

---

# Characterisation of the DNA binding domain of the yeast RAP1 protein

---

Yves A.L.Henry<sup>1</sup>, Alistair Chambers<sup>1</sup>, Jimmy S.H.Tsang<sup>1</sup>, Alan J.Kingsman<sup>1,2</sup> and Susan M.Kingsman<sup>1\*</sup>

<sup>1</sup>Department of Biochemistry, Oxford University, South Parks Road, Oxford OX1 3QU and

<sup>2</sup>Department of Molecular Biology, British Biotechnology Ltd, Watlington Road, Cowley, Oxford OX4 5LY, UK

---

Received February 3, 1990; Accepted March 6, 1990

---

## ABSTRACT

**The 827 amino acid yeast RAP1 protein interacts with DNA to regulate gene expression at numerous unrelated loci in the yeast genome. By a combination of amino, carboxy and internal deletions, we have defined an internal 235 amino acid fragment of the yeast RAP1 protein that can bind efficiently to the RAP1 binding site of the *PGK* Upstream Activation Sequence (UAS). This domain spans residues 361 to 596 of the full length protein and lacks any homology to the DNA binding 'zinc finger' or 'helix-turn-helix' structural motifs. All the RAP1 binding sites we have tested bind domain 361–596, arguing that RAP1 binds all its chromosomal sites via this domain. The domain could not be further reduced in size suggesting that it represents the minimal functional DNA binding domain. The relevance of potential regions of secondary structure within the minimal binding domain is discussed.**

## INTRODUCTION

The RAP1 protein (Repressor Activator Protein 1), also called GRF1 (General Regulatory Factor 1), was initially identified as a factor that binds to the *HMRE* silencer element and to the *MATalpha* UAS (1,2,3). The RAP1 binding site is important for silencing transcription at *HMR* (4). This same RAP1 binding site from the *HMR* locus is however able to activate transcription from heterologous promoters (4). Purified RAP1/GRF1 was also shown to bind *in vitro* to the UASs of ribosomal protein genes (2), of elongation factor gene *TEF2* (2,3,5) and of the gene coding for the small subunit of ribonucleotide reductase, *RNR2* (6). It also binds to repeat sequences found at yeast telomeres (3,5,7). A high affinity RAP1/GRF1 consensus binding site of the form 5'(A/G)(A/C)ACCCANNCA(T/C)(T/C)3', where N is any nucleotide, was proposed (3,5). The fact that purified RAP1/GRF1 binds *in vitro* to ribosomal gene promoters prompted the suggestion that it is identical to a previously identified factor called TUF (8,9), in spite of the differences in

reported molecular weights (2,10). The glycolytic genes *TPI*, *ENO1*, *PYK1*, *PDC*, *ADH1* and *PGK* all contain RAP1 consensus binding sites in their promoters (11,12,13,14,15,16).

Whenever the position of the UAS is known, the consensus RAP1 binding site is located within it (14,16,17,18,19). Purified RAP1 binds *in vitro* to oligonucleotides corresponding to the consensus RAP1 binding sites of the *ENO1* and *PYK1* UASs (5). *In vitro* produced RAP1 binds to the *PGK* UAS subfragment called Z<sup>+</sup> extending from –473 to –409 in the *PGK* promoter, which contains a consensus RAP1 binding site (the AC sequence) (19). RAP1 is an essential component of the transcriptional activation process of the *PGK* gene, as a deletion which removed the AC sequence almost totally abolished transcription from the *PGK* promoter (20). The *PGK* gene is regulated transcriptionally by carbon source (21) and as RAP1 binding activity is not detected in nuclear extracts prepared from cells grown on pyruvate, RAP1 may play a central role in both transcriptional activation and carbon source regulation processes of glycolytic genes (19).

RAP1/GRF1/TUF is thus emerging as a key cellular factor which governs the expression of genes involved in such diverse processes as transcription, translation, nutrient transport, glycolysis, mating type regulation etc... (reviewed in 22). It plays a role in activating transcription at some sites, in repressing transcription at others while at telomeres it may only play a structural role. Furthermore, its activity with respect to certain genes is regulated by growth conditions. How can RAP1 be so versatile? It is conceivable, for instance, that RAP1 DNA binding activity is modulated by post-translational modification(s) affecting its DNA binding affinity for different sites to different extents. RAP1 is such a large protein that there may be different DNA binding domains directing binding to different RAP1 binding sites.

The RAP1 DNA binding domain has not to date been identified. Characterisation of this domain is central to an investigation of the basis of the versatility of RAP1 action. With a knowledge of the location of the binding domain within the primary sequence of the protein, one can assess the relevance for protein binding of elements of secondary structure and of amino acid residues that are potential targets for post-translational

---

\* To whom correspondence should be addressed

modifications. One can determine whether RAP1 binding sites found upstream of genes regulated differently are contacted by the same RAP1 domain and whether binding affinity is solely governed by the binding domain *per se*.

In this paper, we present the identification of the RAP1 DNA binding domain. This was achieved by producing RAP1 derivatives *in vitro* truncated either at their amino or carboxy termini, or deleted internally. These derivatives were then tested for their DNA binding ability by the band retardation method. This enabled us to localise a minimal RAP1 DNA binding domain to a centrally located 235 amino acid fragment.

## MATERIALS AND METHODS

### DNA manipulations, plasmid constructions and protein production

Restriction enzymes, T4 DNA ligase, T4 DNA polymerase and Klenow fragment were purchased from BRL, Inc. and New England Biolabs, Inc.. Radioactive isotopes [<sup>32</sup>P] dTTP (400Ci/mmol) and [<sup>35</sup>S] methionine (1,000Ci/mmol) were purchased from Amersham. DNA manipulations and plasmid constructions followed standard methods (23). All reagents used were analytical grade.

### *In vitro* protein production

All RAP1 mutant proteins were produced by *in vitro* transcription/translation using appropriate linearised templates (see below) driven from the SP6 promoter, using SP6 RNA polymerase (Promega Biotec). *In vitro* transcription was carried out as described in (19). The resultant RNAs were then translated in a cell free rabbit reticulocyte extract (Promega Biotec) in the presence of [<sup>35</sup>S] methionine (Amersham) as described in (19). SDS-PAGE was performed in order to check the amount and quality of the proteins produced. 10% SDS-polyacrylamide gels were run as described and referenced in (24).

### Production of RAP1 carboxy terminal deletions

The template for the production of carboxy terminal deletion mutants was pSP56RT7, described in (19). pSP56RT7 is a derivative of pSP56 containing the entire RAP1 sequence placed downstream and adjacent to the SP6 promoter. Proteins P1-824, P1-700, P1-596, P1-532 were produced by run-off transcription/translation of pSP56RT7 linearised at the *EcoO109I*, *BglII*, *BglI* and *SpHI* restriction sites, respectively (see Figure 1). The names of the proteins specify the RAP1 amino acids they contain with the starting amino terminal methionine as residue 1.

### Production of RAP1 amino terminal deletion mutants

Templates for the production of amino terminal deletion mutants were constructed by inserting restriction fragments from the RAP1 ORF into one of three initiation codon containing plasmids pSP66A1, pSP66A2 and pSP66A3. pSP66 is a derivative of pSP65 (25), where the *SpHI* site has been destroyed by recessing *SpHI* ends with T4 DNA polymerase, followed by blunt-end ligation. pSP66A1, pSP66A2 and pSP66A3 are derivatives of pSP66, containing the synthetic oligonucleotides A1, A2 and A3 respectively. Each oligonucleotide adaptor contains:

(i) an ATG codon placed in a correct nucleotide environment for efficient translation in a mammalian cell free translation extract.

(ii) a *BglII* expression site placed downstream from the ATG. This site is in a different reading frame with respect to the

initiating ATG in the different adaptors, thus allowing any RAP1 DNA fragment to be cloned in frame. Their respective sequences read:

A1 5'GCCATGGCAGATCTGCCATGGC3'  
A2 5'GCCATGGCCAGATCTGGCCATGGC3'  
A3 5'GCCATGGCCAGATCTGGGCCATGGC3'

Proteins were produced *in vitro* from the following templates: Proteins P422-596 and P422-691 were made from pPE420, linearised at *BglI* and *HindIII*, respectively. pPE420 contains a filled-in 837bp *HincII-BglII* fragment from the RAP1 ORF cloned into filled-in *BglII-HindIII* cut pSP66A3. P405-596 and P405-691 were made from pPE419 linearised at *BglI* and *HindIII*. pPE419 contains a filled-in 889bp *BanI-BglII* fragment from the RAP1 ORF cloned into filled-in *BglII-HindIII* cut pSP66A2. P378-596 and P378-691 were made from pPE417 linearised at *BglI* and *HindIII*. pPE417 contains a filled-in 971bp *FokI-BglII* fragment from the RAP1 ORF cloned into filled-in *BglII-HindIII* cut pSP66A1. P361-596 and P361-649 were made from pPE418 linearised at *BglI* and *DraI*. pPE418 contains a filled-in 994bp *HindIII-HindIII* fragment from the RAP1 ORF cloned into filled-in *BglII-HindIII* cut pSP66A2.

Proteins P302-701, P302-596, P302-583, P302-536, P302-532 and P302-497 were produced from plasmid pPE416, which contains a filled-in 1.2kbp *Tth3I-BglII* fragment from the RAP1 ORF cloned into the filled-in *BglII* site of pSP66A1. P302-701 was made by run-off transcription/translation of pPE416 cut downstream from the RAP1 *BglII* junction. P302-596, P302-583, P302-536, P302-532 and P302-497 were made from pPE416 linearised at the *BglI*, *HaeIII*, *DdeI*, *SpHI* and *BamHI* sites, respectively.

### Production of RAP1 internal deletion mutants

P(1-18)(498-827) protein was produced from pSP56RT7B linearised at the *XbaI* site. pSP56RT7B is derived from pSP56RT7 where the internal 1437bp *BamHI* fragment has been excised and the plasmid religated on itself. The reading frame is maintained. P(361-421)(498-649) protein was produced from pPE421, P(361-498)(535-649) from pPE422 and P(361-532)(584-649) from pPE423. These three plasmids were linearised at *DraI* prior to *in vitro* transcription/translation. pPE421 is *HincII*, *BamHI* cut filled-in pPE418 religated on itself. pPE422 is pPE418 cut with *BamHI* and filled-in, then cut with *SpHI*, recessed with T4 DNA polymerase and religated. pPE423 is pPE418 cut with *StuI* and *SpHI*, recessed with T4 DNA polymerase and religated.

### Band retardation assays

Band retardation assays were performed essentially as described previously (20,21). 400ng of calf thymus DNA was used in all binding reactions as non-specific competitor. No attempt was made to purify *in vitro* produced proteins from lysate proteins. 1μl of translated protein/lysate mix was added where indicated. The 90bp Z<sup>+</sup> fragment, which contains *PGK* promoter sequences from -473 to -409, was excised from pSP46-Z<sup>+</sup> (20) by a *Sall*, *BamHI* digest. The *HMRE*, *TEF2* and *RP51* binding sites were excised by *BamHI*, *HindIII* digests from pUC13 based plasmids D5, D6 and D7 (a generous gift from David Shore, Columbia University). The sequence of these binding sites is given in (2). The 87bp fragment from the *ENO1* UAS was excised by a *Sau3A*, *SmaI* digest from pHU38003 (see ref.12, a generous gift from H. Uemura, National Chemical

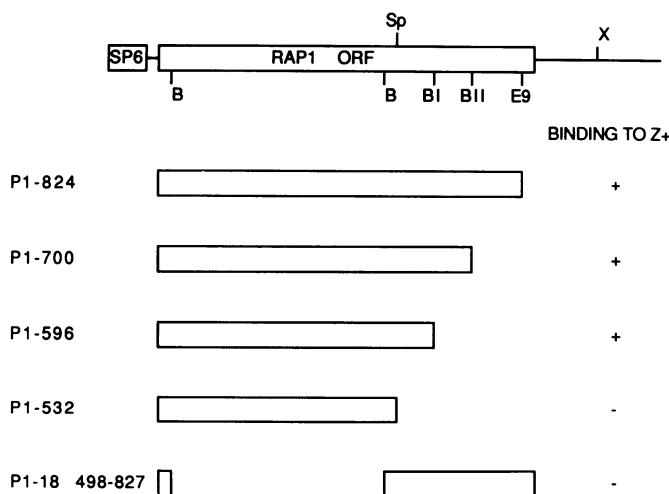
Laboratory for Industry, Yatabe, Japan). All fragments were end-labelled with [<sup>32</sup>P] dTTP and Klenow polymerase. Z<sup>+</sup> fragment corresponding to an activity of 50 counts per second (cps) was added in all binding reactions except for the band retardation assays of Figure 7 where all DNA fragments added correspond to an activity of 10 cps. Autoradiographs of band retardation assays presented were obtained using screens opaque to <sup>35</sup>S label thus protein/DNA complexes are detected only via the <sup>32</sup>P label in the DNA fragments.

**RESULTS**

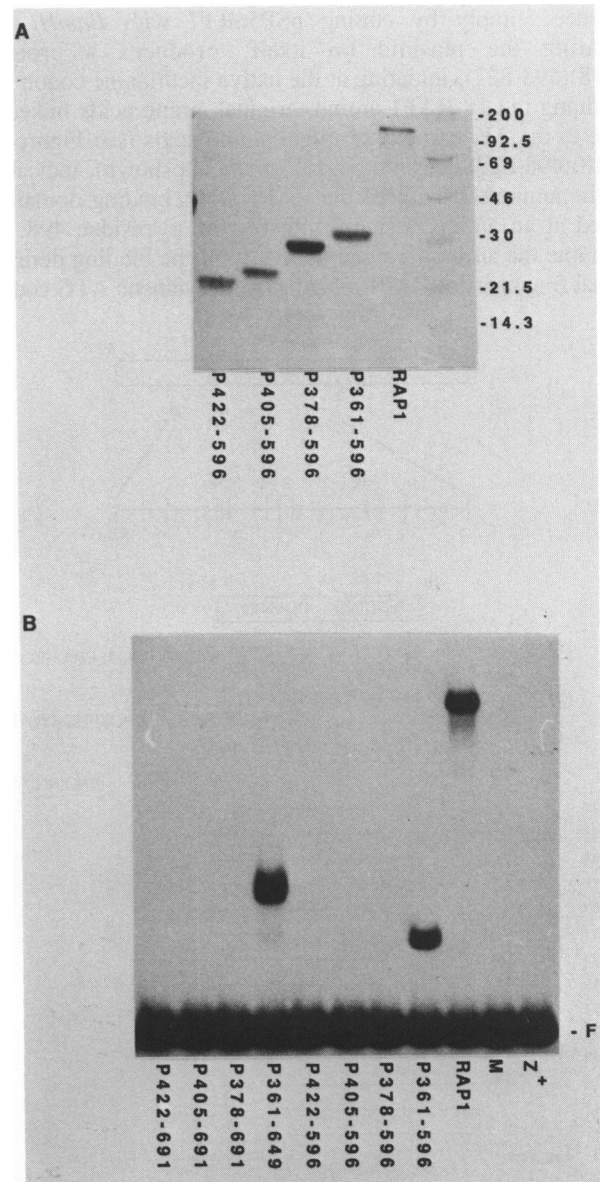
**Determination of the carboxy terminal endpoint of the RAP1 DNA binding domain**

We have previously shown that RAP1 produced *in vitro* binds to the PGK UAS efficiently (19). In order to define the carboxy terminal endpoint of the binding domain, RAP1 mutant proteins were produced *in vitro* with truncations of increasing length at their carboxy termini. Truncated RAP1 proteins were then tested for their ability to bind to the PGK UAS Z<sup>+</sup> subfragment by the band retardation method (26,27). Production of carboxy terminal deletion mutants of RAP1 was carried out by *in vitro* transcription/translation using a full length RAP1 clone driven from the SP6 promoter and linearised at appropriate restriction sites (Materials and Methods and Figure 1). *In vitro* translation was performed in a cell free rabbit reticulocyte extract in the presence of [<sup>35</sup>S] methionine so that the synthesised proteins could be visualised by autoradiography. *In vitro* produced proteins P1-824, P1-700, P1-596 and P1-532 are shown on Figure 2.A. In each case the numbering specifies the RAP1 amino acid residues present where residue 1 corresponds to the starting amino methionine. The molecular weights as determined from mobility on SDS-polyacrylamide gels are in all cases greater than the predicted ones, similar to what is observed for wild type RAP1. Shorter molecular weight products seen on Figure 2.A are probably due to premature termination of transcription and/or translation. Band retardation assays (26,27, Materials and Methods) using the Z<sup>+</sup> fragment as the DNA binding fragment

were performed with these truncated RAP1 derivatives (Figure 2.B). P1-824, P1-700 and P1-596 all gave rise to a retarded complex when incubated with the Z<sup>+</sup> fragment. The extent of the shift was related to the molecular weights of the binding proteins, as expected. Binding with lysate proteins alone (see Mock track, Figure 2.B) did not produce a shift, consistent with the retarded complexes being due to the RAP1 mutant proteins. P1-532, on the other hand, did not give rise to any retarded complex (see Figure 2.B, track P1-532). Judging from the intensity of the protein bands on the SDS-polyacrylamide gel (Figure 2.A), at least the same amount of P1-532 protein was



**Figure 1.** Restriction map of the RAP1 clone in pSP56RT7 and carboxy terminal and internal deletions derived from it. 'SP6' indicates the position of the SP6 promoter; '+' denotes proteins which bind to the PGK Z<sup>+</sup> UAS fragment and '-' denotes proteins which fail to bind to this fragment. Abbreviations for restriction sites are as follows: B=*Bam*HI, BI=*Bgl*II, BII=*Bgl*III, Sp=*Sp*HI, E9=*Eco*O109I, X=*Xba*I.



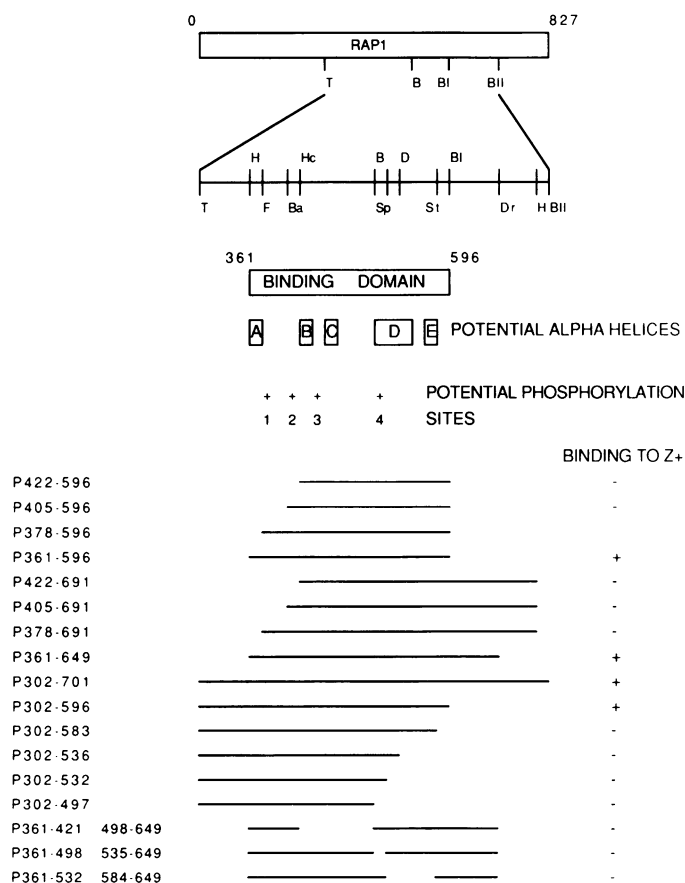
**Figure 2. A.** SDS-polyacrylamide gel of carboxy truncated versions of RAP1. Names of mutants specify the RAP1 amino acids that remain. 'M' refers to mock lysate. Markers are <sup>14</sup>C-labelled Rainbow markers (Amersham), whose molecular weights are given in kilodaltons. **B.** Band retardation assay with the Z<sup>+</sup> PGK UAS subfragment as DNA binding site. Proteins used in binding reactions are named as in Figure 2.A. and are indicated below each corresponding lane. Lane marked 'Z<sup>+</sup>' corresponds to a binding reaction with no protein added and lane marked 'M' to a binding reaction with mock lysate. 'F' indicates unbound fragment and 'C' RAP1 specific complexes.

added in the binding reaction compared to the other deletion derivatives. These results show that the carboxy terminal 230 amino acids of RAP1 can be deleted without affecting the ability of the protein to bind to DNA since P1-596 binds DNA efficiently. P1-532 does not however bind DNA either because the overall structure of the protein has been altered or because we have deleted an essential portion of the binding domain. These data argue that the carboxy terminal endpoint of the RAP1 DNA binding domain is located between residues 532 and 596.

**The DNA binding domain of RAP1 corresponds to an internal fragment spanning residues 361 to 596.**

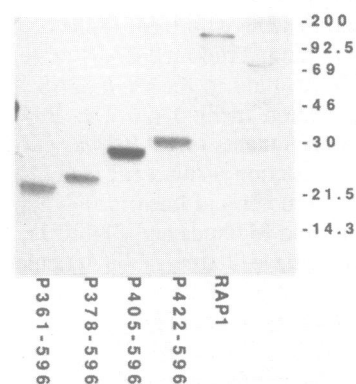
Removing an internal *Bam*HI fragment within the RAP1 sequence, simply by cutting pSP56RT7 with *Bam*HI and religating the plasmid on itself, produces a protein, P(1-18)(498-827), initiating at the native methionine codon and containing the 18 RAP1 amino terminal amino acids linked in frame to the 330 carboxy terminal amino acids (see Figure 1). P(1-18)(498-827) does not bind Z<sup>+</sup> (data not shown), indicating that the amino terminus of the RAP1 DNA binding domain is located at an amino position with respect to residue 498. To determine the amino terminal boundary of the binding domain, internal fragments of RAP1 initiating from synthetic ATG codons

and whose amino termini are sequentially located further towards the native initiation codon, starting from the *Hinc*II site next to residue 498, were tested for binding activity. Proteins with amino termini ranging from positions 422 to 361 and ending at carboxy position 596 were produced (see Figures 3 and 4.A). Binding to the Z<sup>+</sup> fragment was observed for protein P361-596 but not for protein P378-596, or for proteins with amino boundaries downstream from residue 378 (Figure 4.B). This argued that the amino boundary of the binding domain is located between residues 361 and 378 and that the internal RAP1 fragment extending from residues 361 to 596 functions as an independent unit able to bind DNA. We were concerned that proteins that seemed not to bind DNA might in fact be too small to produce

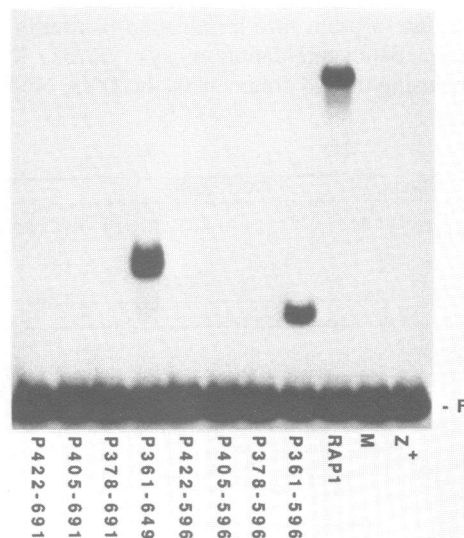


**Figure 3.** Detailed restriction map of the *Tih31-BglII* central fragment of RAP1 shown below the restriction map of the full gene. Abbreviations for restriction enzyme sites are as follows: B=*Bam*HI, Ba=*Ban*I, BI=*Bgl*I, BII=*Bgl*III, D=*Dde*I, Dr=*Dra*I, F=*Fok*I, H=*Hind*III, Hc=*Hinc*II, Sp=*Sp*HI, St=*Stu*I, T=*Tih*3I. The RAP1 internal domains expressed from synthetic ATG oligonucleotides are indicated. Also indicated are potential alpha-helices A,B,C,D and E and potential phosphorylation sites 1,2,3 and 4. '+' denotes proteins which bind to the Z<sup>+</sup> fragment and '-' denotes proteins which fail to bind to this fragment.

A



B

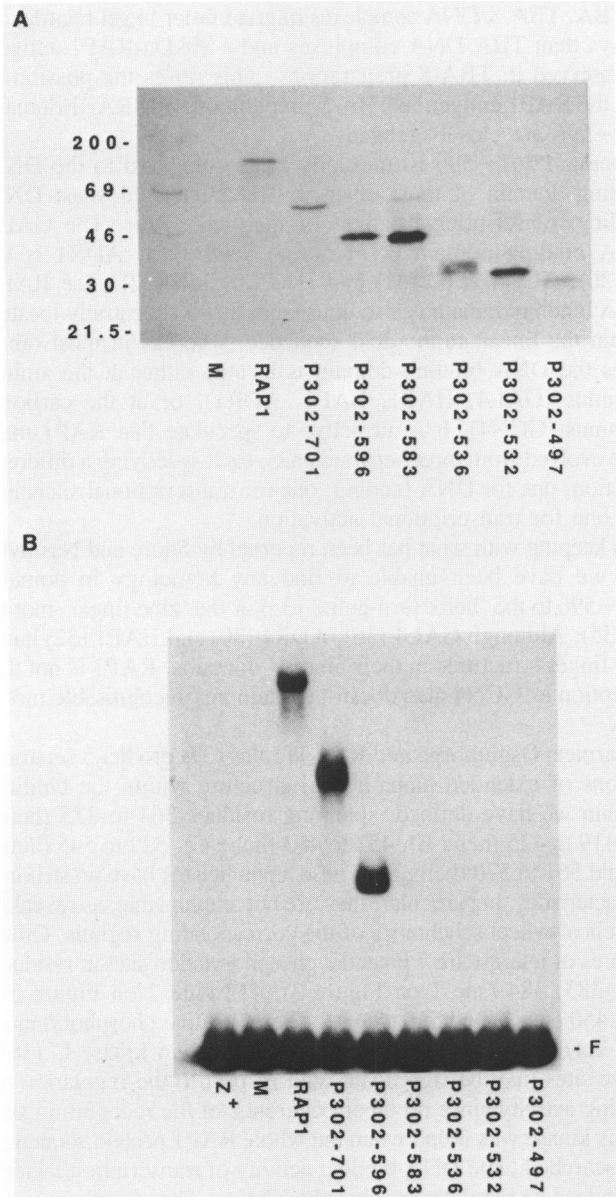


**Figure 4. A.** SDS-polyacrylamide gel of internal fragments of RAP1 produced to characterise the amino terminal boundary of the binding domain. Names of proteins specify the RAP1 amino acid residues they contain and are indicated below each corresponding lane. Markers are <sup>14</sup>C-labelled Rainbow markers (Amersham), whose molecular weights are given in kilodaltons. **B.** Corresponding band retardation assay with the Z<sup>+</sup> fragment and internal RAP1 derivatives. Proteins used in each binding reaction are named as in Figure 4.A and are indicated below each corresponding lane. Lane marked 'Z<sup>+</sup>' corresponds to a binding reaction with no protein added and lane marked 'M' to a binding reaction with mock lysate. 'F' indicates unbound fragment.

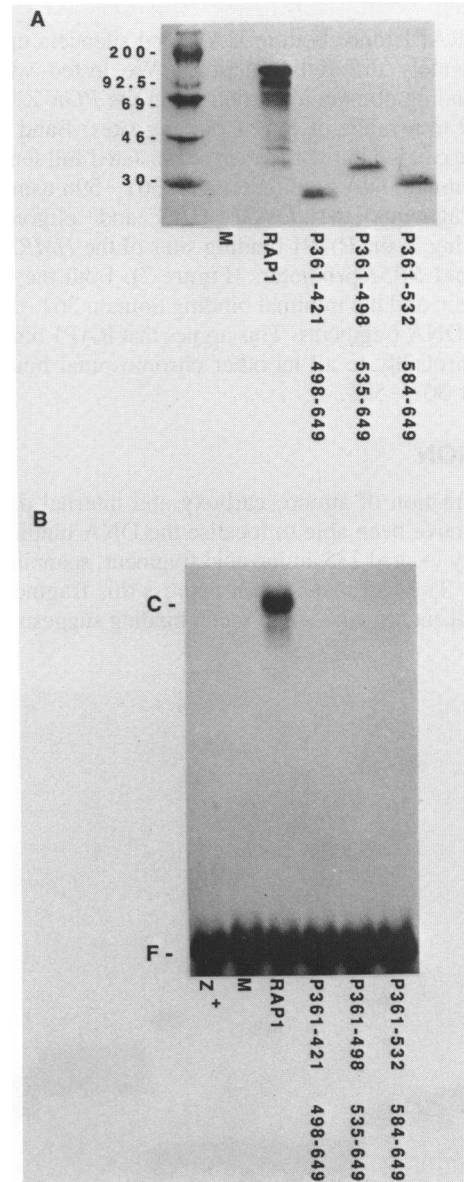
a detectable band shift. We therefore expressed a set of proteins with amino terminal boundaries ranging from positions 422 to 361 but whose molecular weights would be greater than protein P361-596 which produces a detectable band shift (see Figure 3, proteins P422-691 to P361-649). Protein P361-649 did bind  $Z^+$  as expected but P378-691 and proteins with amino termini at positions downstream from residue 378 did not bind (see Figure 4.B). This confirmed that the amino boundary of the binding domain is located between residues 361 and 378.

Construct pPE416 (see Materials and Methods) that contains a RAP1 sequence insert extending from residues 302 to 701 was

used to map the carboxy terminal boundary of the binding domain in greater detail. Proteins with amino termini at positions 302 and with carboxy boundaries ranging from residues 701 to 497 were produced (see Figures 3 and 5.A, proteins P302-701 to P302-497) and tested for their ability to bind DNA in the gel retardation assay. Proteins P302-701 and P302-596 were found to bind DNA whereas protein P302-583 and the other deletion derivatives with carboxy termini upstream of residue 583 did not bind DNA (see Figure 5.B). This indicated that the carboxy terminus of the binding domain is located between the residues 583 and 596.



**Figure 5. A.** SDS-polyacrylamide gel of internal fragments of RAP1 used to define the carboxy terminal boundary of the binding domain precisely. Names of proteins are indicated below each corresponding lane and specify the RAP1 residues they contain. 'M' is mock lysate. Markers are  $^{14}C$ -labelled Rainbow markers (Amersham), whose molecular weights are given in kilodaltons. **B.** Corresponding band retardation assay with the  $Z^+$  fragment and the internal RAP1 derivatives. Proteins used in each binding reaction are named as in Figure 5.A and are indicated below each corresponding lane. Lane marked ' $Z^+$ ' corresponds to a binding reaction with no protein added and lane marked 'M' corresponds to a binding reaction with mock lysate. 'F' indicates unbound fragment.



**Figure 6. A.** SDS-polyacrylamide gel of derivatives of the binding domain P361-596 containing internal deletions. The names of the proteins are indicated below each lane and specify the RAP1 residues they contain. 'M' is mock lysate. Markers are  $^{14}C$ -labelled Rainbow markers (Amersham), whose molecular weights are given in kilodaltons. **B.** Corresponding band retardation assay with the  $Z^+$  fragment. Proteins used in each binding reaction are named as in Figure 6.A and are indicated below each corresponding lane. Lane marked ' $Z^+$ ' corresponds to a binding reaction with no protein added and lane marked 'M' corresponds to a binding reaction with mock lysate. 'F' indicates unbound fragment and 'C' the RAP1 specific complex.

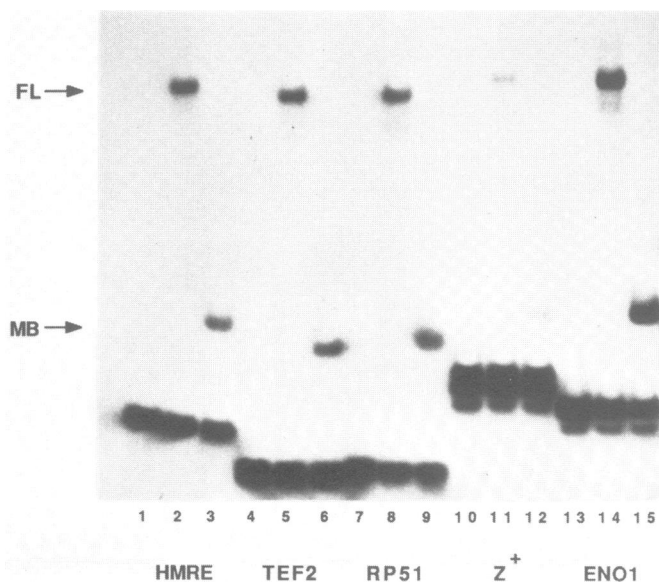
Compared to other yeast transcriptional activators characterised so far (28 to 32), the binding domain we have defined as 'minimal' remains very large. Possibly regions of the 361-596 fragment are not necessary to form a functional DNA binding domain. To test this hypothesis, we produced internal deletions within this fragment, removing residues 422 to 497, 499 to 534 and 533 to 583 (P(361-421)(498-649), P(361-498)(535-649) and P(361-532)(584-649) respectively, see Figures 3 and 6.A). None of these proteins bound the Z<sup>+</sup> fragment (see Figure 6.B). This may mean that the fragments deleted are part of the binding domain *per se*, or that their removal prevents its proper folding.

#### Domain 361-596 binds to both activator and silencer elements

Wild type RAP1 binds both to UASs and silencers upstream of genes of widely different functions. We tested whether the minimal binding domain identified using the *PGK* Z<sup>+</sup> fragment would bind to a range of RAP1 binding sites. Band retardation assays were carried out with *in vitro* translated full length RAP1 and the minimal DNA binding domain 361-596 using as DNA binding fragments the *ENO1* UAS and oligonucleotides corresponding to the RAP1 binding sites of the *HMRE* silencer, the *TEF2* and *RP51* promoters (Figure 7). Both the full length RAP1 protein and the minimal binding domain 361-596 bound to all these DNA fragments. This argues that RAP1 binds to these sites, and probably to all its other chromosomal binding sites, via domain 361-596.

#### DISCUSSION

By a combination of amino, carboxy and internal deletions of RAP1, we have been able to localise the DNA binding domain to a centrally located 235 amino acid fragment, spanning residues 361 to 596. The fact that we can express this fragment *in vitro* without adjacent sequences and retain binding suggests that it can



**Figure 7.** Binding of P361-596 to various RAP1 DNA binding sites. Lanes 1,2,3 correspond to binding to the *HMRE* RAP1 binding site, lanes 4,5,6 to the *TEF2* site, lanes 7,8,9 to the *RP51* site, lanes 10,11,12 to the *PGK* UAS Z<sup>+</sup> subfragment, lanes 13,14,15 to the *ENO1* UAS. 1  $\mu$ l of mock lysate was added in reactions corresponding to lanes 1,4,7,10,13. 1  $\mu$ l of *in vitro* translated RAP1 was added in reactions corresponding to lanes 2,5,8,11,14. 1  $\mu$ l of protein P361-596 was added in reactions corresponding to lanes 3,6,9,12,15.

fold independently into a functional unit. There is speculation that RAP1/GRF1 and TUF are the same factors since their consensus binding sites are identical (5,9). Interestingly, Huet and Sentenac have reported that limited proteolytic digest of TUF releases a domain of 50kDa (as judged from its mobility on an SDS-polyacrylamide gel) that binds DNA with the same affinity as undigested TUF. This is substantially larger than the minimal domain defined here which runs as a protein of molecular weight 33kDa. Recently, a yeast telomere binding activity (TBA) has been shown to be indistinguishable from RAP1 (7). A proteolysed form of TBA (TBA-S) binds to telomere repeats and to various DNA RAP1 binding sites with the same relative binding affinities as TBA. TBA-S/DNA complexes migrate faster in gel retardation assays than TBA/DNA complexes and a 30kDa RAP1 antigen is observed in TBA-S preparations. This raises the possibility that the RAP1 antigen in TBA-S preparations and RAP1 domain 361-596 are closely related.

Domain 361-596 is unusually large compared to the DNA binding domain of most other well characterised yeast DNA binding transcriptional activators analysed so far. The *GAL4* DNA binding domain is 74 amino acids (28), *ADR1* is 96 (29,30), *GCN4* is 60 (31) and *HAP1* is 148 (32). The RAP1 DNA binding domain is also unusual in that it is centrally located within the linear amino acid sequence, whereas in most other cases the DNA binding domain is located either at the amino terminus (*GAL4*, *HAP1*, *BAF1*, *ADR1*), or at the carboxy terminus (*GCN4*). It is attractive to speculate that RAP1 may have evolved from three separate units, each specifying a different function, one for DNA binding, one for transcriptional silencing and one for transcriptional activation.

In keeping with what has been reported by Shore and Nasmyth (2), we have been unable to find any homology in domain 361-596 to the 'helix-turn-helix' (33) or the 'zinc finger' motifs (34,35). Although *GAL4* (36), *ADR1* (30) and *HAP1* (32) have zinc finger structures in their binding domains, RAP1 is not the exception as *GCN4* also doesn't contain any recognisable motif (31).

Garnier, Osguthorpe and Robson rules (37) predict 5 separate regions of extended alpha-helical structure within the binding domain we have defined, spanning residues 364 to 375 (helix A), 419 to 425 (helix B), 452 to 463 (helix C), 522 to 545 (helix D) and 561 to 570 (helix E). These alpha helices have no striking characteristics; in particular, they are not amphipathic, as revealed by helical-wheel schematics of the corresponding regions. Other features of interest are 7 potential phosphorylation sites at residues 380, 383, 384 (site 1 on Figure 3), 412 (site 2 on Figure 3), 444, 450 (site 3) and 516 (site 4). These putative phosphorylation sites correspond to recognition sites of protein kinase C (38). These sites may be significant with respect to the regulation of binding activity since no other occurrence of the recognition site of this kinase was found when the whole RAP1 protein sequence was searched. The DNA binding activity of transcription factors can indeed be regulated by phosphorylation: The phosphoprotein serum response factor (SRF) doesn't bind to its cognate DNA binding site, the serum response element (SRE), after treatment with potato acid phosphatase (39). Likewise, the DNA binding activity of the adenovirus induced E4F transcription factor for its cognate DNA binding site in the adenovirus E4 promoter is abolished *in vitro* after treatment with alkaline phosphatase (40). E4F has not been shown to be a phosphoprotein, however. The transcriptional activation factor CREB, which has been shown to be a phosphoprotein, binds to the cyclic-AMP response element (CRE) as a monomer or as a dimer. Dimer formation on CRE

is reduced after treatment of affinity purified CREB with alkaline phosphatase and augmented following treatment of CREB with protein kinase C (41).

Alpha helices A,B,C,D and E and phosphorylation sites 1,2,3 and 4 are indicated on figure 3. Interestingly, P378-596 which lacks the whole of alpha-helix A, is unable to bind DNA, arguing that this helix might be an essential component of the binding domain, either contacting DNA, or necessary for proper folding, or both. Comparison between P302-596, which binds, and P302-583, which doesn't, clearly demonstrates the importance of residues between amino acids 583 and 596.

The unusual size of the binding domain prompted us to investigate whether some internal fragments could be deleted without affecting DNA binding. We deleted residues 422 to 497, 499 to 534 and 533 to 583. These three deletions abolished binding. It has to be noted that deletion 422 to 497 removes potential alpha-helices B and C and phosphorylation site 3 while deletion 499 to 534 removes phosphorylation site 4 and one half of helix D and deletion 533 to 583 the other half as well as helix E. These data may point to their importance as structural elements of the binding domain. However, definite answers await crystallographic data.

RAP1 binds to a range of different chromosomal locations. Most of these sites show strong sequence homology, arguing that they are contacted by the same domain of the protein. Our data support this notion: We have shown here that domain 361–596 binds to the RAP1 binding sites of the *HMRE* silencer, the *RP51* and *TEF2* gene promoters and to the UASs of the *PGK* and *ENO1* genes. We infer from this that RAP1 binds all its chromosomal sites via domain 361–596, although we cannot rule out the possibility that binding to different RAP1 binding sites is mediated by different parts of this domain.

RAP1 appears to be significantly different from other protein factors that bind to a restricted class of promoters. In particular, RAP1 does not seem to act as a 'conventional' activator protein on its own (42) and its minimal DNA binding domain is large and seem not to contain any of the known DNA binding structures. Interestingly, the protein BAF1/ABF1 (also called SBF-B) also binds to an array of different promoters and is implicated in a range of functions, including replication (1,3) and transcription activation and silencing (4). Like the RAP1 DNA binding domain, the DNA binding domain of BAF1/ABF1 is large although in this case a putative zinc finger appears to be implicated (43,44). It has been reported that RAP1 and ABF1/BAF1 share regions of homology (43), although others have claimed not to find any overall homology between ABF1 and other known proteins (45). The significance of functional and, if any, structural similarities between RAP1 and ABF1/BAF1 is not yet clear but they may represent a new class of control proteins that have a global role in the management of DNA for a variety of functions.

## ACKNOWLEDGMENTS

We would like to thank David Shore for the gift of *HMRE*, *TEF2* and *RP51* oligonucleotides and Hiroshi Uemura for the gift of plasmid pHU38003. We would also like to thank Ken Johnson for photography. Y.H. is a graduate student supported by the United Kingdom M.R.C., A.C. and J.T. are postdoctoral research fellows supported by the United Kingdom A.F.R.C.

## REFERENCES

1. Shore, D., Stillman, D.J., Brand, A.H. and Nasmyth, K.A. (1987) *EMBO J.*, **6**, 461–467.
2. Shore, D. and Nasmyth, K. (1987) *Cell*, **51**, 721–732.
3. Buchman, A.R., Kimmerly, W.J., Rine, J. and Kornberg, R.D. (1988) *Mol. Cell. Biol.*, **8**, 210–225.
4. Brand, A.H., Micklem, G. and Nasmyth, K. (1987) *Cell*, **51**, 709–719.
5. Buchman, A.R., Lue, N.F. and Kornberg, R.D. (1988) *Mol. Cell. Biol.*, **8**, 5086–5099.
6. Hurd, H.K. and Roberts, J.W. (1989) *Mol. Cell. Biol.*, **9**, 5359–5372.
7. Longtine, M.S., Wilson, N.M., Petracek, M.E. and Berman, J. (1989) *Curr. Genet.*, **16**, 225–239.
8. 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.
9. Vignais, M.-L., Woudt, L.P., Wassenaar, G.M., Mager, W.H., Sentenac, A. and Planta, R.J. (1987) *EMBO J.*, **6**, 1451–1457.
10. Huet, J. and Sentenac, A. (1987) *Proc. Natl. Acad. Sci. UAS*, **84**, 3648–3652.
11. Alber, T. and Kawasaki, G. (1982) *J. Mol. Appl. Genetics*, **1**, 419–434.
12. Uemura, H., Shiba, T., Patterson, M., Jigami, Y. and Tanaka, H. (1986) *Gene*, **45**, 67–75.
13. Burke, R.L., Tekamp-Olson, P. and Najarian, R. (1983) *J. Biol. Chem.*, **258**, 2193–2201.
14. Butler, G. and McConnell, D.J. (1988) *Curr. Genet.*, **14**, 405–412.
15. Bennetzen, J.L. and Hall, B.D. (1982) *J. Biol. Chem.*, **257**, 3018–3025.
16. Ogden, J.E., Stanway, C., Kim, S., Mellor, J., Kingsman, A.J. and Kingsman, S.M. (1986) *Mol. Cell. Biol.*, **6**, 4335–4343.
17. Machida, M., Uemura, H., Jigami, Y. and Tanaka, H. (1988) *Nucleic Acids Res.*, **16**, 1407–1422.
18. Nishizawa, M., Araki, R. and Teranishi, Y. (1989) *Mol. Cell. Biol.*, **9**, 442–451.
19. Chambers, A., Tsang, J.S.H., Stanway, C., Kingsman, A.J. and Kingsman, S.M. (1989) *Mol. Cell. Biol.*, **9**, 5516–5524.
20. Chambers, A., Stanway, C., Kingsman, A.J. and Kingsman, S.M. (1988) *Nucleic Acids Res.*, **16**, 8245–8260.
21. Stanway, C., Mellor, J., Ogden, J.E., Kingsman, A.J. and Kingsman, S.M. (1987) *Nucleic Acids Res.*, **15**, 6855–6873.
22. Capiex, E., Vignais, M.-L., Sentenac, A. and Goffeau, A. (1989) *J. Biol. Chem.*, **264**, 7437–7446.
23. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
24. Wilson, W., Braddock, M., Adams, S.E., Rathjen, P.D., Kingsman, S.M. and Kingsman, A.J. (1988) *Cell*, **55**, 1159–1169.
25. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
26. Fried, M. and Crothers, D.M. (1981) *Nucleic Acids Res.*, **9**, 6505–6525.
27. Garner, M.M. and Revzin, A. (1981) *Nucleic Acids Res.*, **9**, 3047–3060.
28. Keegan, L., Gill, G. and Ptashne, M. (1986) *Science*, **231**, 699–704.
29. Bemis, L.T. and Denis, C.L. (1988) *Mol. Cell. Biol.*, **8**, 2125–2131.
30. Thukral, S.K., Taviani, M.A., Blumberg, H. and Young, E.T. (1989) *Mol. Cell. Biol.*, **9**, 2360–2369.
31. Hope, I.A. and Struhl, K. (1986) *Cell*, **46**, 885–894.
32. Pfeifer, K., Kim, K.-S., Kogan, S. and Guarente, L. (1989) *Cell*, **56**, 291–301.
33. Pabo, C.O. and Sauer, R.T. (1984) *Ann. Rev. Biochem.*, **53**, 293–321.
34. Miller, J., McLachlan, A.D. and Klug, A. (1985) *EMBO J.*, **4**, 1609–1614.
35. Berg, J.M. (1986) *Science*, **232**, 485–487.
36. Johnston, M. and Dover, J. (1987) *Proc. Natl. Acad. Sci. UAS*, **84**, 2401–2405.
37. Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.*, **120**, 97–120.
38. Leader, D.P. and Katan, M. (1988) *J. Gen. Virol.*, **69**, 1441–1464.
39. Prywes, R., Dutta, A., Cromlish, J.A. and Roeder, R.G. (1988) *Proc. Natl. Acad. Sci. UAS*, **85**, 7206–7210.
40. Raychaudhuri, P., Bagchi, S. and Nevins, J.R. (1989) *Genes Dev.*, **3**, 620–627.
41. Yamamoto, K.K., Gonzales, G.A., Biggs III, W.H. and Montminy, M.R. (1988) *Nature*, **334**, 494–498.
42. Stanway, C.A., Chambers, A., Kingsman, A.J. and Kingsman, S.M. (1989) *Nucleic Acids Res.*, **17**, 9205–9218.
43. Diffley, J.F.X. and Stillman, B. (1989) *Science*, **246**, 1034–1038.
44. Halfter, H., Kavety, B., Vandekerckhove, J., Kiefer, F. and Gallwitz, D. (1989) *EMBO J.*, **8**, 4265–4272.
45. Rhode, P.R., Sweder, K.S., Oegema, K.F. and Campbell, J.L. (1989) *Genes Dev.*, **3**, 1926–1939.