

Identification of high affinity Tbf1p-binding sites within the budding yeast genome

Catherine Elaine Koering, Genevieve Fourel, Emmanuelle Binet-Brasselet, Thierry Laroche¹, Franz Klein² and Eric Gilson*

Laboratoire de Biologie Moléculaire et Cellulaire de l'Ecole Normale Supérieure de Lyon, UMR 5665 CNRS/ENS, 46 Allée d'Italie, 69364 Lyon Cedex 07, France, ¹ISREC (Swiss Institute for Experimental Cancer Research), 155 Chemin de Boveresses, CH-1066 Epalinges/Lausanne, Switzerland and ²Institute of Botany, Rennweg 14, A-1030, Vienna, Austria

Received March 16, 2000; Revised and Accepted May 6, 2000

ABSTRACT

The yeast *TBF1* gene is essential for mitotic growth and encodes a protein that binds the human telomere repeats *in vitro*, although its cellular function is unknown. The sequence of the DNA-binding domain of Tbf1p is more closely related to that of the human telomeric proteins TRF1 and TRF2 than to any yeast protein sequence, yet the functional homologue of TRF1 and TRF2 is thought to be Rap1p. In this study we show that the Tbf1p DNA-binding domain can target the Gal4 transactivation domain to a (TTAGGG)_n sequence inserted in the yeast genome, supporting the model that Tbf1p binds this sub-telomeric repeat motif *in vivo*. Immunofluorescence of Tbf1p shows a spotty pattern throughout the interphase nucleus and along synapsed chromosomes in meiosis, suggesting that Tbf1p binds internal chromosomal sites in addition to sub-telomeric regions. PCR-assisted binding site selection was used to define a consensus for high affinity Tbf1p-binding sites. Compilation of 50 selected oligonucleotides identified the consensus TAGGGTTGG. Five potential Tbf1p-binding sites resulting from a search of the total yeast genome were tested directly in gel shift assays and shown to bind Tbf1p efficiently *in vitro*, thus confirming this as a valid consensus for Tbf1p recognition.

INTRODUCTION

In *Saccharomyces cerevisiae* the DNA-binding protein Tbf1p (TTAGGG repeat binding factor 1) is encoded by a gene essential for mitotic growth and was originally cloned through its ability to bind human type telomere repeats (1). Interestingly, its DNA-binding domain is closely related to that of telomeric proteins in fission yeast (telomere-associated in *Schizosaccharomyces pombe* protein, Taz1p) (2) and in vertebrates (TTAGGG repeat factors 1 and 2, TRF1 and TRF2) (3). These conserved sequences also exhibit significant homologies to a single Myb repeat and are often referred to as Myb-like sequences (4). On

the other hand, their high degree of similarity identifies them as members of a distinct group of Myb repeats, called the 'telobox' family (3,5,6). The telobox domains of TRF1, TRF2 and Taz1p are sufficient to bind telomeric DNA specifically *in vitro* (3,4,7). The binding specificity of all the members of the telobox family is likely to be highly related to the TTAGGG type of telomeric repeat since Tbf1p, TRF1, TRF2 and Taz1p show a high affinity for TTAGGG-like repeats but not to the irregular telomeric repeats of *S.cerevisiae*. The major telomere repeat DNA-binding protein in budding yeast is Rap1p, which binds DNA through two degenerate Myb repeats, only distantly related to teloboxes.

In yeast, the telomere-associated repeats X and Y' contain multiple TTAGGG- or TTAGGG-like repeats, which are known to bind Tbf1p *in vitro* (1,8,9). This suggests that this protein exerts a function at chromosome ends through binding to sub-telomeric regions. This hypothesis was recently reinforced by the demonstration that Tbf1p sites can counteract telomeric position effects and, more generally, Sir-mediated transcriptional silencing (9). The sub-telomeric regions containing the Tbf1p sites and exhibiting anti-silencing activities were named STARS for sub-telomeric anti-silencing regions (9). Furthermore, unpublished data from our laboratory demonstrate that Tbf1p proteins artificially tethered to DNA recapitulate STAR activities, blocking propagation of the silent chromatin (G.Fourel, C.Boscheron, E.Revardel and E.Gilson, in preparation). The anti-silencing properties of Tbf1p in the X and Y' elements are redundant with Reb1p, a weak transactivation factor whose binding sites are found in a variety of yeast promoters (10–12). Several Tbf1p and Reb1p sites are clustered in the STARS of X and Y' (9). STARS might form boundaries to limit telomeric position effects and/or to protect the telomere from the influence of sub-telomeric elements. Interestingly, reminiscent of the enhancer-blocking activity of metazoan insulators, Reb1p was reported to inhibit transcription when placed between an upstream activating sequence and a TATA element (13). The possibility that Tbf1p and/or Reb1p could be involved in boundary effects throughout the yeast genome makes them an attractive target for further analysis.

In this study we first demonstrate that Tbf1p binds sub-telomeric TTAGGG-like repeats using a one-hybrid system. We then asked whether Tbf1p could bind non-sub-telomeric regions in the yeast genome. We started this analysis by

*To whom correspondence should be addressed. Tel: +33 5727 28453; Fax: +33 5727 28080; Email: eric.gilson@ens-lyon.fr

determining a consensus sequence for high affinity Tbf1p binding sites through a PCR-based selection method. Using the consensus, 52 potential binding sites were found scattered through the yeast genome. The presence of a large number of chromosome-associated Tbf1p proteins was confirmed by the analysis of Tbf1p immunostaining in yeast nuclei and meiotic chromosomes. Overall, these data predict an important role for Tbf1p throughout the entire genome and not only in sub-telomeric regions. Furthermore, the potential Tbf1p-binding sites identified in this work will serve as important starting points in further studies of the possible roles of Tbf1p.

MATERIALS AND METHODS

Yeast strain constructs and β -galactosidase assays

In vitro DNA manipulation, plasmid extraction and Southern blot analysis were carried out as described previously (14). Yeast media, growth conditions and β -galactosidase assays on yeast permeabilised cells were as described (15).

Plasmid pLG-312-178 DNA (16) (here termed pSX178), which contains the *CYC1* upstream region with no UAS fused to *lacZ*, was cut with *XhoI*, rendered flush and ligated with a *HindIII*-*EcoRI* fragment, flush at both ends, either from pHuTel (17), leading to pHTel1 and pHTel2, or from pTel270 (17), leading to pYTel. pHuTel1 and pHuTel2 contain a 110 nt long fragment containing 10 repeats of TTAGGG in opposite orientations (orientations are given in Fig. 1). pYTel contains a 305 nt long fragment containing 270 nt of the yeast telomeric repeats ($(TG_{1-3})_n$). All these plasmids have been checked by sequencing the insert region.

In order to integrate the *CYC1-lacZ* reporter constructs into the yeast genome, a *ScaI* fragment from pSX178, pHTel1, pHTel2 or pYTel was inserted into the *SmaI* and *EcoRV* sites of the integrative plasmid pRS303 (18), leading respectively to pRSX178, pRHTel1, pRHTel2 and pRYTel. Integration of these plasmids into the *his3* locus was performed by transforming *PstI*-linearised plasmid DNA into the S150-2B strain (*MATa leu2-3,112 ura3-52 trp1-289 his3 gal2*, kindly provided by J. Broach). After HIS^+ selection, the correct integration events were screened by a Southern restriction site analysis. The resulting strains are called SX (from pRSX178), HTEL1 (from pRHTel1), HTEL2 (from pRHTel2) and HYTEL (from pRYTel).

Plasmid pGAD424 DNA (purchased from Clontech) was cut with *EcoRI* and ligated with the 800 nt *EcoRI* fragment from pCDS47 (1) carrying the 3'-part of the *TBF1* gene. The resulting plasmid, pA-TBD, encodes the transactivation domain of Gal4 fused to the last 256 amino acids of Tbf1p, known to contain the DNA-binding domain of the protein (1).

Immunofluorescence and yeast meiotic spreads

Immunofluorescence was performed on formaldehyde-fixed yeast spheroplasts of the GA-7 strain (*MATa/ade2/ADE2 his4/HIS4 lys2/LYS2 pep4-3/pep4-3 trp1/TRP1*) grown for 12 h on YPAD at 30°C, then for 2.5 h at 37°C, and on spreads of meiotic yeast nuclei as described (19). Both the TBD serum and the fluorescein-conjugated goat anti-rabbit IgG were diluted 1:50 in phosphate-buffered saline (PBS). Affinity-purified anti-TBD was diluted 1:10. The slides were mounted in 1× PBS, 50% glycerol, 2 μ g/ml ethidium bromide (or 0.2 μ g/ml

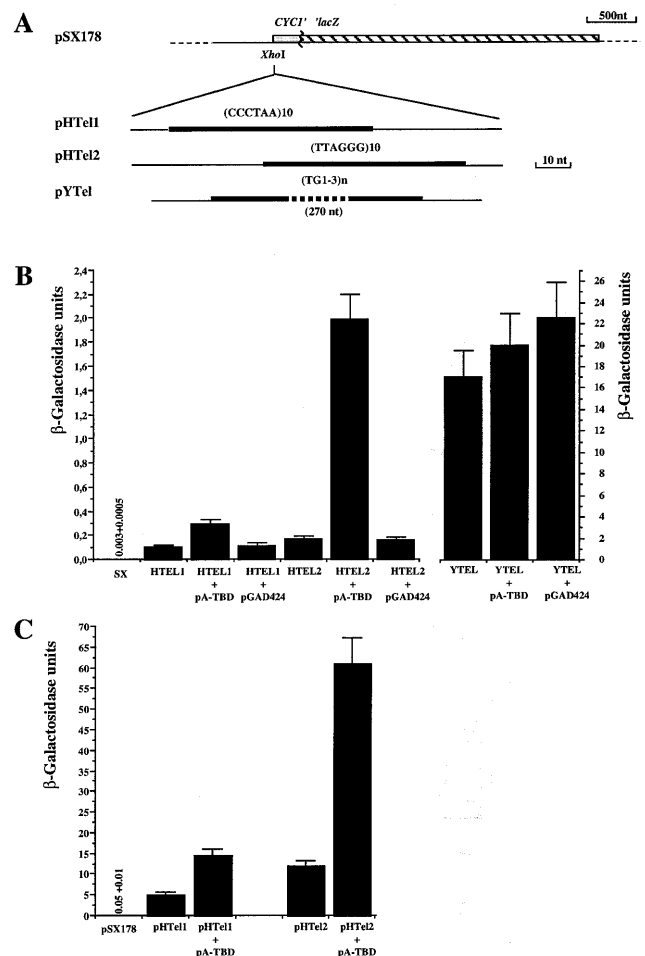


Figure 1. (TTAGGG)_n-dependent one hybrid assays. (A) The structures of the reporter genes used in this study are presented. The thick lines represent the telomeric repeats inserted in front of a UAS-less *CYC1-lacZ* fusion gene. CCCTAA, TTAGGG and TG₁₋₃ indicate the insert orientation 5'→3'. (B and C) β -Galactosidase activities produced in different yeast strains carrying the integrated constructs at the *HIS3* locus (see Materials and Methods) (B) or with the reporters on a 2 μ plasmid (C). Cells were grown in SD minimal medium until they reached 10⁷ cells/ml and β -galactosidase assays were performed in permeabilised cells as described (15). Units are expressed as 10³ OD₄₂₀/OD₆₀₀ ml/min. The error bars represent the SD of the mean, calculated from the results of at least three independent cultures of the same strain. Each culture was assayed in duplicate. Beneath each column, the relevant strain and any additional plasmids in that strain are given. For strains with very low level expression, the actual numerical values are indicated.

DAPI) and 24 μ g/ml 1,4-diazabicyclo[2.2.2]octane (Sigma). Confocal microscopy was performed on a Zeiss Laser Scanning Microscope 410 system with a Plan-Neofluar 100× (1.3 oil) objective. Standard filtering operations were performed on the confocal images.

PCR selection procedure

D10 is a 40mer oligonucleotide containing 10 nt of random DNA sequence flanked by *HindIII* and *XbaI* restriction sites [5'-GTA CGT ATC TAG ATC (N₁₀) GC AAG CTT CTA GCAG-3', where N represents an equal probability of all four bases]. D10 was converted to double-stranded DNA by primer extension with primer D10A (5'-GCT GAC TGC TAG AAG

CTT GC-3') previously labelled with [γ - 32 P]ATP by T4 polynucleotide kinase (PNK). The D10 duplex probe was purified on a 4.5% polyacrylamide gel (29/1) and electroeluted. As a positive control, we used D3, a 48mer oligonucleotide which contains two TTAGGG repetitions flanked by *Eco*RI and *Bam*HI restriction sites [5'-TCG AGA CGG ATC CAG (N₃) (TTAGGG)₂ (N₃) CCG AAT TCG ATA TCG-3']. D3 double-stranded DNA was obtained after primer extension with primer D3At (5'-TGT CTC GAT TAC GAA TTC GG-3') previously labelled with [γ - 32 P]ATP by PNK. The D3 probe was also purified on a 4.5% polyacrylamide gel and electroeluted.

For the selection rounds, band shift assays were performed by incubating 15×10^4 c.p.m. of D10 probe (~2.5 pmol) with 6–30 μ g of a bacterial extract enriched in maltose-binding protein (ETBD) (3) in a final volume of 10 μ l of binding buffer [20 mM KCl, 180 mM NaCl, 15 mM Tris pH 7.4, 0.05 mM spermine, 0.125 mM spermidine, 500 ng poly(dI-dC)]. After 20 min incubation at room temperature the mixture was loaded onto a 4.5% polyacrylamide gel (29/1) with 0.5 \times Tris–borate and run at a constant 170 V for 80 min at 4°C. The shifted complexes were visualised by autoradiography of the gel, after an overnight exposure at room temperature. The shifted complexes were cut out, electroeluted and ethanol precipitated. Half of the recovered DNA was PCR amplified by 20 cycles performed under standard conditions (20), with primers D10A and D10B (5'-CGTACGTACGTATCTAGATC-3') previously labelled with [γ - 32 P]ATP by PNK. The PCR products were purified on a 4.5% polyacrylamide gel, electroeluted and used for another round of affinity selection (probes D10.1 and D10.2). The second half of the recovered affinity-selected DNA obtained after rounds 2–6 was PCR amplified with cold primers, digested with *Xba*I and *Hind*III, purified on a 10% polyacrylamide gel, electroeluted and cloned into the *Xba*I and *Hind*III sites of the pBluescript cloning vector DNA (pBSK). The inserted oligonucleotide sequence was determined in 50 independent clones.

Band shift analysis

The oligonucleotide probes were obtained as follows. Complementary oligonucleotides were labelled with [γ - 32 P]ATP by PNK, annealed together, purified on a 4.5% polyacrylamide gel, electroeluted and precipitated. The 2.1 probe was obtained by PCR labelling of the 2.1 clone plasmid DNA with radioactive primers D10A and D10B.

The binding reactions were performed as indicated for the PCR selection procedure with either extract of bacteria expressing ETBD (designated ETBD) or ETBD purified to near homogeneity on an amylose column (designated ETBDp) or reticulocyte lysate containing the full-length Tbf1p protein (designated Tbf1p). The binding reaction was separated on a 4.5% polyacrylamide gel (29/1) with 0.5 \times Tris–borate. The dried gels were exposed to PhosphorImager screens and, when required, the signals corresponding to the shifted complexes were quantified with ImageQuant Software (Molecular Dynamics).

RESULTS

Tbf1p binds (TTAGGG)₁₀ in a yeast one-hybrid assay

In order to explore whether Tbf1p recognises TTAGGG repeats *in vivo*, we designed a yeast one-hybrid assay (21) with

a reporter strain containing 10 repeats of TTAGGG in front of a UAS-less *CYC1-lacZ* gene (HTEL1 or HTEL2; Fig. 1). As a control, we used the same construct without an insert upstream of the reporter gene (SX) or with inserted (TG₁₋₃)_n sequences (YTEL) which are poorly bound *in vitro* by Tbf1p (1,3,8; Fig. 1).

In strains containing one copy of the reporter gene inserted into the yeast genome, a slight activation of the *CYC1-lacZ* gene due to the presence of (TTAGGG)₁₀ is observed; compare the basal level of the SX strain with that of the HTEL1 and HTEL2 strains (Fig. 1B). As reported previously for (TG₁₋₃)_n sequences (22), the YTEL strain exhibits a high level of activation likely to be due to binding of multiple Rap1p molecules (Fig. 1B). More interestingly, both HTEL1 and HTEL2 show an increase in β -galactosidase activity when carrying the pA-TBD plasmid, which encodes the transactivation domain of Gal4 fused to the Tbf1p DNA-binding domain (abbreviated to TBD) (1) and not when they contain the parental plasmid encoding only the Gal4 moiety (pGAD424) (Fig. 1B). The level of transactivation due to pA-TBD is much stronger for HTEL2 (~12 times the level of the same strain without pA-TBD) than for HTEL1 (~3 times; Fig. 1B). The same reporter constructs present on a multicopy plasmid (pHTel2) also exhibited a much higher level of transactivation due to pA-TBD than that of pHTel1 (Fig. 1C). Thus, a GAD–TBD hybrid protein can activate a promoter carrying an upstream (TTAGGG)₁₀ insert. This finding demonstrates that the Tbf1p DNA-binding domain binds *in vivo* to (TTAGGG)₁₀, which has previously only been demonstrated *in vitro* (1,3,8). This strongly suggests that Tbf1p also binds to sub-telomeric TTAGGG-like repeats *in vivo*.

Tbf1p immunolocalisation in interphase nuclei and in meiotic chromosome spreads

A polyclonal antiserum raised against the TBD polypeptide in rabbit was previously described (3). After affinity purification this serum is monospecific for Tbf1p in western blot analysis of total yeast protein extract (3; data not shown). Indirect immunofluorescence using the TBD antiserum on vegetatively growing yeast cells is shown in Figure 2A. The Tbf1p staining (green) is exclusively nuclear, as shown by its co-localisation (yellow) with ethidium bromide fluorescence (red) (Fig. 2Ac). This staining forms a dispersed network throughout the nucleoplasm (Fig. 2Ab). The intensity of the Tbf1p fluorescence signal correlates with that due to ethidium bromide, suggesting that most Tbf1p is directly associated with chromatin (Fig. 2A, compare a and b). This network pattern was obtained with a variety of wild-type strains (data not shown and Fig. 2A) and with affinity-purified antibodies as well as the antiserum. Omission of the primary antibody showed no FITC signal above background (data not shown), confirming the monospecificity of the antiserum. Interestingly, in nearly all nuclei Tbf1p is clearly excluded from a crescent-shaped domain of the nucleus always located at one pole, which is characteristic of the yeast nucleolus (Fig. 2A). This zone of Tbf1p exclusion contains a discrete DNA loop visible by ethidium staining, marked by a triangle in Figure 2Ac, which is likely to represent the rDNA of chromosome XII when the nucleolar RNA is released (23,24). It appears, therefore, that Tbf1p is not part of the yeast nucleolus. Tbf1p localisation was also studied in meiotic (pachytene) chromosomes to see if its staining is, like

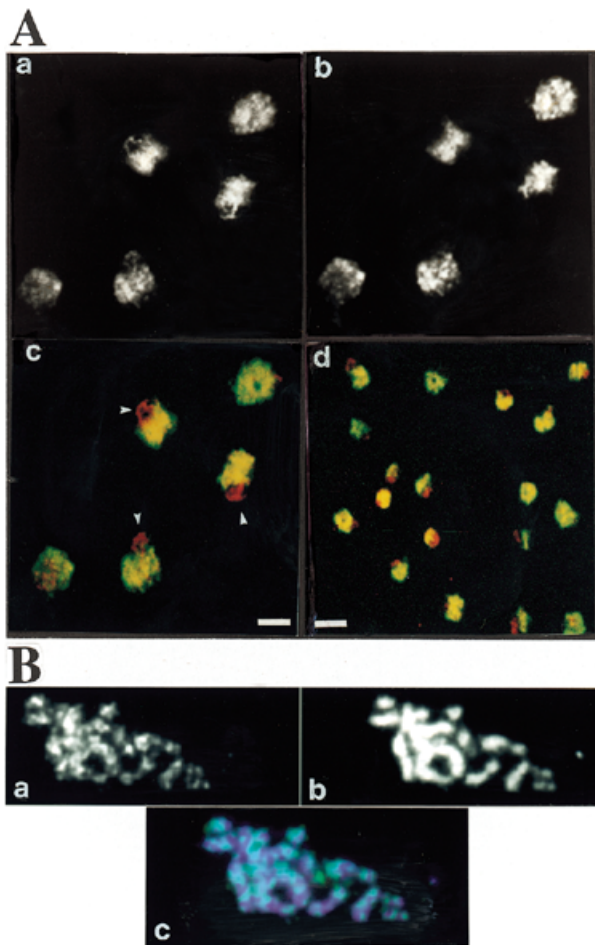


Figure 2. Immunolocalisation of Tbf1p in yeast cells. (A) A monospecific anti-TBD antibody was used on fixed yeast spheroplasts of the diploid strain GA-7, as described in Materials and Methods. FITC-coupled secondary antibody detects Tbf1p staining and ethidium bromide reveals nuclear DNA. (a) Ethidium bromide staining corresponding to five yeast cells. (b) Anti-TBD staining. (c) Superposition of (a) and (b). Arrows indicate the nucleolar region of the yeast nucleus. For (a)–(c) the bar indicates 1.5 μ m. (d) An overview of the same labelling reaction showing merged DNA staining and immunofluorescence. Regions that are red indicate DNA in the nucleolus where no anti-TBD staining is observed. Yellow indicates overlap of the Tbf1p and DNA staining patterns. Bar 3 μ m. (B) Anti-TBD antibodies were used to detect Tbf1p in spreads of meiotic yeast nuclei from strain SK1. The pachytene stage is characterised by the presence of 16 chromosomal structures, the condensed bivalents. (a) The Tbf1 pattern on a single spread pachytene nucleus. (b) Corresponding DAPI staining of the DNA. (c) A merged image of the Tbf1p immunofluorescence (green) with DNA staining (blue).

that of Rap1p, enriched at telomeres (19). A micropunctate pattern was observed all along the meiotic chromosomes, suggesting the presence of numerous Tbf1p sites in yeast chromosomes (Fig. 2B, compare a and b). Unlike Rap1p, Tbf1p was not predominantly found at telomeres. However, brighter immunofluorescence spots appear to be found at the ends of some paired chromosomes (Fig. 2Bc), consistent with the original hypothesis that sub-telomeric regions are also Tbf1p-binding sites *in vivo*.

In summary, immunofluorescence studies with interphase nuclei and with meiotic spreads revealed the presence of many

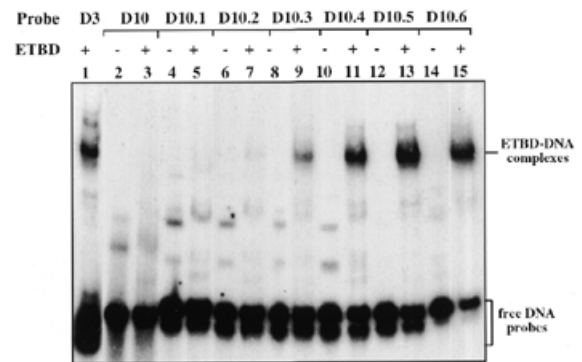


Figure 3. PCR selection of Tbf1p-binding sites. Aliquots of 3×10^4 c.p.m. of each probe were incubated without or with 30 μ g of proteins of bacterial soluble extract containing the ETBD protein, as indicated. The control D3 probe contains two TTAGGG repetitions. The D10 probe contains 10 nt of random DNA sequence flanked by *Hind*III and *Xba*I restriction sites. Probe D10.1 was obtained after the first round of affinity selection; D10.2 after the second round of affinity selection, etc. (see text).

Tbf1p sites along yeast chromosomes, not necessarily exclusively at sub-telomeric locations (Fig. 2). Therefore, in the rest of this work we will investigate the presence of Tbf1p DNA-binding sites outside sub-telomeric regions.

Selection of specific Tbf1p-binding sites

In order to select for high affinity Tbf1p-binding sites from random sequence oligonucleotides we used the Casting method (cyclic amplification and selection for targets) (25). We employed the Tbf1p DNA-binding domain fused to ETBD partially purified from *Escherichia coli* cells (3). The ETBD molecule contains the 236 residues of the C-terminal half of Tbf1p, which was previously demonstrated to contain the entire Tbf1p DNA-binding domain (1). The DNA-binding properties of ETBD and of full-length Tbf1p expressed in a reticulocyte lysate are indistinguishable (1,3,9; data not shown).

The 40 bp oligonucleotide D10 containing a central random sequence of 10 bp was used as template to start the selection (see Materials and Methods). The choice of the N_{10} core was based on previous studies, which indicated that this length would allow a search for single binding sites (8; data not shown). A 48 bp oligonucleotide D3 containing the $(TTAGGG)_2$ sequence was used as a positive control. To ensure that all possible combinations of the 10 nt randomised region ($4^{10} = 10^6$) would be represented, 70 ng (2.5 pmol) of D10, containing $\sim 10^{12}$ molecules, were used in each binding reaction. The protein–DNA complexes were separated by gel electrophoresis, cut out from the gel, electroeluted, amplified by PCR with primers D10A and D10B and used in further rounds of selection. Figure 3 shows the progressive enrichment of ETBD-binding sites. No retarded complexes with the D10 probe nor with the D10.1 probe obtained after the first PCR amplification could be visualised (Fig. 3, lanes 3 and 5) as compared to the complexes formed with D3 (Fig. 3, lane 1). In these cases, the ETBD–DNA complex was excised blind, based on the position of retarded ETBD–DNA complexes.

ETBD to 35 of the 50 cloned sequences was checked by specific competition in band shift experiments (Table 1; data not shown).

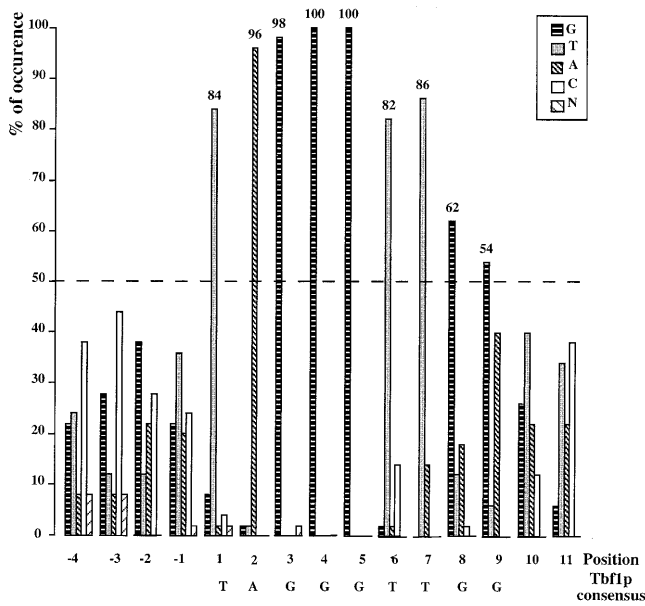


Figure 4. Histogram used to determine the Tbf1p consensus sequence. The high affinity sites were aligned according to the TAGGG motif. The percentage of occurrence of A, T, G, C or N (y-axis) is represented as a function of the position in the sequence (x-axis). The consensus was determined by the nucleotides present at more than 50% in each position. The consensus sequence is shown below the histogram.

After the second round of selection (D10.2, Fig. 3, lane 8), the complexes were detected and their relative amount increased each round, until the sixth round (Fig. 3, lane 15). At this point selection was stopped.

Analysis of the selected binding sites

The DNA recovered after rounds 2, 3, 4, 5 and 6 was amplified by PCR with primers D10A and D10B. The PCR products were digested with *Xba*I and *Hind*III, purified on a 10% polyacrylamide gel, electroeluted and inserted into pBSK. At total of 50 independent clones were isolated and sequenced (Table 1). Examination of these sequences revealed that an exact match to TAGGG appeared in a vast majority (41 of 50). When this sequence, or one-mismatch derivatives, was employed as a reference point, other nucleotide positions displayed marked sequence preferences (Fig. 4). A consensus corresponding to TAGGGTTGG was evident from the alignment of all of these sequences (Fig. 4 and Table 1). For the nine oligonucleotides that do not contain the TAGGG motif, the sequence that conformed best to the consensus was used in the alignment. The consensus is composed of bases that are observed >50% of the time at each position. Inspection of Figure 4 reveals that this cut-off value, although arbitrary, accurately reflects the nucleotide distribution in the observed data set. In a portion of the oligonucleotides, the sequence homologous to the TAGGGTTGG consensus overlaps the cloning sites of the N₁₀ oligonucleotide. We did not recover oligonucleotides containing multiple TAGGG core sequences or obvious repeated consensus sequences. The binding of

Table 1. Sequences of the cloned oligonucleotides

Clone name	N10
2.1*	CC <u>TGGGT</u> TGga
C3.1X	TCCCC <u>TAGGg</u> caag
3.2*	T <u>TAGGG</u> TTNg
3.3	GGG <u>GAGGG</u> TTga
3.3X*	CTG <u>TAGGG</u> TTga
3.4*	ATC <u>TAGGG</u> TTga
3.4X	GGC <u>TAGGG</u> TTga
3.5X	CTA <u>TAGGG</u> TTga
3.7X*	GAG <u>TAGGG</u> CTga
3.8X	CC <u>TGGGG</u> TTGg
3.9X	<u>TAGGG</u> TTGg
4.0	GGGT <u>TAGGg</u> atct
4.1*	CAT <u>TAGGG</u> TTga
4.1X*	GG <u>TAGGG</u> TTga
4.2*	GG <u>TAGGG</u> TTga
C4.2X*	CCACT <u>TAGGg</u> caag
4.3*	CTA <u>TAGGG</u> TTga
4.4*	CTA <u>TAGGG</u> TTga
4.4X	T <u>CAGGG</u> TTAg
4.5*	GGT <u>TAGGG</u> TTga
C4.6X	CANFCT <u>TAGGg</u> caag
4.7X	TGT <u>TAGGG</u> TTga
4.8X	<u>TAGGG</u> TTGG
4.9X*	<u>GAGGG</u> TTTC G
5.1*	T <u>GAGGG</u> TTGA
C5.1X*	CACAT <u>TAGGg</u> caag
5.2X*	AN <u>AAGGG</u> TTGg
5.4X*	AGA <u>TAGGG</u> TTga
5.5	CGT <u>TAGGG</u> TTga
5.5X*	AG <u>TAGGG</u> TTGg
5.6X*	GGG <u>GAGGG</u> TTga
C5.9X*	CAANTC <u>TAGGg</u> caag
6.0X*	A <u>TAGGG</u> TTAT
C6.1*	CACCAT <u>TAGGg</u> caag
6.1X*	AG <u>TAGGG</u> TTGg
6.3X*	GG <u>TAGGG</u> TTGG T
6.4X*	GT <u>TAGGG</u> TTTC
6.6X*	AG <u>TAGGG</u> TTGg
6.7X*	A <u>TAGGG</u> TTGG
C6.8X*	ACACAC <u>TAGGg</u> caag
6.9X*	AAT <u>TAGGG</u> TTga
6.10	GGT <u>TAGGG</u> TTna
6.11*	C <u>TAGGG</u> TTTC
6.12*	T <u>TAGGG</u> TTga
6.13*	<u>TAGGG</u> TTGG G
6.14	A <u>TAGGG</u> TTGG
6.15*	A <u>TAGGG</u> TTPT
6.16	A <u>TAGGG</u> TTTC
6.18*	GG <u>TACGG</u> TTTC
6.19*	A <u>TAGGG</u> TTGG
<hr/>	
CONSENSUS	<u>TAGGG</u> TTGG

The sequences of the variable parts (N₁₀) of the cloned oligonucleotides are shown in capital letters. The sequences matching the Tbf1p consensus (shown at the bottom) are underlined. When the fixed part of the oligonucleotide contributes to the consensus, the corresponding sequence is shown in lower case letters. The clones are named as follows: the first number indicates the round of selection, the second and the third are the number of the clone inside this round. The letter C indicates that the sequence of the clone is presented in inverted orientation, in order to match with the consensus. The clones marked by a star were used in band shift assays with Tbf1p and have been shown to be effective *in vitro* Tbf1p-binding sites (data not shown).

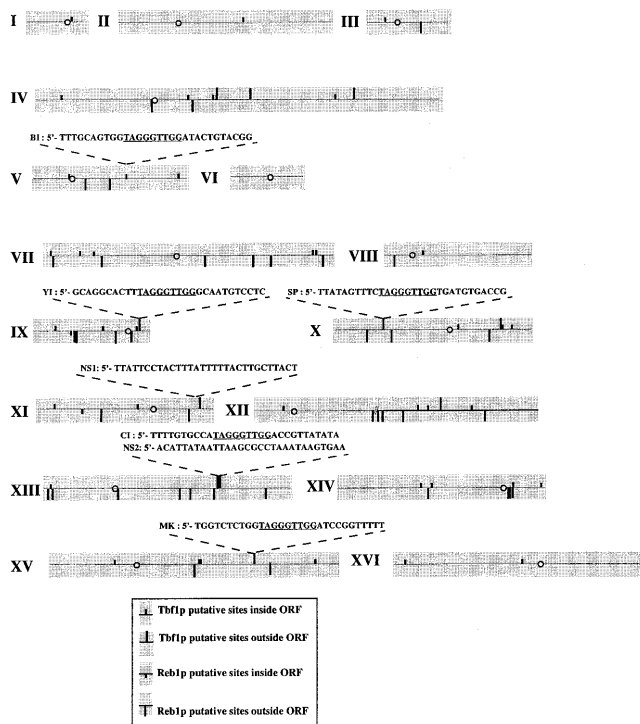


Figure 5. Putative binding sites for Tbf1p and Reb1p in the *S.cerevisiae* genome. The exact matches to the Tbf1p consensus (TAGGGTTGG) are represented for each chromosome as bars above the horizontal axis of each chromosome (I–XVI). Exact matches to the high affinity Reb1p-binding site (CCGGGTAAC) are represented as bars below the horizontal chromosome axis. The long bars indicate occurrences in intergenic regions; the short bars indicate occurrences within open reading frames. The enlarged regions show the oligonucleotide sequences used in band shift assays. The Tbf1p consensus is underlined.

The Tbf1p consensus can be used to accurately predict genomic high affinity sites

We performed computer-assisted searches on the *Saccharomyces cerevisiae* Genome Database (SGD) using PatMatch with the consensus sequence TAGGGTTGG. A total occurrence of 52 exact matches was obtained, including 10 that were not in open reading frames (Fig. 5 and Table 2). To confirm that the sites detected in the yeast genome using the Tbf1p consensus indeed bind Tbf1p in the context of their surrounding sequences, double-stranded oligonucleotides of 30 bp spanning five potential Tbf1p sites were synthesised, four in intergenic regions (BI, SP, CI and MK) and one in a coding region (YI) (Fig. 5). As negative controls we designed two double-stranded oligonucleotides of identical size, but corresponding to regions juxtaposed to two potential Tbf1p sites, named NS1 and NS2 (Fig. 5). As shown in Figure 6A, probes derived from the oligonucleotides containing the Tbf1p consensus were avidly bound by ETBD, whereas the two probes lacking this sequence were not. Similar results were obtained with a highly purified ETBD preparation (Fig. 6B and C, lane 2; data not shown) and with full-length Tbf1p synthesised in a reticulocyte lysate (Fig. 6C, lane 3; data not shown). Furthermore, binding of ETBD appears similar between one of the potential chromosomal Tbf1p sites (YI) and a known Tbf1p-binding site, the STAR motif of the sub-telomeric X element (Fig. 6B). Overall, we conclude that the Tbf1p consensus predicts high affinity sites for bacterially produced ETBD and for full-length Tbf1p obtained from a eukaryotic source. This confirms the validity of the Tbf1p consensus sequence determined by our selection procedure.

The 52 potential Tbf1p binding sites identified in the yeast genome appear to be evenly distributed among the 16 yeast chromosomes and do not show any significant bias towards chromosome ends (Fig. 5). Strikingly, none of the X or Y' elements contains the Tbf1p consensus. However, five near

Table 2. The results of PatMatch searches using the complete *S.cerevisiae* genomic DNA and the indicated nucleotide pattern

Search	Entered nucleotide pattern	Total hits		Intergenic hits		Intergenic hits (%)	
		0m	1m	0m	1m	0m	1m
Reb1	CCGGGTAAC	44	896	41	557	93	62
Reb1 ^{Y'}	CCGGGTAAG	46	880	44	436	95	50
Reb1 ^{Xa}	TAGGGTAAT	89	2333	68	1027	76	44
Reb1 ^{Xb}	CAGGGTAAG	44	1560	22	515	50	33
Reb1 ^{Xc}	CAGGGTAAG	55	1339	46	476	84	36
Reb1 ^M	CAATGGGCC	61	1650	5	240	8	14
Tbf1	TAGGGTTGG	52	1753	10	343	19	19
Tbf1 ^{Y'a}	TAGGGCTAT	69	1767	40	531	58	30
Tbf1 ^{Y'b}	TAGGGCTAA	62	1605	29	435	46	27
Tbf1 ^{Y'c}	CAGGGTTTC	73	1971	28	421	38	21
Tbf1 ^{Xa}	CAGGGTTGG	44	1388	9	182	20	13
Tbf1 ^{Xb}	TAGGGTTAG	49	1544	18	466	38	30
Rap1p	YYACCCANNCAYY	62	2037	18	645	29	31
Abf1p	TRTCRYTNNNACGA	35	1221	8	286	23	23

Y = C or T; R = G or A; N = A, T, G or C. 0m, no mismatches; 1m, 1 mismatch.

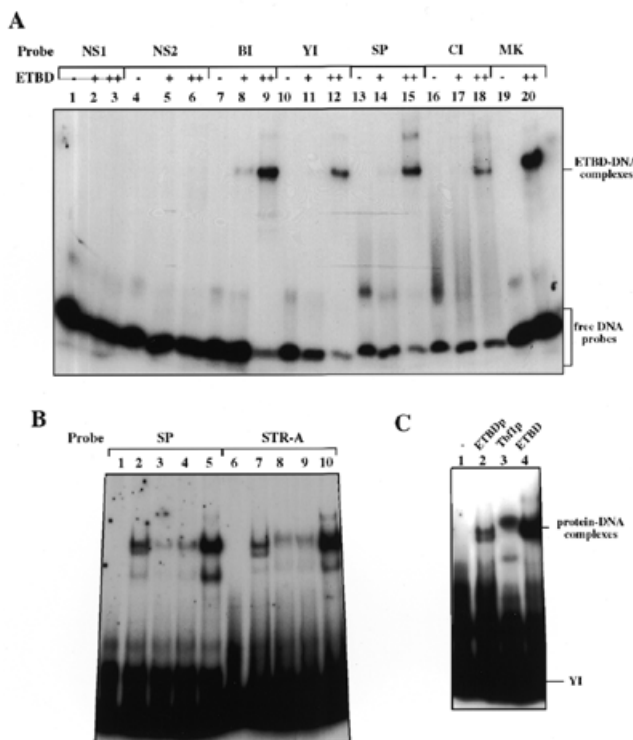


Figure 6. Band shift assays of putative Tbf1p-binding sites. (A) Probes corresponding to the genomic sequences shown in Figure 6 were mixed with increasing amounts of ETBD bacterial extracts (-, none; +, 3 µg; ++, 15 µg) and analysed by band shift assay. (B) The indicated probes (SP and STR-A) were analysed by band shift assay without protein (lanes 1 and 6), with highly purified ETBD (ETBDp, 150 ng) (lanes 2 and 7) and with increasing amounts of the ETBD extracts (3, 15 and 30 µg) (lanes 3-5 and 8-10). (C) The YI probe was mixed with either 150 ng of ETBDp or 1 µl of reticulocyte lysate containing Tbf1p or 15 mg of ETBD extract.

matches to the Tbf1p consensus sequence can be found in these regions, with respect to the AGGG core sequence (designated Tbf1^{Y'a}, Tbf1^{Y'b}, Tbf1^{Y'c}, Tbf1^{X'a} and Tbf1^{X'b} in Table 2; data not shown). These motifs probably account for most of the observed Tbf1p binding to sub-telomeric motifs (1,8,9). Therefore, it is likely that efficient Tbf1p-binding sites can tolerate some sequence variations from the consensus. Exact matches to the sub-telomeric variants of the Tbf1p consensus are found at internal chromosomal locations (data not shown). This suggests the presence of other sets of Tbf1p sites scattered throughout the genome. Interestingly, the occurrences of sub-telomeric Tbf1p motifs are significantly enriched in intergenic regions, reaching 58% of hits (Table 2). This distribution is clearly non-random, since the proportion of intergenic regions in the yeast genome is roughly 22% (26). It is also clearly different from the profile of the Tbf1p consensus perfect matches, which does not appear to exhibit any particular bias towards intergenic region (Table 2). This raises the possibility that subsets of Tbf1p-binding sites have different roles in the genome.

Comparison of the distribution of Tbf1p and Reb1p sites

Since Tbf1p and Reb1p sites are found associated in the yeast sub-telomeric regions, we asked whether a significant number of potential Tbf1p sites co-localise with Reb1p sites. We made

use of a Reb1p consensus sequence, previously defined by a PCR selection procedure (11). Searching for the Reb1p-binding sequence CCGGGTAAAC, 44 occurrences were found (Fig. 5 and Table 2). As for the Tbf1p sites, the known sub-telomeric Reb1p sites do not perfectly match the Reb1p consensus. Therefore, other searches using sub-telomeric variants of the Reb1p sites with respect to the GGGTAA core sequence (designated Reb1^{Y'}, Reb1^{X'a}, Reb1^{X'b} and Reb1^{X'c}) were performed (Table 2). A comparison of the distribution of non-sub-telomeric Tbf1p and Reb1p sites failed to reveal any co-localisation (<500 nt) of these sites (data not shown). This is in contrast to the sub-telomeric regions (9,27). Furthermore, this analysis did not reveal any internal chromosomal regions containing multiple Reb1p or Tbf1p sites.

Strikingly, the Reb1p consensus or its sub-telomeric variants are found almost exclusively in intergenic regions (Table 2). This non-random distribution of Reb1p sites does not result from a particular nucleotide composition of the intergenic regions, since a sequence corresponding to the mirror symmetry of the Reb1p sequence used in the computer search (Reb1M) does not show any bias towards intergenic regions (Table 2). Furthermore, if one mismatch is tolerated during the search, the number of occurrences greatly increases, but still displays a strong bias towards intergenic regions, reinforcing the significance of the non-random distribution observed for the perfect matches (Table 2). This bias towards intergenic regions was also observed for the sub-telomeric variants of the Tbf1p consensus, although to a lesser extent (see above). On the other hand, the bias is not observed for putative sites of two other yeast DNA-binding proteins, Rap1p and Abf1p (Table 2). The implications of this intriguing distribution of Tbf1p and Reb1p sites are considered in the Discussion.

DISCUSSION

The yeast protein Tbf1p, encoded by an essential gene, *in vitro* binds TTAGGG-like repeats which are found in sub-telomeric regions of *S.cerevisiae* chromosomes (1,8,9). In this paper we have investigated the localisation of Tbf1p in the yeast genome. We demonstrate that the Tbf1p DNA-binding domain (TBD) can target the Gal4 transactivation domain to a (TTAGGG)₁₀ sequence to activate transcription *in vivo*. Since the hybrid protein is expressed at a low level (unpublished observation), this transactivation assay is expected to reflect the *in vivo* Tbf1p binding behaviour. Strikingly, the level of transactivation varies according to the orientation of the (TTAGGG)₁₀ sequence (Fig. 1). This effect could simply be due to differences in the junction sequence between the telomeric insert and the rest of the construct, altering the susceptibility of the promoter to be activated. Alternatively, and in agreement with the fact that the telomeric repeats have the same orientation, the effect of Tbf1p binding could be polar, being more capable of transactivation in one orientation than in the other. In any case, these results strongly suggest that the TTAGGG repeats present in sub-telomeric regions of yeast chromosomes are target sites for Tbf1p. This probably accounts for the Tbf1p spots visible at the end of some pachytene chromosomes (Fig. 2). Furthermore, the sequence characteristics of high affinity Tbf1p-binding sites have been investigated by a PCR-based selection procedure. Since the data set consisted of 50 independently selected oligonucleotides and since all of

35 randomly chosen oligonucleotides were confirmed to specifically bind Tbf1p, it is likely that the final consensus TAGGGTTGG faithfully predicts high affinity Tbf1p sites. Indeed, five regions of the yeast genome containing this sequence were confirmed to bind Tbf1p efficiently.

The gene encoding Tbf1p was cloned through its ability to bind the vertebrate type telomeric sequence (TTAGGG)_n (1). In agreement with this property of Tbf1p, the consensus identified in this work can be aligned thanks to only one mismatch to TAGGGTTAG, a sequence included in (TTAGGG)_n. Interestingly, this TAGGGTTAG sequence is identical to the binding site determined for a monomeric telobox domain of TRF1 (4) and highly similar to the YTAGGGTTTR consensus sequence for the full-length TRF1 dimer, as identified by a SELEX procedure (28). These data, together with the recent characterisation of Teb1p, a *S.pombe* telobox protein binding with high affinity to (TTAGGG)_n sequences (7), fully support the proposal that teloboxes bind with high affinity to TTAGGG-like sequences *in vivo* (3,6).

Using the Tbf1p consensus sequence determined in this work, an overall 52 Tbf1p sites were detected in the entire yeast genome, distributed over 15 of the 16 chromosomes. Since only perfect matches to the consensus were sought, a number of Tbf1p sites may have escaped this analysis. For instance, known Tbf1p sites present in sub-telomeric regions were not selected. Motifs closely related to the consensus are found repeated in these regions, which could cooperate to create a high affinity Tbf1p-binding area. By using the perfect consensus, our search is likely to reveal only a subset of the Tbf1p sites in the genome, but will be of high predictive value. The binding of Tbf1p to these sites can be modulated *in vivo* by the presence of chromatin. For example, it is not known if Tbf1p recognises its site when the DNA is wrapped around the nucleosome. Nonetheless, the large number of signals seen in the immunofluorescence studies (Fig. 2) suggests that the yeast genome contains a large number of Tbf1p sites, including those in the sub-telomeric regions.

In the sub-telomeric regions Tbf1p sites are associated with Reb1p sites to form regions that limit the propagation of silent chromatin emanating from the proximal telomere (STARs) (9). This does not seem to be the case for the putative Tbf1p and Reb1p sites found at internal chromosomal locations. For example, no putative Tbf1p site was found in the rDNA repeats, consistent with a lack of nucleolar staining with Tbf1p antibodies in interphase nuclei (Fig. 2). In contrast, it is known that Reb1p binds and is required for expression of the rDNA transcription unit (29), as confirmed by our database search.

The putative Tbf1p and Reb1p sites detected in the yeast genome are found enriched in the intergenic regions. This is particularly pronounced for the Reb1p sites (Table 2). This is in contrast to a more even distribution between the intra- and intergenic regions of potential Rap1p and Abf1p and probably other transcriptional regulators. Indeed, transcriptional regulatory elements can be found within coding regions in yeast (30,31). We believe that the striking bias of Tbf1p or Reb1p sites towards intergenic regions reflects a particular function of these proteins. For example, the binding of Reb1p could create pause sites during elongation by various types of RNA polymerases (32). Because of the capacity of Tbf1p and Reb1p to mediate STAR insulator properties we speculate that the

intergenic localisation of Tbf1p and Reb1p might reflect the delimitation of domains of independent gene expression.

ACKNOWLEDGEMENTS

We are especially grateful to S. Gasser for continuous encouragement and suggestions as well as for critical reading of the manuscript. We thank D. Shore and J. Broach for gifts of plasmids and strains as well as H. Gruffat and A. Sergeant for their help with the Casting procedure. The work in the E. Gilson laboratory is supported by grants from the Ligue Nationale contre le Cancer. F.K. was supported by Austrian Science Foundation grant S8203. The immunostaining of interphase cells was performed in S.M. Gasser's laboratory by T.L., who thanks the Swiss National Science Foundation for continued support.

REFERENCES

- Brigati, C., Kurtz, S., Balderes, D., Vidali, G. and Shore, D. (1993) *Mol. Cell. Biol.*, **13**, 1306–1314.
- Cooper, J.P., Nimmo, E.R., Allshire, R.C. and Cech, T.R. (1997) *Nature*, **385**, 744–747.
- Bilaud, T., Koering, C.E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S.M. and Gilson, E. (1996) *Nucleic Acids Res.*, **24**, 1294–1303.
- Konig, P., Fairall, L. and Rhodes, D. (1998) *Nucleic Acids Res.*, **26**, 1731–1740.
- Bilaud, T., Brun, C., Ancelin, K., Koering, C.E., Laroche, T. and Gilson, E. (1997) *Nature Genet.*, **17**, 236–239.
- Brun, C., Marcand, S. and Gilson, E. (1997) *Trends Cell Biol.*, **7**, 317–324.
- Vassetzky, N.S., Gaden, F., Brun, C., Gasser, S.M. and Gilson, E. (1999) *Nucleic Acids Res.*, **27**, 4687–4694.
- Liu, Z. and Tye, B.K. (1990) *Genes Dev.*, **5**, 49–59.
- Fourel, G., Revardel, E., Koering, C.E. and Gilson, E. (1999) *EMBO J.*, **18**, 2522–2537.
- Ju, Q.D., Morrow, B.E. and Warner, J.R. (1990) *Mol. Cell. Biol.*, **10**, 5226–5234.
- Liaw, P.C. and Brandl, C.J. (1994) *Yeast*, **10**, 771–787.
- Doorenbosch, T., Mager, W.H. and Planta, R.J. (1992) *Gene Expr.*, **2**, 193–201.
- Wang, H., Nicholson, P.R. and Stillman, D.J. (1990) *Mol. Cell. Biol.*, **10**, 1743–1753.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Rose, M.D., Winston, F. and Hieter, P. (1990) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Guarente, L. and Mason, T. (1983) *Cell*, **32**, 1279–1286.
- Gilson, E., Müller, T., Sogo, J., Laroche, T. and Gasser, S.M. (1994) *Nucleic Acids Res.*, **22**, 5310–5320.
- Sikorski, R.S. and Hieter, P. (1989) *Genetics*, **122**, 19–27.
- Klein, F., Laroche, T., Cardenas, M.E., Hofmann, J.F.X., Schweizer, D. and Gasser, S.M. (1992) *J. Cell Biol.*, **117**, 935–948.
- Gruffat, H. and Sergeant, A. (1994) *Nucleic Acids Res.*, **22**, 1172–1178.
- Li, J.J. and Herskowitz, I. (1993) *Science*, **262**, 1870–1874.
- Runge, K.W. and Zakian, V.A. (1990) *Nucleic Acids Res.*, **18**, 1783–1787.
- Guacci, V., Hogan, E. and Koshland, D. (1994) *J. Cell Biol.*, **125**, 517–530.
- Gotta, M., Strahl-Bolsinger, S., Renaud, H., Laroche, T., Kennedy, B.K., Grunstein, M. and Gasser, S.M. (1997) *EMBO J.*, **16**, 3243–3255.
- Wright, W.E., Binder, M. and Funk, W. (1991) *Mol. Cell. Biol.*, **11**, 4104–4110.
- Dujon, B. (1996) *Trends Genet.*, **12**, 263–270.
- Chasman, D.I., Lue, N.F., Buchman, A.R., LaPointe, J.W., Lorch, Y. and Kornberg, R.D. (1990) *Genes Dev.*, **4**, 503–514.
- Bianchi, A., Stansel, R.M., Fairall, L., Griffith, J.D., Rhodes, D. and de Lange, T. (1999) *EMBO J.*, **18**, 5735–5744.
- Planta, R.J., Goncalves, P.M. and Mager, W.H. (1995) *Biochem. Cell Biol.*, **73**, 825–834.
- Sinclair, D.A., Kornfeld, G.D. and Dawes, I.W. (1994) *Mol. Cell. Biol.*, **14**, 214–225.
- Fantino, E., Marguet, D. and Lauquin, G.J. (1992) *Mol. Gen. Genet.*, **236**, 65–75.
- Lang, W.H., Morrow, B.E., Ju, Q., Warner, J.R. and Reeder, R.H. (1994) *Cell*, **79**, 527–534.