Molecular Characterization, cDNA Cloning and Phylogenetic Analysis of Cecropin Gene Isolated from Bacterial-challenged Cotton Leafworm, *Spodoptera littoralis*

Alaa Eddeen M. Seufi

Entomology Department, Faculty of Science, Cairo University, Giza, Egypt

ABSTRACT

Experiments were designed and conducted to isolate and characterize the full length cDNA of cecropin gene from the cotton leafworm, *Spodoptera littoralis*. In addition, sequence and phylogenetic analyses of this gene were studied. Using differential display technique, a typical cationic insect cecropin gene (*Spli*Cec) was isolated from bacterial-challenged *S. littoralis* haemolymph. Five differentially displayed bands were sequenced. Based on sequence analyses of the data, specific primers for the full length of cecropin were designed and successfully amplified 388 bp amplicon. The integration of the results revealed that the 388 bp-PCR product has one open reading frame (*orf*) of 186 bp long, including both start codon (AUG) and stop codon (UGA). The deduced amino acid sequence of *Spli*Cec indicated that the full length prepropeptide consists of a 22-residue signal peptide, a dipeptide prosequence and a 38-residue mature peptide. The *Spli*Cec sequence created significant similarity with many insect cecropin sequences, especially those of *M. sexta* and *A. convolvuli*.

Keywords: *Spodoptera littoralis*, immune response, DD-PCR, cecropin gene.

INTRODUCTION

Recently, a large number of antimicrobial peptides (AMPs) have been identified and characterized from a variety of insects, including the Lepidoptera, Hymenoptera, Diptera, and Coleoptera. It has been generally accepted that insect AMPs suppress bacterial infections, and have minimal toxic and allergic side effects in host cells (Bulet *et al*., 1999). These AMPs are classified on the basis of their amino acid sequences and secondary structures into cecropins, defensins, and peptides with an over representation of proline and/or glycine residues, e.g., lebocins and moricins (Bulet *et al*., 1999). Cecropin was initially isolated from bacterially challenged *Hyalophora cecropia* pupa (Hultmark *et al.,* 1980) from which the term cecropin was derived. Cecropins constitute a main part of the cell-free immunity of insects. They have been given various names; bactericidin, sarcotoxin, etc. All of these peptides are structurally related (Boman *et al.,* 1991). The active mature cecropins without cysteine residues are typically 35 to 39 amino acids in length, and form two amphipathic α-helices connected by a hinge region (Saito *et al*., 2005). They evidenced broad-spectrum activity against a broad range of bacteria (Vizioli *et al*., 2000), as well as certain fungal (Andra *et al*., 2001) and metazoan parasites (DeLucca *et al*., 1997). They lyse bacterial cell membranes; they also inhibit proline uptake and cause leaky membranes (Wade *et al*., 1992).

This study reported on cDNA cloning and characterization of a cationic cecropin gene encoding for antimicrobial peptide, *Spli*Cec, from haemolymph of the Egyptian

cotton leafworm, *S. littoralis*. In addition, sequence and phylogenetic analyses of the *Spli*Cec gene were also reported.

MATERIALS AND METHODS

Insects and bacterial strains

Laboratory colony of the cotton leafworm, *S. littoralis*, used for our experiments was originally collected from okra field at Giza, Egypt in 1995 and maintained in the insectary of the Department of Entomology, Faculty of Science, Cairo University according to the technique described by Levinson and Navon (1969) and kept at 25 °C, 65–70% RH and 14L: 10D photoperiod cycle.

Three gram (+) bacteria: *Staphylococcus aureus, Bacillus subtilis* and *Streptococcus sanguinis* and three gram (-) bacterial strains: *Escherichia coli* (D_{31}), *Proteus vulgaris* and *Klebsiella pneumoniae*, were obtained from the Unit for Genetic Engineering and Agricultural Biotechnology, Faculty of Agriculture, Ain Shams University and used for insect immunization. Bacteria were grown in a peptone medium (1%), supplemented with 1% meat extract and 0.5% NaCl, at 37 °C in a rotary shaker.

Bacterial challenge and haemolymph collection

Bacterial challenge was performed as described by Seufi *et al.* (2011). Haemolymph was collected at 24, 48 and 72 h post-infection (p.i.) at 4 °C (500 µl/ each), containing few crystals of phenylthiourea to prevent melanization. Aliquots of 100 µl each were stored at -80 °C until investigated. Control group was injected with bacteria-free saline solution.

Differential display using primers corresponding to cecropin-like sequences (DD-PCR)

Total RNA of the insect haemolymph (300-500 µl) was extracted using RNeasy kit according to the manufacturer's instructions (Qiagen, Germany). Residual genomic DNA was removed from RNA using RNase-free DNase (Ambion, Germany). RNA integrity and purity were justified by examining 260/280 and 260/230 ratios for protein and solvent contamination. Reverse transcription reaction was carried out according to the ABgene protocol (ABgene, Germany). The cDNA was aliquoted and stored at -80 °C until processed.

A total reaction volume of 25 µl containing 2.5 µl PCR buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 1 U *Taq* DNA polymerase (AmpliTaq, Perkin-Elmer), 2.5 µl of 10 pmol primer (Table 1) and 2.5 µl of each cDNA was cycled in a DNA thermal cycler (Eppendorf, Mastercycler 384, Germany). The amplification program was one cycle at 94 °C for 5 min (hot start), followed by 40 cycles at 94 °C for 1 min, 40 °C for 1 min and 72 °C for 1 min. The reaction was then incubated at 72 °C for 10 min for final extension. PCR product was visualized on 2 % agarose gel and photographed using gel documentation system. For DNA contamination assessment, a no–reverse transcription control reaction was performed.

Five reproducible bacterial-induced bands were eluted, cloned in *PCR-TOPO* vector (Invitrogen, USA) and sequenced using M_{13} universal primer and model 310 automated sequencer (Applied Biosystems, Foster City, CA, USA). Analyses of nucleotide and deduced amino acid sequences were carried out using EditSeq-DNAstar Inc., Expert Sequence Analysis software, Windows 32 Edit Seq 4.00 (1989-1999) and ExPasy database (http://expasy.org/tools/dna.html).

Full length cDNA cloning and sequence analysis

Based on the sequence and alignment data, specific primers for cecropin-related sequences were designed and tried for reverse transcription polymerase chain reaction

(RT-PCR). RT-PCR reaction was performed as previously described in this section regarding to the optimum annealing temperature (T_a) for each specific primer set. Positive PCR products were visualized and eluted from the gel using GenClean Kit (Invitrogen Corporation, San Diego, CA, USA) following the manufacturer's instructions. The purified PCR products were cloned into *PCR-TOPO* vector with TOPO TA cloning kit (Invitrogen, USA) following the manufacturer's instructions. Ligation mix was used to transform competent *E. coli* strain $TOPO_{10}$ provided with the cloning kit. White colonies were screened using PCR as described earlier in this section. Two positive clones of *Spli*Cec fragment were selected and sequenced (to exclude PCR errors certainly) using the specific forward and reverse primers. In addition to the above mentioned analyses, ExPasy Proteomics Server (http://expasy.org/tools) was used to calculate physico-chemical parameters of the translated peptide (ProtParam tool). Furthermore, primary and secondary structure analyses, post-translational modifications and topology predictions were investigated using SignalP, NetGlycate, NetPhos, NetPhosK, NetSurfP, SUSI and TMpred tools.

Phylogenetic analyses

Phylogenetic analyses of the nucleotide sequence and its deduced amino acids (a.a.) were done using Mega4 software. Poorly aligned positions and divergent sequences were eliminated manually. Multiple alignment of 57 published cecropin-related nucleotide sequences was done before phylogenetic analyses to approximate sequence lengths manually. 100% homologous sequences of the same species with different accession numbers were represented by only one sequence. The cloned DNA fragment was deposited in GenBank under the JQ408983 accession number.

RESULTS

Differential display using primers corresponding to well kown cecropins

Differential display technique was used to characterize the transcript variation between bacterial-challenged and control cotton leafworm, *S. littoralis*. Fig. (1A) shows the results of differentially displayed cDNAs of bacterial-challenged and control insects using 8 primers corresponding to previously characterized cecropins (Table 1). Haemolymph samples were differentially displayed at 24, 48 and/ or 72 h p.i. with *S. aureus, S. sanguinis*, *E. coli, P. vulgaris* and *K. pneumoniae* bacterial strains. It was observed that *S. aureus*-challenged insects died 24 h p.i., *E. coli*-challenged insects died 48 h p.i. and *S. sanguinis*-challenged insects died 72 h p.i. All insects died before sampling in the case of *P. vulgaris* and *K. pneumoniae*.

Table 1: Key table for the primers used in DD-PCR study providing their names, origin and sequences.

Primer name	Origin	Sequence (5´ -- 3´)
$CecF_1$	Cecropin-based	CACAATGAATTTCTCAAGAGC
$CecR_1$	Cecropin-based	CCTTTCATTTCCCACGAGCG
$CecF_2$	Cecropin-based	ATGGGCCGCAACATCCG
$CecR_2$	Cecropin-based	GCATCGCGAGGGTCGGGG
Fwd_1	Cecropin-based	CGATGTCCAGTCATCAGTCGC
Fwd_2	Cecropin-based	CCGCAAAATGAATTTC
Fwd_3	Cecropin-based	CAAAATGAACTTCTCTCGCG
Fwd_4	Cecropin-based	AAATGAAATTCTCTCGCC
FLCecF	Full length cecropin	GTCGCCACTGTTAGTAAC
FLCecR	Full length cecropin	AAGAAGGGTCTGAACCTGAAAATGT

All oligonucleotides were synthesized by Invitrogen, USA and HPLC purified.

Differential display results revealed that the average number of bands per sample was 3.4 bands for each amplification reaction. The total number of bands (transcripts) resolved in 1.5 % agarose gel for both control and challenged insects was 124 (molecular size ranged from >1300 to ~80 bp). Thirty two polymorphic bands (34.2 %) were differentially displayed with 5 of the used primers. Five reproducible, infection-induced bands were cloned and sequenced using M_{13} universal primer. Analyses of the results revealed that a cDNA fragment of 388 bp was amplified containing the open reading frame (*orf*) of a cecropin gene.

Primer design, RT-PCR amplification and cloning of cecropin gene

Specific primers for the full length cecropin gene were designed. These primers would be used later in the following reactions during this study. Nucleotide sequence of the used primers was illustrated in Table (1). PCR was optimized for each primer set and primers successfully produced positive PCR amplicons of 186 bp for the full length cecropin sequence (Fig. 1B). The full length fragment includes one open reading frame (*orf*) of cecropin gene (positions: 34 (AUG) - 222 (UAA)). Subsequently this segment (*Spli*Cec) was cloned into PCR-TOPO vector (Fig. 1B, lane 2) and transformed cells were tested with PCR using the same primers (Fig. 1B, lane 4). Using this screening method, clone PCR-TOPO*Spli*Cec was tested as positive (Fig. 1B, lane 4).
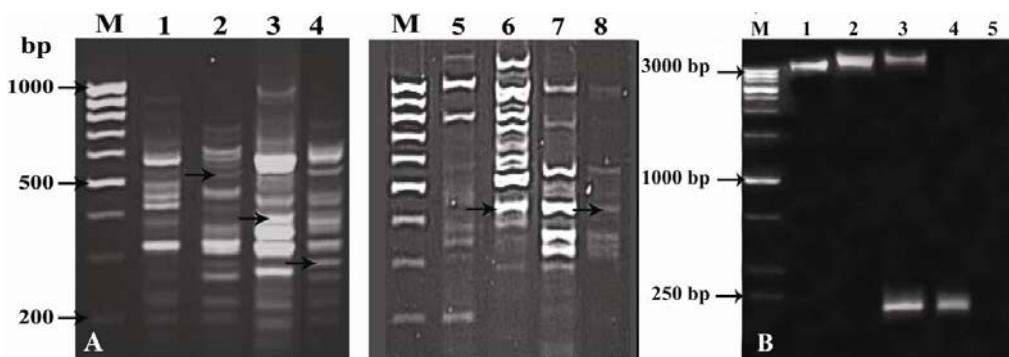
Fig. (1): (A): Representative 2% agarose gels of DD-PCR patterns generated from control and *S. aureus*, *E. coli* and *S. sanguinis*-challenged haemolymph samples using 8 primers corresponding to well known defense genes. (B): Lanes M: DNA marker, lanes 1 and 5: controls of different treatments, lanes 2 and 3: 24 and 48 h p.i. by *E. coli*, Lanes 4, 6 and 7: 24, 48 and 72 h p.i. by *S. sanguinis* and Lane 8: 42 h p.i. by *S. aureus*. Arrows refer to differentially displayed sequenced bands. (B): 2% gel electrophoresis showing positive PCR and cloning products. Lanes 1, 2, 3 and 4 show empty *PCR-TOPO*, *E. coli* harbouring *PCR-Spli*Cec, *PCR-Spli*Cec after digestion with *EcoR*I and positive control (186 bp), respectively. Lane 5: PCR mix without DNA used as negative control. The size of the bands is shown in bp.

Nucleotide sequence and sequence analyses

Nucleotide sequences of the *Spli*Cec and its deduced amino acid sequence are shown in Fig. (2). A single *orf* that could encode a polypeptide of 62 amino acids was detected for *Spli*Cec. One stop codon was found at the 3′ end of the sequence. The flanking region of the initiation codon ATG is ACAATGAAC, and the length of 3′ untranslated region was 166 bp before the poly (A) track (Fig. 2). Two putative polyadenylation sequences (AATAAA) were located 78 and 146 bp downstream from the stop codon (Fig. 2). The identified cecropin *orf* includes signal peptide (66 bp),

propeptide (6 bp) and mature peptide (114 bp). The deduced *Spli*Cec polypeptide (prepropeptide) contains 9 strongly basic, 3 strongly acidic, 38 hydrophobic and 12 polar amino acids. The calculated molecular masses of the full length and mature cecropins, were 6.58 and 3.99 KDa, respectively. The calculated isoelectric points (PIs) were 11.36 and 10.93, respectively and the net charges at pH 7.0 were 6.0 and 4.0, respectively. The cecropin prepropeptide was classified as unstable protein (Instability Index (II): 41.72) and its mature peptide was classified as stable protein (II: 32.61). Ratios of hydrophilic residues were 19 and 26% for prepropeptide and mature cecropin peptides, respectively.


```
5′ UTR-CCTGTGCAGTCGCCACTGTTAGTAACATTCACA (33 bp)
ATG   AAC   TTC   TCA   CGC   GTT   TTG   TTC   TTC   GCC
 M     N     F     S     R     V     L     F     F     A
TTG   GCA   TGC   CTC   AGA   GCA   TTC   GCC   ATG   GCT
 L     A     C     L     R     A     F     A     M     A
TCA   GCT   GCA   CCC   CGC   TGG   CCC   TTC   AAG   GAC
 S     A  ▼  A     P  ▼  R     W     P     F     K     D
TTG   GAG   AAG   GCT   GGT   CAG   CGA   GTC   CGC   GAC
 L     E     K     A     G     Q     R     V     R     D
GCC   GTG   ATC   AGC   GCG   GCT   CCG   GCA   GTC   GCC
 A     V     I     S     A     A     P     A     V     A
ACC   GTT   GGA   CAG   GCG   GCC   GCC   ATG   GCC   AGG
 T     V     G     Q     A     A     A     M     A     R
GGG   AAA   TAA   3′ UTR-GCAACAAATGCTTTATCTATGAATATGT
 G     K     *    TTATTAATTATACAAGTTTCATATAACTTTATTA
CAATGATGATTTGGTAAAATAAATACTTATATACCCTGAAAATGTAATACA
TTCCCAAATTTTCAAAATGTAAAAATAACAACAAAATAAATGCTTTACATA
AGAAAAAAAAAAAA (166 bp + polyA)
```

Fig. 2: Nucleotide sequence and corresponding deduced amino acid sequence of *S. littoralis* cecropin gene (*SpliCec*). Cleavage sites of signal and propeptides are indicated by arrows. Mature peptide is double underlined. Asterisk indicates the stop codon. Boxed sequence represents the two putative polyadenylation signal.

Blast search of nucleotide sequence of the *Spli*Cec to all available cecropin-related sequences in GenBank database created significant alignment with 17 insect-published sequences (2 *M. sexta* bactericidins (Acc# M23661, M23662), *A. convolvuli* CecD (Acc# GQ888768), 3 *B. mori* CecD (Acc# NM_001043368, AB010825, AK383622), 6 *B. mori* enbocins (Acc# AB239447, NM_001098374, FJ373019, AK383331, NM_001099840, AB289653), 2 *A. mylitta* Cec-like (Acc# DQ666493, DQ666494), *A. pernyi* Cec-like (Acc# EU557306), 2 *H. cecropia* CecD (Acc# X06673, M63846)). The percentage identity ranged from 90% for *M. sexta* to 68% for *A. pernyi*. Interestingly, the *Spli*Cec putative peptide showed significant alignment with more than 82 insect-published peptides (64 cecropins, 5 cecropin-like, 2 bactericidins, a hinnavin II, a hyphancin-3E, a papiliocin, a defense protein and 7 enbocin peptides). The percentage identity ranged from 85% for the bactericidin of *M. sexta* (Acc# AAA29306) to 33% for CecB of *D. melanogaster* (Acc# BAA28722).

The full length of lepidopteran preprocecropin peptide ranged from 67 a.a. for Cec-6 of *M. sexta* and CecD of *T. ni* (Acc# AAO74638 and ABV68873, respectively) to 55 amino acids (a.a.) for CecD of *S. litura* (Acc# ABQ51093). Lepidopteran secropins are linear, amphipathic peptides of 35–39 a.a. residues derived from preproproteins of 62–64 residues. These preproproteins contain an amino-terminal signal peptide (21-22 a.a.) and a prosequence of 2-4 a.a. residues.

SignalP analysis showed that the cleavage site for the potential signal peptide of the *Spli*Cec was predicted between 22-Ala and 23-Ala. Further, a cleavage site

between 24-Pro and 25-Arg was also predicted by the alignment of the amino acid sequence of the *Spli*Cec with that of 22 lepidopteran cecropins. Thus in the *Spli*Cec sequence, the signal peptide makes up the first 22 residues, leaving a dipeptide Ala-Pro (AP) before the start of 38 a.a. residues mature ceropin peptide. In addition to the precise conservation of proline (P) at N-terminal domain, the aromatic tryptophan residue (W) at position 2, the lysine residues (K) at positions 5 and 9, the arginine residue (R) at position 13 and the glycine-lysine residues (GK) at C-terminal domain were also observed. The mature cecropin consists of two amphipathic α-helices interrupted by a hinge region composed of Ala-Pro dipeptide (AP). These motifs were found to be conserved at equivalent positions in the sequences of some other insect cecropins (Fig. 2).

Primary, secondary structure analyses, post-translational modifications and topology predictions revealed that amino acid sequence of the putative *Spli*Cec peptide has two potential glycated lysine residues (positions: 29 and 62) and two kinase specific phosphorylation sites (Ser: 1 at position 44, Thr: 1 at position 51) with highest score: 0.66 PKC at position 51. NetSurfP results for the probability for α-helix confirmed that the putative mature *Spli*Cec peptide consists of two α-helices (first: 27-44 and second: 48-59) interrupted by a hinge dipeptide (AP). In addition, two strong transmembrane helices (one primary: 3-23 with inside to outside orientation and one secondary: 41-59 with outside to inside orientation) were predicted.

Phylogenetic analyses of the *Spli*Cec sequence

Phylogenetic analyses have been performed on the *Spli*Cec nucleotide seuquence and its deduced polypeptide and the results of these analyses are shown in Fig. (3).
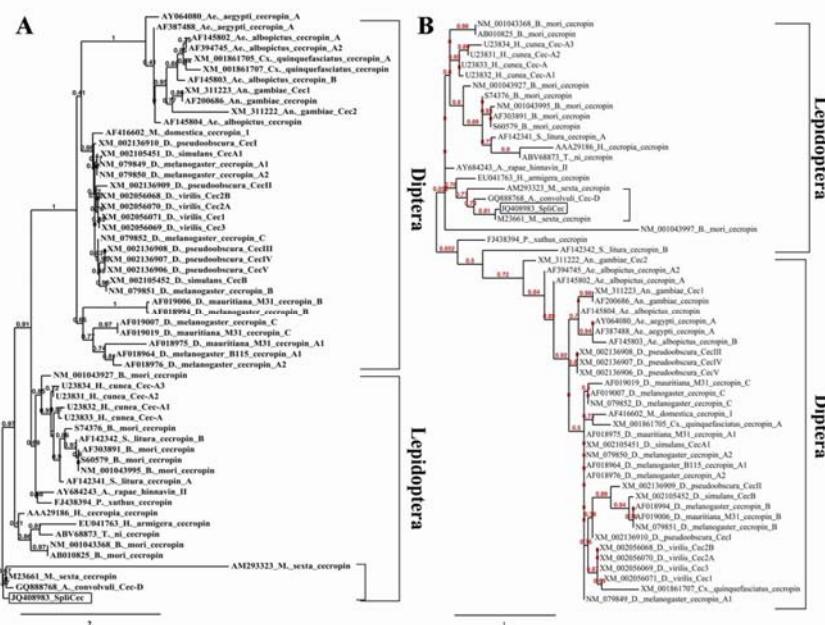
Fig. 3: Phylogenetic analysis of *SpliCec* nucleotide (A) and its deduced amino acid (B) sequences compared to 58 and 57 sequences, respectively. Phylogenetic trees were generated by neighbor-joining distance analysis using Phylogeny.fr web service, One Click mode. Full sequence names and accession numbers are included in the trees.

In the case of *Spli*Cec nucleotide sequence, a phylogenetic tree was generated from 58 cecropin-related sequences (21 insect species including 11 Lepidoptera and 10 Diptera) by neighbor-joining distance analysis with maximum sequence difference 1.0 (Fig. 3A). The topology shows two distinct lineages including 21 (Lepidoptera)

and 37 (Lepidoptera and Diptera) cecropin-related sequences, respectively. The maximum nucleotide sequence divergence was exhibited in the lineage II (21 phylogenetic groups). Meanwhile, the cecropin sequences appear in the other lineage as less divergent clades (14 phylogenetic groups). The *Spli*Cec was clustered with *M. sexta* and *A. convolvuli* cecropins (Acc# M23661 and GQ888768, respectively) in a monophyletic sister clade (Fig. 3A). Surprisingly, two lepidopteran secropins were clustered with the dipteran lineage II (Fig. 3A). In the case of *Spli*Cec deduced amino acid seuquence, a phylogenetic tree was generated from sequence data of 57 published sequences (21 insect species including 11 Lepidoptera and 10 Diptera) by neighbor-joining distance analysis with maximum sequence difference 2.0 (Fig. 3B). The topology shows two distinct lineages including 4 (Lepidoptera) and 53 (Lepidoptera and Diptera) cecropin peptides, respectively. The maximum divergence of amino acid sequences was exhibited in lineage I (33 phylogenetic groups). However, minimum divergence was observed in the other lineage (3 phylogenetic groups). The *Spli*Cec putative peptide was clustered with *M. sexta* and *A. convolvuli* cecropins (Acc# M23661 and GQ888768, respectively) in a monophyletic sister clade (Fig. 3B). Meanwhile, the other lepidopteran sequence was grouped in a separate phylogenetic group (13 sister clades) in the lineage I (Fig. 3B).

DISCUSSION AND CONCLUSION

In the present study, the common bands revealed by DD-PCR in both control and challenged samples may represent the house-keeping genes. Some bands were recorded in control insects and disappeared in challenged ones (genes were turned off). On the other hand, many bands were induced as a result of bacterial-challenge at different time intervals post-infection. DD-PCR technique is considered a powerful genetic screening tool for complicated dynamic tissue processes, to detect and compare altered gene expression in eukaryotic cells, and to screen and to characterize differentially expressed mRNAs (Santana *et al.,* 2006), because it allows for simultaneous amplification of multiple arbitrary transcripts. Many publications described the enhancement of the insect immune system and induction of AMPs due to stress and/ or infection (*e.g.* Seufi *et al*., 2011). Cecropins were isolated from only four insect orders: Lepidoptera (*e.g.* Kim *et al*., 2004, Kaneko *et al*., 2007, Hong *et al*., 2008, Kim *et al*., 2010), Hymenoptera (*e.g.* Orivel *et al*., 2001), Coleoptera (*e.g.* Saito *et al*., 2005) and Diptera (*e.g.* Vizioli *et al*., 2000, Jin *et al*., 2010). The full length cDNA of the *Spli*Cec contained a 186 bp *orf* encoding 62 amino acids, preceded by a 33 bp 5′-UTR and followed by a 166 bp 3′-UTR. The putative translational start site sequence (ACAATGAAC) conformed well to the Kozak consensus sequence and keeps the adenine nucleotide at position -3 as a universal feature in all eukaryotic mRNA (Kozak, 1984). Many insect cecropins also conformed Kozak consensus sequence (ACAATGAAC for *An. gambiae* (Vizioli *et al*., 2000), ACAATGAAT for *B. mori* (Hong *et al*., 2008), AAAATGAAT for *G. mellonella* (Kim *et al*., 2004) and AAAATGAAT for *P. xuthus* (Kim *et al*., 2010). The *Spli*Cec cDNA harbored two putative polyadenylation sequences (AATAAA). Two polyadenylation sequences were reported in the CecB of *H. cecropia* (Xanthopoulos *et al*., 1988, Boman *et al*., 1991). On the other hand, only one polyadenylation sequence was reported in the following cases: cecropins A and D of *H. cecropia* (Gudmundsson *et al*., 1991), cecropin-like of *Pachycondyla goeldii* (Orivel *et al*., 2001), cecropin-like of *B. mori* (Kaneko *et al*., 2007), cecropins of *An. gambiae, G. mellonella, B. mori* and *P. xuthus* (Vizioli *et al*., 2000, Kim *et al*., 2004, Hong *et al*.,

2008, Kim *et al*., 2010). The *Spli*Cec cDNA identified in this study encodes a protein of 62 a.a. showing a high degree of similarity to insect cecropins, particularly to the cecropins of *M. sexta, A. convolvuli* and *H. cecropia*. Comparison of the *Spli*Cec a.a. sequence to 22 lepidopteran cecropins showed some replacements. The possibility that it comes from polymorphism cannot be excluded, because cecropins were identified from different insect species (Kaneko *et al*., 2007). Despite of the identified replacements, many conserved regions were observed throughout the 22 compared sequences. These conserved regions may be very important in designing universal primers to detect lepidopteran cecropin genes. A potential 22-residue signal peptide, a dipeptide prosequence and a 38-residue mature peptide were predicted by SignalP software and by multiple alignment of the amino acid sequence of the *Spli*Cec peptide with that of 22 lepidopteran cecropins. The rule is that signal peptide sequences are not conserved (Von Heijne, 1985). The signal peptides of lepidopteran cecropins probably make up the first 22 a.a., leaving 2- to 4-residue propeptide before the start of a linear, amphipathic mature peptide of 35 to 39 a.a. residues (Boman *et al*., 1991). The deduced amino acid sequence of two *B. mori* cecropin-like sequences were reported to contain 20-residue signal peptides, dipeptide prosequences and 37-residue mature peptides (Kaneko *et al*., 2007). Thus, the *Spli*Cec peptide (62 a.a.) was comparable in size to that of the other lepidopteran cecropins (55-67 a.a.). Moreover, the conserved pattern of the putative *Spli*Cec peptide was consistent with the typical features of invertebrate cecropins.

Reconstruction of the phylogenetic trees of the *Spli*Cec nucleotide seuquence and its deduced polypeptide resulted in two different topologies. In spite of constructing two different tree-topologies, both trees clustered the *Spli*Cec sequence with that of *M. sexta* and *A. convolvuli* to indicate that they descend from a common ancestor. The grouping of some lepidopteran and dipteran cecropins (e.g. *S. litura* with *An. gambiae*) in one sister clade indicated that they may be homologous or share some similarity.

The potential glycated and phosphorylated residues may serve for correct folding and stability of the protein in both prokaryotic and eukaryotic organisms (Barford *et al*., 1998). Reversible phosphorylation (phosphorylation and dephosphorylation) is very important mechanism in protein-protein interaction via recognition domains (Cole *et al*., 2003). It may also result in conformational changes in the structure of many peptides, causing them to become activated, deactivated or degraded (Olsen *et al*., 2006). In addition, two α-helices interrupted by a hinge dipeptide (AP), and two strong transmembrane helices were predicted. Structure prediction studies of *Ang*CecA supports an overall α-helical structure with a strikingly amphipathic N-terminal α-helix (Vizioli *et al*., 2000). Like many insect cecropins, *Spli*Cec also harbors Gly-Lys residues (GK) at the C-terminal end. Contrary to our results, it was reported that the majority of insect cecropin-like peptides harbor amidated C-termini (Saito *et al*., 2005). Additionally, it was previously described that *H. cecropia* CecA activity was enhanced significantly by this C-terminal amidation (Callaway *et al*., 1993). On the other hand, the amidation of *An. gambiae*'s Gly-extended cecropin did not influence the antimicrobial activity (Vizioli *et al*., 2000). However, the amidation of Gly-residue was previously theorized to protect the cecropin-like peptide against carboxypeptidase digestion (Liang *et al*., 2006). Whether the amidating enzyme works on Gly-Lys is not known. If not, a carboxypeptidase H will also have to be invoked for the removal of the terminal Lys-residue.

The *Spli*Cec precursor harbors a dipeptide (AP) between the signal and mature peptides, which is not conserved in terms of the tetrapeptides (APEP, VPEP or APSP)

in cecropins of *B. mori*, *T. ni*, *S. litura*, *H. cunea* and *G. mellonella* or the dipeptides (SP or WD) in cecropins of *P. xuthus* and *H. armigera* (Kim *et al*., 2004, Hong *et al*., 2008, Kim *et al*., 2010). In addition, the precise conservation of Pro-residue (P) at N-terminal domain, the aromatic Trp-residue (W) at position 2, the Lys-residues (K) at positions 5 and 9, and the Arg-residue (R) at position 13 confirmed a possible replacements of Lys^8 and Arg^{13} with Glu^8 and Gln^{13} in the case of *B. mori*, *H. cunea* and *P. rapae* cecropins (Kim *et al*., 2004, Hong *et al*., 2008, Kim *et al*., 2010).

Cecropin was initially isolated from a bacterially challenged *H. cecropia* pupa as a cationic antimicrobial peptide, and then a number of cecropin-like peptides have also been identified in lepidopteran, dipteran, hymenopteran and coleopteran insects.

The current results provide a typical cationic insect cecropin gene (*Spli*Cec) with a possible replacement mutation. *Spli*Cec plays an important immune role in *S. littoralis* by cooperating with other AMPs to control bacterial infection. These findings would be helpful in cecropin studies concerning ELISA, PCR and other related molecular and immunological techniques. Study of the expression profile, potential antimicrobial, hemolytic, apoptotic and antitumor activties of the *Spli*Cec recombinant protein as well as structure-function relationship studies are required for more understanding of its mode of action.

ACKNOWLEDGEMENTS

REFERENCES

Andra, J.; Berninghausen, O. and Leippe, M. (2001): Cecropins, antibacterial peptides from insects and mammals, are potently fungicidal against *Candida albicans*. *Med. Microbiol. Immunol.,* 189: 169–173.

Barford, D.; Das, A. and Egloff, M. (1998): The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.,* 27: 133-164.

Boman, H.; Faye I.; Gudmundsson, G.; Lee, J. and Lidholm, D. (1991): Cell-free immunity in *Cecropia*: a model system for antibacterial proteins. *Eur. J. Biochem.,* 201: 23–31.

Bulet, P.; Hetru, C.; Dimarcq, J. and Hoffmann, D. (1999): Antimicrobial peptides in insects; structure and function. *Dev. Comp. Immunol.,* 23: 329-344.

Callaway, J.; Lai, J.; Haselbeck, B.; Baltaian, M.; Bonnesen, S.; Weickmann, J.; Wilcox, G. and Lei, S. (1993): Modification of the C-terminus of cecropin is essential for broad-spectrum antimicrobial activity. *Antimicrob. Agents Chemother.,* 37: 1614–1619.

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

Cole, P.; Shen, K.; Qiao, Y. and Wang, D. (2003): Protein tyrosine kinases Src and Csk: a tail's tale. *Curr. Opin. Chem. Biol.,* 7: 580-585.

DeLucca, A.; Bland, J.; Jacks, T.; Grimm, C.; Clevel, T. and Walsh, T. (1997): Fungicidal activity of cecropin A. *Antimicrob. Agents Chemother.,* 41: 481–483.

Dimopoulos, G.; Richman, A.; Torre, A.; Kafatos, F. and Louis, C. (1996): Identification and characterization of differentially expressed cDNAs of the vector mosquito, *Anopheles gambiae. Proc. Nat. Acad. Sci.,* 93: 13066-13074.

Gudmundsson, G.; Lidholm, D.; Asling, B.; Gan, R. and Boman, H. (1991): The cecropin locus. Cloning and expression of a gene cluster encoding three antibacterial peptides in *Hyalophora cecropia. J. Biol. Chem.,* 266: 11510–11517.

Hong, S.; Kusakabe, T.; Lee, J.; Tatsuke, T.; Kawaguchi, Y.; Kang, M.; Kang, S.; Kim, K. and Nho, S. (2008): Structure and expression analysis of the cecropin-E gene from the silkworm, *Bombyx mori. Biosci. Biotechnol. Biochem.,* 72: 1992-1998.

Hultmark, D.; Steiner, H.; Rasmuson, T. and Boman, H. (1980): Insect immunity. Purification and properties of three inducible bactericidal proteins from haemolymph of immunized pupae of *Hyalophora cecropia. Eur. J. Biochem.,* 106: 7–16.

Jin, X.; Me, H. and Li, X. (2010): Apoptosis-inducing activity of the antimicrobial peptide cecropin of *Musca domestica* in human hepatocellular carcinoma cell line BEL-7402 and the possible mechanism. *Acta Biochim. Biophys. Sin.,* 259-265.

Kaneko, Y.; Furukawa, S.; Tanaka, H. and Yamakawa, M. (2007): Express of antimicrobial peptide genes encoding enbocin and gloverin isoform in the silkworm, *Bombyx mori. Biosci. Biotechnol. Biochem.,* 71: 2233–2241.

Kim, S.; Hong, M.; Par, S.; Choi, K.; Yun, E.; Goo, T.; Kang, S.; Suh, H.; Kim, I. and Hwang, J. (2010): Characterization and cDNA cloning of a cecropin-like antimicrobial peptide, papiliocin, from the swallow tail butterfly, *Papilio xuthus. Mol. Cells.,* 29: 419-423.

Kim, S.; Lee, K.; Yoon, H.; Park, N.; Lee, S.; Je, Y.; Jin, B. and Sohn, H. (2004): cDNA cloning, expression, and characterization of an arylphorin-like hexameric storage protein, *Age*Hex2, from the mulberry longicorn beetle, *Apriona germari. Arch. Insect Biochem. Physiol.,* 56: 61–72.

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

Levinson, Z. H. and Navon, A. (1969): Ascorbic acid and unsaturated fatty acids in the nutrition of the Egyptian cotton leaf worm. *J. Insect Physiol.,* 15: 591-595.

Liang, Y.; Wang, J.; Zhao, X.; Du, X. and Xue, J. (2006): Molecular cloning and characterization of cecropin from the housefly (*Musca domestica*), and its expression in *Escherichia* coli. *Dev. Comp. Immunol.,* 30: 249–257.

Olsen, J.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P. and Mann, M. (2006): Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell,* 127: 635-648.

Orivel, J.; Redeker, V.; Le Caer, J.; Krieri, F.; Revol-Junellesi, A.; Longeon, A.; Chaffotte, A.; Dejean, A. and Rossier, J. (2001): Ponericins, new antibacterial and insecticidal peptides from the venom of the ant *Pachycondyla goeldii. J. Biol. Chem.,* 276: 17823–17829.

Saito, A.; Ueda, K.; Imamura, M.; Atsumi, S.; Tabunoki, H.; Miura, N.; Watanabe, A.; Kitami, M. and Sato, R. (2005): Purification and cDNA cloning of a cecropin from the longicorn beetle, *Acalolepta luxuriosa. Comp. Biochem. Physiol. B.,* 142: 317–323.

Santana, F.; Nunes, F.; Vieira, C.; Machado, M.; Kerr, W.; Silva, J. and Bonetti, A. (2006): Differentially displayed expressed sequence tags in *Melipona*

scutellaris (Hymenoptera, Apidae, Meliponini) development. *Anais da Academia Brasileira de Ciências*, 78: 69-73.

Seufi, A.; Hafez, E. and Galal, F. (2011): Identification, phylogenetic analysis and expression profile of an anionic insect defensin gene, with antibacterial activity, from bacterial-challenged cotton leafworm, *Spodoptera littoralis*. *BMC Mol. Biol.*, 12: 47.

Vizioli, J.; Bulet, P.; Charlet, M.; Lowenberger, C.; Blass, C.; Muller, H.; Dimopoulos, G.; Hoffmann, J.; Kafatos, F. and Richman, A. (2000): Cloning and analysis of a cecropin gene from the malaria vector mosquito, *Anopheles gambiae*. *Insect Mol. Biol.,* 9: 75–84.

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

Wade, D.; Andreu, D.; Mitchell, S.; Silveira, A.; Boman, A.; Boman, H. and Merrifield, R. (1992): Antibacterial peptides designed as analogs or hybrids of cecropins and melittin. *Int. J. Pept. Protein Res.,* 40: 429–436.

Xanthopoulos, K.; Lee, J.; Gan, R.; Kockum, K.; Faye, I. and Boman, H. (1988): The structure of the gene for cecropin B, an antibacterial immune protein from *Hyalophora cecropia*. *Eur. J. Biochem.,* 172: 371 -376.

ARABIC SUMMARY

التوصيف الجزيئي و انتسال جزئ الدنا المكمل وتحليل النشوء والتطور لجين السيكروبين المعزول من دودة ورق القطن، *سبودوبترا ليتوراليس*، المستحثة مناعيا بالبكتيريا

علاء السيوفى
قسم علم الحشرات ـ كلية العلوم ـجامعة القاهرة ـالقاهرة ـمصر

صممت التجارب و أجريت بهدف عزل وتوصيف جزيئى لجزئ دنا مكمل لجين السيكروبين من دودة القطن، *سبودوبترا ليتوراليس*، وبالإضافة إلى ذلك، تمت دراسة تحليل تسلسل النشوء والتطور لهذا الجين. و باستخدام تقنية العرض التباينى، اختيرت خمس حزم تباينية من دم اليرقات المستحثة مناعيا بالبكتريا، و تم الكشف عن تتابعات قواعدها النيتروجينية. صممت البادئات التخصصية لجين السيكروبين بناء على نتائج كشف التتابعات السابقة، و تم عزل الطول الكامل لهذا الجين، و بعد تحليل تتابعات القواعد النيتروجينية له وجد أنه يمثل التركيب النموذجى لعائلة جين السيكروبين الحشرى. نجحت البادئات المصممة فى تكبير جزئ دنا مكمل طوله 388 زوجا من القواعد النيتروجينية. و قد أظهرت النتائج عند تحليلها أن الجزئ المعزول يحتوى على الطول الكامل للجين من شفرة البداية إلى شفرة النهاية. وقد أوضح استنتاج تسلسل الأحماض الأمينية لهذا الجزئ أن البروبريببتيد يتكون بريببتيد (22 حمض أمينى)، و بروببتيد (2 حمض أمينى)، و ببتيد ناضج يتكون من 38 حمض أمينى. و قد أظهرت الدراسات المقارنة للجين أنه يشبه تركيب جين السيكروبين الحشرى ، و خاصة حشرتى *ماندوكا سيكستا* و *أجريوس كونفولفيولاى*، مع وجود طفرة فى بعض التتابعات.