Final Report TRF BRG/18/2541

Period of Funding 3 Years (30 September 1998 – 29 September 2001)

The Organization and Regulation of Class I Glutathione S-Transferase (GST) Genes in the Thai Malaria Vector *Anopheles dirus*. TRF BRG/18/2541

Associate Professor Albert J. Ketterman Ph.D.
Institute of Molecular Biology and Genetics (IMBG), Mahidol University, Salaya Campus
Tel: 02-4419003-7 ext. 1279 Fax: 02-4419906
Email: frakt@mahidol.ac.th

Final Report TRF BRG/18/2541

Period of Funding 3 Years (30 September 1998 – 29 September 2001)

The Organization and Regulation of Class I Glutathione S-Transferase (GST) Genes in the Thai Malaria Vector *Anopheles dirus*. TRF BRG/18/2541

Associate Professor Albert J. Ketterman Ph.D.
Institute of Molecular Biology and Genetics (IMBG), Mahidol University, Salaya Campus
Tel: 02-4419003-7 ext. 1279 Fax: 02-4419906
Email: fraki@mahidol.ac.th

Acknowledgements

The progress achieved on this project was possible because of the efforts and contributions from the following people:

Institute of Molecular Biology and Genetics, Mahidol University
Technical help:
Jantana Wongsantichon, Umnaj Chanama, Somphop Leetachewa, Chanikarn Boonchauy, Assist. Prof, Chartchai Krittanai.

Ph.D Students:
Saengtong Pongjaroenkit (now completed, funded by the Royal Golden Jubilee), Rungrutai Udomsinprasert (funded by the Royal Golden Jubilee), Jeerang Wongtrakul (funded by the Royal Golden Jubilee)

M.Sc. Students:
Peerada Prommeenate (now completed), Kanya Jirajaroenrat (now completed), Thasaneeya Harnnoi (now completed), Satjana Pattanasak, Ardcharaporn Vararattanavech

B.Sc. Student
Katila Varivarn

Chiang Mai University
Dr. L. Prapanthadara and N. Promtet

University of Western Australia, Australia
Assist. Prof. M.C.J. Wilce and Dr. A.J. Oakley

Table of Contents

Publications from Basic Science Grant BRG/18/2541

1.Ketterman,A.J., Prommeenate,P., Boonchauy,C., Chanama,U., Leetachewa,S., Promtet,N., and Prapanthadara,L. (2001). Single amino acid changes outside the active site significantly affect activity of glutathione S-transferases. Insect Biochem. Molec. Biol. *31*, 65-74.

2.Pongjaroenkit,S., Jirajaroenrat,K., Boonchauy,C., Chanama,U., Leetachewa,S., Prapanthadara,L., and Ketterman,A.J. (2001). Genomic organization and putative promotors of highly conserved glutathione S-transferases originating by alternative splicing in *Anopheles dirus*. Insect Biochem. Molec. Biol. *31*, 75-85.

3.Jirajaroenrat,K., Pongjaroenkit,S., Krittanai,C., Prapanthadara,L., and Ketterman,A.J. (2001). Heterologous expression and characterization of alternatively spliced glutathione S-transferases from a single *Anopheles* gene. Insect Biochem. Molec. Biol. *31*, 867-875.

4.Oakley,A.J., Jirajaroenrat,K., Harnnoi,T., Ketterman,A.J., and Wilce,M.C.J. (2001). Crystallization of two glutathione S-transferases from an unusual gene family. Acta Crystal. D *D57*, 870-872.

5.Udomsinprasert,R. and Ketterman,A.J. (2001). Expression and Characterization of a Novel Class of Glutathione S-Transferase from *Anopheles dirus*. Insect Biochem. Molec. Biol. *In press*.

The Organization and Regulation of Class I Glutathione S-Transferase (GST) Genes in the Thai Malaria Vector *Anopheles dirus*. TRF BRG/18/2541

Glutathione S-transferases (GSTs; EC 2.5.1.18) are a multigene family of multifunctional dimeric proteins that play a central role in detoxication. This detoxication function allows GSTs to contribute to the insecticide resistance problem occurring in mosquito malaria vectors. The main aim of this project was to identify and characterize the organization and regulation of GST genes in the Thai malaria vector *Anopheles dirus*. This involved obtaining GST genomic sequence and identifying putative regulatory elements that could control gene expression. Supplementary goals included heterologous expression of the GST genes obtained and characterization of those recombinant enzymes. The enzyme characterization included substrate specificity as well as inhibitor kinetic studies.

During this Project we obtained genomic sequence for three distinct GST genes. Two of these genes appear to each code for a single protein. The third gene coded for four different proteins through an alternative splicing mechanism. These six proteins from the three genes were heterologously expressed and the recombinant enzymes characterized. In addition, two of these proteins have now been crystallized and the respective structures solved. The above work is described in the following five publications that constitute this report.

Pergamon

Insect Biochemistry and Molecular Biology 31 (2001) 65-74

Insect Biochemistry and Molecular Biology

www.elsevier.com/locate/ibmb

Single amino acid changes outside the active site significantly affect activity of glutathione S-transferases

Albert J. Ketterman [a,*], Peerada Prommeenate [a], Chanikarn Boonchauy [b], Umnaj Chanama [b], Somphop Leetachewa [a], Nongkran Promtet [c], La-aied Prapanthadara [c]

[a] Institute of Molecular Biology and Genetics, Mahidol University, Salaya Campus, Nakorn Pathom 73170, Thailand
[b] Institute of Science and Technology for Research and Development, Mahidol University, Salaya Campus, Nakorn Pathom 73170, Thailand
[c] Research Institute for Health Sciences, Chiang Mai University, Chiang Mai 50200, Thailand

Received 6 March 2000; received in revised form 26 May 2000; accepted 1 June 2000

Abstract

Glutathione S-transferases (GSTs: E.C. 2.5.1.18) are a multigene family of multifunctional dimeric proteins that play a central role in detoxication. Four allelic forms of the mosquito *Anopheles dirus* GST, adGST1-1, were cloned, expressed and characterized. The one or two amino acid changes in each allelic form was shown to confer different kinetic properties. Based on an available crystal structure, several of the residue changes were not in the putative substrate-binding pocket. Modeling showed that these insect Delta class GSTs also possess a hydrophobic surface pocket reported for Alpha, Mu and Pi class GSTs. The atom movement after replacement and minimization showed an average atom movement of about 0.1 Å for the 0 to 25 Å distance from the alpha carbon of the single replaced residue. This does not appear to be a significant movement in a static modeled protein structure. However, 200–500 atoms were involved with movements greater than 0.2 Å. Dynamics simulations were performed to study the effects this phenomenon would exert on the accessible conformations. The data show that residues affecting nearby responsive regions of tertiary structure can modulate enzyme specificities, possibly through regulating attainable configurations of the protein.

Keywords: Glutathione transferase; Mosquito; Molecular modeling; Molecular dynamics; Allelic isoenzymes; Structure–function study

1. Introduction

Glutathione S-transferases (GSTs, EC 2.5.1.18) are ubiquitous enzymes that constitute a major cellular detoxication mechanism (for a recent review see Hayes and Pulford, 1995). GSTs catalyze the nucleophilic attack of glutathione on the electrophilic center of a range of endogenous or xenobiotic hydrophobic molecules. The GSTs are a superfamily of currently nine major classes with >50% amino acid sequence identity among the class members. Not only is there polymorphism within each class, but the individual GST genes also occur as multiple alleles (Pemble et al., 1994; Widersten et al., 1991; Board, 1981). The possession of specific variants of GST has been shown to correlate with tumor cell resistance to chemotherapy and predisposition to cancer in humans and insecticide resistance in insects (Daniel, 1993; Tsuchida and Sato, 1992; Hayes and Pulford, 1995).

Currently there are nine classifications of cytosolic GSTs although more groupings are possible as shown by analysis of the known sequences (Snyder and Maddison, 1997). These include the four familiar mammalian classes Alpha, Mu, Pi, Theta (Mannervik et al., 1992) and the new classes of Kappa (Pemble et al., 1996) and Zeta (Board et al., 1997). Generally, GSTs within a class have ≥50% amino acid sequence identity and between classes <30% identity (Mannervik et al., 1992). The other classes include the Sigma in cephalopods and arthropods, the Phi in plants and the Delta class in insects (Board et al., 1997). Previously the Delta class in insects has been referred to as Theta class (Pemble and Taylor, 1992). In addition, although named Theta class at the time, crystal structure has been solved for a

* Corresponding author. Fax: +66-2-441-9906.
E-mail address: frakt@mucc.mahidol.ac.th (A.J. Ketterman).

0965-1748/01/$ - see front matter
PII: S0965-1748(00)00106-5

Delta class GST from the Australian sheep blowfly, *Lucilia cuprina* (Wilce et al., 1995). The availability of the crystal structure of this GST has allowed us to perform homology modeling with our GSTs from the mosquito malaria vector *Anopheles dirus*. These GSTs are 66% identical to the *L. cuprina* amino acid sequence. Although primary amino acid sequence varies considerably between the GST classes there are highly conserved residues (Mannervik et al., 1985; Dirr et al., 1994a). There also appears to be a GST fold for the tertiary structure that allows comparison between the classes (Reinemer et al. 1992, 1996; Dirr et al. 1994b, 1991; Xiao et al., 1998; Ji et al., 1995; Krengel et al., 1998; Wilce et al., 1995; Dirr et al., 1994a). We have isolated several cDNA sequences that appear to be allelic variants of the single gene locus for the previously reported *adgst* 1-1 (Prapanthadara et al., 1998). Among these recombinant GSTs the sequence variation is only one or two amino acids. However, there are significant effects on the properties of the enzymes. Therefore, some of the amino acid changes influence enzyme specificity through structural effects.

2. Materials and methods

2.1. Mosquito strains

The Department of Parasitology, Faculty of Medicine, Chiangmai University, established the *Anopheles dirus* (species B) colony that was used in this study. The colony was identified as species B on the basis of its morphological and chromosomal characteristics (Baimai, 1989).

2.2. RNA extraction and PCR amplification

Approximately 200–400 mg of 4th instar larvae of *Anopheles dirus* (species B) was snap frozen in liquid nitrogen and homogenized using a mortar and pestle to a powder. The RNA was extracted by using the RNeasy Mini Kit (QIAGEN) as described in the manufacturer's instructions. The total RNA was eluted in 30–50 l of DEPC-treated water. The RNA solution was stored at −70 C for amplification by the RT-PCR technique. The primers used in these experiments were designed based on the 5′ and 3′ sequence of *An. dirus*, *adgst* 1-1 (Prapanthadara et al., 1998). Two sets of primers were used for cloning into the different vectors, pUC18 was used as the cloning vector and pET3a (Novagen) was used as the expression vector.

For the first strand cDNA synthesis, Superscript™ II RNase H⁻ Reverse Transcriptase (Gibco BRL) was used. Following the manufacturer's instructions for a 20 l reaction, 5 g of total RNA was used. After the cDNA synthesis, 5 l of the Superscript II reaction mixture was used as the template for the PCR reaction. For cloning into the pUC18 vector, the PCR was performed using [5′-CCGGCGGGATCCATGGATTTTTATT-ACCTA-CCC-3′] as the 5′ primer and [5′-CCGGCGGTCGACG-GACATGAATTTCGCCTTGAA-3′] as the 3′ primer. The 50 µl PCR reaction consisted of 50 pmole of each forward and reverse primer, 1×Vent® DNA Polymerase buffer (20 mM Tris–HCl (pH 8.8), 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% (v/v) Triton X-100), 200 µM each of dATP, dGTP, dCTP and dTTP, 1 unit of Vent® DNA Polymerase. Amplification, 40 cycles, was carried out after the initial denaturation step at 94°C for 5 min. The PCR reaction consisted of the following steps: the denaturing step performed at 94°C for 30 s, followed by the annealing step at 45°C for 1 min and then the extension step at 72°C for 1 min. Each reaction also had a final extension step at 72°C for 7 min. The PCR products were purified with a QIAquick Gel Extraction Kit (Qiagen). After purification, the products were double digested with *Bam*HI (GIBCO BRL) and *Sal*I (Promega) and ligated into pUC18. The inserts were sequenced in both directions at least twice using an ABI Prism 377 DNA Sequencer (Perkin Elmer).

For subcloning into pET3a, PCR was used to obtain the *Nde* I restriction site on the GST inserts. The 5′ primer used was [5′-CGCGGGATCCATATGATGGAT TTTTATTA-CCTA-3′] and the 3′ primer was [5′-CG CGGGATCCATATGTTAGGACA-TGAATTTCGC-3′]. The PCR was performed in 50 l reactions consisting of 50 pmole of each forward and reverse primer, 1 cloned *Pfu* DNA Polymerase buffer (20 mM Tris–HCl (pH 8.8), 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% (v/v) Triton X-100, 100 g/ l nuclease-free BSA), 200 M each of dATP, dGTP, dCTP and dTTP, 2.5 units of Cloned Pfu DNA Polymerase (Stratagene). Amplification, 40 cycles, was carried out after the step of an initial denaturation at 94 C for 5 min. A cycle contained the following steps: the denaturing step, performed at 94 C for 30 s; followed by the annealing step at 65 C for 1 min; and finally, the extension step at 72 C for 1 min. Each reaction also had a final extension step at 72 C for 7 min. The PCR products were digested and ligated into pET3a vector at the *Nde* I site. The inserts were sequenced at least twice in both directions to confirm that no changes had been introduced by the PCR. The pET3a clones were transformed into *E. coli* BL21(DE3)pLysS competent cells for protein expression. A colony of *E. coli* BL21(DE3)pLysS which contained a recombinant plasmid was grown at 37 C until the $O.D._{600}$ was approximately 0.6. After induction with 0.1 mM IPTG (final concentration) for 3 h, the cells were placed on ice for 20 min, and collected by centrifugation at 5000 rpm, 4 C for 10 min. The cell pellets from 50 ml of culture were suspended by mixing with 4.8 ml of 50 mM Tris–HCl (pH 7.4), 1 mM EDTA, 200 l of 100 mg/ml lysozyme and 3.6 l of 1.4 M -

mercaptoethanol. The suspension was incubated on ice for 20 min, then 50 l of 1 M DTT was added and the suspension lysed at 900 p.s.i. in a French Pressure cell. The lysate was then centrifuged at 10,000*g*, 4 C for 20 min. The supernatant containing the soluble form of the recombinant protein was separated from the pellet.

2.3. Protein purification method

The soluble target protein was purified at 4 C by using S-hexylglutathione immobilized on agarose (Sigma). The soluble target protein was mixed with the S-hexylglutathione agarose gel and incubated on ice for 5–10 min. The gel was then washed 6 with 50 mM Tris–HCl (pH 7.4) containing 0.2 M NaCl and 1 mM EDTA. The GST recombinant protein was eluted from the S-hexylglutathione agarose gel with 4 gel volumes of 50 mM Tris–HCl (pH 7.4), 0.2 M NaCl, 1 mM EDTA, containing 5 mM S-hexylglutathione and 10 mM DTT. After a concentration step using Centriprep-10 (Amicon), S-hexylglutathione that was bound to the recombinant GST was eliminated by gel filtration using PD-10 columns (Pharmacia) equilibrated with 50 mM phosphate buffer pH 6.5. The concentrated supernatant containing the target protein was applied to the column and the purified protein eluted in the same buffer supplemented with 10 mM DTT. This S-hexylglutathione elimination step was performed twice. All the steps were carried out at 4 C. Glycerol was then added to a final concentration of 40% (v/v) and the purified concentrated GSTs stored at −20 C.

2.4. Enzyme and protein assays

The GST activity assays were performed as previously described (Prapanthadara et al., 1996). Protein concentration was determined by the method of (Bradford, 1976) using the Bio Rad protein reagent with bovine serum albumin as the standard protein.

2.5. Molecular modeling

Coordinates of the dimeric *Lucilia cuprina* crystal structure, including a glutathione molecule in each of the two active sites and 256 water molecules, were used for homology modeling of the *Anopheles dirus* protein. The amino acid sequences of the two proteins, *Lucilia cuprina* GST and *Anopheles dirus* GST1-1, are 66.7% identical. Modeling of the tertiary structure was performed using the InsightII software suite including the Biopolymer and the Discover modules (Biosym/Molecular Simulations Inc. of San Diego, CA). The model structure was minimized using the CVFF force field initially for several thousand iterations followed by minimization using the class II force field CFF91. The minimization algorithm Steepest Gradient was employed until the average absolute derivative was 0.003. Then the Conjugate Gradient algorithm was employed until convergence with the average absolute derivative of 1.9e-6; a r.m.s. of 2.7e-6 and the maximum absolute derivative of 4.3e-5. The amino acid at position 188 was replaced and the model structure was minimized using the CFF91 force field and the Conjugate Gradient algorithm until convergence with an average absolute derivative of 1.1e-6; a r.m.s. of 1.6e-6 and the maximum absolute derivative of 1.8e-5. Also the amino acid at position 25 was replaced and the model was minimized as above until convergence with an average absolute derivative of 1.0e-6; a r.m.s. of 1.5e-6 and the maximum absolute derivative of 2.7e-5. Dynamics simulation of the three models was performed for 14 to 15 picoseconds. An integration timestep of 1 fs, a temperature of 300 K, and the same random number seed were used for all three models. No constraints were employed.

3. Results

The PCR primers used in this study were for the 5′ and 3′ termini of the coding sequence for *adgst 1-1*. The 5′ terminus sequence is highly conserved within the Delta class, however the 3′ terminus is highly variable and will be specific for the individual GST within the class. Previously we have shown DDT resistant and susceptible strains of *An. gambiae* to possess GST isoforms with different enzyme properties (Prapanthadara et al., 1993). However the *An. dirus* colony used in the present study was not selected for insecticide resistance and would be a more heterogeneous population. Approximately 70–80 larvae would have been used in the RNA isolation. We therefore expected to obtain allelic variations of the *adgst 1-1* sequence. We were interested in observing and studying how small changes in sequence might affect enzyme characteristics. All the GST recombinant enzymes in the present study originated from cloning the PCR product from a single RT-PCR. Upon cloning the PCR product into pUC18, 85 white colonies were screened by restriction digest analysis. From this analysis, 43 clones were obtained that were positive for a GST insert in the proper orientation. From these 43 GST clones, 17 were sequenced and 11 clones were found to code for an identical amino acid sequence, although there were a total of six base changes in the nucleic acid sequence of four clones. The remaining six clones contained one or two amino acid residues that were changed relative to the sequence of the 11 clones. The clone for "GST Base" was chosen as the baseline sequence from among the 11 clones with identical amino acid sequence. GST Base was used as the standard amino acid sequence for purposes of comparison. GST Base, as well as the six clones coding for amino acid changes were subcloned into the pET3a expression vec-

tor. All of the GST sequences were inserted into the *Nde* I cloning site of the pET3a vector and were expressed without the pET3a 11 amino acid N-terminal fusion tag that is present on expressed products cloned into the *Bam*HI site.

The nucleic acid and translated amino acid sequence for GST Base is shown (Fig. 1). We propose this sequence is an allelic variant of the previously reported *adgst* 1-1 (Prapanthadara et al., 1998). The initial *adgst* 1-1 clone as well as the clones reported in the present study have been obtained from the same *An. dirus* colony. However, there was approximately three years between the RNA isolations and the colony gene pool has shifted with respect to GSTs. We have also obtained genomic sequence from the current colony that is identical to the GST Base sequence and not to the initial *adgst* 1-1. GST Base, as well as the other clones, was obtained using PCR primer sequence that had originally yielded the *adgst* 1-1. A comparison between GST Base sequence and *adgst* 1-1 show that there are only 10 nucleotide changes, some of which code for the five

```
          *         20         *         40         *
ATGGATTTTTATTACCTACCCGGATCTGCGCCGTGCCGTGCCGTTCAGATGACGGCG
 M  D  F  Y  Y  L  P  G  S  A  P  C  R  A  V  Q  M  T  A

 60         *         80         *        100         *
GCCGCCGTTGGCGTCGAGCTGAACCTGAAGCTCACCAACCTGATGGCGGGCGAGCAC
A  A  V  G  V  E  L  N  L  K  L  T  N  L  M  A  G  E  H

   120         *        140         *        160         *
ATGAAGCCGGAATTCCTGAAGCTTAACCCCCAGCACTGCATTCCGACGCTGGACGAC
M  K  P  E  F  L  K  L  N  P  Q  H  C  I  P  T  L  D  D

     180         *        200         *        220
AACGGTTTCTCGCTGTGGGAGTCGCGCGCCATCCAGATCTATCTGGTCGAGAAGTAC
N  G  F  S  L  W  E  S  R  A  I  Q  I  Y  L  V  E  K  Y

 *       240         *        260         *        280
GGCAAGGACGACAAGCTGTACCCGAAGGACCCCCAGAAGCGCGCCGTCGTCAACCAG
G  K  D  D  K  L  Y  P  K  D  P  Q  K  R  A  V  V  N  Q

    *        300         *        320         *        340
CGGCTGTTCTTCGACATGGGCACGCTGTACCAGCGCTTCGGGGACTACTGGTACCCG
R  L  F  F  D  M  G  T  L  Y  Q  R  F  G  D  Y  W  Y  P

       *        360         *        380         *
CAGATCTTCGCCAAGCAGCCGGCCAACGCGGAGAACGAGAAGAAGATGAAGGAGGCG
Q  I  F  A  K  Q  P  A  N  A  E  N  E  K  K  M  K  E  A

)0         *        420         *        440         *
GTCGGCTTCCTGAACACGTTCCTCGAGGGCCAGGAGTACGCGGCCGGCAGTGATCTC
V  G  F  L  N  T  F  L  E  G  Q  E  Y  A  A  G  S  D  L

 460         *        480         *        500         *
ACGATCGCCGATCTCAGTCTGGCGGCGTCGATCGCCACCTACGAGGTGGCCGGGTTC
T  I  A  D  L  S  L  A  A  S  I  A  T  Y  E  V  A  G  F

    520         *        540         *        560         *
GACTTCGCCCCCTACCCGAACGTGGCCGCGTGGCTGGCCCCGCTGCAAGGCGAACGCC
D  F  A  P  Y  P  N  V  A  A  W  L  A  R  C  K  A  N  A

       580         *        600         *        620
CCCGGCTATGCCCTCAACCAGGCCGGTGCCGATGAATTCAAGGCGAAATTCATGTCC
P  G  Y  A  L  N  Q  A  G  A  D  E  F  K  A  K  F  M  S

  *
TAA  :  630
```

Fig. 1. Nucleic acid and translated amino acid sequence for *Anopheles dirus* GST 1-1 base line sequence. This sequence is "GST Base" in this report.

Table 1
Amino acid sequence identity for *Anopheles* GSTs[a]

	GST Base	Ad1-1	Ag1-6	Ag1-5	Ag1-4	Ag1-3	Ag1-2	Ag1-7
GST Base	100	97	93	77	64	62	50	46
Ad1-1		100	91	76	63	60	48	45
Ag1-6			100	82	66	63	50	44
Ag1-5				100	64	64	50	41
Ag1-4					100	60	47	41
Ag1-5						100	49	38
Ag1-2							100	36
Ag1-7								100

[a] The abbreviations used are: GST Base sequence from present study (AF 273041); Ad1-1 for *An. dirus* 1-1 (U78784); Ag for *An. gambiae*, Ag1-6 for AgGST 1-6 (2842718); AgGST 1-5 (2842717); AgGST 1-4 (3549274); AgGST 1-3 (071163); AgGST 1-2 (1495237) and AgGST 1-7 (3549272). The numbers shown in parentheses are the GenBank Accession Numbers.

amino acid changes. These few changes entail a high percent amino acid sequence identity for GST Base and AdGST 1-1 as well as with AgGST 1-6, the orthologous GST in *An. gambiae* (Table 1). As shown by a comparison of AgGST 1-6 with the other reported *An. gambiae* GSTs, the high percent identity of GST Base with AdGST 1-1 supports the allelic variant model. In addition to the GST Base sequence, which is used as the baseline sequence in the present study, six other allelic sequences were obtained from the same PCR reaction as GST Base. The nucleic acid and amino acid residue changes for these other GST sequences are shown in Table 2. It was observed that although clones 7, 36 and 41 expressed the recombinant protein to the same extent as the other clones no GST 1-chloro-2,4-dinitrobenzene (CDNB) activity was detectable in the *E. coli* lysate. These clones were not studied further but the remaining

Table 2
Nucleic acid and amino acid residue changes for the recombinant GSTs relative to "GST Base" (see Fig. 1)

GST	Nucleic Acid	Amino Acid
GST A188T	C153T	C51
	G562A	A188T
GST G109W	C153T	C51
	G325T	G109W
GST E25Q	G73C	E25Q
	C348T	I116
	G459C	T153
Clone 7	G152A	C51Y
	G424C	E142Q
	G486C	S162
Clone 36	A457G	T153A
	C547A	L183M
Clone 41	G40A	A14T
	C112T	H38Y

Table 3
Summary of the kinetic constants for the four recombinant allelic GSTs from *An. dirus*[a]

Constant	GST Base	GST A188T	GST G109W	GST E25Q
V_m	12.9 0.63	31.8 1.06	8.65 0.53	28.5 2.00
$K_{m\ CDNB}$	0.104 0.028	0.104 0.030	0.156 0.054	0.172 0.046
$K_{m\ GSH}$	0.858 0.179	0.755 0.111	0.877 0.228	0.734 0.103
k_{cat}	5.03	12.4	3.39	11.1
$k_{cat}/K_{m\ CDNB}$	48.4	119	21.9	64.6
$k_{cat}/K_{m\ GSH}$	5.86	16.4	3.87	15.1

[a] The units used in the table are: V_m: mol/min/mg; K_m: mM; K_{cat} s^{-1}; k_{cat}/K_m: $s^{-1}.mM^{-1}$. The data are the mean standard error of at least three separate experiments.

recombinant enzymes were purified and further characterized. These four clones included the baseline sequence clone, GST Base; a clone containing a single amino acid change of A188T, GST A188T; a clone containing a single amino acid change of G109W, GST G109W; and a clone containing a single amino acid change of E25Q, GST E25Q.

Using the substrates CDNB and GSH to determine kinetic constants demonstrates that the four different recombinant GSTs possess different V_m and K_m (Table 3). The V_m and K_m for both CDNB and GSH appeared similar in a pairwise comparison of GST A188T and GST E25Q or GST Base and GST G109W. Between the two pairs the V_m was significantly different with GST A188T and GST E25Q possessing a 2–3-fold greater V_m. However, considering k_{cat}/K_m highlights the substrate specificity differences of all four of the enzymes.

Using DDT as the electrophilic substrate another pattern of specificity was observed (Table 4). The ratio of CDNB activity to DDTase activity was similar for GST A188T and GST E25Q. Recombinant GST Base had the lowest DDTase activity with an apparent relative preference for CDNB substrate. Of particular interest is the recombinant GST G109W with a relatively high DDTase activity and a relatively low CDNB activity to DDTase ratio. This demonstrates that the single residue change G109W is important for interaction with DDT.

Testing the interaction of other insecticides showed differences amongst the four recombinant GSTs (Table 5). GST A188T was only completely inhibited by chlorpyrifos, whereas GST G109W still retained 26% of its activity. In general, GST G109W appeared to have the lowest degree of interaction with the studied insecticides. This suggests that G109W is important for interaction with these types of compounds although other amino acids do have a role in modulating the enzyme interaction as shown by the data for the other recombinant GSTs. The data suggest that GST Base and GST G109W bind DDT to a similar extent thereby inhibiting CDNB activity. However, GST G109W catalyses dehydrochlorination at a 5-fold greater rate than GST Base. The interactions of the isoenzymes with the insecticides appear to parallel the pairing similarity shown by V_m. That is, GST A188T is similar to GST E25Q and GST Base is similar to GST G109W. In addition, the interactions of GST Base and GST G109W with the studied compounds are less than those of GST A188T and GST E25Q.

Table 4
DDTase activity for the expressed recombinant GSTs[a]

GST	DDTase activity nmol/mg	Units GST/DDTase
Base	0.95	3.0
A188T	2.77	1.9
G109W	4.83	0.4
E25Q	1.20	1.6

[a] Units of GST activity are with 1 mM CDNB and 10 mM glutathione as substrates.

The IC_{50} was determined for three GST inhibitors with the four recombinant allelic GSTs (Table 6). Although it was difficult to show statistical significance in all cases, there appeared to be differences in the interaction for some of the recombinant enzymes with the tested inhibitors. All four recombinant GSTs showed simple linear inhibition in plots of the data for the three inhibitors (data not shown). This type of plot can be used to distinguish between single and multiple sites of inhibitor interaction (Tahir and Mannervik, 1986).

As shown by the tertiary structure model, the positions of the amino acid residues 25 and 188 are outside of the putative substrate-binding pocket (Fig. 2). These observations of the single amino acid changes conferring different kinetic properties to the recombinant GSTs prompted us to model how these changes affect tertiary structure. Within the modeled tertiary structure the amino acids are found in the following secondary structures: residues 1–6 in β_1 sheet; residues 9–23 in α_1 helix; residues 26–31 in β_2 sheet; residues 40–47 in α_2 helix; residues 53–58 in β_3 sheet; residues 64–76 in α_3 helix; residues 86–119 in α_4 helix; residues 123–142 in α_5 helix; residues 153–170 in α_6 helix and residues 176–190 in α_7 helix. After the residue 188 replacement and minimization, the dihedral of the alpha carbon backbone of helix 7 where residue 188 is located, was changed by 1.25 Å. Relative to the alpha carbon of residue 188, the average atom movement in 209 atoms in the helix was 0.10 Å but with movement in 23 atoms of 0.2 Å to 0.43 Å. Residues of helix 7 interact with residues of both helices 5 and 6 through hydrogen bonding and van der Waals contact. These changes in helix 7 were reflected in a dihedral change of the alpha carbon backbone of 0.55 Å in helix 6 and 0.46 Å in helix 5. However, residue 25 is not within a helix therefore within this model no dihedral changed more than 0.17 Å.

Superimposing 1608 atoms in aligned positions of the

Table 5
Insecticide inhibition of CDNB activity of the four recombinant allelic GSTs from *An. dirus*[a]

Insecticide	% Inhibition			
	GST Base	GST A188T	GST G109W	GST E25Q
DDT	51.4±29.4	86.6±1.6	66.6±13.5	71.0±8.1
DDE	70.8±11.3	87.6±1.2	54.6±36.0	83.1±6.5
Lindane	44.7±11.2	69.1±7.3	55.7±9.6	75.0±5.2
Permethrin	45.6±30.3	88.2±2.2	50.4±19.0	77.7±9.4
□ cyhalothrin	30.7±18.8	77.3±16.7	36.1±17.8	60.7±3.8
Deltamethrin	46.3±14.0	78.6±3.3	26.3±5.9	76.0±4.4
Diazinon	47.6±24.1	86.6±12.4	43.8±14.5	89.3±9.9
Fenitrothion	74.4±14.9	86.7±4.0	64.4±21.3	92.1±6.9
Pirimiphos-methyl	80.9±26.9	90.6±3.7	65.6±15.4	76.7±31.6
Chlorpyrifos	81.5±16.6	98.3±2.4	74.1±8.4	96.2±1.7
Temephos	56.8±12.1	77.5±12.4	55.0±21.0	49.8±32.8
Bendiocarb	55.2±9.6	80.4±7.5	36.7±15.2	77.4±9.5

[a] Final concentrations of all insecticides were 100 □ M except diazinon which was 10 □ M. The GSH and CDNB concentrations were 10.0 and 1.0 mM, respectively. The data are the mean±standard deviation of at least three separate experiments, each of which was performed in duplicate.

Table 6
IC_{50} for the interaction of three GST inhibitors with the four recombinant allelic GSTs from *An. dirus*[a]

Compound	IC_{50} (□ M)			
	GST Base	GST A188T	GST G109W	GST E25Q
Bromosulphothalein	0.204±0.048	0.291±0.045	0.192±0.027	0.205±0.041
Cibracron Blue	0.416±0.256	0.237±0.014	0.107±0.009	0.242±0.043
Ethacrynic Acid	0.173±0.075	0.180±0.149	0.400±0.151	0.249±0.090

[a] The data are the mean±standard deviation of at least three separate experiments.

carbon backbone of the polypeptide chain before residue 188 replacement onto the polypeptide chain after residue 188 replacement had a r.m.s. deviation of 0.030. In the minimized polypeptide chains, atom movement at a distance of 0 Å to 25 Å relative to the alpha carbon of residue 188 was monitored with an average atom movement of 0.108 Å. Of the 1955 atoms included within this distance, 696 atoms moved from 0.1 to 0.2 Å and 249 atoms moved from 0.2 to 0.43 Å. This does not appear to be a significant movement in a static modeled protein structure.

Superimposing the baseline carbon backbone onto the minimized carbon backbone after E25Q replacement had a r.m.s. deviation of 0.053. Modeling the atom movement after replacement and minimization of E25Q showed an average atom movement of 0.149 Å for the 0 to 25 Å distance from the alpha carbon of residue 25. Of the 1732 atoms included within this distance 500 atoms moved from 0.1 to 0.2 Å and 514 atoms moved from 0.2 to 1.13 Å. This does not appear to be a significant movement in a static modeled protein structure.

In both models of GST A188T and GST E25Q, 200–500 atoms were involved with movements greater than 0.2 Å. Dynamics simulations were performed to study the effects this phenomena would exert on the conformational flexibility of the structures. Dynamics provides insight into the accessible conformational states of the molecule by modeling the time evolution of the system. Initial velocities are randomly generated at the beginning of a dynamics run according to the temperature and random number seed. To repeat the exact same dynamics run, it is important to specify the random number seed used in the previous run. This guarantees generating the same set of initial velocities in the current run (Biosym/Molecular Simulations Inc. of San Diego, CA). In analyzing the dynamics simulation data it became clear that parts of the GST flex significantly, especially in helix 4. Several residues of helix 4 are thought to directly interact with substrates in the active site (Wilce et al., 1995; Dirr et al., 1994a).

Animation of the trajectories from the dynamics simulations showed some change in the residues positioned around the previously proposed active site pocket. This pocket consists of residues from both subunits (for a crystal structure review see Dirr et al., 1994a). Plots of atom to atom distance spanning the active site pocket from subunit A to subunit B show little difference between the models (Fig. 3(A)). Additionally, as suggested in the static models, most of the helices dihedrals remained very similar except for helix 5 (Fig. 3(B)). The

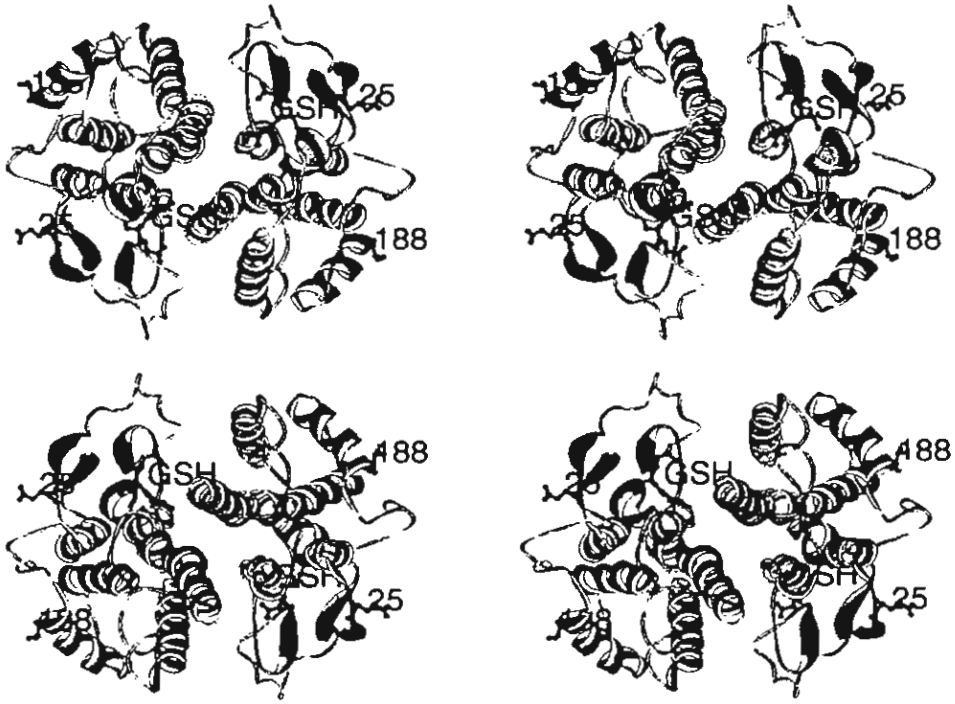

Fig. 2. Stereo views of GST showing the positions of the residues changed outside the active site. The top view is down the molecular two-fold axis of the assembled dimer with glutathione (GSH; shown in black) in the active site. The bottom view is down the molecular two-fold axis rotated 180° from the top view. Residues 25 and 188 are labeled and shown in black. The images were made with Weblab Viewer Pro 3.5.

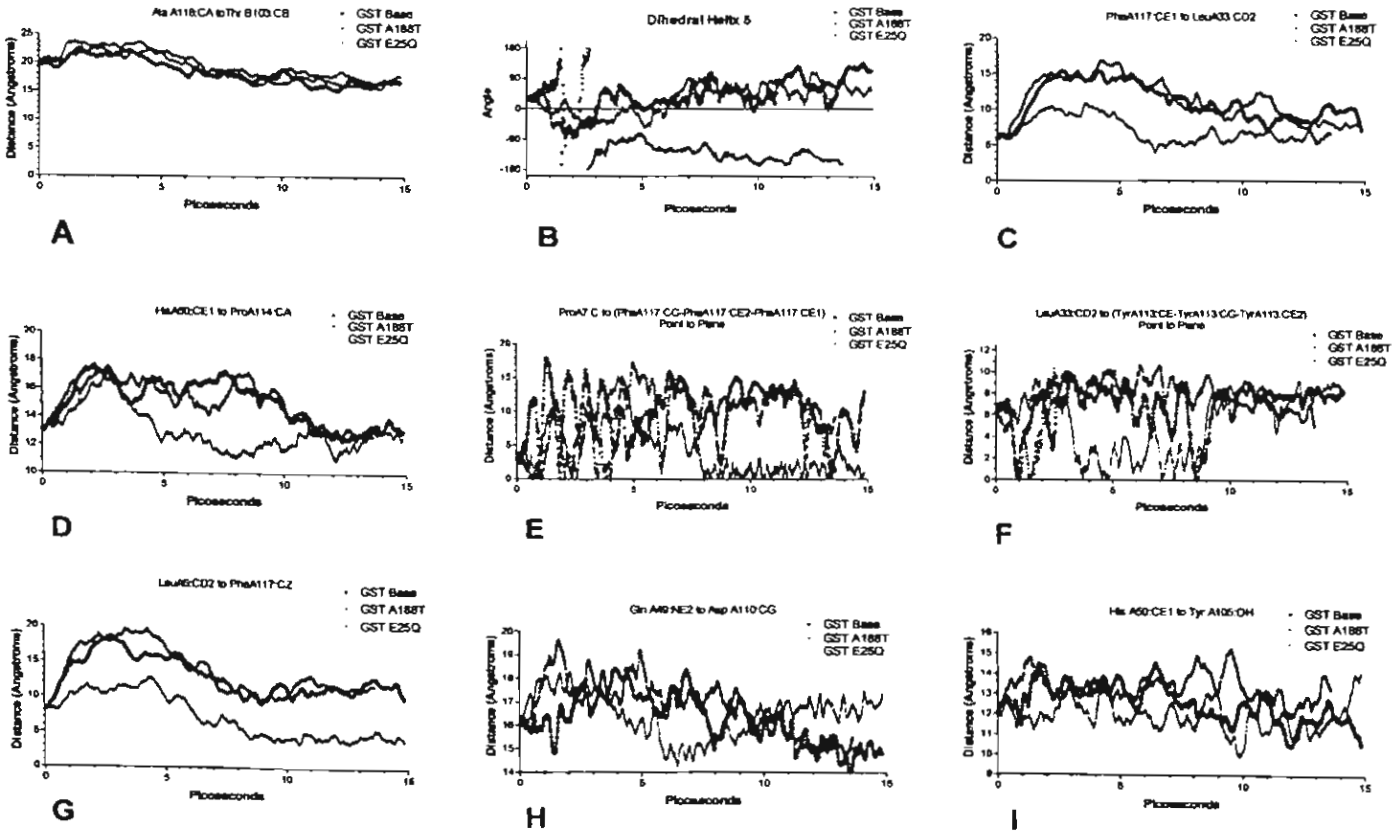

Fig. 3. Comparison of the trajectory data from the three GST models. Panels (A) and (C)–(I) show changes in the distances between the indicated atoms over the time of the dynamics simulation. Panel (B) shows changes in the dihedral of helix 5 over time.

dihedral of the alpha carbon backbone for helix 5 of GST A188T changed significantly compared to both GST Base and GST E25Q. Several residues of helix 5 contribute to the hydrophobic core formed between helix 5, 6 and 7. This core is thought to be involved in stabilizing GST tertiary structure as well as modulating specificity towards effector molecules such as inhibitors (Dirr et al., 1994a).

An observation of further interest is the identification of a second substrate pocket. This pocket has been recently reported as a hydrophobic surface pocket in Pi class GSTs (Oakley et al., 1999). This was shown also in several other GST classes but was unrecognized in the insect static crystal structure because the insect structure appeared to be in the closed position (Fig. 4). In *An. dirus* GST these residues are Leu6, Pro7, Gly8, Ala10, Pro11, Cys12, Leu30, Leu33, Ala35, Ile52, Met101, Tyr105, Phe108, Ile116, Phe117, Ala118, Val168 and Ala169. This pocket opens and closes with atom movements up to 10 Å (Fig. 3C–I). This pocket opening varies significantly in the models as shown in the trajectory plots (Figs. 3C–I and 4). The simulation data shows that the pocket opening varies between the models not only in relative distance but absolute distance (Fig. 3G) as well as the relative length of time the residues remain apart or in proximity (Fig. 3D–F).

4. Discussion

We have obtained cDNA sequences for GST from RT-PCR of mosquito larvae total RNA. We propose that these sequences represent allelic variations present in mosquito populations for the single GST gene *adgst* 1-1. The translated amino acid sequences for the clones obtained vary by only one or two amino acids. Heterologous expression and enzyme characterization of these recombinant proteins demonstrated kinetic differences between the forms. Using an available crystal tertiary structure, we observed that in two of the enzymes the changed amino acid residues were on the exterior surface of the protein. We then proceeded to model how amino acid residues outside the active site could affect enzyme kinetics.

Several of the recombinant enzymes had no detectable CDNB activity in *E. coli* lysates and were not studied further at this time. Clone 7 most likely had no detectable activity due to the change C51Y. The two residue

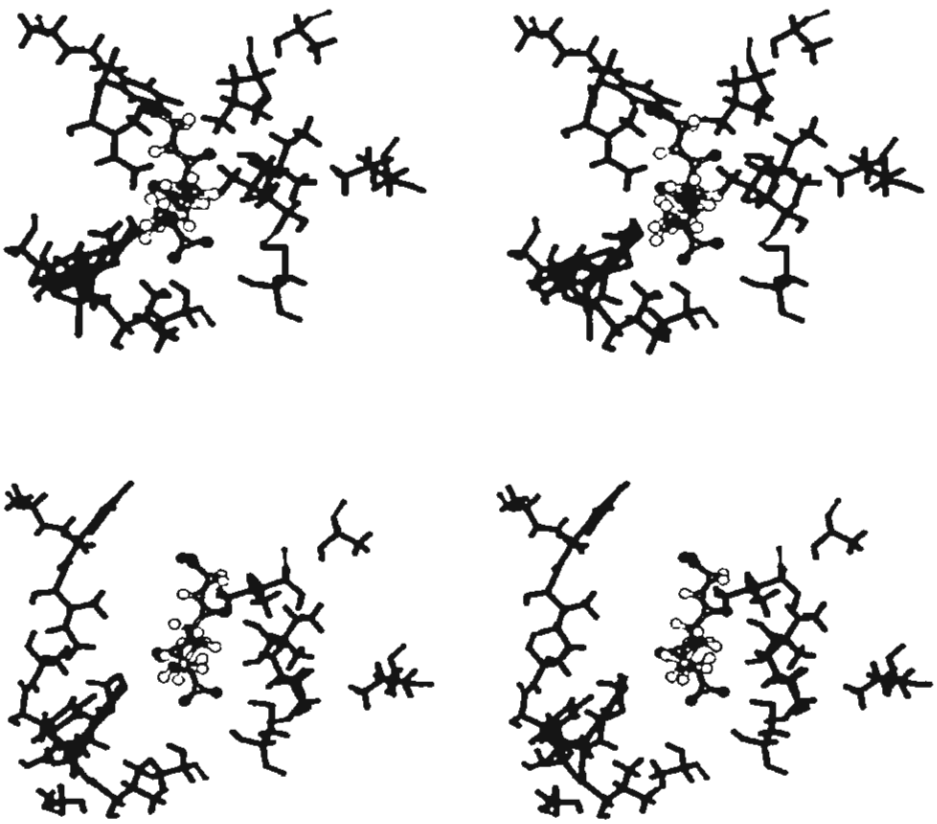

Fig. 4. Stereo view of the hydrophobic surface pocket in both closed and opened configurations. The top view is the closed configuration and the bottom view is the opened configuration. The glutathione in the active site is shown in ball and stick representation, the sulfur atom faces out of the page. The amino acid residues of GST are shown in a stick representation. The images were made with Weblab Viewer Pro 3.5 from the trajectory data generated by Insight II.

on either side of 51, His50 and Ile52, have been shown by crystal structure to directly interact with glutathione (Wilce et al., 1995). The loss of activity may be due to steric interference by the much larger tyrosine side chain as well as the concomitant adjustments in the neighboring residues. In clone 41, tyrosine replaced His38, which has been shown to directly interact with glutathione. Two residue changes for Clone 36, T153A at the N-terminus of helix 6 and L183M in helix 7, are both outside of the active site pocket. These changes involve two of the helices involved in formation of a hydrophobic core that is important for stabilizing GST tertiary structure (Dirr et al., 1994a). Structural changes to this hydrophobic core are thought to modulate specificity of GST substrate/inhibitor interactions (Manoharan et al., 1992b,a; Stenberg et al., 1991; Kong et al., 1992). The reason for the apparent loss of CDNB activity for Clone 36 needs to be further studied but one may speculate it will be for similar reasons as discussed below for the changes for GST A188T and GST E25Q.

The residue change for GST G109W is in the active site pocket in helix 4. This residue would have a direct interaction with the substrates and its effects are reflected in changes in specificity (Tables 3 and 4). In comparison to the GST Base enzyme the GST G109W k_{cat}/K_{mCDNB} decreased two-fold and the ratio of CDNB activity to DDT activity also decreased 7.5-fold.

GST A188T is near Arg185 which appears equivalent to an invariant arginine found in Alpha, Mu and Pi class GSTs (Dirr et al., 1994a). This arginine forms an important salt bridge necessary for structure stabilization in those classes. In an Alpha class GST, four arginines conserved across the GST classes were studied (Stenberg et al., 1991). These residues were thought to stabilize one of the conformational states for the enzyme. It was proposed that a shift in the relative distribution of different conformations might affect binding of substrates or ligands. The GST E25Q residue is located between helix 1 and β sheet 2 and appears to affect tertiary structure around the hydrophobic core in domain I. This hydrophobic core appears to be structurally important and changes affect stability, activity and binding affinities (Widersten et al., 1992; Stenberg et al., 1991; Kong et al., 1992, 1993; Manoharan et al., 1992a,b). Our data therefore suggests that different conformational states may also occur due to changing influences from neighboring residues of structurally important regions as well as amino acids.

A comparison of amino acid sequence for the *An. dirus* GST to the other GST classes Alpha, Mu and Pi shows less than 30% identity. However, the insect Delta class GST adopts the canonical GST fold (Wilce et al., 1995). To achieve this state some residues have been functionally or completely conserved. From crystal structure data 26 invariant amino acids were found in the Alpha, Mu and Pi classes (Dirr et al., 1994a). Many of these residues or conservative equivalents were found in the appropriate secondary structure positions in the *An. dirus* model. These residues would thus appear to fulfill similar functions. In *An. dirus*, 16 of these residues are Tyr5, Arg13 (equivalent to Arg11 in Pi), Pro53, Ile68, Val73 (Leu76 in Pi), Gly77, Leu104, Leu141, Gly149, Asp156 (next to a conserved Ala155 and three residues from a conserved Leu159), Ile163 (replacing Leu157 in Pi), Phe171, Val179 (for Leu174 in Pi) and Arg185. Further examination may identify even more functionally equivalent residues. This conservation of residues to maintain tertiary structure suggests that the enzymes of the different GST classes function in a similar manner, albeit with some class specific modulation.

In conclusion, we obtained recombinant GSTs for which we could determine differences in kinetic properties although the changed amino acids were outside of the active site. Atom movements to accommodate the replaced residues in models showed only small displacements. However, these small displacements occurred for hundreds of atoms. Dynamics simulation using these models showed differences existed in the configurations attained by each of the different models. The data demonstrate that single amino acid changes outside the active site and not directly involved in structural maintenance can influence the accessible conformational states that affect enzymatic properties. The conservation of tertiary structure as shown by the hydrophobic surface pocket, the invariant residues and the GST fold, suggests that these observations in the Delta class GST will also be applicable to the other GST classes. The implications of these data are that it will be possible to subtly modulate enzyme specificities by controlling the accessible conformational states of the protein through residues that affect sensitive regions of the tertiary structure. These data also demonstrate how single amino acid changes, either in or out of the active site, may change the properties of allelic forms of enzymes. These single amino acid changes would then confer insecticide resistance capabilities to individual members of a field population.

Acknowledgements

We thank Dr Michael W. Parker for kindly providing the coordinates of the *L. cuprina* glutathione transferase. This work was funded by the Thailand Research Fund (TRF).

References

Baimai, V., 1989. Speciation and species complexes of the anopheles malaria vectors in Thailand. In: The Third Conference on Malaria Research, Thailand, pp. 146–152.

Board, P., Baker, R.T., Chelvanayagam, G., Jermiin, L.S., 1997. Zeta,

a novel class of glutathione transferases in a range of species from plants to humans. Biochem. J. 328, 929–935.

Board, P.G., 1981. Biochemical genetics of glutathione S-transferase in man. Am. J. Hum. Genet. 33, 36–43.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.

Daniel, V., 1993. Glutathione S-transferases: gene structure and regulation of expression. CRC Crit. Rev. Biochem. Molec. Biol. 28, 173–207.

Dirr, H., Reinemer, P., Huber, R., 1994a. X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. Eur. J. Biochem. 220, 645–661.

Dirr, H., Reinemer, P., Huber, R., 1994b. Refined crystal structure of porcine class pi glutathione S-transferase (pGSTP1-1) at 2.1 Å resolution. J. Mol. Biol. 243, 72–92.

Dirr, H.W., Mann, K., Huber, R., Ladenstein, R., Reinemer, P., 1991. Class 0 glutathione S-transferase from pig lung. Purification, biochemical characterization, primary structure and crystallization. Eur. J. Biochem. 196, 693–698.

Hayes, J.D., Pulford, D.J., 1995. The glutathione S-transferase supergene family: Regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. CRC Crit. Rev. Biochem. Molec. Biol. 30, 445–600.

Ji, X., Von Rosenvinge, E.C., Johnson, W.W., Tomarev, S.I., Paitigorsky, J., Armstrong, R.N., Gilliland, G.L., 1995. Three-dimensional structure, catalytic properties, and evolution of a sigma class glutathione transferase from squid, a progenitor of the lens S-crystallins of cephalopods. Biochemistry 34, 5317–5328.

Kong, K.-H., Inoue, H., Takahashi, K., 1992. Site-directed mutagenesis of amino acid residues involved in the glutathione binding of human glutathione S-transferase P1-1. J. Biochem. 112, 725–728.

Kong, K.-H., Inoue, H., Takahashi, K., 1993. Site-directed mutagenesis study on the roles of evolutionally conserved aspartic acid residues in human glutathione S-transferase P1-1. Protein Engineering 6, 93–99.

Krengel, U., Schröter, K.-H., Hoier, H., Arkema, A., Kalk, K.H., Zimniak, P., Dijkstra, B.W., 1998. Crystal structure of a murine 0 -class glutathione S-transferase involved in cellular defense against oxidative stress. FEBS Lett. 422, 285–290.

Mannervik, B., Ålin, P., Guthenberg, C., Jensson, H., Tahir, M.K., Warholm, M., Jörnvall, H., 1985. Identification of three classes of cytosolic glutathione transferase common to several mammalian species: correlation between structural data and enzymatic properties. Proc. Natl. Acad. Sci. USA 82, 7202–7206.

Mannervik, B., Awasthi, Y.C., Board, P.G., Hayes, J.D., Di Ilio, C., Ketterer, B., Listowsky, I., Morgenstern, R., Muramatsu, M., Pearson, W.R., Pickett, C.B., Sato, K., Widersten, M., Wolf, C.R., 1992. Nomenclature for human glutathione transferases. Biochem. J. 282, 305–306.

Manoharan, T.H., Gulick, A.M., Puchalski, R.B., Servias, A.L., Fahl, W.E., 1992b. Structural studies on human glutathione S-transferase 0 . Substitution mutations to determine amino acids necessary for binding glutathione. J. Biol. Chem. 267, 18940–18945.

Manoharan, T.H., Gulick, A.M., Reinemer, P., Dirr, H.W., Huber, R., Fahl, W.E., 1992a. Mutational substitution of residues implicated by crystal structure in binding the substrate glutathione to human glutathione S-transferase 0 . J. Mol. Biol. 226, 319–322.

Oakley, A.J., Bello, M.L., Nuccetelli, M., Mazzetti, A.P., Parker, M.W., 1999. The ligandin (non-substrate) binding site of human Pi class glutathione transferase is located in the electrophile binding site (H-site). J. Mol. Biol. 291, 913–926.

Pemble, S., Schroeder, K.R., Spencer, S.R., Meyer, D.J., Hallier, E., Bolt, H.M., Ketterer, B., Taylor, J.B., 1994. Human glutathione S-transferase Theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. Biochem. J. 300, 271–276.

Pemble, S.E., Taylor, J.B., 1992. An evolutionary perspective on glutathione transferases inferred from class-Theta glutathione transferase cDNA sequences. Biochem. J. 287, 957–963.

Pemble, S.E., Wardle, A.F., Taylor, J.B., 1996. Glutathione S-transferase class kappa: characterization by the cloning of rat mitochondria GST and identification of a human homologue. Biochem. J. 319, 749–754.

Prapanthadara, L., Hemingway, J., Ketterman, A.J., 1993. Partial purification and characterization of glutathione S-transferases involved in DDT resistance from the mosquito *Anopheles gambiae*. Pestic. Biochem. Physiol. 47, 119–133.

Prapanthadara, L., Koottathep, S., Promtet, N., Hemingway, J., Ketterman, A., 1996. Purification and characterization of a major glutathione S-transferase from the mosquito *Anopheles dirus* (species B). Insect Biochem. Molec. Biol. 26, 277–285.

Prapanthadara, L., Ranson, H., Somboon, P., Hemingway, J., 1998. Cloning, expression and characterization of an insect class I glutathione S-transferase from *Anopheles dirus* species B. Insect Biochem. Molec. Biol. 28, 321–329.

Reinemer, P., Dirr, H.W., Ladenstein, R., Huber, R., Lo Bello, M., Federici, G., Parker, M.W., 1992. Three-dimensional structure of class 0 glutathione S-transferase from human placenta in complex with S-hexylglutathione at 2.8 Å resolution. J. Mol. Biol. 227, 214–226.

Reinemer, P., Prade, L., Hof, P., Neuefeind, T., Huber, R., Zettl, R., Palme, K., Schell, J., Koelln, I., Barrunik, H.D., Bieseler, B., 1996. Three-dimenstional structure of glutathione S-transferase from *Arabidopsis thaliana* at 2.2 Å resolution: structural characterization of herbicide-conjugating plant glutathione S-transferases and a novel active site architecture. J. Mol. Biol. 255, 289–309.

Snyder, M.J., Maddison, D.R., 1997. Molecular phylogeny of glutathione-S-transferases. DNA and Cell Biology 16, 1373–1384.

Stenberg, G., Board, P.G., Carlberg, I., Mannervik, B., 1991. Effects of directed mutagenesis on conserved arginine residues in a human class alpha glutathione transferase. Biochem. J. 274, 549–555.

Tahir, M.K., Mannervik, B., 1986. Simple inhibition studies for distinction between homodimeric and heterodimeric isoenzymes of glutathione transferase. J. Biol. Chem. 261, 1048–1051.

Tsuchida, S., Sato, K., 1992. Glutathione transferases and cancer. CRC Crit. Rev. Biochem. Molec. Biol. 27, 337–384.

Widersten, M., Kolm, R.H., Björnestedt, R., Mannervik, B., 1992. Contribution of five amino acid residues in the glutathione binding site to the function of human glutathione transferase P1-1. Biochem. J. 285, 377–381.

Widersten, M., Pearson, W.R., Engström, Å., Mannervik, B., 1991. Heterologous expression of the allelic variant Mu-class glutathione transferases 0 and 0 . Biochem. J. 276, 519–524.

Wilce, M.C.J., Board, P.G., Feil, S.C., Parker, M.W., 1995. Crystal structure of a theta-class glutathione transferase. EMBO J. 14, 2133–2143.

Xiao, G., Parsons, J.F., Tesh, K., Armstrong, R.N., Gilliland, G.L., 1998. Conformation changes in the crystal structure of rat glutathione transferase M1-1 with global substitution of 3-fluorotyrosine for tyrosine. J. Mol. Biol. 281, 323–339.

Pergamon

Insect Biochemistry and Molecular Biology 31 (2001) 75–85

Insect Biochemistry and Molecular Biology

www.elsevier.com/locate/ibmb

Genomic organization and putative promoters of highly conserved glutathione S-transferases originating by alternative splicing in *Anopheles dirus*

Saengtong Pongjaroenkit [a], Kanya Jirajaroenrat [a], Chanikarn Boonchauy [b], Umnaj Chanama [b], Somphob Leetachewa [a], La-aied Prapanthadara [c], Albert J. Ketterman [a,*]

[a] *Institute of Molecular Biology and Genetics, Mahidol University, Salaya Campus, Nakorn Pathom 73170, Thailand*
[b] *Institute of Science and Technology for Research and Development, Mahidol University, Salaya Campus, Nakorn Pathom, Thailand*
[c] *Research Institute for Health Sciences, Chiangmai University, Chiangmai, Thailand*

Received 6 March 2000; received in revised form 26 May 2000; accepted 1 June 2000

Abstract

The genomic DNA of a GST class I alternative splicing gene has been characterized from *Anopheles dirus*, a Thai malaria vector. This gene organization is highly conserved in *An. dirus* and *Anopheles gambiae* (*aggst1α*), with >80% nucleotide identity in the coding region. Their gene organization contains six exons for four mature GST transcripts, which share exon 1 and exon 2 but vary between four different exon 3 sequences (exon 3A–3D). The deduced amino acid sequence of the GST transcripts from these two genes also shows very high conservation, with 85–93% identity for each orthologous gene. Two putative promoters and possible regulatory elements were predicted by a combination of the TSSW and MatInspector programs. The Ad214 promoter is proposed to be involved in developmental stage regulation. The Ad2112 promoter would appear to respond to intra- or extracellular stimuli. These two Anopheline species appear to have diverged in the distant past based on gene neighbors and phylogenetic data, yet these GST genes are still conserved. Therefore it is highly probable that this GST gene organization has one or more important roles.

Keywords: Glutathione transferase; Mosquito; Gene regulation; Promoter; Genomic organization; mRNA alternative splicing

1. Introduction

Glutathione S-transferases (GSTs; EC 2.5.1.18) are a supergene family of dimeric multifunctional proteins that are essential in xenobiotic metabolism and protection against peroxidative damage (Mannervik and Danielson, 1988; Pickett and Lu, 1989; Tsuchida and Sato, 1992). GSTs are ubiquitous and widely distributed in most forms of life. Most organisms have the genetic capacity to encode multiple GST isoforms to fulfill these diverse physiological functions. Mammalian cytosolic GSTs have been divided into five classes (Alpha, Mu, Pi, Theta and Sigma) based on their distinct primary structure (Hayes and Pulford, 1995).

The insect class I GSTs are of interest because of their potential role in insecticide resistance. Although not as well characterized as the mammalian GSTs, insect GSTs have been identified and reported in multiple forms from house fly (Clark and Dauterman, 1982; Clark et al., 1984; Motoyama and Dauterman, 1978), grass grub (Clark et al., 1985), *Drosophila melanogaster* (Toung et al. 1990, 1993) and mosquito (Grant and Matsumura, 1989; Prapanthadara et al., 1993). Insect GSTs have been divided into two classes (Class I and Class II) based on their amino acid sequence homology (>40% identity to other members within the class) and immunological properties (Fournier et al., 1992; Toung et al., 1991). The insect GST class I cDNAs and the genomic DNA fragments have been cloned from several species. The

* Corresponding author. Fax: +66-2-441-9906.
E-mail address: frakt@mucc.mahidol.ac.th (A.J. Ketterman).

0965-1748/01/$ - see front matter
PII: S0965-1748(00)00107-7

data for class I GST reveals a sequentially arranged cluster for these genes (Toung et al., 1993; Ranson et al. 1998, 1997). In the *gstD* cluster, *D. melanogaster* contains seven *gst* genes that were intronless in the coding region (Toung et al., 1993). The intronless feature also was reported for *aggst1-2* of *Anopheles gambiae* (Ranson et al., 1997) and *MdGST-3* in adult *Musca domestica* (Zhou and Syvanen, 1997). In contrast, introns were identified in a class I GST of *An. gambiae* gene (*aggst1α*) that contained three exons which coded for a mature mRNA (Ranson et al., 1998). However an intron was identified in the 5′ untranslated region (5′UTR) of several intronless GST genes from *gstD* in drosophila (Lougarre et al., 1999) and from *MdGST-3* and *MdGST-4* in house fly (Zhou and Syvanen, 1997). These results demonstrate that the insect class I GST mRNA may be transcribed from more than one exon.

Class I GST enzymes have also been purified from *Anopheles dirus*, a Thai malaria vector (Prapanthadara et al., 1996). However, only one cDNA sequence (*adgst1-1*) has been reported from this mosquito (Prapanthadara et al., 1998). The present report describes the gene structure of a class I GST gene from genomic DNA that contains coding sequence for four class I GSTs of *An. dirus*, one of which is *adgst1-1*. This genomic region appears to be orthologous, that is, the same region in a different species, to a sequence obtained from *An. gambiae* (Ranson et al., 1998). Putative promoters and the possible regulatory elements of this conserved anopheline class I GST gene were identified.

2. Methods

2.1. Mosquito strains

The *Anopheles dirus* (species B) colony that was established by the Department of Parasitology, Faculty of Medicine, Chiangmai University was used for the source of the genomic DNA. The colony was identified as species B on the basis of its morphological and chromosome characteristics (Baimai, 1989).

2.2. Genomic DNA extraction, library construction and screening

Genomic DNA was extracted from 4th instar larvae of *An. dirus*. The extraction procedure was as described by (Vaughan et al., 1995) with some modifications. In brief, the proteinase K, with a higher optimal temperature, was used and the incubation time was extended until the homogenate was not viscous. The centrifugation steps were eliminated for all DNA precipitation. The 50 μg of genomic DNA was partially digested with *Sau3AI* at 0.125 U/μg DNA for 1 h. Genomic DNA fragments, 9–23 kb in size were partially filled-in with dGTP and dATP. Then the partially filled-in fragments were ligated into partially filled-in *XhoI* digested arms of Lambda FixII for the genomic DNA library construction system (Stratagene).

The library (4.9×10^5 plaque forming units) was screened with ^{32}P-labeled *adgst1-1* (630 bp) which is the coding region of *An. dirus* class I GST cDNA sequence (Prapanthadara et al., 1998). The ^{32}P-labeled *adgst1-1* probe was prepared by using Random Primers DNA Labeling system (Gibco BRL). The 50% formamide hybridization was performed as described in the Colony/Plaque Screen™ Hybridization Transfer Membranes, Transfer and Detection Protocols (NEN) for plaque hybridization. Positive recombinant bacteriophage DNA were extracted according to a standard protocol (Sambrook et al., 1989).

2.3. Analysis of positive recombinant plaque clones

Positive clones were digested with *EcoRI*, *SalI*, and *XhoI* and used for a Southern blotting. Southern hybridization was performed as described in aqueous hybridization, Colony/Plaque Screen™ Hybridization Transfer Membranes, Transfer and Detection Protocols (NEN) using ^{32}P-labeled *adgst1-1* as the probe at high stringency conditions (65°C for hybridization and washing temperature). The Digoxigenin (DIG)-labeled 5′ part of *adgst1-1* (the conserved part of insect GSTs; 200 bp) was used as the probe in low stringency hybridization (50°C for hybridization and washing temperature). The DIG-labeled probe was prepared by polymerase (PCR) using Vent® polymerase with a ratio of DIG-dUTP:TTP of 1:4. The amplification (30 cycles of 94°C for 1 min, 65°C for 30 s and 72°C for 30 s) was performed on a PE system 2400 (Perkin-Elmer). The hybridization and detection procedures were as described in The DIG System User's Guide for Filter hybridization (Boehringer-Mannheim).

The restriction enzyme digest pattern of single or double digestion was also used to generate a restriction map. The positive signal fragments were subcloned into pBluescript II SK$^+$/KS$^+$ (Stratagene) for sequencing. DNA sequencing reactions were performed using the BigDye terminator cycle sequencing system (Perkin-Elmer). Sequencing was performed using an automated sequencer (ABI PRISM ™ 377, Perkin-Elmer). The contiguous sequence was assembled using GCG FRAGMENT ASSEMBLY.

2.4. Total RNA extraction, reverse-transcription and polymerase chain reaction (RT-PCR)

The 4th instar larvae of *An. dirus* were ground to powder in liquid nitrogen. Total RNA was isolated by using RNeasy Mini Kit (QIAGEN) or TRIzol Reagent as described in the manufacturer's instructions. Then first

strand cDNA was synthesized by using Superscript II reverse transcriptase (Gibco/BRL) and the oligo-$(dT)_{15}$-adaptor primer. Primers designed from genomic sequence were used in the PCR. For each primer set, separate PCRs were performed by using Vent®polymerase to amplify the coding sequence of each transcript (*adgst1-2, adgdt1-3* and *adgst1-4*). The PCR products were purified and subcloned into pUC19 and sequenced.

2.5. 5′ rapid amplification of cDNA ends (5′RACE)

Total RNA was extracted from 4th instar larvae of *An. dirus*. Primer oligo 3 (5′-CCGGCGGTCGACGATGGCGCGCGACTCCCACAG-3′), which is to the conserved region of the four different exon 3 sequences, was used to synthesize first strand cDNA with Superscript II reverse transcriptase (Gibco/BRL). The 5′RACE was performed according to the 5′RACE System for Rapid Amplification of cDNA Ends, Version 2.0 instruction manual (Gibco/BRL). The 5′RACE product was subcloned into pBluescript II SK+ (Stratagene) and sequenced as described above.

3. Results

3.1. Isolation and genomic organization of adgst1AS1 gene

Using the ^{32}P-labeled *adgst1-1* cDNA as a probe, 4.9×10^5 plaques of an *An. dirus* genomic DNA library were screened and two independent positive clones were isolated and designated as 5A.1 and 8A.2. Most of the genomic organization of the *adgst1AS1* gene (Genbank accession AF 251478) was determined from the phage clone 8A.2, which is approximately 14 kb in size (Fig. 1).

A restriction map of *EcoRI, SalI* and *XhoI* was generated. From Southern hybridization at high stringency of the *SalI* digestion, four positive fragments were detected which were 1.2, 0.8, 0.2 and 3.0 kb fragments as show in Fig. 1. These four *SalI* fragments were completely sequenced and the overlapping restriction sites were used to confirm the *SalI* junction sequence. The continuous sequence was compared with sequences in the database to identify putative GST genes by the BLAST program. The result showed the greatest identity to *An. gambiae* GSTs.

The 4078 bp of these four *SalI* fragments contains six exons that are sequentially arranged and just five exons for the GST coding region. These GST coding regions show great identity to *aggst1α* as shown in Fig. 1 (Ranson et al., 1998) that when expressed has three exons which yield a mature GST transcript. The first exon encodes a 5′-untranslated region which is shown by only one 5′RACE product and no other 5′UTR were obtained from 4th instar larvae. The 5′UTR is exon 1 in both *An. dirus* and *An. gambiae*. The second exon encodes the amino-terminus of insect class I GSTs which is 45 amino acids in both mosquitoes. The four different third exons in both mosquitoes each encode a different carboxyl-terminus of the full-length GSTs, which are 164–174 amino acids in *An. dirus*. Therefore this GST gene was named *adgst1AS1* for *An. dirus* GSTclass I Alternatively Spliced gene 1.

The mRNA product of exon 1, exon 2 and exon 3D has previously been reported as *adgst1-1* (Prapanthadara et al., 1998) and was used as the genomic library probe in this study. The mRNA splicing product of exon 1,2-3C, 1,2-3B, 1,2-3A are designated as *adgst1-2, adgst1-3* and *adgst1-4*, respectively. All splice junction sequences for the donor and the acceptor are in agreement with the GT/AG rule as show in Table 1.

RT-PCR was performed to identify the mRNA splicing products by using primers based on the genomic sequences. The RT-PCR products for the coding sequences of *adgst1-2, adgst1-3* and *adgst1-4* were obtained. These PCR products were sequenced to confirm that they were the *adgst1-2, adgst1-3* and *adgst1-4* sequences and to determine whether splicing has occurred at the predicted sites. These cDNA sequences reveal all four alternatively spliced transcripts were expressed in 4th instar larvae and spliced at the predicted sites. These cDNA sequences were compared to each predicted coding sequence from the genomic clone (data not shown). The result showed two base differences in *adgst1-2* (2/651) and *adgst1-3* (2/627). One base change occurred in *adgst1-4* (1/657). But the deduced amino acid sequence of these cDNAs remained unchanged. The *adgst1-1* that was used as the probe showed 13 base changes from the predicted coding sequence (13/627) of the genomic clone and these coded for five amino acid changes. The initial *adgst* 1-1 clone as well as the sequences reported in the present study have been obtained from the same *An. dirus* colony. However, there was approximately three years between the RNA isolations and the colony gene pool has shifted with respect to GSTs. We have also obtained a cDNA sequence from the current colony that is an identical amino acid sequence to the translated genomic sequence and not to the initial *adgst* 1-1. Currently an allelic form of *adgst1-1* has been obtained by RT-PCR that has five base changes but does not code for any change in the predicted amino acid sequence compared to the genomic clone of *adgst1-1*. Therefore these differences are most likely due to the presence of allelic forms, which is a common characteristic of GST (Board, 1981; Pemble et al., 1994; Widersten et al., 1991).

3.2. The homology of adgst1AS1 and aggst1α

Nucleotide sequence comparison was performed to determine the homology of *adgst1AS1* and *aggst1α* as

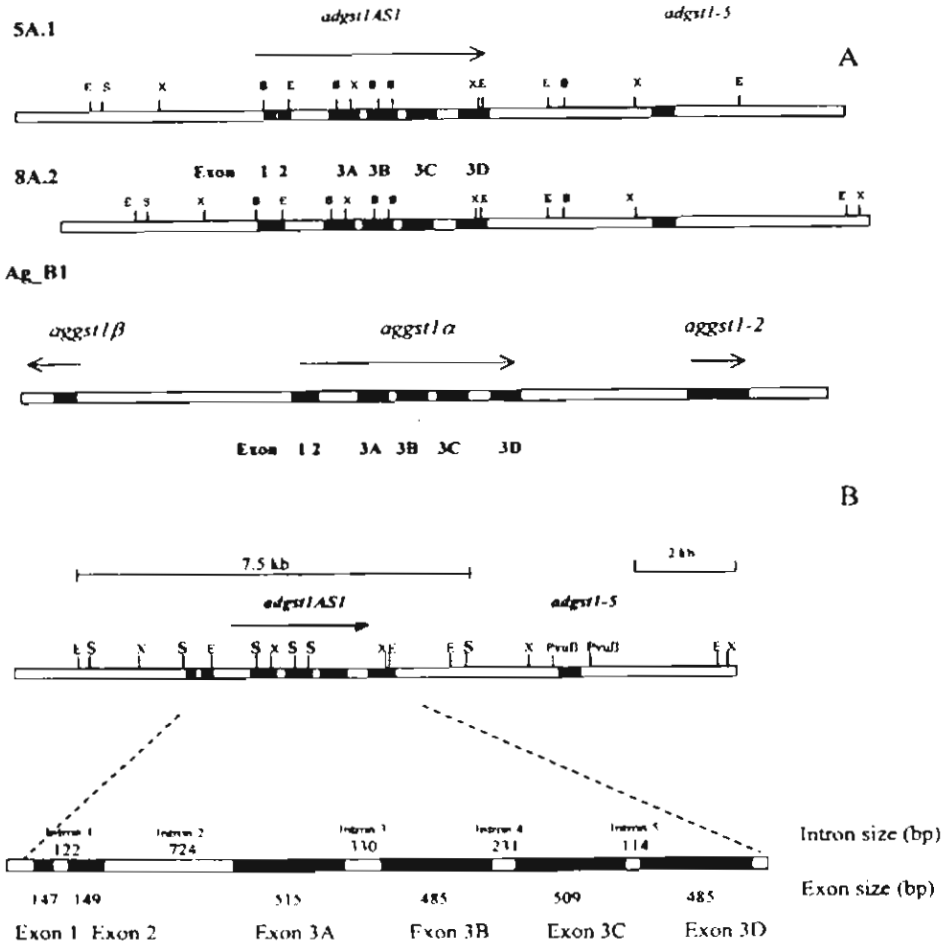

Fig. 1. (A) The comparison of recombinant phage clone of *An. dirus* (5A.1 and 8A.2) and *An. gambiae* (Ag_B1) (Ranson et al., 1998). The restriction map of clone isolated from an *An. dirus* genomic library. The enzyme abbreviations are E; *EcoRI*, S; *SalI* and X; *XhoI*. Southern hybridization positive fragments from *SalI* digestion that were sequenced are shown in bold. The *adgst1AS1* was identified and the arrow indicates the direction of *adgst1AS1*. (B) The sizes of the exons and introns of *adgst1AS1* are shown in bp. The 7.5 kb genomic sequence (Genbank accession AF 251478) has been analyzed for GST coding sequence and putative regulatory elements.

Table 1
The exon composition and splice sites of alternative transcripts of *adgst1AS1* and the splice site of the putative 5′region of *adgst1-5* (*aggst1β* orthologous gene)*

Gene	Exon composition **Consensus**	5′splice site (exon/intron) **AG/GTRAGT**	3′splice site (intron/exon) **Y$_n$NYAG/NN**
adgst1AS1	exon1 ...exon2	CG/GTGAGT	CTCGCAG/AA
	exon2....exon3A	AG/GTAAGT	TTTAAAG/CT
	exon2....exon3B		CCCTCAG/AT
	exon2....exon3C		TCCGCAG/CT
	exon2....exon3D		ATTACAG/CT
adgst1-5	1st coding sequence	AG/GTAAGC	

* (R=A or G, Y=C or T, N=A,C,G or T).

shown in Fig. 2. The comparison reveals a very high level of identity in all coding regions; they are, 93.3% for exon 2, 81% for exon 3A, 85.4% for exon 3B, 78.5% for exon 3C and 86.8% for exon 3D. However, there were differences observed in exon 1 (41.56%) and in the intron sizes and sequences of these two genes (Fig. 2).

The comparison revealed *adgst1AS1* and *aggst1α* have an identical arrangement and exon composition. The deduced amino acid sequences of these two genes were also compared as shown in Fig. 3. The comparison showed a very high identity for the deduced amino acid sequences of the orthologous proteins that are 93% identical for AdGST1-1, 85% identical for AdGST1-2, 92% identical for AdGST1-3 and 86% identical for