FEBS 25983

FEBS Letters 517 (2002) 67–71

Novel cDNAs encoding salivary proteins from the malaria vector *Anopheles gambiae*

Alessandra Lanfrancotti[a], Fabrizio Lombardo[a], Federica Santolamazza[a], Massimiliano Veneri[a], Tiziana Castrignanò[b], Mario Coluzzi[a], Bruno Arcà[a,*]

[a] *Department of Public Health, Parasitology Unit, University 'La Sapienza', P.le Aldo Moro 5, P.O. Box 6 – Roma 62, 00185 Rome, Italy*
[b] *CASPUR – Supercomputing Center for University and Research, c/o University 'La Sapienza', P.le Aldo Moro 5, 00185 Rome, Italy*

Received 24 January 2002; revised 6 March 2002; accepted 7 March 2002

First published online 22 March 2002

Edited by Takashi Gojobori

Abstract Several genes encoding salivary components of the mosquito *Anopheles gambiae* were identified using a selective trapping approach. Among these, five corresponded to genes expressed specifically in female glands and their role may possibly be linked to blood-feeding. Our collection included a fourth member of the D7 protein family and two polypep tides that showed weak similarity to anti-coagulants from distantly related species. Moreover, we identified two additional members of a novel group of proteins that we named glandins. The isolation of tissue-specific genes represents a first step toward a deeper molecular analysis of mosquito salivary secretions.

Key words: Mosquito; Salivary gland; Saliva; Hematophagy

1. Introduction

Several parasitic and viral diseases that represent a severe threat to human health, especially in developing countries, are transmitted through the bites of arthropod vectors. The African mosquito *Anopheles gambiae* is the most important vector of human malaria, a disease that is still a global public health problem with devastating consequences. After ingestion by the mosquito with an infected blood meal, the malaria parasite *Plasmodium* undergoes, within the insect midgut, complex developmental processes leading to maturation of the asexual infective stage, the sporozoite. Once released into the hemolymph, sporozoites infect the salivary glands and can be transmitted to the vertebrate host with the saliva during the next blood meal [1]. Recognition and invasion of the glands is a species-specific process that is essential for transmission and requires interactions between ligand(s) on the parasite surface and receptor(s), yet to be identified, on the basal lamina and/or plasma membrane of gland cells [2,3].

Mosquito salivary glands produce a wide array of secreted compounds that are delivered with the saliva and help blood-feeding by affecting the host hemostatic response. Anti-hemostatic activities, such as platelet inhibitors, vasodilators and anti-coagulants, are the better-studied components in the saliva of hematophagous arthropods [4,5]. However, immunomodulatory factors are also present and possibly account for an enhancement of pathogen transmission [6]. In the mosquitoes *Aedes aegypti* and *A. gambiae*, female salivary glands show morphological and biochemical differentiation with genes involved in blood-feeding being expressed primarily in distal–lateral lobes, the same region preferentially invaded by sporozoites [7,8].

Molecular studies of the salivary glands may enable a more detailed understanding of vector–parasite–vertebrate interactions and also contribute to the design of malaria transmission-blocking strategies that involve the engineering of genetically modified mosquitoes. Practical implications might be the identification of parasite receptor(s) or the isolation of control elements suitable for driving the expression of anti-parasitic effector proteins in transgenic mosquitoes [9,10]. Moreover, as recently shown for the sandfly *Phlebotomus papatasi*, specific salivary antigens may be useful targets for the development of vaccines protecting the vertebrate host from parasite infection [11].

We previously reported the identification of several *A. gambiae* salivary gland genes using the signal sequence trap (SST) technique [12–14]. Here, we describe a further round of SST screening leading to the identification of 17 additional cDNA fragments.

2. Materials and methods

2.1. Mosquito colony and in situ hybridization

The *A. gambiae* used in this study was the GASUA reference strain (*Xag*, *2R*, *2La*, *3R*, *3L*) maintained under standard conditions. Salivary gland dissection and in situ hybridization to polytene chromosomes were performed as described [12].

2.2. Subtractive hybridization and immuno-screening in COS-7 cells

If not otherwise specified, experimental procedures followed the protocols in [15,16]. An aliquot of a 5′-end enriched *A. gambiae* salivary gland cDNA library [12] was used to transform XL1-blue supercompetent cells (Stratagene). Colonies were transferred to nylon membranes and hybridized, under stringent conditions, to a mixture of radiolabelled probes corresponding to the 15 cDNAs previously isolated in our laboratory from the same library. Approximately 1700 non-hybridizing colonies were arrayed on agar plates in 47 pools with 36 clones arranged in a 6×6 matrix. Transfection and immuno-staining of COS-7 cells were performed as previously described [12,17]. Positive pools were divided in 12 smaller pools and rescreened for the identification of single positive clones.

2.3. Expression analysis

Total RNA was extracted using Trizol reagent (Gibco BRL). Approximately 100 ng of DNase-treated total RNA (DNase I RNase-

*Corresponding author. Fax: (39)-06-4991 4644.
E-mail address: b.arca@caspur.it (B. Arcà).

0014-5793/02/$22.00
PII: S0014-5793(02)02578-4

free, Boehringer Mannheim) was used as template for the reverse transcription-polymerase chain reaction (RT-PCR) amplification with the Superscript one-step RT-PCR system (Gibco BRL). First strand cDNA synthesis (50°C for 30 min) and heat inactivation of the reverse transcriptase (2 min at 94°C) were followed by 35 cycles of amplification (94°C for 30 s, 55°C for 30 s and 72°C for 45 s). Twenty-six cycles were used for the amplification of the actin mRNA to keep the reaction below saturation levels. Primer sequences are available on request.

2.4. Cloning of full-length cDNAs

Full-length cDNAs were isolated by screening a thoracic cDNA library [12]. Exceptions were the cDNAs encoded by *gSG6* and *Ag9*, which were obtained by the 3′-rapid amplification of cDNA ends (RACE) system according to the manufacturer's instructions (Gibco BRL). Thirty-five cycles of amplification (94°C for 30 s, 55°C for 30 s, 72°C for 45 or 90 s) with *gSG6*- (5′-ATTCGCTCCTGCCTGTGTCT-3′) or *Ag9*-specific (5′-TAAGACCTCTTCCAAGCAGTCG-3′) primers were used. The amplification products of ~600 bp and ~1600 bp were cloned into the pBCKS(+) vector (Stratagene). Sequences were obtained commercially (MWG-Biotech AG).

2.5. Signal peptide prediction and sequence analysis

Signal peptide prediction analysis was performed by the SIGNALP program [18]. Sequence comparison, database searches and multiple alignment were done by the Wisconsin Package Version 10.2 (Genetics Computer Group, Madison, WI, USA), the BLAST and CLUSTAL W programs [19,20].

3. Results and discussion

3.1. Isolation of cDNAs by the SST

Before embarking on a new round of SST screening, we used subtractive hybridization to remove clones already identified from the salivary gland cDNA library [12]. Then, 47 pools, corresponding to approximately 1700 clones, were screened in COS-7 cells by transfection and immuno-staining with an anti-Tac monoclonal antibody [17]. Twenty-two positive pools were identified and, after rescreening, 34 positive clones, 25 of which represented novel sequences, were isolated. Among these, 17 cDNA fragments were unique and ranged in size from 236 to 551 bp (Table 1). Signal peptide prediction analysis indicated that most cDNA fragments potentially include amino-terminal signal peptides, suggesting that they may encode secreted or membrane-anchored molecules. The only exceptions were clones 1 and 4, which may represent artefacts of the screening procedure.

3.2. Tissue and developmental expression profile

To evaluate the tissue-specificity of expression, gene-specific primers were used to amplify the corresponding mRNAs by RT-PCR. Total RNAs extracted from female salivary glands, carcasses (adult females devoid of the salivary glands) and adult males were used as templates. The 15 cDNAs could be grouped in four main categories (Fig. 1, left panel and Table 1): (A) female salivary gland-specific clones (51.5F, 7.6D, 12.5D, 7 and 11.8); (B) cDNAs whose expression is enriched in female glands (5.5F); (C) clones expressed in both female glands and (whole) adult males and, therefore, presumably transcribed both in female and in male glands (5, 1.1B, 10 and 14); (D) cDNA clones with a ubiquitous expression profile (a3.6B, e1.6B, 8 and 9). Clone 6, showing a distinctive pattern, defined a category per se. To confirm and extend the tissue-specific expression analysis we also assayed total RNA from different developmental stages. It should be emphasized that adult salivary glands originate from a bud of imaginal cells located at the anterior end of the larval salivary glands, completing development only after emergence [21]. As shown on the right panel of Fig. 1 there is a general agreement between tissue-specific and developmental expression analysis. cDNAs expected to be gland-specific (groups A and C) showed an expression profile restricted to adults and, occasionally, to pupal stages. The pupal expression of gland-specific genes may be explained by early expression in the developing adult glands. Clones in group B showed a higher transcriptional activity in adult females while clones in group D again exhibited ubiquitous expression profiles. Finally, clone 6 did not share any of the above-described

Table 1
Properties of the cDNAs isolated by SST

Clone	Accession no.[a]	SP[b]	Size[c]	Similarity[d]	Division[e]	RT-PCR[f]	Gene[g]
1.1B	AJ302654	Y	392 (807)	gSG2	2L25A	sg, m	*gSG2-like*
5.5F	AJ297932	Y	326	gSG1		sg+, m	*gSG1a*
51.5F	AJ302655	Y	551 (1184)	Apo A-I	2R18D	sg	*gSG5*
7.6D	AJ302656	Y	410 (1225)	gSG1		sg	*gSG1b*
12.5D	AJ302657	Y	407 (565)	AcAps	3R34C	sg	*gSG6*
a3.6B	AJ297930	Y	485			sg, c, m	
e1.6B	U28809	Y	343	Lysozyme	2L27	sg, c, m	
1	AJ419878	N	506	Sec61	3R33C	nd	*AgSec61*
4	AJ419879	N	236			nd	
5	AJ302658	Y	315 (550)	Snake venom PLA_2	3R30A	sg, m	*gSG7*
6	AJ297931	Y	533		2L24D	sg, c+, m+	
7	AJ302659	Y	438 (584)	D7-related	3R30B	sg	*D7r4*
8	AJ419880	Y	354			sg, c, m	
9	AJ297933	Y	490 (1824)	CG5276 clone	2R7A	sg, c, m	*Ag9*
10	AJ302660	Y	418 (642)	gSG9	3L44C	sg, m	*gSG10*
11.8	AJ302661	Y	524 (1096)		3L39C	sg	*gSG8*
14	AJ302662	Y	257	gSG10	2R18A	sg, m	*gSG9*

[a]European Molecular Biology Laboratory accession numbers.
[b]Presence (Y) or absence (N) of a signal peptide (SP).
[c]Length of the partial cDNA fragments, and of the full-length if available (in parentheses), expressed in bp.
[d]Similarities to known proteins deposited in the databases.
[e]Locations on polytene chromosomes as obtained by in situ hybridization.
[f]Summary of the expression analysis as determined by RT-PCR. sg, female salivary glands; c, carcasses (adult females with salivary glands removed); m, adult males; nd, not done. The + sign indicates an enriched expression.
[g]Gene names, if assigned.

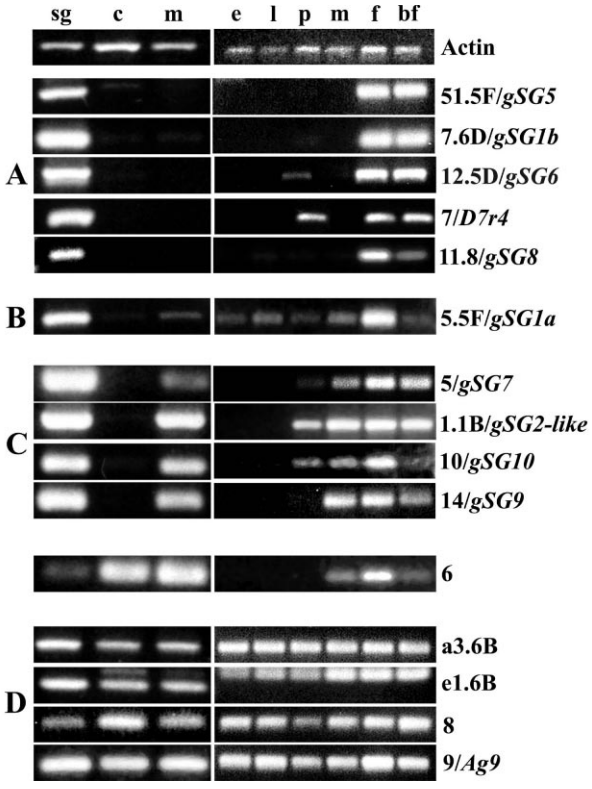

Fig. 1. RT-PCR expression analysis of the clones identified by the SST. Gene-specific primers were used to amplify total RNA from different tissues and developmental stages. sg: adult female salivary glands; c: carcasses (adult females with salivary glands removed); m: adult males; e: 0–48 h embryos; l: first–fourth instar larvae; p: pupae; m: 0–3-day-old adult males; f: 0–3-day-old adult females; bf: adult females 24 h after blood-feeding. A–D refer to classes according to expression profile (see text). Actin mRNA was amplified as control. The different clones and the corresponding gene names are indicated on the right.

expression pattern; it represents the only gene that is transcribed at a reduced level in the female salivary glands as compared to the other tissues and whose expression is restricted to adult stages. In summary, the analysis suggests that we have identified five novel female salivary gland-specific genes, whereas another four are presumably expressed both in female and in male glands.

3.3. Similarity searches and full-length cDNA cloning

Only a few of the isolated clones were similar to known proteins. One of these, e1.6B, was the previously characterized *A. gambiae* lysozyme, whereas clone 1 was the *A. gambiae* homologue of the alpha subunit of Sec61, a ubiquitous component of the protein translocation apparatus. Moreover, clone 7 showed high similarity to three previously identified *A. gambiae* D7-related genes and the corresponding gene was therefore named *D7r4* (Table 1). The properties of *D7r4* and of the other members of this gene family have been reported elsewhere and will not be discussed here [14].

We subsequently cloned, either by screening a thoracic cDNA library or by 3′-RACE [22], most of the full-length cDNAs belonging to classes A, B and C. The size of the full-length transcripts, the similarities obtained by database searches and the localization on polytene chromosomes are shown in Table 1. With one exception, *D7r4*, genes specifically expressed or enriched in the glands were named *gSG* (*gambiae* salivary gene) followed by a progressive number; the corresponding protein products will be accordingly indicated as gSG*n*.

The cDNAs 1.1B, 5.5F and 7.6D showed similarity to the *A. gambiae* salivary gland-specific genes *gSG1* and *gSG2* and the corresponding genes were therefore named respectively *gSG2-like1*, *gSG1a* and *gSG1b* (Table 1). Consequently, five genes encoding gSG1-like proteins are present in the genome of *A. gambiae*: two are female salivary gland-specific (*gSG1* and *gSG1b*), one seems to be ubiquitous (*bD3*) and the expression of the other two appears enriched in female glands (*dF2* and *gSG1a*) [12]. No significant similarities to known proteins could be detected by database searches hence their function is still unknown. In light of their preferential expression in female glands they may somehow be involved in blood-feeding; we propose to refer to this novel group of proteins as glandins. Fig. 2 shows an alignment of the three putative glandins for which complete sequence information is available (gSG1, dF2, gSG1b). Percentages of identity (20–25%) and similarity (approximately 50%) are not very high but the overall conservation suggests that they may be divergent members of the same protein family.

gSG10 contains the eukaryotic aspartyl protease active site signature and is similar to the polypeptide encoded by *gSG9* ($P = e{-}7$). Moreover, the female gland-specific gSG5 showed similarity in the carboxy-terminus to several apolipoprotein A-I precursors ($P = 0.00055$). The significance of these similarities remains to be clarified. However, since *gSG9* and *gSG10* are expressed in both sexes we may speculate that they could encode proteolytic enzymes, perhaps involved in digestive processes.

The cDNAs 12.5D and 5, whose corresponding genes were

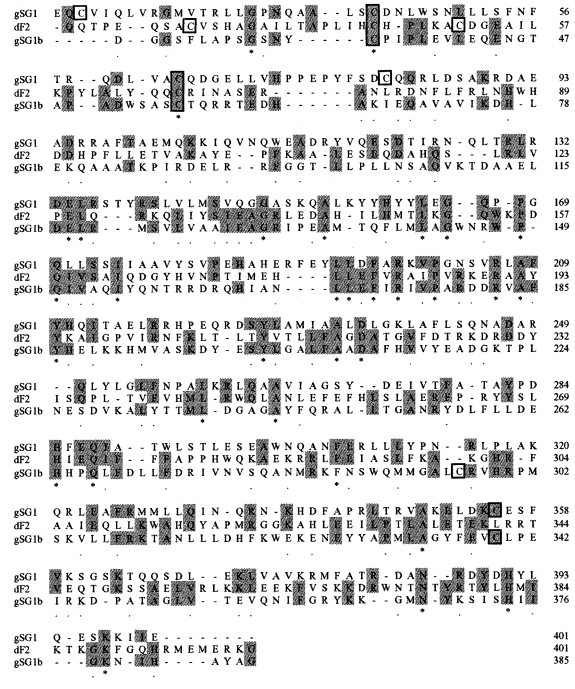

Fig. 2. Sequence alignment of the three *A. gambiae* glandins gSG1, dF2 and gSG1b. Asterisks mark identities in all of the sequences; dots identify conserved substitutions. Residues conserved in at least two of the aligned sequences are shaded. All cysteines are boxed. Secretory signal peptides were not included.

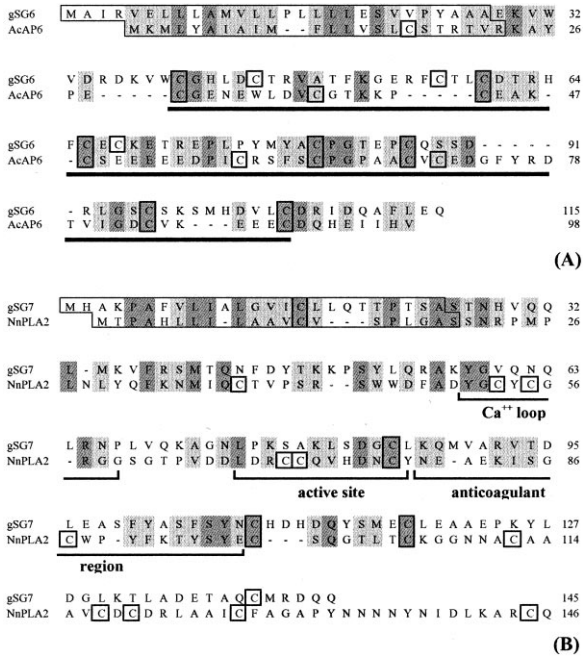

Fig. 3. Alignment of the putative peptides encoded by *gSG6* (A) and *gSG7* (B) to the *A. caninum* anti-coagulant protein 6 (AcAP6) and to the PLA_2 from the venom of the cobra, *Naja naja atra* ($NnPLA_2$). Putative signal peptides and cysteine residues are boxed. Identities and similarities are shaded respectively in dark and in light gray. The tick line under the AcAP6 sequence marks the TIL domain. The calcium loop, the active site and the anti-coagulant region of the PLA_2 from *Naja naja atra* are indicated.

named *gSG6* and *gSG7*, showed weak but potentially meaningful similarity to anti-coagulant proteins from the hematophagous nematode, *Ancylostoma caninum*, and from the venom of the Chinese cobra, *Naja naja atra*, respectively [23,24]. The hookworm *A. caninum* infects a wide range of mammalian hosts and obtains its blood meal from the lacerated vessels of the small intestine. To prevent host hemostasis this parasitic nematode secretes a family of small proteins named AcAPs (*A. caninum* anti-coagulant proteins) that inhibit the activity of the blood coagulation factor Xa (FXa) and the complex between factor VIIa and tissue factor (FVIIa/TF) [23]. It should be emphasized here that the AcAPs carry a TIL domain (trypsin inhibitor-like cysteine rich domain, amino acids 29–90 in AcAP6, Fig. 3A), i.e. 10 cysteine residues found in several serine proteinase inhibitors, and that coagulation is typically achieved through a cascade of activation of serine proteinases. The gSG6-deduced protein shows the highest level of similarity with AcAP6 (24% identity, 65% similarity). The high conservation of both the number and the position of cysteine residues and the overall similar size suggest that gSG6 may share a similar tridimensional structure with these anti-coagulant proteins (Fig. 3A). The putative *gSG7* gene product was weakly similar (19% identity and 45% similarity) to a secreted phospholipase A_2 (PLA_2) from the venom of the Chinese cobra (Fig. 3B). gSG7 lacks most of the typical features of members of the PLA_2 family [25]. However, the similarity to this enzyme, albeit weak, is interesting in view of the FXa-directed anti-coagulant activity of the snake protein, which is dissociated from the enzymatic activity [26] and maps to a region that is distinct from both catalytic and calcium-binding sites (Fig. 3B) [27]. We should point out that previous studies indicated that salivary extracts from anopheline mosquitoes contain potent thrombin-directed anti-coagulants but no inhibitors of FXa [28]. Indeed, the salivary glands of *A. gambiae* may produce a thrombin inhibitor related in sequence to the anophelin of *Anopheles albimanus* [29]. The similarity of the hypothetical gSG6 and gSG7 proteins to anti-coagulants from evolutionarily distant species does not necessarily imply any functional homology, however the coincidence should be noted.

Finally, the putative protein encoded by clone 9 was similar to a family of polypeptides widely conserved among evolutionarily distant organisms from *Drosophila* to humans. Most of these proteins have been identified by different genome sequencing projects and their role is as yet unknown; however, they are similar to the recently characterized *Cimex* family of apyrases isolated from the saliva of the bed bug, *Cimex lectularius*, and of the sandflies, *P. papatasi* and *Lutzomyia longipalpis* [30]. These proteins display apyrase activity and their function, as in other hematophagous arthropods, is to inhibit the ADP-induced platelet aggregation. We cannot rule out the possibility that the *Ag9* gene product plays some role in connection to blood-feeding. However, its ubiquitous expression profile and the higher degree of similarity with CG5276 from *Drosophila melanogaster* (53% identity, 68% similarity, $P = e-109$) rather than with the apyrases of the sandflies (49% identity, 65% similarity, $P = e-72$) support the idea that it may have housekeeping functions. This hypothesis is reinforced by the observation that a female-specific putative apyrase with similarity to 5′-nucleotidases has already been identified from the salivary glands of *A. gambiae* [13].

In conclusion, as a preliminary step toward a detailed molecular and biochemical analysis of the salivary glands of the malaria mosquito *A. gambiae* we have identified several novel salivary genes. A striking feature is that the majority of these genes seem to encode 'unknown' functions and, therefore, we can only hypothesize their possible involvement in blood uptake and digestion. This observation emphasizes the complexity of the salivary secretions of hematophagous insects, suggesting that several pharmacological activities still remain to be identified. Purification from salivary extracts and/or in vitro expression of recombinant proteins should allow further biochemical analyses that will reasonably lead to a better understanding of the physiological functions of mosquito salivary glands.

Acknowledgements: We would like to thank C. Louis for critical reading of the manuscript; F. Spano, G. Luoni and T. Nolan for the useful suggestions during the final revision; G. Pietrangeli and M. Calzetta for technical assistance with the mosquito colony. This work was supported by Grants from the European Union (HPRN-CT-2000-00080), from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), from the Foundation Compagnia di San Paolo (Turin, Italy) and partly from MIUR/COFIN funds.

References

[1] Sinden, R.E. (1999) Parassitologia 41, 139–148.
[2] Rosenberg, R. (1985) Am. J. Trop. Med. Hyg. 34, 687–691.
[3] Barreau, C., Touray, M., Pimenta, P.F., Miller, L.H. and Vernick, K.D. (1995) Exp. Parasitol. 81, 332–343.
[4] Ribeiro, J.M. (1995) Infect. Agents Dis. 4, 143–152.
[5] Stark, K.R. and James, A.A. (1996) in: The Biology of Disease

Vectors (Beaty, B.J. and Marquardt, W.C., Eds.), pp. 333–348, University Press of Colorado, Niwot, CO.
[6] Titus, R.G. and Ribeiro, J.M. (1988) Science 239, 1306–1308.
[7] James, A.A. (1994) Bull. Inst. Pasteur 92, 133–150.
[8] Arcà, B. et al. (1999) Parassitologia 41, 483–487.
[9] James, A.A., Beerntsen, B.T., de Lara Capurro, M., Coates, C.J., Coleman, J., Jasinskiene, N. and Krettli, A.U. (1999) Parassitologia 41, 461–471.
[10] Collins, F.H., Kamau, L., Ranson, H.A. and Vulule, J.M. (2000) Bull. WHO 78, 1412–1423.
[11] Valenzuela, J.G., Belkaid, Y., Garfield, M.K., Mendez, S., Kamhawi, S., Rowton, E.D., Sacks, D.L. and Ribeiro, J.M. (2001) J. Exp. Med. 194, 331–342.
[12] Arcà, B., Lombardo, F., de Lara Capurro, M., della Torre, A., Dimopoulos, G., James, A.A. and Coluzzi, M. (1999) Proc. Natl. Acad. Sci. USA 96, 1516–1521.
[13] Lombardo, F., Di Cristina, M., Spanos, L., Louis, C., Coluzzi, M. and Arcà, B. (2000) J. Biol. Chem. 275, 23861–23868.
[14] Arcà, B., Lombardo, F., Lanfrancotti, A., Spanos, L., Veneri, M., Louis, C. and Coluzzi, M. (2002) Insect. Mol. Biol. 11, 47–55.
[15] Ausubel, F.M., Brent, R., Kingston, R.F., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1991) John Wiley and Sons, New York.
[16] Sambrook, J., Fritsh, E.F. and Maniatis, T. (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY.
[17] Tashiro, K., Hideaki, T., Heilker, R., Shirozu, M., Nakano, T. and Honjo, T. (1993) Science 261, 600–603.
[18] Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Protein Eng. 10, 1–6.
[19] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[20] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[21] Clements, A.N. (1992) in: The Biology of Mosquitoes, pp. 251–262, Chapman and Hall, London.
[22] Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA 85, 8998–9002.
[23] Stassens, P. et al. (1996) Proc. Natl. Acad. Sci. USA 93, 2149–2154.
[24] Pan, F.M., Chang, W.C. and Chiou, S.H. (1994) Biochem. Mol. Biol. Int. 33, 187–194.
[25] Six, D.A. and Dennis, E.A. (2000) Biochim. Biophys. Acta 1488, 1–19.
[26] Kerns, R.T., Kini, R.M., Stefansson, S. and Evans, H.J. (1999) Arch. Biochem. Biophys. 369, 107–113.
[27] Kini, R.M. and Evans, H.J. (1987) J. Biol. Chem. 262, 14402–14407.
[28] Stark, K.R. and James, A.A. (1996) J. Med. Entomol. 33, 645–650.
[29] Valenzuela, J.G., Francischetti, I.M. and Ribeiro, J.M. (1999) Biochemistry 38, 11209–11215.
[30] Valenzuela, J.G., Belkaid, Y., Rowton, E. and Ribeiro, J.M. (2001) J. Exp. Biol. 204, 229–237.