Regulatory Mechanisms Involved in the Control of CYP6M2 Gene in Insecticide Resistant Anopheles gambiae (Diptera: Culicidae)

Balarabe Rabiu Mohammed

DVM, MSc

A thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy in the School of Science, Engineering & Technology, Abertay University, Dundee, United Kingdom

December, 2014

© 2014
ABSTRACT

Cytochrome P450s, including CYP6M2 gene, are involved in the detoxification of permethrin. Some of these genes are regulated by CnCC / dKeap 1 and Spineless / Tango in Drosophila melanogaster. This mechanism is yet to be identified in Anopheles gambiae. Thus, we examine whether there is differential regulation of CYP6M2 gene between permethrin resistant Tiassalé and susceptible Kisumu strains of An. gambiae. Bioinformatics analysis was used to search for cis-acting elements within CYP6M2 (896 bp) region hypothesised to contain the promoter. Isolated and cloned CYP6M2 promoter reporter constructs were transfected into Anopheles gambiae Sua 5.1* cells to measure luciferase activity as a surrogate promoter activity. The WHO adult bioassay was used to expose adult females of the permethrin resistant Tiassalé and susceptible Kisumu strains of An. gambiae to discriminating doses of 0.75% permethrin. Uncharacterised strains from Auyo (Auyo-Nigeria) selected to 4% DDT and 0.1% Bendiocarb as recommended by WHO were also studied. Total RNA was isolated from the respective selected and unselected strains of An. gambiae and cDNA synthesised. Semi and Real time quantitative PCR (qPCR) using SYBR® Green were used to determine the gene expression and regulation levels. Results established the presence of putative AGAP010259 (AhR) and AGAP005300 (Nf2e1) cis-acting elements within Anopheles gambiae CYP6M2 promoters in silico. Luciferase reporter gene assays revealed no promoter activity as confirmed by using CYP9M10 promoter from Culex quinquefasciatus with a known promoter activity as control. There is higher expression of Nf2e1 than AGAP010259 and a variable expression of CYP6M2 in all the insecticide selected individuals, which may potentially be associated with insecticide resistance. This study provides useful information on our understanding of the regulatory mechanisms involved in insecticide resistance. These results have implications for the control of mosquito populations and the global spread of human, livestock and poultry diseases.
DEDICATION

This thesis is dedicated to my late father

Alhaji Rabiu Yaro Mohammed of blessed memory to whom I am

forever grateful

and

my beloved mother

Hajiya Aisha Rabiu Mohammed
I BALARABE RABIU MOHAMMED declare that this thesis is my own original work. It is being submitted for the Degree of Doctor of Philosophy at the Abertay University, Dundee, United Kingdom. It has not been submitted before for any degree or examination in any other University.

____________________________________
(Signature of candidate)
Address:

__________________ day of ____________________ 20________________

B.R.Mohammed - IV- Abertay University, Dundee, UK
I certify that BALARABE RABIU MOHAMMED, a PhD candidate has undertaken all the work described herein and is based on the original work done at the Abertay University, Dundee in partial fulfilment for the requirements for the award of Doctor of Philosophy in Molecular Entomology. This has not previously formed the basis for the award of any degree, diploma, associateship, fellowship or any similar title and represents an independent work on the part of the candidate. This is also a true and accurate version of the thesis approved by the examiners, and that all relevant ordinance regulations have been fulfilled.

(Signature of the Principal Supervisor)

Dr Yusuf Y. Deeni
Second Supervisor

Dr Phillip J. Collier
School of Science, Engineering and Technology (SET),
Dundee, DD1 1HG, United Kingdom.

________________________________________ day of ____________________ 20 __________________

B.R.Mohammed  - V-  Abertay University, Dundee, UK
ACKNOWLEDGEMENTS

First of all, my ultimate gratitude is to Almighty Allah (S.W.A) who created me and provided me with the ability, knowledge, patience and strength to surmount this enormous challenge.

To quote Sir Isaac Newton, an esteemed scientist: “If I have seen further, it is because I have been standing on the shoulders of giants.” This PhD thesis would not have been possible without collaboration with other institutions and important individuals among whom I will like to express my heart felt appreciation:

To Dr Yusuf Deeni my principal supervisor for entrusting me with most of the decisions related to my research project. His insightful, thoughtful and unfailingly wise suggestions and encouragement made me clearly build motivation and enthusiasm to become an independent researcher. He also exposed me to criticisms that always gave me lessons in thinking, reasoning and expressing ideas in simple and clear ways.

To Dr Phillip J. Collier who doubles as my second supervisor and director of research for his excellent guidance and tutelage. He always demonstrated great kindness and patience towards me, and for indefatigable supervision and sage advice.

To Prof. John Palfreyman (now retired) for painstakingly offering suggestions and comments in designing this research especially at the initial take off.

To Dr Mark J. Paine, Vector Group, Liverpool School of Tropical Medicine (LSTM), for receiving me in his lab as a visiting student allowing me to perform most of my experiments. This was fundamental for the success of this PhD research.

To Dr Craig S. Wilding of Liverpool John Moores University (Formerly of Vector Group, Liverpool School of Tropical Medicine) for his constant contribution throughout my research period and for providing me with the opportunity to benefit from his repertoire of knowledge. He kept me on the right path and was always available for support and advice which was especially useful when things were not going according to plan.

To the other minds at both Abertay University and LSTM who made the whole experience a great deal less traumatic. Some provided technical assistance; some provided general chit chat, comfort and others generally just made life easier. A special mention must be given to Dr Hilal Khalil, Sojimade Nuruddeen and Mr Mike Black (Abertay University), Kay Hemmings, Emily Rippon; Jonathan Thornton, Dr Amy Lynd Dr Rodolphe Pouardin (LSTM) and members of the Vector Group. All of them cannot be named (due to space constraints!). I feel extremely fortunate to have been in such a fantastic environment, and can only hope I have contributed in turn to the spirit of these laboratories.

To my family who have been a great source of comfort and encouragement all these years. In particular, I want to express my sincerest gratitude to my mother Hajia Aisha Rabiu Mohammed for her undying concern, prayers and support during this time and always.

To my wife, Hajia Halima for being my rock, my strength and for all the love, patience and encouragement you have always given me. Thank you for consistently giving me your best.

To my lovely kids Amir, Abba, Ummalkhair and Nyass for their patience, understanding and limitless love without which this research would not have been accomplished.

To my sponsors, the Tertiary Education Trust Fund (TETFUND) and my employers University of Abuja, Nigeria for providing the financial support to conduct this research.
In submitting this thesis to the Abertay University, Dundee, I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby.

______________________________________________________________

______________________________________________________________

______________________________________________________________
ACRONYMS/ABBREVIATIONS/SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>Beta</td>
</tr>
<tr>
<td>AhR</td>
<td>Aryl hydrocarbon receptor</td>
</tr>
<tr>
<td>ARNT</td>
<td>Aryl hydrocarbon receptor nuclear translocator</td>
</tr>
<tr>
<td>An.</td>
<td>Anopheles</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ARE</td>
<td>Antioxidant response element</td>
</tr>
<tr>
<td>ATG</td>
<td>Start codon</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>CncC</td>
<td>Cap ‘n’ collar isoform-C</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CYP</td>
<td>Cytochrome P450</td>
</tr>
<tr>
<td>DDT</td>
<td>Dichloro diphenyltrichloroethane</td>
</tr>
<tr>
<td>dH2O</td>
<td>Distilled water</td>
</tr>
<tr>
<td>dKeap1</td>
<td>Drosophila Kelch- like ECH-associated protein 1</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DRE</td>
<td>Dioxin response element</td>
</tr>
<tr>
<td>Ef1</td>
<td>Elongation factor 1 protein</td>
</tr>
<tr>
<td>et al.</td>
<td>et alia (and others)</td>
</tr>
<tr>
<td>FP</td>
<td>Forward primer</td>
</tr>
<tr>
<td>g</td>
<td>relative centrifugal force</td>
</tr>
<tr>
<td>g/L</td>
<td>grams per litre</td>
</tr>
<tr>
<td>gDNA</td>
<td>genomic DNA</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross domestic product</td>
</tr>
<tr>
<td>h</td>
<td>hour (s)</td>
</tr>
<tr>
<td>HAH</td>
<td>Halogenated aryl hydrocarbons</td>
</tr>
<tr>
<td>Hsp</td>
<td>Heat shock protein</td>
</tr>
<tr>
<td>IRS</td>
<td>Indoor residual spray</td>
</tr>
<tr>
<td>ITNs</td>
<td>Insecticide treated nets</td>
</tr>
<tr>
<td>Kb</td>
<td>kilo base</td>
</tr>
<tr>
<td>Keap1</td>
<td>Kelch- like ECH-associated protein 1</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>LITE</td>
<td>Liverpool Insect Testing Establishment</td>
</tr>
<tr>
<td>LLINs</td>
<td>Long-lasting insecticide treated nets</td>
</tr>
<tr>
<td>LSTM</td>
<td>Liverpool School of Tropical Medicine</td>
</tr>
<tr>
<td>Maf</td>
<td>Muscle aponeurosis fibromatosis</td>
</tr>
<tr>
<td>MCS</td>
<td>Multiple cloning site</td>
</tr>
<tr>
<td>Min</td>
<td>minute (s)</td>
</tr>
<tr>
<td>mL</td>
<td>milliliter</td>
</tr>
<tr>
<td>mM</td>
<td>millimole</td>
</tr>
<tr>
<td>µ</td>
<td>micro</td>
</tr>
<tr>
<td>µg</td>
<td>microgram</td>
</tr>
<tr>
<td>µL</td>
<td>microliter</td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mm</td>
<td>millimeter</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Centre for Biotechnology Information</td>
</tr>
<tr>
<td>ng</td>
<td>nanogram</td>
</tr>
<tr>
<td>nm</td>
<td>Nanomole</td>
</tr>
<tr>
<td>Nf2e1</td>
<td>Nuclear factor erythroid 2, invertebrate</td>
</tr>
<tr>
<td>Nrf2</td>
<td>Nuclear factor erythroid-2 related factor-2</td>
</tr>
<tr>
<td>°C</td>
<td>Degrees Celsius</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>PAH</td>
<td>Polycyclic aromatic hydrocarbon</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>pGL3-Basic</td>
<td>Firefly luciferase reporter vector w/o CY6M2 gene promoter</td>
</tr>
<tr>
<td>pGL3-CYP6M2</td>
<td>Firefly luciferase reporter vector w/ CY6M2 putative promoter</td>
</tr>
<tr>
<td>pH</td>
<td>Potential of hydrogen</td>
</tr>
<tr>
<td>pJET1.2_CYP6M2</td>
<td>PCR cloning vector with CYP6M2 putative promoter</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>Rcf</td>
<td>Relative centrifugal force</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RP</td>
<td>Reverse primer</td>
</tr>
<tr>
<td>RPM</td>
<td>Revolution per minute</td>
</tr>
<tr>
<td>Rsp7</td>
<td>Ribosomal protein</td>
</tr>
<tr>
<td>s</td>
<td>Second (s)</td>
</tr>
<tr>
<td>spp</td>
<td>Species</td>
</tr>
<tr>
<td>Ss</td>
<td>Spineless</td>
</tr>
<tr>
<td>TCDD</td>
<td>2, 3, 7, 8-tetrachlorodibenzo-p-dioxin</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris acetate EDTA buffer</td>
</tr>
<tr>
<td>TE</td>
<td>Transposable element</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TFBS</td>
<td>Transcription factor binding site</td>
</tr>
<tr>
<td>Tgo</td>
<td>Tango</td>
</tr>
<tr>
<td>Tm</td>
<td>Primer melting temperature</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>UBQ</td>
<td>Ubiquitin</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>UTR(s)</td>
<td>Untranslated region(s)</td>
</tr>
<tr>
<td>UV(s)</td>
<td>Ultra violet</td>
</tr>
<tr>
<td>V</td>
<td>Volts</td>
</tr>
<tr>
<td>Vol</td>
<td>Volume (s)</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>XRE</td>
<td>Xenobiotic response element</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS

ABSTRACT ........................................................................................................ II
DEDICATION .................................................................................................... III
AUTHOR’S DECLARATION ........................................................................... IV
CERTIFICATE .................................................................................................. V
ACKNOWLEDGEMENTS ............................................................................... VI
UAD- PERMISSION TO COPY ...................................................................... VII
ACRONYMS/ABBREVIATIONS/SYMBOLS ................................................ VIII
LIST OF FIGURES ....................................................................................... XVII
LIST OF TABLES .......................................................................................... XX

CHAPTER ONE
Introduction and Literature Review

1.0 Introduction and Literature Review ..................................................... 2
1.1 Brief history of mosquitoes ................................................................. 2
  1.1.1 Medical and Veterinary importance of mosquitoes ...................... 3
1.2 Mosquitoes as vectors of diseases .................................................... 4
  1.2.1 Transmission of malaria in man ................................................... 4
  1.2.2 Epidemiology of malaria .............................................................. 6
  1.2.3 Mosquito vector control ............................................................... 7
1.3 Classes of insecticides in the control of mosquitoes ......................... 7
  1.3.1 Pyrethroids insecticides ............................................................... 8
  1.3.2 Permethrin .................................................................................. 8
    1.3.2.1 Structure ............................................................................ 8
    1.3.2.2 Mode of action of pyrethroids .............................................. 9
1.4 Insecticide resistance in Anopheles gambiae ..................................... 10
1.5 Biology of P450s ................................................................................. 11
  1.5.1 Historical frame work ................................................................. 11
  1.5.2 Nomenclature ............................................................................. 12
  1.5.3 P450 clans .................................................................................. 12
1.5.4 The CYP6 family ................................................................. 13
1.5.5 Localisation and function .................................................. 13
1.5.6 Cytochrome P450s in Anopheles gambiae and Drosophila melanogaster .......................... 14
1.5.7 Role of P450s in insecticide resistance ................................ 15
1.5.8 Regulation of gene expression ............................................ 15
  1.5.8.1 Promoters ................................................................. 16
  1.5.8.2 Enhancers ............................................................... 17
  1.5.8.3 Repressors ............................................................. 18
    1.5.8.3.1 Silencers ......................................................... 18
  1.5.8.4 Insulators .............................................................. 18
1.5.9 Regulatory mechanism in resistance .................................. 19
1.6 The spineless (Ss) gene ..................................................... 19
  1.6.1 Function and structure of spineless ................................ 20
  1.6.2 Ligands of spineless (AhR) .......................................... 20
  1.6.3 Molecular mechanisms of AhR functions in the regulation of cytochrome P450s in Drosophila .......................................................... 21
    1.6.3.1 Spineless (Ss) / Tango (Tgo) signalling pathway .......... 21
1.7 The cap ‘n’ collar isoform C (CnCC) named after the cnc gene of Drosophila .......................................................... 22
  1.7.1 Function of the CnCC .................................................. 23
  1.7.2 Ligands of the CnCC .................................................. 23
  1.7.3 Molecular mechanisms of the CnCC functions in the regulation of P450 cytochrome genes .................................................. 23
    1.7.3.1 CnCC / dKeap1 signalling pathway .......................... 23
  1.7.4 Other Nuclear receptors in higher mammalian .................... 26
    1.7.4.1 Ligands of PXR and CAR .................................... 26
    1.7.4.2 PXR signalling pathway .................................... 27
Research question ............................................................... 27
Research goal and specific objectives ..................................... 28
CHAPTER TWO
Bioinformatic Analysis of Regulatory Elements within the Promoter Region of the Cytochrome P450 Gene, CYP6M2 in Anopheles gambiae

2.1 Introduction .................................................................................................................. 30

2.2 Experimental Approach ............................................................................................. 32
  2.2.1 Phylogenetic tree of the Anopheles gambiae CYP6M2 gene sequence .................. 32
  2.2.2 Genomic DNA and mRNA sequence retrieval ...................................................... 33
  2.2.3 Insilico analysis of CYP6M2 gene sequence and CYP6M2 gene promoter ............ 34
  2.2.4 Recognition of promoter / enhancer elements within the cloned and sequenced CYP6M2 gene promoters ................................................................. 34
  2.2.5 Prediction of putative regulatory elements within the cloned and sequenced CYP6M2 gene promoters ................................................................. 35
  2.2.6 Insilico Identification of Aryl hydrocarbon receptor (AhR), Nuclear factor erythroid -2 related factor-2 (Nrf2) orthologs and their receptors .......................... 35
    2.2.6.1 Identification of AhR / ARNT orthologs ......................................................... 35
    2.2.6.2 Identification of Nrf2 / ARE orthologs .......................................................... 36
  2.2.7 Multiple amino acid sequence alignments of the CYP6M2 gene promoter ........... 36
    2.2.7.1 Multiple amino acid alignments of CnCC from Drosophila melanogaster, Nf2e1 from Anopheles gambiae and Nrf2 from Homo sapiens ......................... 36
    2.2.7.2 Multiple amino acid sequence alignments of Spineless from Drosophila melanogaster, AGAP010259 from Anopheles gambiae AhR from Homo sapiens ....... 37

2.3 Result ............................................................................................................................ 37
  2.3.1 Phylogenetic tree of CYP6M2 gene sequence ......................................................... 37
  2.3.2 CYP6M2 promoter genomic sequence retrieval .................................................... 38
  2.3.3 Cytochrome P450s promoter searching and analysis of An. gambiae, Aedes aegypti, D. melanogaster, Culex quinquefasciatus and Homo sapiens .......................... 39
  2.3.4 Prediction of putative promoter elements ............................................................. 41
  2.3.5 Prediction of conserved Anopheles gambiae (CYP6M2) and Drosophila melanogaster (CYP6G1) putative transcription factor binding sites (TFBS) in the promoter regions (Consites) ................................................................. 47
  2.3.6 Putative transcription factor binding sites found along 5’ 896 bp up stream region of CYP6M2 in Anopheles gambiae at 70% cut off score ........................................... 49
2.3.7 Identification of the Nrf2 / ARE and AhR / ARNT orthologs ......................... 49

2.3.8 Multiple amino acid sequence alignments of CnCC from Drosophila melanogaster, Anopheles gambiae (Nf2e1) and Homo sapiens (Nrf2) .................. 50

2.3.9 Multiple amino acid sequence alignments of Spineless from Drosophila melanogaster, Anopheles gambiae (AGAP010259) and Homo sapiens (AhR) ....... 51

2.4 Discussion ........................................................................................................... 52

2.5 Conclusions ......................................................................................................... 54

CHAPTER THREE
Cloning of Anopheles gambiae CYP6M2 Gene Promoter and Construction of its Luciferase Gene Reporter System

3.1 Introduction ....................................................................................................... 56

3.2 Experimental Approach .................................................................................... 58

3.2.1 Mosquito strains .......................................................................................... 58

3.2.2 Designing primers for PCR ........................................................................... 58

3.2.3 BlastX sequence analysis .............................................................................. 59

3.2.4 Isolation of genomic DNA and polymerase chain reaction (PCR) ............... 59

3.2.4.1 Cloning and sequencing of CYP6M2 promoter ......................................... 60

3.2.4.2 DNA sequencing ...................................................................................... 60

3.3 Construction of luciferase gene reporter system .............................................. 61

3.3.1 Design of primers for PCR ........................................................................... 61

3.3.2 Isolation of the CYP6M2_Tias (631 bp) and CYP6M2_Kis (591 bp) putative promoter using polymerase chain reaction (PCR) .................................... 61

3.3.3 Colony PCR to screen for successful ligations ............................................ 61

3.3.4 Restriction digests of CYP6M2_Kis, CYP6M2_Tias and pGL3-Enhencer vector ........................................................................................................ 62

3.4 Preparation of positive control plasmids for dual luciferase assay .................. 63

3.4.1 Kpn1 and HindIII double restriction digests of pGL3-CYP9M10 ISOP450 (1504 bp) and pGL3-CYP9M10_CqSF (804 bp) plasmids from Culex quinquefasciatus .................................................. 63

3.5 Results ............................................................................................................... 63
Regulatory Mechanisms Involved in the Control of CYP6M2

3.5.1 Sequence identification of CYP6M2_Kis and CYP6M2_Tias promoter .......... 63
3.5.2 Polymerase chain reaction (PCR) and DNA sequencing of CYP6M2 5' upstream region from susceptible and resistant An. gambiae strains ................. 64
3.5.3 Cloning of CYP6M2 putative promoter ........................................... 64
3.5.3.1 Map and Features of pJET1.2/blunt cloning vector ...................... 65
3.5.4 DNA sequencing input sequence ....................................................... 66
   3.5.4.1.1 >CYP6M2_Kis-1 - First Input Sequence .............................. 66
   3.5.4.1.2 >CYP6M2_Tias-1 - Second Input Sequence ......................... 66
   3.5.4.1.3 >CYP6M2_Kias-2 - First Input Sequence ............................ 67
   3.5.4.1.4 >CYP6M2_Tias-2 - Second Input Sequence ......................... 67
3.5.5 Pairwise sequence alignment of the Tiassalé resistant and Kisumu susceptible (Kis-1 and Tias-1) from EMBL-EBI .............................................. 67
3.5.2.2 DNA output sequence for Kis-1 and Tias-1 .................................... 69
3.6 Construction of luciferase gene reporter assay vectors ....................... 71
   3.6.1 PCR Amplification of the CYP6M2_Kis and CYP6M2_Tias promoters .... 71
3.6.2 Map and Features of pGL3 Enhancer vector .................................... 72
3.6.3 Colony PCR screening using CYP6M2_Kis (FP) & CYP6M2_Kis (RP) and CYP6M2_Tias (FP) & CYP6M2_Tias (RP) ............................................. 72
3.6.4 XhoI and NcoI double enzyme restriction digests of the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias gene constructs ........................................ 73
3.6.5 KpnI and HindIII double restriction digests of pGL3-CYP9M10_ISOP450 (1504 bp) and pGL3-CYP9M10_CqSF (804 bp) plasmids from Culex quinquefasciatus ......................................................... 74
3.7 Discussion ......................................................................................... 75
3.8 Conclusions ..................................................................................... 77

CHAPTER FOUR
Transcriptional response elements in the promoter of CYP6M2, an insect P450 gene regulated in the mosquito Anopheles gambiae

B.R.Mohammed - XIV - Abertay University, Dundee, UK
CHAPTER FIVE
Differential Expression of AGAP010259 (AhR) and Nf2e1 (Nrf2) Candidate Genes in Selected Strains of Anopheles gambiae

B.R.Mohammed - XV- Abertay University, Dundee, UK
5.1 Introduction ............................................................................................................. 94
5.2 Experimental Approach ....................................................................................... 97
  5.2.1 Mosquito strains ............................................................................................... 97
  5.2.2 WHO adult bioassay ......................................................................................... 98
  5.2.3 RNA isolation and cDNA synthesis for qPCR .................................................. 99
    5.2.3.1 Total RNA isolation ..................................................................................... 99
    5.2.3.2 cDNA synthesis ......................................................................................... 99
    5.2.3.3 Primer design ........................................................................................... 99
    5.2.3.4 Quantitative (real-time) PCR (qPCR) using SYBR® Green Master mix ....... 101
5.3 Results .................................................................................................................. 102
  5.3.1 Selection of candidate reference genes in permethrin selected An. gambiae strains using SYBR® Green (qPCR) ................................................................. 102
  5.3.2 Selection of candidate reference genes in Auyo-Selected and Auyo-Base An. gambiae using semi-quantitative qPCR .......................................................... 103
  5.4 Quantitative (real-time) PCR (qPCR) ................................................................. 104
    5.4.1 Differential expression levels of AGAP010259 (AhR), Nf2e1 (Nrf2) and CYP6M2 ................................................................. 104
5.5 Discussion ............................................................................................................ 106
5.6 Conclusions .......................................................................................................... 109

CHAPTER SIX
Discussion, Conclusions and Suggestions for future work

6.0 Discussion .............................................................................................................. 111
  6.1 Introduction ......................................................................................................... 111
  6.2 Insilico Identification of the orthologs to CnCC / dKeap 1 and / Spineless / Tango in Anopheles gambiae ......................................................................................... 113
  6.3 Insilico Establishment of putative transcription binding sites within the 5’ region upstream of CYP6M2 ...................................................................................... 114
  6.4 Cloning of the 5’ 896 bp region upstream of CYP6M2 and construction of its luciferase gene reporter system ................................................................. 114
  6.5 Differential expression of AGAP010259 (AhR) and Nf2e1 (Nrf2) candidate genes in selected strains of Anopheles gambiae ........................................................................ 115
6.5.1 Nuclear factor erythroid 2, invertebrate (Nf2e1) / AGAP003645 signalling pathway in Anopheles gambiae .................................................. 117

6.5.2 Inhibition of the Nf2e1 / AGAP003645 signalling pathway in Anopheles gambiae ................................................................. 118

6.6 General conclusions ............................................................................................................................................................................. 118

6.7 Suggestions for further works ......................................................................................................................................................... 119

References............................................................................................................................................................................................... 123

Refereed Journal articles and conference abstracts publications from this thesis .................................................................................. 151

Appendix I ......................................................................................................................................................................................................... 152

Appendix II .................................................................................................................................................................................................... 160

Appendix III .................................................................................................................................................................................................. 166

Appendix IV ................................................................................................................................................................................................ 173

Appendix V ................................................................................................................................................................................................ 174
LIST OF FIGURES

CHAPTER ONE

**Figure 1.1:** A global map of dominant or potentially important malaria vector species ............................................................... 3

**Figure 1.2:** Malaria transmission by mosquito in Man (*Plasmodium* species) life cycle .................................................................................................................. 6

**Figure 1.3:** Shows the structural formula of permethrin (a pyrethroid) ........ 9

**Figure 1.4:** Schematic presentation of potential physiological changes associated with insecticide resistance in *Anopheles gambiae* vectors ........................................... 12

**Figure 1.5:** Simplified model of transcriptional regulation of protein-encoding genes ................................................................................ 19

**Figure 1.6:** General scheme for the induction of *Spineless (Ss) / Tango (Tgo)-signalling pathway* ......................................................................................................................... 22

**Figure 1.7:** General scheme for the induction of *CnCC / dKeap1-signalling pathway* ......................................................................................................................... 25

CHAPTER TWO

**Figure 2.1:** Phylogenetic tree of *CYP6M2* gene sequence ................... 38

**Figure 2.2:** Shows image of the retrieved *CYP6M2* promoter gene .......... 39

**Figure 2.3A:** McPromoter analysis of the 5' 896 bp region upstream *CYP6M2* _Kis-1* promoter ............................................................... 43

**Figure 2.3B:** McPromoter analysis of the 5' 896 bp region upstream *CYP6M2* _Kis-2* promoter ............................................................... 44

**Figure 2.3C:** McPromoter analysis of the 5' 896 bp region upstream *CYP6M2* _Tias-1* promoter ............................................................... 45

**Figure 2.3D:** McPromoter analysis of the 5' 896 bp region upstream *CYP6M2* _Tias-2* promoter ............................................................... 46

**Figure 2.4A:** Sequence alignment and TFBS analysis of *Anopheles gambiae* (*CYP6M2*) ............................................................... 47
Figure 2.4B: Sequence alignment and TFBS analysis of *Anopheles gambiae* (CYP6G1) ................................................................. 48

Figure 2.5: TFBS analysis on 5’ 896 bp region upstream of *CYP6M2* gene ............... 49

Figure 2.6A: Multiple amino acid sequence alignments of FBgn0262975 (CnCC) sequences in *Drosophila melanogaster* (Dm) ................................................................. 51

Figure 2.6B: Multiple amino acid sequence alignments of FBgn003513 (Ss) sequences in *Drosophila melanogaster* (Dm) ................................................................. 52

CHAPTER THREE

Figure 3.1: 1% Agarose gels showing the amplification of *CYP6M2* promoter using genomic DNA from *Anopheles gambiae* ................................................................. 64

Figure 3.2A&B: Map and feature of pJET1.2/blunt PCR cloning vector (Thermoscientific) ................................................................. 65

Figure 3.3: Pairwise sequence alignment using ClustalW of the sequenced clones of *CYP6M2_Kis-1* with *CYP6M2_Tias-1* ................................................................. 68

Figure 3.4: Pairwise sequence alignment using ClustalW of the sequenced clones of *CYP6M2_Kis-2* with *CYP6M2_Tias-2* ................................................................. 70

Figure 3.5: 1% Agarose gels showing the amplification of *CYP6M2_Kis* and *CYP6M2_Tias* promoter ................................................................. 71

Figure 3.6: Map of pGL3- Enhancer vector (Promega) ................................................................. 72

Figure 3.7: 1% Agarose gel showing the Xhol and Ncol double restriction digests of the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias plasmids ................................................................. 73

Figure 3.8: 1% Agarose gel showing the Xhol and Ncol restriction digests of the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias plasmids ................................................................. 74

Figure 3.9: 1% Agarose gels showing the digested pGL3-CYP9M10_ISO0P450 and pGL3 CYP9M10_CqSF plasmids ................................................................. 75

CHAPTER FOUR

Figure 4.1: Basal promoter activity in the *Anopheles gambiae* Sua 5.1* cell lines ................................................................. 87

Figure 4.2: Permethrin challenged *Anopheles gambiae* Sua 5.1* cell lines ............... 88

Figure 4.3: TCDD challenged *Anopheles gambiae* Sua 5.1* cell lines ............... 89

Figure 4.4: Microscopic images of A. Untransfected healthy *Anopheles gambiae* Sua 5.1* cells ................................................................. 90
CHAPTER FOUR

Figure 5.1: Relative expression levels (linear fold changes) of the target genes AhR, Nrf2 and CYP6M2

CHAPTER SIX

Figure 6.1A: A proposed general scheme for the induction of Nf2e1/AGAP003645-signaling pathway

Figure 6.1B: A proposed general scheme for the inhibition of Nf2e1/AGAP003645-signaling pathway
LIST OF TABLES

CHAPTER ONE

Table 1.1: P450s involved in insecticide metabolism in *Anopheles gambiae* and *Drosophila melanogaster* ................................................................. 14

CHAPTER TWO

Table 2.1: Percent identity matrix - created by Clustal2.1................................. 40
Table 2.2: McPromoter invertebrate analysis of the CYP6M2 cloned sequences... 40
Table 2.3: The identification orthologs for AhR / ARNT and Nrf2 / ARE of *Homo sapiens* in *Drosophila melanogaster* and *Anopheles gambiae*......................... 50

CHAPTER THREE

Table 3.1: Primer design – CYP6M2 5' 896 bp promoter.................................. 59
Table 3.2: Primer design – CYP6M2 _Tias (631 bp) and CYP6M2 _Kis (591 bp) primers................................................................................................. 61

CHAPTER FIVE

Table 5.1: The strains of mosquitoes used in this study................................. 97
Table 5.2: Primers used in quantitative real-time PCR (qPCR) (F = forward, R= reverse) ................................................................................................. 100
Table 5.3: Normalizing genes (Housekeeping gene)........................................ 101
Table 5.4: Selection of candidate reference genes in permethrin selected *An. gambiae* strains using semi-quantitative qPCR........................................... 103
Table 5.5: Selection of candidate reference genes in permethrin selected *An. gambiae* strains using SYBR® Green (qPCR)..................................................... 103
Introduction and Literature Review
1.0 Introduction and Literature Review

1.1 Brief history of mosquitoes

Mosquito is a Spanish or Portuguese word meaning ‘little fly’. The mosquito belongs to the Class Insecta, Order Diptera (i.e. bearing two wings) and Family Culicidae. It occurs throughout temperate & tropical regions of the world and well beyond the Arctic Circle (WHO, 2013a). There are about 3,500 species of mosquitoes of which the most well-known members belong to two subfamilies, Anophilinae (e.g. Anopheles gambiae) and Culicinae (e.g. Culex quinquefasciatus) (Briggs, 2013; Ashfaq et al., 2014; Khalita et al., 2014). Mosquitoes are regarded as public health enemies, because of their biting annoyance and noise nuisance that lead to sleeplessness, allergic reaction and disease transmission (Wilson et al., 2013; Halasa et al., 2014).

The global distribution of the Anopheles mosquito shows Anopheles gambiae found in Sub-Saharan Africa being the most efficient and predominant species responsible for about 90% of malaria-related deaths (Gabbad et al., 2014; Hemingway, 2014; Onyido et al., 2014) (Figure 1.1).
**Figure 1.1**: A global map of dominant or potentially important malaria vector species. The *Anopheles* mosquito is the global vector of the malaria parasite. *Anopheles gambiae* is the predominant species especially in Sub-Saharan Africa. (Adapted and modified from Sinka et al., 2012).

1.1.1 **Medical and Veterinary importance of mosquitoes**

Mosquitoes have been efficient transmitters of human, livestock and poultry diseases due to their repeated blood feedings during their life span, causing millions of deaths worldwide annually (Raveen et al., 2014; Solomon et al., 2014). Mosquitoes transmit human diseases such as malaria (*Anopheles gambiae* and many other species) (Ndiath et al., 2014), yellow fever (*Aedes aegypti*) (Tomé et al., 2014), dengue hemorrhagic fever (*Aedes aegypti* and many other species) (Beaumier et al., 2014), filariasis (*Anopheles spp*, *Culex spp* and *Aedes spp*) (Govindarajan and Sivakumar, 2014), chikungunya (*Aedes spp*) (Agarwal et al., 2014) and encephalitis (*Culex spp*) (Bradford, 2014). They also transmit livestock diseases such as fowl pox of poultry (*Culex tarsalis*) (LaPointe et al.,...
2012), myxomatosis of rabbits (various spp of mosquitoes) (McColl et al., 2014), rift valley fever of sheep and humans (Culex spp) (Sindato et al., 2014), encephalitis of horses (Aedes spp) (Go et al., 2014), birds (Aedes and Culex spp) (Tao et al., 2014) and dirofilariasis of dogs (Culex, Aedes, and Anopheles spp) (Vieira et al., 2014).

1.2. Mosquitoes as vectors of diseases

The genus Anopheles contains vectors of malaria – a disease of man caused by a protozoan parasite of the genus Plasmodium (Alam, 2014) whilst the Culicinae (which includes the genera Aedes, Culex, and Mansonia), contain the mosquito vectors for various other arboviruses of man & animals and parasitic nematodes such as dirofilariasis in dogs (Srivastava and Syed, 2014). Of all the mosquito-borne diseases, malaria is the most perilous one (Okorie et al., 2014).

1.2.1 Transmission of malaria in man

Malaria parasites are transmitted by female mosquitoes belonging to the genus Anopheles (Onyido et al., 2014). It is dominated by the Anopheles gambiae sensu lato complex groups, eight (8) sibling species, including 2 of the most efficient African human malaria vectors, Anopheles arabiensis (Patton, 1905) and Anopheles gambiae sensu stricto (Gile, 1902) (Arnal et al., 2014). The latter comprises 2 incipient species named Anopheles gambiae (formerly, An. gambiae S form) and Anopheles coluzzii (formerly, An. gambiae M form) which are genetically and biologically isolated from one another through assortative mating (Arnal et al., 2014) and the other being Anopheles funestus group of mosquitoes (Mzilahowa et al., 2012; Alout et al., 2014) in Sub-Saharan Africa.
The male *Anopheles* mosquitoes only feed on flower nectar and plant juices as sources of carbohydrate and cannot transmit malaria (WHO, 2013b). Malarial parasites are therefore transmitted by adult female *Anopheles* mosquitoes (WHO, 2013c; Tumwiine *et al*., 2014), as it is only the female that requires a blood meal in order to obtain enough protein so as to develop a batch of eggs (Artis *et al*., 2014). *An. gambiae* reportedly takes its first blood meal 12 hours after emerging, and will take multiple meals throughout its gonotrophic cycle (Briegel and Horler, 1993; Paaijmans *et al*., 2013). It is the female mosquito’s blood-feeding habit that allows uptake of the gametocyte form of the parasite. This developmental stage of *Plasmodium* is the only stage infective to the *Anopheline* mosquito host (Smith *et al*., 2003). Transmission to another human host occurs when an infective female *Anopheles* mosquito bites an uninfected person and during the blood meal, the mosquito injects sporozoites into the blood stream and that person goes ahead to develop malaria (Tumwiine *et al*., 2014)(Figure 1.2).
1.2.2 Epidemiology of malaria

Malaria continues to be a major global public health problem with about half of the global population (3.3 billion people) at risk in more than 106 endemic countries (Singh et al., 2014). With estimated 0.65 –1.2 million deaths annually, malaria accounts for about 40 to 45 million DALYs (Disability-Adjusted Life Years) (Ursing et al., 2014; Karunamoorthi et al., 2013; 2014; Zofou et al., 2014) and the cost to Africa alone in lost Gross domestic product (GDP) is estimated at £7.13 billion annually which accounts for 40% of the continent’s public health spending (Knox et al., 2014, Abdulai
and Haadi, 2014). Therefore, it imposes not only high morbidity and mortality in human and animal populations, but also negative socio-economic development in the resource-poor settings of Africa, Asia, and Latin America (Karunamoorthi et al., 2014; Singh et al., 2014). One of the approaches for the control of mosquito-borne diseases is the interruption of disease transmission by killing or preventing mosquitoes from biting man and animals (Fanello et al., 1999; Eguyomi et al., 2010).

1.2.3 Mosquito vector control

Vector control is a major component of the global strategy for malaria control which aims to prevent parasite transmission mainly through interventions targeting adult Anopheline vectors (Singh et al., 2014). The major mosquito control methods: indoor residual spray (IRS) and long lasting insecticide–treated bed nets (LLINs) rely heavily on the use of insecticides (Yasuoka et al., 2014; Hemingway, 2014; Killeen et al., 2014).

1.3 Classes of insecticides used in the control of mosquitoes

There are four classes of insecticides licenced for public-health, namely carbamate (esters of carbamic acid), organophosphates (phosphoric acid derivatives), organochlorines (chlorinated hydrocarbons) and pyrethroids (synthetic pyrethroids) (Brooke et al., 2013). Malarial vector control is currently dependent upon pyrethroids, because of their safe, cheap, effective, long lasting nature and their minimal mammalian toxicity (Butler, 2011; WHO, 2013c). Therefore as the focus of this study, more detail will be given on this class of insecticide only.
1.3.1 Pyrethroids insecticides

Pyrethroids are synthesized derivatives of naturally occurring pyrethrins, which are taken from pyrethrum, an extract of dried Chrysanthemum flowers prepared from C. cinerariafolium or less frequently C. coccineum (Ramirez, 2013). Pyrethroids are generally fast-acting poisons if ingested or internalized through direct tarsal contact. They act by paralyzing the nervous system of insects producing the “knockdown” effect, a phenomenon where the pyrethroids target the transmembrane voltage gated sodium channel from the insect nervous system thereby triggering rapid convulsions followed by death (Chandra et al., 2013; Linss et al., 2014).

1.3.2 Permethrin

1.3.2.1 Structure

Almost all pyrethroids including permethrin are esters, composed of an acid and alcohol groups connected by an ester bond. The acid tends to be a substituted cyclopropane carboxylic acid (Schleier and Peterson, 2011). The alcohol groups consist of pyrethrolone for the pyrethrins and phenoxybenzyl groups for the pyrethroids (Figure 1.3).
**Figure 1.3**: Shows the structural formula of permethrin (a pyrethroid) IUPAC: 3-phenoxybenzyl (1RS) -cis, Trans-3-(2, 2-dichlorovinyl)-2, 2 dimethylcy clopropane carboxylate, Molecular formula: C_{12}H_{20}O_{3} and has a molecular weight of 391.29 Source: (http://www.inchem.org/documents/pds/pds/pest51.gif).

### 1.3.2.2 Mode of action of pyrethroids

Pyrethroids function as neurotoxins that act by prolonging sodium channel activation in the nervous system and other excitable cells (Chang *et al*., 2014). However, this remains ideal for treating bed nets due to their safety for humans and other mammals, whilst being effective in killing and repelling mosquitoes even at lower doses (Adams, 2014; N'Guessan *et al*., 2014). Hence, there is no safety concern associated with synthetic pyrethroids because the recommended concentration of the insecticide for treatment of mosquito nets is quite safe with LD50 value of 1.02 ng/mg female for permethrin (Overgaard *et al*., 2014; Surya *et al*., 2014).
1.4 Insecticide resistance in Anopheles gambiae

The resistance to pyrethroids is mainly due to three mechanisms; reduction in sensitivity of the target site, reduced penetration due to altered cuticle and increase in enzyme metabolism (Nardini et al., 2012; Kasai et al., 2014). However, control of mosquitoes remains a challenge even after continuous use of synthetic insecticides such as pyrethroids in public health (Singh et al., 2014). These insecticides are the only class approved for use on insecticide treated nettings (Zaim and Guillet, 2002; Ranson et al., 2011) and are increasingly deployed in IRS (Indoor residual spray) programs in Africa (Ranson et al., 2011). A large number of studies have indicated that multiple insecticide resistance mechanisms involving many genes exist in many insect species, including mosquitoes (Raymond et al., 1989; Hemingway et al., 2002 & 2004; Liu and Scott, 1995, 1996, 1997 & 1998; Liu and Yue, 2000 & 2001; Ranson et al., 2002; Liu et al., 2005, 2007 & 2011; Vontas et al., 2005; Xu et al., 2005; Zhu and Liu 2008; Zhu et al., 2008a & 2008b).

Detoxification enzyme-based resistance occurs when increased activity of cytochrome P450 monooxygenases, GSTs etc results in sequestration or detoxification of the insecticide thereby impairing the toxicity of the insecticide before it reaches its target site (Liu et al., 2006; Stradi, 2012). The cytochrome P450s are the primary enzyme family responsible for pyrethroid metabolism in insects and metabolize insecticides through O-, S-, and N-alkyl hydroxylation, aliphatic hydroxylation & epoxidation, aromatic hydroxylation, ester oxidation, nitrogen and thioether oxidation (Wilkinson, 1976; Brogdon and McAllister, 1998; Zomuanpuii, 2012). It has been well documented that P450-mediated pyrethroid resistant insects have higher levels or more efficient enzyme forms of one or more P450s compared to susceptible Anopheles gambiae (Feyereisen, 1999; Félix and Silveira, 2012) (Figure 1.4).
Cytochrome P450s are generally associated with the enzymatic metabolism of insecticides (Abdalla et al; 2014; Harrop et al., 2014). The detailed studies on these cytochrome P450s essential for this research are therefore enumerated below:

### 1.5 Biology of P450s

#### 1.5.1 Historical framework

The term P450 enzyme is used to describe the enzymes encoded by cytochrome P450 genes, which derive their name from the discovery of a liver microsomal pigment (P) (Omura and Sato, 1962) (Fatma and Wajidi, 2013). Cytochrome P450s are named after their absorption at 450 nm in the UV spectrum that appears when these
enzymes are reduced and saturated with carbon monoxide (Lamb and Waterman, 2013). They are a large superfamily of heme-containing enzymes found in vertebrates, bacteria, plants and insects including *Anopheles gambiae* (Tuan et al., 2014; Kawashima and Satta, 2014).

### 1.5.2 Nomenclature

Cytochrome P450s are designated various names, including cytochrome P450 monooxygenases, polysubstrate monooxygenases (PSMOs), microsomal oxidizes, heme-thiolate proteins and mixed function oxidases (MFOs) (Omura and Sato, 1962; Fatma and Wajidi, 2013). Cytochrome P450s are organised into families (those with >40% amino acid sequence homology are usually grouped into the same family) and subfamilies (those with >55% amino acids sequence homology) (Félix and Silveira, 2012; Nakamura et al., 2012). Gene family numbers are prefixed by the word “CYP” (short for cytochrome P450). For instance those from family 6 are named CYP6. After the family number designation, a letter is used to represent the subfamily e.g. M in CYP6M. Following the subfamily is a number that represents the individual genes present in the specific subfamily e.g. CYP6M2.

### 1.5.3 P450 clans

CYP genes are further classified into clans, families and subfamilies based on phylogenetics as well as sequence identity (Kawashima and Satta, 2014). Clans are defined as “groups of P450 families that consistently cluster together on a phylogenetic tree” (Chen et al., 2014). These are higher-order groups and are basically similar to clades, although clades technically refer to species with a common ancestor and not to sequences (Fatma and Wajidi, 2013). The insect CYP6 and CYP9 families belong in a
clan with vertebrate CYP3 and CYP5. This has been named the CYP3 clan for the lowest family number in the group. Insects have four clans comprising CYP2, CYP3, CYP4, and mitochondrial CYP11, CYP24, CYP27 families (Fatma and Wajidi, 2013).

1.5.4 The CYP6 family

The CYP6 family has been implicated in insecticide resistance more often than any other CYP family to date (Corbel and N’Guessan, 2013). It is found exclusively in insects and is the most extensively studied P450 group in insects (Liu and Scott, 1998; Kasai et al., 1998; Winter et al., 1999; Wen and Scott, 2001; Nikou et al., 2003; Rodpradit et al., 2005; Edi et al., 2012; 2014, Mitchell et al., 2012; 2014). Genes within this family have also been shown to be associated with pyrethroid resistance in An. funestus (Wondji et al., 2009; Matambo, 2008; Matambo et al., 2010; Christian et al., 2011). Other cases of involvement of this family in insecticide resistance have also been documented in the house fly (CYP6A1) (Feyereisen et al., 1989; Andersen et al., 1994; Murtaliev et al, 2008); (CYP6D1) (Tomita et al., 1995, Kasai et al., 2014 and (CYP6A2) Brun et al., 1996; Vázquez-Gómez et al., 2010; Mitchell et al., 2014).

1.5.5 Localisation and function

Cytochrome P450s are categorized as mitochondrial P450s found in the mitochondria and microsomal CYPs located in the endoplasmic reticulum (Zhu et al., 2013). These types of P450s differ in the amino-terminal sequence, which determines the subcellular localization and influences the electron donor system (Werck-Reichhart and Feyereisen, 2000; Omura, 2010). Insect microsomal P450 genes are expressed in many tissues, including the digestive tract and a rich source of P450s is the fat body. Some P450s are larval stage specific, whereas others are expressed only in adults.
(Feyereisen, 1999; Fatma and Wajidi, 2013). P450s also participate in the biosynthesis of juvenile hormone, ecdysteroids, cuticular hydrocarbons and pheromones (Feyereisen, 1999; Félix and Silveira, 2012). Some of these enzymes are involved in the detoxification or in the activation of xenobiotics including insecticides (Nardini et al., 2012; Gong et al., 2013).

1.5.6 Cytochrome P450s in Anopheles gambiae and Drosophila melanogaster

The An. gambiae genome contains 111 P450 genes of which seven are pseudo genes whilst the fruit fly Drosophila melanogaster has 83 P450 genes (Ranson et al., 2002; Fatma and Wajidi, 2013). Amenya (2005) identified 12 CYP4, 12 CYP6 and 7 CYP9 partial genes in Anopheles funestus that have a high percentage sequence similarity to that of An. gambiae. Only one of these was unique to An. funestus (CYP9J14) and showed a low similarity (55%) to An. gambiae CYP9 ortholog. Amenya et al (2008) have further shown that CYP6P9 is overexpressed in a pyrethroid resistant An. funestus laboratory colony. Some common P450s involved in insecticide resistance in Anopheles gambiae and Drosophila melanogaster are highlighted (Table 1.1).

<table>
<thead>
<tr>
<th>Species</th>
<th>P450s</th>
<th>Insecticide substrates</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>CYP6A2</td>
<td>DDT, aldrin, dieldrin, diazinon</td>
<td>1, 2</td>
</tr>
<tr>
<td></td>
<td>CYP6G1</td>
<td>DDT, imidacloprid</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>CYP12A4</td>
<td>Lufeneron</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>CYP12D1</td>
<td>DDT and dicyclanil</td>
<td>4</td>
</tr>
<tr>
<td>An. gambiae</td>
<td>CYP6Z1</td>
<td>DDT, carbaryl</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>CYP6Z2</td>
<td>Carbaryl</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>CYP6P3</td>
<td>Trans- and cis-permethrin, deltamethrin</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>CYP6M2</td>
<td>Permethrin, bendiocarb, DDT</td>
<td>6,7</td>
</tr>
</tbody>
</table>

1.5.7 Role of P450s in insecticide resistance

Apart from their involvement in the endogenous biosynthesis and metabolism, CYPs also play a major role in the metabolism of foreign or xenobiotic compounds including insecticides (Feyereisen, 1999; Kuruganti, 2006). Several studies have identified specific P450s involved in the detoxification of insecticides, plant allelochemicals and promutagens using heterologous expression, reconstitution experiments and isoform specific antibodies (McKenzie and Batterham, 1998). Some of these cytochrome P450 genes are known to be upregulated by the transcription factors Aryl hydrocarbon receptor (AhR) / Aryl hydrocarbon receptor nuclear translocator (ARNT) and or Nuclear factor erythroid-2 related factor-2 (Nrf2) / Kelch-like ECH-associated protein 1 (Keap 1) in higher mammals (Cederbaum, 2013; Deng & Kerppola, 2013; Giantin et al., 2013). Interestingly, Nrf2 / Keap 1 and AhR / ARNT exhibit a multilevel crosstalk where the former is also a target of the later (Pang et al., 2012; Bock, 2014). Previous studies also revealed the orthologs to AhR / ARNT and Nrf2 / Keap 1 in *Drosophila melanogaster* to be Spineless (Ss) / Tango (tgo) and Cap 'n' collar isoform C (CnCC) / *Drosophila* Kelch-like ECH-associatesed protein 1 (dKeap 1) pathways respectively (Misra et al., 2013; Reitzel et al., 2014). Therefore to understand the mechanism by which these P450s are regulated in *Drosophila melanogaster* (an insect model) the basics underpinning this regulation are hereby enumerated.

1.5.8 Regulation of gene expression

The essential part of the control of gene expression is achieved at the transcriptional level (Bhattacharjee et al., 2013). This level of regulation integrates the contribution of multiple types of cis-acting genomic element, which are important molecular switches involved in the transcriptional regulation of a dynamic network of gene activities.
controlling various biological processes, including abiotic stress responses, hormone responses and developmental processes (Yamaguchi-Shinozaki and Shinozaki, 2005; Symmons and Spitz, 2013). This transcriptional regulation takes place within a complex genomic milieu in which enhancers, promoters, and insulators are closely connected both along the one-dimensional linear chromosome and within the three-dimensional nuclear chromatin environment (Stees et al., 2012; Atkinson and Halfon, 2014; Hernandez-Garcia & Finer, 2014).

1.5.8.1 Promoters

Regulation of gene expression at the promoter level is mainly controlled by the cis-acting elements localized upstream of the transcriptional start site (Huang et al., 2014). The physical interaction between regulatory proteins and the basic transcriptional machinery is straightforward during initiation of transcription due to the location of proximal elements to the core promoter. Upstream of the core promoter region are the proximal and distal regions of promoters. Distal promoter elements, located far away from the transcriptional start site can also impact gene expression. The mechanisms of how distal elements come into close proximity to the core promoter to modulate gene expression during transcription involve DNA folding mediated by conformational changes in the 3-dimensional structure of DNA and chromatin (Hernandez-Garcia and Finer, 2014). The core promoter is required for the transcription of eukaryotic RNA polymerase II transcribed genes, typically defined as “consisting of the DNA approximately 35-40 bp upstream and downstream of the transcription start site (TSS)” (Stees et al., 2012). This is at least in part a functional definition in that this region is usually sufficient to mediate gene expression in a reporter gene assay, versatile and sensitive methods of assaying numerous targets in
high-throughput drug-screening programs (Liu et al., 2009; Atkinson and Halfon, 2014). The core promoter contains sequence elements, referred to as “core promoter motifs,” which interact with the basal transcription machinery, including RNA polymerase II and the TFIID (Transcription factor II D, one of several general transcription factors that make up the RNA polymerase II pre-initiation) complex (Figure 1.5).

**Figure 1.5:** Simplified model of transcriptional regulation of protein-encoding genes (Adapted from Hernandez-Garcia and Finer, 2014).

### 1.5.8.2 Enhancers

In contrast to promoters, enhancers are usually located far away from the genes they regulate (Stees et al., 2012; Symmons and Spitz, 2013). Although a promoter is absolutely required for gene transcription, a significant part of metazoan transcriptional regulation occurs via the action of distal cis-regulatory modules (Maston et al., 2006; Atkinson and Halfon, 2014). The best studied of these are transcriptional enhancers, distal non-coding sequences that positively regulate transcription. As originally defined, “enhancers act without regard to orientation, distance or placement (5'/3’) relative to the transcribed gene” (Maston et al., 2006). Enhancers are functionally similar to proximal promoter elements, and the distinction between the
two classes is somewhat vague (Maston et al., 2006). Enhancers therefore serve as a platform for the assembly of transcription factors, including activators and repressors (Atkinson and Halfon, 2014).

1.5.8.3 Repressors

Repressors appear to function by blocking the binding of a nearby activator, or by directly competing for the same site (Gaston & Jayaraman, 2003; Maston et al., 2006; Nie et al., 2013). It has been suggested that the difference between the two may relate to the recruitment of different cofactors.

1.5.8.3.1 Silencers

Silencers are binding sites for negative transcription factors called repressors (Maston et al., 2006). Silencers are sequence-specific elements that confer a negative (i.e., silencing or repressing) effect on the transcription of a target gene (Vooght et al., 2009). Repressor function can require the recruitment of negative cofactors, also called co-repressors, and in some cases, an activator can switch to a repressor by differential cofactor (Maston et al., 2006). In Drosophila, two classes of silencers have been observed: short-range silencers, which generally must reside within ~ 100 bp of their target gene to have a repressive effect, and long-range silencers, which can repress multiple enhancers or promoters over a span of a few kilo base pairs (Maston et al., 2006).

1.5.8.4 Insulators

A third critical component contributing to gene expression is the insulator. Originally defined in Drosophila, and still best understood in that organism, insulators were so
named due to their ability to “insulate” genes from position effects in transgenic assays. Historically, two major roles have been ascribed to insulator elements: the ability to serve as boundary elements preventing the spread of heterochromatin, and the ability to prevent enhancer activity when interposed between an enhancer and promoter (Atkinson and Halfon, 2014).

1.5.9 Regulatory mechanism in resistance

*Drosophila melanogaster* has been used extensively as a model system to understand the molecular mechanisms underlying insecticide resistance (Misra et al., 2013). Further studies also addressed the mechanisms that underlie this regulation, mapping critical promoter elements that are required for P450 gene induction in response to pesticides or the well-studied xenobiotic phenobarbital (PB) (Brun et al., 1996; Maitra et al., 1996; Danielson et al., 1997; Dunkov et al., 1996; Dombrowski et al., 1998; McDonnell et al., 2004; Brown et al., 2005; Maitra et al., 2010). The Spineless / Tango and CnCC / dKeap1 signaling pathways whose orthologs in *Anopheles gambiae* constitute part of this study are hereby enumerated below:

1.6 The Spineless (Ss) gene

The *Drosophila* gene spineless (Ss) is the ortholog of vertebrate AhR (Duncan et al., 1998; Reitzel et al., 2014). These proteins share extensive sequence identity, especially in their basic helix-loop-helix (bHLH) regions, and must share common ancestry, as several of the splice sites in the Ss and AhR genes are precisely conserved (Kuzin et al., 2014).
1.6.1 Function and structure of Spineless

Like other invertebrate homologs of AhR, spineless does not bind prototypical xenobiotic ligands of the vertebrate receptor such as 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD) (McMillan and Bradfield, 2007). Studies revealed that spineless, functions as a heterodimer with Tango, the D. melanogaster ortholog of ARNT and both appear to recognise the same DNA sequence, the xenobiotic response element (XRE), a core nucleotide sequence at the upstream of inducible target genes for the transcription factor aryl hydrocarbon receptor (AhR) that is responsible for recognition of exogenous environmental pollutants in eukaryotic cells (Emmons et al., 1999; Kuramoto et al., 2002). In addition, Tango heterodimerize with Trachealess and Single minded (both bHLH-PAS family members) and regulates transcription in the trachea and central midline, respectively (Zhang et al., 2011).

1.6.2 Ligands of Spineless (AhR)

Spineless (Ss) in D. melanogaster does not show the ability to bind to toxic agonists such as dioxin congener (TCDD) and polycyclic aromatic hydrocarbons (PAHs) (Céspedes et al., 2010). It is therefore conceivable that, as a result of their toxic action, an endogenous ligand capable of activating the Ss protein is generated. Such a role could be played by one of the endogenous AhR ligands– a toxic tryptophan derivative formyl-indolo-carbazole (FICZ) (Kuzin et al., 2014). This protein can however bind to the XRE and stimulate transcription from genes containing this cis-acting element (Emmons et al., 1999, Misra et al., 2011). Moreover, it regulates normal morphogenesis of the leg or antenna and bristles, all of which are major Drosophila sensor organs or tissues that respond to environmental chemicals (Adachi et al., 2001).
1.6.3 Molecular mechanisms of AhR functions in the regulation of cytochrome P450s in Drosophila

1.6.3.1 Spineless (Ss) Tango (Tgo) signalling pathway

The spineless (Ss) protein is known to interact with Tango (Tgo), the fly homolog of mammalian ARNT and through this interaction proteins are transported to the nucleus from the cytoplasm where it binds another bHLH-PAS-protein, the Aryl hydrocarbon receptor nuclear translocator (ARNT). The Ss: Tango heterodimer binds to a specific motif, XRE in the promoters of its target genes and controls their transcription (Swanzon et al., 1995; Kuzin et al., 2014). The Ss: tgo heterodimer can both suppress and activate specific genes, indicating the heterodimer’s interaction with other transcription or nucleosome assembly factors (Kuzin et al., 2014). AhR gene is highly conserved between vertebrates and invertebrates (Kuzin et al., 2014) (Figures 1.6 A and B).
1.7 The Cap ‘n’ Collar Isoform C (CnCC) (named after the cnc gene of *Drosophila*)

Studies of the *Drosophila* orthologs to *Nrf2* and *dKeap1* have provided insights into the functions of this protein. The *Drosophila* Cap’n’collar locus encodes CnCC, which contains a bZIP domain homologous to that of *Nrf2* and N-terminal DTG (Asp-Thr-Gly), a low-affinity motif & a high-affinity ETGE (Glu-Thr-Gly-Glu) motif , separated by a central lysine-rich α-helix and are homologous to those that mediate *Nrf2* interaction with *Keap1* (a member of the Kelch family of actin binding proteins, named...
after the fruit fly’s Kelch protein (a component of the egg chambers) (McMahon et al., 2006).

1.7.1 Function of the CnCC

CncC-Keap1 (dKeap 1) protein complexes regulate both transcriptional responses to xenobiotic compounds as well as native cellular and developmental processes (Grimberg et al., 2011; Deng and Kerpolla, 2013).

1.7.2 Ligands of CnCC

DKeap1 can function as a sensor of oxidants and electrophiles, which react with its redox sensitive cysteine residues (Zhang, 2006; Sykiotis and Bohmann, 2008). Oxidative stresses or electrophilic xenobiotics abolish the inhibition of CnCC by dKeap1 (Itoh et al., 2004; Ma and He, 2012). CnCC is then stabilized and accumulates in the nucleus, where it binds to the Antioxidant Response Element (ARE) in the enhancers of its target genes (Jaiswal, 2004; Sykiotis and Bohmann, 2008).

1.7.3 Molecular mechanisms of CnCC functions in the regulation of cytochrome p450 genes

1.7.3.1 CnCC / dKeap1 signalling pathway

ARE-mediated response to oxidative stress pathway is conserved from flies to humans. In unstressed conditions, Nrf2 (Nuclear factor erythroid-2 related factor- 2) in mammals, and CncC (Cap ‘n’ collar isoform C) in Drosophila are repressed by dKeap1 (Drosophila Kelch-like ECH-associated protein 1), which also functions as a sensor of oxidants and electrophilic compounds (Nioi et al., 2003; Sykiotis & Bohmann, 2008). Under oxidative stress conditions, the inhibition of CncC by dkeap1 are abolished allowing these transcription factors to bind, together with other
proteins, to ARE sequences upregulating downstream genes such as P450s. The *Drosophila dKeap1* contains Kelch repeats homologous to those that mediate *Keap1* interaction with *Nrf2* as well as a sequence motif that is required for mammalian *Keap1* export from the nucleus (Deng and Kerpolla, 2013). Overexpression of *CncC* and depletion of *dKeap1* in *Drosophila melanogaster* activates the transcription of many genes including *CYP6G1* and *CYP6A2* that protect cells from xenobiotic compounds, whereas *dKeap1* overexpression represses their transcription, indicating that the functions of these protein families in the xenobiotic response are conserved between mammals and *Drosophila* (Deng and Kerpolla, 2013, Misra et al., 2013) (Figure 1.7 A&B).
Figure 1:7: General scheme for the induction of CnCC/dKeap 1-signaling pathway. The antioxidant response element (ARE) in the promoter region of select genes allows the coordinated upregulation of antioxidant and detoxifying enzymes in response to oxidative/electrophilic stress. This upregulation is mediated through Cap ‘n’ collar isoform C nuclear (CnCC) that may be activated by endogenous and exogenous molecules or stressful conditions. These agents disrupt the association between CnCC and dKeap1 with subsequent nuclear translocation of CnCC. In the cell nucleus CnCC interacts with small MAF protein, forming a heterodimer that binds to the ARE sequence in the promoter region and upregulates transcription of many genes encoding detoxifying enzymes such as CYP6G1. We therefore speculate that this signalling pathway is constitutively upregulated in long-lived individuals providing extension of longevity and health span. A and B show the conservation of Nrf2 and Keap 1 in Drosophila (Adapted and Modified from Sykiotis and Bohmann, 2008; Guio et al., 2014).

In contrast to the detailed studies in higher mammalians and Drosophila melanogaster, the regulatory mechanism in Anopheles gambiae is yet to be identified.
1.7.4 Other Nuclear receptors in higher mammalian

Nuclear receptors, such as NR1I2, the pregnane X receptor (PXR; also known as the steroid and xenobiotic receptor, SXR) and the related NR1I3, constitutive androstane receptor (CAR) are ligand-activated transcription factors often referred to as “xenobiotic sensors.” (Kubota et al., 2014). These PXR and CAR receptors regulate not only drug metabolism and disposition but also various human diseases such as cancer, diabetes, inflammatory disease, metabolic disease and liver diseases, suggesting that PXR and CAR are promising targets for drug discovery (Banerjee et al., 2014; Anna et al., 2014). Together, CAR and PXR orchestrate an adaptive network for handling toxic derivatives of both endogenous and exogenous compounds, triggering pleiotropic effects in physiological or pathological functions (Masi et al., 2009; Drozdzik et al., 2014).

1.7.4.1 Ligands of PXR and CAR

PXR and CAR can interact with co-activators or co-repressors, depending on the ligands the receptors bind to. Binding of agonistic ligands causes receptor conformational changes that expose the hydrophobic surface within the ligand-binding domain (LBD) for co-activator binding. By contrast, antagonistic ligands induce co-repressor binding, resulting in receptor deactivation (Banerjee et al., 2014). Likewise, AhR, CAR, and PXR regulate the transcription of enzyme and membrane transporter genes, namely a large set of overlapping xenobiotic metabolizing genes–CYP genes (i.e. CYP3A4, CYP2B6, CYP2C, and CYP2A6) (Tolson and Wang, 2010; Drozdzik et al., 2014).
1.7.4.2 PXR signalling pathway

As a member of the nuclear receptor superfamily and ligand-activated transcription factor, PXR forms a heterodimer with retinoid X receptor (RXR)-α and binds to the cognate DNA motifs (PXR responsive element, PXRE) in the regulatory regions of the target genes. Upon the activation by a broad range of xenobiotics, PXR transcriptionally up-regulates the genes for detoxification, including the phase I cytochrome P450 (CYP) enzymes, phase II conjugating enzymes and transporters (Wang et al., 2014). For instance, CXR binds to a previously identified phenobarbital-responsive enhancer unit (PBRU) in the 5′-flanking region of the chicken CYP2H1 gene (Handschin et al., 2000). Phenobarbital (PB) and other drugs of this class of prototypical inducers display a distinct activation pattern of CYPs (Handschin et al., 2000). PXR activation also results in up-regulation of CYP3A4, leading to the excessive metabolism of certain medications, the enhanced toxicity of many xenobiotics, or drug-induced toxic events (Zhang et al., 2002; Lim et al., 2014; Banerjee et al., 2014). Interestingly, crosstalk between AhR, CAR, PXR and other nuclear receptors or transcription factors (including Nrf2) are also involved in controlling endogenous and exogenous signalling pathways (Drozdzik et al., 2014).

Research question

Like other animals, insects including Anopheles gambiae appear to regulate the transcription of detoxification genes in response to xenobiotic challenge. Studies have addressed the mechanisms that underlie this regulation, mapping critical promoter elements that are required for P450 gene induction in response to pesticides or the well-studied xenobiotic phenobarbital (PB) in Drosophila melanogaster (Brun et al., 1996; Maitra et al., 1996; Danielson et al., 1997; Dunkov et al., 1996; Dombrowski et
al., 1998; McDonnell et al., 2004; Brown et al., 2005; Maitra et al., 2010; Misra et al., 2013). It is conceivable that the CnCC / dKeap 1 and/ Spineless / Tango pathways play similar role in upregulating insecticide detoxification in Anopheles gambiae. However, the way, manner and control of this upregulation is not clear. The molecular mechanisms involved in insecticide resistance and the detailed and specific control mechanisms leading to the transcriptional upregulation of P450s, including CYP6M2 are yet to be identified.

**Research goal and specific objectives are:**

To address the need for understanding the regulatory mechanisms involved in the control of cytochrome P450, CYP6M2 gene in insecticide resistant *Anopheles gambiae*, Three objectives will be addressed as follows;

1. To identify and establish the number of putative promoter elements or transcription factor binding sites (TFBS) present in insecticide detoxification P450 (CYP6M2) of *Anopheles gambiae insilico*, especially with respect to Nuclear factor erythroid-2 related factor-2 (Nrf2) and / or Aryl hydrocarbon receptor (AhR) orthologs.

2. To isolate and clone the cis-acting regulatory elements (promoter elements) of CYP6M2 from *Anopheles gambiae*.

3. To examine the expression and regulation of CYP6M2, Nrf2 (CnCC) and AhR (Spineless) orthologs in *Anopheles gambiae* by using semi-quantitative (gel densitometry) and quantitative polymerase chain reaction (qPCR).
CHAPTER TWO

Bioinformatic Analysis of Selected Regulatory Elements within the Promoter Region of the Cytochrome P450 Gene, CYP6M2 in Anopheles gambiae
2.1 Introduction

Cytochrome P450 (CYPs) are known to play a central role in insecticide resistance, allowing resistant insects to detoxify insecticides at a higher rate (David et al., 2013). Among them, members of CYP6M2 gene have been repeatedly implicated in resistance to insecticides (Edi et al., 2014). The development of insecticide resistance to all the commonly used insecticides in *Anopheles gambiae* necessitates the evolvement of novel strategies to combat malaria in Africa (Migliani and Gakhar, 2013; Nkya et al., 2014), often through microarray analyses of resistant strains. Whilst specific genes such as CYP6M2 and CYP6P3 in *An. gambiae* and CYP6P9 in *Anopheles funestus* are repeatedly identified as upregulated in resistant strains and have been demonstrated to be capable of insecticide metabolism *in vitro*, the specific sequence elements responsible for driving the upregulation and hence resistance remain to be identified. Both cis-elements (on the same chromosome, typically immediately upstream of the transcription start site) and trans-elements (distant from the gene—including on separate chromosomes) may have evolved to increase gene expression. However, identifying these is a challenging process necessitating both bioinformatics and experimental techniques. As available sequence resources increase (e.g. the sequencing of 16 *Anopheles* genomes (Neafsey et al., 2014) and over 1000 *An. gambiae* genomes (Holt et al., 2002) therefore bioinformatic methods will aid in understanding the complex mechanisms underlying gene regulation and expression in insecticide resistance (Pavesi et al., 2004). And the first step towards the recognition of appropriate promoters to carry on the expression and regulation of genes is the identification of a transcription factor binding sites (TFBS), which are 5-20 bp in length (Migliani and Gakhar 2013). Here the term promoter is defined ‘as the DNA sequences, usually between 100-1000 base pairs (bp) located upstream of gene coding regions proximal to transcription start
site (TSS) of a gene’. The promoter region contains cis-acting elements, which are specific binding sites for proteins involved in the initiation and regulation of gene transcription (Qiu, 2003; Amit et al., 2011). This transcription is controlled primarily by transcription factors (TFs) which recognize and bind to specific short DNA sequence motifs (Talebzadeh and Zare-Mirakabad, 2014). The most familiar of these motifs is the TATA box (found in 30-40% of core promoters) and located in the region upstream of the genes (Ohler et al., 2002). Transcription factor binding leads to the activation or repression of transcription in response to changes in the environment, as well as during development (Pavesi et al., 2004). Each gene contains a set of unique combination of transcription factor binding sites (TFBSs) in the promoter that determines its temporal and spatial expression (Hu and Gallo, 2010; Pettinato et al., 2014). The TFBS only carry the potential for the gene of interest to bind to the gene promoter. However, the TFBS binding can occur everywhere within the genome and are therefore, not restricted to regulatory regions (Cartharius et al., 2005). The control of gene transcription initiation is a significant mechanism for the determination of gene expression and how much mRNA and consequently protein is being produced (Nain et al., 2011). Therefore, one major task of understanding transcriptional regulation networks is to identify all TFBSs bound by all TFs encoded in a genome, which eventually will provide the information necessary to construct models for describing transcriptional regulatory networks (Lee et al., 2002; Qiu, 2003; Nain et al., 2011). Although recent studies in mosquitoes revealed that, the vast majority of described promoters are from Anopheles and Aedes species (Chen and Rasgon, 2014); no study appears to have been made in understanding the complex mechanisms regulating CYP6M2 gene expression in insecticide resistance in Anopheles gambiae. In this chapter therefore, as a prerequisite for understanding the molecular
mechanism involved in the regulation of CYP6M2, characterization of its promoter sequence and searches for potential regulatory elements were made. Similar searches for the regulatory elements of 5’ upstream region of CYP6G1, a single P450 allele responsible for insecticide resistance in Drosophila melanogaster were also made. This CYP6G1 gene is also known to be upregulated by the orthologs of Nuclear factor erythroid-2 related factor-2 (Nrf2) / Kelch-like ECH-associated protein 1 (Keap 1) and Aryl hydrocarbon receptor (AhR) / Aryl hydrocarbon receptor nuclear translocator (ARNT) (Liu, 2012; Good et al., 2014). The recognition of these TFBSs has the potential to improve knowledge on how wild populations of Anopheles gambiae become resistant to insecticides and are activated by different endogenous and exogenous challenges. This promoter analysis only infers the binding potential and not the functionality of a site. The functionality can only be proven through wet-laboratory experiments with predetermined parameters, particularly since a potential binding site in a promoter can be functional in certain cells and non-functional under different conditions (Cartharius et al., 2005).

2.2 Experimental Approach

2.2.1 Phylogenetic tree of the Anopheles gambiae CYP6M2 gene sequence

In order to analyse the evolutionary similarities, differences and relationships between various insect species and Anopheles gambiae, phylogenetic analysis was carried out. The organization of various taxa into a tree formation presents a hierarchy, which implies common ancestry. Comparative phylogenetic and bioinformatics enable sequence homology within several species to be identified. The phylogenetic analysis of CYP6M2 orthologs in Anopheles gambiae was carried out using Ensembl Metazoa Genome Browser (http://metazoas.ensembl.org/Anophelesgambiae/Info/Index). The
Ensembl Metazoa used in this study is built primarily from data maintained by two community-based genomics resources for different types of invertebrate metazoa: FlyBase (Drysdale, 2008) (focused on *Drosophila*) and VectorBase (Lawson et al., 2009) (National Institute of Allergy and Infectious disease (NIAID) resource for invertebrate vectors of human pathogens) (focused on *Anopheles gambiae*). The nucleotide sequences of the *Anopheles gambiae* (AGAP008212), *Culex* spp (GU188856), *Drosophila* spp (AY010608), *Anopheles darlingi* (ADMH00000000) were retrieved using the aforementioned respective data bases. Sequence alignment was performed by using CLUSTAL-X software (<http://www.clustal.org>), version 2.1.

### 2.2.2 Genomic DNA and mRNA sequence retrieval

In order to identify the promoter sequences with respect to the *CYP6M2* from both the Tiassalé and Kisumu strains, VectorBase database was used to retrieve sequences from the *Anopheles gambiae* genome (<http://www.vectorbase.org>). The National Centre for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) and the Ensembl Metazoa Genome Browser (<http://www.ensembl.org/index.html>) data bases were also used. Furthermore, the 5' upstream region of (*CYP6G1*) (896 bp) in *Drosophila melanogaster* was also retrieved and analysed along with the *CYP6M2* promoter sequences using the same tools to determine the degree of conservation in both hypothesised promoter sequences from the insecticide resistant Tiassalé & susceptible Kisumu strains of *Anopheles gambiae* and perhaps function as described below. In order to successfully clone the 5' upstream *CYP6M2* gene, it was cloned into pJET1.2/blunt cloning vector, pJET-CYP6M2_Tiassalé (Tias) (930 bp) and pJET-CYP6M2_Kisumu (Kis) (896 bp) in Chapter Three of this study.
2.2.3 *Insilico analysis of CYP6M2 gene sequence and CYP6M2 gene promoter*

In order to establish the regulatory regions within the CYP6M2 gene promoter, the presence and location of transcription factor binding sites in the 5' regulatory regions of the retrieved CYP6M2 gene sequence and promoter sequence analysis was done using several bioinformatics tools as described below. The sequenced constructs were also analysed for the presence of putative GATA transcription binding factors considered as positive cis-acting regulatory elements sequences (Kadalayil et al., 1997; Félix, 2011).

2.2.4 *Recognition of promoter / enhancer elements within the cloned and sequenced CYP6M2_(Tias) and CYP6M2_(Kis) gene promoters*

In order to identify putative regulatory motifs within the cloned and sequenced CYP6M2_(Kis) and CYP6M2_(Tias) gene promoter sequences in Chapter three of this study, several online tools were used. The CpG Finder tool was used to predict the existence of cPG dinucleotides within the gene promoter sequence (cPGisland.islands.USC.edu). Clusters of cPGs / GC rich regions called cPG islands, (regions containing >200 cPG sequences) are often found close to genes, with cPG sequences appearing in the promoters and first exons but also in downstream regions. These so-called cPG islands (cPG dinucleotides in invertebrates), first defined by Bird in 1986, ‘are on average 400–500 bp of length, have a C+G content of 0.5 or higher and an observed to expected cPG ratio of 0.6 or higher within a range of 200 bp or greater’ (Krinner et al., 2014). The sequenced CYP6M2_Kis-1 & CYP6M2_Kis-2 and CYP6M2_Tias-1 and CYP6M2_Tias-2 promoters were analysed using cPG island searcher (http://cPGislands.usc.edu/) (Sharif et al., 2010). The TSSW (Recognition of human PolII promoter region and start of transcription) tool predicts the existence of enhancer and general transcription factor binding sites (e.g. TATA box). Further analysis using McPromoter
defined as “a probabilistic promoter prediction system that identifies likely TSSs in large genomic sequences tool” were also made (Ohler et al., 2002) (McPromoter MM: II http://genes.mit.edu/McPromoter.html).

2.2.5 Prediction of putative regulatory elements within the cloned and sequenced CYP6M2 gene promoters

In order to predict the putative transcription factor binding sites (TFBS) within the cloned plasmids in Chapter Three of this study and the sequenced CYP6M2_Kis and CYP6M2_Tias gene promoters, an online analysis using ConSite (http://asp.ii.Uib.no:8090/cgi-bin/CONSITE/consite/) was performed. The putative transcription factor binding sites for AhR / ARNT and Nrf2 / ARE were searched using Consite web site (http://consite.genedreg.net/cgi-bin/consite?rm=tinput) (Sandelin et al., 2004; Miglani and Gakhar, 2013). ConSite is a user friendly web-based interface that is used conventionally to identify cis-regulatory elements within genomic sequences. Similarly, prediction of conserved Anopheles gambiae (CYP6M2) and Drosophila melanogaster (CYP6G1) putative TFs in the promoter region (Consite). In order to predict the conservation between the regulatory regions of CYP6M2 gene (896 bp) and CYP6G1 gene (896 bp) putative TFs in the promoter regions, the ConSite tool was used (http://consite.genedreg.net/cgi-bin/consite?rm=tinput) (Sandelin et al., 2004; Neafsey et al., 2014).

2.2.6 Insilico Identification of Aryl hydrocarbon receptor (AhR), Nuclear factor erythroid -2 related factor-2 (Nrf2) orthologs and their receptors

2.2.6.1 Identification of AhR / ARNT orthologs

In order to search for the orthologs of AhR in Anopheles gambiae, Ensembl genome browser (http://www.ensembl.org/Multi/Search/Results?q=AhR;y=3;site=ensembl
was used to search for the AhR gene in Homo sapiens. Whilst Fly base (http://flybase.org/reports/FBgn0003513.html) data base was used to search for its ortholog in Drosophila melanogaster (Dm) and VectorBase data base (https://www.vectorbase.org/Anopheles_gambiae/Gene) was used to search for the orthologs in Anopheles gambiae (Ag).

2.2.6.2 Identification of Nrf2 / ARE orthologs

In order to search for the orthologs of Nrf2 / ARE in Anopheles gambiae, Ensembl genome browser (http://www.ensembl.org/Human/Search/Results?q=Nrf2;m%20y=%203;site=ensemblall;x=3;;page=1;facet_species=Human) was used to search for the Nrf2 gene in Homo sapiens. Whilst Fly base (http://flybase.org/reports/FBgn0262975.html) data base was used to search for its orthologs in Drosophila melanogaster (Dm) and VectorBase data base (https://www.VectorBase.org/Anopheles_gambiae/Gene) was used to search for the orthologs in Anopheles gambiae (Ag).

2.2.7 Multiple amino acid sequence alignments of the Nrf2/AhR orthologs

2.2.7.1 Multiple amino acid alignments of CnCC from Drosophila melanogaster, Nf2e1 from Anopheles gambiae and Nrf2 from Homo sapiens

Multiple sequence alignments demonstrate conserved amino acid sequences. The online gene sequence alignment tool VectorBase (http://www.vectorbase.org) was used to align previously isolated CnCC from Drosophila melanogaster, Nf2e1 from Anopheles gambiae and Nrf2 from Homo sapiens. This programme aligned the promoters of these three gene orthologs to show their level of conservation.
2.2.7.2 Multiple amino acid sequence alignments of Spineless from Drosophila melanogaster, AGAP010259 from Anopheles gambiae AhR from Homo sapiens

Multiple sequence alignments demonstrate conserved amino acid sequences. The online gene sequence alignment tool VectorBase (http://www.vectorbase.org) was used to align previously isolated Spineless from Drosophila melanogaster, AGAP010259 from Anopheles gambiae and AhR from Homo sapiens. This programme aligned the promoters of these three gene orthologs to show their level of conservation.

2.3 Results
2.3.1 Phylogenetic tree of CYP6M2 gene sequence

In order to determine the evolutionary similarities, differences and relationships between various species of mosquitoes, phylogenetic analysis was carried out. The organization of various taxa into a tree formation presents a hierarchy, which implies common ancestry. Comparative phylogenetics and bioinformatics enables sequence homology within several species to be identified. A phylogenetic analysis with respect to Anopheles gambiae CYP6M2 gene sequence is between different species of insects, a phylogenetic tree of the Anopheles gambiae CYP6M2 gene sequence was obtained from Ensembl (Ensembl Genome Browser).

Figure 2.1 shows the inferred evolutionary relationship between various species based on CYP6M2 gene sequence similarities. The Anopheles gambiae CYP6M2 gene sequence is the reference sequence in this regard. This result indicates that the Anopheles gambiae CYP6M2 gene sequence shares particular homology with members of the Diptera order.
Figure 2.1: Phylogenetic tree of CYP6M2 gene sequence (Ensembl Genome Browser). The above tree was generated by using the Ensembl data base and takes gene orthologs and paralogs into account. The phylogenic tree highlights “CYP6M2, Anopheles gambiae” in red font. Paralogues have been shown in blue.

For instance, in Drosophila melanogaster and Anopheles gambiae, between 41% and 73% of the known orthologs genes remain linked in the respective homologous chromosomal arms (Bolshakov et al., 2002; Krinner et al., 2014).

2.3.2 CYP6M2 promoter genomic sequence retrieval from the genomic database

An 896 bp sequence which is located upstream of the CYP6M2 TSS was retrieved from VectorBase. Upstream of this region is a large transposable element (TE) also known as "jumping genes" or transposons, which are sequences of DNA that move (or jump) from one location in the genome to another either directly by a cut-and-
paste mechanism (transposons) or indirectly through an RNA intermediate (retro transposons) (Pray, 2008; Fedoroff, 2012). This TE is located upstream of the CYP6M2 gene between 6,928,002 to 6,928,908 bp and it is hypothesised that the promoter elements must be between this TE and the TSS (Figure 2.2).

![Image of CYP6M2 promoter gene](https://www.vectorbase.org/Anopheles_gambiae/Location/View?g=AGAP008212;r=3R:6928012-6928907;t=AGAP008212-RA)

Figure 2.2: Shows image of the retrieved CYP6M2 promoter gene indicating the CYP6M2 protein coding (Red boxes) entire forward and reverse strands of 896 bp (Black arrows).

Figure 2.2 shows the image of the region in details on the location of the 896 bp CYP6M2 gene on the 3R chromosomes (3R:6928012-6928907) while tRNA is the forward strand.

### 2.3.3 Cytochrome P450s promoter searching and analysis of Anopheles gambiae, Aedes aegypti, Drosophila melanogaster, Culex quinquefasciatus and Homo sapiens

The sequences comparison of 896 bp promoter region of Anopheles gambiae, Aedes aegypti, Drosophila melanogaster, Culex quinquefasciatus and Homo sapiens showed similarity using percent identity matrix (Table 2.1).
Table 2.1: Percent identity matrix - created by Clustal2.1

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>AAEL009496</th>
<th>AGAP008212</th>
<th>Fbgn0000473</th>
<th>CPIJ016809</th>
<th>ENSG00000059377</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAEL009496</td>
<td>100.00</td>
<td>44.99</td>
<td>37.26</td>
<td>35.27</td>
<td>35.18</td>
</tr>
<tr>
<td>AGAP008212</td>
<td>44.99</td>
<td>100.00</td>
<td>36.18</td>
<td>37.77</td>
<td>36.79</td>
</tr>
<tr>
<td>Fbgn0000473</td>
<td>37.26</td>
<td>36.18</td>
<td>100.00</td>
<td>42.40</td>
<td>40.75</td>
</tr>
<tr>
<td>CPIJ016809</td>
<td>35.27</td>
<td>37.77</td>
<td>42.40</td>
<td>100.00</td>
<td>43.55</td>
</tr>
<tr>
<td>ENSG00000059377</td>
<td>35.18</td>
<td>36.79</td>
<td>40.75</td>
<td>43.55</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Nucleotide homologies between CYP6M2 896 bp Anopheles gambiae (AGAP08212). The CYP6M2 gene was compared with Aedes (AAE009496), Culex (CPIJ016809), Drosophila (Fbgn0000473) (CYP6A2) and Homo sapiens (ENSG000000059377) (CYP5A1) species 896 bp promoters. Matrix indicates the percentage in identities of the aligned nucleotide sequences from Anopheles, Aedes, Culex, Drosophila and Homo sapiens promoters.

Table 2.1 shows that the CYP6M2 sequences were 44.9% identical between ENSG000000593 (CYP5A1) 35.18% and Anopheles gambiae, 37.26% identical between Anopheles gambiae and Drosophila, and so forth. This can read either on the first line of the matrix or down the first column. By the way, the scores are decided solely on the basis of the number of nucleotides that are different. The comparison is more important between Aedes, Culex and Drosophila than in Humans.

Table 2.2: The McPromoter invertebrate analysis and prediction of TSSs within the 5’ upstream region in the cloned CYP6M2 sequences

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Length (bp)</th>
<th>TATA box (TSS)</th>
<th>Threshold cPG count</th>
<th>Ratio of O/E (%GC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP6M2_Kis-1</td>
<td>915</td>
<td>-91</td>
<td>-97</td>
<td>+ 0.06703</td>
</tr>
<tr>
<td>CYP6M2_Kis-2</td>
<td>896</td>
<td>-78</td>
<td>-108</td>
<td>+ 0.04599</td>
</tr>
<tr>
<td>CYP6M2_Tias-1</td>
<td>929</td>
<td>-91</td>
<td>-50</td>
<td>+ 0.05115</td>
</tr>
<tr>
<td>CYP6M2_Tias-2</td>
<td>929</td>
<td>-91</td>
<td>-50</td>
<td>+ 0.05115</td>
</tr>
</tbody>
</table>

Key: Length in base pairs (bp), TATA box position, Transcription start site (TSS)-position relative to the first coding base in the sequence, cPG count- Total cPG count within the sequence, Ratio of observed over expected cPG dinucleotide and % G+C (GC content)= 50.50 - 53.00.
Tables 2.2 shows the prediction of putative CpG dinucleotides performed using the CpG finder at the European Bioinformatics Institute (EBI), (http://www.ebi.ac.uk/).

The following parameters: CpG length >200 bp, G+C >50% and a “CpG value” of at least 0.75 were used. Further analysis of promoter region and transcription start site prediction using McPromoter invertebrate analysis tool (Ohler et al., 2002; Ryazansky et al., 2011) predicted the existence of a promoter at position 360 and a TATA box at positions -91 bp (CYP6M2_Kis-1), -78 bp (CYP6M2_Kis-2), -91bp (CYP6M2_Tias-1) and -91bp (CYP6M2_Tias-2) respectively.

The cPG count is the number of CG dinucleotides. The percentage cPG is the ratio of cPG nucleotide bases (twice the cPG count) to the length. The ratio of observed to expected cPG is calculated according to the formula cited in (Gardner-Garden and Frommer, 1987). CpG dinucleotide analysis by carrying out the cPG dinucleotide search, we generated data on the cPG dinucleotide content in the Tiassalé (Tias-1 & Tias-2) resistant and Kisumu (Kis-1 & Kis-2) susceptible strains of Anopheles gambiae (Table 2.2). The region of cPG dinucleotide content is more or less similar with each other, although the length of the cPG is different from each other. The GC contents of these strains range from 50.50-53.00%. The transcription start site (TSS) of the susceptible Kisumu and resistant Tiassalé strains is different from each other with CYP6M2_Tias-1 and CYP6M2_Tias-2 (-50 bp) at a lower position than CYP6M2_Kis-1 (-97bp) and CYP6M2_Kis-2 (-108 bp) respectively.

2.3.4 Prediction of putative promoter elements

McPromoter was used to search for promoter elements within the Kisumu and Tiassalé clones. McPromoter is a probabilistic promoter prediction system that identifies likely location of eukaryotic RNA polymerase II transcription start sites.
(TSSs) in large genomic sequences (Ohler et al., 2002). The McPromoter analysis incorporates data on DNA structure and binding site information thereby generating a graph that displays the likelihood of a promoter occurring at each TSS position. In each of the cloned CYP6M2_Tias and CYP6M2_Kis promoter regions, the CpG dinucleotide (enclosed by the highest threshold Figures 2.3 A-D) were identified as the most likely location of a promoter element. The TSS has a predictive score of 0.00999, on a range of -0.5 to 1 in which higher values indicate a greater likelihood of promoter activity (Figures 2.3A-D).
Figure 2.3A: McPromoter analysis of the 5' 896 bp region upstream CYP6M2_Kis-1 promoter. Green lines indicate the boundaries of CpG dinucleotides identified using the UCSC browser (http://tools.genome.duke.edu/generegulation/McPromoter/). The height of the graph at any point indicates the likelihood of a promoter activity occurring at that site, on a scale of -0.5 to 1 with 1 being optimal. The threshold used is for an intermediate sensitivity of 50%, for which the threshold range for all the clones is between +0.04599 to +0.06703. However, the CpG dinucleotide content is identified as the most likely to contain a promoter element. Lowering the threshold will increase the sensitivity and the number of promoters predicted, but will also increase the chance of false-positives.
Figure 2.3B: McPromoter analysis of the 5' 896 bp region upstream CYP6M2_Kis-2 promoter. Green line indicates the boundaries of CpG dinucleotides identified using the UCSC browser (http://toolsgenome.duke.edu/generegulation/McPromoter/). The height of the graph at any point indicates the likelihood of a promoter activity occurring at that site, on a scale of -0.5 to 1 with 1 being optimal. The threshold used is for an intermediate sensitivity of 50%, for which the threshold range for all the clones is between +0.04599 to +0.06703. However, the CpG dinucleotide content is identified as the most likely to contain a promoter element. Lowering the threshold will increase the sensitivity and the number of promoters predicted, but will also increase the chance of false-positives.
Figure 2.3C: McPromoter analysis of the 5’ 896 bp region upstream CYP6M2_Tias-1 promoter. Green line indicates the boundaries of CpG dinucleotides identified using the UCSC browser (http://tools.genome.duke.edu/generegulation/McPromoter/). The height of the graph at any point indicates the likelihood of a promoter activity occurring at that site, on a scale of -0.5 to 1 with 1 being optimal. The threshold used is for an intermediate sensitivity of 50%, for which the threshold range for all the clones is between +0.04599 to +0.06703. However, the CpG dinucleotide content is identified as the most likely to contain a promoter element. Lowering the threshold will increase the sensitivity and the number of promoters predicted, but will also increase the chance of false-positives.
Figure 2.3D: McPromoter analysis of the 5’ 896 bp region upstream CYP6M2_Tias 2 promoter. Green line indicates the boundaries of CpG dinucleotides identified using the UCSC browser (http://tools.genome.duke.edu/generegulation/McPromoter/). The height of the graph at any point indicates the likelihood of a promoter activity occurring at that site, on a scale of -0.5 to 1 with 1 being optimal. The threshold used is for an intermediate sensitivity of 50%, for which the threshold range for all the clones is between +0.04599 to +0.06703. However, the CpG dinucleotide content is identified as the most likely to contain a promoter element. Lowering the threshold will increase the sensitivity and the number of promoters predicted, but will also increase the chance of false-positives.
2.3.5 Prediction of conserved *Anopheles gambiae* (CYP6M2) and *Drosophila melanogaster* (CYP6G1) putative transcription factor binding sites (TFBS) in the promoter regions (Consite)

Previous studies have shown that sequence conservation between species can be used to predict functional TFBS. Pairwise alignment and subsequent analysis have shown the two following transcription factor binding sites using Consite website (http://consite.genereg.net/cgi-bin/consite?rm=tin put) as being conserved between the *Anopheles gambiae* and *Drosophila melanogaster* (Figures 2.4A and 2.4B).

![Conservation profile of *Anopheles gambiae*](image)

**Figure 2.4A:** Sequence alignment and TFBS analysis of *Anopheles gambiae* (CYP6M2) gene promoter using conservation cut off score of 30% and TF score threshold of 80% (Phylofoot.org). The *Anopheles gambiae* (CYP6M2 5’ 896 bp) sequence was reverse complimented (The Sequence Manipulation Suite). The “Analyze orthologs pairs of genomic sequences” tool (http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite?rm=t_input) from ConSite was used to visualise TFBS conservation (Phylofoot.org). The sequences were transcribed into the ConSite tool and the *AhR/ARNT* and *Nrf2/ARE* were selected to produce the above graphs.
Figure 2.4B: Sequence alignment and TFBS analysis of *Drosophila melanogaster* (CYP6G1) gene promoter using conservation cut off score of 30% and TF score threshold of 80% (Phylofoot.org). The *Drosophila melanogaster* (CYP6G1 5’ 896 bp) promoter sequences were reverse complimented (The Sequence Manipulation Suite). The “Analyze orthologs pairs of genomic sequences” tool (http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite?rm=t_input) from ConSite was used to visualise TFBS conservation (Phylofoot.org). The sequences were transcribed into the ConSite tool and the AhR/ARNT and Nrf2/ARE were selected to produce the above graphs.
2.3.6 Putative transcription factor binding sites found along 5’ 896 bp upstream region of in Anopheles gambiae at 70% cut off score

The 5’896 bp region upstream of CYP6M2 sequence analysis at 70% TF cut off score revealed the following transcription factor binding sites (TFB’s), 39 AhR-ARNT and 10 Nrf2. The TFBS are the putative binding sites where the AhR/ARNT and or Nrf2/Keap 1 orthologous bind to drive the transcription of CYP6M2 ( Figure 2.5).

Figure 2.5: TFBS analysis on 896 bp CYP6M2 promoter. Reverse compliment (http://www.Bioinformatics.org/sms/index.html) was performed on the sequenced CYP6M2. The reverse compliment sequence was analysed using the “analyze single sequence” tool available from ConSite (Phylofoot.org,). Sequence analysis at 70% cut off score or stringency revealed a total of 39 AhR/ARNT and 10 Nrf2/ARE putative transcription binding sites were found within the 5’ 896 bp CYP6M2 gene.

2.3.7 Identification of the Nrf2/ARE and AhR/ARNT orthologs

The orthologs to mammalian ARNT is Tango in Drosopila melanogaster and AGAP009748 in Anopheles gambiae. Similarly, orthologs to Nrf2 in higher mammalian is identified as CnCC in Drosophila melanogaster and Nf2e1 in Anopheles gambiae
respectively. The orthologs to AhR in higher mammalian is Spineless (Ss) and AGAP010259 in *Drosophila melanogaster* and *Anopheles gambiae* respectively (Table 2.4). Searches were made using VectorBase (focused on *Anopheles gambiae*), Flybase database (focused on *Drosophila melanogaster*) and Ensembl (focused on *Homo sapiens*) to identify the orthologous for the respective gene promoters (Table 2.3).

### Table 2.3: The orthologs to AhR /ARNT and Nrf2 / ARE of Homo sapiens in *Drosophila melanogaster* and *Anopheles gambiae*.

<table>
<thead>
<tr>
<th>Mammalian</th>
<th><em>Drosophila melanogaster</em></th>
<th><em>Anopheles gambiae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>AhR</td>
<td>Ss</td>
<td>AGAP010259</td>
</tr>
<tr>
<td>ARNT</td>
<td>Tgo</td>
<td>AGAP009748</td>
</tr>
<tr>
<td>Nrf2</td>
<td>CnCC</td>
<td>Nf2e1</td>
</tr>
<tr>
<td>Kea 1</td>
<td>dKeap 1</td>
<td>AGAP003645</td>
</tr>
</tbody>
</table>

**Key:** AhR (Aryl hydrocarbon receptor), ARNT (Aryl hydrocarbon receptor nuclear translocator), Ss (Spineless), Tgo (Tango), Nrf2 (Nuclear factor erythroid-2 related factor-2), CnCC (Cap ‘n’ collar isoform C), Nf2e1 (Nuclear factor erythroid 2, invertebrate).

Whilst the orthologs to higher mammalian Keap 1 in *Drosophila melanogaster* is dKeap 1 and AGAP003645 in *Anopheles gambiae* respectively.

#### 2.3.8 Multiple amino acid alignments of CnCC from *Drosophila melanogaster*, Nf2e1 from *Anopheles gambiae* and Nrf2 from *Homo sapiens*

Amino acid sequences of Nf2e1 from *Anopheles gambiae*, CnCC from *Drosophila melanogaster* and Nrf2 from *Homo sapiens* were made to show the conservation between the promoters of these three genes (Figure 2.6A).
Chapter Two

B.R.Mohammed

Abertay University, Dundee, UK

**Figure 2.6A:** Multiple amino acid sequence alignment of FBgn0262975 (CnCC) sequences *Drosophila melanogaster* (Dm), with AGAP005300 *Anopheles gambiae* (Ag) and ENSG00000116044.11 (Nrf2) *Homo sapiens* (Hs) species. Where * indicate identical residues; : conserved substitutions; . semi-conserved substitutions. The alignment was performed using ClustalX (http://www.ebi.ac.uk/). Dashes are used to denote gaps introduced for a maximum alignment.

**2.3.9 Multiple amino acid sequence alignments of Spineless from Drosophila melanogaster, AGAP010259 from Anopheles gambiae AhR from Homo sapiens**

Amino acid sequences of AGAP010259 from *Anopheles gambiae*, spineless from *Drosophila melanogaster* and AhR from *Homo sapiens* were made to show the conservation between the promoters of these three genes (Figure 2.6B).
## 2.4 Discussion

The focus of this chapter was to identify the region upstream of *CYP6M2* gene in *Anopheles gambiae* involved in its regulation, characterization of its promoter sequence and searches for its selected potential regulatory elements. The recognition of these AhR/ARNT and Nrf2/ARE TFBSs has the potential to improve knowledge on how wild populations of *Anopheles gambiae* become resistant to insecticides and are activated by different endogenous and exogenous challenges. *Anopheles gambiae* genomic resources were used to assess the 5' upstream region hypothesised to contain the promoter region of *CYP6M2*, a gene associated with the detoxification of insecticides including permethrin (Mitchell *et al.*, 2012; Edi *et al.*, 2014). The P450

---

**Figure 2.6B**: Multiple amino acid sequence alignment of FBgn003513 (Ss) sequences *Drosophila melanogaster* (Dm), with AGAP010259 *Anopheles gambiae* (Ag) and ENST000242057 (AhR) in *Homo sapiens* (Hs) species. Where * indicate identical residues; : conserved substitutions; . semi-conserved substitutions. The alignment was performed using ClustalX (http://www.ebi.ac.uk/). Dashes are used to denote gaps introduced for a maximum alignment.
superfamily in insects has a diverse sequence and the overall homology may be less than 40% (Wang et al., 2014a). Conserved putative TFBS within the cloned and sequenced *Anopheles gambiae* promoter were therefore identified. To establish the differences between the pJET-CYP6M2_Tias and pJET-CYP6M2_Kis promoters, two separate analyses were carried out to firstly predict the position of the promoter, TATA box and Enhancer sequences, and secondly to predict the position of CpG dinucleotides in the cloned and sequenced DNA. The CYP6M2 sequence contains a number of CpG dinucleotides, TATA box and a promoter sequence. CpG dinucleotides could stimulate gene expression in several ways due to their binding affinity to powerful transcription factors such as Sp1 (Specificity protein 1) (Kadalayil et al., 1997). Organisms such as the invertebrates *Drosophila melanogaster*, *Caenorhabditis elegans* and the fungus *Saccharomyces cerevisiae* have little or no DNA methylation and, as a result, CpG dinucleotides occur at the expected frequency throughout the genome (Deaton and Bird, 2011). DNA methylation in invertebrates is associated with transcription level, alternative splicing and genome evolution (Rivière, 2014). Findings from this study are in accord with the conclusions that ‘The promoters of vertebrate and invertebrate organisms differ in that invertebrate genomes do not contain CpG islands, a feature for more than half of vertebrate genes’ (Gardiner-Garden and Frommer, 1987), but consistent with the presence of CpG dinucleotides in invertebrates including *Anopheline* spp (Elsango et al., 2009; Feliciello et al., 2013). Comparison of *Homo sapiens* AhR / Nrf2 with their orthologs in *Drosophila melanogaster* and *An. gambiae* through multiple peptide sequence alignment show significant degree of conservation in a number of regions (Figures 2.5 A and B). According to the Bioinformatics analysis of *An. gambiae* cytochrome P450 related genes including CYP6M2, significant progress in understanding the complex process of regulation of this gene by AhR (AGAP010259)
and Nrf2 (Nf2e1) (AGAP005300) have been achieved. It appears that this work only focuses on CYP6M2, but the significance of the work is far more important than the CYP6M2 itself. The approach in this work can be carried out with other cytochrome P450 genes to gather more information on the pathways involved in the regulation of genes involved in the metabolism of insecticides. This research therefore is a fundamental work in the fields of the bio-informational analysis, and also put forward a new way for the bioinformatics analysis of other detoxification genes.

2.5 Conclusions

The present findings established in silico presence of putative AhR / ARNT (AGAP010259 / AGAP009748) and / or Nrf2 / Keap 1 (AGAP005300 / AGAP003645) binding sites in Anopheles gambiae. This potentially suggests that the cytochrome P450 CYP6M2 is under the control of Nrf2 through Antioxidant response element (ARE) and or AhR through Xenobiotic response element (XRE) in the 5'-flanking region. It is therefore proposed that AhR / ARNT or Nrf2 / Keap 1 orthogous pathways are involved in the cellular network that maintains redox homeostasis in order to protect cells from oxidative stress induced by their respective ligands in Anopheles gambiae. Verification could be made by subsequent wet-lab experiments such as dual luciferase assay in chapter four and / or quantitative polymerase chain reaction (qPCR) in chapter five of this study. Cloning of the analysed 5'896 bp upstream region of CYP6M2 gene and construction of its luciferase gene assay system is a prerequisite for the dual luciferase assay studies.
CHAPTER THREE

Cloning of *Anopheles gambiae* CYP6M2 Gene Promoter and Construction of its Luciferase Gene Reporter System
3.1 Introduction

Cytochromes P450 (CYP) is one of the largest and most functionally diverse classes of heme-containing enzymes found in nature (Estabrook, 2003; Guo et al., 2013). These P450s constitute one of the oldest enzyme superfamilies’ and are present in all living organisms including mammals, plants, bacteria, and insects (Zhu et al., 2013; Chen et al., 2014; Qin et al., 2014). Cytochromes P450 are found in almost all tissues and are important in the biosynthesis of several endogenous compounds, as well as metabolism of xenobiotics (Chauhan et al., 2013). Insecticides are used widely in the control of insects of public health importance including control of mosquito vectors of malaria (Hemingway et al., 2002). Increased insecticide detoxification mediated by over-expressed P450s is a common mechanism of insecticide resistance (Wan-Norafikah et al., 2013a). Only a limited number of cytochromes P450 have been identified through microarray studies of the insecticide resistance phenotype as being repeatedly implicated in the resistance phenotype. With the increased threat to malaria vector control caused by insecticide resistance attributed to P450s, there has been an interest in understanding the role and underlying mechanisms of this resistance (Munhenga and Koekemoer, 2011) which will be useful in the development of more sensitive diagnostic tests for effective monitoring of metabolic based resistance development. In the *Anopheles gambiae* genome alone, there are 111 annotated P450 cytochromes (Ranson et al., 2002; Félix and Silveira, 2012; Zhu et al., 2013). Studies have revealed that the *CYP6M2* gene in *Anopheles gambiae* is established to be directly involved in the acquirement of insecticide resistance (Stevenson et al., 2011; Félix and Silveira, 2012; Mitchell et al., 2012; Witzig et al., 2013; Zimmer et al., 2014). Whilst P450s such as *CYP6M2* have been identified as having a role in insecticide resistance, the regulatory mechanisms underpinning the over-expression of
this gene (and other P450s) is yet to be ascertained. Some cytochrome P450 genes are
known to be up regulated by Cap ‘n’ collar isoform C (CnCC) / Drosophila Kelch-like
ECH-associated protein 1 (dKeap 1) and Spineless (Ss) / Tango (tgo) in Drosophila
melanogaster (McMillan and Bradfield, 2007; Misra et al., 2013). These are orthologs to
Nuclear factor erythroid-2 related factor-2 (Nrf2) / Kelch-like–ECH associated protein
1 (Keap 1) and Aryl hydrocarbon receptor (AhR) / Aryl hydrocarbon receptor nuclear
translocator (ARNT) signalling pathways in mammals respectively (Wang et al., 2013;
Maayah et al., 2013; Phillipson, 2014; Vorrink et al., 2014). This mechanism is yet to be
identified in Anopheles gambiae and is discussed further in Chapter Five of this study.

An essential part of gene regulation involves the attachment of a regulatory protein
complex to the gene promoter. Bioinformatic study of the promoter region is detailed
in the previous chapter of this study (Chapter Two). Promoter regions can be also
studied in vitro using reporter assays where the promoter activity is used to drive
expression of a reporter gene, usually expression of luciferase. Previous studies in
insects have used the dual luciferase gene reporter assay to demonstrate the role of
promoter variants in the up regulation of cytochrome P450 genes involved in
insecticide resistance (Wilding et al., 2012; Wan–Nokarifah et al., 2013b). In this
chapter, as a prerequisite towards understanding the molecular mechanisms involved
in insecticide resistance in Anopheles gambiae, the cloning of the putative CYP6M2 gene
promoter and construction of its luciferase gene reporter system were undertaken.
These are quintessential tools in the dual luciferase reporter assay designated as
further work.
3.2 Experimental Approach

3.2.1 Mosquito strains

The Kisumu strain of *Anopheles gambiae*, a laboratory insecticide susceptible strain originally colonised from Kisumu, western Kenya in 1953 (Alout et al., 2013) and a wild multiple insecticide resistant Tiassalékro (Tiassalé) (Konan et al., 2012) strain originally collected from northwest of Abidjan in Ivory Coast (Chouaibou et al., 2012) were used in this study. These mosquito strains were kindly provided by the Liverpool Insect Testing Establishment (LITE) unit of the Liverpool School of Tropical Medicine (LSTM), United Kingdom.

3.2.2 Designing primers for PCR

In order to isolate the putative promoter sequence of the *CYP6M2* gene, specific primers to amplify this region were designed and analysed for physical properties using primer 3.0 (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and OligoAnalyzer 3.1 (Integrated DNA technologies, IDT) (https://eu.idtdna.com/analyzerApplications/OligoAnalyzer/). In order to maximise the chance of capturing the promoter elements upstream of *CYP6M2*, the reverse primer was constrained to be within the 5' UTR and situated as close as possible to the start codon. The forward primer was constrained to be within a tRNAlys sequence approximately 900bp 5' of *CYP6M2*. From the PEST genome, the primers designed would amplify an 896bp sequence contiguous with the *CYP6M2* coding sequence. In order to ensure primer specificity, stringent primer design parameters were utilised. The Nucleotide BLAST tool from VectorBase was used to perform a genome wide homology and alignment searches using (https://www.vectorbase.org/clustalw) to verify the specificity of the designed primers to the *Anopheles gambiae CYP6M2* upstream sequence (Table 3.1).
Table 3.1: Primer design – CYP6M2 896 bp putative promoter

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Length (bp)</th>
<th>GC (%)</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA-1 (FP)</td>
<td>5’ GAACCCACGACCCTGAGAT 3’</td>
<td>19</td>
<td>57.9</td>
<td>56.6</td>
</tr>
<tr>
<td>START (RP)</td>
<td>5’ ATTTTGGAAACGCGAGGAG 3’</td>
<td>19</td>
<td>47.4</td>
<td>53.3</td>
</tr>
</tbody>
</table>

Key: FP-Forward primer, RP-Reverse primer

These primers were designed with no internal restriction enzyme; this is because the pJET1.2/ cloning vector has BglII restriction sites on either side of the multiple cloning sites (MCS) (Figure 3.2).

3.2.3 BlastX sequence analysis

In order to search for sequences that are similar to the retrieved CYP6M2 promoter (query) sequence, a wide range of web-based tools were used for the analysis of CYP6M2 gene putative promoter sequences as described below. The upstream sequences of CYP6M2 were submitted for analysis through BLASTX. BLAST was accessed through NCBI (http://www.ncbi.nlm.nih.gov/) to search for sequence similarity. To compare reverse and forward strand sequences, the reverse strand was reverse complemented using the reverse-complement program web site (http://www.Bioinformatics.org/sms/revcomp.html) and the pairwise alignment undertaken using the VectorBase database (https://www.Vectorbase.org/search/site/CYP6M2) respectively.

3.2.4 Isolation of genomic DNA and polymerase chain reaction (PCR)

Genomic DNA was extracted from both Kisumu and Tiassalé strains of the Anopheles gambiae using the Qiagen DNEasy kit (Cat. # 6956, Qiagen, Valencia, CA, USA). The concentration and purity of the isolated DNA were analysed using a Nanodrop™
1000 Spectrophotometer. The specific primers designed (Section 3.2.2) for the region upstream of \textit{CYP6M2} were used to amplify this region from the genomic DNA of susceptible Kisumu (Kis) and resistant Tiassalé (Tias) strains of \textit{Anopheles gambiae} (Table 3.1). The PCR conditions using HOT start DNA polymerase (Cat. \# F-549S / L, Thermoscientific, Pittsburgh, PA, USA) were as follows: Initial denaturation at 98 °C for 30 s, followed by 30 cycles at 98 °C for 10 s, 57 °C for 15 s, and 72 °C for 30 s with a final 5 min extension at 72 °C.

3.2.4.1 Cloning and sequencing of \textit{CYP6M2} promoter

The PCR products obtained as a result of the amplification of the genomic DNA from both the Tiassalé and Kisumu strains templates were purified using a Thermoscientific DNA purification kit (Cat. \# K0701, Thermoscientific, Pittsburgh, PA, USA) and ligated into CloneJET™ PCR cloning system (Cat. \# K1231, Pittsburgh, PA, USA) and sequenced. Ligation reactions were carried out by incubating 10 µL of 2X reaction buffer, 1 µL vector (50 ng / µL), 2 µL purified PCR product (100 ng / µL), 1 µL T4 DNA ligase (1U / µL ) and 6 µL sterile distilled water (to 20 µL) at 22 °C for 30 min. Ligation mixtures were transformed into XL1-Blue competent cells (Cat. \# 200249, Stratagene, Santa Clara, CA, USA) and positive colonies were identified by colony PCR using vector specific primers. Sequencing was undertaken by Source Biosciences, Nottingham, United Kingdom. Consensus sequences for both strains were aligned using EMBOSS (http://www.ebi.ac.uk/emboss/align/).

3.2.4.2 DNA Sequencing

DNA sequence homology was further examined in terms of percentage, using the ClustalW2 online tool from EMBL-EBI (http://www.ebi.ac.uk/) (EMBL-EBI). This
programme aligned the susceptible Kisumu (pJET\textsubscript{CYP6M2}-Kis) (3.5.4.1.1 and 3.5.4.1.2) clones with the resistant Tiassalé (pJET\textsubscript{CYP6M2}-Tias) (3.5.4.1.3 and 3.5.4.1.4) clones promoter sequences (retrieved from the previous sequencing results) to produce an alignment profile (EMBL-EBI).

### 3.3 Construction of luciferase gene reporter vectors

#### 3.3.1 Design of primers for PCR

In order to insert the putative promoters for \textit{CYP6M2\textsubscript{Kis}} and \textit{CYP6M2\textsubscript{Tias}} inserts into reporter vectors it was necessary to add restriction enzyme recognition sequences using modified primers designed from the sequences of pJET-CYP6M2\textsubscript{Kis} and pJET-CYP6M2\textsubscript{Tias}. \textit{Xho}1 and \textit{Nco}1 restriction sites (bold and underlined) were inserted at the forward and reverse primers respectively to maintain correct insert orientation during ligation into the pGL3-Enhancer Vector (Table 3.2).

#### Table 3.2: Primer design – \textit{CYP6M2\textsubscript{Tias}} (631 bp) and \textit{CYP6M2\textsubscript{Kis}} (591 bp) promoter

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Length (bp)</th>
<th>GC (%)</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{CYP6M2\textsubscript{Tias}(FP)}</td>
<td>5’ GCT\textsubscript{CTCGAG} GTTGATAA\textsubscript{ACTGAGCCGC} 3’</td>
<td>26</td>
<td>57.7</td>
<td>62.8</td>
</tr>
<tr>
<td>\textit{CYP6M2\textsubscript{Tias}(RP)}</td>
<td>5’ GACGC\textsubscript{CCATGG} TGATAG\textsubscript{TTCGCTGCG} 3’</td>
<td>26</td>
<td>61.5</td>
<td>65.2</td>
</tr>
<tr>
<td>\textit{CYP6M2\textsubscript{Kis}(FP)}</td>
<td>5’ GCG\textsubscript{CTCGAG} AGATCT\textsubscript{TCATGCGCACACG} 3’</td>
<td>28</td>
<td>60.7</td>
<td>66.1</td>
</tr>
<tr>
<td>\textit{CYP6M2\textsubscript{Kis}(RP)}</td>
<td>5’ GG\textsubscript{CCTCGAG} GGCAAGG\textsubscript{GCGTCTAAAGG} 3’</td>
<td>27</td>
<td>66.7</td>
<td>68</td>
</tr>
</tbody>
</table>

Key: FP-Forward primer, RP-Reverse primer

#### 3.3.2 Isolation of the \textit{CYP6M2\textsubscript{Tias}} (631 bp) and \textit{CYP6M2\textsubscript{Kis}} (591 bp) putative promoter using polymerase chain reaction (PCR)

The designed specific primers (\textit{CYP6M2\textsubscript{Tias}} (631 bp) and \textit{CYP6M2\textsubscript{Kis}} (591 bp) (Table 3.2) were used to amplify this region from the pJET-CYP6M2\textsubscript{Kis} and pJET-CYP6M2\textsubscript{Tias} clones. The PCR conditions using My Taq Red Mix polymerase (Cat. #
BIO-25043, Bioline, London, NW, UK) were as follows: Initial denaturation at 95°C for 1 min, followed by 35 cycles at 95°C for 15 s, 64°C for 15 s, and 72°C for 10 s with a final 4 min extension at 72°C. The CYP6M2_Kis and CYP6M2_Tias inserts were cut out from the pJET1.2/blunt cloning vector by XhoI and NcoI double enzyme restriction digests and gel purified (Cat. # 28704, Qiagen, Hilden, Germany) before ligation into the linear pGL3-Enhancer vector (Cat. # E1741, Promega, Madison, WI, USA). Recombinant pGL3-Enhancer vector was transformed into E. coli and spread onto 5 mL LB agar plates containing 100 µg/mL ampicillin and incubated overnight for colonies to grow. Individual colonies (Section 3.3.3) were then picked and cultured overnight in LB broth containing 50 µg/mL ampicillin at 37°C with shaking in order to grow up for storage as glycerol stock.

3.3.3 Colony PCR to screen for successful ligations

In order to ascertain the presence of the insert within the pGL3-Enhancer vector and in the correct orientation, CYP6M2_Tias (FP) & CYP6M2_Tias (RP) and CYP6M2_Kis (FP) & CYP6M2_Kis (RP) (Table 3.2) were respectively used for the Kisumu and Tiassalé strains of Anopheles gambiae. The PCR conditions used were as in Section 3.3.2 of this chapter.

3.3.4 Restriction digests of CYP6M2_Kis, CYP6M2_Tias and pGL3-Enhancer vector

The pGL3-Enhancer vector was digested using XhoI and NcoI restriction enzymes. The pGL3-Enhancer vector was purified using Qiagen midi-prep kit (Cat. # 28704, Qiagen, Hilden, Germany) before ligation with the digested CYP6M2_Kis and CYP6M2_Tias inserts.
3.4 Preparation of positive control plasmids for dual luciferase assay

Two clones from the insecticide susceptible (pGL3-CYP9M10_CqSF) and insecticide resistant (pGL3-CYP9M10_ISOp450) strains of Culex quinquefasciatus with a known promoter activity were also prepared. These clones will be used as positive control in the dual luciferase assay in chapter four of this study.

3.4.1 KpnI and HindIII double restriction digests of pGL3-CYP9M10_ISOP450 (1504 bp) and pGL3-CYP9M10_CqSF (804 bp) plasmids from Culex quinquefasciatus

In order to confirm if the pGL3-CYP9M10_ISOP450 and pGL3-CYP9M10_CqSF plasmids are carrying the CYP9M10 promoter inserts in the plasmids from Culex quinquefasciatus clones, a double restriction digest was set up using KpnI (Cat # FD0524, Thermoscientific, Pittsburgh, PA, USA) and HindIII (Cat # FD0504, Thermoscientific, Pittsburgh, PA, USA) restriction enzymes according to manufacturer's conditions. These enzymes were selected because these promoters were modified to have the HindIII and KpnI restriction sites which cut the pGL3 Enhancer vector at the multiple cloning.

3.5 Results

3.5.1 Sequence identification of CYP6M2_Kis and CYP6M2_Tias promoter

The sequence upstream of CYP6M2 (Accession number AGAP008212) hypothesised to contain the putative promoter sequence was retrieved from VectorBase. Only approximately 1000 bp upstream of CYP6M2 was considered as likely to contain the putative upstream regulatory region as a result of possible presence of transposable elements further 5’ of this. Using Primer3 (version 0.40), primers to amplify this upstream region from wild and laboratory colonies of An. gambiae were designed
(Table 3.1) which would amplify an 896bp fragment (as adjudged from the PEST sequence in VectorBase).

### 3.5.2 Polymerase chain reaction (PCR) and DNA sequencing of CYP6M2 5’ upstream region from susceptible and resistant An. gambiae strains

The obtained PCR products from the amplification of the genomic DNA of Kisumu and Tiassalé strains of *Anopheles gambiae* as template using the specific primers designed for the CYP6M2 gene (Table 3.1) produced amplicons from the Kisumu and Tiassalé DNA of 896 bp and 930 bp respectively (Figure 3.1).

![Figure 3.1. 1% Agarose gels showing the amplification of CYP6M2 promoter using genomic DNA from Anopheles gambiae. Five (5µL) each of (Kisumu -L2 & L3 (White box) and Tiassalé- L4 & L5 (Yellow box) mixed with 1 µL of loading dye (Bioline) all with a molecular weight of about 1 Kb. M is 5 µL molecular weight maker 10 Kb (Bioline) loaded. The gel was run for 40 min at 120V.](image)

### 3.5.3 Cloning of CYP6M2 putative promoter

In order to undertake cloning of CYP6M2_Kis (896 bp) and CYP6M2_Tias (930 bp) PCR products, the amplified sequences were respectively ligated into the Clone JET PCR cloning vector.
3.5.3.1 Map and Features of pJET1.2/blunt cloning vector

Figures 3.2 shows the map of pJET1.2/blunt cloning vector with BglII restriction sites on both flanks of the multiple cloning sites (MCS)(A) and the feature of the map showing the DNA sequences of the MCS region (B).

Figure 3.2: A. Map of pJET1.2/blunt cloning vector (Thermoscientific). The pJET1.2/blunt cloning vector has been linearized with Eco32I (EcoRV) (GenBank/EMBL Accession number EF694056). The blunt ends of the vector contain 5’-phosphoryl groups. B. The DNA sequence of multiple cloning system (MCS) region with the forward and reverse sequencing primers. Source: (http://www.thermoscientificbio.com/ligation/).
3.5.4 DNA sequencing

The PCR products were cloned, purified and sent with pJET1.2 forward and pJET1.2 reverse primers for DNA sequencing and the cloned sequences were analysed using the NCBI blast tool. Two sequences from the Kisumu strain and two from the Tiassalé strain were generated (Section 3.3.3.1 - 3.3.3.4). These sequences varied in length and sequence from that of the PEST strain of An. gambiae retrieved from VectorBase. CYP6M2_Kis 1 (914 bp) has 94% and CYP6M2_Kis 2 (878 bp) has 91.5% similarity and identity to Anopheles gambiae PEST chromosome 3 (R). Whilst CYP6M2_Tias 1 (928 bp) has 92.6% and CYP6M2_Tias 2 (928 bp) has 92.6 % similarity and identity to Anopheles gambiae PEST chromosome 3 (R).

3.5.4.1 DNA Sequence inputs

3.5.4.1.1 >CYP6M2_Kis-1- First Input Sequence

GAACCCACGACCCTCTAGATAGTTAGCTATCTATCTCTATCCAGCTAGCTAGCTAGGGGCTGT
GATAACCTAGCCGCTCTTTACACGATAGTTAGCTATCTATCTCTATCCAGCTAGCTAGCTAGGGGCTGT
GCCACAGCTTTTGGCGAAGATACAGCTGCACGGCGCAAGAAGGCAAGGCGCAGCTTAAGGAT
AGATTAGAGCGATGCTGCTACCTACACGATAGTTAGCTATCTATCTCTATCCAGCTAGCTAGGGGCTGT
ATACACACTACCAACCCCAAAACACTACTGTTTACTCTCTGCGCAGTTTTACACACTTACAC
ATACACACTACCAACCCCAAAACACTACTGTTTACTCTCTGCGCAGTTTTACACACTTACAC
CATTGCAGCTCTCTCTCTATACTACTGTTTACTCTCTGCGCAGTTTTACACACTTACAC
ACACACACACCGAATAGTGGTGCTGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC

3.5.4.1.2 >CYP6M2_Tias-1- Second Input Sequence

GAACCCACGACCCTCTAGATAGTTAGCTATCTATCTCTATCCAGCTAGCTAGCTAGGGGCTGT
GATAACCTAGCCGCTCTTTACACGATAGTTAGCTATCTATCTCTATCCAGCTAGCTAGCTAGGGGCTGT
GCCACAGCTTTTGGCGAAGATACAGCTGCACGGCGCAAGAAGGCAAGGCGCAGCTTAAGGAT
AGATTAGAGCGATGCTGCTACCTACACGATAGTTAGCTATCTATCTCTATCCAGCTAGCTAGGGGCTGT
ATACACACTACCAACCCCAAAACACTACTGTTTACTCTCTGCGCAGTTTTACACACTTACAC
ATACACACTACCAACCCCAAAACACTACTGTTTACTCTCTGCGCAGTTTTACACACTTACAC
CATTGCAGCTCTCTCTCTATACTACTGTTTACTCTCTGCGCAGTTTTACACACTTACAC
ACACACACACCGAATAGTGGTGCTGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC

3.5.4 DNA sequencing

The PCR products were cloned, purified and sent with pJET1.2 forward and pJET1.2 reverse primers for DNA sequencing and the cloned sequences were analysed using the NCBI blast tool. Two sequences from the Kisumu strain and two from the Tiassalé strain were generated (Section 3.3.3.1 - 3.3.3.4). These sequences varied in length and sequence from that of the PEST strain of An. gambiae retrieved from VectorBase. CYP6M2_Kis 1 (914 bp) has 94% and CYP6M2_Kis 2 (878 bp) has 91.5% similarity and identity to Anopheles gambiae PEST chromosome 3 (R). Whilst CYP6M2_Tias 1 (928 bp) has 92.6% and CYP6M2_Tias 2 (928 bp) has 92.6 % similarity and identity to Anopheles gambiae PEST chromosome 3 (R).
3.5.4.1.3 >CYP6M2_Kias-2- First Input Sequence

CTAGCCGGCGCTTTGTAATACTGAGCCGCGCTTTGCTACACCACTGTTAGCTAACAGACAGCATGAGAGATCTTCATGCGCACACGTTTTGCGACGCATCACTGCGCAGCCGCAAGAAGGCCGGTCTAAAGGTAGATTGAAGCCGATGCTTGAGATAAGAAAAAATCAGTTACAGTGTTTATTATTACCTAAAATTTGAATTATAGAGAATAGCAGTTACTTAATCAATGTATGTAGTTACTTAATCATCTATTATTTACTGCATTTATTCATTATGTATTCCTTTATCAATTTTTGCTTCAAAGTCATTGTCCCTATCAATACACTCACACCCCCAAAAC

GTAAGATTTCTAACTCTTCGCGGATATTTACATATTACTCACCACATTCCTATGTTTGATCTACGCACACCCTTCTCTAACCCTCCTCAATACCACAAGCTCAATTTGATACACTTCAAAATTTATGCCTATGTCATCGTTTCGCACCCGGTTCCCATCGTTCACTTGGACCTTCTTCTTAATCTTATCTTTTTTTATATTTTGTGCAAACGACTCATAGCACCCCCCGGGAATACACCACGCGGACGGACTCTTCACGGTGGGATCATGCTTACTTTATTGA

The EMBOSS needle nucleotide alignment tool (http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html) was used to examine the homology between the sequenced DNA (Kis-1 kb and Tias-1 promoter) (Ensemble Genome Browser). Figure 31 shows 878/934 bp (94.0%) identity and a 94.0% sequence similarity. Discrepancy only occurred in 27/934 base pairs accounting for 2.9% of the entire sequence. The discrepancy demonstrated by the sequence alignments shows that there are differences between the Tiassalé (Tias-1) and Kisumu (Kis-1) sequences (Figure 3.3 & Figure 3.4).
Cloning of *Anopheles gambiae CYP6M2* Chapter Three

Figure 3.3. Pairwise sequence alignment using ClustalW ([http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html)) of the sequenced clones of CYP6M2_Kis-1 with CYP6M2_Tias-1. The discrepancies in the alignment of the nucleotides from the sequenced clones have been indicated by boxes.
3.5.5.2 DNA input sequence for Kis-2 and Tias-2

The EMBOSS needle nucleotide alignment tool (http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html) was used to examine the homology between the sequenced DNA (Kis-2 and Tias-1 promoter) (Ensemble Genome Browser). Figure 3.8 shows 847/929bp (91.2%) identity and a 91.2% sequence similarity. Discrepancy only occurred in 50/929 base pairs accounting for 5.4% of the entire sequence. The discrepancy demonstrated by the sequence alignments shows that there are differences between the susceptible Kisumu (Kis-2) and resistant Tiassalé (Tias-2) sequences. Sequence data revealed the putative promoter sequence of the region upstream of CYP6M2 has 94% (pJET-CYP6M2_Kis 1; 914 bp), 91.5% (pJET-CYP6M2_Kis 2; 878 bp), 92.6% (pJET-CYP6M2_Tias 1; 928 bp) and 92.6% (pJET-CYP6M2_Tias-2; 928 bp) similarity to the retrieved Anopheles gambiae PEST strain data base sequence.
Figure 3.4. Pairwise sequence alignment using ClustalW (http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html) of the sequenced clones of CYP6M2_Kis-2 with CYP6M2_Tias-2. The discrepancies in the alignment of the nucleotides from the sequenced clones have been indicated by boxes.
Chapter Three

3.6 Construction of luciferase gene reporter assay vectors

3.6.1 PCR Amplification of the CYP6M2_Kis and CYP6M2_Tias promoters

The designed primers (Table 3.2) were used to amplify and isolate the CYP6M2_Kis and CYP6M2_Tias putative promoters using the sequenced pJET-CYP6M2_Kis and pJET-CYP6M2_Tias sequenced promoters respectively as template.

Figure 3.5 shows the 1% agarose gel image of the successful isolation of the CYP6M2_Kis (631 bp) and CYP6M2_Tias (591 bp) promoter fragments.

Figure 3.5: 1% Agarose gels showing the amplification of CYP6M2_Kis and CYP6M2_Tias promoter using the sequenced pJET-CYP6M2_Kis and pJET-CYP6M2_Tias from Anopheles gambiae. Five (5µL) each of CYP6M2_Kis (L1) and CYP6M2_Tias (L2)(Yellow box) mixed with 1 µL of loading dye (Bioline) all with a molecular weight of about 1 Kb. M is 5 µL molecular weight maker 10 Kb (Bioline) loaded. The gel was run for 40 min at 120V.

NcoI and XhoI restriction enzymes were used to perform double enzyme restriction digests of pGL3-Enhancer vector. This was possible owing to the presence of XhoI and NcoI recognition sites on both sides of the multiple cloning regions in the pGL3-Enhancer vector (Figure 3.6).
3.6.2 Map and Features of pGL3 Enhancer vector

Figures 3.6 shows the map of pGL3 Enhancer vector with XhoI and NcoI restriction sites on either flanks of the multiple cloning sites (MCS).

![Map of pGL3 Enhancer vector (Promega).](http://www.promega.co.uk/~media/Files/Resources/Protocols/Technical%20Manuals/0/pGL3%20Luciferase%20Reporter%20Vectors%20Protocol.pdf)

The concentration of each plasmid was adjusted to 600 ng/µL and determined using a Nanodrop™ 1000 Spectrophotometer.

3.6.3 Colony PCR screening using CYP6M2_Kis (FP) & CYP6M2_Kis (RP) and CYP6M2_Tias (FP) & CYP6M2_Tias (RP) primers

In order to confirm the presence of the CYP6M2_Kis and CYP6M2_Tias inserts within the pGL3-Enhencer vector and in the right orientation the designed primers in Table
3.2 were used to amplify the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias plasmids respectively. The presence of bands of about 600 bp after the colony screening further confirms the presence of the insert and verifies its orientation within the pGL3-Enhancer vector (Figure 3.7).

![Figure 3.7: 1% Agarose gel showing the XhoI and NcoI double restriction digests of the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias plasmids. L1 and L12 showed bands of about 630 bp and 591 bp for the CYP6M2_Kis and CYP6M2_Tias inserts (Red box) and pGL3-Enhancer vector (Orange box) respectively. While M is 5 μL of the molecular weight maker (10 Kb from Bioline) loaded. The gel was run for 40 min at 120 V.](image)

3.6.4 XhoI and NcoI double enzyme restriction digests of the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias gene constructs

In order to confirm the presence of the insert within the constructed plasmids, XhoI and NcoI double restriction digests of the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias gene constructs was carried out on the Mini prep plasmid samples to further verify the successful ligation of the gel purified plasmids. This was made possible due to the presence of XhoI and NcoI recognition sites on either side of the multiple cloning sites (Figure 3.2) (Figure 3.8).
Cloning of Anopheles gambiae CYP6M2  

Chapter Three

B.R. Mohammed - 74 - Abertay University, Dundee, UK

Figure 3.8: 1% Agarose gel showing the Xho1 and Nco1 restriction digests of the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias plasmids. L1 and L2 showed bands of about 600 bp for the CYP6M2_Kis and CYP6M2_Tias inserts (Orange box) and pGL3-Enhencer vector (White box) respectively. While M is 5 µL of the molecular weight maker (10 Kb from Bioline) loaded. The gel was run for 40 min at 120 V.

3.6.5 KpnI and HindIII double restriction digests of pGL3-CYP9M10_ISOP450 (1504 bp) and pGL3-CYP9M10_CqSF (804 bp) plasmids from Culex quinquefasciatus

Double restriction digest of the plasmid DNA samples was carried out to verify the successful ligation of insert DNA into the pGL3-Basic vector. Both pGL3-ISOP450 (1504 bp) and pGL3-CqSF (804 bp) constructs were double digested. KpnI and HindIII restriction enzymes were used. (Red box - 804 bp) are clearly visible on the agarose gel (Figure 3.9).
Cytochromes P450 (CYP) genes are often involved in the detoxification of insecticides associated with insecticide resistance in *Anopheles gambiae* ([Kasai et al.](#), 2014; [Mitchell et al.](#), 2014; [Edi et al.](#), 2014). *CYP6M2*, a member of the CYP6 family has been repeatedly identified as upregulated in insecticide resistant mosquitoes. It is demonstrated to be capable of insecticide metabolism leading to insecticide resistance, thereby making this gene an attractive target for further studies. However, prior to this study no previous works with the notable exception of Félix (2011) had been undertaken about the *CYP6M2* gene regulation in *Anopheles gambiae*. In order to confirm that *CYP6M2* is involved in insecticide resistance, it is necessary to demonstrate that this P450 gene is able to metabolise pyrethroid insecticides. This has been done ([Mitchell, Edi](#)). Through further study of the mechanisms involved in the regulation of this gene it may be possible to identify regulatory variants which can be used in diagnostic testing. As a prerequisite towards understanding this mechanism in
Anopheles gambiae, the cloning of the putative CYP6M2 gene promoter and construction of its luciferase gene reporter assay system are quintessential. To date, P450s belonging to the CYP6 family have been cloned from many insects, such as the cotton bollworm, Helicoverpa armigera (CYP6B7) (Zhao et al., 2014), (CYP6B6) (Liu et al., 2013), the beetle Dastarcus helophoroides (CYP6BQ) (Wang et al., 2014b), Silkworm, Bombyx mori (CYP6U1) (Zeng et al., 2010), housefly (CYP6D1) (Lin and Scott, 2011), fruit fly Drosophila melanogaster (CYP6A2) (Wan et al., 2013), Oriental Fruit fly, Bactrocera dorsalis (CYP6G2) (Zuo and Chen, 2014). The vast majority of these cloned CYP genes from different insect species are implicated in the metabolism of exogenous compounds such as plant toxins (Zhang et al., 2011) through gene expression studies. In mosquitoes, Culex quinquefasciatus (CYP9M10) (Wilding et al., 2012), Culex pipiens pallens (CYP6F1) (Gong et al., 2005) and Aedes aegypti (CYP4H28v2) (Fatma and Wajidi, 2013), CYP4 genes were all successfully cloned for in vitro studies of metabolic potential.

In this study, the 5’ upstream region of CYP6M2 gene hypothesised to contain the promoter was amplified by PCR using genomic DNA from both Kisumu and Tiassalé strains of Anopheles gambiae. The concentration of the genomic DNA was 100 ng / μL and 110 ng / μL for the Kisumu and Tiassalé strains respectively, though concentration measurement is not always reliable and reproducible using the Nanodrop in comparison to fluorescent nucleic acid stain (Pico Green) (Wilding et al., 2009) but were more than the required 5 to 10 ng / μL for a successful PCR. In order to confirm the size of the CYP6M2-Kis and CYP6M2_Tias within the vector, XhoI and NcoI restriction digests of the pGL3-CYP6M2_Kis and pGL3- CYP6M2_Tias plasmids were carried out as shown in Figure 3.6. The cloning of CYP6M2 and construction of pGL3-CYP6M2_Tias and pGL3-CYP6M2_Kis luciferase gene reporter system for dual
Cloning of *Anopheles gambiae* CYP6M2

**Chapter Three**

Luciferase assay in chapter four of this study are therefore confirmed. This procedure is consistent with previous works of Wilding et al., (2012) for *CYP9M10* in *Culex quinquefasciatus*, Wan et al., (2013) for *CYP6A2* in *Drosophila melanogaster*, and Félix (2011) for *CYP6M2* in *Anopheles gambiae*. This study has provided a basis for further studies on the gene function associated with insecticide resistance, which will improve the understanding of the regulatory mechanism in insecticide resistance.

### 3.8 Conclusions

In conclusion, the core promoters of the *CYP6M2* gene in *Anopheles gambiae* are reported for the first time in this study. The 5' region upstream of *CYP6M2* gene from both Kisumu and Tiassalé strains were successfully isolated from *Anopheles gambiae*. The insilico analysis of the sequenced *CYP6M2_Kis* and *CYP6M2_Tias* has confirmed several consensus motifs in the aligned *CYP6M2* 5' regions which are hypothesised to contain the promoter motifs involved in the regulation of *CYP6M2* expression. The *CYP6M2* luciferase gene reporter systems were also successfully constructed. In future work therefore, luciferase reporter assays will be conducted in order to investigate the effects of xenobiotic substances such as permethrin on *CYP6M2* up stream region hypothesised to contain the promoter.
Transcriptional Response Elements in the Promoter of CYP6M2, an Insect P450 Gene Regulated in the Mosquito Anopheles gambiae
4.1 Introduction

The control of disease vectors of human and veterinary importance including mosquitoes relies heavily on the use of synthetic chemical insecticides (Diabaté et al., 2013; Karunamoorthi and Sabesan, 2013; Ho and Zairi, 2013; Dhanasekaran and Thangaraj, 2014). In addition to the persistent nuisance, they also transmit a number of diseases through blood feeding causing morbidity, mortality, discomfort, and heavy economic losses (Land and Miljand, 2014; Thuilliez et al., 2014; Schäfer and Lundström, 2014). Insecticides play a central role in the control of mosquito vectors, including *Anopheles gambiae*, the vector of malaria. However, this is increasingly threatened by the development of resistance to every chemical class of insecticides, such as organochlorines, carbamate, organophosphate pyrethroid and insect growth regulators (IGR) (Tan et al., 2011; David et al., 2014; Aïzoun et al., 2014). Pyrethroid insecticides in particular are widely used both to treat bed nets and in indoor residual spray programs in efforts to control the transmission of malaria (David et al., 2013; Chang et al., 2014; White et al., 2014, Strode et al., 2014). Insecticide resistance to pyrethroids in *Anopheles gambiae* is a major public health concern (Abdalla et al., 2014; Japheth et al., 2014). The mechanisms by which *Anopheles gambiae* acquires resistance to insecticides are primarily elevated levels of detoxifying enzymes (metabolic resistance) including cytochrome P450 monooxygenases (P450s) and target-site insensitivity (Ranson et al., 2011; Kabula et al., 2014, Mitchell et al., 2014). Cytochrome P450 monooxygenases (P450 or CYP) constitute the largest gene superfamily of structurally diverse and functionally versatile haem-containing enzymes with more than 15,000 known genes distributed across all living organisms including insects (Lamb and Waterman, 2013; Sezutsu et al., 2013). Previous microarray studies have shown that overexpression of P450s has been observed in several pyrethroid
resistant insects such as increased expression of CYP6D1 in pyrethroid resistant *Musca domestica* (Højland et al., 2014), CYP9J28 in pyrethroid resistant *Drosophila melanogaster* (Pavlidi et al., 2012; Poupardin et al., 2014), CYP6B7 in a pyrethroid resistant strain of *Helicoverpa armigera* (Rashid et al., 2013), CYP6E1 in *Culex pipiens quinquefasciatus* (Gong et al., 2005), CYP6P3 in pyrethroid resistant *Anopheles gambiae* (Edi et al., 2014; Kabula et al., 2014; Mitchell et al., 2014) and overexpression of CYP6M2 in pyrethroid resistant *Anopheles gambiae* (Edi et al., 2014; Mitchell et al., 2014). The increased threat to malaria control due to insecticide resistance associated with Cytochrome P450s necessitates the need to understand the mechanisms involved in this resistance (Edi et al., 2014) and this will enhance the provision of a more effective diagnostic monitoring of metabolic based resistance development (Bass et al., 2010; Hemmingway, 2014). Previous studies revealed that CYP6M2 is established to be directly involved in the acquirement of insecticide resistance in *Anopheles gambiae* (Hardstone et al., 2010; David et al., 2013; Abdalla et al., 2014; Nkya et al., 2014). Whilst CYP6M2 has been identified as having a role in insecticide resistance, the regulatory mechanisms involved in the over-expression of this gene is yet to be identified. Chapter Two of this study identified the *Anopheles gambiae* genes AGAP005300 (Nf2e1) / AGAP003645 and AGAP010259 / AGAP009748 as orthologs to Cap ‘n’ Collar isoform C (CnCC) / dKelch-like ECH-associated protein 1 (dKeap 1) and Spineless (Ss) / Tango (tgo) known to up regulate Cytochrome P450s in *Drosophila melanogaster* (Guio et al., 2014; Kuzin et al., 2014). These are also orthologs to Nuclear factor erythroid 2- related factor 2 (Nrf2) / Kelch-like-ECH associated protein 1 (Keap 1) and Aryl hydrocarbon nuclear translocator (ARNT) signalling pathways in mammals respectively (Siller et al., 2014; Das et al., 2014). The role of the orthologs of these genes in cytochrome P450 regulation in *Anopheles gambiae* is yet to be studied. Previous research in other insect
species addressed the mechanisms that underpin this regulation, mapping critical promoter elements that are required for P450 gene induction in response to the xenobiotic phenobarbital (PB) in Drosophila melanogaster (MacDonnell et al., 2004; Brown et al., 2005; Morra et al., 2010; Misra et al., 2011). Insect cell lines have been important working models for research in applied entomology since the first cell line from a moth was developed (Grace, 1962; Smagghe et al., 2009; Barletta et al, 2012). Cells from Anopheles gambiae (5.1* and Sua 5B), Aedes albopictus (C6 /36 cells), Drosophila melanogaster (S2 cells) have also been used to investigate different aspects of cellular response to stimuli and insect immunity (Pitaluga et al., 2008; Hughes et al., 2011; Moon et al., 2011). Cell lines from the malaria vector Anopheles gambiae display powerful humoral cellular response to xenobiotics (Smagghe et al., 2009). In this chapter, we employed dual Luciferase reporter assays in insect cell lines of An. gambiae to investigate the regulation of cytochrome P450 CYP6M2 gene expression, determine the activity of putative promoters from insecticide resistant and insecticide susceptible strains, and to investigate the induction of expression by exposure to insecticides in vitro. This chapter will therefore provide an insight into the regulatory mechanisms involved in permethrin insecticide resistance in Anopheles gambiae.

4.2 Experimental Approach

4.2.1 Construction of the expression vector

In order to facilitate construction of luciferase gene reporter assay vectors, the 5’ upstream region of CYP6M2 amplified from the Tiassalé and Kisumu strains of Anopheles gambiae were earlier cloned into the pJET1.2 cloning vector to produce, pJET-CYP6M2 Tiassalé (Tias) (930 bp) and pJET-CYP6M2_Kisumu (Kis) (896 bp). Primers were designed from the sequenced promoter sequences of both Tiassalé (631
bp) and Kisumu (591 bp), amplified, digested with NcoI and XhoI and ligated into pGL3-Enhencer in Chapter Three of this study. pGL3-CYP6M2_Tias & pGL3-CYP6M2_Kis constructs were grown in *E. coli* (200 mL), and purified using a Qiagen midi prep kit (Cat. # 12143, Qiagen, Hilden, Germany).

### 4.2.2 Anopheles gambiae Sua 5.1* cell culture

The *An. gambiae* Sua 5.1* cell line was established with minced neonate *Anopheles gambiae* (within an hour of hatching) using modified protocols (Pudney et al., 1979; Müller et al., 1999) and received from the Kafatos Lab, Imperial College London, UK. These cells were kindly provided by Dr Amy Lynd of the Vector Group, Liverpool School of Tropical Medicine (LSTM). The *Anopheles gambiae* Sua 5.1* cells were kept in continual culture in a 5% CO₂ humified incubator at 28°C in Schneider’s Drosophila medium (Cat. # 21720-024, Life Technologies, Paisley, UK) supplemented with 10% fetal bovine serum (FBS) and 100 U/mL Penicillin and 100 mg/mL Streptomycin (Cat. # P0781, Sigma-Aldrich, St. Louis, MO, USA). Approximately 5 × 10⁴ cells were seeded on 24-well plates and cultured for 24 h before transfection and allowed to reach 60-70% confluence.

### 4.2.3 Transient transfection

The *Anopheles gambiae* Sua 5.1* cell line were transfected with the pGL3-CYP6M2_Tias & pGL3-CYP6M2_Kis constructs using Qiagen Effectene Transfection Reagent (Cat. # 301425, Qiagen, Hilden, Germany) with modifications. Briefly, for each of the 24 wells, 600 ng DNA promoter constructs were added (600 ng in 60 µL DNA condensation buffer, containing 50 ng / µL of Actin Renilla plasmid (used to normalize for transfection efficiency), 6 µL Effectene (Cat. # 301425, Qiagen, Hilden, Germany).
and 4.8 µL Enhancer were sequentially added to 80 µL of serum free opti-MEM medium (Cat. # 31985-062, Life technologies, Paisely, UK) to bring the volume to 150 µL, the two solutions were combined and mixed gently. The DNA/Effectene complex was then applied to the appropriate wells with 250 µL complete opti-MEM medium (Cat. # 31985-062, Life technologies, Paisely, UK) in triplicate, followed by incubation at 28 °C for 5 h. The medium was then replaced with 250 µL Drosophila of fresh media and 20% FBS culture medium. Transfected cells were incubated at 28°C for 24 h prior to the insecticide challenge.

### 4.2.4 Cell challenges

In order to check for induction of the CYP6M2 gene promoter by insecticide, twenty-four (24) hours after transfection, the cells were stimulated with 2 µM and 20 µM of Schneider Drosophila medium supplemented with the insecticide permethrin (Cat. # 1667, FMC Corp., Philadelphia, PA, USA) and 2,3,7,8-Tetrachlorodibenzo-p-dioxin solution (TCDD) (Cat. # 328/1, Sigma-Aldrich, St. Louis, MO, USA) diluted in methanol (Cat. # 1209-22, ReAgent chem. Cheshire, UK) to a final concentration of 2 µM and 20 µM per well respectively. CYP6M2 gene promoter activity was examined quantitatively by dual luciferase reporter assay (Cat. # E1910, Promega, Madison, WI, USA). Controls had to be treated appropriately to ensure only the desired compound was inducing or repressing promoter activity. An initial basal activity of the promoter was performed.

### 4.2.5 Microscopy of the Anopheles gambiae Sua 5.1* cell line

Microscopic images of the seeded untransfected cells (A), cells challenged with 2 µM permethrin (B), (C) 20 µM permethrin and 2 µM TCDD were made.
4.2.6 Treatment of transfected Anopheles gambiae Sua 5.1* cells with xenobiotics

4.2.6.1 The effect of 2 µM and 20 µM concentrations of permethrin on the CYP6M2_Kis and CYP6M2_Tias promoter activity

The effect of permethrin at the two different concentrations was examined to characterise the existence of the antioxidant response element (ARE) within the CYP6M2 gene promoter sequence. These if activated by ligands such as permethrin have the ability to activate Nrf2 (Nf2e1) in the cytoplasm, which translocates into the nucleus and heterodimerase with muscle apoptosis fibromatosis (Maf). The Nrf2-Maf complex binds to the antioxidant ARE potentially resulting in the transcription of the cytochrome p450s including CYP6M2 (Guio et al., 2014).

4.2.6.2 The effect of 2 µM and 20 µM concentrations of TCDD on the CYP6M2_Kis and CYP6M2_Tias promoter activity

The effect of TCDD at the two different concentrations was examined to characterise the existence of the Dioxine response element (DRE) within the CYP6M2 gene promoter sequence. These if activated by ligands such as TCDD have the ability to activate AhR (AGAP010259) in the cytoplasm, which translocates into the nucleus and dimerases with ARNT. The AhR (AGAP010259) -ARNT complex potentially binds to the DRE resulting in the transcription of the cytochrome p450s including CYP6M2 (Kuzin et al., 2014).

4.2.7 Passive Cell lysis

Briefly, Twenty four hours (24 h) after the cell challenge, growth media was removed and cells were washed with 800 µL of sterile PBS (Cat. # D8537, Sigma Aldrich, USA), then harvested in 100 µL of 1 X Passive Lysis Buffer (PLB) (Cat. # E1910, Promega, Madison, WI, USA) (squirted on to aid dislodging of cells) and placed on a rocking
platform for 20 min at room temperature for cell lysis. The cell lysate (20 μL) were transferred into white, opaque, luminescence microplate reader, 96 well plates in duplicates of each of the 24 wells.

4.2.8 The Dual-luciferase reporter assays

The Dual-Luciferase Reporter (DLR) Assay System from Promega was used to examine promoter activity of the pGL3-CYP6M2_Kis and pGL3-CYP6M2_Tias constructs. Briefly, 24 hours after the cell challenge, 50 μL of Luciferase Reagent II (LARII) was injected into 20 μL lysate and luciferase activity measured for 10 sec on a luminometer (Turner Biosystems Luminometer). Then, 50 μL of Stop and Glo® reagent whose application quenches firefly luciferase activity thereby allowing bioluminescence produced by Renilla luciferase to be examined and measured (Cat. # E1910, Promega, Madison, WI, USA) was injected to the lysate to stop the luciferase activity and catalyse the Renilla reaction, incubated and then Renilla activity was measured for 10 sec. Luminescence of firefly luciferase was normalized by luminescence of Renilla luciferase. The normalized firefly luminescence represented the promoter activity driven by the CYP6M2 putative promoter. Data were collected from at least two experiments of 3 replicates and standard deviation was calculated. Two independent transfections (2 μM permethrin, 20 μM permethrin, 2 μM TCDD, 20 μM. Luciferase activity was normalised to the corresponding pRL-CMV activity for each sample. Results have been presented as the mean ± SD of transfections (pGL3-Enhencer, pGL3- CYP6M2_Kis & CYP6M2_Tias) and (pGL3-CYP9M10_CqSF & pGL3-CYP9M10_ISOP450) performed in triplicate using the aforementioned plasmid preparations and assayed in triplicate by dual luciferase reporter assay.
4.3 Results

4.3.1 Cloning of putative promoter elements upstream of the CYP6M2 gene for reporter assay measurement

In order to undertake functional analysis of the regulation of CYP6M2 expression, the 5' flanking regions of the two strains of Anopheles gambiae (previously cloned into pJET and called pJET-CYP6M2_Tias (631 bp) and pJET-CYP6M2_Kis (591 bp) were subcloned (in Chapter Three of this study) into the luciferase expression vector pGL3-Enhencer and used these in dual-luciferase reporter assays.

4.3.2 Examination of the basal CYP6M2 promoter Activity in the Anopheles gambiae Sua cell line

Figure 4.2 shows a higher fold change in basal relative luciferase activity in the Anopheles gambiae Sua cells transfected with pGL3-CYP6M2_Tias and pGL3-CYP9M10_1S0P50 compared to the respective pGL3-Enhencer vector controls (negative control). This verified the functionality of the four promoters; however the pGL3-Enhencer vector generated luciferase activity in the absence of a promoter regulatory element perhaps due to the presence of an enhancer within the vector. Furthermore, the pGL3-CYP9M10_1S0P450 construct has higher fold change than the pGL3-CYP9M10_CqSF compared to the Kisumu susceptible strain of Anopheles gambiae and the promoterless pGL3-Enhencer control.
However, pGL3-CYP6M2_Kis (531 bp) and pGL3-CYP9M10_CqSF (804 bp) have a lower fold change in basal luciferase activity and this further showed that the pGL3-Enhencer vector (Promega) generated luciferase activity in the absence of a promoter regulatory element due to the presence of an enhancer element within the vector. The CYP9M10_ISOP450 used as positive control has approximately 700 bp of sequence more than CYP9M10_CqSF and as such it also has a greater number of response elements. Under basal conditions, the lack of additional TFBS in the shorter promoter sequence results in less luciferase activity. Luciferase activity was normalised to the corresponding pRL-CMV activity for each sample. Results have been presented as the mean ± SD of transfections (pGL3-Enhencer, pGL3-CYP6M2_Kis & CYP6M2_Tias)
and (pGL3-CYP9M10_CqSF & pGL3-CYP9M10_ISOP450) performed in two separate experiments in triplicates using the aforementioned plasmid preparations and assayed in triplicate by dual luciferase reporter assay.

4.3.3 Treatment of transfected Anopheles gambiae Sua 5.1* cells with xenobiotics

4.3.3.1 The effect of 2 μM and 20 μM permethrin on the CYP6M2_Kis and CYP6M2_Tias promoter

Figure 4.3 shows that following treatment with permethrin, only pGL3-CYP6M2_Tias treated with 2μM permethrin has a higher fold change in luciferase activity compared to the Kisumu strain of Anopheles gambiae. Permethrin is therefore suggested to inhibit the induction of CYP6M2 in the Tiassalé strain.

![Figure 4.2: Permethrin challenged Anopheles gambiae Sua 5.1* cell lines. Cell lines were transfected with 50 ng of pRL-CMV vector and 0.6µg of pGL3-CYP6M2_Kis, pGL3-CYP6M2_Tias, pGL3-CYP9M10_CqSF and pGL3-CYP9M10_ISOP450 promoter construct or promoter-less pGL3-enhancer vector. Fold change in the permethrin treated plasmids relative luciferase activity was expressed as a percentage of pGL3-enhancer luciferase expression. Results have been presented as the mean ± SD of transfections (pGL3-enhancer, pGL3-CYP6M2_Kis, pGL3-CYP6M2_Tias, pGL3-CYP9M10_CqSF and pGL3-CYP9M10_ISOP450) performed in duplicate using the aforementioned plasmid preparations and assayed in triplicate by dual luciferase reporter assay (Promega).]
4.3.3.2 The effect of 2 µM and 20 µM TCDD on the CYP6M2_Kis and CYP6M2_Tias promoter

Figure 4.4 shows that following treatment with TCDD there was no induction of CYP6M2 in the Tiassalé resistant strain as compared to the Kisumu strain of Anopheles gambiae. pGL3-CYP9M10_CqSF used as positive control treated with 2 µM TCDD has a higher fold change in luciferase activity compared to all the other treatments.

Figure 4.3: TCDD challenged Anopheles gambiae Sua 5.1* cell lines. Cell lines were transfected with 50 ng of pRL-CMV vector and 0.5µg of pGL3-CYP6M2_Kis, pGL3-CYP6M2_Tias, pGL3-CYP9M10_CqSF and pGL3-CYP9M10_ISOP450 promoter construct or promoter-less pGL3-Enhencer vector. Fold change in the TCDD treated relative luciferase activity was expressed as a percentage of pGL3-Enhencer luciferase expression. Luciferase activity was normalized to the corresponding pRL-CMV activity for each sample. Results have been presented as the mean ± SD of transfections (pGL3-Enhencer, pGL3-CYP6M2_Kis, pGL3-CYP6M2_Tias, pGL3-CYP9M10_CqSF and pGL3-CYP9M10_ISOP450) performed in duplicate using the aforementioned plasmid preparations and assayed in triplicate by dual luciferase reporter assay (Promega).

4.3.4 Microscopy of the Anopheles gambiae Sua 5.1* cell line

Figure 4.4 shows microscopic images of healthy viable cells following A) seeding seeding of the cells but prior to infection B) transfection and treatment with 2 µM
permethrin. C transfection and treatment with 20 μM permethrin whilst D) treatment with 2 μM TCDD.

Figure 4.4: Microscopic images of A. Untransfected healthy *Anopheles gambiae* Sua 5.1* cells B. *Anopheles gambiae* Sua cells transfected and challenged with 2 μM Permethrin C. Twenty (20) μM permethrin D. Two (2) μM TCDD. The cells were observed at X 40 magnification.

4.4 Discussion

The emergence of insecticide resistance due to widespread use of insecticides has become a major public health problem, which presents a great challenge to the control of vector borne diseases (Karunamoorthi and Sabesan, 2013; Muthusamy et al., 2014). In this research *CYP6M2*, a member of the Cytochrome P450 super family was studied, because of its potential role in the metabolism of insecticides during xenobiotic exposure. The interest in this gene lies in the understanding of the regulatory mechanisms involved in the control of *CYP6M2* expression, particularly as it’s over expression can result in insecticide resistance (Mitchell et al., 2012; Edi et al., 2013).
2014). In the current study, no promoter activity was detected using the dual luciferase assay, with some indication of a reduction in luciferase activity in the presence of the CYP6M2 promoter possibly indicative of the inhibition by both the permethrin and TCDD (Figures 4.2 and 4.3). This is counter to expectations given the up regulation of CYP6M2 frequently seen in studies of insecticide resistant An. gambiae including those of the Tiassalé strain (Edi et al., 2014). However, it is consistent with previous studies in Drosophila melanogaster which showed that six chemically distinct insecticides did not induce the expression of P450s with the exception of DDT, and even DDT exposure only led to a weak induction of a single P450, Cyp12d1. This therefore indicates that exposure to insecticides results in very little induction of transcription of CYP genes (Willoughby et al., 2006; Rand et al., 2012; Harrop et al., 2014). Zhou et al. (2010) also suggest that due to their toxicity, insecticides generally are unlikely to cause induction of P450s. Another study using microarrays showed that permethrin exposure of larvae of the mosquito (Aedes aegypti) resulted in a significant over expression of three CYP genes (CYP9M8, CYP9M9 and CYP314A1) (Poupardin et al., 2008), 16 different CYP genes were induced in response to caffeine in Drosophila melanogaster (Willoughby et al., 2006).

4.5 Conclusions

This work provides a first step in identifying the regulatory mechanisms involved in the expression of CYP6M2. Further more, the work indicates that both permethrin & TCDD at 2 µM and 20 µM concentrations do not induce the putative CYP6M2 promoter cloned in this study. Additional experimental work will be required to investigate the expression of CYP6M2 in the insecticide susceptible Kisumu strain and resistant Tiassalé strain in Anopheles gambiae. Therefore in Chapter Five of this study,
quantitative polymerase chain reaction (qPCR) was employed to measure the differential expression of $CYP6M2$, $AhR$ (AGAP010259) and ($Nf2e1$) $Nrf2$ in the Tiassalé, Kisumu and Auyo (Nigeria) strains of $Anopheles gambiae$ phenotypes for insecticide resistance.
CHAPTER FIVE

Differential Expression of AGAP010259 (AhR) and Nf2e1 (Nrf2) Candidate Genes in Some Selected Strains of Anopheles gambiae
5.1 Introduction
Mosquitoes can be efficient transmitters of human, livestock and poultry pathogens including malaria (David et al., 2013; Huo and Qiu, 2014; Chang et al., 2014), rift valley fever (Rodolakis, 2014), myxomatosis (McColl et al., 2014) and fowl pox (Lawson et al., 2012). The major mosquito control methods, indoor residual spraying (IRS) and long lasting insecticide-treated bed nets (LLINs), rely heavily on the use of insecticides (Okumu and Moore, 2011; West et al., 2014; Ngufor et al., 2014). However, resistance to the limited number of available insecticides is developing rapidly, with some populations now showing resistance to all four classes (pyrethroids, organochlorines, organophosphates and carbamates) (Ranson et al., 2009; Edi et al., 2012; Witzig et al., 2013; Strode et al., 2014). Permethrin, discovered in 1973 is the first potent, photo stable pyrethroid (Soderlund and Knipple, 2003) and has been widely used due to its consistent effectiveness in mosquito control, high toxicity to mosquitoes even at low doses and low toxicity to higher mammals (Pridgeon et al., 2009; Wan-Norafikah et al., 2013b; Matowo et al., 2014a). Therefore, this resistance to pyrethroid insecticides is of enormous concern as this is the major class of insecticides used in public health and the only class of insecticides permitted for impregnation of mosquito nets (WHO, 2012; Witzig et al., 2013; Nkya et al., 2014). If pyrethroid insecticides were to fail due to resistance of target populations, WHO estimates a loss of 50% of the current impact on vector control in Africa (WHO, 2012). Resistance to pyrethroids is mainly due to two mechanisms; increase in enzyme metabolism or reduction in sensitivity to target site. Previous microarray studies on insecticide-resistant mosquitoes, including Anopheles gambiae has identified a relatively small number of up-regulated CYP genes after exposures of the mosquitoes to different concentrations of insecticides (Liang et al., 2015). Upregulation of these
Cytochrome oxidases (P450s) including CYP6M2 has been associated with resistance and enzymatic metabolism of insecticides by CYP6M2 has been shown in vitro (Stevenson et al., 2011; Mitchell et al., 2012; Edi et al., 2014). In *Drosophila melanogaster*, some genes involved in metabolic activity are known to be upregulated by the transcription factors Spineless (Ss) / Tango (Tgo) and / or Cap ‘n’ collar isoform (CnCC) / *Drosophila* Kelch-like-ECH-associated protein 1 (dKeap 1). These are orthologs to Nuclear factor erythroid -2 related factor-2 (Nrf2) / Kelch-like–ECH-associated protein 1 (Keap 1) and Aryl hydrocarbon receptor (AhR)/ Aryl hydrocarbon receptor nuclear translocator (ARNT) signalling pathways in higher mammals respectively (Cao et al., 2013; Misra et al., 2013; Das et al., 2014; Siller et al., 2014). Recent studies have demonstrated that the evolutionarily conserved AhR / ARNT and Nrf2 / Keap 1 pathways play a central role in regulating the coordinate transcriptional response to xenobiotic compounds in *D. melanogaster* (Misra et al., 2011; Jones et al., 2013; Guio et al., 2014; Kuzin et al., 2014). Under normal physiological conditions, AhR is normally dormant within the cytoplasm forming a complex with Heat shock protein 90 (HSP90), xap2, and p23 co-chaperone. However, when it binds exogenous or endogenous toxins such as dioxin, it is activated and trans located into the nucleus where it dimerizes with another basic helix–loop–helix (bHLH)-PAS protein and its hetero-dimer partner, Aryl hydrocarbon receptor nuclear translocator (ARNT). The AhR: ARNT heterodimer binds to a specific motif (Xenobiotic Response Element; XRE) in the promoters of its target genes and controls their transcription (Misra et al., 2013; Hansen et al., 2014; Kuzin et al., 2014). TCDD (2, 3, 7, 8-tetrachlorodibenzo-p-dioxin) or benzo[a]pyrene act as a ligand for AhR, and functional studies of AhR revealed although it is not directly genotoxic, but binds to the aryl hydrocarbon receptor nuclear translocator (ARNT), an intracellular protein that is a transcriptional enhancer.
affecting the expression of important genes (Faust et al., 2013; Le Vee et al., 2010; Moskalev et al., 2014). Similarly, for the \( Nrf2 / Keap1 \) pathway, in the absence of stress, \( Nrf2 \) is retained in the cytoplasm by the actin-binding protein \( Keap1 \), which also functions as an E3 ubiquitin ligase to promote \( Nrf2 \) degradation by the 26S proteasome. Activation of this pathway through oxidative stress disrupts the \( Nrf2–Keap1 \) interaction, allowing \( Nrf2 \) to translocate to the nucleus, where it can heterodimerize with the small Maf (muscle aponeurosis fibromatosis) proteins and bind to antioxidant response elements (AREs) in the genome (Atia and Bin Abdullah, 2014). \( Nrf2, Maf, \) and \( Keap1 \) are all conserved in \( D. \ melanogaster \) and appear to wield the same regulatory interactions as described in vertebrates (Si and Liu, 2014; Dhanoa et al., 2013). Activation of this pathway through electrophilic xenobiotics / oxidative stress is necessary and sufficient for xenobiotic-induced transcription of a wide range of detoxification genes in \( Drosophila \) species (Misra et al., 2011; Deng and Kerpolla, 2013). In contrast to the detailed studies in higher mammalians and \( D. \ melanogaster \), the regulatory mechanism in \( A. \ gambiae \) is yet to be identified. Here we use a combination of bioinformatics and molecular biology to show that AGAP010259 (AhR)/ \( Tango \) and Nuclear factor erythroid 2, invertebrate) (Nf2e1) \( (Nrf2) / dKeap1 \) pathways play central roles in the regulation of xenobiotic responses in \( A. \ gambiae \). Consistent with the previous studies on the \( Drosophila \) model pathway, these studies have established that the \( Nrf2 / Keap1(Nf2e1 / dKeap1) \) pathway is differentially active as a key regulator of xenobiotic responses in both the insecticide resistant Tiassalé strain and the previously uncharacterised insecticide selected strains (Auyo-Nigeria) strains of \( A. \ gambiae \). We therefore conclude that inhibition of this \( Nf2e1 / dKeap1 \) in particular and AGAP010259 / \( Tango \) response may potentially improve the efficacy of insecticides. These studies have implications for understanding the regulatory
mechanisms of acquirement of insecticide resistance and its impact in the control of mosquito-borne diseases.

5.2 Experimental Approach

5.2.1 Mosquito strains

Two established strains of *An. gambiae* were used (obtained from the Liverpool Insect Testing Establishment (LITE) unit of the Liverpool School of Tropical Medicine (LSTM). The multiple insecticide resistant Tiassalé strain was derived from material collected in northwest of Abidjan, Ivory Coast and the laboratory susceptible strain (Kisumu) was derived from material collected in western Kenya. Additionally, 3 other uncharacterised strains from Auyo, Sudan savannah region of northern Nigeria selected on 3 different insecticides according to the recommendation by the WHO (Table 5.1).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Selected on:</th>
<th>Resistant to:</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kisumu</td>
<td>Permethrin (0.75%)</td>
<td>-</td>
<td>Kenya</td>
</tr>
<tr>
<td>Tiassalé</td>
<td>Permethrin (0.75%)</td>
<td>all insecticides</td>
<td>Ivory Coast</td>
</tr>
<tr>
<td>Auyo</td>
<td>Unselected (Base)</td>
<td>unknown</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Auyo</td>
<td>Permethrin (0.75%)</td>
<td>unknown</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Auyo</td>
<td>Bendiocarb (0.1%)</td>
<td>unknown</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Auyo</td>
<td>DDT (4%)</td>
<td>unknown</td>
<td>Nigeria</td>
</tr>
</tbody>
</table>

The selected and unselected Auyo strains of *An. gambiae* from Nigeria were kindly provided by Habibu Abdu Usman of the School of Science, Engineering and Technology (SET), Abertay University, Dundee.
The Kisumu and Tiassalé strains of An. gambiae were maintained under standard insectary conditions of 25°C with a relative humidity of 80 % and a 12 hour day/night light and 45-min dusk/dawn cycles (Hunt et al., 2005). The selection of insecticide resistance in all the strains were performed by separating newly emerged male and female adults in to cages supplied with 10% sucrose solution. This ensured that mating did not occur prior to exposure so as to select resistant strains prior to bioassay (below).

5.2.2 WHO adult bioassay

A diagnostic test using standard WHO Test Kits tube was conducted by means of tarsal exposure to papers impregnated with discriminating concentration of permethrin 0.75% (WHO, 2013d). Briefly, two to five day old female mosquitoes (n=10) each of the Tiassalé and Kisumu strains and uncharacterised strains from Auyo were transferred into a standard holding tube (containing untreated filter paper). The mosquitoes were then transferred into exposure tubes (containing 0.75% permethrin impregnated paper) and exposed for 1 hour after which the knockdown rate was recorded. The mosquitoes were then transferred back into holding tubes, supplied with 10% sucrose solution (soaked in a cotton wool pad) and left for 24 hours. A final mortality was recorded after 24 hours. All tests were undertaken at 25 ± 2 ºC and 55–78% relative humidity, and a natural photoperiod was maintained (WHO, 2013d; Vogelweith et al., 2014). Cotton pads soaked in 10% sugar solution were provided during the 24 hours holding period.
5.2.3  RNA isolation and cDNA synthesis for qPCR

5.2.3.1  Total RNA isolation

In order to prepare cDNA for the amplification of AGAP010259 (AhR) and Nf2e1 (Nrf2) using gene specific primers designed for this study (Table 5.2), total RNA was isolated from the 10 pool of Tiassalé, Kisumu, Permethrin, Bendiocarb and DDT exposed and non-exposed (from Auyo) strains of *An. gambiae* mosquitoes using the RNAqueous®-4PCR Kit according to the manufacturers’ recommendation (Cat. # AM1914, Ambion-life technologies, Paisley, PA, UK). The quality and quantity of the RNA was then determined by running on 1.5 % agarose gel electrophoresis and NanoDrop™ 1000 spectrophotometer (Nanodrop Technologies, Oxfordshire, UK).

5.2.3.2  cDNA synthesis

Superscript III (Cat. # 18080-051, Invitrogen, Life technologies, Invitrogen, Paisley, PA, UK) and oligo dT<sub>20</sub> was used to synthesise cDNA from RNA. Samples were incubated for 5 min at 65°C followed by 50 min at 65°C, 5 min at 85°C and 20 min at 37°C in a thermocycler with a 105°C heated lid. The cDNA was cleaned using Qiagen purification kit (Cat. # 28106, Qiagen, Valencia, CA, USA) and quantified using NanoDrop™ 1000 spectrophotometer (Nanodrop Technologies, Oxfordshire, UK). All cDNAs were diluted 10-fold with nuclease free water and 5 µL were used in each PCR reaction.

5.2.3.3  Primer design

Designing optimal primer pairs is critical for successful quantitative real-time PCR (qPCR). The primer pairs should be able to yield a highly specific PCR product and have minimal nonspecific annealing. Three pairs of exon-exon boundary crossing
primers for \( Nf2e1 \) (Nrf2) and AGAP010259 (AhR) (Table 4.2) were designed using Primer 3 software (Broad Institute, USA) (http://www.ncbi.nlm.nih.gov/tools/primer-sblast/).

**Table 5.2: Primers used in quantitative real-time PCR (qPCR)**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession number</th>
<th>Sequence (5′-3′)</th>
<th>Annealing temperature (°C)</th>
<th>Transcript length (bp)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRFFP001</td>
<td>AGAP005300</td>
<td>CCGCATTTTCGTTCTAGCC</td>
<td>64.7</td>
<td>210</td>
<td>This study</td>
</tr>
<tr>
<td>NRFRP002</td>
<td></td>
<td>ATCCATTTCCGTGTTGGGG</td>
<td>68.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRFFP003</td>
<td></td>
<td>GGGCGACTTTTTACCCGCG</td>
<td>67.5</td>
<td>120</td>
<td>This study</td>
</tr>
<tr>
<td>NRFRP004</td>
<td></td>
<td>ACCAACGCAACCACAAAGC</td>
<td>67.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRFFP005</td>
<td></td>
<td>ACCAACGCAACCACAAAGC</td>
<td>63.9</td>
<td>159</td>
<td>This study</td>
</tr>
<tr>
<td>NRFRP006</td>
<td></td>
<td>GATCCATTTCCGTGTTGGGG</td>
<td>69.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHRFP007</td>
<td>AGAP010259</td>
<td>CAGCACACTTTCCCGGTACT</td>
<td>64.0</td>
<td>122</td>
<td>This study</td>
</tr>
<tr>
<td>AHRFP008</td>
<td></td>
<td>ATTCCATAACCTTGGGCAGCCG</td>
<td>66.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHRFP009</td>
<td></td>
<td>CCGAACAGGTTCGCCTAGT</td>
<td>61.7</td>
<td>251</td>
<td>This study</td>
</tr>
<tr>
<td>AHRFP010</td>
<td></td>
<td>ATTCCATAACCTTGGGCAGCCG</td>
<td>59.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHRFP011</td>
<td></td>
<td>TGCAGCTGCTATTCAACA</td>
<td>66.8</td>
<td>166</td>
<td>This study</td>
</tr>
<tr>
<td>AHRFP012</td>
<td></td>
<td>ATTCCATAACCTTGGGCAGCCG</td>
<td>66.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP6M2FP 013</td>
<td>AGAP008212</td>
<td>CATGACACAAAACGACAAAGG</td>
<td>60.0</td>
<td>235</td>
<td>Nardini et al., 2012</td>
</tr>
<tr>
<td>CYP6M2RP 014</td>
<td></td>
<td>GGTGAGGAGAGCTGACAAAG</td>
<td>60.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Key:** F= forward and R= reverse

Gene expression must be normalized against a housekeeping gene (HKG) which is constitutively expressed in all cell types and tissues being used (Thellin et al., 1999). The reference gene should not be regulated or influenced by the experimental conditions (Yilmaz et al., 2012). The most commonly used reference genes considered to normalise data include; the Elongation factor 1 protein (AGAP003541), Ribosomal protein s7 (AGAP010592) and Ubiquitin (AGAP007927) primers (Table 5.3).
Table 5.3: Normalizing genes (Housekeeping gene)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Symbol</th>
<th>GenBank</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubiquitin</td>
<td>[UBQ]</td>
<td>AGAP007927-RA</td>
<td>Protein degradation</td>
<td>(1) [2]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Transcription regulation</td>
<td></td>
</tr>
<tr>
<td>Elongation factor 1</td>
<td>[Ef1]</td>
<td>AGAP003541-RA</td>
<td>Protein synthesis</td>
<td>(1) [3] [4]</td>
</tr>
<tr>
<td>Protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


5.2.3.4 Quantitative PCR (qPCR) using SYBR® Green Master mix

qPCR reactions were carried out in 96-well polypropylene plates Stratagene (Cat # 40133, Carlsbad, CA, USA) in a volume of 20 µL and in triplicate (Wilding et al., 2012). Briefly, 5 µL 1/10 cDNA was added to 15 µL reaction mix containing 10 µL SYBR ® SELECT Master Mix (Cat # 4472908, Life technologies, Carlsbad, CA, USA), 200 µM of each specific primer pair specific for each gene. SYBR ® Green is an intercalating dye which binds to double stranded DNA and results in fluorescence as the amount of PCR product increases (Thornton and Basu, 2011). Confirmation that the product detected by SYBR ® Green is a single amplicon can be obtained by generating a melt curve at the end of the PCR amplification cycles. PCRs were run on a MX3000P system (Stratagene, Santa Clara, CA, USA) with the following cycling conditions: initial denaturation at 95 °C for 10 min followed by 40 cycles of denaturation at 95 °C for 30 sec, annealing of 55 °C for 1 min, and a primer extension of 72°C for 1 min with melting curves run after each end point amplification. Following these 40 cycles reactions were heated to 95 °C for 1 min, followed by 30 s increments of 1 °C from 55 °C to 95 °C for the melting curve analysis. All reactions were run in duplicate and the
mean Ct-values were used for further analysis. Three independent experiments were performed. The mean of the threshold cycles (Ct) for each gene were normalised against the average values for Ubiquitin. To compare gene expression between treatments, we calculated the $\Delta\Delta$Ct. The fold-change in gene expression for each target gene, normalized to the ubiquitin (UBQ) relative to a laboratory susceptible strain of Kisumu (Base) and Auyo (Base) from Nigeria was calculated according to the $2^{\Delta\Delta C_t}$ method incorporating PCR efficiency (Schmittgen and Livak, 2008). The $2^{\Delta\Delta C_t}$ method may be used to calculate relative changes in gene expression determined from real-time quantitative PCR experiments. Basic data analysis regression and t-tests were performed in Excel with $p < 0.05$ used to assess significant difference between treatments for the t-tests. The relative expressions of CYP6M2 and AhR (AGAP010259) & Nrf2 (Nf2e1) were expressed as fold changes of the three biological replicates. The one-way analysis of variance (ANOVA) was used to further statistically analyse data to establish any significance or correlation between data sets. This was followed by Tukey’s HSD test for multiple comparisons. These statistical tests were carried out using online IBM SPSS Software version 22.

5.3 Results

5.3.1 Selection of candidate reference genes in permethrin selected An. gambiae strains using SYBR® Green (qPCR)

Table 5.4 shows that using SYBR® Green (qPCR) of the ubiquitin (UBQ) gene has a constant level of expression calculated by the comparative CT method.
Table 5.4: A brief summary on the selection of candidate reference genes in permethrin selected *An. gambiae* strains using SYBR® Green (qPCR) (Ct values).

<table>
<thead>
<tr>
<th>Candidate Reference gene</th>
<th>Tiassalé (Mean ± SE)</th>
<th>Kisumu (Mean ± SE)</th>
<th>Auyo-Perm (Mean ± SE)</th>
<th>Auyo-Base (Mean ± SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubiquitin (UBQ)</td>
<td>26.64±0.06</td>
<td>26.04 ±0.26</td>
<td>26.50 ±0.02</td>
<td>26.39±0.02</td>
</tr>
<tr>
<td>Elongation factor 1 (EF1)</td>
<td>16.41±0.54</td>
<td>16.34±0.37</td>
<td>16.17±0.29</td>
<td>16.75±0.66</td>
</tr>
<tr>
<td>Ribosomal protein (Rsp7)</td>
<td>29.68±0.05</td>
<td>28.23 ±1.37</td>
<td>28.07±0.18</td>
<td>28.27±0.16</td>
</tr>
</tbody>
</table>

The expression levels calculated were Tiassalé (26.64±0.06), Kisumu (26.04 ±0.26), Auyo-Perm (26.39±0.02) and Auyo-Base (26.39±0.02). This is in comparison to the Elongation factor protein and Ribosomal protein for the same strains of *An. gambiae*.

Ubiquitin (UBQ) gene was therefore selected for the subsequent qPCR (Section 5.2).

### 5.3.2 Selection of candidate reference genes in Auyo-Selected and Auyo-Base *An. gambiae* using semi-quantitative qPCR

Table 5.5 shows that using SYBR® Green (qPCR), ubiquitin (UBQ) gene has a constant level of expression calculated by the comparative CT method.

<table>
<thead>
<tr>
<th>Candidate reference gene</th>
<th>Auyo-DDT (D) (Mean±SE)</th>
<th>Auyo-DDT (S) (Mean±SE)</th>
<th>Auyo-Bendi (D) (Mean±SE)</th>
<th>Auyo-Bendi (S) (Mean±SE)</th>
<th>Auyo-Base (Mean±SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubiquitin [UBQ]</td>
<td>25.92±1.16</td>
<td>25.06 ±0.57</td>
<td>24.93±1.62</td>
<td>24.61±0.09</td>
<td>24.86±1.40</td>
</tr>
<tr>
<td>Elongation factor 1 [EF1]</td>
<td>16.24±0.78</td>
<td>17.56±1.16</td>
<td>16.14±0.33</td>
<td>15.91±0.23</td>
<td>24.86±1.40</td>
</tr>
<tr>
<td>Ribosomal protein [Rsp7]</td>
<td>23.19±0.60</td>
<td>24.19 ±0.38</td>
<td>24.39±0.59</td>
<td>24.19± 0.38</td>
<td>15.79±0.25</td>
</tr>
</tbody>
</table>

**Key:** D= Dead, S= Survived, Auyo-DDT (Auyo strains exposed with DDT), Auyo-Bendi (Auyo strains exposed with Bendiocarb), Auyo-Base (Auyo strain non-exposed).
5.4 Quantitative real-time PCR (qPCR)

5.4.1 Differential expression levels of AGAP010259 (AhR), Nf2e1 (Nrf2) and CYP6M2

Quantitative real-time PCR (qPCR) was used to detect the relative expression levels of (AGAP010259) AhR, (Nf2e1) Nrf2 and CYP6M2 in insecticide resistant Tiassalé, susceptible Kisumu and Auyo strains of *An. gambiae*. Ubiquitin (UBQ) gene was used as a normalization control (Figure 5.1).

![Graph showing relative expression levels of AhR, Nrf2, and CYP6M2](graph.png)

**Figure 5.1:** Relative expression levels (linear fold changes) of the target genes AhR, Nrf2 and CYP6M2. There was a significant statistical difference between *An. gambiae* strains expressions as determined by one-way ANOVA (P≤0.01). Data were expressed as mean SEM, n= 3 per group. Error bars are 95% confidence interval. Stars indicate the significance level: one star for p<0.05; two stars for p<0.01; three stars for p<0.001.

**Key:** T-Perm (Tiassalé strain selected with permethrin), K-Perm (Kisumu strain selected with permethrin), Auyo-Perm (Auyo strain exposed with permethrin), A-Base (Auyo strain non-exposed), A-Be (D) (Auyo strain exposed to Bendiocarb)(Dead), A-Be (S) (Auyo strain exposed to Bendiocarb)(Survived), A-D (D) (Auyo strain exposed to DDT)(Dead) and A-D (S) (Auyo strain exposed to DDT)(Survived).
Here, analysis was made on the relative expression of \textit{AhR}, \textit{Nrf2} and \textit{CYP6M2} in the Auyo, Kisumu and Tiassalé strains of \textit{An. gambiae} after exposure to Permethrin, Bendiocarb and DDT insecticides. The expression levels of \textit{AhR}, \textit{Nrf2} and \textit{CYP6M2} were compared between i) Wild population Auyo strains exposed to Permethrin, Bendiocarb and DDT ii) non-exposed susceptible Kisumu and Auyo strains iii) Laboratory recently colonised resistant Tiassalé and susceptible Kisumu strains exposed to Permethrin. There were notable differences in relative expression levels between the wild population Auyo strains and the laboratory susceptible Kisumu and resistant Tiassalé strains of \textit{An. gambiae} as described below: \textit{CYP6M2} expression levels were significantly different in all the \textit{An. gambiae} strains except between Auyo strains exposed to permethrin (Fold change (FC) =3.5, \(p<0.05\)) and those that survived DDT exposure (FC=7, \(p< 0.05\)) compared to the strains non-exposed to permethrin K-Base (FC=1.8, \(p< 0.05\)). \textit{Nrf2} gene was overexpressed in Auyo strains that survived Bendiocarb A-Bendi(S) (FC=14, \(p< 0.001\)) and DDT A-D(S) (FC=11, \(p<0.001\)) wild population permethrin exposed strains compared to those that died due to the exposure of permethrin A-Bendi(D) (FC=2.8, \(p<0.001\)) and A-DDT(D) (FC=2.8, \(p<0.001\)). Colonised resistant Tiassalé-Permethrin (FC=3, \(p<0.001\)) colonised resistant strain challenged with permethrin however did not show high expression of \textit{Nrf2} (\textit{Nfr2e1}). Additionally, there is no correlation between the exposed and non-exposed strains of \textit{An. gambiae} in the expression of \textit{AhR} gene. No significant differences in the \textit{CYP6M2} expression between the permethrin, Bendiocarb and DDT exposed and non-exposed strains observed in any populations (\(p>0.05\)).

The transcription level of \textit{Nrf2} is higher in the Auyo strains that survived the exposure of both Bendiocarb and DDT. These transcription levels are strongly correlated with high expression of cytochrome p450s including \textit{CYP6M2} and the rate of insecticide
metabolism therefore leading to insecticide resistance in An. gambiae. Over expression of CYP6M2 was however not observed in the Auyo strains that survived the exposure to Bendiocarb in comparison with the Auyo strains that survived the exposure to DDT and K-Base control.

5.5 Discussion

Mosquito borne disease control programs globally rely heavily on vector control. The recent emergence of insecticide resistance to all the WHO recommended insecticides has had a major impact on the spread of vector-borne human, livestock and poultry diseases, with a disproportionate effect on developing countries. Consequently, an effort has been aimed at understanding the regulatory mechanisms by which mosquitoes acquire insecticide resistance. Our studies here indicate that there were low levels of (AGAP010259) AhR in the Tiassalé insecticide resistant selected strains but higher in the susceptible Kisumu strain exposed to permethrin. This is consistent with the earlier genetic studies of spineless (Ss) orthologs in D. melanogaster, which indicates that it plays an essential developmental role in leg and antennal specification, and suggests that it has no function in xenobiotic pathways (Duncan et al., 1998; McMillan and Bradfield 2007). However, interaction of this spineless and CG5017 gene is established to be important in detoxification during oxidative stress in Drosophila (Kuzin et al., 2014). Moreover, an interesting but yet to be understood phenomenon occurs in higher mammals, where AhR and Nrf2 exhibit multilevel crosstalk with the former gene also being a target of the latter (Bock, 2014). In the presence of xenobiotics, metabolic resistance can be related to constitutive or induced detoxification process or both (Boyer et al., 2006; Matowo et al., 2014). Here analysis was made on the induction effect of AhR, Nrf2 and CYP6M2 in the wild population
Auyo, laboratory colonised Kisumu and Tiassalé strains of *An. gambiae* after exposure to Permethrin, Bendiocarb and DDT insecticides. Results showed that mosquito exposure to DDT induce over-expression of *Nrf2* and *CYP6M2*, suggesting that the transcription of *CYP6M2* expression is driven by the expression of *Nrf2*. Additionally, we show here propose that differential expression of the *Nf2e1 / dKeap 1* pathway is potentially central to the overexpression of detoxifying genes in insecticide wild population exposed Auyo strains that survive the exposure of Bendiocarb (Auyo-Bendi(S) and DDT (Auyo-DDT(S)) *An. gambiae* exposed from Nigeria. This is consistent with the previous finding of Misra *et al.*, (2011 & 2013) that CnCC / dKeap 1 pathway plays a key role in the coordinated induction of detoxification gene expression in response to xenobiotic treatment and ectopic activation of this pathway is sufficient to confer resistance to Malathion. Similarly, the AhR–XRE pathway was shown to be conserved in insects such as the black swallowtail caterpillar (*Papilio polyxenes*) which responds to xanthotoxin by inducing *CYP6B1* via XRE-like binding sites (McDonnell *et al.*, 2004; Brown *et al.*, 2005) and Putative XRE binding sites were also found upstream of *An. gambiae* CYP6 genes induced by permethrin (Poupardin *et al.*, 2008). Wild population and laboratory-selected strains of insecticide-resistant *An. gambiae* overexpress a number of detoxifying genes including *CYP6M2*, demonstrating a correlation between their resistance and detoxification gene expression (Ffrench-Constant *et al.*, 2004; Li *et al.*, 2007). *CYP6M2* was included in this chapter to demonstrate its expression in the selected strains of *An. gambiae* which was not established in Chapter Four of this study as it was previously identified by Muller *et al.*, (2008) and Nardini *et al.*, (2012). Moreover, previous studies have demonstrated that *CYP6M2* is capable of metabolizing the organochlorine insecticide DDT in *An. gambiae*, hence demonstrating the first evidence for a metabolic cross resistance in malaria.
Differential Expression of AGAP010259

Chapter Five

B.R. Mohammed

-108-  Abertay University, Dundee, UK

vectors (Mitchell et al., 2014; Matowo et al., 2014). Following qPCR analysis therefore, CYP6M2 was over-expressed in the wild population Auyo strains that survived the exposure of DDT (A-D(S)) compared to the Auyo (FC=7, p< 0.05) strains that died of the DDT (FC=1.5, p<0.001) exposure. The high expression of CYP6M2 (FC=7, p<0.001) in DDT (An. gambiae that survived the exposure to the insecticide) is consistent with another study on Bendiocarb (Carbamate) resistance in An. gambiae (Aïkpon et al., 2013) and Tiassalé An. gambiae also show exceptionally high-level of carbamate resistance (Edi et al., 2014). The high variability in CYP6M2 expression among biological replicates, especially evident in qPCR, suggests that the regulatory mechanisms generating overexpression is far from fixation in the selected strains. The low levels of expression of the, CYP6M2 in the pyrethroid resistant Tiassalé strain implies that the observed pyrethroid resistance results from other mechanisms rather than insecticide metabolism. Results in this study therefore demonstrate that there was induction of CYP6M2 fold change (FC=7, p< 0.05) in the Auyo strains that survived the exposure to DDT compared to all the other strains. This is consistent with studies which demonstrated that CYP6M2 is capable of metabolizing the organochlorine insecticide DDT in An. gambiae, hence demonstrating the first evidence for a metabolic cross resistance in malaria vectors (Mitchell et al., 2014; Matowo et al., 2014).

Previous studies in D. melanogaster, suggest that a single resistant CYP6G1 allele has spread globally (Daborn et al., 2002). If this rapid spread of a resistant CYP6M2 allele is mirrored in An. gambiae, the success of ITN-based malaria control programmes could be jeopardised (Nikou et al., 2003). Although, these results suggest that the Nf2eI / dKeap1 pathway may contribute for the overexpression of CYP6M2 gene as seen in Auyo strains that survived the exposure to DDT and Auyo strain that survived the
exposure to Bendiocarb, activation of this pathway is unlikely to be the only factor that contributes to insecticide resistance in these strains. Resistance modelling has suggested that selection with sub-lethal concentrations of insecticides favours the development of polygenic resistance in *D. melanogaster* (Schlipalius et al., 2008).

### 5.6 Conclusions

The study compared the AhR, *Nrf2* (*Nf2e1*) and *CYP6M2* genes expression levels of permethrin-resistant Tiassalé, Kisumu susceptible and uncharacterized Auyo strains of *An. gambiae* using quantitative real-time PCR (qPCR). Based on the qPCR results, *CYP6M2* and *Nrf2* (*Nf2e1*) were over transcribed in the Auyo strain that survived the exposure to DDT. However, further investigation is needed to clarify the role of *Nrf2* (*Nf2e1*) in resistance as there was no corresponding transcription of *CYP6M2* in the Auyo stains that survived exposure to Bendiocarb. There is no real evidence that *Nrf2* (*Nf2e1*) is linked to the observed resistance in the Auyo strains that survived the DDT exposure, although significant lower expression of this gene was observed in both the susceptible Kisumu and Auyo Base (non-exposed) strains of the *An. gambiae*. These results therefore, suggest that the inhibition of the *Nf2e1* / dKeap1 pathway should potentially sensitize insects to insecticides application. Accordingly, *Nf2e1* inhibitors might also act as effective synergists that could improve our use of insecticides to control mosquito-borne diseases such as malaria.
Discussion, Conclusions and Suggestions for Further Work
6.0 Discussion

6.1 Introduction

The effective usage of indoor residual spray (IRS) and insecticide treated nets (ITNs) has demonstrated beneficial effects in the control of malaria (Anderson et al., 2014; Boene et al., 2014; Chang et al., 2014; Lwetoijera et al., 2014; Ngufor et al., 2014; Riveron et al., 2014; Rulisa, 2014; West et al., 2014; Toé et al., 2014). However, the development of resistance to the common insecticides recommended by the World Health Organisation (WHO) has now become a challenge despite the enormous effort put into this topic over the last few decades. The original hypothesis set out at the beginning of this thesis was that Cytochrome P450s, which are known to have key roles in the insecticide resistance phenotype including CYP6M2, are up-regulated by the orthologs of Cap ‘n’ Collar Isoform C (CnCC) / Drosophila Kelch like-ECH associated protein I (dKeap 1) and or Spineless (Ss) / Tango (tgo) in Anopheles gambiae (Diptera: Culicidae). A series of laboratory-based experiments and bioinformatic approaches were designed in order to test this hypothesis. Previous studies revealed that cytochrome P450s including CYP6M2 gene is established to be directly involved in the acquirement of insecticide resistance through the detoxification of pyrethroids and other WHO approved insecticides in the control of Anopheles gambiae (Antonio-Nkondjio et al., 2014; Balmert et al., 2014; Djègbè et al., 2014; Edi et al., 2014; Kabula et al., 2014; Lumjuan et al., 2014; Mitchell et al., 2014; Pang et al., 2014; Toé et al., 2014). Furthermore, some information is known about the up-regulation of these P450s by (CnCC) / (dKeap 1) and or Spineless (Ss) / Tango (tgo) signalling pathways in Drosophila melanogaster (Baanannou et al., 2013; Huang et al., 2013; Montenegro et al., 2013; Pickering et al., 2013; Kuzin et al., 2014; Li et al., 2014). Whether the mechanism underpinning this up-regulation operates in Anopheles gambiae is yet to be identified.
To this end, the work presented here aims to develop knowledge of mechanisms involved in the regulation of the cytochrome P450 CYP6M2 in insecticide resistant *Anopheles gambiae*.

In order to achieve this aim, specific goals/objectives were therefore designed as stated below:

- To identify and establish the nature and organization of the putative promoter elements or transcription factor binding sites (TFBS) present in insecticide detoxification P450 (CYP6M2) of *Anopheles gambiae*, especially with respect to Nuclear factor erythroid-2 related factor-2 (Nrf2) and/or Aryl hydrocarbon receptor (AhR) orthologs.
- To isolate, clone and characterize the cis-acting regulatory elements (promoter elements) of CYP6M2 from *Anopheles gambiae*.
- To examine the expression and regulation of CYP6M2, Nrf2 (CnCC) and AhR (Spineless) orthologs in *Anopheles gambiae* by using semi-quantitative (gel densitometry) and quantitative polymerase chain reaction (qPCR).

Subsequently, these specific goals were addressed in four interconnected and mutually complementary chapters. Chapter Two reported on the bioinformatic analysis of the selected regulatory elements within the 5' region upstream of CYP6M2 hypothesised to contain the promoter for this gene in *Anopheles gambiae*. Chapter Three covered the cloning of this gene into a pJET1.2 blunt cloning vector and then sub-cloning into a pGL3-Enhence luciferase vector. These bioinformatic analyses, cloning into pJET1.2 and sub-cloning of the CYP6M2 gene into pGL3 Enhencer expression vector form a pre-requisite for dual luciferase assay studies to demonstrate the up-regulation of CYP6M2 in *Anopheles gambiae*. Chapter Four investigated the dual luciferase assay...
studies using the constructs described in Chapter Three and conducted in the Anopheles gambiae Sua 5.1* cell line. Chapter Five investigated the differential expression of AGAP010259 (AhR), Nf2e1 (Nrf2) and CYP6M2 in the Tiassalé, Kisumu and Auyo strains of Anopheles gambiae using both semi-quantitative and SYBR® Green qPCR. Overall the work in this thesis has shown the following:

6.2 Identification of the orthologs to CnCC / dKeap 1 and / Spineless / Tango in Anopheles gambiae

Literature survey suggests that the identification of Anopheles gambiae orthologs to CnCC / dKeap 1 and / Spineless / Tango genes undertaken in this study was the first such identification and these findings are presented in Chapter Two (Section 2.3.9). Thus, through the use of bioinformatic resources, such as VectorBase (focused on Anopheles gambiae) and FlyBase (focused on Drosophila melanogaster), the orthologs of CnCC / dKeap 1 and / Spineless / Tango were established as Nuclear factor erythroid 2, invertebrate (Nf2e1) AGAP005300 / AGAP003645 & AGAP010259 / AGAP000748 respectively (Chapter Two, Table 2.4). Throughout the course of this thesis, the Drosophila melanogaster P450, CYP6G1 and its up regulators CnCC / dKeap1 & Ss / tgo have been used as reference points and for direct comparison with Anopheles gambiae data, when utilising bioinformatics resources such as ClustalX, multiple Clustal alignment and ConSite web based prediction site. These Drosophila orthologs have been reasonably well studied (Misra et al., 2013; Pickering et al., 2013; Deng, 2014; Deng & Kerppola, 2014; Kuzin et al., 2014; Li et al., 2014) and show reasonable degree of peptide sequence conservation (Chapter Two, Figures 2.5 A & B) by identity (Chapter Two, Table 2.1) to those of Anopheles gambiae.
6.3 Establishing putative transcription binding sites within the 5’ region upstream of CYP6M2

Literature review revealed that there was no previous data on the identification of location of putative transcription factor binding sites (TFBS) for (AGAP005300) (Nf2e1) (Nrf2) and (AGAP010259) (AhR) within the 5’ upstream region of CYP6M2 in Anopheles gambiae. Hence, the identification of these TFBSs within the 5’ region upstream of CYP6M2 in Anopheles gambiae is also presented in Chapter Two (Figure 2.6). In this study therefore, 39 AhR (AGAP010259) and 10 Nrf2 (Nf2e1) TFBS were established within the 5’ 896 bp region upstream of CYP6M2. Binding of AGAP010259 and or Nf2e1 genes at the TFBS could drive the transcription of CYP6M2 through their respective signalling pathways thereby influencing the metabolism of insecticides in Anopheles gambiae as discussed in Chapter One (Figures 1.7 & 1.8) and Chapter Five (Section 5.1) of this study. Prediction of conservation of CYP6M2 in Anopheles gambiae and CYP6G1 in Drosophila melanogaster was also established using ConSite web based predictions in Chapter Two (Section 2.3.5). This is consistent with the methodologies applied in the studies of other genes in the upstream regions of Bari-Jheh and Jheh in Drosophila melanogaster (Guio et al., 2014). The studies in Chapter Two (Sections 2.3.6 & 2.3.7) showing high conservation of protein primary structures between the Anopheles gambiae CYP6M2 and in Drosophila melanogaster CYP6G1 has implication for the cellular functionality equivalent though not necessarily identical.

6.4 Cloning of the 5’ 896 bp region upstream of CYP6M2 and construction of its luciferase gene reporter system

No other studies have been reported on the cloning of 5’ region upstream of CYP6M2 in the insecticide resistant Tiassalé and susceptible Kisumu strains of Anopheles gambiae. Here, using molecular biology techniques 5’ 896 bp region upstream of
CYP6M2 putative regulatory sequences were amplified, cloned and sequenced from both laboratory colony specimens (the Kisumu strain) and a recently established field strain (Tiassalé) to yield constructs for transfection into the Anopheles gambiae Sua 5.1* cell line for use in luciferase reporter assays (Chapter Three, Sections 3.6.2 to 3.6.6). These form a pre-requisite for dual luciferase assay studies to demonstrate the up-regulation of CYP6M2 by the D. melanogaster CnCC / dKeap 1 and or Spineless/ Tango orthologs (AGAP005300) (Nf2e1) / AGAP003645 and or AGAP010259 / AGAP000748 in Anopheles gambiae.

6.5 Differential expression of AGAP010259 (AhR) and Nf2e1 (Nrf2) candidate genes in selected strains of Anopheles gambiae

The experiments described in Chapter Five (Sections 5.3.1 & 5.3.2) of this study were performed in order to examine the expression and regulation of AGAP010259 (AhR) (Spineless) and Nf2e1 (Nrf2) (CnCC) in the Tiassalé, Kisumu and Auyo strains of Anopheles gambiae. This study, using SYBR® Green qPCR, was the first such study reported for these genes in Anopheles gambiae. The data gathered including that of CYP6M2 expression has served to highlight the differences in the expression levels of AGAP010259, Nf2e1 and CYP6M2 between the Tiassalé resistant, Kisumu susceptible, Auyo-selected and Auyo-unselected strains of Anopheles gambiae. The Anopheles gambiae strain from Nigeria challenged with various WHO recommended discriminating concentrations of insecticides, 0.1% Bendiocarb (Auyo-Bendi), 4%DDT (1, 1, 1-trichloro-2, 2-di (4-chlorophenyl) ethane) (Auyo-DDT) and 0.75% Permethrin (Auyo-Perm) were included in Chapter Five (Figure 5.4) of this study to address the earlier hypothesis and research questions raised in this study on the demonstration of CYP6M2 in Anopheles gambiae. Moreover, little is known about insecticide resistance in
the Sudan savannah region of northern Nigeria. With a notable exception of Ibrahim et al., (2014) who reported the exposure of Auyo strains of *Anopheles gambiae* to lamda cyclodithrin, a pyrethroid, no other studies were reported before this study.

The data from Chapter Five (Section 5.4.1) of this work suggests that both (AGAP010259) AhR and (Nf2e1) Nrf2 were expressed in all the selected strains of *An. gambiae*. Comparison between the Kisumu, Tiassalé and Auyo strains indicated that the level of expression of CYP6M2 as described in Chapter Five (Section 5.4.3) were higher in the wild population Auyo strains that survived the exposure to DDT (A-D(S)) (FC=7, p< 0.05) whilst Nf2e1 (Nrf2) expression levels were higher in the Auyo strains that survived exposure to Bendiocarb (A-Bendi(S)) (FC=14, p< 0.001) respectively.

The differential expressions of these genes in all the selected strains of *Anopheles gambiae* therefore provide a valuable insight into the potential roles of AGAP010259 / tango and Nf2e1 / dKeap 1 signalling pathways in insecticide resistance in *Anopheles gambiae*.

The over-expression of Nf2e1 and depletion of AGAP003645 (dKeap 1) in *Anopheles gambiae* potentially activates the transcription of many P450 genes including CYP6M2 and protect cells from xenobiotic compounds, whereas dKeap1 overexpression potentially represses their transcription, indicating that the functions of these protein families in the xenobiotic response are conserved between mammals, *Drosophila melanogaster* and *Anopheles gambiae* (Chapter Three, Sections 2.3.6 & 2.3.7). The differential expression of Nf2e1 in the selected insecticide resistant Tiassalé and Auyo strains of *Anopheles gambiae* in particular therefore has revealed a connection between Nf2e1 / dKeap1 signalling pathways and CYP6M2 in insecticide resistance. A proposed scheme for the Nf2e1/ AGAP003645 pathway and that of its inhibition are described below>
6.5.1 Nuclear factor erythroid 2 invertebrate (Nf2e1) / AGAP003645 signalling pathway in Anopheles gambiae

Under oxidative stress conditions, the inhibition of Nf2e1 by AGAP003645 is suppressed allowing these transcription factors to bind together with other proteins, as maf (muscle aponeurosis fibromatosis) to ARE sequences up regulating downstream P450 genes such as CYP6M2 responsible for detoxification of insecticides thereby conferring protection and possibly resistance to insecticides in Anopheles gambiae (Figure 6.1A).

Figure 6.1A: A proposed general scheme for the induction of Nf2e1/ AGAP003645- (dKeap 1) signalling pathway. The antioxidant response element (ARE) in the promoter region of select genes allows the coordinated up-regulation of antioxidant and detoxifying enzymes in response to oxidative/electrophilic stress. This up-regulation is mediated through Nuclear factor erythroid 2, invertebrate (Nf2e1) that may be activated by endogenous and exogenous molecules or stressful conditions. These agents disrupt the association between Nf2e1 and AGAP003645 with subsequent nuclear translocation of Nf2e1. In the cell nucleus, Nf2e1 interacts with small MAF (Muscle apoptosis fibromatosis) protein, forming a heterodimer that binds to the ARE (Antioxidant response element) sequence in the promoter region and up-regulates transcription of many genes encoding detoxifying enzymes such as CYP6M2.
6.5.2 Inhibition of the Nf2e1/ AGAP003645 pathway in Anopheles gambiae

Figure 6.1B is an Illustration of the hypothetical inhibition of the Nf2e1/ AGAP003645 signalling pathway which potentially hinders the up-regulation CYP6M2 responsible for insecticide metabolism in Anopheles gambiae.

Figure 6.1B: A proposed general scheme for the inhibition of NF2e1/ AGAP003645 (dKeap 1) -signalling pathway. This up-regulation is mediated through Nuclear factor erythroid 2 invertebrate (Nf2e1) that may be activated by endogenous and exogenous molecules including xenobiotics (ligands) or stressful conditions. When these agents are inhibited, the association between Nf2e1 and AGAP003645 remains intact in the cytoplasm. This disrupts the up-regulation and transcription of many genes encoding detoxifying enzymes such as CYP6M2.

6.6 General conclusions

These novel results suggest the possibility that the Nf2e1 / (AGAP003645) dKeap1 pathway is necessary for the overexpression of CYP6M2 in the selected and unselected Tiassalé, Kisumu and Auyo strains of Anopheles gambiae. However, activation of this
pathway is unlikely to be the only factor that contributes to insecticide resistance in the Tiassalé, Auyo-Perm, Auyo-Bendi and Auyo-DDT strains. Nevertheless, inhibition of the \( Nf2e1 / dKeap 1 \) (Figure 6.4 B) in particular and AGAP010259 / Tango response may potentially improve the efficacy of insecticides and development of methods to knock down or inhibit these pathways may prove fruitful. This work has increased our knowledge of the regulatory mechanisms involved in the control of \( CYP6M2 \) in insecticide resistance in \textit{Anopheles gambiae}. Even if the underlying mechanisms are still not very clear, this work shows the importance of these regulatory genes in the control of \( CYP6M2 \) in response to insecticide selection. These findings have implications in the ability to control the spread of malaria due to the reduction in insecticide resistance in \textit{Anopheles gambiae}.

### 6.7 Suggestions for further work

The study presented here, albeit being generally successful, was not without limitations. In this section the main limitations and suggestions of how to address them in order to confirm the aforementioned hypothesis in future research are hereby enumerated;

- It appears that this work only focuses on \( CYP6M2 \), but the significance of the work is far more important than just \( CYP6M2 \). The use of bioinformatic database resources and series of experimental approach in Chapters Two to Five of this work can be carried out with other Cytochrome P450 genes, such as \( CYP6P3 \) and \( CYP6Z2 \), which are also associated with insecticide resistance in order to gather more information on the AGAP010259 / Tango and \( Nf2e1 / dKeap 1 \) pathways involved in the regulation of these detoxification genes in \textit{Anopheles gambiae}. 
The study of gene functions using mosquito transgenic and RNA interference techniques would also be of interest. This is in order to reveal the expression of the \textit{CYP6M2} gene. Using this approach, the mutant \textit{CYP6M2} gene can be introduced into Tiassalé, Kisumu and Auyo strains of \textit{Anopheles gambiae} and their functions studied \textit{in vivo}. Moreover, specific gene expressions can also be silenced using double stranded RNA (dsRNA)-mediated interference (RNAi) techniques (van Cleef et al., 2014) to reveal the function of the AGAP010259 (AhR) and \textit{Nf2e1} (Nrf2) genes in the Tiassalé, Kisumu and Auyo strains of \textit{Anopheles gambiae} using this approach (Deng and Kerpolla, 2014; Wang et al., 2014b).

It would be interesting to investigate the differential expression of \textit{CYP6M2} between Tiassalé resistant and Kisumu susceptible strains using cDNA-AFLP (Amplified Fragment Length Polymorphisms) and microarray techniques. cDNA-AFLP method (Schuler and Berenbaum, 2013) is a commonly used, robust, and reproducible tool for genome-wide expression analysis in any species, without requirement of prior sequence knowledge. Here mRNAs of the Kisumu susceptible and Tiassalé resistant strains of the \textit{Anopheles gambiae} need to be prepared, and then reverse-transcripted to single strand cDNAs. Single strand cDNAs will be used to synthesize double stand cDNAs, and then subjected to enzyme digestion. The products of enzyme digestion will be used as templates for AFLP analysis using different primer sets. Microarray method (Wilding et al., 2014) is a collection of microscopic DNA spots attached to a solid surface used to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. In this technique, normalized mosquito cDNA library needs to be constructed. Then, a number
of PCR experiments will be carried out using cDNA library as templates. PCR products need then to be arrayed on chips. These chips need to be hybridized with cDNA from the insecticide susceptible Kisumu and resistant Tiassalé strains of *Anopheles gambiae* that are labelled with two different fluorophores. Using this method, the differentially expressed *CYP6M2* gene between Tiassalé resistant and Kisumu susceptible strains will be identified.


Bradford, J. 2014. Examining Culex tarsalis (Diptera: Culicidae) population changes with satellite vegetation index data.


References


References


Deng, H. 2014. Multiple roles of Nrf2-Keapl signalling. Fly, 8(1), 7-12.


References


References


References

Handschin, C., Podvinec, M., & Meyer, U. A. 2000. CXR, a chicken xenobiotic-sensing orphan nuclear receptor, is related to both mammalian pregnane X receptor (PXR) and constitutive androstane receptor (CAR). *Proceedings of the National Academy of Sciences, 97*(20), 10769-10774.


References


References


References


Ma, Q. & He, X. 2012. Molecular basis of electrophilic and oxidative defence: promises and perils of Nrf2. Pharmacological reviews, 64(4), 1055-1081.


Masi, A.D., Marinis, E.D., Ascenzi, P., Marino, M. 2009. Nuclear receptors CAR and PXR: Molecular, functional, and biomedical aspects, Molecular Aspects of Medicine, 30(5): 297-343.


References


Nie, Y., Liu, H., & Sun, X. 2013. The patterns of histone modifications in the vicinity of transcription factor binding sites in human lymphoblastoid cell lines. PloS one, 8(3), e60002.


References


Stradi, M. 2012. Genetic and biochemical characterization of the cytochrome P450, CYP6P9, associated with pyrethroid resistance in the African malaria vector Anopheles funestus (Doctoral dissertation), University of the Witwatersrand, Johannesburg, South Africa.


Symmons, O., & Spitz, F. 2013. From remote enhancers to gene regulation: charting the genome’s regulatory landscapes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620), 20120358.


References


Vázquez-Gómez, G., Sánchez-Santos, A., Vázquez-Medrano, J., Quintanar-Zúñiga, R., Monsalvo-Reyes, A. C., Piedra-Ibarra, E., ... & Heres-Pulido, M. E. 2010. Sulforaphane modulates the expression of *Cyp6a2* and *Cyp6g1* in larvae of the ST and HB crosses of the *Drosophila* wing spot test and is genotoxic in the ST cross. *Food and Chemical Toxicology*, 48(12), 3333-3339.


References


References


World Health Organisation (WHO) 2013b. Effective Malaria Control through Durable Housing Improvements: Can we learn new strategies from past experience? “World Malaria Report 2013 Fact Sheet.”


REFEEREED JOURNAL ARTICLES AND
CONFERENCE ABSTRACTS PUBLICATIONS FROM
THIS THESIS

B.R.Mohammed
-150-          Abertay University, Dundee, UK
The content of Appendices I-III, pages 151-173, have been removed to comply with copyright law. The content included 3 journal articles and a conference abstract. The citations to the articles are listed below.


Appendix V

GERMAN CONFERENCE ON BIOINFORMATICS, 2014
http://www.gcb2014.de/icndex.php/proceedings

ABSTRACT


Bioinformatic Analysis of Regulatory Elements within the Promoter Region of the Cytochrome P450 gene, CYP6M2 in Anopheles gambiae

Abstract: Cytochrome P450s including CYP6M2 have been demonstrated to be involved in the metabolism of insecticides, typically through up regulation in resistant individuals. The difference in gene expression levels seen in insecticide resistant mosquitoes may result from sequence differences in the 5’ upstream region of CYP6M2 including those in the promoter elements. Understanding the complex mechanisms regulating CYP6M2 expression in insecticide resistant Anopheles gambiae remains a great challenge. In this study, extensive bioinformatics resources were used to predict regulatory elements and cross-species comparison in the cis-acting elements within CYP6M2 (896 bp) and CYP6G1 (896 bp) known to be upregulated by the orthologs of Nuclear factor-erythroid 2-related factor-2 (Nrf2)/Kelch-like ECH-associated protein 1 (Keap1) and Aryl hydrocarbon receptor (AhR)/Aryl hydrocarbon receptor nuclear translocator (ARNT) in Drosophila melanogaster. Searches were also made for the cis-acting elements within an 896 bp region upstream of CYP6M2 hypothesised to contain the promoter for this gene in both the Tiassalé multiple insecticide-resistant and Kisumu susceptible strains of An. gambiae. Results revealed the presence of Nrf2/Keap1 and AhR/ARNT as putative transcription factor binding sites (TFBS) within the CYP6M2 promoters. Further, we identified the orthologs of those transcription factors which bind to these elements (Cap ‘n’ collar isoform C (CnCC)/dKeap1) & Spineless (Ss) / Tango (tgo) in Drosophila melanogaster as AGAP010259/AGAP009748 & AGAP005300 d/ AGAP003645 in An. gambiae. These data suggest the presence of putative AGAP010259/AGAP009748 and AGAP005300 /AGAP003645 binding sites in the promoter of An. gambiae CYP6M2, which may potentially be associated with the up regulation of CYP6M2 involved in insecticide resistance. These if established have implications in the control of malaria.