Use of a selection technique to identify the diversity of binding sites for the yeast RAP1 transcription factor

Ian R.Graham and Alistair Chambers*
Department of Genetics, University of Nottingham, Queen's Medical Centre,
Nottingham NG7 2UH, UK

Received November 18, 1993; Revised and Accepted December 10, 1993

ABSTRACT

We have used the technique known as selected and amplified binding (SAAB) to isolate binding sites for the yeast transcription factor RAP1 from a degenerate pool of oligonucleotides. A total of 47 sequences were isolated, of which two were shown to be contaminating non-RAP1 binding sites. After excluding these two sequences the remainder of the sequences were used to derive a new consensus binding site for RAP1. The new consensus 5′ A/G T A/G C A C C C A N N C C/A C C 3′ is a significant extension of the existing consensus (4). It is longer by two base pairs at the 5′ end and is significantly more constrained at the 3′ end. An analysis of the combinations of mis-matches in individual SAAB sequences, compared to the consensus RAP1 binding site, has allowed us to analyse the structure of the RAP1 binding site in some detail. The binding site can be sub-divided into three regions; a core binding site, a 5′ flanking region and a 3′ flanking region. The core binding site, consisting of the sequence 5′CACCCA3′, is critical for recognition by RAP1. The less conserved flanking regions are not as important. Interactions between RAP1 and these regions probably stabilise the interaction between RAP1 and the core binding site. Each of the sequences isolated in the SAAB analysis was used to search release 78 of the EMBL + GenBank DNA data base. The searches identified 102 potential binding sites for RAP1 within promoters of yeast genes.

INTRODUCTION

The yeast transcription factor RAP1 (1) (also known as TUF (2, 3) and GRF1 (4)) is a key multi-functional protein in the yeast cell. It plays a role in activating transcription of a large number of yeast genes, including genes encoding glycolytic enzymes, components of the translational machinery and other proteins involved in general house-keeping activities (5–7). In addition to its role in activating transcription, RAP1 is also involved in the opposite process, transcriptional silencing. It interacts with the silencers at *HML* and *HMR,* where it is one component of the protein assembly which mediates repression of the silent mating type loci (1, 8–10). RAP1 is also important in maintaining chromosome structure. It binds to the $C_{(1-3)}A$ repeat sequence found at yeast telomeres and it appears to be critical for the maintenance of the correct length of this telomeric repeat sequence (4, 9, 11–15).

Although RAP1 binds strongly to DNA, little is known about the details of this interaction. The DNA binding domain of the protein has been characterised by a series of deletion experiments *in vitro* . It extends from position 361 to position 596 in a primary amino acid sequence of 827 amino acids (16). This is a large region of the protein but it contains no significant homologies to previously identified DNA binding motifs (16, 17).

RAP1 has been shown to interact *in vitro* with a variety of gene promoters, sequences from the silent mating type loci and telomeres (1, 3, 4, 6, 7, 11–14, 18, 19). RAP1, under the name TUF, was originally proposed to interact with two conserved sequences, designated HOMOL1 (5′ A A C A T C T/C A/G T A/G C A 3′) and the RPG box (5′ A C C C A T A C A T T T/A 3′) located in the promoters of genes encoding ribosomal proteins (2, 5). These sequences were later combined to produce a high affinity RPG box, 14 bp in length, consisting of the sequence 5′ A C A C C C A T A C A T T T 3′(5, 20). On binding to DNA, TUF (RAP1) has been shown to induce DNA bending upstream of the RPG box (20). The effect of individual point mutations within the RPG box on the binding and bending activities of TUF (RAP1) has been investigated (20). These experiments demonstrated that a point mutation at position 1 in the RPG box had little effect on the strength of TUF (RAP1) binding but that mutations at positions 3, 7 and 10 reduced the strength of binding by between 88% and 96% (20).

A consensus high affinity binding site for RAP1 (under the name GRF1) has been derived by comparison of RAP1 binding sites at *HMR E*, *HML E*, *Mat*α UAS, *TEF2* UAS and a telomeric binding site with the original RPG box (4). This 13 bp consensus 5′ A/G A/C A C C C A N N C A T/C T/C 3′ has been widely adopted as a description of the RAP1 binding site. However, although this sequence may be a good description of a strong RAP1 binding site it is clear that not all strong RAP1 binding sites are perfect matches to this consensus. For example, the strong RAP1 binding site at *HMR E* consists of the sequence 5′ A A A C C C A T C A A C C 3′(1). This has an A instead

* To whom correspondence should be addressed

of a C at the 10th position of the consensus and requires either a mismatch, or a gap plus a mismatch, to align it to the existing consensus.

There are other known RAP1 binding sites which do not conform completely to the established consensus. UAS1 of the *PMA1* gene contains a RAP1 binding site which has a T instead of A/G at the 1st position, whilst UAS2 of the same gene has a RAP1 binding site with a G instead of T/C at the 13th position (7). The UAS of the *PGK* gene contains a strong RAP1 binding site which has a G at the 13th position instead of T/C (6). The *HML E* silencer and the *HIS4* gene promoter contain RAP1 binding sites which have an A instead of T/C at the 13th position (1, 19). More dramatic variations also exist, a recently described RAP1 binding site within the coding region of the *SRP1* gene has three mismatches to the current consensus at the 11th, 12th and 13th positions (21).

In order to investigate the interaction between RAP1 and its binding site in more detail we have undertaken a detailed analysis of the DNA binding site for RAP1 using the technique known as selected and amplified binding (SAAB), also known as CAST (cyclic amplification and selection of targets) and SELEX (systematic evolution of ligands by exponential enrichment). This technique has been used to identify the binding sites for the *Drosophila* morphogen *dorsal* (22), the yeast transcription factor MCM1 (23), and the mammalian transcription factors SRF (24), E12 (25), E2A and MyoD (26). Using the SAAB technique, we have identified a range of binding sites for RAP1 *in vitro* from which we have developed a new consensus RAP1 binding site and a set of rules for predicting whether RAP1 will interact with any particular sequence. We have searched the EMBL + GenBank DNA database to identify close matches to each of the SAAB sequences we isolated and we have applied the new set of rules to predict which of these sequences are good candidates for RAP1 binding sites.

MATERIALS AND METHODS

Strains and media

The yeast strain used throughout this series of experiments was *Saccharomyces cerevisiae* DBY745 (α *ade-100 leu2-3 leu2-112 ura3-52*), grown in YEPD medium (27).

Plasmid manipulations were carried out using *Escherichia coli* MC1061 (*F⁻ araD139* Δ(ara-leu)7696 Δ(lac)X74 galU galK hsdR2 ($r_K^- m_K^+$) mcrB1 rpsL (Str^r)), grown in LB medium, with or without ampicillin (50 μg/ml) (28).

Selection and amplification of RAP1 binding sites

Two cycles of SAAB were performed, largely as described by Blackwell and Weintraub (26). Briefly, a pool of oligodeoxynucleotides of 57 nt in length were synthesised, such that there was total degeneracy at the central 13 nt. This 'N13' sequence was flanked by restriction sites, and by binding sites for a pair of PCR oligonucleotides (see below). Following second strand synthesis, the DNA was labelled using [γ-^{32}P] ATP, in the presence of T4 polynucleotide kinase, and subjected to gel retardation analysis, using RAP1 produced by *in vitro* translation (IVT). A fragment of the resulting dehydrated gel, which comigrated with the RAP1-specific complex in a positive control lane, was excised, and the DNA eluted. The 56 bp positive control used comprised a 48 bp fragment of the promoter from the yeast *RAP1* gene, extending from positions −98 to −51, and 8 bp of adjacent polylinker sequence. We have previously shown that this fragment contains a binding site for RAP1 (IRG and AC, submitted).

Thirty cycles of PCR were performed using SP6 promoter (5′-GATTTAGGTGACACTATAG-3′) and pUC/M13 reverse (5′-AACAGCTATGACCAT-3′) primers. After the second SAAB cycle, the amplified material was digested with *Bam* HI and *Bgl* II, cloned into the polylinker of pSP46 (29), and transformed into *E.coli*. Fifty colonies were selected at random, and the plasmid DNA from them was analysed by the dideoxy sequencing method (30).

Preparation of yeast total protein extracts

DBY745 were grown to mid-log phase (4 to 6×10^6 cells per ml) in YEPD medium (27). Pelleted cells were washed twice with 25 mM sodium phosphate (pH 7.5), then resuspended in PMSF (1 mM in 25 mM sodium phosphate, pH 7.5; 0.3 ml). Subsequent manipulations were performed at 4°C in 1.5 ml microcentrifuge tubes. Glass beads (0.4 mm diameter; BDH Merck) were added to within 3 mm of the surface of the liquid, then vortexed for 90 seconds. The beads and cell debris were pelleted by 10 seconds at 13000×g in a MicroCentaur (MSE). The supernatant was transferred to a fresh tube, and the residue washed once with 0.2 ml PMSF solution, as above. Following another 10 second spin, the supernatant fractions were pooled, then spun at 13000×g for 10 minutes. The supernatant was transferred to a fresh tube, then assayed for protein content (31).

DNA fragments

N13 clones were digested with *Eco* RI and *Bgl* II to give fragments of 37 bp in length. For use in gel retardation assays, each of these fragments was labelled using [γ-^{32}P] ATP (>5000 Ci/mmol; Amersham International plc) and T4 polynucleotide kinase (Life Technologies, Inc.).

Gel retardation analysis

Gel retardation analysis was performed on 10000 cpm of each DNA fragment, as described previously (6), using an aliquot of a yeast total protein extract containing 1−2 μg of protein. In experiments using IVT RAP1, 1−2 μl of rabbit reticulocyte lysate, primed with RNA encoding RAP1, was used (6).

RESULTS

Isolation of RAP1 binding sites by SAAB

We have performed a two-cycle SAAB analysis of RAP1 binding sites, using RAP1 produced by *in vitro* translation (IVT RAP1) to select oligonucleotides from a degenerate pool (N13). The IVT RAP1 has previously been shown to interact with a range of binding sites with the same specificity as RAP1 in yeast total protein extracts (6). RAP1 was incubated with the radioactively labelled pool of oligonucleotides and RAP1 binding sites selected by a gel retardation assay. Due to the low abundance of RAP1 binding sites in the degenerate oligonucleotide pool, a radioactive complex was not detected in the initial cycle of the analysis. We had to rely on a positive control fragment of similar size, containing a known RAP1 binding site, to localise the position of the RAP1−N13 complex. However, the selection which had taken place in this first cycle resulted in a RAP1−N13 complex in the subsequent cycle which was visible after overnight autoradiography (data not shown). The oligonucleotides selected by the two rounds of SAAB were sub-cloned into pSP46 (29) and sequenced.

Table 1. Alignment of sequences derived by two cycles of SAAB

Position:								*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*		
Consensus:								**A/G**	**A/C**	**A**	**C**	**C**	**C**	**A**	**N**	**N**	**C**	**A**	**T/C**	**T/C**		
1 MISMATCH																						
N13.01						t	c	c	A	A	C	C	C	A	C	A	C	A	C	C		
N13.17						a	t	c	c	A	C	C	C	A	G	G	C	A	C	C	C	G
N13.30			A	C	C	A	T	A	C	A	C	C	C	A	G	a	g	a	t	c		
1 MISMATCH + 1 BULGE																						
N13.07						a	t	c	c	A	C	C	C	A	G	AA	C	A	T	T	T	
2 MISMATCHES																						
N13.11			G	C	T	A	A	C	C	A	C	C	C	A	G	a	g	a	t	c		
N13.12	G	A	G	C	C	T	A	A	C	A	C	C	C	a	g	a	t	c	t	t		
N13.25						a	t	c	c	A	C	C	C	C	T	G	C	A	T	C	G	
N13.31						c	t	G	C	A	C	C	C	A	G	A	C	C	C	a		
N13.35						a	t	c	c	G	C	C	C	A	T	A	C	A	T	T	C	C
N13.36						a	t	c	c	A	C	C	C	A	T	T	C	A	A	C	C	C
N13.42						a	t	c	c	A	C	C	C	A	A	C	C	C	C	C	G	A
N13.43	C	A	A	C	G	A	T	A	C	A	C	C	C	a	g	a	t	c	t	t		
N13.45						a	t	c	c	A	C	C	C	A	C	G	C	A	C	G		
N13.47						t	c	c	C	A	C	C	C	A	G	G	C	C	C	C	A	
2 MISMATCHES + 1 GAP																						
N13.29						G	G	A	C	A	C	C	□	A	C	C	T	A	C	a		
3 MISMATCHES																						
N13.20					G	T	G	A	C	A	C	C	C	A	G	A	T	a	g	a		
N13.03					C	G	T	A	C	A	C	C	C	A	C	C	a	g	a	t		
N13.16						G	T	A	C	A	C	C	C	A	A	C	C	C	a	g		
N13.24						a	t	c	c	G	C	C	C	A	T	A	C	A	C	A	A	C
N13.39						G	T	G	C	A	C	C	C	A	T	A	A	A	a	g		
3 MISMATCHES + 1 GAP																						
N13.13						G	T	G	C	A	C	C	□	A	C	C	C	T	G	a		
N13.14						G	T	G	C	A	C	C	□	A	G	A	C	T	G	a		
N13.19						G	T	G	C	A	C	C	□	A	T	C	A	G	T	a		
N13.26						A	C	A	C	A	C	C	□	A	C	G	C	C	A	a		
N13.28						A	T	A	C	A	C	C	□	A	C	G	C	C	a	g		
N13.33						c	c	G	G	G	C	C	□	A	G	A	C	A	T	T	C	
N13.40						A	G	A	C	A	C	C	□	A	C	T	A	C	C	a		
3 MISMATCHES +1 BULGE																						
N13.06						G	T	G	C	AT	C	C	T	C	C	a	g	a	t	c		
N13.15						G	T	A	C	A	C	CG	C	A	T	G	T	a	g	a		
4 MISMATCHES																						
N13.02						A	G	A	C	A	C	C	C	T	A	G	C	T	a	g		
N13.04	C	A	T	C	A	C	T	A	C	A	C	C	C	a	g	a	t	c	t	a		
N13.05	C	C	G	T	G	A	A	A	C	A	C	C	C	a	g	a	t	c	t	a		
N13.08						a	t	c	c	G	C	C	C	A	G	A	C	C	G	C	C	T
N13.18						G	T	G	C	A	C	C	G	A	C	G	C	C	a	g		
N13.22						G	T	G	C	A	C	C	C	T	G	C	C	T	a	g		
N13.23						G	T	G	C	A	C	C	C	C	G	C	G	A	a	g		
N13.27						G	T	G	C	A	C	C	C	G	T	A	C	C	a	g		
N13.44						G	G	A	C	A	C	C	T	A	C	G	G	a	g	t		
N13.48						a	t	c	c	A	C	C	C	A	A	T	A	C	C	A	A	T
4 MISMATCHES + 1 GAP																						
N13.34						G	T	G	C	A	C	C	□	A	T	G	T	T	a	g		
5 MISMATCHES																						
N13.09	G	C	T	G	C	A	G	C	C	G	C	C	C	g	g	a	t	c	c	c		
N13.32						G	T	A	C	C	C	C	C	G	C	C	T	A	a	g		
N13.41	A	T	A	C	A	T	A	C	G	A	C	C	a	g	a	t	c	t	t	t		
N13.37	A	G	A	C	A	T	C	T	T	A	C	C	G	a	g	a	t	c	t	t		
5 MISMATCHES + 1 GAP																						
N13.46						a	t	c	c	G	C	C	□	A	T	A	G	A	G	C	A	T
NON-BINDERS																						
N13.21						G	A	G	C	A	A	T	C	A	A	A	A	C	g	g		
N13.38						T	G	G	C	A	C	G		A	C	T	T	C	g	g	a	t

Products of the PCR step in the second cycle of a SAAB procedure were digested with *Bam* HI and *Bgl* II, and cloned into the polylinker of plasmid pSP46. The sequences of 47 recombinant clones were determined. A manual alignment was performed, with the CCC or CC motif present in the sequences being used as a weight. Gaps and bulges (shown as open boxes) were introduced to improve the fit of a particular sequence to the established RAP1 consensus (shown in line 2). The number of mismatches between the consensus and each sequence was determined, and the sequences grouped according to this classification. Positions 1 to 13 of the consensus are referred to in the text. Nucleotides which differ from the consensus are shaded.

A total of 50 clones were analysed, out of which 47 were found to have inserts containing SAAB-derived sequences.

Alignment of SAAB derived sequences and comparison to the established consensus RAP1 binding site

Using the doublet or triplet C as a weight, all 47 sequences were aligned (Table 1). The alignment was performed by eye, since multiple alignment packages on computer were found to be unable to align this number of sequences faithfully. Six of the 47 sequences contained two copies of the CCC sequence. In each of these cases, the triplet closest to the 5′ end of the sequence was used as the basis for the alignment. This allowed the best fit of a RAP1 binding site within the small amount of sequence available in these oligonucleotides. Whilst all of the cloned sequences contained short regions corresponding to the restriction sites flanking the degenerate region of the N13 oligonucleotide (shown in lowercase letters in Table 1), which may have some influence on the strength of a particular site, these regions were ignored for the purposes of the alignment. Each of the aligned sequences was then compared to the established consensus binding site for RAP1 (5′ A/G A/C A C C C A N N C A T/C T/C 3′) and the number of mismatches to the consensus identified for each sequence. In Table 1, the sequences are listed in order of similarity to the established consensus RAP1 binding site, with the mismatched bases indicated.

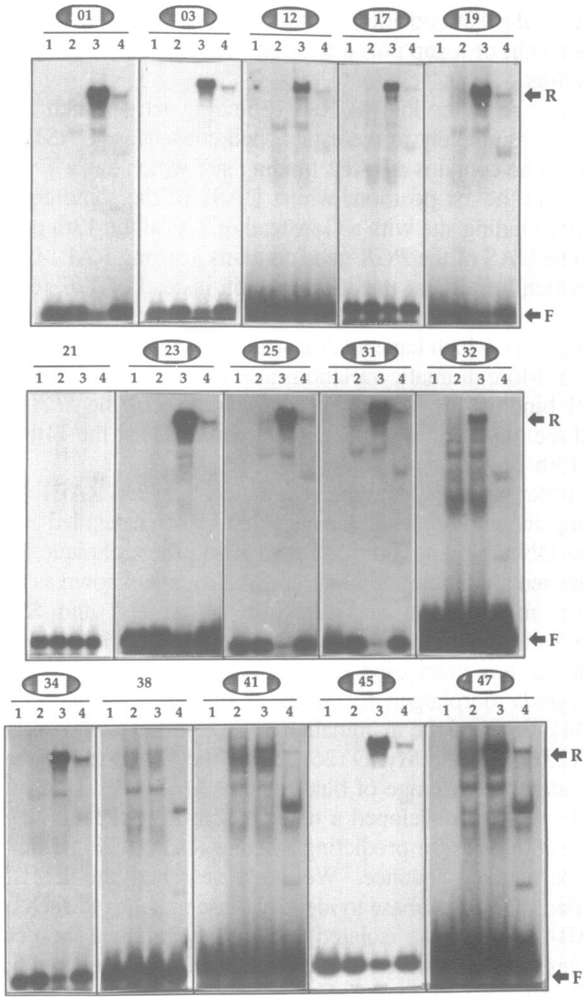

Figure 1. Gel retardation analysis of potential RAP1 binding sites. Radioactively labelled DNA fragments were incubated alone (lanes 1), with a mock lysate (lanes 2), with a lysate primed with *RAP1* mRNA (lanes 3), or with 1 μg of a total protein extract of yeast (lanes 4). Samples treated in this way (indicated above the appropriate lanes) are as designated in Table 1. Fragments which bind RAP1 in these assays are indicated by shading. The positions of migration of unbound DNA (F) and RAP1–DNA (R) complexes are indicated.

Interactions between SAAB isolated sequences and RAP1

Because some of the sequences isolated as a result of the SAAB were markedly different from the established consensus binding

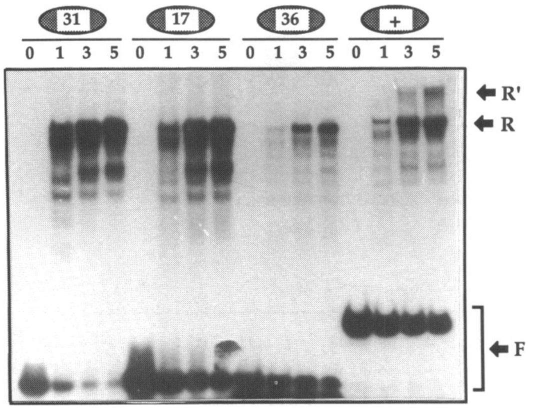

Figure 2. Analysis of the potential for duplicate RAP1 binding sites. Radioactively labelled fragments of DNA containing two CCC motifs were incubated alone (lanes 0), or in the presence of 1 μg, 3 μg or 5 μg of IVT RAP1 (lanes 1, 3 and 5, respectively). SAAB sequences treated in this way are indicated above the appropriate lanes as designated in Table 1. A positive control fragment for double RAP1 binding, isolated from the *RAP1* promoter, was also used (Panel +). The positions of migration of unbound DNA (F), and of complexes formed between the labelled fragment and one molecule or two molecules of RAP1 (R and R', respectively) are indicated.

site for RAP1, we were concerned to determine the level of contaminating non-RAP1 binding sites in the set of sequences we had isolated. To do this, we selected a total of ten of the SAAB sequences (N13.01, N13.12, N13.19, N13.23, N13.25, N13.31, N13.32, N13.41, N13.45 and N13.47) at random, and tested each individually for RAP1 binding. Each sequence was tested for binding by gel retardation assay using both IVT RAP1 and RAP1 in a yeast total protein extract (Figure 1). IVT RAP1 was found to interact with all ten of these randomly selected sequences to produce a DNA-protein complex (R) (Figure 1, lanes 3). This complex was not produced in control assays using a mock IVT lysate not primed with *RAP1* mRNA (Figure 1, lanes 2). All of the sequences also formed a complex of identical mobility in the yeast total protein extract (Figure 1, lanes 4). This complex is almost certainly the result of RAP1 within the extract interacting with the SAAB sequence.

Some variation was observed between the strengths of the RAP1 interactions with the different sequences, and whilst these assays are not quantitative, they do give some idea as to the relative strengths of the RAP1 binding sites. Because ten sites selected at random were all found to interact with RAP1 we concluded that the level of non-RAP1 binding sites isolated was very low.

When all of the seqences were arranged in order of similarity to the established RAP1 consensus binding site it was observed that although the majority of them were similar to the consensus, a number of them had five or six mismatches (Table 1). We therefore tested another selection of the SAAB sequences to include two of these extreme variants, N13.38 and N13.21. N13.38 has five mismatches and a gap when compared to the established consensus whilst N13.21 has six mismatches.

We also tested three other selected sequences: N13.17, N13.03 and N13.34. N13.17 has one mismatch to the established consensus, N13.03 has three mismatches and N13.34 has four mismatches plus a gap. We performed gel retardation assays using IVT RAP1 and using a yeast total protein extract (Figure 1). Incubation of these SAAB sequences with IVT RAP1 gave rise to varying levels of protein−DNA complexes. Sequences N13.03, N13.17 and N13.34 all gave strong complexes (R) with IVT RAP1 (Figure 1, lanes 3), which were not seen when these DNA fragments were incubated with a control IVT extract lacking RAP1 (Figure 1, lanes 2). Sequences N13.21 and N13.38 did not form detectable RAP1 specific complexes (Figure 1, lanes 3). The results upon incubation of each of the sequences with yeast total protein extract confirmed the results obtained with IVT RAP1. Sequences N13.03, N13.17 and N13.34 all produced a retarded complex (R) with identical mobility to the complex they produced with the IVT RAP1 (Figure 1, compare lanes 3 and 4). In each case this complex is almost certainly the result of binding by RAP1 present in the total protein extract. Sequences N13.21 and N13.38 did not produce detectable levels of this complex.

Six of the sequences isolated by the SAAB contain two copies of the sequence CCC which has previously been shown to be highly conserved at positions 4 to 6 of the established consensus binding site for RAP1 (4). This raised the possibility that these SAAB sequences could contain two different binding sites for RAP1. We have tested three of the SAAB sequences (N13.31, N13.17 and N13.36) to determine whether this is the case. Higher than normal concentrations of IVT RAP1 were incubated with the individual SAAB sequences before standard gel retardation assays. When a DNA fragment containing two binding sites for RAP1 and a high concentration of RAP1 are used in gel retardation assays, it is possible to produce a super-shifted complex consisting of two RAP1 molecules bound to the same DNA fragment. In our experiments, a control fragment from the promoter of the *RAP1* gene, containing two RAP1 binding sites, was used (IRG and AC, submitted). On incubation with the standard amount of IVT RAP1 extract (1μg) this control fragment (+) formed a single DNA−protein complex (R) (Figure 2, lane 1). On incubation with 3μg and 5μg of IVT RAP1 extract, the control fragment produced a supershifted complex (R') (Figure 2, lanes 3 and 5). None of the three SAAB sites used in this experiment gave rise to a supershifted complex, even in the presence of 5μg of IVT RAP1 extract (Figure 2, lanes 5). This suggested that if these fragments do contain two RAP1 binding sites these sites cannot be occupied simultaneously.

Potential RAP1 binding sites in the yeast genome

The sequence from each SAAB clone (except for the two known non-binders) which contained the best alignment to the new consensus RAP1 binding site, and which was derived from the degenerate region of the oligonucleotide, was used to search Release 78 of the EMBL+GenBank database. The database entries established by this search were subsequently analysed for the presence of the SAAB-derived sequences in regions upstream of the coding sequence, since we were most interested in the presence of potential RAP1 binding sites in promoter regions. In this way, we identified 145 yeast genes which have sequences closely related to one or more of the SAAB clones in their upstream regions. Each of these sequences was then analysed using the set of rules outlined in the discussion to produce a list of potential RAP1 binding sites in the yeast genome which are likely targets for RAP1 *in vivo*.

DISCUSSION

We have isolated 47 sequences by amplification of the DNA component of a complex formed between the yeast transcription factor RAP1 and a pool of degenerate oligonucleotides.

Table 2. Conservation of nucleotides at each position of the consensus

Position	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
Consensus	A/G	A/C	A	C	C	C	A	N	N	C	A	T/C	T/C
Mismatches	17	4	7+1B	0	1B	5+10G	9	-	-	23	25	22	25

Following the alignment shown in Table 1, the number of sequences that differ from the established consensus (line 2) at each position was determined (line 3). B, bulged nucleotide; G, gapped position. The two sequences in Table 1 which were shown not to interact with RAP1 were excluded from this analysis.

Table 3. Comparison of SAAB sequences identified as RAP1 binders

							-2	*-1*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*		
N13.01										A	A	C	C	C	A	C	A	C	A	C	C		
N13.03						C	G	T	A	C	A	C	C	C	A	C	C						
N13.12	G	A	G	C	C	T	A		A	C	A	C	C	C									
N13.17											A	C	C	C	A	G	G	C	A	C	C	C	G
N13.19							G	T	G	C	A	C	C		A	T	C	A	G	T			
N13.23							G	T	G	C	A	C	C	C		C	G	C	G	A			
N13.25											A	C	C	C	G	T	G	C	A	T	C	G	
N13.31									G	C	A	C	C	C	A	G	A	C	C	C			
N13.32							G	T		C	C	C	C	C		G	C	C	T	A			
N13.34							G	T	G	C	A	C	C		A	T	G	T	T				
N13.36											A	C	C	C	A	T	T	C	A	A	C	C	C
N13.41			A	T	A	C	A	T	A	CG	A	C	C										
N13.45											A	C	C	C	A	C	G	C	A	C	G		
N13.47										C	A	C	C	C	A	G	G	C	C	C	C	A	
Consensus							A/G	T	A/G	C	A	C	C	C	A	N	N	C	A	C	C		

Each of the sequences that were shown in Figure 1 to bind RAP1, *in vitro*, were aligned, and adjusted by up to one gap or one bulged nucleotide. The consensus to which these aligned sequences conform is shown.

The set of sequences isolated contains very few contaminating sequences which are not RAP1 binding sites. When ten sequences were selected from the set at random, all ten were found to interact with RAP1. When the 47 sequences were compared with the established consensus RAP1 binding site (4) it was clear that most of them resembled this consensus (Table 1). Only two of the sequences had no apparent features in common with the consensus (N13.38 and N13.21). When tested in gel retardation assays neither of these two sequences was found to interact with RAP1 (Figure 1). We believe that only these two sequences out of the whole set are not genuine RAP1 binding sites.

In Table 1, the SAAB sequences are arranged in groups, in order of similarity to the established consensus RAP1 binding site. It is immediately clear that considerable variation from this consensus does not preclude RAP1 binding. We have isolated a sequence (N13.32) which varies from the consensus by as many as five mismatches, but can still interact with RAP1. However, the position and combination of mismatched bases are both important. In Table 2 we have listed the frequency of mismatched bases at each position within the SAAB sequences compared to the established consensus binding sequence.

Bases at positions which varied often are probably less critical for strong RAP1 binding than bases at positions which varied only infrequently. Variation at positions where the sequence is apparently well conserved, either alone or in combination with a variation elsewhere within the binding site, probably led to a dramatic reduction in the strength of that binding site. Consequently, that sequence was represented only poorly in the range of sequences isolated.

The results suggest that positions 2 to 7 form the core of the RAP1 binding site.

Within this region only two positions appear to be absolutely critical for RAP1 binding: positions 4 and 5. Excluding the two known non-binders, all of the SAAB sequences isolated, had a C residue at each of these positions.

In addition to the invariant positions 4 and 5, the core binding site contains the other positions which varied only infrequently from the consensus. Position 6 in the consensus RAP1 binding site is the position of the third C residue of the core C triplet. This position appears to be less important than positions 4 and 5 because a total of 15 sequences were isolated which either had a mismatched base or a gap at this position. Perhaps surprisingly, position 2 is also very important, with only four of the sequences isolated having a mismatch at this position. In the established consensus, position 2 is defined as A or C. Of the sequences isolated in the SAAB analysis, 40 had a C, and only one had an A in this position. At positions 3 and 7, some variation is tolerated, with seven of the SAAB sequences lacking an A at position 3 and nine lacking an A at position 7.

Sequences with mismatches at positions 10, 11, 12 and 13 occurred relatively frequently, suggesting that this region of the binding site may not be essential for strong RAP1 binding *in vitro*. Similarly, seventeen of the sequences did not contain the consensus purine residue (A or G) at position 1, suggesting that this position, too, is not critical for a strong RAP1 interaction.

Potentially the most important information regarding the nature of the RAP1 binding site is provided by an inspection of the combinations of mismatches which could occur without abolishing RAP1 binding. Most striking here is the fact that a mismatch in position 1 was relatively common in sequences with few mismatches at the 3′ end of the sequence (positions 10 to 13), but was relatively uncommon in sequences which also had two or more mismatches at the 3′ end. Position 1 is separated from the 3′ end of the consensus by approximately one helical turn of DNA. This observation suggests that RAP1 may bind to the core of the consensus and this interaction may then be stabilised

Table 4. Potential RAP1 binding sites in the yeast genome

Gene	SAAB site	Position	Sequence														
ADR6	47	-648	T	T	C	C	A	C	C	C	A	G	G	C	C	T	C
AFR1	30	-869	A	T	A	C	A	C	C		A	A	G	A	T	T	G
"	36	-199	T	C	T	C	A	C	C	C	T	T	T	C	A	A	C
"	48	-845	T	A	A	A	A	C	C	A	A	T	A	C	C	A	C
BLH1	39	-40	G	G	G	C	A	C	C	C	A	T	A	A	A	T	A
CARG B	17	-264	T	A	G	A	A	C	C	C	A	T	G	C	A	C	C
CDC43	05	-147	A	A	A	C	A	C	C	C	G	C	C	A	G	T	T
CDC63	34	-95	G	T	G	C	A	C	C		A	T	C	T	T	A	T
CHC1	42	-228	T	A	T	T	A	C	C	C		T	A	C	C	C	C
COX8	46	-46	T	A	G	A	G	C	C		A	C	A	G	A	G	C
CPH1	08	-298	G	G	G	C	G	C	C	C	A	G	C	C	C	G	C
CPS1	05	-292	A	A	A	C	A	C	C	C	A	C	T	T	C	T	A
CYS3	48	-299	C	T	G	G	A	C	C	C	C	A	T	A	C	C	A
DBF20	01	-161	A	T	T	A	A	C	C	C	A	T	A	C	A	C	C
EF1αB	08, 30, 43	-419	A	T	A	C	A	C	C	C	A	G	A	C	C	G	C
ENO2	34	-496	G	T	G	C	A	C	C		A	T	T	T	T	T	G
ERD2	05	-607	A	A	A	C	A	C	C	C	A	C	A	T	C	G	C
FRE1	30, 43	-224	A	T	A	C	A	C	C	C	A	A	T	T	T	C	T
FRS1	05	-945	A	A	A	C	A	C	C	C	C	A	A	A	G	G	A
G3PD	45	-261	A	T	T	G	A	C	C	C	A	C	G	C	A	T	G
GAL11	47	-285	T	T	C	C	A	C	C	C	C	G	G	C	C	C	C
GCD1	32	-144	G	G	T	C	A	C	C	C		G	C	C	T	A	T
GCD2	05	-302	A	A	A	C	A	C	C	C	A	T	A	G	G	A	A
GCD7	26	-39	A	C	A	C	G	C	C		A	A	C	T	T	T	C
GDH1	35	-395	A	C	C	G	G	C	C	C	A	T	G	C	A	T	T
GGS1	03	-744	G	T	A	T	A	C	C	C	A	C	C	T	A	T	A
GUT2	03	-44	G	T	A	C	A	C	C	C	C	C	C	C	C	C	C
"	05	-392	A	A	A	C	A	C	C	C	A	C	A	C	C	C	C
HSF1	34	-434	G	T	A	C	A	C	C		A	T	G	T	T	A	T
"	36	-200	A	G	A	A	A	C	C	C	A	T	T	C	A	C	C
HXK1	07, 30	-643	A	G	A	C	A	C	C	C	A	G	A	A	A	A	T
HXK2	07, 30	-643	A	G	A	C	A	C	C	C	A	G	A	A	A	A	T
HXT1	12	-176	T	A	A	C	A	C	C	C	G	A	A	G	T	T	G
ILV1	05	-561	A	A	A	C	A	C	C	C	G	C	T	T	G	T	C
ILV5	30	-126	T	T	A	C	A	C	C	C	A	G	T	A	T	T	T
INH1	20	-212	T	G	A	C	A	C	C	C	A	T		C	G	T	T
ITR1	11	-533	A	A	G	C	A	C	C	C	A	G	T	C	T	T	G
KIN3	05	-791	A	A	A	C	A	C	C	C	T	A	A	G	A	A	A
L12eIB	30	-352	A	T	A	C	G	C	C	C	A	G	A	C	A	T	A
LEU1	24	-568	A	T	G	C	G	C	C	C	A	T	A	C	A	G	A
LYP1	05	-333	A	A	A	C	A	C	C	C	A	C	G	A	G	T	G
"	36	-515	A	A	T	T	A	C	C	C	A	T	G	C	A	A	C
MET2	35	-634	A	T	G	T	G	C	C	C	A	T	A	T	A	T	T
MGT1	30	-224	A	T	A	A	A	C	C	C	A	G	T	T	C	A	G
MRS3	20	-543	T	G	A	C	A	C	C	C	C	C	C	C	A	C	A
MSD1	25	-486	A	C	T	C	A	C	C	C	C	T	G	C	A	T	C
NPR1	36	-38	T	G	A	T	A	C	C	C	A	T	T	C	A	A	A
NUP1	05	-970	A	A	A	C	A	C	C	C	A	A	T	C	A	T	A
OPI3	30	-314	A	T	A	A	A	C	C	C	A	G	G	T	G	G	G
PDR1	05	-889	A	A	A	C	A	C	C	C	A	A	A	A	T	G	T
PDR4	05	-697	A	A	A	C	A	C	C	C	G	A	A	C	T	A	T
PEP3	45	-376	T	G	T	G	A	C	C	C	A	C	G	T	A	C	G
PHO84	05	-497	A	A	A	C	A	C	C	C		G	T	T	C	C	T
PLC1	31	-269	G	T	G	G	A	C	C	C	A	G	A	C	C	C	A
POLA1	05	-643	A	A	A	C	A	C	C	C	A	T	A	C	A	C	C
"	43	-635	A	T	A	C	A	C	C	C		T	C	A	T	C	G
PPH3	05	-573	A	A	A	C	A	C	C	C		T	T	T	T	T	A
PUT3	43	-256	A	T	A	C	A	C	C	C	A	T	A	C	C	T	C
PYK1	30	-657	G	T	A	C	A	C	C	C	A	G	A	C	A	T	C
QRI8	34	-453	G	T	G	C	A	C	C		A	T	T	T	T	C	G
RAD2	48	-259	G	T	G	A	C	C	C	C	A	A	T	A	C	C	A
RAD16	01	-337	A	C	C	A	A	C	C	C	A	C	A	C	A	G	C
REV3	12	-166	T	A	A	C	A	C	C	C	C	T	T	T	A	T	T
RNR2	45	-447	C	C	A	C	A	C	C	C	A	C	G	C	G	C	G
ROX3	30, 43	-284	A	T	A	C	A	C	C	C	A	A	A	C	A	T	A
RP39A	35	-236	G	A	G	A	A	C	C	C	A	T	A	C	A	T	T
RP51A	35	-296	T	T	A	C	A	C	C	C	A	T	A	C	A	T	T
RPB4	36	-97	A	G	C	A	A	C	C	C	A	T	T	C	A	A	G
RPC53	30, 43	-316	A	T	A	C	A	C	C	C	A	C	T	C	T	A	A
RPL1	05	-276	A	A	A	C	A	C	C	C	A	T	A	A	A	G	C
RPL4	05	-242	A	A	A	C	A	C	C	C	A	A	A	C	A	T	A
RPL25	12	-551	T	A	A	C	A	C	C	C		G	G	A	G	G	A
RPL44	30	-351	A	T	A	C	G	C	C	C	A	G	A	C	A	T	A
RPS4	43	-384	C	T	A	G	A	C	C	C	A	T	A	C	A	C	C
"	43	-405	A	T	A	C	A	C	C	C	A	A	C	C	A	T	A
RPS7	25	-385	A	A	G	C	A	C	C	C	A	T	G	C	A	T	C
RPS10	35	-427	T	T	G	A	A	C	C	C	A	T	A	C	A	T	T
RPS24	35	-268	G	T	G	C	G	C	C	C	A	T	A	C	A	T	C
RPS28A	30, 43	-372	A	T	A	C	A	C	C	C	A	T	A	C	C	C	T
"	30	-226	A	G	A	C	A	C	C	C	A	G	C	G	G	A	G
RPS28B	05	-312	A	A	A	C	A	C	C	C	A	T	A	C	A	T	A
SGA5	12	-301	T	A	A	C	A	C	C	C	A	T	T	C	A	T	A
SGV1	30, 43	-180	A	T	A	C	A	C	C	C	A	T	A	C	A	C	C
SIN4	04	-318	C	T	A	C	A	C	C	C		T	C	T	T	T	G
SNF4	01	-236	A	A	G	A	A	C	C	C	A	C	A	C	A	C	C
STE18	48	-806	A	T	G	C	A	C	C	C	A	A	T	G	C	C	A
SUI1	05	-446	A	A	A	C	A	C	C	C		T	T	T	T	T	G
TEF1	05	-336	A	A	A	C	A	C	C	C	A	A	G	C	A	C	A
"	42	-443	T	C	A	C	A	C	C	C	A	A	T	C	C	C	C
TFC3	05	-711	A	A	A	C	A	C	C	C		G	A	G	G	C	A
"	32	-525	G	C	A	G	T	C	C		A	C	C	G	C	C	T
TOA1	24	-30	C	A	A	G	G	C	C		A	T	A	C	A	C	A
TRP5	12	-466	T	A	A	C	A	C	C	C		C	G	C	T	A	C
VAT C	30	-754	A	T	A	C	A	C	C		A	A	G	T	T	G	G
VMA6	35	-517	T	A	T	G	A	C	C	C	A	T	A	C	A	T	T
VPS15	05	-928	A	A	A	C	A	C	C	C	A	T	T	A	T	T	C
YAK1	30	-136	A	T	A	C	C	C	C	C	A	G	T	A	A	T	G
"	36	-485	A	G	C	A	A	C	C	C	A	T	T	C	A	A	G
YL3	05	-275	A	A	A	C	A	C	C	C	A	T	A	A	A	G	C
YL16	17	-453	T	A	T	C	A	C	C	C	A	T	G	C	A	C	C
YS11	25	-242	G	C	A	C	A	C	C	C	A	T	G	C	A	T	C
YTA3	05	-911	A	A	A	C	A	C	C	C		G	A	A	C	C	A

Each of the sequences generated by the SAAB analysis was used to screen the EMBL+GenBank database of cloned yeast DNA sequences (Release 78), allowing for a maximum of one mismatch between the sequences. The search identified a total of 160 sequences that were very similar to the SAAB sequences, in the promoters of 145 different yeast genes. Using the rules which we have developed (see Discussion), 58 of the sequences were found to contain combinations of mismatches which would be predicted to abolish RAP1 binding. The remaining 102 sequences are listed in Table 4. Bases which differ from our newly derived consensus are shaded.

by interactions with the bases at position 1 and the bases at positions 10 to 13. When recognition of position 1 is prevented by the presence of a mismatched base, interactions with positions 10 to 13 become more important for strong binding. Conversely, when positions 10 to 13 are mismatched, the interaction with position 1 becomes critical.

We have used the information from the sequences isolated to derive a new consensus binding site for RAP1 (Table 3). In order to prevent biasing our consensus due to the presence of common flanking sequences in all the oligonucleotides we compared only bases at each position contributed by the degenerate region. Initially, we compared all 47 of the SAAB sequences, which gave the consensus 5′ A/G T A/G C A C C C A N N C C/A C C 3′. However, as these sequences included two sequences which were subsequently shown not to interact with RAP1, and a number of sequences which were not proven to be binding sites, we also compiled a consensus using only the 14 sequences which have been shown to interact with RAP1 in gel retardation assays. This comparison gave the slightly modified consensus 5′ A/G T A/G C A C C C A N N C A C C 3′ (see Table 3). The 3′-most four positions of the latter consensus were derived from a comparison of just 11 clones. There is no evidence that the sequences which have not been tested for binding deviate from this region of the consensus to any greater extent than those which bind RAP1. Therefore, we feel it is appropriate to adopt the consensus which is derived from the larger population of sequences. This consensus is two base pairs longer than the established consensus which reflects the fact that the two bases immediately preceding position 1 of the established consensus showed significant conservation. Position −1 is a T in 18 out of 30 possible sequences and position −2 is an A or G in 25 out of 30 possible sequences. This suggests that the RAP1 binding site is more extensive than was previously thought. This correlates with methylation interference footprinting data using the RAP1 binding site from the *PGK* UAS, which suggested that RAP1 underwent a contact with the nucleotide at position −1 (6). When the original consensus (4) was used to analyse the yeast sequences within Release 78 of the EMBL+GenBank database, 26 exact matches were found. The 5′ ends of 20 out of the 26 sequences conform to the newly derived consensus at all but one position. This suggests to us that the extension of the original consensus is indeed valid. Our new consensus, 5′ A/G T A/G C A C C C A N N C C/A C C 3′, has a C replacing A/C at position 2 of the established consensus and A/C replacing the A at position 11. Both of these changes are well supported by the SAAB data. The new consensus is also slightly more constrained at the 3′ end. In the established consensus positions 12 and 13 are described as T or C. In our sequences, although these positions showed considerable variation, C was significantly more common than any other base and there is no evidence that T is preferred over A or G.

When the RAP1 binding site is considered as an extended 15 bp sequence instead of the existing 13 bp sequence it can be separated into three distinct regions: a core binding site, extending from positions 2 to 7; a 5′ flanking region, extending from positions −2 to 1; and a 3′ flanking region, extending from positions 10 to 13. The main role of the non-conserved bases at positions 8 and 9 may be to position the 3′ flanking sequence correctly, in order to help to stabilise the binding of RAP1 to the core sequence.

Six of the sequences isolated by the SAAB contain two copies of the CCC sequence found at the core of the RAP1 binding site. We have tested three of these sequences in order to determine whether they could be bound simultaneously by two molecules of RAP1. None of the sequences tested could interact with RAP1 in this way. In each case the second CCC sequence is either too close to the end of the oligonucleotide to be part of a complete RAP1 binding site or it overlaps with part of the better match to the RAP1 consensus binding site within the same sequence.

The results from this set of experiments allow us to predict which sequences within the yeast genome are likely to be good candidates for an interaction with RAP1. When a sequence is compared to our newly derived consensus, a 100% match is not essential. One mismatch within the core binding site, except for the invariant positions 4 and 5, is permissible. In addition, there can be up to four mismatches in either the 5′ or 3′ flanking sequences, provided the other flanking sequence in each case is not extensively mismatched (more than 2 mismatches).

Each of the SAAB-derived sequences (except for the 2 known non-binders) was used to look for matches in the promoters of cloned yeast genes. Searches were performed at a minimum percentage match of 90%. This ensured a maximum of one mismatch between any two sequences. The search identified a total of 160 sequences that were very similar to the SAAB sequences (that is, potential binding sites for RAP1) in the promoters of 145 different yeast genes.

When the above rules for predicting RAP1 binding sites were applied to this set of sequences, 58 of them were found to contain combinations of mismatches which would be predicted to abolish RAP1 binding. The remaining 102 sequences are listed in Table 4. Some genes which contain previously identified binding sites for RAP1 (*PGK*, for example (6)) are not represented in this table. These genes contain binding sites which match the consensus, but which are not sufficiently similar to any sequence within the sub-set of binding sites identified in this work. The sequences in Table 4 would all appear to be good candidates for RAP1 binding sites *in vitro*. Further experiments will be necessary to determine their *in vivo* significance. However, the results raise the possibility that RAP1 is involved in controlling expression of a very large number of yeast genes.

ACKNOWLEDGEMENTS

We thank Sue and Alan Kingsman for generously allowing the use of many materials generated in their laboratory, and Margaret Duffy, Louise Jenkins and Liz Packham for helpful discussions. This work was supported by the Agricultural and Food Research Council.

REFERENCES

1. Shore, D. and Nasmyth, K. (1987) *Cell*, **51**, 721–732.
2. Huet, J., Cottrelle, P., Cool, M., Vignais, M.-L., Thiele, D., Marck, C., Buhler, J.M., Sentenac, A. and Fromageot, P. (1985) *EMBO J.*, **4**, 3539–3547.
3. Huet, J. and Sentenac, A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3648–3652.
4. Buchman, A.R., Kimmerly, W.J., Rine, J. and Kornberg, R.D. (1988) *Mol. Cell. Biol.*, **8**, 210–225.
5. Vignais, M.L., Woudt, L.P., Wassenaar, G.M., Mager, W.H., Sentenac, A. and Planta, R. (1987) *EMBO J.* **6**, 1451–1457.
6. Chambers, A., Tsang, J.S.H., Stanway, C., Kingsman, A.J. and Kingsman, S.M. (1989) *Mol. Cell. Biol*, **9**, 5516–5524.
7. Capieaux, E., Vignais, M.-L., Sentenac, A. and Goffeau, A. (1989) *J. Biol. Chem.*, **264**, 7437–7446.
8. Kurtz, S. and Shore, D. (1991) *Genes Dev.* **5**, 616–628.
9. Sussel, L. and Shore, D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7749–7753.
10. Hardy, C.F.J., Sussel, L. and Shore, D. (1992) *Genes Dev.* **6**, 801–814.
11. Buchman, A.R., Lue, N.F. and Kornberg, R.D. (1988) *Mol. Cell. Biol.*, **8**, 5086–5099.
12. Longtine, M.S., Wilson, N.M., Petracek, M.E. and Berman, J. (1989) *Curr. Genet.*, **16**, 225–239.
13. Lustig, A.J., Kurtz, S. and Shore, D. (1990) *Science* **250**, 549–553.
14. Conrad, M.N., Wright, J.H., Wolf, A.J. and Zakian, V.A. (1990) *Cell*, **63**, 739–750.
15. Kyrion, G., Liu, K., Liu, C. and Lustig, A.J. (1993) *Genes Dev.* **7**, 1146–1159.
16. Henry, Y.A.L., Chambers, A., Tsang, J.S.H., Kingsman, A.J. and Kingsman, S.M. (1990) *Nucleic Acids Res.* **18**, 2617–2623.
17. Diffley, J.F.X. and Stillman, B. (1989) *Science* **246**, 1034–1038.
18. Chambers, A., Stanway, C.A., Kingsman, A.J. and Kingsman, S.M. (1988) *Nucleic Acids Res.* **16**, 8245–8260.
19. Devlin, C., Tice-Baldwin, K., Shore, D. and Arndt, K. (1991) *Mol. Cell. Biol.* **11**, 3642–3651.
20. Vignais, M.-L. and Sentenac, A. (1989) *J. Biol. Chem.* **264**, 8463–8466.
21. Fantino, E., Marguet, D. and Lauquin, G.J.-M. (1992) *Mol. Gen. Genet.* **236**, 65–75.
22. Pan, D. and Courey, A.J. (1992) *EMBO J.*, **11**, 1837–1842.
23. Wynne, J. and Treisman, R. (1992) *Nucleic Acids Res.*, **20**, 3297–3303.
24. Pollock, R. and Treisman, R. (1990) *Nucleic Acids Res.*, **18**, 6197–6204.
25. Sun, X.-H. and Baltimore, D. (1991) *Cell*, **64**, 459–470.
26. Blackwell, T.K. and Weintraub, H. (1990) *Science*, **250**, 1104–1110.
27. MacKay, V.L. (1983) *Meths. Enzymol.*, **101**, 325–343.
28. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual (2nd edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
29. Ogden, J.E., Stanway, C., Kim, S., Mellor, J., Kingsman, A.J. and Kingsman, S.M. (1986) *Mol. Cell. Biol.*, **6**, 4335–4343.
30. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA.*, **74**, 5463–5467.
31. Bradford, M.M. (1976) *Anal. Biochem.*, **72**, 248–254.