Journal of Cell Science 113, 4523-4531 (2000)
Printed in Great Britain © The Company of Biologists Limited 2000
JCS1689

Human Cdc5, a regulator of mitotic entry, can act as a site-specific DNA binding protein

Xiang-He Lei, Xun Shen, Xiao-Qin Xu and Harold S. Bernstein*

Department of Pediatrics, Cardiovascular Research Institute and Cancer Center, University of California, San Francisco, 505 Parnassus Avenue, Box 0130, San Francisco, California 94143-0130, USA
*Author for correspondence (e-mail: hsbernstein@pedcard.ucsf.edu)

Accepted 10 October; published on WWW 16 November 2000

SUMMARY

G_2/M progression requires coordinated expression of many gene products, but little is known about the transcriptional regulators involved. We recently identified human Cdc5, a positive regulator of G_2/M in mammalian cells. We also demonstrated the presence of a latent activation domain in its carboxyl terminus, suggesting that human Cdc5 regulates G_2/M through transcriptional activation. Despite the presence of a DNA binding domain, studies by others have failed to identify a preferential binding site for Cdc5 family members. In addition, Cdc5 recently has been associated with the splicesome in several organisms, suggesting that it may not act through DNA binding. We now report the identification of a 12 bp sequence to which human Cdc5 binds specifically and with high affinity through its amino terminus. We show that this DNA-protein interaction is capable of activating transcription. We also used a selection system in yeast to identify human genomic fragments that interact with human Cdc5. Several of these contained sequences similar to the binding site. We demonstrate that these bind human Cdc5 with similar specificity and affinity. These experiments provide the first evidence that Cdc5 family members can act as site-specific DNA binding proteins, and that human Cdc5 may interact with specific, low abundance sequences in the human genome. This raises the possibility that Cdc5 proteins may participate in more than one process necessary for regulated cell division.

Key words: Cdc5, DNA binding, Mitotic entry, Cell cycle

INTRODUCTION

The eukaryotic cell division cycle includes a series of checkpoints, which ensure that some events are completed before others begin (Murray and Hunt, 1993). One set of controls determines whether the cell replicates its genetic material in preparation for division (G_1/S), while another checks that DNA replication is complete and that the genome has not been damaged (G_2/M) (Nasmyth, 1996). Cyclins and cyclin-dependent kinases mediate these events, and are themselves regulated through multiple mechanisms (Morgan, 1995).

The biochemical events controlling G_2/M transit in mammalian cells are remarkably similar to those in lower eukaryotes. Mammalian Cdc2 kinase accumulates in S phase (Shimizu et al., 1995) and is regulated by Wee1 kinase (Parker and Piwnica-Worms, 1992) and Cdc25 phosphatase (Lee et al., 1992; Millar et al., 1991). Recent studies also have begun to elucidate the pathways by which G_2 arrest occurs in response to DNA damage in mammalian cells (Nurse, 1997). Chk1 kinase, which is activated by phosphorylation in response to DNA damage, in turn phosphorylates Cdc25C (Sanchez et al., 1997). This potentiates the binding of 14-3-3 to Cdc25C, which prevents Cdc25C from dephosphorylating and activating Cdc2 (Peng et al., 1997). In addition, it has been shown that expression of 14-3-3 is regulated by p53, a DNA binding protein already implicated in $p21^{WAF1/CIP1}$-mediated G_1 arrest (Hermeking et al., 1997).

While G_2/M progression clearly requires the coordinated expression of many gene products, less is known about how these are regulated at the level of transcription. Characterization of transcription factors regulating G_2 progression and mitotic entry would significantly advance our understanding of the mechanisms controlling this portion of the cell cycle.

S. pombe cdc5p was first described as a putative DNA binding protein implicated in G_2/M transit (Nasmyth and Nurse, 1981; Ohi et al., 1994). We subsequently identified a cDNA encoding a protein with limited homology to *S. pombe* cdc5p (Bernstein and Coughlin, 1997). Its widespread expression in human tissues and homology with expressed sequences in other eukaryotes suggested an evolutionarily conserved general function (Bernstein and Coughlin, 1997). Nuclear import upon serum stimulation of mammalian cells suggested a possible role in cell proliferation (Bernstein and Coughlin, 1997).

More recently, we described the function of human Cdc5 (hCdc5) in mammalian cells (Bernstein and Coughlin, 1998). Overexpression of hCdc5 shortened G_2 and reduced cell size. A dominant negative mutant of hCdc5 consisting of the

putative DNA binding domain slowed G_2 progression and delayed entry into mitosis. These results demonstrated that hCdc5 regulates G_2 progression and mitotic entry, and suggested that it functions through DNA binding (Bernstein and Coughlin, 1998).

While hCdc5 appears to positively regulate G_2/M in mammalian cells, its function through DNA binding and transcriptional regulation remains to be confirmed. Effector genes regulated by hCdc5 or other members of the Cdc5 family in *S. cerevisiae* (Ohi et al., 1998), *A. thaliana* (Hirayama and Shinozaki, 1996), *C. elegans* (Bernstein and Coughlin, 1997), *D. melanogaster* (Katzen et al., 1998; Ohi et al., 1998), and *M. musculus* (Bernstein and Coughlin, 1997) have not been identified. hCdc5 and its relatives contain tandem helix-turn-helix DNA binding motifs at their amino termini, similar to that seen in Myb-related proteins (Bernstein and Coughlin, 1997). In contrast with c-Myb, however, Cdc5-related proteins contain only two repeats of the helix-turn-helix motif, whereas Myb family members possess three (Bernstein and Coughlin, 1997; Hirayama and Shinozaki, 1996; Katzen et al., 1998; Ohi et al., 1998; Ohi et al., 1994). Moreover, within this domain hCdc5 and its relatives bear a Val→Leu substitution at a position critical for DNA binding specificity (Carr et al., 1996; Ogata et al., 1996). Cdc5 family members, therefore, likely differ from Myb in their DNA binding properties.

Recently, a 7-base-pair (bp) nucleotide sequence identified by random oligonucleotide binding site selection was shown to interact with the DNA binding domain of *A. thaliana* Cdc5 in vitro; however, binding with this sequence was reduced with non-specific, competitor DNA (Hirayama and Shinozaki, 1996). Similar experiments to identify a consensus binding site for the highly conserved DNA binding domains of *D. melanogaster* Cdc5 and the Cdc5 homologue in *S. cerevisiae*, Cef1p, failed to identify any preferential site, nor did they interact with the 7 bp sequence identified for *A. thaliana* (Ohi et al., 1998). In addition, others have shown that Cdc5 does not activate the transcription of candidate genes known to be upregulated during G_2/M, for example *cdc2* and *String* in *D. melanogaster* (Katzen et al., 1998), *Clb1* and *Swi5* in *S. cerevisiae* (Ohi et al., 1998), and $cdc13^+$ and $cdc25^+$ in *S. pombe* (Ohi et al., 1998). These results raised the question of whether Cdc5 family members serve functions other than as site-specific DNA binding proteins (Ohi et al., 1998), even though they possess a Myb-like DNA binding domain and the human homologue experimentally has been shown to contain a latent transcriptional activation domain (Bernstein and Coughlin, 1998). In addition, recent studies in fission and budding yeast have implicated a role for cdc5p and Cef1p, respectively, in pre-mRNA splicing (McDonald et al., 1999; Tsai et al., 1999), and hCdc5 has been identified as a component of the mammalian splicesome (Burns et al., 1999; Neubauer et al., 1998).

We now report that human Cdc5 may serve as a site-specific, double-stranded DNA binding protein capable of forming high affinity complexes with specific sequences through its DNA binding domain. We employed cyclic amplification and selection from a pool of random oligonucleotides to identify preferential binding sites for the hCdc5 DNA binding domain. A high-affinity, double-stranded, consensus sequence was identified, which contains a helix-turn-helix binding motif. We also showed that this DNA-protein interaction is capable of activating transcription. We then used a selection system in yeast to identify human genomic sequences that bind hCdc5 through its helix-turn-helix domain. These experiments demonstrate that a member of the Cdc5 family interacts with specific sequences through direct DNA-protein interaction. They also suggest that mammalian Cdc5 family members may participate in more than one process necessary for regulated cell division.

MATERIALS AND METHODS

Materials and strains

Yeast expression plasmids pBM2389 and pBM2463, and *S. cerevisiae* strain YM4271 (*MAT**a** ura3-52 his3-_200 ade2-101 lys2-801 trp1-901 gal4-Δ512 gal80-Δ538 ade5::hisG*), were generously provided by Mark Johnston (Washington University School of Medicine).

Expression and purification of recombinant hCdc5

In order to express the DNA binding domain of hCdc5, unique *Eco*RV and *Sal*I sites were inserted into pSK67 (Bernstein and Coughlin, 1997) using site-directed mutagenesis (Kunkel et al., 1987). The 1500 bp cDNA fragment encoding 500 amino acids from the amino terminus of hCdc5 was then subcloned into *Eco*RV/*Sal*I sites in pET29c(+) (Novagen). The resulting plasmid was transformed into *E. coli* strain BL21 (DE3), and individual colonies were inoculated into 500 ml liquid LB medium containing 50 μg/ml kanamycin. Cultures were grown at 37°C to OD_{600} 0.2-0.6. Cultures were induced with isopropyl β-D-thiogalactopyranoside (IPTG) at final concentration of 1 mM for 2 hours. Bacterial pellets were collected by centrifugation and lysed in 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mg/ml lysozyme at 4°C for 1 hour followed by sonication. The lysate supernatant then was added to 0.5 ml 50% Ni-NTA resin (Qiagen) and mixed gently at room temperature for 60 minutes. The resin was washed with 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 20 mM imidazole, and the His_6-tagged protein eluted with 50 mM NaH_2PO_2, pH 8.0, 300 mM NaCl, 70 mM imidazole. The protein then was dialyzed in 10 mM Tris, pH 7.9, 60 mM KCl, 4 mM $MgCl_2$ at 4°C overnight. Dialysed protein was analyzed by SDS-polyacrylamide gel electrophoresis and immunoblot using monoclonal anti-His antibody (Invitrogen) according to previously described methods (Bernstein and Coughlin, 1997).

Oligonucleotide selection

Oligonucleotide selection was performed with a DNA fragment that contained a 15 bp random sequence flanked on either side by 20 bp of nonrandom sequence (5′-CGCTCGAGGGATCCGAATTC(N_{15})T-CTAGAAAGCTTGTCGACGC-3′). The single-stranded oligonucleotide was made double-stranded with *Taq* DNA polymerase and a primer complementary to the 3′ flanking sequence (5′-GCGTCGA-CAAGCTTTCTAGA-3′). After extraction with phenol/chloroform/isoamyl alcohol and precipitation with glycogen/ethanol, 60 picomoles of the double-stranded oligonucleotide pool were incubated with 0.5 μg purified His_6-hCdc5 in binding buffer (10 mM Hepes, pH 7.8, 75 mM KCl, 2.5 mM $MgCl_2$, 3% Ficoll, 40 μg/ml poly(dI-dC), in the presence of 1 μg/ml phenylmethylsulfonyl fluoride, 0.4 ng/ml aprotinin, 1 ng/ml leupeptin, 0.4 ng/ml soybean trypsin inhibitor, 0.4 ng/ml pepstatin A, 10 ng/ml benzamidine) at room temperature for 30 minutes. The binding solution then was incubated with 50% Ni-NTA agarose at room temperature for 10 minutes. After washing the resin extensively with binding buffer, the DNA-protein complexes were eluted with binding buffer containing 250 mM imidazole.

Bound oligonucleotides were amplified with *Taq* polymerase over 15-20 cycles by annealing to flanking primers (5′-GCGTCGAC-AAGCTTTCTAGA-3′ and 5′-CGCTCGAGGGATCCGAATTC-3′) at

55°C for 20 seconds and extending at 72°C for 10 seconds. Subsequent rounds of binding were accomplished as above with 0.1 μg recombinant protein and 1/10th of the unpurified PCR product. After 3-8 cycles of binding, selection and amplification, PCR products were cloned into a TA vector (Invitrogen) and sequenced.

Electrophoretic mobility shift assay

DNA-protein binding was accomplished by incubating 0.8 pmoles recombinant protein with 0.01 pmole double-stranded oligonucleotide, labeled with γ-^{32}P to a specific activity of 10^6 cpm/pmole, in binding buffer at room temperature for 30 minutes. The same molar ratio of protein/DNA was used in all samples. After separation of bound and free DNA by 5% non-denaturing acrylamide gel electrophoresis, DNA binding was quantitated with a Storm 860 Phosphor/Fluor Imager and ImageQuant software (Molecular Dynamics) and calculated as a percentage of DNA bound divided by total DNA (bound/bound+free). Each DNA binding experiment was repeated at least three times, and representative assays are shown.

Equilibrium dissociation constants (K_d) were determined by incubating a constant amount of labeled DNA (0.01 pmole at 10^6 cpm/pmole) with increasing amounts of protein (0.2-2.0 pmoles) at room temperature for 30 minutes. After binding activity was quantitated as described above, the K_d was calculated with the equation K_d=[D][P]/[DP], where [D] is the concentration of free DNA, [P] is the concentration of free protein, and [DP] is the concentration of the DNA-protein complex. The DNA concentration was limiting relative to the protein concentration to allow the approximation [DP]≈[DP$_{total}$].

Transcriptional reporter assay

p5xCdc5 was constructed by cloning five, tandem copies of the 12 bp, hCdc5 binding sequence in place of the *GAL4* binding element in pFR-Luc by PCR using primers 5′-CAAGCTTGCATGCCTGCAGGTGATTTAACATAAGATTTAACATAAGATTTAACATAAGATTAACATAAGATTTAACATAAACTCTAGAG-3′ and 5′-CGTGTACATCGACTGAAATCCC-3′. p3x*GAL4*-3xCdc5 was constructed by cloning three copies of the *GAL4* binding site in tandem with three copies of the hCdc5 binding site in place of the *GAL4* binding site of pFR-Luc using primers 5′-CAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGCCGGAGTACTGTCCTCCGCCGGAGTACTGTCCTCCGCGATTTAACATAAGATTTAACATAAGATTTAACATAAACTCTAGAG-3′ and 5′-CGTGTACATCGACTGAAATCCC3′. pFR-Luc (Stratagene) contains the entire coding region of *Photinus pyralis* luciferase downstream of five tandem repeats of the *GAL4* binding element and a basic promoter (TATATA).

HeLa cells were transiently transfected using LipofectAMINE Plus (Gibco) according to the manufacturer's instructions, and luciferase activity was assayed in whole cell lysates using the Dual-Luciferase Reporter Assay System (Promega) as previously described (Bernstein and Coughlin, 1998). pM3-VP16 (Clontech), encoding a *GAL4*-VP16 fusion, was used as a positive control. A plasmid encoding the Δ675 hCdc5 mutant has previously been described (Bernstein and Coughlin, 1998). pRL-TK (Promega), encoding *Renilla reniformis* luciferase downstream of a Herpes TK promoter, was included in each transfection for normalization of transfection efficiency.

Binding site selection in yeast

The method for selection of human genomic binding sites was adapted from a similar system used to define the binding specificity of NGFI-B (Wilson et al., 1991). The library or UAS (upstream activating sequence)-reporter plasmid was constructed by inserting *Sau*3AI-digested, human genomic DNA into the unique *Bam*HI site of pBM2389 (Liu et al., 1993), upstream of an inactive *GAL1* promoter. The activator plasmid was constructed by cloning sequence encoding the first 120 amino acids of hCdc5 (Bernstein and Coughlin, 1997) into unique *Not*I/*Xho*I sites of pBM2463 (Liu et al., 1993), placing it in-frame between the *LexA* DNA binding domain and *GAL4* activating domain.

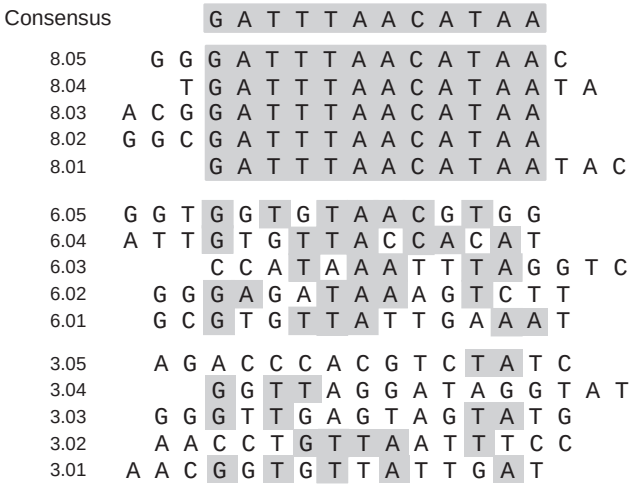


Fig. 1. Selection and amplification of hCdc5 targets. Representative samples of selected 15-bp oligonucleotides after 3 (3.01-3.05), 6 (6.01-6.05) and 8 (8.01-8.05) cycles of selection and amplification demonstrated enrichment for the consensus 12 bp sequence (shaded) with subsequent cycles.

Yeast were manipulated according to standard protocols (Guthrie and Fink, 1991). UAS and activator plasmids were cotransformed into *S. cerevisiae* strain YM4271 (Liu et al., 1993), to allow for red-white selection with *ADE5*, and counterselection with 5-FOA. Non-sectoring, His$^+$ colonies were patched to medium containing 0.1% 5-FOA, and library plasmids that allowed activator-dependent growth in His$^-$ medium, as confirmed by growth failure in 5-FOA, were selected for sequencing.

RESULTS AND DISCUSSION

The DNA binding domain of hCdc5 binds specifically and with high affinity to a 12-bp, double-stranded DNA sequence

To determine whether hCdc5 interacts with specific consensus sequences through its consensus DNA binding domain, we performed selection and amplification from a pool of random oligonucleotides (Blackwell, 1995). We first expressed and purified the amino terminus of hCdc5 (amino acids 1-500) in *E. coli*, then used this to select a preferential binding site from a degenerate pool of 15 bp oligonucleotides flanked by conserved sequences to facilitate amplification and cloning. After 8 cycles, a consensus 12 bp sequence, GATTTAACATAA, was identified (Fig. 1).

We then used an electrophoretic mobility-shift assay to evaluate the DNA-protein interaction (Fig. 2). Competition with unlabeled, consensus oligonucleotide demonstrated that the DNA binding domain of hCdc5 binds the double-stranded target sequence specifically (Fig. 2A). Binding did not occur with single-stranded sense or antisense oligomer (data not shown). Titration with increasing amounts of purified, recombinant protein enabled us to calculate an equilibrium dissociation constant of approx. 10^{-8} M (Fig. 2B). This value is comparable to other helix-turn-helix DNA binding domains, such as that published for c-Myb (Krieg et al., 1995). These experiments

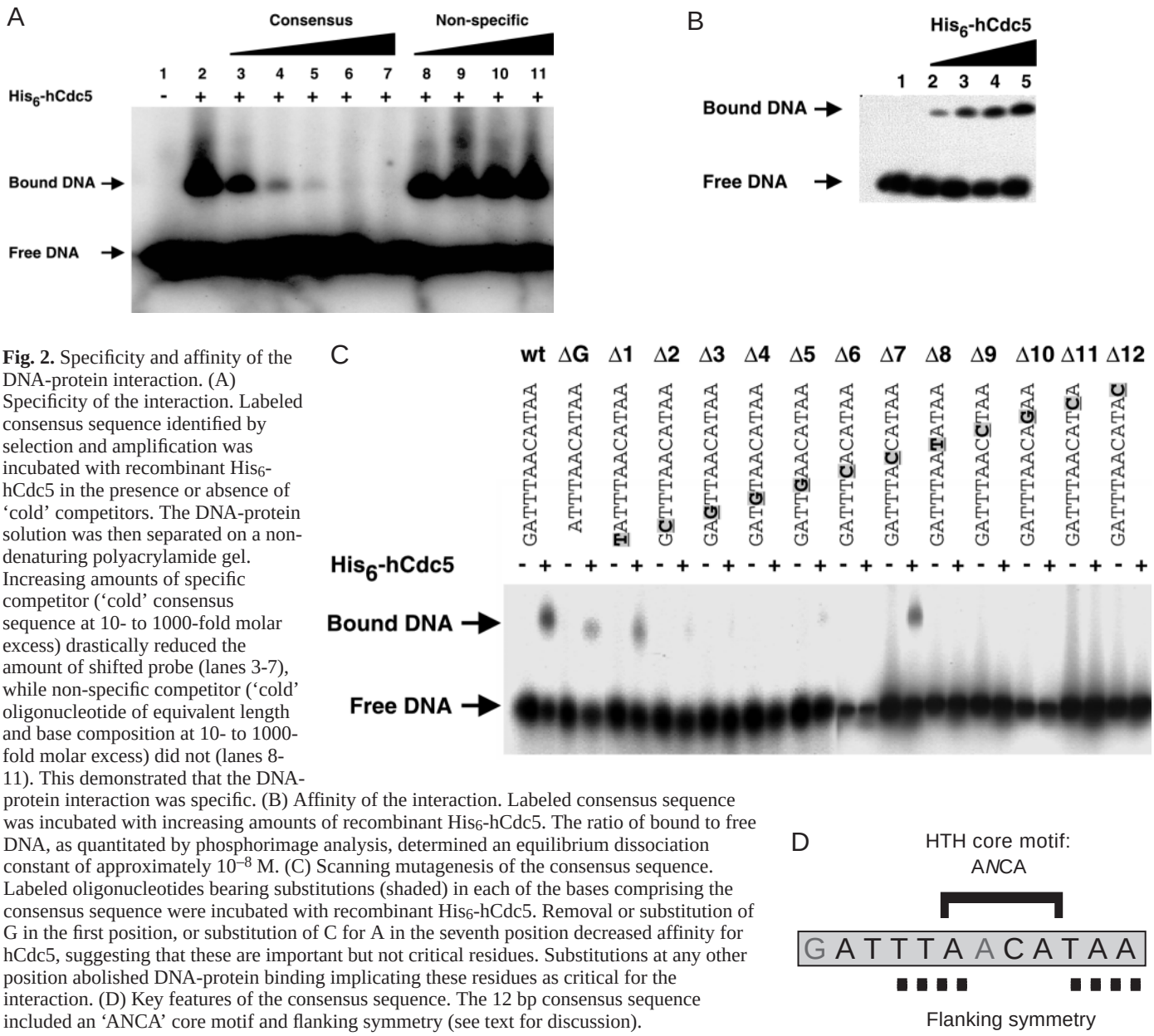


Fig. 2. Specificity and affinity of the DNA-protein interaction. (A) Specificity of the interaction. Labeled consensus sequence identified by selection and amplification was incubated with recombinant His_6-hCdc5 in the presence or absence of 'cold' competitors. The DNA-protein solution was then separated on a non-denaturing polyacrylamide gel. Increasing amounts of specific competitor ('cold' consensus sequence at 10- to 1000-fold molar excess) drastically reduced the amount of shifted probe (lanes 3-7), while non-specific competitor ('cold' oligonucleotide of equivalent length and base composition at 10- to 1000-fold molar excess) did not (lanes 8-11). This demonstrated that the DNA-protein interaction was specific. (B) Affinity of the interaction. Labeled consensus sequence was incubated with increasing amounts of recombinant His_6-hCdc5. The ratio of bound to free DNA, as quantitated by phosphorimage analysis, determined an equilibrium dissociation constant of approximately 10^{-8} M. (C) Scanning mutagenesis of the consensus sequence. Labeled oligonucleotides bearing substitutions (shaded) in each of the bases comprising the consensus sequence were incubated with recombinant His_6-hCdc5. Removal or substitution of G in the first position, or substitution of C for A in the seventh position decreased affinity for hCdc5, suggesting that these are important but not critical residues. Substitutions at any other position abolished DNA-protein binding implicating these residues as critical for the interaction. (D) Key features of the consensus sequence. The 12 bp consensus sequence included an 'ANCA' core motif and flanking symmetry (see text for discussion).

demonstrated for the first time that hCdc5 can bind with high affinity to a specific, double-stranded DNA sequence. It also confirms that the binding specificity for Cdc5 in higher eukaryotes is markedly distinct from even the extended consensus sequences reported for c-Myb, YAACKGHH (Weston, 1992) and YGRCVGTTR (Howe and Watson, 1991).

It is not clear why others were unable to identify preferential binding sites for Cef1p, the putative Cdc5 homologue in *S. cerevisiae* (Ohi et al., 1998). While Cef1p shares sequence homology at its amino terminus with Cdc5 family members, it was unable to complement the temperature-sensitive growth defect of *S. pombe cdc5-120* at its restrictive temperature (Ohi et al., 1998). In addition, human and fission yeast Cdc5 proteins were unable to rescue the lethality of *S. cerevisiae cef1Δ* cells (Ohi et al., 1998). These findings suggest that Cdc5 family members perhaps have multiple or even divergent functions in lower eukaryotes.

The hCdc5 target sequence contains a helix-turn-helix binding motif

To determine which nucleotides within the consensus sequence were necessary for protein binding, we next performed scanning mutagenesis of the sequence (Fig. 2C). This revealed that the core 'ANCA' motif, commonly found in the binding site for other Myb-related, helix-turn-helix DNA binding proteins (McIntosh et al., 1998; Oda et al., 1998; Ording et al., 1996; Sala et al., 1999; Suzuki et al., 1998) was essential for binding. It also demonstrated that the symmetrical, flanking sequences, TTA/TAA, increased binding affinity (Fig. 2C).

To further investigate whether DNA binding was specific for the helix-turn-helix motif, as suggested by the 'ANCA' core sequence, we expressed a truncation of the hCdc5 amino terminus (amino acids 1-120), containing only the helix-turn-helix domain (Bernstein and Coughlin, 1997), as well as a Cdc5 mutant in which the helix-turn-helix motif was disrupted

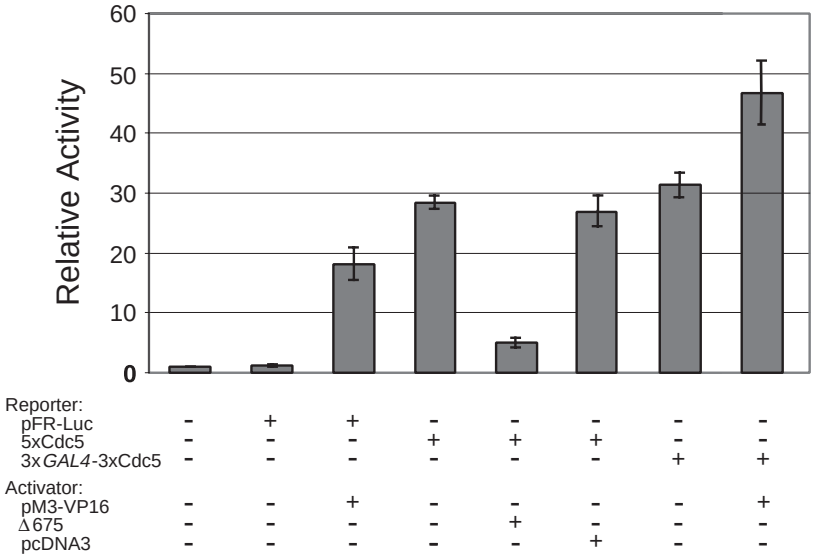


Fig. 3. Interaction of endogenous hCdc5 with its binding site activates transcription. HeLa cells were transfected with plasmids containing the *GAL4* binding element (pFR-Luc), the hCdc5 binding element (p5xCdc5) or both (p3x*GAL4*-3xCdc5), upstream of the *P. pyralis* luciferase gene, in combination with plasmids encoding a *GAL4* binding-VP16 activating domain fusion (pM3-VP16), a dominant negative hCdc5 mutant (Δ675), or empty vector. All cells were cotransfected with pRL-TK, which constitutively expresses *R. reniformis* luciferase. Data shown are mean ratios of *P. pyralis* luciferase to *R. reniformis* luciferase activity ± s.e.m. (n=5). A 28-fold increase in normalized luciferase activity was seen in cells transfected with the reporter plasmid containing the hCdc5 binding element. This activity was abolished by coexpression of mutant Δ675, but significantly increased from a reporter plasmid containing both hCdc5 and *GAL4* binding sites in the presence of the *GAL4* binding-VP16 activating domain fusion protein. These data suggested that endogenous hCdc5 was capable of activating transcription through interaction with its selected binding sequence.

by W→G substitutions at positions 31, 53 and 82 in the human peptide sequence (Bernstein and Coughlin, 1997). These substitutions replace the core tryptophan residues essential to the helix-turn-helix structure, and correspond to the W33G W52G W84G *cef1* mutant previously reported as unable to rescue growth of a *cef1-Δ1::HIS3* mutant (Ohi et al., 1998). While the helix-turn-helix domain alone bound the target sequence with the same specificity and affinity as the longer, amino-terminal peptide (amino acids 1-500), the W31G W53G W82G mutant did not bind the consensus sequence by mobility shift assay (data not shown). Together with the data demonstrating that the 'ANCA' core sequence was essential for hCdc5 binding (Fig. 2C), these results suggest that binding is specific for the helix-turn-helix motif.

Our observation that a T→G substitution at position 4 or A→C substitution at position 12 abolished the DNA-protein interaction suggested that the symmetry of the flanking sequences was important. To determine whether the specific sequences or the symmetry provided by the palindrome was critical for binding, we tested additional mutants that created new palindromes flanking the core sequence, namely GAT*A*TAACATA*T* and GAT*G*TAACATA*C*. Neither of these mutants bound Cdc5 in a mobility-shift assay (data not shown), suggesting that the specific, palindromic sequence was critical for binding.

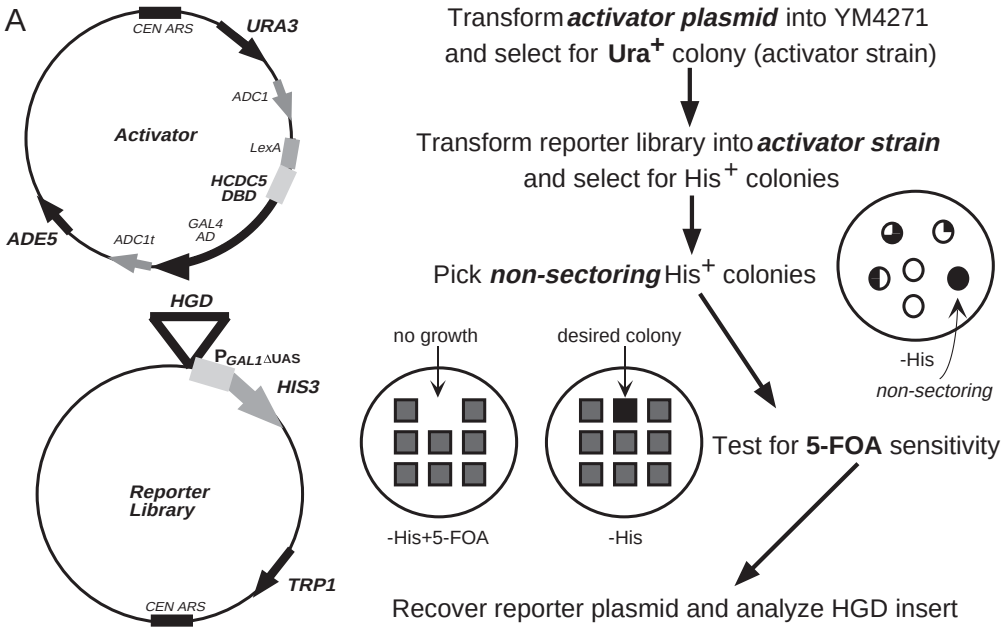
The presence of a core, helix-turn-helix binding motif in the target sequence is consistent with a DNA binding function for hCdc5. The symmetry observed in the sequence suggests that hCdc5 may adopt a higher order structure in binding to its target. We have demonstrated that specific domains of hCdc5 can homodimerize in vitro (X.-H. Lei and H. S. Bernstein, unpublished observations); however, whether dimerization is important for hCdc5 function in vivo remains to be demonstrated.

hCdc5 binding to its target sequence can activate transcription

To test whether interaction between hCdc5 and its binding site could activate transcription, we transfected HeLa cells (known to express hCdc5; Bernstein and Coughlin, 1997) with a plasmid (p5xCdc5) containing the hCdc5 binding site upstream of a luciferase reporter (Fig. 3). Luciferase activity was increased 28-fold in HeLa cells transfected with the plasmid containing the hCdc5 binding site, while activity was not detected in cells transfected with the same plasmid in which the binding site for *GAL4* was substituted for the hCdc5 binding site (pFR-Luc). To determine whether binding by endogenous hCdc5 was responsible for the observed reporter activity, we coexpressed a dominant-negative hCdc5 mutant (Bernstein and Coughlin, 1998) containing only the amino-terminal DNA binding domain (Δ675). We previously demonstrated that this mutant was transcriptionally inactive, and delayed G_2/M in transfected cells (Bernstein and Coughlin, 1998). In HeLa cells, the Δ675 mutant abrogated reporter activity, presumably by competing with endogenous hCdc5 for binding to the consensus site, and blocking activation of the reporter. These data demonstrated that specific interaction of endogenous hCdc5 with its selected binding sequence is capable of activating transcription.

hCdc5 interacts preferentially with specific sequences in the human genome

To determine whether hCdc5 binds specific sites in the human genome, we employed a selection system in yeast that previously was used to define the binding specificity of the nerve growth factor induced-B gene product (Wilson et al., 1991). We constructed two plasmids to express in *S. cerevisiae* (Fig. 4A). The library or UAS (upstream activating sequence)-reporter plasmid consisted of the *HIS3* gene placed downstream of an inactive *GAL1* promoter missing the *GAL1* UAS (Lorch and Kornberg, 1985). We then inserted size-selected, human genomic DNA (1-4 kb) upstream of the inactive *GAL1* promoter as a source of potential DNA binding sites. The activator plasmid encoded an epitope-tagged fusion of the amino terminus of hCdc5 with the activating domain of *GAL4* (Brent and Ptashne, 1985) under control of a strong, constitutive promoter (i.e., *ADH1*). Yeast cells expressing the hybrid transcriptional activator and carrying a library plasmid



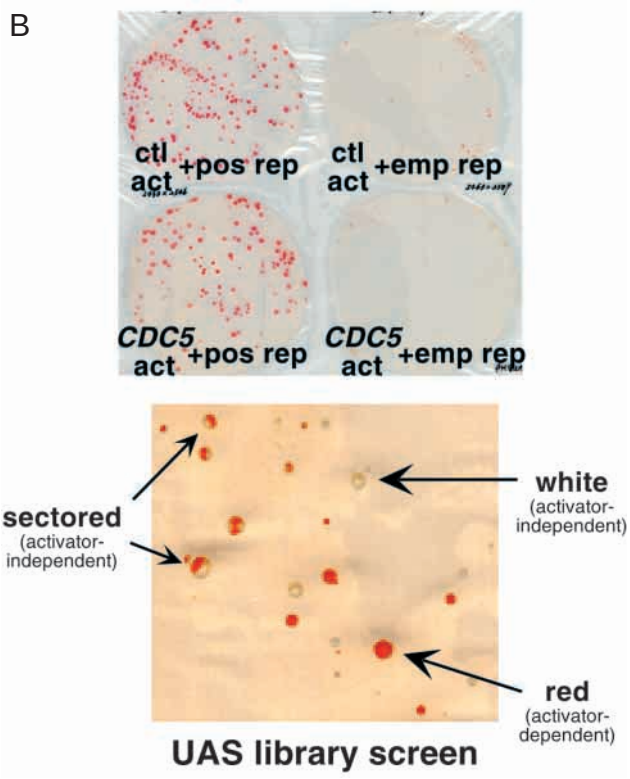


Fig. 4. Genomic binding site selection in yeast. (A) Strategy for selection (see text for discussion). (B) Upper panel, control transformations. *S. cerevisiae* strain YM4271 was cotransformed with control activator (ctl act) or hCdc5 activator (*CDC5* act) and positive control reporter (pos rep) or empty reporter (emp rep) plasmids. Ctl act+pos rep transformants grow as red (activator-dependent) colonies on His^- medium due to interaction between *LexA* DNA binding domain on activator and *lexO* site on reporter. Similarly, *CDC5*+pos rep transformants are red due to *LexA*/*lexO* interaction, and confirm that the *GAL4* activating domain is expressed in-frame with the *LexA*/hCdc5 DNA binding domain fusion. Predominantly white (activator-independent) colonies are seen with negative control transformations (i.e. emp rep). Lower panel, typical library screen. Colony growth on His^- medium of human genomic UAS library cotransformed with *CDC5* act plasmid is shown. Mostly activator-independent (sectored or white) colonies, with few activator-dependent (red) colonies, were seen. The latter were counterscreened for sensitivity to 5-FOA, as described in the text.

containing binding sites for hCdc5 expressed *HIS3* and were able to grow in histidine-deficient medium (Fig. 4A,B).

To avoid the anticipated expression of *HIS3* as a result of UAS recognition by an endogenous yeast transcriptional activator, we took advantage of the *ADE5* and *URA3* markers on the activator plasmid. Yeast colonies bearing null mutations in the *ade2* and *ade5* genes turn red in the presence of *ADE5* (Liu et al., 1993). Colonies that were able to grow in histidine-deficient medium in the absence of the activator plasmid formed red colonies with white sectors, indicating loss of the activator plasmid. We discarded such 'sectoring' colonies as false positives (Fig. 4B). In addition, yeast expressing the *URA3* gene product die in the presence of 5-fluoro-orotic acid (5-FOA), which is metabolized to a toxic metabolite in the presence of the enzyme encoded by *URA3* (Sikorski and Boeke, 1991). Yeast transformants that required the presence of the activator plasmid for function of the library UAS grew in histidine-deficient medium only in the absence of 5-FOA. We also discarded transformants that grew in histidine-deficient medium in the presence of 5-FOA as false positives.

Of approx. 5×10^5 transformants screened (representing approximately one-third of the human genome), 12 out of 218 non-sectoring colonies were sensitive to 5-FOA. These were sequenced and analyzed for homology with known sequences, as well as with each other. All 12 plasmids contained novel human genomic sequence without homology to any sequences available in GenBank, including the Eukaryotic Promoter Database. A subset of these had short regions of limited homology (<500 bp) between them, and two in this subset appeared to be overlapping clones from the same genomic region with >3kb of overlap. The observation that selected clones had extended regions of homology suggested that hCdc5 interacts preferentially with certain sequences within the human genome.

Interaction of hCdc5 with specific sites in the human genome is mediated by binding to its target sequence

To determine whether hCdc5 interacts with specific genomic sequences through binding to its consensus target sequence, we first surveyed the 12 genomic sequences (totaling approx.

A

Clone	Protein binding	Sequence
Consensus	+	GATTTAACATAA
6	+	AATAAAATCAAAATT
30	+	AATAAAATCAAAATT
12(1)	+	AAAGGGGAACACTTT
12(2)	-	CCCCCCTTTAAACCAGCGTGGAGGGGGG
25	-	AATTCCCCGGATCATTGCAAACAATT
8	-	AATGAACGAATCAAATT

B

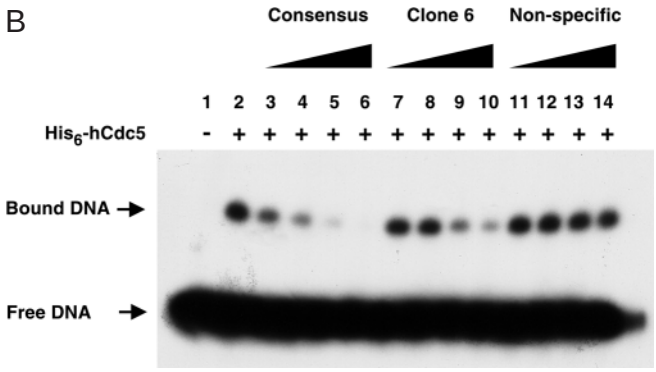


Fig. 5. hCdc5 binds specifically to low abundance sequences in the human genome. (A) Potential binding sites in human genomic clones. Human genomic fragments isolated by binding site selection in yeast (see Fig. 4) were sequenced. Sites containing an 'ANCA' core flanked by palindromes were identified in five clones as shown. Sites found in clones 6, 30 and 12 bound hCdc5, as determined by gel-shift assay (see Fig. 5B). (B) Specific interaction between genomic sequence and hCdc5. Labeled sequence from clone 6, identified by binding site selection in yeast (see Fig. 5A), was incubated with recombinant His_6-hCdc5 in the presence or absence of 'cold' competitors. Increasing amounts of specific competitors ('cold' clone 6 sequence or consensus sequence at 1- to 30-fold molar excess) drastically reduced the amount of shifted probe from clone 6 (lanes 3-6, 7-10), while non-specific competitors ('cold' oligonucleotide of equivalent length and base composition at 1- to 30-fold molar excess) did not (lanes 11-14). Similar data were obtained for the sequence from clone 30 (not shown). This demonstrated that the genomic DNA-protein interaction was specific, and suggested that the consensus sequence could be refined to include an 'AWCA' core flanked by AT-rich palindrome within 3 bp of the core (see Fig. 5A).

36,200 bp) identified by binding site selection in yeast for the consensus sequence identified by cyclic amplification and selection (Fig. 2E). An exact match was not found, although this was not surprising as the likelihood of finding a unique 12 bp sequence is 6×10^{-8}, and is predicted to occur by chance approx. 180 times in the human genome.

We then used the information obtained by scanning mutagenesis (Fig. 2C,D) to look for sites containing the core 'ANCA' motif flanked by AT-containing palindromes within the genomic clones. Six such sites were identified in clones 6, 8, 12, 25 and 30 (Fig. 5A), of which three were able to interact with the amino terminus of hCdc5, as demonstrated by electrophoretic mobility gel-shift assay (Fig. 5B). Two of these were found in the two clones that represented overlapping sequences from the same region of the genome (clones 6 and 30). All three sites shared features of the consensus binding sequence: a core 'ANCA' motif, AT-rich flanking palindrome, and a distance between core and palindrome of no more than 3 bp.

To test whether the sequence we abstracted from clone 6 was the one responsible for Cdc5/*GAL4*-dependent activation in yeast, we cloned subfragments of clone 6 into the UAS-reporter plasmid, and tested their ability to confer activator-dependent growth in selective media (data not shown). A 500 bp subfragment containing the identified 15 bp sequence was necessary for Cdc5/*GAL4*-dependent activation in yeast, while removing the 15 bp sequence from this larger fragment abolished activator-dependent growth. Smaller fragments containing the 15 bp sequence failed to confer growth in selective media, suggesting that while the 15 bp sequence is necessary for activation, other elements contained in the 500 bp clone 6 subfragment may be required for Cdc5/*GAL4*-dependent activation in yeast.

These experiments suggested that the binding site for hCdc5 contains a core 'AWCA' flanked within 3 bp by an AT-rich region of symmetry, although deduction of the 'real' consensus binding sequence for hCdc5 will require the identification of its downstream target genes and the analysis of their promoters. This remains, however, the first demonstration that a member of the Cdc5 family can interact with specific genomic sequences through direct DNA-protein binding.

The role of Cdc5 in cell division

Genetic and biochemical studies of Cdc5 family members in fission yeast, fly and human confirm a positive regulatory role at the G_2/M transition in dividing cells (Bernstein and Coughlin, 1998; Ohi et al., 1998; Ohi et al., 1994). The presence of a consensus DNA binding domain, and experiments with the human protein that have identified a latent transcriptional activation domain, suggest that Cdc5 might function in the transcriptional regulation of gene expression (Bernstein and Coughlin, 1998; Ohi et al., 1998).

We now demonstrate that the amino terminus of hCdc5 binds specifically and with high affinity to a 12 bp, double-stranded DNA sequence, and that this sequence contains a helix-turn-helix binding motif, as well as rotational symmetry. We demonstrate that binding of this sequence by endogenous hCdc5 in cell extracts is capable of activating transcription. We also demonstrate that hCdc5 may interact with specific sequences in the human genome by binding to its consensus target sequence. These studies provide the first evidence that human Cdc5 can act as a site-specific DNA binding protein.

Recently, studies with hCdc5 and homologues in *S. cerevisiae* and *S. pombe* have shown that members of this protein family associate with the pre-mRNA splicing complex, or 'spliceosome' (Burns et al., 1999; McDonald et al., 1999; Tsai et al., 1999). The spliceosome is known to consist of small nuclear ribonucleic acids (Varani and Nagai, 1998) in complex with a number of essential, associated proteins (Staley and Guthrie, 1998). These proteins have been

postulated to function in a variety of regulatory roles, including substrate specificity, splice site recognition, catalytic activity and developmental and tissue-specific control of alternative splicing (Lopez, 1998; Will and Luhrmann, 1997). The heterogeneous nuclear ribonucleoprotein complex also has been implicated in transcriptional regulation through site-specific DNA binding (Krecic and Swanson, 1999). The specific role of Cdc5 in this complex, or the mechanism by which spliceosome function regulates the G_2/M transition, are not yet known.

Our previous finding that overexpression of hCdc5 accelerates G_2/M (Bernstein and Coughlin, 1998) is difficult to explain solely from its putative role in pre-mRNA splicing. In addition, we have now identified a high-affinity, consensus DNA sequence that specifically binds hCdc5 and appears at low frequency in the human genome. We also have found this sequence in the 5′ untranslated regions of several cell cycle regulatory protein genes in both *D. melanogaster* (*Uba1*; Canning and Finnegan, 1999) and *mts* (Berry and Gehring, 1999; Berkeley Drosophila Genome Project) and *C. elegans* (*cya-2* and *cyb-2*; Kreutzer et al., 1995; Durbin and Mieg (1996). Taken together, these observations leave open the possibility that Cdc5 family members have multiple, perhaps convergent functions, including the control of cell division through both transcriptional and post-transcriptional mechanisms.

This work was support by grants from the National Institutes of Health, the American Heart Association National Center, the March of Dimes Birth Defects Foundation, and the Child Health Research Center at UCSF to H.S.B. X.H.L. was supported by a postdoctoral fellowship from the National Institutes of Health. We thank Dr Mark Johnston for providing us with plasmids, yeast strains and advice; Dr Raghavendra Mirmira for sharing technical expertise; Drs Shaun Coughlin, Ira Herskowitz, Synthia Mellon, Michael German and Brian Black for helpful discussions; and the anonymous referees for their insightful comments on the manuscript.

REFERENCES

Bernstein, H. S. and Coughlin, S. R. (1997). *Pombe* Cdc5-related protein. A putative human transcription factor implicated in mitogen-activated signaling. *J. Biol. Chem.* **272**, 5833-5837.

Bernstein, H. S. and Coughlin, S. R. (1998). A mammalian homolog of fission yeast Cdc5 regulates G_2 progression and mitotic entry. *J. Biol. Chem.* **273**, 4666-4671.

Berry, M. and Gehring, W. J. (1999). The phosphorylation status of the SCR homeodomain determines its functional activity: Essential role for protein phosphatase 2A, B′. *Eur. Dros. Res. Conf.* **16**, 268.

Blackwell, T. K. (1995). Selection of protein binding sites from random nucleic acid sequences. *Meth. Enzymol.* **254**, 604-618.

Brent, R. and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729-736.

Burns, C. G., Ohi, R., Krainer, A. R. and Gould, K. L. (1999). Evidence that Myb-related Cdc5 proteins are required for pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA* **96**, 13789-13794.

Canning, M. and Finnegan, D. J. (1999). The role of ubiquitination in *Drosophila* development. *Eur. Dros. Res. Conf.* **16**, 152.

Carr, M. D., Wollborn, U., McIntosh, P. B., Frenkiel, T. A., McCormick, J. E., Bauer, C. J., Klempnauer, K. H. and Feeney, J. (1996). Structure of the B-Myb DNA-binding domain in solution and evidence for multiple conformations in the region of repeat-2 involved in DNA binding: Implications for sequence-specific DNA binding by Myb proteins. *Eur. J. Biochem.* **235**, 721-735.

Durbin, R. and Mieg, J. T. (1996). *A* C. elegans *database* (ACEDB). [Online.] http:www-hgc.lbl.gov/Genome/macace.html. [8 November 2000, last date accessed.]

Guthrie, C. and Fink, G. (1991). Guide to yeast molecular genetics and molecular biology. In *Methods in Enzymology*, vol. 194, 933 pp. San Diego: Academic Press, Inc.

Hermeking, H., Lengauer, C., Polyak, K., He, T.-C., Zhang, L., Thiagalingam, S., Kinzler, K. and Vogelstein, B. (1997). 14-3-3σ is a p53-regulated inhibitor of G_2/M progression. *Mol. Cell* **1**, 3-11.

Hirayama, T. and Shinozaki, K. (1996). A *cdc5*$^+$ homolog of a higher plant, *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **93**, 13371-13376.

Howe, K. M. and Watson, R. J. (1991). Nucleotide preferences in sequence-specific recognition of DNA by C-Myb protein. *Nucleic Acids Res.* **19**, 3913-3919.

Katzen, A. L., Jackson, J., Harmon, B. P., Fung, S. M., Ramsay, G. and Bishop, J. M. (1998). *Drosophila* Myb is required for the G_2/M transition and maintenance of diploidy. *Genes Dev.* **12**, 831-843.

Krecic, A. M. and Swanson, M. S. (1999). hnRNP complexes: Composition, structure, and function. *Curr. Opin. Cell Biol.* **11**, 363-371.

Kreutzer, M. A., Richards, J. P., De Silva-Udawatta, M. N., Temenak, J. J., Knoblich, J. A., Lehner, C. F. and Bennett, K. L. (1995). *Caenorhabditis elegans* cyclin A- and B-type genes: A cyclin a multigene family, an ancestral cyclin B3 and differential germline expression. *J. Cell Sci.* **108**, 2415-2424.

Krieg, J., Oelgeschlager, M., Janknecht, R. and Luscher, B. (1995). High affinity DNA binding of native full length C-Myb and differential proteolytic sensitivity of its N- and C-terminal domains. *Oncogene* **10**, 2221-2228.

Kunkel, T. A., Roberts, J. D. and Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Meth. Enzymol.* **154**, 367-382.

Lee, M. S., Ogg, S., Xu, M., Parker, L. L., Donoghue, D. J., Maller, J. L. and Piwnica-Worms, H. (1992). *cdc25*$^+$ encodes a protein phosphatase that dephosphorylates p34^{cdc2}. *Mol. Biol. Cell.* **3**, 73-84.

Liu, J., Wilson, T., Milbrandt, J. and Johnston, M. (1993). Identifying DNA-binding sites and analyzing DNA-binding domains using a yeast selection system. *Methods* **5**, 125-137.

Lopez, A. J. (1998). Alternative splicing of pre-mRNA: Developmental consequences and mechanisms of regulation. *Ann. Rev. Genet.* **32**, 279-305.

Lorch, Y. and Kornberg, R. D. (1985). A region flanking the *gal7* gene and a binding site for *GAL4* protein as upstream activating sequences in yeast. *J. Mol. Biol.* **186**, 821-824.

McDonald, W. H., Ohi, R., Smelkova, N., Frendewey, D. and Gould, K. L. (1999). Myb-related fission yeast *cdc5p* is a component of a 40S snRNP-containing complex and is essential for pre-mRNA splicing. *Mol. Cell. Biol.* **19**, 5352-5362.

McIntosh, P. B., Frenkiel, T. A., Wollborn, U., McCormick, J. E., Klempnauer, K. H., Feeney, J. and Carr, M. D. (1998). Solution structure of the B-Myb DNA-binding domain: A possible role for conformational instability of the protein in DNA binding and control of gene expression. *Biochem.* **37**, 9619-9629.

Millar, J. B., Blevitt, J., Gerace, L., Sadhu, K., Featherstone, C. and Russell, P. (1991). p55^{cdc25} is a nuclear protein required for the initiation of mitosis in human cells. *Proc. Natl. Acad. Sci. USA* **88**, 10500-10504.

Morgan, D. O. (1995). Principles of CDK regulation. *Nature* **374**, 131-134.

Murray, A. and Hunt, T. (1993). *The Cell Cycle.* Oxford: Oxford University Press.

Nasmyth, K. (1996). Viewpoint: Putting the cell cycle in order. *Science* **274**, 1643-1645.

Nasmyth, K. and Nurse, P. (1981). Cell division cycle mutants altered in DNA replication and mitosis in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* **182**, 119-124.

Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A. and Mann, M. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex [see comments]. *Nature Genet.* **20**, 46-50.

Nurse, P. (1997). Checkpoint pathways come of age. *Cell* **91**, 865-867.

Oda, M., Furukawa, K., Ogata, K., Sarai, A. and Nakamura, H. (1998). Thermodynamics of specific and non-specific DNA binding by the C-Myb DNA-binding domain. *J. Mol. Biol.* **276**, 571-590.

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., Nishimura, Y., Ishii, S. and Sarai, A. (1996). The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Biol.* **3**, 178-187.

Ohi, R., Feoktistova, A., McCann, S., Valentine, V., Look, A. T., Lipsick, J. S. and Gould, K. L. (1998). Myb-related *Schizosaccharomyces pombe cdc5p* is structurally and functionally conserved in eukaryotes. *Mol. Cell. Biol.* **18**, 4097-4108.

Ohi, R., McCollum, D., Hirani, B., Den Haese, G. J., Zhang, X., Burke, J. D., Turner, K. and Gould, K. L. (1994). The *Schizosaccharomyces pombe cdc5*$^{+}$ gene encodes an essential protein with homology to C-Myb. *EMBO J.* **13**, 471-483.

Ording, E., Bergholtz, S., Brendeford, E. M., Jamin, N. and Gabrielsen, O. S. (1996). Flexibility in the second half-site sequence recognised by the C-Myb R2 domain – in vitro and in vivo analysis. *Oncogene* **13**, 1043-1051.

Parker, L. L. and Piwnica-Worms, H. (1992). Inactivation of the p34^{cdc2}-cyclin B complex by the human Wee1 tyrosine kinase. *Science* **257**, 1955-1957.

Peng, C.-Y., Graves, P. R., Thoma, R. S., Wu, Z., Shaw, A. S. and Piwnica-Worms, H. (1997). Mitotic and G_2 checkpoint control: Regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science* **277**, 1501-1505.

Sala, A., Saitta, B., De Luca, P., Cervellera, M. N., Casella, I., Lewis, R. E., Watson, R. and Peschle, C. (1999). B-Myb transactivates its own promoter through SP1-binding sites. *Oncogene* **18**, 1333-1339.

Sanchez, Y., Wong, C., Thoma, R. S., Richman, R., Wu, Z., Piwnica-Worms, H. and Elledge, S. J. (1997). Conservation of the Chk1 checkpoint pathway in mammals: Linkage of DNA damage to CDK regulation through Cdc25. *Science* **277**, 1497-1501.

Shimizu, M., Ichikawa, E., Inoue, U., Nakamura, T., Nakajima, T., Nojima, H., Okayama, H. and Oda, K. (1995). The G_1/S boundary-specific enhancer of the rat *cdc2* promoter. *Mol. Cell. Biol.* **15**, 2882-2892.

Sikorski, R. and Boeke, J. (1991). *In vitro* mutagenesis and plasmid shuffling: From cloned gene to mutant yeast. In *Guide to Yeast Molecular Genetics and Molecular Biology*, vol. 194 (ed. C. Guthrie and G. Fink), pp. 302-318. San Diego: Academic Press, Inc.

Staley, J. P. and Guthrie, C. (1998). Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92**, 315-326.

Suzuki, A., Wu, C. Y., Washida, H. and Takaiwa, F. (1998). Rice Myb protein OSMYB5 specifically binds to the AACA motif conserved among promoters of genes for storage protein glutelin. *Plant Cell Physiol.* **39**, 555-559.

Tsai, W. Y., Chow, Y. T., Chen, H. R., Huang, K. T., Hong, R. I., Jan, S. P., Kuo, N. Y., Tsao, T. Y., Chen, C. H. and Cheng, S. C. (1999). Cef1p is a component of the Prp19p-associated complex and essential for pre-mRNA splicing. *J. Biol. Chem.* **274**, 9455-9462.

Varani, G. and Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing. *Ann. Rev. Biophys. Biomol. Struct.* **27**, 407-445.

Weston, K. (1992). Extension of the DNA binding consensus of the chicken C-Myb and V-Myb proteins. *Nucl. Acids Res.* **20**, 3043-3049.

Will, C. L. and Luhrmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9**, 320-328.

Wilson, T. E., Fahrner, T. J., Johnston, M. and Milbrandt, J. (1991). Identification of the DNA binding site for NGFI-B by genetic selection in yeast. *Science* **252**, 1296-300.