

A novel yeast system for *in vivo* selection of recognition sequences: defining an optimal c-Myb-responsive element

Tone Berge, Stine L. Bergholtz, Kristin Brevik Andersson and Odd S. Gabrielsen*

Department of Biochemistry, University of Oslo, PO Box 1041, Blindern, N-0316 Oslo 3, Norway

Received July 9, 2001; Revised and Accepted August 28, 2001

ABSTRACT

Yeast (*Saccharomyces cerevisiae*) has proved to be a highly valuable tool in a range of screening methods. We present in this work the design and use of a novel yeast effector–reporter system for selection of sequences recognised by DNA-binding proteins *in vivo*. A dual *HIS3–lacZ* reporter under the control of a single randomised response element facilitates both positive growth selection of binding sequences and subsequent quantification of the strength of the selected sequence. A galactose-inducible effector allows discrimination between reporter activation caused by the protein under study and activation due to endogenous factors. The system mimics the physiological gene dosage relationship between transcription factor and target genes *in vivo* by using a low copy effector plasmid and a high copy reporter plasmid, favouring sequence selectivity. The utility of the novel yeast screening system was demonstrated by using it to refine the definition of an optimal recognition element for the c-Myb transcription factor (MRE). We present screening data supporting an extended MRE consensus closely mimicking known strong response elements and where a sequence of 11 nt influences activity. Novel features include a more strict sequence requirement in the second half-site of the MRE where a T-rich sequence is preferred *in vivo*.

INTRODUCTION

Yeast (*Saccharomyces cerevisiae*) has proved to be a highly valuable tool in a range of screening methods. In particular, the yeast two-hybrid technique is a remarkably powerful approach for cloning cDNAs based merely on protein–protein interactions (1,2). Several later variants have also been developed, including one-hybrid systems, reverse two-hybrid systems, two-bait systems and RNA-based and ligand-based three-hybrid systems (3–5). Although less exploited, yeast should in principle also be a useful tool to identify functional *cis*-elements that *in vivo* act as optimal recognition sites for

DNA-binding proteins like transcription factors. Although *in vitro* methods dominate among approaches allowing selection of recognition sequences (6–8), successful examples of *in vivo* screening for recognition elements in yeast have been reported (9,10). We have previously demonstrated that yeast cells have clear advantages in the study of sequence recognition by the transcription factor c-Myb (11,12). In particular, we observed that sequences that are equally well recognised by c-Myb *in vitro* resulted in major differences in Myb response *in vivo* when monitored in a yeast effector–reporter system. The sequence selectivity observed in yeast was highly dependent on the Myb expression level; low expression of Myb protein resulted in a system with a high sequence selectivity (12).

The c-Myb protein is a transcription factor encoded by the *c-myb* proto-oncogene (reviewed in 13–16). A large body of evidence suggests that c-Myb is involved in regulating cell growth and differentiation in haematopoietic cells, a function that is consistent with phenotypes observed in mice with a *c-myb*^{null} mutation and in homozygous null *c-Myb/Rag1* chimeric mice (17,18). As a transcription factor c-Myb exerts its function through recognition of specific response elements in the promoters of its target genes. A highly conserved DNA-binding domain is located in the N-terminal region of the protein and is composed of three imperfect repeats, R₁, R₂ and R₃, each with a W-rich signature motif. The R₂ and R₃ repeats alone are sufficient for sequence-specific DNA binding (19). Several *in vitro* approaches have been used to define the consensus core DNA-binding site for c-Myb, YAACNG (Myb recognition element, MRE). This consensus was first derived from a direct comparison of sequences in DNA fragments bound by Myb (20,21). Subsequently, PCR-based binding site selection methods resulted in minor extensions of the MRE consensus sequence to YAACBGYCR or YAACKGHH (22,23). We have previously shown that the MRE is functionally bipartite. The first half-site [YAAC] is absolutely required for DNA binding, while the second half-site [NGHH] can adopt a range of sequences mainly affecting the half-life of the protein–DNA complex *in vitro* (24). The orientation of the c-Myb–DNA complex is such that the first half-site has the majority of specific contacts to R₃, and the less well-defined second half-site has mainly specific contacts with the R₂ subdomain (24–26). A more detailed analysis of the second half-site revealed a flexible sequence requirement, in particular with respect to configuration of G residues at positions 5 and 6, possibly caused by a flexible structure in R₂ (11).

*To whom correspondence should be addressed. Tel: +47 22 85 73 46; Fax: +47 22 85 44 43; Email: o.s.gabrielsen@biokjemi.uio.no

Table 1. Oligonucleotides used for plasmid constructions

Oligo designation	Sequence (5'→3') [MRE bold, positions 5 and 6 underlined]
HisiF	CCAAGCGATTAAATACTAGTGGCCGGCCTCTAGAAATTCCTGGCATTATCAC
HisLacR	TCGGCAATGCATGCGGATCCATAAGAACACCTTTGGTGGAGGG
1×MRE-GG	GCATTATAACGGTCTTTTAGCGCC
1×MRE-TG	GCATTATAACTGTCTTTTAGCGCC
1×MRE-GT	GCATTATAACGTCTTTTAGCGCC
1×MRE-TT	GCATTATAACTTTCTTTTAGCGCC
3×MRE-GG	GCATTATAACGGTCTTTAACGGTCTTTAACGGTCTTTTAGCGCC
3×MRE-TG	GCATTATAACTGTCTTTAACTGTCTTTAACTGTCTTTTAGCGCC
3×MRE-GT	GCATTATAACGTCTTTAACGTCTTTAACGTCTTTTAGCGCC
3×MRE-TT	GCATTATAACTTTCTTTAACTTTCTTTAACTTTCTTTTAGCGCC
2×MRE[GG+GG]	CTTGGACTAGTCTCGAGATTATAACGGTCTTTAACGGTCTTTTGCCGG
2×MRE[GG+TG]	CTTGGACTAGTCTCGAGATTATAACTGTCTTTAACGGTCTTTTGCCGG
2×MRE[GG+GT]	CTTGGACTAGTCTCGAGATTATAACGTCTTTAACGGTCTTTTGCCGG
2×MRE[GG+TT]	CTTGGACTAGTCTCGAGATTATAACTTTCTTTAACGGTCTTTTGCCGG
2×GG-TT	CTTGGACTAGTCTCGAGATTATAACGGTCTTTAACGGTCTTTTGCCGG
2×GG-AA	CTTGGACTAGTCTCGAGATTATAACGGTCTAAATAACGGTCTTTTGCCGG
2×MRE[GG+NN]	CTTGGACTAGTCTCGAGNNNNNAACNNNNNNNTAACGGTCTTTTGCCGG
iv-TDApr	GGAATTTCTAGAGCCGGCCAAAAGA CCGTTA
Mim primer	GGCGCTAAAAG

One problem in predicting the presence of MREs in promoters of putative target genes is the degenerate nature of the consensus sequence. MREs fitting the consensus sequences, like YAACKG or YAACKGHH, are expected to occur on average every 1024 or 1820 bp, respectively, much too frequent to have strong predictive value. In the present work we have developed a novel yeast system for *in vivo* selection of recognition sequences and applied the system to define a better MRE consensus for c-Myb. We present screening data supporting an extended MRE consensus where a sequence of 11 bases influences activity. The derived consensus closely resembles the classical *mim-1* A-site (27). This novel consensus reflects a more stringent requirement in the second half-site of MRE where a T-rich sequence is preferred *in vivo* and illustrates the utility of the novel yeast screening system.

MATERIALS AND METHODS

Constructs and plasmids

The low copy (centromeric) yeast expression plasmid pDBD₁₁R₂R₃ (11), derived from pDBD₁₁ (28), expresses human/chicken c-Myb R₂R₃ as a VP16 fusion under control of the inducible *GAL1* promoter. The high copy yeast reporter plasmid pYHLfus is derived from pYES2 (Invitrogen). It contains a fusion of two reporter genes, *HIS3* and *lacZ*, and was constructed as follows. The yeast *HIS3* gene together with its own promoter, *minHisi-1*, was amplified and modified by PCR from the plasmid pHisi-1 (Clontech) using the primers HisiF and HisLacR (primer sequences given in Table 1). The product was digested and inserted between the *SwaI* and *SphI*

sites of pYES2. Digestion of pYES2 with *SwaI* and *SphI* removed the *GAL1* promoter from the plasmid. A 3 kb *BamHI* fragment from pMC1871 (Amersham Pharmacia Biotech) containing the *lacZ* gene was ligated in-frame into this construct. The yeast reporter constructs pYHLfus/3×GG, pYHLfus/3×TG, pYHLfus/3×GT, pYHLfus/3×TT, pYHLfus/1×GG, pYHLfus/1×TG, pYHLfus/1×GT and pYHLfus/1×TT were made by inserting double-stranded oligos with either one or three copies of different MREs (see Table 1) into a filled-in *SpeI* site upstream of the *HIS3* minimal promoter. The different oligos were first annealed to a common oligo (mim-primer, see Table 1), filled in with Klenow DNA polymerase I and then ligated into pYHLfus. The yeast reporter constructs pYHLfus/GG-GG, pYHLfus/GG-TG, pYHLfus/GG-GT, pYHLfus/GG-TT, pYHLfus/2×GG-TT, pYHLfus/2×GG-AA and pYHLfus/2×MRE[GG+NN] were made by inserting double-stranded oligos with two copies of different MRE sequences (see Table 1) between the *SpeI* and *FseI* sites upstream of the *HIS3* minimal promoter. The different oligos were first annealed to a common oligo ivTDApr (see Table 1), filled in with Klenow DNA polymerase I, digested with *SpeI* and *FseI* and then ligated into digested pYHLfus.

Yeast strain, growth media and transformation

The yeast strain INVSC1 (MAT α , *his3- Δ 1*, *leu2*, *trp1-289*, *ura3-52*; Invitrogen) was transformed with both effector and reporter plasmids using the one-step method (29). When the cells were transformed with the plasmid library, we used a high efficiency protocol adopted from the 2 Hybrid System TRAF0 Protocol (<http://www.umanitoba.ca/academic/faculties/medicine/biochem/gietz/2HS.html>). Yeast were grown in

complete (YPD) or synthetic (SC) (2% w/v glucose, 0.17% w/v nitrogen base and 0.5% w/v ammonium sulphate) medium supplemented with amino acids as described (30). Transformants were selected on plates with medium selective for the plasmids.

Transactivation assays in yeast

Transactivation assays measuring β -galactosidase (β -Gal) activity in permeabilised cells were performed with yeast INVSC1 transformed with both effector and reporter plasmids in either glucose medium (repressive state for *GALI*-controlled genes) or galactose medium (activating state for *GALI*-controlled genes) (31,32). The β -Gal assays were performed according to the protocol of Kippert (33) as previously described (34). Transactivation assays, measuring growth ability of the transformants, were done on plates with galactose medium selective for the plasmids and lacking histidine.

Site selection

The oligo library (oligo 2 \times MRE[GG+NN], see above) was inserted into pYHLfus, amplified in *Escherichia coli* and then transformed into INVSC1 yeast cells. Clones that grew on galactose medium lacking histidine, but not on control plates with glucose as carbon source, were also tested in a β -Gal activity assay. Plasmids from clones that grew on galactose medium lacking histidine and were β -Gal-positive were rescued (35) and the rescued plasmids were retransformed into INVSC1 cells for verification of the β -Gal activity measured. Inserts of double positive plasmids were sequenced.

Electrophoretic mobility shift assay (EMSA)

In vitro DNA binding was monitored by EMSA as described (36). Recombinant c-Myb-R₂R₃ (residues 88–192 of chicken c-Myb) was expressed and purified as previously described (19). When binding to different EMSA probes was compared, the various MRE oligos (see Table 2) were labelled to identical specific activities as described (11). Oligos were first annealed to an end-labelled common oligo upMIMpr (for the studies of MRE positions 10 and 11) or MimPr-ny (for the studies of the other MRE positions), then filled in with Klenow DNA polymerase I and purified on MicroSpin G-25 columns (Pharmacia Biotech). Unlabelled double-stranded oligos for competition were similarly annealed, filled in and purified on the same columns. Phospho-imaging (GS-250 Molecular Imager; Bio-Rad) was used to quantify the intensities of the bands on the EMSA gel.

RESULTS

Design of a yeast effector–reporter system for selection of optimal binding sites for transcription factors *in vivo*

The high sequence selectivity of the transcription factor c-Myb observed in our previously described yeast effector–reporter system (12) led us to further develop this into a selection system to isolate optimal DNA binding sites *in vivo* for c-Myb. The system is designed to be generally applicable to define binding sites for any non-yeast DNA-binding protein. As illustrated in Figure 1, the basic design of the selection system depends on an effector expressing the transcription factor (here c-Myb) and a reporter monitoring DNA binding through a simple readout. The reporter is driven by a promoter where the

Table 2. Oligonucleotides used for EMSA

Primer designation	Primer sequence 5'→3'
Mim-1A*	GCGCTAAAGACCGTTATAATGC
Mim-1A*-AA	GCGCTATTAGACCGTTATAATGC
UpMIMpr	GCATTATAACGG
MimPr-ny (E025)	CCAGGCGCTAAA
MRE-consens	G GCTCTTAACGGTTTTTTAGCGCCTGG
MRE-cons[A9]	G GCTCTTAACGGTTATTTAGCGCCTGG
MRE-cons[A1]	G GCTCTAAACGGTTTTTTAGCGCCTGG
MRE-cons[A-1]	G GCTCATAACGGTTTTTTAGCGCCTGG
MRE-cons[G-1]	G GCTCGTAACGGTTTTTTAGCGCCTGG
MRE-cons[A-2]	G GCTATTAACGGTTTTTTAGCGCCTGG
MRE-cons[G-2]	G GCTGTTAACGGTTTTTTAGCGCCTGG

response element can be randomised to allow selection for optimal binding sequences for the factor. We wanted the selection system to meet the following three criteria: (i) high sequence selectivity, being able to discriminate between strong and moderately strong binding sites; (ii) reporter gene activation dependent on the transcription factor being studied should be easily distinguished from reporter activation caused by endogenous yeast factors; and (iii) two reporter genes should be under the control of the same single response element.

The first criterion is easily met (at least for c-Myb) by using a centromeric effector plasmid to express the transcription factor, as previously documented for c-Myb (12). The second requirement can be met by using an inducible *GALI* promoter in the effector plasmid. Thereby, activation of a reporter plasmid will be effector-dependent when observed only in galactose medium. When activation is observed in the absence of galactose (in glucose medium), it must be caused by endogenous factors in yeast. Because we wanted to keep the stringency of the system high, we had to keep expression of the effector protein low, and hence we risked obtaining only weak reporter activation. We therefore compensated for this by fusing the DNA-binding domain to a strong constitutive transactivation domain, the VP16 transactivation domain from herpes simplex virus. Hence, in the version of the system analysed here, the low copy centromeric effector expressed a chicken c-Myb R₂R₃–VP16 fusion. We have recently shown that there are only subtle differences in the MRE preferences *in vitro* between c-Myb R₂R₃–VP16 and recombinant c-Myb R₂R₃, as well as between c-Myb R₂R₃–VP16 and full-length c-Myb (12).

The third criterion was motivated by a wish to combine the advantages of the two reporter genes, *HIS3* and *lacZ*, commonly used separately in classical two-hybrid screenings. *HIS3* is ideal for positive growth selection among a large number of clones, while *lacZ* is more amenable to quantification, which was needed to identify the elements giving the highest reporter activation. However, the two reporters should here be driven by a single randomised recognition element to allow a simple two-step screening procedure: first the rapid isolation of candidate clones from growth on medium without histidine, then a simple quantification of the *lacZ* reporter activation in the same clones to evaluate the strength of the

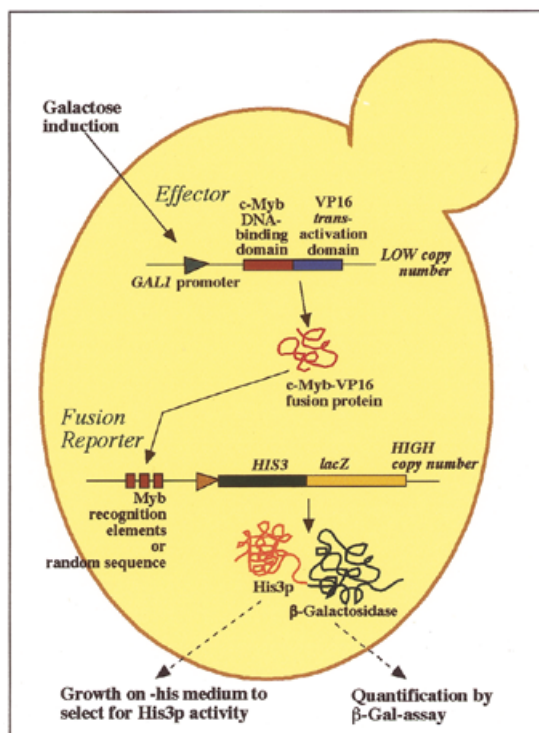


Figure 1. The yeast effector-reporter system for *in vivo* selection of optimal Myb-binding sites. The centromeric effector plasmid pDBD₁₁-R₂R₃ encodes the chicken c-Myb-R₂R₃ DNA-binding domain fused to the VP16 transactivation domain and its expression is under control of the *GAL1* promoter. The high copy reporter plasmid pYHLfus contains a fusion of two reporter genes, *HIS3* and *lacZ*. A single randomised recognition element allows a simple two-step screening procedure as described in the text.

element. To obtain such a dual reporter three different strategies were considered. Insertion of an internal ribosomal entry site (IRES) between the two reporters was discarded because several groups have reported that the IRES is unable to direct translation in *S.cerevisiae* (37,38). We tried a strategy with a bi-directional promoter. A reporter plasmid with *HIS3* and *lacZ* under the control of a common divergent promoter was made, but was found to give unacceptably high levels of Myb-independent reporter activation and hence was discarded (data not shown). A more successful third strategy was to design a fusion between the two reporter genes. A fusion gene combining the two reporters, *HIS3* and *lacZ*, was constructed as described in Materials and Methods and placed downstream of a minimal *HIS3* promoter in a high copy number plasmid (2 μ). Response elements for the effector could then be inserted upstream of the minimal *HIS3* promoter.

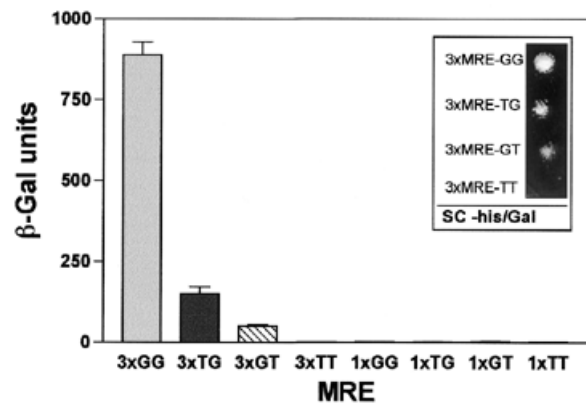
Validation of the system

To validate the system and test its behaviour, triple copies of various MREs (Fig. 2A) were inserted in front of the *HIS3-lacZ* fusion reporter gene. These reporters were combined with the Myb effector plasmid encoding c-Myb R₂R₃-VP16 and introduced into yeast strain INVSC1. In the absence of effector activation (glucose as carbon source), the *HIS3-lacZ* fusion reporter did not promote growth on media lacking histidine (not shown), but did so after induction of c-Myb (galactose as

A

Different MREs used in the validation of the yeast system:															
MRE position	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
MRE-GG	A	T	T	A	T	A	A	C	G	G	T	C	T	T	T
MRE-TG	A	T	T	A	T	A	A	C	T	G	T	C	T	T	T
MRE-GT	A	T	T	A	T	A	A	C	G	T	T	C	T	T	T
MRE-TT	A	T	T	A	T	A	A	C	T	T	T	C	T	T	T

B



C

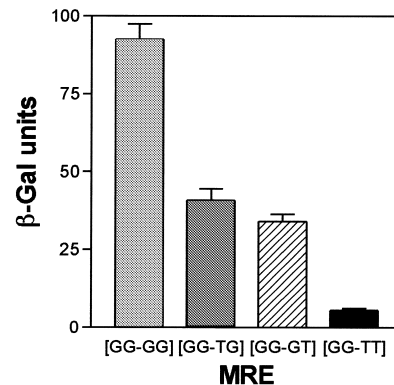


Figure 2. Validation of the yeast effector-reporter system. (A) To evaluate the system, four different MREs with characterised c-Myb responses (11,24) were inserted in front of the *HIS3-lacZ* fusion reporter gene. The MREs are designated MRE-GG, MRE-TG, MRE-GT or MRE-TT depending on the nucleotides in MRE positions 5 and 6. After co-transformation of the yeast strain INVSC1 with the Myb effector and one of the different reporter plasmids, growth on medium lacking histidine and β -Gal activity were measured as described in Materials and Methods. (B) Activation of *lacZ* reporters with single or triple MRE inserts by the chicken c-Myb-R₂R₃-VP16 effector. (Insert) Activation of *HIS3* by chicken c-Myb-R₂R₃-VP16. Aliquots of 10⁴ cells were spotted on SC galactose medium selective for the plasmids and lacking histidine. (C) Activation of *lacZ* reporters with inserts containing one constant MRE and one variable MRE (described in the text). The β -Gal assays were performed on exponentially growing cells as described in Materials and Methods. The data in (B) and (C) are presented as mean β -Gal values \pm SEM for three independent experiments, each carried out in triplicate.

carbon source, Fig. 2B). This behaviour eliminated the need to add the inhibitor 3-amino-1,2,4-triazole (3-AT), which is often necessary to obtain histidine auxotrophy because of the leaky

behaviour of *HIS3* in many commonly used reporters. Possibly, the specific activity of the encoded imidazolglycerol phosphate dehydratase (His3p) fusion protein is lower than that of the wild type so that the protein threshold needed to achieve sufficient His3p activity for growth is higher with the fusion than with normal His3p. Parallel to the Myb-dependent growth on SC –His medium was induction of β -Gal activity (Fig. 2B), thus demonstrating that both partners in the fusion were enzymatically active.

Two additional questions were addressed in this experiment: whether the system revealed the differences in strength known to exist between the MREs employed and whether a single copy of a strong response element would be sufficient for reporter activation. As shown in Figure 2B, comparison of single and triple copies of four already known MREs of variable strength showed that the triple copy reporters behaved as expected with gradual differences in activation, while the single copy constructs harbouring the same MREs did not activate significantly above background. We reasoned that randomisation of three binding sites would complicate the final sequence analysis after screening, so we tried an intermediate strategy of using double MREs, one of which was constant (the strong MRE-GG element) and the other variable. The variable element could then be randomised in the screening. To test this strategy we inserted in the variable position the four MREs used above (MRE-GG, MRE-TG, MRE-GT or MRE-TT) and the reporters were analysed for Myb response. These reporters with two MRE copies gave sufficient Myb-dependent activation to induce growth on medium lacking histidine (not shown) and sufficient β -Gal activation to be quantified and ranked in the predicted order (Fig. 2C). This design of the reporter was therefore chosen for screening purposes.

Screening

Having then achieved a design of the system that seemed to behave correctly when tested on known MREs, we were ready to use the system in a screening experiment to select optimal binding sites for the minimal DNA-binding domain of c-Myb (effector, c-Myb R₂R₃-VP16). A partially randomised oligo library was inserted in the variable position in front of the *HIS3-lacZ* fusion reporter gene (Fig. 3A). The inserted oligo contained 12 randomised nucleotides surrounding a fixed AAC core sequence. This core has been shown to be necessary for binding of c-Myb to DNA (24,26,39). Due to this fixed AAC core, we expected a high percentage of positive clones. The plasmid library was amplified in *E.coli*, giving 75 000 independent clones. This number of clones did not represent a full coverage of the complexity of the randomised region. However, due to the presence of a constant AAC core in the middle of the randomised insert a high number of positive clones was anticipated, and screening from 75 000 sequences would give enough data to refine the optimal flanking region of the AAC core. The library was transformed into yeast together with the effector plasmid. Out of 2000 independent yeast clones picked, 178 grew on selective SC –His medium containing galactose, as illustrated in Figure 3B. Since 58 of these also grew on glucose SC –His medium, only 120 clones required c-Myb R₂R₃-VP16 to activate the *His3-lacZ* dual reporter for growth. When β -Gal activity was measured in the same collection, 89 clones were found to activate *lacZ* in a Myb-dependent manner [five did not activate and 26 clones

activated the *lacZ* reporter in glucose (Myb-independent)]. Plasmids from the 89 remaining selected clones were rescued and retransformed (twice) into yeast and β -Gal activities were measured in the retransformed cells to verify that the β -Gal levels obtained were reproducible. The rescued plasmids were then sequenced, after which 20 more clones were discarded due to the presence of more than one insert. The sequences of the remaining 69 single insert clones are shown in Table 3. The collection should contain only functional MREs and be free of sequences corresponding to 'background noise'. More importantly, the selected MREs resulted in different β -Gal activities, reflecting different binding strengths, and could be arranged in groups according to these values. Based on their β -Gal activities obtained in transformed yeast cells, the 69 clones selected were divided into six different groups, as illustrated in Figure 3C. The four model MREs used above (Fig. 2A) fall into distinct groups, MRE-TT in group 1, MRE-GT in group 2, MRE-TG in group 3 and the strong MRE-GG in group 6.

Derivation of an *in vivo* consensus sequence for MRE

The particular feature of the collection of sequences generated by the present method is that a value is linked to each sequence reflecting how efficient the Myb induction is with that particular sequence in our yeast system. It was therefore important to exploit these values in the generation of an *in vivo* consensus sequence for c-Myb. We examined different approaches to incorporate transactivation values (see Discussion), but the results were very similar and we present in Table 4 the derivation of a consensus sequence from weighted sequences. Skipping group 1 sequences (the weakest) as well as counting a few identical sequences only once, the remaining 46 different single Myb-binding sites were aligned according to the invariant AAC core. Nucleotides at each position in all the selected sequences were assigned a weight (1–5) depending on the β -Gal group to which the sequence belonged (group 2, weight 1; group 3, weight 2; etc.). Based on the resulting matrix and the rules given in the legend to Table 4, we deduced the following *in vivo* MRE consensus sequence:

–4	+1	56	7	11
NNCN	TAACGG		TTTTT	

It is reassuring to note that this consensus has an almost exact match (except for a C→T difference at –2) to the strong A-site found in the *mim-1* gene promoter, one of the best established c-Myb target genes (40). In addition, the following features were apparent.

The hexamer core of the MRE consensus sequence. We observed a clear preference for pyrimidines, especially T residues, in MRE position 1, consistent with previous conclusions from *in vitro* data (22,23). Additional nucleotide preferences were also observed in MRE positions 5 and 6, with a double G as the preferred variant. These two positions are not independent of each other (11). When dinucleotide configurations in these two positions were counted, the following frequencies were found GG (14) > CG (8) > GT (5) > TG (4) > CC (3). All other combinations occurred less than twice.

Nucleotides in the 5'-end of the MRE consensus sequence. From the consensus sequence c-Myb does not seem to have very strong nucleotide preferences 5' for the hexamer core,

Table 3. Sequences of the selected binding sites

Clone no.	Insert sequence	β -Gal units	Clone no.	Insert sequence	β -Gal units
139	CC TGC AAC CGG TTT C	6.4	66	CG CTA AAC CGT TTC T	42
2	GG TTC AAC GTT TGT T	8.9	124	GT TGT AAC TGA TCG T	43
13	GT TCT AAC TTT TTC T	10	31	CT GTT AAC GGT TGG T	44
16	TG CGC AAC TTC TCT G	12	146	TT CGA AAC CGT TTT C	44
102	TG ATT AAC CGG CCC G	12	64	TT TCT AAC GGC CCC T	46
117	TT TCC AAC GCC TGT T	12	135	AT TAT AAC TGT CTT T	46
147	TC GGT AAC GTG TGC G	12	159	AT CTC AAC GGT TCT T	49
149	TC GGA AAC TCG CGG G	12	167	TT TCT AAC GGT CGG C	50
56	CC TCC AAC GAA TGT T	14	107	CT CAT AAC GGC GGT T	51
140	CC TTG AAC CGG TTT T	15	119	CT TTT AAC CGT TTC C	51
173	GC CCT AAC CCG GCT T	18	131	CG TTT AAC CTT CCG T	52
88	CC GTT AAC TCT GTT T	20	152	TT CGA AAC CGT TTT C	54
166	GT CTT AAC GTC CGT G	20	78	CG CCG AAC TGT TTT T	55
122	GA CGA AAC GGG GGT G	21	115	AT TAT ACC TGT CTT T	55
160	GT CTT AAC GTG CGT G	21	34	CG AGT AAC GCT TTT T	57
63	GG GCT AAC ACG TTT T	23	60	TC CTT AAC GGT TCT T	62
163	TT CGA AAC GCC TTT T	23	36	AT TAT AAC TGT CTT T	63
67	AT CCG AAC CTC GGC G	24	100	GC CAG AAC GTG CCT T	63
87	CT ACC AAC CCG TTT T	24	101	CT CAT AAC GGC GGT T	64
17	TC GCC AAC CCC TTG T	27	142	AT TAT AAC TGT CTT T	64
161	GA CGC AAC CCG TTT T	28	176	TC TGC AAC GGT TTT C	64
58	TG TTT AAC GGT CCA G	32	104	AG CAC AAC GAG ATT G	85
93	CA TGC AAC TTT TTT T	32	80	AG CGT AAC GGC AAT T	90
128	TC GGC AAC GTC CGT T	34	157	AG CCC AAC GGT TAT A	94
137	TC GTT AAC TTG CCT A	34	165	AG CCC AAC GGT TAT A	98
113	CT TTT AAC CGT TTC C	36	132	GC GCT AAC CGT TAC T	101
129	TC CGT AAC CGT TCG T	38	110	TC GTT AAC GAC CGT T	102
99	GG GGT AAC CAT CGC G	39	89	GT CCT AAC GGC TTT T	124
21	AT TAT AAC GTT CTT T	40	8	GT CGT AAC GGC TTT G	125
72	CT TGT AAC GTT TGT T	40	83	AA ATT AAC CGT TAA G	130
82	GC TTC AAC CGG TAA T	40	114	CC CAG AAC TGG CGG G	133
109	AT TAT AAC TGT CTT T	40	20	CT TTT AAC GTT GTG T	148
130	GT TGT AAC GGA TCG T	40	3	GA TCC AAC ATC GCT T	170
125	CG TTT AAC CTT CCG T	41	50	CC GTT AAC GTT TTT G	173
			55	CC GTT AAC GTT TTT G	190

The 69 sequences from the screening are arranged according to increasing β -Gal activities obtained in transformed yeast cells.

positions -4 to -1, except for a reduced frequency of A residues and an enhanced preference for C or Y at position -2.

Nucleotides in the 3'-end of the MRE consensus sequence. A clear nucleotide preference was observed downstream of the MRE hexamer core. Here thymidine was the preferred nucleotide at all the MRE positions from 7 to 11. The lowest frequency

observed was at position 9 and the highest at positions 10 and 11. This preference for a T-rich flanking sequence is a novel feature, since no such preference was apparent in previous site selection studies performed with *in vitro* methods (22,23). In addition, the non-random distribution of bases down to position 11 indicated that the c-Myb-binding site could be more extended than previously noted. In the following sections we

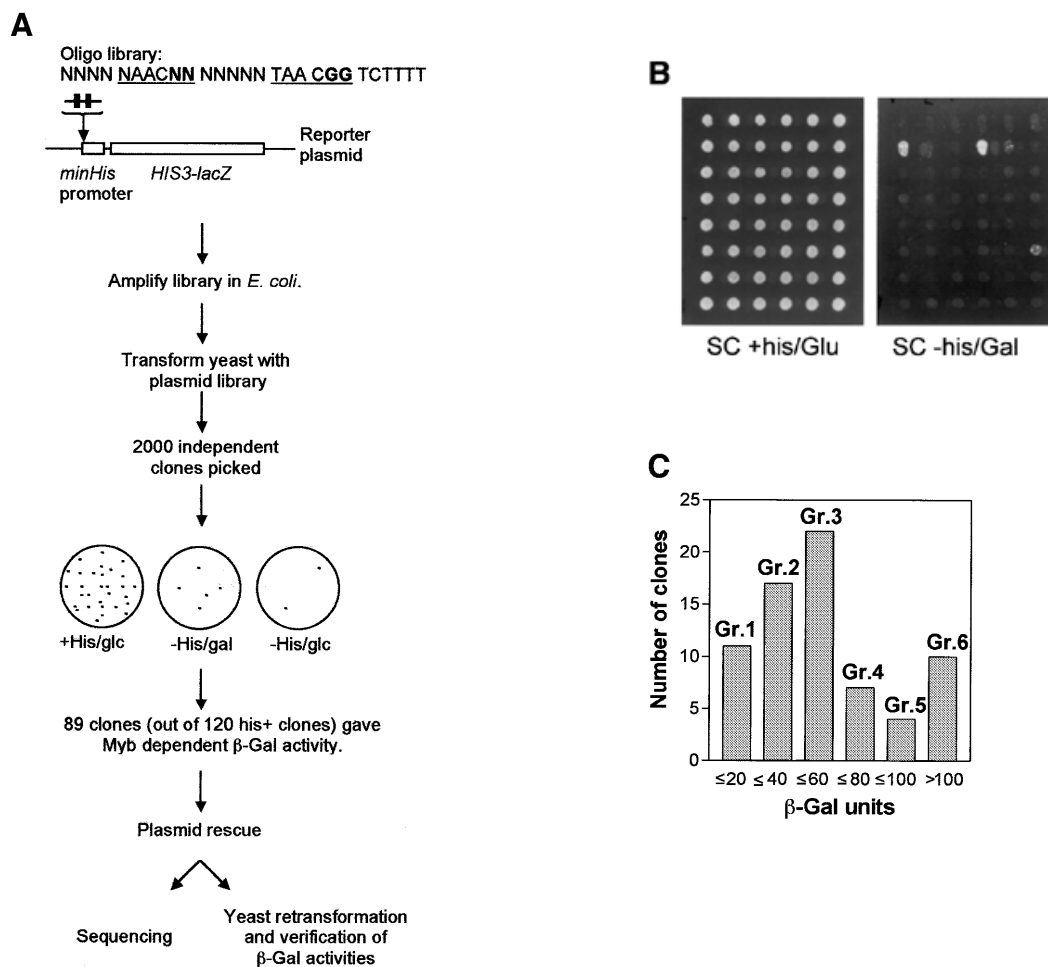


Figure 3. Screening for optimal MREs *in vivo*. (A) A schematic view of the screening (see text for details). (B) Clones with a reporter plasmid bearing Myb-binding sites were selected on galactose medium without histidine (SC -His/Gal). An example from 48 independent clones (growing on medium \pm histidine) is shown. (C) The β -Gal activities for growth-selected clones were measured as described in the legend to Figure 2 and in Materials and Methods. Based on their β -Gal activity, the clones were divided into six different groups as depicted.

Table 4. A weighted matrix to determine the *in vivo* consensus sequence for c-Myb

	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	18	10	7	14	7	100	100	0	4	8	4	7	15	7	4
C	32	30	42	24	23	0	0	100	25	9	25	23	24	13	11
G	27	22	18	28	10	0	0	0	58	62	17	12	20	20	21
T	23	38	33	34	61	0	0	0	13	22	53	59	41	61	64
	N	N	C	N	T	A	A	C	G	G	T	T	T	T	T

To derive a consensus MRE, a selection of the sequences shown in Table 3 was used. The weakest group 1 sequences (β -Gal units < 20) were skipped and a few identical sequences were counted only once. The remaining 46 different single Myb-binding sites were aligned according to the invariant AAC core. After alignment nucleotides at each position in all the selected sequences were assigned a weight (1–5) depending on the β -Gal group to which the sequence belonged: group 2 ($20 < \beta$ -Gal < 40), weight 1; group 3 ($40 < \beta$ -Gal < 60), weight 2; etc. Then for each position and nucleotide the weights were added together and normalised (scale 0–100). The deduced consensus sequence for c-Myb is shown in the last row marked in bold and was derived using the following rules: (i) every nucleotide that obtains a high score (≥ 60) is specified directly; (ii) when a nucleotide obtains a moderately high score (≥ 40 and < 60), the second ranked nucleotide is also considered, to judge whether it also is preferred over the remaining nucleotides. When the score of the second ranked nucleotide is larger than twice the score of the third ranked nucleotide, we also consider the second ranked nucleotide to be a preferred one and this is indicated in the consensus by specification of both the first and second nucleotides. When the score of the second ranked nucleotide does not show this preference, only the first ranked nucleotide is specified.

present further analysis of the effect of changes at positions 10 and 11. In addition we analysed how deviations from the deduced consensus affected protein–DNA complex stabilities.

Extension of the MRE analysed by alterations at positions 10 and 11

The overall consensus presented here is longer than the previously proposed consensus binding sequences for c-Myb. To test the validity of an extended region with nucleotide preferences, the importance of T residues present at MRE positions 10 and 11 was assayed by EMSA and in transactivation assays in yeast. In these tests, we compared a binding sequence closely related to the *mim-1* A-element (24), having or not a double T at positions 10 and 11 (Fig. 4A). EMSA studies of DNA binding showed that both efficiently bound c-Myb-R₂R₃ (data not shown). But when the time course of complex dissociation under competition with an unlabelled specific oligo was monitored, a distinct difference in complex stability between the Mim-1A* and the Mim-1A*-AA probes was observed (Fig. 4B). The same two variants were then tested for Myb-dependent transactivation in yeast after insertion into the same reporter as used for screening. As shown in Figure 4C, transactivation obtained with the reporter harbouring the 2×GG-AA variant was consistently 40–50% lower than that obtained with the 2×GG-TT reporter, showing that these residues in fact contributed to the strength of the MRE sequence. We therefore conclude from both *in vivo* and *in vitro* experiments that the MRE preference extends to position 11 and that T residues are the preferred nucleotides at MRE positions 10 and 11. This preference is not a simple AT effect since the introduced change was from TT to AA.

Analysis of systematic deviations from the derived consensus

In addition to the analysis of positions 10 and 11 above and previous studies of positions 5 and 6 (11), we also investigated effects of specific bases in other positions to see if they might affect the Myb–DNA complex. Specifically, we wanted to confirm that nucleotides in the 5'-end of the MRE consensus sequence, where the consensus is less specific, did not play a role in determining complex stability. In addition, since many selected sequences had GTTR elements (antiparallel YAAC), we wanted to see whether the creation of a pseudo-palindromic sequence generated by T9→A was favourable for complex formation. To simplify the analysis we compared the sequences by analysing decay rates upon competition in EMSA, an approach that both earlier (11,41) and in the present work (Fig. 4B and C) was found to nicely correlate with activities *in vivo*. Hence, from a reference sequence fitting the overall consensus (GGCTCTTAACGGTTTTT-TAGCGCCTGG, consensus underlined, core hexamer in bold), we systematically changed several positions (–2, –1, +1 and +9). As illustrated in Figure 5, it appeared that destabilisation occurred in all cases where the sequence deviated from the consensus, while changes that were neutral relative to the overall consensus had no effect. Thus changes at position –1 had no significant effects (T–1→A and T–1→G), while a change from the consensus at position –2 (C–2→G) had a detectable effect (decay rate doubled). A marked effect was observed when position +1 was altered (T1→A, 7-fold increase in decay rate). This is consistent with the overall

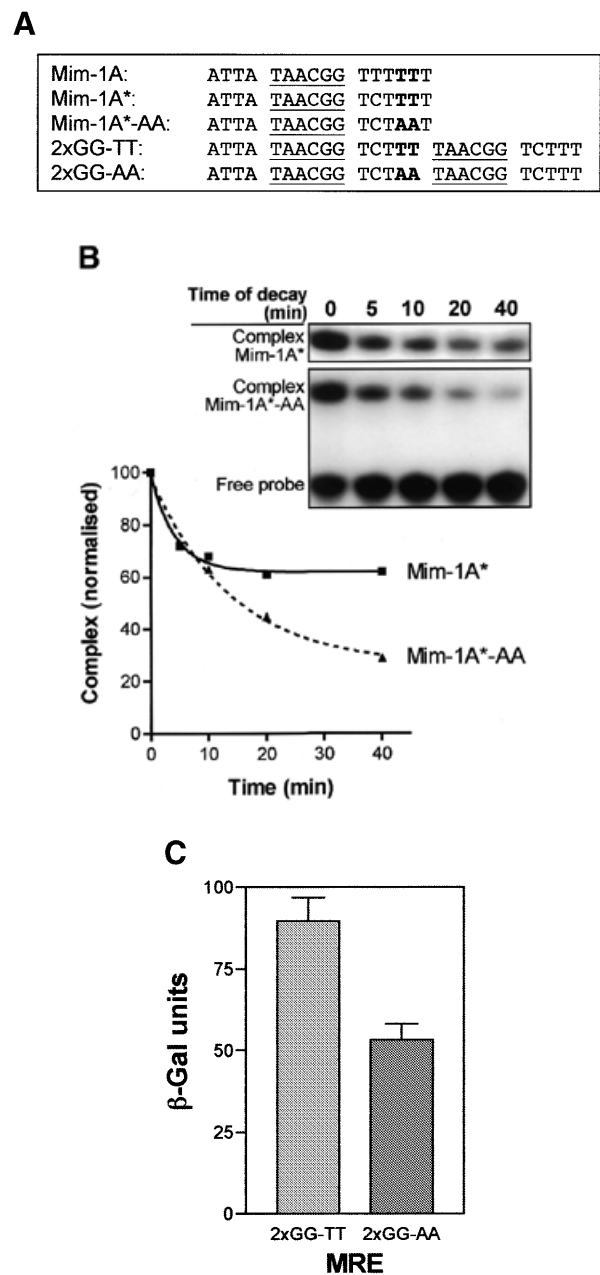


Figure 4. Studies of the nucleotides at MRE positions 10 and 11. To determine the importance of the nucleotides at MRE positions 10 and 11 for Myb DNA binding, both *in vitro* (EMSA) and *in vivo* (effector–reporter assays in yeast) experiments were performed. (A) The design of the oligos for these studies was based on the *mim-1* A-site in the *mim-1* promoter (40). The oligos for the EMSA experiments, Mim-1A* and Mim-1A*-AA, contain a C instead of a T in MRE position 8 to prevent the creation of a second MRE in the antiparallel direction. The 2×GG-TT and 2×GG-AA oligos were designed in the same way as the library clones (see Fig. 3A) and were used in the yeast experiments. (B) An EMSA gel showing a time course of decay of the c-Myb-R₂R₃–DNA complex upon competition with excess non-labelled probe. Myb–DNA complexes were generated using 10 fmol c-Myb-R₂R₃ and 10 fmol MRE probe. The c-Myb-R₂R₃–DNA complexes were allowed to form for 15 min at 25°C before they were exposed to a 75-fold excess of non-labelled specific probe for 0, 5, 10, 20 and 40 min. The densities of the complexes were determined, normalised and graphically displayed. (C) Two oligos, 2×GG-TT and 2×GG-AA, were inserted in the pYHLfus reporter plasmid to test Myb-dependent transactivation in yeast. A β-Gal activity assay was performed as described in the legend to Figure 2 and in Materials and Methods. The data are presented as mean β-Gal values ± SEM of three independent experiments, each carried out in triplicate.

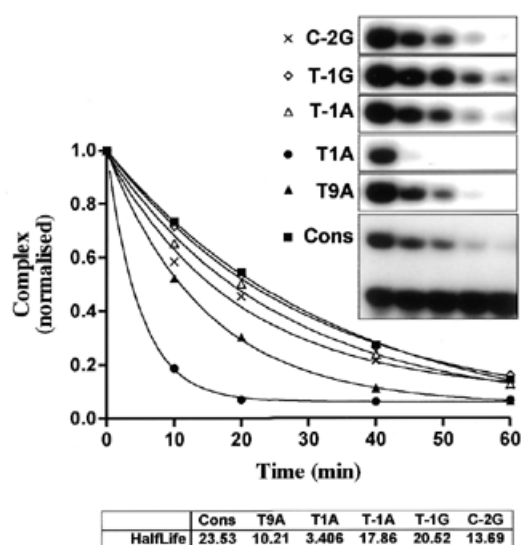


Figure 5. Myb-DNA complex dissociation: comparative analysis of variant MREs. Time course of complex dissociation upon competition. The oligos are designated T9A, T1A, T-1A, T-1G and C-2G. They all contain a single base pair deviation from the derived consensus sequence, with the first letter indicating the base specified in the consensus, the number giving the position and the last letter indicating the novel base introduced. Complexes of c-Myb and DNA were generated using 20 fmol c-Myb-R₂R₃ and 10 fmol different MRE probes (Consensus, T9A, T1A, T-1A, T-1G and C-2G) and allowed to form for 15 min at 25°C with 0.1 µg poly(dI-dC) present. The complexes were then exposed to 750 fmol unlabeled MRE-mim-1A probe for 0, 10, 20, 40 and 60 min before samples were analysed by EMSA as described in Materials and Methods. The densities of the complexes were determined using a phosphorimager, normalised and graphically displayed. The half-lives of the complexes were determined after fitting the points by non-linear regression to a one-phase exponential decay curve.

consensus being optimal for binding. It is noteworthy that the T9→A variant did not form more stable complexes than the reference sequence *in vitro*. Rather, it destabilised the complex (decay rate doubled), consistent with its deviation from the consensus.

DISCUSSION

We present in this work the design and use of a novel yeast system for *in vivo* selection of sequences recognised by DNA-binding proteins. The DNA-binding protein is expressed from an inducible *GAL1* promoter making it easy to distinguish activation caused by the factor under study and activation due to endogenous factors. The main novel feature of the effector-reporter system is the use of a dual *His3-lacZ* fusion reporter under the control of a single set of response elements. This design facilitates both positive growth selection of binding sequences ideal for screening and subsequent quantification of the strength of the selected sequences. The latter generates a transactivation value linked to each of the selected sequences, values that can be used in the derivation of a consensus binding sequence. It is also important that we have tried to mimic the gene dosage relationship between transcription factor and target genes *in vivo* by using a low copy effector plasmid and a high copy reporter plasmid. The normal gene dosage relationship is one where the transcription factor is encoded from a

gene present in one or two copies, and the gene product will interact with a large number of recognition sites in diverse target genes. Previous studies of c-Myb have shown that keeping this normal gene dosage relationship improves sequence selectivity compared to experiments where c-Myb is overexpressed from a high copy plasmid (12). This aspect is, however, not an essential feature of the novel selection system. For specific purposes it could be as appropriate to use high copy number effector plasmids, as discussed below. Although the system has been tested on the c-Myb transcription factor, it should be easy to adapt it to any sequence-specific DNA-binding protein as far as it binds a sequence distinct from endogenous yeast transcription factors.

Few reports have been published where yeast has been exploited to screen for *cis*-elements. To our knowledge the first successful example was the work of Wilson and co-workers used to identify the DNA-binding site for NGFI-B protein (9). Compared to their design, the present system is distinguished by the use of a dual reporter that also allows quantification, simple galactose induction to eliminate sequences responsive to endogenous factors and randomised inserts rather than genomic fragments.

The relevance of using a *S.cerevisiae* transactivation system for selecting protein-binding sites requires that the DNA-binding protein under study has very similar behaviour in yeast and mammalian cells. Even if this is often questionable with respect to interaction with cooperating factors and coactivators, the DNA-binding properties are usually intrinsic properties of the factor where relevant behaviour is expected in yeast. Hence, it is more reasonable to anticipate differences in the transcriptional output of a specific factor between yeast and mammalian cells than in DNA binding specificity *per se*. By adding a strong transactivation domain, like VP16 used in the present work, such transactivation differences are easily concealed. The present system is best regarded as a supplement to *in vitro* methods that until now have dominated when a protein-binding site is to be defined. The evident advantage of our system is that selection is performed *in vivo* and thus takes place under biologically relevant conditions. A yeast cell will in several ways mimic a mammalian cell better than any *in vitro* experiment: with regard to the availability of binding sites packed in chromatin, the concentrations of DNA-binding proteins, the competitive situation with many factors interacting with a multitude of sites, the general milieu of ionic composition and pH, etc. In addition, a practical advantage is that yeast is a simple model organism upon which several screening methods are already based.

The most obvious limitation of this novel method is that it cannot be applied to DNA-binding proteins having the same specificity as an endogenous yeast transactivator. In the case of c-Myb this was no problem since a Myb-responsive reporter shows only background level activation in yeast in the absence of co-expressed c-Myb. A second limitation, less expected, was that at least two binding sites were required to obtain sufficient transactivation to surpass the growth threshold when c-Myb was expressed from a centromeric vector. For this reason, screening had to be performed with one fixed and one randomised binding site. In general, the threshold for growth set by the *His3-lacZ* reporter has to be overcome through a correct balance between several factors, like the strength of DNA binding, the concentration of the DNA-binding protein,

the number of sites in the promoter and the strength of the transactivation domain. We expect the sequence selectivity to be highest when this threshold is just surpassed, but that the conditions for this may vary with each individual protein studied. Hence, we expect that some proteins with modest binding affinities may require high copy effectors to achieve proper growth response. This may also be the case when genomic fragments, rather than randomised oligos, are used for screening.

However, we realised that the selection *in vivo* also introduced additional unexpected complexities. First, the selection for multiplicity of binding sites was very strong. Inspection of the sequence collection (Table 3) revealed an unforeseen high frequency of GTT sequences (antiparallel AAC, the MRE core sequence), suggesting that multiple MREs had been selected for within the small randomised region. About 60% of the selected sequences contained one or more extra GTT sequences, compared to only ~20% among the reported *in vitro* selected sequences using PCR-based methods (21,22). We do not interpret this as multiplicity being an inherent feature of an optimal MRE. We rather explain this as an unintended consequence of the selective pressure imposed by the limited randomised region. It is known from genome-wide analyses of binding sites that multiplicity can be a very important selection factor *in vivo* to discriminate binding sites in promoters from irrelevant sites (42). In the present system, sequences that happen to incorporate more than one binding site will cause higher activation and therefore be selected. However, this phenomenon of multiple MREs complicated the derivation of a consensus sequence, since only one of the multiple MREs would in each case be properly aligned and it was not evident which one was the most important for activity, nor how much each contributed to the overall activity of the sequence. Attempts to simply correlate the number of GTT sequences (or potential MREs) with activity did not reveal any obvious correlation. We tried several approaches to exploit the activity data and avoid the multiplicity problem in derivation of the consensus sequence, but with only minor differences in the final result. One approach was to use all the putative MREs as independent entries (79 MRE entries from 46 selected sequences), reasoning that the best MREs would have the highest frequencies and thus be visible above a background generated by multiple MREs. Even when this was applied to different activity sets, the derived consensus was almost indistinguishable from the one presented. We think that increased 'noise' from this multiplicity of putative binding sites is an inherent feature of the *in vivo* screening approach that could only be overcome by analysis of many sequences and defining rules by which enrichment over a specified level is used to define preferred nucleotides. Using this approach, we obtained a consensus sequence that according to our analysis of deviating variants clearly represents one of the best sequences with respect to complex stability.

The screening method presented could be adopted to answer two different types of binding site problems. First, the method could be useful in refining the sequence requirement for DNA-binding proteins in cases where some knowledge of core sequences already exists. This is the type of problem we have addressed for c-Myb in this work. Since many families of transcription factors have common core sequences, like the E-box for the helix-loop-helix proteins, this use should be widely

applicable. The more demanding task is to use the system to define totally unknown binding sequences for factors that are expected to be sequence-specific DNA-binding proteins. The complexity of the randomised promoter will differ in the two cases. In the first case, the core sequence can be kept fixed and only the flanking sequences need to be randomised, thus creating a modest complexity and the need to screen only a limited number of clones. In the second case, one has to use fully randomised elements and the number of clones screened will be larger. With a protein binding to a site n bp long, the frequency of hitting a single site in a randomised sequence of N bp will be $2(N - n)/4^n$. One specific hexamer in a 25 bp randomised region would be found in 1% of the sequences and two hexamers in ~0.01% of the clones. To obtain a collection of 50 sequences would in theory require screening of 5000 or 500 000 clones in the two cases, respectively. Even if the real number had to be doubled due to endogenous activation in a significant fraction of the clones, the numbers are well within the range normally screened in two-hybrid experiments and do not represent any technical limitation of our method. In cases of fully random screening it might, however, be more difficult to find the correct balance of variables to just surpass the threshold for growth, as discussed above. It might be useful in those cases to start with a high copy version of the effector and a randomised promoter large enough to encompass more than one binding site and then, if necessary, increase the stringency after having obtained sequence information from the primary screen.

In the present work, the novel screening method was used to refine the consensus sequence for an optimal MRE. The sequence obtained closely mimics the *mim-1* A-site in the *mim1* gene promoter, one of the best established c-Myb target genes (40). A matching MRE is also found in the *myc* promoter, another c-Myb target gene (43). This lends trustworthiness to the method. We are also confident that the deduced consensus represents an optimal MRE since the mutant analysis revealed that deviations from the derived consensus sequence created protein-DNA complexes with lower stability, while neutral mutations had no effect. Compared to previous MRE consensus sequences derived from classical *in vitro* amplification approaches, the present consensus contains both expected similarities and distinct features. The 5'-flanking sequences of the hexamer core sequence YAACGG seem to be of minor importance. The sequence of the hexamer core itself supports previous conclusions that a double G at positions 5 and 6 creates a particularly strong site. We have previously argued for a flexible sequence recognition in these positions, proposing that c-Myb requires a G at *either* position 5 or 6 to bind efficiently, with a double GG as the optimal configuration (11). The present data are consistent with this model. Of the selected sequences, 76% contained a G at position 5 or 6 or both. Moreover, most of the sequences lacking a G at both positions 5 and 6 had an alternative putative MRE in the antisense direction that contained one or two G residues in these positions. Interestingly, 37% of the selected sequences contained a C at position 5 or 6 or both, with CG being the dominant combination. One might speculate whether a flexible recognition of G residues in this region of the MRE also extends to the opposite strand, implying that CG could be almost as favourable as GG at positions 5 and 6. This could explain why CG showed up as the second most

frequent dinucleotide here. A combination of a G and a T at these two positions gives a reasonable activation of the reporter using a GG-TG or a GG-GT reporter, but occurred in only 20% of the selected sequences, with GT and TG being equally frequent.

The most novel feature of the present consensus MRE is in the 3'-flanking sequence, where a clear sequence preference was observed and where the extent of nucleotides affecting binding was expanded. In this region, expected to be contacted by the second repeat of c-Myb, we observed a strong preference for T residues. In fact, a stretch of five T residues was the preferred configuration, implying a more extended binding site than previously delineated and resembling the T-stretch found in this region in the *mim-1* A-site (40). Previous reports have underscored the absence of G residues downstream of the MRE hexamer core, in particular at positions 7 and 8 (23). Our data also suggest under-representation of purines in this region, but the lowest frequency was of A rather than of G residues. This is noteworthy, because it also suggests that it is not simply a preference for AT-rich DNA that is critical in this part of the MRE. The functional importance of this configuration in the 3'-flanking region was supported by both decay studies of complexes with MRE variants and by *in vivo* studies in yeast. Recently, Ganter and Lipsick reported an *in vitro* selection study where the same region in the MRE differed from previous reports (44). Their finding of a favourable effect *in vitro* of a T-stretch 3' of the hexamer core is consistent with the present results, while their inference that this stretch could be present on either strand clearly contradicts our observations. Curiously, in their study the T-stretch had to be on the same strand as deduced in the present work to enhance transcriptional activation in transiently transfected QT6 cells.

ACKNOWLEDGEMENTS

We thank Oddmund Nordgård and Tor Øyvind Andersen for skilful assistance in parts of the work. This work was supported by The Norwegian Research Council, The Norwegian Cancer Society and the Anders Jahres Foundation.

REFERENCES

- Brent, R. and Finley, R.L.Jr (1997) Understanding gene and allele function with two-hybrid methods. *Annu. Rev. Genet.*, **31**, 663–704.
- Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340**, 245–246.
- Drees, B.L. (1999) Progress and variations in two-hybrid and three-hybrid technologies. *Curr. Opin. Chem. Biol.*, **3**, 64–70.
- Vidal, M. and Legrain, P. (1999) Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.*, **27**, 919–929.
- Mendelsohn, A.R. and Brent, R. (1999) Protein interaction methods—toward an endgame. *Science*, **284**, 1948–1950.
- Blackwell, T.K. and Weintraub, H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science*, **250**, 1104–1110.
- Blackwell, T.K. (1995) Selection of protein binding sites from random nucleic acid sequences. *Methods Enzymol.*, **254**, 604–618.
- Thiesen, H.J. and Bach, C. (1990) Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res.*, **18**, 3203–3209.
- Wilson, T.E., Fahrner, T.J., Johnston, M. and Milbrandt, J. (1991) Identification of the DNA binding site for NGFI-B by genetic selection in yeast. *Science*, **252**, 1296–1300.
- Lei, X.H., Shen, X., Xu, X.Q. and Bernstein, H.S. (2000) Human Cdc5, a regulator of mitotic entry, can act as a site-specific DNA binding protein. *J. Cell Sci.*, **113**, 4523–4531.
- Ording, E., Bergholtz, S., Brendeford, E.M., Jamin, N. and Gabrielsen, O.S. (1996) Flexibility in the second half-site sequence recognised by the c-Myb R2 domain—in vitro and in vivo analysis. *Oncogene*, **13**, 1043–1051.
- Andersson, K.B., Berge, T., Matre, V. and Gabrielsen, O.S. (1999) Sequence selectivity of c-Myb in vivo. Resolution of a DNA target specificity paradox. *J. Biol. Chem.*, **274**, 21986–21994.
- Graf, T. (1998) Leukemogenesis: small differences in Myb have large effects. *Curr. Biol.*, **8**, R353–R355.
- Weston, K. (1999) Reassessing the role of C-MYB in tumorigenesis. *Oncogene*, **18**, 3034–3038.
- Oh, I.H. and Reddy, E.P. (1999) The myb gene family in cell growth, differentiation and apoptosis. *Oncogene*, **18**, 3017–3033.
- Ganter, B. and Lipsick, J.S. (1999) Myb and oncogenesis. *Adv. Cancer Res.*, **76**, 21–60.
- Mucenski, M.L., McLain, K., Kier, A.B., Swerdlow, S.H., Schreiner, C.M., Miller, T.A., Pietryga, D.W., Scott, W.J.Jr and Potter, S.S. (1991) A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell*, **65**, 677–689.
- Allen, R.D., Bender, T.P. and Siu, G. (1999) c-Myb is essential for early T cell development. *Genes Dev.*, **13**, 1073–1078.
- Gabrielsen, O.S., Sentenac, A. and Fromageot, P. (1991) Specific DNA binding by c-Myb: evidence for a double helix-turn-helix-related motif. *Science*, **253**, 1140–1143.
- Biedenapp, H., Borgmeyer, U., Sippel, A.E. and Klempnauer, K.H. (1988) Viral myb oncogene encodes a sequence-specific DNA-binding activity. *Nature*, **335**, 835–837.
- Nakagoshi, H., Nagase, T., Kanai, I., Ueno, Y. and Ishii, S. (1990) Binding of the c-myb proto-oncogene product to the simian virus 40 enhancer stimulates transcription. *J. Biol. Chem.*, **265**, 3479–3483.
- Howe, K.M. and Watson, R.J. (1991) Nucleotide preferences in sequence-specific recognition of DNA by c-myb protein. *Nucleic Acids Res.*, **19**, 3913–3919.
- Weston, K. (1992) Extension of the DNA binding consensus of the chicken c-Myb and v-Myb proteins. *Nucleic Acids Res.*, **20**, 3043–3049.
- Ording, E., Kvavik, W., Bostad, A. and Gabrielsen, O.S. (1994) Two functionally distinct half sites in the DNA-recognition sequence of the Myb oncoprotein. *Eur. J. Biochem.*, **222**, 113–120.
- Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, **79**, 639–648.
- Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S. and Sarai, A. (1993) Recognition of specific DNA sequences by the c-myb protooncogene product: role of three repeat units in the DNA-binding domain. *Proc. Natl Acad. Sci. USA*, **90**, 9320–9324.
- Ness, S.A. (1996) The Myb oncoprotein: regulating a regulator. *Biochim. Biophys. Acta*, **1288**, F123–F139.
- Bonner, J.J. (1991) Vectors for the expression and analysis of DNA-binding proteins in yeast. *Gene*, **104**, 113–118.
- Chen, D.C., Yang, B.C. and Kuo, T.T. (1992) One-step transformation of yeast in stationary phase. *Curr. Genet.*, **21**, 83–84.
- Sherman, F., Fink, G.R. and Hicks, J.B. (1986) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schneider, J.C. and Guarente, L. (1991) Vectors for expression of cloned genes in yeast: regulation, overproduction and underproduction. *Methods Enzymol.*, **194**, 373–388.
- Romanos, M.A., Scorer, C.A. and Clare, J.J. (1992) Foreign gene expression in yeast: a review. *Yeast*, **8**, 423–488.
- Kippert, F. (1995) A rapid permeabilization procedure for accurate quantitative determination of beta-galactosidase activity in yeast cells. *FEMS Microbiol. Lett.*, **128**, 201–206.
- Pinson, B., Sagot, I., Borne, F., Gabrielsen, O.S. and Daignan-Fornier, B. (1998) Mutations in the yeast Myb-like protein Bas1p resulting in discrimination between promoters *in vivo* but not *in vitro*. *Nucleic Acids Res.*, **26**, 3977–3985.
- Robzyk, K. and Kassir, Y. (1992) A simple and highly efficient procedure for rescuing autonomous plasmids from yeast. *Nucleic Acids Res.*, **20**, 3790.
- Gabrielsen, O.S., Matre, V. and Bergholtz, S. (2000) Protein-oligonucleotide interactions. In Meyers, R.A. (ed.), *Encyclopedia of*

- Analytical Chemistry*, John Wiley & Sons, Chichester, UK, pp. 5997–6017.
37. Das,S., Kenan,D.J., Bocskai,D., Keene,J.D. and Dasgupta,A. (1996) Sequences within a small yeast RNA required for inhibition of internal initiation of translation: interaction with La and other cellular proteins influences its inhibitory activity. *J. Virol.*, **70**, 1624–1632.
 38. Evstafieva,A.G., Beletsky,A.V., Borovjagin,A.V. and Bogdanov,A.A. (1993) Internal ribosome entry site of encephalomyocarditis virus RNA is unable to direct translation in *Saccharomyces cerevisiae*. *FEBS Lett.*, **335**, 273–276.
 39. Deng,Q.L., Ishii,S. and Sarai,A. (1996) Binding site analysis of c-Myb: screening of potential binding sites by using the mutation matrix derived from systematic binding affinity measurements. *Nucleic Acids Res.*, **24**, 766–774.
 40. Ness,S.A., Marknell,A. and Graf,T. (1989) The v-myb oncogene product binds to and activates the promyelocyte-specific mim-1 gene. *Cell*, **59**, 1115–1125.
 41. Brendeford,E.M., Myrset,A.H., Hegvold,A.B., Lundin,M. and Gabrielsen,O.S. (1997) Oncogenic point mutations induce altered conformation, redox sensitivity and DNA binding in the minimal DNA binding domain of avian myeloblastosis virus v-Myb. *J. Biol. Chem.*, **272**, 4436–4443.
 42. Clarke,N.D. and Berg,J.M. (1998) Zinc fingers in *Caenorhabditis elegans*: finding families and probing pathways. *Science*, **282**, 2018–2022.
 43. Schmidt,M., Nazarov,V., Stevens,L., Watson,R. and Wolff,L. (2000) Regulation of the resident chromosomal copy of c-myc by c-Myb is involved in myeloid leukemogenesis. *Mol. Cell. Biol.*, **20**, 1970–1981.
 44. Ganter,B., Chao,S.T. and Lipsick,J.S. (1999) Transcriptional activation by the myb proteins requires a specific local promoter structure. *FEBS Lett.*, **460**, 401–410.