Identification of New Genes Regulated by the Crt1 Transcription Factor, an Effector of the DNA Damage Checkpoint Pathway in Saccharomyces cerevisiae*

Jolanta Zaim, Elżbieta Speina, and Andrzej M. Kierzek‡

From the Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warsaw, Poland

The Crt1 (RFX1) protein in Saccharomyces cerevisiae is an effector of the DNA damage checkpoint pathway. It recognizes a 13-bp cis-regulatory element in the 5′-untranslated region (5′-UTR) of the ribonucleotide reductase genes RNR2, RNR3, and RNR4; the HUG1 gene; and itself. We calculated the weight matrix representing the Crt1p binding site motif according to analysis of the 5′-UTR sequences of the genes that are under its regulation. We subsequently searched the 5′-UTR sequences of all the genes in the yeast genome for the occurrence of this motif. The motif was found in regulatory regions of 30 genes. A statistical analysis showed that it is unlikely that a random gene cluster contains the motif conserved as well as the Crt1p binding site. Analysis of microarray data provided supporting evidence for five putative Crt1p targets: YFL345W, YBR255C, YLR345W, UBC5, NDE2, and NTH2. We used reverse transcription-PCR to compare the expression levels of these genes in wild-type and crt1Δ strains. Our results indicated that YFL345W, YBR255C, and NTH2 are indeed under the regulation of Crt1p. Sequence analysis of the Crt1p3p indicated that this protein may be involved in folate metabolism either by carrying serine hydrolase activity required for the novel metabolic pathway involving dihydrofolate reductase (DHFR) or by directly interacting with the DHFR enzyme. We postulate that Crt1p may influence deoxyribonucleotide synthesis not only by regulating expression of the RNR genes but also by modulating DHFR activity. YFL345Wp and NTH2p are enzymes involved in the central metabolism under stress conditions.

High fidelity of transmission of genetic information is crucial for the survival of living organisms. To assure faithful DNA replication, cells have evolved complex mechanisms that activate a multifaceted response to occasional DNA damage resulting from both environmental factors and cellular processes. In eukaryotes DNA damage activates evolutionarily conserved checkpoint pathways that induce apoptosis, arrest the cell cycle, and activate transcription of genes whose products are involved in repair processes. Signal transduction in these pathways proceeds through a kinase cascade involving products of the ATR, ATM, Chk1, and Chk2 genes in mammals and their yeast homologs MEC1, TEL1, CHK1, and RAD53 (1, 2). The DNA damage response pathways and their evolutionary conservation in eukaryotic cells have been reviewed elsewhere (1, 2).

The product of the CRT1 gene in the yeast Saccharomyces cerevisiae is one of the effector proteins that take part in the DNA damage checkpoint pathway (3). CRT1 encodes a DNA-binding protein that acts by recruiting Ssn6p and Tup1p, general repressors to the promoters of damage-inducible genes. In response to DNA damage Crt1p becomes phosphorylated via the Dun1p pathway downstream of the Rad53p checkpoint kinase. The hyperphosphorylated form of Crt1p does not bind DNA, preventing formation of a repressor complex, and leads to activation of damage-inducible genes.

So far there are five experimentally verified targets of Crt1p action: three ribonucleotide reductase genes, RNR2, RNR3, and RNR4 (3), and the hydroxyurea/UV light/γ radiation-induced gene HUG1 (4). Crt1p also regulates expression of its own gene (3). In the case of the RNR and CRT1 genes, Crt1p has been experimentally shown to recognize a 13-nucleotide-long cis-regulatory element (3). The element resembles the mammalian X-box motif recognized by RFX transcription factors that take part in regulation of the major histocompatibility complex genes (5, 6). This homology is not surprising considering that the DNA-binding domain of Crt1p shares significant sequence similarity with the DNA-binding domain of RFX proteins. In the case of the HUG1 gene, transcription factor binding has not been studied experimentally, but the upstream region of the gene contains 13-nucleotide-long motifs strongly resembling the Crt1p binding sites (4).

The Crt1p binding sites, experimentally studied in the RNR and CRT1 genes, have been classified as strong or weak based on the relative binding affinity of Crt1p and the degree of sequence conservation (3). In RNR2, RNR4, and CRT1, one strong and one weak site have been found. RNR3 contains one strong and two weak sites. Huang et al. (3) postulated that multiple binding sites allow graded response to the DNA damage signal. When the pool of Crt1p molecules decreases, Crt1p dissociates first from weak sites allowing gradual transcription induction.

Knowledge about experimentally verified Crt1p binding sites has never been used to define the binding site motif of the Crt1 transcription factor and study its distribution in regulatory regions of yeast genes. In this study we used a sequence motif discovery approach to derive optimal weight matrix representing the Crt1p binding site motif. The search of yeast gene regulatory regions with the resulting motif representation yielded a cluster of 30 putative Crt1p targets. Further analysis of expression profile data allowed us to select the set of most probable additional Crt1p targets. We experimentally verified that three of them, YLR345W, NTH2, and FSH3, are indeed regulated by Crt1p. Analysis of liter-

* This work was supported by Grant QLK3-2000-00785 from the European Union 5th framework program. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
‡ To whom correspondence should be addressed. Tel.: 48-22-5922348; Fax: 48-22-658-46-36; E-mail: andrzejk@ibb.waw.pl.

Received for publication, April 27, 2004, and in revised form, October 13, 2004
Published, JBC Papers in Press, October 19, 2004, DOI 10.1074/jbc.M404669200

Published, JBC Papers in Press, October 19, 2004, DOI 10.1074/jbc.M404669200

Printed in U.S.A.
nature data and the protein sequence of $\text{FSH3}$ suggests that the mec1p-rad53p-dun1p-crt1p signal transduction pathway may control deoxyribonucleotide synthesis not only at the level of ribonucleotide reductase activity but also by altering one-carbon group metabolism and consequently purine ring synthesis. The fact that $\text{YLR345W}$ and $\text{NTH2}$ are regulated by Crt1p indicates the regulatory link between DNA damage response and central metabolism.

**EXPERIMENTAL PROCEDURES**

**Data Sources—**Sequences of 1-kb-long 5'-untranslated regions (5'-UTRs) of all ORFs in the $\text{S. cerevisiae}$ genome were downloaded from the Saccharomyces Genome Database (8). 2 Experimentally confirmed sequences of Crt1p binding sites were taken from Huang et al. (3). For expression profile analysis rough microarray data of Gasch et al. (9) were obtained from Stanford Microarray Database, and the data of Jelinsky et al. (10) were obtained from ExpressDB WWW server (arep.med.harvard.edu/cgi-bin/ExpressDB yeast/EXDStart). Data of Hughes et al. (11) were obtained from WWW site www.rri.com/tech/pubs/cell.hughes.htm. The search for microarray data relevant to this work was aided by the yeast Microarray Global Viewer service (12). The Yeast Protein Database (13) and Swiss-Prot Database (14) were used for functional annotation of protein sequences. Sequence homology searches were performed with the BLAST program run at Swiss-Prot Database server.

**MEME Analysis of 5'-UTR Sequences of Genes Regulated by Crt1p—**

The MEME method (15) was used to find conserved motifs in the 1-kb-long 5'-UTR sequences of genes experimentally shown to be regulated by Crt1p ($\text{RNR2}$, $\text{RNR3}$, $\text{RNR4}$, $\text{CRT1}$, and $\text{HUG1}$). MEME is a motif discovery algorithm, i.e. it searches for motifs that occur in a training set of unaligned sequences more frequently than expected by chance. Motifs are represented as weight matrices. A default background model uses single base composition of the sequences analyzed. As an option, frequencies of words may be used. We found that similarly to other DNA sequence analysis problems, application of the fifth order Markov model of random sequence improves the quality of the results. To apply this approach we computed frequencies of all 1-, 2-, . . . , 6-nucleotide-long words in the data base of yeast 5'-UTR sequences. The oligonucleotide composition of yeast 5'-UTR sequences was computed with the WOCCOUNT program from the EMBOSS package (16).

MEME algorithm searches for motifs of predefined length. It also uses constraints on the number of occurrences of the motif per sequence. We searched for motifs ranging in length from 6 to 50 nucleotides and tested all kinds of constraints on the number of motif occurrences, namely one occurrence per sequence (OOPS mode), zero or one occurrence per sequence (ZOOPS mode), and any number of occurrences per sequence (TCM mode).

We used the MEME program from stand alone distribution of the MEME 3.0 package. The MAST program from the same package was applied to search the data base of 5'-UTR sequences of yeast genes for sites matching weight matrices obtained in MEME analysis (E-value cut-off of 10 was used).

**Computation of the Sequence Logo—**To compute information content of every position in the Crt1p binding site motif, we used the relative entropy formula (17): $R_i = \sum_{j} p_{ij} \log(p_{ij}/\bar{p}_j)$, where $R_i$ is the information content of the $i$th position in the motif, $p_{ij}$ is the probability of finding residue $i$ in the $j$th position of the motif, and $\bar{p}_j$ is the data set frequency of residue $i$. We used the relative entropy method instead of the more frequently used algorithm of Schneider and Stephens (18) to take into account background nucleotide frequencies that are highly biased in our data set. The sequence logo was drawn as follows. The height of every column in the logo was set to $R_i$. Relative heights of individual residues in the same column were computed according to their $\log(p_{ij}/\bar{p}_j)$ contribution to $R_i$. Residues for which the absolute value of $p_{ij}/\bar{p}_j$ was less than 0.1 were not drawn. If the residue column frequency was less than background frequency, the information given by the individual residue was negative. In this case the letter was drawn below the $x$ axis to indicate negative information content.

**RNA Isolation and RT-PCR—**Total RNA was isolated from yeast by a single step guanidinium thiocyanate/phenol-chloroform extraction using the TRIzol reagent (Invitrogen) according to the manufacturer’s protocol. One microgram of total RNA was reverse transcribed using an oligo(dT)$_{18}$ as primer (BD Biosciences) and the BD Biosciences Advantage RT kit according to the manufacturer’s instruction. Five microliters of this reaction was amplified by PCR in a Peltier Thermal Cycler apparatus (MJ Research, Watertown, MA) using Taq polymerase (Invitrogen). The primers used for RT-PCR listed in Table I were produced using an Applied Biosystems 3800 HT automated DNA synthesizer. To ensure that PCR signals were not the result of contaminating genomic DNA, control samples containing either water or RNA, in which the reverse transcriptase was omitted during cDNA synthesis, were run. For RT-PCR, the number of cycles was chosen so that amplification remained well within the linear range. An initial denaturation step at 96 °C for 2 min was followed by 25–28 amplification cycles (30 s at 96 °C, 30 s at 52 °C, and 90 s at 72 °C) and a final extension period of 7 min at 72 °C. As an internal control, primers for the detection of yeast pyruvate dehydrogenase (PDA1) were used. Aliquots of each amplification reaction were separated on 1.2% agarose gels containing ethidium bromide, and images of the gels were recorded using the UVP Gel Documentation System (UVP Inc., Upland, CA). The bands on the images were quantified using ImageQuant software (Amersham Biosciences, Version 5.2), and the densities were normalized with respect to the values obtained for PDA1.

**RESULTS AND DISCUSSION**

**MEME Analysis of 5'-UTR Regions of Genes Regulated by Crt1p—** We applied the MEME motif discovery algorithm (15) to build weight matrices modeling over-represented sequence motifs in 1-kb-long 5'-UTRs of the genes experimentally shown to be under Crt1p regulation. Various parameter configur-
tions of the algorithm were tested. Different parameter sets used for motif calculation were evaluated by the ability of the resulting weight matrices to retrieve experimentally known Crt1p binding sites (Fig. 1) in the MAST search of the data base of yeast 5'-UTRs. The rationale of using MEME rather than direct computation of weight matrices from the alignment of known binding sites is the following. First, experimental procedures used to study regulatory regions of Crt1p target genes do not guarantee that all the binding sites will be found. The MEME algorithm could, in principle, find additional Crt1p binding sites if they are present. Moreover the site of Crt1p action within the HUG1 gene has not been experimentally confirmed. Therefore, one cannot a priori exclude the possibility that the regulatory region of the HUG1 gene contains the site that would better conform to a Crt1p binding site motif than the putative site proposed by Basrai et al. (4). Second, MEME analysis would show whether the sequences analyzed contain a motif that is more conserved than the Crt1p cis-regulatory element and would give an estimate of the expected number of its occurrence in random sequences. This helps to judge the specificity of the motif.

We used MEME to find motifs ranging from 5 to 60 nucleotides in the 1-kb-long 5'-UTR sequences of RNR2, RNR3, RNR4, CRT1, and HUG1 genes. The influence of the random sequence model on the results was tested. When data set letter frequencies were used as a random sequence model, two low complexity A/T-rich regions of 27 and 26 bases were found as the most conserved regions. The Crt1p binding site motif was found as the third region with respect to E-value computed by MEME. The fact that a low complexity sequence was found as the best motif should be considered an artifact taking into account the very high frequency of consecutive A/T base occurrences in yeast 5'-UTRs.

To solve this problem we applied the fifth order Markov model of random sequence computed according to the frequencies of 1-, 2-, ..., 6-nucleotide-long words in all 5'-UTRs of yeast genes. MEME analysis with this background model was no longer biased toward A/T-rich sequence stretches, and the Crt1p binding site was the most conserved motif.

MEME uses one of the three modes of constraints on the number of motif occurrences per sequence (see “Experimental Procedures”). As there are multiple Crt1p binding sites in the regulatory regions under investigation the most reasonable choice seems to be the TCM mode allowing any number of occurrences of the motif per sequence. Surprisingly the analysis run with the assumption of exactly one occurrence of the motif per sequence (OOPS mode) produced better results than analysis run in TCM mode. When the data base of the yeast 5'-UTRs was searched with the MAST program using weight matrix calculated in MEME analysis with the TCM mode, the Crt1p gene was not retrieved. When the motif defined by MEME analysis using the OOPS mode was used in MAST search, all five known targets of Crt1p action were found. Moreover all experimentally verified sites of Crt1p action, except a weak site of the CRT1 gene, were correctly positioned by the MAST program. The CRT1 gene weak site has the lowest DNA binding affinity of all experimentally verified sites. In the regulatory region of the HUG1 gene only the putative strong site proposed in Ref. 4 was found in the MAST search.

Fig. 2 shows our statistical model of the Crt1p binding site motif. Matching sites in RNR2, RNR3, RNR4, CRT1, and HUG1 genes are shown in Table II. We conclude that the best strategy to define this motif is to perform MEME analysis in the OOPS mode using the fifth order Markov model of yeast regulatory regions as the reference state. The analysis run in the OOPS mode is forced to search for strong sites only. The resulting weight matrix is able to correctly position not only strong but also most of the weak sites when used in the MAST search of yeast regulatory regions.

**Putative Crt1p Binding Sites Are Found in Regulatory Regions of 30 Yeast Genes**—In the MAST search of yeast 5'-UTR sequences 30 genes were found that matched the best weight matrix derived in MEME analysis. In addition to the five genes known to contain the Crt1p binding site, 25 novel potential Crt1p targets were identified. The genes are listed in Table II together with their functional annotation according to the Yeast Protein Database (13) and the putative Crt1p binding sites. As shown by the ORF identifiers in the first column of

![Fig. 1. Crt1p binding sites in experimentally confirmed Crt1p targets. Strong sites are denoted XS, and weak sites are denoted XW. All sites in the RNR2, RNR3, RNR4, and CRT1 genes were subjected to in vitro binding studies by Huang et al. (3). The strong site in the HUG1 gene was proposed by Basrai et al. (4) on the basis of sequence similarity. Alignment columns are shaded on 100, 80, and 60% identity levels.](http://www.jbc.org/)

![Fig. 2. Crt1p binding site motif defined by MEME analysis. Sequence logo, computed as described under “Experimental Procedures,” and probability matrix representing the motif are shown. The last, separated column of the matrix shows data set letter frequencies. Perfect conservation of A/T carries less information than perfect conservation of G/C as A/T base pairs are much more frequent in the regulatory regions of the genes under investigation.](http://www.jbc.org/)
### Table II

Occurrence of Crt1p binding site motif in 1-kb-long 5′-UTR sequences of yeast genes

| ORF   | Gene   | Cellular role (YPD) | Site<sup>a</sup>                    | Strand | Position |
|-------|--------|---------------------|-------------------------------------|--------|----------|
| YJL026W | RNR2   | Nucleotide metabolism | TCGCCATGGCAAC                       | +1     | −428     |
|        |        |                     | TCGCATGCAAC                         | +1     | −787     |
|        |        |                     | TCCTGTGGCAAC                        | −1     | −355     |
| YIL066C | RNR3   | Nucleotide metabolism | GTCGCTGCAAC                         | −1     | −359     |
|        |        |                     | GTCGCTGCAAC                         | −1     | −369     |
| YIL064W |        | Unknown             | TCGCCATGCAAC                        | +1     | −964     |
|        |        |                     | GTCGCTGCAAC                         | +1     | −854     |
|        |        |                     | GTCGCTGCAAC                         | +1     | −915     |
| YGR180C | RNR4   | Nucleotide metabolism | TGGCCATGCAAC                        | −1     | −750     |
|        |        |                     | GTCGCTGCAAC                         | −1     | −381     |
| YDL085W | NDE2   | Energy generation   | TCCGCTGCAAC                         | −1     | −423     |
| YDL085C-A |        | Unknown             | TCCGCTGCAAC                         | +1     | −106     |
| YLR177W |        | Unknown             | TCGCCATGCAAC                        | +1     | −775     |
| YLR176C | CRT1/RFX1 | DNA repair, Polymerase II transcription | TCCGCTGCAAC | −1     | −35     |
| YOR282W |        | Unknown             | GTCGCTGCAAC                         | +1     | −902     |
| YOR280C | FSH3   | Unknown             | GTCGCTGCAAC                         | −1     | −294     |
| YLL034C |        | Protein folding, cytokinesis, membrane fusion | TCCGCTGCAAC | −1     | −171     |
| YLL033W |        | Unknown             | TCCGCTGCAAC                         | +1     | −80      |
| YPL030W |        | Unknown             | TCCGCTGAAAC                         | +1     | −976     |
|        |        |                     | TTTCGGTGGCAAC                        | −1     | −944     |
| YLR345W |        | Carbohydrate metabolism | TGGCTATGCAAC                       | −1     | −3       |
| YML059C |        | Unknown             | GTCGCTGAAAC                         | +1     | −305     |
| YML058W-A | HUG1/NORF5 | Cell cycle control   | GTCGCTGAAAC                         | −1     | −185     |
| YML058W | SML1   | Nucleotide metabolism, DNA repair, chromatin/chromosome structure | GTCGCTGAAAC | −1     | −808     |
| YBR003W | COQ1   | Other metabolism, energy generation | GTCGCTGAAAC | −1     | −668     |
|        |        |                     | ACCTATGAAAC                         | −1     | −236     |
| YBR001C | NTH2   | Carbohydrate metabolism | GTCGCTGAAAC                         | +1     | −846     |
| YOR269W | PAC1   | Mitosis, cell structure | TACGATGGCAAC                       | +1     | −358     |
| YOR268C |        | Unknown             | TACGATGGCAAC                        | −1     | −83      |
| YGL167C | PMR1   | Protein modification, small molecule transport, vesicular transport | TCCGCTGCTAC | +1     | −590     |
| YGL166W | CUP2   | Polymerase II transcription, cell stress | TCCGATGGCATC | −1     | −59      |
|        |        |                     | CTCTATGCGTCATC                     | +1     | −727     |
| YOR242C | SSP2   | Differentiation     | TACGGTGGCCAC                        | +1     | −381     |
| YLR187W |        | Unknown             | TACGGTGGCCAC                        | +1     | −603     |
|        |        |                     | GGCCTATGGCAAA                      | −1     | −500     |
| YNR049C | MSO1   | Vesicular transport | GTCGCAATTGCAAC                     | −1     | −527     |
| YIL079C | AIR1   | Protein modification | TGCGATGGCAAC                       | −1     | −380     |
| YIL078W | THS1   | Protein synthesis   | TGCGATGGCAAC                        | +1     | −102     |
|        |        |                     | TTGAATGGCAAC                        | +1     | −906     |
| YDR060W | MAK21/NOC1 | Protein synthesis, protein complex assembly | TCCGATGGCAAC | +1     | −680     |
| YDR059C | UBC5   | Protein degradation | TCCGATGGCAAC                        | −1     | −186     |

<sup>a</sup> Yeast Protein Database.

<sup>b</sup> In several cases the sites are composed of two Crt1p binding site motifs located on opposite strands and shifted by one base. In these cases the 14-nucleotide-long DNA stretch containing both motifs is shown. One of them is shown in bold, whereas the second one located on the opposite strand is underlined.
Table II in many cases the regulatory site is shared by two neighboring genes positioned on opposite strands (W and C in the ORF names are strand identifiers; in neighboring genes the numbers in ORF names differ by 1). The ORF of the HUG1 gene is so short that a putative Crt1p binding site in the regulatory region of this gene is also within 1000 bp from the gene is so short that a putative Crt1p binding site in the regulatory region of this gene is also within 1000 bp from the SML1 gene positioned downstream of HUG1 (expression of SML1 gene does not depend on the Ssn6p-Tup1p-Crt1p repressor complex (4)).

There are two putative Crt1p targets that contain sites identical to those of the genes experimentally shown to be under Crt1p regulation. The site within the regulatory region of NTH2 is identical to the putative strong site of the HUG1 gene. NDE2 contains a site identical to the experimentally verified strong binding site within the 5′-UTR of the RNR4 and RNR3 genes.

Crt1p recognizes sequences resembling palindromes. For this reason in several cases listed in Table II the sites identified are composed of two overlapping 13-nucleotide-long DNA stretches located on opposite strands and shifted by one base. For example, one of the weak binding sites of the RNR3 gene is TTGCTGTGACAA at position −369 on the minus strand. The second overlapping site of the sequence matching the Crt1p binding site motif was found by MAST at position −370 on the plus strand. Therefore, oligonucleotide TTGCTGTGACAA contains two overlapping sequences matching the Crt1p binding site motif located on opposite strands and shifted by one base.

In most of the genes listed in Table II only a single putative Crt1p binding site was detected. However, we did not reject genes that do not have multiple sites from the list of putative Crt1p targets for the following reasons. According to the mechanism of Crt1p action proposed in Ref. 3, multiple sites allow the modulation of timing and strength of the response rather than reflecting structural requirements for oligomeric protein-DNA complexes to be formed. The variable number of weak sites and wide distribution of distances between them support this assumption. The RNR3 gene contains three experimentally confirmed Crt1p binding sites, whereas RNR2, RNR4, and CRT1 have only two. The distance between the sites in RNR4 is 369 bp, while in RNR2 it is only 73 bp. Moreover the example of the CRT1 gene shows that Crt1p may bind with low affinity to the sites that are significantly different from the consensus sequence. The weak site of the CRT1 gene is well conserved only in the last five positions of the motif (Fig. 1). The nucleotides in positions 1, 3, and 8 are not present in any other experimentally confirmed Crt1p binding site. For this reason the weak site of the CRT1 gene was not found in the MAST search of yeast 5′-UTR sequences. Multiple sites were likewise not found in the regulatory region of the HUG1 gene, an experimentally verified Crt1p target. The example of the CRT1 gene weak sites shows that possible sequence diversity of weak sites is too large to include all these sites into a statistical model that would still be able to discriminate target regions of Crt1p action. Therefore, one cannot exclude the possibility that other weak sites could be present in the regulatory regions of the genes listed in Table II that were not found in our search. For example, 25 of the genes listed in Table II contained multiple copies of RRCAAC (where R is purine) sequences in their regulatory regions (data not shown). Taking into account the variability of the weak sites of the CRT1 gene (Fig. 1), there could be a potential weak site in the vicinity of most of RRCAAC sites.

Statistical Significance of the Crt1p Binding Site Motif—Analysis of the genomic distribution of DNA sequence motifs is known for its high rate of false positive hits. Therefore, one should consider whether the cluster of 30 genes obtained in our analysis is the result of the random occurrence of sequences similar to Crt1p binding sites or is a cluster of genes containing an evolutionarily conserved Crt1p binding site motif. To address this question on a statistical basis we tried to reject two null hypotheses. In both tests we eliminated, from the cluster of 30 genes, one of each pair of genes that have overlapping regulatory regions. It was necessary to assure that each regulatory region contributes equally to the final result of statistical analysis. The resulting data set contained 18 sequences.

The first null hypothesis considered states that the set of 18 random, 1-kb-long sequences would contain a 13-nucleotide-long motif that is equally or better conserved than the Crt1p binding site motif found in the cluster of 18 genes. To test this hypothesis the MEME program (with the same parameters as for the calculation of final motif representation) was run on the set of 18 regulatory regions. The Crt1p cis-regulatory element was again found as the most conserved motif with an E-value of 3.5 × 10^{-29}. The E-value calculated by MEME is the expected number of occurrences of equally or better conserved motifs in the similarly sized training set of random sequences. Therefore, the value of 3.5 × 10^{-29} allows us to reject this hypothesis with very high confidence.

The second null hypothesis states that the cluster of 18 randomly chosen yeast regulatory regions would contain a motif of 5–60 bp that would be equally or better conserved than the Crt1p binding site. The purpose of this analysis was to directly test how likely it is that a random group of yeast genes contains a “stronger” signal for the regulatory factor than the cluster of genes studied here. To this end 1000 random clusters of 18 yeast 5′-UTR regions were generated. In every cluster the most conserved motifs of 6–50 bp were found by the MEME method. In none of the experiments was a motif with an E-value lower than 3.5 × 10^{-29} found. The smallest E-value observed for random clusters was 3.5 × 10^{-10}. Therefore, it is very unlikely that the random cluster of yeast regulatory regions would contain a more conserved motif than the Crt1p binding site. In other words, the probability that MEME analysis would find a motif of better quality than the Crt1p binding site in a random set of yeast regulatory regions is low. We conclude that it is very unlikely that either the cluster of randomly generated 1-kb-long DNA sequences or the cluster of randomly picked yeast 5′-UTR regions would contain a regulatory signal of comparable strength (degree of conservation) to the Crt1p binding site motif found in the genes listed in Table II. Analysis of Expression Profiles of Putative Crt1p Target Genes—We used publicly available yeast genomic microarray data to examine whether the genes listed in Table II show expression profiles characteristic of genes regulated by the Mec1p-Rad53p-Dun1p-Crt1p pathway. The data available permitted us to check which of the candidate genes are significantly induced in response to DNA damage in cells treated with MMS or exposed to ionizing radiation. We could also examine the response of candidate genes to inactivation of the components of the Tup1p-Ssn6p-Crt1p repressor complex and to deficiency of the Mec1p-Rad53p-Dun1p signal transduction pathway. Therefore, we examined whether the genes with putative Crt1p binding sites show the following alterations in expression profiles: (i) a significant induction as the result of treatment with DNA-damaging agents, (ii) a decrease in the magnitude of this induction in strains with inactivated MEC1 or DUN1 genes, and (iii) a significant induction in strains with inactivated Ssn6p, Tup1p, and CRT1 genes.

Fig. 3 shows the response of several genes listed in Table II to treatment with 0.02% MMS as measured in Ref. 9. Only
genes that exceeded the expression ratio of 2 in at least one experiment (significance criterion used by the authors) are shown. As can be seen, the ribonucleotide reductase genes RNR2 and RNR4 show a very characteristic response (data for RNR3 are missing in the data set). In the MMS-treated wild-type cells the genes are gradually induced, reaching expression levels about 8 times higher than in untreated control cells. The magnitude of induction is much lower in mec1Δ and dun1Δ strains.

Four of the putative Crt1p targets listed in Table I show significant response to MMS treatment according to Ref. 9. They include FSH3, YLR345W, NTH2, and UBC5. As seen in Fig. 3, FSH3, NTH2, and UBC5 show a gradual induction similar to that observed with the ribonucleotide reductase genes. In the

![Figure 3](http://www.jbc.org/)
Novel Crt1-regulated Genes

The data discussed above are summarized in Table III. The genes FSH3, YLR345W, NTH2, and UBC5 are likely targets of Crt1p regulation, supported by several microarray experiments. The genes listed in the bottom row of Table III show significant induction only in a single microarray data set. Among these genes, NDE2 may still be considered as a likely Crt1p target because its 5′-UTR contains a site identical to the RNR4 strong site.

Experimental Verification of Putative Crt1p Targets—To provide experimental evidence corroborating our theoretical predictions, the RT-PCR method was used to compare transcript levels of the most likely Crt1p targets in crt1Δ and wild-type strains. Genes regulated by the Crt1 repressor should exhibit induction (derepression) in the crt1Δ strain relative to the wild-type strain. Fig. 5 shows results of this approach for the five most likely Crt1p targets, FSH3, YLR345W, NTH2, NDE2, and UBC5. The genes CRT1, RNR2, RNR3, and RNR4 were used as a positive control, and constitutively expressed PDA1 was used as an example of a gene obviously not regulated in a Crt1p-dependent manner.

Our results indicated that FSH3, YLR345W, and NTH2 are indeed regulated by Crt1p because their transcription was significantly induced in the untreated crt1Δ strain relative to the untreated wild-type control. One should also note that in the case of YLR345W and NTH2 MMS treatment of crt1Δ strain significantly (p = 0.012 and 0.002) increased expression with respect to untreated crt1Δ. Moreover induction of YLR345W and NTH2 resulting from MMS treatment of the wild-type strain was higher than that caused by Crt1 deletion alone. This effect was statistically significant for NTH2 (p = 0.003) and was close to the 5% limit of statistical significance in the case of YLR345W (p = 0.09). These observations suggested that the response of both genes to MMS treatment is regulated not only via Crt1 inactivation but also by an additional mechanism. This scenario is likely because we found stress response elements in the 5′-UTR of both YLR345W and NTH2 (two AGGGG sequences at positions −154 and −44) and NTH2 (AGGGG at position −688). The stress response element motifs, present in genes regulated in regions, were shown to recruit transcription activators such as Msn2 and Msn4 to a large number of yeast multistress response genes (20, 21). Involvement of YLR345W in multistress response is further supported by the fact that the microarray experiments (22, 23) indicate that expression of this gene is induced by a variety of known environmental and metabolic stressors (H2O2, menadione, diamide, nitrogen starvation, and growth in stationary phase). Causton et al. (22) assigned YLR345W to the cluster of Common Environmental Response genes. Experiments published in Ref. 23 also indicate significant induction of NTH2 by both environmental and metabolic stress factors. Moreover there is direct experimental evidence that this gene is required for the heat shock recovery and induced by temperature and the presence of toxic chemicals (24).

According to our results the genes NDE2 and UBC5 were induced by MMS treatment. To the best of our knowledge this is the first experimental evidence, other than microarray data, indicating that these genes are activated by MMS treatment. However, both genes clearly did not require the Crt1 transcription factor for MMS response. Their expression levels did not significantly differ between wild-type and crt1Δ strains (p > 0.05). Therefore, NDE2 and UBC5 cannot be considered as genes regulated by the Crt1 transcription factor.
Comparison with Genome-wide Location Studies—Crt1p was one of the 203 yeast transcription factors that have been subjected to genome-wide location studies (25, 26). From the genes, which were experimentally shown to be regulated by Crt1p, only RNR2, RNR3, RNR4, and FSH3 were detected as statistically significant Crt1p targets in the data set of Lee et al. (25). Other genes, for which regulation by Crt1p was experimentally confirmed by us and other authors (HUG1, RNR2, CRT1, NTH2, and YLR345W), have not been detected even when a very high significance threshold ($p = 0.05$) was used. A recent genome-wide location experiment (26), published when this manuscript was in the final stages of revision, includes RNR2, RNR3, RNR4, and FSH3, although the significance of the result obtained for RNR2 ($p = 0.002$) is still formally above the threshold of 0.001 used by the authors. According to the recent results (26) Crt1p binds in the regulatory region of PAC1; this gene was found in our MAST searches (see Table I). However, the PAC1 gene did not exhibit an expression pattern characteristic of the genes regulated by Crt1p and was not examined in our experimental studies. Another interesting case is YMR279C, which shows significant response to MMS, γ radiation, and CRT1 deletion in microarray experiments. However, we could not find a Crt1p binding site motif in the regulatory region of this gene. In total, the genome-wide location experiment detected 21 Crt1p targets ($p < 0.001$) for which the Crt1p binding site motif could not be found by the methods used in our work.

Our theoretical and experimental studies, performed independently of genome-wide location data, resulted in detection of three new genes regulated by Crt1p. Experimental results presented in this work confirm the influence of Crt1p on the transcription of the genes under investigation. The chromatin immunoprecipitation on chip data estimates the probability of its binding within the gene regulatory region. Moreover genome-wide location data sets did not list YLR345W and NTH2 as significant Crt1p targets.

**FSH3 Encodes an Evolutionary Conserved Domain Implicated in Folate Metabolism**—The FSH3 open reading frame encodes a 228-amino acid-long protein. According to BLAST searches of the Swiss-Prot/TREMBL data bases, the yeast genome contains two paralogous proteins (YMR229Cp, 47% identity, E-value = $2 \times 10^{-53}$, YHR049Wp, 28% identity, E-value = $5 \times 10^{-18}$). All three proteins have been shown to belong to a novel serine hydrolase family on the basis of their chemical reactivity and the geometrical analysis of their active site in structural models (25). Fsh3p is significantly similar to the N-terminal part of the Schizosaccharomyces pombe DFR1p. The alignment found by BLAST (E-value = $2 \times 10^{-25}$, 33% identity) contains residues 1–230 in Fsh3p and 1–236 in DFR1p. The C-terminal part of DFR1p encodes the catalytic domain with dihydrofolate reductase (DHFR) activity (26). The BLAST search with Fsh3p identified similar sequences also in other eukaryotic genomes. Proteins about 200 amino acids that are significantly (E-value < $10^{-5}$) similar to the product of the FSH3 gene are found in Caenorhabditis elegans, Drosophila melanogaster, Arabidopsis thaliana, and Homo sapiens. The human homologue OVC2 was found by genetic studies to be a tumor suppressor gene implicated in ovarian cancer (27, 28).

It was frequently observed that the proteins that act in subsequent steps of metabolic pathways or form non-covalent complexes in certain organisms are encoded as single polypeptide chains in other genomes. On this basis, the so-called Rosetta stone methodology (29) for detecting protein-protein interactions and functional relationships was formulated. DHFR is one of the archetypal Rosetta stone proteins since catalytic domains carrying this activity are frequently connected with thymidylate synthase, an enzyme catalyzing a subsequent step of the metabolic pathway. Therefore, the presence of a protein domain significantly similar to Fsh3p in the same polypeptide chain as the DHFR domain in the genome of S. pombe indicates that FSH3 encodes a protein that interacts with DHFR of S. cerevisiae or takes part in a novel metabolic pathway requiring coordination of serine hydrolyase and DHFR enzymatic activities.

DHFRs reduce dihydrofolate to tetrahydrofolate (30). Tetrahydrofolate is a donor of a one-carbon group in purine ring synthesis. It is also converted to N$^{10}$-methylene tetrahydrofolate, which in turn takes part in deoxythymidylate synthesis as a methyl group donor in the methylation of dUTP catalyzed by thymidylate synthase. Therefore, our findings imply that the Mec1p-Rad53p-Dun1p-Crt1p pathway may regulate the deoxyribonucleotide pool not only at the level of ribonucleotide reductase activity but also at the level of DHFR. An attractive scenario is that in response to DNA damage, signal activation of the pathway results in derepression of the FSH3 gene and that the activity of Fsh3p leads to an increase in tetrahydrofolate and N$^{10}$-methylene tetrahydrofolate pools enhancing deoxyribonucleotide synthesis. If this assumption is correct, the Mec1p-dependent pathway regulates the deoxyribonucleotide pool by influencing two key metabolic processes, ribonucleotide reduction and one-carbon pool synthesis.

**YLR345W and NTH2 Encode Enzymes Involved in the Regulation of Central Metabolism**—There is no experimental evidence concerning the function of the YLR345W gene. However, the YLR345Wp sequence is significantly similar to that of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase enzymes (EC 2.7.1.105/EC 3.1.3.46). The most similar enzyme of this group, whose enzymatic activity has been experimentally confirmed (31), is the product of the rat PFKFB4 gene (BLAST
Alignment of the YLR345W and PFKFB4 gene products covers both functional domains required for enzymatic activity.

The enzymatic activity of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase is responsible for the phosphorylation of fructose 6-phosphate to fructose 2,6-bisphosphate and dephosphorylation of fructose 2,6-bisphosphate. Fructose 2,6-bisphosphate is an activator of 6-phosphofructo-1-kinase. In mammals it is one of the major signaling molecules in the regulation of the glycolytic and gluconeogenic pathways (30). It is therefore
likely that YLR345W encodes an isoform of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, an enzyme taking part in the regulation of glycolysis and gluconeogenesis under DNA damage conditions.

The product of NTH2 gene is involved in the metabolism of trehalose, a crucial protector of protein and membranes against a variety of stresses, including heat, cold, starvation, and exposure to toxic substance. The NTH1 and NTH2 genes encode neutral trehalases, enzymes responsible for the recycling of trehalose to glucose. Activation of trehalase immediately after cells are subjected to the stress factor seems to be counterintuitive as high trehalose concentrations are observed under stress conditions. However, this mechanism of NTH1 and NTH2 gene regulation has been experimentally confirmed. Recent kinetic studies have shown that genes encoding trehalases should be induced at early stages of the stress response to allow quick return of trehalose concentration to the low level after the stress ceases (32). Rapid trehalose degradation is essential for stress recovery because the presence of high amounts of this metabolite, although advantageous under stress conditions, severely disrupts normal physiological functions (7).

Both YLR345W and NTH2 encode enzymes involved in the regulation of central metabolism under stress conditions. Regulation of these enzymes by Crt1p may help the cell to adjust its central metabolism to the biosynthetic demands of intensive DNA repair processes and survival under the presence of DNA-damaging factors.

Conclusions—We used the 5'-UTR sequences of the known regulatory targets of the Crt1 transcription factor to determine the weight matrix representing its binding site sequence motif. Using this weight matrix, a statistically significant cluster of 30 putative Crt1p targets was identified. Yeast genomic microarray data provided supporting evidence that five candidate genes, FSH3, UBC5, NTH2, NDE2, and YLR345W, are subject to regulation by Crt1p. We showed by RT-PCR comparison of transcript levels in crt1Δ and wild-type strains of S. cerevisiae that FSH3, YLR345W, and NTH2 are indeed under the control of Crt1p. It is also likely that YLR345W and NTH2 are under the control of the stress response element regulatory sequence motif, constituting an example of the cross-talk between the DNA damage and general stress response pathways.

Sequence analysis of FSH3 indicated that this protein may be involved in folate metabolism either by carrying serine hydrolase activity required for a novel metabolic pathway involving DHFR or by directly interacting with DHFR. If this hypothesis is correct Crt1p regulates synthesis of the dNTP pool not only at the level of ribonucleotide reduction but also at the level of purine and pyrimidine ring synthesis. These findings suggest a direction for further functional studies of a large family of eukaryotic proteins involving the human tumor suppressor gene OVCA2. YLR345Wp shares significant sequence similarity with 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase enzymes responsible for the metabolism of fructoso-2,6-bisphosphate, an important signaling molecule responsible for the regulation of glycolysis and gluconeogenesis. NTH2 encodes neutral trehalase involved in the precise regulation of trehalose concentration. We postulate that both NTH2p and YLR345Wp are enzyme isoforms responsible for alteration of central metabolism during the stress response.

The fact that Crt1p regulates the ribonucleotide reductase genes suggests that this effector of the DNA damage checkpoint pathway is responsible for the adjustment of the biosynthetic capabilities of the cell for intensive DNA repair processes. The putative functions of the three new target genes discovered here corroborate this view emphasizing the influence of the DNA damage checkpoint pathway on folate and sugar metabolism. These findings should help to establish the complex wiring of the network of molecular interactions responsible for DNA damage response in eukaryotic cells.

Acknowledgments—We are grateful to Dr. Zygmunt Ciesla for bringing the subject of the DNA damage checkpoint pathway to our attention and for critical reading of the manuscript. We are also grateful to Dr. Marta Fikus for many useful discussions. We are also grateful to Dr. Anne Lise Haenni for critical reading of the manuscript.

REFERENCES

1. Zhou, B. B., and Elledge, S. J. (2000) Nature 408, 433–439
2. Wahl, G. M., and Carr, A. M. (2001) Nat. Cell Biol. 3, E277–E286
3. Huang, M., Zhou, Z., and Elledge, S. J. (1998) Cell 94, 595–605
4. Basrai, M. A, Velcrocius, V. E., Kinzler, K. W., and Hieter, P. (1999) Mol. Cell Biol. 19, 7941–7949
5. Reith, W., Herrero-Sánchez, C., Kobr, M., Silacci, P., Berte, C., Barras, E., Fey, S., and Mach, B. (1999) Genes Dev. 13, 1528–1540
6. Durand, B., Kobr, M., Reith, W., and Mach, B. (1994) Mol. Cell. Biol. 14, 6839–6847
7. Sebilla, A., Louzada, P. O., Sola-Penna, M., Sarone-Williams, V., Coelho-Sampaio, T., Sérgio, T., and Ferreira, S. T. (2004) Int. J. Biochem. Cell Biol. 36, 900–908
8. Ball, C. A., Dolinski, K., Dwight, S. S., Harris, M. A., Issel-Tarver, L., Kasarskis, A., Scafe, C. R., Sherlock, G., Binkley, G., Jin, H., Kaloper, M., Orr, S. D., Schroeder, M., Weng, S., Zha, Y., Botstein, D., and Cherry, J. M. (2000) Nucleic Acids Res. 28, 77–80
9. Gasch, A. P., Huang, M., Metzenr, S., Botstein, D., Elledge, S. J., and Brown, P. O. (2001) Mol. Cell Biol. 12, 2987–3003
10. Polisky, S. A., Estep, P., Church, G. M., and Samson, L. D. (2000) Mol. Cell. Biol. 20, 8157–8167
11. Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P. Y., Stepaniants, S. B., Shoemaker, D. D., Gachotte, D., Kahara, B., Kemon, N., Sabatier, P., and Rouns, S. H. (2000) Cell 102, 109–126
12. Le Cron, S., Devaux, P., Jaqc, C., and Marc, P. (2002) Nucleic Acids Res. 30, 76–79
13. Costancz, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Roberton, L. S., Skryapke, M. S., Braun, B. R., Hopkins, K. L., Kondo, P., Lengieza, C., Lew-Smith, J. E., Tillberg, M., and Garrelis, J. I. (2001) Nucleic Acids Res. 29, 75–79
14. Bairoch, A., and Apweiler, R. (2000) Nucleic Acids Res. 28, 45–48
15. Bailey, T. L., and Elkan, C. (1994) Proceeding of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28–36, AAAI Press, Menlo Park, CA
16. Lee, P., Longden, J., and Bleasby, A. (2000) Trends Genet. 16, 276–277
17. Darbin, R., Eddy, S., Kragh, A., and Mitchinson, G. (1998) Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, p. 308, Cambridge University Press, Cambridge, UK
18. Schonheit, P. D., and Stepphan, S. J. (1990) Nucleic Acids Res. 18, 6097–6100
19. Mercier, G., Denis, Y., Marc, P., Picard, L., and Dutreux, M. (2001) Mutat. Res. 487, 157–172
20. Schmitt, A. P., and Meenzee, K. (1996) Proc. Natl. Acad. Sci. U.S.A. 93, 5777–5782
21. Moskvin, E., Schuller, C., Maurer, C. T., Mager, W. H., and Ruis, H. (1998) Yeast 14, 1041–1450
22. Causton, H. C., Res, R., Koh, S. H., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Landers, E. S., and Young, R. A. (2001) Mol. Cell Biol. 13, 323–337
23. Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz G., Botstein, D., and Brown, P. O. (2000) Mol. Cell Biol. 21, 4241–4257
24. Nwaka, S., and Holzer, H. (1998) Prog. Nucleic Acids Res. Mol. Biol. 58, 197–237
25. Baxter, S. M., Vanderree, L., Berman, D. B., Hamilton, T. C., Wong, A. J., and Godwin, A. K. (1996) Cancer Res. 56, 1997–2002
26. Jensen, M. R., and Helin, K. (2004) Genes Dev. 18, 245–248
27. Marcaro, K. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yvates, T. O., and Eisenberg, D. (1999) Science 285, 751–753
28. Stryer, L. (1998) Biochemistry, 4th Ed., p. 748, W. H. Freeman and Co., New York
29. Kitamura, K., Uyeda, K., Hartman, F. C., Kangawa, K., and Matsuo, H. (1989) J. Biol. Chem. 264, 6344–6348
30. Voit, E. O. (2003) J. Theor. Biol. 223, 55–78
