

Nuclear Factors Bind to a Conserved DNA Element That Modulates Transcription of *Anopheles gambiae* Trypsin Genes*

Received for publication, June 23, 2000, and in revised form, August 23, 2000
Published, JBC Papers in Press, October 2, 2000, DOI 10.1074/jbc.M005540200

Federico Giannoni‡§, Hans-Michael Müller¶||, Jacopo Vizioli§**, Flaminia Catteruccia‡, Fotis C. Kafatos¶||, and Andrea Crisanti‡ §§

From the ‡Imperial College of Science, Technology and Medicine, Department of Biology, Imperial College Road, London SW7 9AX, United Kingdom, §Università 'La Sapienza,' Istituto di Parassitologia, P. le Aldo Moro 5, 00185 Rome, Italy, and the ¶European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg D-69117, Germany

The *Anopheles gambiae* trypsin family consists of seven genes that are transcribed in the gut of female mosquitoes in a temporal coordinated and mutually exclusive manner, suggesting the involvement of a complex transcription regulatory mechanism. We identified a highly conserved 12-nucleotide motif present in all *A. gambiae* and *Anopheles stephensi* trypsin promoters. We investigated the role of this putative trypsin regulatory element (PTRE) in controlling the transcription of the trypsin genes. Gel shift experiments demonstrated that nuclear proteins of *A. gambiae* cell lines formed two distinct complexes with probes encompassing the PTRE sequence. Mapping of the binding sites revealed that one of the complex has the specificity of a GATA transcription factor. Promoter constructs containing mutations in the PTRE sequence that selectively abolished the binding of either one or both complexes exerted opposite effects on the transcriptional activity of trypsin promoters in *A. gambiae* and *Aedes aegypti* cell lines. In addition, the expression of a novel GATA gene was highly enriched in *A. gambiae* guts. Taken together our data prove that factors binding to the PTRE region are key regulatory elements possibly involved in the blood meal-induced repression and activation of transcription in *early* and *late* trypsin genes.

After female mosquitoes feed on a vertebrate host the blood meal is digested by the combined action of several enzymes including trypsin serine proteases (1). In *Anopheles gambiae*, seven trypsin genes have been identified, *Antryp1* to *Antryp7*, all clustered together in a relatively small chromosomal locus of 11 kb¹ (2). These genes were arranged in the following two distinct groups on the basis of their expression pattern: constitutively expressed “*early*” trypsins (*Antryp3*, -4, -5, -6, and -7) and blood meal induced “*late*” trypsins (*Antryp1* and -2). Unfed female mosquitoes were shown to constitutively transcribe in

their gut *Antryp1* and *Antryp4* and to a lesser extent *Antryp3*, -5, -6, and -7. After blood feeding, the transcription of *Antryp1* and -2 was rapidly induced, whereas the expression of *Antryp3*, -4, -5, -6, and -7 appeared to be down-regulated (3). The identification of the DNA elements and transcription factors that regulate the expression of the *early* and *late* trypsin genes has important implications to understand the molecular mechanisms that control gene expression in the mosquito gut. This information is anticipated to shed new light on how distinct members of complex gene families in different organisms are differentially transcribed in a spatial and temporal regulated manner. Moreover, the functional characterization of *early* and *late* trypsin promoters may prove invaluable for the development of control measures against vector-borne diseases based on transgenic mosquitoes expressing anti-parasitic or antiviral agents in the gut.

Transformation experiments of *Drosophila melanogaster* with *Anopheles gambiae* trypsin promoter constructs revealed that promoter sequences of *Antryp1* and -2 were constitutively active in the gut of the fruit fly (4). The *cis*-acting elements were mapped to DNA sequences encompassing nucleotides -360 to -153 and -418 to -168 of the *Antryp1* and *Antryp2* promoter, respectively (4), and binding sites to *D. melanogaster* nuclear factors were identified (5). In this experimental system the trypsin promoters were transcribed in the gut of both female and male flies. These findings indicated that all *A. gambiae* and *D. melanogaster* share DNA elements and factors that selectively control gene expression in their gut, additional transcription control mechanisms are responsible for the temporal, sex-restricted, and mutual exclusive transcription of the trypsin genes in *A. gambiae*.

Sequence analysis of the trypsin locus revealed the presence of a highly conserved motif of 12 nucleotides upstream to the TATA box of all seven trypsin genes outside the region controlling tissue-specific transcription in transgenic flies (Fig. 1). The sequence and the structural organization of this motif and its upstream region suggested that it could be implicated in controlling the transcription of the *early* and *late* trypsin genes in a mutually exclusive manner. This putative trypsin regulatory element (PTRE) encompassed a TATCA-conserved motif and a TTAATC consensus sequence, which are known to function as binding sites for GATA transcription factors and homeobox genes, respectively. The analysis of the DNA sequences immediately upstream to the PTRE motifs showed a structural organization that substantially differed in *early* and *late* trypsin genes. In the *late* trypsin genes, *Antryp1* and -2, the PTRE is part of a large palindromic sequence (4). This organization generates a second GATA-binding site oriented in opposite direction to the PTRE. With the exception of *Antryp7*, the palindromic structure of the *early* trypsin genes appears

* This work was supported in part by a World Health Organization grant (to A. C. and F. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

|| Supported by a fellowship of the Deutsche Forschungsgemeinschaft.

** Present address: Institut de Biologie Moléculaire et Cellulaire, UPR 9022 du Centre National de la Recherche Scientifique, 15 Rue René Descartes, F67084 Strasbourg, France.

§§ To whom correspondence should be addressed. Tel.: 44-171-5945426; Fax: 44-171-5945439; E-mail: a.drcrisanti@ic.ac.uk.

¹ The abbreviations used are: kb, kilobase pairs; PTRE, putative trypsin regulatory element; PCR, polymerase chain reaction; DTT, dithiothreitol; PMSF, phenylmethylsulfonyl fluoride; EMSA, electrophoretic mobility shift assays; 5'-Flk, 5'-flanking sequence; h, human.

largely degenerated and contains only a few conserved nucleotides. The PTRE palindrome of *Antryp7* is highly conserved; however, this sequence contains a single nucleotide substitution that disrupts the GATA consensus sequence upstream to the PTRE thus suggesting that *Antryp7* may have switched from *late* to *early* trypsin after losing this second GATA site.

We report here on a series of experiments aimed at identifying factors able to bind to the PTRE and its palindrome and at investigating the ability of these DNA elements to regulate the transcription of the trypsin genes. Oligonucleotides spanning the PTRE of *early* and *late* trypsin genes and a series of PTRE mutated versions were used in electrophoretic mobility shift assays together with nuclear extracts from a collection of *A. gambiae* cell lines. Different *early* and *late* trypsin promoter constructs carrying either single nucleotide substitutions and/or deletions in the PTRE and its palindromic sequence were assessed in transient transfection experiments to analyze their transcriptional activity.

MATERIALS AND METHODS

Plasmid Constructs—All plasmid constructs were generated by subcloning the DNA inserts into pBluescript SK plasmid (Stratagene). Plasmid pBLuc contains the luciferase cDNA (1.7 kb) followed by 0.9 kb of SV40 polyadenylation signal. Plasmid pT4-Luc contains the luciferase cDNA and SV40 polyadenylation signal under the control of 1.375 kb of the *Antryp4* putative promoter sequences, from nucleotide -1 to -1375. Plasmids pAct-Luc and pAct-LacZ contain 2.7 kb of *D. melanogaster* 5C actin promoter followed by the luciferase or the LacZ-coding sequences and SV40 polyadenylation signal. In order to manipulate the DNA motifs of *Antryp4* described in the Introduction, oligonucleotides T4-Eco-S and T4-Eco-AS were used to engineer two *EcoRI* sites in positions -46 and -124, positions chosen to minimize the number of introduced mutations, using a SculptorTM *in vitro* mutagenesis system (Amersham Pharmacia Biotech). The intervening sequences encompassed the putative regulatory elements of the *Antryp4* promoter. This procedure generated a DNA cassette that could be easily removed and replaced with a given DNA sequence flanked by two *EcoRI* sites. Sequence substitutions and point mutations were introduced in the DNA regions encompassing the two *EcoRI* sites using an overlap PCR approach. Briefly, two complementary oligonucleotides containing the desired mutations and a short 5'-overlapping sequence were used in two separate amplifications using external primers. The products thus generated were gel-purified, mixed at equal concentrations in a PCR mix containing the external primers, and amplified to obtain a product carrying the desired mutations. This product was then digested with *EcoRI* and inserted into *EcoRI*-digested pT4. All mutated constructs were verified by sequencing the DNA region using a T7 Sequencing kit following the manufacturer's instructions (Amersham Pharmacia Biotech).

Cloning of *Anopheles stephensi* Trypsin Genes—A *Sau3AI* partially digested *A. stephensi* genomic library was prepared in λ EMBL3A as described previously (2). For library screening, a PCR-amplified 772-base pair DNA fragment, corresponding to the complete coding sequence of *Antryp1*, was labeled with digoxigenin (Roche Molecular Biochemicals) and used as a probe, according to the manufacturer's protocol. Four positive clones were identified, and the 14.5-kb clone λ T3 was used for subcloning and sequencing the *A. stephensi* trypsin genes *Astyp1* (GenBankTM accession number U52359) and *Astyp3* (GenBankTM accession number AF012809).

Establishment of Cell Lines, Culture, and Transfection—The *A. gambiae* lines Sua 1.0, Sua 1B, Sua 4.0, Sua 5.0, Sua 5B, 4a-3A, 4a-2.4, and L35-5 were established as described for cell line 4a-3A (6) using mosquito strains Suakoko 2La, 4a r/r, and L35. Insect cells were grown at 22–26 °C in Schneider's medium prepared as recommended by Sigma, supplemented with 5–10% fetal calf serum, 100 units/ml penicillin, and 100 μ g/ml streptomycin. MOS 20 cells were grown at 22–26 °C in M199 medium containing 11 g/liter M199, 4 g/liter lactalbumin, 12.5 ml/liter yeastolate, 0.35 g/liter sodium bicarbonate, 5–10% fetal calf serum (Sebam), 100 units/ml penicillin, and 100 μ g/ml streptomycin (Life Technologies, Inc.). For transfection, 1–4 $\times 10^5$ per cm² cells were plated in 24-well (2 cm² each) plates (Costar) in M199 supplemented with 10% fetal calf serum and antibiotics. After approximately 24–48 h, when optimal cell density was reached, cell medium was replaced with 0.9 ml of incomplete M199 medium, lacking fetal calf serum and antibiotics. In 1.5-ml sterile microcentrifuge tubes, a mixture containing 95

μ l of incomplete M199, 5 μ l of Lipofectin reagent (Life Technologies, Inc.), 1.5 μ g of DNA trypsin promoter constructs, and 1.5 μ g of pAct-LacZ for normalization purposes was incubated for 15 min at room temperature and then added to each well. The cells were incubated in transfection mixture for 6 h up to overnight at 26 °C. After transfection, medium was replaced with complete M199 or Schneider's medium, and cells were incubated for approximately 24 h before lysis. Cells were then washed once with phosphate-buffered saline and lysed in Cell Culture Lysis Reagent (Promega). Plates were stored at -80 °C before determination of reporter activity. Luciferase activity was assessed according to manufacturer's protocol (Luciferase Assay System, Promega) with a Berthold luminometer and measuring the light emission for 20 s. The activity of β -galactosidase was measured by incubating 5–20 μ l of cell lysate diluted in a total of 150 μ l of Cell Culture Lysis Reagent (Luciferase Assay System, Promega) in a mixture containing 150 μ l of 2 \times Assay Buffer (β -galactosidase Enzyme Assay System, Promega) at 37 °C. Reaction was stopped by adding 500 μ l of 1 M Tris, and absorbance was read at 420 nm. Only values within the linear range were considered. Relative luciferase units were normalized to β -galactosidase units and expressed relative to the control plasmid in each assay. All transfection experiments were performed in triplicate wells for each construct, and all experiments were repeated at least twice.

Preparation of Nuclear Extracts—*A. gambiae* and *Aedes aegypti* cells were grown at 26 °C as described above in 75-cm² flasks. Cultured cells were harvested either by vigorous pipetting or by scraping. The cells were washed in phosphate-buffered saline, centrifuged at 1000 $\times g$, and resuspended in hypotonic buffer, containing 10 mM Hepes, pH 7.9, at 4 °C, 1.5 mM MgCl₂, 10 mM KCl, 0.2 mM PMSF, 0.5 mM DTT, 10 μ g/liter leupeptin and antipain (Sigma). PMSF, DTT, leupeptin, and antipain were added just before use. The cells were allowed to swell on ice for 10 min and then lysed in a Dounce homogenizer, with 20 strokes of a type B pestle. Homogenates were centrifuged 15 min at 3300 $\times g$. The supernatants were collected and directly dialyzed for cytoplasmic extract. Nuclear pellets were resuspended in half-packed volume of low salt buffer, containing 20 mM Hepes, pH 7.9, at 4 °C, 1.5 mM MgCl₂, 20 mM KCl, 0.2 mM EDTA, pH 8, 25% glycerol, 0.2 mM PMSF, 0.5 mM DTT, 10 μ g/liter leupeptin and antipain. An equal volume of high salt buffer, containing 20 mM Hepes, pH 7.9, at 4 °C, 1.5 mM MgCl₂, 1.2 M KCl, 0.2 mM EDTA, pH 8, 25% glycerol, 0.2 mM PMSF, 0.5 mM DTT, 10 μ g/liter leupeptin and antipain, was slowly added to the nuclei. After 30 min of incubation with gentle stirring, the extracted nuclei were centrifuged for 30 min at 17,000 $\times g$. The supernatant, containing the nuclear proteins, was recovered and dialyzed against dialysis buffer, containing 20 mM Hepes, pH 7.9, at 4 °C, 100 mM KCl, 0.2 mM EDTA, pH 8, 20% glycerol, 0.2 mM PMSF, and 0.5 mM DTT. All incubations and centrifugations were carried out at 4 °C. Nuclear extracts were distributed in aliquots and stored at -80 °C until use. Protein concentrations was determined according to Bradford (7).

Electrophoretic Mobility Shift Assay (EMSA)—Oligonucleotides designed according to the putative regulatory sequences of *Antryp1*, -2, and -4 were annealed to the complementary strand, in order to constitute a double strand probe. Each oligonucleotide contained a 5'-protruding tail of four deoxythymidine residues that did not anneal to the complementary strand, to allow labeling by fill-in procedure. Approximately 10 pmol of double strand oligonucleotides were radiolabeled in 10 μ l of labeling buffer containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 5 μ M each dCTP, dGTP, dTTP, 2–3 μ l of α -³²S-dATP (10 μ Ci/ μ l, 1000 Ci/mmol, Amersham Pharmacia Biotech), and 1 unit of *Taq* DNA polymerase (Roche Molecular Biochemicals) and incubated at 37 °C for 45 min. Labeled probes were purified from the non incorporated [α -³²S]dATP using ProbeQuant G-50 Micro Columns (Amersham Pharmacia Biotech). EMSA were performed according to standard procedures (8). A suspension containing approximately 1.5 μ g of nuclear extract was incubated for 20 min at room temperature with 0.2 pmol of radiolabeled probe, 0.5 μ g of denatured herring sperm DNA, 1 μ g of polydeoxyinosinic-deoxycytidylic acid, in 10% glycerol. The binding buffer contained 10 mM Hepes, pH 7.6, 20 mM KCl, 1 mM EDTA, pH 8.0, 2.5 mM DTT, in a final volume of 15 μ l. For competition experiments, a 150-fold molar excess of cold competitor was added. The samples were run on a 4% non-denaturing polyacrylamide-bisacrylamide (30:0.4) gel containing 6.8 mM Tris-HCl, pH 7.9, 1 mM EDTA, pH 8.0, 2.5% glycerol in Low Ionic Strength Electrophoresis buffer, containing 6.8 mM Tris-HCl, pH 7.9, 1 mM EDTA, pH 8.0, 3.3 mM sodium acetate, pH 7.9. Gels were dried and exposed for autoradiography.

Cloning of *AgGATA*—Genomic DNA from Sua 4.0 cells was extracted according to a standard procedure (7). Approximately 300 ng of genomic DNA was used in PCR amplification with primers BAC-ZFS (5'-GAG-GGACGAGAGTGCCTCAAC-3') and GATA-Rev-Deg-Xba (5'-GATCTC-

TAGADCCRCANGCRRTRCANACNGGRTCDCC-3'). Primer BAC-ZFS was designed from a sequence found in the *A. gambiae* Genoscope data base and sharing high homology to the first zinc finger motif of GATA factors. The degenerate GATA-Rev-Deg-Xba primer encompassed the amino acid sequence GDPVCNACG and was previously used by Drevet *et al.* (9) to identify two GATA factors from the silkworm *Bombyx mori*. PCR was carried out for 35 cycles using an annealing temperature of 50 °C for the first 3 cycles and 55 °C for the next 32 cycles. A PCR product of the expected size (376 base pairs) was cloned in pGEM-T Vector (Promega) and sequenced (GenBank™ accession number AJ404478).

RNA Extraction and RT-PCR Procedure—Total RNA was extracted from guts and carcasses of young male and freshly emerged female *A. gambiae* mosquitoes as well as from Sua 4.0 and Sua 5.0 cells as described previously (10). DNase I-treated total RNA (2–5 µg) was incubated for 5 min at 65 °C with 50 ng of random hexamer oligonucleotides (Promega) before adding RT buffer (Promega), 40 units of RNasin Ribonuclease Inhibitor (Promega), 1.25 µl of 10 mM each dNTP, and 200 units of Moloney murine leukemia virus reverse transcriptase (Promega)-reverse transcriptase. The equivalent of approximately 200 ng of reverse-transcribed RNA was processed for PCR using the primers 3/10-RT1S (5'-CGTGC GCGCTATACACGAGACA-3'), 3/10-AS2 (5'-ATGCGTTGCAGACCGATCGCC-3'), and 3/10-AS3 (5'-AGTTGGCACAGGTCACGCCCCGAA-3'), annealing to different regions of the AgGATA sequence. Products obtained with 3/10-RT1S and 3/10-AS2 were cloned and sequenced. Alternatively, the identity to the AgGATA product was confirmed by PCR with nested primer 3/10-AS3. As control the cDNA was amplified using the primers Act-F1 (5'-ACCCCATCTCACACAC-TTC-3') and Act-R1 (5'-ATGTCTTTCATTGCGGCC-3') annealing to the *A. gambiae* actin 5C gene. A duplicate of each RNA sample was processed for PCR without the reverse transcriptase, to detect genomic DNA contamination.

The oligonucleotides used for EMSA are as follows: T4-Eco-S, CCC-CAATCTTCTGAATTCACGT; T4-Eco-AS, TGATACGATGAATTCTGATTTTCTACATG; T4 PTRE, TTTTAGTATCAGTTAATCAG; T4 5'-Flk, TTTTACGATCATTACGATC; T2 PTRE, TTTTAGTATCAATAATCAG; T2 5'-Flk, TTTTGCAGTATTTGATAGTG; T2 Ups, TTTTGTCTATCAGTATTTTCG; T1 PTRE, TTTTAGTATCAATAATAAG; T1 5'-Flk, TTTTGCAGCAATTGATATCG; T1 Ups, TTTTACGATCAGCAATTGCG; h-GATA, TTTTGGCCTTATCTCTTTACCCCCAC; h-GATA CT, TTTTGGCCTTAAGTCTTTACCCCCAC.

RESULTS

Distribution of the PTRE Sequence in Trypsin Genes of Different Mosquito Species—To facilitate the structural and functional analysis of the trypsin promoters in *Anopheles* mosquitoes, we have analyzed the DNA regions flanking the trypsin genes of *A. stephensi*. We have isolated the genomic clones of two trypsin genes from *A. stephensi*. Sequence analysis allowed the identification of two genes named *Ast1* and *Ast3* for the sequence similarities with their *A. gambiae* homologues *An1* and *An3*. The putative promoter proximal regions of *Ast1* and *Ast3* contained a sequence showing a high degree of similarity to the PTRE regions of *A. gambiae* trypsin genes (Fig. 1). The PTRE and 5'-Flk regions of *Ast1* and *Ast3* showed the same structural organization found in *An1*. Sequence alignment of the upstream regions of *A. aegypti* trypsin genes (11) with the corresponding region of anopheline genes revealed the presence of a DNA motif showing some homology to the PTRE. Similarly to what has been observed in *A. gambiae*, early and late trypsin genes of *A. aegypti* differed for the distribution of potential GATA sites in front of the PTRE sequence. These findings indicated that early and late trypsin genes share a similar structural organization in the distribution and orientation of PTRE sequences near the TATA box thus suggesting the presence of a common transcription regulatory mechanism in anopheline species, possibly conserved in distantly related mosquitoes.

Nuclear Proteins of *A. gambiae* Cells Bind to the PTRE Sequence—To search for *A. gambiae* nuclear proteins selectively binding to the PTRE and its 5'-flanking sequence (5'-Flk) we have designed a series of double strand oligonucleotide probes that encompassed the PTRE and the 5'-Flk region of

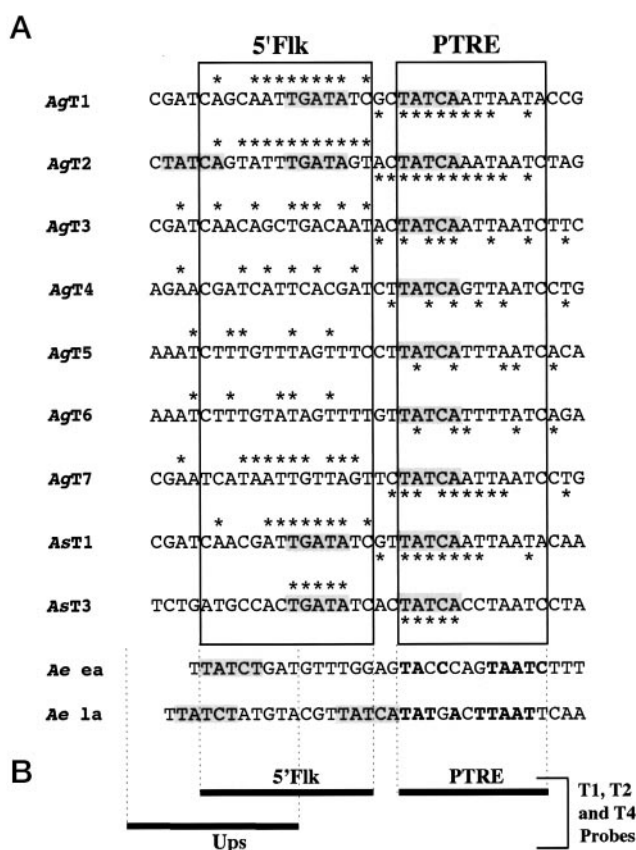


FIG. 1. Sequence alignment of the PTRE regions of *A. gambiae*, *A. stephensi*, and *A. aegypti* trypsin genes. A, the conserved PTRE and 5'-Flk sequences are boxed. Palindromic nucleotides are indicated with an asterisk. All potential GATA sites are highlighted in gray boxes. Non-conserved nucleotides in the TTAATC motif are underlined. The *A. gambiae* and *A. stephensi* sequences are abbreviated as AgT and AsT, respectively. *Aedes* sequences are abbreviated Ae ea (early) and Ae la (late). B, schematic representation of the regions encompassed by the T1, T2, and T4 probes used in EMSA experiments.

different early and late trypsin genes (Fig. 1). These probes were used in EMSAs using nuclear extracts from a collection of *A. gambiae* cell lines that have been recently developed (6). The cell lines were employed in this study as an alternative to mosquito guts that proved not to represent a useful source of nuclear extracts for EMSA experiments. After dissecting thousands of *A. gambiae* guts, we could not recover suitable material to study DNA-protein interactions, possibly because of the low number of epithelial cells and the presence of high concentrations of activated proteases (12). EMSA experiments revealed that the nuclear extract of the larvae-derived *A. gambiae* Sua 4.0 cells contained factors forming two different complexes (A and B) with a probe spanning the PTRE of *An1* (T4 PTRE) (Fig. 2A). A single band that showed an electrophoretic mobility similar to complex A was also observed when incubating the same nuclear extracts with a probe encompassing 12 nucleotides upstream to *An1* PTRE (T4 5'-Flk) (Fig. 2A). Proteinase K treatment prevented the formation of complexes A and B thus demonstrating that the electrophoretic mobility shift of T4 PTRE and T4 5'-Flk were due to the presence of nuclear proteins that bound to these probes. To rule out the possibility that additional complexes were binding to target sites spanning across the PTRE and the 5'-Flk probes, we have carried out EMSAs using a probe that encompassed both sequences. This analysis revealed that no additional complexes bound to the PTRE and the 5'-flanking sequences of *An1* (data not shown).

We have extended the search of factors binding to the PTRE

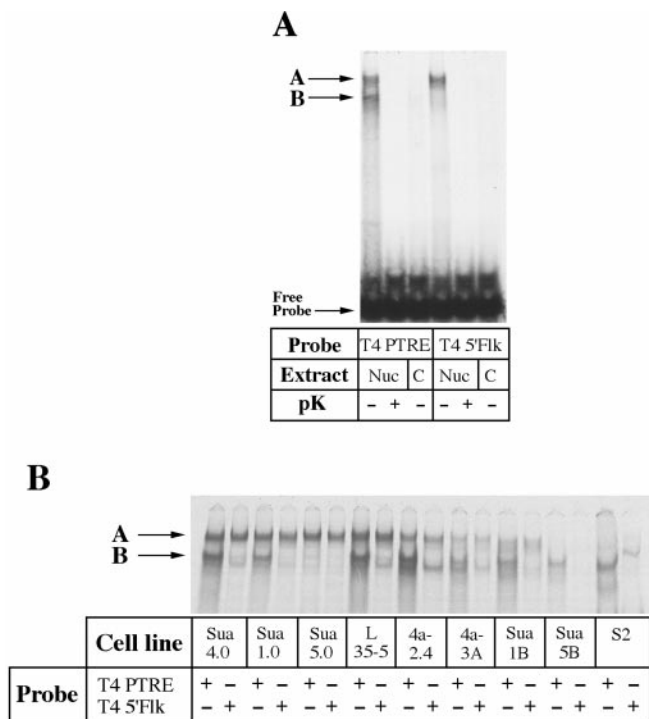


FIG. 2. Nuclear proteins of *A. gambiae* cells binding to the PTRE and 5'-Flk probes of *Antryp4* (T4). A, EMSA showing the binding of factors present in the nuclear (Nuc) and cytoplasmic (C) extracts of Sua 4.0 cells. Proteinase K (pK) was added (+) at a final concentration of 60 ng/ μ l. B, binding pattern of nuclear extracts of *A. gambiae* (Sua 4.0, Sua 1.0, Sua 5.0, L35-5, 4a-2.4, 4a-3A, Sua 1B, and Sua 5B) and *D. melanogaster* (S2) cell lines to T4 PTRE and T4 5'-Flk probes. The arrows indicate the migration of complex A, B, and unbound probe.

to a vast repertoire of *A. gambiae* cell lines and S2 cells from *D. melanogaster*. Our results indicated that nuclear proteins from most of the *A. gambiae* cell lines with the exception of Sua 5.0 and Sua 5B showed a binding pattern to T4 PTRE and T4 5'-Flk similar to that observed with Sua 4.0 cells (Fig. 2B). *D. melanogaster* S2 nuclear extracts formed a complex showing an electrophoretic motility similar to complex B. Nuclear extracts from Sua 5.0 and Sua 5B cells formed with the T4 PTRE probe only complex A and complex B, respectively. Moreover, nuclear extract from the *A. aegypti* cell line MOS 20 incubated with a probe encompassing both PTRE and 5'-Flk motifs formed two complexes showing the electrophoretic mobility of complex A and complex B of *A. gambiae* cells (not shown). These findings would indicate that the expression of the two complexes is not ubiquitous and can be independently regulated. These experiments also provided the rationale to employ Sua 4.0, Sua 5.0, and Sua 5B cells for further analyzing the transcriptional activity of complex A and B on promoter constructs.

Complex A and B Interact with Different Residues of the PTRE Sequence—To investigate the specificity of the interaction between nuclear proteins and the T4 PTRE and T4 5'-Flk probes, we carried out EMSA experiments in the presence of different competitor oligonucleotides. The binding of Sua 4.0 nuclear proteins to T4 PTRE was abolished by the presence of an excess of the corresponding unlabeled oligonucleotide. A mutated T4 PTRE probe, carrying two nucleotide substitutions at position 3 and 9 (G3C9), was unable to inhibit the formation of complex A and B (Fig. 3A); accordingly, nuclear extracts of Sua 4.0 cells failed to bind to this probe. Interestingly, an excess of unlabeled T4 5'-Flk oligonucleotide efficiently inhibited the formation of complex A with T4 PTRE, whereas complex B was not affected. Similarly, the formation of complex A

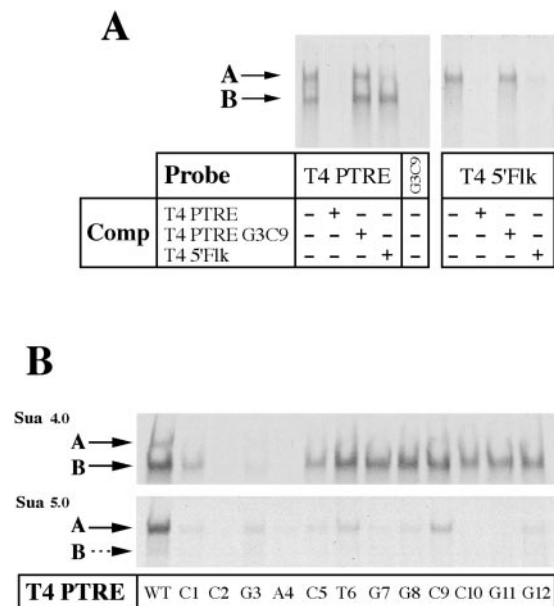


FIG. 3. Binding specificity of complex A and B to the PTRE and 5'-Flk sequences of *Antryp4* (T4). A, Sua 4.0 nuclear extracts were incubated with T4 PTRE and T4 5'-Flk in the presence of an excess (Comp) of T4 PTRE, double mutant T4 PTRE G3C9, containing T-G and an A-C substitutions in positions 3 and 9 of the PTRE, and T4 5'-Flk oligonucleotides. Lane G3C9 shows incubation with the labeled probe T4 PTRE G3C9. B, mapping of the binding sites of complex A and B to the PTRE. Sua 4.0 (upper panel) and Sua 5.0 (lower panel) nuclear extracts were incubated with T4 PTRE probes containing point mutations in each of the 12-nucleotide motifs. The letter and the number indicate the nucleotide introduced and its position in the motif, respectively. Binding of non-mutated (WT) T4 PTRE is also shown. The migration of complex A and B bands is indicated.

with the T4 5'-Flk probe was inhibited by an excess of unlabeled T4 PTRE and T4 5'-Flk oligonucleotides (Fig. 3A), thus indicating that complex A proteins could bind to both T4 PTRE and T4 5'-Flk probes. These experiments also showed that complex B differed from complex A in terms of size and DNA sequence specificity.

To map exactly the nucleotide residues forming the binding sites of complex A and B within the PTRE sequence, we carried out EMSAs using a series of oligonucleotides that contained point mutations in each of the 12 residues of T4 PTRE. The mutated probes were incubated with nuclear extracts of Sua 4.0 and Sua 5.0 cells. Mutations in the residues spanning the TATCA motif either reduced or abolished the ability of the oligonucleotides to form complex B with nuclear extracts of Sua 4.0 cells. The formation of complex A with Sua 4.0 nuclear extracts was severely impaired by mutations in any of the 12 residues of the PTRE sequence (Fig. 3B). When using Sua 5.0 extracts, which lacked complex B, some differences could be detected in the binding of complex A to the mutated probes (Fig. 3B). Substitutions in the residues 2, 4, 5, 7, 10, and 11 completely inhibited the formation of complex A with the PTRE, whereas mutations at the other positions still allowed a reduced binding to occur. Sequence comparison revealed that most of these residues form a motif (ATCANTNA) that is present in both the PTRE and 5'-Flk regions of *Antryp4* thus providing the basis for understanding the binding of complex A to both T4 PTRE and T4 5'-Flk. Competition experiments using mutated T4 5'-Flk oligonucleotides confirmed that the ATCA sequence is implicated in the formation of complex A with the 5'-Flk region (data not shown).

The identification of complex B-binding site revealed that this complex had the binding specificity of GATA transcription factors (13). To investigate this property further, we have an-

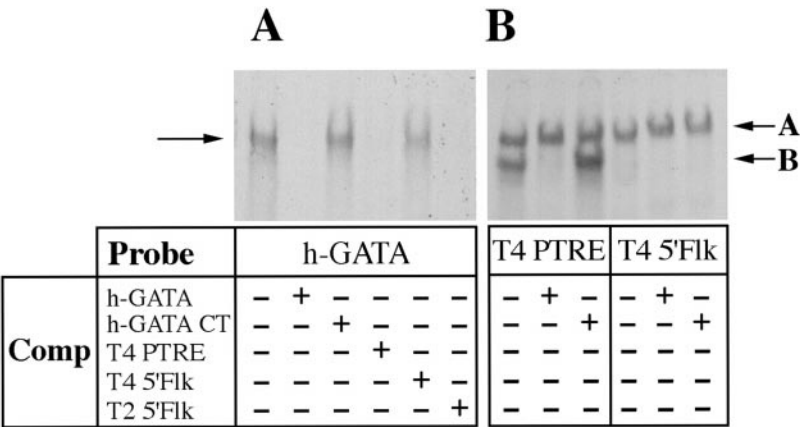


FIG. 4. **Binding of Sua 4.0 nuclear factor(s) to the h-GATA probe.** A, nuclear extracts were incubated with the h-GATA probe, encompassing a sequence of the human porphobilinogen deaminase promoter. EMSA was carried out either in the absence or in the presence of an excess (Comp) of the oligonucleotides h-GATA, h-GATA CT, T4 PTRE, T4 5'-Flk, and T2 5'-Flk. The probe h-GATA CT contains two nucleotide substitutions in the GATA site of the h-GATA oligonucleotide. B, binding of the A and B complexes to the T4 PTRE and T4 5'-Flk probes in the presence of the competitor (Comp) oligonucleotides h-GATA and h-GATA CT. Arrows indicate the migration of complex A and B.

alyzed the ability of Sua 4.0 nuclear extracts to bind to a probe (h-GATA) designed to encompass a sequence of the human porphobilinogen deaminase promoter. This probe contained a GATA site (14) and flanking sequences totally unrelated to the PTRE region. Our results indicated that the h-GATA probe formed with nuclear extracts of Sua 4.0 cells a single complex that showed the electrophoretic mobility of complex B and was competed by an excess of T4 PTRE and T2 5'-Flk, (Fig. 4A). Moreover, the h-GATA oligonucleotide selectively competed the binding of complex B to T4 PTRE (Fig. 4B), whereas oligonucleotides either lacking a GATA sequence (T4 5'-Flk) or containing mutations in this motif (h-GATA-CT) had no effect (Fig. 4, A and B).

While the PTRE showed little sequence variation among the trypsin genes, the 5'-Flk sequences significantly differed in the *early* and *late* trypsin genes for the presence of additional GATA sites. This observation prompted us to compare the PTRE and the 5'-Flk sequence of *early* (*Antryp4*) and *late* (*Antryp1* and 2) trypsin genes for their binding pattern to nuclear extracts of Sua 4.0 cells (Fig. 5). Our results indicated that the PTRE and the 5'-Flk probes of *Antryp1* and *Antryp2* efficiently bound to complex B which is in agreement with the presence of palindromic GATA sites in the 5'-Flk sequences of the *late* trypsin promoters. The T2 Ups probe, encompassing a sequence immediately upstream to the 5'-Flk of T2, showed a strong binding to complex B, whereas the corresponding T1 Ups probe that lacked a GATA sequence did not bind to any factor (Fig. 5). The binding of complex A to the T1 and T2 5'-Flk and T2 PTRE probes was detectable but weak (Fig. 7).

Mosquito Guts and Sua 4.0 Cells Express the Same GATA Transcription Factor—The observation that nuclear extracts of distinct *A. gambiae* cell lines contained proteins with the binding specificity of GATA factors raises questions about the nature of these molecules and their expression in the mosquito gut. To address these issues, we have attempted to clone *A. gambiae* GATA factors and investigate their transcription pattern in mosquito tissues, organs, and cell lines. Primers designed to anneal to the conserved DNA binding domain of GATA factors were employed in PCR experiments using as template genomic DNA extracted from the *A. gambiae* cell line Sua 4.0. We cloned a 376-base pair PCR product (AgGATA) that showed a strong sequence homology to GATA factors and encoded two conserved zinc finger structures Cys-X₂-Cys-X₁₇-Cys-X₂-Cys (Fig. 6A). Analysis of the genomic locus of AgGATA revealed that the coding sequence contained a putative intron region. We have used oligonucleotides annealing within the

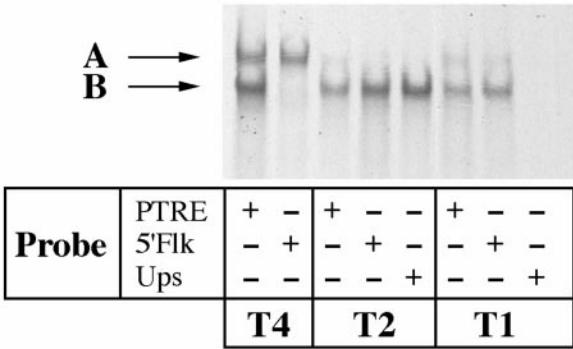


FIG. 5. **Binding pattern differences in early and late trypsins.** Sua 4.0 nuclear extracts were incubated with probes encompassing the PTRE and 5'-Flk motifs of *Antryp4* (T4), *Antryp2* (T2), and *Antryp1* (T1) and the Ups sequence of *Antryp2* and *Antryp1*. Arrows indicate complex A and B.

AgGATA sequence to investigate the expression of the corresponding gene in cell lines as well as in gut and carcasses of female and male mosquitoes. As control all cDNA samples were also processed for RT-PCR using primers annealing to the actin gene of *A. gambiae*. The RT-PCR analysis showed that AgGATA was expressed in Sua 4.0 cells but not in Sua 5.0 cells that were previously shown to be negative for the complex B in gel shift experiments (Fig. 6B). Specific RT-PCR products were detected in both female and male guts, whereas carcasses were mostly negative (Fig. 6B). Sequence analysis confirmed that the PCR products amplified from mosquito tissues encompassed the AgGATA coding region, and the smaller RT-PCR product size was explained by the lack of the intron sequence. The faint AgGATA product occasionally amplified from mosquito carcasses could either be due to a contamination of the sample preparation with gut RNA or to a low level of AgGATA transcription in other mosquito tissues.

Transcriptional Activity of the PTRE and the 5'-Flk Sequences in A. gambiae and A. aegypti Cells—We have analyzed the transcriptional activity of the PTRE and 5'-Flk sequences by transfecting Sua 4.0 cells and the *A. aegypti* cell line MOS20 with constructs encompassing the corresponding sequences of *Antryp4* (Fig. 7). To correlate the binding of complex A and B with the expression pattern of *early* and *late* trypsin genes, we have developed a series of promoter constructs in which the PTRE and 5'-Flk sequences have been mutated. The transcription of the wild type *Antryp4* promoter in Sua 4.0 and MOS 20

A



B

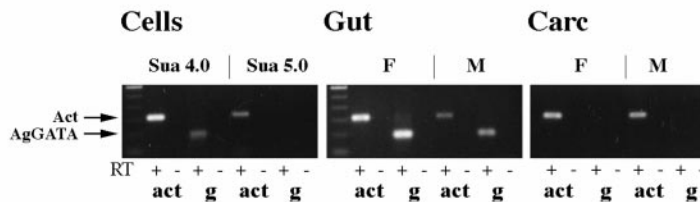


FIG. 6. **Cloning and expression of AgGATA.** *A*, alignment between the predicted amino acid sequence of AgGATA and GATA factors from *D. melanogaster* (*D*), *Xenopus laevis* (*X*), mouse (*M*), and human (*H*) showing the highest homology in the zinc finger domain. Identical residues (*) and conserved amino acid substitutions (:) are indicated. The position of the intron in AgGATA is highlighted ▼. Conserved cysteines of the zinc finger domain are boxed. *B*, RT-PCR of total RNA extracted from Sua 4.0 and 5.0 cells, midguts (*Gut*), and carcasses (*Carc*) dissected from female (*F*) and male (*M*) *A. gambiae* mosquitoes. The amplification reactions were carried out using primers specific either to the actin (*act*) or the AgGATA (*g*) coding sequence in absence (–) or in the presence (+) of Moloney murine leukemia virus reverse transcriptase. *Arrows* indicate the migration of the actin and AgGATA PCR products.

cells was compared with the activity of promoter constructs in which the entire PTRE, the 5'-Flk motif, or a sequence immediately upstream to it had been replaced by unrelated sequences. This mutation-scanning analysis revealed that in Sua 4.0 and MOS 20 cells the disruption of the PTRE (construct pT4 -76-87) resulted in an increase of transcription (Fig. 7A). This activation effect appeared to be dependent on the integrity of the 5'-Flk motif, since mutations or the substitution of this region with the corresponding *Antryp1* sequence restored the base-line transcription activity of the *Antryp4* promoter (constructs pT4 -91-105/76-87) and pT4 T1 5'-Flk/-76-87) (Fig. 7A). EMSA experiments would predict construct pT4 -76-87 to bind only to complex A at the 5'-Flk sequence. This notion suggested that the binding of complex A to the 5'-Flk sequence resulted in a positive transcriptional activity that may be down-regulated by either complex A or B binding to the PTRE. To distinguish between these two possibilities, we have analyzed the transcriptional activity of *Antryp4* promoter constructs in which the PTRE has been selectively mutagenized to allow the binding of complex B only (Fig. 7B). This analysis revealed that mutations in the GATA site, which are known to disrupt the binding of complex A and B to the PTRE, had little effect on transcription of *Antryp4* promoter in Sua 4.0 cells. The constructs pT4-(G3) and pT4-(T4) showed only a moderate decrease of the activity of the reporter gene when compared with control construct pT4-luc (Fig. 7B). On the contrary the constructs pT4-(C8) and pT4-(C9), carrying mutations that prevented the binding of complex A to the PTRE, showed a marked increase of reporter gene activity. In MOS 20 these mutated constructs showed a moderate increase of transcription. The constructs were also analyzed in Sua 5B cells that lack complex A, according to EMSA analysis. As expected, in these cells the constructs pT4-(C8) and pT4-(G9) did not show the increased transcriptional activity observed in Sua 4.0.

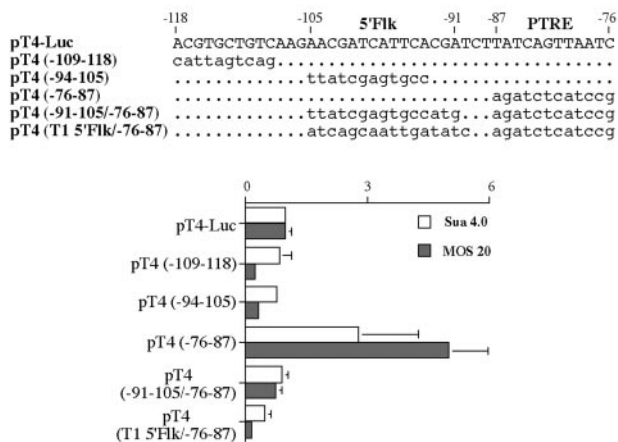
These findings together with the information concerning the differential expression of complex A and B in different cell lines and their binding specificity suggested the following conclusion. The binding of complex A to the 5'-Flk may have a positive transcriptional activity that could be either enhanced or down-regulated by the binding of complex B and complex A to the PTRE, respectively.

DISCUSSION

We have demonstrated that a conserved motif of 12 nucleotides found upstream to all *A. gambiae* and *A. stephensi* trypsin genes, the putative trypsin regulatory element, or PTRE bound to two nuclear protein complexes (A and B) that showed different DNA specificity and electrophoretic mobility. Competition and binding experiments to mutated oligonucleotides revealed that complex B selectively bound to the TATCA sequence of the PTRE thus indicating that a specific *A. gambiae* GATA factor could account for the formation of this complex. This hypothesis was supported by experiments demonstrating that Sua 4.0 cells expressed a novel gene AgGATA with sequence homology to GATA transcription factors. Accordingly, Sua 5.0 cells, which in EMSA were shown to lack complex B, did not express AgGATA. We have also shown that AgGATA is selectively transcribed in the gut of *A. gambiae* mosquitoes. These observations indicated that Sua 4.0 cells represent a valuable experimental system to study the transcriptional activity of trypsin promoters and that AgGATA is a strong candidate for playing a role in the regulation of *A. gambiae* trypsin genes. This notion is in agreement with experimental evidence indicating that GATA transcription factors are widely conserved key regulators of cell-specific functions. They are expressed in different tissues in a variety of organisms including vertebrates, insects (9, 15), *Caenorhabditis elegans* (16), and fungi (17). GATA factors were shown to control the transcription of genes expressed in a sex-restricted, tissue-specific, and time-dependent manner, and in some cases, such as the mammalian globin genes, in a mutually exclusive pattern (18–20). GATA-1, which contributes to the control of the globin locus genes, has been shown to act either as an activator (21, 22) or a repressor (23), depending on specific interactions with additional proteins. AgGATA showed a high degree of homology with MGATA-6 and HGATA-6, members of the GATA-6 subfamily, that are expressed by cells of different organs and tissues including gut epithelial cells (24, 25).

Nuclear extracts of *A. gambiae* cell lines were also shown to contain proteins forming with the PTRE of *early* (*Antryp4*) and *late* (*Antryp1* and 2) trypsin genes complex A, which migrated slower than complex B in EMSA. Mapping experiments suggested that the binding site of complex A could span most of the

A



B

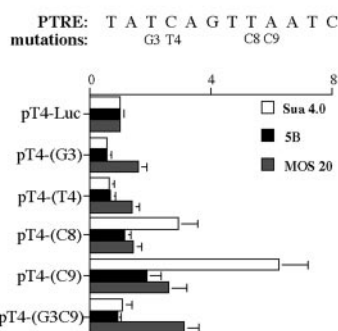


FIG. 7. Transient transfection of Sua 4.0, 5B, and MOS 20 cells with constructs containing mutations in the *Antryp4* PTRE region. A, schematic representation of the plasmid pT4-Luc and the scanning mutations introduced in the PTRE region of *Antryp4*. The pT4-Luc mutated constructs were named with numbers in brackets indicating the DNA sequences that have been substituted. In pT4 (T1 5'-Flk/-76-87) the 5'-Flk motif of *Antryp4* has been replaced with the corresponding motif of *Antryp1* (T1). In the diagram, bars indicate the mean value of two independent experiments performed in triplicate. Relative luciferase units are normalized to β -galactosidase for transfection efficiency, and values of pT4-Luc are standardized to 1. Standard deviation is also shown. B, schematic representation of the point mutations introduced in the PTRE of *Antryp4*. The letters indicate the nucleotides that were substituted, and the numbers indicate their position in the motif. Diagram shows normalized relative luciferase activity of pT4-Luc and pT4-Luc containing single and double point mutations in the PTRE sequence.

PTRE sequence including the GATA sequence. This conclusion was supported by the observation that nucleotide substitutions in most of the 12 residues of the PTRE impaired the binding of complex A to the mutated probes. To rule out the possibility that the binding of complex A to the mutated probes was altered by the presence of complex B, we repeated the mapping experiment using nuclear extracts of Sua 5.0. Sequence analysis revealed that the (ATCANTNA) motif was found in the flanking region of *Antryp4* PTRE, suggesting that the binding pattern of complex A to the PTRE and to the 5'-flanking sequence of *early* and *late* trypsin genes may reflect the distribution of these conserved nucleotide residues, forming its binding site. Complex A efficiently bound to the sequence immediately upstream to the PTRE of the *early* trypsin gene *Antryp4*, whereas it showed only a weak interaction with the corresponding regions of the *late* trypsin gene *Antryp1* and *Antryp2*.

Our findings indicated that the PTRE encompasses two partially overlapping sites for complex A and B. In the 5'-flanking

region the binding sites of complex A and B are distributed in a mutually exclusive manner in the *early* and *late* trypsin genes. We have shown that complex B bound to the 5'-Flk region of the *late* trypsin genes *Antryp1* and *Antryp2*, whereas complex A preferentially bound to probes encompassing the corresponding sequences of the *early* trypsin gene *Antryp4*. The differential distribution of the binding sites for complex A and B in the 5'-Flk region could be of functional significance for regulating *early* and *late* trypsin genes in a mutually exclusive temporal manner. This hypothesis is consistent with the observation that in a number of promoters, palindromic GATA sequences were shown to bind to GATA factors with higher affinity than single GATA sites, due to the involvement of the N-terminal zinc finger (26). Moreover, our data indicate that mutations or substitutions, anticipated to change the binding of complex A and B to the PTRE and to its 5'-flanking sequence, exerted opposite effects on the transcriptional activity of the *Antryp4* promoter in mosquito cell lines. In Sua 4.0 and MOS 20 cells the optimal activation of *Antryp4* promoter over its base-line activity was achieved by constructs that exclusively allowed the binding in tandem of complex A and complex B to the 5'-flanking sequence and the PTRE, respectively. Mutated constructs allowing the binding of either complex A or B to the 5'-flanking sequence and the PTRE showed very low activity.

These findings all together provide a framework for a transcriptional model that could explain the molecular mechanisms regulating the differential expression of the *early* and *late* trypsin genes in the mosquito gut. This model is based on the assumption that relative abundance of complex A and complex B determines the combinations of complexes binding to the 5'-flanking sequence and to the PTRE of the *early* and *late* trypsin genes. An excess of complex B is anticipated to displace complex A from the PTRE and to bind to the 5'-flanking sequence of *Antryp1* and *Antryp2*. Under these conditions complex A could still bind to the 5'-flanking sequence of *Antryp4* in tandem with complex B bound to the PTRE thus inducing the expression of this *early* trypsin gene. On the contrary, an excess of complex A would activate the *late* trypsin genes by allowing the binding in tandem of complex A and B to the PTRE and to the 5'-flanking sequence of *Antryp1* and *Antryp2*.

In conclusion, our findings provide evidence for a key role of the PTRE region in transcriptional regulation, possibly linked to the response to activation and repression of *late* and *early* trypsin genes induced by the blood meal. Newly developed *Anopheles* transformation protocol (27) will be helpful to elucidate the role of these elements and factors *in vivo*, and trypsin promoters will prove an invaluable tool toward the generation of transgenic mosquitoes expressing disease-blocking genes.

Acknowledgments—We thank the Institut Pasteur-Genoscope Anopheles Gambiae STC project for providing *A. gambiae* sequences. We thank Manlio Di Cristina, Furio Spano, Bruno Arcà, and Fabrizio Lombardo for their critical discussions and suggestions; Silvia Naitza and Roberta Spaccapelo for their precious help in providing and dissecting *A. gambiae* mosquitoes; and Simona Panni for her initial contribution to this work.

REFERENCES

- Billingsley, P. F., and Hecker, H. (1991) *J. Med. Entomol.* **28**, 865–871
- Müller, H. M., Crampton, J. M., della Torre, A., Sinden, R., and Crisanti, A. (1993) *EMBO J.* **12**, 2891–2900
- Müller, H. M., Catteruccia, F., Vizioli, J., della Torre, A., and Crisanti, A. (1995) *Exp. Parasitol.* **81**, 371–385
- Skavdis, G., Siden-Kiamos, I., Müller, H.-M., Crisanti, A., and Louis, C. (1996) *EMBO J.* **15**, 334–350
- Shen, Z., and Jacobs-Lorena, M. (1998) *Insect Biochem. Mol. Biol.* **28**, 1007–1012
- Müller, H. M., Dimopoulos, G., Blass, C., and Kafatos, F. C. (1999) *J. Biol. Chem.* **274**, 11727–11735
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. D., Smith, J. A., and Struhl, K. (1992) *Short Protocols in Molecular Biology*, 2nd Ed., Greene Publishing Associates and John Wiley & Sons Publishers, New York

9. Drevet, J. R., Skeiky, Y. A., and Iatrou, K. (1994) *J. Biol. Chem.* **269**, 10660–10667
10. Holmes, D. S., and Bonner, J. (1973) *Biochemistry* **12**, 2330–2338
11. Barillas-Mury, C., Noriega, F. G., and Wells, M. A. (1994) *Insect Biochem. Mol. Biol.* **25**, 241–246
12. Müller, H. M., Vizioli, J., della Torre, A., and Crisanti, A. (1993) *Parassitologia* **35**, (suppl.) 73–76
13. Orkin, S. H. (1992) *Blood* **80**, 575–581
14. Mignotte, V., Wall, L., deBoer, E., Grosveld, F., and Romeo, P. H. (1989) *Nucleic Acids Res.* **17**, 37–54
15. Romain, P., Heitzler, P., Haenlin, M., and Simpson, P. (1993) *Development* **119**, 1277–1291
16. Spieth, J., Shim, Y. H., Lea, K., Conrad, R., and Blumenthal, T. (1991) *Mol. Cell. Biol.* **11**, 4651–4659
17. Kudla, B., Caddick, M. X., Langdon, T., Martinez-Rossi, N. M., Bennett, C. F., Sibley, S., Davies, R. W., and Arst, H. N., Jr. (1990) *EMBO J.* **9**, 1355–1364
18. Evans, T., Reitman, M., and Felsenfeld, G. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5976–5980
19. Tsai, S. F., Martin, D. I., Zon, L. I., D'Andrea, A. D., Wong, G. G., and Orkin, S. H. (1989) *Nature* **339**, 446–451
20. Wall, L., deBoer, E., and Grosveld, F. (1988) *Genes Dev.* **2**, 1089–1100
21. Reitman, M., and Felsenfeld, G. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6267–6271
22. Seshasayee, D., Gaines, P., and Wojchowski, D. M. (1998) *Mol. Cell. Biol.* **18**, 3278–3288
23. Raich, N., Clegg, C. H., Grofti, J., Romeo, P. H., and Stamatoyannopoulos, G. (1995) *EMBO J.* **14**, 801–809
24. Huggon, I. C., Davies, A., Gove, C., Moscoso, G., Moniz, C., Foss, Y., Farzaneh, F., and Towner, P. (1997) *Biochim. Biophys. Acta* **1353**, 98–102
25. Laverriere, A. C., MacNeill, C., Mueller, C., Poelmann, R. E., Burch, J. B., and Evans, T. (1994) *J. Biol. Chem.* **269**, 23177–23184
26. Trainor, C. D., Omichinski, J. G., Vandergon, T. L., Gronenborn, A. M., Clore, G. M., and Felsenfeld, G. (1996) *Mol. Cell. Biol.* **16**, 2238–2247
27. Catteruccia, F., Nolan, T., Loukeris, T. G., Blass, C., Savakis, C., Kafatos, F. C., and Crisanti, A. (2000) *Nature* **405**, 959–962