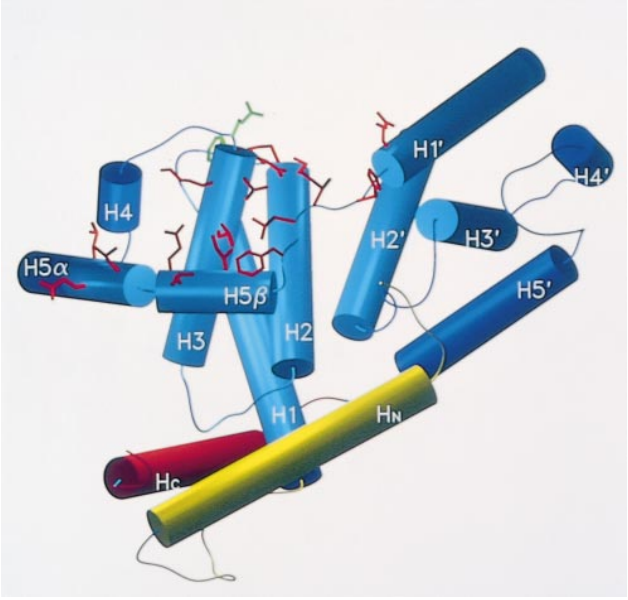


Fig. 6. Residues involved in cdk binding. Regions of cyclin H which are structurally equivalent (see Materials and methods) to cyclin A are shown in dark blue. Non-homologous regions are drawn in light blue. From the structure of the cyclin A–cdk2 complex, residues of cyclin H which should interact with cdk7 were identified. Most of the residues are in a comparable conformation in cyclins A and H (red). Two residues, Glu117 and Phe118, might undergo a significant conformational change upon kinase binding (green).

for the cyclin H–cdk7 pair. The residues of cyclin H in contact with the cdk7 kinase would be located mainly at the C-terminus of H3 and within the segment H5 to H1′. They constitute the most conserved region of the cyclins and include the invariant Lys114 and Glu147 (Figure 3A). Although the structure of cyclin H was solved in its free state, most side chains of the amino acids participating in kinase recognition are oriented in a conformation close to that observed in the complexed cyclin A.

The interacting parts of the two structures fit almost perfectly except for two loop residues, Glu117 and Phe118 (Glu269 and Ile270 in cyclin A) (Figure 6). Phe118 packs against Pro71, Pro76 and Phe71 in a conformation different from that observed for Ile270 in cyclin A (Brown *et al.*, 1995; Jeffrey *et al.*, 1995). Since in cyclin A Glu269 interacts strongly with the conserved Arg150 in the T-loop of cdk2, it is legitimate to predict a similar conformation of the loop in the cyclin H–cdk7 complex.

Cyclin A can functionally substitute for cyclin H

To analyse the functional significance of the structural conservation of cyclins, we tested the ability of cyclin A to substitute for cyclin H. Cyclin A was able to activate cdk7 only in the presence of MAT1 (Figure 7A, lanes 1 and 2). Phosphorylation of the ctd4 substrate is highly specific since neither cyclin A, cdk7 nor MAT1 alone or in combination present any kinase activity (Figure 7A, lanes 3, 4, 7 and 8). Although TFIIB harbours similar helical structures (Nikolov *et al.*, 1995), it is unable to interact with (data not shown) and subsequently activate cdk7 in the presence of MAT1 (Figure 7A, lane 6). Having demonstrated a functional interaction between cyclin A, the partner of cdk2, and cdk7, the partner of cyclin H, we performed the converse experiment. The GST–cdk2 recombinant kinase was tested for its ability to retain either recombinant cyclin H or cyclin A on a glutathione resin (Figure 7B, lane 5). After extensive washing with 250 mM KCl, the adsorbed proteins were tested by Western blotting. Cdk2 was shown to recognize specifically cyclin H (Figure 7B, lane 5) as well as cyclin A its cognate regulatory subunit (data not shown). Control experiments show that neither GST nor protein G–Sepharose beads interact with either cyclin H or cyclin A. Together, these experiments suggest a common mode of cyclin recognition within the kinase family and demonstrate the interchangeability of both cyclins.

Discussion

The first striking feature of the molecular structure of human cyclin H is the existence of a highly conserved core domain made of two repeats (H1–H5 and H1′–H5′ respectively) which form the canonical cyclin fold. Despite the large sequence variability in Repeat 2, we observe a high structural homology between cyclin H (this work) and cyclin A (Brown *et al.*, 1995; Jeffrey *et al.*, 1995), two distantly related cyclins. The signature sequence derived from the structure analysis rests largely within Repeat 1 and allows cyclin-type sequences to be retrieved from the various databanks. Residues from the two cyclins located at the interface with the kinase, as deduced from a comparison of cyclin H with the structure of the cyclin A–cdk2 complex (Jeffrey *et al.*, 1995), correlate almost perfectly. Our biochemical data provide a direct confirmation of the importance of this conservation, since both cyclin H and cyclin A can interact with and activate cdk7. Conversely, cyclin H was shown to be capable of binding to cdk2.

The second interesting feature lies in the differences between cyclins H and A, which concern the regions located outside the cyclin fold. In cyclin H, these residues form a third domain comprising two large helices H_N and H_C. Our biochemical data show their absolute requirement for cdk7 activation. Both interact with Repeat 1 but are located opposite the kinase binding site and distant from it. In cyclin A, residues at positions 182 and 185 of the H_N helix interact directly with the kinase without conformational change (Brown *et al.*, 1995). By analogy with cyclin A–cdk2, it is tempting to implicate the N-terminal helix H_N of cyclin H in a specific association with cdk7. An interaction similar to that observed in the cyclin A–cdk2 complex would require a major conformational rearrangement to bring H_N in close contact with the cdk7 kinase, whereas maintaining H_N in its present location would limit the interactions to the first residues of the cyclin (disordered in the crystal structure). In both cases, the situation would be different from that of cyclin A, thus underscoring the role of H_N in specificity.

In addition to the putative interactions of H_N of cyclin H with the kinase, the most plausible role for H_N and H_C is to stabilize the correct conformation of Repeat 1 and to fix the relative position of Repeat 2. Removing H_N results in the loss of numerous stabilizing contacts with H1 and the H2–H3 loop of Repeat 1 and in the exposure of numerous previously buried hydrophobic residues. This could explain the inactivation of the complex by a distortion of the cyclin framework. A similar explanation

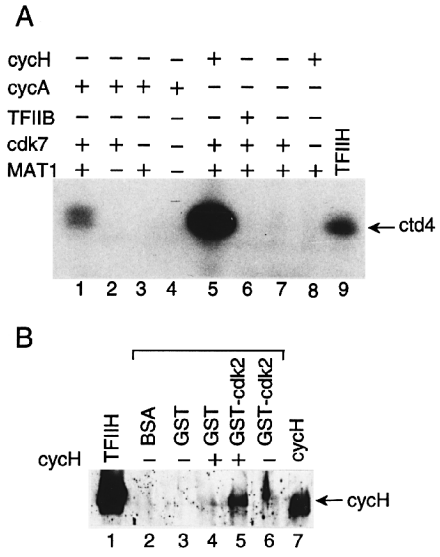


Fig. 7. Ctd kinase activity and interacting property of various cyclin fold-containing proteins. (**A**) One hundred µl of baculovirus-infected Sf9 cell extract MAT1, 100 ng of partially purified recombinant cdk7 and 300 ng of either cyclin H, cyclin A or TFIIB were pre-incubated in various combinations as indicated at the top of the panel, immunoprecipitated with Ab-cdk7 and tested as indicated in Materials and methods. Forty µl of TFIIH (heparin fraction from our purification procedure; Gerard *et al.*, 1991) were processed under the same conditions (lane 9). (**B**) Three hundred ng of cyclin H were pre-incubated with the same amount of GST–cdk2 or GST, and 300 ng of BSA, GST and GST–cdk2 were pre-incubated alone. Then, 20 µl of glutathione–Sepharose beads were added. After shaking for 1 h at 4°C and extensive washing, protein complexes were analysed by Western blotting with the polyclonal antibodies directed against cyclin H (lanes 2–6); 5 µl of TFIIH or 10 ng of cyclin H were used as control.

holds for H_C, which sits in a symmetrical position around H1.

As shown by the cyclin interchange experiments, the functioning of the cdk7 kinase could depend partially on the topological position of H_N and H_C. Indeed, the specific activity of the resulting non-cognate complex cyclin A–cdk7 is much weaker than the cognate cyclin H–cdk7 (Figure 7), which implies a modulator effect of the two helices. It could also be emphasized that the two helices H_N and H_C may participate in the choice of the substrate to be phosphorylated, either cdk, a substrate involved in cell cycle regulation, or the largest subunit of RNA polymerase II, the main participant in transcription of protein-coding genes (data not shown; D.Bentley personal communication).

Another putative role for the two terminal helices can be clearly ruled out, i.e. the recognition of MAT1, the third partner of the CAK complex. Indeed, previous studies have demonstrated that MAT1 was required for stimulating phosphorylation of ctd and cdk substrates. MAT1 was supposed to stabilize the pre-formed binary complex cyclin H–cdk7 (Fisher and Morgan, 1994; Fisher *et al.*, 1995). Bacculovirus co-infection experiments show that MAT1 interacts directly and specifically with cdk7, independently of the presence or absence of cyclin H (our unpublished results). Moreover, since cyclin A–cdk7 is able to phosphorylate the ctd substrate only in the presence of MAT1, it is likely that MAT1 activates cdk7 kinase independently of the cyclin type.

A general scheme can be proposed in the light of the present structure–function study, and using already published work on other cyclin–kinase pairs (for review, see Meijer, 1996). Our results show that cdk7 may function, at least *in vitro*, not only with cyclin H but also with cyclin A, demonstrating that cyclins and cdks can functionally interchange. Indeed, cdk2 functionally interacts not only with cyclin A, but also with cyclin E and cyclin B; the latter may also interact with cdk1 (for review, see Meijer, 1996). In addition, the specificity of the cyclin and/or the substrate has to take into account the nature and the position of the elements located outside the cyclin fold, e.g. H_N and H_C, and to a certain extent the interface between cdk and cyclins. Whether or not the interaction between non-cognate cyclin–cdk molecules is physiologically relevant will depend on their simultaneous presence in a precise compartment of the cell during various stages of the cell cycle.

The present study on cyclin H and cyclin A will be of great help in designing some potential drugs that may modulate the kinase activity (Meijer, 1996) and/or, by extension, TFIIH transcription activity. In these circumstances, negative regulation of cell proliferation, that may affect transcription regulation, cell cycle checkpoint control and tumour suppression, could occur through interference of the anti-tumour agent with the cyclin–cdk complex.

Materials and methods

Crystallographic data

The structure of human cyclin H, as reported in Andersen *et al.* (1996), was solved with the multiple isomorphous replacement (MIR) method using two derivatives. The model was refined between 14.0 and 2.6 Å. Within this resolution range, the *R*-factor is 22.2% for 18 836 reflections larger than 2σ, and the free *R*-factor is 27.1%. It includes 277 amino acid residues and 55 water molecules. The stereochemistry of the model was inspected with PROCHECK (Laskowski *et al.*, 1993). For the main chain dihedral angles, 89.8% of the residues are in the most favoured regions and 9.4% are in the additional allowed regions of the Ramachandran plot. The PROCHECK G-value is 2.7.

Structure analysis

Structural superpositions were performed by least-squares fitting using O (Jones *et al.*, 1991). Superposition were done by superposing an initial set of equivalent Cα atoms. The resulting structural alignment was then improved by allowing all Cα atoms from the regions of interest to be included in the search procedure, in order to derive the lowest r.m.s.d. automatically. Residues were considered as structurally equivalent if they were within 3.8 Å of each other and if the superposition was topologically meaningful. Surface interaction calculations (interactions between cyclins and kinases) and buried surface calculations (interactions between helices HN and HC and the cyclin fold helices) were performed with the Grasp package (Nicholls *et al.*, 1991). The areas were obtained in calculating the molecular surface–surface contacts which are within a 3.5 Å cut-off.

Sequence analysis

The sequence analysis and database search were done with the ClustalW 1.6 (Thompson *et al.*, 1994) and UWGCG (Devereux *et al.*, 1984) packages, with default parameters. ClustalW 1.6 allows incorporation of secondary structure information in the sequence alignment process, in order to keep the integrity of secondary structure elements (helices and sheets). The secondary structure data were extracted from the 3D structures with DSSP (Kabsch and Sander, 1983).

We first aligned cyclins H and C by taking into account the secondary structure. We then applied the same procedure for A-BDE cyclin types with cyclin A as the reference for the secondary structure information. The two sequence alignments were then merged. From this alignment,

patterns of residues conserved in H-C (pH-C) and A-BDE (pA-BDE) cyclin types were extracted (see Figure 3A).

To generate a global alignment, a representative set of cyclins was retrieved based on the 'cyclin' keyword search. From this list, 100 sequences were selected in order to encompass the sequence variability of the cyclins. The different cyclin subgroups, as obtained from blast scoring statistics, were first aligned and then a complete profile alignment was generated, aligning each group relative to the master profile. The secondary structure data, as seen in cyclin H, was used throughout the profile alignment process. From this alignment, a general pattern was extracted encompassing a region of 60 residues. The complete derived pattern including residues from H1, H2 and H3, is the following:

[VILM]xx[VILMFYW]x<4,10> (H1)

[**ASC**TVLIMF]x3[ASCTVILMF]x3[GASTC]xx[**C**VILMFYW]
[VILMFYW]x[RKWSLQ][VILMFYW] (H2)

x<10,25>[GASTVC][ASTCV][**A**VILMF][**S**VILMFYW][**A**TVILM]
[GASC][ASTCV][KR] (H3)

where x is any amino acid; [XiXj] is Xi or Xj amino acids in this position and x<i,j> is any amino acid between i and j times.

At some positions, the small and hydrophobic residue classes require additional residues types (e.g. the Ser in the hydrophobic set at the end of H3, in bold in the above pattern) to describe sequence variability among cyclins.

Database search

Using the pattern of conserved residues defined for all cyclins (Figure 3A, motif), a database search (translated EMBL database V47.0) retrieved 275 sequences, of which 236 are annotated as cyclins. The already published cyclin motif (motif PS00292, Bairoch *et al.*, 1996), based on residues from H1 and H2, retrieved 184 sequences from the database, which were also retrieved by our motif, with the exception of one. In this latter entry (accession No. APCYCLI_1), the conserved positively charged residue is deleted. In addition to these 183 sequences, 55 additional cyclin sequences were found [e.g. UME3 (SC16248_1) a C-type cyclin, SRB11 (SC20221) an RNA polymerase II subunit, SSN8 (SC20635_1) and PUC1 (SPUC1_1)]. Due to the degeneracy of the motif, the remaining 37 sequences (10 gene products and 27 characterized sequences) were found.

Out of the 10 gene products found in the previous set, eight are cyclins, as found with a blast search with a probability score $>1.0e^{-20}$ (Altschul *et al.*, 1994). For the two last gene products, no significant homologous sequence is found. Of the remaining 27 characterized sequences, some are either predicted to adopt a cyclin-like fold (e.g. the ECSBCA8 family; five sequences) or are known to share a common fold with cyclins, like TFIIB (Nikolov *et al.*, 1995) (10 sequences). From the last 12 sequences, two are clearly false positives due to the pattern degeneracy.

Oligonucleotide primer design and PCR

A PCR strategy was used to generate the cyclin H mutants. Briefly, two primers (~30mers) were used to generate each of the N- and C-terminal deletion mutants. The 5′ primer contained a restriction site for *Nde*I. The 3′ primer included a restriction site for *Bam*HI. This ensured that, after being cut with *Nde*I and *Bam*HI and ligated into the pET15b vector (Novagen) digested with the same restriction enzymes, the sequence would be in-frame.

The PCR mixture included 0.2 mM deoxynucleoside triphosphate, 50 pmol of each of two oligonucleotide primers, 400 ng of template (pET15b-cyclin H vector), 10 mM KCl, 20 mM Tris–HCl pH 8.8, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100 and 2 U of Vent DNA polymerase (New England Biolabs) in a final volume of 100 µl. After the mixtures were covered with 50 µl of mineral oil, the reactions were carried out in a DNA thermal cycler (Perkin Elmer) at three temperatures (94°C for 1 min, 55°C for 2 min, 72°C for 1 min) for 25 cycles. The PCR-amplified DNA fragments were analysed by agarose gel electrophoresis, and the appropriate sized fragments were cut from the gel and purified with a Geneclean kit (Bio 101). After digestion with *Nde*I and *Bam*HI, the fragments were cloned into the same sites of the pET15b vector.

Expression in Escherichia coli and in Sf9 cells infected by baculovirus

GST–cdk2 protein was expressed in *Escherichia coli* cells transformed by a pGEX-derived family vector (Pharmacia). N- and C-terminal mutants of human cyclin H were expressed in *E.coli* cells transformed by pET15b-derived vectors. These cells were grown at 25°C to an OD_{600} of 0.6–1.0 and then induced by addition of 0.8 mM IPTG. After 4 h, cells were harvested and washed in phosphate-buffered saline (cells expressing GST–cdk2) or in 50 mM Tris–HCl buffer pH 7.5, 500 mM NaCl, 5 mM imidazole (cells expressing cyclin H deletion mutants).

The *Nde*I–*Bam*HI fragments containing the entire coding sequences of human cyclin H and bovine cyclin A were inserted into pET15b and pET16b vectors respectively. These cyclins were expressed and purified as previously described for cyclin H in Poterszman *et al.* (1996).

The *Eco*RI fragments containing the entire coding sequences of human cdk7 and MAT1 were inserted into the pVL1392 vector (PharMingen). Monolayer Sf9 cells (~1.2×10^8 cells) were infected by cdk7 or MAT1 baculoviruses with a multiplicity of infection of 2 or 10 plaque-forming units per cell respectively. The cells were washed in 20 mM Tris–HCl buffer pH 7.5, 50 mM NaCl, 2 days after infection, and lysed at 4°C in the same buffer containing 5 mM β-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride, 2.5 µg/ml aprotinin, leupeptin, pepstatin, antipain and chymostatin. The lysate was clarified by ultracentrifugation and the resulting supernatant was used for assays with cyclin H deletion mutants and for MAT1.

Human cdk7 protein was purified from the clarified lysate, first on ion exchange DEAE–Sepharose resin (Pharmacia) then by S-Sepharose resin (Pharmacia) followed by TSK-phenyl resin (TosoHaas).

GST–cdk2 protein was purified from the clarified lysate on glutathione–Sepharose resin (Pharmacia) by elution with 50 mM Tris–HCl buffer pH 7.5, 150 mM NaCl, 10 mM glutathione.

Protein–protein interaction and kinase assays

The fractions to be tested for protein–protein interactions were incubated for 1 h at room temperature in buffer A (20 mM Tris–HCl pH 7.5 buffer, 0.1 mM EDTA, 10% glycerol and 0.1% NP-40) plus 50 mM KCl. Then 20 µl of protein G–Sepharose beads (Pharmacia) cross-linked with Ab-cdk7 or 20 µl of glutathione–Sepharose beads (Pharmacia) washed in buffer A plus 100 mM KCl and 0.5 mg/ml of bovine serum albumin were added. After 1 h of shaking at 4°C, the beads were washed extensively in buffer A plus 100 or 250 mM KCl. Then the beads were either boiled in SDS loading buffer and analysed by SDS–PAGE or subjected to a kinase assay using a synthetic ctd4 peptide corresponding to four copies of the wild-type ctd consensus sequence (YSPTSPS) as substrate as previously described (Roy *et al.*, 1994).

The incubated mixture contains 100 µl of clarified Sf9 cell extract containing the overexpressed MAT1, 100 ng of partially purified recombinant cdk7 (Adamczewski *et al.*, 1996), 300 µl of either wild-type or truncated *E.coli* cyclin H or cyclin A recombinant in various combinations, before immunoprecipitation.

The monoclonal antibody Ab-cdk7 (2F8) was directed against a cdk7-derived peptide (PVETLKEQSNPALAIKRK) and Ab-MAT1 (2D3) against a micro-sequenced peptide (LEEALYEYQPLQIETYG). Polyclonal antibodies raised against the entire cyclin H molecule were produced by immunizing rabbits. For Western blot analysis, we used the ECL chemiluminescent method (Amersham, UK) according to the manufacturer's instructions.

Coordinates

The coordinates of human cyclin H have been deposited at the Brookhaven Protein Data Bank under the accession No. 1JKW.

Acknowledgements

We are thankful to J.A.Kirk for the gift of pET16b-cyclin A vector and to J.C.Cavadore for cyclin A polyclonal antibodies. We thank J.A.Endicott and N.Pavletich for providing the coordinates of the free and complexed cyclin A prior to release. We also are particularly indebted to M.Rossignol for the expression vectors, to L.Moulinier and S.Ourjoumtsev for help with refinement and to J.Arnez for comments and critical reading of the manuscript. We thank J.M.Chipoulet and I.Kolb-Cheynel for expertise in protein overexpression and purification. G.A. was supported by the Danish Natural Science Research Council. This work was supported by grants from the INSERM, the CNRS, the Ministère de la Recherche et de l'Enseignement Supérieur, the Association pour la Recherche sur le Cancer and the Direction des Recherches Etudes et Techniques.

References

Adamczewski,J.P., Rossignol,M., Tassan,J.P., Nigg,E.A., Moncolin,V. and Egly,J.M. (1996) MAT1, cdk7 and cyclin H form a kinase complex

which is UV light-sensitive upon association with TFIIH. *EMBO J.*, **15**, 1877–1884.
Altschul,S.F., Boguski,M.S., Gish,W. and Wootton,J.C. (1994) Issues in searching molecular sequence databases. *Nature Genet.*, **6**, 119–129.
Andersen,G., Poterszman,A., Egly,J.M., Moras,D. and Thierry,J.C. (1996) The crystal structure of human cyclin H. *FEBS Lett.*, **397**, 65–69.
Bairoch,A., Bucher,P. and Hofmann,K. (1996) The PROSITE database, its status in 1995. *Nucleic Acids Res.*, **24**, 189–196.
Bazan,J.F. (1996) Helical fold prediction for the cyclin box. *Proteins*, **24**, 1–17.
Brown,N.R., Noble,M.E., Endicott,J.A., Garman,E.F., Wakatsuki,S., Mitchell,E., Rasmussen,B., Hunt,T. and Johnson,L.N. (1995) The crystal structure of cyclin A [see comments]. *Structure*, **3**, 1235–1247.
Cleaver,J.E. and Kraemer,K.H. (1996) Xeroderma pigmentosum and Cockayne syndrome. In Scriver,C.R., Beaudet,A.L., Sly,W.S. and Valle,D. (eds), *The Metabolic and Molecular Basis of Inherited Diseases.* Mc Graw-Hill Book Company, New York, pp. 4393–4419.
Devereux,J., Haeberli,P. and Smithies,O. (1984) A comprenhensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387–395.
Drapkin,R., Reardon,J.T., Ansari,A., Huang,J.C., Zawel,L., Ahn,K., Sancar,A. and Reinberg,D. (1994) Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature*, **368**, 769–772.
Evans,S.V. (1993) SETOR: hardware-lighted three-dimensional solid model representation of macromolecules. *J. Mol. Graphics*, **11**, 134–138.
Feaver,W.J., Svejstrup,J.Q., Henry,N.L. and Kornberg,R.D. (1994) Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell*, **79**, 1103–1109.
Fisher,R.P. and Morgan,D.O. (1994) A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell*, **78**, 713–724.
Fisher,R.P., Jin,P., Chamberlin,H.M. and Morgan,D.O. (1995) Alternative mechanisms of CAK assembly require an assembly factor or an activating kinase. *Cell*, **83**, 47–57.
Flores,T.P., Orengo,C.A., Moss,D.S. and Thornton,J.M. (1993) Comparison of conformational characteristics in structurally similar protein pairs. *Protein Sci.*, **2**, 1811–1826.
Gerard,M., Fischer,L., Moncollin,V., Chipoulet,J.M., Chambon,P. and Egly,J.M. (1991) Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *J. Biol. Chem.*, **266**, 20940–20945.
Jeffrey,P.D., Russo,A.A., Polyak,K., Gibbs,E., Hurwitz,J., Massague,J. and Pavletich,N.P. (1995) Mechanism of CDK activation revealed by the structure of a cyclin A–CDK2 complex. *Nature*, **376**, 313–320.
Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) 'O'. *Acta Crystallogr.*, **A47**, 110–119.
Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
Meijer,L. (1996) Chemical inhibitiors of cyclin-dependent kinases. *Trends Cell Biol.*, **6**, 393–397.
Miller,S., Lesk,A., Janin,J. and Cyrus,C. (1987) The accessible surface area and stability of oligomeric proteins. *Nature*, **328**, 834–836.
Morgan,D.O. (1995) Principles of CDK regulation. *Nature*, **374**, 131–134.
Nicholls,A., Sharp,K. and Hoenig,B. (1991) Protein folding and association: insight from the interfacial and thermodynamics properties of hydrocarbons. *Proteins*, **11**, 282–290.
Nigg,E.A. (1995) Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. *BioEssays*, **17**, 471–480.
Nigg,E.A. (1996) Cyclin-dependent kinase 7: at the cross-roads of transcription, DNA repair and cell cycle control? *Curr. Opin. Cell Biol.*, **8**, 312–317.
Nikolov,D.B., Chen,H., Halay,E.D., Usheva,A.A., Hisatake,K., Lee,D.K., Roeder,R.G. and Burley,S.K. (1995) Crystal structure of a TFIIB–TBP–TATA-element ternary complex. *Nature*, **377**, 119–128.
Poterszman,A., Andersen,G., Busso,D., Rossignol,M., Egly,J.M. and Thierry,J.C. (1996) Expression in *E.coli*: purification and characterisation of cyclin H, a subunit of the human general transcription/DNA repair factor TFIIH. *Protein Expr. Purification*, in press.
Roy,R., Adamczewski,J.P., Seroz,T., Vermeulen,W., Tassan,J.P., Schaeffer,L., Nigg,E.A., Hoeijmakers,J.H. and Egly,J.M. (1994) The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell*, **79**, 1093–1101.
Russo,A., Jeffrey,P. and Pavletich,N. (1996a) Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nature Struct. Biol.*, **3**, 696–700.
Russo,A.A., Jeffrey,P.D., Patten,A.K., Massaggué,J. and Pavletich,N.K. (1996b) Crystal structure of the p27kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A–Cdk2 complex. *Nature*, **382**, 325–331.
Schaeffer,L., Roy,R., Humbert,S., Moncollin,V., Vermeulen,W., Hoeijmakers,J.H., Chambon,P. and Egly,J.M. (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor [see comments]. *Science*, **260**, 58–63.
Svejstrup,J., Vichi,P. and Egly,J.M. (1996) The multiple roles of transcription/repair factor TFIIH. *Trends Biol. Sci.*, **21**, 346–350.
Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

Received on October 9, 1996; revised on November 18, 1996

Note added

An independent determination of the human cyclin H structure was published in *Nature Structural Biology*, **3**, 849–855 after this present paper was sent for review. As can be judged from both studies, the two structures are identical.