

DNA-binding requirements of the yeast protein Rap1p as selected in silico from ribosomal protein gene promoter sequences

Romeo F. Lascaris, Willem H. Mager and Rudi J. Planta

Department of Biochemistry and Molecular Biology, IMBW, Biocentrum Amsterdam, Vrije Universiteit, de Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

Received on October 16, 1998; revised on December 15, 1998; accepted on December 16, 1998

Abstract

Motivation: High-level transcriptional activation of most ribosomal protein (*rp*) genes in *Saccharomyces cerevisiae* is promoted by the global DNA-binding factor Rap1p. The creation of the complete database of yeast *rp* gene promoter sequences enabled us to develop a computational selection strategy aimed at acquiring detailed information concerning the DNA-binding specificity of Rap1p.

Results: Rap1p sites in *rp* gene promoters are often found in duplicate, exhibiting strong preferences in both spacing and orientation. Using these preferences, a weight matrix was selected that represents the *in vivo* binding requirements of Rap1p. The resulting matrix renders the identification of functional Rap1p binding sites more accurate and allowed us to re-evaluate previous *in vitro* data. Tandemly arranged Rap1p binding sites appear to be typical for *rp* gene promoters and differ in preferred spacing from sites occurring in (sub)telomeric repeats. The preferred spacing that is found in duplicate Rap1p binding sites of *rp* gene promoters is restricted to a small window between 15 and 26 bp. This is proposed to reflect the borders within which binding co-operativity operates. The data presented clearly illustrate that computational selection strategies provide information that reaches beyond experimental data.

Availability: The *rp* database is available at the url: <http://www.chem.vu.nl/BMB/Database.html>.

Contact: mager@chem.vu.nl

Introduction

The genome of *Saccharomyces cerevisiae* contains 137 non-clustered genes coding for ribosomal proteins (*rp* genes) (Mager *et al.*, 1997). Since 59 genes are duplicated, the *rp* gene family encodes 78 different ribosomal proteins, 32 of the small and 46 of the large ribosomal subunit. The promoters that regulate these genes are highly efficient: recent data from SAGE analysis (Velculescu *et al.*, 1997) indicate that *rp* gene transcripts account for ~20% of the total mRNA content of a rapidly growing yeast cell. Previous promoter analyses indicated that transcription activation of the major-

ity of the yeast *rp* genes is mediated by the DNA-binding protein Rap1p, a so-called global regulator involved in multiple cellular functions. In addition, transcription of a subset of *rp* genes is promoted by the global regulator Abf1p (Mager and Planta, 1990). T-rich elements have been demonstrated to serve as auxiliary elements in reaching the high transcriptional yields of *rp* gene promoters (Buchman and Kornberg, 1990; Gonçalves *et al.*, 1995). The actual mechanism of transcription activation, however, and the way in which this process is regulated in response to varying growth conditions, remain to be elucidated.

Rap1p, the major factor binding to *rp* gene promoters, is an abundant multifunctional protein that is essential for cellular viability. It has strong DNA-bending properties (Vignais and Sentenac, 1989) and it creates a nucleosome-free region by binding to its cognate sites (Devlin *et al.*, 1991). These data suggest that Rap1p plays a role in organizing the chromatin structure in order to allow gene-specific regulatory proteins access to their binding sites (Planta *et al.*, 1995). In addition, Rap1p can recruit other factors, such as the Sir and Rif proteins, by protein–protein interactions, resulting in silencing at the silent mating type loci and telomeres and in telomere length regulation (Conrad *et al.*, 1990; Lustig *et al.*, 1990; Aparicio *et al.*, 1991; Hardy *et al.*, 1992; Kyrion *et al.*, 1992; Moretti *et al.*, 1994; Shore, 1997; Wotton and Shore, 1997).

Rap1p has been shown to bind to sequences that vary considerably. Several consensus sequences have been published: AACATCYRTRCA (Teem *et al.*, 1984), RMAVC-CRTRCMYY (Mager, 1988; Planta and Raué, 1988), RMAC-CCANNCAYY (Buchman *et al.*, 1988) and ACACCCATACATTT (Nieuwint *et al.*, 1989). Mutational analysis of the high-affinity binding site ACACCCATACATTT revealed that each nucleotide in this sequence, except for the underlined position, could be replaced by at least one other nucleotide without significant loss in transcription of a reporter gene (Nieuwint *et al.*, 1989). Similar results have been obtained from *in vitro* binding studies using synthetic oligonucleotides (Vignais *et al.*, 1990; Graham and Chambers, 1994). This makes the identification of a functional Rap1p binding site troublesome, which is further aggravated by the fact that the binding affinity of Rap1p for a

particular sequence *in vitro* does not necessarily reflect its functional efficiency *in vivo* (Nieuwint *et al.*, 1989). The latter might be due to post-translational modifications (Tsang *et al.*, 1990), the presence of neighbouring binding sites or local chromatin structure. In order to define the sequence requirements for Rap1p binding more precisely and, thus, be able to identify functional Rap1p binding sites in *rp* gene promoters, we developed a computational selection strategy in which non-sequence-related criteria are incorporated (see also Bucher, 1990). We started by scanning a database of all *rp* gene promoters with the low-stringency string AYCCRNNCM and subsequently generated matrices by incorporating additional parameters evolved from this analysis. The matrix we finally obtained contains more information and exhibits more detail than can be deduced from experimental data. This approach sheds new light on Rap1p binding sequences selected from randomized oligonucleotides (Graham and Chambers, 1994) and makes the identification of functional Rap1p binding site in the yeast genome more accurate. The likely upstream activation sequences in *rp* gene promoters are presented and it is concluded that duplicate Rap1p sites are typical for these promoters. The spacing between tandemly arranged Rap1p sites in *rp* gene promoters is shown to be confined to a small window of ~12 bp, which differs strongly from the spacing in multiple Rap1p sites found in (sub)telomeric repeat sequences.

Materials and methods

A total of 137 *rp* gene sequences, spanning from 600 bp upstream to 100 bp downstream from the ATG, were obtained from the *Saccharomyces* Genome Database (SGD) to generate an *rp* gene promoter database. For a few *rp* genes which have been shown to contain an intron in the leader sequence (Planta and Mager, 1998) or for which intron-specific sequences were detected in the leader, the promoter was represented by a sequence from 600 bp upstream to 100 bp downstream of the 5'-splice site. The sequence of the yeast genome was obtained from SGD by anonymous FTP.

A computational search strategy was designed, based on the idea that sequence information about the sequence requirements of a particular DNA-binding protein can be obtained by selection using sequence-unrelated parameters. From a group of binding sites that show similarity to the consensus binding site, certain binding sites can be selected that can be expected to be biologically functional according to the position with respect to other landmarks in the sequence.

For matrix analyses, the freeware programs MatInd and MatInspector (Quandt *et al.*, 1995) were used. MatInd generates, from DNA sequences or matrices, specific nucleotide distribution matrices for subsequent use as input file for the program MatInspector. The latter program performs the actual matrix search. In the matrix library of MatInspector, a Rap1p matrix (RAP_C) is included. However, this matrix

exhibits five nucleotide positions, corresponding to 4A, 5C, 6C, 7C and 8A in Figure 1, which have a maximum score of 100%. Since these high scores are not all expected from experimental data and not consistent with all consensus sequences (see Introduction), the RAP_C matrix was only used for comparison (see Results and discussion).

Defining the threshold at which the *rp* database should be searched with the MR2 matrix, use was made of a random sequence file with the same nucleotide composition as the *rp* database, A and T = 31%, G and C = 19%, which was obtained using the Random Sequence Generator (<http://copan.cifn.unam.mx/~yeast>).

As shown in Table 1, we also tried to define the putative binding sites for Abf1p and Reb1p, using the consensus sequences RTMRYB(N₄)ACG for Abf1p and CGGGTARNNR for Reb1p (Remacle and Holmberg, 1992). For Abf1p, we chose not to use previously published matrices (Quandt *et al.*, 1995), because the 5' part of the consensus sequence RTCRYB(N₄)ACG has been shown to be inconsistent with the binding sites that have been shown to bind Abf1p *in vivo*, viz. sequences containing either RTAR (Della Seta *et al.*, 1990; Kraakman *et al.*, 1991) or even RGCR (Biswas and Biswas, 1990; Einerhand *et al.*, 1995) (M.de Boer, unpublished data from this laboratory). Consequently, it is expected that more Abf1p binding sites may be found in *rp* gene promoters by computational analysis when the binding characteristics of Abf1p, and also of Reb1p, are investigated more thoroughly.

Table 1. Computational prediction of the UASs of *rp* genes. Distances relative from the ATG. An asterisk indicates *rp* genes containing an intron in their leader sequence (Mager *et al.*, 1997)

RPL01A	
RPL01B	Rap ⁺ Rap ⁻ (-314-294)
RPL02A	Rap ⁺ (-514)
RPL02B	Rap ⁺ Rap ⁺ (-397+73)
RPL03	Abf1p (-232)
RPL04A	Abf1p (-181)
RPL04B	Abf1p (-189)
RPL05	Rap ⁺ Rap ⁺ (-293-190)
RPL06A	Rap ⁺ Rap ⁻ (-354-335)
RPL06B	Rap ⁻ (-440)
RPL07A	Rap ⁺ Rap ⁺ (-254-236)
RPL07B	Rap ⁺ (-221)
RPL08A	Rap ⁺ (-242)
RPL08B	Rap ⁺ (-287)
RPL09A	Rap ⁻ (-312)
RPL09B	
RPL10	Rap ⁺ Rap ⁺ (-271-254)
RPL11A	Rap ⁺ Rap ⁻ (-385-368)Reb1(-246)
RPL11B	Rap ⁺ Rap ⁺ Rap ⁺ (-255-240-34)
RPL12A	Rap ⁺ Rap ⁻ (-530-513)
RPL12B	Rap ⁺ Rap ⁻ (-508-491)
RPL13A	Rap ⁻ Rap ⁺ Rap ⁺ (-283-221-201)
RPL13B	4xRap ⁺ (-424-408-349-334)Abf1p (-297)

Table 1. Continued

RPL14A	Rap ⁺ Rap ⁺ (-413-394)
RPL14B	Rap ⁺ (-291)
RPL15A	Abf1p (-268)Rap ⁻ Rap ⁺ (-235-101)
RPL15B	Reb1(-161)
RPL16A	Rap ⁺ (-407)
RPL16B	Rap ⁺ Rap ⁺ (-407-390)
RPL17A	Rap ⁺ Rap ⁺ (-367-349)
RPL17B	Rap ⁺ Rap ⁺ Rap ⁻ (-540-526-209)
RPL18A	Rap ⁻ Rap ⁻ (-280-245)
RPL18B	
RPL19A	Rap ⁺ Rap ⁺ (-434-418)
RPL19B	Rap ⁺ (-389)
RPL20A	
RPL20B	Rap ⁺ (-372)
RPL21A	Rap ⁺ Rap ⁺ (-355+61)
RPL21B	4xRap ⁺ (-266-250-150-146)
RPL22A	Rap ⁺ Rap ⁺ (-250-230)
RPL22B	Rap ⁺ Rap ⁺ (-247-224)
RPL23A	Rap ⁺ Rap ⁺ (-351-191)
RPL23B	Rap ⁺ Rap ⁺ (-391-373)
RPL24A*	Rap ⁺ (-837)
RPL24B*	Rap ⁺ Rap ⁺ (-997-763)
RPL25	Rap ⁻ (-396)
RPL26A	Rap ⁺ Rap ⁻ (-314-272)
RPL26B	Rap ⁺ Rap ⁻ (-420-403)
RPL28	Rap ⁺ Rap ⁺ (-282-267)
RPL27A	Rap ⁺ Rap ⁺ (-419-405)
RPL27B	Rap ⁺ Rap ⁺ (-353-338)
RPL29*	Rap ⁺ Rap ⁺ (-633-615)
RPL30	Rap ⁻ (-333)
RPL31A	Rap ⁺ (-428)
RPL31B	5xRap ⁺ (-558-533-525-517-509)Abf1p(+83)
RPL32*	Rap ⁺ Rap ⁻ (-809-790)
RPL33A	Rap ⁺ Rap ⁺ (-373-357)
RPL33B	Rap ⁺ (-473)
RPL34A	Rap ⁻ Rap ⁻ (-353-314)
RPL34B	Rap ⁻ Rap ⁻ (-483-455)
RPL35A	Rap ⁺ (-295)
RPL35B	Rap ⁺ Rap ⁺ Rap ⁻ (-287-182+5)
RPL36A	Rap ⁺ (-165)
RPL36B	Rap ⁺ (-317)
RPL37A	Rap ⁺ Rap ⁺ (-386-361)
RPL37B	Rap ⁺ Rap ⁺ (-243-219)
RPL38	Rap ⁺ Rap ⁺ (-290-272)
RPL39	Rap ⁺ Rap ⁻ Rap ⁻ (-573-415-373)
RPL40A	Rap ⁺ Rap ⁻ (-359-340)
RPL40B	Rap ⁻ (-245)
RPL41A	Rap ⁻ Rap ⁺ (-594-278)
RPL41B	Rap ⁺ (-347)Reb1(-213)
RPL42A	Rap ⁺ Rap ⁺ (-373-353)
RPL42B	Rap ⁺ Rap ⁺ Rap ⁺ (-380-363-342)
RPL43A	Rap ⁺ (-226)
RPL43B	Rap ⁻ Rap ⁺ Rap ⁺ (-417-310-292)Abf1p (-192)
RPP00	Abf1p (-599) Rap ⁺ Rap ⁺ (-381-265)
RPP01α	Abf1p (-232)
RPP01β	Rap ⁺ Rap ⁺ (-277-253)
RPP02α	Rap ⁺ Rap ⁻ (-351-319)

Table 1. Continued

RPP02Bβ	Abf1p (-232)
RPS01A	Rap ⁺ (-248)
RPS01B	Rap ⁺ (-399)
RPS00A	Rap ⁺ Rap ⁺ (-504-486)
RPS00B	Rap ⁺ Rap ⁺ (-426-405)
RPS02	Abf1p (-465)Rap ⁻ Rap ⁺ (-412-391)Abf1p (-295)
RPS03	Rap ⁺ (-260)
RPS04A	Rap ⁺ Rap ⁺ (-396-367)Abf1p (-303)
RPS04B	Rap ⁺ Rap ⁺ (-483-470)
RPS05	Abf1p (-522)Rap ⁺ (-455)
RPS06A	Rap ⁺ Rap ⁺ (-458-433)
RPS06B	Abf1p (-568)Rap ⁺ Rap ⁺ (-324-307)
RPS07A	Rap ⁺ Rap ⁺ (-501-476)Abf1p (-217)
RPS07B	Rap ⁺ Rap ⁺ (-319-303)
RPS08A*	Rap ⁺ Rap ⁺ (-689-669)
RPS08B*	Rap ⁺ Rap ⁺ (-774-751)Reb1(-608)
RPS09A	Rap ⁺ (-245)
RPS09B	Rap ⁺ Rap ⁺ (-224-208)Abf1p (+10)
RPS10A	Abf1p (-560)Rap ⁺ Rap ⁻ (-361-342)
RPS10B	Rap ⁺ Rap ⁻ (-209-192)
RPS11A	Rap ⁺ (-385)
RPS11B	Abf1p (-597)Rap ⁺ Rap ⁺ (-307-291)
RPS12	Rap ⁺ Rap ⁺ (-295-269)
RPS13	Rap ⁺ (-233)
RPS14A	Rap ⁺ (-269)
RPS14B	Rap ⁻ Rap ⁺ (-555-206)
RPS15	Rap ⁺ Rap ⁻ Rap ⁺ (-414-382-63)
RPS16A	
RPS16B	Rap ⁻ Rap ⁻ Rap ⁺ (-447-427-365)
RPS17A	Rap ⁺ Rap ⁺ (-358-345)
RPS17B	Rap ⁺ Rap ⁺ (-323-299)
RPS18A	Rap ⁺ Rap ⁺ (-328-313)
RPS18B	Rap ⁺ Rap ⁺ (-400-375)
RPS19A	Rap ⁺ Rap ⁺ (-275-249)
RPS19B	Rap ⁺ Rap ⁻ (-388+59)
RPS20	Rap ⁺ Rap ⁺ (-260-243)Reb1(-187)
RPS21A	Rap ⁺ (-226)Abf1p (-141)
RPS21B	Rap ⁺ Rap ⁺ (-345-328)
RPS22A	Rap ⁺ Rap ⁺ Rap ⁻ (-271-229-71)
RPS22B*	Abf1p (-737)
RPS23A	Rap ⁺ Rap ⁺ (-388-371)
RPS23B	Rap ⁺ Rap ⁺ (-329-311)
RPS24A	Rap ⁺ (-339)
RPS24B	Rap ⁺ (-483)
RPS25A*	Rap ⁺ (-559)
RPS25B*	
RPS26A*	Rap ⁺ Rap ⁺ Rap ⁺ (-743-735-731)
RPS26B*	Rap ⁺ (-681)
RPS27A	Rap ⁺ (-193)
RPS27B	
RPS28A	Rap ⁺ (-495)Abf1p (-163)
RPS28B	Abf1p (-181)
RPS29A*	
RPS29B	Rap ⁻ Rap ⁺ (-465-239)
RPS30A	Rap ⁺ Rap ⁺ (-317-227)
RPS30B	Rap ⁺ (-230)
RPS31	

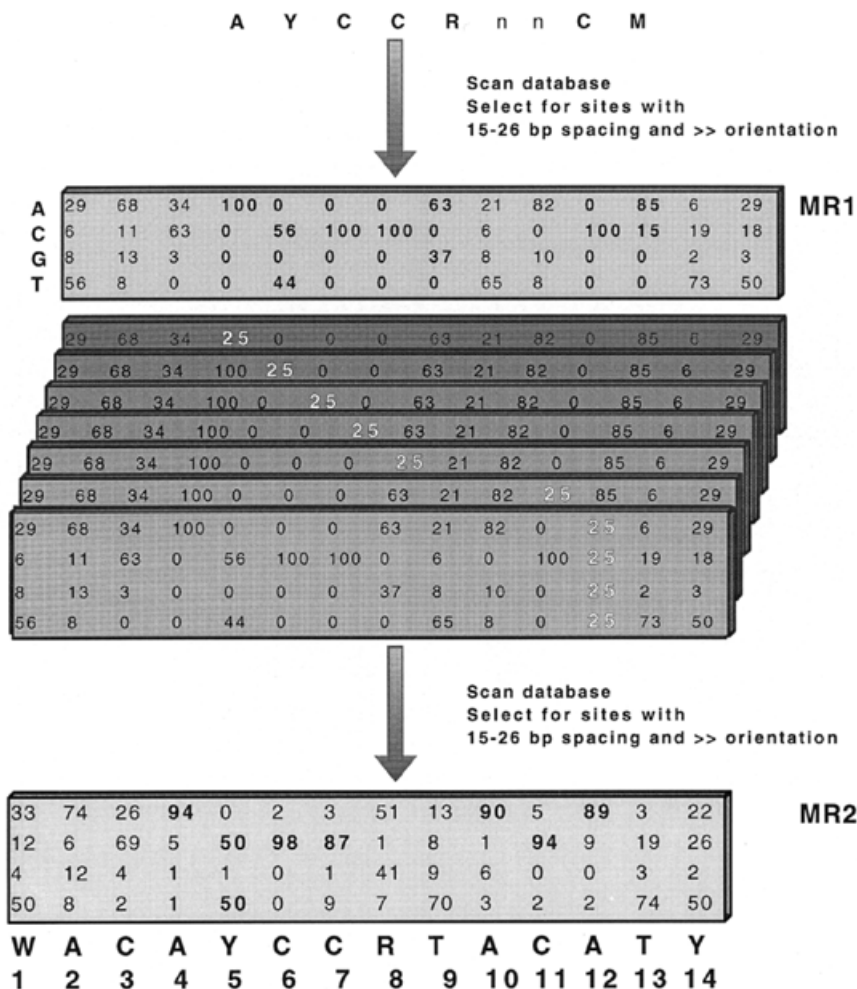


Fig. 1. The computational selection strategy which has been followed. Each panel represents a matrix. Starting with the AYCCRNNCM string, the 14 bp MR1 matrix was created by selecting for a distance and orientation that was found to be preferred in duplicate Rap1p sites. In order to decrease the influence of the first string search (100% scores at the MR1), seven MR1-derived matrices were created and equalized at the respective positions. After scanning the rp database, 129 matches to these matrices were obtained and used to generate matrix MR2. Values are percentages. Sequence preferences with strong conservation are indicated in bold.

Results and discussion

The binding specificity of Rap1p as determined by matrix analysis

In order to define the sequence requirements of a functional Rap1p binding site more precisely, we constructed a database of rp gene promoter regions, using the complete sequence of the *S.cerevisiae* genome. This database, encompassing 137 sequences, was then scanned using the low-stringency consensus string AYCCRNNCM for a Rap1p site. A total of 204 matches was found. These potential Rap1p binding sites predominantly occur within a distance of 200–500 bp upstream from the translational start codon ATG. Furthermore, a large portion is arranged in tandem:

51% of the sites identified are part of a duplicate set separated by an arbitrarily chosen distance of <50 bp between the upstream ends of the two members. Surprisingly, the duplicate sites that obey this criterion show a strong bias in their spacing: in 85% of the cases, the separation is 15–26 bp. Moreover, there also is a preferred orientation of the two sites: in 70% of the cases, the two sites are oriented 5' to 3'. Although at present the biological significance of these preferences is unknown, it is obvious that they can be used as additional selection criteria in a computational analysis aimed at a more precise definition of functional Rap1p binding sites. In order to generate a nucleotide distribution matrix that may better reflect the nucleotide requirements of Rap1p, we followed the strategy that is depicted in Figure 1. First, 62 sites from

the above-mentioned string search, that fulfil the criteria of spacing and orientation, were selected. From these sites, a 14 bp matrix was constructed, denoted as MR1, which represents the strongest preferences having an extra nucleotide position on either side. In order to decrease the influence of the first string search, seven additional matrices were generated, in which the nucleotide distribution at each of the seven positions biased by the AYCCRNNCM string, was randomized (see Figure 1). All MR1-derived matrices and MR1 itself were used to scan the rp database (threshold = 0.85) and 129 matches were selected using the additional criteria of 5'-3' tandem orientation and a distance of 15–26 bp. From these sequences, the final matrix, MR2, shown in Figure 1, was generated. As can be seen, the MR2 matrix shows only minor deviations from MR1 at the positions where the latter has a 100% score (4A, 6C, 7C and 11C). This indicates that the AYCCRNNCM string, on which the original search was based, already exhibits a high specificity for Rap1p binding sites. Nevertheless, the mismatches that seem to be allowed at these positions are mainly confined to particular nucleotides (4A=>C 5%, 7C=>T 9% and 11C=>A 5%), which have been shown to be allowed according to *in vivo* analysis (Nieuwint *et al.*, 1989; see hereafter). These observations not only underline that the MR2 matrix better reflects the nucleotide requirements of Rap1p than MR1, but also indicate that the number of false-positive sites that are incorporated in the matrix is small. In fact, 6C, which has been shown experimentally to be the only position at which no mismatch is allowed (Nieuwint *et al.*, 1989; Vignais *et al.*, 1990), is present in 98% of the binding sites that were used to generate the MR2 matrix.

In retrospect, the MR2 matrix was compared with the existing Rap1p matrix RAP_C (Quandt *et al.*, 1995), which we decided not to use (see Materials and methods). In this matrix, a core region is present in which five nucleotide positions (4A, 5C, 6C, 7C and 8A) have a maximum score of 100%. The MR2 matrix displays a higher variability at these nucleotide positions (Figure 1). Although the nucleotide positions in MR2 are 'weaker', the matrix is more specific for Rap1p binding, as can be observed from experimental data (see hereafter). Further, in almost all other nucleotide positions, 5' and 3' of the mentioned region, the MR2 matrix is actually 'stronger', exhibiting higher scores (2A 59=>74%; 3C 46=>69%, 9T 62=>70%, 10A 80=>90%, 12A 77=>89% and 13T 48=>74%) and better conservations [higher C_i value (Quandt *et al.*, 1995), not shown] than the RAP_C matrix. This finding indicates that Rap1p sites which are incorporated in the MR2 matrix compensate for mismatches in the core region (position 4–8) by increased specificity at the flanking regions. This is, in fact, the advantage of our approach, since the strategy not only results in incorporation of information in the matrix concerning the tolerance for mismatches at certain nucleotide positions, but also

captures compensating specificities at other nucleotide positions.

The specificity of the MR2 matrix is in agreement with in vivo and in vitro data

In general, the sequence preferences shown by matrix MR2 are in agreement with *in vivo* data from previous analysis in our laboratory (Nieuwint *et al.*, 1989) showing that each of the nucleotides of the Rap1p binding sequence ACACCCA-TACATTT can be replaced by one or more other nucleotides, with the exception of 6C. Of the five substitutions that were found to abolish transcription of the reporter gene, four do not, or only marginally, occur in the matrix (4A=>T 1%, 6C=>T 0%, 6C=>G 0%, 7C=>G 1%), whereas substitution 7C=>A is found in 3% of the sites used to generate the MR2 matrix. We do not know whether the latter sites do indeed function as Rap1p binding sites *in vivo* or whether they should be considered as false positives. On the other hand, the majority of nucleotide replacements that still support wild-type transcription levels are also found in (a minority of) the sites that were used to create the MR2 matrix (4A=>G 1%, 4A=>C 5%, 5C=>T 50%, 7C=>T 9%, 8R=>C 1%, 11C=>A 5%, 11C=>T 2%, 12A=>T 2% and 12A=>C 9%). Thus, the MR2 matrix indicates a higher sequence specificity than was deduced from mutational analysis. Probably, this is due to the fact that individual Rap1p sites in rp gene promoters generally have more than one deviation from the consensus sequence, and hence require more 'preferred nucleotides' at the stringent threshold used, than the high-affinity (Vignais *et al.*, 1990) Rap1p binding sequence ACACCCA-TACATTT in which the nucleotide substitutions have been made.

Data from *in vitro* binding of Rap1p (Vignais *et al.*, 1990) are also in agreement with the nucleotide distribution (data not shown), although we believe the *in vivo* function of Rap1p at promoter regions to be more reliable for this analysis.

The sequences WACAYC and TACATY in the 5' and 3' part of the consensus sequence established by the MR2 matrix (Figure 1), respectively, show a strong similarity. This is consistent with the finding that the DNA-binding domain of Rap1p contains two structurally similar and tandemly oriented subdomains, i.e. subdomain 1 and 2, which recognize the same sequence in telomeric repeats (König *et al.*, 1998). It could be speculated that the linker between subdomain 1 and 2 allows a larger spacing between the two protein subdomains, which would result in binding of Rap1p when 1 bp is inserted between the recognition sequences of the subdomains. In this respect, it would be expected that more Rap1p sites can be found when an equally distributed nucleotide position is inserted between position 8R and 9T of the MR2 matrix. However, scanning the rp gene promoter database with this MR2-derived matrix at a threshold of 0.85,

only a small number of 11 matches were found, of which the vast majority correlate to sites that are found with the MR2 matrix itself. Similar results are obtained by deleting nucleotide position 8R, corresponding to a gap between the sequence repeats. Thus, these data again suggest that the MR2 matrix represents the binding requirements of Rap1p, in which no gap or bulge is allowed.

We believe that the MR2 matrix represents the *in vivo* sequence requirements for binding of Rap1p, on the basis of: (i) the large number of sequences in *rp* gene promoters that show a strong similarity to the consensus established by this matrix; (ii) the fact that the sites have been selected for Rap1p-specific parameters not related to DNA-sequence, a preferred tandem orientation and a preferred 15–26 bp spacing; (iii) the finding that the preferences represented by the MR2 matrix are consistent with *in vivo* and *in vitro* mutational analyses; (iv) the fact that we do not find any indication that the matrix should allow for a gap or a bulge between the subsequences recognized by subdomain 1 and 2.

Redefining Rap1p binding sites in oligonucleotides selected from a randomized oligonucleotide pool for their ability to bind Rap1p in vitro

The prime data that can be used to validate the obtained matrix come from a study in which 47 sequences have been selected from a pool of random oligonucleotides by their ability to bind Rap1p *in vitro* (Graham and Chambers, 1994). In order to fit the sequences to the consensus sequence RMACCCANNCAYY, 1–5 mismatches had to be allowed and gaps and bulges needed to be introduced. When the Rap1p-interacting oligonucleotides were subjected to Matrix analysis using the MR2 matrix, to our surprise a number of matches were found, the core of which corresponds to the ATCC sequence of the *Bam*HI site flanking the oligonucleotides. When the oligonucleotides were aligned with respect to the *Bam*HI site (Figure 2), they could be subdivided into four groups, three of which show a strong homology at the first five positions of the randomized oligonucleotides. The 3' consensus sequences found are either RBACA, GTGCA or RCCCA. Apparently, the fourth and fifth nucleotides in these three groups are in nearly all cases C and A, which corresponds to the strong preferences revealed by the nucleotide position 11C (94%) and 12A (89%) in the MR2 matrix, respectively (see Figure 1). Group 1 and 2 not only show considerable homology to the preferences (positions 8–12) of the MR2 matrix, but in addition display preferences for Cs downstream to the homologous sequences. Although these preferred nucleotides show a moderate bias in the MR2 matrix (13C 19% and 14C 26%), this observation is nevertheless not surprising, since Cs at these positions are also found

in telomeric repeat sequences. Group 3 has the lowest score in a search using the MR2 matrix (0.71). Since the *Bam*HI-linked consensus RCCCA of this group also corresponds to the core sequence AYCCR of the MR2 matrix, it is suggested that two low-affinity binding sites in the same orientation, which are overlapping and therefore mutually exclusive, could result in an overall high-affinity binding sequence. This latter insight is supported by recent data from Stormo and Fields (1998). Group 4 consists of 10 oligonucleotides without an homologous region adjacent to the *Bam*HI site. In these oligonucleotides, however, an ACCC core is present. Since in six of the 10 oligonucleotides this core corresponds to a match with the MR2 matrix (threshold = 0.75), these motifs are thought to be responsible for Rap1p binding. Finally, in five oligonucleotides, no match to the MR2 matrix could be found, not even at a low threshold of 0.65. Since two of these oligonucleotides could not bind Rap1p *in vitro* (Graham and Chambers, 1994), all five nucleotides are considered as non-binders.

In summary, it seems that the selection procedure used by Graham *et al.* in fact reflects an optimization of the 3'-recognition sequence for Rap1p binding. This new insight strongly suggests that Rap1p binding requires a higher specificity than could be deduced from the alignment made by the respective authors.

Defining Rap1p sites in rp gene promoters by a search with the MR2 matrix

Using matrix MR2, which represents the sequence requirements for Rap1p binding, we wished to define all putative Rap1p binding sites in the *rp* database. Since at a threshold of 0.82 the increase in matches in the *rp* database is still higher than the increase in the random sequence file (see Materials and methods), it indicates that at least up to this threshold the MR2 matrix recognizes Rap1p binding sites specifically. Therefore, the database was scanned using the MR2 matrix at a threshold of 0.82.

In the *rp* database, 222 Rap1p sites were found with the MR2 matrix at this threshold: 17 *rp* genes, no Rap1p site; 37 *rp* genes, one site; 68 *rp* genes, two sites; 12 *rp* genes, three sites; two *rp* genes, four sites; one *rp* gene, five sites. These Rap1p sites with their orientation and distance from the translational start ATG are listed in Table 1. A number of the obtained sites have been experimentally shown to bind Rap1p *in vitro* [Raué and Planta (1991) and references therein].

As is shown in Table 1, we also tried to define the putative binding sites for Abf1p and Reb1p, using the consensus sequences RTMRYB(N₄)ACG for Abf1p and CGGGTARNNR for Reb1p (Remacle and Holmberg, 1992), respectively.

	WACAYCCRTACATY (MR2 consensus)		
	<i>Bam</i> HI	<i>Bgl</i> III	
Group 1:			
N13.02	gggatccAGACCCCTAGCTagatcttt		
N13.15	gggatccGTACACCCGATGTagatcttt		
N13.16	gggatccGTACACCCCAACCagatcttt		
N13.26	gggatccACACACCAGCCAAgagatcttt		
N13.28	gggatccATACACCAGCCagatcttt		
N13.29	gggatccGGACACCCTACagatcttt		
N13.32	gggatccGTACCCCGCTAagatcttt	+	
N13.37	gggatccAGACATCTTACCgagatcttt		
N13.40	gggatccAGACACCCTACagatcttt		
N13.41	gggatccATACATACGACCagatcttt	+	
N13.44	gggatccGGACACCTACGgagatcttt		
consensus	RBACA 0.80 mean		
Group 2:			
N13.06	gggatccGTGCATCTCCagatcttt		
N13.13	gggatccGTGCACCACCTGagatcttt		
N13.14	gggatccGTGCACCAGACTgagatcttt		
N13.18	gggatccGTGCACCAGCCagatcttt		
N13.19	gggatccGTGCACCATCAGTagatcttt	+	
N13.22	gggatccGTGCACCCTGCCagatcttt		
N13.23	gggatccGTGCACCCTGCAagatcttt	+	
N13.27	gggatccGTGCACCCTGACagatcttt		0.78
N13.34	gggatccGTGCACCATGTTgagatcttt	+	
N13.39	gggatccGTGCACCATAAAagatcttt		0.76
consensus	GTGCA 0.73 mean		
Group 3:			
N13.07	gggatccACCAGAACATTTagatcttt		
N13.08	gggatccGCCAGACCGCTagatcttt		
N13.17	gggatccACCAGCCAGCCagatcttt	+	0.76
N13.24	gggatccGCCATACACAagatcttt		0.78
N13.25	gggatccACCCTGACATGagatcttt	+	0.84
N13.35	gggatccGCCATATATTCagatcttt		0.84
N13.36	gggatccACCATTTCACCagatcttt		0.79
N13.42	gggatccACCCAACCCCGAagatcttt		
N13.45	gggatccACCAGCCAGagatcttt	+	0.75
N13.48	gggatccACCCAATACCAATagatcttt		
consensus	RCCCA 0.71 mean		
Group 4:			
N13.01	gggatccACCACACACCagatcttt	+	0.81
N13.47	gggatccACCAGCCCAagatcttt	+	
N13.03	gggatccCGTACACCACagatcttt	+	
N13.20	gggatccGTGACACCAGATagatcttt		0.75
N13.11	gggatccGCTAACACCAGagatcttt		0.78
N13.30	gggatccACCATACACCagatcttt		0.83
N13.04	gggatccCATCAGTACACCagatcttt		0.76 0.75
N13.05	gggatccCGGTGAACACCagatcttt		
N13.12	gggatccGAGCCTAACACCagatcttt	+	
N13.43	gggatccCAACGATACACCagatcttt		0.75
NON-BINDERS:			
N13.21	gggatccGTTTTGATGTGCTagatcttt	-	
N13.38	gggatccGAAGTCGTGCCagatcttt	-	
N13.09	gggatccGGGGGCTGCAGCagatcttt		
N13.33	gggatccGGCCAGAGATTgagatcttt		
N13.46	gggatccATGCTCTATGGCagatcttt		
NOT DETERMINED:			
N13.31	ctGCACCAGACCA	+	0.77

Fig. 2. Sequences as published by Graham and Chambers (1994) aligned relative to the *Bam*HI restriction site. The sequences represent 47 clones that were obtained from a pool of double-stranded oligonucleotides which were randomized over a region of 13 nucleotides and were selected for the ability to bind Rap1p *in vitro* using the SAAB procedure. The *Bam*HI and *Bgl*III restriction sites were added according to the sequences published (the restriction sites in N13.31 could not be deduced). For groups 1–3, which contain homologous sequences 3' to the *Bam*HI core atcc, the first five nucleotides of the randomized nucleotides are displayed beneath each group together with a mean matrix similarity score. Additional matches to the MR2 matrix having a matrix similarity score of >0.75 are underlined with the respective scores displayed at the right side. (+) indicates sequences that have been shown to bind Rap1p *in vitro* and (–) indicates sequences not able to bind Rap1p *in vitro*.

Table 2. Duplicate Rap1p binding sites (15–26 bp) found by scanning the genome using the MR2 matrix with a threshold of 0.82. The distance from the translation initiation codon is calculated from the 5' side of the Rap1p site which is most close to the ATG

ORF	Name	Orientation	Distance from ATG	CAI ^b
YMR242c	RPL20A	>>	805	0.66
YBR118w	TEF2	>>	456	0.88
YGL008c	PMA1	>> ^a	762	0.73
YNR017w	TIM23 ^c	<<	422	0.12
YNR016c	ACC1 ^c		1094	0.33

^aIn coding region of YGL007w (MIPS ID: questionable ORF).

^bCodon Adaptation Index, reflects the expression level of the gene.

^cGenes with head-to-head orientation.

Seventeen rp genes were found to have no match to MR2 (threshold = 0.82). Inspection of these sequences revealed five rp genes that contain missed single or duplicate Rap1p sites which are expected to be biologically functional (see Figure 3). Hence, 11 rp genes are presumed to be not regulated by Rap1p. Of these genes, eight were found to contain a single Abf1p binding site. Generally, all these sites contain an ACGCA motif of yet unknown function, in addition to a T-rich region with which Abf1p has been shown to function synergistically (Gonçalves *et al.*, 1995). Further, one gene was found to harbour a single Reb1p binding site, whereas for four rp genes we could not define the UAS according to the criteria used, but which might be regulated by non-consensus Reb1p, Rap1p or Abf1p binding sites (see Figure 3).

We propose that the listed sites in Table 1 and Figure 3 reflect the UASs of rp gene promoters.

Duplicate Rap1p sites in the yeast genome

Because the presence of duplicate Rap1p sites in the rp gene promoters is very pronounced, and since many genes other than rp genes are regulated by Rap1p, viz. glycolytic genes (Chambers *et al.*, 1990; Scott and Baker, 1993), it is interesting to analyse whether more genes or classes of genes may be regulated by a duplicate Rap1p site. The whole genome was searched for putative functional duplicate Rap1p sites by selecting for matches to the MR2 matrix (threshold = 0.82) having a distance of 15–26 bp.

In addition to (sub)telomeric sequences and the 56 duplicate Rap1p sites with a distance of 15–26 bp that can be assigned to 57 rp genes (see Table 1), only four other promoters were found to contain a duplicate Rap1p site. In fact, one of these genes was found to be an rp gene which was missed in the previous analysis because it is likely to contain an intron in its leader (an intron-specific TACTAAC motif at –26 from the ATG). Two other genes that have been found in this search are *TEF2*, encoding translation elongation factor 2, and *PMA1*, encoding the major isoform of the proton-trans-

17 RP-gene promoters without any match to MR2 (threshold>8.2)
from ATG

Missed Rap1p-sites:

1	RPL01A	-177	<u>GGTGCACGGATTTT</u> AGCAGTCTTTTTCTTTCTTGGCTTTTGCGA rap1 0.80
2	RPL09B	-248	<u>GGTGCAC</u> TGGTGTGTTTCCAAGACTGCACATTTAACTGGGAAT ₁₁ rap1 0.75
3	RPL20A	duplicate	Rap1p-site: putative intron in leader (see genome search)
4	RPS27B	-439,-425	<u>TTAGCCCATACATTTT</u> CACCCATGCACCATTGGATTATAAAGA ₇ 0.80 0.80

Single Abf1p-binding sites:

5	RPL03	-232	<u>ATCGTTTTGTACGTTTT</u> TCAAGAAGCG <u>ACGCACA</u> ACTGTTTTCCAT ₁₈
6	RPL04A	-181	<u>ATCACGTACACAG</u> <u>ACGCACA</u> AGTGAGAAGTGAAAAATTTTTTTTCAATCTGA ₂₅
7	RPL04B	-189	<u>ATAATATTTACGAAAGC</u> ATCGCGCAG <u>ACGCACA</u> ATTTTTTTTAAACAT ₁₂
8	RPP01alpha	-232	<u>ATCGTGAGGTACGA</u> ATATCGTAAGATGATACC <u>ACGCACA</u> ACTTTGTATATGAT ₁₂ CAT ₆
9	RPP02Bβ	-232	<u>CGTGTAAGATTAC</u> CAAGCAGAAAAATGTAATAATTTTTCTTTTTCCAAATTTT
10	RPS22B	-173*	<u>CGTCAAAAATGACGG</u> <u>ACGCACA</u> CTTTTCTCAGAAATTTTTTTTTTTCATTTCTC
11	RPS28B	-181	<u>CGTACAAAATGATC</u> <u>ACGCACA</u> CTTTTGACAATTTTTTTTTTTTCATCTGTTTTT
	RPS28A#	-163	<u>CGTCTAGAGTGACC</u> <u>ACGCACA</u> CTTTTTTGATAAATTTTTTTTTTCTTGGTCGTTG
12	RPS16A	-228	<u>GTAAATGAGTA</u> <u>ACGCACA</u> TAGTGTATTTATCCAAAGGAAAGAATTGTTATTTTTTA
	RPL08A#	-157	<u>ATAATAACCGA</u> <u>ACGCACA</u> ACAAATTTGGAAAAACCA <u>ACGCACA</u> AAAAAAAAA
	RPL08B#	-176	<u>ATCATACTATACG</u> ATAGAA <u>ACGCACA</u> TTGAAACTTTTCCCATCTCAAAAT
	RPS02^	-256	<u>TTAGTATTGCA</u> <u>ACGCACA</u> GCTTCCCAGGACGCCTAGCTATTTTTT

Single Reb1p-binding sites:

13	RPL15B	-161	<u>TCGTTACCCGAAAAC</u> ATCTTTCTTGTTCCTTTCGATACTTTTT
----	--------	------	---

Miscellaneous:

14	RPL18B	-176	<u>ACGTTACCCGACCT</u> CGTTATTTTACGACAACATATGATAAAATT Reb1p?
15	RPS29A	-297*	<u>AACACCTGTACCT</u> CAAAAAACCCATTTCTTTTCTGTTTTT Rap1p? 0.80
16	RPS25B	-168	<u>GACCACCATACAGT</u> TTAGTCTGCAGGCCAATTTGCTTTTGCTTTGACT ₈ Rap1p? 0.77
17	RPS31	-110	<u>CGTTATATCAGA</u> AACTTCCCTTACTTCTATCTTTTATTCCAATACAAGAAG Abf1p?

Fig. 3. Summary of the UASs (underlined) of 17 rp genes that do not contain a match to the MR2 matrix at a threshold of 0.82. *from 5' splice site. # additional match to MR2. RPS02^ contains a duplicate Rap1p site and two matches to an Abf1p site. Mismatches are indicated in italics.

porting P-type ATPases. Both genes have indeed shown to be regulated by Rap1p (Huet and Sentenac, 1987; Vignais *et al.*, 1990; Rao *et al.*, 1993). Finally, one other promoter contains a duplicate Rap1p site. Whether this duplicate set is involved in the regulation of these genes is, to our knowledge, not known. In conclusion, duplicate Rap1p sites apparently are typical for rp gene promoters.

The MR2 search on genomic DNA (threshold = 0.82) revealed telomeric and subtelomeric stretches of Rap1p sites of which the vast majority match the sequence CACACCACACACC/A (score = 0.89). This supports the conclusion of Gilson *et al.* (1993), who have shown that sites that are both reactive to KMnO_4^- and protected from DNase I coincide strongly with the mentioned telomeric Rap1p binding site. This suggests that this is the main Rap1p binding site in (sub)telomeric tracts. In the same study, an additional site having the sequence CACACCCACACCAC was identified. The fact that this site has a lower affinity for Rap1p and is less

protected from DNase I is in agreement with its lower degree of similarity to the MR2 matrix (score = 0.78).

Using a threshold of 0.82 and excluding the incomplete (sub)telomeric sequences in the database, we found a total of 3569 Rap1p sites in the yeast genome. The number of Rap1p molecules per haploid cell has been estimated at ~4000–5500 (Buchman *et al.*, 1988; Gilson *et al.*, 1993) and telomeric sequences have been estimated to bind a minimum of 600 molecules (Gilson *et al.*, 1993). These estimations indicate that the vast majority of Rap1p molecules may be DNA bound or that the Rap1p sites found are not occupied continuously.

The distance constraint of multiple Rap1p sites in rp gene promoters and (sub)telomeric sequences

Figure 4A shows the distances of duplicate Rap1p sites (≤ 50 bp) found in the rp gene database search using the MR2

matrix (threshold = 0.82). As has been discussed for the matches found with the AYCCRNNCM string, such duplicate sites show a strong preference for a tandem 5'-3' orientation and a 15-26 bp spacing. Thus, when Rap1p binding sites are found in duplicate, they are in close proximity to each other. From Figure 4, however, it is evident that the distances are not distributed evenly. The distribution seems to be phased, with maxima at 17 and 25 bp, and a minimum at 22 bp. Both the close distance between binding sites and the apparent phasing are notably similar to the action of Rap1p and Gcr1p binding sites in glycolytic genes. Drazinic *et al.* (1996) have shown that a Gcr1p binding site should be in close proximity to a Rap1p binding site in order to sustain high transcriptional yields of a reporter gene and to ascertain the *in vivo* occupancy of the Gcr1p binding site. Furthermore, a strong effect of the phasing between a Rap1p and a Gcr1p binding site was reported. From these observations, it has been proposed that Rap1p facilitates binding of Gcr1p by either protein-protein interactions between Rap1p and Gcr1p or by increasing the affinity of the Gcr1p binding site for Gcr1p due to Rap1p-induced bending of DNA (Drazinic *et al.*, 1996). The same mechanism could explain the distance and phasing dependency between duplicate Rap1p sites in *rp* gene promoters. The former model would imply that Rap1p interacts with itself, but to our knowledge, this has never been shown. Nevertheless, the distance distribution in Figure 4A is likely to reflect co-operativity in binding, which is further supported by the observation that two Rap1p binding sites activate transcription synergistically (Woudt *et al.*, 1986). It has been shown that binding co-operativity is of minor importance in telomeric DNA (Gilson *et al.*, 1993). This is consistent with the finding that the distance distribution of Rap1p sites in telomeric DNA is remarkably different from the distribution in *rp* gene promoters (see Figure 4B). The preferred spacing of 13 bp in (sub)telomeric sequences may result from competition between telomere length increase—by telomerase activity—and telomere degradation. A distance of 13 bp therefore probably reflects the minimal distance which stably binds an additional Rap1p molecule, increasing the protection of newly formed DNA against degradation.

In conclusion, we propose that the distance distribution of duplicate Rap1p sites in *rp* gene promoters reflects the borders within which binding co-operativity operates.

Acknowledgements

We thank Ellen de Groot and Dr Walter Ravenek for their support in constructing the database. Furthermore, we are indebted to Dr Dick Raué (this laboratory) and Drs Christopher Workman and Gary Stormo (University of Colorado) for their advice and critical reading of the manuscript. The work presented in this paper was financially supported by the Dutch Organization for Scientific Research (N.W.O).

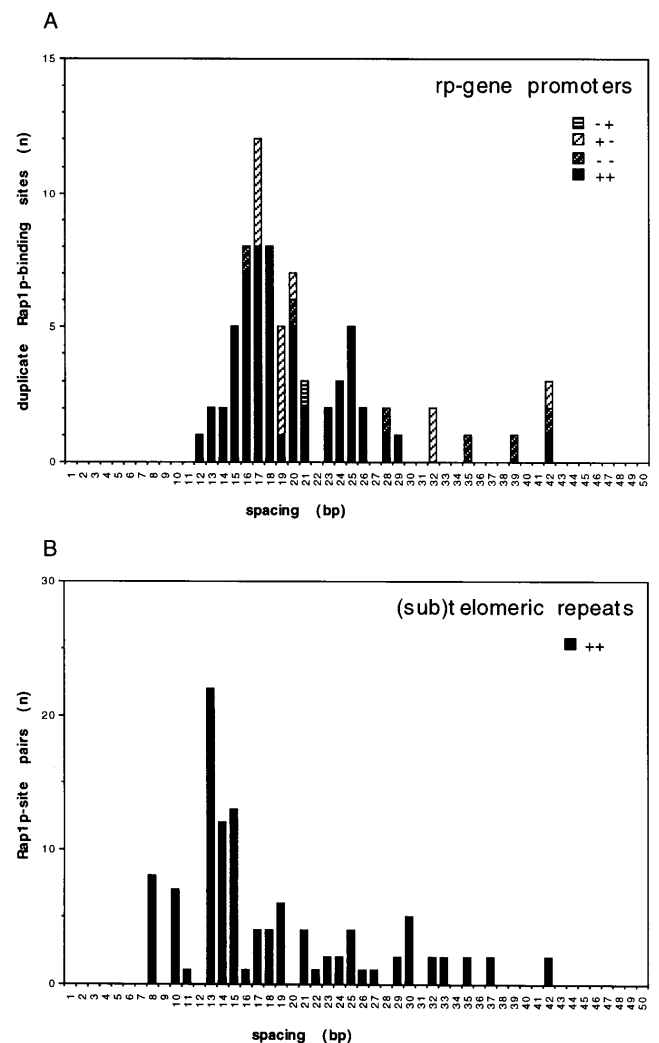


Fig. 4. Distances between the matches to the MR2 matrix (threshold = 0.82) as were found in multiple Rap1p sites in (A) *rp* gene promoters or (B) telomeric sequences as obtained from the genome search. In *rp* gene promoters, two triplicate Rap1p sites with distances of 8 + 8 and 8 + 4 were counted as duplicate Rap1p sites with distances of 16 and 12 bp, respectively. In (sub)telomeric sequences, 8 and 10 bp distances were isolated and could therefore not be added.

References

- Aparicio, O.M., Billington, B.L. and Gottschling, D.E. (1991) Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. *Cell*, **66**, 1279-1287.
- Biswas, S.B. and Biswas, E.E. (1990) ARS binding factor I of the yeast *Saccharomyces cerevisiae* binds to sequences in telomeric and nontelomeric autonomously replicating sequences. *Mol. Cell. Biol.*, **10**, 810-815.

- Bucher,P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.*, **212**, 563–578.
- Buchman,A.R. and Kornberg,R.D. (1990) A yeast ARS-binding protein activates transcription synergistically in combination with other weak activating factors. *Mol. Cell. Biol.*, **10**, 887–897.
- Buchman,A.R., Kimmerly,W.J., Rine,J. and Kornberg,R.D. (1988) Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences and telomeres in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **8**, 210–225.
- Chambers,A., Stanway,C., Tsang,J.S., Henry,Y., Kingsman,A.J. and Kingsman,S.M. (1990) ARS binding factor 1 binds adjacent to RAP1 at the UASs of the yeast glycolytic genes PGK and PYK1. *Nucleic Acids Res.*, **18**, 5393–5399.
- Conrad,M.N., Wright,J.H., Wolf,A.J. and Zakian,V.A. (1990) RAP1 protein interacts with yeast telomeres *in vivo*: overproduction alters telomere structure and decreases chromosome stability. *Cell*, **63**, 739–750.
- Della Seta,F., Ciafre,S.A., Marck,C., Santoro,B., Presutti,C., Sentenac,A. and Bozzoni,I. (1990) The ABF1 factor is the transcriptional activator of the L2 ribosomal protein genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **10**, 2437–2441.
- Devlin,C., Tice-Baldwin,K., Shore,D. and Arndt,K.T. (1991) RAP1 is required for BAS1/BAS2- and GCN4-dependent transcription of the yeast HIS4 gene. *Mol. Cell. Biol.*, **11**, 3642–3651.
- Drazinic,C.M., Smerage,J.B., Lopez,M.C. and Baker,H.V. (1996) Activation mechanism of the multifunctional transcription factor repressor-activator protein 1 (Rap1p). *Mol. Cell. Biol.*, **16**, 3187–3196.
- Einerhand,A.W., Kos,W., Smart,W.C., Kal,A.J., Tabak,H.F. and Cooper,T.G. (1995) The upstream region of the FOX3 gene encoding peroxisomal 3-oxoacyl-coenzyme A thiolase in *Saccharomyces cerevisiae* contains ABF1- and replication protein A-binding sites that participate in its regulation by glucose repression. *Mol. Cell. Biol.*, **15**, 3405–3414.
- Gilson,E., Roberge,M., Giraldo,R., Rhodes,D. and Gasser,S.M. (1993) Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites. *J. Mol. Biol.*, **231**, 293–310.
- Gonçalves,P.M., Griffioen,G., Minnee,R., Bosma,M., Kraakman,L.S., Mager,W.H. and Planta,R.J. (1995) Transcription activation of yeast ribosomal protein genes requires additional elements apart from binding sites for Abf1p or Rap1p. *Nucleic Acids Res.*, **23**, 1475–1480.
- Graham,I.R. and Chambers,A. (1994) Use of a selection technique to identify the diversity of binding sites for the yeast RAP1 transcription factor. *Nucleic Acids Res.*, **22**, 124–130.
- Hardy,C.F., Sussel,L. and Shore,D. (1992) A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes Dev.*, **6**, 801–814.
- Huet,J. and Sentenac,A. (1987) TUF, the yeast DNA-binding factor specific for UASrpg upstream activating sequences: identification of the protein and its DNA-binding domain. *Proc. Natl Acad. Sci. USA*, **84**, 3648–3652.
- König,P., Fairall,L. and Rhodes,D. (1998) Sequence-specific DNA recognition by the Myb-like domain of the human telomere binding protein TRF1: a model for the protein–DNA complex. *Nucleic Acids Res.*, **26**, 1731–1740.
- Kraakman,L.S., Mager,W.H., Grootjans,J.J. and Planta,R.J. (1991) Functional analysis of the promoter of the gene encoding the acidic ribosomal protein L45 in yeast. *Biochim. Biophys. Acta*, **1090**, 204–210.
- Kyrion,G., Boakye,K.A. and Lustig,A.J. (1992) C-terminal truncation of RAP1 results in the deregulation of telomere size, stability and function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **12**, 5159–5173.
- Lustig,A.J., Kurtz,S. and Shore,D. (1990) Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science*, **250**, 549–553.
- Mager,W.H. (1988) Control of ribosomal protein gene expression. *Biochim. Biophys. Acta*, **949**, 1–15.
- Mager,W.H. and Planta,R.J. (1990) Multifunctional DNA-binding proteins mediate concerted transcription activation of yeast ribosomal protein genes. *Biochim. Biophys. Acta*, **1050**, 351–355.
- Mager,W.H., Planta,R.J., Ballesta,J.G., Lee,J.C., Mizuta,K., Suzuki,K., Warner,J.R. and Woolford,J. (1997) A new nomenclature for the cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **25**, 4872–4875.
- Moretti,P., Freeman,K., Coodly,L. and Shore,D. (1994) Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. *Genes Dev.*, **8**, 2257–2269.
- Nieuwint,R.T., Mager,W.H., Maurer,K.C. and Planta,R.J. (1989) Mutational analysis of the upstream activation site of yeast ribosomal protein genes. *Curr. Genet.*, **15**, 247–251.
- Planta,R.J. and Mager,W.H. (1998) The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Yeast*, **14**, 471–477.
- Planta,R.J. and Raué,H.A. (1988) Control of ribosome biogenesis in yeast. *Trends Genet.*, **4**, 64–68.
- Planta,R.J., Gonçalves,P.M. and Mager,W.H. (1995) Global regulators of ribosome biosynthesis in yeast. *Biochem. Cell Biol.*, **73**, 825–834.
- Quandt,K., Frech,K., Karas,H., Wingender,E. and Werner,T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.*, **23**, 4878–4884.
- Rao,R., Drummond-Barbosa,D. and Slayman,C.W. (1993) Transcriptional regulation by glucose of the yeast PMA1 gene encoding the plasma membrane H (+)-ATPase. *Yeast*, **9**, 1075–1084.
- Raué,H.A. and Planta,R.J. (1991) Ribosome biogenesis in yeast. *Prog. Nucleic Acid Res. Mol. Biol.*, **41**, 89–129.
- Remacle,J.E. and Holmberg,S. (1992) A REB1-binding site is required for GCN4-independent ILV1 basal level transcription and can be functionally replaced by an ABF1-binding site. *Mol. Cell. Biol.*, **12**, 5516–5526.
- Scott,E.W. and Baker,H.V. (1993) Concerted action of the transcriptional activators REB1, RAP1 and GCR1 in the high-level expression of the glycolytic gene TPI. *Mol. Cell. Biol.*, **13**, 543–550.
- Shore,D. (1997) Telomere length regulation: getting the measure of chromosome ends. *Biol. Chem.*, **378**, 591–597.
- Stormo,G.D. and Fields,D.S. (1998) Specificity, free energy and information content in protein–DNA interactions. *Trends Biochem. Sci.*, **23**, 109–113.
- Teem,J.L. *et al.* (1984) A comparison of yeast ribosomal protein gene DNA sequences. *Nucleic Acids Res.*, **12**, 8295–8312.

- Tsang,J.S., Henry,Y.A., Chambers,A., Kingsman,A.J. and Kingsman,S.M. (1990) Phosphorylation influences the binding of the yeast RAP1 protein to the upstream activating sequence of the PGK gene. *Nucleic Acids Res.*, **18**, 7331–7337.
- Velculescu,V.E., Zhang,L., Zhou,W., Vogelstein,J., Basrai,M.A., Bassett,D.E., Jr, Hieter,P., Vogelstein,B. and Kinzler,K.W. (1997) Characterization of the yeast transcriptome. *Cell*, **88**, 243–251.
- Vignais,M.L. and Sentenac,A. (1989) Asymmetric DNA bending induced by the yeast multifunctional factor TUF. *J. Biol. Chem.*, **264**, 8463–8466.
- Vignais,M.L., Huet,J., Buhler,J.M. and Sentenac,A. (1990) Contacts between the factor TUF and RPG sequences. *J. Biol. Chem.*, **265**, 14669–14674.
- Wotton,D. and Shore,D. (1997) A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*. *Genes Dev.*, **11**, 748–760.
- Woudt,L.P., Smit,A.B., Mager,W.H. and Planta,R.J. (1986) Conserved sequence elements upstream of the gene encoding yeast ribosomal protein L25 are involved in transcription activation. *EMBO J.*, **5**, 1037–1040.