The crystal structure of human cyclin H

G. Andersen, A. Poterszman, J.M. Egly, D. Moras, J.-C. Thierry*

Institut de Génétique et Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, 1 rue Laurent Fries, BP163, 67404 Illkirch Cédex, C.U. de Strasbourg, Strasbourg, France

Received 11 September 1996; revised version received 1 October 1996

Abstract The crystal structure of human cyclin H has been solved at 2.6 Å resolution by the MIR method and refined to an R-factor of 23.1%. The core of the molecule consists of two helical repeats adopting the canonical cyclin fold already observed in the structures of cyclin A [Brown et al. (1995) Structure 3, 1235–1247; Jeffrey et al. (1995) Nature 376, 313–320; Russo et al. (1996) Nature 382, 325–331] and TFIIB [Nikoilov et al. (1995) Nature 377, 119–128]. The N-terminal and C-terminal residues form a new domain built on two long helices interacting essentially with the first repeat of the molecule.

Key words: Human; Cyclin; Cell cycle; TFIIH; Transcription factor; X-ray structure

1. Introduction

Cyclins are the positive regulatory subunits of cyclin-dependent kinases (CDK). They play a crucial role in the coordination of the eucaryotic cell cycle by binding to the catalytic subunits of CDKs [5,6]. They share distant sequence homology over a 100 amino acid region called the cyclin box and have been classified in two phylogenetically divergent subfamilies. The first contains A, B, D, E, F, G, ... cyclins and the second the C and H cyclins [7].

Cyclin A is so far the only cyclin whose 3-dimensional structure has been determined. A member of the first subfamily, it has been extensively studied in its free state [1], complexed with CDK2 [2] and in a ternary complex with CDK2 and the inhibitor $p27^{Kip1}$ [3]. These structures show that the cyclin box is associated with a structural domain of five helices repeated once in the structure. The same topology, now referred to as the cyclin fold, was later found in the structure of the general transcription factor TFIIB [4]. Upon the formation of the CDK2–cyclin A complex, cyclin A does not undergo significant conformational changes [1,2] whereas very large ones occur in CDK2 when compared to its free state [8].

No structural information has so far been available for the second subfamily composed of H and C cyclins. Human cyclin H, a protein of 323 amino acids, is well conserved from yeast to human. It shares significant homologies with C-type cyclins [9,10]. However, only 23.1% identity and 43% similarity is observed with human cyclin A. Cyclin H was originally identified as a member of the Cyclin-dependent activating kinase complex, composed of cyclin H, CDK7 and MAT1 [11–15]. Further investigations revealed that cyclin H is also a subunit of the general transcription factor TFIIH, a multiprotein complex involved in three important mechanisms, transcription, DNA repair and cell cycle regulation, containing at least nine subunits and shown to possess several enzymatic activities including helicase, ATPase and kinase [16–21].

As a first step towards structural investigation of TFIIH, we present here the crystal structure of human cyclin H determined at 2.6 Å resolution by the multiple isomorphous replacement method.

2. Materials and methods

Recombinant human cyclin H was expressed in *Escherichia coli* and purified according to the protocol described in [22]. For crystallisation the protein was desalted into a solution containing 300 mM ammonium phosphate, 30 mM sodium citrate, 10 mM β-ME at pH 5.8 and concentrated to 15–20 mg/ml in Centricon cells. Crystals were grown by vapour diffusion in hanging drops at 4°C by mixing equal volumes of the protein and of a reservoir solution containing 1 M ammonium phosphate, 0.1 M sodium citrate, 2% glycerol, 10 mM β-ME at pH 5.8. Brick-shaped crystals with maximum dimensions of $0.6\times0.4\times0.2$ mm^3 grew in 1–3 weeks. They belong to the space group I4(1)22 with cell parameters $a=b=84.04$ Å and $c=373.3$ Å. The asymmetric unit contains one molecule of cyclin H and has a very high solvent content of 72%.

In order to allow preparation of heavy atom derivatives and data collection at cryogenic temperature, crystals were transferred stepwise into a stabilising solution containing 1 M ammonium sulphate, 50 mM MES, 10 mM β-ME at pH 5.8. The quality of the diffraction pattern and the resolution limit of diffraction from stabilised crystals were similar to that obtained from crystals mounted in the mother liquor containing ammonium phosphate. Native (20–3.0 Å) and derivative data sets (12–3.5 Å) were collected at 4°C to solve the structure. Data used for structure refinement were collected at a cryogenic temperature (Table 1). For this purpose, crystals were transferred stepwise to the stabilising solution containing in addition 32% v/v glycerol and then flash-frozen in a nitrogen stream at −160°C [23]. All data were integrated and scaled using the program XDS [24].

Heavy atom derivatives were obtained by soaking crystals in the stabilising solution (without β-ME) containing 10 μM of either $HgBr_2$ (16 h) or EtHgCl (4 h). Two sites were detected for the $HgBr_2$ derivative and one site for the EtHgCl derivative. Reflections between 12 and 4.5 Å were phased to an overall figure of merit of 0.44, using programs from the CCP4 suite [25]. The phasing powers were 1.03 and 1.26 for the $HgBr_2$ and EtHgCl derivative, respectively. In spite of the low value of the figure of merit, the quality of the MIR map calculated in the range 12–4.5 Å allowed us to locate the molecule in the unit cell and to build long helices in the electron density with the modelling program 'O' [26]. Phases were extended to 3.5 Å by solvent flattening with the program DM [27]. The resulting map enabled us to build most of the molecule. Several cycles of combination of MIR and model phases with SIGMAA [28] followed by density modification with DM were used to extend the phases to 3.0 Å and improve the model. The data collected at cryogenic temperature were then used to carry on further model building and refinement. The final model was subjected to minimisation, simulated annealing/torsion angle dynamics, B-factor refinement and bulk solvent correction with the package X-PLOR [29], and the resolution was gradually extended from 3.0 to 2.6 Å. The final R-factor is 23.1% for 18836 reflections

*Corresponding author. Fax: (33) 03-88-65-32-76.

Abbreviations: MES, 2-(*N*-morpholino)-ethanesulfonic acid; β-ME, β-mercaptoethanol; MIR, multiple isomorphous replacement; EtHgCl, ethylmercury chloride; CDK, cyclin-dependent kinase

PII S0014-5793(96)01143-X

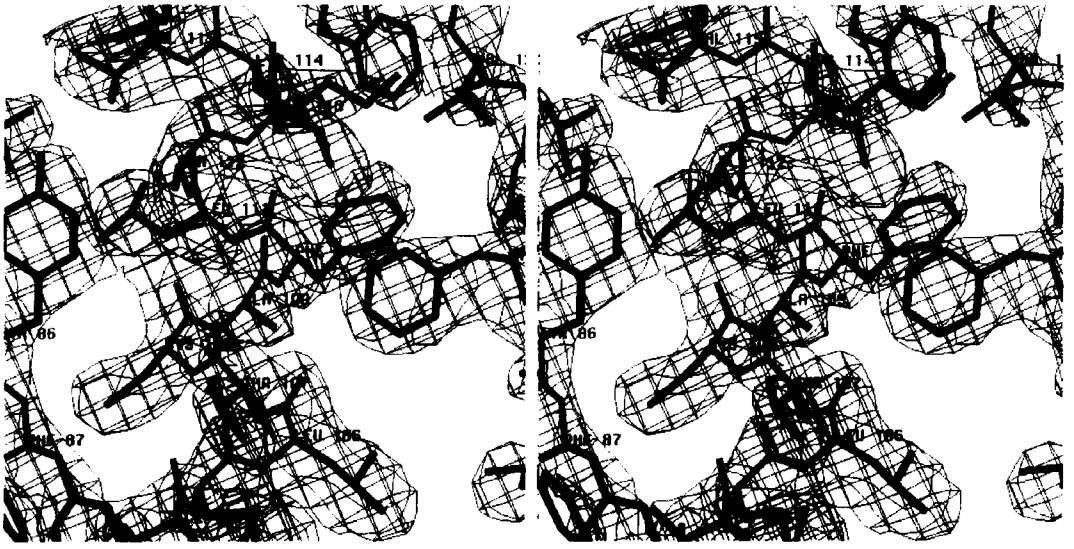

Fig. 1. Stereo view of the experimental $2F_o$-F_c electron density map contoured at 1.7σ after the final refinement of the cyclin H model at 2.6 Å resolution. This region of the first domain shows a cluster of hydrophobic amino acids, rich in aromatic residues, located close to the centre of helix H3, characteristic of the cyclin box.

larger than 2σ, and the free R-factor is 29.0% (Table 1). The average B-factor of the model as given by the bulk solvent correction is 50 Å^2 except for two disordered loops (between amino acids 181–183 and 234–243) where B-factors can exceed 100 Å^2. This is in agreement with the overall B-factor determined by Wilson plot on the measured intensities. It can be explained by the high solvent content in the crystal and the few packing interactions between the molecules in the crystal lattice. The stereochemistry of the model was inspected with PROCHECK [30]. For the main chain dihedral angles, 83% of the residues were in the most favoured regions and 17% were in the additional allowed regions of the Ramachandran plot. The PROCHECK G-value was 0.0.

3. Results and discussion

The final model includes 277 amino acid residues and 45 water molecules. The first 10 residues at the N-terminus and the last 35 residues at the C-terminus (288–323), could not be built with confidence due to a probable disorder in the crystals. Fig. 1 shows a portion of the final $2F_o$-F_c electron density map around a central helix of the model. It is of interest to note that limited proteolysis experiments performed with chymotrypsin generate a stable product with an approximate molecular weight of 31 kDa with the N-terminal starting at Ser^{12} (our unpublished results). This proteolytic fragment corresponds to what can be traced in the electron density map. As expected, the mercury sites are located close to Cys^{57}, Cys^{84} and Cys^{280} in the refined structure.

The structure of cyclin H (Fig. 2) comprises two α-helical domains which form two lobes linked by a short hinge region giving the molecule its elongated shape. A third essentially

Table 1
Statistics for data used in the structure determination

	Number of crystals	Resolution (Å)	Mean I/σ(I)	R_{merge} (%)	Data coverage (%)	Multiplicity	Temperature (°C)	Source and detector
Native 1	1	20–3.8	16.1	4.4	91.5	2.4	4	Rot. anode
		(4.0–3.8)	(5.1)	(9.2)	(62.7)	(1.4)		area detector
Native 2	2	12–3	25.2	5.4	95.3	2.6	4	LURE W32
		(3.2–3.0)	(6.5)	(22.8)	(95.9)	(2.7)		MAR IP
$HgBr_2$	2	12–3.5	26.8	6.1	94.3	3.2	4	LURE W32
		(4–3.5)	(13.3)	(13.2)	(95.9)	(3.0)		MAR IP
EtHgCl	2	12–3.5	32.1	7.5	82.9	4.0	4	LURE W32
		(4–3.5)	(17.9)	(15.9)	(85.7)	(4.1)		MAR IP
Cryo 1	1	14–3.1	21.2	6.3	93.3	3.5	−154	Rot. anode
		(3.5–3.1)	(7.6)	(18.9)	(94.6)	(3.5)		MAR IP
Cryo 2	1	10–2.6	23.2	5.9	91.5	2.8	−160	LURE W32
		(2.7–2.6)	(4.5)	(28.6)	(95.0)	(2.6)		MAR IP

						r.m.s	
Refinement	Resolution (Å)	Reflections (with $\|F\| > 2\sigma$)	Number of non-hydrogen atoms	Number of waters	R_{conv}/R_{free} (%)	Bonds (Å)	Angles (°)
	14–2.6	18836	2288	45	23.1/29.1	0.017	2.07

In the resolution range 14–2.6 Å there are 21 133 possible unique reflections. From a total of 20 809 measured reflections at 14–2.6 Å 1575 reflections were selected for the test set. The values in parentheses are for the outer resolution shell. Abbreviations: $R_{merge} = (\Sigma_h \Sigma_{j=1,N} |I_h - I_h(j)|)/\Sigma N \times I_h$ for the intensity of reflection h observed N times, r.m.s. = root-mean-square deviations from ideal geometry, $R_{conv} = \Sigma_h ||F_o| - |F_c|| / \Sigma_h |F_o|$ where $|F_c|$ is the calculated structure factor scaled to F_o, R_{free} is identical to R_{conv} on a subset of test reflections not used in refinement.

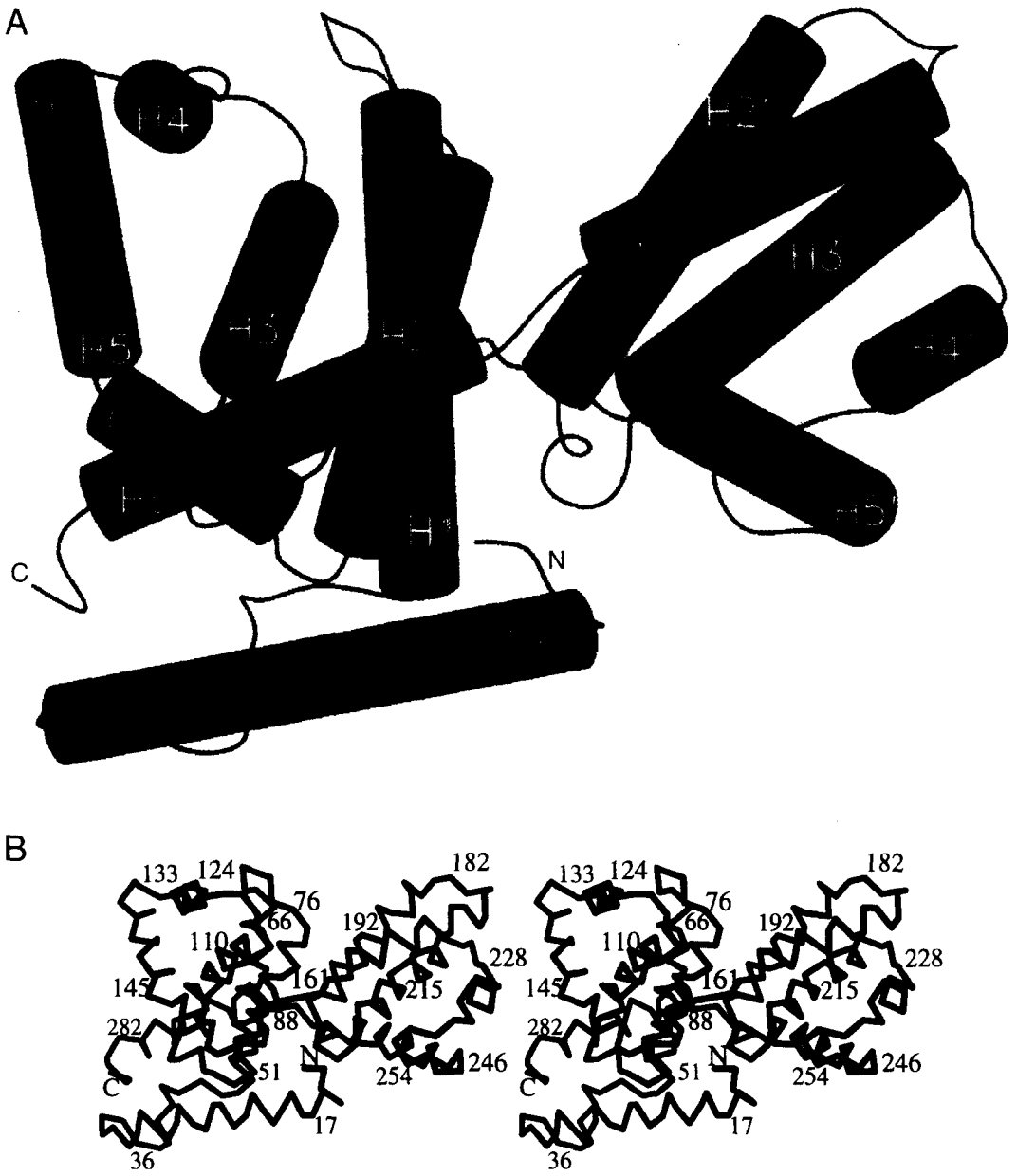

Fig. 2. The fold of human cyclin H. (A) Schematic diagram of human cyclin H using the program SETOR [32]. The secondary structures, helices represented as cylinders and random coil stretches of the protein, were defined using the program DSSP [33]. The first two compact domains (blue) share the common cyclin fold already observed in the structure of cyclin A [1,2]. They are built of five helices identified as H1 to H5 for the first repeat and H1′ to H5′ for the second. The long helix H5 kinks at Glu^{145} into two helices, H5α and H5β. HN (magenta) and HC (pink) designate the N-terminal and C-terminal helices which form a third domain characteristic of cyclin H. The view suggests the existence of a pseudo-two fold axis superposing the two helical domains. (B) Stereo view of the molecule using the same orientation with the numbering of the amino acids.

helical (helices HN and HC) domain associates the N- and C-terminal regions of the protein. The overall dimensions of the molecule are roughly 65×40×40 Å. Only a few residues adopt a conformation characteristic of β-strands but no β-sheets are present in the structure. Hence the structure agrees with CD data which predicted a mostly α-helical protein [22].

The first two domains consist of five helices each (H1–H5 and H1′–H5′). Both exhibit the cyclin fold topology, a topology also found in cyclin A and TFIIB [2,4]. The central helix H3 (H3′ in the second repeat) plays a key role in the organisation of the structure. On one side H3 (H3′) is flanked by helices H1 and H2 (H1′ and H2′) located at the interface between the two lobes of the molecule. Towards the outside of the repeat H3 (H3′) contacts helices H4 and H5 (H4′ and H5′). All these helices pack through clusters of hydrophobic residues conserved in sequences of all cyclin H. The structures of cyclin A, free [1] or complexed with CDK2 [2] as well as sequence alignments [7,31] have suggested the high conservation of the 'cyclin box' domain observed in all cyclins. This 100 amino acid region essentially corresponds to the first helical repeat (H1–H5). The structure of cyclin H allows its extension to more distantly related members of the subfamily including cyclin C [7].

Superposition of cyclin H and cyclin A illustrates the struc-

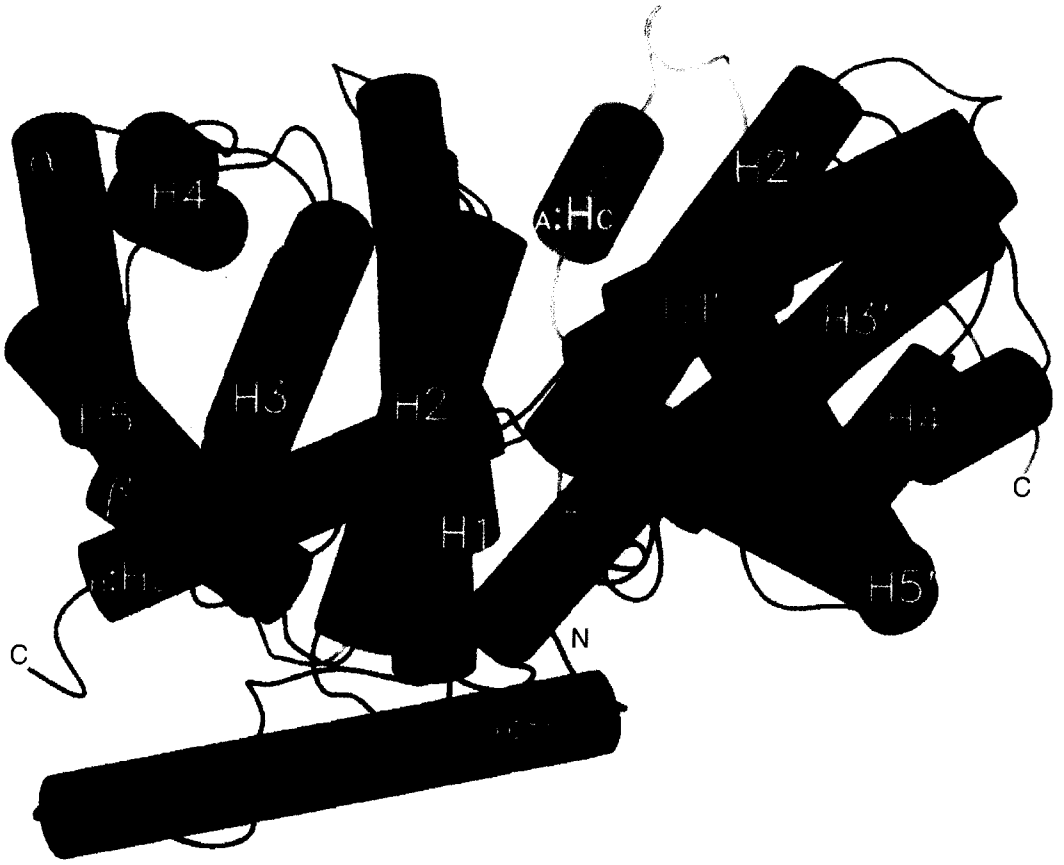

Fig. 3. Superposition of free human cyclin H and bovine cyclin A [1], with the helical repeats coloured blue and green, respectively. The backbone atoms of both cyclins are represented schematically using the program SETOR [32]. The molecules are superimposed using helices H1, H2 and H3 of the first repeat. The two repeats superimpose with a r.m.s. deviation of 1.2 Å (for 48 Cα atoms). If helices H1 to H3 of the two repeats overlap well, helices H4, H5, H4′ and H5′ are moved and differently oriented in the two structures. The N- and C-terminal helices of cyclin H and cyclin A are respectively labelled H:HN (magenta), H:HC (pink) and A:HN (red) and A:HC (yellow). Their locations are characteristic of the two cyclins, one essentially close to the first repeat (cyclin H), the other at the interface between the two repeats (cyclin A).

tural similarity of the two repeats in both structures (Fig. 3). The three helices H1, H2 and H3 from the core of the first module of cyclin H superimpose with a root-mean-square (r.m.s.) value of 1.2 Å for 48 Cα atoms on the corresponding helices of cyclin A. However, significant differences can be observed between the two cyclins. Helices H1 and H5 are respectively one turn and two turns longer at their N-terminal ends in cyclin H than in cyclin A. A kink in helix H5 occurs at Glu^{145} in the middle of the helix. It is stabilised by two water molecules so that the helix segments forms an angle of approximately 30°.

Two extended structures located between helices H5 and H1′ and H5′ and HC, form a hinge holding the two lobes of the molecule. A set of interactions, including salt bridges, involving amino acids from the hinge, the two first helices of each lobe (H1, H2, H1′ and H2′) as well as water molecules, imposes the relative positioning of the two lobes. This orientation is different from that observed in TFIIB and similar to that in cyclin A despite a different set of interactions in the interface region of this molecule.

The third domain, specific to cyclin H, is essentially in contact with the first repeat and the hinge region. It is built of two long helices, HN and HC and the connecting loop from HN to H1. Helix HN packs with one side of H1 and its N-terminus points into the cleft between the two repeats. Helix HC sits on the other side of H1 and contacts the loop connecting helices H2 and H3. This domain plays a major role in the structural integrity of cyclin H and distinguishes it from cyclin A. It is located on the side opposite to the kinase recognition surface as seen in the cyclin A/CDK2 complex [2]. In cyclin A, the N- and the C-terminal ends are separated and fill two clefts between the repeats (Fig. 3). This gives to cyclin A a more compact shape compared to cyclin H.

The regions of cyclin A containing the amino acids shown to interact with the kinase in the structure of the complex [2], superpose almost perfectly on the structure of cyclin H. Most of these are highly conserved within the cyclin families. The similar location of these residues implies that the CDK2–cyclin A structure provides a good model for most if not all CDK–cyclin pairs. The interface between CDK2 and cyclin A, which involves residues from the first repeat and the linker to the second [2], is likely to be similar in all complexes. The residues Tyr^{185} and Ile^{182} in the N-terminal helix of cyclin A are also involved in kinase recognition [2]. In the structure of cyclin H there are no equivalent residues. However, it is possible that the 10 N-terminal residues, which cannot be traced, play an analogous role in the cyclin H–CDK7 complex since the Cα atom of Trp^{11} is within 3 Å of the Cα atom of Lys^{192} of cyclin A when the two structures are superposed.

Coordinates have been deposited at the Brookhaven Protein Data Bank.

Acknowledgements: We thank A. Mitschler for his assistance with data collection, R. Fourme and staff at LURE for beamline support, S. Ourjoumsev and L. Moulinier for helping during the determination and the refinement of the structure, R. Ripp for help with the drawings and J. Arnez for critical reading of the manuscript. G.A. was supported by a fellowship from the Danish Natural Science Research Council. We also thank J. Endicott for providing the coordinates of cyclin A and N.P. Pavletich for the coordinates of the cyclin A–CDK2 and the cyclin A–CDK2-p27^{Kip1} complexes. This work was supported by a grant from the Association pour la Recherche sur le Cancer.

References

[1] 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) Structure 3, 1235–1247.
[2] Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J. and Pavletich, N.P. (1995) Nature 376, 313–320.
[3] Russo, A.A., Jeffrey, P.D., Patten, A.K., Massaggué, J. and Pavletich, N.K. (1996) Nature 382, 325–331.
[4] Nikoilov, D.B., Chen, D.B., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G. and Burley, S.K. (1995) Nature 377, 119–128.
[5] Nigg, E.A. (1995) Bioessays 17, 471–480.
[6] Morgan, D.O. (1995) Nature 374, 131–134.
[7] Bazan, J.F. (1996) Proteins 24, 1–17.
[8] De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O. and Kim, S.H. (1993) Nature 363, 595–602.
[9] Fisher, R.P. and Morgan, D.O. (1994) Cell 78, 713–724.
[10] Makela, T.P., Tassan, J.P., Nigg, E.A., Frutiger, S., Hughes, G.J. and Weinberg, R.A. (1994) Nature 371, 254–257.
[11] Tassan, J.P., Schultz, S.J., Bartek, J. and Nigg, E.A. (1994) J. Cell Biol. 127, 467–478.
[12] Yee, A., Nichols, M.A., Wu, L., Hall, F.L., Kobayashi, R. and Xiong, Y. (1995) Cancer Res. 55, 6058–6062.
[13] Tassan, J.P., Jaquenoud, M., Fry, A.M., Frutiger, S., Hughes, G.J. and Nigg, E.A. (1995) EMBO J. 14, 5608–5617.
[14] Fisher, R.P., Jin, P., Chamberlin, H.M. and Morgan, D.O. (1995) Cell 83, 47–57.
[15] Devault, A., Martinez, A.M., Fesquet, D., Labbe, J.C., Morin, N., Tassan, J.P., Nigg, E.A., Cavadore, J.C. and Doree, M. (1995) EMBO J. 14, 5027–5036.
[16] 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) Cell 79, 1093–1101.
[17] Adamczewski, J.P., Rossignol, M., Tassan, J.P., Nigg, E.A., Moncolin, V. and Egly, J.M. (1996) EMBO J. 15, 1877–1884.
[18] Feaver, W.J., Svejstrup, J.Q., Henry, N.L. and Kornberg, R.D. (1994) Cell 79, 1103–1109.
[19] Nigg, E.A. (1996) Curr. Opin. Genet. Dev. 8, 312–317.
[20] Seroz, T., Hwang, J.R., Moncollin, V. and Egly, J.M. (1995) Curr. Opin. Genet. Dev. 5, 217–221.
[21] Shiekhattar, R., Mermelstein, F., Fisher, R.P., Drapkin, R., Dynlacht, B., Wessling, H.C., Morgan, D.O. and Reinberg, D. (1995) Nature 374, 283–287.
[22] Poterzman, A., Andersen, G., Busso, D., Rossignol, M., Egly, J.M. and Thierry, J.C. (1996) Protein Expression and Purification, in press.
[23] Teng, T.Y. (1990) J. Appl. Cryst. 23, 387–391.
[24] Kabsch, W. (1988) J. Appl. Cryst. 21, 916–920.
[25] Collaborative Computational Project, Number 4 (1994) Acta Cryst. D50, 760–763.
[26] Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Cryst. A47, 110–119.
[27] Cowtan, K. (1996) Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31, 34–38.
[28] Read, R. (1986) Acta Cryst. A42, 140–149.
[29] Brünger, A. (1992) X-PLOR Manual Version 3.1. Yale University, New Haven, CT.
[30] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) J. Appl. Cryst. 26, 283–291.
[31] Gibson, T.J., Thompson, J.D., Blocker, A. and Kouzarides, T. (1994) Nucl. Acids Res. 22, 946–952.
[32] Evans, S.V. (1993) J. Mol. Graphics 11, 134–138.
[33] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.