



Identification *in silico* of putative damage responsive elements (DRE) in promoter regions of the yeast genome

Walkiria Luckwu de Santana Silva¹, Andre Ricardo de Oliveira Cavalcanti², Katia Silva Guimarães¹ and Marcos Antonio de Morais Jr.^{3*}

¹Universidade Federal de Pernambuco, Centro de Informática, Laboratório de Bioinformática, Recife, PE, Brazil.

²Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Departamento de Química Fundamental, Recife, PE, Brazil.

³Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Genética, Recife, PE, Brasil.

Abstract

We report an *in silico* analysis to identify nucleotide sequence motifs in DNA repair genes that may define a binding site for regulatory proteins during the induction of those genes by mutagens. The damage responsive elements (DRE) weight matrix generated in this analysis was used to search for homologous sequences in the promoter region of all genes, including putative gene and hypothetical open reading frames (ORFs), in the *Saccharomyces* Genome Data Base (SGD). The results demonstrated that over one third of the yeast genes in the database presented at least one 15-bp sequence in their promoter region with 85% or more of similarity to the DRE consensus sequence. The presence of the DRE sequence in the promoter region of regulatory genes and its high similarity to other well reported DNA binding sites points to its involvement in the general regulation of not only DNA repair genes but yeast genes in general.

Key words: DNA binding site, DNA repair, gene promoter, gene regulation, Transfac, weight matrix.

Received: September 24, 2004; Accepted: April 12, 2005.

Introduction

Mutagenic agents that damage DNA can result in increased transcription of a variety of *Saccharomyces cerevisiae* genes, including DNA repair genes such as the *RAD2*, *RAD7*, *RAD18*, *RAD23*, *RAD51*, *RAD54*, *PHR1* and *MAG1* as well as genes involved in DNA metabolism and protein modification such as genes *RAD6*, *RNR1*, *RNR2*, *RNR3*, *CDC9*, *POL1* and *UBI4* (Friedberg, 1991). Moreover, four damage inducible (DIN) and six DNA damage responsive (DDR) genes have so far been identified (McClanahan *et al.*, 1984; Ruby and Szostak, 1985). The induction of these genes is dependent on the presence of a set of *cis*-regulatory sequences in their promoter region, as has been shown by detailed studies of the *RAD2* and *RNR2* gene promoters (Friedberg, 1991). The promoter region of the *RAD2* gene contains two upstream activation sequence (UAS) elements known as damage responsive elements 1

(DRE1) and 2 (DRE2), essential for DNA-damage induced expression (Siede and Friedberg, 1992), while the *RNR2* gene promoter contains three UAS elements (Elledge and Davis, 1989).

Lowndes and Murguia (2000) have schematically described a working model for checkpoint gene regulation in response to different types of DNA damage. In brief, the presence of pre-replicative bulk-lesions or double-strand breaks generated by exposure to mutagenic agents activates parallel upstream sensor mechanisms (UAS elements or upstream repressor sequence (URS) elements) by means of effector proteins that use modification mechanisms such as phosphorylation to activate specific DNA binding proteins that bind to target genes in the UAS or URS elements and arrest both transcription and the cell cycle. Therefore, cell cycle and some metabolic genes are repressed, while DNA repair, DNA metabolism and other metabolic genes are induced. Recently, microarray analysis has been employed to survey the whole *S. cerevisiae* genome in order to form a complete picture of cellular responsiveness to DNA injuries (Jelinsky *et al.*, 2000; Gasch *et al.*, 2001), although it

Send correspondence to Marcos Antonio de Morais Jr. Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Genética, Av. Moraes Rego s/n, 50732-970 Recife, PE, Brazil. E-mail: morais@lika.ufpe.br.

still seems that the search for such inducible genes is far from complete.

The *SNM1/PSO2* (henceforth called the *SNM1* gene) gene belongs to the excision repair pathway (RAD3 pathway) and its product is involved in the repair of interstrand DNA cross-links caused by bifunctional mutagens (Henriques and Brendel, 1990). The *SNM1* gene can be induced by cross-linking agents or ultraviolet light (Wolters *et al.*, 1996), serial deletions in the *SNM1* gene promoter having showed that a 15-bp sequence homologous to the *RAD2* gene and the DRE2 element is essential for *SNM1* induction (Wolters *et al.*, 1996). Sequences similar to DRE2 have been found in other genes involved in DNA repair and nucleotide synthesis (Siede and Friedberg, 1992). Since *SNM1* gene expression seems to be tightly controlled in yeasts by the presence of DNA lesions, it may offer a good platform to study the DNA repair regulatory circuitry that can be predicted by functional genomics and computational tools. Similarly, deletion of the DRE-like element present at the *MAG1* gene promoter decreased the level of mutagen-induced gene expression five times (Xiao *et al.*, 1993) in a manner similar to that seen for the *SNM1* gene (Wolters *et al.*, 1996).

During the *in silico* study reported in this paper we carried out computational analysis of the promoter region of yeast genes to identify base sequence patterns that could be targeted by activator or repressor proteins in response to DNA damage. We used the DRE1 and DRE2 sequences found in some well-known DNA repair genes such as *RAD2* and *SNM1* to identify DRE-like sequences in a variety of related and unrelated DNA repair genes.

Materials and Methods

Matrix sequence preparation

The DRE-like sequences in the promoter region of the *SNM1*, *RAD2*, *PHR1*, *RAD18*, *RNR3*, *RAD23* and *RAD7* genes have been described as essential for the expression of these genes (Wolters *et al.*, 1996). We submitted these sequences to multiple alignments using the MegAlign tool of the DNASTar software package (DNASTar Inc., USA) to generate a consensus sequence containing 15 bp. In parallel, a set of sequences containing the -1000 bp upstream region of 13 yeast genes described as related to repair in the *Saccharomyces* Genome Database (SGD) (www.yeastgenome.org) was selected after pairwise alignment of each sequence with the 15 bp consensus generated by us. The SGD set and the consensus sequence were inputted into the Clustal-W program (<http://clustalw.genome.ad.jp/>) using the default parameters, yielding an alignment which contained several 15 bp sequences which were similar to the 15 bp consensus sequence. The ten highest scored were chosen to produce a new set with the seven original sequences. This new set was then used to generate a weight matrix by using the genome exploring and modeling soft-

ware (GEMS) launcher (www.genomatix.de/). The matrix generated was first checked against the whole yeast genome and produced a random expectation of 0.12 matches per 1000 nucleotides. The matrix was also used to scan the -500 bp upstream region of the yeast genes by using the default conditions (core similarity of 0.9) of the MatInspector Professional Program (Quandt *et al.*, 1995; Wingender *et al.*, 2000). This program contains the Transcription Factor Database (TRANSFAC) that allows the identification of the query matrix by comparing to those deposited in the database. The sequences resulting from this search being re-analyzed to estimate the degree of conservation by increasing the core similarity to 0.95 and 1.00.

Gene annotation and DNA binding motifs analysis

The genes matching the matrix sequence were first grouped according to their metabolic families by using the MIPS (<http://www.mips.biochem.mpg.de/>) and YPD (<http://www.proteome.com/database/YPD>). The gene clusters and DNA binding sequences used were those available on-line at http://arep.med.harvard.edu/network_discovery/ (Tavazoie *et al.*, 1999) and <http://cgsigma.cshl.org/jian/> (Zhu and Zhang, 1999). The microarray data from mutagen-treated cells was taken from <http://www.hsph.harvard.edu/geneexpression/> (Jelinsky *et al.*, 2000) and <http://www-genome.stanford.edu/Mec1/> (Gasch *et al.*, 2001).

Results

Identification of DRE-matrix element in yeast DNA repair genes

The sequence of the DRE-like sequence was 5'-GNRRAKGNATTGAAA-3' (The bold-face letters N, R, R, K and N correspond to five ambiguous nucleotide positions) as established by the alignment of the DNA repair genes described by Wolters *et al.* (1996) and 13 other DNA repair genes collected from SGD. Unless otherwise indicated, references to DRE-like sequences refer to sequences conforming to this sequence henceforth. The nucleotide codes used were according to the International Union of Pure and Applied Chemistry (IUPAC). The consensus index for each nucleotide position is shown in Figure 1.

When we scanned the -500 bp sequence of the yeast genome we found many genes containing sequences which were similar to the DRE-like consensus sequence with complete conservation in the core TGAAA sequence at the 3' end (Table 1). Other nucleotides also seem to be conserved, includes the nucleotide G at position 1, A at position 5, G at position 7 and the dinucleotide AT just before the core. All genes identified in Table 1 were induced by at least one mutagen according to microarray data produced by Jelinsky *et al.* (2000), although the induction level and the nature of the mutagen were variable.

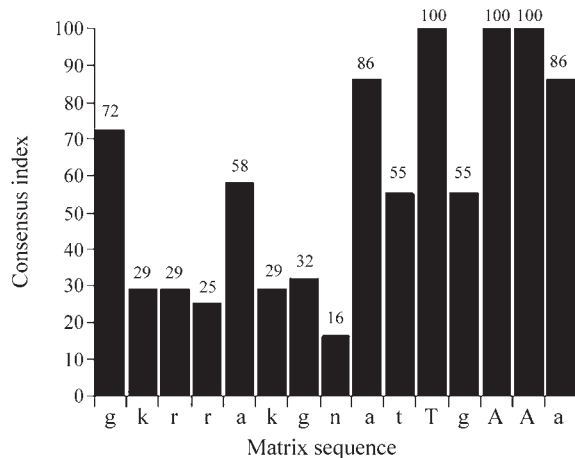


Figure 1 - Matrix sequence output by MatDefine from GEMS Launcher suite v3.6 using damage repair elements (DRE) identified from yeast DNA repair genes. Capital letters in the DNA sequence represent nucleotide positions with high consensus index score (Ci-value > 60) as indicated by MatInspector. Ci-values are shown above the gray bars.

Our SGD search for homologous sequences in the -500 bp promoter region of yeast genes yielded a total of 1645 matches with at least 85% matrix similarity and 100% similarity to the core TGAAA. In the rest of this paper we will use the term 'gene' to refer not only to both identified and putative genes but also in some cases even to hypothetical ORFs described on the SGD web site. The MatInspector result suggested that this element might be widespread in the promoter regions of the yeast genome, these regions accounting for a third of the total genes described in the SGD.

We also found two matches in the promoter regions of the ORFs YER041w (*YEN1* gene) and YEL018w

(Table 1), with the last ORF presenting two DRE-like consensus sequences in its promoter region. Both ORFs have been described in the Yeast genome Database as putative DNA repair genes, which were induced by treatment with methyl methane sulfoxide (MMS) and with 4-nitroquinoline 1-oxide (4-NQO).

When we increased the matrix similarity threshold to 95% the number of matches decreased substantially to 15 a group of genes (Table 2), with the *SNMI* gene being the only gene in this group involved in some type of DNA metabolism and six (including the glutamate 5-kinase encoding *PRO1* gene containing the complete DRE-like consensus in the promoter region) were not induced by any mutagen (Table 2). These results indicate that despite its function in the induction of the *SNMI* gene the DRE-like sequence may act as a regulatory element for other genes.

Additionally, the *SNZI* gene was the only match containing four DRE-like consensus sequences and it was also described by Jelinsky *et al.* (2000) as being very susceptible to induction by mutagens with an induction factor of about 82 after treatment with MMS. This gene is induced in response to nitrogen limitation and growth arrest, and its protein product is part of the glutamine amidotransferase complex together with Sno1 protein (Dong *et al.*, 2004). We found another 20 matches containing three DRE-like consensus sequences presenting from 85% to 95% matrix similarity, none of which were directly involved in DNA repair. In our analysis, three ORFs (YLR060W, phenylalanyl-tRNA synthetase; YLR224W, hypothetically a protein; and YLR059C, putative 3'-5' exonuclease) containing three DRE-like consensus sequences were not known to be induced by any mutagen yet tested, so there was no exact correlation between the presence or the num-

Table 1 - Computational analysis of DNA repairs genes from *S. cerevisiae* identified by MatInspector algorithm containing DRE-matrix elements in their promoters.

Gene/SGD access	Biological function ^a	Upstream position ^b	DNA strand ^b	Matrix similarity ^c	Sequence (5'-3')	Induction factor ^d
<i>RAD50</i> /YNL250W	Strand-break repair protein	213-227	(-)	0.852	GGCAAAGC CGTGAAA	2.5 (BCNU)
<i>PSO2</i> /YMR137C	ICL repair protein	384-398	(+)	0.957	GGAAACGG ACTGAAA	3.4 (8-MOP) ^e
<i>MGT1</i> /YDL200C	O ⁶ -Methyltransferase	182-196	(-)	0.862	TCAGATGT AATGAAA	3.3 (MMS)
<i>RAD16</i> /YBR114W	NER protein	482-496	(+)	0.906	GTGTACCA ACTGAAA	3.4 (t-BuOOH)
<i>RAD18</i> /YCR066W	Error-prone repair protein	22-36	(-)	0.924	GAAAAGAA AATGAAA	2.9 (MMS)
<i>RAD23</i> /YEL037C	NER protein	207-221	(+)	0.907	GTGGCGAA ATTGAAA	2.5 (BCNU)
<i>YEN1</i> /YER041W	Weak similarity to Rad2p	93-107	(-)	0.872	TCAGAGGT AATGAAA	3.3 (MMS)
		432-446	(+)	0.905	GTCCATCC AATGAAA	
<i>EAF5</i> /YEL018W	Weak similarity to Rad50p	39-53	(+)	0.850	GAGGATTG TTTGAAA	2.5 (4NQO)
					GKRRAKGN ATTGAAA	

^aICL: inter-strand cross link; NER: nucleotide excision repair.

^bUpstream nucleotide position relative to +1 and Watson (+) or Crick (-) strands relative to SGD.

^cDRE sequence similarity relative to the matrix.

^dMicroarray data published by Jelinsky *et al.* (2000), considering the highest induction value with a specific mutagen in brackets. Abbreviations: MMS (methyl metanosulfonate), BCNU (1.3-bis-2-chloroethyl-1-nitrosourea), t-BuOOH (*tert*-butyl hydroperoxide), 4-NQO (4-nitroquinoline 1-oxide) and 8-MOP (8-methoxy-psoralen).

^eData from Wolters *et al.* (1996) using *SNM1-lacZ* fusion.

^fHighly conserved (80%) nucleotide sequences are in bold. Ambiguous codes are according to IUPAC: N (any), R (A or G) and K (G or T).

Table 2 - List of the yeast genes showing 95% or more sequence similarity in their upstream sequences to DRE-matrix^a.

Gene/SGD access	Biological function	Upstream position	DNA strand	Matrix similarity	Sequence (5'-3')	Induction factor
<i>PSO2/YMR137C</i>	ICL repair protein	384-398	(+)	0.957	GGAAACGG ACTGAAA	3.4 (8-MOP)
<i>PAC11/YDR488C</i>	Required in the absence of Cin8p	479-493	(+)	0.952	GAAGAAGG ATTGAAA	5.5 (4-NQO)
<i>PIN4/YBL051C</i>	Similar Z66568Cp of <i>S.pombe</i>	109-123	(-)	0.958	GTAGAGTT ATTGAAA	not induced
<i>PRE3/YJL001W</i>	Endopeptidase multicatalytic complex subunit	396-410	(+)	0.967	GGAAAAGC ATTGAAA	4.9 (t-BuOOH)
<i>RTS1/YOR014W</i>	Phosphatase A regulatory subunit	318-332	(-)	0.962	GGGAAGGA AATGAAA	2.5 (t-BuOOH)
<i>YLR282C</i>	Questionable ORF	446-460	(-)	0.962	GGCAATGG ATTGAAA	not induced
<i>OST1/YJL002C</i>	Oligosacaryltransferase alfa subunit	80-94	(-)	0.967	GGAAAAGC ATTGAAA	not induced
<i>NIF3/YGL221C</i>	Ngg1p-interacting factor 3	399-413	(-)	0.968	GTGAAAGA ATTGAAA	2.0 (4-NQO)
<i>PUF6/YDR496C</i>	Similar to hypothetical human protein	141-155	(+)	0.962	GTTCATGA ATTGAAA	not induced
<i>PEX22/YAL055W</i>	Hypothetical protein	235-249	(-)	0.970	GTGAAAGC ATTGAAA	2.9 (t-BuOOH)
<i>MUB1/YMR100W</i>	Hypothetical protein	393-407	(-)	0.976	GGAAATGA ATTGAAA	5.5 (4-NQO)
<i>CLP1/YOR250C</i>	Similarity to SPAC22H10.05c protein of <i>S.pombe</i>	59-73	(+)	0.974	GAAAAGGC ATTGAAA	3.5 (MMS)
<i>DRS2/YAL026C</i>	Type-P amino-phospholipid-ATPase transmembrane	92-106	(-)	0.974	GTAAAAGA ATTGAAA	not induced
<i>CFT1/YDR301W</i>	Pre-mRNA 3'-terminal processing factor	126-140	(-)	0.995	GTAAAGGA ATTGAAA	4.9 (4-NQO)
<i>PRO1/YDR300C</i>	Glutamate 5-kinase	301-315	(+)	0.995	GTAAAGGA ATTGAAA	not induced

^aDescriptions, symbols and abbreviations are as described in Table 1.

ber of DRE-like consensus sequences and the ability of a gene to be induced by a mutagen.

We were also able to identify 19 transcription factor-encoding genes containing in their promoter regions upstream sequences which were more than 90% similar to the DRE-like consensus element, 15 of these genes being inducible by treatments which damage DNA (Table 3). Ten of the 15 mutagen-inducible genes encode subunits of the

so-called basal transcription complex, which either binds to the TATA box or correctly localize the RNA polymerase complex at the initiation site during initiation of transcription. Additionally, DRE-like consensus sequences were identified at the promoter region of genes encoding specific transcription factors, such as Gal4p, Leu3p and Gcr3p (Table 3). Therefore, the presence of DRE-like consensus sequences in the promoter region of regulatory genes may

Table 3 - Yeast transcription factor-encoding genes identified by MatInspector for the presence of a DRE-matrix motif in their -500 bp promoter sequence^a.

Gene/SGD access	Biological function	Upstream position	DNA strand	Matrix similarity	Sequence (5'-3')	Induction factor
<i>DST1/YGL043W</i>	TFIIS elongation factor	75-89	(-)	0.86	ATCAATAC ATTGAAA	3.4 (MMS)
		358-372	(-)	0.87	GGGTAAGA GTTGAAA	
		415-429	(+)	0.92	GACTACGT ATTGAAA	
<i>GAL4/YPL248C</i>	Galactose induction factor	209-223	(-)	0.87	GCGGACGT TTTGAAA	not induced
		279-293	(-)	0.88	GTAAATAC TTTGAAA	
<i>GCR3/YMR125W</i>	Glycolytic genes induction factor	457-471	(-)	0.89	GTAATTTG AATGAAA	2.7 (4-NQO)
<i>LEU3/YLR451W</i>	Branched-chain amino acids genes induction factor	342-356	(+)	0.89	GAGCAAAA AATGAAA	4.0 (MMS)
<i>MBP1/YDL056W</i>	MBF complex subunit	330-344	(+)	0.86	GATGAAAA AATGAAA	2.0 (MMS)
<i>MGAI/YGR249W</i>	Similar to heat shock factor	153-167	(-)	0.89	GGTCAACA ACTGAAA	not induced
<i>REB1/YBR049C</i>	Transcription factor	150-164	(-)	0.89	TTCAATGT ATTGAAA	2.6 (MNNG)
		151-165	(+)	0.86	TTCAATAC ATTGAAA	
<i>RF42/YNL312W</i>	36kDa DNA replication factor	180-194	(-)	0.85	GTAGAATC TCTGAAA	5.1 (MMS)
<i>RRN10/YBL025W</i>	RNA polymerase I-specific initiation transcription factor	373-387	(+)	0.86	GCGAATGT CATGAAA	3.5 (MMS)
<i>RRN6/YBL014C</i>	RNA polymerase I-specific initiation transcription factor	63-77	(+)	0.89	GAGAAAT ACTGAAA	2.3 (4-NQO)
		210-224	(-)	0.89	GTGAGAAG ATTGAAA	
<i>RRN7/YJL025W</i>	RNA polymerase I-specific initiation transcription factor	346-360	(+)	0.87	TTAAAAA ATTGAAA	2.6 (BCNU)
<i>SKN7/YHR206W</i>	Similar to heat shock factor	207-221	(+)	0.89	GTGACAAG ATTGAAA	not induced

Table 3 (cont.)

Gene/SGD access	Biological function	Upstream position	DNA strand	Matrix similarity	Sequence (5'-3')	Induction factor
<i>TAF40/YML015C</i>	TBP-associated 40 kDa TFIID subunit	407-421	(+)	0.88	TGAAAGAA ATTGAAA	4.0 (MMS)
<i>TAF61/YDR145W</i>	TBP-associated 61 kDa TFIID subunit	34-48	(+)	0.87	AATAAGGT ATTGAAA	5.2 (4-NQO)
<i>TAF67/YMR227C</i>	TBP-associated 67 kDa TFIID subunit	114-128	(-)	0.86	CCAGAAGA ATTGAAA	8.8 (MMS)
<i>TAF90/YBR198C</i>	TBP-associated 90 kDa TFIID subunit	241-255	(-)	0.86	CTGCATGT AGTGAAA	3.2 (MMS)
<i>TFB1/YDR311W</i>	TFIIH subunit	256-270	(+)	0.86	GGGAAAGA TATGAAA	2.6 (4-NQO)
<i>TFC1/YBR123C</i>	TFIIIC subunit	401-415	(+)	0.88	AGACAGGA ACTGAAA	3.2 (MMS)
<i>TFC2/YPR186C</i>	TFIIIA subunit	383-397	(-)	0.86	GATATTA ATTGAAA	not induced

^aDescriptions, symbols and abbreviations are as described in Table 1.

support the idea of the complex inter-connection of metabolic networks.

Homology between the DRE-like consensus sequence and known yeast DNA binding sites

Searching for homologous sequences was performed in the *Saccharomyces cerevisiae* Promoter Database (SCPD) using the consensus sequence and admitting six possible mismatches corresponding to the five ambiguous nucleotide positions in the matrix **GNRRAKGNAT** TGAAA (in bold) plus one mismatch at any of the other positions. A total of 16 transcriptional factor-binding sites were identified during our analysis (Table 4). The heat shock element (HSE) controls the stress response of several yeast genes encoding heat shock proteins by binding the heat shock transcription factor (HSF) described by Sakurai and Fukasawa (2001). All *HSP* genes were induced by mutagens with the exception of the HSE-containing *CUP1* gene that was mutagen-repressed (Jelinsky *et al.*, 2000).

High homology was also observed between the DRE-like consensus sequence and the stress response element (STRE) present at the promoter region of the *DDR2* gene (Table 4). This gene is induced by a variety of mutagens as well as by heat shock but the function of its protein product is still unknown (Treger *et al.*, 1998). A previous computer-generated pattern also identified a STRE motif for *DDR48* gene (Treger *et al.*, 1998), because of which we also performed sequence alignment of the *DDR48* gene promoter and identified the presence of the homologous element 5'-**GGCCAGCACCGGAAA**-' (conserved positions in bold) at position -318 to -304 at the Crick strand. The third member of this group, the polyubiquitin encoding gene *UBI4* is induced in response to a variety of environmental stresses such as mutagens and heat shock due to the presence of both HSE and STRE sequences in its promoter region (Simon *et al.*, 1998). Our pairwise alignment analysis with the DRE sequence identified the sequence 5'-**TAAAAAAGATTG AAC**-3' (conserved nucleotides in bold) at positions -301 to -315 in the promoter region of the *STE12* gene encoding a transcription factor involved in pheromone and pseudophal growth signal transduction pathways.

Table 4 - Motifs in the yeast genome homologous to the DRE sequence.

Transcription factor	Binding motif	Sequence (5'-3')	Cell metabolism
-	DRE	GKRRAKGN ATTGAAA	
Hsf1p	HSE	GGAACGTT CTGGAAA	Stress response
Msn2/4p	STRE	ATAAGGGG TGAGAAA	Cell death and aging
Ime1p	IRE	GGGAAAGG ATCAAAG	Pre-meiotic DNA synthesis
Ino2p/Ino4p	-	CATGTGAAAAATGTGAAA	Phospholipid biosynthesis
Dal82p	-	GAAAAGCA AACGCAA	Purine biosynthesis
Abf1p	-	GTAAGGCG CTATCAA	DNA replication
Adr1p	-	GTGACAGA ATTGGAG	Cell aging and fermentation
Mot3p	-	GTAATAGG GATCAAT	Mating and transposition
Rap1p	-	CCATACTT TTTGAAA	Chromatin maintenance
Cbf1p	-	AATGATTC ATTGAAA	Chromosome segregation
-	MCB	GGGTAACG CCTGAAA	Cell cycle control
-	ECB	GGAAACGT AATGAAA	Cell cycle control
-	SCB	—TGTAATTGAAA	Cell cycle control
-	M13	ACCGAACA AATGAAA	n.d.
-	M27	GAAAGAAA TTTGAAA	n.d.
-	M3a	TAAAGAAA ATTGAAA	n.d.

n.d.: not defined within a recognizable pathway by Tavazoie *et al.* (1999).

Discussion

Our homology similarity-based searches in the promoter regions of the whole yeast genome using a consensus sequence based on the 5'-**GNRRAKGNATTGAAA**-3' DRE-like consensus led to some intriguing results. Firstly, our data shows that not all DNA repair genes that contain the DRE-like consensus sequence analyzed in this work were induced by the mutagens tested by Jelinsky *et al.* (2000), although, of course, induction by other mutagens not yet tested cannot be ruled out. Microarray analysis also revealed interesting data in relation to gene regulation in that of the 12 genes in the microarray profile with signifi-

cant mutagen induction only *RNR3* is involved in DNA metabolism (Jelinsky *et al.*, 2000). Ren *et al.* (2000) have demonstrated that the expression of the *FUR4* gene is increased by the presence of galactose despite the fact that its promoter region does not contain any Gal4p consensus sequence binding site, although yet other yeast genes exist which contain the Gal4p binding site but are not galactose-inducible. Therefore, we cannot discard the possibility that our 5'-GNRRAKGNATTGAAA-3' DRE-like sequence is involved in some kind of gene expression regulation. This suggests that the simple presence of a DRE-like sequence alone may not be a good enough indicator for gene induction by DNA-damage agents, although its presence is essential for induction of the *SNM1/PSO2* (Wolters *et al.*, 1996) and *MAG1* (Xiao *et al.*, 1993) genes. Ren *et al.* (2000) and Iyer *et al.* (2001) have indicated the need for additional empirical data and combined, and perhaps improved, search algorithms in order for investigators to accurately predict genuine binding sites.

Tavazoie *et al.* (1999) have proposed transcriptional regulatory networks based on clustering yeast genes according to the presence of putative *cis*-regulatory sequences in their promoter regions, these authors showing that the STE factor is more frequent in genes belonging to cluster 8 which is rich in genes related to carbohydrate and tricarboxylic acid (TCA) metabolism. Intriguingly, the only genes in our analysis belonging to Tavazoie's cluster 8 were *RAD14* and *UBI4* and no other stress or DNA damage responsive genes. Genes involved in DNA synthesis and replication, cell cycle control and mitosis, recombination and DNA repair were allocated to Tavazoie's cluster 2, while genes responsive to stress and involved in cell rescue, defense and cell death were allocated to cluster 5 (Tavazoie *et al.*, 1999). Gene cluster 2 is characterized by the presence of MCB and SCB sequence motifs (Table 4), and, to a lesser extent, the M13 motif, which were all also recognized as homologous to our DRE-like sequence (Table 4). The presence of MCB binding sites was reported to confer irradiation-replication specific regulation on many DNA repair genes (Mercier *et al.*, 2001). Cluster 5 genes, to which DNA repair genes *MAG1* and *RAD10* belong, also contain mainly the M13 motif (Tavazoie *et al.*, 1999). This motif contains the pentanucleotide TGAAA that is recognized by the ADR1p transcriptional factor. The homology we found between M13 and our DRE-like sequence (Table 4) suggests that the 5'-GNRRAKGNATTGAAA-3' DRE sequence can be also recognized by a transcription factor. Similarly, Ettwiller *et al.* (2003) combined information from metabolic networks with genome information to predict *cis*-regulatory sequences in yeast promoters, their analysis producing 42 motifs with motif 19 showing high similarity to our DRE-like sequence and since motif 19 is not a recognized yeast transcription factor binding motif it is possible that it may represent a new regulatory motif.

In an attempt to describe a new method for identification of regulatory sequences, Harrison and DeLisi (2001) identified the consensus sequence AWGAAA as a target for binding of the Ste12 transcription factor by using the anchor motif generation method. The AWGAAA motif is similar to those previously described in TRANSFAC (ATGAAC) and the SCPD (ATGAAA) databases, these sequences being identical to the core sequence (ATGAAA) of our 5'-GNRRAKGNATTGAAA-3' DRE-like sequence. The homology found between our DRE-like sequence and the previously identified regulatory motifs, either experimentally or by *in silico* analysis, suggests that these sequences belong to a family of regulatory sequences participating in the general mechanism of gene regulation. Further experimental data should confirm this hypothesis.

Acknowledgments

The authors are grateful to Dr. K. Quandt (Genomatix, Germany) for permission to use GEMS Launcher professional to prepare the DRE-matrix.

References

- Dong YX, Sueda S, Nikawa J and Kondo H (2004) Characterization of the products of the genes *SNO1* and *SNZ1* involved in pyridoxine synthesis in *Saccharomyces cerevisiae*. *Eur J Biochem* 271:745-52.
- Elledge SJ and Davis RW (1989) Identification of the DNA damage responsive element of *RNR2* and evidence that four distinct cellular factors bind it. *Mol Cell Biol* 9:5373-5386.
- Errede B and Ammerer G (1989) STE12, a protein involved in cell-type-specific transcription and signal transduction in yeast, is part of protein-DNA complexes. *Genes Develop* 3:1349-1361.
- Ettwiller LM, Rung J and Birney E (2003) Discovering novel *cis*-regulatory motifs using regulatory networks. *Genome Res* 13:883-895.
- Friedberg EC (1991) Yeast genes involved in DNA repair process: New looks on old faces. *Mol Microbiol* 5:2303-2310.
- Gasch AP, Huang M, Metzner S, Botstein D, Elledge SJ and Brown PO (2001) Genomic expression responses to DNA-damage agents and the regulatory role of the yeast *ATR* homolog Mec1p. *Mol Cell Biol* 12:2987-3003.
- Harrison H and DeLisi C (2001) Condition specific transcription factor binding site characterization in *Saccharomyces cerevisiae*. *Bioinformatics* 18:1289-1296.
- Henriques JAP and Brendel M (1990) The role of *PSO* and *SNM* genes in DNA repair of the yeast *Saccharomyces cerevisiae*. *Curr Genet* 18:387-393.
- Hughes JD, Estep PW, Tavazoie S and Church GM (2000) Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J Mol Biol* 296:1205-1214.
- Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M and Brown PO (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409:533-538.
- Jelinsky S, Estep P, Church G and Samson L (2000) Regulatory networks revealed by transcriptional profiling of damaged

- Saccharomyces cerevisiae* cells: RPN4 links base excision repair with proteasomes. *Mol Cell Biol* 20:8157-8167.
- Kiser GL and Weinert TA (1996) Distinct roles of yeast *MEC* and *RAD* checkpoint genes in transcriptional induction after DNA damage and implications for function. *Mol Cell Biol* 7:703-718.
- Lowndes NF and Murguia JR (2000) Sensing and responding to DNA damage. *Curr Opin Genet Dev* 10:17-25.
- McClanahan T and McEntee K (1984) Specific transcripts are elevated in *Saccharomyces cerevisiae* in response to DNA damage. *Mol Cell Biol* 4:2356-2363.
- Mercier G, Denis Y, Marc P, Picard L and Dutriex M (2001) Transcriptional induction of repair genes during slowing of replication in irradiated *Saccharomyces cerevisiae*. *Mut Res* 487:157-172.
- Quandt K, Frech K, Karas H, Wingender E and Werner T (1995) MatFind and MatInspector: New fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucl Acids Res* 23:4878-4884.
- Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP and Young RA (2000) Genome-wide location and function of DNA binding proteins. *Science* 290:2306-2309.
- Ruby SW and Szostak JW (1985) Specific *Saccharomyces cerevisiae* genes are expressed in response to DNA-damaging agents. *Mol Cell Biol* 5:75-84.
- Sakurai H and Fukasawa T (2001) A novel domain of the yeast heat shock factor that regulates its activation function. *Biochem Biophys Res Commun* 285:696-701.
- Siede W and Friedberg EC (1992) Regulation of the yeast *RAD2* gene: DNA damage-dependent induction correlates with protein binding to regulatory sequences and their deletion influences survival. *Mol Gen Genet* 232:247-256.
- Simon JR, Treger JM and McEntee K (1998). Multiple independent regulatory pathways control *UBI4* expression after heat shock in *Saccharomyces cerevisiae*. *Mol Microbiol* 31:823-832.
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ and Church GM (1999) Systematic determination of genetic network architecture. *Nat Genet* 22:281-285.
- Treger JM, Schmitt AP, Simon JR and McEntee K (1998). Transcriptional factor mutations reveal regulatory complexities of heat shock and newly identified stress genes in *Saccharomyces cerevisiae*. *J Biol Chem* 273:26875-26879.
- Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüß M, Reuter I and Schacherer F (2000) TRANSFAC: An integrated system for gene expression regulation. *Nucl Acids Res* 28:316-319.
- Wolter R, Siede W and Brendel M (1996) Regulation of *SNM1*, an inducible *Saccharomyces cerevisiae* gene required for repair of DNA cross-links. *Mol Gen Genet* 250:162-168.
- Xiao W, Singh KK, Chen B and Samson L (1993) A common element involved in transcriptional regulation of two DNA alkylation repair genes (*MAG* and *MGT1*) of *Saccharomyces cerevisiae*. *Mol Cell Biol* 13:7213-7221.
- Zhu J and Zhang MQ (1999) SCPD: A promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics* 15:607-611.

Associate Editor: Sérgio Olavo Pinto da Costa