African Journal of Biotechnology Vol.11(4), pp. 758-770, 12 January, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.573
ISSN 1684–5315 © 2012 Academic Journals

Full Length Research Paper

Cloning, expression and characterisation of a novel gene encoding a chemosensory protein from *Bemisia tabaci* Gennadius (Hemiptera: Aleyrodidae)

Yaofa Li[1,2], Yuchuan Qin[1]*, Zhanlin Gao[2], Zhihong Dang[2], Wenliang Pan[2] and Guiming Xu[2]

[1]Department of Entomology, China Agricultural University, Beijing 100193, China.
[2]Plant Protection Institute of Hebei Academy of Agricultural and Forestry Sciences, Baoding 071000, China.

Accepted 12 August, 2011

Chemosensory proteins (CSPs) are thought to play an important role in olfactory mediating perception, identification, transportation, and transduction of semio-chemicals. These proteins are also associated with the regulation of circadian rhythms and maturation of certain tissues or appendages. In this study, a cDNA from *Bemisia tabaci* encoding a CSP (GU250808), denoted BtabCSP was cloned by RT-PCR and rapid amplification of cDNA ends (RACE) technique. The genomic DNA sequence comparisons revealed a 1490 bp intron flanking the full length cDNA. Sequencing and structural analyses of the full length cDNA indicated that BtabCSP is 381 bp in length, encoding 126 amino acid residues of which a 22 amino acid residue coded for a signal peptide. The predicted molecular weight of BtabCSP is 14.17 kDa. The BtabCSP amino acid residues deduced from the respective full-length cDNA shares four conserved cysteine motifs with known CSPs from other insects. Homology modelling indicated a very good fit between the structural conformation of BtabCSP and a moth CSP molecule. The results of phylogenetic analyses indicates that the CSPs gene of Hemipteran insects have two more sub-families. The recombinant BtabCSP was successfully expressed in *Escherichia coli* cells. This is the first report on the existence of chemosensory protein-coding gene in whiteflies. It will help us to elucidate the molecular basis of whitefly behaviour, and explore new approach for the management of this major pest.

Key words: Whitefly, chemosensory protein, rapid amplification of cDNA ends (RACE), genomic organisation, sequence analysis.

INTRODUCTION

Chemosensory proteins (CSPs) are small, soluble acidic proteins that are distributed in sensory appendages and tarsi (Vogt, 2005; Pelosi et al., 2005; Zhou et al., 2006). The insect CSP family includes the olfactory specific-D-like protein (OS-D-like) (McKenna et al., 1994), CSPs (Angeli et al., 1999), sensory appendage proteins (SAPs) (Robertson et al., 1999) and pherokines (PHKs) (Sabatier et al., 2003).

The first member of the CSP family was discovered more than a decade ago in *Drosophila melanogaster*, named as olfactory-specific protein D (OS-D) (McKenna et al., 1994). Since then, other members of this family were discovered in Diptera, Hymenoptera, Lepidoptera, Hemiptera, and Orthoptera (Wanner et al., 2004; Liu et al., 2005). Homologue BLAST searches (http://www.ncbi.nlm.nih.gov) identified more than 50 CSPs in Hemipteran insects such as *Acyrthosiphon pisum* and *Myzus persicae*. These CSPs are highly-conserved with identities ranging from 30 to 90% at the amino acid level (Liu et al., 2005). These CSPs share common characteristics such as a relatively low molecular weight of about 13 kD, amino acid residues ranging from 100 to 115 and four conserved cysteine motifs (Liu et al., 2005). In addition, many members of the OS-D-like protein family have several conserved sequence motifs, including the N-terminal YTTKYDN (V/I) (N/D) (L/V) DEIL, central

*Corresponding author. E-mail: qyuchuan@cau.edu.cn.

DGKELKXX (I/L) PDAL and C-terminal KYDP (Wanner et al., 2004).

Insect CSPs are small polypeptides that fold to yield a compact structure that is rich in α-helical domains. The α-helix is more flexible than the β-barrel. Therefore, the CSPs can exhibit a ligand-accommodation mechanism based on not only internal side-chain fluidity but also largely on protein backbone flexibility (Campanacci et al., 2003). On the basis of their binding ability with chemical molecules, the function of these proteins was proposed to involve in chemoreception (Steinbrecht, 1998; Campanacci et al., 2003). The conformational changes suggest that CSPs might be considered as a primary acceptor for chemical compounds (food or odours), and that the special conformation achieved upon ligand binding might be used to trigger receptor recognition and activation. Related studies have elucidated some physiological functions of insect CSPs, which are focused on perception and the delivery of chemical signals. Their specific localisation to the sensillum lymph and the capability to bind to a variety of chemicals leads to the hypothesis that OS-D-like proteins may be a new and different type of odorant-binding protein (OBP) (Angeli et al., 1999; Monteforti et al., 2002; Briand et al., 2002; Lartigue et al., 2002; Campanacci et al., 2003), another small, soluble protein concentrated in the sensillum lymph. Extensive study has revealed that CSPs can bind reversibly to N-phenyl-1-naphthylamine (Ban, 2002), 12-bromo-dodecanol (Campanacci et al., 2003), and medium-length fatty acid derivatives with low specificity (Briand et al., 2002; Lartigue et al., 2002). Furthermore, the CSPs are expected to have additional non-olfactory functions, as they are found also in none sensory tissues (Anholt and Mackay, 2001; Shyamala and Chopra, 1999).

Further researches on CSPs are significant to help elucidate the essence of insect behavior, and explore the new approach for pest management and develop new insect behavior regulators. The whitefly, *Bemisia tabaci* Gennadius (Hemiptera: Aleyrodidae), is an important agricultural insect pest causing crop damage by piercing the leaves, sucking sap and transmitting numerous plant viruses worldwide (Jones, 2003). *B. tabaci* is generally considered to be a species complex, which includes several genetically differentiated populations. Some populations have been labeled as different biotypes; among them, the biotypes B and Q are most invasive (Wan et al., 2009). The two biotypes share common characteristics, such as having broad feeding habits, higher survival rate, more spawning amount, strong resistance and transmitting numerous viruses. *B. tabaci* can transmit more than 100 viral diseases, which could lead to devastating losses to the affected crop (Wan et al., 2009). Research on *B. tabaci* and hostplant interaction were focused on identification of different host plant volatile substances and their functions, but studies of molecular mechanism of sensory perception of the whitefly were less.

To better explore the recognition mechanism of this insect to plant volatile or environmental substances, we studied the CSPs of the whitefly. In this study, a full-length sequence of *B. tabaci* CSP (BtabCSP) gene was cloned and the encoded protein was expressed successfully using a prokaryotic expression vector. The expressed protein was characterised using sequence alignment, phylogenetic relationships, gene splicing patterns and homology modelling using bioinformatics. Lastly, the relationship between the BtabCSP and the environmental suitability of the whitefly were discussed.

MATERIALS AND METHODS

The whiteflies

The whitefly *B. tabaci* population was collected from tomatoes, that is, *Lycopersicum esculentum*, at the Hebei Plant Protection Institute, Baoding, China. The plants were maintained at 22 to 24°C, 60 to 80% relative humidity, and exposed to a 16 h light: 8 h dark photoperiod. The whitefly used in this study belonged to the biotype Q, as determined using the RAPD-PCR method (Qiu et al., 2009).

Cloning of the full-length cDNA sequences

Total RNA was isolated from the adult whitefly using Trizol reagent according to the manufacturer's instructions (Invitrogen, China). The M-MLV Reverse transcription cDNA Synthesis Kit (Takara Biotechnology, Dalian, China) was used to generate cDNA from total RNA based on manufacturer's instructions.

Specific primers were designed based on the expressed sequence tags (EST) of *B. tabaci*, having high identity with the known CSPs. The BtabCSP primers are listed in Table 1. The cDNA encoding BtabCSP was amplified using TaKaRa LA Taq^{TM} according to the manufacturer's instructions. Briefly, 25 µl amplification reaction included 0.5 µl 5 U TaKaRa LA Taq^{TM} , 2.5 µl 10×Taq enzyme buffer, and 1 µl (10 µmol/l) of each primer. The touchdown PCR (TD-PCR) procedures consisted of a 94°C cycle denaturation for 3 min followed by 24 cycles at incrementally decreasing annealing temperature (62°C in the first cycle, which was decreased to 51°C in 1°C increments at every second cycle) and extension at 72°C for 1 min. The TD-PCR was followed by a conventional PCR, consisting 15 cycles at 94°C for 1 min, 50°C for 1 min and 72°C for 1 min with 10 min final extension at 72°C. The 5' and 3' ends of the BtabCSP cDNA were obtained using the rapid amplification of cDNA ends (RACE) procedure according to the instructions for TaKaRa LA Taq. Using BtabCSP gene specific primers, the amplification was performed under the following conditions: 94°C for 3 min, followed by 30 cycles of 30 s at 94°C, 30 s at 60°C and 30 s at 72°C, and a final extension for 10 min at 72°C. The 3'- and 5'- RACE PCR products were sequenced. The fragments were then assembled by TaKaRa Biotechnology Co., Ltd.

Phylogenetic analysis and homology modelling

The full-length cDNA sequences were analysed using online bioinformatics tools from NCBI (http://www.ncbi.nlm.nih.gov). Selected CSPs of Hemipteran insect with complete open reading frames (ORF) were obtained from GenBank (Table 2). Putative

Table 1. Primers used for cloning cDNA and genomic DNA sequences for *B. tabaci* CSP.

Specificity	Name	Orientation	Sequence (5'→3')
cDNA	Pr1	Forward	TGTCCTGGTCGGCTGTGC
	Pr2	Reverse	AACGGAGGATTTGTTCTGG
	Pr3	Forward	GACGAGGGACCCTGCACCAAC
	Pr4	Reverse	GGTCGTATTTCTGCTGGAGTT
	Pr5	Reverse	GCAAGCGGTTTTCAAAGCGTC
	Pr6	Forward	cgggatccATGCAGGTTTTGACTTTAGTTGT
	Pr7	Reverse	cccaagcttTTACAGGGTTTGTCCGAAGAGG
Genomic DNA	Pr8	Forward	ATCTTACGATTGTTAGTGACTTGTG
	Pr9	Reverse	TCTTTGGTTCATTTAGTTGAGGAT

Pr1 and Pr2 are BtabCSP gene specific primers designed for RT-PCR; Pr3, Pr4, and Pr5 are the specific primers designed for (5'/3')-RACE; Pr6 and Pr7, contains added restriction enzyme sites; *Bam* HI and *Hind* III, to facilitate the cloning, sequencing and expression of the entire *B. tabaci* CSP.

Table 2. Sequences sources of homology analysis and phylogenetic inference.

Code	Species	Accession number
AfabOS-D1	*Aphis fabae*	CAI64036.1
AfabOS-D2	*Aphis fabae*	CAG25440.1
AfabOS-D3	*Aphis fabae*	CAI64031.1
AgosCSP1	*Aphis gossypii*	ACJ64044.1
AgosCSP2	*Aphis gossypii*	ACJ64045.1
AgosCSP4	*Aphis gossypii*	ACJ64046.1
AlinCSP2	*Adelphocoris lineolatus*	ACZ58021.1
AlinCSP3	*Adelphocoris lineolatus*	ACZ58020.1
ApisCSP1	*Acyrthosiphon pisum*	BAH71609.1
ApisCSP2	*Acyrthosiphon pisum*	BAH71872.1
ApisCSP3	*Acyrthosiphon pisum*	BAH71897.1
ApisCSP4	*Acyrthosiphon pisum*	BAH72056.1
ApisCSP5	*Acyrthosiphon pisum*	NP_001119649.1
ApisCSP6	*Acyrthosiphon pisum*	NP_001119650.1
ApisCSP7	*Acyrthosiphon pisum*	NP_001119651.1
ApisCSP8	*Acyrthosiphon pisum*	NP_001119652.1
ApisCSP9	*Acyrthosiphon pisum*	NP_001128404.1
ApisCSP10	*Acyrthosiphon pisum*	NP_001156200.1
BtabCSP1	*Bemisia tabaci*	ADG56568.1
LeryCSP1	*Lipaphis erysimi*	ABU56011.1
MdirOS-D1	*Metopolophium dirhodum*	CAI64037.1
MdirOS-D2b	*Metopolophium dirhodum*	CAG25442.1
MdirOS-D2a	*Metopolophium dirhodum*	CAG25441.1
MdirOS-D3	*Metopolophium dirhodum*	CAI64032.1
MperCSP1	*Myzus persicae*	ACJ64047.1
MperCSP4	*Myzus persicae*	ACJ64048.1
MperCSP5	*Myzus persicae*	ACJ64049.1
MperOS-D1	*Myzus persicae*	CAI64038.1
MperOS-D2b	*Myzus persicae*	CAG25445.1
MperOS-D2a	*Myzus persicae*	CAG25444.1
MperOS-D3	*Myzus persicae*	CAI64033.1
MvicOS-D1	*Megoura viciae*	CAG25434.1

Table 2. Continues.

MvicOS-D2	*Megoura viciae*	CAG25435.1
MvicOS-D2b	*Megoura viciae*	CAG25436.1
MvicOS-D2c	*Megoura viciae*	CAG25437.1
MvicOS-D2d	*Megoura viciae*	CAG25438.1
MvicOS-D2e	*Megoura viciae*	CAG25439.1
MvicOS-D3	*Megoura viciae*	CAI64034.1
MvicOS-D4	*Megoura viciae*	CAI64039.1
NlugCSP3	*Nilaparvata lugens*	ACJ64050.1
NlugCSP4	*Nilaparvata lugens*	ACJ64051.1
NlugCSP5	*Nilaparvata lugens*	ACJ64052.1
NlugCSP6	*Nilaparvata lugens*	ACJ64053.1
NlugCSP8	*Nilaparvata lugens*	ACJ64054.1
NlugCSP9	*Nilaparvata lugens*	ACJ64055.1
NlugCSP10	*Nilaparvata lugens*	ADN06871.1
NlugCSP11	*Nilaparvata lugens*	ADN06872.1
NlugCSP12	*Nilaparvata lugens*	ADN06873.1
NribOS-D1	*Nasonovia ribis-nigri*	CAI64035.1
NribOS-D2	*Nasonovia ribis-nigri*	CAG25443.1
NribOS-D3	*Nasonovia ribis-nigri*	CAI64040.1

signal peptides were identified by comparison with mature CSPs reported in the literature. For analysis and calculation of protein sequence identities, proteins were truncated at point 25 residues upstream of the first conserved Cys residue to eliminate the highly variable signal peptides. All the truncated protein sequences were aligned using ClustalX 2.0 (Thompson et al., 1997). Phylogenetic analyses were performed using Mega V. 4 software (Kumar et al., 2001). Additional tree-building algorithms were based on 1000 bootstrap replicates.

The three-dimensional homology of BtabCSP was modelled using a moth CSP (MbraA6CSP, Protein Data Bank ID code 1KX8) with crystal structure as a template. The identity between the model and the raw sequence of BtabCSP was determined at 33.90%. The initial model was generated using DeepView/ Swiss-Pdb Viewer V 4.0 standalone software (Guex and Peitsch, 1997). The molecular 3-D images were generated using ViewerLite V .5.0 (Accelyrs Inc.).

Genomic DNA sequences and analyses of gene-splicing patterns

Genomic DNA was extracted from the whole adult whitefly using a salting-out procedure with minor modifications (Teng et al., 2009). Adult whiteflies were initially frozen in liquid nitrogen for 10 min and then homogenised on ice with TENS solution (0.05 mol/l Tris-HCl pH 8.0, 0.02 mol/l EDTA, 0.40 mol/l NaCl, 1% SDS). After the addition of 20 µg/µl proteinase K (4~6 µl) and incubation for 30 min at 50°C, 0.4 mol/l NaCl was added and the lysate was centrifuged at 14,000 r/min for 10 min. The DNA was precipitated from the supernatant using equal volume of 100% ethanol and collected through centrifugation at 14,000 rpm for 10 min. The pellet was then washed once with 70% ethanol and resuspended in ddH_2O.

Gene specific primers were used to amplify the CSP fragments from the genomic DNA of *B. tabaci* (Table 1). The amplification reactions were conducted using 1 µl of genomic DNA, 1 µl (10 µmol/l) of each specific primer, and 0.5 µl (5U) of High Fidelity Taq DNA polymerase, 2.5 µl 10× PCR buffer and the final volume was made up to 25 µl using ddH_2O. The PCR was performed with the following conditions: initial denaturation at 94°C for 3 min followed by 30 cycles of 30 s at 94°C, 30 s at 56°C, 30 s at 72°C, and a final extension for 10 min at 72°C. The PCR products were sequenced by Takara Biotechnology (Dalian) Co., Ltd.

Construction of expression vector

The full-length BtabCSP cDNA was amplified by PCR using the pr6 and pr7 specific primers (Table 1). A *Bam*HI restriction enzyme site was incorporated in the forward primers and a *Hind*III site was incorporated in the reverse primer. TD-PCR procedures (the same as those used earlier in the cloning of the full-length cDNA sequences) were used to amplify the full length gene. The coding region of the BtabCSP cDNA was cloned, in sense with a GST-tag and a His-tag at the N-terminus in an expression vector pET41a, digested with respective restriction enzymes. The fidelity of the inserts was confirmed by sequencing.

Protein expression

The plasmid was extracted and inserted into *Escherichia coli* BL21 (DE3) cells. Single-colony transformants harbouring the pET41a-BtabCSP expression vectors were cultured at 37°C and 200 rpm in a 10 ml LB medium containing 30 µg/ml kanamycin. The volume was increased to 50 ml using LB medium containing respective antibiotic. The culture was allowed to grow until OD_{600} was at ~0.80 and then shifted to 25°C; isopropyl thio-β-D-galactoside (IPTG) was added to a final concentration of 1 mM to induce protein expression. After induction for 8 h, the cells were harvested from liquid culture by centrifugation at 10,000 ×*g* for 10 min. The cell pellet was resuspended at room temperature using 5 ml of PBS by gentle vortexing. The cell suspension was incubated on a shaking platform at 80 rpm for about 15 min at room temperature. When the extract became clear, the insoluble cell

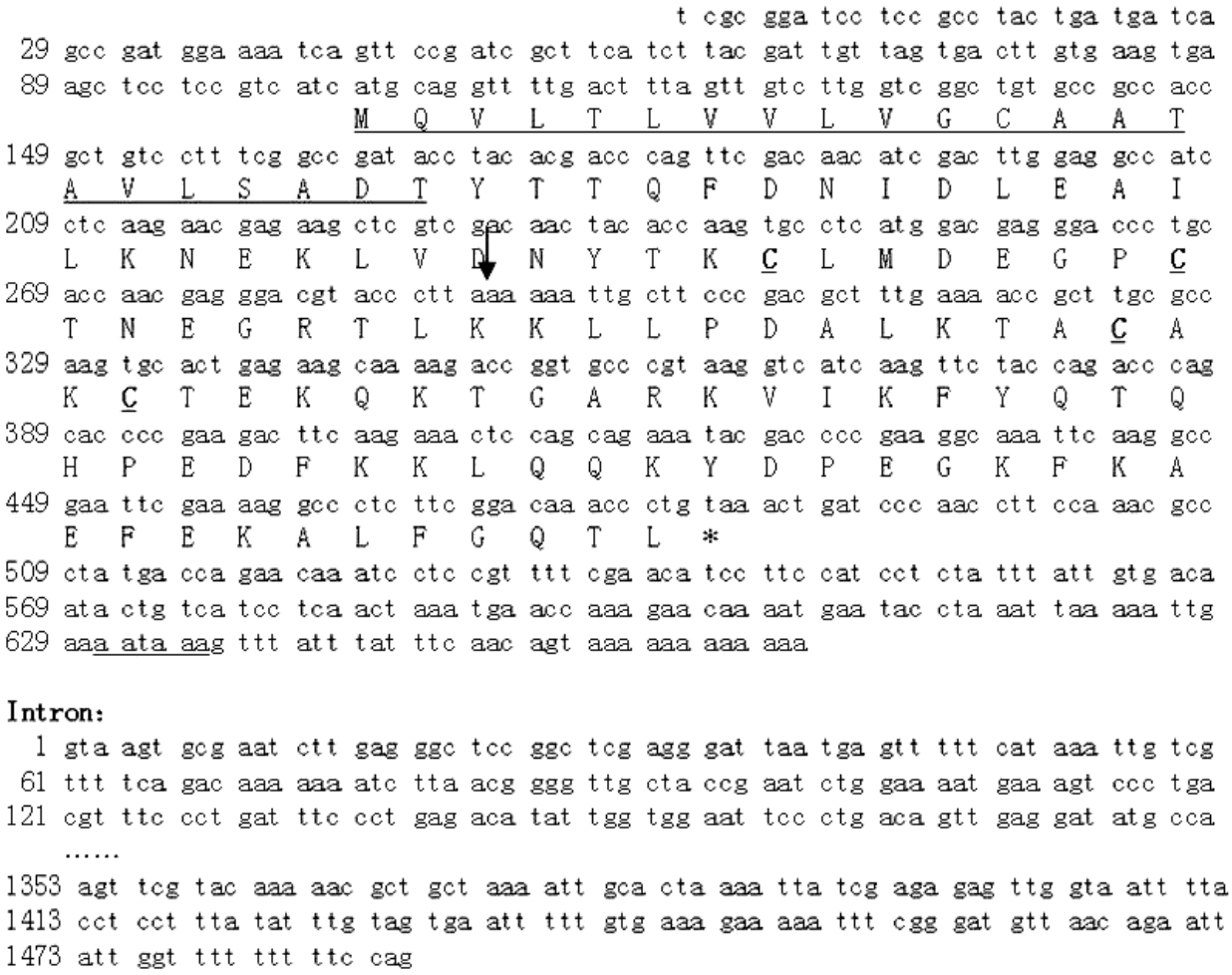

Figure 1. Nucleotide, inferred amino acid sequence of BtabCSP and the intron sequence. The intron splice site is marked with a "↓", located one nucleotide past the 62nd amino acid, a conserved lysine codon. Capital letters represent the deduced amino acid sequence. The conserved cysteine residues are bold and underlined. The signal peptides and the polyadenylation signals are underlined. The stop codon is marked with asterisks. The intron has the important GT-AG characteristic structure and is 1490 bp in length.

debris was separated from the supernatant by centrifugation at 16,000 ×*g* for 20 min. The insoluble pellet was resuspended at room temperature in 5 ml of PBS. The total fraction and the insoluble pellet were then analysed using SDS-PAGE.

RESULTS

The cDNA sequence analysis

The full-length cDNA sequence encoding BtabCSP was cloned from whole adult whiteflies and the 5' and 3' UTR sequences were generated using 5'- and 3'- RACE. The full length cDNA of BtabCSP was 667 bp in length, with an ORF of 381 bp. The full length sequence was deposited in GenBank with the accession number GU250808. The 3' end of BtabCSP contains polyA signals typical for eukaryotes (Max et al., 1985; Michael et al., 1990) and the AATAAA sequence included 17 bases at the 5' end of the GT portion that lead to a polyA stretch. The 5' UTR contained about 100 bases before the initial ATG codon. The BtabCSP cDNA encodes a predicted 126 amino acid protein with a calculated molecular mass of 14.17 kDa (http://www.expasy.org/cgi-bin/pi_tool). It was found that the first 22 amino acid residues have the characteristic features of a signal peptide common among the secretory proteins (von Heijne, 1986). Correspondingly, the mature protein was predicted to contain 104 amino acids and have a calculated molecular mass of 12.01 kDa. The deduced mature amino acid sequence had four conserved cysteine residues, which were also found in CSPs from other insect species (Wanner et al., 2004) (Figure 1).

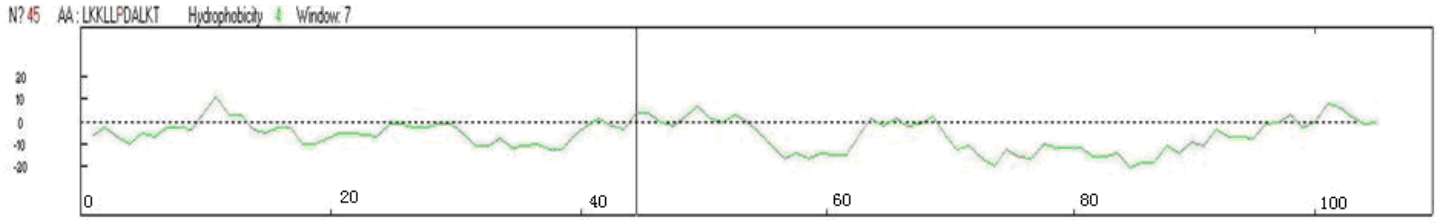

Figure 2. The hydrophobicity of BtabCSP mature protein. Analysis using Antheprot 4.3 software. Three distinct hydrophobic domains are seen between amino acid residues 10 and 13, 42 and 55, and 97 and 103, respectively.

Analysis using Antheprot 4.3 software showed that the secondary structure of BtabCSP was mainly composed of α-helices, similar to other CSPs from different insect species (Wanner et al., 2004). The deduced amino acid sequence for the mature BtabCSP was analysed for local hydrophobicity based on the procedure adapted from Kyte and Doolittle (1982). The results demonstrate a distinct hydrophobic domain at amino acid residues at 10 to 13, 42 to 55 and 97 to 103 that could bind to hydrophobic volatile compounds (Figure 2).

Sequence comparisons and phylogenetic analysis

Comparison analyses between the amino acid sequences of CSPs (Figure 3) indicate the presence of highly conserved sequence motifs. The conserved structure is related to the function of this protein. For example, the proteins had four conserved Cys positions as well as conserved sequence residues, including the N-terminal YTTK(Y/F)D(H/N)ID(V/I) DQ(I/V)L, the central EGXELKXLPDAL, and the C-terminal K(Y/W)DP. The aromatic residues at positions 26, 81, and 94 may have important functions since they are well conserved. The hydrophobic domain in BtabCSP identified between the amino acid residues at 10 to 13 and 42 to 55 were conserved in all CSPs, but the amino acid residues at 97 to 103 was conserved partially.

A neighbour-joining tree (Figure 4) was constructed to depict the hypothesised relationship among the CSPs of Hemipteran. It was clearly seen that the CSPs gene of this order have two more sub-families. The first sub-family included 14 proteins, such as BtabCSP1, NlugCSP4, ApisCSP9, MperOS-D2a, MvicOS-D2e, MvicOS-D2d. The second sub-family contained 30 proteins. The AlinCSP3, ApisCSP6, NlugCSP3, MperCSP1, NlugCSP11, NlugCSP8 and NlugCSP10, which differentiated from each other, could not be classified to any sub-family. It is interesting that among the nine CSPs from *Nilaparvata lugens*, two proteins were classified to the first sub-family, three to the second sub-family, and four proteins have much difference from each other. The same as *N. lugens*, the ten CSPs of *Acythosiphon pisum* had much difference also. The differences in structure of the same order may indicate the diversity of function of the CSPs.

Homology modelling

The moth *Mamestra brassicae* CSP crystal structure was used as a template to reconstruct the three-dimensional structure of BtabCSP. The identity between the model and the raw sequence of BtabCSP was determined at 33.90%. The major structural feature of the BtabCSP is its α-helices. Helices A and B as well as helices D and E form two V-shaped structures, whereas helix C was perpendicular and in between them. The C-terminal helix F is packed against the external face of the D-E helices and does not take part in core assembly. The conserved Tyr-26 blocks the tunnel opening and a narrow channel extends from the surface region to the core of the protein (Figure 5). The homology modelling results suggest a very good fit between the structural conformation of BtabCSP and the moth template molecule.

Gene sequences and splicing patterns of BtabCSP

To elucidate the evolutionary relationship between the BtabCSP gene and genes from other insect species, the CSP gene was cloned from the genomic DNA of *B. tabaci* and the gene structure was analysed. The BtabCSP gene had two exons on each end, and the intron is situated in the middle (1490 bp in length). The intron had characteristic GT-AG structure (Figure 1).

Most of the other CSP genes have a single (typically small) intron, of about 700 bp in length with several exceptions lacking introns, such as CSP genes from *Bombyx mori* (BmorCSP4 and BmorCSP 16), and *A. gambiae* (AgamSAP1 and Agameaa12591) and Drosophila (Dmelphk3). However, the introns of several CSPs are appreciably larger; for example, the intron in AgamEAA12702 is 19.6 Kb; those in BmorCSP2, BmorCSP9 and BmorCSP15 are approximately 2000 bp in length (Wanner et al., 2004; Gong et al., 2007). In this study, the intron related to the BtabCSP gene was relatively large with a length of 1490 bp.

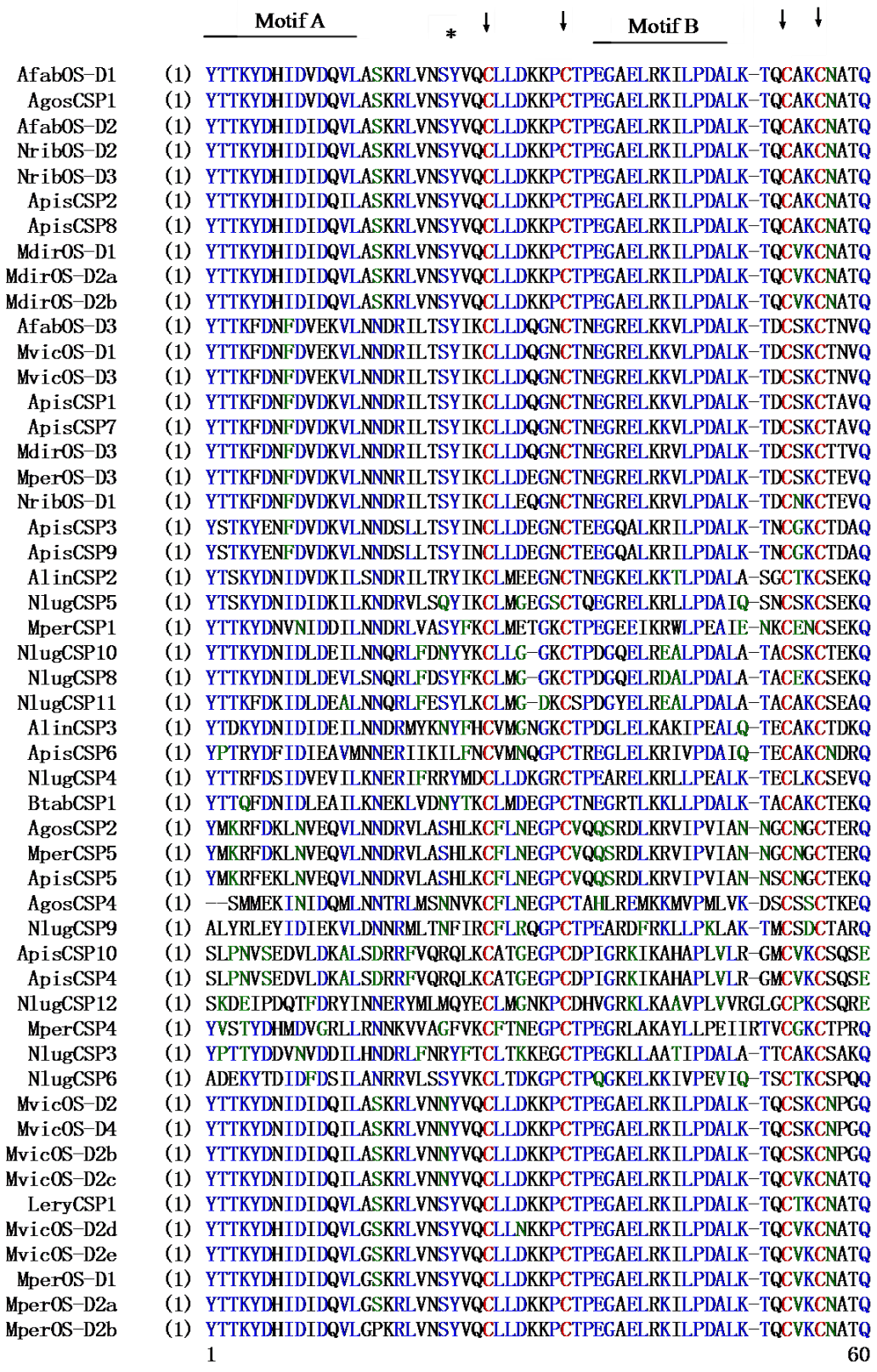

Figure 3. Alignment of the CSP family. Shaded regions indicate conserved motifs; aromatic residues at positions 26, 81, and 94 are marked with an "*"; the four conserved Cys residues are marked by an "↓". Motif A, Motif B, and Motif C are the conserved residues.

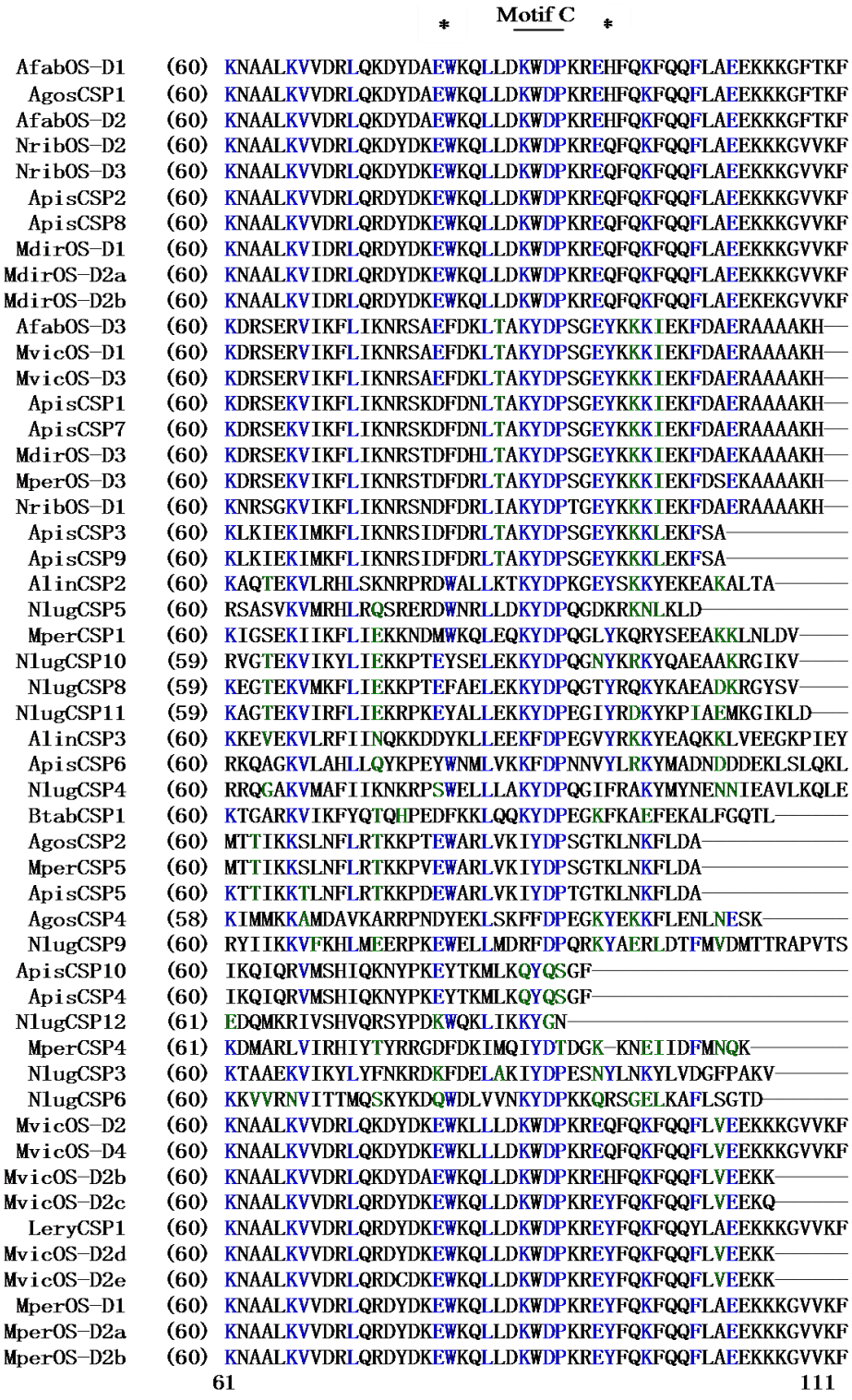

Figure 3. Contd.

The intron splice site was also conserved, which is a general characteristic of the CSP family. It is always located past a conserved lysine codon (Wanner et al., 2004; Gong et al., 2007). In this study, the intron splice site was located at one nucleotide past a conserved lysine codon, the 62nd amino acid.

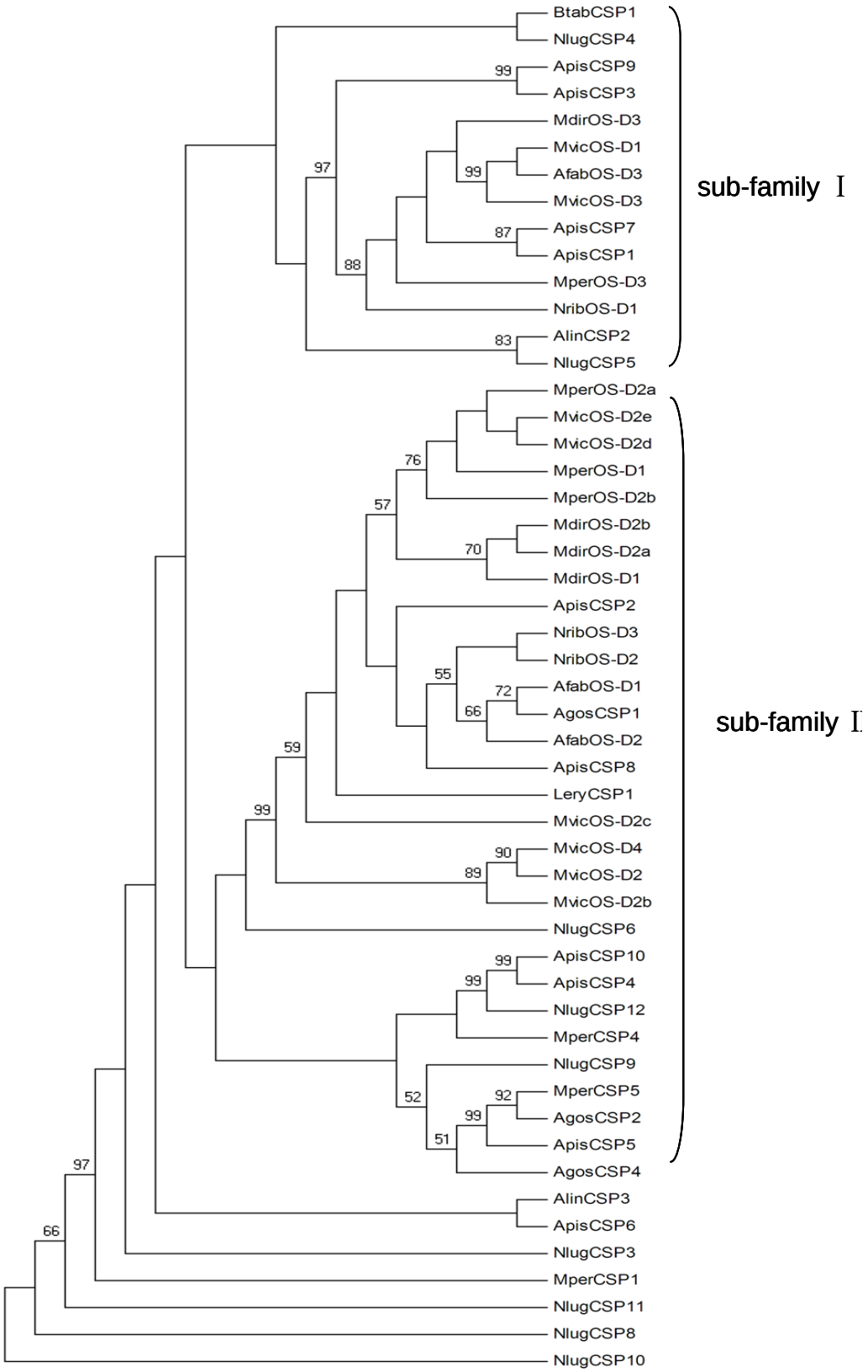

Figure 4. Neighbour-joining tree constructed in MEGA v.4 (bootstrap=1000 replications), using the amino acid sequences for CSPs from Hemipteran insects.

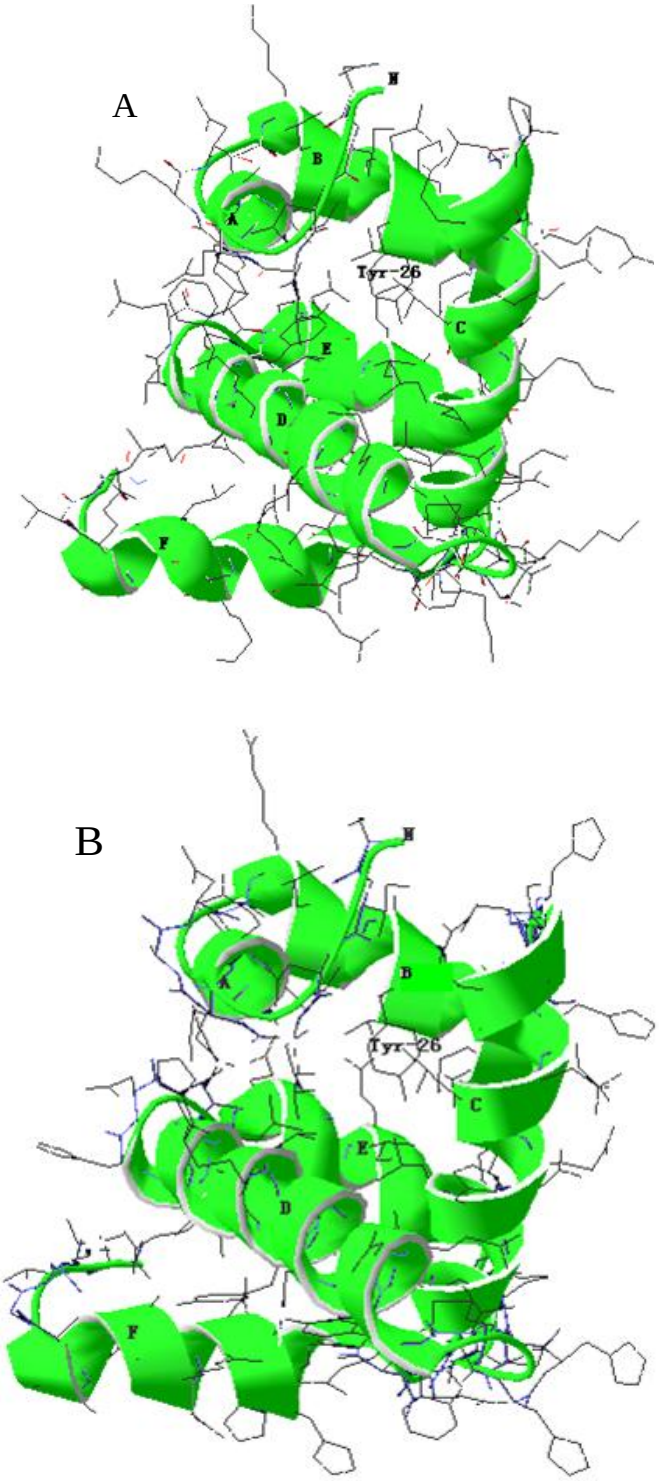

Figure 5. Three-dimensional BtabCSP model structure and the MbraCSPA6 crystal structure. The BtabCSP model structure (A) and the MbraCSPA6 crystal structure (B) both had 6 α-helices, as well as the trigger amino acid Tyr-26. Helices A and B as well as helices D and E form two V-shaped structures, whereas helix C is perpendicular and in between them. The C-terminal helix F was packed against the external face of the D-E helices and does not take part in core assembly.

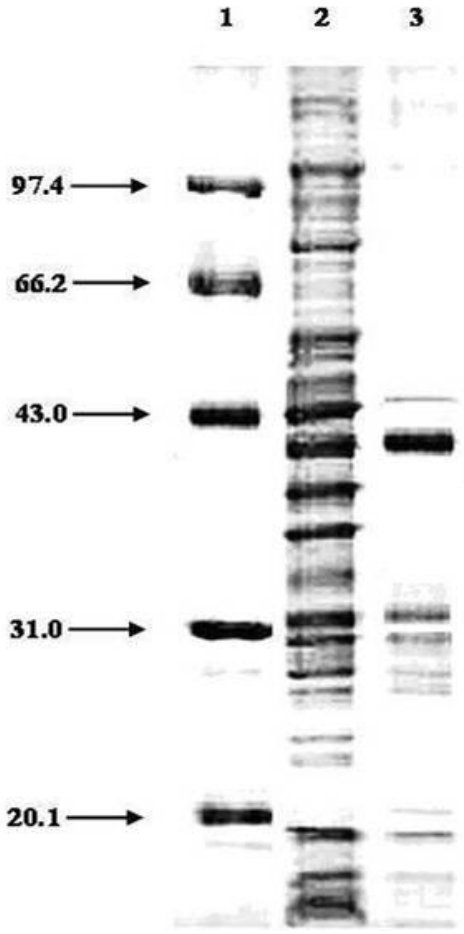

Figure 6. SDS-PAGE analysis of expressed product for pET-41a/BtabCSP in E.coli BL21 (DE3); lane 1, low molecular mass markers (Fajiete Biotechnology, Beijing, China); lane 2, total fraction of non-induced cells transformed by pET-41a/BtabCSP; lane 3, insoluble fraction of induced cells transformed by pET-41a/BtabCSP.

Recombinant expression

Using *E. coli* BL21(DE3) cells, BtabCSP was successfully expressed as a GST-tagged fusion protein. Significant amounts of insoluble recombinant BtabCSP proteins were obtained under optimised conditions. The recombinant protein was analysed using sodium dodecyl sulphate-polacrylamide gel electrophoresis (SDS-PAGE) (Figure 6). The SDS-PAGE analysis showed that the pET-41a/ BtabCSP (about 40kD) was successfully expressed in *E. coli* BL21(DE3) cells.

DISCUSSION

The RT-PCR and RACE technique were used to sequence the full-length cDNA encoding BtabCSP and was successfully cloned from the whole adult whitefly. The deduced amino acid sequence displayed the features allowing classification of the corresponding protein as a member of insect CSPs. The conserved characteristics include four conserved cysteine motifs ($CX_6CX_{16-19}CX_2C$, where X represents any amino acid) (Wanner et al., 2004). The genes in this family had conserved sequence residues in the N-terminal, central, and C-terminal regions (Figure 3). In addition, the aromatic residues at positions 26, 81, and 94 with possible functional importance were also highly conserved, along with residues at position 62.

The physiological functions of insect CSPs depend on their structures. Insect CSPs are small polypeptides and fold to yield a compact structure that is rich in α-helical domains. The α-helix is more flexible than the β-barrel. Therefore, the CSPs can exhibit a ligand-accommodation mechanism based on not only internal side-chain fluidity but also in large part on protein backbone flexibility. For example, the MbraCSPA6 exhibits a novel type of α-helix fold with six helices connected by α-α loops. A narrow channel extends within the protein hydrophobic core and it is able to bind such compounds with C12-18 alkyl chains (Lartigue et al., 2002). The Tyr-26 residue, which is located in the amino residue 23 in this paper, MbraCSPA6 seems to block tunnel opening; its rotation might very well be the trigger for the conformational change induced by the binding of the first ligand (Figure 5) (Campanacci et al., 2003). The results of this research indicate that the amino acid residues at 10 to 13 might not be the binding site with hydrophobic volatile compounds, as it was located at the gate of the narrow channel. The amino acid residues at 42 to 55 could well be the binding site, which is in the channel and have more hydrophobic amino acid than others. The amino acid residues at 97 to 100 might have specific physiological functions, because of the drastic difference from CSPs of this same order.

The conserved properties of the CSPs in different insect species play an important role in insect physiology. At the same time, the presence of different CSPs in the same insect and the wide tissue distribution and specificity indicates that CSP genes might have different functions (Gong et al., 2007; Anholt and Mackay, 2001; Shyamala and Chopra, 1999). Many researches on the non-olfactory function for CSPs were summarized by Wanner et al. (2004), such as the function of regulating circadian rhythms and tissue development, being immune responsive when challenged with virus and bacteria. Later studies indicate that CSPs may be involved in diverse cellular processes ranging from embryonic development to chemosensory signal transduction, and may function in cuticle synthesis (Foret et al., 2007).

The BtabCSP gene is the first reported CSP for the insect whitefly, *B. tabaci* supporting the view that these genes predate the divergence of the Neoptera (Wanner et al., 2004). This work provides the necessary background

for expression of these proteins, which is a prerequisite for structure-function research by means of X-ray crystallography and NMR, as well as for measuring and demonstrating the affinity of these proteins to several potential ligands.

The whitefly *B. tabaci* has a large diversity of host plants, including vegetable and field crops as well as woody plants. It is important that the volatile matter helps insects to find its' host plants (Du, 2001). It is also important to report the first CSP in *B. tabaci*. As other insects expressing multiple CSPs, there should be more CSPs in *B. tabaci*. However, it is difficult to obtain more sequences according to BLAST in EST database. Only the availability of the complete genome can provide definite information on the maximum number that can be expressed in a given species (Pelosi et al., 2006). But it is unknown from our result whether BtabCSP directly participates in transportation of odor molecules to the receptors. Therefore, the study of CSPs from whitefly will help to lay a foundation for better evaluation of the behaviour mechanism, elucidate the essence of its' behaviour and semiochemicals from surroundings and explore the new approaches for its' management. Future experiments will be required to explain the relationships among BtabCSP, odorant receptors, odor binding proteins and some related odorants of *B. tabaci*. We hope that the information presented here will be useful in future studies concerning the roles of BtabCSP in Hemiptera and other insect orders.

ACKNOWLEDGEMENTS

We thank Suting Zhong (China Agricultural University) for the assistance in this experiment. The research was funded by the National Sustaining Foundation of Science and Technology of China (No.2006BAD08A02) and The Professional Science and Technology Foundation (No.200803032) from The Ministry of Agriculture of China.

REFERENCES

Angeli S, Ceron F, Scaloni A, Monti M, Monteforti G, Minnocci A, Petacchi R, Pelosi P (1999). Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. Eur. J. Biochem. 262(3): 745-754.

Anholt RR, Mackay TF (2001). The genetic architecture of odor-guided behavior in *Drosophila melanogaster*. Behav. Genet. 31(1): 17-27.

Ban LP, Zhang L, Yan YH, Pelosi P (2002). Binding properties of a locust's chemosensory protein. Biochem. Biophys. Res. Commun. 293(1): 50-54.

Briand L, Swasdipan N, Nespoulous C, Bezirard V, Blon F, Huet JC, Ebert P, Pernollet JC (2002). Characterization of a chemosensory protein (ASP3c) from honeybee (*Apis mellifera* L.) as a brood pheromone carrier. Eur. J. Biochem. 269(18): 4586-4596.

Campanacci V, Lartigue A, Hallberg BM, Jones TA, Giudici-Orticoni MT, Tegoni M, Cambillau C (2003). Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. Proc. Natl. Acad. Sci. USA. 100(9): 5069-5074.

Du JW (2001). Plant-insect Chemical Communication and Its Behavior Control. Acta Phytophysiol. Sinica, 27(3): 193-200

Foret S, Wanner KW, Maleszka R (2007). Chemosensory proteins in the honey bee: Insights from the annotated genome, comparative analyses and expressional profiling. Insect Biochem. Mol. Biol. 37(1): 19-28.

Gong DP, Zhang HJ, Zhao P, Lin Y, Xia QY, Xiang ZH (2007). Indentification and expression pattern of the chemosensory protein gene family in the silkworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 37 (3): 266-277.

Guex N, Peitsch MC (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis, 18(15): 2714-2723.

Jones DR (2003). Plant viruses transmitted by whiteflies. Eur. J. Plant Pathol. 109 (3): 195-219.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics, 17(12): 1244-1245.

Kyte J, Doolittle RF (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157(1): 105-132.

Lartigue A, Campanacci V, Roussel A, Larsson AM, Jones TA, Tegoni M, Cambillau C (2002). X-ray structure and ligand binding study of a moth chemosensory protein. J. Biol. Chem. 277(35): 32094-32098.

Liu JX, Zhong GH, Xie JJ, Guan S, Hu MY (2005). Recent advances in chemosensory proteins of insect. Acta Entomol. Sinica, 48(3): 418-426.

Max LB, Meinrad B, Katharina S (1985). Transcription termination and 3' processing: the end is in site. Cell, 41(2): 349-359.

McKenna MP, Hekmat-Scafe DS, Gaines P, Carlson JR (1994). Putative Drosophila pheromone-binding proteins expressed in a subregion of the olfactory system. J. Biol. Chem. 269(23): 16340-16347.

Michael DS, Stephen CO, Marvin PW (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation *in vitro*. Nucleic Acids Res. 18(19): 5799-5805.

Monteforti G, Angeli S, Petacchi R, Minnocci A (2002). Ultrastructural characterization of antennal sensilla and immunocytochemical localization of a chemosensory protein in *Carausius morosus* Brunner (Phasmida: Phasmatidae). Arthropod. Struct. Dev. 30(3): 195-205.

Qiu BL, Chen YP, Liu L, Peng WL, Li XX, Ahdem MZ, Mathur V, Du YZ, Ren SX (2009). Identification of three major *Bemisia tabaci* biotypes in China based on morphological and DNA polymorphisms. Progress natural sci. 19(6): 713-718.

Pelosi P, Calvello M, Ban LP (2005). Diversity of odorant-binding proteins and chemosensory proteins in insects. Chem. Senses, 30(1): 291-292.

Pelosi P, Zhou JJ, Ban LP, Calvello M (2006). Soluble proteins in insect chemical communication. Cell. Mol. Life Sci. 63(14): 1658-1676.

Robertson HM, Martos R, Sears CR, Todres EZ, Walden KK, Nardi JB (1999). Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. Insect Mol. Biol. 8(4): 501-518.

Sabatier L, Jouanguy E, Dostert C, Zachary D, Dimarcq JL, Bulet P, Imler JL (2003). Pherokine-2 and -3. Eur. J. Biochem. 270(16): 3398-3407.

Shyamala BV, Chopra A (1999). *Drosophila melanogaster* chemosensory and muscle development: identification and properties of a novel allele of *scalloped* and of a new locus, SG18.1, in a Gal4 enhancer trap screen. J. Genet. 78(2): 87-97.

Steinbrecht RA (1998). Odorant-binding proteins: expression and function. Ann. N. Y. Acad. Sci. 855: 323-332.

Teng X, Wu Q, Wan FH (2009). A salting-out method for genomic DNA extraction from *Bemisia tabaci* (Gennadius). Biotechnol. Bullet. 8: 166-168.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The Clustal_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25(24): 4876-4882.

Vogt RG (2005). Molecular basis of pheromone detection in insects. In:

Gilbert LI, Iatro K, Gill S. editors. Comprehensive Insect Physiology, Biochemistry, Pharmacology and Molecular Biology. New York. Elsevier. 3: 753-804.
von Heijne G (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14(11): 4683-4690.
Wan FH, Guo JY, Zhang F (2009). Research on biological invasions in China. Beijing, China. Sci. Press, pp. 28-29
Wanner KW, Willis LG, Theilmann DA, Isman MB, Feng QL, Plettner E (2004). Analysis of the os-d-like gene family. J. Chem. Ecol. 30(5): 889-991.
Zhou JJ, Kan YC, Antoniw J, Pickett JA, Field LM (2006). Genome and EST analysis and expression of a gene family with putative functions in insect chemoreception. Chem. Senses, 31(5): 453-465.