Gene, 138 (1994) 67–77
© 1994 Elsevier Science B.V. All rights reserved. 0378-1119/94/$07.00
SSDI 0378-1119 (93) E0656-X

GENE 07660

Structure and genomic organization of the *ipiB* and *ipiO* gene clusters of *Phytophthora infestans*

(Filamentous fungi; gene structure; glycine-rich protein; pathogenicity gene; potato late blight; oomycetes)

Corné M.J. Pieterse*, Pieter van West, Henk M. Verbakel*, Paul W.H.M. Brassé, Grardy C.M. van den Berg-Velthuis and Francine Govers

Department of Phytopathology, Wageningen Agricultural University, Wageningen, The Netherlands

Received by J.R. Kinghorn: 20 June 1993; Revised/Accepted: 24 September/26 September 1993; Received at publishers: 11 October 1993

SUMMARY

Two *in planta*-induced (*ipi*) genes, designated *ipiB* and *ipiO*, of the potato late blight fungus, *Phytophthora infestans* (Mont.) de Bary, were isolated from a genomic library by a differential hybridization procedure [Pieterse et al., Physiol. Mol. Plant Pathol. (1993a) in press]. Both genes are expressed at high levels in the early phases of the pathogenic interaction of *P. infestans* with its host plant potato, suggesting that their gene products have a function in the early stages of the infection process. Here, we describe the nucleotide (nt) sequence and genomic organization of *ipiB* and *ipiO*. The *ipiB* gene belongs to a small gene family consisting of at least three genes, designated *ipiB1*, *ipiB2* and *ipiB3*, which are clustered in a head-to-tail arrangement. The three *ipiB* genes are highly homologous throughout the coding regions and 5′ and 3′ flanking regions. The *P. infestans* genome contains two very similar *ipiO* genes, *ipiO1* and *ipiO2*, which are closely linked and arranged in an inverted orientation. The *ipiB* genes encode three novel, highly similar Gly-rich proteins of 301, 343 and 347 amino acids (aa), respectively. The Gly-rich domains of the IPI-B proteins are predominantly composed of two repeats with the core sequences, A/V-G-A-G-L-Y-G-R and G-A-G-Y/V-G-G. The *ipiO* genes code for two almost identical 152-aa proteins which do not have any homology with sequences present in data libraries. IPI-B, as well as IPI-O, contains putative signal peptides of 20 and 21 aa, respectively, suggesting that they are transported out of the cytoplasm. In the promoter regions of *ipiB* and *ipiO*, a 16-nt sequence motif, matching the core sequence, GCTCATTYYNCAWTTT (where N = A or C or G or T; W = A or T; Y = C or T), was found. This sequence motif appears to be present around the transcription start point (*tsp*) of seven out of eight oomycetous genes for which the *tsp* have been determined, suggesting that oomycetes have a sequence preference for transcription initiation.

INTRODUCTION

The oomycetous fungus *Phytophthora infestans* (Mont.) de Bary is the causal agent of the devastating late blight disease on potato (*Solanum tuberosum* L.) and tomato (*Lycopersicon esculentum* Mill.). Molecular studies on the potato-*P. infestans* interaction have demonstrated that pathogen attack activates genes in the

Correspondence to: Dr. F. Govers, Department of Phytopathology, Wageningen Agricultural University, Binnenhaven 9, P.O. Box 8025, 6700 EE Wageningen, The Netherlands. Tel. (31-8370) 83138; Fax (31-8370) 83412; e-mail: GOVERS@rcl.wau.nl

*Present addresses: (C.M.J.P.) Department of Plant Ecology and Evolutionary Biology, University of Utrecht, P.O. Box 800.84, 3508 TB Utrecht, The Netherlands. Tel. (31-30) 537432; e-mail: CORNEP@boev.biol.ruu.nl; (H.M.V.) Keygene nv, P.O. Box 216, 6700 AE Wageningen, The Netherlands. Tel. (31-8370) 24141.

Abbreviations: aa, amino acid(s); bp, base pair(s); DHC, differentially hybridizing clone; *ipiB* and *ipiO*, *in planta*-induced genes *B* and *O*; IPI-B and IPI-O, proteins encoded by *ipiB* and *ipiO*; kb, kilobase(s) or 1000 bp; N, any nucleoside; nt, nucleotide(s); ORF, open reading frame; *P.*, *Phytophthora*; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na_3-citrate pH 7.6; *tsp*, transcription start point(s); W, A or T; Y, pyrimidine (C or T).

host plant (Choi et al., 1992; Fritzemeier et al., 1987; Hahlbrock et al., 1989; Martini et al., 1993; Schröder et al., 1992; Taylor et al., 1990). Many of these genes encode products which are thought to be involved in the inhibition of pathogen development. Also in the pathogen, interaction with the host plant is accompanied by the activation of certain genes (Pieterse et al., 1991; 1992; 1993a,b). Products of these so-called *in planta*-induced (*ipi*) genes may be necessary for establishment and maintenance of basic pathogenicity or for the increase of disease severity. Characterization of *P. infestans* genes of which the expression is specifically induced *in planta* may, therefore, lead to the identification of so far unknown pathogenicity factors.

Recently, we described the selection of nine *in planta*-induced genes by differential screening of a genomic library of *P. infestans* DNA using first-strand cDNA probes synthesized on (*i*) mRNA isolated from *P. infestans*-infected potato leaves and (*ii*) mRNA isolated from *P. infestans* grown in vitro (Pieterse et al., 1993a). A detailed characterization of two of these *in planta*-induced genes, *ubi3R* and *calA*, showed that they encode polyubiquitin and calmodulin, respectively (Pieterse et al., 1991; 1993b). Ubiquitin plays a key role in several cellular processes such as selective degradation of intracellular proteins, maintenance of chromatin structure, regulation of gene expression and modification of cell-surface receptors (Monia et al., 1990). Calmodulin is a calcium-binding protein which is known to play an essential role in basic cellular processes such as signal transduction, ion transport and cytoskeleton function (Cheung, 1980). Both *ubi3R* and *calA* are expressed during growth of the fungus in vitro but during pathogenesis on potato, the expression levels are consistently fivefold higher. In contrast to *ubi3R* and *calA*, two other *in planta*-induced genes, *ipiB* and *ipiO*, show a transient expression pattern during pathogenesis with the highest expression level in the early stages of infection (Pieterse et al., 1993a). It appears that both *ipiB* and *ipiO* belong to small, clustered gene families. In this paper we describe the molecular characterization and genomic organization of the members of the *ipiB* and *ipiO* gene clusters.

RESULTS AND DISCUSSION

(a) Isolation and genomic organization of the *ipiB* and *ipiO* genes

With the aim to select *P. infestans* genes whose expression is induced or significantly increased during pathogenesis on potato, a genomic library of *P. infestans* DNA was constructed in λEMBL3 and differentially screened as described previously (Pieterse et al., 1993a). The differential screening resulted in the selection of several genomic clones. Two of these differentially hybridizing clones (DHCs), DHC-B and DHC-O, contain *in planta*-induced genes which are highly expressed in the early stages of infection (Pieterse et al., 1993a). The *in planta*-induced genes located on DHC-B and DHC-O were designated *ipiB* and *ipiO*, respectively. The approximate location of the coding regions of the *ipiB* and *ipiO* genes on DHC-B and DHC-O was assessed by Southern blot analyses. Blots containing digested DNA of DHC-B and DHC-O were hybridized with a labeled first-strand cDNA probe which was synthesized on poly(A)$^+$RNA isolated from *P. infestans*-infected potato leaves, two days post-inoculation (interaction cDNA probe). In this way, the DHC-B and DHC-O fragments containing transcribed sequences which correspond to the coding regions of the *ipiB* and *ipiO* genes were identified (indicated with a closed bar in Fig. 1a and 1c). These fragments were subcloned and a detailed restriction endonuclease profile of the DNA surrounding the *ipiB* and *ipiO* genes was determined (Fig. 1b and 1d). In both cases, repetition of specific endonuclease profiles was observed. Cross-hybridization experiments showed that these repeated areas are highly homologous suggesting the presence of a cluster of similar genes.

To determine which restriction fragments contain the coding regions of the *ipiB* and *ipiO* genes, Southern blots of digested DHC-B and DHC-O subclones were hybridized using labeled interaction cDNA as probe. Of the DHC-B subclones, one *Sst*I-*Hin*cII fragment of 0.9 kb and two 0.98-kb *Sst*I-*Pst*I fragments hybridized with the interaction cDNA probe, whereas of the DHC-O subclones two 0.63-kb *Sst*I-*Xba*I fragments hybridized (indicated with dotted lines in Fig. 1b and 1d). This indicates that the coding regions of the *ipiB* and *ipiO* genes are constrained within these respective DNA fragments. On Northern blots containing RNA isolated from *P. infestans*-infected potato leaves, probes derived from all three DHC-B fragments hybridized to a mRNA transcript of 1200 nucleotides (nt) in length (Pieterse et al., 1993a). Probes derived from the two DHC-O fragments both hybridized to a mRNA transcript of 600 nt in length. Considering the repetitive restriction endonuclease profile, the cross-hybridization, and the size of the DNA fragments in comparison to the length of the hybridizing mRNA transcripts, it can be concluded that DHC-B, as well as DHC-O, contains gene clusters with three and two highly homologous genes, respectively. The genes located on DHC-B were designated *ipiB1*, *ipiB2* and *ipiB3* (Fig. 1b), the ones on DHC-O *ipiO1* and *ipiO2*, respectively (Fig. 1d). There is no cross-hybridization between the *ipiB* and *ipiO* genes.

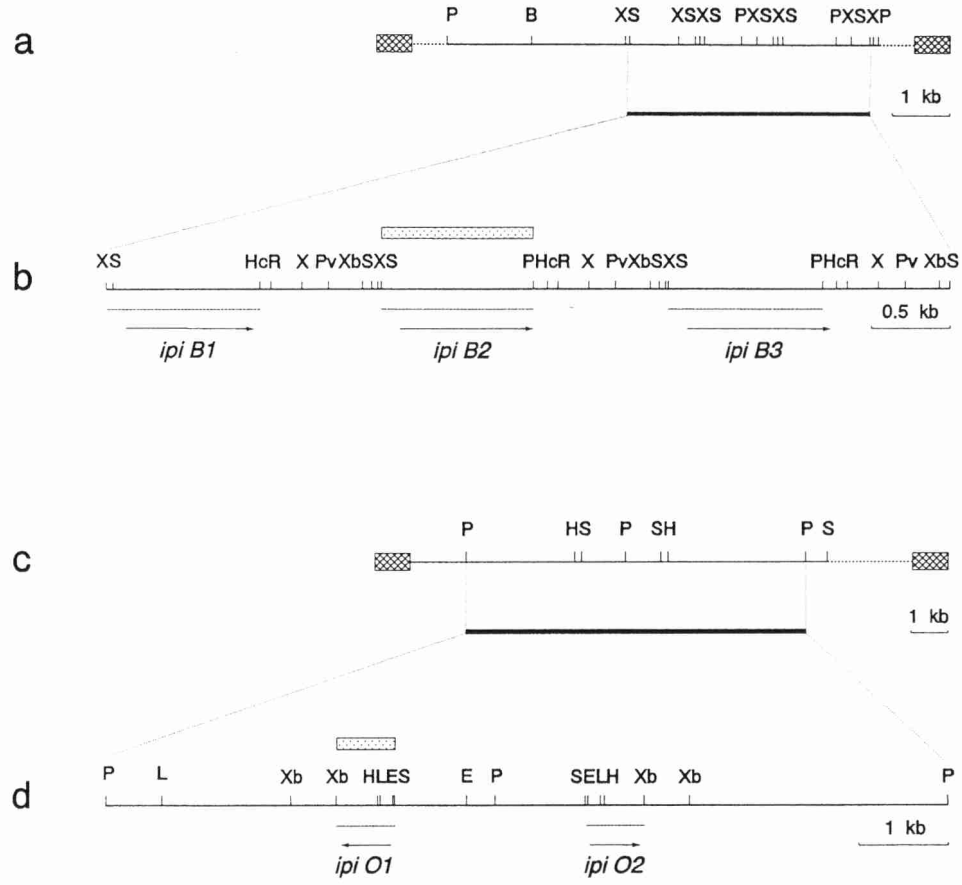

Fig. 1. Organization of the *P. infestans ipiB* and *ipiO* genes. **a** and **c**; partial restriction map of λEMBL3 recombinant phages DHC-B (**a**) and DHC-O (**c**) which were isolated from a genomic library of *P. infestans* DNA by differential hybridization as described by Pieterse et al. (1993a). Crosshatched bars represent λEMBL3 arms. Discontinuous lines represent DNA regions of unknown length and restriction endonuclease profile. Closed bars indicate the approximate position of the coding regions of the *ipiB* and *ipiO* genes as identified by Southern blot analysis of DHC-B and DHC-O restriction fragments, using as probe first strand cDNA synthesized on poly(A)$^+$RNA isolated from *P. infestans*-infected potato leaves, two days post-inoculation (interaction cDNA probe). RNA isolation and cDNA synthesis were performed as described previously (Pieterse et al., 1993a). **b** and **d**; Restriction map of DNA regions from DHC-B and DHC-O in which *ipiB* and *ipiO* genes are located. Dotted lines show the positions of the coding regions of the *ipiB* (**b**) and *ipiO* (**d**) genes, respectively. This was determined by Southern blot analysis of blots containing insert DNA of a number of subclones derived from the shown DNA region, hybridized with labeled interaction cDNA as probe. Arrows indicate positions and directions of the coding sequences of the *ipiB* and *ipiO* genes as assessed by dideoxy sequencing. Stippled bars represent the DNA fragments which were used as probes for hybridization of genomic Southern blots and for the isolation of the *ipiO1* cDNA clone from the λZAP cDNA library. Restriction sites: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*cII; L, *Sal*I; P, *Pst*I; Pv, *Pvu*II; R, *Eco*RV; S, *Sst*I; X, *Xho*I; Xb, *Xba*I.

(b) Southern blot analyses of genomic *P. infestans* DNA

To determine the copy number of the *ipiB* and *ipiO* genes in the *P. infestans* genome, Southern blot analyses were performed. Blots containing digested genomic *P. infestans* DNA were hybridized with the *ipiB2* containing 0.98-kb *Sst*I-*Pst*I fragment from DHC-B and with the *ipiO1* containing 0.63-kb *Sst*I-*Xba*I fragment from DHC-O, respectively (indicated with dotted bars in Fig. 1b and 1d). The *ipiB2* probe hybridized to approximately ten *Pst*I, *Xho*I and *Sst*I fragments (Fig. 2a). Only the 7.0- and 1.8-kb *Pst*I, the 1.4- and 1.2-kb *Xho*I and the 1.8- and 1.7-kb *Sst*I fragments correspond to restriction fragments present in DHC-B, indicating that there are other *ipiB* genes or *ipiB*-like sequences present in the *P. infestans* genome. The *ipiO1* probe hybridized to two *Pst*I fragments of 4.3 kb and 5.0 kb in length and to two *Sst*I fragments of 6.5 kb and 4.5 kb in length (Fig. 2b). These restriction fragments match with those found in DHC-O and two overlapping λ clones, DHC-O′ and DHC-O″ (data not shown). It can thus be concluded that the *ipiO* gene cluster present in DHC-O is unique in the *P. infestans* genome. Under the hybridization conditions used, the *ipiB* nor the *ipiO* probe hybridized to potato DNA (Fig. 2), indicating that there are no highly similar sequences present in the potato genome.

(c) Nucleotide sequence of the *ipiB* genes

The nt sequence of the *Xho*I-*Sst*I fragment of 5424 bp in length (Fig. 1b), comprising the coding regions of *ipiB1*, *ipiB2* and *ipiB3*, was determined by dideoxy sequencing and is shown in Fig. 3. Three highly homologous open reading frames (ORFs) of 903, 1029 and 1041

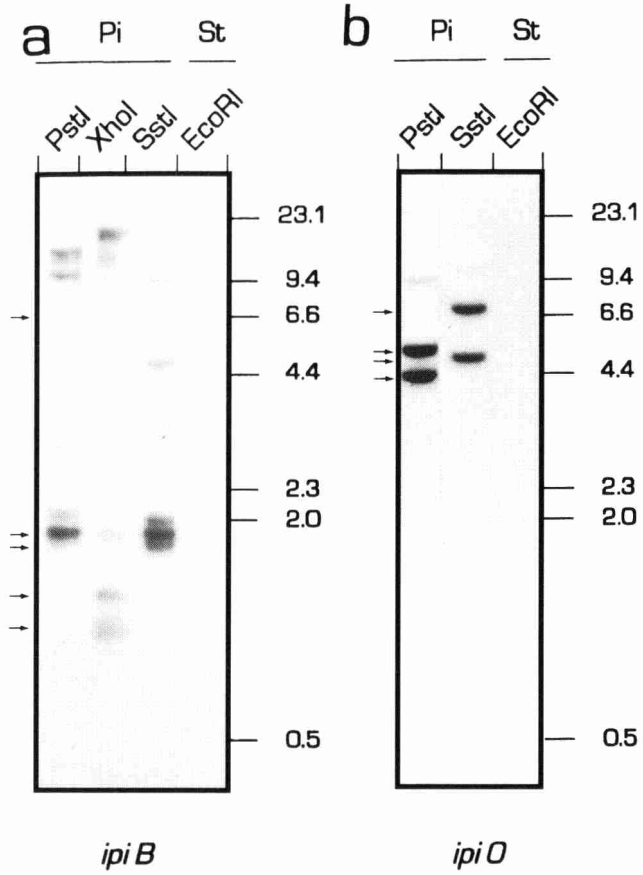

Fig. 2. Autoradiographs of Southern blots containing genomic DNA of (Pi) *P. infestans* (strain 88069) and (St) potato (cultivar Ajax), hybridized with the *ipiB* and *ipiO* probes. Genomic DNA was isolated as described by Pieterse et al. (1991). Genomic DNA (10 µg) was digested with restriction endonucleases *Eco*RI, *Pst*I, *Sst*I or *Xho*I. After electrophoresis on a 0.7% agarose gel, Southern blotting was performed on Hybond-N$^+$ membranes (Amersham) according to the manufacturers instructions. Blots were hybridized with probes derived from the 0.98-kb *Sst*I-*Pst*I fragment of DHC-B (**a**) and the 0.63-kb *Sst*I-*Xba*I fragment of DHC-O (**b**) as indicated by the stippled bars in Fig. 1b and 1d, respectively. Probes were labeled by random primer labelling (Feinberg and Vogelstein, 1983). Hybridization was performed overnight in 0.5 M Na_2HPO_4/NaH_2PO_4 (pH 7.2), 7% SDS/1 mM EDTA at 65°C. Blots were subsequently washed in 0.2 × SSC, 0.1% SDS at 65°C and exposed to Kodak X-OMAT S film. Molecular size markers are indicated in kb. Arrows indicate hybridizing DNA fragments present in DHC-B and DHC-O, respectively.

nt were found at positions where the coding regions of the *ipiB* genes were predicted (Fig. 1b). The lengths of the ORFs are in agreement with the size of the 1200-nt *ipiB* mRNA when adding 5′ and 3′ non-translated regions. The distances between the ORFs of *ipiB1* and *ipiB2*, and between *ipiB2* and *ipiB3* are 820 and 819 nt, respectively. When allowing gaps for optimal alignment, the coding regions of the *ipiB* genes are 96% identical whereas the 0.82-kb intergenic DNA sequences are for 98% the same. A DNA region highly homologous to the 0.82-kb intergenic DNA sequences is also present immediately downstream from the *ipiB3* coding region, suggesting the presence of a fourth gene succeeding the *ipiB3* gene. However, the DNA regions surrounding the sequenced 5424-bp *Xho*I-*Sst*I fragment do not hybridize to the *ipiB2* probe (data not shown), indicating that there are no additional *ipiB* genes in the direct vicinity of this *ipiB* gene cluster. The 2.2-kb *Bam*HI-*Sst*I fragment preceding the 5424-bp *Xho*I-*Sst*I fragment shows cross-hybridization with a probe derived from the 0.82-kb intergenic region, suggesting that the 5′ regulatory sequences of *ipiB1* are similar to those of *ipiB2* and *ipiB3* (data not shown).

(d) Nucleotide sequence of the *ipiO* genes

The nt sequence of the 3440-bp *Xba*I fragment, comprising the coding regions of *ipiO1* and *ipiO2* was determined by dideoxy sequencing (Fig. 4). In addition, the DNA sequence of a partial *ipiO* cDNA clone was assessed. This *ipiO* cDNA clone was isolated from a λZAP cDNA library representing poly(A)$^+$ RNA from *P. infestans*-infected potato leaves, two days post-inoculation. The library was screened with the 0.63-kb *Sst*I-*Xba*I fragment from DHC-O (Fig. 1d). In Fig. 4, lines indicated by (a) show the nt sequence of the *Xba*I-*Pst*I fragment on which *ipiO1* is located. Lines indicated by (b) show the nt sequence of the adjacent *Pst*I-*Xba*I fragment containing *ipiO2*. Two ORFs of 456 nt were found at positions where the coding regions of the *ipiO* genes were predicted (Fig. 1d). The 518 nt sequence of the partial *ipiO* cDNA clone is identical to the *ipiO1* sequence from nt +51 relative to the ATG start codon up to 110 nt downstream from the TAG stop codon (Fig. 4). The ORF representing *ipiO1* is located at a distance of 2224 bp upstream from the ORF of *ipiO2*. The orientations of the ORFs are inverted (Fig. 1d). The nt sequences of *ipiO1* and *ipiO2* show 99% identity from 637 nt upstream from the ATG start codon, throughout the coding sequence, up to at least 152 nt downstream from the TAG stop codon. The restriction endonuclease profile suggests that the similarity extends even further since both genes have a cross-hybridizing 0.6-kb *Xba*I fragment downstream from their coding regions (Fig. 1d). In the 5′ regions, the DNA sequences diverge upstream from positions −637 resulting in a unique 950-nt intergenic DNA region.

(e) Structural features of the *ipiB* and *ipiO* genes

The transcription start point (*tsp*) of the *ipiO* genes was determined by primer extension. A 5′ end labeled oligonucleotide complementary to nt +86 to +102 in *ipiO1* and *ipiO2* was annealed to and extended on poly(A)$^+$ RNA isolated from infected potato leaves, two days post-inoculation. A single primer extension product of 128 nt was found indicating that the *tsp* corresponds to the A at position −26 relative to the ATG start codon (*tsp* is

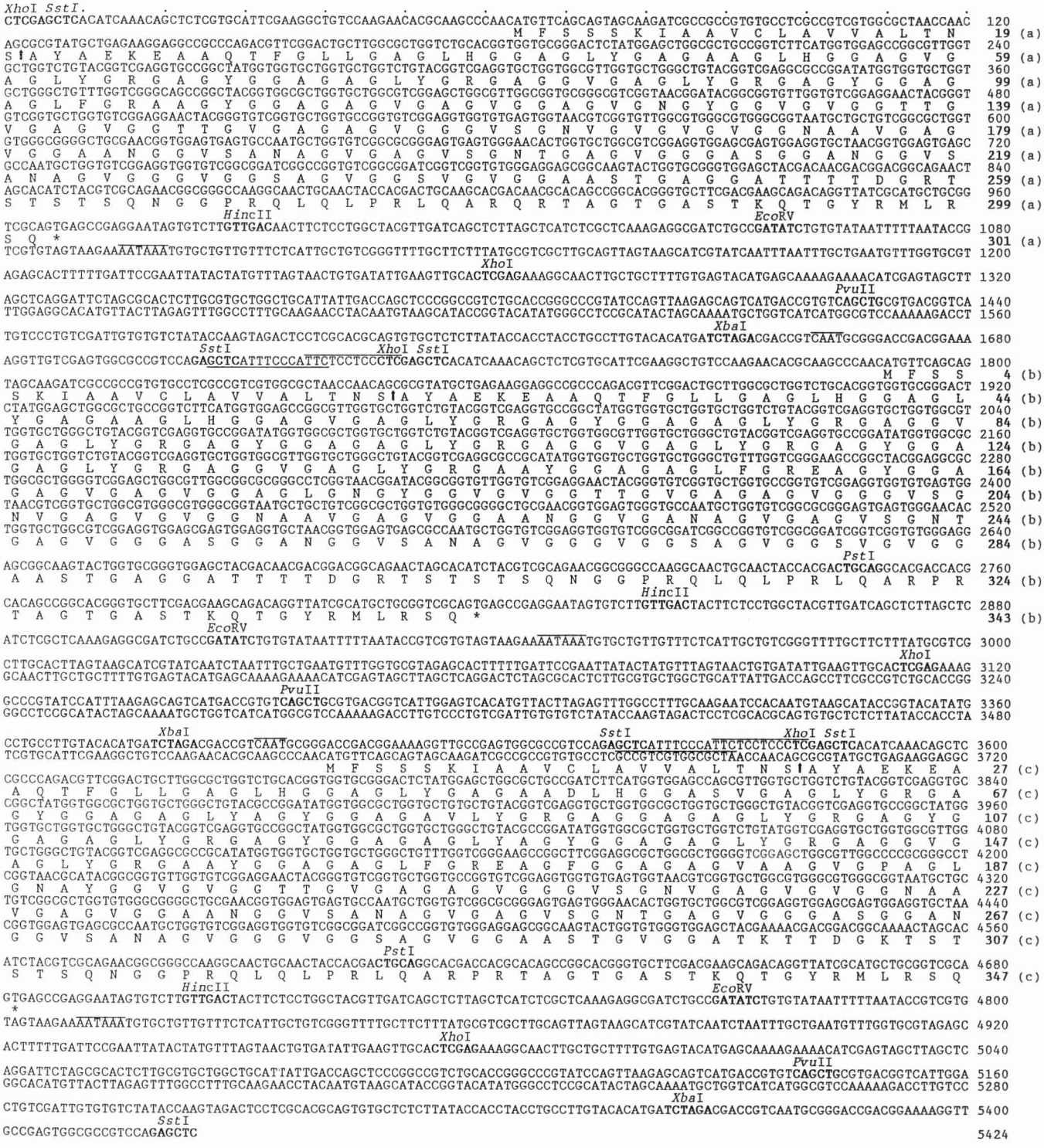

Fig. 3. Nucleotide sequence of the *P. infestans ipiB* gene cluster. The nt sequence starts at the most left *Xho*I site shown in Fig. 1b and extends up to the most right *Sst*I site depicted in the same figure. Deduced aa sequences of IPI-B1 (**a**), IPI-B2 (**b**) and IPI-B3 (**c**) are indicated below the ORFs of *ipiB1*, *ipiB2* and *ipiB3*, respectively. Noted in the figure are: CAAT-motifs (overlined); sequences matching the conserved sequence motif GCTCATTYYNCAWTTT (underlined); CT-rich regions (overlined); putative signal sequence cleavage sites (↑); and potential polyadenylation signals AATAAA (overlined). The *ipiB1*, *ipiB2* and *ipiB3* sequences will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession Nos. L24206, L23937 and L23936, respectively. For sequencing, overlapping subclones were made in pTZ19U using standard procedures (Sambrook et al., 1989). Sequencing was performed on double stranded DNA by the dideoxynucleotide termination method (Sanger et al., 1977) using the Multiwell Microtitre Plate Sequencing System (Amersham) and [α-^{35}S]dATP as a label. Analyses of sequence data and alignment of nt and aa sequences as shown in Figs. 3–7 were performed using the Sequence Analysis Software Package, version 7.1 of the Genetics Computer Group (GCG, Madison, WI, USA; Devereux et al., 1984).

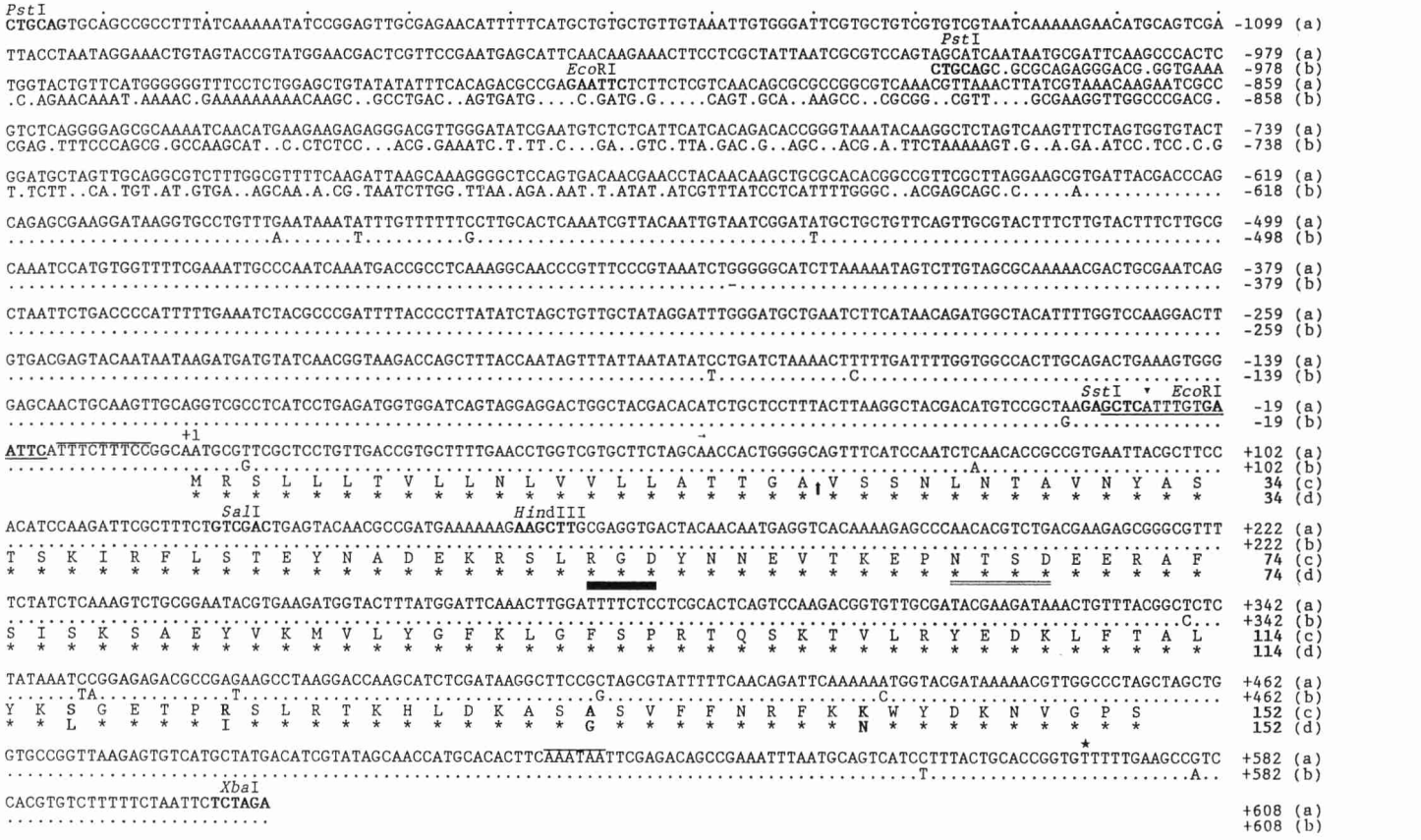

Fig. 4. Nucleotide sequence of the *P. infestans ipiO* gene cluster and deduced aa sequences of the IPI-O1 and IPI-O2 proteins. (**a**) The nt sequence of *ipiO1* starting from the *Pst*I site in the middle of the intergenic region and extending to the *Xba*I site downstream from *ipiO1* (Fig. 1d). (**b**) The nt sequence of *ipiO2* starting from the same *Pst*I site and extending to the *Xba*I site downstream from *ipiO2* (Fig. 1d). Nucleotides identical to those in *ipiO1* are indicated by dots. To allow optimal alignment, one dash (-) is introduced at nt position −427 in the *ipiO2* nt sequence; (**c**) Deduced aa sequence of the IPI-O1 protein; (**d**) Deduced aa sequence of the IPI-O2 protein. Amino acid residues identical to those in the IPI-O1 sequence are indicated by asterisks. Noted in the figure are: sequence matching the conserved sequence motif GCTCATTYYNCAWTTT (single underline); *tsp* (▼); CT-rich region (overlined); putative polyadenylation signal AAATAA (overlined); 5′ end of the partial *ipiO1* cDNA clone (→); start poly(A) tail in partial *ipiO1* cDNA clone (★); putative signal sequence cleavage site (↑); RGD cell adhesion motif (closed bar); putative *N*-glycosylation site (double underline). The nt sequences of *ipiO1* and *ipiO2* will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession Nos. L23939 and L23938. For sequencing, overlapping subclones and deletion clones were made in pTZ19U using standard procedures (Sambrook et al., 1989). Deletion clones were generated by partial *Sau*3AI digestion of master subclones followed by electrophoresis of digestion products on a 0.7% agarose gel along with linearized master subclone DNA as marker. Singly digested, linearized DNA was isolated from the gel and digested to completion with *Bam*HI. DNA fragments were then circularized by ligation and transferred to *Escherichia coli* cells. The interaction cDNA library from which the partial *ipiO1* cDNA clone was isolated was constructed in λZAP (Stratagene, La Jolla, CA, USA) according to the manufacturers instructions. As template, poly(A)$^+$ RNA isolated from *P. infestans* infected potato leaf tissue was used. The infected tissue was obtained from a zone of 1 cm in width at the outer edge of lesions surrounding the infection site on leaves of potato cultivar Ajax, 3 days after spot inoculation with 10 µl of a suspension of sporangia from *P. infestans* strain 88069 (5×10^5 sporangia/ml).

marked by a ▼ in Fig. 4). A direct comparison of the nt sequence surrounding the major *tsp* of eight distinct oomycetous genes in which the *tsp* have been determined, revealed that these oomycetous genes have a sequence preference for transcription initiation. In seven out of eight genes, the major *tsp* is located within the sequence motif 5′-GCTCATTYYNCAWTTT (Table I) which is invariably situated within the first 100 nt upstream from the ATG start codon. The *P. infestans actB* gene encoding actin (Unkles et al., 1991), does contain the conserved motif but the five *tsp* of this gene are located 70 to 114 nt upstream from this sequence. The conserved sequence motif is also present in the 5′ flanking regions of the *ipiB* genes, 69 to 84 nt upstream from the ATG start codon (Table I). However, in this study the *tsp* of the *ipiB* genes were not determined. The *P. infestans* genome may contain more *ipiB*-like genes in addition to the ones characterized here (see section **b**) and so far it is not known whether the *ipiB* genes we have isolated are the ones which are transcribed. The presence of a conserved sequence motif surrounding the major *tsp* of the oomycetous genes suggests that the motif is important for transcription initiation. Since this motif is not conserved in genes of higher fungi, plants or animals, the

TABLE I

Alignment of the DNA regions surrounding the conserved sequence motif GCTCATTYYNCAWTTT which is present within the first 100 nt upstream of the ATG start codon of eight oomycetous genes[a]

Gene	nt sequence surrounding the GCTCATTYYNCAWTTT motif[b]			
P. infestans ipiB[c,d]	−94 GCCGTCCAGA	GCTCATTTCCCATTCT * * * * * * * * * * * * * * * *	CCTCCCT	−62
P. infestans ipiO1[e]	−40 TCCGCTAAGA	GCTCATTTGTGAATTC * * * * * * * * * * * * * * * *	ATTTCTT	−8
P. infestans ubi3R	−66 CGCCTCCTTT	GCTCATTTTCCATTTT * * * * * * * * * * * * * * * *	GAGCGGA	−34
P. infestans calA	−62 TTTTGGATGG	GATCATTGTTGGATTT * * * * * * * * * * * * * * * *	CCCTCGA	−30
P. infestans actA	−84 TCCCTCTTTG	GCTCATTTCCC/TTTT * * * * * * * * * * * * * * * *	CTTCCAG	−53
P. infestans actB[f]	−60 GTGTCAAAGT	TCTCATTCTGCATTTT * * * * * * * * * * * * * * * *	GTCTCGA	−28
P. megasperma actin	−71 GGACCTTGCT	CGTCATTCCGCAATTT * * * * * * * * * * * * * * * *	GCTGCCA	−39
B. lactucae ham34	−85 CGATCGGAAG	GCTCATTCTCC/TTTT * * * * * * * * * * * * * * * *	CACTCTC	−54
B. lactucae hsp70	−78 TCTCAAGTTT	GCTCACTTTGAAATTT * * * * * * * * * * * * * * * *	TCCATCT	−46

[a] *Phytophthora infestans* genes *ipiB2* and *ipiB3* (Fig. 3), *ipiO1* and *ipiO2* (Fig. 4), *ubi3R* (Pieterse et al., 1991), *calA* (Pieterse et al., 1993b), *actA* and *actB* (Unkles et al., 1991), the *P. megasperma* actin gene (Dudler, 1990), and the *Bremia lactucae* genes *ham34* (Judelson and Michelmore, 1990) and *hsp70* (Judelson and Michelmore, 1989).

[b] Numbers refer to the position of the first and last nt of the depicted sequence relative to the ATG start codon of the respective gene. Nucleotides marked with an asterisk match with the consensus sequence GCTCATTYYNCAWTTT. In the *P. infestans actA* sequence and the *B. lactucae ham34* sequence, a one nt gap, indicated by a slash (/), is introduced in the conserved motif to obtain optimal alignment. Underlined nt show(s) the position of the *tsp*.

[c] The nt sequence shown is present in the promoter region of the *P. infestans ipiB2* and *ipiB3* gene. The nt sequence at positions −62 to −94 relative to the ATG start codon of *ipiB1* was not determined.

[d] The position of the *tsp* has not been determined.

[e] The nt sequence at position −8 to −40 relative to the ATG start codon of *ipiO1* and *ipiO2* is nearly identical. Only at position −34, the A in *ipiO1* is a G in *ipiO2*.

[f] Five *tsp* have been found in the *P. infestans actB* promoter region which are all located more upstream of the depicted sequence between positions −120 and −164 relative to the ATG start codon (Unkles et al., 1991).

5′-GCTCATTYYNCAWTTT sequence motif can be considered to be a consensus sequence for transcription initiation in oomycetous genes.

In the majority of genes identified in higher eukaryotes, the consensus 'core promoter' sequences TATAAA and CAAT are found around 30 bp and 70–90 bp upstream from the major transcription initiation site, respectively. However, the significance of these motifs in transcription initiation in filamentous fungi has never been convincingly established (Gurr et al., 1987; Unkles, 1992). In the 5′ flanking regions of *ipiB2* and *ipiB3* there are no typical TATAAA-like motifs. A CAAT motif is present at nt −139 to −136 relative to the ATG start codon (overlined in Fig. 3). In the 5′ flanking regions of the *ipiO* genes no TATAAA or CAAT-like motifs are present near the transcription start. Both *ipiB* and the *ipiO* genes contain a C+T-rich region directly downstream from the transcription initiation consensus sequence (overlined in Fig. 3 and 4). C+T-rich regions are commonly found in the vicinity of transcription initiation sites of filamentous fungal genes and are thought to be important for determining the position of transcription initiation (Unkles, 1992). The nt sequence surrounding the translation start codons of the *ipiB* and *ipiO* genes (CCAAC*ATG*TT and CGGCA*ATG*CG, respectively) follow the Kozak consensus sequence for translation initiation (Kozak, 1984), the most highly conserved nt at position −3 being a purine.

The 3′ terminus of the *ipiO* genes was determined by dideoxy sequencing of a partial *ipiO1* cDNA clone. In the *ipiO1* cDNA sequence, the poly(A) tail starts 110 nt downstream from the TAG stop codon which corresponds to nt +569 of the *ipiO1* genomic sequence (indicated by a ★ in Fig. 4). A putative polyadenylation signal AAATAA (consensus AATAAA) was found 48 nt upstream from the poly(A) tail in the cDNA sequence (overlined in Fig. 4). In the genomic sequence of the *ipiB* genes, potential polyadenylation signals (AATAAA) are present 126 nt downstream from the TGA stop codons (overlined in Fig. 3).

The ORFs of the three *ipiB* genes, as well as the ORFs

of the two *ipiO* genes, are not interrupted by introns. In contrast to most filamentous fungal genes, 68% of which contains introns (Gurr et al., 1987; Unkles, 1992), the oomycetous genes studied to date predominantly lack introns. Among all oomycetous genes reported so far (Dudler, 1990; Judelson and Michelmore, 1989; Judelson and Michelmore, 1990; LéJohn, 1989; Moon et al., 1992; Pieterse et al., 1991; 1993b; Unkles et al., 1991) there is only one gene which contains introns, i.e., the *P. parasitica trp1* gene encoding indole-3-glycerolphosphate synthase-*N*-(5′-phosphoribosyl)anthranilate isomerase (Karlovsky and Prell, 1991). The presence of introns in this gene is surprising, since all homologous *trp* genes from other eukaryotes tend to lack introns. In addition, the nt sequence of the *P. infestans niaA* gene encoding nitrate reductase (C.M.J.P., unpublished) suggests the presence of an intron in this gene as well.

(f) Analysis of the IPI-B aa sequences

The IPI-B1 (301 aa), IPI-B2 (343 aa) and IPI-B3 (347 aa) proteins encoded by the ORFs of *ipiB1*, *ipiB2* and *ipiB3*, have a calculated molecular mass of 25.8, 29.6, and 30.2 kDa and a predicted p*I* of 11.2, 11.1, and 10.9, respectively. When allowing gaps for optimal alignment the proteins are 96–98% identical (Fig. 5). The IPI-B proteins have a high content of Gly residues and their Gly-rich domains show up to 47% identity to the Gly-rich domains of several plant Gly-rich proteins. Most plant Gly-rich proteins are characterized by their repetitive primary structure consisting of up to 70% Gly residues which are arranged in short aa repeats. Furthermore, they usually have an N-terminal signal sequence for transport out of the cytoplasm (Showalter, 1993). Analysis of the predicted aa sequences revealed that the IPI-B proteins share these characteristics with the plant Gly-rich proteins. The IPI-B proteins are composed of four domains. The first 20 aa at the N-terminal end (domain I) comprise a putative signal sequence for secretion, which can be recognized by a hydrophobic region in the hydropathy plot shown in Fig. 6a. The signal sequence cleavage site, predicted according to von Heijne (1986), is located between a serine (aa 20) and an alanine residue (aa 21). The hydropathy plot shows that the putative signal sequence is followed by a short hydrophilic region of 10 aa (domain II). Domain III, comprising 74–80% of the protein, consists of a large Gly-rich region with up to 47% Gly residues (IPI-B1 43%; IPI-B2 47%; IPI-B3 46%). The primary structure of the Gly-rich domain is highly repetitive with two typical repeats (Fig. 7). The first repeat is characterized by the core sequence A/V-G-A-G-L-Y-G-R, the second repeat by G-A-G-Y/V-G-G. The C termini of the predicted IPI-B proteins are composed of a 46 aa hydrophilic region (domain IV).

To our knowledge, this is the first report on genes encoding fungal Gly-rich proteins. Many plant genes encoding Gly-rich proteins have been characterized (Condit and Keller, 1990; Showalter, 1993) and most of them have been implicated to be cell wall proteins (Condit et al., 1990; Keller et al., 1988; 1989a,b). Whether the *P. infestans* IPI-B proteins are structural proteins associated with the fungal cell wall needs to be investigated. Since the *ipiB* mRNA is detectable at high levels in the early stages of infection, it is tempting to speculate that the IPI-B proteins are involved in the development of infection structures. Although the IPI-B proteins share some characteristics with fungal hydrophobins (Stringer and Timberlake, 1993; Wessels, 1992), e.g., putative cell-wall location, hydrophobicity and the presence of a signal peptide, they do not contain the conserved Cys motif and their lengths are not within the known size range for hydrophobins (96 to 157 aa).

(g) Analysis of IPI-O aa sequences

The *ipiO* genes encode a 152-aa 17-kDa protein (Fig. 4) with an p*I* of 10.5. The deduced aa sequences of IPI-O1

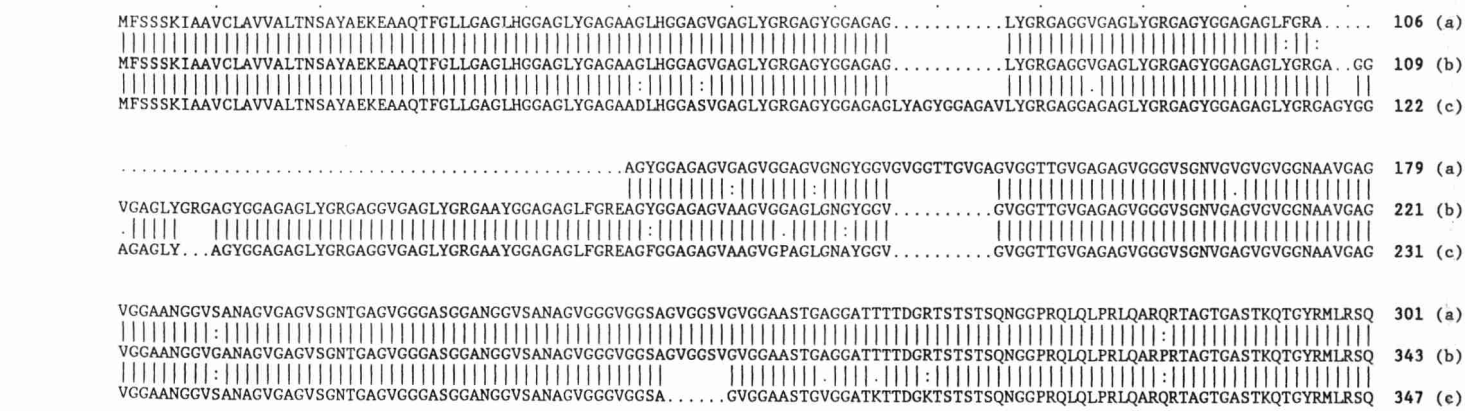

Fig. 5. Comparison of the predicted aa sequences of IPI-B1 (**a**), IPI-B2 (**b**) and IPI-B3 (**c**). Gaps are introduced in the sequence to obtain optimal alignment. Numbers indicate position of last aa.

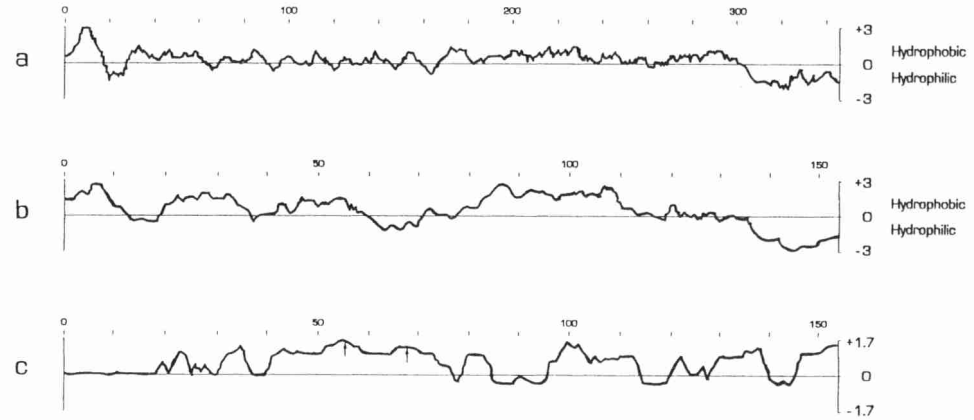

Fig. 6. Hydropathy plot of the deduced aa sequence of IPI-B2 (**a**) and IPI-O1 (**b**), and antigenic index of the predicted aa sequence of IPI-O1 (**c**). Hydrophobicity was determined by the method of Kyte and Doolittle (1982). Along the coordinate, aa positions in the proteins are indicated. In (**a**) and (**b**), regions above the base line are hydrophobic, regions below this line are hydrophilic. The antigenic index was predicted according to the algorithm described by Jameson and Wolf (1988). Positions of the 'RGD' cell attachment sequence and the putative *N*-glycosylation site are indicated by arrows. The hydropathy plots of IPI-B1, IPI-B3 and IPI-O2, as well as the antigenic index of IPI-O2, are not shown, since they are highly similar to those of IPI-B2 and IPI-O1, respectively.

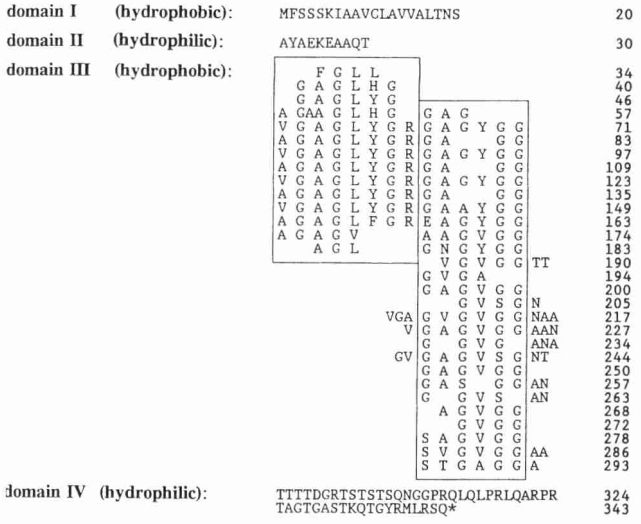

Fig. 7. The four domains in the IPI-B2 protein. In domain III, the Gly-rich repeats with the core sequence A/V-G-A-G-L-Y-G-R and G-A-G-Y/V-G-G are boxed. Numbers indicate position of last aa in each line.

and IPI-O2 are 97% homologous. They differ in only 4 single aa residues at positions 117, 122, 134 and 143. The predicted IPI-O1 and IPI-O2 aa sequences show no significant similarity with any sequence present in data libraries. The proteins contain a putative N-terminal signal sequence which suggests that they are secreted (see Fig. 6b for hydropathy plot). According to the rules of von Heijne (1986), the signal sequence cleavage site was predicted between an alanine (aa 21) and a valine residue (aa 22). A putative cell attachment sequence, recognized by the Arg-Gly-Asp (RGD) tripeptide, is located at positions 54–56. At positions 66–69, a potential *N*-glycosylation site (NTSD) is present. From the surface probability plot (Jameson and Wolf, 1988) it appears that the RGD tripeptide, as well as the NTSD sequence, are located on potentially exposed surface peaks (Fig. 6c), indicating that their function as cell attachment sequence or *N*-glycosylation site, respectively, is not hindered.

The RGD motif was first identified in the animal cell-substratum adhesion proteins fibronectin, fibrinogen, vitronectin and the von Willebrand coagulation factor (Ruoslathi and Pierschbacher, 1986), and was found to be crucial for the interaction with their cell surface receptors, the so-called integrins (D'Souza et al., 1988; Hynes, 1987). These specific ligand-receptor interactions mediate adhesion of cells to the extracellular matrix and are essential for normal development of animal cells. Also in lower eukaryotes and bacteria, cell adhesion molecules with a functional RGD tripeptide have been found, e.g., discoidin I which plays a role in aggregation of the slime mold *Dictyostelium discoideum* (Gabius et al., 1985). Moreover, in pathogenicity factors of some bacterial human pathogens, a functional RGD tripeptide was found to interact with integrins on the surface of host cells, thereby mediating uptake into the host cells (Finlay, 1990; Relman et al., 1990; Russell and Wright, 1988). The RGD tripeptide is also included in the attachment site of the foot-and-mouth disease virus (Fox et al., 1989). In plants, RGD-dependent cell wall-cell membrane adhesions have been demonstrated to occur (Schindler et al., 1989; Zhu et al., 1993) and proteins immunologically related to human vitronectin and its receptor have been detected (Quatrano et al., 1991; Sanders et al., 1991; Schindler et al., 1989). Whether the *P. infestans* IPI-O proteins have cell adhesion properties is currently under investigation.

(h) Conclusions

(*1*) The phage clones DHC-B and DHC-O, which were isolated from a *P. infestans* genomic library in a screen

for *in planta*-induced genes, both contain a small gene cluster. The *ipiB* gene cluster contains three members, whereas on other locations in the genome additional *ipiB* or *ipiB*-like genes are present. The *ipiO* gene cluster consists of only two members. Within these gene clusters, the members are highly homologous throughout the coding sequences and the regulatory 5′ and 3′ flanking regions.

(*2*) In seven out of the eight distinct oomycetous genes in which the *tsp* have been determined, transcription initiation occurs in the conserved sequence motif 5′-GCTCATTYYNCAWTTT.

(*3*) The coding regions of the *ipiB* and the *ipiO* genes are not interrupted by introns, a feature which is observed in most oomycetous genes studied so far.

(*4*) The members of the *ipiB* gene cluster encode three novel, highly homologous Gly-rich proteins. The IPI-B proteins have a putative signal sequence for transport out of the cytoplasm and a highly repetitive Gly-rich domain, both features which are often found in plant cell-wall Gly-rich proteins.

(*5*) The two *ipiO* genes code for two almost identical proteins which have no significant similarity with any sequence in the data libraries. The IPI-O proteins have an N-terminal signal sequence. In addition, they contain a RGD motif which might function as a cell attachment sequence, and a putative *N*-glycosylation site.

ACKNOWLEDGEMENTS

The authors wish to thank Pierre de Wit for continuous support and for critically reading the manuscript and Guido van den Ackerveken for helpful comments on the manuscript.

REFERENCES

Cheung, W.Y.: Calcium and Cell Function, Vol. I. Calmodulin. Academic Press, New York, NY, 1980.

Choi, D., Ward, B.L. and Bostock, R.M.: Differential induction and suppression of potato 3-hydroxy-3-methylglutaryl coenzyme A reductase genes in response to *Phytophthora infestans* and to its elicitor arachidonic acid. Plant Cell 4 (1992) 1333–1344.

Condit, C.M. and Keller, B.: The glycine-rich cell wall proteins of higher plants. In: Adair, W.S. and Mecham, R.P. (Eds.), Organization and Assembly of Plant and Animal Extracellular Matrix. Academic Press, New York, NY, 1990, pp. 119–135.

Condit, C.M., McClean, B.G. and Meagher, R.B.: Characterization of the expression of the petunia glycine-rich protein-1 gene product. Plant Physiol. 93 (1990) 596–602.

Devereux, J., Haeberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12 (1984) 387–395.

D'Souza, S.E., Ginsberg, M.H., Burke, T.A., Lam, S.C.-T. and Plow, E.F.: Localization of an Arg-Gly-Asp recognition site within an integrin adhesion receptor. Science 242 (1988) 91–93.

Dudler, R.: The single-copy actin gene of *Phytophthora megasperma* encodes a protein considerably diverged from any other known actin. Plant Mol. Biol. 14 (1990) 415–422.

Feinberg, A.P. and Vogelstein, G.: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132 (1983) 6–13.

Finlay, B.B.: Cell adhesion and invasion mechanisms in microbial pathogens. Curr. Opin. Cell Biol. 2 (1990) 815–820.

Fox, G., Parry, N.R., Varnett, P.V., McGinn, B., Rowlands, D.J. and Brown, F.: The cell attachment site on foot-and-mouth disease virus includes the amino acid sequence RGD (arginine-glycine-aspartic acid). J. Gen. Virol. 70 (1989) 625–637.

Fritzemeier, K.-H., Cretin, C., Kombrink, E., Rohwer, F., Taylor, J., Scheel, D. and Hahlbrock, K.: Transient induction of phenylalanine ammonia-lyase and 4-coumarate:CoA ligase mRNAs in potato leaves infected with virulent or avirulent races of *Phytophthora infestans*. Plant Physiol. 85 (1987) 34–41.

Gabius, H.-J., Springer, W.R. and Barondes, S.H.: Receptor for the cell binding site of Discoidin I. Cell 42 (1985) 449–456.

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 Eukaryotic Microbes. IRL Press, Oxford, 1987, pp. 93–139.

Hahlbrock, K., Arabatzis, N., Becker-André, M., Joos, H.-J., Kombrink, E., Schröder, M., Strittmatter, G. and Taylor, J.: Local and systemic gene activation in fungus-infected potato leaves. In: Lugtenberg, B.J.J. (Ed.), Signal Molecules in Plants and Plant-Microbe Interactions, NATO ASI Series, Vol. H36, Springer, Berlin, 1989, pp. 241–249.

Hynes, R.O.: Integrins, a family of cell surface receptors. Cell 48 (1987) 549–554.

Jameson, B.A. and Wolf, H.: The antigenic index: a novel algorithm for predicting antigenic determinants. CABIOS 4 (1988) 181–186.

Judelson, H.S. and Michelmore, R.W.: Structure and expression of a gene encoding heat-shock protein *Hsp70* from the oomycete fungus *Bremia lactucae*. Gene 79 (1989) 207–217.

Judelson, H.S. and Michelmore, R.W.: Highly abundant and stage-specific mRNAs in the obligate pathogen *Bremia lactucae*. Mol. Plant-Microb. Interact. 3 (1990) 225–232.

Karlovsky, P. and Prell, H.H.: The *TRP1* gene of *Phytophthora parasitica* encoding indole-3-glycerolphosphate synthase-*N*-(5′-phosphoribosyl)anthranilate isomerase: structure and evolutionary distance from homologous fungal genes. Gene 109 (1991) 161–165.

Keller, B., Sauer, N. and Lamb, C.J.: Glycine-rich cell wall proteins in bean: gene structure and association of the protein with the vascular system. EMBO J. 7 (1988) 3625–3633.

Keller, B., Schmid, J. and Lamb, C.J.: Vascular expression of a bean cell wall glycine rich protein-β-glucuronidase gene fusion in transgenic tobacco. EMBO J. 8 (1989a) 1309–1314.

Keller, B., Templeton, M.D. and Lamb, C.J.: Specific localization of a plant cell wall glycine-rich protein in protoxylem cells of the vascular system. Proc. Natl. Acad. Sci. USA 86 (1989b) 1529–1533.

Kozak, M.: Composition and analysis of sequences upstream from the translational start site in eukaryotic mRNA. Nucleic Acids Res. 12 (1984) 857–872.

Kyte, J. and Doolittle, R.F.: A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157 (1982) 105–132.

LéJohn, H.B.: Structure and expression of fungal calmodulin gene. J. Biol. Chem. 264 (1989) 19366–19372.

Martini, N., Egen, M., Rüntz, I. and Strittmatter, G.: Promoter sequences of a potato pathogenesis-related gene mediate transcriptional activity selectively upon fungal infection. Mol. Gen. Genet. 236 (1993) 179–186.

Monia, B.P., Ecker, D.J. and Crooke, S.T.: New perspectives on the structure and function of ubiquitin. Bio/Technology 8 (1990) 209–215.

Moon, R.P., Unkles, S.E., Duncan, J.M., Hawkins, A.R. and Kinghorn, J.R.: Sequence of the *Phytophthora infestans* glyceraldehyde-3-phosphate dehydrogenase-encoding gene (*gpdA*). Plant Mol. Biol. 18 (1992) 1209–1211.

Pieterse, C.M.J., Risseeuw, E.P., and Davidse, L.C.: An *in planta* induced gene of *Phytophthora infestans* codes for ubiquitin. Plant Mol. Biol. 17 (1991) 799–811.

Pieterse, C.M.J., de Wit, P.J.G.M. and Govers, F.: Molecular aspects of the potato-*Phytophthora infestans* interaction. Neth. J. Plant Pathol. 98 (Suppl. 2) (1992) 85–92.

Pieterse, C.M.J., Riach, M.B.R., Bleker, T., van den Berg-Velthuis, G.C.M. and Govers, F.: Isolation of putative pathogenicity genes of the potato late blight fungus *Phytophthora infestans* by differential hybridization of a genomic library. Physiol. Mol. Plant Pathol. (1993a), in press.

Pieterse, C.M.J., Verbakel, H.M., Hoek Spaans, J., Davidse, L.C. and Govers, F.: Increased expression of the calmodulin gene of the late blight fungus *Phytophthora infestans* during pathogenesis on potato. Mol. Plant-Microb. Interact. 6 (1993b) 164–172.

Quatrano, R.S., Brian, L., Aldridge, J. and Schultz, T.: Polar axis fixation in *Fucus* zygotes: components of the cytoskeleton and extracellular matrix. Development 1 (Suppl.) (1991) 11–16.

Relman, D., Tuomanen, E., Falkow, S., Golenbock, D.T., Saukkonen, K. and Wright, S.D.: Recognition of a bacterial adhesin by an integrin: macrophage CR3 ($\alpha_M\beta_2$, CD11b/CD18) binds filamentous hemagglutinin of *Bordetella pertussis*. Cell 61 (1990) 1375–1382.

Ruoslathi, E. and Pierschbacher, M.D.: Arg-Gly-Asp: a versatile cell recognition signal. Cell 44 (1986) 517–518.

Russell, D.G. and Wright, S.D.: Complement receptor type 3 (CR3) binds to an Arg-Gly-Asp-containing region of the major surface glycoprotein, gp63, of *Leishmania* promastigotes. J. Exp. Med. 168 (1988) 279–292.

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

Sanders, L.C., Wang, C.-S., Walling, L.L. and Lord, E.M.: A homolog of the substrate adhesion molecule vitronectin occurs in four species of flowering plants. Plant Cell 3 (1991) 629–635.

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

Schindler, M., Meiners, S. and Cheresh, D.A.: RGD-dependent linkage between plant cell wall and plasma membrane: consequences for growth. J. Cell Biol. 108 (1989) 1955–1965.

Schröder, M., Hahlbrock, K. and Kombrink, E.: Temporal and spatial patterns of 1,3-β-glucanase and chitinase induction in potato leaves infected by *Phytophthora infestans*. Plant J. 2 (1992) 161–172.

Showalter, A.W.: Structure and function of plant cell wall proteins. Plant Cell 5 (1993) 9–23.

Stringer, M.A. and Timberlake, W.E.: Cerato-ulmin, a toxin involved in Dutch elm disease, is a fungal hydrophobin. Plant Cell 5 (1993) 145–146.

Taylor, J.L., Fritzemeier, K.-H., Häuser, I., Kombrink, E., Rohwer, F., Schröder, M., Strittmatter, G. and Hahlbrock, K.: Structural analysis and activation by fungal infection of a gene encoding a pathogenesis-related protein in potato. Mol. Plant-Microb. Interact. 3 (1990) 72–77.

Unkles, S.E.: Gene organization in industrial filamentous fungi. In: Kinghorn, J.R. and Turner, G. (Eds.), Applied Molecular Genetics of Filamentous Fungi. Blackie Academic & Professional, Glasgow, 1992, pp. 28–53.

Unkles, S.E., Moon, R.P., Hawkins, A.R., Duncan, J.M. and Kinghorn, J.R.: Actin in the oomycetous fungus *Phytophthora infestans* is the product of several genes. Gene 100 (1991) 105–112.

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

Wessels, J.G.H.: Gene expression during fruiting in *Schizophyllum commune*. Mycol. Res. 96 (1992) 609–620.

Zhu, J.-K., Shi, J., Singh, U., Wyatt, S.E., Bressan, R.A., Hasegawa, P.M. and Carpita, N.C.: Enrichment of vitronectin- and fibronectin-like proteins in NaCl-adapted plant cells and evidence for their involvement in plasma membrane-cell wall adhesion. Plant J. 3 (1993) 637–646.