Abstract

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 five publications that constitute this report.

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.

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* (*aggst1*a), 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.

Three cDNA sequences of glutathione *S*-transferase (GST), *adgst1–2*, *adgst1–3* and *adgst1–4*, which are alternatively spliced products of the *adgst1AS1* gene, were

obtained from fourth instar larvae of *Anopheles dirus* mosquito by reverse transcriptase PCR reactions. The nucleotide sequences of these three cDNAs share 67% identity and the translated amino acid sequences share 61– 64% identity. A comparison of the *An. dirus* to the *An. gambiae* enzymes shows that adGST1–2 versus agGST1–4, adGST1–3 versus agGST1–5 and adGST1–4 versus agGST1–3 have 85, 92 and 85% amino acid sequence identity, respectively, which confirms that orthologous isoenzymes occur across anopheline species. These three proteins were expressed at high levels, approximately 15–20 mg from 200 ml of *E. coli* culture. The recombinant enzymes were purified by affinity chromatography on an *S*-hexylglutathione agarose column. The subunit sizes of adGST1–2, adGST1–3 and adGST1–4 are 24.3, 23.9 and 25.1 kDa. The recombinant enzymes have high activities with 1-chloro-2,4-dinitrobenzene (CDNB), detectable activity with 1,2-dichloro-4-nitrobenzene but markedly low activity with ethacrynic acid and *p*-nitrophenethyl bromide. adGST1–3 was shown to be the most active enzyme from the kinetic studies. Permethrin inhibition of CDNB activity, at varying concentrations of CDNB, was significantly different, being uncompetitive for adGST1–2, noncompetitive for adGST1–3 and competitive for adGST1–4. In contrast, permethrin inhibition with varying glutathione concentrations was noncompetitive for all three GSTs. Despite the enzymes being splicing products of the same gene and sharing identical sequence in the N-terminal 45 amino acids, these GSTs show distinct substrate specificities, kinetic properties and inhibition properties modulated by the differences in the C-terminus.

Two glutathione S-transferase isozymes from the mosquito Anopheles dirus (AdGST1-3 and AdGST1-4) from an alternately spliced gene family have been expressed, purified and crystallized. The isozymes share an N-terminal domain derived from a single exon and C-terminal domains from unique exons. Despite the high level of sequence identity (64% overall), the two isozymes crystallize in different space groups, the 1-3 isozyme in P3121 or P3221 (unit-cell parameters a = 49.9, c = 271.8 A at 100 K) and the 1-4 isozyme in P41 or P43 (unit-cell parameters a = 87.8, c = 166.1 at 100 K). Determination of these structures will advance our understanding of how these enzymes inactivate pesticides and the structural consequences of alternate splicing.

A new *Anopheles dirus* glutathione S-transferase (GST) has been obtained and named adGST4-1. Both genomic DNA and cDNA for heterologous expression were acquired. The genomic sequence was 3188 bp and consisted of the GST gene as well as flanking sequence. The flanking sequence was analyzed for possible regulatory elements that would control gene expression. In *Drosophila* several of these elements have been shown to be involved in development and cell differentiation. The deduced amino acid sequence has low identity compared with the four alternatively spliced enzymes, adGST1-1 to 1-4, from another *An. dirus* GST gene *adg st1AS1*. The percent identities are 30–40% and 11–12% comparing adGST4-1 to insect GSTs from Delta and Sigma classes, respectively. Enzyme characterization of adGST4-1 shows it to be distinct from the other *An. dirus* GSTs because of low enzyme activity for customary GST substrates including 1-chloro-2, 4-dinitrobenzene (CDNB). However, this enzyme has a greater affinity of interaction with pyrethroids compared to the other *An. dirus* GSTs.

Keywords: glutathione transferase, *Anopheles dirus*, mosquito, molecular modeling, isoenzymes, structure-function study, gene regulation, genomic organization, alternatively spliced proteins, crystallization

กลูตาไธโอน เอส ทรานสเฟอเรส (Glutathione S-transferases หรือเรียกย่อๆ ว่า GSTs) เป็นกลุ่มของเอนไซม์ที่เป็น multigene family ซึ่งอยู่ร่วมกันเป็นคู่และมีหน้าที่หลากหลาย (multifunctional dimeric form) โดยมีบทบาทหน้าที่หลักในการขจัดสารพิษ GSTs จึงมีส่วนเกี่ยวข้องในการดื้อต่อยาฆ่าแมลงในยุงซึ่งเป็นพาหะของโรคมาลาเรีย จุดประสงค์หลักของโครงการวิจัยนี้ คือเพื่อศึกษาจำแนกการจัดเรียงตัวและกลไกการควบคุมของยีน GSTs ในแมลงพาหะของโรคมาลาเรียในไทย คือ *Anopheles dirus* รวมทั้งศึกษาเพื่อให้ได้มาซึ่งลำดับเบสในจีโนมของ GSTs และเพื่อจำแนกหาหน่วยควบคุมการแสดงออกของยีน (regulatory element) นอกจากนี้รวมถึงการศึกษาการแสดงออกของยีน GST ต่างๆ ที่ได้ และการศึกษาคุณลักษณะเฉพาะของเอนไซม์ในเชิงความจำเพาะเจาะจงต่อ substrate และในเชิง inhibition kinetics

ผลงานที่ได้ในโครงการวิจัยนี้ เราสามารถทราบลำดับเบสในจีโนมที่จำเพาะต่อยีน GST จำนวน 3 ยีน โดย 2 ยีนเป็นรหัสสำหรับสร้างโปรตีนเดี่ยว 2 ชนิด ส่วนยีนที่ 3 เป็นรหัสสำหรับสร้างโปรตีน 4 ชนิดที่แตกต่างกัน ซึ่งเป็นผลจากกลไกแบบ alternative splicing โดยโปรตีนทั้ง 6 ชนิดจาก 3 ยีนนี้สามารถถูกแสดงออกได้เป็นลักษณะเฉพาะของเอนไซม์ที่ได้ นอกจากนี้ เรายังสามารถตกผลึกโปรตีน (crystallized) และได้โครงสร้าง 3 มิติของโปรตีน GST 2 ชนิดแล้ว โดยที่งานทั้งหมดที่กล่าวมาได้อธิบายอยู่ในผลงานตีพิมพ์ทั้ง 5 ฉบับ ซึ่งประกอบอยู่รายงานฉบับนี้