946–952 *Nucleic Acids Research, 1994, Vol. 22, No. 6*

Evidence for a protein domain superfamily shared by the cyclins, TFIIB and RB/p107

Toby J.Gibson*, Julie D.Thompson, Ariel Blocker and Tony Kouzarides[1]
European Molecular Biology Laboratory, Postfach 102209, Meyerhofstrasse 1, D-69012 Heidelberg, Germany and [1]Wellcome/CRC Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

Received January 6, 1994; Revised and Accepted February 11, 1994

ABSTRACT

Cyclins, TFIIB and RB play major roles in cell cycle and/or gene regulation. Earlier work has suggested common ancestry for the TFIIB repeats and RB pocket B which share 20% sequence identity. We now report that database searches with profiles based on a multiple alignment of cyclin core regions (the 'cyclin box') detect the TFIIB repeats with equivalent scores to divergent cyclins. Several features of the sequences support the notion of common ancestry: e.g. cyclins A/B, C and D share ~20–30% identity but each have ~15–20% identity with vertebrate TFIIB, showing that conserved cyclin features underlie the match. These results suggest the presence of a domain superfamily, which we term the TR domain, in nuclear regulatory proteins belonging to the TFIIB, cyclin and RB families, that has been duplicated many times during eukaryotic evolution. The TR domain appears to function in protein–protein interactions.

INTRODUCTION

A recurring problem in molecular biology is the evaluation of weak but genuine similarities between proteins. In a classic example, many TIM barrell enzymes have no detectable sequence similarity yet equivalent positioning of the active sites strongly imply common ancestry (1). Biological similarity and score ranking order can both help to justify borderline hits from database searches, but are themselves unreliable yardsticks. The types of matched residues are critical, since most conserved positions in globular domains are core-forming hydrophobic residues. Well known spurious matches include highly charged regions picking up the charge-rich coiled-coil tails of myosin, and proline-rich or hydrophilic random coil regions scoring highly against each other regardless of common ancestry. In multidomain proteins, it is vital for hydrophobic matches to span the full length of structural units. The prediction of homology between the peptide-binding domains of hsp70 and HLA-A2 illustrates these points (2).

Multiple sequence alignments, together with the PAM250 amino acid replacement matrix (3), are often used to prepare family specific matrices of residue preference, known as profiles (4), for comparison to sequence databases using the Smith–Waterman local similarity algorithm (5). We have modified profiles to upweight sequences by divergence (6) using the branch lengths of Neighbour-Joining trees (7). The weights increased sensitivity as did a newer substitution matrix, BLOSUM62 (8), and excision of INDELs (sites of insertion and deletion) while marking their positions with reduced gap penalties. *De novo* searches with RNA-associated KH domains (9) and with signalling protein PH domains (10,11) both revealed many new examples (12–14). We now report results stemming from profile searches with the cyclin family.

Cyclins (gene name *CYC*, *CCN* or *CLN*) were first discovered as proteins that gradually accumulate prior to mitosis, only to be abruptly destroyed (15). Subsequently several subtypes have been catalogued, reaching peak abundance at either the G1/S or the G2/mitosis transition control points in the cell cycle. Most, perhaps all, cyclins associate with cdc2-like kinases to form functional complexes which are thought to play major roles in cell cycle control. A- and E-type cyclins have also been found (with cdc2-related kinases) in complexes with transcription regulators, such as the adenovirus E1A protein, the protein product of the retinoblastoma gene, RB, and its relative p107, as well as the transcription factors E2F and DRTF1 (16–21). These associations imply that cyclins are involved directly in gene regulation.

Sequence similarity between cyclins extends over ~200 residues, the 'cyclin box'. Different families can have unrelated N- and/or C-terminal extensions. Within the cyclin box, divergence is so extensive that only a single Glu residue remains absolutely conserved. Many cyclins share only 10–20% identity, e.g. yeast CLN3 and human cyclin C score 12.1% identity, and the extremes drop below 10%. The difficulty of alignment is compounded by large insertions in several cyclin boxes, e.g. 77 residues in the case of yeast CLN1. Therefore, in preparing a cyclin alignment, profiles were used as aids to bring in the more divergent sequences.

Rather surprisingly, the cyclin profiles ranked TFIIB database entries highly. TFIIB is a general transcription factor and one of the mediators linking the TATA-binding protein TBP and RNA

*To whom correspondence should be addressed polymerase II (reviewed by 22,23). While the similarity, at about 15–20% identity, does not guarantee relatedness, certain characteristics of the match point to common ancestry. In this manuscript, we outline the logic underpinning the proposed TFIIB/cyclin homology and discuss implications for the roles and evolution of these two protein families, together with the retinoblastoma protein (RB) family found earlier to possess 20% sequence similarity to the TFIIB repeat (24).

MATERIALS AND METHODS

Cyclin box alignment

Cyclins were extracted by keyword searches with SRS (25) from the SWISSPROT v. 26 database (26) and by database searches with FASTA (27) or the EMBL BLITZ network service (28). CLUSTALV (29) was used to provide initial alignments of the most related cyclins, which were visually corrected in the GDE multiple alignment editor (S.Smith, Harvard University) guided by dotplots (6) and the sequences themselves. Gaps were aligned on gaps so that only a single gap was allowed for each INDEL region. INDELs closer than seven amino acids apart could always be imposed, in accordance with observed INDEL behaviour in aligned homologous structures (30). More divergent sequences were subsequently brought in using either CLUSTALV or profiles prepared from previously aligned cyclins. In this manner, 37 cyclins were reliably aligned, but for others, uncertainties remained that precluded these sequences from contributing to profiles.

Profile searches

Profiles were prepared by the program PROFILEWEIGHT (6) using the BLOSUM62 substitution matrix (8) and excision of positions with more than 80% gaps. Three different sequence weighting methods were employed. The sequences were given either: 1, equivalent weights; 2, branch-proportional tree-based weights; or 3, the weighting system of Altschul *et al.* (31) which penalises sequences that are more distant from the tree root, and by implication, from the common ancestor.

The GCG programs PROFILESEARCH, PROFILESEGMENTS and PROFILEGAP were used for protein database searches and profile-to-sequence alignments (4,32). TPROFILESEARCH was used to search 6-frame translations of DNA databases (P.Rice, EMBL, unpublished). Default normalisations for amino acid compositions and sequence length were turned off. Output scores are dependent on length of the aligned sequences, residue composition, gap penalties and choice of substitution matrix. As with all sequence searches, probability scores cannot unambiguously distinguish weak true hits from noise and the ranking order can be helpful.

Dotplots using profiles

The program PROPLOT produces diagonal similarity plots for visual inspection using any combination of sequence or profile (6). To compare a sequence against a profile, the residue in each position is used to look up the score in the profile. To compare a profile against a profile, it is desirable to assess the relative scores for each amino acid at every position of the profiles. This is done by calculating the normalised r.m.s. deviation of the individual amino acid frequencies between any two positions in the profiles (6). User-specified windows and cutoffs determine which points are actually plotted.

Single sequence database search comparisons

These were conducted against SWISSPROT using the EMBL BLITZ network service (28). BLITZ compares sequences using MPsrch (J.Collins and S.Sturrock, Edinburgh) which conducts a full Smith–Waterman (5) local similarity search. A search of SWISSPROT with a 200 amino acid query sequence takes under 1 minute. MPsrch is implemented on a MASPAR parallel computer. The scoring system is similar to that of earlier software developed by the Edinburgh group for an ICL DAP and offered as a service accessible by network (33). The BLOSUM62 matrix together with gap penalty 8 were used in all searches.

Availability of programs

PROFILEWEIGHT, TPROFILESEARCH and PROPLOT can be obtained from the EMBL fileserver (28).

RESULTS

Detection of TFIIB sequences by 'cyclin box' profiles

A sequence-weighted profile from 37 aligned cyclins was used to search the SWISSPROT protein sequence database. Several TFIIB entries ranked highly in the output list, with scores comparable to the most divergent cyclins (Table I). The output local alignments revealed that the cyclin box profile was aligned to the two divergent ~90 residue repeats in TFIIB. The scores were substantially based upon matches between hydrophobic positions including the most conserved hydrophobic segments. These favourable features warranted further investigation. Therefore profiles were also prepared using the Altschul tree-based weighting scheme which, in addition to upweighting by long branch length, favours sequences which are closer to the tree root relative to those which are distant (31). This weighting scheme can be more sensitive in detecting related sequences which diverged before the common ancestor of the aligned set (6). The

Table I. Scores and ranking positions resulting from searching the SWISSPROT database with cyclin profiles

	Unweighted Profile		Tree-Weighted Profile		Altschul-Weighted Profile		Entry is in	SwissProt Accession
Entry	Position	Score	Position	Score	Position	Score	Profile	Number
Cgb1_human	1	81.87	1	71.75	5	48.59	*	P14635
Cg2a_human	15	67.77	16	63.40	3	49.92	*	P20248
cg2b_medsa	26	61.79	25	58.52	17	45.27	*	P30278
Cgd1_human	34	45.50	35	42.59	27	43.01	*	P24385
Cg1e_human	35	45.35	33	44.89	31	41.40	*	P24864
Cg1c_human	48	17.38	48	17.69	41	34.37	*	P24863
Cg1c_drome	47	18.73	46	18.64	39	35.20	*	P25008
Cg13_yeast	43	26.49	43	27.62	45	28.76	*	P13365
Cg12_yeast	44	20.45	45	19.75	47	16.47		P20438
Cg11_yeast	45	20.41	44	19.97	46	17.28		P20437
Cg16_yeast	53	12.08	52	11.14	50	9.01		P24867
Cg17_yeast	184	9.49	85	8.80	53	7.44		P25693
Tf2b_xenla	49	12.74	49	11.85	49	9.03		P29054
Tf2b_rat	50	12.43	50	11.54	51	8.60		P29053
Tf2b_human	51	12.43	51	11.54	52	8.60		Q00403
Tf2b_drome	85	10.29	69	9.26	59	6.92		P29052
Tf2b_yeast	>1000	5.91	>1000	5.31	>1000	4.41		P29055
Tf3b_yeast	>1000	7.78	>1000	6.87	>1000	4.81		P29056
Tf2b_pyrwo[1]	>1000	7.07	>1000	6.73	198	5.92		P29095
Pw TFIIB[2]	(962)	8.23	(609)	7.28	(198)	5.92		
Kl TFIIB[2]	(186)	9.48	(321)	7.72	(>1000)	4.86		
Top false hit	52	12.11	53	10.45	54	7.43		

[1] Partial sequence of *Pyrococcus woesei* TFIIB, missing part of repeat 1.

[2] The equivalent ranking and scores are given for two sequences, not yet in SWISSPROT, the completed sequence of *Pyrococcus woesei* TFIIB and the sequence of *Kluyveromyces lactis* TFIIB.

Altschul-weighted cyclin profiles detected TFIIB sequences more strongly than the branch-proportional weighted profile (Table I). An unweighted profile was worse than either weighting scheme at detecting both TFIIB and divergent cyclin entries (Table I).

Scores depend on TFIIB sequence divergence

Seven TFIIB sequences are known; three vertebrates, *Drosophila*, the budding yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis* (34) and an archaebacterium *Pyrococcus woesei* (35,36). An eighth sequence, yeast TFIIIB (also called BRF-1), is an RNA polymerase III activator with ~25% identity to the vertebrate TFIIB sequences (37). The ranking of the TFIIB sequences by the cyclin profiles varied quite widely (Table I), which might have implied that the match was spurious but could also be a consequence of variation in TFIIB molecular clocks (38).

The region common to all TFIIB sequences consists of two divergent ~90 residue repeats (e.g. sharing 18% identity in human TFIIB). An alignment of the ~180 residue repeat regions from the TFIIB sequences was prepared, from which a Neighbour-Joining tree was calculated (7). The tree has a branching order consistent with phylogenetic expectation (Figure 1). Branch length variations show that the yeasts, especially *S.cerevisiae*, have incorporated the most substitutions while the archaebacterial sequence has changed least. If TFIIB and cyclins share a common ancestor in a proto-eukaryote, the sequences most similar to the cyclin branch point can be estimated (by measuring the branch lengths) as being animals, followed by *Pyrococcus*. If however the last common ancestor was in archaebacteria, *Pyrococcus* should be closest, followed by animals. The ranking of the TFIIB scores in Table I is animals > *Pyrococcus* > yeast spp. As well as being completely consistent with the distances in the tree, these scores may imply an origin of the cyclins after the archaebacterial/eukaryotic split.

Table II. Scores and ranking positions resulting from searching the SWISSPROT database with TFIIB profiles

	Unweighted Profile		Tree-Weighted Profile		Altschul-Weighted Profile		Entry is in Profile	SwissProt Accession Number
Entry	Position	Score	Position	Score	Position	Score		
Tf2b_xenla	1	102.35	1	82.15	1	79.20	*	P29054
Tf2b_rat	2	101.90	2	81.41	2	78.32	*	P29053
Tf2b_human	3	101.90	3	81.41	3	78.32	*	Q00403
Tf2b_drome	4	97.99	4	79.17	4	76.93	*	P29052
Tf2b_yeast	5	73.94	5	74.99	6	65.69	*	P29055
Tf3b_yeast	7	53.77	6	64.68	5	71.80	*	P29056
Tf2b_pyrwo	6	56.99	7	59.18	7	63.52	*	P29095
Cg2a_bovin	8	8.74	8	8.05	10	7.82		P30274
Cg2a_human	9	8.74	9	8.05	11	7.82		P20248
Cg2a_drome	10	8.29	15	7.24	21	6.85		P14785
Cg2b_medsa	11	8.06	33	6.41	42	6.22		P30274
Cgd1_human	35	6.77	12	7.30	8	7.87		P24385
Cgd1_mouse	40	6.68	14	7.25	9	7.85		P25322
Cg1c_drome	29	6.92	17	7.15	14	7.29		P25008
Cg1c_human	37	6.72	23	6.68	20	6.85		P24863
P107_human	24	6.99	29	6.52	23	6.32		P28749
Rb_human	227	5.81	122	5.72	97	5.80		P06400
Rb_mouse	465	5.54	433	5.21	206	5.37		P13405
Top false hit	12	8.05	10	7.38	12	7.36		
No. Cyclins[1]	24		30		33			

[1] Number of cyclin entries in top 100 hits.

Table III. Scores and ranking positions[1] resulting from searching the SWISSPROT database by MPsrch using individual cyclin box and TFIIB repeat region sequences

Search Sequence:	human cyclin C		alfalfa cyclin B		yeast CLN 3		yeast HCS26		*Xenopus* TFIIB		*Dros.* TFIIB	
Residues:	56-251		102-281		107-317		50-268		114-293		113-292	
No. of detections	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Cyclins in top 50	43	7	43	7	22	28	8	42	19	31	11	39
Cyclins in top 100	44	6	44	6	38	12	9	41	21	29	16	34
TFIIBs in top 50	2	5	4	3	0	7	0	7	7	0	7	0
TFIIBs in top 100	5	2	4	3	0	7	0	7	7	0	7	0
Entry	Pos.	Score	Pos.	Score	Pos.	Score	Pos.	Score	Pos.	Score	Pos.	Score
Cgd1_human	1	815	35	257	71	115	-	(<106)	20	128	33	111
Cg2b_medsa	11	257	1	912	51	121	-	(<106)	8	161	9	132
Cg13_yeast	61	119	74	121	1	1071	-	(<106)	-	(<101)	-	(<101)
Cg16_yeast	-	(<109)	-	(<114)	-	(<107)	1	1127	-	(<101)	-	(<101)
Cg17_yeast	-	(<109)	-	(<114)	-	(<107)	2	414	-	(<101)	-	(<101)
Tf2b_xenla	48	130	43	165	-	(<107)	-	(<106)	1	892	4	767
Tf2b_rat	53	124	45	155	-	(<107)	-	(<106)	2	859	2	777
Tf2b_human	54	124	46	155	-	(<107)	-	(<106)	3	859	3	777
Tf2b_drome	66	116	48	134	-	(<107)	-	(<106)	4	767	1	899
Tf2b_pyrwo	-	(<109)	-	(<114)	-	(<107)	-	(<106)	5	308	5	284
Tf2b_yeast	-	(<109)	-	(<114)	-	(<107)	-	(<106)	6	277	6	276
Tf3b_yeast	47	135	-	(<114)	-	(<107)	-	(<106)	7	218	7	201
Top False Hit	42	152	47	139	4	154	3	155	9	158	8	144

[1] Ranking positions are given only for entries in the top 100 hits. Scores in parentheses are those for the 100th ranked hit in a search which defines the upper limit for the possible entry score.

Profile searches using aligned TFIIB sequences

Profiles were prepared for the core TFIIB repeats. These embody less information than the cyclin profiles as there are only seven TFIIB sequences (since rat and human are identical) and they are less divergent. This is reflected in much higher self scores for TFIIB entries relative to all non-TFIIB entries in the SWISSPROT database.

All three TFIIB profiles ranked the seven TFIIB entries top, followed directly by two to four cyclins (Table II). Of the 50 cyclin entries, 24 were ranked in the top 100 by the unweighted profile, 30 by the tree-weighted profile and 33 by the Altschul tree-weighted profile, correllating with the increased sensitivity of the weighted profiles. Highly divergent cyclins were not detected by the TFIIB profiles—which might have implied spurious matches.

Depending on the profile, the top scoring cyclins were either bovine cyclin A or human cyclin D. Cyclin C is also consistently well detected (Table II). These three cyclin classes range between 15–30% identity, compared with each other. The detection of three classes of cyclins, not in themselves closely related, is consistent with the expansion of the cyclin family after the divergence of TFIIB and cyclins from a shared common ancestor.

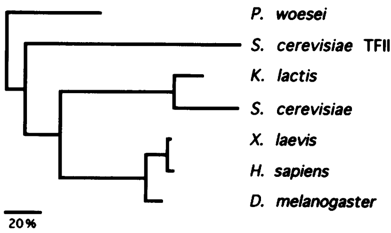


Figure 1. Neighbour-Joining tree for TFIIB sequences, using only the repeated regions. The tree was calculated in CLUSTALV (29) using pairwise distances between sequences corrected for multiple substitutions (54). Rat TFIIB is identical to human. The tree is shown rooted at an arbitrary point on the archaebacterial branch. If the TFIIIB sequence diverged before the archaebacteria/eukaryote split, the root would be on its branch. Given either root, the tree is consistent with phylogenetic expectation. Horizontal branch lengths in NJ trees are proportional to sequence divergence. The branch length variation ranks the sequences for the number of incorporated mutations in the descending order *S.cerevisiae* > *K.lactis* > animals > *P.woesei*. The deep nodes correspond to events whose times are not accurately known, therefore explicit values for substitution rates cannot be safely inferred from the tree.

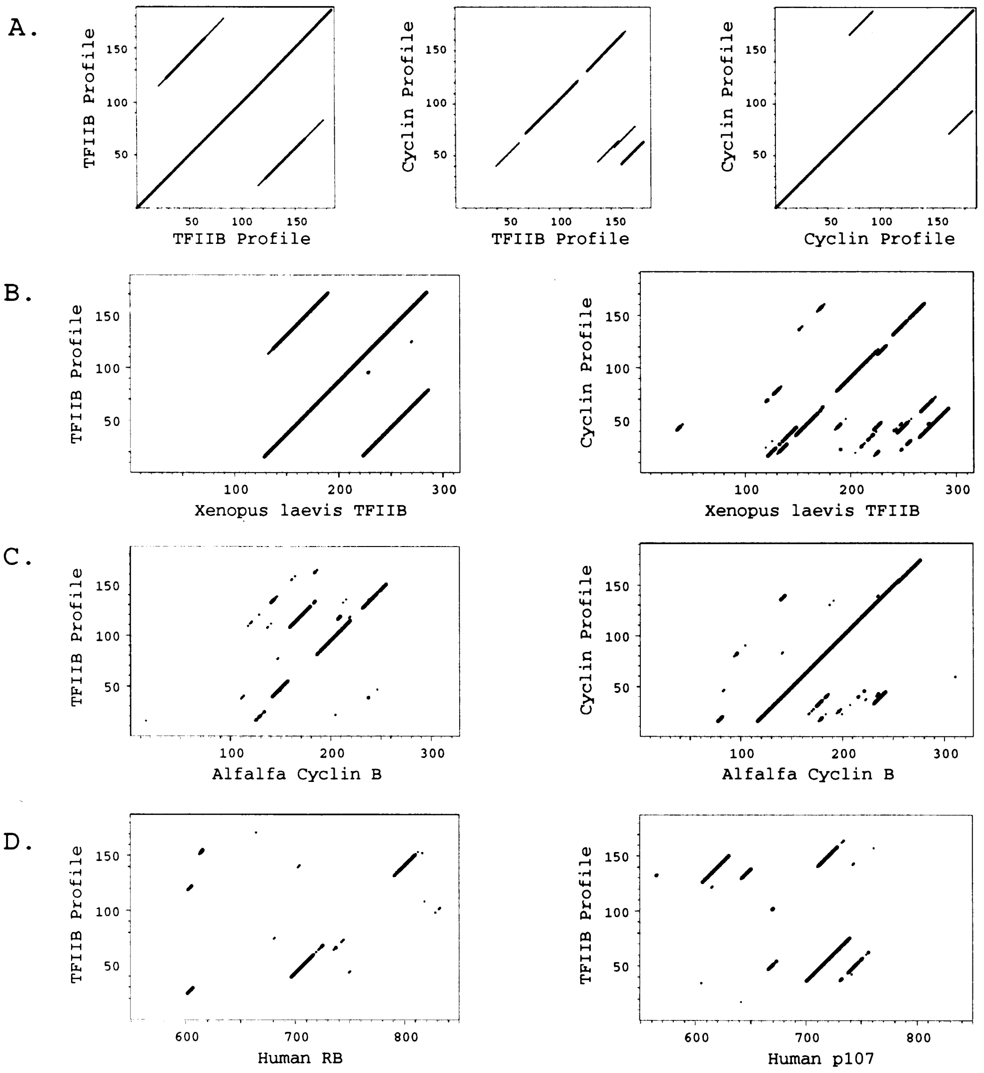


Figure 2. Dot matrix comparisons with TFIIB and cyclin sequence-weighted profiles and cyclin, TFIIB and RB/p107 sequences. (**A**) Profile v. Profile dotplots for TFIIB v. TFIIB, TFIIB v. cyclin and cyclin v. cyclin. Normalised r.m.s.d. values between profile columns were summed over 21 residue windows. Thick lines were plotted for the top 0.05% of scores, thinner lines for the top 0.10%. The triple diagonal pattern in the symmetrical TFIIB self-comparison is the signature of a twice repeated sequence. The TFIIB v. cyclin plot shows an interrupted central diagonal, suggesting colinearity of the profiles, and a high scoring segment between TFIIB repeat 2 and the proposed cyclin repeat 1. The cyclin self-comparison shows a short segment of similarity between the C-termini of the proposed repeats. (**B**) *Xenopus* TFIIB plotted against the TFIIB and cyclin profiles. Values were taken from the profile matrix according to residues in the sequence and summed over a window of 31 residues. Large dots are the top 0.05% of the scores, small dots the top 0.10%. Comparison with the TFIIB profile reveals the repeated region. Comparison with the cyclin profile reveals an interrupted diagonal spanning the TFIIB repeats, supporting colinearity of these regions. (**C**) Alfalfa cyclin B plotted against the TFIIB and cyclin profiles, with the same parameters. Comparison with the cyclin profile reveals the cyclin box. The TFIIB profile shows an interrupted diagonal spanning the cyclin box. (**D**) The TFIIB profile plotted against residues 550–850 of the RB and p107 sequences with the same parameters. Common to both plots is a single strong diagonal centred on the most conserved region of TFIIB repeat 1. There is no indication of a second repeat in RB/p107.

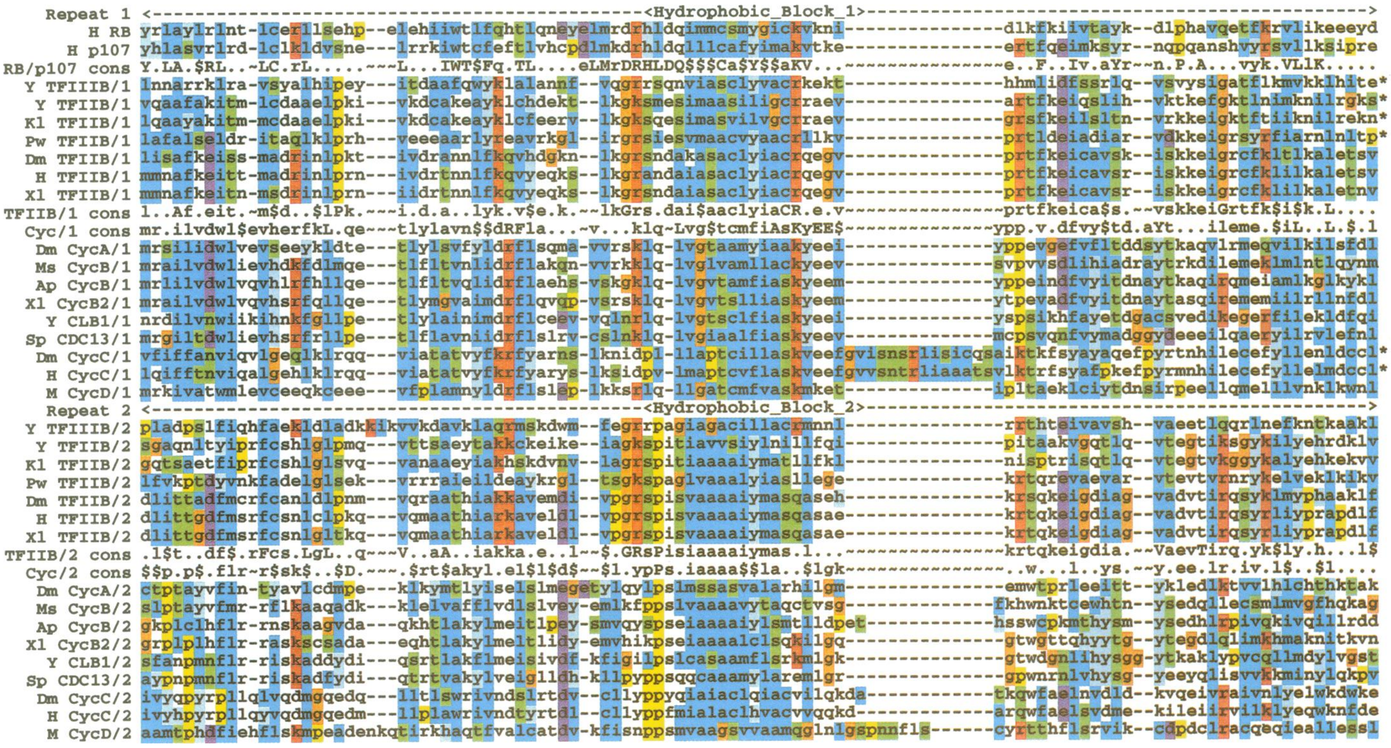
Reciprocal detection in searches with single sequences

The *Xenopus* TFIIB repeat region was compared against SWISSPROT using MPsrch (28,33). The seven TFIIB entries were ranked top, with cyclins constituting eleven of the following thirteen entries and nineteen of the top 50 (Table III). High scoring cyclins were again G1-specific cyclin D and G2-specific cyclins A/B, themselves only ~25% identical. This indicates that it is conserved features of these cyclins which are being detected. Again, none of the likely highly diverged cyclins were detected (Table III).

Reciprocal database searches were conducted with two cyclins, alfalfa cyclin B and human cyclin D, that ranked highly in the TFIIB searches and two cyclins which ranked poorly, yeast CLN3 and HCS26. The cyclin B search ranked *Xenopus* and mammalian TFIIB entries higher than seven known cyclins and above the first false positive (Table II). The cyclin D search scored TFIIBs nearly as highly. By contrast, the searches with the highly divergent cyclins failed to detect any TFIIBs (Table II). Moreover, HCS26 is much worse at detecting other cyclins than are the animal TFIIBs, with only 8 of the top 50 sequences being cyclins.

Dotplots using TFIIB and cyclin profiles

Smith–Waterman (5) database searches routinely output a single best hit per sequence entry. Therefore dotplots, which show high



```
     Repeat 1 <---------------------------------------------<Hydrophobic_Block_1>------------------------------------------------->
         H RB yrlaylrlnt-lcerllsehp--elehiiwtlfqhtlqneyelmrdrhldqimmcsmygickvkni---------------dlkfkiivtayk--dlphavqetfkrvlikeeeyd
       H p107 yhlasvrlrd-lclkldvsne---lrrkiwtcfeftlvhcpdlmkdrhldqlllcafyimakvtke---------------ertfqeimksyr--nqpqanshvyrsvllksipre
 RB/p107 cons Y.LA.$RL..~LC.rL.....~~~L...IWT$Fq.TL....eLMrDRHLDQ$$$Ca$Y$$aKV...~~~~~~~~~~~~~~~e..F..Iv.aYr~~n.P.A...vyk.VLlK.....
  Y TFIIIB/1 lnnarrklra-vsyalhipey---itdaafqwyklalannf--vqgrrsqnviasclyvacrkekt---------------hhmlidfssrlq--vsvysigatflkmvkklhite*
   Y TFIIB/1 vqaafakitm-lcdaaelpki---vkdcakeayklchdekt--lkgksmesimaasiligcrraev---------------artfkeiqslih--vktkefgktlnimknilrgks*
  Kl TFIIB/1 lqaayakitm-mcdaaelpki---vkdcakeayklcfeerv--lkgksqesimasvilvgcrraev---------------grsfkeilsltn--vrkkeigktftiiknilrekn*
  Pw TFIIB/1 lafalseldr-itaqlklprh---veeeaarlyreavrkgl--irgrsiesvmaacvyaacrllkv---------------prtldeiadiar--vdkkeigrsyrfiarnlnltp*
  Dm TFIIB/1 lisafkeiss-madrinlpkt---ivdrannlfkqvhdgkn--lkgrsndakasaclyiacrqegv---------------prtfkeicavsk--iskkeigrcfkltlkaletsv
   H TFIIB/1 mmnafkeitt-madrinlprn---ivdrtnnlfkqvyeqks--lkgrandaiasaclyiacrqegv---------------prtfkeicavsr--iskkeigrcfklilkaletsv
  Xl TFIIB/1 mmnafkeitn-msdrinlprn---iidrtnnlfkqvyeqks--lkgrsndaiasaclyiacrqegv---------------prtfkeicavsr--iskkeigrcfklilkaletnv
TFIIB/1 cons l..Af.eit.~m$d..$lPk.~~~i.d.a..lyk.v$e.k.~~lkGrs.dai$aaclyiaCR.e.v~~~~~~~~~~~~~~~prtfkeica$s.~~vskkeiGrtfk$i$k.L....
  Cyc/1 cons mr.ilvdwl$evherfkL.qe~~~tlylavn$$dRFla...~v...klq~Lvg$tcmfiAsKyEE$~~~~~~~~~~~~~~ypp.v.dfvy$td.aYt...ileme.$iL..L.$.l
  Dm CycA/1 mrsilidwlvevseeykldte---tlylsvfyldrflsqma-vvrsklq-lvgtaamyiaakyeei--------------yppevgefvfltddsytkaqvlrmeqvilkilsfdl
   Ms CycB/1 mrailvdwlievhdkfdlmqe---tlfltvnlidrflakqn-vvrkklq-lvglvamllackyeev--------------svpvvsdlihiadraytrkdilemeklmlntlqynm
   Ap CycB/1 mrlilvdwlvqvhlrfhllqe---tlfltvqlidrflaehs-vskgklq-lvgvtamfiaskyeem--------------yppeindfvyitdnaytkaqirqmeiamlkglkykl
 Xl CycB2/1 mrailvdwlvqvhsrfqllqe---tlymgvaimdrflqvqp-vsrsklq-lvgvtslliaskyeem--------------ytpevadfvyitdnaytasqirememiilrllnfdl
    Y CLB1/1 nrdilvnwiikihnkfgllpe---tlylainimdrflceev-vqlnrlq-lvgtsclfiaskyeei--------------yspsikhfayetdgacsvedikegerfilekldfqi
 Sp CDC13/1 mrgiltdwlievhsrfrllpe---tlflavniidrflslrv-cslnklq-lvgiaalfiaskyeev--------------mcpsvqnfvymadggydeeeilqaeryilrvlefnl
   Dm CycC/1 vfiffanviqvlgeqlklrqq---viatatvyfkrfyarns-lknidpl-llaptcillaskveefgvisnsrlisicqsaiktkfsyayaqefpyrtnhilecefyllenldccl*
    H CycC/1 lqifftnviqalgehlklrqq---viatatvyfkrfyarys-lksidpv-lmaptcvflaskveefgvvsntrliaaatsvlktrfsyafpkefpyrmnhilecefyllelmdccl*
    M CycD/1 mrkivatwmlevceeqkceee---vfplamnyldrflslep-lkksrlq-llgatcmfvaskmket--------------ipltaeklciytdnsirpeellqmelllvnklkwnl
     Repeat 2 <---------------------------------------------<Hydrophobic_Block_2>------------------------------------------------->
  Y TFIIIB/2 pladpslfiqhfaekldladkkikvvkdavklaqrmskdwm--fegrrpagiagacillacrmnnl---------------rrthteivavsh--vaeetlqqrlnefkntkaakl
   Y TFIIB/2 sgaqnltyiprfcshlglpmq---vttsaeytakkckeike--iagkspitiavvsiylnillfqi---------------pitaakvgqtlq--vtegtiksgykilyehrdklv
  Kl TFIIB/2 gqtsaetfiprfcshlglsvq---vanaaeyiakhskdvnv--lagrspitiaaaaiymatllfkl---------------nisptrisqtlq--vtegtvkggykalyehkekvv
  Pw TFIIB/2 lfvkptdyvnkfadelglsek---vrrraieildeaykrgl--tsgkspaglvaaalyiasllege---------------krtqrevaevar--vtevtvrnrykelveklkikv
  Dm TFIIB/2 dlittadfmcrfcanldlpnm---vqraathiakkavemdi--vpgrspisvaaaaiymasqaseh---------------krsqkeigdiag--vadvtirqsyklmyphaaklf
   H TFIIB/2 dlittgdfmsrfcsnlclpkq---vqmaathiarkaveldl--vpgrspisvaaaaiymasqasae---------------krtqkeigdiag--vadvtirqsyrliyprapdlf
  Xl TFIIB/2 dlittgdfmsrfcsnlgltkq---vqmaathiarkaveldl--vpgrspisvaaaaiymasqasae---------------krtqkeigdiag--vadvtirqsyrliyprapdlf
TFIIB/2 cons .l$t..df$.rFcs.LgL..q~~~V..aA..iakka.e..l~~$.GRsPisiaaaaiymas.l...~~~~~~~~~~~~~~~krtqkeigdia.~~VaevTirq.yk$ly.h...l$
  Cyc/2 cons $$p.p$.flr~r$sk$..$D.~~~.$rt$akyl.el$l$d$~~$l.ypPs.iaaaa$$la..$lgk~~~~~~~~~~~~~~~..w...l..ys.~~y.ee.lr.iv.l$..$l....
   Dm CycA/2 ctptayvfin-tyavlcdmpe---klkymtlyiselslmegetylqylpslmssasvalarhilgm---------------emwtprleeitt--ykledlktvvlhlchthktak
   Ms CycB/2 slptayvfmr-rflkaaqadk---klelvafflvdlslvey-emlkfppslvaaaavytaqctvsg--------------fkhwnktcewhtn--ysedqllecsmlmvgfhqkag
   Ap CycB/2 gkplclhflr-rnskaagvda---qkhtlakylmeitlpey-smvqyspseiaaaaiylsmtlldpet------------hsswcpkmthysm--ysedhlrpivqkivqillrdd
 Xl CycB2/2 grplplhflr-raskscsada---eqhtlakylmeltlidy-emvhikpseiaaaalclsqkilgq---------------gtwgttqhyytg--ytegdlqlimkhmaknitkvn
    Y CLB1/2 sfanpmnflr-riskaddydi---qsrtlakflmeisivdf-kfigilpslcasaamflsrkmlgk---------------gtwdgnlihysgg-ytkaklypvcqllmdylvgst
 Sp CDC13/2 aypnpmnflr-riskadfydi---qtrtvakylveiglldh-kllpyppsqqcaaamylaremlgr---------------gpwnrnlvhysg--yeeyqlisvvkkminylqkpv
   Dm CycC/2 ivyqpyrpllqlvqdmgqedq---lltlswrivndslrtdv--cllyppyqiaiaclqiacvilqkda------------tkqwfaelnvdld--kvqeivraivnlyelwkdwke
    H CycC/2 ivyhpyrpllqyvqdmgqedm---llplawrivndtyrtdl--cllyppfmialaclhvacvvqqkd-------------arqwfaelsvdme--kileiirvilklyeqwknfde
    M CycD/2 aamtphdfiehflskmpeadenkqtirkhaqtfvalcatdv-kfisnppsmvaagsvvaamqglnlgspnnfls------cyrtthflsrvik--cdpdclracqeqieallessl
```

Figure 3. Aligned sequences from seven TFIIB homologues, a representative set of cyclin boxes and RB/p107 pocket B, colour coded to highlight the similarity both within and between the repeats. The critical conserved hydrophobic block underpinning the similarity is delineated. Insertions between repeats 1 and 2 occur in some sequences, indicated by *. A consensus is given for each family repeat. Upper case indicates strongly conserved residues and $, conserved hydrophobicity. Residue colouring is used to highlight important features. All Gly (orange) and Pro (yellow) residues are coloured. Other colouring is by conserved property occuring in more than 40% of a column: uncoloured residues lack a sufficiently conserved property. Blue, hydrophobic; light blue, partially hydrophobic; red, positive; purple, negative; green, hydrophilic. Species are coded: Ap, *Arbacia punctata*; Dm, *Drosophila melanogaster*; H, human; Kl, *Kluyveromyces lactis*; M, mouse; Ms, *Medico sativa*; Pw, *Pyrococcus woesei*; Sp, *Schizosaccharomyces pombe*; Xl, *Xenopus laevis*; Y, yeast. The figure was prepared with the GDE multiple alignment editor (S.Smith, Harvard University) in conjunction with COLORMASK (J.Thompson, unpublished) providing POSTSCRIPT output for a colour laser printer.

scoring segments wherever they occur, are useful to provide an overview of matching subsegments. Figure 2A shows cyclin and TFIIB profile-to-profile dotplots (6). The TFIIB/cyclin comparison detects five high scoring segments, of which three are close to the central diagonal, together extending over most of the profiles. This behaviour indicates that similar amino acids tend to be preferred at many equivalent positions throughout the cyclin box and TFIIB repeats, consistent with the proposal that these regions are colinear. Dotplots comparing *Xenopus* TFIIB and alfalfa cyclin sequences against the two profiles are shown in Figure 2B,C. In both cases interrupted diagonals from the cross-comparisons superimpose on stronger diagonals in the self-comparisons supporting colinearity of the cyclin box and TFIIB repeats. The dotplot comparisons show that the similarity between cyclins and TFIIB fulfils the requirement that domains must match throughout their length to be considered to be homologous.

Merging the TFIIB and cyclin alignments

In order to examine the potential fit of the cyclin box and TFIIB repeat region further, it was necessary to merge the aligned sets. The merge was guided by ensuring that the highest matching segments of the dotplots and profile search outputs were aligned and that INDELs were superimposed. Counterparts for conserved blocks in the alignments were present. Each block was checked individually for the best fit. The most strongly matching blocks were aligned by the most conserved identical residues. Less well matched blocks were aligned to maximise the superposition of columns that behaved similarly: Conserved hydrophobic columns that were likely to be core forming were superimposed, as were the least conserved positions, expected to be solvent exposed.

Figure 3 shows the TFIIB repeat sequences and a representative set of cyclin box sequences aligned throughout their lengths. The most conserved motif in the TFIIB repeats is a block that has strong hydrophobic preference for eleven consecutive residues, seven of which also have a strong preference for small sidechains like alanine. The two most highly conserved motifs in the cyclins show very similar patterns. The first conserved cyclin block also has eleven consecutive hydrophobic positions of which six have a preference for small residues that superimpose on equivalent preferences in TFIIB repeat 1. Also superimposed in this region are several positions with charged or aromatic preferences. The second conserved cyclin block has nine consecutive hydrophobic residues, of which five have a preference for small residues that superimpose on equivalent residues in TFIIB repeat 2. Also superimposed in this region are positions with hydrophobic, aromatic or proline preferences. The two motifs in the cyclin box can themselves be aligned (Figure 3) but are more similar to the respective TFIIB motifs than they are to each other, in agreement with the colinearity suggested by the dotplots. These motifs are indicated in the alignment in Figure 3 which superimposes both the TFIIB repeats and the putative cyclin repeats.

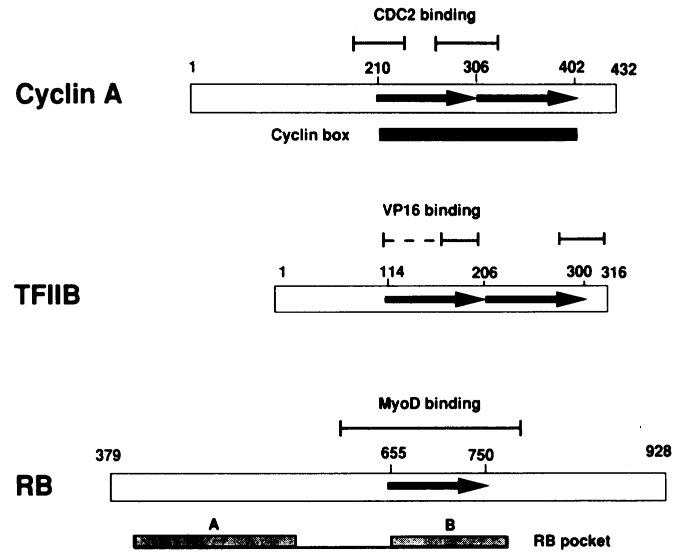


Figure 4. Diagrammatic representation of the relationship between cyclins, TFIIB and RB. The region of similarity between cyclin A and TFIIB is marked by two arrows, which represent the direct repeats first identified in TFIIB. The RB protein has similarity only with the first repeat (arrow in RB). Two predefined regions are also indicated, the cyclin box, which is the region in cyclin A homologous to other cyclins, and pocket B of RB. The experimentally determined protein–protein interacting regions spanning the first repeat in each protein are indicated.

Similarity of aligned TFIIB and cyclin sequences

Table IV records percent identities within and between cyclins (of different classes and divergence) and TFIIBs. The minimum reliable score between cyclins is 12.1% for human cyclin C against Yeast CLN3. The lowest pairwise cyclin identity is below 10%, but involves cyclins such as Yeast HCS26 that could not be reliably aligned. The cyclin Cs are less than 20% identical to all other cyclin families. The TFIIBs exhibit stronger conservation with a low of 20.1% for yeast TFIIB and TFIIIB.

Xenopus TFIIB and alfalfa cyclin B share 18.8% identity, higher than many pairwise scores for divergent cyclins. While this might be a chance score, values of 14.5% against cyclin C and 15.4% against cyclin D and cyclin E, themselves between 16.0–24.6% identical, show that it is the conserved positions in cyclins which match to the TFIIB sequence.

Table IV shows that the least conserved cyclin and TFIIB sequences are also those that share the fewest identities, e.g. 8.3% between archaebacterial TFIIB and yeast CLN3. This systematic fall in similarity is consistent with evolution from a common ancestor, but would otherwise be awkward to explain.

DISCUSSION

Cyclin and TFIIB sequences share common ancestry

The limited sequence identity shared by cyclins and TFIIB bears the hallmarks of both common ancestry and common structural fold. The shared identity is underpinned by hydrophobic positions whose conservation likely results from structural requirements. It extends over the full length of the cyclin box and the TFIIB repeats. Those cyclins that are closest to the common ancestral cyclin are also the most TFIIB-like, while the reciprocal also holds. The prediction that the cyclin box consists of a hitherto undetected duplication will be verified (or otherwise) by a future determination of the cyclin box tertiary structure.

Table IV. Percentage of identical residues shared by pairs of TFIIB repeat region and cyclin box sequences

H	H	H	Sp	Y	Y	Ss	Ms	Xl	H	Dm	Y	Pw	Y	
CCN C	CCN D	CCN E	PUC1	CLN3	HCS26*	CCN A	CCN B	TFIIB	TFIIB	TFIIB	TFIIB	TFIIB	TFIIIB	
100.0	16.6	16.0	13.4	12.1	9.6	18.5	16.3	14.5	13.4	13.4	10.0	12.2	12.8	H CCN C
	100.0	24.6	18.3	19.2	13.4	29.4	26.7	15.4	14.8	14.8	9.3	12.1	15.1	H CCN D
		100.0	18.8	17.3	12.5	31.0	27.1	15.4	15.4	12.6	10.4	14.3	9.3	H CCN E
			100.0	27.6	15.0	23.2	25.3	10.4	10.5	11.1	7.6	12.9	9.4	Sp PUC1
				100.0	12.1	20.5	19.5	11.7	12.8	10.0	12.2	8.3	12.5	Y CLN3
					100.0	13.6	12.0	7.3	7.9	7.9	6.2	9.6	5.0	Y HCS26*
						100.0	41.7	17.7	18.2	16.6	11.6	16.0	13.3	Ss CCN A
							100.0	18.8	18.2	18.2	10.5	14.4	13.8	Ms CCN B
								100.0	96.2	80.3	31.1	33.3	24.6	Xl TFIIB
									100.0	82.0	30.6	32.2	24.0	H TFIIB
										100.0	32.2	31.7	24.6	Dm TFIIB
											100.0	26.5	20.1	Y TFIIB
												100.0	26.7	Pw TFIIB
													100.0	Y TFIIIB

*Parts of the HCS26 alignment with other cyclins are uncertain because of the extreme mismatch, but the lack of conserved residues means that the scores are approximately correct.

The cyclin-TFIIB common ancestor

A priori models for the common ancestor are: 1, The common ancestor is neither cyclin nor TFIIB; 2, TFIIB arose from a branch of a preexisting cyclin family; or 3, cyclins arose from a preexisting TFIIB. While model 2 is disfavoured, either model 1 or 3 may hold. TFIIB may be older than the cyclins, which are only known in eukaryotes. The lack of a single cyclin branch especially close to TFIIB is inconsistent with model 2. Since the TFIIB repeats are more closely related to one another than are the cyclin repeats, they are more similar to the preduplication common ancestor, which may favour model 3 (depending on substitution rates). The correlation between TFIIB detections and the branch lengths in the TFIIB tree (Figure 1) favour a split of cyclins from a TFIIB-like lineage in a proto-eukaryote. In sum, the available data suggest that the common ancestor may have been TFIIB and was certainly more like the extant TFIIBs than cyclins.

A domain superfamily in nuclear regulatory proteins

Recently, one of our groups reported a distant (~20% identity) similarity between TFIIB and a conserved domain in the RB/p107 family, RB pocket B, in which disruptions are associated with tumourigenesis (24). This match, which spans a single TFIIB repeat, aligns the most conserved hydrophobic segment in RB/p107 'pocket B' with the critical hydrophobic motif in TFIIB (Figure 3). The match is strongest to TFIIB repeat 1 as indicated in the dotplots in Figure 2D.

The presence of a shared domain (which we designate TR for TFIIB Repeat) in TFIIB, cyclins and RB/p107 unites these proteins in a nuclear regulatory protein superfamily. Evolutionary parsimony suggests that this shared structural unit may be used to accomplish related functions. Figure 4 shows that the first TR domains in cyclin, TFIIB and RB are already defined as regions involved in protein–protein interactions. In cyclin A, TR1 is within the region required to contact the cdc2 kinase (39). TFIIB repeat 1 can bind the activation domain of the transactivator VP16 (40). The TR domain of RB corresponds to the pocket B required to contact the transcriptional activator MyoD (41). In addition, RB pocket B, along with pocket A, are required to bind a number of viral (E1A, TAg, E7) and cellular (E2F, c-myc, Pu.1) transcription factors (reviewed in 42–44). RB pocket B is also essential for correct RB phosphorylation, although it is not itself phosphorylated (45).

A common link between these families of proteins can be envisaged from what is known about their function. RB and TFIIB are both involved in transcription control. RB acts as a transcriptional repressor by binding to and sequestering the activation domain of E2F (46,47). TFIIB is a general transcription factor which is likely to be the target for the binding of transcriptional activators, of which the best studied example is VP16. We have shown that the activation domain of E2F (which contacts RB) can also contact the TFIIB protein, suggesting that RB and TFIIB may bind similar transcription factors (C.Hagemeier and T.K., unpublished results).

The cyclins are not, strictly speaking, transcription factors but they are intimately involved in regulating their function. Cyclins A and E have been found in a DNA-bound complex with E2F and RB/p107 (26,48,49), whereas cyclin D has been found to mediate RB phosphorylation via a cyclin D-specific kinase (50,51). Additional evidence that transcription may be regulated by cell cycle events comes from the revelation that the transcription factor TAF_{II} 250 is identical to the genetically defined cell cycle regulator CCG1 (52,53).

Although there is no evidence that TFIIB or RB directly contact cdc2-like kinases, the possibility is highlighted by the sequence similarities pointed out in this paper. Since cdc2-like kinases require cofactors, it is possible that certain kinases exist which require TFIIB or RB for their activation. Given that RB pocket B is essential for correct RB phosphorylation, it will be interesting to see whether the RB TR domain can act as an internal adaptor for a kinase.

Evolutionary origins of eukaryotic cell cycle regulation

The requirements during division of the eukaryotic cell are very different from the prokaryotic precursor with its rigid cell wall, lack of subcellular partitioning and simplicity of chomosome segregation. It follows that the origin of the eukaryotic cell necessitated the development of entirely new layers of regulation of cell state. It is becoming clear that these act through, and are overlayed upon, the basic transcriptional machinery, controlling the expression of genes involved in cell replication. If, as the data suggest, TFIIB is close to the common ancestor of cyclins, it follows that cell cycle regulatory elements have arisen by duplication and divergence of more basic transcriptional control elements.

ACKNOWLEDGEMENTS

We thank Des Higgins, Iain Mattaj and Paul Clarke for discussions concerning these results.

REFERENCES

1. Farber, G. K. and Petsko, G. A. (1990) *Trends Biochem. Sci.*, **15**, 228–234.
2. Rippmann, F., Taylor, W. R., Rothbard, J. B. and Green, N. M. (1991) *EMBO J.*, **10**, 1053–1059.
3. Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C. (1978) In *Atlas of Protein Sequence and Structure.* Dayhoff, M. O. (Ed.) **5**, suppl. 3, 345–358.
4. Gribskov, M., McLachlan, A. D. and Eisenberg, D. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4355–4358.
5. Smith, T. F. and Waterman, M. S. (1981) *Adv. Appl. Math.*, **2**, 482–489.
6. Thompson, J. D. Higgins, D. G. and Gibson, T. J. (1994) *Comput. Applic. Biosci.*, **10**, 19–30.
7. Saitou, M. and Nei, M. (1987) *Mol. Biol. Evol.*, **4**, 406–425.
8. Henikoff, S. and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10915–10919
9. Siomi, H., Matunis, M. J., Michael, W. M. and Dreyfuss, G. (1993) *Nucleic Acids Res.*, **21**, 1193–1198.
10. Haslam, R. J., Kolde, H. B. and Hemmings, B. A. (1993) *Nature*, **363**, 309–310.
11. Mayer, B. J., Ren, R., Clark, K. L. and Baltimore, D. (1993) *Cell*, **73**, 629–630.
12. Gibson, T. J., Thompson, J. D. and Heringa, J. (1993) *FEBS Lett.*, **3**, 361–366.
13. Gibson, T. J., Rice, P. D., Thompson, J. D. and Heringa, J. (1993) *Trends Biochem. Sci.*, **18**, 331–333.
14. Musacchio, A., Gibson, T., Rice, P., Thompson, J. and Saraste, M. (1993) *Trends Biochem. Sci.*, **18**, 343–348.
15. Evans, T., Rosenthal, E. T., Joungblom, J., Distel, D. and Hunt, T. (1983) *Cell*, **33**, 389–396.
16. Giordano, A., Whyte, P., Harlow, E., Franza Jr., B. R., Beach. D. and Draetta, G. (1989) *Cell*, **58**, 981–990.
17. Ewen, E.W., Faha, B., Harlow, E. and Livingston, D.W. (1992) *Science*, **255**, 85–87.
18. Faha, B., Ewen, M. E., Tsai, L.-H., Livingston, D. M. and Harlow, E. (1992) *Science*, **255**, 87–90.
19. Mudryj, M., Devoto, S. H., Hiebert, S. W., Hunter, T., Pines, J. and Nevins, J. R. (1991) *Cell*, **65**, 1243–1253.
20. Lees, E., Faha, B., Dulic, V., Reed, S. I. and Harlow, E. (1992) *Genes Dev.*, **6**, 1874–1885.
21. Bandara, L.R., Adamczewski, J.P., Hunt, T. and La Thangue, N.B. (1991) *Nature*, **352**, 249–251.
22. Lew, D.J. and Reed, S.I. (1992) *Trends Cell Biol.*, **2**, 77–81.
23. Weinmann, R. (1992) *Gene-Expr.*, **2**, 81–91.
24. Hagemeier, C., Bannister, A. J., Cook, A. and Kouzarides, T. (1993) *Proc. Natl. Acad. Sci USA*, **90**, 1580–1584.
25. Etzold, T. and Argos, P. (1993) *Comput. Applic. Biosci.*, **9**, 49–57.
26. Bairoch, A. and Boeckmann, B. (1993) *Nucleic Acids Res.*, **21**, 3093–3096.
27. Pearson, W. R. and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
28. Rice, C. M., Fuchs, R., Higgins, D. G., Stoehr, P. J. and Cameron, G. N. (1993) *Nucleic Acids Res.*, **21**, 2967–2971.
29. Higgins, D. G., Bleasby, A. J. and Fuchs, R. (1992) *Comput. Applic. Biosci.*, **8**, 189–191.
30. Pascarella, S. and Argos, P. (1992) *J. Mol. Biol.*, **224**, 461–471.
31. Altschul, S. F., Carroll, R. J. and Lipman, D. J. (1989) *J. Mol. Biol.*, **207**, 647–653.
32. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
33. Coulson, A. F. W., Collins, J. F. and Lyall, A. (1987) *Computer J.*, **30**, 420–424.
34. Na, J. G. and Hampsey, M. (1993) *Nucleic Acids Res.*, **21**, 3413–3417.
35. Ouzounis, C. and Sander, C. (1992) *Cell*, **71**, 189–190.
36. Creti, R., Londei, P. and Cammarano, P. (1993) *Nucleic Acids Res.*, **21**, 2942.
37. Lopez-de-Leon, A., Librizzi, M., Puglia, K. and Willis, I. M. (1992) *Cell*, **71**, 211–220.
38. Olsen, G. J. (1987) *Cold Spring Harbour Symp. Quant. Biol.*, **LII**, 825–837.
39. Kobayashi, H., Stewart, E., Poon, R., Adamczewski, J. P., Gannon, J., Hunt, T. (1992) *Mol. Biol. Cell.*, **3**, 1279–1294.
40. Roberts, S. G. E., Ha, I., Maldonado, E., Reinberg, D. and Green, M. R. (1993) *Nature*, **363**, 741–744.
41. Gu, W., Schneider, J. W., Condorelli, G., Kaushai, S., Mahdavi, V. and Nadal-Ginard, B. (1993) *Cell*, **72**, 309–324.
42. Kouzarides, T. (1993) *Trends Cell Biol.*, **3**, 211–213.
43. Levine, A. J. (1993) *Annu. Rev. Biochem.*, **62**, 623–651.
44. Hamel, P. A., Gallie, B. L. and Phillips, R. A. (1992) *Trends Genetics*, **8**, 180–185.
45. Qiau, Y., Luckey, C., Horton, L., Essev, M. and Templeton, D. J. (1992) *Mol. Cell. Biol.*, **12**, 5363–5372.
46. Flemington, E. K., Speck, S. H. and Kaelin, W. G. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6914–6918.
47. Hagemeier, C., Cook, A. and Kouzarides, T. (1993) *Nucleic Acids Res.*, **21**, 4998–5004.
48. Shirodkar, S., Ewen, M., DeCaprio, J. A., Morgan, J., Livingston, D. M. and Chittenden, T. (1992) *Cell*, **68**, 157–166.
49. Devoto, S. H., Mudryj, M., Pines, J., Hunter, T. and Nevins, J. R. (1992) *Cell*, **68**, 167–176.
50. Kato, J.-Y., Matsushime, H., Hiebert, S. W., Ewen, M. E. and Sherr, C. J. (1993) *Genes Dev.*, **7**, 331–342.
51. Dowdy, S. F., Hinds, P. W., Louie, K., Reed, S. E., Arnold, A. and Weinberg, R. A. (1993) *Cell*, **73**, 499–511.
52. Hisatake, K., Hasegawa, S., Takeda, R., Nakatini, Y., Horikoshi, M. and Roeder, R. G. (1993) *Nature*, **362**, 179–181.
53. Ruppert, S., Wang, E. H. and Tjian, R. (1993) *Nature*, **362**, 175–179.
54. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, England, p. 75.