Cloning and Localization of a Glutathione S-transferase Class I Gene from *Anopheles gambiae*

(Received for publication, August 14, 1996, and in revised form, October 25, 1996)

Hilary Ranson‡, Anthony J. Cornel§, Didier Fournier¶, Ashley Vaughan‡, Frank H. Collins§, and Janet Hemingway‡‡

From the ‡Department of Pure and Applied Biology, University of Wales, Cardiff CF1 3TL, Wales, United Kingdom, the §Entomology Branch, Division of Parasitic Diseases, Centers for Disease Control and Prevention, Chamblee, Georgia 30341, and the ¶Universite Paul Sabatier, Laboratoire d’Entomologie appliquee, 31062 Toulouse Cedex, France

1,1,1-Trichloro-2,2-bis(p-chlorophenyl)ethane (DDT) resistance in both adults and larvae of *Anopheles gambiae* is mediated by stage-specific glutathione S-transferases (GSTs). On the basis of their biochemical characteristics the larval resistance-associated GSTs are likely to be insect class I GSTs. *Aggst1–2*, a class I GST gene, which is expressed in larvae, has been cloned from the malaria vector *A. gambiae*. The gene was inserted into a bacterial expression system, and the detection of 1-chloro-2,4-dinitrobenzene (CDNB) conjugating activity in *Eschericia coli* expressing the recombinant enzyme confirmed that *aggst1–2* encodes a catalytically active GST. The gene encodes a 209 amino acid protein with 46% sequence similarity to a *Drosophila melanogaster* class I GST (GST-D1), 44% similarity with a *Musca domestica* class I GST (MdGST-1), but only low levels of homology with class II insect GSTs, including the adult specific AgGST2–1 from *A. gambiae*. Southern analysis of genomic DNA indicated that *A. gambiae* has multiple class I GSTs. *In situ* hybridization of class I genomic and cDNA clones to polytene chromosomes identified a single region of complementarity on chromosome 2R division 18B, suggesting that these class I GSTs in *A. gambiae* are arranged sequentially in the genome. Three positive overlapping recombinant clones were identified from an *A. gambiae* genomic library. Mapping and partial sequencing of these clones suggests that there are several GSTs and truncated GST pseudogenes within the 30kb of DNA that these clones span.

Glutathione S-transferases (GSTs); EC 2.5.1.18 are a large family of enzymes abundant in most organisms (1–6). They protect the cell from attack by a wide range of reactive electrophilic compounds by conjugating these compounds with the tripeptide, glutathione, thereby increasing their solubility and aiding excretion from the cell. GSTs are also involved in the glutathione-dependent dehydrochlorination of insecticides (7). They can also act as binding proteins and may be important in the intracellular transport and excretion of hydrophobic compounds (8). In mammals, five cytosolic classes (alpha, mu, pi, sigma, and theta) of GSTs have been distinguished on the basis of amino acid sequence and substrate specificity (6, 9, 10). The GSTs are differentially regulated with specific enzymes being expressed in different tissues during each developmental stage and in response to various xenobiotics. In insects, GSTs are classified as either class I or class II on the basis of their immunological cross-reactivity and amino acid sequences (11).

Elevated levels of GST activity are associated with insecticide resistance in many insects. Extensive past and current use of the organochlorine insecticide DDT in malaria control programs has selected for DDT-resistant mosquito populations throughout the world. In the principal sub-saharan malaria vector, *Anopheles gambiae*, this resistance is mediated by GST-based metabolism of DDT (12). Partial purification of the GSTs from this species resolved multiple peaks of GST activity with differential abilities to metabolise DDT (13). Further characterization of *A. gambiae* GSTs was complicated by the presence of multiple GST isoenzymes, each with overlapping substrate specificities and physical properties that made purification to homogeneity difficult to achieve. We therefore set out to characterize the genes encoding *A. gambiae* GSTs in order to determine the genetic changes that have occurred in response to exposure to insecticides. We targeted the *A. gambiae* class I GSTs since earlier biochemical analysis (14) and Northern blots suggested that class I enzymes were more likely candidates for involvement in resistance.

**EXPERIMENTAL PROCEDURES**

Two strains of *A. gambiae sensu stricto* were used. G3 was colonized from *The Gambia* in 1975 and has been maintained in the United Kingdom since then without exposure to insecticides. Suakoko 2L, originated from Suakoko in Liberia. This strain was used since it is now being employed internationally as the reference strain for *A. gambiae* mapping work (15).

**PCR Amplification of *A. gambiae* Class I GSTs**—Total RNA was extracted from *A. gambiae* G3 fourth instar larvae as described previously (16). mRNA was isolated using the Poly(A)Tract mRNA isolation system (Promega). Reverse transcription of mRNA to cDNA was achieved with the Promega Riboblonc cDNA synthesis system using an oligo(dT) adapter primer (5’-GACTCGAGTCGACATCGA(dT)17-3’). Degenerate primers for insect class I GSTs were designed based on the sequence of *Musca domestica* GST-1 (11), and these were used to amplify a homologous region from *A. gambiae* cDNA. The 50-μl PCR reaction contained 20 ng of first strand cDNA, 50 ng of each primer (5’-TA/C/T/GAAAGCCTT/C/T/CGNGA/C/T/TA/C/T/CG/CC-3’ and 5’-TCNGCNACNGTNAA/G/N/C/G/A/T/A/G/T/C/TCNCNGC-CAGCA/G/T/TA/C/T/TC/AG/T/G/CT-3’), 0.5 mM dNTPs, 2 mM MgCl₂, 1.5 units of Taq DNA polymerase, and Taq DNA polymerase buffer (50 mM KCl, 10 mM Tris-HCl (pH 9.0 at 25 °C), 0.1% Triton X-100). Thirty-five

This paper is available on line at http://www-jbc.stanford.edu/jbc/
couples of amplification were carried out (95°C for 1 min, 56°C for 2 min, and 72°C for 3 min). The 155-bp PCR product was subcloned into pBluescript (Stratagene) and sequenced in both directions. An 18-bp antisense oligonucleotide (5′-GGGCTGAGCGCCGCGGCTC-3′) specific to the 155-bp GST fragment was designed and used in a modified 5′-RACE (rapid amplification of cDNA ends) procedure (16) to obtain the 5′-end of the cDNA. The 546-bp product, *aggst1*-1, was subcloned and sequenced. For each plasmid insert, both strands of DNA were sequenced at least twice. Alignments of the nucleotide sequences were carried out using the LASENGERE package (DNASTAR, Inc.). This package was also used to align multiple insect GSTs using the clustal method. Analysis of *A. gambiae* Genomic DNA—Genomic DNA was isolated from G3 as described previously (16). Separate 10 μg aliquots of G3 genomic DNA were digested to completion with *SaI*, EcoRV, and *Hind*III, separated on an 0.8% agarose gel, transferred to a nylon membrane (Amersham Life Sciences) and hybridized with a 32P-labeled probe (specific activity > 1 × 10^9 dpm/μg) at 60°C overnight in hybridization buffer (6 × SSC (20 °C SSC contains 175 g of NaCl and 88.2 g of sodium citrate in 1 liter of water adjusted to pH 7.0 with NaOH), 0.1% (w/v) SDS, 0.1% (w/v) sodium pyrophosphate, 5% (w/v) polyethylene glycol 8000, and 1% blocking reagent (Boehringer Mannheim)).

A genomic library was made from the Suakoko 2La strain by partially digested genomic DNA with *Sau*III and *ligating* into *Bam*HI digested arms of Lambda Dash II (Stratagene). The library (150,000 plaque forming units) was screened with 32P-labeled genomic DNA were digested to completion with *Sau*III and partially digesting genomic DNA with *Hin*dIII. The extension time was decreased to 1 min.

**RESULTS**

Alignment of the 155-bp fragment from *A. gambiae* G3 cDNA, isolated by PCR using degenerate primers for insect class I GSTs, to the sequence of *aggst1*-1 from *A. gambiae* was obtained as described by Chang and Natori (18). The restriction digested hybrid was hybridized with *aggst1*-1 as described above. An open reading frame encoding *aggst1*-2, a GST which was related to but distinct from *aggst1*-1, was identified in the recombinant clone Ag_B1. A 22 bp primer (5′-GGGCTGAGCGCCGCGGCTC-3′) encompassing the predicted initiation codon (underlined) of *aggst1*-2 was constructed. This was used to amplify a cDNA containing the full length open reading frame of *aggst1*-2 from *A. gambiae* G3 in a PCR reaction utilizing the oligo-(dT) adaptor primer described above. The PCR reaction was performed as described above with the following modifications: the Taq DNA polymerase was added after an initial denaturation step at 94°C for 5 sec. The temperature was dropped to 50°C and the extension time was decreased to 1 min.

In *vitro* Expression of *aggst1*-2. *Aggst1*-2 was expressed in *vitro* by inserting the cDNA isolated above into the plasmid expression vector, pET 3a (Novagen) which contains the bacteriophage T7 promoter. The coding region of *aggst1*-2 was reamplified in a PCR reaction using primers which contained the initiation and termination codons of the gene preceded by *Bam*HI sites. The 650 bp PCR product obtained was digested with *Bam*HI, ligated into the *Bam*HI site of pET 3a and the resultant plasmid was used to transform *E. coli* BL21(DE3)pLyS. Colonies containing the appropriate insert in the correct orientation were identified by restriction digestion and grown at 30°C to an O.D. of 0.6. Expression of the recombinant enzyme was induced by the addition of 0.4 mM isopropyl β-d-thiogalactoside (IPTG) and the incubation was continued for a further 3 h at 30°C. The cells were harvested by centrifugation for 10 min at 5000 g, freeze thawed, resuspended in 50 mM Tris HCl pH 7.4, 1 mM EDTA, 10 mM β-mercaptoethanol, and disrupted by sonication. After the addition of 10 mM DTT, the cell debris was removed by centrifugation (30000 g 20 min) and the supernatants were assayed. Protein concentration was determined using Bio-Rad protein reagent (19) and GST activity was assayed spectrophotometrically by measuring the conjugation of glutathione to the standard GST substrate, 1-chloro-2,4-dinitrobenzene (CDNB) (20).

In *situ* Hybridization—*In situ* hybridizations were conducted on *A. gambiae* ovarian nurse cell polytene chromosomes as described earlier (21). Separate hybridizations were conducted using the recombinant bacteriophage clone, Ag_B1, and the 5′-GST cDNA clone, *aggst1*-1.

**Alignment**

Alignment of the 155-bp fragment from *A. gambiae* G3 cDNA, isolated by PCR using degenerate primers for insect class I GSTs, to the sequence of *M. domestica* GST-1 (11) indicated this encoded a partial GST cDNA (Fig. 1). An anti-sense oligonucleotide based on the sequence of the PCR product was used to obtain the 5′-end of the cDNA by RACE techniques. A 546-bp cDNA, *aggst1*-1, with an open reading frame of 426-bp was obtained (Fig. 2). This partial cDNA encoded a protein whose sequence shared 70–75% homology with other insect class I GSTs, indicating that *aggst1*-1 is part of a class I GST gene (Table 1). A comparison of the nucleotide sequence of *aggst1*-1 with the sequence of the initial 155-bp PCR fragment (Fig. 1) revealed several differences between the two clones.
suggesting that these cDNAs represent different alleles or are the products of different genes.

To determine whether multiple class I GST genes occur in the insecticide-susceptible G3 strain of *A. gambiae*, a Southern blot of *EcoRV-, SalI-, and HindIII*-digested genomic DNA was probed with a class I cDNA (Fig. 3). Multiple bands were detected with all three restriction enzymes. The size and number of the bands detected suggests that either *aggst1–1* is part of a large gene containing one or more introns or that multiple class I GST genes are present. In order to decide between these two interpretations, the genomic organization of class I GST genes from *A. gambiae* was investigated. To identify the structural gene encoding *aggst1–1* and to determine whether this forms part of a contiguous gene family, an *A. gambiae* 2La genomic library was screened with the cDNA probe. Arrangement of the *A. gambiae* class I GST genes was studied in the...
2La strain rather than G3 since this is the standard reference strain for large scale A. gambiae genome work (15).

Three recombinant bacteriophage clones were isolated, and partial restriction mapping showed that they spanned a contiguous region of > 30 kb of A. gambiae 2La genomic DNA. The clone Ag_B1 was mapped and partially sequenced. A 3-kb SalI fragment contained two partial GST sequences, and a 5-kb ScuI fragment encompassing the SalI fragment was then sequenced. Three open reading frames, which encoded GSTs, were identified within this fragment. The arrangement of these genes is shown in Fig. 4. Aggst1–3 and aggst1–4 are thought to be pseudogenes since, based on homology with other class I GST genes, they are truncated at the 5'-end. Aggst1–2 encodes a full-length GST. The nucleotide sequence and deduced amino acid sequences of aggst1–2 are shown in Fig. 5. The open reading frame contains no introns and encodes a 209 amino acid protein with a predicted molecular mass of 23.5 kDa. Two putative polyadenylation signals were identified.

The isolation, by 3' -RACE PCR techniques, of a full-length open reading frame of aggst1–2 from A. gambiae G3 larval cDNA confirmed that this gene is actively transcribed. To establish whether aggst1–2 encoded a catalytically active GST, the cDNA was expressed in vitro, and the CDNB conjugating activity of crude protein homogenates containing the recombinant protein was compared with similar homogenates containing the non-recombinant pET vector. Growth of the E. coli cultures containing the expression constructs at 37 °C resulted in all the recombinant protein being produced in insoluble inclusion bodies. Lowering the incubation temperature to 30 °C reduced but did not eliminate this sequestration, making comparisons between separate cultures difficult. Nevertheless, the GST activity of replicate cultures expressing recombinant agGST1–2 was always >15-fold higher than the GST activity of control cultures. The results of a representative experiment are shown in Fig. 6.

An alignment of the agGST1–2 amino acid sequence with other insect GSTs was used to calculate percentage similarities (Table I). AgGST1–2 has 41–46% similarity indices with class I GSTs from other insect species but less than 14% similarity indices with class II GSTs, indicating that agGST1–2 is a class I insect GST. The four invariant residues that appear to be of critical importance in ensuring the correct folding of all GSTs (Pro-54, Leu-143, Gly-151, and Asp-158) are conserved in agGST1–2 (indicated with an asterisk in Fig. 5). In addition, agGST1–2 possesses a serine residue near the N terminus (Ser-10) and an asparagine (Asn-48) (shown in bold in Fig. 5) near the invariant proline, both of which are characteristics of insect class I GSTs (22).

To determine the physical location of class I GSTs in A. gambiae, the aggst1–1 cDNA and genomic clones containing aggst1–2 were independently hybridized to adult ovarian nurse cell polytene chromosomes. All the probes hybridized to a single region on chromosome 2R, division 18B (Fig. 7).

**DISCUSSION**

The GSTs of A. gambiae are primarily of interest because of their role in DDT resistance. Multiple GSTs have been partially purified from this mosquito, and resistance is associated with both qualitative and quantitative changes in a number of different GST enzymes (13). The biochemical characteristics of these enzymes suggested they were likely to be insect class I GSTs. This prediction is supported by the metabolism of DDT by a recombinant Drosophila class I GST (GST D1) (23), the only expressed GST to date where DDT metabolism has been
demonstrated. We have now cloned and sequenced the first class I GST gene, \textit{aggst1–2} from \textit{A. gambiae}. The gene encodes a catalytically active GST as demonstrated by the ability of recombinant agGST1–2 to conjugate glutathione to the general GST substrate, CDNB. The open reading frame of \textit{aggst1–2} is uninterrupted by introns, which have also been observed in the \textit{Drosophila} class I GSTs (24). The \textit{aggst1–2} gene is distinct from the \textit{aggst2–1} reported earlier (25). The two genes belong to different GST classes as determined by sequence homology. They also map to different regions of the polytene chromosome.

In the ZANDS resistant strain of \textit{A. gambiae}, although DDT resistance in both adults and larvae is GST based, the same enzymes are not involved in the different life stages (13). The class II GST cDNA, \textit{aggst2–1}, was isolated from an adult female cDNA library (25). Attempts in our laboratory to isolate \textit{aggst2–1} cDNA from \textit{A. gambiae} larvae were unsuccessful and Northern analysis using an AgGST2–1 probe provided by R. Reiss (University of Irvine, CA) detected an AgGST2–1 transcript in adult mosquitoes but not in larvae. This suggests that AgGST2–1 is an adult specific cDNA. In contrast, the agGST1–2 is clearly expressed in larvae as demonstrated by our ability to amplify it by PCR from larval derived cDNA. This stage specific expression of GSTs is not unique to \textit{A. gambiae}. In \textit{D. melanogaster}, certain class I GSTs are not expressed in adults but are expressed in earlier developmental stages (24).

In \textit{D. melanogaster}, a cluster of class I GSTs, the GST D genes, are closely linked within a 60-kb DNA region. This GST family comprises at least eight genes although two of these are likely to be pseudogenes (24). Similarly, the genes encoding the multiple class I GSTs present in the housefly, \textit{Musca domestica}, are thought to be arranged sequentially within the genome (26). Southern blots of \textit{A. gambiae} genomic DNA suggested that multiple class I GST genes are also present in this species. The identification of three overlapping genomic clones spanning an area >30 kb, which were shown by hybridization experiments to contain class I GSTs, indicated that the mosquito genes are also closely linked.

\textit{In situ} hybridization with A. gambiae class I GST cDNAs or genomic clones highlighted a single region on chromosome arm 2R. The failure to detect any other region of complementarity with any of the probes indicated that the class I GST genes in \textit{A. gambiae} are arranged sequentially along chromosome 2R. As in \textit{Drosophila}, this region also contains pseudogenes as two 5’ truncated class I GST genes were found upstream of the full-length \textit{aggst1–1} gene.

REFERENCES

1. Migmogna, G., Allocati, N., Aceto, A., Piccolomini, R., Ilio, C. D., Barra, D., and Martini, F. (1993) Eur. J. Biochem. 211, 421–425
2. Wood, E., Casashe, N., Melgar, F., and Zebra, E. (1986) Comp. Biochem. Physiol. A 84B, 607–617
3. Bartling, D., Radzio, R., Steiner, U., and Weiler, E. W. (1993) Eur. J. Biochem. 216, 579–586
4. Clark, A. G., Dick, G. L., Martindale, S. M., and Smith, J. N. (1985) Insect Biochem. 15, 35–44
5. Chien, C., and Dauteraman, W. C. (1991) Insect Biochem. 21, 857–864
6. Hayes, J. D., and Pullford, D. J. (1995) Crit. Rev. Biochem. Mol. Biol. 30, 445–609
7. Clark, A. G., and Shamaan, N. A. (1984) Pest. Biochem. Physiol. 22, 249–261
8. Pembie, S. E., and Taylor, J. B. (1992) Biochem. J. 287, 957–963
9. Meyer, D. J., Coles, B., Pembie, S. E., Gilmore, K. S., Fraser, G. M., and Ketterer, B. (1991) Biochem. J. 274, 409–414
10. Meyer, D. J., and Thomas, M. (1995) Biochem. J. 311, 739–742
11. Fournier, D., Bride, J. M., Poirie, M., Berge, J.-B., and Plapp, F. W., Jr. (1992) J. Biol. Chem. 267, 1840–1845
12. Hemingway, J., Malcolm, C. A., Kisson, K. E., Boddington, R. G., Curtis, C. F., and Hill, N. (1985) Pest. Biochem. Physiol. 24, 68–76
13. Prapanthadara, L., Hemingway, J., and Ketterman, A. J. (1995) Bull. Ent. Res. 85, 267–274
14. Prapanthadara, L., Hemingway, J., and Ketterman, A. J. (1993) Pest. Biochem. Physiol. 47, 119–133
15. Favia, G., Dimopoulos, G., Della Torre, A., Toure, Y. T., Coluzzi, M., and Louis, C. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 10315–10319
16. Vaughan, A., Rodriguez, M., and Hemingway, J. (1985) Biochem. J. 215, 651–658
17. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory, Cold Spring Harbor, New York
18. Chang, P. K. and Natori, M. (1994) J. Bacteriol. 136, 247–351
19. Bradford, M. M. (1976) Anal. Biochem. 72 248–254
20. Habig, W. H., Pabst, M. J., and Jakoby, W. B. (1974) J. Biol. Chem. 249 7130–7139
21. Kumar, V., and Collins, F. H. (1994) Insect Mol. Biol. 3, 41–47
22. Wilce, M. C. J., Board, P. G., Feil, S. C., and Parker, M. W. (1995) EMBO J. 14, 2133–2143
23. Tang, A. H., and Tu, C.-P. D. (1994) J. Biol. Chem. 269, 27876–27884
24. Teung, Y.-P. S., Hsieh, T., and Tu, C.-P. D. (1993) J. Biol. Chem. 268, 9737–9746
25. Reiss, R. A., and James, A. A. (1993) Insect Biochem. Mol. Biol. 2, 25–32
26. Syvanen, M., Zhou, Z., and Wang, J. (1994) Mol. & Gen. Genet. 245 25–31
27. Teung, Y.-P. S., Hsieh, T.-S., and Tu, C.-P. D. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 31–35
28. Board, P., Russell, R. J., Marano, R. J., and Oakeshott, J. G. (1994) Biochem. J. 299, 425–430
29. Beall, C., Fyrberg, C., Song, S., and Fyrberg, E. (1992) Biochem. Genet. 30, 515–527
30. Snyder, M. J., Walding, J. K., and Feyereisen, R. (1995) Insect Biochem. Mol. Biol. 25, 455–465