Insect Molecular Biology (2009) **18**(1), 77–85 doi: 10.1111/j.1365-2583.2008.00850.x

Genomic structure and promoter analysis of pathogen-induced *repat* genes from *Spodoptera exigua*

C. S. Hernández-Rodríguez, J. Ferré and S. Herrero

Department of Genetics, Universitat de Valencia, Burjassot, Spain

Abstract

The *repat* gene family encodes midgut proteins over-expressed in response to pathogen infection in the lepidopteran *Spodoptera exigua.* Up-regulation of *repat* genes has been observed after challenging the larvae with both *Bacillus thuringiensis* toxins and after infection with the baculovirus *Autographa californica multiple nucleopolyhedrovirus*. In our study, PCR amplification of the genomic region and genome walking were used to obtain the genomic structure and the sequence of the 5′-upstream region of *repat1* and *repat2*, two of the most phylogenetically distant members of the *repat* family. A similar gene structure between *repat1* and *repat2* has been found, with conserved exon-intron positions and junction sequences, suggesting a common origin for these genes. Recombinant baculoviruses carrying the firefly *luciferase* gene under the control of different 5′-upstream regions of the *repat1* gene were constructed to elucidate the influence of these regions in gene expression. Infection of *Helicoverpa zea* gut-derived cells with the recombinant baculoviruses revealed the upstream regions of the *repat1* gene which are involved in gene transcription and demonstrated the role of an intron located in the 5′-untranslated region in the enhancement of gene expression.

Keywords: *Bacillus thuringiensis*; nucleopolyhedrovirus; immune response; insect pathogens; intron-mediated enhancement.

First published online 8 December 2008.

Correspondence: Salvador Herrero. Department of Genetics, Universitat de Valencia, Dr Moliner 50, 46100 Burjassot, Spain. Tel.: + 34 963543006; fax: + 34 963543029; e-mail: sherrero@uv.es

Introduction

During the larval stage, holometabolous insects spend most of their time eating, and the digestive tract becomes the primary entrance point for a large variety of pathogens and their toxic compounds. In order to overcome or, at least, minimize the pathological effect, insects possess an innate immune system involving different mechanisms acting simultaneously. These mechanisms include the activation of detoxification and damage repair systems (Loeb *et al*., 2001; Li *et al*., 2006), the synthesis and secretion of anti-microbial peptides and degradation enzymes (Cheng *et al*., 2006), phagocytosis, melanization (Ma *et al*., 2005; Rivkin *et al*., 2006), cell apoptosis (Blissard, 1996) and cell sloughing (Washburn *et al*., 2003), among others. All of these defence mechanisms are rapid, lasting up to a few days, and provide particularly powerful resistance to infections (Vilmos & Kurucz, 1998). As a result of the high energetic cost that would imply a constitutive expression of the genes associated with these defence mechanisms, their response is usually inducible, and it is mediated through the activation of several signalling pathways. So far, three of these pathways are known in the immune response: the Toll, the immune deficiency (IMD), and the Janus Kinases (JAK)/Signal Tranducers and Activators of Transcription (STAT) pathways (Boutros *et al*., 2002; Cherry & Silverman, 2006; Wang & Ligoxygakis, 2006). These pathways have been mainly studied in *Drosophila melanogaster,* although several studies have suggested that similar systems also work in other insect orders, including Lepidoptera (Cheng *et al*., 2008).

In a previous study, we identified a novel family of midgut proteins expressed in response to pathogens (REPAT proteins) in larvae of the lepidopteran *Spodoptera exigua* (Herrero *et al*., 2007). Members of this gene family were up-regulated after feeding *S. exigua* larvae with different toxins from the entomopathogenic bacteria *Bacillus thuringiensis*. The same study revealed that this gene family was also up-regulated in the midgut of the larvae during infection with *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV). Furthermore, recombinant baculoviruses expressing one of these proteins were constructed and used to infect *S. exigua* larvae; the results

Table 1. Nucleotide sequences and localization of primers in *repat1* and *repat2*

Gene	Name	Position from translation start site	Orientation	Sequence (5′–3′)
repat1	1-ATG	+1/+23	+	ATGAGGAGTTTAATTATCATCGC
	1-STOP	+733/+767	–	GGAAGCTTACTGCCGTCTGAATAATTCAATAGTGG
	1-GSP1	+163/+189	–	CTCAGAAAGCAGCACCCAACCCTCGTA
	1-GSP2	+18/+45	–	TGCTGCGAGTGCTGACAACACTGCGATG
	1-GSP3	–572/–546	–	CTGTGGTACGAACGTGACTAAATGTCT
	1-GSP4	–642/–616	–	TTCTTTAAGGTAGAAAACAGTTCGAGA
repat2	2-ATG	+1/+28	+	ATGAGGAGCTTCATCATCATCACCGTGC
	2-STOP	+828/+854	–	CTCTCTTCCATATGTGGAAATAGTTGC
	2-GSP1	+406/+434	–	AAGGAAGGGGTTGGATTTGAGTTATTAAG
	2-GSP2	+228/+255	–	CTAGGAATTACGGAGAATTTTACAGAAC

showed that recombinant baculoviruses expressing REPAT1 had reduced virulence, which suggested a role of this protein in the insect's defence against pathogens.

Knowledge of the genomic structure and the *cis*-regulatory region of *repat* genes may offer some clues about the role of *repat* proteins in the response to pathogen infection. In this work, PCR amplification of the genomic region and genome walking were used to obtain the genomic structure and the sequence of the 5′-upstream region of two phylogenetically distant *repat* genes (*repat1* and *repat2)*. Additionally, several recombinant baculoviruses carrying the firefly *luciferase* gene under the control of different 5′-regions of the *repat1* gene were used to elucidate the influence of these regions on the expression level of this gene.

Results

Cloning of repat1 *and* repat2 *open reading frames*

Four clones for each of the coding regions of *repat1* and *repat2* were obtained using the primers described in Table 1 and genomic DNA obtained from two different populations of *S. exigua* (Oxford and Almería populations). Multiple alignment of the sequences of the amplified genes, from ATG to the stop codon, including intronic sequences, revealed both intra- and interpopulation sequence heterogeneity. A nucleotide identity from 88.0 to 96.3% was observed among *repat1* clones, and from 86.3 to 98.4% among *repat2* clones. Amplicons of these clones also showed length variability, ranging from 727 to 767 base pairs (bp) for *repat1* and from 857 to 944 bp for *repat2*. As the nucleotide differences were mainly found in noncoding regions, there was less heterogeneity among larvae in the protein sequences. Consequently, translated *repat* genes showed a degree of amino acid identity from 93.4 to 99.2% for *repat1* and from 95.5 to 98.5% for *repat2*.

Structure of repat1 *and* repat2 *genes*

Genome walking was used to obtain the 5′-upstream region of the *repat* genes. Comparison of the genomic sequence with cDNA allowed the identification of three exons and two introns in both genes (Fig. 1A). All exon-intron junctions retained the conserved sequence GT/AG (Table 2). The analysis of several cDNA clones, obtained in a previous

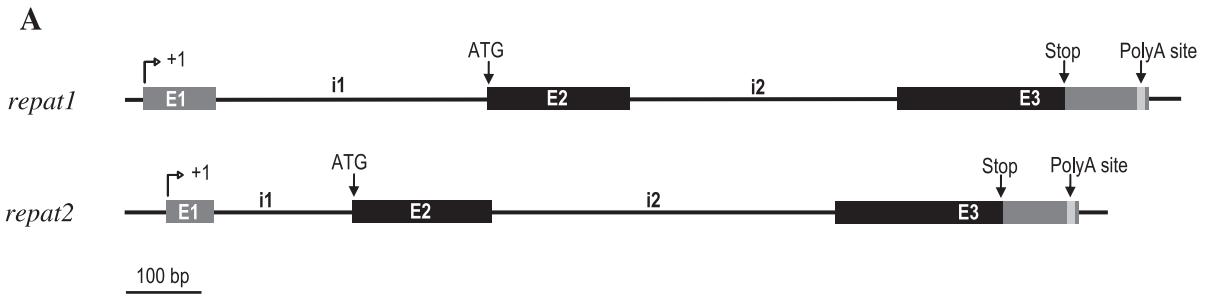

Figure 1. Genomic structure and nucleotide sequence of *repat* genes. (A) Schematic representation of the gene structure for *repat1* and *repat2*. Exons (E1–E3) are represented by rectangles and introns (i1, i2) by lines. cDNA 5′-UTR and 3′-UTR are coloured in grey. A bent arrow indicates the putative transcription start sites. (B) Gene sequences and 5′-upstream regions of *repat1* (EU797187) and *repat2* (EU797188). The transcription start sites, based on the EF153740 and EF153741 cDNAs, are marked with a bent arrow and bold letter. GAS elements in *repat1* and GATA elements in *repat2* are typed in bold and italic letters. The consensus sequences of the branchpoint site in intron 2 are underlined. Within the transcribed region, exon and intron sequences are shown in upper case and lower case, respectively. Asterisks indicate nucleotide identities.

B

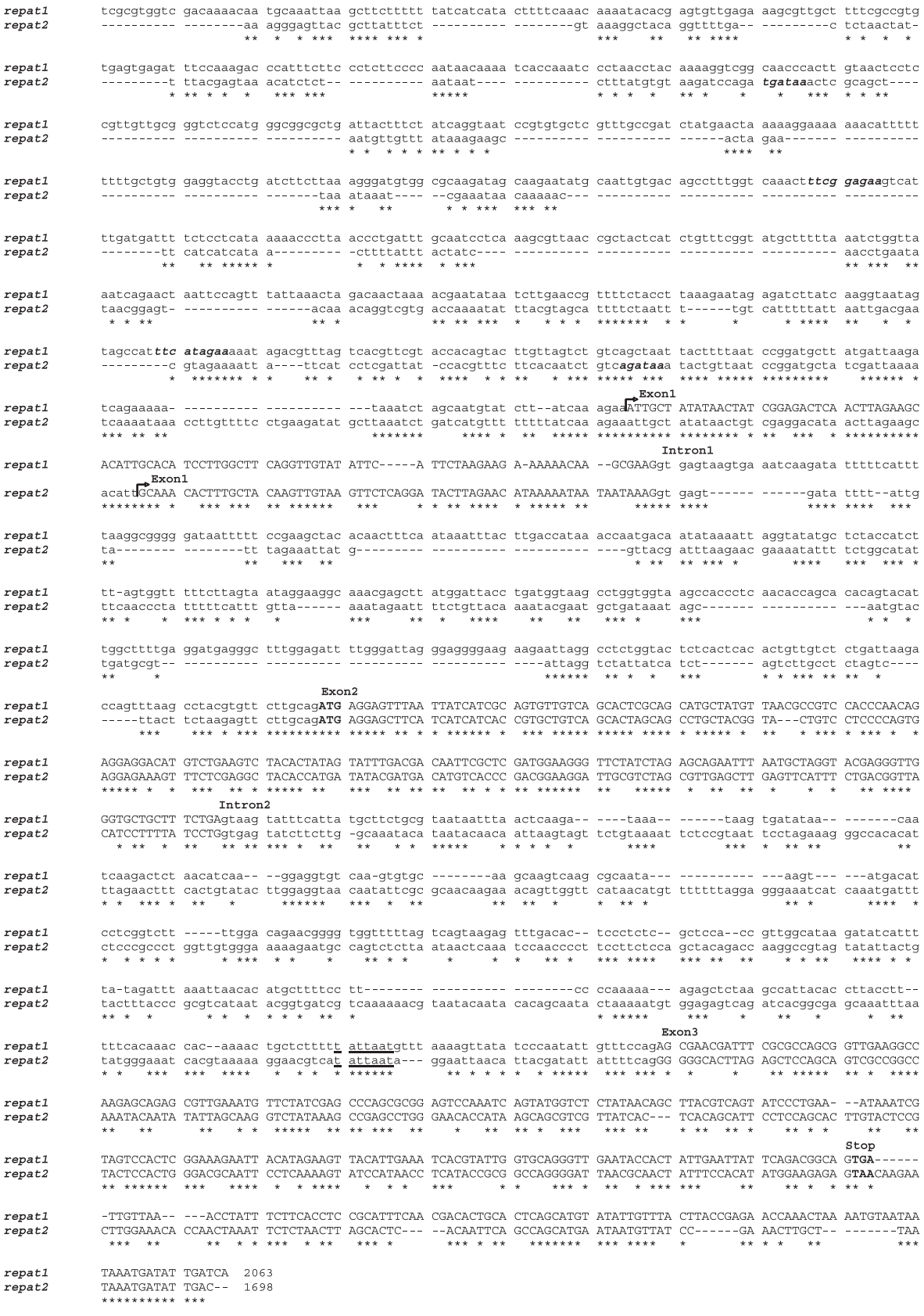

Figure 1. *Continued.*

Table 2. Boundaries, sizes and sequence identity of the *repat* genes

Exon	Gene	Intron/exon	Exon size	Identity (%)	Exon/intron	Intron size	Identity (%)	Residues§
1	*repat 1*	aaagaa*/ATTGCT	96	39	GCGAAG/gtgagt	358	28	–
	repat 2	aaagaa*/ATTGCT	63		ATAAAG/gtgagt	208		–
2	*repat1*	ttgcag/ATGAGG	188	65	TTCTGA/gtaagt	350	35	1–63
	repat2	ttgcag/ATGAGG	185		ATCCTG/gtgagt	449		1–62
3†	*repat1*	ttccag/AGCGAA	223	53				63–136
	repat2	tttcag/GGGGGC	223					62–135
3‡	*repat1*		110	47				
	repat2		100					

*Upstream region from the transcription start site.
†Refers only to the coding region (up to the stop codon) from exon 3.
‡Refers only to 3′-Untranslated Region (UTR) region from exon 3.
§Refers to the protein sequences: ABO64231 (*repat1*) and ABO64232 (*repat2*).

study as products of 5′-Rapid Amplification of CDNA ends (RACE) (Herrero *et al.*, 2007), revealed one major transcription start site for both genes (Fig. 1B). Gene structure is conserved between *repat1* and *repat2* (Fig. 1A, Table 2). Both have exon 1 (96 bp in *repat1* and 63 bp in *repat2*) encoding the 5′-UTR and exon 2 (188 bp in *repat1* and 185 bp in *repat2*) starting at the ATG translation initial codon. For both genes, exon 3 contained 223 bp up to the stop codon and also encoded the 3′-UTR (110 bp in *repat1* and 100 bp in *repat2*) (Table 2). A putative polyadenylation signal was found in position +1308 in *repat1* and in position +1214 in *repat2*. A consensus sequence of a branchpoint site (TATTAAT) was identified in both genes in intron 2 (Fig. 1B). No other homologies between the *repat* genes were found in intronic sequences, in contrast to the high degree of sequence identity observed in the exonic regions (Table 2). Analysis of the DNA sequence of both genes revealed the absence of tandem repeats. In *repat1*, a GC-rich region (CpG island) with fewer than 100 bp and an Observed:Expected (Obs:Exp) ratio of 0.6 was found in exon 2 and exon 3. The search for retroelements or transposon-like sequences in these genes showed a fragment of a *mariner*-like transposable element (MLE) in intron 1 of *repat1*. No other insertions, such as Short Interspersed Nucleotide Element (SINE) or Long Interspersed Nucleotide Element (LINE) elements, were detected in these genes.

Analysis of the 5′-upstream region

Sequencing of the upstream region of *repat1* and *repat2* yielded 1192 and 741 bp, respectively (counting from the ATG start codon). An intron of 358 bp in *repat1* and of 208 bp in *repat2* was identified in the 5′-UTR, with the donor consensus sequence immediately preceding the ATG codon (Fig. 1B). No promoter region containing all the typical RNA-polymerase II promoter elements (GC box, CCAAT box, and TATA box) was found in these upstream sequences. A search for transcription factor recognition elements upstream of the *repat* genes did not show any common putative binding site for *repat1* and *repat2*. Searching for binding elements involved in the immune response, we found in *repat1* two sequences of the consensus STAT binding sites TTCN$^{2-4}$GAA (GAS element) (Decker *et al.*, 1997) and in *repat2* two GATA consensus sequences WGATAA (Yamakawa & Tanaka, 1999) at positions –373 and –142 (Fig. 1B).

Analysis of promoter activity of repat1 *upstream sequence in insect cells*

In order to determine the promoter activity of the *repat1* upstream region, we constructed recombinant baculoviruses carrying the firefly *luciferase* gene under the control of this region [baculovirus A (–738/+464)]. As a negative control, a recombinant virus was constructed lacking any promoter region [baculovirus F (ΔP)] (Fig. 2A). HzGUT (gut-derived) and Sf21 (ovary-derived) cells were used to test the promoter activity of the *repat1* upstream region in the recombinant baculoviruses. After subtraction of the basal transcriptional activity of the *luciferase* gene obtained by the construct F (ΔP), the level of luciferase activity of the full-length construct A (–738/+464) was calculated at different time points after infection. An increase of luciferase activity was observed in HzGUT cells starting at 22 hours post infection (h.p.i.) and up to 34 h.p.i, the time at which it reached a plateau (Fig. 2B). However, when the same virus was used to infect Sf21 cells, no luciferase activity was found up to 46 h.p.i.

Determination of the promoter activity of reporter gene constructs

To characterize the transcriptional activity of the *repat1* upstream region, four additional recombinant baculoviruses were generated with different fragments of the *repat1* upstream region (Fig. 2A). HzGUT cells were infected with recombinant baculoviruses and the transcriptional activity of each construct was determined from the level of luciferase activity (Fig. 2C). Construct F (ΔP) was used to determine the basal transcription of the *luciferase* gene. Luciferase activity of the full-length construct A (–738/+464) was set to 100%. The construct B (+47/+464), which lacked the upstream region of 5′-UTR and the first 47

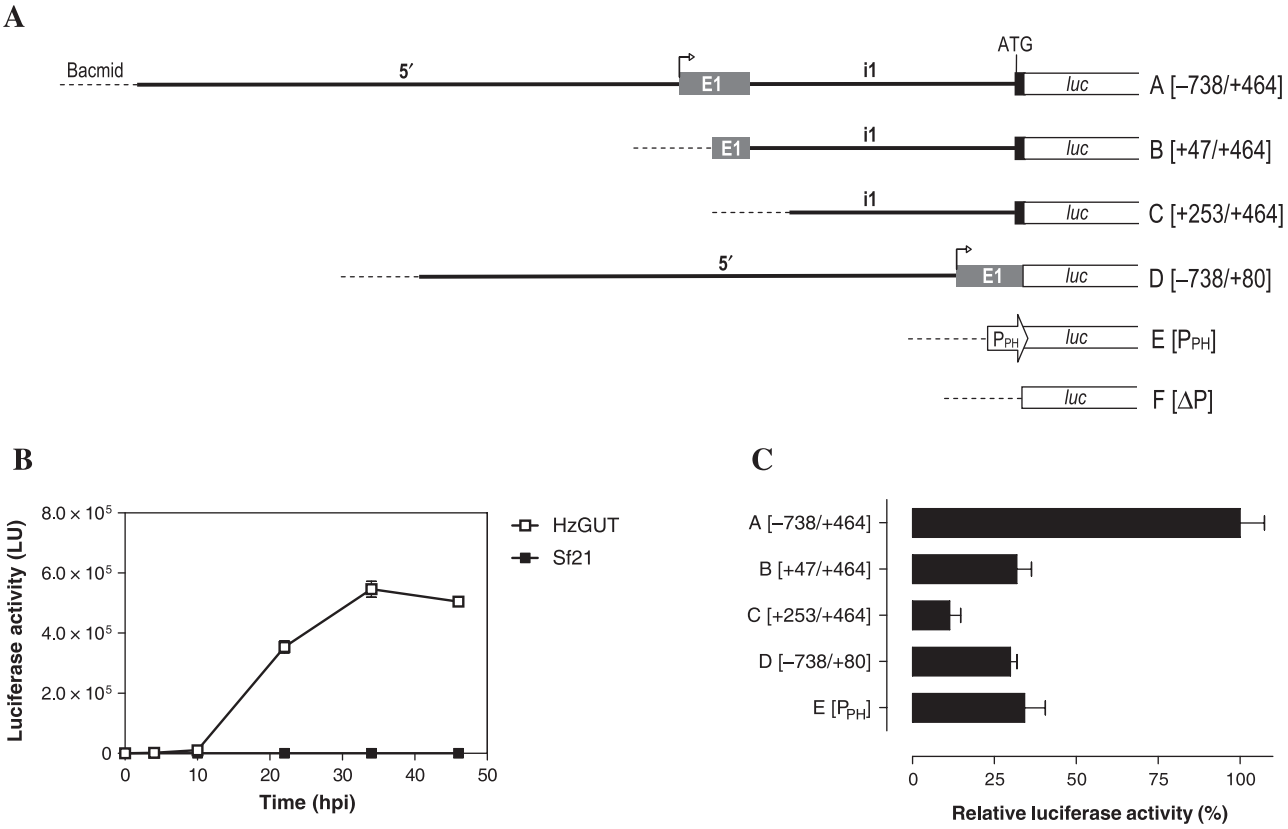

Figure 2. Luciferase activity in cell cultures infected with recombinant baculoviruses carrying upstream regions of *repat1*. (A) Structure of recombinant baculoviruses carrying different upstream regions fused with the *luciferase* reporter gene. Grey boxes represent exon 1; i1 represents intron 1; 5′ refers to the upstream region of the transcriptional initiation. The white arrow in E (P_{PH}) is the polyhedrin promoter. Bent arrows indicate the transcriptional start site and ATG refers to the translational start site. The white boxes represent the *luciferase* gene. Dotted lines correspond to the recombinant baculovirus (bacmid) sequences. (B) Time course of luciferase expression in HzGUT and Sf21 cells measured from 4 to 46 hours post infection (h.p.i.). Results are expressed in luciferase units (LU) of the full-length construct [A (–738/+464)] after subtraction of the basal transcriptional activity obtained by the promoter-less construct [F (ΔP)]. Each data point is the mean of two independent experiments. (C) Luciferase activity of HzGUT cells infected with different recombinant baculoviruses. Luciferase activities were measured at 30 h.p.i. and normalized to Green Fluonescent Protein (GFP) values. Results were expressed as relative luciferase activity of the full-length construct A (–738/+464) after subtraction of the basal activity obtained with the baculovirus control F (ΔP). Values are the means of two duplicate independent experiments. Error bars represent the standard error of the mean.

nucleotides of exon 1, showed approximately three times less luciferase activity than the full-length construct A (–738/+464), suggesting that one or various elements important for promoter activity are located in this region. Activity of construct C (+253/+464), that only contained the major part of the intronic region of the 5′-UTR sequence but missing the 5′-splicing junction and the adjacent area, was also measured. The great reduction in luciferase activity for this construct (up to ~11% of maximum activity) indicated that either the 3′ region in exon 1 (the fragment remaining in the B construct) also exercises some regulatory function or that splicing of intron 1 may have a role in gene expression. Construct D (–738/+80) was generated by deletion of the 5′-UTR intron whilst retaining exon 1 (Fig. 2A). Deletion of this region reduced the level of luciferase activity to ~30% of the maximum activity, suggesting that the 5′-UTR intron enhances the expression of the reporter gene. As a positive control, a recombinant baculovirus carrying the luciferase gene under the control of the late polyhedrin promoter from the baculovirus was generated [construct E (P_{PH})]. Luciferase activity of this construct was around three times lower than that of the *repat1* construct A (–738/+464) at 30 h.p.i.

Discussion

In this study, the genomic structure of the *repat1* and *repat2* genes was obtained by means of amplification of their open reading frame (ORF) sequences and genome walking to the 5′-upstream regions. Nucleotide and amino acid analysis of cloned sequences showed allelic heterogeneity both among larvae from the same population and among larvae from the three populations used in this study. The genomic sequences of *repat1* and *repat2*, two of the most

phylogenetically distant members of the *repat* gene family (Herrero *et al.*, 2007), showed an overall identity of around 45%. Despite this moderate homology, the exon-intron positions and junction-flanking sequences for both genes are highly conserved, including the location of an intron in the 5′-untranslated region with the donor sequence immediately before the start ATG codon (Fig. 1B, Table 2). The similarity in the genomic structure of *repat1* and *repat2* may indicate a common origin for both genes. As *repat1* is phylogenetically more related to *repat3* and *repat4* than to *repat2* (Herrero *et al.*, 2007), all genes belonging to the *repat* gene family are likely to have a common origin and the *repat* gene diversification could be the result of several gene duplication events.

Computational analysis of the upstream regions from the start transcription site did not show a common core promoter containing typical consensus elements such as a TATA box, CAAT box, GC box and transcriptional initiation elements. The upstream regions of the *repat* genes were also screened for the presence of transcription factor binding sites. A refined search was carried out for binding sites involved in the immune-related pathways. In *repat1*, two hypothetical GAS sites, which are responsible for the transcription mediated by the JAK/STAT pathway, were found. The JAK/STAT pathway has been associated with insect response to septic injury, cellular stress and viral infection (Dostert *et al.*, 2005; Agaisse, 2007) fitting very well with the *repat* gene expression profile (up-regulated after the ingestion of a toxin from *B. thuringiensis* that leads to pore formation in the midgut cells, and also after infection with baculovirus; Herrero *et al.*, 2007).

The capacity of the upstream sequence of *repat1* to drive gene transcription in midgut-derived cells (HzGUT), but not in ovary-derived cells (Sf21), is in agreement with a previous observation that the expression of *repat* genes was restricted to the gut tissue (Herrero *et al.*, 2007). However, the activity of the reporter gene in HzGUT cells (derived from *Helicoverpa zea*) suggests the presence of proteins with regulation similar to that of REPAT in this insect species. So far, no REPAT homologues have been found in *Helicoverpa* spp. Future sequence information from the *Helicoverpa armigera* genome sequencing project (K. Gordon, pers. comm.) could reveal new REPAT homologues, as well as provide information about their regulatory regions. Interestingly, the transfection of HzGUT cells with the reporter *luciferase* gene under the control of the full upstream sequence of *repat1* in the pFBD-GFP-derived vector did not show any promoter activity (data not shown). This observation reveals that the promoter activity obtained in our assays is not just a result of the cell type but also of the activation of the *repat1* promoter after infection with baculovirus.

Several recombinant baculoviruses carrying different fragments of the upstream sequence of *repat1* were used to identify the main regions involved in promoter activity. The reduction of reporter activity in the recombinant baculovirus B (+47/+464) suggests that one or several regulatory elements are located in the upstream region of *repat1* or in the initial sequence downstream of the transcription start site. The high level of homology in exon 1 and the adjacent 5′-upstream sequences also supports the presence of regulatory sequences in these regions. Analysis of these regions in other members of the *repat* family, together with mutational analysis of the conserved sequences, could reveal the presence of novel regulatory sequences in these genes. The 30% of activity found in baculovirus B suggests that the region contained in this construct can still initiate transcription. Baculovirus D (–738/+80), which contains the full sequence up to the transcription start site, but lacks intron 1 (located in the 5′-UTR), showed three times less luciferase activity than the full-length construct, revealing an important role of this intron in the regulation of expression. The effect of the introns in the enhancement of gene expression has been described for several decades (Le Hir *et al.*, 2003). Although specific regulatory elements are found within some introns, more general intron-mediated enhancement is thought to result from synergistic interaction among the factors involved in the various steps of gene expression from transcription to translation (Le Hir *et al.*, 2003; Rose, 2004). It has been suggested that splicing of 5′-UTR introns would lead to the exon junction complex (EJC) assembly, increasing the translational efficiency (Le Hir *et al.*, 2000; Nott *et al.*, 2004). Chung *et al.* (2006) proposed that closer introns would recruit the EJC, facilitating the interaction among the RNA, *trans*-factors and the ribosome. As intron 1 is next to the ATG codon, enhancement carried out by this intron could be likely to be related to the splicing complex interactions. A slight ability of this intronic sequence to direct transcription by itself is observed in the activity reported by construct C (+253/+464), which lacks the first exon-intron junction required for splicing and, therefore, the intronic sequence is not removed in the transcript. Although the differences in the levels of expression between the B and C constructs can be attributed to differences in gene transcription, it is likely that these post-transcriptional mechanisms could be responsible for these differences in activity. Description of introns in the UTR sequences in Lepidoptera species is restricted to very few genes (Bel & Escriche, 2006; Sonoda *et al.*, 2006a,b). No evidence of their role in the enhancement of gene expression has been shown. In other insect orders, as far as we know, transcription enhancement by an intron in a 5′-UTR has only been reported in tropomyosin II from *Drosophila melanogaster* (Meredith & Storti, 1993).

In this paper, the genomic structure of two *repat* genes has been reported by the first time. Further experiments may prove that the similar gene organization found in *repat1*

and *repat2* may also be present in the rest of the genes belonging to the *repat* family and, therefore, corroborate a common origin. The involvement of an intron in the 5′-UTR region in *repat1* gene expression suggests a similar regulation for *repat2*, which also contains an intron in the same region, and for those *repat* genes that conserve the described genomic structure. The identification of *cis*- and *trans*-elements and their role in the up-regulation of *repat* genes may be useful to elucidate the contribution of *repat* proteins to the response mechanisms after cell damage or infection.

Experimental procedures

Insects and cells

Three colonies of *S. exigua* from different origins (Oxford, UK; Almería, Spain; Adra, Spain) were continuously reared on an artificial diet at 25 °C under a 16 h light : 8 h dark photoperiod. A *Spodoptera frugiperda* ovary-derived cell line (Sf21) (Vaughn *et al.*, 1977) was maintained at 27 °C in Grace's Medium (GIBCO, Invitrogen Corp., Grand Island, NY, USA) supplemented with 10% foetal bovine serum, 100 U penicillin, and 0.1 mg/ml streptomycin. A *H. zea* gut-derived cell line, RP-HzGUT-AW1 (HzGUT) (Goodman *et al.*, 2004), was maintained at 27 °C in Ex-Cell 420 Medium (SAFC Biosciences, Kansas City, MO, USA) supplemented with 10% foetal bovine serum, 100 U penicillin, and 0.1 mg/ml streptomycin.

Cloning of the repat1 *and* repat2 *open reading frames*

Genomic DNA was obtained from whole last instar larvae of *S. exigua* using the DNeasy Tissue Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. The coding region of the *repat1* and *repat2* genes was amplified by PCR using two primer sets designed on the 5′- and 3′-ends of the ORF (Table 1) based on the published sequences (Herrero *et al.*, 2007). PCR conditions were 5 min at 94 °C, 30 cycles of 1 min at 94 °C, 45 s at 50 °C and 2 min at 72 °C, followed by a final extension of 72 °C for 7 min. The reaction products were purified with a High Pure PCR Product Purification Kit (Roche, Indianapolis, IN, USA) and cloned into the pGEM-T Easy cloning system (Promega, Madison, WI, USA). Plasmid DNA from the clones was purified using the High Pure Plasmid Isolation Kit (Roche) and sequenced using an ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit v.1.1 (Applied Biosystems, Foster City, CA, USA) and T7 and SP6 primers.

Cloning of the repat *5′-upstream regions*

A Universal Genome Walker Kit (Clontech, Mountain View, CA) was used for the sequencing of the promoter region of the *repat* genes. For this purpose, three genomic DNA libraries were constructed using *Dra*I, *EcoR*V and *Pvu*II to digest the genomic DNA from *S. exigua* (Adra, Spain). The resulting DNA fragments were ligated to the Genome Walker adaptors and subjected to primary and secondary PCR rounds using the primers provided with the kit and specific primers from the *repat* sequences (Table 1) following the manufacturer's protocol. Amplified products were purified, cloned in a pGEM-T Easy cloning system (Promega) and sequenced as described above.

Sequence analysis

SEQUENCHER™ 4.0.5 software (Gene Codes Corp. Ann Arbor, MI, USA) was used for sequence contig assembly. Nucleotide sequence alignments were performed using the CLUSTALW program (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Searches in databases were carried out with the BLAST program (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). Sequence motifs of gene regulation and expression in 5′- and 3′-UTRs were searched with UTRscan and UTRblast at UTResource (http://www.ba.itb.cnr.it/UTR/). Prediction of promoter regions in the upstream sequence of *repat1* was obtained with the NEURAL NETWORK PROMOTER PREDICTION program (http://www.fruitfly.org/seq_tools/promoter.html) and the PROMOTER SCAN program (http://www-bimas.cit.nih.gov/molbio/proscan). Searching for transcription factor binding sites was carried out with the TFSEARCH program (http://www.cbrc.jp/research/db/TFSEARCH.html) and the MOTIF search program (http://motif.genome.jp). DNA repeats were analysed using the TANDEM REPEATS FINDER program (http://tandem.bu.edu/trf/trf.html). Detection of regions of genomic sequences rich in the CpG pattern (CpG islands) was reported with the EMBOSS CpGPlot/CpGReport/Isochore utility (http://www.ebi.ac.uk/emboss/cpgplot/). Screening of insertion elements was performed with the RepeatMasker Web Server (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker).

Sequences have been deposited in the GENEBANK database with the following accession numbers: repat1 gene: EU797187; repat2 gene: EU797188.

Plasmid construction with repat1 *upstream sequences*

In order to obtain the final constructs, a sequential cloning process was carried out with pGEM-T Easy (Promega), pGL3-Enhancer (Promega) and pFBD-GFP (Kaba *et al.*, 2002) vectors. Upstream regions were obtained by PCR with the primers described in Table 3. The amplified products were subcloned into the pGEM-T

Table 3. Nucleotide sequences and localization of primers used in *repat1* constructs

Name	Position from transcription start site	Orientation	Restriction enzyme site	Sequence (5′–3′)
138	–738/–707	+	*Sac*I	CTGGAGCTCTCGCGTGGTCGACAAAACAATG
140	+48/+79	+	*Sac*I	CCTGAGCTCCAGGTTGTATATTCATTCTAAG
141	+253/+282	+	*Sac*I	AAGGAGCTCGAGCTTATGGATTACCTGAT
143	+464/+435	–	*Bgl*II	AAAAGATCTATCTGCAAGAACACGTAGGC
145	+464/+435	–	*Xho*I	AAACTCGAGATCTGCAAGAACACGTAGGC
146	+47/+80	–	*Xho*I	CTTCTCGAGATGAATATACAACCTGAAGCCAAG

Underlined letters indicate restriction sites used for cloning.

Easy vector and double-digested with *Sac*I/*Xho*I or with *Sac*I/*Bgl*II. Purified digestion products were subcloned in the PGL3-Enhancer vector upstream of the firefly *luciferase* gene. Double digestions with *Sac*I/*Xba*I released the constructs containing the upstream sequence of *repat1* followed by the *luciferase* gene. These *Sac*I/*Xba*I fragments were cloned in a pFBD-GFP plasmid in which the late polyhedrin (PH) promoter was removed from the *Bst*1107I to *Stu*I site. From the PGL3-Enhancer vector, the *luciferase* gene was also obtained by digestion with *Sac*I/*Xba*I and cloned in the pFBD-GFP plasmid downstream of the PH promoter.

Generation of recombinant baculoviruses

Recombinant baculovirus (bacmids) carrying the *GFP* gene and the luciferase constructs were obtained as previously described (Herrero *et al*., 2005). Isolated bacmid DNA was used to transfect Sf21 cells using the Insect Genejuice® transfection reagent (Novagen, Darmstadt, Germany), resulting in recombinant viruses with upstream regions of the *repat1* gene (baculoviruses A–D, Fig. 2A) or control viruses (baculoviruses E and F, Fig. 2A). Recombinant viruses were grown to high titre stock using standard procedures (King & Possee, 1992).

Luciferase assays

The ability of the upstream region of the *repat1* gene to promote transcription was tested by luciferase expression as a reporter gene in Sf21 and HzGUT cell cultures. The full-length construct [A (–738/+464)] and promoter-less construct [F (ΔP)] baculoviruses were used to infect culture cells. Culture wells were seeded to a 70–80% confluence and incubated for 30 min to complete adhesion. Multiplicity of infection (MOI) values were calculated in Sf21 cells. In order to get a homogenous infection, a MOI value of 20 was used for both cell types. Cells were infected with each virus in the corresponding serum free medium. After 3 h of infection, the medium was removed and fresh complete medium was added. Cell cultures were collected at 4, 10, 22, 34 and 46 h.p.i. and gently washed in phosphate-buffered saline (8 mM Na_2HPO_4, 2 mM KH_2PO_4, 150 mM NaCl; pH 7.4). Luciferase activity was determined using the Bright-Glo™ Luciferase Assay System (Promega) and a Victor2 multiplate-reader (model: 1420-041, Perkin-Elmer, Waltham, MA, USA). Results were expressed in luciferase units of the full-length construct [A (–738/+464)] after subtraction of the basal transcriptional activity of the *luciferase* gene obtained by the promoter-less construct [F (ΔP)]. Reported values are the average of duplicate values from two independent replicates.

Recombinant baculoviruses carrying different constructs of the upstream region of *repat1* plus the *luciferase* gene were used to infect HzGUT cells as described above. Luciferase activities were measured at 30 h.p.i and normalized to GFP values. Results were expressed as relative luciferase activity of the full-length construct A (–738/+464) after subtraction of the basal activity obtained with the baculovirus control F (ΔP).

Acknowledgements

We would like to thank Cynthia L. Goodman for providing us with the RP-HzGUT-AW1 cell line. Research was supported by the Spanish Ministry of Education and Science (Research contract from the 'Ramón y Cajal' program and Project AGL2005-07909-C03-03), and the Generalitat Valenciana (APOSTD/2007/041).

References

Agaisse, H. (2007). An adaptive immune response in *Drosophila*? *Cell Host Microbe* **1**: 91–93.

Bel, Y. and Escriche, B. (2006). Common genomic structure for the Lepidoptera cadherin-like genes. *Gene* **381**: 71–80.

Blissard, G.W. (1996). Baculovirus–insect cell interactions. *Cytotechnology* **20**: 73–93.

Boutros, M., Agaisse, H. and Perrimon, N. (2002). Sequential activation of signaling pathways during innate immune responses in *Drosophila*. *Dev Cell* **3**: 711–722.

Cheng, T., Zhao, P., Liu, C., Xu, P., Gao, Z., Xia, Q. *et al*. (2006). Structures, regulatory regions, and inductive expression patterns of antimicrobial peptide genes in the silkworm *Bombyx mori*. *Genomics* **87**: 356–365.

Cheng, T.C., Zhang, Y.L., Liu, C., Xu, P.Z., Gao, Z.H., Xia, Q.Y. *et al*. (2008). Identification and analysis of Toll-related genes in the domesticated silkworm, *Bombyx mori*. *Dev Comp Immunol* **32**: 464–475.

Cherry, S. and Silverman, N. (2006). Host-pathogen interactions in *Drosophila*: new tricks from an old friend. *Nat Immunol* **7**: 911–917.

Chung, B.Y., Simons, C., Firth, A.E., Brown, C.M. and Hellens, R.P. (2006). Effect of 5′UTR introns on gene expression in *Arabidopsis thaliana*. *BMC Genomics* **7**: 120.

Decker, T., Kovarik, P. and Meinke, A. (1997). GAS elements: a few nucleotides with a major impact on cytokine-induced gene expression. *J Interferon Cytokine Res* **17**: 121–134.

Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana-Arnoux, D., Hetru, C. *et al*. (2005). The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*. *Nat Immunol* **6**: 946–953.

Goodman, C.L., Wang, A.A., Nabli, H., McIntosh, A.H., Wittmeyer, J.L. and Grasela, J.J. (2004). Development and partial characterization of heliothine cell lines from embryonic and differentiated tissues. *In Vitro Cell Dev Biol Anim* **40**: 89–94.

Herrero, S., Combes, E., van Oers, M.M., Vlak, J.M., de Maagd, R.A. and Beekwilder, J. (2005). Identification and recombinant expression of a novel chymotrypsin from *Spodoptera exigua*. *Insect Biochem Mol Biol* **35**: 1073–1082.

Herrero, S., Ansems, M., van Oers, M.M., Vlak, J.M., Bakker, P.L. and de Maagd, R.A. (2007). REPAT, a new family of proteins induced by bacterial toxins and baculovirus infection in *Spodoptera exigua*. *Insect Biochem Mol Biol* **37**: 1109–1118.

Kaba, S.A., Nene, V., Musoke, A.J., Vlak, J.M. and van Oers, M.M. (2002). Fusion to green fluorescent protein improves expression levels of *Theileria parva* sporozoite surface antigen p67 in insect cells. *Parasitology* **125**: 497–505.

King, L.A. and Possee, R.D. (1992). *The baculovirus expression system: a laboratory guide*. Chapman and Hall, London.

Le Hir, H., Moore, M.J. and Maquat, L.E. (2000). Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev* **14**: 1098–1108.

Le Hir, H., Nott, A. and Moore, M.J. (2003). How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* **28**: 215–220.

Li, X., Schuler, M.A. and Berenbaum, M.R. (2006). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu Rev Entomol* **52**: 231–253.

Loeb, M.J., Martin, P.A., Hakim, R.S., Goto, S. and Takeda, M. (2001). Regeneration of cultured midgut cells after exposure to sublethal

doses of toxin from two strains of *Bacillus thuringiensis*. *J Insect Physiol* **47**: 599–606.

Ma, G., Roberts, H., Sarjan, M., Featherstone, N., Lahnstein, J. Akhurst, R. *et al*. (2005). Is the mature endotoxin Cry1Ac from *Bacillus thuringiensis* inactivated by a coagulation reaction in the gut lumen of resistant *Helicoverpa armigera* larvae? *Insect Biochem Mol Biol* **35**: 729–739.

Meredith, J. and Storti, R.V. (1993). Developmental regulation of the *Drosophila tropomyosin II* gene in different muscles is controlled by muscle-type-specific intron enhancer elements and distal and proximal promoter control elements. *Dev Biol* **159**: 500–512.

Nott, A., Le, H.H. and Moore, M.J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev*. **18**: 210–222.

Rivkin, H., Kroemer, J.A., Bronshtein, A., Belausov, E., Webb, B.A. and Chejanovsky, N. (2006). Response of immunocompetent and immunosuppressed *Spodoptera littoralis* larvae to baculovirus infection. *J Gen Virol* **87**: 2217–2225.

Rose, A.B. (2004). The effect of intron location on intron-mediated enhancement of gene expression in *Arabidopsis*. *Plant J* **40**: 744–751.

Sonoda, S., Ashfaq, M. and Tsumuki, H. (2006a). Genomic organization and developmental expression of *glutathione S-transferase* genes of the diamondback moth, *Plutella xylostella*. *J Insect Sci* **6**.

Sonoda, S., Ashfaq, M. and Tsumuki, H. (2006b). Cloning and nucleotide sequencing of three heat shock protein genes (*hsp90, hsc70, and hsp19.5*) from the diamondback moth, *Plutella xylostella* (L.) and their expression in relation to developmental stage and temperature. *Arch Insect Biochem Physiol* **62**: 80–90.

Vaughn, J.L., Goodwin, R.H., Tompkins, G.J. and McCawley, P. (1977). The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In Vitro* **13**: 213–217.

Vilmos, P. and Kurucz, E. (1998). Insect immunity: evolutionary roots of the mammalian innate immune system. *Immunol Lett* **62**: 59–66.

Wang, L. and Ligoxygakis, P. (2006). Pathogen recognition and signalling in the *Drosophila* innate immune response. *Immunobiology* **211**: 251–261.

Washburn, J.O., Trudeau, D., Wong, J.F. and Volkman, L.E. (2003). Early pathogenesis of *Autographa californica* multiple nucleopolyhedrovirus and *Helicoverpa zea* single nucleopolyhedro virus in *Heliothis virescens*: a comparison of the 'M' and 'S' strategies for establishing fatal infection. *J Gen Virol* **84**: 343–351.

Yamakawa, M. and Tanaka, H. (1999). Immune proteins and their gene expression in the silkworm, *Bombyx mori*. *Dev Comp Immunol* **23**: 281–289.

