Volume 15 Number 1 1987 Nucleic Acids Research

The sequence of the *Saccharomyces cerevisiae* gene *PHO2* codes for a regulatory protein with unusual aminoacid composition

Christian Sengstag and Albert Hinnen

Ciba-Geigy AG, Department of Biotechnology, CH-4002 Basel, Switzerland

Received November 17, 1986; Accepted December 8, 1986

ABSTRACT

A new centromere vector for the construction of a Saccharomyces cerevisiae gene library, allowing direct selection for DNA insert, will be described. From that library the gene for the regulatory protein PHO2 involved in PHO5 induction has been cloned by complementation of a pho2 mutation. The complementing activity was shown to be located on a 3.6 kb HindIII fragment. This fragment was used to evict the genomic copy and with appropriate genetic crosses we proved, that the cloned gene is PHO2. The DNA sequence of PHO2 was determined. Analysis of the sequence data uncovered striking homology regions with PHO4, another protein necessary for the induction of PHO5. The relevance of the observed homology will be discussed.

INTRODUCTION

The phosphatase system in yeast Saccharomyces cerevisiae is a network consisting of structural genes and regulatory genes (1). The phosphatases can be divided into alkaline (PHO8) and acid phosphatases (APase) (PHO5, PHO3, PHO11). The PHO3 gene, known as the constitutive APase, is dependent on the regulatory genes PHO6 and PHO7 and seems to be repressed by the PHO5 gene product (2). PHO5 and PHO11 are in turn regulated APases - transcriptionally repressed in high Pi medium - that are induced via the gene products of PHO4 and PHO2. The presence of the PHO81 gene is central for the APase induction circuit. In low Pi medium it binds the negative factors coded by PHO80 and/or PHO85 dissociating them from the PHO2 and/or PHO4 gene products thus allowing activation of PHO5 transcription.

In analogy to GAL4 (3) a model proposes that PHO4 binds as a complex with PHO2 directly to the upstream activator sequence (UAS) of PHO5. Although genetic data put PHO2 on the same hierarchy level as PHO4, it cannot be ruled out that the action

of PHO2 is posttranscriptional. With the aim of learning more about the function of PHO2, we cloned the PHO2 gene from a new gene library of S. cerevisiae. Here we report the DNA sequence of PHO2 and show that it exhibits distinct homology regions with PHO4.

MATERIALS AND METHODS

Strains

E. coli HB101 (r_k^-, m_k^-, leu, pro, recA) and JM 109 (recA, endA1, gyrA96, thi, hsdR17, supE44, relA, Δ(lac-pro), F'(traD36, proAB, $lacI^q$, ZΔM15)) were used for standard transformations. B15 (trp, pyrF, r_k^-, m_k^-) was used to select for the URA3 gene in E. coli on minimal plates containing tryptophane.

Yeast S288C is our prototrophic wild-type strain. YAT104 (a, trp1, pho3, pho2) was a generous gift of A. Toh-e.
YS104 is a ura3Δ derivative of YAT104. YS18 is a ura3Δ derivative of GRF18 (α, his3-11, his3-15, leu2-3, leu2-112, can^R).

Media

YPD medium contained 2% Bacto Peptone (Difco), 1% Bacto Yeast Extract (Difco) and 2% glucose. Ura^+ selection was done on synthetic complete medium (4) lacking only uracil. Antibiotics were used at 50µg/ml for Ampicillin and 4µg/ml for Tetracyclin. E. coli M9 minimal plates have been described (5).

Yeast transformation

Competent cells were obtained by Li-acetate treatment (5). An exponentially growing culture in YPD was harvested at OD_{600}=2, washed with TE buffer (10 mM Tris, 1mM EDTA, pH 8), resuspended in 1/2 vol 1M Li-acetate and shaked at 30° for 1 1/2 h. Then the cells were resuspended in 1/50 vol 1M Li-acetate and 1-10 µg transforming DNA was added to 0.2 ml competent cells. After 10' incubation at 30°C, PEG 4000 was added to 40% final volume and incubation continued for 1h. Cells were harvested by centrifugation, washed with 0.8M sorbitol and spread on appropriate selection plate.

Isolation of DNA fragments from gel

DNA bands to recover from agarose gels were electrophoresed into a suspension of hydroxylappatite. From there the DNA was eluted with 0.5M phosphate buffer pH 7.5 and purified on a Sephadex G100 mini column in TE buffer.

Introducing a deletion into the chromosomal URA3 gene of YAT104

Plasmid pUC12::URA3, containing the URA3 gene cloned as 1.1kb HindIII fragment, was digested with NcoI and StuI, the ends were filled in with Klenow polymerase and the plasmid was religated. Deletion derivative pUC12::ura3Δ was digested to completion with HindIII and used to cotransform YAT104 together with 1µg YRp7. Trp^+ transformants were selected on plate, washed off and Ura^- mutants were selected in 5-fluoro-orotic acid as described (7, 8). Correct integration of the ura3Δ allele at the URA3 locus has been checked by Southern hybridization.

In analogy the URA3 gene was mutated in GRF18, except that Yep13 was used as cotransforming plasmid.

Construction of pCS19

The 1.1kb HindIII fragment with the URA3 gene was inserted into the HindIII site of YRp7-PHO5 (8). Subsequently the plasmid was digested with XbaI/BamHI, and the large fragment was ligated with a 1kb XbaI/BglII fragment carrying CEN15 (9). The ClaI site was converted to a XhoI site by linker insertion giving pCS14. In parallel the PstI/BamHI fragment of pUN121 (10) carrying the cI gene was cloned into PstI/BamHI digested pUC18 followed by conversion of the PstI site to a XhoI site by linker insertion. The short PvuI/XhoI fragment was then replaced by the PvuI/XhoI fragment of pCS14, carrying CEN15, followed by replacing the short PvuI/BamHI fragment by the ori containing PvuI/BamHI fragment of pUN121. Finally the short PvuI/SalI fragment was replaced by the long PvuI/SalI fragment of pBR322, in which a 158bp HincII/SmaI fragment containing the E. coli trp-terminator had been inserted at the AvaI site behind the Tet^r gene. This vector (pCS19) has been used for cloning random yeast DNA.

Construction of pCS21

To evict the chromosomal PHO2 gene, the 3.6kb HindIII fragment containing PHO2 (see Results) was subcloned into the HindIII site of pUC19. Subsequently the HincII site located in the vector part was eliminated by digesting the plasmid at the overlapping SalI site and religating the plasmid upon filling in the SalI ends by Klenow polymerase. The resulting plasmid was digested with HincII, religated in the presence of end-repaired 1.1kb HindIII fragment carrying URA3 and E. coli strain B15 (pyrF) was transformed. Clones expressing URA3 were selected on minimal plates (+ tryptophane) and the correct construction (plasmid pCS21) was confirmed by restriction enzyme digests. The yeast DNA part of pCS21 is shown in Fig. 2.

Introducing a frameshift mutation in the N-terminal part of PHO2

The 3.6kb HindIII fragment was cloned into pUC19. This plasmid was then digested with EcoRI and the protruding 5' ends were filled in by Klenow polymerase. One half of this end-repaired DNA was digested with SalI and the 950 bp fragment (Fig. 2) was isolated from an agarose gel. The other half was digested with BglII and the shortest fragment was isolated. Finally, the original plasmid was digested with SalI/BglII and the large fragment was ligated with the two DNA fragments isolated before. Restriction digestion showed, that the EcoRI site at the N-terminal end of PHO2 got lost, and DNA sequencing using a synthetic oligodeoxynucleotide proved that we had introduced a +4 frameshift at that particular site. The mutated HindIII fragment was then subcloned into pDP39 (see vectors) and transformed in YS104.

Acid phosphatase assay

The staining assay for APase has been described (11).

Vectors

Plasmid pDP39, a generous gift from D. Pridmore, was used for subcloning of PHO2 DNA fragments. It was constructed by inserting URA3, ARS1 and CEN14 into pUC19. All PHO2 fragments have been cloned into the unique SalI site, after converting it to a blunt end site by Klenow polymerase.

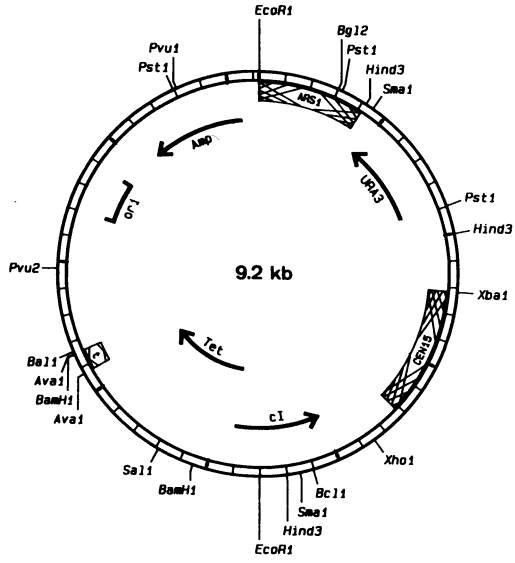


Fig. 1 Restriction map of vector pCS19. The λcI gene product, coded by the vector, represses the λpL promoter which controls Tet^r expression. Upon cloning into the unique BclI site, Tet^r becomes expressed allowing direct selection for recombinant plasmids. The introduced trp terminator is represented by the letter t.

RESULTS

Construction of a S. cerevisiae gene library in a centromere vector

Gene dosage may strongly influence gene expression which can lead to severe problems in cloning regulatory genes. Therefore we constructed vector pCS19 for cloning random yeast DNA. Due to its centromere, this vector (Fig. 1) is present at one to two copies per yeast cell and is mitotically stable. By cloning into its unique BclI site, it allows direct selection for recombinant plasmids by expressing tetracyclin-resistance in E. coli. For details of construction see Materials and Methods.

DNA of the wild-type yeast strain S288C was partially digested with Sau3A, fragments in the range of 10kb were purified from gel and ligated together with BclI cleaved pCS19. 5000 individual tet^r E. coli transformants were picked and stored. Three yeast genome equivalents should be contained in

this library, giving about 95% probability for a specific gene to be represented.

Cloning of yeast DNA sequences that complement the pho2 mutation

Plasmid DNA was purified from pools of each 384 E. coli gene bank clones and used to transform YS104 (pho2, ura3) to Ura^+. Transformants selected on synthetic complete medium plates lacking uracil were replica plated on low Pi and high Pi medium. Elevated APase activity was monitored by a colony staining procedure (Materials and Methods). Two transformants showed correctly expressed APase activity and were chosen for further analysis. Total DNA was isolated and used to transform E. coli. Subsequently, plasmid DNA was recovered from Amp^rTet^r tranformants and used to transform YS104. Correct expression of APase by the yeast transformants proved that both plasmids could complement the pho2 mutation.

A 3.6 kb HindIII fragment can complement the pho2 mutation

Restriction enzyme digestion of the two plasmids revealed that they carried similar but not identical 10 kb long yeast sequences. In particular a 3.6 kb HindIII fragment was present in both plasmids. Cloning of this HindIII fragment in pDP39 and subsequent transformation of YS104 demonstrated that this DNA was sufficient for complementing pho2. A restriction map of the HindIII fragment is shown in Fig. 2.

DNA sequence of the 3.6 kb HindIII fragment

Figure 2 shows the strategy used for sequencing. In brief, various subfragments of the 3.6 kb HindIII fragment were subcloned into M13mp18 or M13mp19 and sequenced according to the dideoxy method (12) using the M13 universal primer. Subsequently, sequencing was continued by using synthetic deoxyoligonucleotide-primers.

One long open reading frame was found starting nearby the leftmost EcoRI site and extending about 200 bp to the right of a HincII site. It can code for a 63390 D protein consisting of 559 aminoacids. In the flanking regions open reading frames of only less than 74 aminoacids are found. The sequence of the long open reading frame and its 5' and 3' flanking regions is shown in Fig. 3 together with the deduced protein sequence.

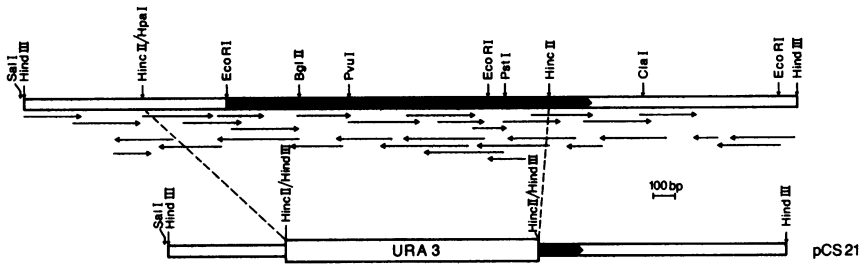
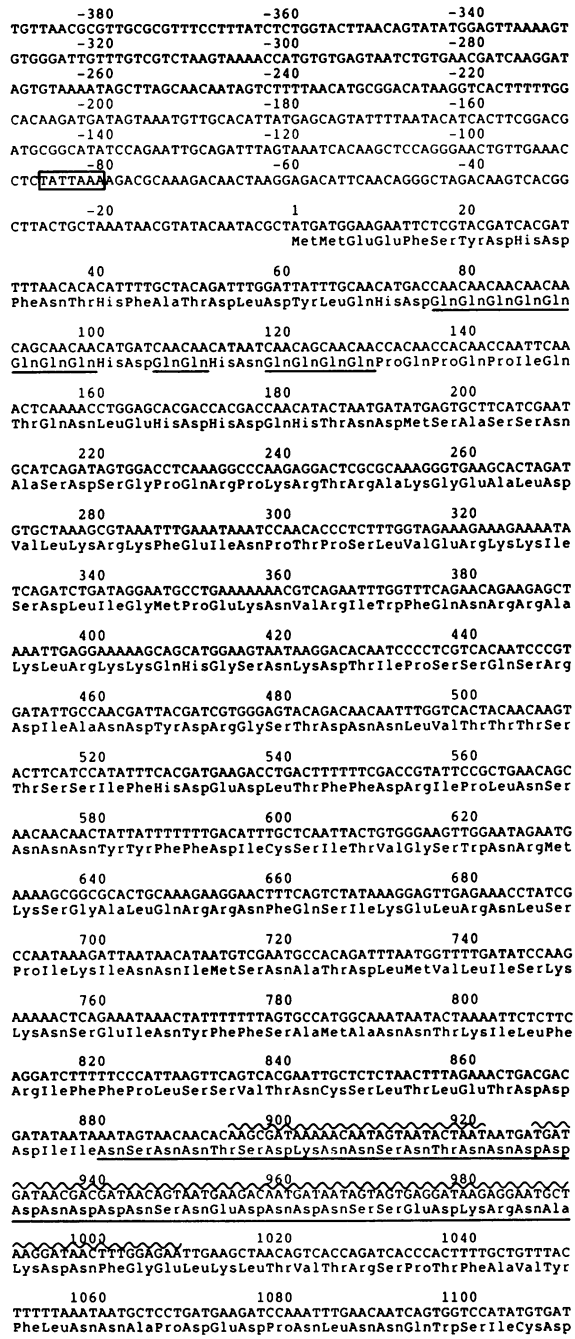


Fig. 2 A restriction map of the 3.6kb HindIII fragment is shown in the upper part. The PHO2 open reading frame is represented by the filled region. Arrows indicate the extent of DNA sequence determination. Both strands have been sequenced between the HpaI and ClaI sites. The lower part shows the yeast DNA segment present in pCS21, where the major part of the PHO2 gene is exchanged against the URA3 gene. The HincII/HindIII fusion does not restore a HindIII site. The SalI site of vector pUC19 is present just left of the HindIII fragment.

The long open reading frame codes for PHO2

To prove that we had cloned the PHO2 gene and to exclude that the observed APase expression was due to a cloned unrelated factor, different from PHO2, we decided to evict the putative chromosomal PHO2 region. The 3.6 kb HindIII fragment was subcloned into pUC19 and and the internal HincII fragment was exchanged against the URA3 gene, giving pCS21 (Fig. 2) (see Materials and Methods). Yeast strain YS18 (PHO2, ura3) was transformed with HindIII digested pCS21 plasmid DNA and selected for Ura$^+$. Total DNA of ten transformants was isolated, digested with HindIII and electrophoresed. The gel was blotted to nitrocellulose and probed with the nick-translated 3.6 kb HindIII fragment (putative PHO2). All ten transformants showed a unique band at 2.9 kb that was shifted with respect to the 3.6 kb band of the parental strain YS18, indicating an efficient gene replacement (data not shown). Subsequent colony staining clearly showed that APase expression became eliminated upon transformation. Furthermore the APase deficient transformants (YS18 pho2::URA3) were backcrossed with YS104 and the resulting diploids turned out to lack APase activity.


-380 -360 -340
TGTTAACGCGTTGCGCGTTTCCTTTATCTCTGGTACTTAACAGTATATGGAGTTAAAAGT
-320 -300 -280
GTGGGATTGTTTGTCGTCTAAGTAAAACCATGTGTGAGTAATCTGTGAACGATCAAGGAT
-260 -240 -220
AGTGTAAAATAGCTTAGCAACAATAGTCTTTTAACATGCGGACATAAGGTCACTTTTTGG
-200 -180 -160
CACAAGATGATAGTAAATGTTGCACATTATGAGCAGTATTTTAATACATCACTTCGGACG
-140 -120 -100
ATGCGGCATATCCAGAATTGCAGATTTAGTAAATCACAAGCTCCAGGGAACTGTTGAAAC
-80 -60 -40
CTCTATTAAAAGACGCAAAGACAACTAAGGAGACATTCAACAGGGCTAGACAAGTCACGG
-20 1 20
CTTACTGCTAAATAACGTATACAATACGCTATGATGGAAGAATTCTCGTACGATCACGAT
MetMetGluGluPheSerTyrAspHisAsp
40 60 80
TTTAACACACATTTTGCTACAGATTTGGATTATTTGCAACATGACCAACAACAACAACAA
PheAsnThrHisPheAlaThrAspLeuAspTyrLeuGlnHisAspGlnGlnGlnGlnGln
100 120 140
CAGCAACAACATGATCAACAACATAATCAACAGCAACAACCACAACCACAACCAATTCAA
GlnGlnGlnHisAspGlnGlnHisAsnGlnGlnGlnGlnProGlnProGlnProIleGln
160 180 200
ACTCAAAACCTGGAGCACGACCACGACCAACATACTAATGATATGAGTGCTTCATCGAAT
ThrGlnAsnLeuGluHisAspHisAspGlnHisThrAsnAspMetSerAlaSerSerAsn
220 240 260
GCATCAGATAGTGGACCTCAAAGGCCCAAGAGGACTCGCGCAAAGGGTGAAGCACTAGAT
AlaSerAspSerGlyProGlnArgProLysArgThrArgAlaLysGlyGluAlaLeuAsp
280 300 320
GTGCTAAAGCGTAAATTTGAAATAAATCCAACACCCTCTTTGGTAGAAAGAAAGAAAATA
ValLeuLysArgLysPheGluIleAsnProThrProSerLeuValGluArgLysLysIle
340 360 380
TCAGATCTGATAGGAATGCCTGAAAAAAACGTCAGAATTTGGTTTCAGAACAGAAGAGCT
SerAspLeuIleGlyMetProGluLysAsnValArgIleTrpPheGlnAsnArgArgAla
400 420 440
AAATTGAGGAAAAAGCAGCATGGAAGTAATAAGGACACAATCCCCTCGTCACAATCCCGT
LysLeuArgLysLysGlnHisGlySerAsnLysAspThrIleProSerSerGlnSerArg
460 480 500
GATATTGCCAACGATTACGATCGTGGGAGTACAGACAACAATTTGGTCACTACAACAAGT
AspIleAlaAsnAspTyrAspArgGlySerThrAspAsnAsnLeuValThrThrThrSer
520 540 560
ACTTCATCCATATTTCACGATGAAGACCTGACTTTTTTCGACCGTATTCCGCTGAACAGC
ThrSerSerIlePheHisAspGluAspLeuThrPhePheAspArgIleProLeuAsnSer
580 600 620
AACAACAACTATTATTTTTTTGACATTTGCTCAATTACTGTGGGAAGTTGGAATAGAATG
AsnAsnAsnTyrTyrPhePheAspIleCysSerIleThrValGlySerTrpAsnArgMet
640 660 680
AAAAGCGGCGCACTGCAAAGAAGGAACTTTCAGTCTATAAAGGAGTTGAGAAACCTATCG
LysSerGlyAlaLeuGlnArgArgAsnPheGlnSerIleLysGluLeuArgAsnLeuSer
700 720 740
CCAATAAAGATTAATAACATAATGTCGAATGCCACAGATTTAATGGTTTTGATATCCAAG
ProIleLysIleAsnAsnIleMetSerAsnAlaThrAspLeuMetValLeuIleSerLys
760 780 800
AAAAACTCAGAAATAAACTATTTTTTTAGTGCCATGGCAAATAATACTAAAATTCTCTTC
LysAsnSerGluIleAsnTyrPhePheSerAlaMetAlaAsnAsnThrLysIleLeuPhe
820 840 860
AGGATCTTTTTCCCATTAAGTTCAGTCACGAATTGCTCTCTAACTTTAGAAACTGACGAC
ArgIlePhePheProLeuSerSerValThrAsnCysSerLeuThrLeuGluThrAspAsp
880 900 920
GATATAATAAATAGTAACAACACAAGCGATAAAAACAATAGTAATACTAATAATGATGAT
AspIleIleAsnSerAsnAsnThrSerAspLysAsnAsnSerAsnThrAsnAsnAspAsp
940 960 980
GATAACGACGATAACAGTAATGAAGACAATGATAATAGTAGTGAGGATAAGAGGAATGCT
AspAsnAspAspAsnSerAsnGluAspAsnAspAsnSerSerGluAspLysArgAsnAla
1000 1020 1040
AAGGATAACTTTGGAGAATTGAAGCTAACAGTCACCAGATCACCCACTTTTGCTGTTTAC
LysAspAsnPheGlyGluLeuLysLeuThrValThrArgSerProThrPheAlaValTyr
1060 1080 1100
TTTTTAAATAATGCTCCTGATGAAGATCCAAATTTGAACAATCAGTGGTCCATATGTGAT
PheLeuAsnAsnAlaProAspGluAspProAsnLeuAsnAsnGlnTrpSerIleCysAsp

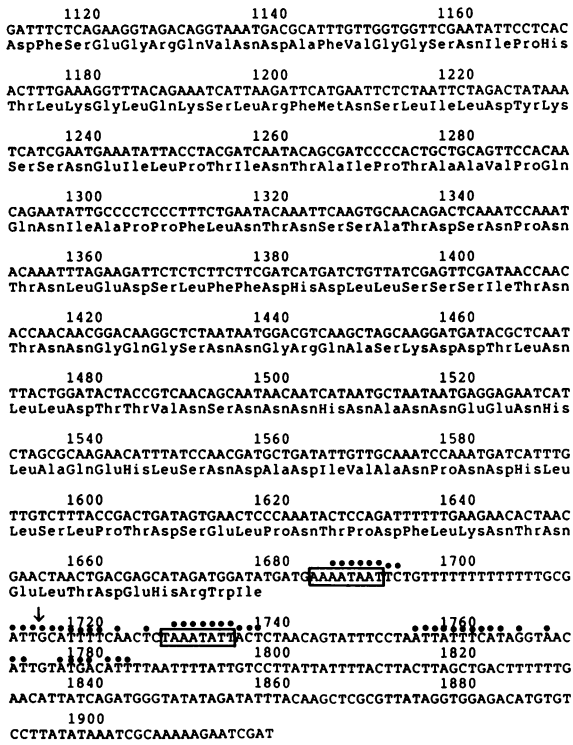


Fig. 3 DNA sequence of the HpaI/ClaI fragment and deduced amino acid sequence of the PHO2 gene. The possible TATA element at -87 and the two motifs at 1683 and 1727 assumed to be involved in transcription termination/poly(A) addition are boxed. Dots above the sequence represent the strong homology to the transcription termination region of the yeast ribosomal protein L17a and the arrow points to the 3' end of the L17a mRNA (19). The region between position 76 and 129 containing almost exclusively Gln residues is underlined as well as the acidic region between 880 and 987 containing only hydrophilic, mainly Asp and Asn residues. The region showing strong homology to PHO4 is denoted by a wavy line.

Tetrad analysis of one of these diploids is shown in Table I. The outcome of a vast majority of Pho^- spores indicates a tight linkage between the evicted gene and PHO2. In addition YS18 pho2::URA3 was crossed with IH22 (PHO2, ura3). Tetrad analysis showed that in 14 out of 15 tetrads Pho^- segregated with Ura^+, proving that the URA3 gene had inserted into the PHO2 gene. From these results we conclude that the long open

Table I Segregation of pho2 in a genetic cross YS18 pho2::URA3 x YS104 (pho2)

	Pho⁻	:	Pho⁺
	4:0	3:1	2:2
Expected if URA3 replaced PHO2	100	0	0
expected if URA3 replaced a DNA segment unlinked to PHO2	17	66	17
observed	18	5	1

Some of the unexpected Pho⁺ spores showed a constitutive APase activity. Those have not been furter analyzed. In all tetrades the Ura⁺ phenotype segregated 2:2.

reading frame on the 3.6 kb HindIII fragment codes for the PHO2 protein.

Localization of the PHO2 gene within the 3.6kb HindIII fragment

No sequences further upstream, than the HpaI site, and no sequences further downstream than the ClaI site (Fig. 2) are essential for PHO2 expression. This has been demonstrated by deleting these segments individually in a pDP39 clone carrying the 3.6kb HindIII fragment.

Neither construction gave rise to a reduced APase activity upon transformation of YS104.

Furthermore, a frameshift mutation has been introduced *in vitro* at the EcoRI site which is present in the putative N-terminal end of the protein at position 10 (see Fig. 3). The fact that this mutation abolishes APase activity proves that the PHO2 protein begins with one of the two met residues at position 1 or 4 (for DNA construction see Materials and Methods).

Deduced primary structure of the PHO2 protein

The protein starts at position +1 with two met residues, and two TGA codons in tandem determine the carboxyl end. The protein is very rich in Asn (78 residues), Asp (54), Gln (34) and His (17). On the other hand the aminoacids Cys, Gly, Val and Tyr are underrepresented with regard to an average aminoacid composition. Two hydrophilic regions of an extraordinary sequence content emerge. The N-terminal 80

Table II. Codon usage of the PHO2 gene

Phe	TTT	18	Ser	TTC*	7	Tyr	TAT	5	Cys	TGT*	1
Phe	TTC*	8	Ser	TCC*	5	Tyr	TAC*	3	Cys	TGC	2
			Ser	TCA	14						
Leu	TTA	12	Ser	TCG	9	His	CAT	12	Trp	TGG	4
Leu	TTG*	13	Ser	AGT	16	His	CAC*	5			
Leu	CTT	0	Ser	AGC	5				Arg	CGT	5
Leu	CTC	4				Gln	CAA*	25	Arg	CGC	1
Leu	CTA	9	Pro	CCT	6	Gln	CAG	9	Arg	CGA	0
Leu	CTG	8	Pro	CCC	6				Arg	CGG	0
			Pro	CCA*	12	Asn	AAT	48	Arg	AGA*	11
Ile	ATT*	12	Pro	CCG	2	Asn	AAC*	30	Arg	AGG	6
Ile	ATC*	4									
Ile	ATA	15	Thr	ACT*	19	Lys	AAA	12	Gly	GGT*	5
			Thr	ACC*	4	Lys	AAG*	14	Gly	GGC	2
Met	ATG	9	Thr	ACA	14				Gly	GGA	7
			Thr	ACG	3	Asp	GAT	38	Gly	GGG	1
Val	GTT*	5				Asp	GAC	16			
Val	GTC*	5									
Val	GTA	2	Ala	GCT*	10	Glu	GAA*	18			
Val	GTG	2	Ala	GCC*	4	Glu	GAG	6			
			Ala	GCA	9						
			Ala	GCG	2						

The numbers denote the use of each codon. The 22 preferred codons that preferentially occur in highly expressed genes are marked with an asterix. Underlined are "rare" codons that are preferentially used in the PHO2 gene.

aminoacids are very hydrophilic with a continuous stretch of 14 Gln only interrupted by 4 additional aminoacids (position 76-129). A second hydrophilic region occurs at position 880 to 987 (Fig. 3), is very acidic and contains continuous stretches of Asp and Asn.

Codon usage in the PHO2 gene

Table II shows the occurrence of each codon used in the PHO2 gene. Preferred codons used for highly expressed genes (13) are marked with an asterix. Evidently many "rare" codons occur in the PHO2 sequence. Rather the codon selection approximates very closely the one determined for lowly expressed yeast genes (14). The codon bias index (CBI) which has a value of 1, if only the 22 preferred codons are used and a value of 0 if all codons are used randomly (13), was calculated to be 0.04.

The 5' and 3' flanking regions of the PHO2 gene

The 5' region of most eukaryotic genes including those of yeast shows the presence of the model sequence $TATA^{T}_{A}A^{T}_{A}$ involved in mRNA start-site selection (15-18). In the PHO2 upstream region the sequence TATTAAA occurs at -87 constituting

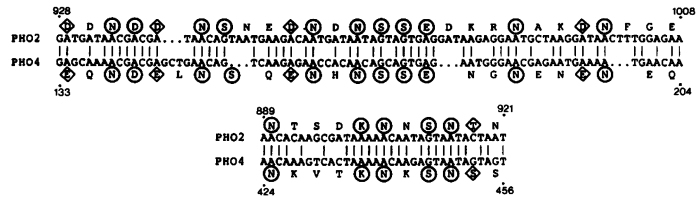


Fig. 4 Two homology regions between the PHO2 and PHO4 genes. Identical bases are denoted by a slash. Identical amino acids are surrounded by a circle and homologous aminoacids by a square. In both genes position 1 represents the first base of the ATG codon.

a possible TATA box. There is no ATG codon present between this potential promoter element and the ATG at +1, consistent with the observation that eukaryots use the 5' proximal ATG codon in the mRNA as start codon.

The 3'end of the gene contains two derivatives of the consensus sequence TAAATAA$^{A}_{G}$ (Fig. 3). This motif is found in many yeast genes and is assumed to function as a signal for transcription termination and/or poly(A)addition (13). Apart from that, the same region shows striking homology with the transcription termination region of yeast ribosomal protein L17a (19) where the 3'end of the mRNA has been mapped (Fig. 3). In analogy we suggest that the PHO2 mRNA ends at position 1713.

DISCUSSION

The 3.6 kb HindIII fragment with its 559 codon long ORF carries the PHO2 gene, since it is able to complement the pho2 mutation and inactivate the wild-type PHO2 gene upon transplacement of a mutated pho2 sequence into the chromosome. A frameshift introduced in vitro in the N-terminal part of the protein abolished PHO2 function. The gene could be narrowd down to a 2.3kb HpaI/ClaI fragment.

The deduced protein sequence shows a very unusual structure with a segment in the N-terminal part consisting mainly of Gln residues and another in the middle of PHO2 consisting of a continuous sequence of 36 aminoacids (mainly Asp and Asn) that are either charged or uncharged, but hydrophilic. Computer search in a datalibrary revealed that this second charged

region is strongly homologous with two regions of the PHO4 protein (20) on the aminoacid as well as on the DNA level. These homology regions are shown in Fig. 4.

Four models can be envisaged: the first states that the homologous regions are domains within the PHO2/PHO4 proteins that are involved in DNA binding at any of the defined (21) UASp elements of the PHO5 promoter. This hypothesis, however, is very unlikely due to the strong negative charges contained in these regions. Second, the homology regions could constitute interaction sites with other still unknown protein(s) (e.g. initiator proteins), or third, with the RNA polymerase. Fourth they could be interaction sites with the gene product of the negative factor PHO80. In any case, the observed homology rather argues for a transcriptional control involvement of PHO2 than for a post transcriptional action.

Two observations suggest that PHO2 is a lowly expressed gene. First, the codon bias index (CBI) is near to zero as was calculated for other regulatory genes like PHO4, GAL4 (20, 13) also expressed at a low level. Second, the 5' flanking region (100 bp 5' to ATG) does not show a higher A-T content (61%) than the coding region (63%) as was observed in the structural genes of the acid phosphatase gene family that are highly expressed (22).

Our motif for cloning and sequencing PHO2 was the intention to study the involvement of this regulatory protein in the phosphatase pathway. With the presented data we are now able to manipulate the PHO2 gene which should help us to obtain more insight in the nature of protein/DNA interactions in eukaryotic gene regulation.

ACKNOWLEDGEMENTS

We like to thank H. Rudolph for fruitful discussions, D. Primdore for supplying appropriate vector molecules and H. Rink for quick synthesis of oligonucleotides. Furthermore we wish to thank C. Widmer for typing the manuscript.

REFERENCES

1. Oshima, Y. (1982) in the Molecular Biology of the Yeast Saccharomyces. Strathern, J.N., Jones, E.W. and Broach, J.R. Eds., pp 159-180, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

2. Tait-Kamradt, A.G., Parent, S.A., LeVitre, J., Lifanova, O. and Bostian, K.A. (submitted for publication)
3. Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Cell, 40, 767-774.
4. Sherman, F., Fink, G.R. and Lawrence, C.W. (1983) in Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 61-64.
5. Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) J. Bact. 153, 163-168.
6. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., p. 440.
7. Boeke, J.D., Lacroute, F. and Fink, G.R. (1984) Mol. Gen. Genet. 197, 345-346.
8. Rudolph, H., Koening-Rauseo, I. and Hinnen, A. (1985) Gene 36, 87-95.
9. Hieter, P., Pridmore, D., Hegemann, J.H., Thomas, M., Davis, R.W. and Philippsen, P. (1985) Cell 42, 913-921.
10. Nilsson, B., Uhlén, M., Josephson, S., Gatenbeck, S. and Philipson, L. (1983) Nucl. Acids Res. 11, 8019-8030.
11. Meyhack, B., Bajwa, W., Rudolph, H. and Hinnen, A. (1982) EMBO J. 1, 675-680.
12. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
13. Bennetzen, J.L. and Hall, D.H. (1982) J. Biol. Chem. 257, 3018-3031.
14. Sharp, P.M., Tuohy, T.M.F. and Mosurski, R. (1986) Nucl. Acids Res. 14, 5125-5143.
15. Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J. and Kingsman, S.M. (1982) Nucl. Acids Res. 10, 2625-2637.
16. Nasmyth, K.A., Tatchell, K., Hall, B.D., Astell, C. and Smith, M. (1981) Nature 289, 244-250.
17. Corden, J., Wasylyk, B., Buckwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon P. (1980) Science 209, 1406-1414.
18. Holland, J.P. and Holland, M.J. (1980) J. Biol. Chem. 255, 2596-2605.
19. Leer, R.J., van Raamsdonk-Duin, M.M.C., Hagendoorn, M.J.M., Mager, W.H. and Planta, R.J. (1984) Nucl. Acids Res. 12, 6685-6700.
20. Legrain, M., De Wilde, M. and Hilger, F. (1986) Nucl. Acids Res. 14, 3059-3073.
21. Rudolph, H. and Hinnen, A. in press.
22. Bajwa, W., Meyhack, B., Rudolph, H., Schweingruber, A.M. and Hinnen, A. (1984) Nucl. Acids Res. 12, 7721-7739.