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

Alessandra Lanfrancotti\(^a\), Fabrizio Lombardo\(^a\), Federica Santolamazza\(^a\), Massimiliano Veneri\(^a\), Tiziana Castignano\(^b\), Mario Coluzzi\(^a\), Bruno Arcà\(^a\),*\(^\text{\textsuperscript{\textast}}\)

\(^a\)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 polypeptides 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. © 2002 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

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 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 [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 papatasii*, 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 (*Xq*, 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 immunostaining 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 © 2002 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

PII: S 0 0 1 4 - 5 7 9 3 ( 0 2 ) 0 2 5 7 8 - 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 s or 90 s) with gSG6- (5’-ATTC-GCTTCTGCTGTGCT-3’) or Ag9-specific (5’-TAAAGCCT-CTTCAAGCAGTCG-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.

Table 1: Properties of the cDNAs isolated by SST

<table>
<thead>
<tr>
<th>Clone</th>
<th>Accession no.</th>
<th>SP</th>
<th>Size</th>
<th>Similarity</th>
<th>Division</th>
<th>RT-PCR</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1B</td>
<td>AJ302654</td>
<td>Y</td>
<td>392</td>
<td>gSG2</td>
<td>2L25A</td>
<td>sg, m</td>
<td>gSG2-like</td>
</tr>
<tr>
<td>5.5F</td>
<td>AJ297932</td>
<td>Y</td>
<td>326</td>
<td>gSG1</td>
<td>2R18D</td>
<td>sg, m</td>
<td>gSG1a</td>
</tr>
<tr>
<td>51.5F</td>
<td>AJ302655</td>
<td>Y</td>
<td>551</td>
<td>Apo A-I</td>
<td>2R18D</td>
<td>sg</td>
<td>gSG1b</td>
</tr>
<tr>
<td>7.6D</td>
<td>AJ302656</td>
<td>Y</td>
<td>410</td>
<td>gSG1</td>
<td>3R34C</td>
<td>sg</td>
<td>gSG6</td>
</tr>
<tr>
<td>12.5D</td>
<td>AJ302657</td>
<td>Y</td>
<td>407</td>
<td>AcAps</td>
<td>3R34C</td>
<td>sg, c, m</td>
<td></td>
</tr>
<tr>
<td>a3.6B</td>
<td>AJ297930</td>
<td>Y</td>
<td>485</td>
<td>Lysozyme</td>
<td>2L27</td>
<td>sg, c, m</td>
<td></td>
</tr>
<tr>
<td>e1.6B</td>
<td>U28809</td>
<td>Y</td>
<td>343</td>
<td>Sec61</td>
<td>3R33C</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>AJ441978</td>
<td>N</td>
<td>506</td>
<td>D7-related</td>
<td></td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>AJ302658</td>
<td>Y</td>
<td>315</td>
<td>Snake venom PLA₂</td>
<td>3R30A</td>
<td>sg, m</td>
<td>gSG7</td>
</tr>
<tr>
<td>6</td>
<td>AJ297931</td>
<td>Y</td>
<td>533</td>
<td></td>
<td>2L24D</td>
<td>sg, c+, m+</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>AJ302659</td>
<td>Y</td>
<td>438</td>
<td>D7-related</td>
<td>3R30B</td>
<td>sg</td>
<td>D7r4</td>
</tr>
<tr>
<td>9</td>
<td>AJ297933</td>
<td>Y</td>
<td>354</td>
<td></td>
<td></td>
<td>sg, c, m</td>
<td>Ag9</td>
</tr>
<tr>
<td>10</td>
<td>AJ302660</td>
<td>Y</td>
<td>418</td>
<td>gSG9</td>
<td>3L44C</td>
<td>sg, m</td>
<td>gSG10</td>
</tr>
<tr>
<td>11.8</td>
<td>AJ302661</td>
<td>Y</td>
<td>524</td>
<td>gSG10</td>
<td>3L39C</td>
<td>sg</td>
<td>gSG8</td>
</tr>
<tr>
<td>14</td>
<td>AJ302662</td>
<td>Y</td>
<td>257</td>
<td></td>
<td>2R18A</td>
<td>sg, m</td>
<td>gSG9</td>
</tr>
</tbody>
</table>

*aEuropean Molecular Biology Laboratory accession numbers.

bPresence (Y) or absence (N) of a signal peptide (SP).

cLength of the partial cDNA fragments, and of the full-length if available (in parentheses), expressed in bp.

dSimilarities to known proteins deposited in the databases.

eLocations on polytene chromosomes as obtained by in situ hybridization.

fSummary of the expression analysis as determined by RT-PCR. sg, female salivary glands; c, 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
expression pattern; it represents the only gene that is tran-
scribed at a reduced level in the female salivary glands as
compared to the other tissues and whose expression is re-
stricted to adult stages. In summary, the analysis suggests
that we have identified five novel female salivary gland-speci-
cic 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 com-
ponent of the protein translocation apparatus. Moreover,
clone 7 showed high similarity to three previously identi-
ced 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 corre-
sponding protein products will be accordingly indicated as
gSGn.

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 expres-
sion 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 expres-
sion in female glands they may somehow be involved in
blood-feeding; we propose to refer to this novel group of proteins as gSGin.

Fig. 1. 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 ac-
cording to expression profile (see text). Actin mRNA was ampli-
fied as control. The different clones and the corresponding gene names
are indicated on the right.

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.

The cDNAs 12.5D and 5, whose corresponding genes were

and to the PLA2 from the venom of the cobra, 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 PLA2 from the venom of the cobra, Naja naja atra (NnPLA2). 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 PLA2 from Naja naja atra are indicated.

named gSG6 and gSG7, showed weak but potentially meaningful similarity to anti-coagulant proteins from the hematothaphous nematode, Ancylostoma caninum, and from the venom of the Chinese cobra, Naja naja atra (NnPLA2). 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 PLA2 from Naja naja atra are indicated.

References
