The yeast *ADR6* gene encodes homopolymeric amino acid sequences and a potential metal-binding domain

Patrick J.O'Hara, Heidi Horowitz¹, Greg Eichinger and Elton T.Young¹

ZymoGenetics, Inc., 4225 Roosevelt Way NE, Seattle, WA 98105 and ¹Department of Biochemistry, SJ-70, University of Washington, Seattle, WA 98195, USA

Received August 1, 1988; Revised and Accepted October 12, 1988

Accession no. X12493

ABSTRACT

The *ADR6* gene of *Saccharomyces cerevisiae* has an open reading frame which could encode a polypeptide of 1314 amino acids. The predicted mRNA encodes a protein with homopolymeric stretches of asparagine and threonine, particularly near its amino terminus and contains additional sequences consisting of polyglutamine repeats. The predicted protein also contains a potential metal binding (Cys)₄-type finger near its carboxy-terminus. An *ADR6*/β-galactosidase fusion protein was predominantly nuclear in location, consistent with its role as an activator of *ADH2* transcription.

INTRODUCTION

The alcohol dehydrogenase (ADH) II isozyme (encoded by the *ADH2* gene) of *Saccharomyces cerevisiae* is regulated by carbon catabolite levels. In the presence of fermentable carbon sources such as glucose, ADH II is strongly repressed; in media containing only non-fermentable carbon sources such as ethanol, ADH II activity is markedly increased. The regulation of *ADH2* expression at the level of transcription (1; Yu and Young, submitted) is mediated at least in part by the *ADR1* gene product. The *ADR1* gene (2) encodes a transcriptional activator (3) whose DNA binding site is a palindromic sequence in the *ADH2* promoter (Eisen et al., in press). The *ADR1* gene product contains metal-binding, nucleic acid-binding finger sequences similar to those in the xenopus transcription factor TFIIIA (4; Eisen et al., in press). Recently, other functional regions of *ADR1* have been identified (5).

Another gene involved in the expression of *ADH2* is *ADR6* (6, 7). This gene appears to encode a second transcription factor whose site of action is unknown. Here, we report the primary structure and potential gene product of *ADR6*. The gene contains an open reading frame which can encode a protein which has a potential metal-

binding finger domain near its carboxy terminus. The open reading frame also encodes several regions of homopolymeric amino acid sequences similar to those found in a number of other nucleic acid-binding proteins, such as the Engrailed gene product of Drosophila (8). Consistent with its proposed role as an activator of ADH2 transcription, an ADR6/β-galactosidase protein was localized to the nucleus of yeast cells.

MATERIALS AND METHODS

DNA Sequencing

Restriction fragments of the plasmid YRp7-1, containing ADR6 sequences were subcloned into phage M13 mp18 and mp19 for DNA sequence determination. Sequencing was done by the dideoxy chain termination method (9) from overlapping restriction fragments, using ³⁵S-dATP from NEN (10). Sequencing reactions were catalyzed by the Klenow fragment of E. coli DNA polymerase I (Pharmacia) or modified T7 DNA polymerase (U.S. Biochemicals) (11) and were primed with a universal M13 primer or with oligonucleotides complementary to the ADR6 region which were synthesized on an Applied Biosystems (Foster City, CA) model 380A DNA synthesizer.

Computer Analysis

The DNA sequence of ADR6 and the predicted ADR6 gene product sequence were analyzed on Sun 100 and Compaq 286 computers using Intelligenetics (Mountain View, CA) software and homology search programs by Lipman and Pearson (12) and Gribskov et al. (13). The sequences were compared to the GenBank (14) and EMBL (15) nucleotide sequence databanks and the PIR protein identification resource protein databank (16).

5' and 3' analysis of ADR6 mRNA

The 5' and 3' ends of the ADR6 mRNA were localized using the S1 nuclease-mapping technique of Berk and Sharp (17). Total yeast RNA was isolated from the adr6-1 mutant strain, m4:5b (6) which had been transformed with an autonomously replicating plasmid (YRp7-1) (7) bearing the wild-type ADR6 gene. Optimal hybridization of the various DNA probes to the yeast RNA was achieved by varying the temperature. S1 nuclease digestions (1000-1500 units/ml) were performed at 37°C; S1 nuclease was from Boehringer Mannheim.

ADR6-lacZ Gene Fusion

Plasmid YRp7-1 was digested with Bam HI which cuts the plasmid

uniquely at nucleotide 3408 in the ADR6 gene. This vector was ligated with a Bam HI fragment from pMC1871 (18) which contains the lacZ gene. As predicted from the ADR6 nucleotide sequence, this lacZ fragment was translated in the same reading frame as ADR6 as indicated by the β -galactosidase activity present in both E. coli and yeast transformants.

Indirect Immunofluorescence

The procedure utilized for indirect immunofluorescence has been described (19). Five to 10 mls of cells were grown to late logarithmic phase, harvested, fixed with formaldehyde and spheroplasted. Spheroplasts were affixed to the polylysine-treated wells of toxoplasmosis slides (Bellco Glass) and then incubated with 10 μ l of a 1/10 dilution of rabbit anti- β -galactosidase antibody (Cappel) that had been affinity-purified using an Affi-Gel 10 resin (Biorad) coupled to E. coli β -galactosidase (Sigma). After 90 min. of incubation at room temperature, slides were washed and 10 μ l of a 1:100 dilution of FITC-conjugated goat anti-rabbit IgG antibody was added. Following 90 min. of incubation at RT, slides were washed and 20 μ l of 1 mg/ml DAPI was added to each well for 1 min. Slides were washed and mounted as described (19) and viewed under a Nikon fluorescent microscope using a 100x objective. Photomicroscopy was performed using a Nikon FX-35A camera and Kodak black and white TX 135-36 film, ASA 400.

RESULTS

DNA Sequencing

The cloning and characterization of the plasmid YRp7-1, containing the ADR6 region was described earlier (6). Here we report the determination of 4,830 nucleotides of sequence from YRp7-1 encompassing the putative ADR6 coding region. Overall, the sequence is 65% A+T, which is approximately 5% greater than average for yeast genes. The sequence contains an open reading frame 3.94 kb in length (Figure 1). The first methionine codon in this reading frame is preceded by a sequence which is a four out of six match with the yeast optimal AUG context consensus sequence ([A/T]A[A/C]A[A/C]A) (20), including the particularly conserved A at position -3. The sequence upstream of the open reading frame is A-T rich, as is typical of yeast genes. The codon usage in the open frame closely follows the relative degrees of usage observed in yeast genes expressed

Figure 1. The nucleotide sequence of the ADR6 gene and the predicted amino acid sequence of the ADR6 gene product. Nucleotide position numbers are given in the left column and amino acid positions are given in the right column. The amino acid translation is given in the single-letter code and is shown under the nucleotide sequence. Homopolymeric or repetitive amino acid sequences are underlined. A putative Cys₄ zinc finger motif is double underlined. The experimentally determined range of possible 5' and 3' end(s) of the message are overlined between asterisks.

CODON USAGE							
No. %		No. %		No. %		No. %	
TTT-Phe	32 2.4	TCT-Ser	36 2.7	TAT-Tyr	17 1.3	TGT-Cys	5 .4
TTC-Phe	22 1.7	TCC-Ser	20 1.5	TAC-Tyr	14 1.1	TGC-Cys	3 .2
TTA-Leu	44 3.3	TCA-Ser	28 2.1	TAA-	0 .0	TGA-	1 .1
TTG-Leu	44 3.3	TCG-Ser	11 .8	TAG-	0 .0	TGG-Trp	5 .4
CTT-Leu	13 1.0	CCT-Pro	18 1.4	CAT-His	8 .6	CGT-Arg	4 .3
CTC-Leu	9 .7	CCC-Pro	10 .8	CAC-His	5 .4	CGC-Arg	2 .2
CTA-Leu	18 1.4	CCA-Pro	22 1.7	CAA-Gln	71 5.4	CGA-Arg	3 .2
CTG-Leu	12 .9	CCG-Pro	13 1.0	CAG-Gln	40 3.0	CGG-Arg	1 .1
ATT-Ile	37 2.8	ACT-Thr	31 2.4	AAT-Asn	131 10.0	AGT-Ser	24 1.8
ATC-Ile	23 1.7	ACC-Thr	18 1.4	AAC-Asn	83 6.3	AGC-Ser	13 1.0
ATA-Ile	26 2.0	ACA-Thr	20 1.5	AAA-Lys	55 4.2	AGA-Arg	18 1.4
ATG-MET	17 1.3	ACG-Thr	6 .5	AAG-Lys	24 1.8	AGG-Arg	8 .6
GTT-Val	23 1.7	GCT-Ala	24 1.8	GAT-Asp	37 2.8	GGT-Gly	13 1.0
GTC-Val	9 .7	GCC-Ala	13 1.0	GAC-Asp	15 1.1	GGC-Gly	8 .6
GTA-Val	11 .8	GCA-Ala	20 1.5	GAA-Glu	39 3.0	GGA-Gly	10 .8
GTG-Val	8 .6	GCG-Ala	6 .5	GAG-Glu	12 .9	GGG-Gly	2 .2

AMINO ACID COMPOSITION

Residue	No.	%	d%
Asparagine	214	16.3	+200
Leucine	140	10.7	+ 10
Serine	132	10.0	+ 40
Glutamine	111	8.4	+ 70
Isoleucine	86	6.5	0
Lysine	79	6.0	- 20
Threonine	75	5.7	0
Alanine	63	4.8	- 30
Proline	63	4.8	0
Phenylalanine	54	4.1	0
Aspartic Acid	52	4.0	- 30
Glutamic Acid	51	3.9	- 40
Valine	51	3.9	- 60
Arginine	36	2.7	- 30
Glycine	33	2.5	- 60
Tyrosine	31	2.4	- 40
Methionine	17	1.3	- 50
Histidine	13	1.0	- 50
Cysteine	8	.6	- 40
Tryptophan	5	.4	- 50

Figure 2. Codon usage of the predicted ADR6 mRNA coding region and the amino acid composition of the predicted protein. The codon usage is similar to that found for other yeast genes expressed at a low level, except for the polyasparagine-polythreonine sequences near the amino terminus. The predicted ADR6 protein has an unusual amino acid composition. The percent difference (d%) between the occurrence of each amino acid in ADR6 and all of the yeast proteins in the PIR database is shown. Asparagine and glutamine are markedly elevated.

at a low level (21) (Figure 2). The open reading frame ends with a TGA at position 4626. Nucleotides 4634-4672 contain the elements of the [(T-rich)-TAG-TA(T)GT-(AT-rich)-TTT] consensus sequence proposed by Zaret and Sherman (22) to be involved in transcription termination in yeast.

The DNA sequence in the open frame contains several repetitive regions sequence similar to that observed in Trypanosomal (23, 24), Plasmodial (25, 26), and yeast mitochondrial genes (27-32). As will be discussed below, these regions encode unusual polypeptide sequences.

If translation starts at the first Met codon in the open reading frame, the nucleotide sequence predicts a 1314 residue protein with a molecular weight of 148 kilodaltons. The amino acid composition predicted is rather unusual (Figure 2). Most amino acids in the ADR6 gene product appear less often than in the average yeast gene product in the PIR database. This is particularly true for glycine and valine, which are down 62% and 56%, respectively. Four amino acids appear more often than usual: asparagine, glutamine, serine and leucine. Asparagine and glutamine appear 200% and 70%, respectively, more often than average. Taken together, asparagine, glutamine, serine and leucine make up nearly half of the protein; the combined average for these four amino acids in yeast proteins in the PIR database is 25%. In addition to the unusual amino acid composition, the open reading frame predicts a skewed amino acid distribution, with extensive clustering of identical residues in specific subregions. Several areas are characterized by stretches of homopolymeric sequences, particularly polyasparagine, polythreonine, and polyglutamine. In addition, there are clusters of sequence made up principally of only two or three amino acids, e.g., alanine, serine and proline, or glutamine and lysine.

Reflecting the lower than average abundance of charged amino acids in the predicted protein, there are several stretches of sequence which have an overall hydrophobic character. Five of these regions are predicted to contain alpha helices, most of which have a hydrophobic moment characteristic of globular protein helices (33). The predicted protein does not contain a secretory signal sequence, consistent with nuclear localization (see below). Near the carboxyl terminus of the predicted protein, there is a sequence containing four cysteines which is similar to some predicted metal-binding motifs (34).

5' end determination

A 941 base-pair DNA fragment of ADR6 extending from the Pst I

Figure 3. *ADR6* mRNA 5' end mapping with a 941 bp *Pst* I (position 2, Figure 1)-*Hinf* I (position 943) probe. The probe was 5'-end-labelled at the *Hinf* I site, hybridized to RNA from m4:5b/YRp7-1 (6), digested with S1 nuclease, and electrophoresed on a 4.25% denaturing polyacrylamide gel. Lanes: 1: 944 bp probe 2: probe hybridized to tRNA and digested with S1 nuclease (control showing no protection); 3: probe hybridized to m4:5b/YRp7-1 RNA and digested with S1 nuclease (shows protection of 640 bp fragment). The position and sizes of marker fragments are indicated to the right.

site at position two of the nucleotide sequence to the *Hinf* I site at position 943 was radioactively labelled at the *Hinf* I 5' end with T4 polynucleotide kinase and γ -ATP (virtually no labelling occurred at the recessed 5' end of the *Pst* I site). This probe was hybridized at 42°C to RNA obtained from the strain m4:5b/YRp7-1, digested with S1 nuclease, and subjected to electrophoresis on a denaturing polyacrylamide gel alongside marker DNA fragments of known size. A single S1 nuclease digestion product was observed. Its mobility corresponds to a DNA fragment approximately 640 nucleotides long, placing the 5' end of the *ADR6* message at approximately nucleotide 305 of the sequence (Figure 3). In order to obtain a more accurate estimate of the 5' end of the message, a shorter probe, which extended from the *Pst* I site at position 2 of the sequence to a 5' end-labelled *Xba* I site at position 433 was utilized. Upon S1 nuclease digestion, a single broad signal was obtained, corresponding to fragments ranging in size from approximately 133 to 146 nucleotides in length. A similar sequence has been observed using RNA derived from a strain carrying only the chromosomal copy of the *ADR6*

gene. This finding places the 5' end(s) of the ADR6 message between positions 287 and 300 of the sequence, in good agreement with the results obtained from the Hinf I probe (data not shown).

The deduced 5' end for ADR6 mRNA corresponds to an untranslated leader sequence nearly 400 nucleotides in length. The first ATG encountered in the RNA is the one predicted to be utilized as the start codon for the translated portion of the gene. There is no TATA sequence within 300 nucleotides 5' of the RNA start site but several sequences similar to canonical yeast TATA boxes (35, 36) can be found.

3' end mapping

A Dde I fragment extending from position 4125 in the ADR6 sequence to a second Dde I site approximately 1100 base pairs promoter distal to this was 3' end-filled with $\alpha^{32}\text{P}$ -dATP, non-radioactive dTTP and the Klenow fragment of DNA polymerase I. This served to radioactively label only the Dde I site (CTNAG) at position 4125, where the central base is a T. Hybridization of this probe to m4:5b/YRp7-1 RNA at 37°C, followed by S1 nuclease analysis revealed the presence of four bands of 0.52, 0.54, 0.55 and 0.63 kb in length (Figure 4). These signals correspond to a series of 3' ends for the ADR6 message within the region extending approximately from 4648 to 4754 of the sequence. A similar pattern of S1 signals derived from DNA fragments of appropriately increased length was observed when a fragment end-labelled at the Eco RI site (position 3763 of the sequence) was used as a probe (Figure 4). These findings place the 3' ends of the ADR6 message between approximately 20 and 126 nucleotides beyond the predicted stop codon of the ADR6 gene.

The deduced 5' and 3' ends of the ADR6 message should be considered to be putative, as they were determined using a strain harboring a multicopy ADR6 plasmid. However, it should be noted that a similar start site for the message has been identified in a strain lacking any extrachromosomal copies of ADR6. Furthermore, the deduced 5' and 3' ends of the ADR6 message correspond to a transcript approximately 4.4 kb long. This size is in good agreement with the length of the chromosomally-derived ADR6 message previously estimated by Northern analysis (7).

Figure 4. Mapping the 3' ends of ADR6 mRNA. The 1.1 Kb Dde I probe extends from the Dde I site at position 4125 (Figure 1) to a second Dde I site approximately 0.4 Kb beyond the end of the DNA sequence shown in Figure 1. The Eco RI probe extends from the Eco RI site at position 3763 of the ADR6 sequence through 2.6 Kb of ADR6-specific DNA in plasmid YRp7-1 (including approximately 1.5 Kb of DNA beyond the sequence in Figure 1) to an Eco RI site in pBR 322 (6.6 Kb total size). The probes were 3' end labeled, hybridized to the indicated RNA, digested with S1 nuclease and electrophoresed on a 4% denaturing polyacrylamide gel. 1: Dde I probe untreated; 2: Dde I probe hybridized to tRNA and digested with S1 nuclease (no S1-specific signal produced); 3: Dde I probe hybridized to m4:5b/YRp7-1 RNA (6) and digested with S1 nuclease. Four S1-specific signals are produced, ranging from 0.52 to 0.63 Kb in length; 4: Eco RI probe hybridized to tRNA and digested with S1 nuclease (no S1-specific signals observed); 5: Eco RI probe hybridized to m4:5b/YRp7-1 RNA and S1-digested. S1-specific bands range from 0.9 to 1.0 Kb in length. The two middle signals observed with the Dde I probe, which differ in length by only 13 nucleotides, are not resolved in lane 5. 6: Eco RI probe, untreated. The lanes shown are from two autoradiographic exposures of the same gel. The position and sizes of marker fragments are indicated to the right.

Figure 5. Nuclear localization of an ADR6- β -galactosidase gene fusion product. The haploid yeast strain 2000-6A (*ura3*, *trp1*, *leu2*, *adh3*, *MATa*) was transformed with pHH73, a YRp7-based plasmid bearing the ADR6- β -galactosidase protein fusion described in Results. Two independent TRP⁺ transformants were selected for analysis by indirect immunofluorescence. Cells were grown to exponential phase in glucose-supplemented synthetic medium lacking tryptophan then treated as described in Materials and Methods. A, B: DAPI staining of DNA to localize the nucleus. C, D: Immunofluorescence localizing the β -galactosidase protein present in the cell. A comparison of A to C and B to D indicates that the immunofluorescence corresponds to the nucleus.

Subcellular localization of an ADR6/ β -galactosidase fusion protein

In order to determine the subcellular localization of ADR6, a gene fusion approach was used. *E. coli* β -galactosidase protein when expressed in yeast is excluded from the nucleus. If, however, it is fused to a protein containing a nuclear targeting signal, it is present predominantly in the nucleus (37). LacZ was fused to the ADR6 gene at the unique Bam HI site of plasmid YRp7-1. This construction places the fusion at nucleotide 3408 in the DNA sequence (Figure 1), which corresponds to amino acid 908. After transfor-

mation of yeast cells, indirect immunofluorescence using anti- β -galactosidase anti-serum revealed a signal which was more intense over the nuclear region than elsewhere in the cell (Figure 5) indicating that the ADR6 gene product is a nuclear protein.

DISCUSSION

The most striking feature of the primary structure of ADR6 is the presence of stretches of homopolymeric amino acid sequence, most notably polyasparagine, polyglutamine and polythreonine. There are also repeats of three or more residues in tandem of serine, alanine, lysine, tyrosine and leucine. Examples of homopolymeric peptide sequences made up of each of the amino acids are known (for instance, polyglutamine in gliadin, the yeast STE18 gene product [Whiteway et al., submitted] and mouse interleukin 2, polyserine in phosvitin, polyarginine in histones, polyalanine in antifreeze proteins, polycysteine in amylase, polyglycine in keratin as well as numerous homopolymeric stretches in nuclear proteins, discussed below), although polymers of methionine and the aromatic amino acids are uncommon (e.g., 38-46).

The homopolymeric sequence found most extensively in ADR6 is polyasparagine. In searches comparing ADR6 sequence with various sequence databases, the highest scores found were matches with other polyasparagine-containing sequences from Trypanosomes (23, 24), Plasmodia (25, 26) and yeast mitochondrial VAR1 genes (28, 31). The polyasparagine region of the VAR1 gene is associated with polymorphism at this locus due to variation in the tandemly repeated AA(T/C) codon repeat copy number. In addition, polyasparagine is found in the DNA-binding regulatory product of the LAC9 gene of Kluyveromyces lactis (47). The polyasparagine stretches in the putative ADR6 gene product is encoded by A-T rich repeats which extend upstream of the ADR6 open reading frame (Figure 1).

The ADR6 open reading frame also contains [CA(A/G)] repeats encoding polyglutamine. These repeats are similar to the opa (or M or strep) repeats of Drosophila (48-50). The opa repeats encode polyglutamine in a number of Drosophila homeotic DNA-binding proteins, including Engrailed (8), Notch (51) Fushi Tarazu (52) and also in the yeast regulatory protein encoded by PHO2 (53), which also contains a homeo box (54). Tandem repeats, encoding a variety of homo-

polymeric stretches, appear to be particularly common in some types of DNA binding proteins. Besides polyglutamine stretches in homeotic proteins, homopolymers in proteins which interact with nucleic acids include polyalanine and polyserine in Engrailed, polyglycine in Engrailed-related (8), polyalanine, polyserine and polyasparagine in Caudal (55), polyalanine and polyserine in Even-skipped (56) polyasparagine in Deformed (57), polyglutamine in Antennapedia (58) and polyglutamic acid in mouse M6 (59) and HOX3 (60) homeotic proteins. The pen repeat in a Drosophila cDNA encodes a polyglycine-containing domain similar to rat helix destabilizing protein (61).

Several potential metal-binding and nucleic acid binding proteins contain homopolymeric stretches. HAP2 (62) contains polyglutamine, NGFI-A (63) contains polyserine and polyglycine, CDC16 (64) contains polythreonine, LEU3 (65) contains polyglutamic acid and HAP3 (66) contains short stretches of polyserine and polyglutamine. Some nuclear proteins with no homeobox or finger homologies contain extensive homopolymeric sequences. The Drosophila single-minded (sim) gene (67) encodes a protein with regions of polyvaline, polyserine, polyasparagine, polyhistidine, polyglutamine, polyglycine, polytyrosine and poly-alanine/glutamine repeats. The Drosophila Zeste gene (68) encodes stretches of polyglutamine, polyalanine, polyglycine and poly-alanine/glutamine.

The variety in composition of homopolymeric sequences has led some authors to suggest that the repeats function at the nucleic acid level rather than the protein level, perhaps representing protein binding site(s) in the mRNA (63). For instance the opa repeat (CAG) might be read in any frame, producing polyglutamine, polyserine or polyglycine. The [A(A/C)T] repeat in ADR6 extends upstream of the putative initiation codon and thus would not be translated at all in this region. Once past the putative initiator, it encodes polyasparagine and polythreonine. Imperfect repeats of this tandem repeat are found in several places throughout the ADR6 coding region (Figure 6). Also, the codon usage in the polyasparagine-polythreonine regions is skewed, favoring AAT (Asn) and ACT (Thr) to a much greater extent than in the rest of the molecule. The codon usage in the Drosophila sim gene product is also altered in a repeated region (67). These observations are consistent with the sug-

On the other hand, some repetitive regions in ADR6 code for protein regions with some interesting predicted structural features. For instance, in addition to the homeo box, some homeotic proteins have regions of predicted α helix- β turn- α helix structure

ADR6:	C	VQLIK	C	LVEKSI	C	FEN	C
GAL4:	C	RLKKLK	C	SKEKPK	C	AK	C
PPRI:	C	RLKKIK	C	DQEFPS	C	KR	C
CDC16:	C	YTQ	C	KFNE	C	LEL	C
ARGRII:	C	RGRVKV	C	DLRHPH	C	QR	C
LEU3:	C	RQKSK	C	DAHERAPEP	C	TK	C
LAC9:	C	RKKWK	C	SKTVPT	C	TN	C
qa-1F	C	RAAREK	C	DGIQPA	C	FP	C
HTLVIII tat:	C	TNCY	C	KKCCFH	C	QV	C
T ag HPV:	C	RLPIT	C	LRNKGIST	C	N	C
T ag MPV:	C	RHPLT	C	LVNVKYSS	C	S	C
ST ag SV40:	C	KOMPE	C	AKKNSAN	C	I	C
ST ag BK:	C	KEWPI	C	SKKPSVH	C	P	C
ST ag JC:	C	KEWPN	C	ATNPSVH	C	P	C

typical of DNA binding regions in proteins (69). For example, in the Engrailed gene product, the region that scores highest in a helix-turn-helix detection computer program (70) is not the homeobox, but a polyglutamine-containing region near the amino terminus. The region that scores highest in ADR6 is a polyglutamine-containing region (amino acids 340 to 359) similar to this Engrailed sequence (amino acids 31-50). A region just downstream of this polyglutamine motif in ADR6 shows some similarity with another nucleic acid binding protein, the nif transcriptional activator (71). The codon usage in the polyglutamine regions of ADR6 is not skewed as in the polyasparagine or polythreonine regions. The usage of the glutamine codons CAA and CAG in the polyglutamine regions is similar to the overall usage of these codons in ADR6.

ADR6 contains a group of cysteine residues near the carboxyl terminus that resembles potential metal binding domains in nucleic acid binding proteins (reviewed in 34, 72). Other yeast regulatory genes containing demonstrated or potential metal-binding nucleic "finger" regions include ADR1 (3, 5), GAL4 (73), PPRI (74), CDC16 (64), ARGRII (75), LEU3 (65), LAC9 (47) SWI5 (76) and HAP3 (66). A comparison of the putative metal binding domain of ADR6 with other Cys_x potential metal binding domains with nucleic acid binding potential, such as the HIV virus transcriptional activator and several T antigens, is shown in Figure 7.

The observation that an ADR6/β-galactosidase fusion protein is found in the nucleus of yeast cells is consistent with the genetic data suggesting a role for ADR6 in ADH2 transcription. The functional implications of a potential nucleic acid binding finger domain are also consistent with these results.

ACKNOWLEDGEMENTS

Funding for this research was provided in part by NIH grant GM26079. The authors thank Margo Rogers for preparation of the manuscript and Frank Grant for discussions and technical advice.

REFERENCES

1. Beier, D.R. and Young, E.T. (1985). Mol. Cell. Biol. 5: 1743-1749.
2. Denis, C.L. (1983). Mol. Cell. Biol. 3: 360-370.
3. Hartshorne, T.A., Blumberg, H. and Young, E.T. (1986). Nature 320: 283-287.
4. Miller, J., McLachlan, D. and Klug, A. (1985). EMBO J. 4: 1609-1614.

-
5. Bemis, L.T. and Denis, C.L. (1988). Mol. Cell. Biol. 8: 2125-2131.
 6. Taguchi, A.K.W. and Young, E.T. (1987a). Genetics 116: 523-530.
 7. Taguchi, A.K.W. and Young, E.T. (1987b). Genetics 116: 531-540.
 8. Poole, S.J., Kauver, L.M., Drees, B. and Kornberg, T. (1985). Cell 40: 37-43.
 9. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA 74: 5463-5467.
 10. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983). Proc. Natl. Acad. Sci. USA 80: 3963-3965.
 11. Tabor, S. and Richardson, C.C. (1987). Proc. Natl. Acad. Sci. USA 84: 4767-4771.
 12. Lipman, D.J. and Pearson, W.R. (1985). Science 277: 1435-1441.
 13. Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987). Proc. Natl. Acad. Sci. USA 84: 4355-4358.
 14. Bilofsky, H.S., Burks, C., Fickett, J.W., Goad, W.B., Lewitter, F.I., Rindone, W.P., Swindell, C.D. and Tung, C-S. (1986). Nucl. Acids. Res. 14: 1-4.
 15. Hamm, G.H. and Cameron, G.N. (1986). Nucl. Acids. Res. 14: 5-9.
 16. George, D.G., Barker, W.C. and Hunt, L.T. (1986). Nucl. Acids. Res. 14: 11-15.
 17. Berk, A.J. and Sharp, P.A. (1977). Cell 12: 721-732.
 18. Casadaban, M.J., Martinez-Arias, A., Shapria, S.K. and Chou, J. (1983). Methods Enzymol. 100: 293-308.
 19. Kilmartin, J.V. and Adams, A.E.M. (1984). J. Cell. Biol. 98: 922-933.
 20. Hamilton, R., Watanabe, C.K., de Boer, H.A. (1987). Nucl. Acids. Res. 15: 3581-3593.
 21. Sharp, P.M., Tuohy, M.F. and Mosurski, R. (1986). Nucl. Acids. Res. 14: 5125-5139.
 22. Zaret, K.S. and Sherman, F. (1982). Cell 28: 563-573.
 23. Campbell, D.A., van Bree, M.P. and Boothroyd, J.C., (1984). Nucl. Acids Res. 12: 2759-2774.
 24. Aline, R., MacDonald, G., Brown, E., Allison, J., Myler P., Rothwell, V. and Stuart, K. (1985). Nucl. Acids Res. 13: 3161-3177.
 25. Wahlgren, M., Aslund, L., Franzen, L., Sundvall, M., Wahlin, B., Berzins, K., McNicol, L.A., Bjorkman, A., Wigzell, H., Perlman, P. and Pettersson, U. (1986). Proc. Nat. Acad. Sci. USA 83: 2677-2681.
 26. Stahl, H.D., Bianco, A.E., Crewther, P.E., Burkot, T., Coppel, R.L., Brown, G.V., Anders, R.F. and Kemp, D.J. (1986). Nucl. Acids Res. 14: 3089-3102.
 27. Tzagoloff, A. Nobrega, M., Akai, A. and Macino, G. (1980). Curr. Gen. 2: 149-157.
 28. Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S., Butow, R.A. and Grossman, L.I. (1982). Cell 30: 617-626.
 29. Miller, D.L., Folse, J.R., Benson, P.J. and Martin, N.C. (1983). Biochemistry-USA 22: 1709-1714.
 30. Zassenhaus, H.P., Farrelly, F., Hudspeth, M.E.S., Grossman, L.I. and Butow, R.A. (1983). Mol. Cell. Biol. 3: 1615-1624.
 31. Ainley, W.M., Macreadie, I.G. and Butow, R.A. (1985). J. Mol. Biol. 184: 565-576.
 32. Clark-Walker, G.D., McArthur, C.R. and Sriprakash, K.S. (1985). EMBO J. 4: 465-473.
-

33. Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R. (1984). J.Mol. Biol. 179: 125-142.
34. Berg, J.M. (1986). Science 232: 485-487.
35. Struhl, K. (1982). Proc. Natl. Acad. Sci. USA 79: 7385-7389.
36. Guarente, L.P. and Mason, T. (1983). Cell 32: 1279-1286.
37. Hall, M.W., Hereford, L. and Herskowitz, I. (1984). Cell 36: 1057-1065.
38. Byrne, B.M., van het Schip, A.D., van de Klundert, J.A.M., Arnberg, A.C., and Gruber, M. (1984). Biochemistry 23: 4275-4279.
39. Rafalski, J.A., Scheets, K., Metzler, M., Peterson, D.M., Hedgcoth, C. and Soll D.G. (1984). EMBO J. 3: 1409-1415.
40. Kleene, K.C., Distel, R.J. and Hecht, N.B. (1985). Biochemistry 24: 719-722.
41. Lin, Y. and Gross, J.K. (1981). Proc. Nat. Acad. Sci. USA 78: 2825-2829.
42. Gorecki, M. and Zeelon, E.P. (1979). J. Biol. Chem. 254: 525-529.
43. Isemura, S., Saitoh, E. and Sanada, K. (1979). J. Biochem. 86: 79-86.
44. Steinert, P.M., Parry, D.A.D., Idler, W.W., Johnson, L.D., Steven, A.C. and Roop, D.R. (1985). J. Biol. Chem. 260: 7142-7149.
45. Fuse, A., Fujita, T., Yasumitsu, H., Kashima, N., Hasegawa, K. and Taniguchi, T. (1984). Nucl. Acids Res. 12: 9323-9331.
46. Garfinkle, M.D., Pruitt, R.E. and Meyerowitz, E.M. (1983). J. Mol. Biol. 168: 765-789.
47. Wray, L.V., Witte, M.M., Dickson, R.C. and Riley, M.I. (1987). Mol. Cell. Biol. 7: 1111-1121.
48. Wharton, K.A., Yedvobnick, B., Finnerty, V.G. and Artavanis-Tsakonas, S. (1985a). Cell 40: 55-62.
49. Kidd, S., Lockett, T.J. and Young, M.W. (1983). Cell 34: 421-433.
50. McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984). Nature 308: 428-433.
51. Wharton, K.A., Johansen, K.M., Xu, T. and Artavanis-Tsakonas, S. (1985b). Cell 43: 567-581.
52. Laughon, A. and Scott, M.P. (1984). Nature 310: 25-31.
53. Sengstag, C. and Hinnen, A. (1987). Nucl. Acids. Res. 15: 233-246.
54. Burglin, T.R. (1988). Cell 53: 339-340.
55. Mlodzik, M. and Gehring, W.J. (1987). Cell 48: 465-478.
56. MacDonald, P.M., Ingham, P. and Struhl, G. (1986). Cell 47: 721-734.
57. Regulski, M., McGinnis, N., Chadwick, R. and McGinnis, W. (1987). EMBO J. 6: 767-777.
58. Schneuwly, S., Kuroiwa, A., Baumgartner, P. and Gehring, W.J. (1986). EMBO J. 5: 733-739.
59. Colberg-Poley, A.M., Voss, S.D., Chowdhury, K. and Gruss, P. (1985). Nature 314: 713-718.
60. Breier, G., Dressler, G.R. and Gruss, P. (1988). EMBO J. 7: 1329-1336.
61. Haynes, S.R., Rebbert, M.L., Mozer, B.A., Forquignon, F. and Dawid, I.B. (1987). Proc. Natl. Acad. Sci. USA 84: 1819-1823.
62. Pinkham, J.L., Olesen, J.T. and Guarente, L.P. (1987). Mol. Cell. Biol. 7: 578-585.
63. Milbrandt, J. (1987). Science 238:797-799.

-
64. Icho, T. and Wickner, R.B. (1987). Nucl. Acids. Res. 15: 8439-8450.
 65. Friden, P. and Schimmel, P. (1987). Mol. Cell. Biol. 7: 2708-2717.
 66. Hahn, S., Pinkham, J., Wei, R., Miller, R. and Guarente, L. (1988). Mol. Cell. Biol. 8: 655-663.
 67. Crews, S.T., Thomas, J.B. and Goodman, C.S. (1988). Cell 52: 143-151.
 68. Pirrotta, V., Manet, E., Hardon, E., Bickel, S.E. and Benson, M. (1987). EMBO J. 6: 791-799.
 69. Pabo, C.O. and Sauer, R.T. (1984). Ann. Rev. Biochem. 53: 293-321.
 70. Moore, J., Engelberg, A. and Bairoch, P. (1988). Biotechniques 6: 566-572.
 71. Buikema, W.J., Szeto, W.W., Lemley, P.V., Orme-Johnson, W.H. and Ausubel, F.M. (1985). Nucl. Acids Res. 13: 4539-4555.
 72. Evans, R.M. and Hollenberg, S.M. (1988). Cell 52: 1-3.
 73. Laughon, A. and Gesteland, R.F. (1984). Mol. Cell. Biol. 4: 260-267.
 74. Kammerer, B., Guyonvarch, A. and Hubert, J.C. (1984). J. Mol. Biol. 180: 239-250.
 75. Messenguy, F., Dubois, E. and Descamps, F. (1986). Eur. J. Biochem. 157: 77-81.
 76. Stillman, D.J., Bankier, A.T., Seddon, A., Groenhout, E.G. and Nasmyth, K.A. (1988). EMBO J. 7: 485-494.
 77. Baum, J.A., Geever, R. and Giles, N.H. (1987). Mol. Cell. Biol. 7: 1256-1266.
 78. Arya, S.K. and Gallo, R.C. (1986). Proc. Nat. Acad. Sci. USA 83: 2209-2213.
 79. Delmas, V., Bastien, C., Scherneck, S. and Feunteun, J. (1985). EMBO J. 4: 1279-1286.
 80. Soeda, E., Arrand, J.R., Smolar, N., Walsh, J.E. and Griffin, B.E. (1980). Nature 283: 445-453.
 81. Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L. and Weissman, S.M. (1978). Science 200: 494-502.
 82. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., van de Voorde, A., van Heuverswyn, H., van Herreweghe, J., Volckaert, G. and Ysebaert, M. (1978). Nature 273: 113-120.
 83. Seif, I., Khoury, G. and Dhar R. (1979). Cell 18: 963-977.
 84. Frisque, R.J., Bream, G.L. and Cannella, M.T. (1984). J. Virol. 51: 458-469.
-