Gene, 113 (1992) 129–133

GENE 06383

Isolation and analysis of the *Penicillium chrysogenum phoA* gene encoding a secreted phosphate-repressible acid phosphatase

(Recombinant DNA; gene structure; filamentous fungi; intron, signal sequence; transcript mapping)

Hubertus Haas[a], Bernhard Redl[a], Ernst Friedlin[b] and Georg Stöffler[a]

[a] Institut für Mikrobiologie (Med. Fak.), Universität Innsbruck, 6020 Innsbruck (Austria); and [b] Abt.F+E, Biochemie Ges.m.b.H., 6330 Schaftenau (Austria) Tel. (43-5372)4931282

Received by J.R. Kinghorn: 3 October 1991
Revised/Accepted: 8 December/9 December 1991
Received at publishers: 16 January 1992

SUMMARY

We have isolated the genomic sequence encoding a secreted phosphate-repressible acid phosphatase (PHOA) from *Penicillium chrysogenum* using synthetic oligodeoxyribonucleotide probes. Nucleotide sequence data revealed that this gene consists of two exons of 192 and 1047 bp separated by an intron of 52 bp in length. A sequence encoding a putative signal peptide, resembling known signal sequences of fungi, was identified at the 5′-end of the coding sequence. Northern blot analysis of total cellular RNA indicated that the *phoA* gene codes for a 1.6-kb transcript. The expression of this gene is regulated at the transcriptional level and is markedly affected by the inorganic phosphate concentration of the growth medium.

INTRODUCTION

Under conditions of limited phosphate (P_i) microorganisms have to utilize organic compounds as an extracellular phosphate source. P_i must be removed from the organic substrate by action of various phosphatases (Pho). Several Pho have been isolated from yeasts, fungi and bacteria including non-secreted and secreted forms with acidic and alkaline pH-optima. Beside their general role in physiology, Pho have become important tools for investigation of protein secretion and gene regulation. Molecular and genetic analysis of Pho in lower eukaryotes have been mainly concerned with those of yeasts and *Aspergillus*. For instance, the secreted P_i-repressible acid Pho (PHO5) from *Saccharomyces cerevisiae* has been rigorously studied (Tohe et al., 1973; Arima et al., 1983; Bajwa et al., 1984). At the 5′-flanking region of the gene encoding this enzyme an upstream activating sequence responsible for transcriptional regulation was found (Bergman et al., 1986). Several genes encoding factors involved in this transcriptional regulation, have been identified and a genetic model has been proposed concerning the complex regulatory mechanism of acid Pho synthesis in *S. cerevisiae* (Oshima, 1982; Yoshida et al., 1989). Another yeast gene *PHO1* encoding a secreted acid Pho has been isolated from *Schizosaccharomyces pombe* (Elliot et al., 1986). It was analysed, but no signifi-

Correspondence to: Dr. B. Redl, Institut für Mikrobiologie (Med. Fak.), Universität Innsbruck, Fritz Pregl Str. 3, A-6020 Innsbruck (Austria) Tel. (43-512)5072255; Fax (43-512)507-2235.

Abbreviations: aa, amino acid(s); bp, base pair(s); HPLC, high-performance liquid chromatography; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame(s); *P.*, *Penicillium*; PA, polyacrylamide; PAGE, PA-gel electrophoresis; Pho, phosphatase(s); PHOA, *P. chrysogenum* P_i-repressible acid Pho; *phoA*, gene encoding PHOA; P_i, inorganic phosphate; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na_3·citrate pH 7.6; TMAC, tetramethylammonium chloride; *tsp*, transcription start point(s).

cant homology with *PHO5* was found and it was induced only two–three fold by P_i starvation. Since the 5′-flanking region has not yet been sequenced the presence of upstream activating sequences is unknown.

The gene *pacA* encoding a secreted P_i-repressible acid Pho from *Aspergillus niger* was isolated and characterized (MacRae et al., 1988). Although no homology with yeast acid Pho was found, this gene exhibited a sequence within the 5′-flanking region which resembled the *S. cerevisiae* upstream activator sequence, thus indicating a similar regulatory mechanism.

We have recently purified and characterized an acid Pho from crude filtrate of *Penicillium chrysogenum* (Haas et al., 1991). Total activity of this enzyme was low from mycelia grown in medium of high P_i, but was increased when mycelia were starved for P_i, indicating a P_i-mediated repression of the encoding gene. Here we describe the isolation and sequencing of this P_i-repressible acid Pho-encoding gene. It represents, to our knowledge, the first example of a *Penicillium* gene encoding a secreted enzyme and may be the basis for further investigation of the molecular mechanism of P_i-mediated regulation in *Penicillium*.

EXPERIMENTAL AND DISCUSSION

(a) Construction of probes and isolation of the acid Pho-encoding gene *phoA*

We have found that the N terminus of the purified PHOA protein was blocked (Haas et al., 1991). Therefore 2 nmol of the protein was digested with endoproteinase LysC and resulting fragments were separated by reversed-phase-HPLC. Peak fractions were collected and six peptides were subjected to aa sequencing. The following three partial aa sequences of different peptides have been chosen for further experimental work: P1–GluTyrGlnGluAspMetPro; P2–MetSerThrAspAsnTyr; P3–TyrValAsnTrpGluValAsp. Based on these peptides, oligo probes with degenerated sequences have been synthesized; oligoP1 was a 20-mer of 32-fold degeneration, oligoP2 was a 20-mer of 128-fold degeneration and oligoP3 was an 18-mer of 64-fold degeneration.

Genomic DNA from *P. chrysogenum* was digested with several restriction enzymes, electrophoresed on an agarose gel, Southern blotted, and hybridized using the three oligo mixtures. Hybridizations were performed at 37°C in $6 \times$ SSC/0.1% SDS (Sambrook et al., 1989) and the filters were washed in 3 M TMAC according to Wood et al. (1985). Since oligoP1 gave a strong signal this oligo mixture was used for screening filter replicas of a λEMBL 3 genomic library of *P. chrysogenum*. Three different strongly hybridizing clones could be isolated out of 6×10^4 clones. Subsequent mapping and hybridization experiments using the oligoP1 localized the *phoA* gene on a 3.3-kb *Hin*dII-*Xho*I segment. Two of the λ clones contained this whole fragment whereas the third carried only part of it.

(b) Sequence analysis

The 2.7-kb *Nde*I-*Xho*I fragment and in addition 0.2 kb of the 5′-flanking region of the 0.6-kb *Hin*dII-*Nde*I fragment were subjected to restriction with suitable enzymes, subcloned into pUC-vectors and sequenced by the dideoxy-termination method (Sanger et al., 1977). The strategy for sequencing these fragments is summarized in Fig. 1, the nt sequence and deduced aa sequence is shown in Fig. 2. The entire nt sequence comprised 1291 nt of the coding region including a 52 nt intron as well as 772 nt of the upstream region and 929 nt of the downstream region. As the coding region contained several aa sequences identical to those determined by protein sequencing of peptides of the purified PHOA protein it could be identified as the gene encoding this enzyme.

The 5′-untranslated region of the *phoA* gene shows characteristic features of a lower eucaryotic promoter. There is a perfect TATA-box (Gannon et al., 1979) at position −73 relative to the A of the putative start codon and a consensus region around the translational start site. This region includes the −3 nt position which is most frequently an A residue (Kozak et al., 1987). Within the *phoA* gene this position is also A. A putative CAAT consensus sequence is present at nt −287.

A twice repeated sequence, 5′-TTCCAAGGTT, is found at nt −560 and −402 (relative to the A of the start codon). In addition, the same sequence is found in duplicate but in inverse orientation starting at nt position −103 relative to the start codon. In *S. cerevisiae* a twice repeated sequence 10 nt in length is present in the 5′-flanking region of the P_i-repressible *PHO5* gene (Bergman et al., 1986). This sequence, starting at nt positions −391 and −332 was shown to be involved in transcriptional regulation of *PHO5* (Bergman et al., 1986). A twice repeated sequence, 12 nt in

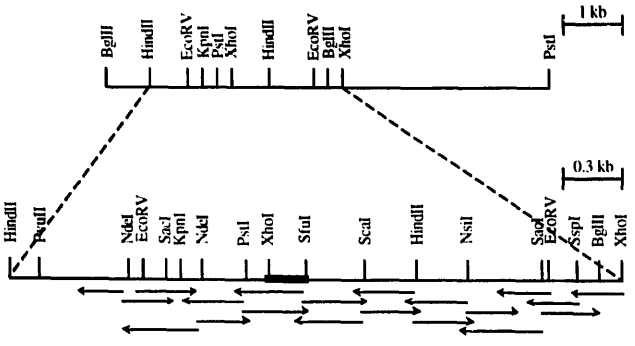

Fig. 1. Restriction map of a 8-kb *Bgl*II-*Pst*I fragment and a 3.3-kb *Hin*dII-*Xho*I fragment of *P. chrysogenum* genomic DNA containing the *phoA* gene. The oligoP1 hybridization region is indicated by a heavy line. Arrows below the restriction map denote sequencing strategy.

length, beginning at nt −306 and −332 is also found in the *A. niger* acid Pho gene *pacA* (McRae et al., 1990). Whether these sequences present in *Aspergillus* and *Penicillium* have a function as upstream activators similar to yeast has to be determined.

The sequence of the first 20 N-terminal aa residues of PHOA exhibits the properties of a typical signal sequence (Kreil, 1981; Von Heijne, 1985). This sequence contains a positively charged Lys near the N-terminal Met followed by 11 consecutive hydrophobic aa. Since the N terminus of the mature PHOA was blocked the putative signal peptide cleavage site was assigned according to the weight-matrix method of Von Heijne (1986). Using this method a cleavage site in the Gly^{19}/Thr^{20} position of the *phoA* gene product is most highly favoured, while Ala^{17}/Thr^{18} is a second, but slightly less favoured, possibility.

The coding region is interrupted by a small intron, 52 nt in length, extending from nt 193 to nt 244. This intron was assigned on the basis of the aa sequence of one of the PHOA peptides as indicated in Fig. 2. It has features consistently found in introns of genes of filamentous fungi (Ballance, 1986): (*1*) it is a small intron; (*2*) it begins with the sequence GTAGGT, ends with the TAG and within the intron lies the sequence GCTCAC which resembles the consensus RCTRAC (Rambosek and Leach, 1987) sequence.

A near approximation of the AATAAA motif, which is thought to be involved in polyadenylation appears in the 3′-flanking regions of several fungal genes (Gurr et al., 1987). With respect to this, the sequence AATAAT found 10 nt downstream from the stop codon may represent a possible polyadenylation signal. Additional polyadenylation signals may be represented by the sequences TATAAA and ATTAAA found 133 nt and 169 nt downstream from the stop codon.

An inverted repeat representing a putative stem-loop structure could be found 46 nt downstream from the stop codon. It is of interest to note that the sequence 5′–TTCCAAGGTT found in the 5′-flanking region is also present 561 nt downstream from the stop codon.

Based on the aa sequence derived from the nt sequence presented here, the calculated M_r of the mature PHOA is 42000. Six potential N-glycosylation sites, Asn-Xaa-Thr/Ser (Pless and Lennarz, 1977) were found in the deduced aa sequence.

When comparing the aa sequence of PHOA with that of acid Pho from different organisms, no significant homology could be found. However, PHOA shares 59% aa identity with a putative protein from *Kluyveromyces lactis*. Statistical analysis by the program ALIGN (Dayhoff et al., 1983)

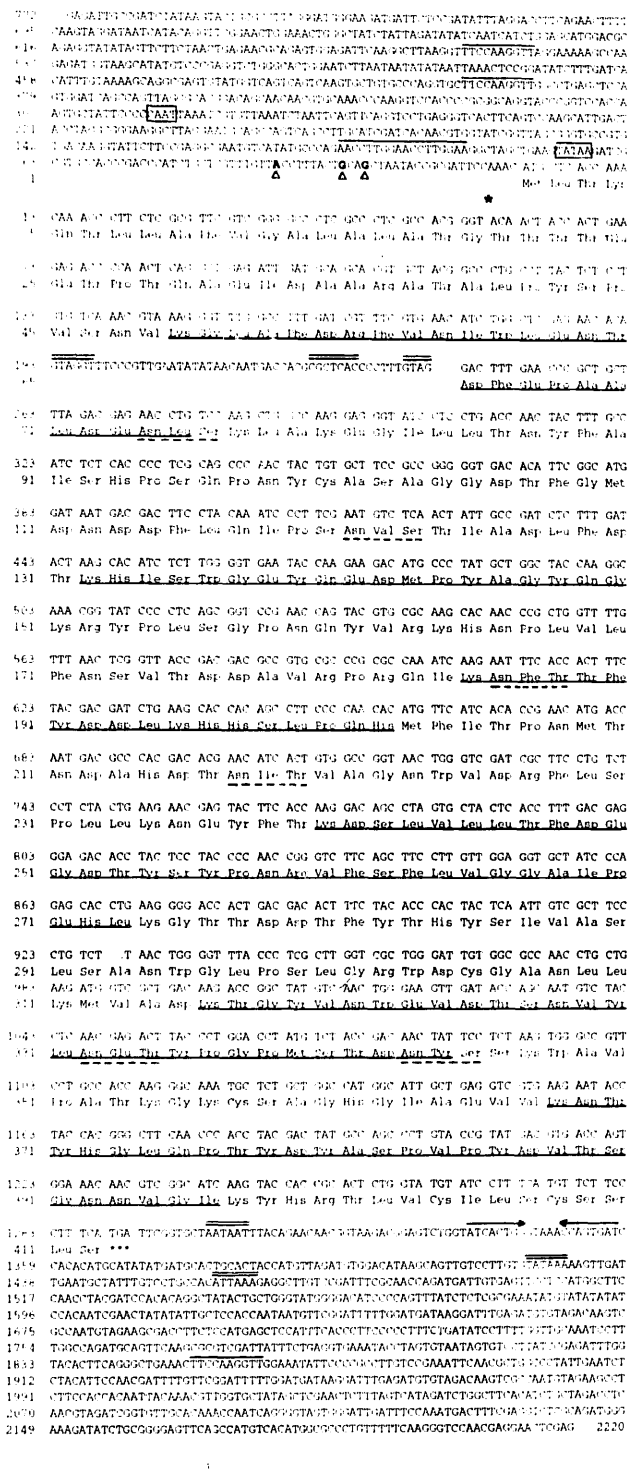

Fig. 2. The nt sequence of the *phoA* gene and the deduced aa sequence (GenBank accession No. M80366). The nt numbering begins from the A in the start codon. The aa residues 1–19 comprise a putative signal peptide and the most favoured signal-peptide-cleavage site is indicated by an asterisk. The three *tsp* are marked in bold letters and open arrowheads. The CAAT and TATA sequences upstream from *tsp* are boxed. The consensus sequences of the intron and for transcription termination are doubly overlined. The repeated sequence 5′–TTCCAAGGTT is overlined. The inverted repeat downstream from the stop codon is indicated by convergent arrows. Underlined aa sequences have been confirmed by direct aa sequencing. Putative N-glycosylation sites are underlined by dashes. Non-coding sequences are shown in lower-case letters. Three asterisks mark stop codon.

gives a highly significant similarity (32 standard deviation units) between these two aa sequences. The gene encoding this putative protein was found to be transcribed by a 4.7-kb mRNA species together with an ORF coding for an inducible lactose permease (Chang and Dickson, 1988).

(c) Codon usage

The codon usage exhibited by *phoA* is similar to that observed in other *P. chrysogenum* genes, e.g., orotidine-5-phosphate decarboxylase (Cantoral et al., 1988), isopenicillin N synthetase (Carr et al., 1986). In all 57 out of 61 possible codons are used. In general, it supports the observation that filamentous fungi show a less marked degree of bias in codon usage than yeast (Ballance, 1986; Gurr et al., 1987). As seen in highly expressed genes of *Aspergillus* and *Neurospora* a number of generalizations (Gurr et al., 1987) also hold for the codon usage pattern of *phoA*. Pyrimidines, especially C dominate the third position (72%). When a purine is used in the third position, G is preferred over A. AGN codons for Ser and Arg are infrequently used.

(d) Transcript mapping and regulation of transcription

The size of the *phoA* mRNA in vivo was estimated by Northern blot analysis and the transcript was found to be

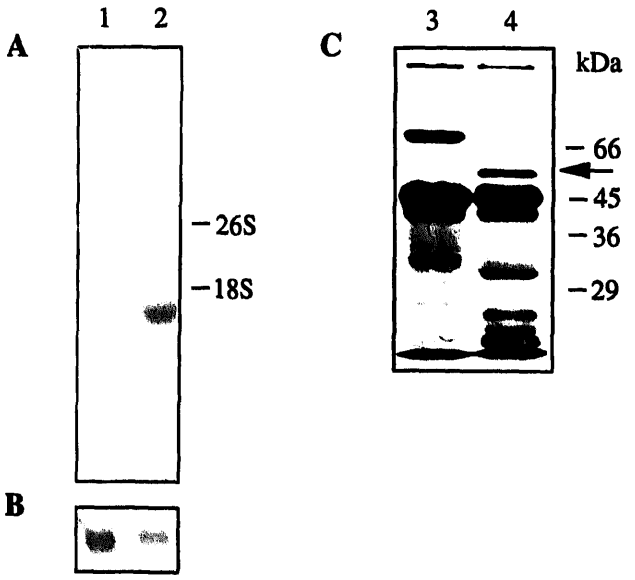

Fig. 3. Regulation of the *phoA* gene expression. (Panel A) Northern blot of *P. chrysogenum* RNA. RNA was isolated from mycelia (Chomczynski and Sacchi, 1987) grown in the presence of 30 mM P_i (lane 1) or in the absence of P_i (lane 2). Total RNA was electrophoresed on a 1.5% agarose-2.2 M formaldehyde gel, blotted onto Hybond-N and hybridized (Sambrook et al., 1989) with the ^{32}P-labelled 0.8-kb *Pst*I-*Sac*I fragment. (Panel B) The blot was subsequently stripped and rehybridized with a probe specific for γ-actin from *Aspergillus nidulans* (Fidel et al., 1988). (Panel C) SDS–PAGE (Laemmli, 1970) of proteins of cleared filtrate from mycelia grown in the presence of 30 mM P_i (lane 3) or in the absence of P_i (lane 4). PA-gels were stained with Coomassie Blue. PHOA is indicated by an arrow.

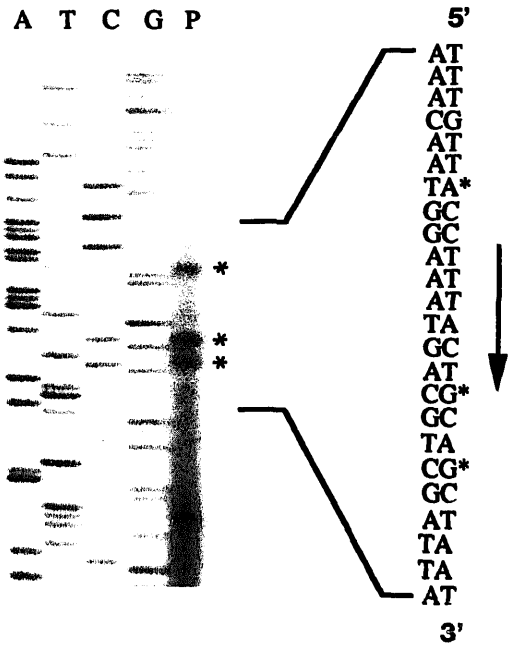

Fig. 4. Primer extension. Analysis of *tsp* was carried out as described by Geliebter et al. (1986) using primer 5'-GTTGGGGTCTCTTCAGTGGTAGTT. Lanes: **P**, primer extension products; **A**, **T**, **C** and **G**, dideoxy sequencing ladders of the *phoA* gene using the same primer. Asterisks mark the major *tsp* and their positions on the corresponding nt sequence. The direction of transcription is indicated by an arrow.

1.6 kb (Fig. 3, panel A). Northern blot analysis of RNA isolated from mycelia grown in medium of high or low P_i indicated a regulation by P_i at the transcriptional level. Transcription of *phoA* was repressed in mycelia grown in high P_i medium (Fig. 3, panel A). This result is consistent with observations obtained after analysis of proteins secreted into the medium by SDS–PAGE. No PHOA was found when mycelia were grown in high P_i medium, but the enzyme was present when mycelia were starved for P_i (Fig. 3, panel C). The regulation of the *P. chrysogenum* enzyme therefore seems to be similiar to acid Pho from *S. cerevisiae* and *A. niger*. To define the *tsp*, in vivo transcripts from *P. chrysogenum* were characterized by primer extension (Fig. 4) and mung bean nuclease mapping (H.H., unpublished). Both methods determined the *tsp* to be at 33, 24 and 21 nt upstream from the start codon. A tendency towards multiple transcriptional initiation has been observed in most of the genes of filamentous fungi (Gurr et al., 1987).

(e) Conclusions

(*1*) The genomic sequence from *P. chrysogenum* encoding a secreted acid Pho has been isolated by using synthetic oligo probes of peptide derived aa sequences of the purified enzyme.

(*2*) The *phoA* gene was located to a 3.3-kb *Hin*dII-*Xho*I

fragment and the coding region comprised 1291 bp including a 52-bp intron. It encodes a transcript of approx. 1.6 kb.

(*3*) At the N terminus of the proposed translation product the presence of a signal peptide was suggested with all properties of a typical signal sequence. The putative signal peptidase cleavage site was assigned at Gly^{19}/Thr^{20}.

(*4*) No significant aa sequence homology of the *P. chrysogenum* enzyme with acid Pho from several organisms was found. However, it shares 59% aa identity with a putative protein from *Kluyveromyces lactis*.

(*5*) Expression of the gene was found to be regulated at the level of mRNA accumulation and was dependent on the P_i concentration in the growth medium.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. G. Weber for kindly providing us with a λEMBL3 genomic library of *P. chrysogenum* and Dr. P. Loidl for his comments on the manuscript. We thank Dr. J.R. Kinghorn for the plasmid containing the γ-actin sequence from *A. nidulans*.

REFERENCES

Arima, K., Oshima, T., Kubota, I., Nakamura, N., Mizunaga, T. and Toh-e, A.: The nucleotide sequence of the yeast *PHO5* gene: a putative precursor of repressible acid phosphatase contains a signal peptide. Nucleic Acids Res. 11 (1983) 1657–1672.

Bajwa, W., Meyhack, B., Rudolph, H., Schweingruber, A.-M. and Hinnen, A.: Structural analysis of the two tandemly repeated acid phosphatase genes in yeast. Nucleic Acids Res. 12 (1984) 7721–7739.

Ballance, D.J.: Sequences important for gene expression in filamentous fungi. Yeast 2 (1986) 229–236.

Bergman, L.W., McClinton, D.C., Madden, S.L. and Preis, L.H.: Molecular analysis of the DNA sequence involved in the transcriptional regulation of the phosphate-repressible acid phosphatase gene (*PHO5*) of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 83 (1986) 6070–6074.

Cantoral, J.M., Barredo, J.L., Alvarez, E., Diaz, B. and Martin, J.F.: Nucleotide sequence of the *Penicillium chrysogenum pyrG* (orotidine-5'-phosphate decarboxylase) gene. Nucleic Acids Res. 16 (1988) 8177–8177.

Carr, L.G., Skatrud, P.L., Scheetz, M.E., Queener, S.W. and Ingolia, T.D.: Cloning and expression of the isopenicilin N synthase gene from *Penicillium chrysogenum*. Gene 48 (1986) 257–266.

Chang, Y.D. and Dickson, R.C.: Primary structure of the lactose permease gene from the yeast *Kluyveromyces lactis*. J. Biol. Chem. 263 (1988) 16696–16703.

Chomczynski, P. and Sacchi, N.: Single-step method of RNA isolation by guanidinium thiocyanat-phenol-chloroform extraction. Anal. Biochem. 162 (1987) 156–157.

Dayhoff, M.O., Barker, W.C. and Hunt, L.T.: Establishing homologies in protein sequences. Methods Enzymol. 91 (1983) 542–545.

Elliot, S., Chang, C.-W., Schweingruber, M.E., Schaller, J., Rickli, E.E. and Carbon, J.: Isolation and characterization of the structural gene for secreted acid phosphatase from *Schizosaccharomyces pombe*. J. Biol. Chem. 261 (1986) 2936–2941.

Fidel, S., Doonan, J.H. and Morris, N.R.: *Aspergillus nidulans* contains a single actin gene which has unique intron locations and encodes a gamma-actin. Gene 70 (1988) 283–293.

Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P.: Organization and sequences at the 5' end of cloned complete ovalbumin gene. Nature 278 (1979) 428–434.

Geliebter, J., Zeff, R.A., Melvold, R.W. and Nathenson, S.G.: Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis. Proc. Natl. Acad. Sci. USA 83 (1986) 3371–3375.

Gurr, S.J., Unkles, S.E. and Kinghorn, J.R.: The structure and organization of nuclear genes of filamentous fungi. In: Kinghorn, J.R. (Ed.), Gene Structure in Eucaryotic Microbes. IRL Press, Oxford, 1987, pp. 93–193.

Haas, H., Redl, B., Leitner, E. and Stöffler, G.: *Penicillium chrysogenum* extracellular acid phosphatase: purification and biochemical characterization. Biochim. Biophys. Acta 1047 (1991) 392–397.

Kozak, M.: An analysis of 5'-noncoding sequences from 699 vertebrate mRNAs. Nucleic Acids Res. 15 (1987) 8125–8148.

Kreil, G.: Transfer of proteins across membranes. Ann. Rev. Biochem. 50 (1981) 317–348.

Laemmli, U.K.: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227 (1970) 680–685.

MacRae, W.D., Buxton, F.P., Sibley, S., Garven, S., Gwynne, D.I., Davies, R.W. and Arst Jr., H.N.: A phosphate-repressible acid phosphatase gene from *Aspergillus niger*: its cloning, sequencing and transcriptional analysis. Gene 71 (1988) 339–384.

Oshima, Y. Regulatory circuits for gene expression: the metabolism of galactose and of phosphate. In: Strathern J.N., Jones E.W. and Broach, J.R. (Eds.), The Molecular Biology of the Yeast *Saccharomyces cerevisiae*: Metabolism and Gene Expression. Cold Spring Harbor, NY, 1982, pp. 159–180.

Pless, D.D. and Lennarz, W.J.: Enzymatic conversion of proteins to glycoproteins. Proc. Natl. Acad. Sci. USA 74 (1977) 134–138.

Rambosek, J. and Leach, J.: Recombinant DNA in filamentous fungi: progress and prospects. Crit. Rev. Biotechnol. 6 (1987) 357–393.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Sambrook, J., Fritsch, E.F. and Maniatis, T.: Molecular Cloning. A Laboratory Manual, 2rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Toh-e, A., Ueda, Y., Kakimoto, S. and Oshima, Y.: Isolation and characterization of acid phosphatase mutants in *Saccharomyces cerevisiae*. J. Bacteriol. 120 (1973) 608–617.

Von Heijne, G.: Signal sequences: the limits of variation. J. Mol. Biol. 184 (1985) 99–105.

Von Heijne, G.: A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14 (1986) 4683–4690.

Wood, W.I., Gitachier, J., Lasky, L.A. and Lawn, R.M.: Base composition-independent hybridization in tetramethylammonium chlorid: a new method for oligonucleotide screening of highly complex gene libraries. Proc. Natl. Acad. Sci. USA 82 (1985) 1585–1588.

Yoshida, K., Ogawa, N. and Oshima, Y.: Function of the *PHO* regulatory genes for repressible acid phosphatase synthesis in *Saccharomyces cerevisiae*. Mol. Gen. Genet. 217 (1989) 40–46.