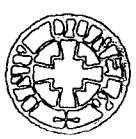
Receptors of peptide hormones in the gut of rice brown planthopper (Nilaparvata lugens) as targets for novel insecticides

Price, Daniel Richard Glenfield

Receptors of Peptide Hormones in the Gut of Rice Brown Planthopper (*Nilaparvata lugens*) As Targets For Novel Insecticides

A thesis submitted by Daniel Richard Glenfield Price, B.Sc in accordance with the requirements of the University of Durham for the degree of Doctor of Philosophy.

Department of Biological and Biomedical Sciences
University of Durham
June 2004



Abstract

RECEPTORS OF PEPTIDE HORMONES IN THE GUT OF RICE BROWN PLANTHOPPER (NILAPARAVTA LUGENS) AS TARGETS FOR NOVEL INSECTICIDES

Daniel.R.G.Price

The rice brown planthopper, *Nilaparvata lugens*, is a pest of rice where it can be responsible for single season crop losses of up to 38% (IRRI statistic). Due to the absence of an effective *Bacillus thuringiensis* (*Bt*) toxin against these insect pests current control methods are reliant on synthetic insecticides. However, these broad acting insecticides can be harmful to beneficial insects, and resistance to these synthetic insecticides amongst *N.lugens* populations is now common. Therefore, other novel approaches are needed for the control of this rice pest species, which is addressed in this thesis. This project aims to characterise proteins expressed in the *N.lugens* gut that could act as potential 'targets' for the design of insecticidal proteins. G protein coupled receptor (GPCR) proteins were identified as 'target' proteins and *N.lugens* candidates were identified and characterised.

A cDNA predicted to encode a diuretic hormone receptor (DHR) in *Nilaparvata lugens* was cloned using RNA extracted from gut tissue as a template in a PCR-based strategy. The coding sequence (639 amino acids, M_r 69.7 kDa) has seven predicted transmembrane domains and is a member of the calcitonin/secretin/corticotropin releasing factor family of G-protein coupled receptors. The protein has greatest similarity to *Acheta domesticus* DHR (Q16983) with 51% sequence identity. *N.lugens* DHR has a large N-terminal extracellular domain (amino acids 1-260) putatively involved in hormone binding. This domain was expressed as a recombinant protein in *E.coli*, purified under denaturing conditions, refolded and used to raise a polyclonal antibody in rabbit. Purified anti-DHR IgG bound specifically to a putative DHR polypeptide extracted from *N.lugens* gut tissues on western blots. Immunolocalisation experiments using dissected guts showed that anti-DHR antibody bound specifically to the Malpighian tubules. The N-terminal hormone-binding domain is located on the cell surface and is exposed to the haemolymph *in vivo*. Anti-DHR antibodies delivered to insects via artificial diet showed no binding to gut or tubule tissue, and had no effect on survival. Ingested antibodies were not detected in the haemolymph.

Another GPCR protein with high similarity to insect allatostatin receptors (ALSTR) was isolated by PCR using highly degenerate primers, from *N.lugens* gut specific cDNA. The coding sequence (383 amino acids, M_r 42.3 kDa) has seven predicted transmembrane domains and is a member of the SST/galanin/opoid receptor family of G-protein coupled receptors. The putative *N.lugens* ALSTR has greatest similarity to *Periplaneta americana* ALSTR (AF336364) with 69% overall sequence identity. Isolation of the full-length *N.lugens* ALSTR is the first step necessary for the design of novel insecticidal proteins.

Declaration

No material here has previously been submitted for any other degree. Except where acknowledged, all material is the work of the author.

Statement of Copyright

Date....28/10/04.....

Signed....Daniel R.G. Price.....

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION 1

1.1 PROJECT BACKGROUND 1

1.2 SAP SUCKING INSECTS-PLANT: INSECT INTERACTIONS 2

1.2.1 Constitutive Plant Defence 2

1.2.2 Inducible Plant Defence 3

1.3 *N.LUGENS* AND RICE-THE PROBLEM 4

1.3.1 The Importance of Rice 4

1.3.2 Promotion of N.lugens *as a Major Pest* 5

1.3.3 Life Cycle of N.lugens 6

1.3.4 Crop Damage 7

1.3.5 N.lugens *as a Virus Vector* 7

1.3.5.1 Rice Grassy Stunt Virus (RGSV) 8

1.3.5.2 Rice Ragged Stunt Virus (RRSV) 8

1.4 CONTROL OF *N.LUGENS* 9

1.4.1 Synthetic Chemical Control 9

1.4.2 Plant-Derived Insecticidal Chemicals 11

1.4.3 Plant-Derived Insecticidal Genes 12

1.4.4 Selective Breeding and N.lugens *Resistance* 14

1.4.5 Integrated Pest Management (IPM) 14

1.5 THE FUTURE OF *N.LUGENS* CONTROL-NOVEL APPROACHES 15

1.5.1 Insecticidal Proteins - The Plantibody Approach 15

1.6 G-PROTEIN COUPLED RECEPTOR (GPCR) PROTEINS 19

1.6.1 GPCR Structural Classification 20

1.7 DROSOPHILA GENOME SEQUENCE AND PEPTIDE GPCRS 21

1.7.1 Drosophila Peptide GPCRs – Family A 22

1.7.2 Drosophila Peptide GPCRs – Family B 22

1.8 GPCR RECEPTOR SIGNALLING 23

1.9 NEUROHORMONAL AND MODULATORY ACTIONS OF INSECT PEPTIDES 24

1.9.1 Neurosecretory Cells 25

1.9.2 Corpora Cardiaca 25

1.9.3 Prothoracic Glands 25

1.9.4 Corpora allata 26

1.9.5 Midgut Endocrine Cells 26

1.10 DIVERSITY OF INSECT NEUROPEPTIDES AND RECEPTORS 26

1.10.1 Proctolin Receptor 28
1.10.2 Pyrokinin Receptor 29
1.10.3 Ecdysis Triggering Hormone Receptors 30
1.10.4 Neuropeptide F receptor 30
1.10.5 Allatostatin Receptor (ALSTR) 31
1.11 WATER REGULATION IN INSECTS-WHAT IS KNOWN? 33
1.11.1 Gut Physiology and Adaptation For Water Regulation 34
1.11.2 The CRF-like Diuretic Peptides (CRF-DPs) 35
1.11.3 The CRF-Like Diuretic Peptide Receptor (DPRs) 36
1.11.4 The Kinins 38
1.11.5 The Leucokinin Receptor 40
1.11.6 Synergism Between DPS (CRF-Like and Kinins) 43
1.12 RESEARCH OBJECTIVES 45

CHAPTER 2. MATERIALS AND METHODS. 46

2.1 *N.LUGENS* INSECT CULTIVATION 46
2.2 *N.LUGENS* ARTIFICIAL DIET FEEDING TRIALS 46
2.2.1 Planthopper Artificial Diet (MMD-1) 46
2.2.2 Artificial Diet Feeding Trials 47
2.3 *ESCHERICHIA COLI* STRAINS 47
2.4 INSECT CELL EXPRESSION CELL LINES 49
2.5 CHEMICALS AND REAGENTS 49
2.6 STANDARD MOLECULAR BIOLOGY TECHNIQUES 50
2.6.1 Bacterial Culture 50
2.6.2 Competent Cells 50
2.6.3 Transformation of Competent Cells 51
2.6.4 Bacterial Glycerol Stocks 51
2.6.5 Plasmid DNA Mini-Preparation 52
2.6.6 Restriction Endonuclease Digestion 52
2.6.7 Agarose Gel Electrophoresis of DNA 52
2.6.8 Recovery of DNA Fragments From Agarose Gels 53
2.6.9 Nucleic Acid Quantification and Qualification 53
2.6.10 Ethanol Precipitation of Nucleic Acid 53
2.6.11 DNA Ligation 54
2.6.12 DNA Sequencing and Sequence Analysis 54
2.6.13 Oligonucleotides 55
2.6.14 Purification of Total RNA 55
2.6.15 Purification of Poly-A$^+$ RNA 56

2.6.16 Formaldehyde Gel Electrophoresis of RNA 56
2.6.17 Standard Polymerase Chain Reaction (PCR) 57
2.6.18 Reverse Transcription-PCR (RT-PCR) 58
2.6.19 Subcloning PCR Products 59
2.6.20 Rapid Amplification of cDNA Ends (RACE) PCR 59
2.6.21 Transfer of DNA From Agarose Gels to Nitrocellulose 63
2.6.22 Radioactive Labelling of DNA Using Random Primers 63
2.6.23 Hybridising Labelled DNA Probes to Filters Containing DNA 64
2.6.24 N.lugens *Whole Insect cDNA Library* 64
2.6.25 N.lugens *Gut Specific cDNA Library* 65
2.7 STANDARD PROTEIN ANALYSIS 65
2.7.1 Estimation of Protein Concentration 65
2.7.2 SDS PAGE Electrophoresis 66
2.7.3 Staining Gels With Coomassie Blue 67
2.7.4 Immunoblotting 67
2.7.5 Protein Extraction From Whole Tissues 69
2.7.6 Expression of Recombinant Proteins in E.coli and Purification under Denaturing Conditions using Nickel Affinity Chromatography 70
2.7.7 SELDI-TOF Mass Spectrometry 71
2.7.8 Expression of Recombinant Proteins in Insect Cells 71
2.7.9 Synthetic Peptide Synthesis 72
2.7.10 Polyclonal Antibody Production 72
2.7.11 IgG Antibody Purification 72
2.8 DIURETIC HORMONE RECEPTOR METHODS 73
2.8.1 Purification of Gut Specific Total RNA From N.lugens 73
2.8.2 RT-PCR Cloning of a DH-R Fragment 75
2.8.3 RACE for Generation of Full Length DH-R cDNA 75
2.8.4 Insect Expression Construct: Full Length DHR 76
2.8.5 Bacteria Expression Construct: DHR N-Terminal Domain 77
2.8.6 Over Expression and Purification of the N.lugens *DHR N-terminal Domain* 78
2.8.7 anti-DHR Antibodies 79
2.8.8 anti-DHR IgG Feeding Trials 80
2.8.9 Immunolocalisation of DHR in N.lugens *using Confocal Microscopy* 81
2.8.10 N.lugens *Feeding Trials with Recombinant DHR N-terminal Domain* 83
2.9 ALSTR METHODS 83
2.9.1 RT-PCR Cloning of an ALSTR Fragment 84
2.9.2 RACE for Generation of Full Length ALSTR cDNA 84

2.9.3 Bacteria Expression Construct: ALSTR N-Terminal Domain *86*
2.9.4 Over Expression of the ALSTR N-Terminal Region *88*
2.9.5 anti-ALSTR Antibodies *88*

CHAPTER 3. IDENTIFICATION OF *N.LUGENS* GUT SPECIFIC TARGET PROTEINS 90

3.1 INTRODUCTION 90
3.2 RESULTS 91
3.2.1 Degenerate Primer Design *91*
3.2.2 General GPCR Primers *92*
3.2.3 DHR Degenerate Primers *95*
3.2.4 ALST-R Degenerate Primer Design *97*
3.2.5 Verification of N.lugens Gut Specific cDNA Library *102*
3.2.5.1 Insert Size Distribution *102*
3.2.5.2 PCR and RT-PCR-Cathepsin and Trypsin *106*
3.2.6 Verification of N.lugens cDNA Library *109*
3.2.7 Mass Sequencing From N.lugens cDNA Library *109*
3.3 DISCUSSION 109

CHAPTER 4. THE *N.LUGENS* DIURETIC HORMONE RECEPTOR 114

4.1 INTRODUCTION 114
4.2 RESULTS 115
4.2.1 Isolation of a cDNA encoding a DHR from the gut of N.lugens *115*
4.2.2 DHR cDNAs (DHR-I and DHR-II) *117*
4.2.3 DHR Sequence Analysis *119*
4.2.4 Bacterial and Insect Cell Expression Constructs *125*
4.2.5 Variable cDNA Sequences Encoding N.lugens DHR *125*
4.2.6 Production of anti-DHR antibody by means of expression of recombinant DHR *132*
4.2.7 Feeding trials with recombinant DHR N-terminal domain *134*
4.2.8 Insect Cell Expression of the Full-Length N.lugens DHR *136*
4.2.9 Anti-DHR Antibodies *140*
4.2.10 Immunolocalisation of DHR in the N.lugens Gut *142*
4.2.11 Feeding trials with anti-DHR antibody *145*
4.3 DISCUSSION 148
4.3.1 Cloning N.lugens *DHR* *148*
4.3.2 N.lugens *DHR Sequence Analysis* *149*
4.3.3 Ligand Binding and Receptor Activation *151*
4.3.4 N.lugens *Feeding Trial with Purified Recombinant DHR N-terminal domain* *153*
4.3.5 N.lugens *DHR Coding Sequence Expression* *154*

4.3.6 Anti-DHR Antibody Characterisation *154*
4.3.7 Immunolocalisation of N.lugens *DHR in Dissected Gut Tissue* *155*
4.3.8 anti-DHR Feeding Trials *156*

CHAPTER 5. THE *N.LUGENS* ALLATOSTATIN RECEPTOR 158

5.1 INTRODUCTION 158
5.2 RESULTS 158
5.2.1 Isolation of a cDNA encoding a ALSTR from the gut of N.lugens *158*
5.2.2 Bacterial Expression Constructs *169*
5.2.3 Production of anti-ALSTR antibodies *170*
5.2.4 Characterisation of anti-ALSTR Antibodies *172*
5.3 DISCUSSION 175
5.3.1 Cloning N.lugens ALSTR *175*
5.3.2 N.lugens *ALSTR Sequence Analysis* *175*
5.3.3 Alternative ALSTR 3' Ends *178*
5.3.4 Recombinant Protein Expression and Anti-ALSTR Antibodies *179*

CHAPTER 6. GENERAL DISCUSSION 180

REFERENCES 185

List of Figures

Figure 1.1 Antibody mediated modulation of target proteins.....17

Figure 1.2 Amino acid alignment of insect CRF-like diuretic hormones.....37

Figure 1.3 Amino acid alignment of insect diuretic hormone receptors (DHRs).....39

Figure 1.4 Amino acid sequence alignment of invertebrate kinin receptors.....44

Figure 2.1 Planthopper artificial diet feeding trial.....48

Figure 2.2. 3' RACE overview.....61

Figure 2.3. 5' RACE overview.....62

Figure 2.4. Semi-dry western blot set up.....68

Figure 3.1 Amplification of GPCR coding sequences with 'generic' degenerate primers.....94

Figure 3.2 Amino acid alignment of insect Family B peptide receptor proteins.....96

Figure 3.3 Nucleotide alignment of insect Family B peptide receptor proteins.....98

Figure 3.4 GPCR Family B degenerate primer design.....99

Figure 3.5 Amino acid alignment of insect Family A peptide receptor proteins.....101

Figure 3.6 Nucleotide alignment of insect Family A peptide receptor proteins.....103

Figure 3.7 GPCR Family A degenerate primer design.....104

Figure 3.8 GPCR Family A nested degenerate primer design.....105

Figure 3.9 Size distribution of cDNA inserts within *N.lugens* gut specific library.....107

Figure 3.10 PCR amplification of coding sequences from *N.lugens* gut specific library.....107

Figure 3.11a Hydrophobicity analysis of the *N.lugens* rhodopsin-like cDNA.....110

Figure 3.11b cDNA sequence of *N.lugens* rhodopsin-like cDNA.....111

Figure 4.1 PCR amplification of *N.lugens* DHR fragment using degenerate primers.....116

Figure 4.2 Size prediction of *N.lugens* DHR 3' and 5' amplification products.....116

Figure 4.3 DHR 3' RACE on *N.lugens* gut specific cDNA.....118

Figure 4.4 DHR 5' RACE on *N.lugens* gut specific cDNA.....118

Figure 4.5a Sequence of *N.lugens* DHR-like cDNA....120

Figure 4.5b Hydrophobicity analysis of the *N.lugens* DHR-like cDNA....120

Figure 4.6a Amino acid sequence alignment of related DHR sequences....123

Figure 4.6b Phylogenetic relationship of the previously aligned sequences....123

Figure 4.7 Amplification of the *N.lugens* DHR coding sequence....126

Figure 4.8 Amplification of the predicted *N.lugens* DHR N-terminal domain....126

Figure 4.9 Restriction analysis of the cloned *N.lugens* coding sequence....128

Figure 4.10a Sequence data from 'long' DHR isoform....129

Figure 4.10b Sequence data from 'short' DHR isoform....129

Figure 4.11 PCR amplification of the variable DHR from genomic DNA....130

Figure 4.12 Sequence analysis of cDNA and genomic DNA from around the variable....131

Figure 4.13a Expression and purification of the N-terminal *N.lugens* DHR in *E.coli*....133

Figure 4.13b SELDI-TOF MS analysis of the purified recombinant N-terminal domain....133

Figure 4.14 Artificial diet feeding trial with purified recombinant DHR N-terminal domain....135

Figure 4.15a SDS-PAGE analysis of expressed *N.lugens* DHR in Sf21 insect cell line....137

Figure 4.15b Western blot analysis of expressed *N.lugens* DHR in Sf21 insect cell line....137

Figure 4.16 Expression analysis of *N.lugens* DHR in a transformed Sf21 insect cell line....139

Figure 4.17 Western blot using anti-DHR antibodies against gut proteins from *N.lugens*....141

Figure 4.18 Confocal analysis of dissected *N.lugens* gut tissue with anti-DHR IgG....143

Figure 4.19 *N.lugens* artificial diet feeding trial with purified anti-DHR IgG....146

Figure 4.20 The fate of ingested anti-IgG in *N.lugens*....147

Figure 5.1a Location of degenerate ALSTR primers and expected PCR product sizes....160

Figure 5.1b Cloning a putative *N.lugens* ALSTR fragment with degenerate primers....160

Figure 5.2 ALSTR 5' RACE on *N.lugens* gut specific cDNA....162

Figure 5.3 ALSTR 3' RACE on *N.lugens* gut specific cDNA....163

Figure 5.4a Sequence of *N.lugens* ALSTR-like cDNA....165

Figure 5.4b Hydrophobicity analysis of the *N.lugens* ALSTR-like cDNA....165

Figure 5.5a Amino acid sequence alignment of related ALSTR sequences.....168

Figure 5.5b Phylogenetic relationship of the previously aligned sequences.....168

Figure 5.6 Amplification of *N.lugens* ALSTR coding sequence.....171

Figure 5.7 Amplification of *N.lugens* ALSTR N-terminal region.....171

Figure 5.8a Expression of the N-terminal *N.lugens* ALSTR in *E.coli*.....173

Figure 5.8b Western blot analysis of N-terminal *N.lugens* ALSTR in *E.coli*.....173

List of Tables

Table 1.2 Insect CRF-like diuretic peptides.....37

Table 1.2 Insect and related kinin neuropeptides.....41

Table 2.1. Planthopper artificial diet composition.....48

Table 2.2. DHR primers.....74

Table 2.3. ALSTR primers.....85

Table 5.1 ALSTR regions and synthetic peptides.....174

Abbreviations

AD	-	artificial diet
ALSTR	-	allatostatin receptor
ATP	-	adenosine triphosphate
Bis-acrylamide	-	bis (N, N' –methylene-bis-acrylamide)
bp	-	base pair
BPH	-	brown planthopper (*Nilaparvata lugens*)
BSA	-	bovine serum albumin
^{o}C	-	degrees centigrade
cDNA	-	complementary DNA
CDS	-	coding sequence
Ci	-	Curie
cpm	-	counts per minute
CRF		corticotropin releasing factor
DEPC	-	diethyl pyrocarbonate
DHR	-	diuretic hormone receptor
DMSO	-	dimethylsulphoxide
DMF	-	dimethylformamide
DNA	-	deoxyribonucleic acid
ds	-	double stranded
DTT	-	dithiothreitol
ECL	-	enhanced chemiluminescence
EDTA	-	ethylenediaminotetracyclic acid
EtBr	-	ethidium bromide
g	-	gram
GDP	-	guanosine diphosphate
GM	-	genetically modified
GPCR	-	G-protein coupled receptor

GSP	-	gene specific primer
GTP	-	guanosine triphosphate
GRK	-	G-protein coupled receptor kinase
HRP	-	horseradish peroxidase
hr	-	hours
I	-	inosine
IgG	-	immunoglobulin G
IPM	-	integrated pest management
IPTG	-	isopropyl-β-d-thiogalactoside
kbp	-	kilobase pair
kDa	-	kilo Dalton
L	-	litre
LB	-	Luria-Bertani (culture medium)
m	-	metre
M	-	molar
M-MLV	-	Moleney Murine Leukemia Virus
min	-	minute
M_r	-	molecular weight
mRNA	-	messenger RNA
MT		Malpighian tubule
OD	-	optical density
O/N	-	overnight
ORF	-	open reading frame
PAGE	-	poly acrylamide gel electrophoresis
PBS	-	phosphate-buffered saline
PCR	-	polymerase chain reaction
pfu	-	plaque forming unit
RACE	-	rapid amplification of cDNA ends

rcf	-	relative centrifugal force
RER	-	rough endoplasmic reticulum
r.h	-	relative humidity
RNA	-	ribonucleic acid
RNAi	-	RNA interference
rpm	-	revolutions per minute
RT	-	reverse transcription
rt	-	room temperature
SDS-PAGE	-	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	-	seconds
ss	-	single stranded
SSC	-	saline sodium citrate
T_a	-	annealing tamperature
T_m	-	melting temperature
Taq	-	*Thermus aquaticus*
TBS	-	tris-buffered saline
TCP	-	total cell protein
Temed	-	*N,N,N,'N'* -tetramethylethylenediamide
TM	-	transmembrane
TSP	-	total soluble protein
UTR	-	untranslated region
UVB	-	ultraviolet-B (λ 302 nm)
v/v	-	volume for volume
w/v	-	weight for volume
X-Gal	-	5-bromo-4-chloro-3-indoyl-β-D-galactose

Amino Acid Abbreviations

A	Ala	-	alanine
R	Arg	-	arginine
N	Asn	-	asparagine
D	Asp	-	aspartic acid
C	Cys	-	cysteine
Q	Gln	-	glutamine
E	Glu	-	glutamic acid
G	Gly	-	glycine
H	His	-	histidine
I	Ile	-	isolucine
L	Leu	-	leucine
K	Lys	-	lysine
M	Met	-	methionine
F	Phe	-	phynylalanine
P	Pro	-	proline
S	Ser	-	serine
T	Thr	-	threonine
W	Trp	-	tryptophan
Y	Tyr	-	tyrosine
V	Val	-	valine

Nucleic Acid Abbreviations

A	-	adenine
T	-	thymine
G	-	guanine
C	-	cytosine

U	-	uracil

Degenerate Nucleic Acid Abbreviations

B	-	C/G/T
D	-	A/G/T
H	-	A/C/T
K	-	G/T
M	-	A/C
N	-	A/T/G/C
R	-	A/G
S	-	G/C
V	-	A/C/G
W	-	A/T
Y	-	C/T

Acknowledgements

I am very grateful for the supervision, constructive criticism and general support provided by Dr. John. A. Gatehouse (University of Durham) and Drs. Jane Cayley and Simon Warner (Syngenta Plc) throughout this project. CASE Funding by BBSRC (Biotechnology and Biological Sciences Research Council) and Syngenta Plc is also gratefully acknowledged.

Thanks to the past and present members of lab 1 who created a friendly working environment, and offered endless advice, particularly Dr. Dave Bown, Miss Hillary Wilkinson, Dr. Elaine Fitches, Dr. Jinping Du and Dr. Martin Edwards. My appreciation is also extended to staff at the School of Biological and Biomedical Sciences, University of Durham; in particular, Mr John Gilroy for protein freeze-drying and Mrs Christine Richardson, Dr Tijs Ketelaar for advice on confocal microscopy.

Thanks must go to those who I have made friends with during my stay in Durham, particularly, Karima Soutsane, Chris Wicker, Dave Page, Emma Clough, Gareth Marlow, Despina Kaloriti, Alessandra Badari, Hannah Hughes, Melissa, Janna, Ruchika Gajwani, Shankar, Vicky O'Brien; who have all made Durham an enjoyable place to be.

Finally, and by no means least, I am indebted to my family for their moral and financial support, guidance, and encouragement throughout my education. A special thanks to my brothers Dave and Steve who provided constant encouragement.

CHAPTER 1

INTRODUCTION

1.1 Project Background

The sap-sucking insects belong to the order Homoptera, and include the leafhoppers, planthoppers and aphids. Many of these insects reach pest status and attack economically important food crops. Control of these insect pests by conventional methods is difficult due to their rapid reproduction rate and the development of resistance to insecticides. Despite the synthesis of improved insecticides, integrated pest management (IPM) strategies and changes in farming practice, Homopteran pests are still proving difficult to control. This difficulty is extended into the genetic engineering approach due to the absence of effective *Bt*-toxins (*Bacillus thuringiensis* toxins). As an alternative strategy, plant-derived insecticidal proteins have been isolated, the most effective of which when fed in artificial diet experiments is the lectin from *Galanthus nivalis*, GNA (Powell *et al.*, 1995). Expression of this protein in rice plants has been shown to confer partial resistance to *Nilaparvata lugens* (brown plant hopper, BPH) and *Nephotettix virescens* (green plant hopper, GPH) (Foissac *et al.*, 2000). However, the insecticidal properties of GNA are unlikely to offer commercially viable crop protection against *N.lugens* and *N.virescens* at the field level (Gatehouse *et al.*, 2000).

It is clear that a new approach is needed for production of insecticidal proteins against Homopteran pest species. This need will be addressed by the present study, the aim of

which is to identify gut specific proteins in *N.lugens* that have a key role in regulation of insect homeostasis, metabolism and/or development. It can be anticipated that antibody-mediated interference of such key proteins and modulation of their function will be detrimental to insect survival. With the recent advances in genetic modification (GM), expression of such insecticidal proteins within target crops could serve as an effective delivery system for highly specific insecticidal proteins. Such an approach has already been used to confer plant resistance to otherwise virulent strains of virus, and a pathogenic nematode worm (De Jaeger *et al.*, 2000).

1.2 Sap Sucking Insects-Plant: Insect Interactions

The diversity and abundance of phloem feeding insects stands in contrast with the limited amount of information available on the constitutive and adaptive response of plants to this form of biotic stress. However, characterisation of the dynamic interactions that exist between plant and phloem feeding insect is a rapidly expanding field of research (Gatehouse, 2002). Knowledge of these plant-insect interactions will be crucial for the development of successful control strategies.

1.2.1 Constitutive Plant Defence

Passive anatomical features, such as the cuticle and the cell wall represent highly effective obstacles to penetration by most insects. However, the Homopteran insects penetrate plant tissue by probing intercellularly through epidermal and mesophyll cell layers using their stylet-like mouthparts. By tapping into the phloem sieve elements, sap-sucking insects have access to a sugar-rich diet. The diversion of photoassimilates

by insect feeding inflicts considerable fitness costs in many plant species (Moran and Thompson, 2001).

Plants synthesise and sequester a vast array of secondary metabolites, many of which are toxic to attacking insect pests. Compounds, such as phenols, alkaloids, and glycosides have known insecticidal activity (Bennett and Wallsgrove, 1994); therefore it is probable that they contribute to plant insect resistance. The distribution of these insecticidal compounds is often tissue specific and restricted to peripheral cell layers (Bennett and Wallsgrove, 1994), an ideal location for deterring attacking insects. Insecticidal phytochemicals are often toxic to plant material; therefore in healthy plants these chemicals are stored within vacuoles and organelles, as inactivated or cytoplasmic precursors (Osbourn, 1996). Tissue damage as a result of insect attack results in the massive release of stored toxic phytochemicals. For example, sulphur-rich glucosinolates are present in plants, sequestered in vacuoles in an inactive form. Disruption of vacuoles by insect attack exposes glucosinolates to cytosolic hydrolyses (myrosinases) and highly toxic products are released. The myrosinase-glucosinolate system is known to affect the activities of numerous insect herbivores (Peterson *et al.*, 1998). In *Arabidopsis thaliana*, glucosinolate-rich cells, or S-cells (sulphur containing cells), are found between the endodermis and the vascular bundle. This may be an evolutionary adaptation to prevent sap-sucking insects locating phloem elements and feeding (Koroleva *et al.*, 2000).

1.2.2 Inducible Plant Defence

A dynamic biochemical interaction exists between Homopteran insects and their host plants, usually balanced in the favour of the insect. For example, during feeding sap-sucking insects secrete watery saliva containing a mix of peroxidases, β-glucosidases and other potential signal-generating molecules; these compounds divert or counter responses at the stylet interface. Some of the non-enzymatic chemicals are thought to oxidise and inactivate defensive phytochemicals that are released by the plant as a result of tissue disruption (Miles, 1999). The best biochemically characterised interaction between a phloem-feeding insect and plant, is the one between *Myzus persicae* (green peach aphid) and *A.thailana* (Moran and Thompson, 2001). This interaction between aphid and host plant suggests that phloem feeding could induce a subset of responses associated with wounding and pathogen infection, bridging the gap between these two independent forms of stress. Salicylic acid (SA) dependent responses to aphid feeding are similar to those associated with virulent pathogen attack. Wound associated responses are triggered as a result of tissue damage caused by aphid feeding (Stout *et al.*, 1999). Further research in this area promises to generate many novel genes and mechanisms involving plant perception, tolerance and in some cases, resistance to phloem-feeding insects.

1.3 *N.lugens* and Rice-The Problem

1.3.1 The Importance of Rice

Cultivated rice *Oryza sativa* L., is mainly a subtropical and tropical crop, but can be grown at latitudes of 53°N and altitudes up to 2500m. Rice is cultivated in diverse

conditions: in flooded fields, uplands, deep-water areas and tidal wetlands. Approximately 150 million ha^{-1}, around 10% of the world's total arable land is cultivated with rice, equating to an annual production of 520 million tons. About 92% of the world's rice production comes from Asia, where it is a staple food supplying 36% of human total calorie consumption, compared to 20% worldwide (all statistics were derived from the International Rice Research Institute, IRRI, web page, http://www.irri.org/). It is therefore of no surprise that rice is the world's most important cereal crop. Unfortunately, rice is host for more diseases and pests than any other commercially grown crop. In the field, over 100 insect species attack rice, although only 20 are of major economic importance. Included in this group are the sap-sucking plant hoppers and leafhoppers (hemiptera).

1.3.2 Promotion of N.lugens *as a Major Pest*

In the Asian tropics, high yielding varieties (HYV) of rice were introduced in the mid-1960s. These high yielding, semi-dwarf varieties rose in popularity and have been planted in many rice growing areas. Such cultivar improvements have had an enormous impact on Asian agriculture and have resulted in huge increases in rice yield. This transition is commonly referred to as the 'Green-Revolution'. HYVs of rice are short, photo-insensitive, have high tillering capabilities and a 4 month growing season. The short generation time of this rice allows two crops per season. This double-cropping strategy allows rice to be grown all year round. However, increases in yield are not without significant drawbacks, the continuous growth of a monoculture rice crop has promoted the spread and development of previously unimportant pests and pathogens. For example, *N.lugens* has been promoted from a secondary to a primary pest of rice,

now causing serious and persistent reductions in rice yields (Pingali and Gerpacio, 1997). The rise of *N.lugens* to pest status coincides with the introduction of HYV rice by the IRRI and the associated intensification in farming practice. The intensification of farming practice post-1969 saw an upsurge in broad-spectrum pesticide use, during the period 1969-1971 there was a 33-fold increase (Holt *et al.*, 1996). This change in farming practice resulted in an exponential growth of the *N.lugens* population, due to the slow recovery of natural *N.lugens* predators and parasitoids following insecticide treatment (Pingali and Gerpacio, 1997). This resurgence is responsible for elevating *N.lugens* from a secondary to a primary pest. Worse still, over use of insecticides has encouraged insecticide resistant strains of *N.lugens* to appear. Because of the global dependence and economic importance of rice, *N.lugens* poses a serious threat to the ever increasing demand for improvements in rice yield.

1.3.3 Life Cycle of N.lugens

For normal behaviour and optimal development, *N.lugens* requires a temperature around 25-29°C, and a relative humidity between 70-95%. Temperature variations either side of this optima can affect hatchability, survival rate, egg period, nymphal period, and preoviposition period of *N.lugens* insects (Pingali and Gerpacio, 1997). During development, *N.lugens* pass through 5 nymphal stages, with each stadia lasting between 7 and 11 days, instars can be differentiated by their body shape. When mature, adult *N.lugens* measure between 2-3.5mm, where the females are slightly larger than the male. As implied by the name, fifth instar *N.lugens* are brown in colour and can hop up to 21 cm, although they can also fly and walk (Gullan and Cranston, 1998). Adult *N.lugens* prefer the lower half of rice plants where they feed and lay eggs, although in crowded

conditions hoppers will move to higher regions of the plant (Rubia and Heong, 1989). Females oviposit an average of 150 eggs, with a preoviposition period of 2-4 days. During her lifetime female *N.lugens* can mate twice. Males can copulate 24 hours after emergence, and up to 9 times in 24 hours (Gullan and Cranston, 1998). It is this high fecundity and ability to live in crowded environments that makes the *N.lugens* a serious pest of rice.

1.3.4 Crop Damage

Heavy plant hopper infestations can devastate rice crop yields, single season losses up to 38% have been recorded and attributed to *N.lugens* feeding (Lee and Park, 1977). Rice plants heavily infested with *N.lugens* show wilting and desiccation (described as 'hopperburn') followed by plant death. Rice plant death is due to the removal of photoassimilates by feeding *N.lugens* and reduction in photosynthesis rate, not the disruption of translocation as was originally thought (Watanabe and Kitagawa, 2000). In Southeast Asia, rice yield loss due to *N.lugens* feeding is a predicted 53.8 kg Ha^{-1}, equating to an annual loss of (US)$248.5 million (Tanaka, 1999). Also, after *N.lugens* feeding, the establishment of secondary infection is more likely, either by viral, bacterial or fungal pathogens. Sugar-rich honeydew secretions by feeding hoppers provide a nutrient source for microorganism proliferation and stylet wounding allows easy entry into the plant. Thus, endogenous resistance of the plant to opportunistic pathogens is impaired as a result of *N.lugens* feeding.

1.3.5 N.lugens *as a Virus Vector*

Although small populations of sap-sucking insects may not cause direct plant death, they are responsible for the transmission of viral diseases. Rice is host to 25 viruses that have a direct impact on production. These viruses are transmitted by either leafhoppers or planthoppers (Abo and Sy, 1998). Rice viruses that cause so many problems today rarely reached epidemic proportions under traditional cropping regimes; intensification has created a new threat to rice production. Transmission of plant viral disease is dependent on the vector, in this case, planthoppers. Therefore, there is a close link between vector levels and transmission rates (Hibino, 1996). Transmission rates are also extenuated by the ability of the vector (*N.lugens*) to fly long distances and migrate to distant crops. Long distance transoceanic migrations have been recorded up to distances of 2500 km (Holt *et al.*, 1996).

N.lugens is responsible for transmission of 2 economically important viruses, the Rice Grassy Stunt Virus (RGSV) and the Rice Ragged Stunt Virus (RRSV), (Hibino, 1996). Measures to control *N.lugens* populations will not only increase rice yield by decreasing pest feeding (a direct result), but also indirectly by reducing viral transmission.

1.3.5.1 Rice Grassy Stunt Virus (RGSV)

In the tropics, RGSV is endemic in areas where rice is planted all year round. *N.lugens* is the principal vector, although two other *Nilaparvata* spp., transmit the disease. The virus is propagative in the vector but is not transmitted via eggs. Infected plants are characteristically shorter and show a proliferation in short erect, narrow leaves, which are pale green or yellow in colour. Plants eventually die prematurely.

1.3.5.2 Rice Ragged Stunt Virus (RRSV)

Again, in the tropics, endemic *N.lugens* populations are associated with RRSV transmission and infection throughout the year. *N.lugens* is the principal vector, although one other *Nilaparvata* sp., may transmit the disease. RRSV is propagative in the vector but not transmitted in the eggs. Infected rice plants show stunting, have twisted leaves with serrated edges, and galls on the underside of the leaf blades. Infected plants have significantly reduced yields of rice.

1.4 Control of *N.lugens*

The characterisation of *Bacillus thuringiensis* toxin (*Bt*) has been a huge success story in recent insect control strategies. However, to date, no insecticidal *Bt* toxin has been isolated for *N.lugens*. Therefore, different approaches are needed for the development of *N.lugens* specific control agents.

1.4.1 Synthetic Chemical Control

Control of pests by chemical means has been practised for centuries, and insecticide development continues to be at the forefront of agrochemical research. Insecticides can be either novel plant derivatives, or entirely synthetic compounds. Synthetic insecticides include the carbamates, organophosphates and the pyrethroids (which are synthetic analogues of pyrethrin, an extract from chrysanthemum flowers). All of these synthetic insecticides exert their neurotoxic effects on the insect central nervous system (CNS), although the mode of action for each class of synthetic insecticide is subtly different. Organophosphates exert their neurotoxic effects by binding and phosphorylating acetylcholinesterase (AChE), resulting in the inhibition of this enzyme. This causes a

build up and over stimulation by the neurotransmitter acetylcholine (Gullan and Cranston, 1998). A similar mode of action underlies carbamate neurotoxicity; although post-synaptic hyperstimulation by acetylcholine is achieved by a different mechanism. Carbamate insecticides act by competing with acetylcholine in binding to the active site of acetylcholinesterase. As a result, acetylcholine is not degraded by acetylcholinesterase, and levels accumulate, again leading to increased firing of postsynaptic neurons (Gullan and Cranston, 1998). The most diverse group of insecticides are synthetic pyrethroids, which account for >30% of the world market for insecticides. Pyrethroids target insect voltage-gated sodium channels, where they act as sodium channel neurotoxins. As the sodium channel is the key structure of excitability in biological systems loss of function is obviously detrimental to insect survival (Zlotkin, 1999).

Control of *N.lugens* populations has been reliant on these synthetic chemicals since the introduction of HYVs of rice in the 1960s. However, resistance to synthetic insecticides is now common, rapid adaptation has been encouraged by insecticide overuse throughout long growing seasons (Pingali and Gerpacio, 1997). It is known that resistance to organophosphorous insecticides (and presumably carbamates) in *N.lugens* occurs by over expression of carboxylesterase, a detoxification gene that is present in susceptible strains (Small and Hemingway, 2000). This elevated-esterase based resistance mechanism has been documented in *N.lugens* (Chen and Sun, 1994), *Laodelphax striatellus* (small brown plant hopper) (Sakata and Miyata, 1994) and *Nephotettix cinticeps* (green rice leaf hopper) (Hasui and Ozaki, 1984). Resistance to

pyrethroid insecticides in *N.lugens* is well documented (Dai and Sun, 1984), and more recent studies have resolved this mechanism of resistance (Vontas *et al.*, 2001; Vontas *et al.*, 2002). As expected, due to the different mode of action, the resistance mechanism towards pyrethroid insecticides is different to that of organophosphate resistance. As well as neurotoxic effects, pyrethroids also induce severe oxidative stress and lipid peroxidation in insects (Vontas *et al.*, 2001). It has been proposed that elevated levels of glutathione S-transferases (GSTs) in pyrethroid resistant *N.lugens* protects tissues from oxidative damage, and thereby reduce insect mortality rates (Vontas *et al.*, 2001).

Chemical pesticides offer effective but far from ideal control of *N.lugens* populations; well-documented economic, agronomic, social and environmental problems exist (Pingali and Gerpacio, 1997). It is the improper use of these broad-spectrum insecticides that is responsible for the resurgence of the *N.lugens* population and promotion from a secondary to a primary pest of rice. As well as having detrimental effects towards beneficial predators and parasitoids, some insecticides are toxic towards mammals (Storm *et al.*, 2000). At the farm level, up to six applications of insecticide per growing season are not uncommon (Warburton *et al.*, 1995). However, it is questionable whether this prophylactic approach is necessary. Across Asia, a change in farming practice is being encouraged, moving from prophylactic insecticide application towards minimal usage of more environmentally benign insecticides from the pyrethroids family (Pingali and Gerpacio, 1997).

Due to the problems associated with chemical pesticides there is renewed interest in

plant-derived insecticidal compounds (section 1.4.2), many of which are insecticidal towards otherwise resistant strains of *N.lugens*.

1.4.2 Plant-Derived Insecticidal Chemicals

Novel derivatives from plants have been isolated and insecticidal characteristics assessed. A mass screening of 52 domestic plant species, 49 Indian plant species and 21 African plant species, revealed foliar extracts from *Ginkgo biloba* L. (Ginkgoaceae) as highly toxic towards *N.lugens* (Ahn *et al.*, 1997). The active chemicals were isolated and demonstrated to be ginkgolides A, B and C, and bilobalide. Further analysis revealed bilobalide not only to be more potent than commonly used carbamate insecticides, but also insecticidal to carbamate resistant *N.lugens* strains. Topical application of bilobalide causes insect tremors and paralysis; insects die within 30 minutes. Large-scale isolation and purification of naturally occurring insecticides from *G.biloba* could be useful as new control agents in the fight against *N.lugens* (Ahn *et al.*, 1997). Also, extracts from the neem tree, *Azadirachta indica*, are known for their insecticidal properties. The active ingredient azadirachtin (AZ), a limonoid, acts against a broad range of insect pests, including sap sucking Homopteran spp., and has been used successfully in field trials. Extracts from neem seed kernels and leaves are known to act as repellants, anti-feedants and growth disturbants (Gullan and Cranston, 1998).

1.4.3 Plant-Derived Insecticidal Genes

In order to facilitate and encourage screening of potential plant derived insecticidal proteins an *in vitro* assay system has been developed (Mitsuhashi, 1974). Artificial diet containing sugar, amino acids and minerals is enclosed in a membrane, from which

N.lugens feed. Potential insecticidal proteins can be incorporated into the diet to assess their activity. In this way, two protein groups with toxicity to *N.lugens* have been identified: lectins and oxidative enzymes. The snowdrop (*Galanthus nivalis*) agglutinin (GNA); wheat germ agglutinin (WGA) and the oxidative enzyme, soybean lipoxygenase (LPO), proved to be insecticidal in artificial diet feeding trials (Powell *et al.*, 1995). Since the initial characterisation of these plant-derived insecticidal proteins, GNA has been the subject of most research. GNA is known to be both an anti-feedant and an insecticide. When fed at 0.1% (w/v) levels, honeydew excreta is reduced by 96% for the first 24 hours, suggesting no artificial diet consumption is taking place. However, *N.lugens* decrease in survival is not entirely due to this anti-feedant effect. Eventual ingestion of artificial diet after 24 hours allows the toxic effect of GNA to manifest (Powell *et al.*, 1995). The mechanism of GNA toxicity towards *N.lugens* is poorly understood. However, immunohistochemical studies reveals that GNA ingestion causes a disruption in the midgut epithelial brush border cells. Also, post-ingestion detection of GNA in the *N.lugens* haemolymph means passage across the gut wall must occur (Powell *et al.*, 1998). GNA, and other lectins from species of the Amaryllidaceae family, are known to bind α-D-mannose, via α-1,3 or α-1,6 glycosidic bonds (Van Damme *et al.*, 1987). Using this knowledge, gut glycoproteins were purified through affinity to GNA. This revealed ferritin as the most abundant target in the *N.lugens* midgut (Du *et al.*, 2000). The low affinity association between GNA and ferritin makes ferritin an ideal transporter, possibly facilitating transport of GNA into the haemolymph. This could explain the post-ingestion systemic appearance of GNA. Because ferritin is crucial for iron homeostasis, direct binding of GNA may cause toxicity towards the insect by

disrupting iron homeostasis. However, in order to determine the precise mechanism of GNA toxicity further research is needed. More recently, the ability of GNA to enter the haemolymph has been exploited as a delivery system for insecticidal peptides. After oral delivery of GNA-peptide fusions, active peptides have been detected in the insect haemolymph, where they can exert their toxic effect (Fitches *et al.*, 2002).

Another approach towards generating increased resistance is the incorporation of naturally expressed plant defence proteinase inhibitors into the rice genome. Proteinase inhibitors are found throughout the plant kingdom and have been studied in Leguminosae, Gramineae and Solanaceae families. Transgenic rice, strongly expressing a soybean Kunitz trypsin inhibitor (SKTI), was toxic to *N.lugens* and lessened the damage caused by insect feeding (Lee *et al.*, 1999). The introduction of *SKTI* into rice plants could be used to control sap-sucking insect pests, although further field level feeding trials are necessary.

1.4.4 Selective Breeding and N.lugens *Resistance*

To date, breeding of rice cultivars that are resistant to *N.lugens* attack offers the most effective way of controlling pest populations, at least at the field level. Intensive research and development, primarily at the IRRI, has lead to the discovery of ten *N.lugens* resistance genes, five of which are dominant and the others recessive. Commercial rice varieties carrying *BPH1* and *BPH2* resistance genes have been bred and released since the 1970s. However, the major gene resistance offered by these cultivars has been overcome by the development of polygenic 'virulence' mechanisms in *N.lugens.* Three biotypes of *N.lugens*, designated 1, 2 and 3 have been described. The

population or individual that cannot infest any resistant rice variety is designated Biotype 1. Biotype 2 and 3 can attack rice varieties *BPH 1* and *BPH 2*, respectively (Tanaka, 1999).

1.4.5 Integrated Pest Management (IPM)

IPM is the amalgamation of field information and ability to make an informed response in relation to prevailing conditions. Within an IPM scheme, insecticides are considered destructive, and are not a fundamental component. Instead, IPM relies on the farmer's knowledge of, and confidence in, crop protection offered by naturally occurring enemies of *N.lugens*. Natural biological control, as opposed to the introduction of foreign *N.lugens* predators offers greater crop protection, due to the extensive co-evolution between *N.lugens* and their natural enemy fauna (Matteson, 2000). However, within such a programme substantial losses in crop yields must be tolerated, therefore the implementation of IPM programmes has been slow.

1.5 The Future of *N.lugens* Control-Novel Approaches

Although extensive research has been applied to various control strategies, *N.lugens* still remains a serious pest of rice. Over the past two decades, yield loss resulting from *N.lugens* feeding has increased (IRRI). Therefore, in order to achieve the 2020 projected rice yields of 840 million tonnes (an increase of 320 million tonnes) a new approach for controlling *N.lugens* must be addressed. My project aims to develop and assess a novel approach for design of insecticidal proteins, using the *N.lugens* and rice plant interaction as a model system. However, in theory such an approach could be applied to any sap sucking insect pest-plant association.

1.5.1 Insecticidal Proteins - The Plantibody Approach

The ability to produce transgenic plants expressing antibodies and antibody fragments is well documented (Peeters *et al.*, 2001). Complete antibodies or antibody fragments, such as Fab or single-chain variable fragments (scFv), have been expressed in plants where they are termed 'plantibodies'. To date, antibody and antibody fragments in plants have been targeted to the cytosol, the endoplasmic reticulum, and the nucleus (De Jaeger *et al.*, 2000). Expression levels of plantibodies are dependent on cellular location and species of plant, but levels up to 5% TSP (total soluble protein) are not uncommon (De Jaeger *et al.*, 2000). This thesis aims to identify 'target' proteins in the *N.lugens* gut and exploit antibodies to modulate their function. It is proposed that antibody mediated modulation of 'target' proteins will have a detrimental effect on insect survival (Fig. 1.1). In theory, an antibody that is detrimental to insect survival could be expressed in a plant and used to confer resistance. Antibody mediated modulation of the target can by accomplished by several mechanisms. The binding of an antibody to the active site of the target protein, preventing ligand binding (termed competitive inhibition, Fig. 1.1B); or, by structurally induced changes by the antibody binding the target protein, preventing substrate binding (termed allosteric inhibition, Fig. 1.1C). Also, direct binding of the substrate is possible, again preventing an interaction of the ligand with the target protein (Fig. 1.1D) (De Jaeger *et al.*, 2000). In 'proof of concept' style experiments this immunomodulation approach has already been used successfully to confer plant resistance to otherwise pathogenic viruses (Tavladoraki *et al.*, 1993; Voss *et al.*, 1995; Fecker *et al.*, 1997; Zimmermann *et al.*, 1998). Plant expression of antibodies towards specific viral coat proteins has proved efficient for inhibition of virus

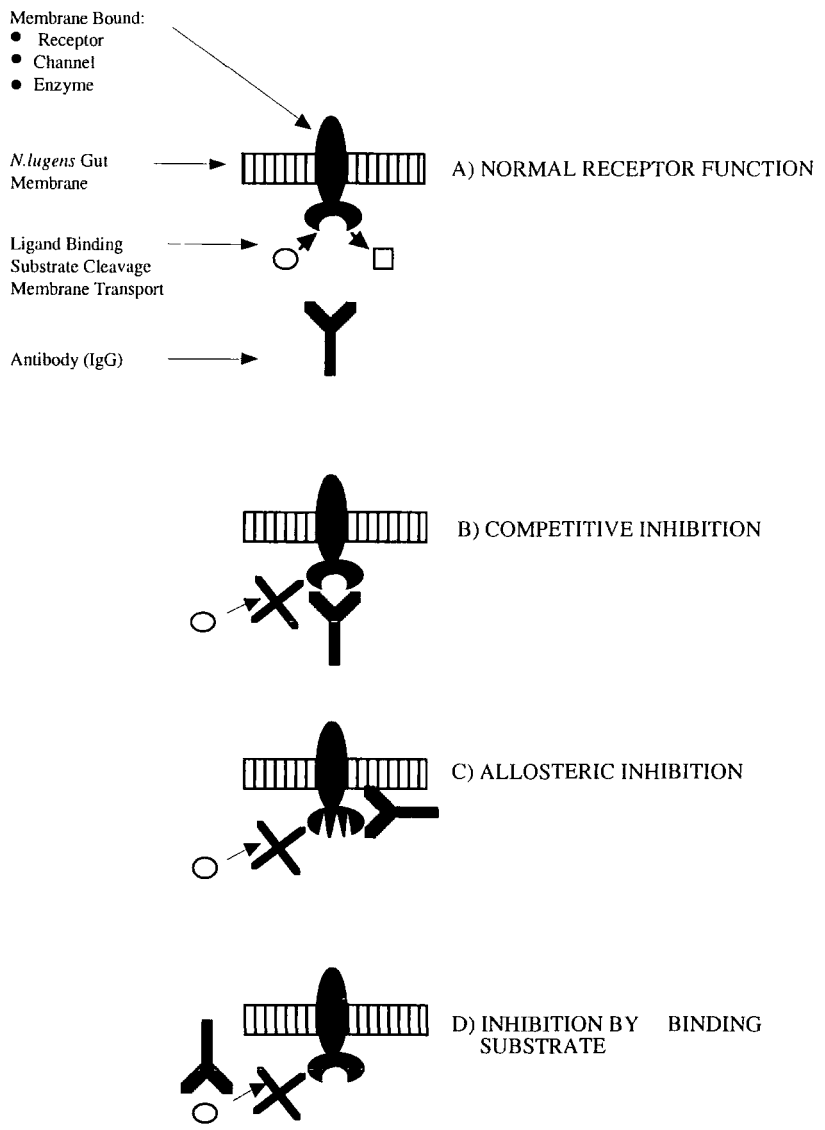


Figure 1.1. Mechanisms of antibody mediated modulation of protein 'targets' in the *N.lugens* gut lumen. It is proposed that antibodies (Y) can interfere with normal receptor activity (A) and prevent ligand binding by competing for receptor binding sites (B, competitive inhibition); or by binding the receptor and creating a conformational change again preventing ligand binding (C, allosteric inhibition); or through direct binging of the ligand, preventing an interaction with the receptor protein (D).

transmission. Admittedly, plantibody-mediated resistance against complex eukaryotic, multicellular pathogens, such as nematodes, fungi and insects, still remains a major challenge. However, resistance against root-knot nematodes, which is one of the world's most damaging agricultural pests, has been tested upon accumulation of the IgM antibody 6D4 (Baum *et al.*, 1996; Rosso *et al.*, 1996). IgM 6D4 binds a glycoprotein of unknown function in the stylets secretions of the root-knot nematode *Meloidogyne incognita*. Transgenic tobacco expressing intracellular 6D4 antibody-derived scFv fragments were produced (Rosso *et al.*, 1996), and nematode reproduction after plant infection was significantly reduced (De Jaeger *et al.*, 2000).

Antibody-mediated toxicity in an insect has been demonstrated in the haematophagous sheep blow fly, *Lucilla cuprina*. Larval development of this insect is inhibited when sheep are immunised with peritrophic membrane extracts from larval *L.cuprina* (East *et al* 1993). During larval development, ingested antibodies from the immunised vertebrate host remain functional in the insect gut and a small fraction cross the gut wall (by an unknown mechanism) and enter the haemolymph (Lehane, 1996); both enabling the antibodies to bind their antigen, and to interfere with insect metabolism. However, for rice plant 'immunisation' to be successful, and offer antibody-mediated protection against otherwise destructive phloem feeding insect species, antibodies must be ingested by phloem feeding insect pests. Towards the approach of phloem-specific protein expression, a promoter from the rice sucrose synthase gene (RSs1) (Rao *et al.*, 1998) has already been used to drive phloem-specific transgene expression. Therefore, this promoter may be useful in a plant genetic modification (GM) approach to direct

expression of insecticidal antibodies for ingestion by phloem feeding insect pests. This type of plant modification would offer highly specific resistance to phloem feeding insect plant pests, and remove the environmental problems associated with toxic broad-spectrum insecticides. Demands for improvements in rice yield to support the burden of a growing population is fuelling agronomic research in this area. However, the first step in this 'proof of concept' experiment is the characterisation of target proteins in the chosen pest species. Therefore, the remainder of this review will consider proteins which are likely to be present in *N.lugens* and which offer attractive 'targets' for the validation of antibody mediated insect control.

1.6 G-Protein Coupled Receptor (GPCR) Proteins

G-protein-coupled receptors (GPCR) are seven-transmembrane proteins (7-TM) that transduce extracellular signals into cellular physiological responses through the activation of heterotrimeric G proteins (Section 1.8). The chemical diversity among endogenous ligands is exceptional. They include biogenic amines, peptides, glycoproteins, lipids, nucleotides, ions, and proteases. Moreover, the sensation of exogenous stimuli such as light, odour and taste, is also mediated via this class of receptor (Kolakowski, 1994; Hoon *et al.*, 1999). The first GPCR was described about 20 years ago (Dixon *et al.*, 1986), and now more than 1000 GPCRs have been described in higher animals (Kolakowski, 1994). This superfamily of proteins constitutes the largest, most ubiquitous and most versatile family of membrane receptors, and have been the subject of several excellent reviews (Probst *et al.*, 1992; Vaughan, 1998; Gether, 2000; Pierce *et al.*, 2002). The recent sequencing of the *C.elegans* genome, as well as genomes from the insects *Drosophila* and *Anopheles* has fuelled research into invertebrate cell

signalling. It is now apparent that invertebrates also possess a vast number of GPCRs. For example, it is predicted that the *C.elegans* genome encodes >1000 GPCRs (Bargmann, 1998); while the *Drosophila* genome is predicted to encode approx. 160 GPCRs (Hewes and Taghert, 2001). Due to the importance of GPCRs in the regulation of numerous physiological events it is of no surprise that drugs which target these receptors (either directly or indirectly) account for most of the medicines sold worldwide (Pierce *et al.*, 2002). Therefore, it can be anticipated that a thorough understanding of future discoveries in the field of insect cell signalling will yield 'targets' for the design of highly effective and specific insecticides.

1.6.1 GPCR Structural Classification

GPCRs do not share any significant overall sequence homology (Probst *et al.*, 1992; Kolakowski, 1994). The defining feature common to all GPCR proteins is the presence of seven membrane-spanning α-helices connected by alternating intracellular and extracellular loops, with an extracellular N-terminus, and a intracellular C-terminal region. The helices contain stretches of 20-30 hydrophobic amino acids and are referred to as transmembrane domains 1-7 (TM1-TM7). This structure can be predicted by hydrophobicity analysis, making identification of proteins from this family straightforward. Significant sequence homology is found, however, within several subfamilies. The four major families include GPCRs related to the rhodopsin and the β_2–adrenergic receptor (family A), the receptors related to the glucagon receptors (family B), receptors related to the metabotropic neurotransmitter receptors (family C), and finally the atypical 7-TM proteins. The only residues conserved across all classes of GPCR are single cysteine residues situated on extracellular loops 2 and 3 (Probst *et al.*,

1992; Gether, 2000). Due to the huge diversity of GPCRs, the remainder of this introduction will only consider insect peptide GPCRs. However, there are numerous reviews considering aspects of GPCR classification (Probst *et al.*, 1992; Kolakowski, 1994; Brody and Cravchik, 2000; Broeck, 2001; Hewes and Taghert, 2001), and receptor signalling (Vaughan, 1998; Gether, 2000; Pierce *et al.*, 2002; Karnik *et al.*, 2003).

1.7 Drosophila Genome Sequence and Peptide GPCRs

Genes encoding putative GPCRs constitute an important fraction of all protein-encoding genes present in metazoan genomes (estimated at 6% in *C.elegans*, 1% in *Drosophila* and 2% in mammals) (Broeck, 2001). The *Drosophila* genome contains approx. 160 genes coding for G-protein coupled receptors (GPCRs), including neurotransmitter and hormone receptors, and olfactory and putative taste receptors (Adams *et al.*, 2000; Clyne *et al.*, 2000; Rubin *et al.*, 2000). Based on sequence similarity to well-characterised higher animal GPCRs the *Drosophila* receptor proteins are classified into 4 families: rhodopsin like (family A); glucagon like (family B); metabotropic neurotransmitter receptors (family C); and atypical 7-TM proteins. This classification is based on primary and secondary structure predictions. From these 160 GPCR genes in *Drosophila* only 44 are predicted to encode peptide GPCRs, which is likely to represent the vast majority, if not all peptide GPCRs in *Drosophila* (Brody and Cravchik, 2000; Hewes and Taghert, 2001). Together, the set of known and predicted *Drosophila* peptide GPCRs contains representatives of at least 15 monophyletic vertebrate GPCR subgroups (Hewes and Taghert, 2001). Although many putative GPCRs have been identified from genome sequence data, a large proportion have unidentified bioactive ligands. These so called

'orphan receptors' are subject to much research in order to identify their natural ligands (Milligan, 2002).

1.7.1 Drosophila Peptide GPCRs – Family A

Drosophila family A contains the largest number of peptide GPCRs, this family is predicted to have 40 members. Also, in higher mammals family A is also by far the largest group, and most widely studied. The overall homology in all family A receptors is low and restricted to a number of key residues. The high degree of conservation among these residues suggests that they have an essential role for either structural or functional integrity of the receptor. The only residue that is well conserved among all family A receptors is the arginine in the Asp-Arg-Tyr (DRY) motif at the cytoplasmic side of TM3 (Probst *et al.*, 1992; Kolakowski, 1994). Many *Drosophila* genes coding for putative family-A related GPCRs show sequence similarity to characterised mammalian peptide receptors. These include insect orphan GPCRs that show sequence similarity to vertebrate receptors for gonadotropin releasing hormone (CG10698), vasopressin (CG6111), somatostatin and opioids (CG7285, CG13702), galanin (CG10001), neuropeptide Y and prolactin releasing peptide (CG1147, CG7395/CG18639), bombesin/neuromedin B (CG18192, CG14593), growth hormone secretagogue and neuromendin U (CG9918, CG8795, CG8784), thyrotropin releasing hormone (CG16726, CG5911) and gastrin/cholecystokinin (CG6894, CG6857, CG6861).

1.7.2 Drosophila Peptide GPCRs – Family B

Except for the disulphide bridge connecting extracellular loop regions 2 and 3, family B receptors do not contain any of the structural features characterising family A receptors. Notably, the DRY motif, which is well conserved in family A GPCRs, is absent from family B GPCRs. The most prominent feature of family B receptors is a large extracellular amino terminus containing several cysteine residues, presumably forming a network of disulphide bridges (Ulrich *et al.*, 1998). In higher animals, family B receptors include approximately 20 different receptors for a variety of peptide hormones and neuropeptides. Higher animal family-B GPCRs include receptors for peptide hormones, such as CRF, calcitonin, glucagons, parathyroid hormone, secretin, vasoactive intestinal peptide, pituitary adenylate cyclase activating peptide, gastric inhibitory peptide, and growth hormone releasing hormone. However, the *Drosophila* genome contains only 5 distinct gene sequences encoding proteins that are very similar to vertebrate B-family peptide receptors, especially to the CRF and calcitonin receptors. These are gene products CG13758, CG12370, CG8422, CG17415 and CG4395. All family B receptors seem to couple to activation of the effector adenylyl cyclase through the G protein G_s (Section 1.8).

1.8 GPCR Receptor Signalling

As previously described, a diverse range of ligands can bind and activate GPCRs (section 1.6). Heterotrimeric G proteins transduce ligand binding to these receptors into intracellular responses, which underlie cellular physiological responses. Ligand binding is thought to stabilise an active receptor conformation that selectively interacts with a heterotrimeric G protein, consisting of -α, -β and -γ subunits, to form a complex of

agonist, receptor, and G-protein. This promotes the exchange of GDP for GTP on the Gα-subunit, which results in a dissociation of Gα-GTP and Gβγ from the receptor. These subunits (Gα and/or Gβγ) then regulate the activity of certain cellular effector proteins, such as enzymes (e.g., adenyly cyclase, phospholipase C beta, cGMP phosphodiesterase, phospholipase A_2) or ion channels, until they reassociate upon hydrolysis of GTP to GDP caused by the intrinsic GTPase activity of Gα. Four main classes of G proteins can be distinguished: G_s, which activates adenyl cyclase; G_i, which inhibits adenylyl cyclase; and G_q, which activates phospholipase C; and G_{12} and G_{13}, of unknown function (Hamm, 1998)

1.9 Neurohormonal and Modulatory Actions of Insect Peptides

In insects neuropeptides can act as either neurohormones or neuromodulators, but there is no evidence for neurotransmitter function. By definition, an insect neurohormone is released into the circulating haemolymph and acts after transport over considerable distances; whereas a neurotransmitter is released at a neuronal synapse and produces a rapid and transient response by acting directly on ligand-gated ion channels. As stated previously no insect neuropeptide has been assigned a neurotransmitter function. However, recently, a peptide-gated ion channel from snail *Helix* provides evidence that neuropeptides can act as fast neurotransmitters (Cottrell, 1997), although similar peptide activated channels have not been identified in insects. A neuromodulator can be defined as a peptide that is either released synaptically of non-synaptically, however they cannot act directly on gating of ion channels. These peptides are often co-released with classical neurotransmitters, and produce a variety of effects, commonly by interacting with G protein-coupled receptors. The modulator actions can be presynaptic or postsynaptic and

serve to prolong or shorten the effects of co-released classical neurotransmitter (Nassel, 2002). The hormones of the insect body are produced at neuronal, neuroglandular and glandular centres, these hormone production sites will be described below (section 1.9.1-1.9.5).

1.9.1 Neurosecretory Cells

Neurosecretory cells (NSC), also called neuroendocrine cells, are modified neurons found throughout the nervous system (within the CNS, peripheral nervous system, and the stomodeal nervous system), but the majority of NSC are found in major groups in the brain. These cells produce most of the known insect hormones, the notable exceptions being the production of ecdysteroids and juvenile hormones by non-neural tissues. However, the synthesis and release of the latter hormones are regulated by neurohormones from NSC.

1.9.2 Corpora Cardiaca

The corpora cardiaca are a pair of neuroglandular bodies located on either side of the aorta and behind the brain. They store and release neurohormones, including prothoracicotropic hormone (PTTH), originating from the NSC of the brain, as well as producing their own neurohormones. PTTH stimulates the secretory activity of the prothoracic glands.

1.9.3 Prothoracic Glands

The prothoracic glands are diffuse paired glands generally located in the thorax or the back of the head. The prothoracic glands secrete an ecdysteroid, which, after hydroxylation, elicits the moulting process of the epidermis.

1.9.4 Corpora allata

The corpora allata are small, discrete, paired glandular bodies derived from the epithelium and located on either side of the foregut. Their function is to secrete juvenile hormone, which has regulatory roles in both metamorphosis and reproduction.

1.9.5 Midgut Endocrine Cells

The endocrine cells of insect the midgut are of great importance, as implied by their widespread occurrence in the majority of representative insect orders, but their physiological roles are very poorly understood. Although the majority of neurohormones circulating the haemolymph are produced in the CNS several are expressed in endocrine cells of the insect gut (Zitnan *et al.*, 1993; Yu *et al.*, 1995; Sehnal and Zitnan, 1996).

1.10 Diversity of Insect Neuropeptides and Receptors

Genome sequencing has revealed that even simple nervous systems of nematodes produce huge variety of different neuropeptides (Bargmann and Kaplan, 1998; Brownlee and Fairweather, 1999). For example, in the nematode *Caenorhabditis elegans* more than 40 genes encoding peptide precursors have been identified, each of which produce numerous neuropeptide isoforms (Bargmann and Kaplan, 1998; Brownlee and Fairweather, 1999). Sequence analysis of the *Drosophila* genome reveals at least 30 genes encoding neuropeptide precursors, although this is likely to be an underestimate (Hetru *et al.*, 1991; Adams *et al.*, 2000; Hewes and Taghert, 2001; Nassel, 2002), and about 44 putative peptide G protein-coupled receptors proteins have

been identified (Brody and Cravchik, 2000; Hewes and Taghert, 2001). It is therefore apparent that neuropeptide signalling is complex, even in invertebrates.

As can be expected the functional roles of insect neuropeptides are diverse, encompassing the regulation of homeostasis, organisation of behaviours, initiation and coordination of developmental processes and modulation of neuronal and muscular activity, all of which are crucial for insect survival. The neuropeptides that coordinate these actions can be grouped into 23 structurally distinct neuropeptide families. Although, the exact role of many of the insect neuropeptides remains to be elucidated, it is clear that distribution patterns vary extensively between types of neuropeptides. Distinct neuropeptides are present in specific sets of neurons and/or neurosecreatory cells and in some cases in cells of the insect midgut. To add to the complexity of insect endocrinology many insect neuropeptides appear to be multifunctional; a neuropeptide may act both in the CNS and as a circulating hormone and play different functional roles at central and peripheral targets (Nassel, 2002). Furthermore, the diversity of insect peptide hormones can increased post-translationally. After translation, neuropeptide precursors enter the secretory pathway, where bioactive neuropeptides are generated via enzymatic cleavage of larger neuropeptide precursors. Often the peptide precursors contain multiple neuropeptides, each flanked by dibasic cleavage sites (or sometimes other cleavage sites) that direct enzyme cleavage and liberation of the peptides. After cleavage the peptides can undergo post-translational modifications, of which carboxy terminus amidation is the most common (Kolhekar *et al.*, 1997). Thus, insect

neuropeptide chemical diversity is amplified through processes of enzymatic cleavage of precursors and post-translational modification.

Although *Drosophila* is an excellent model insect for genetic analysis, it is less suited to physiological experiments, solely due to the small size of the insect. Therefore, other more suitable insects have been used for characterising biologically active proteins and peptides, and for studying insect physiology. Insects amenable to physiological experiments include the cockroach *Periplaneta americana*, the locust *Locusta migratoria* and the moths *Manduca sexta* and *Bombyx mori*. As a consequence, many neuropeptides were first isolated from these species. Twenty-nine years ago, the first insect neuropeptide was purified to homogeneity from 125 kg of *Periplaneta americana* to obtain 180 μg of pure peptide (Starratt and Brown, 1975). Since this initial discovery numerous insect neuropeptides have been identified by genome sequencing and classical peptide purification, however, the exact biological role of the majority still needs to be elucidated.

1.10.1 Proctolin Receptor

The *Drosophila* GPCR cDNA corresponding to gene product CG6986 has been cloned and expressed, and the natural bioactive ligand identified (Egerod *et al.*, 2003). When the receptor was expressed in a functional form in animal cells only the insect neuropeptide proctolin (RYLPT) activated the receptor protein. Proctolin was isolated from cockroaches in 1975, and was the first invertebrate neuropeptide to be isolated and fully sequenced (Starratt and Brown, 1975). Proctolin has since been isolated from a variety of insects and crustaceans, and the RYLPT structure is conserved in all cases. In

both insect and crustacean proctolin has myo- and neurostimulatory effects. Because proctolin has been known for such a long time, and its structure is well conserved within insects, a large range of proctolin peptide and nonpeptide mimetics have been synthesised with the aim of developing insecticides (Konopinska, 1997; Noronha *et al.*, 1997; Plant *et al.*, 2001; Scherkenbeck *et al.*, 2002), however, this approach remains unsuccessful.

1.10.2 Pyrokinin Receptor

The first insect pyrokinin receptors were first isolated from *Drosophila* (Rosenkilde *et al.*, 2003), (gene products CG8784 and CG8795). Gene silencing, using RNA-mediated interference (RNAi), showed that CG8784 gene silencing caused lethality in embryos, whereas CG8795 gene silencing resulted in strongly reduced viability for both embryos and first instar larvae (Rosenkilde *et al.*, 2003). It was further demonstrated that both receptors (CG8784 and CG8795) were activated by low concentrations of the *Drosophila* neuropeptide pyrokinin-2 (SVPFKPRLamide). However, the exact biological role of pyrokinin-2 in *Drosophila* is unknown. In other insects, pyrokinins have diverse myotropic actions and are also important for initiating sex pheromone biosynthesis (pheromone biosynthesis activating neuropeptide, PBAN) (Kitamura *et al.*, 1989; Raina *et al.*, 1989; Rafaeli, 2002), and embryonic diapause (Imai *et al.*, 1991). Entomologists have paid much attention to PBAN, which belongs to structurally related pyrokinins. PBAN has been isolated from a variety of moth species (Lepidoptera). Many attempts have been made to characterise PBAN receptors and to synthesise PBAN mimetics, in order to interfere with mating and other physiological processes in lepidopterans (Altstein, 2001; Teal and Nachman, 2002). Recently the first PBAN

receptors have been cloned in the moth *Helicoverpa zea* (Choi *et al.*, 2003), with a known function of stimulating pheromone biosynthesis in females. Identification of such an important receptor will almost certainly create renewed interest in the synthesis of PBAN receptor agonists and antagonists with the aim of disrupting reproduction in moth pest species.

1.10.3 Ecdysis Triggering Hormone Receptors

Insects and other arthropods have an external skeleton (cuticle) that they need to exchange during growth or metamorphosis. This shedding of the old cuticle is called moulting or ecdysis. Insect ecdysis is a hormonally programmed sequence of events that enables insects to escape their old cuticle at the end of each developmental stage (Truman *et al.*, 1996). The immediate events leading to ecdysis, which are initiated upon release of ecdysis-triggering hormone (ETH) into the bloodstream, include respiratory inflation and sequential stereotypic behaviours that facilitate shedding of the cuticle. *Drosophila* mutants carrying the *eth* gene knockout showed lethal ecdysis events at the end of the first larval instar (Park *et al.*, 2002). However, lethality was reversed by injection of ETH peptides (Park *et al.*, 2002). The experiments of Park *et al.*, (2002) demonstrate that the ETH signalling system is necessary and sufficient for insects to survive the earliest ecdysis. Recently, the *eth* receptor has been identified in *Drosophila* (CG5911), this gene encodes two functionally distinct subtypes of G protein-coupled receptors through alternative splicing (CG5911a and CG5911b) (Iversen *et al.*, 2002; Park *et al.*, 2003). Identification of the *Drosophila eth* receptor should facilitate the identification of similar receptor proteins in other arthropods.

1.10.4 Neuropeptide F receptor

The first insect neuropeptide F receptors have been isolated from *Drosophila* (Feng *et al.*, 2003) (gene products NPFR76F). Although the biological function remains unknown they have been implicated in a signalling system that modulates food response and social behaviour. Maximum receptor activation is achieved while challenging functionally expressed receptor with *Drosophila* neuropeptide F. *Drosophila* neuropeptide F (DmNPF) is a 36 amino acid peptide found in midgut endocrine cells and the CNS of larvae and adults (Brown *et al.*, 1999). DmNPF is structurally related to vertebrate regulatory peptides of the neuropeptide Y (NPY) family, which includes pancreatic polypeptide (PP) and peptide YY (PYY). The NPY superfamily also includes NPF's from molluscs (Hazelwood, 1993) and platyhelminths (Halton *et al.*, 1994). In vertebrates, functions for NPY, PP, and PYY include regulation of food intake (Gerald *et al.*, 1996), circadian rhythms (Hazelwood, 1993), gut enzyme secretion (Hu *et al.*, 1996) and motility (Wettstein *et al.*, 1995). Insects possess the components typical of vertebrate NPY family regulatory systems, including peptides, receptors, and actions related to feeding and digestion. For *Drosophila, in situ* hybridisation indicates DmNPF RNA occurs in midgut endocrine cells and the brain (Garczynski *et al.*, 2002). Although such positioning suggests this peptide may regulate feeding behaviour and digestion, no biological functions have yet been established (Brown *et al.*, 1999).

1.10.5 Allatostatin Receptor (ALSTR)

Allatostatins (ALSTs) are a family of peptides first isolated from the cockroach *Diploptera punctata* (Pratt *et al.*, 1989; Woodhead *et al.*, 1989), but have since been shown to be distributed across many insect orders (Bendena *et al.*, 1997). Their best

documented role is the rapid and reversible inhibition of juvenile hormone (JH) production by the endocrine glands of the corpora allata (CA). JH inhibits the expression of adult features so that a high haemolymph titre of JH is associated with a larval-larval moult, a lower titre with a larval-pupal moult and absence at a pupal-adult moult. The multi-functional allatostatins are referred to as the brain-gut peptides because in addition to genuine allatostatic activity (the inhibition of JH biosynthesis), they can block muscle contraction in the gut of insects. This dual action seems common for all ALSTs across many representative insect orders (Bendena *et al.*, 1997). Also, an ALST from *Diploptera punctata* has been shown to stimulate invertase and α-amylase activity in the midgut, possibly improving the uptake of nutrients (Fuse *et al.*, 1999). It appears that ALSTs are diverse in function, however, to date the inhibition of JH has been characterised most thoroughly (Nassel, 2002).

Three families of allatostatin exist in insects, allatostatin-A (ALST-A; cockroach type), allatostatin-B (ALST-B; cricket type), and allatostatin-C (ALST-C; *M.sexta* type). In *Drosophila* three genes that encode ALST-A (Lenz *et al.*, 2000a; Lenz *et al.*, 2000b; Lenz *et al.*, 2000c), ALST-B (Williamson *et al.*, 2001b) and ALST-C (Williamson *et al.*, 2001a) have been identified. This suggests that three types of ALST are likely to be present in across insect orders. The allatostatins are short peptides (between 8 and 15 amino acids in *Drosophila*) and have well conserved C-terminal regions, Family A Tyr/Phe-X-Phe-Gly-Leu-NH_2, Family B Trp-$(x)_6$-Trp-NH_2, Family C Glu-Val-Arg-Tyr/Phe-Arg-Gln-Cys-Tyr-Phe-Asn-Pro-Ile-Ser-Cys-Phe-Arg. Apparently, only one type of ALST inhibits juvenile hormone biosynthesis (is allatostatic) in a given insect

species and this type varies between insects. For example, the A-type ALSTs are allatostatic in cockroaches, but not in flies (Bendena *et al.*, 1997). The B-type ALSTs are only known to act as allatostatins in crickets (Lorenz *et al.*, 1995), whereas the C-type are only known as genuine ALSTs in moths (Kramer *et al.*, 1991).

The receptors for allatostatins long remained unknown, but recently a receptor from *Drosophila* was cloned (DAR-1) and shown to be intrinsic for *Drosophila* A-type ALST (Birgul *et al.*, 1999). Shortly afterwards a second putative ALSTR (DAR-2) from *Drosophila* was cloned (Lenz *et al.*, 2000a), and has since been functionally characterised. DAR-2 is expressed in the gut and is activated by *Drosophila* A-type ALSTs (Lenz *et al.*, 2001). In addition to ALSTRs from *Drosophila* similar receptors have been cloned from *Periplanta* americana (AF336364) and *Bombyx mori* (BAR, AF254742) (Secher *et al.*, 2001). In both of these cases ALSTR are cognate receptors for A-type ALSTs, and expression of the receptor has been demonstrated in gut tissue. The *B.mori* receptor (BAR) is predominantly expressed in the midgut but also in the brain and other parts of the intestine (Secher *et al.*, 2001). The receptors for type-B and type-C ALSTs remain unidentified at present.

1.11 Water Regulation In Insects-What Is Known?

Ionoregulation, osmoregulation and excretion in insects involves the secretion of primary urine by the Malpighian tubules, followed by selective reabsorbtion of solutes and water in the hindgut. Homeostasis in this excretion system is maintained by the action of two antagonistic hormone groups. Diuretic peptides (DPs) control the secretion of primary urine by the Malpighian tubules and the anti-diuretic hormone (ADH)

controls the reabsorbtion of water and solutes from the primary urine present in the anterior (ileum) and posterior (rectum) hindgut segments. DPs can be broadly divided into two groups: those with varying degrees of homology to the vertebrate corticotropin releasing factor (CRF), and the smaller kinins (O'Donnell and Spring, 2000).

1.11.1 Gut Physiology and Adaptation For Water Regulation

The Malpighian tubules have a role analogous to that of the glomerulus of the vertebrate nephron. However, insects produce primary urine by osmotic filtration, as opposed to ultrafiltration in the vertebrate system. Osmotic filtration is driven by the active transport of KCl and/or NaCl into the tubule lumen and low molecular weight solutes follow by passive diffusion. The primary urine then enters the hindgut and mixes with material entering from the midgut. Before excretion, water, ions and essential metabolites are scavenged in the ileum and rectum, which is functionally analogous to the tubular segments of the vertebrate nephron. DPs are released in response to feeding or at the time of emergence (eclosion) from the pupal stage, and stimulate primary urine production. Feeding is also thought to be a stimulus for the release of ADH, which triggers reabsorbtion in the hindgut (Coast, 1996). Osmotic homeostasis is maintained by the tight control of fluid secretion by the Malpighian tubules and subsequent reabsorbtion by the hindgut; both mechanisms are under neuroendocrine control. High levels of fluid recycling between the Malpighian tubules, hindgut and haemolymph allow clearance of toxic wastes without causing excessive water loss (O'Donnell and Spring, 2000).

Phloem feeding insects in the order Homoptera face a crisis while ingesting large volumes of hyper-osmotic plant fluids, one of water homeostasis. Therefore many Homopteran phloem feeders have evolved an adaptation known as a filter chamber that allows removal of superabundant food molecules. Typically the anterior and posterior part of the gut are in intimate contact allowing passage of excess small molecules, such as simple sugars which are then excreted as a sugar rich honeydew. This short-circuiting eliminates the absorptive portion of the midgut, thus freeing the digestive region of excess water (xylem feeders) or superabundant food molecules (phloem feeders). In phloem feeders the haemolymph is maintained at an osmotic pressure of about 300mOsm/kg, which is lower than that of osmotic pressure of ingested phloem (450mOsm/kg). Without osmo-regulatory control the insect would loose water to the gut contents, this however is not the case (Hamm, 1998). Through the processes of short-circuiting through the filter chamber, as well as assimilation and enzymatic transformation (disaccharide to oligosaccharides) phloem feeding insects are able to reduce the osmotic pressure of ingested fluids so they are iso-osmotic with their haemolymph (Ashford *et al.*, 2000).

1.11.2 The CRF-like Diuretic Peptides (CRF-DPs)

Insect CRF-like diuretic peptides (CRF-like DPs) show greatest structural similarity to the vertebrate corticotropin releasing factor/urotensin/sauvagine family of peptides. The first insect CRF-like DP was isolated from head extracts of *Manduca sexta*, Mas-DP1 (Kataoka *et al.*, 1989). This 41-residue *C*-terminally amidated peptide shares 29-35% sequence identity with the CRF family of vertebrate peptides. *M.sexta* was also the first species found to possess two CRF-related DPs, Mas-DP1 and Mas-DP2. Mas-DP2 is 30

amino acids long and only has 33% sequence identity with Mas-DP1 (Blackburn *et al.*, 1991). Since the initial discovery of Mas-DP1 many more insect CRF-like DPs have been isolated and sequenced (Table 1.1 and Fig. 1.2).

Insect CRF-like DPs range in size from 30-46 amino acids and have sequence homologies varying between 30 and 63% (Fig. 1.2). Despite the relatively low sequence homology between CRF-like DPs the mode of action is thought to be similar, and high levels of interspecific CRF-DP bioactivity has been demonstrated. It has been shown that *M.sexta* CRF-like DP, Mas-DP1 and related peptides from *A.domesticus, L.migratoria* and *P.americana* all cause maximal increases in the production of cAMP by the MTs of both larval and adult *M.sexta* (Audsley *et al.*, 1995). It is known that the N-terminal region of CRF peptides is involved with signal transduction and receptor binding, and not surprisingly it is this region which is best conserved and possibly responsible for the cross reactivity observed. Thus, in Acd–DP deletion of the first six amino acids does not alter function, but deletion of the next five causes a substantial loss in activity (Coast and Kay, 1994). The majority of diuretic hormones are peptides, but so far there is one exception found in the blood sucking bug *Rhodnius prolixus*, the biogenic amine serotonin (5-hydroxtryptamine; 5-HT) has a hormonal function and can stimulate Malpighian tubule secretion, but a CRF-like DP is still required for maximal fluid secretion (O'Donnell and Spring, 2000).

1.11.3 The CRF-Like Diuretic Peptide Receptor (DPRs)

To date three insect CRF-like DPRs have been identified and cloned, the first from *M.sexta* (tobacco hornworm) (Reagan, 1994) and now from cricket, *A.domesticus*

Insect Species	Accession No	Name	Size (amino acids)	Reference
Manduca sexta	1506520A	Mas-DP1	41	(Kataoka *et al.*, 1989)
Manduca sexta	P24858	Mas-DP2	30	(Blackburn *et al.*, 1991)
Locusta migratoria	P23465	Lom-DP	46	(Kay *et al.*, 1991b)
Periplanta americana	P41538	Pea-DP	46	(Kay *et al.*, 1992)
Acheta domesticus	P23834	Acd-DP	46	(Kay *et al.*, 1991a)
Diploptera punctata	P82373	Dip-DP2	46	(Furuya *et al.*, 2000)
Musca domesticus	P41537	Mus-DP1	44	(Clottens, 1994)

Table 1.1. A selection of isolated and sequenced CRF-like diuretic peptides from insect species. Amino acid sequences of the bioactive hormones are aligned in Fig. 1.2.

```
P23834    T G A - Q S L S I V A P L D V L R Q R L M N E L N R R R M R E L Q G S R I Q Q N R Q L L T S I    46
P41538    T G S G P S L S I V N P L D V L R Q R L L L E I A R R R M R Q S Q - D Q I Q A N R E I L Q T I    46
P41537    N K - - P S L S I V N P L D V L R Q R L L L E I A R R Q M K E N T - R Q V E L N R A I L K N V    44
P23465    M G M G P S L S I V N P M D V L R Q R L L L E I A R R R L R D A E - E Q I K A N K D F L Q Q I    46
P82373    T G T G P S L S I V N P L D V L R Q R L L L E I A R R R M R Q T Q - N M I Q A N R D F L E S I    46
P24858    - - - - - S F S V N P A V D I L Q H R Y M E K V A - - - - - - - - - - - - - Q N N R N F L N R V    30
1506520A  R M - - P S L S I D L P M S V L R Q K L S L E - - - - K E R K V H A L R A A A N R N F L N D I    41
```

Figure 1.2. Amino acid sequence alignment (Clustal V) of CRF-DPs from *Periplanta americana* (P41538), *Acheta domesticus* (P23834), *Diploptera punctata* (P82373), *Manduca sexta* (1506520A and P24858), *Locusta migratoria* (P23465) and *Musca domesticus* (P41537).

(Reagan, 1996) and silkworm, *Bombyx mori* (Ha *et al.*, 2000). These receptors are members of the calcitonin/secretin receptor family and bind CRF-like DPs with high affinity. In *Drosophila* 5 proteins belong to Family B GPCRs (glucagon like receptors), from which genes CG8422 and CG12370 are most similar to CRF-like DPRs isolated from other insect species (Hewes and Taghert, 2001). Clustal V amino acid alignment of related CRF-like DPRs show highly conserved blocks in and around membrane spanning regions (Fig. 1.3). The *A.domesticus* CRF-like DPR protein is 441 amino acids long, with a molecular weight of 49.5 kDa, hydrophobicity analysis suggests there are seven membrane-spanning regions. The large N-terminal region has homology to known hormone receptor domains (E-value = 61.2; P = 3e-11). It is this region that interacts with haemolymph borne CRF-like DPs, which in turn triggers intracellular signalling which ultimately results in water secretion. To date several insect CRF-like DPRs have been functionally expressed in animal cells, allowing binding properties of corresponding ligands to be determined. It has been shown that CRF-like DPs are active at nM concentrations and capable of eliciting an intracellular increase of cAMP (Reagan, 1994; Reagan, 1995; Reagan, 1996). In *Aedes* intracellular cAMP increases the Na^+ conductance across the primary cells of the Malpighian tubules, thereby increasing transport through the apical Na^+/H^+ anti-porter. Other low molecular weight solutes follow passively with water (Beyenbach, 1995).

1.11.4 The Kinins

The members of the other major group of DPs are the kinins, first discovered due to their myotropic activity, a spontaneous contraction of the insect hindgut. The first kinin was identified from the cockroach *Leucophaea maderae*, leucokinin (Meola *et al.*,

Figure 1.3. Amino acid sequence alignment (Clustal V) of diuretic hormone receptors from cricket, *A.domesticus* (Q16983); *Drosophila* (CG8422 and CG12370), *Manduca sexta* (P35464) and *Bombyx mori* (BOMBYX, sequence is not deposited in NCBI database, amino acid represents published sequence (Ha *et al.*, 2000). Highlighted regions indicate the transmembrane regions of the *Manduca sexta* diuretic hormone receptor (P35464).

1994). Since the initial discovery the myotropic bioassay has been used to isolate and characterise 23 other insect kinins. This includes eight from cockroach *L.maderae*, five from the cricket *A.domesticus*, one from the locust *L.migratoria*, three from the mosquito *Culex salinarius*, three from the mosquito *Aedes agepti*, and three from the earworm *Helicoverpa zea* (Table 1.2). The kinins are characteristically small peptides, ranging from 6-15 amino acids and have a highly conserved C-terminal domain consisting of a pentapeptide sequence of Phe-X^1-X^2-Trp-Gly-NH_2, where X^1 is Phe, Tyr, Ser, His, or Asn, and X^2 can be Pro, Ser, or Ala. Kinins act as potent stimulants of tubule secretion with EC_{50} values in the 0.3-0.02nM range, at physiological concentrations they can initiate a tubule secretion response 30-75% of that produced by a CRF-DP from the same insect (Coast *et al.*, 1990). The kinin group of diuretic peptides have been shown to stimulate fluid secretion in isolated Malpighian tubules, but in contrast with CRF-DPs this takes place through a cAMP independent response. Instead the cellular response is mediated through an increase in intracellular Ca^{2+}. Progressive truncation of the kinin peptide has revealed that only the conserved pentapeptide sequence in necessary for full biological activity on cockroach hindgut and cricket Malpighian tubules (Coast *et al.*, 1990).

1.11.5 The Leucokinin Receptor

To date, the insect kinin receptor remains unidentified, but it is probable that kinins initiate diuresis by first binding a GPCR located on the insect Malpighian tubule. An iodine labelled kinin analogue (^{125}I-Kinin) cross links with cricket Malpighian tubule membrane proteins, after SDS-PAGE an autoradiograph of the gel revealed 3 bands with M_r 53 to 57 kDa, a size range appropriate for GPCRs (Coast, 1996).

Species	Name	Amino Acid Sequence	Reference
Acheta domesticus	Achetakinin-I	SGADFYPWG-NH_2	(Holman *et al.*, 1990)
	Achetakinin-II	AYFSPWG-NH_2	(Holman *et al.*, 1990)
	Achetakinin-III	ALPFSSWG-NH_2	(Holman *et al.*, 1990)
	Achetakinin-IV	NFKFNPWG-NH_2	(Holman *et al.*, 1990)
	Achetakinin-IV	AFHSWG-NH_2	(Holman *et al.*, 1990)
Aedes aegypti	I	NSKYVSKQKFYSWG-NH_2	(Veenstra, 1999)
	II	NPFHAWG NH_2	(Veenstra, 1999)
	III	NNPNVFYPWG NH_2	(Veenstra, 1999)
Locusta migratoria	Locustakinin	AFSSWG-NH_2	(Schoofs *et al.*, 1992)
Lymnaea stagnalis	Lymnokinin	PSFHSWS-NH_2	(Cox *et al.*, 1997)

Table 1.2. Amino acid sequences from a selection of the 23 members of the insect kinin family and a related kinin from pond snail, *Lymnaea stagnalis*. Conserved amino acids are highlighted.

Immunocytochemical studies using antisera raised to insect leucokinins have indicated the presence of kinin-like peptides in the CNS of the parasitic nematode *Ascaris suum* (Smart *et al.*, 1993), the spider *Cupiennius salei* (Schmid and Becherer, 1996), and also in the mollusc *Helix pomata* (Elekes *et al.*, 1994). It is therefore possible that kinin-like peptides play an important role in water homeostasis across many invertebrate species. Using molecular techniques neuropeptide receptors were cloned from the CNS of the pond snail, *Lymnaea stagnalis*. One of the 'orphan' receptors was shown to have 33% sequence homology with the vertebrate neuropeptide Y_1 receptor protein. The cloned receptor was functionally expressed in animal cells and used in a 'reverse physiology' experiment to identify its natural ligand. In the reverse physiology experiment, extracts from *L.stagnalis* were separated by HPLC and assayed against cell lines stably expressing the receptor. Peptide binding was monitored indirectly by observing calcium motility. Peptides from the active HPLC fractions were purified to homogeneity and then sequenced. The active peptide was classified as a molluscan kinin-like peptide. The peptide, PSFHSWS-NH_2, (Table 1.2) was shown to bind to the receptor with high affinity and was able elicit an increase of intracellular calcium when applied in nM amounts. This is the first example of the cloning and functional expression of an invertebrate neurokinin receptor, NKR (Cox *et al.*, 1997). Since this initial identification another kinin like receptor has been cloned in the Southern cattle tick, *Boophilus microplus*. The *B.microplus* kinin-like receptor cDNA encodes a protein with 40% amino acid identity with the *L. stagnalis* kinin receptor (Holmes *et al.*, 2000). (Holmes *et al.*, 2000) also propose that the *Drosophila* genes AAF50775.1 and CG10626 encode the first insect kinin receptor. However, this speculation can only be confirmed through

functional expression of these receptor proteins and identification of their bioactive ligands. Proteins sequence alignments of kinin receptors from *B.microplus* and *L.stagnalis*, as well as putative *Drosophila* and *Anopheles* kinin receptors reveal a significant amount of sequence similarity between proteins (Fig. 1.4).

1.11.6 Synergism Between DPS (CRF-Like and Kinins)

The amounting body of DP sequence information suggests that both CRF-DPs and kinins commonly coexist in the same species (Table 1.1 & 1.2), which is likely to be the case for all insects. It is thought that kinins act in a cooperative fashion with CRF-peptides to control MT fluid secretion. Evidence for this cooperative behaviour comes from the study of CRF-DPs and kinins on fluid secretion rates in isolated Malpighian tubules of *Locusta migratoria*. When the locustakinin (Table 1.2) and Lom-DP (Table 1.1) were assayed together at physiological concentrations the response was greater than that of the sum of their individual activities, which is by definition synergism (O'Donnell and Spring, 2000). This cooperative action between locustakinin and *Locusta*-DP is the first description of synergism between two insect neuropeptides. Synergistic regulation of fluid secretion allows the insect to conserve metabolically expensive peptides and efficiently modulate fluid secretion with minimal expense. A further advantage to such a system is the speed the system can be switched on and off, allowing tight and accurate control of water regulation, only a small change in stimulant concentration is needed to markedly alter tubule function.

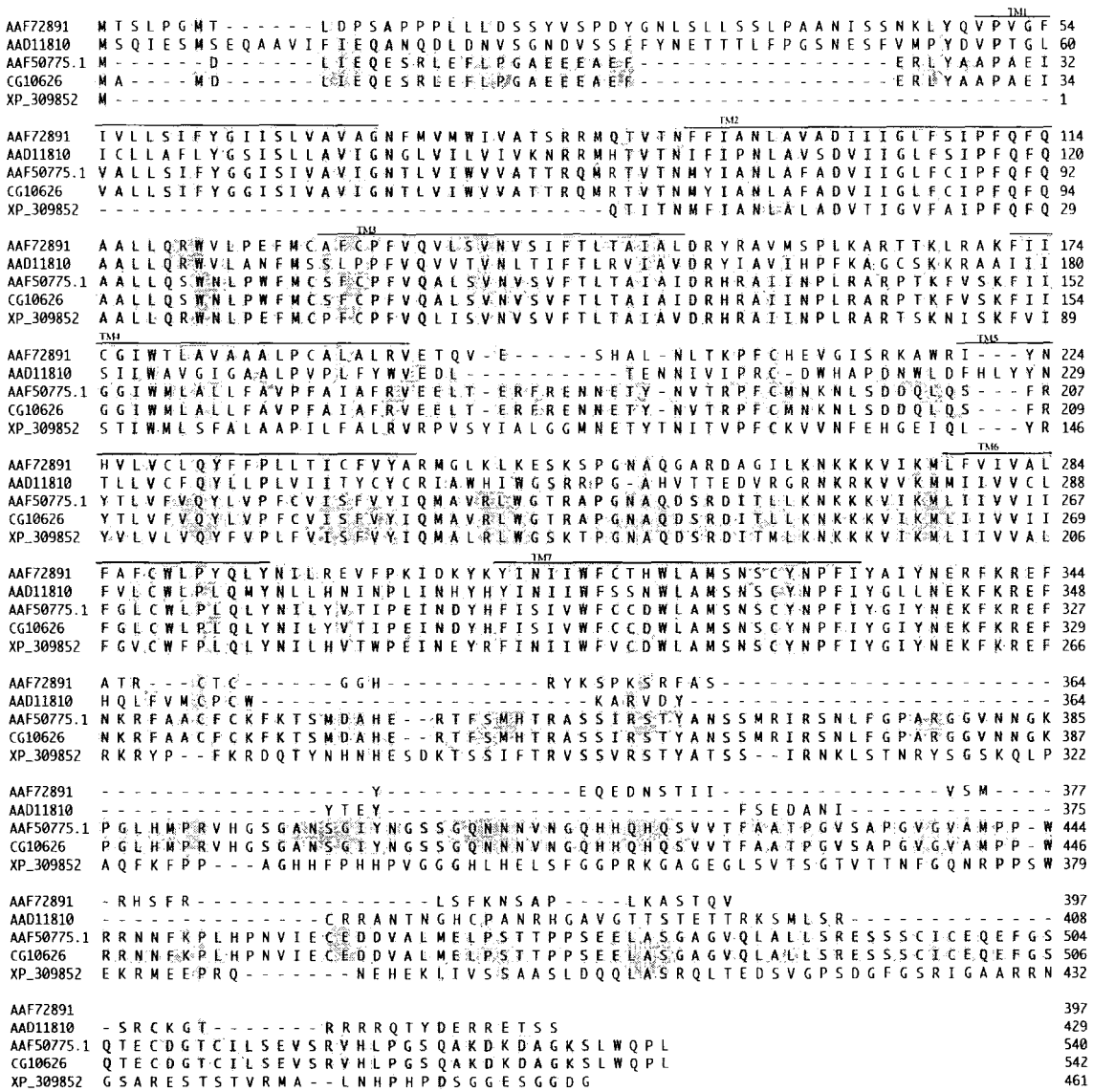


Figure 1.4. CLUSTAL V amino acid sequence alignment of the kinin receptors cloned from *Lymnaea stagnalis* (Lymnokinin-R, AAF72891), *Boophilus microplus* (tick kinin-R, AAD11810) and related neuropeptide receptors from *Drosophila* (CG10626 and AAF50775.1) and *Anopheles gambiae* (XP_309852). The predicted transmembrane domains of the AAF72891 are highlighted above the sequence.

1.12 Research Objectives

The overall aim of this project is to identify and validate gut specific 'target' proteins from *N.lugens*; where a target is defined as a protein that can be anticipated to be crucial for insect survival, and a modulation in function of such a protein is potentially detrimental towards insect survival. Once target sequences from *N.lugens* are cloned, cDNAs will be over expressed in an appropriate system and antibodies will be raised against specific epitopes of purified proteins. Purified antibodies will be orally delivered to *N.lugens* in artificial diet feeding trails and insect survival will be monitored.

CHAPTER 2.

MATERIALS AND METHODS

2.1 *N.lugens* Insect Cultivation

A colony of rice brown planthoppers (*Nilaparvata lugens*) was reared on susceptible two to three month old rice plants (*Oryza sativa* var. Taichung Native 1, TN1) in insect proof cages. *N.lugens* colonies and rice plants were held under controlled environmental conditions; 27°C, >80% r.h., with a 16h : 8h light : dark regime. New rice plants were introduced into *N.lugens* cages at weekly intervals to maintain the colony.

2.2 *N.lugens* Artificial Diet Feeding Trials

2.2.1 Planthopper Artificial Diet (MMD-1)

The planthopper artificial diet MMD-1 (Mitsuhashi, 1974) was determined to be the most suitable diet formulation for *N.lugens* feeding trials (Powell, 1993). Therefore, all feeding trials utilised this diet, the formulation of which is given in Table 2.1. The ingredients cysteine and tyrosine were first dissolved in a minimal volume of 1M HCl and riboflavin was dissolved by gentle heating in distilled water prior to addition to the diet. The artificial diet was filter sterilised (Merk, 0.2 μm pore), and sterility was maintained by carrying out subsequent manipulations in a sterile laminar flow hood. Sterile diet was aliquoted into 10 ml fractions and stored at –20°C for a maximum of 4 weeks, due to the instability of sodium ascorbate. To prevent deleterious effects on

insect survival, artificial diet ingredients were kept separate from common laboratory chemicals, and no antibiotics were incorporated into the diet.

2.2.2 Artificial Diet Feeding Trials

N.lugens were removed from the rice plants by gently tapping the plant stem and were collected in large petri dishes. Using a fine paintbrush five newly emerged first instar nymphs were carefully transferred to a feeding chamber (Fig. 2.1). *N.lugens* nymphs fed from 200 μl of artificial diet (MMD-1), which was sandwiched between two layers of parafilm M stretched over an open petri dish (diameter 30mm). For initial feeding trials test proteins were incorporated in the artificial diet at 0.1% (w/v), a concentration that reflects what is achievable in transgenic rice plant protein expression (Maqbool *et al.*, 2001). For every bioassay two control feeding trials were run in parallel, which included artificial diet only (positive control) and a no diet control (negative control), where the artificial diet was replaced with water. Individual feeding chambers were transferred to a growth room with conditions optimised for *N.lugens* survival (Section 2.1). Ten replicates per treatment were set up and artificial diet was replaced every other day to prevent bacterial contamination. At the same time every day the number of surviving nymphs from each treatment was recorded.

2.3 *Escherichia coli* Strains

Escherichia coli strains were used for plasmid propagation and prokaryotic protein expression. Plasmid DNAs were maintained at high copy numbers in the following *E.coli* strains: DH5α (genotype: F$^-$ ϕ80d*lac*ZΔM15, Δ(*lac*ZYA-*arg*F)U169, *deo*R *rec*A1, *end*A1, *hsd*R17(r_K^-, m_K^+), *pho*A, *sup*E44, , λ$^-$, *thi*-1, *gyr*A96, *rel*A1) and

Ingredient	mg/l	Ingredient	mg/l
K_2HPO_4	7500	L-threonine	1400
$MgSO_4.7H_20$	1230	L-tryptophan	800
Sucrose	50000	L-tyrosine*	400
L-alanine	1000	L-valine	800
L-arginine hydrochloride	2700	Thiamine hydrochloride	25
L-asparagine	5500	Riboflavin#	5
L-aspartic acid	1400	Nicotinic acid	100
L-cysteine*	400	Pyridoxine hydrochloride	25
L-glutamic acid	1400	Folic acid	5
L-glutamine	1500	Calcium pantothenate	50
Glycine	800	Meso-inositol	500
L-histidine	800	Choline chloride	500
L-isolucine	800	Biotin	1
L-leucine	800	Sodium ascorbate	1000
L-lysine hydrochloride	1200	$FeCl_3.6H_2O$	22.28
L-methionine	800	$CuCl_2.2H_2O$	2.68
L-phenylalanine	400	$MnCl_2.4H_2O$	7.93
L-proline	800	$ZnCl_2$	11.88
DL-serine	800	$CaCl_2.2H_2O$	31.15

Adjust the pH to 6.5 with KOH

Table 2.1. Composition of the planthopper artificial diet MMD-1 (Mitsuhashi, 1974). The artificial diet was made up in 200 ml batches, filter sterilised and stored at –20°C. Diet was kept for a maximum of 1 month. 200 μl of artificial diet was added to parafilm feeding sachets and fed to *N.lugens* contained in feeding chambers (Fig. 2.1). *Dissolve in 1N HCl, #Dissolve by gentle heating in distilled water.

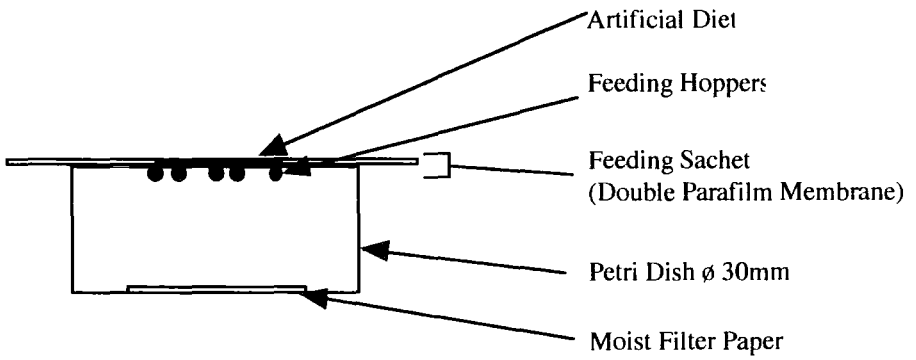


Figure 2.1. The *N.lugens* feeding chamber. Five insects were enclosed in a feeding chamber by a parafilm feeding sachet that contained 200 μl of artificial diet (MMD-1). Humidity was kept high by the presence of moist filter paper at the base of the chamber. To prevent bacterial contamination artificial diet was changed every other day in a sterile laminar flow hood.

TOP10F' (genotype: F'{*lac*I^q, Tn*10*(*Tet*^{R})}, *mrc*A, Δ(*mrr-hsd*RMS-*mcr*BC), ϕ80*lac*ZΔM15, Δ*lac*X74, *deo*R, *rec*A1, *ara*D139, Δ(*ara-leu*)7697, *gal*U, *gal*K, *rps*L(Str^R), *end*A1, *nup*G). DH5α and TOP10F' cells were purchased from Invitrogen.

For over expression of recombinant proteins expression plasmids were transferred to an *E.coli* host containing a chromosomal copy of the T7 RNA polymerase gene under *lacUV5* control. These strains were STAR (DE3) (genotype: F$^-$ *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm rne131* (DE3)) and RosettaBlue(DE3) (genotype: *endA1 hsdR17*($r_{K12}^- m_{K12}^+$) *supE44 thi-1 recA1 gyrA96 relA1 lac* |F' *proA*$^+$*B*$^+$ *lacI*^{q} *ZΔM15* ::Tn*10*(Tc^R)| (DE3) pRARE (Cm^R)). BL21 STAR (DE3) and RosettaBlue (DE3) cells were purchased from Invitrogen and Novagen, respectively.

2.4 Insect Cell Expression Cell Lines

Cloned genes were expressed in either Sf21 cells or Sf9 cells. Both Sf9 and Sf21 cells are traditionally used in baculovirus gene expression, the cell line originated at the USDA Insect Pathology Laboratory. These two cell lines originated from the IPLBSF-21 cell line, derived from pupal ovarian tissue of the fall armyworm, *Spodoptera frugiperda*. Both cell lines were purchased from Invitrogen.

2.5 Chemicals and Reagents

All chemicals and reagents were supplied by either BDH Chemicals Ltd (Poole, Dorset, UK) or Sigma Chemical Company (St Louis, USA). Chemicals and reagents were of analytical grade, or best commercially available.

2.6 Standard Molecular Biology Techniques

All standard techniques are common practice in the Department of Biological and Biomedical Sciences, Durham University, and were based on protocols in Molecular Cloning: A Laboratory Manual (Sambrook and Russell, 2001); unless stated otherwise.

While working with RNA all equipment was treated with 0.1% (v/v) diethyl pyrocarbonate (DEPC) overnight at 37°C and then autoclaved. Equipment for routine DNA work was sterilised by autoclaving.

2.6.1 Bacterial Culture

For routine work, bacteria were propagated in sterile Luria-Bertani Medium (LB: 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl). Liquid cultures were inoculated with a single bacterial clone and grown overnight (approx. 16 hours) at 37°C, with shaking at 220 rpm. For growth of bacteria on solid LB agar, liquid media was prepared and Bacto Agar (15 g/l) was added prior to autoclaving, upon cooling the agar was poured into sterile petri dishes. Inoculated dishes were inverted and incubated at 37°C overnight. When necessary, antibiotics were added to the culture medium. Commonly used antibiotics included carbenicillin (50 μg/ml) and kanamycin (50 μg/ml) and Zeocin (50 μg/ml). When using Zeocin (Invitrogen) Low-Salt LB (LS-LB) was used, where the salt was decreased to 5 g/l.

2.6.2 Competent Cells

Competent *E.coli* (strain TOP10F') were obtained from Invitrogen. Other strains were prepared by using a modified rubidium chloride method (Hanahan, 1985). Briefly, cells

were grown in LB broth supplemented with 20 mM $MgSO_4$, with shaking (220 rpm), until exponential growth phase was reached (OD_{600nm} = 0.4-0.6). Cells were pelleted by centrifugation (4,500 x *g* for 5 minutes, 4°C) and resuspended in 0.4 volumes of ice-cold wash buffer (30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% (v/v) glycerol). Cells were pelleted, as before, and resuspended in 1/25 volume of ice-cold transformation buffer (10 mM MOPS, pH 6.5, 75 mM $CaCl_2$, 10 mM RbCl, 15% (v/v) glycerol). Cells were incubated on ice for 15-60 minutes and divided into 200 μl aliquots and then snap frozen in liquid nitrogen and stored at –80°C.

2.6.3 Transformation of Competent Cells

Competent cells were removed from –80°C storage and thawed on ice. Plasmid DNA for transformation (10 ng plasmid DNA in a maximum of 3 μl) was added to the cells and gently mixed. Cells were incubated on ice for 30 minutes and then heat shocked for 30-60 seconds at 42°C, cells were then placed on ice for 2 minutes. Following transformation 1 ml of LB broth was added to each tube and then incubated at 37°C for 1 hour, with shaking (220 rpm). Transformants were selected by plating cells (100-200 μl) on LB-agar containing an appropriate antibiotic. For plasmid DNAs allowing blue/white screening LB-agar was supplemented with 40 μg/ml 5-bromo-4-chloro-3-indoyl-β-D-galatoside (X-Gal) and 0.1 mM isopropyl-β-D-thiogalatoside (IPTG).

2.6.4 Bacterial Glycerol Stocks

E.coli clones were grown overnight in LB broth with appropriate antibiotic at the specified concentration for the bacterial strain/plasmid combination. An equal volume of overnight culture and sterile 80% (v/v) glycerol was mixed to a final volume of 1 ml in

sterile cryovials. Cryovials were snap frozen in liquid nitrogen and then transferred to –80°C for long-term storage.

2.6.5 Plasmid DNA Mini-Preparation

Plasmid DNAs were isolated from up to 10 ml of overnight *E.coli* culture with the Wizard® *Plus* SV Miniprep kit (Promega), according to manufacturers instructions. Plasmid DNAs were eluted in water and consistently gave plasmid DNA concentrations of 100-250 ng/μl. Isolated plasmid DNAs were stored at –20°C.

2.6.6 Restriction Endonuclease Digestion

Restriction endonuclease (RE) digests were carried out using buffers and temperatures recommended by the manufacturers. Typical digests were carried out at 37°C on 1 μg of DNA using 2-10 units of RE (under optimal conditions 1 U of RE will completely digest 1 μg of DNA in a 50 μl reaction volume in 1 hour). Restriction products were separated by agarose gel electrophoresis and visualised by ethidium bromide staining.

2.6.7 Agarose Gel Electrophoresis of DNA

Mixtures of DNA were separated according to size by submarine agarose gel electrophoresis (NBL medium size gel apparatus and Pharmacia GNA 100 apparatus), as described by Sambrook and Russell, 2001. Gels contained agarose up to a concentration of 2% (w/v) in 1X TAE buffer (20 mM glacial acetic acid, 0.2 mM EDTA, 40 mM Tris, pH 7.2) and ethidium bromide (0.5 μg/ml). Prior to loading, DNA samples were mixed with 1/6 sample volume of 6X loading dye (10 mM Tris-HCl at pH 8.0, 10 mM EDTA, 30% (w/v) glycerol, 0.1% (w/v) orange G). Samples were loaded onto horizontal gels submerged in 1X TAE buffer containing ethidium bromide (0.5

μg/ml) and run at room temperature between 50-100V (constant voltage). Size markers covering the appropriate range were run alongside DNA samples. Size markers included *Hind*III or *Eco*471 digested lambda DNA (MBI-Fermentas); a 100 bp ladder (Bioline) and a 2-log DNA ladder (New England Biolabs, NEB). The ethidium bromide stained gels were viewed using a Bio-Rad Gel Doc 2000 imaging system and images were captured on a PC running Quantity One software (Bio-Rad).

2.6.8 Recovery of DNA Fragments From Agarose Gels

DNA gels were visualised on a trans-illuminator (UVB, λ 302 nm) and appropriate bands were excised from the gel by using a single edged razor blade. DNA was purified from the agarose by using QIAquick gel extraction kit (Qiagen), according to manufacturers instructions. Eluted DNAs were stored at –20°C.

2.6.9 Nucleic Acid Quantification and Qualification

Spectrophotometric methods were used to quantify nucleic acids. Optical density (OD) readings were taken at λ 260 nm using a Beckman DU 7500 spectrophotometer. An OD = 1 was assumed to be equivalent to a concentration of 40 μg/ml for ssRNA, 33 μg/ml for ssDNA and 50 μg/ml for dsDNA. Nucleic acid purity was determined by λ 260 nm / λ 280 nm ratio, nucleic acid solutions with ratios greater than 1.8 were considered free from contaminating proteins. Nucleic acid integrity was checked by agarose gel electrophoresis and ethidium bromide staining.

2.6.10 Ethanol Precipitation of Nucleic Acid

DNA was precipitated from solution by adding 1/10 volume of 3 M sodium acetate (pH 5.2) and 2 volumes of ice-cold ethanol. The tube was vigorously mixed and then placed

at –20°C for at least 1 hour. For small amounts of DNA (<0.1 μg/ml) glycogen was added as a carrier molecule (final concentration 50 μg/ml) and $MgCl_2$ (final concentration 10 mM). Precipitated DNAs were pelleted by centrifugation (13,000 x *g* for 15 minutes, 0°C). The supernatant was removed and the pellet was washed in 70% (v/v) ethanol. Following centrifugation (13,000 x *g* for 5 minutes, 0°C) the supernatant was removed and the pellet was dried in a 37°C block for 10 minutes. The DNA pellet was then dissolved in an appropriate volume of nuclease-free water. RNA was precipitated using the same protocol, except 3 volumes of ice-cold ethanol was added for precipitation.

2.6.11 DNA Ligation

Following RE digestion, DNAs for ligation were separated by agarose gel electrophoresis and purified from the gel (Qiagen). DNAs with compatible ends for ligation were set up in standard 10 μl reactions using commercially available T4 ligase and buffer (Promega), according to manufacturers instructions. A typical ligation reaction used 100 ng of plasmid DNA, in a 1 : 1 molar ratio of vector : insert. Sticky end ligations were incubated at 22°C for a minimum of 3 hours, while blunt end ligations were left overnight. Following ligation the DNA was used to transform *E.coli* cells of an appropriate strain.

2.6.12 DNA Sequencing and Sequence Analysis

DNA sequencing reactions were carried out using BigDye Terminator with AmpliTaq DNA polymerase (ABI Biosciences). Reaction products were analysed on automated sequencers (ABI Prism 373 STRETCH and ABI Prism 377 XL; DBS Genomics, Dept.

of Biological and Biomedical Sciences, University of Durham). Expression constructs were completely sequenced on both strands of the DNA; this was carried out either by subcloning of suitable restriction fragments, or by using primers directed against determined sequence to complete overlaps. Contiguous sequences were produced in Sequencher™ Version 4.1.2 (Gene Codes Corporation) on a Macintosh computer. Edited nucleotide data was used in BLAST similarity searches (Altschul *et al.*, 1990) against the GenBank database (http://www.ncbi.nlm.nih.gov/BLAST/), and identification of sequence features in encoded polypeptides was performed using the CBS prediction servers (http://www.cbs.dtu.dk/services/).

2.6.13 Oligonucleotides

Oligonucleotides were chemically synthesised on a solid support by TAGN Ltd (International Centre for Life, Newcastle; http://www.vhbio.com/tagn/). Upon receipt primers were resuspended in sterile water at a final concentration of 100 pmol/μl and stored at –20°C. In PCR reactions non-degenerate primers were used at a final concentration of 0.2 μM, for degenerate primers this concentration was increased to 1.0 μM. Melting temperature (T_m) was calculated by the following formula: T_m = 69.3+ 0.41x(%G+C)-650/(nA+nG+nC+nT). The annealing temperature (T_a) of a particular oligonucleotide was commonly 3°C below the calculated T_m.

2.6.14 Purification of Total RNA

Prior to RNA isolation, tissue samples were weighed and then preserved in RNA*later* solution (Qiagen). For small amounts of starting material (up to 30 mg) total RNA was isolated using RNeasy Mini purification kit (Qiagen), according to manufacturers

instructions. In brief, tissues were homogenised on ice in a 2 ml glass tube with a PTFE pestle at 1,000 rpm. Cell debris was pelleted by centrifugation (12,000 x *g* for 5 minutes, 22°C) and the supernatant was loaded onto the spin column. Following ethanol wash steps total RNA was eluted in 50 μl RNase free water. RNA was quantified using spectrophotometric methods, and its integrity was determined by denaturing formaldehyde/agarose gel electrophoresis. Eluted RNA was either aliquoted, snap frozen in liquid nitrogen and stored at –80°C; or used as a starting material for cDNA synthesis or poly-A^+ RNA purification.

2.6.15 Purification of Poly-A^+ RNA

Poly-A^+ RNA was purified from 0.1 - 1 mg total RNA using PolyATtract mRNA isolation systems III (Promega), according to the manufacturers protocol. In brief the total RNA solution is mixed with a biotinylated oligo(dT) primer, which binds to poly-A^+ tail of mRNA. Hybrids were recovered using streptavidin coupled to paramagnetic particles and a magnetic separation stand. After several washes poly-A^+ RNA was eluted in RNase free water. RNA was quantified by spectrophotometric methods and its integrity was determined by denaturing formaldehyde/agarose gel electrophoresis.

2.6.16 Formaldehyde Gel Electrophoresis of RNA

RNAs were separated according to size by denaturing formaldehyde agarose gel electrophoresis, using a modified protocol (Fourney *et al.*, 1988). Gels contained agarose up to a concentration of 1.5% (w/v) in 1X MOPS/EDTA (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, pH 7.0), and formaldehyde at a final concentration of 2% (v/v). After pouring, gels were allowed to polymerise for 1 hour before use. Prior to

loading, RNA concentration was determined using a spectrophotometer and a volume equivalent to 5-10 μg of total RNA was prepared for electrophoresis. Dilute RNA samples were precipitated and resuspended in 5 μl RNase free water. RNAs in a 5 μl volume were mixed with 25 μl loading buffer (0.75 ml deionised formamide, 0.15 ml 10X MOPS, 0.24 ml formaldehyde, 0.1 ml deionised RNase free H_2O, 0.1 ml glycerol, 0.08 ml 10% (w/v) bromophenol blue) and placed in a 65°C water bath for 15 minutes. After denaturation 1 μl of ethidium bromide (at 10 mg/ml) was added to each sample. Denaturing gels were submerged in 1X MOPS/EDTA and samples were loaded, and run at room temperature between 50-100V (constant voltage). RNA size markers (Promega) were run alongside RNA samples. Following electrophoresis RNA was directly visualised without further staining or destaining. RNA gels were viewed and imaged as described for DNA gels (Section 2.6.7).

2.6.17 Standard Polymerase Chain Reaction (PCR)

PCR reactions (25-100 μl) were set up on ice, in thin walled PCR tubes. Reactions consisted of 0.2 mM each of dATP, dGTP, dCTP and dTTP, 2.5 mM $MgCl_2$, 50 mM KCl, 0.1% (v/v) Triton X-100, 10 mM Tris-HCl (pH 9.0), DNA template (5-100 ng) and 1.25 Units of *Taq* DNA polymerase (Promega) per 50 μl reaction (were 1U catalyses the incorporation of 10 nmol of dNTP into an acid insoluble form at 74°C). Gene specific oligonucleotide primers were used at 0.2 μM and degenerate primers were used at a higher concentration of 1 μM. For multiple PCR reactions a master mix containing all common components was mixed in a 1.5 ml eppendorf tube, and then aliquoted to individual PCR tubes. Setting up multiple PCR reactions in this way reduced pipette error and ensured consistency across all reactions. PCR thermo-cycling was determined

empirically for each primer combination and was performed on a Biometra 1000 thermal cycler, or a Perkin Elmer 2400 thermal cycler. Where high fidelity PCR was required Advantage 2 polymerase mix (Clontech) was used, this consisted of TITANIUM *Taq* DNA polymerase (Clontech), proof reading polymerase and TaqStart antibody (for hot start PCR). High fidelity PCR was set up using buffers supplied by the manufacturer, according to their protocol. The error rate for high fidelity PCR is 8.5 x 10^{-6} mistakes/base, whereas the error rate for *Taq* polymerase is 2.6 x 10^{-5} mistakes/base.

2.6.18 Reverse Transcription-PCR (RT-PCR)

In a standard procedure total RNA was reverse transcribed into first strand cDNA using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT, Promega). In brief, 2 μg of total RNA was mixed with 0.5 μg of poly-$T_{(24)}$ primer in a 15 μl volume. RNA secondary structure was melted by heating to 70°C for 5 minutes, and then placed on ice to prevent secondary structure reforming. The reaction was completed on ice by adding, 5 μl of 5X reaction buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM $MgCl_2$, 50 mM DTT); dNTPs (final concentration, 0.5 mM) and 200 Units of M-MLV RT (were 1 U catalyses the incorporation of 1 nmol of dNTP into an acid insoluble form in 10 minutes at 37°C), plus additional nuclease free water up to a volume of 25 μl. Reaction tubes were incubated at 42°C for 60 minutes for production of first strand cDNA. Reaction products were diluted 10-fold and 5 μl aliquots were used as template in a standard 50 μl PCR amplification.

2.6.19 Subcloning PCR Products

PCR products were purified from agarose gels (Qiagen), ligated into pCR2.1 TOPO (Invitrogen) and used to transform TOP10F' cells (Invitrogen), according to manufacturers instructions. Transformants were plated on LB-kanamycin (50 μg/ml) with IPTG (0.1 mM) and X-Gal (40 μg/ml). Several independent recombinant clones were picked and grown overnight at 37°C in 10 ml LB broth containing 50 μg/ml kanamycin. Plasmid DNA from positive transformants was isolated by mini-preparation (Promega) and checked by restriction enzyme digest and sequencing.

2.6.20 Rapid Amplification of cDNA Ends (RACE) PCR

RACE experiments were performed on 1 μg of total RNA or poly-A^+ RNA using SMART RACE cDNA amplification kit (Clontech), according to the manufacturers protocol. RACE experiments were used to retrieve the complete 5' and 3' end of partial cDNAs, including any untranslated regions (UTR). Gene specific primers (GSP) from partial cDNA sequences were designed, a sense primer was used in 3' RACE experiments and an antisense primer in 5' RACE experiments.

In 3' RACE poly-A^+ RNA was primed by 3'-RACE CDS primer (5'-AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}N_{-1}$N-3'; N=A, T, G or C; N_{-1}=A, G, C; universal sequence is underlined). This primer anneals to the poly-A^+ tail of mRNA and first strand cDNA is produced by primer extension using PowerScript reverse transcriptase (Clontech). After reverse transcription RACE prepared cDNA was amplified by PCR between the 3' RACE gene specific primer (GSP) and the universal sequence incorporated by the 3'-RACE CDS primer. Amplification between these two

primers ensured that the 3' end of the gene was fully represented, including any 3' UTR (Fig. 2.2a and b).

In 5' RACE poly-A^+ RNA was primed by 5'-RACE CDS primer (5'-$T_{(25)}N_{-1}$N-3'; N=A, T, G or C; N_{-1}=A, G, C). This primer anneals to the poly-A^+ tail of mRNA and first strand cDNA is produced by primer extension using PowerScript reverse transcriptase (Clontech). cDNA extension continues in a 5'-3' direction until the end of the poly-A^+ template is reached. Upon reaching the end of the template PowerScript reverse transcriptase adds 3-5 additional dC residues to the nascent first strand cDNA. Subsequently, a stretch of complementary dG in the SMATR IIA primer base pairs and serves as an extended template for PowerScript reverse transcriptase. After reverse transcription first strand cDNA was amplified by PCR between the 5' RACE GSP and the universal sequence incorporated by cDNA extension across the SMART IIA oligonucleotide. Amplification between these two primers ensured that the 5' end of the gene was fully represented, including any 5' UTR (Figure 2.3a and b).

RACE PCR products were analysed by gel electrophoresis and products of the predicted size were purified from the gel, cloned and sequenced, according to standard protocols. When RACE experiments failed to generate specific product a nested PCR approach was used. Initial RACE amplification reaction products were diluted and reamplified using a pair of internal primers. Specific reaction products were identified by transferring RACE products onto nitrocellulose membranes and probing with a radiolabelled fragment from the gene of interest.

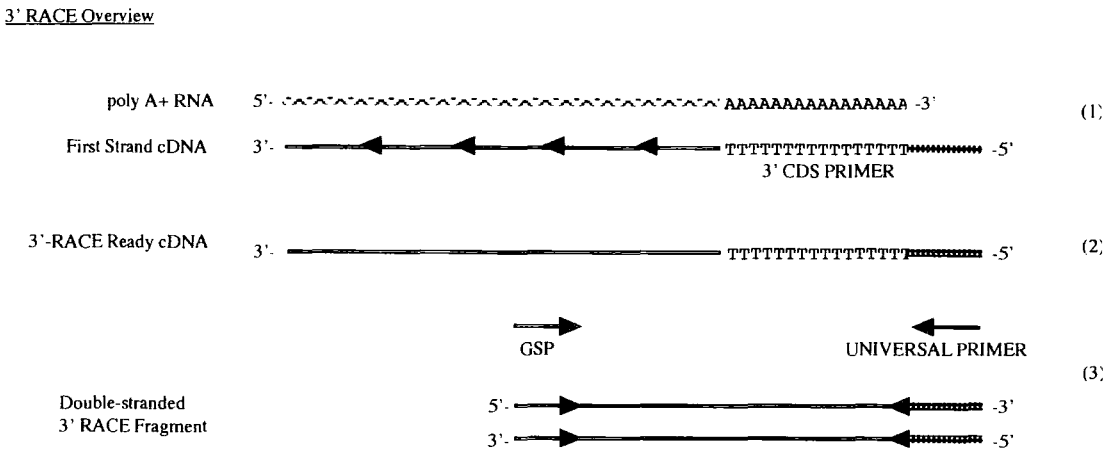


Figure 2.2a. Overview of the 3' RACE (Rapid Amplification of cDNA Ends) reaction.

Poly-A^+ RNA is primed using a poly-$T_{(30)}$ adaptor primer (3' CDS Primer), which incorporates a 5' universal priming site (1). Extension of this primer by PowerScript reverse transcriptase (Clontech) results in first strand cDNA (2). The 3' end of the desired gene is amplified by PCR between a universal primer incorporated in the 3'-CDS primer and a gene specific primer (GSP) corresponding to the gene of interest. Amplification between these two primers ensures that the 3' end of the gene is fully represented including any 3' UTRs (untranslated regions). Primer sequences used in 3' RACE experiments are shown below (Fig. 2.2b). (NOTE: Both Figure 2.2 a and b have been adapted from Clontech RACE manual).

5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'	Long UP
5'-CTAATACGACTCACTATAGGGC-3'	Short UP
5'-AAGCAGTGGTATCAACGCAGAGT-3'	Nested UP
5'-AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}N_{-1}$N-3' (N = A, C, G, or T; N_{-1} = A, G, or C)	3' CDS

Figure 2.2b. Sequence alignment of Clontech primers that were used in 3' RACE reactions

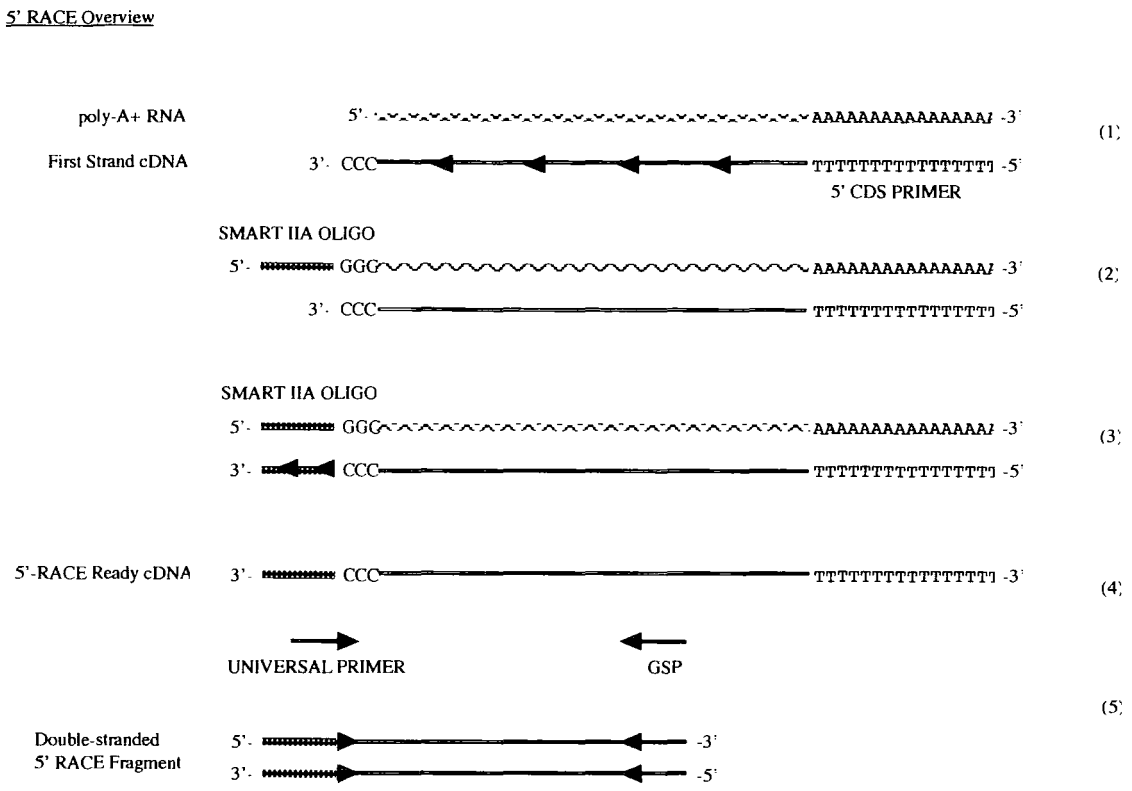


Figure 2.3a. Overview of the 5' RACE (Rapid Amplification of cDNA Ends) reaction.

Poly-A^+ RNA is primed using a poly-$T_{(25)}$ primer (5' CDS Primer), this primer is extended by PowerScript reverse transcriptase (Clontech). Upon reaching the 5' end of the message RNA the reverse transcriptase adds a run of poly C to the nascent first stand cDNA (1). A stretch of complementary G's in the SMART IIA oligo binds this stretch of poly-C (2). The RT switches from the RNA template to the SMART IIA oligo and continues extension across the length of the SMART II A oligo, this ensures full representation of the 5' end of the cDNA including any UTRs (untranscribed regions) (3&4). The 5' end of the gene is amplified by PCR between a universal primer incorporated in the SMART IIA oligo and a gene specific primer (GSP) corresponding to the gene of interest (5). Primer sequences used in 5' RACE experiments are shown below (Fig. 2.3b). (NOTE: Both Figure 2.3 a and b have been adapted from Clontech RACE manual)

5'-AAGCAGTGGTATCAACGCAGAGTACGCGGG-3'	SMART IIA
5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'	Long UP
5'-CTAATACGACTCACTATAGGGC-3'	Short UP
5'-AAGCAGTGGTATCAACGCAGAGT-3'	Nested UP
5'-$(T)_{25}N_{-1}N$-3' (N = A, C, G, or T; N_{-1} = A, G, or C)	5' CDS

Figure 2.3b. Sequence alignment of Clontech Primers that were used in 5' RACE reactions.

2.6.21 Transfer of DNA From Agarose Gels to Nitrocellulose

Gels containing nucleic acid were blotted onto nitrocellulose membranes (Protran BA 85, Schleicher and Schuell) by upward capillary transfer (Sambrook and Russell, 2001). Prior to transfer resolved gels were photographed and then soaked for 2 x 30 minutes in denaturation solution (1.5 M NaCl, 0.5 M NaOH), followed by 2 x 30 minutes in neutralisation solution (1.5 M NaCl, 1 M Tris-HCl, pH 7.4). Gels were then placed on the transfer apparatus as described, and DNA was transferred to a nitrocellulose filter in 20X SSC (3 M NaCl, 300 mM sodium citrate, pH 7.0). After blotting, the position of the lanes was marked on the filter and the DNA was fixed by baking the filter at 80°C for 2 hours. The filter was either stored for later use or was used for DNA hybridisation experiments.

2.6.22 Radioactive Labelling of DNA Using Random Primers

DNA template was prepared by restriction and gel purification (Sections 2.6.6 and 2.6.8, respectively). 25 ng of template was labelled to high specific activity (>10^8 cpm/μg) with 10 mCi/ml[α-^{32}P]dCTP (400 Ci/mmol) through extension of random hexamer primers by Klenow Polymerase (Promega), as described in Sambrook and Russell, 2001. The radiolabelled probe was separated from unincorporated dNTPs by spin column chromatography (QIAquick Nucleotide Removal Kit, Qiagen). Specific activity was determined by mixing a sample of the probe with Ecoscint scintillant and reading on a Packard 1600TR scintillation counter. Probes were stored at –20°C, before use probes were boiled for 5 minutes to render the DNA single-stranded.

2.6.23 Hybridising Labelled DNA Probes to Filters Containing DNA

Nitrocellulose membranes with attached DNA were hybridised with radiolabelled probes in large glass roller bottles (Techne) in a hybridisation oven (Hybridiser HB-1D, Techne). DNA : DNA hybridisations took place at 65°C, with all solutions preheated to this temperature. Filters were prehybridised in 6X SSC and 5X Denhardt's reagent (0.5% (w/v) SDS, 100 μg/ml salmon sperm DNA) for at least 2 hours. This solution was then replaced with hybridisation solution (6X SSC) and 2X Denhardt's reagent (0.5% (w/v) SDS, 100 μg/ml salmon sperm DNA), labelled probe was added and hybridisation was allowed to proceed overnight. Hybridisation solution was then removed and the filters were washed to the required stringency with 30 minute washes in wash solution I (2X SSC, 0.1% (w/v) SDS), wash solution II (0.1X SSC, 0.1% (w/v) SDS) and wash solution III (0.1X SSC). Filters were removed from the roller bottle, blotted dry and exposed to x-ray film with the aid of an intensifying screen. For exposures longer than 6 hours cassettes were stored at –80°C.

2.6.24 N.lugens *Whole Insect cDNA Library*

The *N.lugens* whole insect cDNA library was prepared by Dr Jinping Du using poly-A^+ RNA from final instar insects. The library was constructed using Lambda ZAP II XR library construction kit (Stratagene), according to standard protocols. cDNAs were directionally cloned between *Eco*R I and *Xho* I of the Lambda ZAP II vector (Stratagene) and packaged using Gigapack III Gold packaging extract (Stratagene). cDNAs were excised with helper phage to generate subclones in pBluescript SK(-) phagemid vector.

2.6.25 N.lugens *Gut Specific cDNA Library*

N.lugens guts were dissected out on ice and total RNA was isolated using RNeasy mini purification kit (Qiagen). The library was constructed using SMART cDNA library construction kit (Clontech), according to manufacturers protocol. Due to the limited amount of RNA starting material the cDNA library was constructed from 50 μg total RNA, and the cDNA was enriched using a PCR step. cDNAs were directionally cloned between *Sfi* IA and *Sfi* IB of the λTriplEx2 vector and packaged using Gigapack III Gold packaging extract (Stratagene). cDNAs were excised without helper phage in an *in vivo* recombination event to generate subclones in pTriplEx2 vector.

2.7 Standard Protein Analysis

2.7.1 Estimation of Protein Concentration

Protein concentrations were determined by the Bio-Rad protein assay, based on the methods of Bradford (Bradford, 1976). Upon binding protein the absorbance maximum of the Bradford dye reagent (Coomassie Brilliant Blue G-250) is shifted from 465 nm to 595 nm. The absorbance at 595 nm is proportional to the concentration of the protein solution. Experiments were set up in microtitre plates and bovine serum albumin (BSA) was used as a protein standard, with dilutions prepared in the same buffer as the sample. The linear range of the assay for BSA is 1.2-10 μg/ml, five dilutions within this range were used to plot a standard curve. Concentrations of unknown protein solutions were predicted using the standard curve. In microtitre plates 160 μl of each standard or unknown sample was added to separate wells and then mixed with 40 μl of Bradford

reagent. Absorbance was read at 570 nm using a Dynatech MT 5000 microtitre plate reader.

2.7.2 SDS PAGE Electrophoresis

Prior to loading, protein samples were denatured and reduced by boiling for 2 - 5 minutes in sample buffer diluted to 1X (7.5% (v/v) glycerol, 1% (w/v) SDS, 0.003% (w/v) bromophenol blue, 50mM DTT, 40 mM Tris-HCl, pH 6.8). Denatured proteins were separated according to their size by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Minigels (9 x 10 cm) were cast with a discontinuous pH gradient, which combined a stacking gel (2.5% (w/v) acrylamide, 0.1% (w/v) bisacrylamide, 125 mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate, 0.075% (v/v) *N,N,N',N'*-teramethylethylenediamine) and a resolving gel (8-20% (w/v) acrylamide, 0.21-0.53% (w/v) bisacrylamide, 375 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.075% (w/v) ammonium persulphate, 0.05% (v/v) *N,N,N',N'*-teramethylethylenediamine). Gels were run in electrophoresis buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS, pH 8.3) at a constant voltage (100 - 150 V) in ATTO-AE6450 apparatus. For estimation of protein size a molecular weight marker (SDS7, Sigma) was run alongside protein samples. This marker contained bovine albumin (66 kDa), ovalbumin (45 kDa), glyceraldehyde-3-phosphate dehydrogenase (36 kDa), carbonic anhydrase (29 kDa), trypsinogen (bovine pancreas; 24 kDa), trypsin inhibitor (soybean; 20 kDa) and α-lactalbumin (14.2 kDa). Following electrophoresis gels were either stained with Coomassie blue or transferred to nitrocellulose and probed with antibodies in western blot experiments.

2.7.3 Staining Gels With Coomassie Blue

After polyacrylamide electrophoresis proteins in the μg range were visualised by staining with 0.25% (w/v) Coomassie Brilliant Blue G250, in 10% (v/v) glacial acetic acid, 50% (v/v) methanol, followed by destaining with 5% (v/v) glacial acetic acid, 10% (v/v) methanol. Gels were stained for up to 3 hours, both staining and destaining were carried out at room temperature with gentle agitation.

2.7.4 Immunoblotting

After electrophoresis proteins were transferred to nitrocellulose membrane (Hybond ECL, Amersham) by electroblotting, using a standard semi-dry transfer protocol. Gels for electrotransfer were equilibrated in Bjerrum and Schafer-Nielson buffer (48 mM Tris HCl, 39 mM glycine, 20% (v/v) methanol, 0.0375% SDS, pH 9.2) by soaking for 30 minutes at room temperature. Nitrocellulose membranes and 3MM paper were cut to the same dimensions as the gel and pre-wet in the same buffer. The components of the blot were set up on ATTO AE-6675 blotting apparatus (Figure 2.4). Electroblotting was conducted at constant current at between 125 - 150 mA (~ 2.0 mA/cm^2) for 20 to 60 minutes. Efficiency of transfer was assessed by staining the nitrocellulose membrane with Ponceau S stain (0.2% (w/v) Ponceau S in 3% (v/v) acetic acid), followed by destaining with distilled water. Protein standards were marked on the nitrocellulose membrane prior to destaining. If required, nitrocellulose membranes were placed between 3MM paper, wrapped in cellophane and stored at 4°C.

For immuno detection non-specific protein binding sites were saturated by incubating the nitrocellulose membrane in 5% (w/v) marvel milk powder in Tris-buffered saline

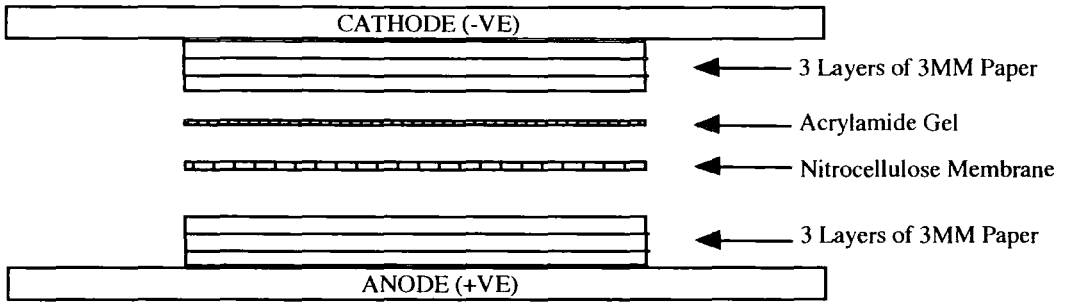


Figure 2.4. Set up of ATTO AE-6675 blotting apparatus for semi-dry transfer of proteins from acrylamide gels to nitrocellulose membranes. Proteins were transferred in Bjerrum and Schafer-Nielson buffer (48mM Tris-HCl, 39mM glycine, 20% (v/v) methanol, 0.0375% SDS, pH9.2). Electroblotting was conducted at constant current at between 125-150 mM (~2.0 mA/cm^2) for 20 to 60 minutes.

with Tween 20, TBST (TBST; 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween 20). Membranes were blocked at room temperature for 30 minutes with gentle agitation. Following blocking, membranes were incubated with TBST containing the appropriate dilution of primary antibody, membranes were reacted with primary antibody at room temperature for 60 minutes with gentle agitation. Unbound primary antibody was removed by washing the membrane three times in TBST for 5 - 10 minutes, each wash. The membrane was transferred to TBST containing the secondary antibody, goat anti-rabbit IgG horseradish peroxidase conjugate (BioRad) at 1 : 3,000 dilution. The membrane was incubated for 30 minutes, followed by three washes in TBST for 5 - 10 minutes each to remove any unbound secondary antibody. Excess Tween 20 was removed from the membranes by briefly washing with distilled water. Enhanced chemiluminescence (ECL) reagents (Amersham) were used to detect specifically bound secondary antibodies, according to manufacturers instructions. Specific antibody binding was visualised by exposing membranes to photo-sensitive film (Fuji-RX). Exposed films were developed with an automatic developer (X-ograph Imaging Systems Compact X4).

2.7.5 Protein Extraction From Whole Tissues

Proteins were extracted from tissues by homogenisation in 5 ml/g protein extraction buffer (20 mM Tris-HCl, pH 7.4, 10 mM NaCl, 10 mM KCl, 3 mM $MgCl_2$), which contained a cocktail of protease inhibitors (0.2 mM PMSF, 1 mM EDTA, 20 μM Leupeptin). Tissues were homogenised on ice in a 2 ml glass tube with a PTFE pestle at 1,000 rpm, antifoam A (Fluka) was added to prevent foaming. After homogenisation Nonidet P-40 was added to 0.5% (v/v) and the mixture was incubated at 4°C for 10-20

minutes with shaking. Insoluble material was pelleted by centrifugation (10,000 x *g* for 15 minutes, 4°C), and the protein-containing supernatant was removed. Protein preparations were either used directly or frozen in liquid nitrogen and stored at –80°C.

2.7.6 Expression of Recombinant Proteins in E.coli and Purification under Denaturing Conditions using Nickel Affinity Chromatography

BL21(DE3) STAR (Invitrogen) or RosettaBlue(DE3) (Novagen) cells were transformed with the pET24 (Novagen) based expression constructs. Clonal transformants were grown in liquid LB broth containing kanamycin (50 μg/ml) at 37°C with shaking (200 rpm) until an absorbance of 0.6 - 0.8 was reached. Mid-log cells were induced for expression with isopropyl β-D-thiogalactoside (IPTG, 1 mM) and growth was continued for 3 hours post induction under similar growth conditions. Cells were collected by centrifugation (10,000 x *g* for 30 min, 4°C) and resuspended in 5 ml/gram of cells of lysis buffer (8 M urea, 0.1 M NaH_2PO4, 0.01 M Tris-HCl, 10 mM imidazole at pH 8.0). To ensure complete lysis cells were mixed at RT for 2 hours followed by sonication. Cell debris was removed by centrifugation (10,000 x *g* for 20 minutes, 4°C) and the supernatant was removed and mixed with 3 ml Ni-NTA Superflow resin (Qiagen). Poly-His proteins were bound for 1 hour at RT. After binding, the Ni-NTA Superflow resin was loaded onto a 5 ml column and washed (8 M urea, 0.1 M NaH_2PO4, 0.01 M Tris-HCl, 20 mM imidazole at pH 8.0) to remove non-specifically bound protein. After washing recombinant proteins were eluted (8 M urea, 0.1 M NaH_2PO4, 0.01 M Tris-HCl, 250 mM imidazole at pH 8.0). Eluted proteins were pooled and protein concentration was estimated by a microtitre based Bradford assay (Biorad) using BSA as a standard. Recombinant protein was diluted to 10-50 μg/ml with lysis buffer and

refolded by dialysis into 10 mM Tris-HCl (pH 7.4) and finally water. After salt removal precipitated proteins were removed by centrifugation (20,000 x *g* for 25 min, 4°C). The supernatant containing renatured proteins was removed, frozen in liquid nitrogen and freeze-dried. Molecular mass of the recombinant protein was determined by either SDS-PAGE or SELDI-TOF mass spectrometry.

2.7.7 SELDI-TOF Mass Spectrometry

Molecular mass of purified proteins was accurately determined by surface enhanced laser desorption/ionisation with time-of-flight detection (SELDI-TOF). For molecular mass determination purified proteins were concentrated to 1 mg/ml in 10 mM Tris-HCl (pH 7.5) and applied to an individual spot of the SELDI protein chip, WCX2 (weak cation exchange, Ciphergen Biosystems Inc.). Samples were left to air dry before analysis. Samples were analysed on a SELDI mass analyser PBS II with a linear time-of-flight mass spectrometer (Ciphergen Biosystems Inc) using time-lag focusing.

2.7.8 Expression of Recombinant Proteins in Insect Cells

Coding sequences for expression in insect cells were cloned into the pIZ/V5-His vector (Invitrogen), where constitutive protein expression is driven by the baculovirus immediate-early promoter, OpIE2. The promoter allows expression in both Sf9 and Sf21 cells. Coding sequences were amplified with primers which incorporated a Kozak translation initiation sequence (G/A)NNATGG around the start codon (Kozak, 1990). Restriction enzyme sites were also included in the forward and reverse primer to facilitate directional cloning of the insert into the expression vector. Expression

constructs were checked by sequencing, constructs with no errors and in frame with the His-tag and V5 epitope were used for over expression in Sf21 insect cells.

For transformation either Sf9 or Sf21 cells were grown to confluence in SF-900 II serum free medium with L-Glutamine (Gibco), according to standard cell culture techniques. Cells with a viability >95% were transfected with the appropriate expression construct with Cellfectin (Invitrogen), according to the manufacturers protocol. Transient protein expression was monitored by analysing cell lysates by western blot, using an anti-V5 antibody (Invitrogen). Upon conformation of transient expression stable cell lines were generated by selection with Zeocin (Invitrogen).

2.7.9 Synthetic Peptide Synthesis

Synthetic peptides were designed to correspond to specific epitopes of a particular protein and were used in the production of polyclonal antibodies in rabbits. Synthetic peptides were synthesised by Dr Andrew Dinsmore (Syngenta CTL, Alderely Park).

2.7.10 Polyclonal Antibody Production

Polyclonal antibodies were raised in rabbits against purified recombinant protein or synthetic peptides conjugated to bovine serum albumin (BSA). Two independent rabbits were used for each immunisation. Polyclonal antibodies were raised according to standard protocols by Dr Andrew Dinsmore (Syngenta CTL, Alderley Park).

2.7.11 IgG Antibody Purification

IgG antibodies were purified from rabbit serum by affinity chromatography on immobilised Protein G. Columns containing 1 ml Protein G Sepharose (Amersham)

were operated with a peristaltic pump (1 ml/minute) in a chromatography system. Prior to loading, rabbit serum was equilibrated with binding buffer (20 mM sodium phosphate, pH 7.0) by overnight dialysis at 4°C. After loading non-specific proteins were removed by washing with binding buffer. IgG proteins were eluted from the column with 0.1 M glycine-HCl buffer, pH 2.7 and collected in 1 ml fractions, pH was immediately neutralised by the addition 200 μl 1 M Tris-HCl, pH 9.0. Fractions containing IgG protein were pooled and concentration determined by Bradford assay. IgGs were concentrated to 10 mg/ml by using a centrifugal concentrator (Vivaspin 20, MWCO 30,000, Vivascience) and purity was determined by SDS-PAGE and Coomassie Blue staining.

2.8 Diuretic Hormone Receptor Methods

2.8.1 Purification of Gut Specific Total RNA From N.lugens

Fifth instar *N.lugens* were selected in batches of 20 and kept on ice throughout dissection. Intact guts were dissected out in RNase free 10 mM Tris-HCl (pH 7.5) using the aid of a dissecting microscope and gauge-5 forceps. Gut tissue was stored at 4°C in RNA*later* solution (Qiagen). Upon collection of 50 full-length guts (~ 15 mg of tissue) total RNA was isolated (Qiagen). RNA was quantified by spectrophotometric methods (Section 2.6.9) and integrity was determined by denaturing formaldehyde/agarose gel electrophoresis (Section 2.6.16). *N.lugens* gut specific total RNA was used as a starting material in first strand cDNA synthesis and PCR cloning experiments. Oligonucleotide sequences of primers used to clone the *N.lugens* DHR and to prepare *E.coli* and insect expression constructs are detailed in table 2.2.

	Name	Sequence (5'→3' Direction)	RE Site	Product (bp)
Degenerate Primers	DHR Degenerate (fwd)	5'-ACIAA(T/C)TT(T/C)TGGATG(T/C)TIG(G/T)IGA(A/G)GG-3'	-	381
	DHR Degenerate (rev)	5'-ACIAGIGC(T/C)IGC(T/C)TTIC(G/T)(A/G)TA(T/C)TG-3'	-	
RACE Primers	DHR 5' RACE	5'-GTTATGGGTAGCCATCCAGGGACAGCT-3'	-	1242 + 5' UTR
	DHR 5' RACE (nested)	5'-CCATCCTATACAGACGTAGGCCCTCAG-3'	-	1119 + 5' UTR
	DHR 3' RACE	5'-ACTGAGGGCCTACGTCTGTATAGGATG-3'	-	829 + 5' UTR
Construct Primers	DHR N-term (fwd)	5'-AGTTCATATGGCGGAGACTGTATCCAATGAAACG-3'	*Nde* I	797
	DHR N-term (rev)	5'-TATGCTCGAGCATAGTGGTAAGTTCTATGCCAGGCTC-3'	*Xho* I	
	DHR CDS (fwd)	5'-AGCGATATC**AAAATGG**CGGAGACTGTATCCAATG-3'	*Eco*R V	1941
	DHR CDS (rev)	5'-TACGCGGCCGCACACGGCATTCTCGAGAGGCGC-3'	*Not* I	

Table 2.2. Oligonucleotide sequence of DHR primers. Some of the primers incorporated restriction endonuclease (RE) sites in order to facilitate downstream cloning, these sites are underlined and the type of the RE site is detailed in the table. Regions shown in bold type are relate to Kozak translation initiation element (G/A)NNATGG (Kozak, 1990), these were incorporated for oligonucletides used for construction of eukaryotic expression constructs.

2.8.2 RT-PCR Cloning of a DH-R Fragment

Degenerate primers were designed to amplify the *N.lugens* DHR (Section 3.2.3). The forward primer (DHR degenerate fwd) and reverse primer (DHR degenerate rev) were expected to amplify a cDNA fragment of around 390 bp (an estimation based on *Manduca sexta* DHR, P35464). The primers were used in a standard PCR reaction on *N.lugens* gut specific cDNA, at a final concentration of 1 μM. Reaction tubes were subjected to an initial hold of 94°C for 3 minutes followed by 40 cycles of: 94°C, 30 seconds; 55°C, 1 minute and 72°C, 1 minute. A final hold of 72°C for 7 minutes ensured full extension of PCR products. PCR products of the predicted size were cloned and sequenced using standard techniques.

2.8.3 RACE for Generation of Full Length DH-R cDNA

The full-length DHR cDNA was amplified from *N.lugens* gut specific cDNA using RACE (Rapid Amplification of cDNA Ends) experiments (Clontech). Gene specific primers (GSPs) were designed corresponding to regions of the *N.lugens* 381 bp DHR fragment. A sense primer (DHR 3' RACE) was designed for retrieval of the complete 3' end of the DHR; and an antisense primer (DHR 5' RACE) was designed for retrieval of the 5' end. These primers were used in separate RACE PCR reactions to amplify DHR cDNA between the GSP and a universal primer situated at the respective 5' end or 3' end of the RACE prepared cDNA. RACE PCR reactions were set up as suggested by the manufacturer and subjected to the following cycles: 3' RACE, an initial hold of 94°C for 3 minutes followed by 30 cycles of: 94°C for 5 seconds, 65°C for 10 seconds and 72°C for 2 min; and 5'RACE, an initial hold of 94°C for 3 minutes followed by 30 cycles at 94°C for 5 seconds, 63°C for 10 seconds and 72°C for 2 minutes. The 5' RACE

experiment produced multiple products so a 5 μl aliquot of this reaction was diluted with 254 μl Tris-HCl (10 mM, pH7.4) and reamplified using a nested universal primer (Clontech) and a nested GSP (DHR 5' RACE nested). The nested PCR was performed under the same conditions as the 5' RACE amplification, except the number of cycles was reduced to twenty. RACE products of the predicted size were cloned and sequenced, using standard procedures. Sequence data was aligned on a computer to produce a contiguous *N.lugens* DHR sequence.

2.8.4 Insect Expression Construct: Full Length DHR

The *N.lugens* DHR coding sequence is 1920 bp, which encodes a protein of 639 amino acids, based upon the longest open reading frame (ORF). Primers were designed to amplify the entire *N.lugens* DHR ORF. The forward primer incorporated an *Eco*R V restriction enzyme site while the reverse primer had a *Not* I site, these sites allowed directional cloning of the DHR ORF into the pIZ/V5/His insect expression vector (Invitrogen). The forward primer also incorporated a Kozak translation initiation sequence (G/A)NNATGG around the start codon (Kozak, 1990) and the reverse primer was designed to be in frame with the poly-His tag, V5 epitope and stop codon of the vector. The primer pair was used with a proof reading polymerase, Advantage2 polymerase (Clontech) on *N.lugens* gut specific cDNA, according to manufacturers instructions. Reaction tubes were subject to an initial hold at 94°C for 3 minutes followed by 35 cycles at 94°C for 30 sec, 60°C for 30 seconds and 72°C for 2 minutes. A 10 μl aliquot of the PCR reaction was removed upon completion of 20, 25, 30 and 35 cycles and the final extension of 72°C for 7 minutes was completed in a pre-heated water bath. PCR products of the correct size and present at the minimal number of

cycles were cloned and sequenced, using standard procedures. An error free clone was excised from the cloning vector by restriction with *Eco*R V and *Not* I in a standard reaction. *N.lugens* DHR cDNA was separated from the cloning vector by agarose gel electrophoresis and gel purified (Qiagen). The expression vector pIZ/V5/His was prepared by following the same procedure. Both insert and vector were quantified by spectrophotometric methods. A ligation reaction containing the excised DHR coding sequence and pIZ/V5/His was set up using a 1 : 1 (vector : insert) ratio, following standard procedures. After transformation clones were selected on low salt LB-Zeocin (50 μg/ml) plates and independent clones were sequenced over the 3' end to check the coding sequence was in frame with the poly-His tag and stop codon from the vector. The expression construct was used to Sf21 insect cells (Invitrogen).

2.8.5 Bacteria Expression Construct: DHR N-Terminal Domain

The TMHMM server, version 2.0; http://www.cbs.dtu.dk/services/TMHMM/ (Krogh *et al.*, 2001) was used to predict transmembrane topology of the putative *N.lugens* DHR protein. The extracellular N-terminal domain was predicted as the N-terminal 260 amino acids of the protein. A forward primer (DHR N-term fwd) and reverse primer (DHR N-term rev) were designed to amplify the 780 bp coding sequence of the *N.lugens* DHR N-terminal domain. The forward primer incorporated a *Nde* I restriction enzyme site while the reverse primer had an *Xho* I site, these sites allowed directional cloning of the DHR N-terminal domain into the pET24(a) expression vector (Novagen). The forward primer was designed to incorporate a start codon and the reverse primer was in frame with the poly-His tag and stop codon of the pET24(a) vector. The primer pair was used with a proof reading polymerase, Advantage 2 polymerase (Clontech) on *N.lugens* gut specific

cDNA, according to manufacturers instructions. Reaction tubes were subject to an initial hold at 94°C for 3 minutes followed by 30 cycles at 94°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 minute. A final hold at 72°C for 7 minutes ensured full extension of PCR products. Products of the predicted size were cloned and several independent clones were completely sequenced, an error free clone was used for all subsequent steps. The bacterial expression construct was prepared by following the procedure detailed in Section 2.8.4, except the coding sequence and the vector were prepared by restriction with *Nde* I and *Xho* I, prior to ligation. After transformation clones were selected on LB-Kanamycin (50 μg/ml) plates and independent clones were sequenced over the 3' end to check the coding sequence was in frame with the poly-His tag and stop codon from the vector. The expression construct was used to transform *E.coli* BL21(DE3) STAR cells (Invitrogen).

2.8.6 Over Expression and Purification of the N.lugens *DHR N-terminal Domain*

The DHR N-terminal domain construct was used to transform BL21 (DE3) STAR cells (Invitrogen). Initial small-scale expression experiments were used to assess the accumulation and solubility of the recombinant DHR N-terminal domain protein. A single clone was used to inoculate 50 ml of LB-kanamycin (50 μg/ml) and cells were grown in shake flasks too mid-log phase (OD λ 600 nm = 0.6 - 0.8). Cells were induced for expression with 1 mM IPTG and 1 ml culture samples were taken at time zero and 1, 2 , 3 hours post induction. Cells from all samples were pelleted by centrifugation (10,000 x *g* for 5 minutes, 4°C) and the supernatant was removed, cells from time zero, 1 and 2 hours post induction were resuspended in 1X sample loading buffer (7.5% (v/v) glycerol, 1% (w/v) SDS, 0.003% (w/v) bromophenol blue, 50 mM DTT, 40 mM Tris-

HCl, pH 6.8). Cells harvested from the 3 hour harvest were resuspended in 20mM Tris-HCl (pH 7.5) and cells were lysed by lysozyme treatment (100 μg/ml at 30°C for 15 minutes) and sonication. Soluble and insoluble cytoplasmic fractions were separated by centrifugation (14,000 x *g* for 10 minutes, 4°C) and each fraction was prepared for SDS-PAGE analysis. Cell lysates were analysed by SDS-PAGE and gels were stained with Coomassie Brilliant Blue G250. Following confirmation of expression level and determination of protein solubility the culture size was increased to 2 l for purification of the recombinant DHR N-terminal region protein. Large volume cultures were grown in shake flasks and inoculated with 10% of the culture volume with a starter culture grown to OD λ 600 nm $\leq$ 0.5. At three hours post induction cells were harvested by centrifugation (10,000 x *g* for 30 minutes, 4°C) and recombinant DHR N-terminal domain protein was purified from cells under denaturing conditions and refolded according to standard procedures (Section 2.7.6). Purified recombinant protein was stored freeze-dried at 4°C and resuspended in 10 mM Tris-HCl (pH 7.5) when required.

2.8.7 anti-DHR Antibodies

Anti-DHR antibodies were raised in rabbits according to standard protocols by Dr Andrew Dinsmore (Syngenta CTL, Alderley Park). At least two independent rabbits were used for each immunization. Antibodies were raised against synthetic peptides conjugated to BSA, and purified recombinant DHR N-terminal domain. The synthetic peptide corresponded to a region on N-terminal domain. The peptide read YTSCKDLSPDQPDLEPGIEVTTM (amino acid 238-260, N-terminal domain). After the final bleed the sensitivity of the antibodies raised against the N-terminal region were tested in western blot experiments against a dilution series of recombinant DHR N-

terminal protein (1.4 ng, 14 ng and 1400 ng). After demonstrating specificity against recombinant protein the IgG fraction was purified from the rabbit serum by affinity chromatography on immobilised protein G, following standard procedures (Section 2.7.11). For western blots purified IgG anti-peptide antibodies were used at a dilution of 1 : 2,000 and the antibodies raised against recombinant protein were used at a dilution of 1 : 10,000. Specific interactions between with *N.lugens* DHR and purified IgG antibodies was demonstrated by western blot against protein extracts from *N.lugens* gut tissue. A protein extract from *N.lugens* guts was prepared (Section 2.7.5) and a volume equivalent to 20 *N.lugens* guts was loaded per lane on a SDS-PAGE gel. Gels were transferred to nitrocellulose membranes and used in western blot experiments using anti-DHR antibodies at their optimal dilution.

2.8.8 anti-DHR IgG Feeding Trials

Antibodies which showed a specific interaction with DHR from *N.lugens* gut protein extract were used in artificial diet feeding trials. Antibodies were incorporated into artificial diet MMD-1 (Mitsuhashi, 1974) at a final concentration of 1 mg/ml, equivalent to 0.1 % (w/v). Diet was replaced every other day and insect survival was recorded daily. During the feeding trial samples of honeydew were collected from the feeding chamber, and after the feeding trial was complete haemolymph samples were collected from surviving insects. For haemolymph collection the thorax of the insect was pierced and the exuding haemolymph from 25 insects was collected in a fine glass capillary. Both honeydew and haemolymph samples were prepared for SDS-PAGE analysis. The honeydew sample was used as a positive control to demonstrate that *N.lugens* were feeding on the artificial diet and ingesting IgG antibodies. The haemolymph samples

were used to determine whether IgG antibodies were capable of crossing the gut wall into the haemolymph, as had been previously reported in the whitefly *Bemisia tabaci* (Morin *et al.*, 1999). Honeydew and haemolymph samples were separated by SDS-PAGE and transferred to nitrocellulose membranes, following standard procedures. Protein blots were incubated with goat anti-rabbit IgG horseradish peroxidase conjugate at 1 : 3,000 dilution. Specifically bound secondary antibodies were detected with ECL reagents and exposure of the membrane to x-ray film (Section 2.4.7).

2.8.9 Immunolocalisation of DHR in N.lugens *using Confocal Microscopy*

The interaction of the DHR antibody with the receptor protein was further investigated by immunolocalisation, using anti-recombinant DHR IgG and a fluorescent secondary antibody Alexa Fluor™ 488 goat anti-rabbit IgG (Molecular Probes). Antibodies were incubated with dissected *N.lugens* gut tissue *in vitro*, or antibodies were delivered *in vivo* by incorporation into liquid artificial diet. These experiments were used to determine the topology of *N.lugens* DHR protein, and whether antibodies delivered in the artificial diet were capable of interacting with the target protein.

2.8.9.1 Immunolocalisation : in vitro antibody incubations

Fifth instar *N.lugens* were collected and immobilised on ice. Full-length guts were dissected out in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10mM Na_2HPO_4 and 2mM KH_2PO_4, pH 7.5) with PMSF (0.2 mM), EDTA (1 mM) and leupeptin (20 μM). After dissection guts were transferred to a tube containing 2% (w/v) paraformaldehyde. Gut tissue was left to fix for 2 hours at room temperature. After fixation gut tissue was washed several times in PBS. Non-specific antibody binding was prevented by incubating gut tissue for 1 hour in a solution containing 4% (w/v) Triton

X-100 with 2% (w/v) bovine serum albumin (BSA) in PBS. After the blocking step the tissue was incubated in primary antibody sera. Anti-DHR was used at a concentration of 1:1000 in anti sera buffer (0.4% (v/v) Triton X-100 with 2% (w/v) BSA in PBS). This antibody incubation was left on a flat bed shaker at 4°C for 24 hours. Gut tissues were then washed in PBS at 4°C for 24 hours. Secondary antibody Alexa Fluor™ 488 goat anti-rabbit IgG (Molecular Probes) was incubated with the tissue at a concentration of 1:200 in antisera buffer at 4°C for 18 hours. The secondary antibody solution was removed and the tissue washed in PBS at 4°C for 18 hours. Just prior to mounting, the gut tissue was counterstained with ethidium bromide (0.5 μg/ml) for 30 minutes at 4°C and unbound stain was removed by several changes of PBS. Immunostained guts were mounted in Vectashield mounting medium (Vector Laboratories) on poly-L-lysine slides. Control experiments were run in parallel, which consisted of a primary antibody only control and a secondary antibody only control. Control experiments were set up following the same procedure, except the appropriate antibody incubation stage was omitted. Slides were observed on laser scanning microscope LSM510 META (Zeiss) under argon laser (λ 488 nm) to give the green fluorescence of localised secondary antibody, and under a helium neon laser (λ 543 nm) to give a tissue background from ethidium bromide counterstaining. Images from red and green channels were handled using LSM 5 Image Examiner (Version 3.1.0.117) run on a PC, and digitally overlaid to produce a final image.

2.8.9.2 Immunolocalisation : in vivo antibody incubations

Experiments were carried out with antibodies being administered to the insect *in vivo*, by incorporation in the artificial diet, so that only accessible DHR could be detected.

Insects (at least 5 replicates per treatment) were first fed on artificial diet containing anti-recombinant DHR IgG at a dilution of 1 : 1,000 for 8 hours, followed by feeding on artificial diet containing fluorescent labelled secondary antibody Alexa Fluor™ 488 goat anti-rabbit IgG (Molecular Probes), at a dilution of 1 : 2,000 for a further 8 hours. The insects were transferred to normal artificial diet without antibody for a further 8 hours, which was sufficient to wash out unbound antibody from the gut contents. Full-length guts were dissected out in PBS with PMSF (0.2 mM), EDTA (1 mM) and leupeptin (20 μM) and then were counterstained with ethidium bromide (0.5 μg/ml) in PBS. Control experiments were run in parallel, which consisted of a primary antibody only control and a secondary antibody only control. Control experiments were set up following the same procedure, except the appropriate antibody feeding step was omitted, and replaced with a diet only feeding step. Immunostained guts were mounted in Vectashield mounting medium (Vector Laboratories) on poly-L-lysine slides and observed on the confocal microscope, as previously described (Section 2.8.9.1).

2.8.10 N.lugens *Feeding Trials with Recombinant DHR N-terminal Domain*

N.lugens were fed artificial diet containing recombinant DHR N-terminal domain at a concentration of 10 mg/ml, equivalent to 1% (w/v). The feeding trial was set up according to standard procedures (Section 2.2), and insect survival was monitored daily.

2.9 ALSTR Methods

All cloning reactions used *N.lugens* gut specific cDNA, which was produced as previously described (Section 2.8.1). Oligonucleotide sequences of primers used to

clone the *N.lugens* ALSTR and to prepare *E.coli* and insect expression constructs are detailed in table 2.3.

2.9.1 RT-PCR Cloning of an ALSTR Fragment

Degenerate primers were designed to amplify the *N.lugens* ALSTR. The forward and reverse primers (ALSTR Degenerate, fwd and rev) were expected amplify a cDNA fragment of around 570 bp, while the nested primer pair (ALSTR Degenerate Nested, fwd and rev) were expected to produce a product of approx. 410 bp. ALSTR degenerate primers were initially optimised on a *Drosophila* cDNA (AF163775) encoding an ALSTR (*Drosophila* allatostatin receptor-1, DAR-1) (Birgul *et al.*, 1999). The clone was a kind gift from Dr Hans-Jürgen Kreienkamp, University of Hamburg. Standard 50 μl PCR on the *Drosophila* cDNA was set up using forward and reverse primers in every possible combination. Following PCR optimisation the degenerate primers were used in the same combinations on *N.lugens* gut specific cDNA. Negative control experiments were run in parallel, which consisted of identical PCR conditions but with the addition of only one primer and a control with both primers but omission of cDNA template. Reaction tubes were subjected to an initial hold of 94°C for 3 min followed by 35 cycles of: 94°C, 30 seconds; 48°C, 2 minutes; 72°C, 2 min. A final hold of 72°C for 10 minutes ensured full extension of PCR products. PCR products of the predicted size were cloned and sequenced using standard techniques.

2.9.2 RACE for Generation of Full Length ALSTR cDNA

The full-length ALSTR cDNA was amplified from *N.lugens* gut specific cDNA using RACE (Rapid Amplification of cDNA Ends) experiments (Clontech). Gene specific primers (GSPs) were designed corresponding to regions of the 587 bp ALSTR fragment.

	Name	Sequence (5'→3' Direction)	RE Site	Product (bp)
Degenerate Primers	ALSTR Degenerate (fwd)	5'-CTGGTICTIATGTC(C/T)ITIGA(C/T)(A/G)G-3'	-	588
	ALSTR Degenerate (rev)	5'-AA(G/A)T(T/G)ITC(G/C)GAIA(G/A)(G/A)AA(G/C)GC(G/A)TA-3'	-	
	ALSTR Degenerate Nested (rev)	5'-GG(G/A)TTIAT(G/C)(C/A)A(G/C)GA(G/A)(T/C)T-3'	-	531
	ALSTR Degenerate Nested (fwd)	5'-T(G/A)GCIGTIGTICA(T/C)CC(C/A)(G/A)T-3'	-	
RACE Primers	ALSTR 5' RACE	5'-GCCAACACGTGACTCACGATTTGCAC-3'	-	977 + 5' UTR
	ALSTR 3' RACE	5'-CTGAAGAGTGTGGACCGTTACGAGATCAC-3'	-	246 + 3' UTR
	ALSTR 3' RACE (nested)	5'-GTGCAAATCGTGAGTCACGTGTTGGC-3'	-	201 + 3' UTR
Construct Primers	ALSTR N-term (fwd)	5'-AGTTCATATGGCGGTACCCCCGAAC-3'	*Nde* I	177
	ALSTR N-term (rev)	5'-TATGCTCGAGTATGGATACGATCTTCTCCATC-3	*Xho* I	
	ALSTR CDS (fwd)	5'-GGATCCGCC**GCCATGG**CGGTACCCCCGAAC-3'	*Bam*H I	1152
	ALSTR CDS (rev)	5'-TCTAGATCACAAGATATCGTTTGTGCGAGT -3'	*Xba* I	

Table 2.3. Oligonucleotide sequence of ALSTR primers. Some of the primers incorporated restriction endonuclease (RE) sites in order to facilitate downstream cloning, these sites are underlined and the type of the RE site is detailed in the table. Regions shown in bold type are relate to Kozak translation initiation element (G/A)NNATGG (Kozak, 1990), these were incorporated for oligonucleotides used for construction of eukaryotic expression constructs.

A sense primer (ALSTR 3' RACE) was designed for retrieval of the complete 3' end of the ALSTR; and an antisense primer (ALSTR 5' RACE) was designed for retrieval of the 5' end. These primers were used in separate PCR reactions to amplify ALSTR cDNA between the GSP and a universal primer situated at the respective 5' end or 3' end of the RACE prepared cDNA. 3' and 5' RACE PCR reactions were set up as suggested by the manufacturer and subjected to an initial hold of 94°C for 3 minutes followed by 40 cycles of: 94°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 min. The 3' RACE experiment produced multiple products so a 5 μl aliquot of this reaction was diluted with 254 μl Tris-HCl (10 mM, pH7.4) and reamplified using a nested universal primer (Clontech) and a nested GSP (ALSTR 3' RACE nested). The nested PCR was performed under the same conditions, except the number of cycles was reduced to twenty. RACE products of the predicted size were cloned and sequenced, using standard procedures. Sequence data was aligned to produce a contiguous *N.lugens* ALSTR sequence.

2.9.3 Bacteria Expression Construct: ALSTR N-Terminal Domain

The membrane topology of the *N.lugens* ALSTR was predicted using the same methods as for the *N.lugens* DHR (Section 2.8.4). The N-terminal of the ALSTR (amino acid 1 – 59) was predicted as extracellular. A forward primer (ALSTR N-term fwd) and reverse primer (ALSTR N-term rev) were designed to amplify the 177 bp coding sequence of the *N.lugens* ALSTR N-terminal region. The forward primer incorporated a *Nde* I restriction enzyme site while the reverse primer had an *Xho* I site, these sites allowed directional cloning of the ALSTR N-terminus into the pET24(a) expression vector (Novagen). The forward primer was designed to incorporate a start codon and the

reverse primer was in frame with the poly-His tag and stop codon of the pET24(a) vector. The primer pair was used with a proof reading polymerase, Advantage 2 polymerase (Clontech) on *N.lugens* gut specific cDNA, according to manufacturers instructions. Reaction tubes were subject to an initial hold at 94°C for 3 minutes followed by 30 cycles at 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 45 seconds. A final hold at 72°C for 7 minutes ensured full extension of PCR products. PCR products of the predicted size were cloned and sequenced, using standard procedures. Several independent clones were completely sequenced and error free clones were used for all subsequent steps. An error free clone was excised from the cloning vector by restriction with *Nde* I and *Xho* I in a standard reaction. ALSTR N-terminus cDNA was separated from the cloning vector by agarose gel electrophoresis and gel purification (Qiagen). The expression vector pET24(a) was prepared by following the same procedure. Both insert and vector were quantified by spectrophotometric methods. A ligation reaction containing ALSTR N-terminus and pET24(a) was set up using a 1 : 1 (vector : insert) ratio, following standard procedures. After transformation clones were selected on LB plates with Kanamycin (50 μg/ml) plates and independent clones were sequenced over the 3' end to check the coding sequence was in frame with the poly-His tag and stop codon from the vector. The expression construct was used to transform expression cells RosettaBlue (DE3) (Novagen) and clones were selected on LB plates with kanamycin (50 μg/ml) and chloramphenicol (34 μg/ml).

2.9.4 Over Expression of the ALSTR N-Terminal Region

The ALSTR N-terminal expression construct was used to transform RosettaBlue DE3 (Novagen) and a single clone was tested for expression. Initial small-scale expression experiments were used to access the accumulation and solubility of the recombinant ALSTR N-terminal region. A single clone was used to inoculate 50 ml of LB-kanamycin (50 μg/ml), chloramphenicol (34 μg/ml) and cells were grown in shake flasks to mid-log phase (OD λ 600 nm = 0.6 - 0.8). Induced cells were left shaking (200 rpm) at 37°C for 3 hours prior to harvest. Cells were harvested by centrifugation (10,000 x *g* for 10 minutes, 4°C) and resuspended in 1/10 volume 10mM Tris-HCl (pH 7.5). Cells were lysed by treatment with chicken egg white lysozyme (100 μg/ml at 30°C for 15 minutes) and sonication. Soluble and insoluble fractions were separated by centrifugation (14,000 x *g* for 10 min, 4°C). The soluble supernatant was removed and mixed with an equal volume of 2X SDS sample buffer. The insoluble pellet was washed several times and resuspended in 1% SDS, and an equal volume of this insoluble fraction and 2X sample buffer was mixed. Samples were prepared for SDS-PAGE and equal amounts of soluble and insoluble material were loaded onto a 15% acrylamide gel and resolved. Gels were run in duplicate, one gel was stained with Coomassie Brilliant Blue G250 and the other gel was transferred to nitrocellulose and probed with INDIA His probe-HRP (Pierce), according to manufacturers instructions.

2.9.5 anti-ALSTR Antibodies

Anti-ALSTR antibodies were raised in rabbits according to standard protocols by Dr Andrew Dinsmore (Syngenta CTL, Alderley Park). Two independent rabbits were used for each immunization. Antibodies were raised against synthetic peptides conjugated to

BSA. The synthetic peptides corresponded to a region on N-terminus and extracellular loops 2 and 3 of the full length *N.lugens* ALSTR. The peptides were ISTPEDNGDPDSFT (amino acid 38-51, N-terminal region); YDPVNRPDGHNKP (amino acid 217-288, extracellular loop 2) and LKSVDRYEITNTSVMVQIVSTCWQG (amino acid 303-328, extracellular loop 3). After the final bleed the sensitivity of the antibodies raised against the N-terminal domain was tested in western blot experiments against *N.lugens* gut proteins.

CHAPTER 3.

IDENTIFICATION OF *N.LUGENS* GUT SPECIFIC TARGET PROTEINS

3.1 Introduction

The identification of *N.lugens* gut specific protein targets used two different approaches. The first was a random sequencing approach using the high throughput sequencing facilities at Torrey Mesa Research Institute (TMRI), La Jolla, California, USA. In total 7680 independent clones were sequenced from the *N.lugens* whole insect cDNA library, under the supervision of Dr David Cotton. Independent clones in the pBluescript SK(-) vector were sequenced over their 5' and 3' end using universal T3 and T7 primers, respectively. Clones were annotated by BLAST homology searches against the GenBank database, and then compiled into an *N.lugens* EST (expressed sequence tag) database. Clones that were considered as potential 'target' proteins were completely sequenced, and missing 5' cDNA ends (as a result of incomplete cDNA extension) were retrieved by rapid amplification of cDNA ends (RACE) PCR experiments.

The second approach to identify 'target' proteins employed a directed approach. A literature review of the function and expression pattern of already identified insect proteins was carried out. Proteins that showed localisation to the gut of other insect species and played a crucial role during insect development, metabolism or homeostasis were considered as potential 'target' proteins. It was considered that the loss of function

of such a protein would be detrimental to insect survival. Proteins that met such criteria were retrieved from the GenBank database, along with the encoding nucleotide data. As well as published sequences the complete genomes of *Drosophila melanogaster* (http://flybase.bio.indiana.edu/) and *Anopheles gambiae* (http://www.ensembl.org/anopheles_gambiae) were searched for similar sequences using BLAST. Retrieved protein sequences were then aligned using the CLUSTAL V algorithm (Higgins and Sharp, 1989) in the MegAlign software package on a Macintosh computer. Regions of strong amino acid conservation were identified. Nucleotide sequence data encoding these well-conserved regions was manually aligned. Using these nucleotide alignments degenerate primers were designed with a bias towards insect codon usage (Nakamura *et al.*, 2000). Primers designed in this way were used in PCR amplification experiments to isolate an *N.lugens* homologue of a particular protein.

This chapter details the alignment of insect diuretic hormone receptor (DHR) and allatostatin receptor (ALSTR) proteins, and the design of low redundancy degenerate primers for the use in PCR cloning experiments. Verifications of the *N.lugens* cDNA library and the gut specific cDNA library and the identification of potential 'target' proteins arising from the random sequencing of the *N.lugens* cDNA library are also presented.

3.2 Results

3.2.1 Degenerate Primer Design

Primers were designed to have a minimal amount of degeneracy in order to increase the specificity of the PCR amplification. In all cases primer degeneracy was kept below 128-fold by a combination of allowing base mismatches and incorporating inosine residues. In order for primers to amplify a specific product base mismatches and inosine residues were kept away from the 3' end of the oligonucleotide. Also, whenever possible, to allow efficient primer extension, oligo dC or dG residues were preferentially incorporated at the 3' end of the primer. Due to the degenerate nature of the primers it was necessary to increase primer concentration in PCR reactions, so primers matching the target sequence were present at a sufficient concentration to allow amplification. Degenerate primers were added to PCR reactions at a concentration of 1.0 μM, instead of the standard 0.2 μM.

3.2.2 General GPCR Primers

Degenerate primers have previously been designed against conserved residues from TM 3 and TM 6 of many higher animal GPCR proteins (Libert *et al.*, 1989; Probst *et al.*, 1992). Such primers have been used to amplify GPCR sequences from numerous invertebrates; for example, a leucokinin-like peptide receptor from Southern cattle tick, *Boophilus microplus* (Holmes *et al.*, 2000); a neuropeptide Y-like receptor (Tensen *et al.*, 1998) and a leucokinin-like peptide receptor (Cox *et al.*, 1997) from the pond snail, *Lymnaea stagnalis*, have all been amplified using such primers. Therefore, primers described by (Libert *et al.*, 1989) were synthesised and used in PCR on *N.lugens* gut

specific cDNA (Fig. 3.1). After confirmation of the integrity of the cDNA (section 3.2.5.2) (Fig. 3.1, lane 1) a PCR reaction using previously published 'generic' GPCR primers (Libert *et al.*, 1989) was assembled. After 40 amplification cycles the products from single primer controls, forward only (Fig. 3.1, lane 3) and reverse only (Fig. 3.1, lane 4) as well as forward and reverse primer combinations (Fig. 3.1, lane 2) were analysed by agarose gel electrophoresis. Unfortunately, the forward primer resulted in high levels of non-specific amplification (Fig 3.1, lane 3), which was also apparent in the forward and reverse primer PCR (Fig. 3.1, lane 2). Therefore, the PCR amplification products were deemed non-specific. Nevertheless bands corresponding to those of the predicted size (150 bp – 850 bp) were purified from the gel, cloned and sequenced. As expected, from the 10 independent clones sequenced none had homology to GPCR sequences, confirming non-specific amplification. Further optimisation of the PCR amplification was attempted, but this did not result in a specific amplification (data not shown). Since the primers described by (Libert *et al.*, 1989) failed to amplify GPCR sequences from *N.lugens* gut specific cDNA an alternative methodology was employed for primer design.

Insect peptide receptors with homology to vertebrate calcitonin/secretin/cortiocotropin releasing factor family (Family B) of G-protein coupled receptors were retrieved from the databases, aligned and degenerate primers were designed (Section 3.2.3). Similarly, degenerate primers were also designed for insect receptor proteins with homology to mammalian somatostatin/galanin/opioid family GPCRs (Family A/Group V) (Section 3.2.4). It was anticipated that degenerate primers designed from such protein alignments

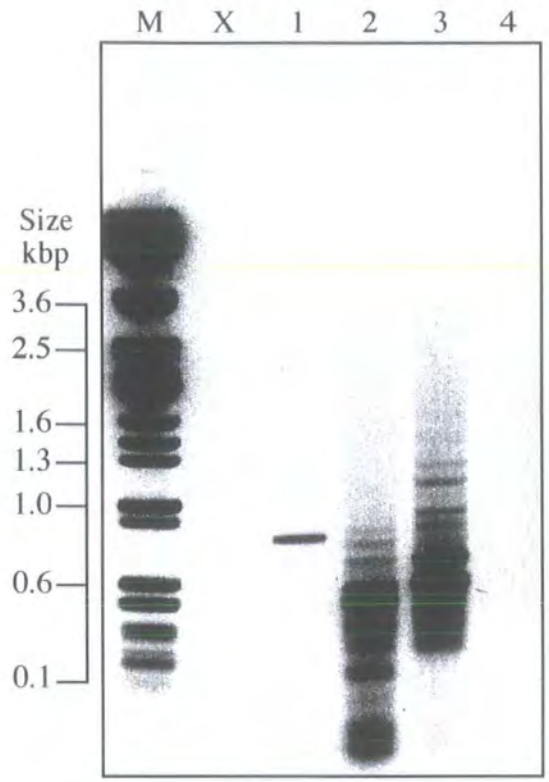


Figure 3.1. PCR amplification products from *N.lugens* gut specific cDNA using degenerate primers designed against conserved residues in TM3 and TM6 of many aligned GPCR sequences. Primer sequences and reaction conditions were as described by (Libert *et al.*, 1989). PCR products were resolved on a 1.2% agarose gel. M, DNA size marker; X, empty lane; 1, *N.lugens* cathepsin B-like protease coding sequence amplification; 2, forward and reverse primers; 3, forward primer only; 4, reverse primer only.

would result in the amplification of insect diuretic hormone receptor sequences and insect allatostatin receptor sequences, which are insect representatives of Family B and Family A GPCRs, respectively.

3.2.3 DHR Degenerate Primers

PCR amplification of a partial DHR coding sequence from *Bombyx mori* has been described using degenerate primers designed from the *Manduca sexta* DHR cDNA (Ha *et al.*, 2000). A similar set of degenerate primers was designed to amplify the *N.lugens* DHR. Except, primers were designed from all published insect DHR sequences (*Manduca sexta*, P35464; *Acheta domesticus* Q16983; and *Bombyx mori*), and related *Drosophila* family B peptide receptor proteins (Brody and Cravchik, 2000; Hewes and Taghert, 2001). *Drosophila* protein sequences, from Family B/Group I (Secretin like receptors) include gene products CG4395, CG17415, CG8422 and CG12370). Sequence CG4395 and CG17415 are closely related to vertebrate calcitonin receptors. In vertebrates calcitonin receptors are responsible for regulation of intracellular Ca^{2+} homeostasis. Two receptors encoded by CG8422 and CG12370 are more closely related to insect diuretic hormone receptors. Alignment of these sequences (Fig. 3.2) reveals conservation is greatest within the predicted transmembrane domains. A conserved region from TM 3 which read (T/S/V)N(F/Y)(F/S)WM(F/L)(V/C)EG was identified as a region of high similarity. Sequence similarity between any two sequences from TM 3 ranged from 50%-100%. A conserved region from TM 6 which read (Q/T/A)(Y/V/I)(R/L/Y)(K/Q)A(T/A/V/F/S)(R/K)A(L/T)(L/I)(L/I/V) was also identified as a region of high sequence similarity. Sequence similarity between any two sequences from TM 6 ranged from 27%-100%. Due to the relatively high degree of sequence

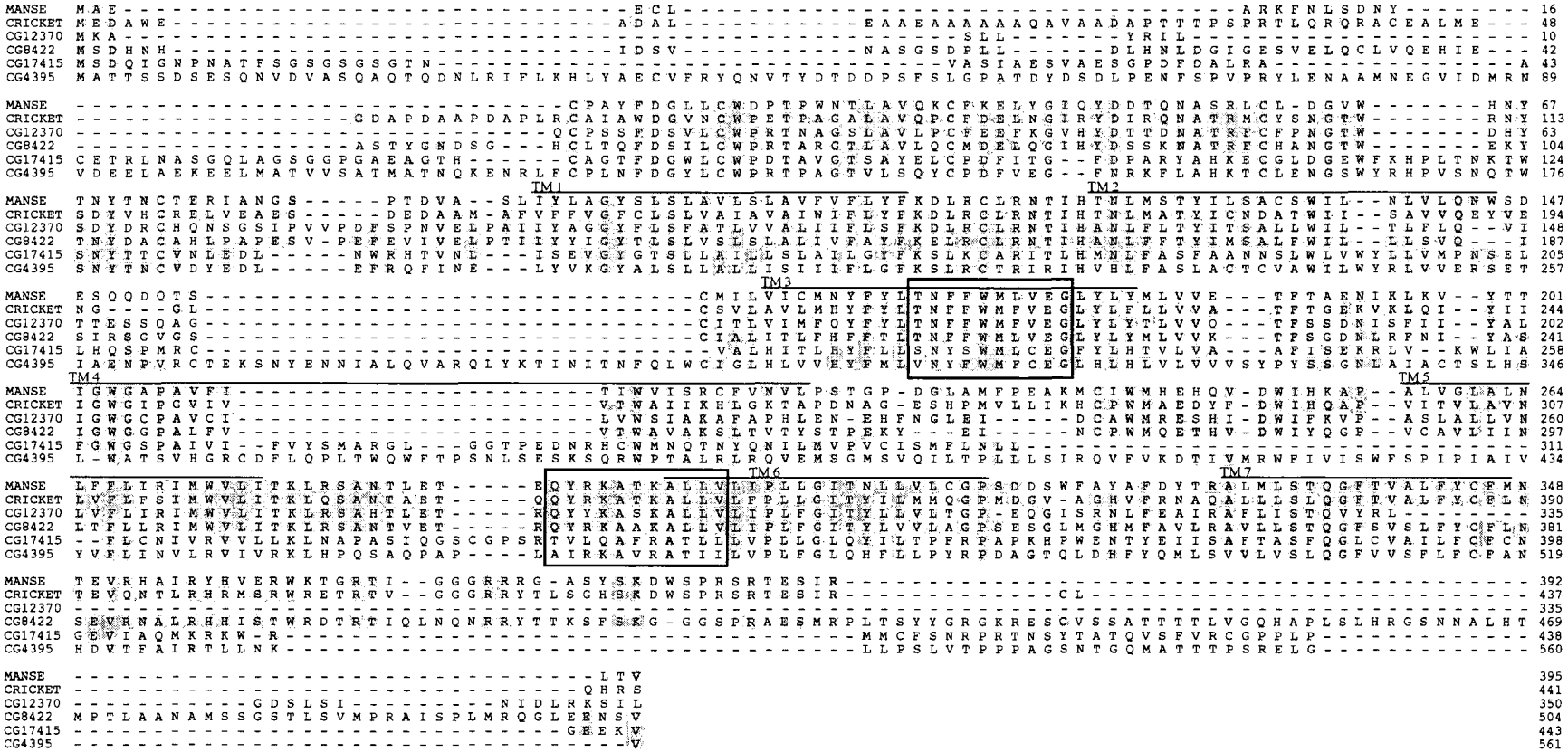


Figure 3.2. Alignment of *Drosophila* peptide receptor proteins from Family B/Group I (Calcitonin and Diuretic Hormone Receptors). Included in the alignment were *Drosophila* diuretic hormone receptor-like proteins (CG12370 and CG8422) and calcitonin receptor-like proteins (CG17415 and CG4395). Diuretic hormone receptors from *Manduca sexta* (MANSE, P35464) and *Acheta domesticus* (CRICKET, Q16983) were also included in the protein alignment. Shaded areas denote amino acids conserved between receptor proteins. Putative transmembrane domains (TM 1-7), as defined by hydropathy analysis, are indicated by lines above the MANSE sequence. The boxed areas show regions for which nucleotide sequence data was retrieved from the appropriate databases. Nucleotide sequence data was manually aligned (Fig. 3.3) and degenerate primer were designed with a bias towards codon use and diuretic hormone receptor sequences (Fig. 3.4). Protein alignment was performed using the Clustal V algorithm in the MegAlign software package.

conservation TM 3 and TM 6 were used to design a forward and reverse degenerate primer, respectively (Fig. 3.2).

Nucleotide sequence data from TM 3 and TM 6 for each receptor, was retrieved from the appropriate database and manually aligned (Fig. 3.3). Degenerate primers were designed from the aligned nucleotide data (Fig. 3.4). The overall degeneracy of each primer was minimised by giving an increased bias to preferentially used codons within the aligned sequences and sequences known to encode diuretic hormone receptors. The forward primer (DHR Degenerate fwd, Table 2.2) consists of a mixture of 64 different 29-mers, with three inosine residues. This primer allowed a match of 87% or better with any of the aligned receptor sequences. Whereas, the reverse primer (DHR Degenerate rev, Table 2.2) consists of a mixture of 64 different 32-mers, with six inosine residues. This primer allowed a match of 67% or better with any of the aligned receptor sequences. These primers were designed to hybridise to opposite strands of the target sequence and allow amplification between TM 3 and TM 6 (Fig. 3.2). Based on aligned sequences amplification products using this primer pair were expected to be between 372 bp and 468 bp long.

3.2.4 ALST-R Degenerate Primer Design.

Drosophila protein sequences from Family A/Group V receptors (Galanin/Allatostatin and Opoid/Somatostatin receptors) were retrieved from the FlyBase database (http://flybase.bio.indiana.edu/). Sequence CG13702 and CG7285 are most similar to vertebrate opioid and somatostatin receptors, while ALSTR-1 (CG2872) and ALSTR-2 (CG10001) are most similar to vertebrate galanin receptors. Additional sequences

Family B/Group I; 5′ Primer

MANSE	ACT	AAT	TTC	TTC	TGG	ATG	TTG	GTG	GAA	GGT
CRICKET	ACC	AAC	TTC	TTC	TGG	ATG	TTC	GTT	GAA	GGT
CG12370	ACC	AAC	TTT	TTC	TGG	ATG	TTT	GTG	GAG	GGC
CG8442	ACC	AAC	TTC	TTT	TGG	ATG	CTG	GTC	GAG	GGT
CG17415	**T**CC	AAT	**T**AC	**T**CC	TGG	ATG	CTC	**T**GC	GAG	GGA
CG4395	**G**TC	AAC	**T**AT	TTC	TGG	ATG	TTT	**T**GC	GAG	GGC
	T	N	Y	F	W	M	F	C	E	G
	S		F	S			L	V		
	V									

Family B/Group I; 3′ Primer

MANSE	CAA	TAC	CGG	AAG	GCT	ACC	AAA	GCG	CTT	CTA	GTC
CRICKET	CAG	TAC	CGC	AAG	GCG	ACG	AAG	GCG	CTG	CTG	GTG
CG12370	CAG	TAT	**T**AC	AAG	GCC	**T**CG	AAG	GCG	CTG	CTG	GTG
CG8442	CAG	TAT	AGG	AAA	GCT	GCT	AAG	GCA	CTG	CTG	GTG
CG17415	**A**CG	**G**TT	**T**TG	**C**AA	GCA	**T**TT	**CG**G	GCA	**A**CG	CTG	CTT
CG4395	**GC**C	**A**TC	CGG	AAG	GCG	GT**T**	**CG**G	GCC	**AC**C	**A**TT	**A**TT
	Q	Y	Y	K	A	S	K	A	L	L	V
	T	V	L	Q		F	R		T	I	L
	A	I	R			V					I
						A					
						T					

Figure 3.3. Alignment of nucleotides corresponding to the regions of conserved amino acids from *Drosophila* peptide receptor proteins from Family B/Group I (Calcitonin and Diuretic Hormone Receptors). Included in the alignment were *Drosophila* diuretic hormone receptor-like proteins (CG8422 and CG12370) and calcitonin receptor-like proteins (CG4395 and CG17415). Diuretic hormone receptors from *Manduca sexta* (MANSE, P35464) and *Acheta domesticus* (CRICKET, Q16983) were also included in the nucleotide alignment. The aligned nucleotide sequences allowed primers to be designed that reflected codon usage, thereby reducing the overall degeneracy. Nucleotide positions that are not represented in the primer sequences are shown in bold type.

Family B/Group I Primers

5′ PRIMER

CONSERVED AMINO ACIDS

V									
S		Y	S			L	C		
T	N	F	F	W	M	F	V	E	G

```
ACC AAC TTC TTC TGG ATG CTC GGC GAA GG      =    5′-ACI AAC TTC TTC TGG ATG CTI GGI GAA GG-3′
  T   T   T   T         T G  TG   G                      T   T   T         T    T    G
                          T   T
```

3′ PRIMER

CONSERVED AMINO ACIDS

					T					
					A					
A	I	R			V					I
T	V	L	Q		F	R		T	I	L
Q	Y	Y	K	A	S	K	A	L	L	V

```
CAA TAC AGC AAA GCA GCC AAG GCA CTC CTA GT   =    5′-ACI AGI AGI GCC TTI GCI GCC TTI CGA TAC TG-3′
  G   T C G   G   C A G   A   C   G   G                    T      T    T      TG   T
                  G   T       G   T   T
                  T
```

Figure 3.4. Degenerate primers were designed from conserved amino acids from *Drosophila* Family B/Group I receptors; Calcitonin receptors (CG17415 and CG4395) and Diuretic Hormone Receptors (CG12370, CG8422), plus diuretic hormone receptors from *Manduca sexta* (P35464) and *Acheta domesticus* (Q16983). Protein sequences were aligned (Fig. 3.2) and TM 3 and TM 6 regions showed strong amino acid conservation. Nucleotide sequences corresponding to these regions was retrieved and manually aligned (Fig. 3.3). Using the nucleotide sequence data a degenerate forward primer and reverse primer was designed, corresponding to TM 3 and TM 6, respectively. Degenerate primers were designed with a bias towards codon usage and diuretic hormone receptor sequences, inosine residues (I) were used to reduce the overall degeneracy of the primers. Based on aligned sequences amplification products using this primer pair would be predicted to be 372-468 bp.

encoding ALSTRs from *Bombyx mori* (AF254742) and *Periplaneta americana* (AF336364) were retrieved from the GenBank database (http://www.ncbi.nlm.gov) and included in the protein alignment (Fig. 3.5).

From the alignment of protein sequences from *Drosophila* Family A/Group V Receptors and two other insect allatostatin receptors it is apparent that sequence conservation is strongest within the predicted transmembrane domains. A conserved region from TM 3 which read L(V/L)(L/I)MS(A/I/F/L)DR(F/Y)(L/M/I)AV(V/C)HP(V/I) was identified as a highly conserved region. Sequence similarity between any two sequences from TM 3 ranged from 69%-94%. A conserved region from TM 7 which read (N/S)S(C/A)(V/I)NP(I/L/V)(L/I)Y(A/D)(F/I)LS(E/D)(N/H)FR(V/K)(A/S)F was also identified as a region of high sequence similarity. Sequence similarity between any two sequences from TM 7 ranged from 80%-90%. Due to the high degree of sequence conservation TM 3 and TM 7 were used to design a forward and reverse degenerate primer, respectively (Fig. 3.5). Nucleotide sequence data for each receptor, corresponding to TM 3 and TM 7 was retrieved and manually aligned (Fig. 3.6). Due to large conserved regions there was sufficient run of conserved nucleotides to allow the design of a 5' primer, and a nested 5' primer; and similarly a 3' primer, and a nested 3' primer (Fig. 3.6). By using a nested PCR strategy the specificity of the amplification could be maximised.

In an effort to reduce the degeneracy of the primers, bias was given to preferred codons and allatostatin receptor sequences (Fig. 3.7 and Fig. 3.8). The forward primer,

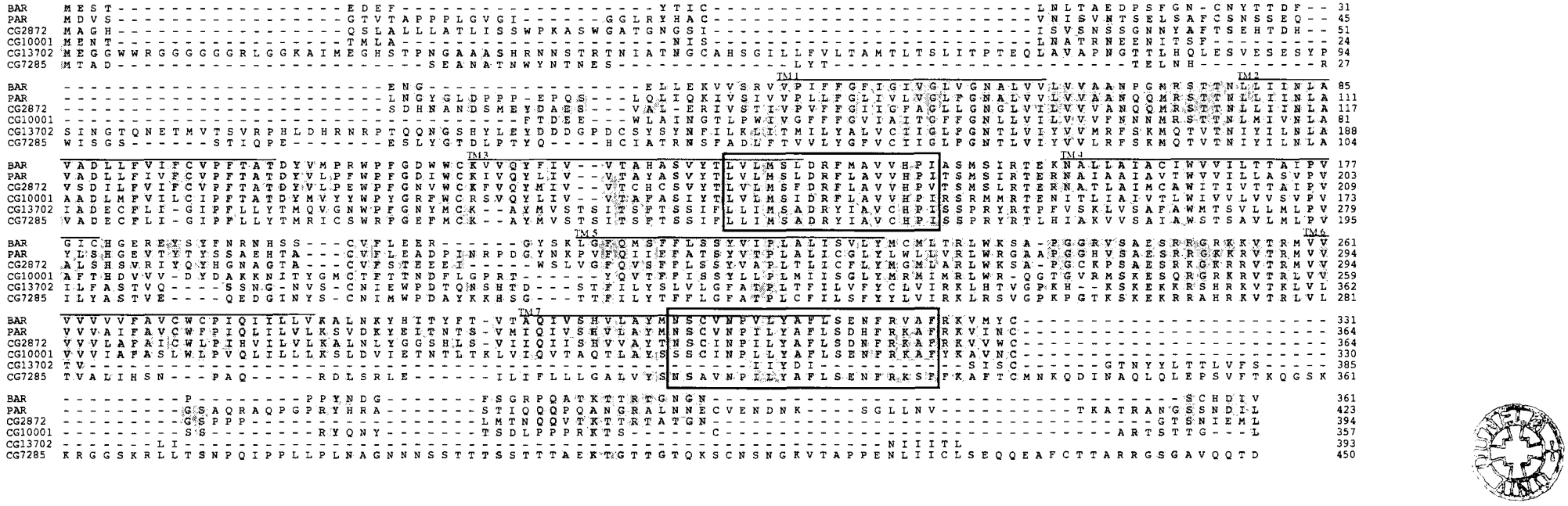


Figure 3.5. Alignment of *Drosophila* peptide receptor proteins from Family A/Group V (Galanin/Allatostatin and Opioid/Somatostatin Receptors). Included in the alignment were *Drosophila* allatostatin receptor-like proteins (CG2872 and CG10001) and somatostatin receptor-like proteins (CG13702 and CG7285). Allatostatin receptors from *Bombyx mori* (BAR, AF254742) and *Periplaneta americana* (PAR, AF336364) were also included in the protein alignment. Shaded areas denote amino acids conserved between receptor proteins. Putative transmembrane domains (TM 1-7), as defined by hydropathy analysis, are indicated by lines above the BAR sequence. The boxed areas show regions for which nucleotide sequence data was retrieved from the appropriate databases. Nucleotide sequence data was manually aligned (Fig. 3.6) and degenerate primer were designed with a bias towards codon use and allatostatin receptor sequences (Fig. 3.7 and 3.8). Protein alignment was performed using the Clustal V algorithm in the MegAlign software package.

corresponding to TM 3 (ALSTR Degenerate fwd, table 2.3) consists of a mixture of 8 different 23-mers, with four inosine residues. This primer allowed a match of 87% or better with any of the aligned receptor sequences. The nested forward primer, again corresponding to TM 3 (ALSTR Degenerate Nested fwd, table 2.3) consists of a mixture of 16 different 19-mers, with 3 inosine residues. This primer allowed a match of 84% or better with any of the aligned receptor sequences. The reverse primer, corresponding to TM 7 (ALSTR Degenerate rev, table 2.3) consists of a mixture of 128 different 23-mers, with two inosine residues. This primer allowed a match of 87% or better with any of the aligned receptor sequences. The nested reverse primer, again corresponding to TM 7 (ALSTR Degenerate Nested rev, table 2.3) consists of a mixture of 64 different 17-mers, with one inosine residue. This primer allowed a match of 82% or better with any of the aligned receptor sequences. The primers were designed to hybridise to opposite strands of the target sequence and allow amplification between TM 3 and TM 7 (Fig. 3.5). Based on aligned sequences amplification products using this primer pair were expected to be between 564 bp and 600 bp.

3.2.5 Verification of N.lugens Gut Specific cDNA Library

3.2.5.1 Insert Size Distribution

The size distribution of cDNA inserts within the excised pTriplEx2 vector was checked by PCR amplification across the multiple cloning site (MCS), using primers supplied by the manufacturer (Clontech). Ten excised clones were chosen at random and plasmid DNA was prepared and used in standard PCR amplification. Inserts were distributed between 800 bp and 1900 bp (Fig. 3.9), which was deemed representative for cDNAs.

Family A/Group V; 5′ PRIMER

	5′PRIMER(#1)								5′PRIMER(#2)						
BAR	CTG	GTA	CTC	ATG	TCC	TTG	GAC	AGA	TTC	ATG	GCA	GTA	GTG	CAC	CCA	ATA
PAR	CT**C**	GTT	CTC	ATG	TCT	CTG	GAT	CGA	TTC	TTG	GCT	GTG	GTT	CAC	CCC	ATC
CG2872	CTG	GTG	CTG	ATG	TCC	TTT	GAT	CGC	TTC	CTG	GCC	GTC	GTT	CAT	CCC	GTG
CG10001	CTG	GTG	CTA	ATG	TCC	ATC	GAT	CGG	TTC	CTG	GCG	GTG	GTT	CAT	CCC	ATT
CG13702	C**GT**	**T**TG	**A**TC	ATG	TC**G**	**GC**G	GAT	CGC	TAC	ATA	GCC	GTT	**TG**C	CAT	CCC	ATA
CG7285	CTG	CTC	**A**TC	ATG	TCC	G**C**G	**C**AT	CGA	TAT	ATA	GCG	GTA	**TG**C	CAC	CC**G**	ATT
	L	V	L	M	S	I	D	R	F	L	A	V	V	H	P	I
		L	I			A			Y	I			C			V
						F				M						
						L										

Family A/Group V; 3′ PRIMER

	3′PRIMER(#2)									3′PRIMER(#1)												
BAR	AAC	**AG**C	TGC	GTC	AAC	CCG	GTG	CTC	TAC	GCG	TTC	CTC	TCG	GAA	AAC	TTC	CGG	AAG	AGC	TTC
PAR	AAC	TCG	TGC	GTC	AAT	CCC	ATC	CTG	TAC	GCC	TTT	CTG	TCG	GAC	CAC	TTC	CGC	AAA	GCC	TTC
CG2872	AAT	TCG	TGC	ATC	AAT	CCG	ATA	CGT	TAT	GCC	TTT	CTA	TCC	GAC	AAC	TTT	CGC	AAG	GCA	TTC
CG10001	AGC	TCG	TGT	ATC	AAT	CCG	CTG	CTC	TAC	GCC	TTC	CTC	TCC	GAG	AAT	TTC	CGG	AAG	GCC	TTC
CG13702	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CG7285	AAT	TCG	**GCG**	GTG	AAT	CCC	ATA	CTT	TAT	GCC	TTC	CTA	**AGT**	GAG	AAC	TTC	CGG	AAG	AGC	TTC
	N	S	C	I	N	P	I	L	Y	A	F	L	S	D	N	F	R	K	A	F
	S		A	V			V	I		D	I			E	H			V	S	
							L													

Figure 3.6. Alignment of nucleotides corresponding to the regions of conserved amino acids from *Drosophila* peptide receptor proteins from Family A/Group V (Galanin/Allatostatin and Opioid/Somatostatin Receptors). Included in the alignment were *Drosophila* allatostatin receptor-like proteins (CG10001 and CG2872) and somatostatin receptor-like proteins (CG7285 and CG13702). Allatostatin receptors from *Bombyx mori* (BAR, AF254742) and *Periplaneta americana* (PAR, AF336364) were also included in the nucleotide alignment. The aligned nucleotide sequences allowed primers to be designed that reflected codon usage, thereby reducing the overall degeneracy. Nucleotide positions that are not represented in the primer sequences are shown in bold type.

Family A/Group V Primers

5′ PRIMER #1

CONSERVED AMINO ACIDS

					L		
					F		
	L	I			A		
L	V	L	M	S	I	D	R
CTG	GTA	CTC	ATG	TCC	ATC	GAC	AG
	C	G		T	C G	T C	
	G	A			G T		
	T				T		

= 5′-C T G G T I C T I A T G T C C I T I G A C A G-3′

T T C

23-MER, 8X (19bp fit, 4 inosine)

3′ PRIMER #1

CONSERVED AMINO ACIDS

	D	I			E	H	
Y	A	F	L	S	D	N	F
TAC	GCC	TTC	CTA	TCC	GAA	AAC	TT
T	G	T T	C	G	C	C T	
			G		G		

= 5′-A A G T T I T C G G A I A G G A A G G C G T A-3′

A G C A A C A

23-MER, 128X (21bp fit, 2 inosine)

Figure 3.7. Degenerate primers were designed from conserved amino acids from *Drosophila* Family A/Group V receptors; Galanin/Allatostatin receptors (CG2872 and CG10001) and Opioid/Somatostatin Receptors (CG13702 and CG7285), plus allatostatin receptors from *Bombyx mori* (AF254742) and *Periplaneta americana* (AF336364). Protein sequences were aligned (Fig. 3.5) and TM 3 and TM 7 showed strong sequence conservation. Nucleotide sequences corresponding to these regions was retrieved and manually aligned (Fig. 3.6). Using the nucleotide sequence data a degenerate forward primer and reverse primer was designed, corresponding to TM 3 and TM 7, respectively. Degenerate primers were designed with a bias towards codon usage and allatostatin receptor sequences. Based on aligned sequences amplification products using this primer pair would be predicted to be 564-600 bp.

Family A/Group V Nested Primers

5' PRIMER #2 (NESTED)

```
        CONSERVED AMINO ACIDS
      ┌──────────────────────┐
      │ M                    │
      │ I         C          V │
      │ L  A   V  V   H   P  I │
      └──────────────────────┘
TG GCA GTA GTC CAT CCC GT        =  5'-T G G C I G T I G T I C A T C C C G T-3'
 A  C   C   G   C   A A                A                       C     A A
    G   G   T                                  19-MER, 16X (16bp fit, 3 inosine)
    T   T
```

3' PRIMER #2 (NESTED)

```
        CONSERVED AMINO ACIDS
      ┌──────────────────────┐
      │ S      A   V         │
      │ N  S   C   I   N   P │
      └──────────────────────┘
AAC TCC TGC ATG AAC CC           =  5'-G G G T T I A T G C A G G A G T T-3'
 GT   G   T G C   T                     A         C A     C     A C
                                          17-MER, 64X (16bp fit, 1 inosine)
```

Figure 3.8. A pair of nested degenerate primers (nested with respect to 5' Primer # 1 and 3' primer #1) were also designed. Nested primers were used to reampliy cDNA in an effort to increase the specificity of PCR. Nested degenerate primers were designed from conserved amino acids from Drosophila Family A/Group V receptors; Galanin/Allatostatin receptors (CG2872 and CG10001) and Opioid/Somatostatin Receptors (CG13702 and CG7285), plus allatostatin receptors from *Bombyx mori* (AF254742) and *Periplaneta americana* (AF336364). Protein sequences were aligned (Fig. 3.5) and TM 3 and TM 7 showed strong sequence conservation. Nucleotide sequences corresponding to these regions was retrieved and manually aligned (Fig. 3.6). Using the nucleotide sequence data a degenerate forward primer and reverse primer was designed, corresponding to TM 3 and TM 7, respectively. Degenerate primers were designed with a bias towards codon usage and allatostatin receptor sequences. Based on aligned sequences amplification products using this primer pair would be predicted to be 564-600 bp.

The presence of coding sequences in the library was checked by PCR using primers against known *N.lugens* cDNAs (section 3.2.5.2).

3.2.5.2 PCR and RT-PCR-Cathepsin and Trypsin

The *N.lugens* gut specific library was checked for the presence of coding cDNAs by PCR using primers directed against a known *N.lugens* Trypsin-like protease (AJ316142), and an *N.lugens* Cathepsin B-like protease (AJ316141). Both proteases have been cloned from *N.lugens* and are known to be gut specific (Foissac *et al.*, 2002). The primer pair for trypsin like protease was designed to amplify a 953 bp coding region of the gene. The forward primer was 5'-TGGGTGCGTCTGTGTG-3' and the reverse primer read 5'-GTATTTATTAGGTATATAC-3'. The primer pair for cathepsin-B like protease was designed to amplify a 1108 bp coding region of the gene. The forward primer was 5'-GCTATTAGTGCACTTCCAGA-3' and the reverse primer read 5'-TTTCTTGTATTTGGGCAG-3'. Separate reactions were set up on both freshly prepared cDNA and a phenol chloroform extract from the packaged *N.lugens* gut specific λTriplEx2 cDNA library. PCR were performed according to standard conditions. The amplification of a cathepsin B-like protease and a trypsin-like protease from both freshly prepared cDNA (Fig. 3.10; lane 2 and 4, respectively) and the *N.lugens* gut library extract (Fig. 3.10; lane 1 and 3, respectively) confirmed the presence of coding inserts within the constructed library. All amplification products were checked by sequencing, and were shown to be an exact match to the *N.lugens* protease sequences reported in the GenBank database. The presence these coding regions in the RNA transcripts, from which the library was based, and in the constructed λTriplEx2 cDNA library verifies the integrity of both the cDNA and the cDNA library.

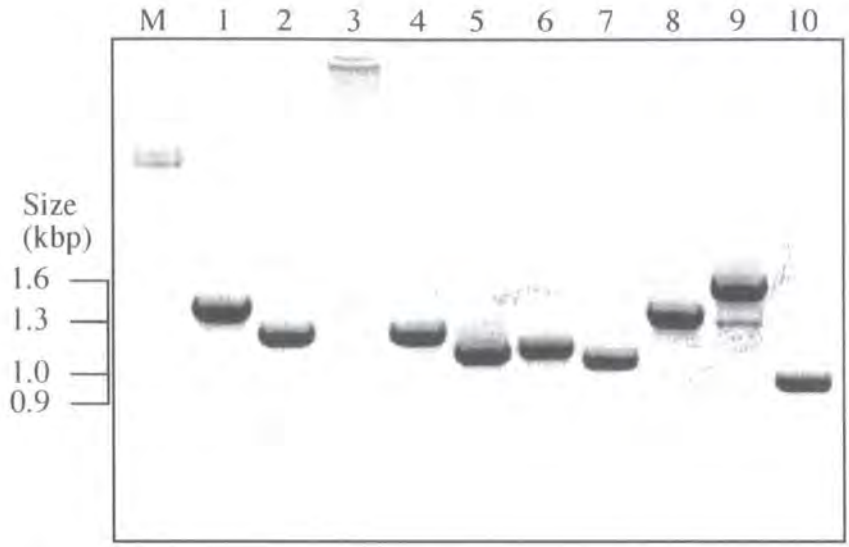


Figure 3.9. Size distribution of cDNA inserts within the *N.lugens* pTriplEx2 gut specific library. After mass excision 10 clones were picked at random and size distribution of the inserts were determined by PCR across the MCS using 5' and 3' sequencing primers (Clontech).

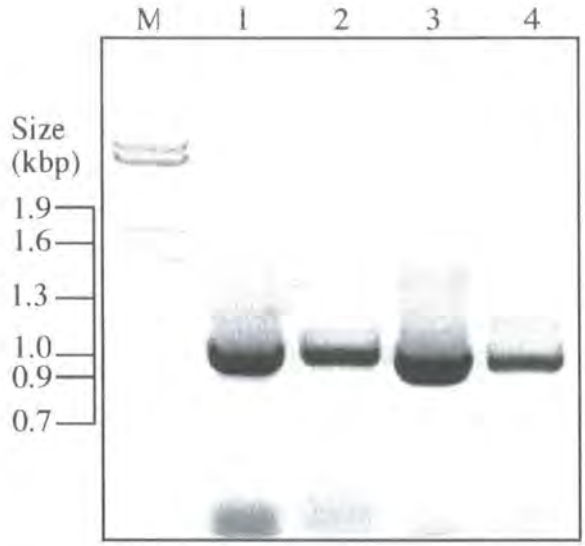


Figure 3.10. PCR on *N.lugens* gut specific λTrplEx2 cDNA library and *N.lugens* gut cDNA using primers directed against the coding sequence (CDS) of *N.lugens* Trypsin-like proteases cDNA (AJ316142); and a *N.lugens* cathepsin B-like protease (AJ316141). Using the Cathepsin B-like protease primers PCR was performed on both packaged *N.lugens* gut specific cDNA library (lane 1) and freshly prepared cDNA (lane 2). The amplification product was expected to be 1108 bp. The PCR product generated from library PCR was confirmed by sequencing to be a *N.lugens* Cathepsin-B like protease. Using the Trypsin primers PCR was performed on packaged *N.lugens* gut specific cDNA library (lane 3) and freshly prepared cDNA (lane 4). The expected amplification product was 953 bp. The DNA products generated from the library PCR were sequenced and shown to match the *N.lugens* cathepsin B-like protease cDNA and the *N.lugens* trypsin-like protease cDNA.

Both the 5' trypsin and cathepsin primers were designed to be at the extreme 5' end of the respective cDNAs. Due to the presence of products it is apparent at the cDNA synthesis stage extension occurred across the full length of the cDNAs; and it is therefore likely that other full-length clones are present in the library.

3.2.5.3 Excision and Mini-Preparation of N.lugens gut *λTriplEx2 cDNA library*

The release of a λTriplEx2 clone to a pTriplEx2 plasmid involves *in vivo* excision and circularisation of a complete plasmid from the recombinant phage. This conversion takes place in BM25.8 cells and successful conversion can be selected by the ability of BM25.8 to grow in the presence of carbenicillin (50 μg/ml). Ten recombinant phage were chosen randomly, converted to pTriplEx2 in BM25.8 cells and transformed cells were selected on LB agar with carbenicillin (50 μg/ml). Independent clones were picked and grown overnight in LB with carbenicillin (50 μg/ml). Alkaline lysis mini-preparation of up to 10 ml of overnight culture of transfected BM25.8 cells resulted in poor recovery and low purity plasmid DNA, which was unsuitable for downstream DNA sequencing or restriction analysis (data not shown). However, when pTriplEx2 plasmid was transferred to DH5-α cells, overnight growth and plasmid preparation was successful. The yield of plasmid DNA was sufficient to allow restriction analysis of the inserts and DNA sequencing. This suggests that there is an incompatibility problem between the pTriplEx2 plasmid and the BM25.8 bacterial strain used for *in vivo* excision. Clontech technical help was contacted but they could not offer a solution to this problem.

3.2.6 Verification of N.lugens cDNA Library

The *N.lugens* insect cDNA library prepared by Dr Jinping Du has been used in previous cloning experiments and has been shown to contain valid coding sequences (Du *et al.*, 2000). Therefore no further verification of this library was necessary.

3.2.7 Mass Sequencing From N.lugens cDNA Library

Due to the low yield of plasmid DNA following *in vivo* conversion of a λTriplEx2 clone into pTriplEx2 in BM25.8 cells (section 3.2.5.3), mass sequencing of the gut specific library was not a viable option. Therefore the whole *N.lugens* library was used for the EST sequencing project. Of the clones sequenced only a single clone was of the GPCR type. BLAST similarity searches revealed that the clone had a 74% sequence identity with a rhodopsin receptor protein from the Carpenter ant, *Camponotus abdominalis* (Q17292). The sequence of which is shown in Fig 3.11.

3.3 Discussion

Due to their central role in cellular regulation GPCRs have proven among the most successful drug targets in higher animals (Howard *et al.*, 2001), and thus are likely to yield important targets for the control of insect species (VandenBroeck *et al.*, 1997). Therefore GPCR proteins were targeted in the insect pest *N.lugens* as candidates for the design of protein antagonists or agonists, which when ingested may be detrimental to insect survival. The identification of such receptor proteins took two approaches, namely a random sequencing approach, and a PCR based approach using degenerate primers. However, for the discovery of potential 'target' proteins both approaches have advantages and disadvantages, as will be discussed in the following section. Due to the

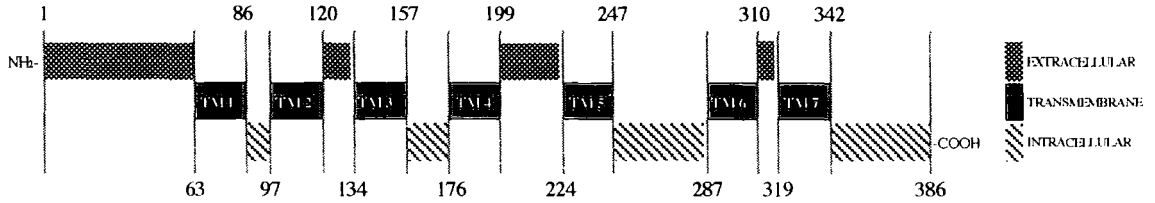


Figure 3.11(a). A schematic showing the predicted location of the extracellular, transmembrane and intracellular regions of the *N.lugens* rhodopsin. Numbers represent amino acid residues, labelled for methionine, residue 1. Membrane topology of the protein was predicted using TMHMM prediction server (http://www.cbs.dtu.dk/services/TMHMM/). Predicted transmembrane regions (TM 1-7) are marked on the *N.lugens* rhodopsin consensus sequence (Fig. 3.11b)

Figure 3.11(b). (opposite page) Nucleotide and deduced amino acid sequence of the *N.lugens* rhosopsin receptor. A methionine at position +1 indicates the start of the open reading frame of 1161 bp, which translates into 387 amino acids. The mature protein has a predicted molecular weight of 43.1 kDa. Putative transmembrane regions are underlined and labeled TM 1-7. Full circles indicate potential N-linked glycosylation sites. Nucleotides shaded in light grey represent 5' and 3' untranslated regions.

```
                                                GCCATCCTTTAGTGTTCTAAATCCTCGTCTGTCAGCAAATTTGGCTCCAA -122
CAATCCTGAAGTAGTTCGTCTGTCAAAACCAAACGATTTTCTTCAGCGATACGACTAACATCTGTCGCAGCC -72

ATGTCTTTGATTTCTGAACCGTCTTACAATGCCTACTCGGCCTGGTACGGAGCTCCAGGTGGAGGGGGTCAT +72
 M  S  L  I  S  E  P  S  Y  N  A  Y  S  A  W  Y  G  A  P  G  G  G  G  H

GGTATGGGCATGGGGGGAAACCAGACCGTGGTTGACAAGGTCCTGCCAGAAATGCTGCACATGGTCGACGCT +144
 G  M  G  M  G  G  N  Q  T  V  V  D  K  V  L  P  E  M  L  H  M  V  D  A

CATTGGTACCAATTCCCCCCGATGAACCCTCTATGGCACGACCTTCTGGGTTTCGCCATTGGGTTCTTGGGT +216
 H  W  Y  Q  F  P  P  M  N  P  L  W  H  D  L  L  G  F  A  I  G  F  L  G
                                                  TM1

TTCCTGTCGGTGTCGGGTAACGCGATGGTGGTCTACATCTTCTGCTCGACCAAGTCGCTGCGGACTCCATCC +288
 F  L  S  V  S  G  N  A  M  V  V  Y  I  F  C  S  T  K  S  L  R  T  P  S

AACCTGTTCGTCGTCAACCTCGCCTTCTCCGACTTCCTCATGATGTTCTGCATGTCGCCGACCATGGTCATC +360
 N  L  F  V  V  N  L  A  F  S  D  F  L  M  M  F  C  M  S  P  T  M  V  I
             TM2

AACTGCTACTACGAGACTTGGGTACTAGGACCTTTCATGTGTGAATTATATGGAATGCTGGGATCACTATTT +432
 N  C  Y  Y  E  T  W  V  L  G  P  F  M  C  E  L  Y  G  M  L  G  S  L  F
                                                  TM3

GGATGTGCCTCCATTTGGTCCATGACTATGATTGCTCTGGATAGGTACAATGTCATCGTCAAGGGTCTCTCA +504
 G  C  A  S  I  W  S  M  T  M  I  A  L  D  R  Y  N  V  I  V  K  G  L  S

GCGAAGCCTCTGACCAACAAGCGTGCTATTCTGAAGCTGTTGTTCGTCTGGCTCTTCTCTCTTGGCTGGACC +576
 A  K  P  L  T  N  K  R  A  I  L  K  L  L  F  V  W  L  F  S  L  G  W  T
                                                TM4

CTGGCCCCATTCTTCGGATGGAACAGGTATGTGCCAGAAGGAAACATGACCGCCTGTGGAACTGACTACTTG +648
 L  A  P  F  F  G  W  N  R  Y  V  P  E  G  N  M  T  A  C  G  T  D  Y  L

ACCAAGGACTGGTTCAGCCGCAGCTACATCATCGTCTACGGATTCTTCGTCTACTTCATGCCTCTCTTCATC +720
 T  K  D  W  F  S  R  S  Y  I  I  V  Y  G  F  F  V  Y  F  M  P  L  F  I
                                                TM5

ATTATCTACTCTTACTGGTTCATTGTGCAGGCTGTGTTCGCCCACGAGAAACAGATGAGGGAACAGGCCAAG +792
 I  I  Y  S  Y  W  F  I  V  Q  A  V  F  A  H  E  K  Q  M  R  E  Q  A  K

AAGATGAACGTTGCTTCCCTGCGATCGGCTGAGTCTCAGACCACCAGCCAGGAAATGAAATTGGCTAAGGTT +864
 K  M  N  V  A  S  L  R  S  A  E  S  Q  T  T  S  Q  E  M  K  L  A  K  V

GCTCTGATGACCATTGCCCTGTGGTTCTTCGCGTGGACCCCATACCTGGTGATCAACTTCTCGGGCATCTTC +936
 A  L  M  T  I  A  L  W  F  F  A  W  T  P  Y  L  V  I  N  F  S  G  I  F
             TM6

GAGACAGCCACCATCAGCCCCCTCTTCACCATCTGGGGATCCGTCTTTGCCAAGGCCAACGCTGTCTACAAT +1008
 E  T  A  T  I  S  P  L  F  T  I  W  G  S  V  F  A  K  A  N  A  V  Y  N
                                                TM7

CCTATCGTCTACGGTATCAGCCATCCTAAGTACCGTGCCGCCCTAGGCAAGAAGTTCCCCAGCCTGGTCTGC +1080
 P  I  V  Y  G  I  S  H  P  K  Y  R  A  A  L  G  K  K  F  P  S  L  V  C

GCGGCCGCTCCATCCGACGACGCCACATCCGTCGCCTCGGGCGCCACAACCGTCACTGAAGAGAAGGGACCA +1152
 A  A  A  P  S  D  D  A  T  S  V  A  S  G  A  T  T  V  T  E  E  K  G  P

AGCGCATAA                                                                  +1161
 S  A  .

GCTTTTATACTGCGACTGCAACGCTATCTGGCGGTCGAAAACGGAACCAGCGTTTCTCGCTCTAACCTACTG +1233
CTGCAATCTTCAGTCGCGTCACTGTCGTTCTACATCTTTTCTCTAGGTTAGGTTAGACGTTGAGCAAGCGTT +1305
GTTAATATTTTCTAACAAAATCTTCCACCGTCCGTTCAGCATGCACTTGTAAATATCCAGGGTCAATATTAA +1377
GAAATGGGTGGGTAATAATTTAATTGTAATTAATGTGTAAGGGGGTGATAGTTTATGTTTTCAGCTTTGAGT +1449
TGTCACCGTTTAGTTCTGTTTTTGAAGGGGGGTTTACCTGTGTTTGTTTTGTAAATTAATAAAGATCCTAGC +1521
AAGTAAAAAAAAAAAAAAAAACTCAAAAAAAAAAAACTCGA                                  +1563
```

rarity and low abundance of GPCR transcripts in message isolated from *N.lugens* the identification of such sequences by random chance is inherently difficult. Within a 'typical' cell mRNAs can be classified as abundant, intermediate or rare. By definition abundant messages represent 20% of the mRNA population, and normally comprised of 5-10 unique messages. Intermediate messages account for 40-60% of the mRNA mass, with 200-2000 distinct mRNAs. Whereas, rare messages represent <20%-40% of the mRNA population which encode 10,000 to 20,000 unique sequences (Carninci *et al.*, 2000). Therefore, sequencing cDNAs from standard cDNA libraries is ineffective for the discovering rarely expressed genes, where intermediately and highly expressed cDNAs are repetitively sequenced. Protocols for normalisation of cDNA libraries have been described (Carninci *et al.*, 2000). However, these procedures were not followed during construction of the *N.lugens* cDNA library, as the levels between abundant and rare message was considered to be to great. Any normalisation procedure would risk loosing underrepresented sequences. Therefore, to identify low abundance receptors large numbers of independent clones were sequenced, annotated by BLAST homology searches and compiled into an *N.lugens* EST database. Of the 7680 *N.lugens* ESTs only a single sequence was of the GPCR type, and this sequence represented the more abundant rhodopsin light sensing receptor. The *N.lugens* rhodopsin receptor was most similar to that from the Carpenter ant, *Camponotus abdominalis* (Q17292), with a 74% sequence identity. Although the mass sequencing approach was successful in retrieving a GPCR sequence the approach failed to generate receptor proteins that fitted the model of an ideal 'target' protein for modulation of function.

The cloning and characterisation of number of insect GPCRs is well documented in the literature (Section 1.10). Therefore, it is possible to establish the expression pattern of such receptor proteins and predict whether ingested antibodies would be capable of interacting with such a receptor. This is a distinct advantage over the random sequencing approach, as the tissue specificity and localisation of a particular protein can be predicted prior to cloning. For peptide hormone receptor proteins, there are numerous reports on the cloning and characterisation of receptors for allatostatin (ALSTR) and diuretic hormone (DHR). For both receptor proteins a gut specific expression pattern has already been demonstrated in several insect species (ALSTR, section 1.10.6 and DHR section 1.11.3). Therefore degenerate primers were designed to retrieve *N.lugens* homologues. The design of degenerate primers was straightforward due to the abundance of similar insect protein and nucleotide sequences in public databases. In all cases it was possible to limit the degeneracy of the primer by considering preferred codon usage within the aligned sequences. Although this approach was effective in retrieving *N.lugens* homologues of potential target proteins it is reliant on, and restricted by, the availability of functionally characterised GPCR sequences in public databases. The sequencing of the *Drosophila* genome has yielded 44 putative peptide receptors (Hewes and Taghert, 2001), however, many still remain 'orphan receptors' as they have not been functionally characterised. Although some of these sequences may represent ideal 'targets' they remain inappropriate for the design of antagonists and agonists until they have been functionally characterised. Characterisation of such 'orphan' GPCR proteins together with a more thorough understanding of insect signalling will lead to the development of novel insecticidal agents.

CHAPTER 4.

THE *N.LUGENS* DIURETIC HORMONE RECEPTOR

4.1 Introduction

An attractive target for the design of novel insecticidal proteins is the diuretic hormone receptor (DHR), due to the central role played in the regulation of fluid and ion homeostasis (Section 1.11). The modulation of DHR function by protein antagonists and agonists could prove detrimental to *N.lugens* survival, thereby useful in the control of this pest species. The ability of antibodies to act as antagonists and compete with hormone for receptor binding sites has been reported (Unson *et al.*, 1996), as well as the capability of ingested IgG antibodies to enter the haemolymph of the insect and retain their antigenic properties (Morin *et al.*, 1999). Therefore, it should be possible for orally delivered antibodies to enter the haemolymph and modulate the function of a 'target' protein. Here we report on the molecular cloning of the *N.lugens* diuretic hormone receptor (DHR) and present evidence to show that the protein is localised to the Malpighian tubules of the midgut. We also demonstrate that anti-DHR antibodies fed to *N.lugens* in liquid artificial diet are unable to enter the haemolymph, and therefore cannot interact with the predicted extracellular N-terminal hormone-binding domain of the DHR.

4.2 Results

4.2.1 Isolation of a cDNA encoding a DHR from the gut of N.lugens

Degenerate primers corresponding to conserved residues from transmembrane helices 3 and 6 of aligned insect DHRs and *Drosophila* family-B peptide receptors were designed and synthesised (Section 3.2.3). DHR degenerate forward and reverse primers (Table 2.2) were used to amplify a specific product from *N.lugens* gut specific cDNA (Fig. 4.1). The size of the amplified product (approx. 400 bp on gel) coincided with size predictions based on related insect DHR sequences. The amplified product was cloned and sequenced; comparison to proteins in the GenBank database using BLAST homology searching (Altschul *et al.*, 1990) revealed the amplified product had greatest similarity to insect diuretic hormone receptor proteins, in particular the *Acheta domesticus* diuretic hormone receptor (Reagan, 1996). Using the partial *N.lugens* DHR sequence gene specific primers were designed for retrieval of the extreme 5' and 3' end of the gene by rapid amplification of cDNA ends (RACE) PCR. The sequence of the 5' and 3' RACE oligonucleotides is detailed in Table 2.2. Based on estimates from the similar *Acheta domesticus* diuretic hormone receptor cDNA (Q16983) the 5' RACE product was predicted to be 873 bp in length, and the 3' RACE product 609 bp; plus additional untranslated regions at either end (Fig. 4.2). As the initial 5' RACE resulted in non-specific amplification the reaction products were reamplified using a nested gene specific primer (which was the reverse and complement of the 3' RACE primer), and this truncated amplification product was expected to be 744 bp, plus any UTR message (Fig. 4.2). Primers were used in RACE reactions on *N.lugens* gut specific cDNA, according to the manufacturers protocol (Clontech). The 3' RACE reaction yielded a

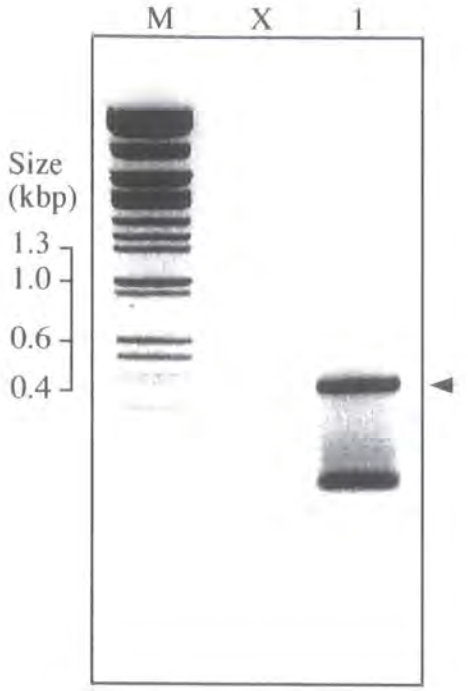


Figure 4.1. PCR amplification of a partial *N.lugens* diuretic hormone receptor (DHR) from gut specific cDNA using degenerate primers. The partial PCR fragment of approx. 400bp (lane 1, arrow) was cut from the gel, purified and cloned. DNA sequencing revealed this amplification product to be a partial sequence with similarity to other insect diuretic hormone receptors. Based on the sequence data gene specific primers were designed and used in 5' and 3' RACE experiments for retrieval of the full-length gene, including untranslated regions (UTRs), as shown in Fig. 4.2.

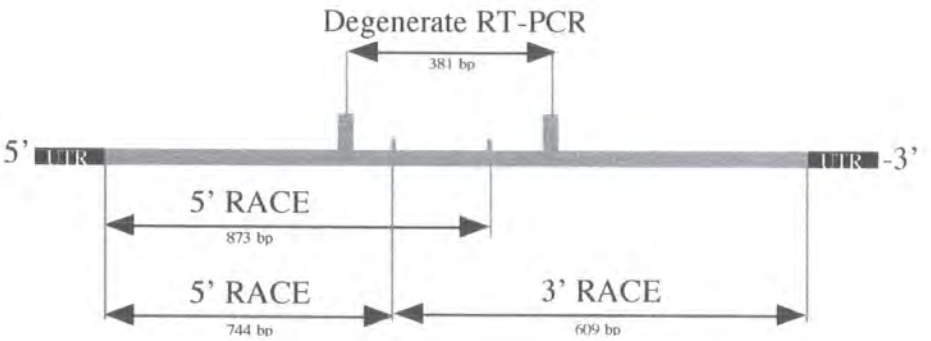


Figure 4.2. Based on the partial *N.lugens* DHR sequence gene specific primers were designed for the use in 5' RACE and 3' RACE experiments, for retrieval of the full-length *N.lugens* DHR. The 3' RACE reaction was expected to produce a PCR product of 609 bp + UTR and the 5' RACE reaction was expected to produce a PCR product of 744 bp + UTR. Expected size of PCR products was based on the *Acheta domesticus* DHR sequence (Reagan, 1996), which the partial *N.lugens* cDNA sequence was most similar to.

strong product of approx. 1300 bp (Fig. 4.3, lane 1). Similar products did not appear in any of the control RACE PCR reactions using single primers (Fig. 4.3, lane 3 & 4). Therefore the reaction product was deemed specific and 1.3 kbp was cloned and sequenced. The initial 5' RACE reaction resulted in non-specific amplification (Fig. 4.4, lane 1). Therefore, PCR products generated by the first amplification were diluted and reamplified using a nested gene specific primer (which was the reverse and complement of the 3' RACE primer). Reamplification of the initial 5' RACE PCR produced strong products at approx. 1.1 kbp and 1.2 kbp (Fig. 4.4, lane 2). Both products appeared to be a result of specific amplification, as products of similar sizes were absent from the single primer controls (Fig. 4.4, lane 4 & 5). DNA sequencing and BLAST homology searches of the GenBank database revealed that the smaller product had homology to insect diuretic hormone receptor sequences, whereas the larger product was a result of non-specific amplification. Multiple clones of PCR products from the two RACE reactions were sequenced, and PCR amplifications were also carried out using primers covering the whole predicted ORF in the DHR cDNA (section 4.2.4). This stratergy allowed two cDNA sequences to be assembled.

4.2.2 DHR cDNAs (DHR-I and DHR-II)

The cDNA represented most often in the clones obtained (DHR-I) was 2219 bp in length from the 5' end of the coding sequence, including 270 bp of the 3' UTR with a 29 bp poly(A) sequence attached (Fig. 4.5a). The longest open reading frame (ORF) consisted of 1920 bp, and encoded a protein of 639 amino acids with a predicted M_r of 69,756. The cDNA was truncated by 22 bp at the 5' end, and was reconstructed to the start codon by PCR based on the sequence of DHR-II (q.v.). The nucleotide sequence and

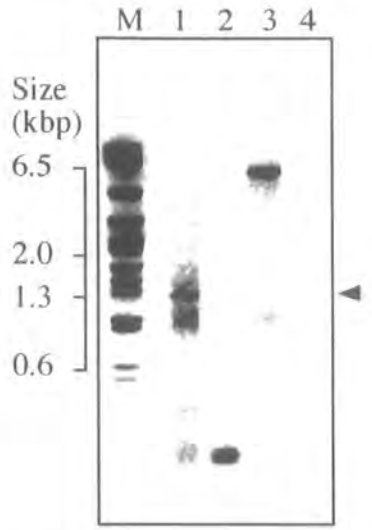


Figure 4.3. *N.lugens* DHR 3' RACE. The 3' end of the *N.lugens* DHR was amplified using 3' RACE PCR with a gene specific primer (GSP) and a 3' RACE primer (lane 1). The 3' RACE cDNA was checked by amplification of an internal fragment of the DHR gene using two gene specific primers (lane 2). Primers were checked for miss-priming by using a standard RACE reaction with either the single GSP (lane 3) or the single 3' RACE primer (lane 4).

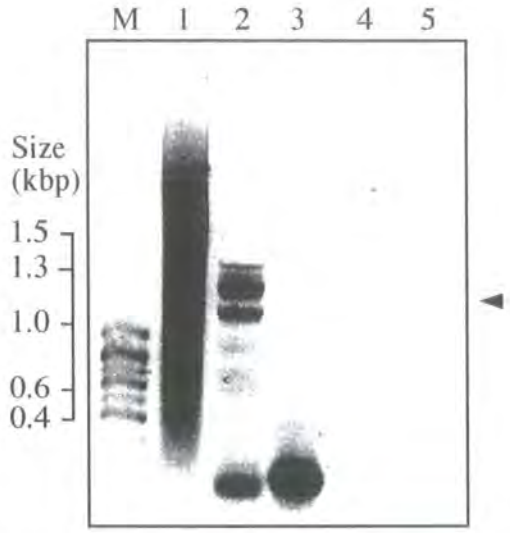


Figure 4.4. *N.lugens* DHR 5' RACE. Initial 5' RACE amplification produced multiple PCR products, which looked like a smear on the gel (lane 1). Due to the non-specific amplification an aliquot of this reaction was diluted 50-fold in Tris-HCl buffer (pH 7.5) and reamplified using a nested GSP and a nested 5' RACE primer. The resulting band of approx. 1.3kbp (lane 2) was purified from the gel, cloned and sequenced. The 5' RACE cDNA was checked by amplification of an internal fragment of the DHR gene using two gene specific primers (lane 3). Primers were checked for miss-priming by using a standard RACE reaction with either the single GSP (lane 4) or the single 5' RACE primer (lane 5). M is the DNA size marker.

translation of DHR-I is shown in Fig. 4.5a. A second sequence (DHR-II) was also co-amplified, although this cDNA appeared to be less abundant than DHR-I. The DHR-II transcript was 2344 bp in length, and contained a 5' UTR, of 167 bp. The longest ORF was of 1926 bp, and encoded a protein of 641 amino acids with a predicted Mr of 69,996. The 3' UTR of this clone, of 229 bp, with a poly(A) sequence attached, was dissimilar to that of the other cDNA; however, the sequence was similar over the coding region, with only minor differences to DHR-I. The DHR-II sequence is not presented, but differences in the encoded amino acid sequence are noted in Fig. 4.5a. Both sequences have been filed with the EMBL sequence database under the accession numbers AJ634780 (DHR-I) and AJ634781 (DHR-II).

4.2.3 DHR Sequence Analysis

BLAST searches of the GenBank database showed that the longest open reading frame of the gene sequence had greatest sequence similarity (51% overall sequence identity) with the *Acheta domesticus* DHR (Reagan, 1996). Based on hydrophobicity analysis (Krogh *et al.*, 2001) the protein is predicted to have an extracellular N-terminal domain (amino acid 1-260), seven transmembrane helices connected by short peptide loops (29 amino acids or less) and an intracellular C-terminal region (Fig. 4.5b). Phosphorylation sites are also a common feature on GPCR proteins, therefore potential serine and threonine phosphorylation sites in the putative *N.lugens* DHR were predicted using NetPhos, version 2.0 software programme (http://www.cbs.dtu.dk/services/NetPhos/) (Blom *et al.*, 1999). Potential serine phosphorylation sites were predicted at ser 581, ser 587, ser 590, ser 608, ser 611 and ser 627; and potential threonine phosphorylation sites at thr 588, thr 554 and thr 545 (Fig. 4.5a). The *N.lugens* DHR protein sequence has 4

Figure 4.5(a) (opposite). Consensus nucleotide sequence, and amino acid translation of the putative diuretic hormone receptor (DHR-I) from *N.lugens* (GenBank accession number AJ634780). A methionine at position +1 indicates the start of the open reading frame of 1920 bp, which translates into 639 amino acids. Differences in amino acid sequence predicted by DHR-II (AJ634781) are shown in brackets. Putative transmembrane regions are underlined and labelled TM 1 – 7. Full circles indicate potential N-linked glycosylation sites, ^ indicate cysteine residues conserved in Family B GPCRs located in the N-terminal domain and predicted extracellular loops, * indicate predicted serine phosphorylation sites, and ~ indicate predicted threonine phosphorylation sites.

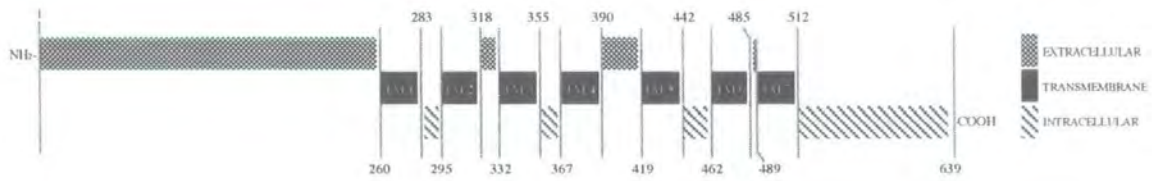


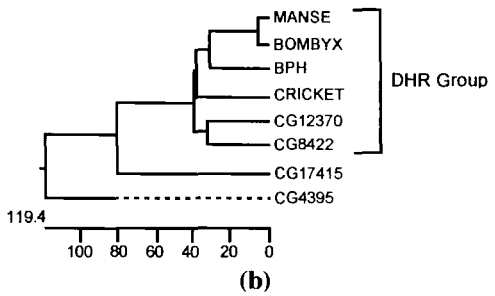
Figure 4.5(b). A schematic showing the predicted location of extracellular, transmembrane and intracellular regions of the putative *N.lugens* DHR. Numbers represent amino acid residues, labelled from methionine, residue 1. Membrane topology of the protein was predicted using TMHMM prediction server (http://www.cbs.ctu.dk/services/TMHMM/). Predicted transmembrane regions (TM 1-7) are marked on the *N.lugens* DHR consensus sequence (Fig. 4.5a).

```
      ACGCGGGGGCAGACTTCATGTGTGTCTCCAGCGTTATTGCCCTGTCACCAGCTGGCTCAAATCTTTGAATAGTAGACTTG  -167
AGATGCGCGGGAACACTAGCGTCAGGTGATCAATTTTCATAACGGCAAGGAAAGTTGTATTGACCGATTGACACGAAACTTTACAAA  -87

ATGGCGGAGACTGTATCCAATGAAACGATGCATTACATGAACGAGGAGTTCGAATGCCAACTGCTCATCTACCGCGAGCAAGAACTG  +87
<M  A  E  T  V  S  N  E  T  M  H  Y  M  N  E  E  F  E  C  Q  L  L  I  Y  R  E  Q  E  L
                  •
GACAGCATGCTATCCTGGCCTCCAACCGATGACGTCACAATCAACCATCCACCCAGTGCTGACGTCATGGGTGGAAGACTGAGACCA  +174
 D  S  M  L  S  W  P  P  T  D  D  V  T  I  N  H  P  P  S  A  D  V  M  G  G  R  L  R  P
(E)  (I)
TCTGTAGGTGACGTCACAATGGGGAACCCCAATGCTGACACCATACAAAATATACTGAAGCATCCTGTGGGTGACGTCACAATGGGG  +261
 S  V  G  D  V  T  M  G  N  P  N  A  D  T  I  Q  N  I  L  K  H  P  V  G  D  V  T  M  G

AATCCCAATACTGGAAGTATATCAGCTGGAATGAGACATTTTGCCGGTGACGTCACAATGGGGAATCCCAGCGCTGACACCATACGA  +348
 N  P  N  T  G  S  I  S  A  G  M  R  H  F  A  G  D  V  T  M  G  N  P  S  A  D  T  I  R
          (G)                    (T)         (V)
GGGATACTGAAACAGTCTGTGGGTGACGTCACAACGGGGAACCAGAATACTGGAAGTGTAGTGTCAGGAATGAGACATTTCGCTGGT  +435
 G  I  L  K  Q  S  V  G  D  V  T  T  G  N  Q  N  T  G  S  V  V  S  G  M  R  H  F  A  G
                                                        (I)(A)(A)                    (S)
GACGTCACAGGGGGGAATCCCAATGCTGATGTCATCGGCGGACGACTAAGGCCGCCTAGCATGGATGTGCCCGTTGGTGACGTCACA  +522
 D  V  T  G  G  N  P  N  A  D  V  I  G  G  R  L  R  P  P  S  M  D  V  P  V  G  D  V  T
       (V)                                                          (H)
CAGGGTAGAGCGCAGTATGCTTCACACTGTCCCAGGAGTTGGGACTCTCTACTTTGTTGGCCACCCACTAAACCTGGACATACGGCA  +609
 Q  G  R  A  Q  Y  A  S  H  C  P  R  S  W  D  S  L  L  C  W  P  P  T  K  P  G  H  T  A
                           ^                          ^
ACGATGGCTTGCTTCCCTGAGCTGAATGGGATCAAGTATGATACTACATTGAACGCAACCAGAAGATGCCTGCTGAATGGTACCTGG  +696
 T  M  A  C  F  P  E  L  N  G  I  K  Y  D  T  T  L  N  A  T  R  R  C  L  L  N  G  T  W
          ^                                    •              ^              •
GACAACTACACAGACTACACGTCCTGCAAAGACCTGAGTCCCGATCAACCAGATTTGGAGCCTGGCATAGAAGTTACCACTATGATA  +783
 D  N  Y  T  D  Y  T  S  C  K  D  L  S  P  D  Q  P  D  L  E  P  G  I  E  V  T  T  M  I
 •                       ^
TACTCAGCTGGATATGCCCTTTCCCTCATAGCTCTTGTACTGGCTGTCTCTATATTCCTCTATTTCAAGGATTTGAGATGTCTGCGG  +870
 Y  S  A  G  Y  A  L  S  L  I  A  L  V  L  A  V  S  I  F  L  Y  F  K  D  L  R  C  L  R
          TM1
AATACAATCCACACCAACCTGATGTGTACATATATAATGGCCGACTTCATGTGGATACTAAATATCACTGTACAGATGTCGATGCCA  +957
 N  T  I  H  T  N  L  M  C  T  Y  I  M  A  D  F  M  W  I  L  N  I  T  V  Q  M  S  M  P
                   TM2
ACCAACGTGTCAGCCTGTGTTATACTAGTCGTCCTCTTACACTACTTCTATTTGACCAATTTCTTCTGGATGTTTGTAGAGGGTCTC  +1044
 T  N  V  S  A  C  V  I  L  V  V  L  L  H  Y  F  Y  L  T  N  F  F  W  M  F  V  E  G  L
                                        TM3
TATCTATATATGCTGGTGGTGGAAACCTTCTCTAGTAAAAACATAAAACTGAGGGCCTACGTCTGTATAGGATGGGGTGTGCCTTGT  +1131
 Y  L  Y  M  L  V  V  E  T  F  S  S  K  N  I  K  L  R  A  Y  V  C  I  G  W  G  V  P  C
                                                          TM4
GCAGTGATAATCATATGGACAATTGTCAGAAGTCTTATGCCTCCTACCTCTGATAATATGGGTCTTTCGGGCATAGTTCTGAAAAGC  +1218
 A  V  I  I  I  W  T  I  V  R  S  L  M  P  P  T  S  D  N  M  G  L  S  G  I  V  L  K  S
TGTCCCTGGATGGCTACCCATAACATTGATTGGATCTACCAGGGCCCAGCAGCTGTCATACTATTCGTCAATGTCATTTTCCTAGTC  +1305
 C  P  W  M  A  T  H  N  I  D  W  I  Y  Q  G  P  A  A  V  I  L  F  V  N  V  I  F  L  V
                                        TM5
ATGATCATGTGGGTTCTGATAACCAAACTAAGGTCAGCTAATACAGTGGAAACTCAACAATATAGAAAGGCAGCTAAAGCATTATTG  +1392
 M  I  M  W  V  L  I  T  K  L  R  S  A  N  T  V  E  T  Q  Q  Y  R  K  A  A  K  A  L  L
GTGCTTATACCTTTGTTGGGTATCACATATATTCTAGTCATATATGTACCCAGTCATGGAGTAACTGCTAATCCATTGGCGTATTGC  +1479
 V  L  I  P  L  L  G  I  T  Y  I  L  V  I  Y  V  P  S  H  G  V  T  A  N  P  L  A  Y  C
          TM6
CGGGCTATACTGCTGTCAACACAGGGTTTCACAGTAGCACTGTTCTACTGTTTCCTGAATTCAGAAGTGCAGCACACGCTCAGTTCG  +1566
 R  A  I  L  L  S  T  Q  G  F  T  V  A  L  F  Y  C  F  L  N  S  E  V  Q  H  T  L  S  S
          TM7
CACTTTGAACGCTGGAAGGAGGCGAGGCAGATAGGAGGAGGAGGTGGTGGTGGTGGGAGGAGGTATACCTATAGCAAGGACTGGTCC  +1653
 H  F  E  R  W  K  E  A  R  Q  I  G  G  G  G  G  G  G  G  R  R  Y  T  Y  S  K  D  W  S
                (E)                 (+R)                 ~
CCCAATACTAGAACAGAAAGTATCCGCCTCTGCAGCACTAACCAGGTGGGAGGAGGAGGTGGTGGAGGCGGAGGAGGAGGAGGCATG  +1740
 P  N  T  R  T  E  S  I  R  L  C  S  T  N  Q  V  G  G  G  G  G  G  G  G  G  G  G  G  M
       ~                                    (+A)
AGTTACAAGAAGCGCGAATCAACAGTTAGTGAGACAACCACAATGACTCTGGTTGGCGGTGGTGTGGGGTCGAATCGGGTGAGCAAC  +1827
 S  Y  K  K  R  E  S  T  V  S  E  T  T  T  M  T  L  V  G  G  G  V  G  S  N  R  V  S  N
 *                    *  ~     *                                                    *
GGGTCTATGGGACCCACTATGGGGTCTCTCAGACCCCCACAGGGGTCCACCAGTCCCTATTTGGAGGTGGCGCCTCTCGAGAATGCC  +1914
 G  S  M  G  P  T  M  G  S  L  R  P  P  Q  G  S  T  S  P  Y  L  E  V  A  P  L  E  N  A
    *                                                  *
GTGTGA                                                                                  +1920
 V  .>

AGAGGTAGAAGTTAAAAAAAATTCAGAGAAGACTAGCCTATAAATTTGGAATAATCAGGTAATAAGAAAGGATTGTAAGCAACAGTG  +2010
AAAGAAGAATGGGAATCAATGTAGAAGTAGGAGGAGGAGGACGAGCAGCAGGTGAAGGGAAGAAGGAGAGGAGAAGGAGAGAGAG  +2100
AGAACGAGAGAGAGAGAGAGACGAAGAAGAGAATCATCCATGAATCAATTATGATGAAGACAATGTTTCCCGAAACATTGGTTTA  +2190
ATGAATGATAAAAAAAAAAAAAAAAAAAAAAAAAAAA                                                   +2219
```

potential glycosylation sites on the N-terminal domain, matching the N-X-S/T consensus at asn 7, asn 221, asn 229 and asn 234 (Fig. 4.5a). Analysis of the protein using SignalP, version 2.0 software programme (http://www.cbs.dtu.dk/) (Nielsen *et al.*, 1997) did not reveal a conventional signal sequence for rough endoplasmic reticulum (RER) membrane translocation.

The *N.lugens* receptor protein was aligned with diuretic hormone receptors from *Manduca sexta* (P35464), *Acheta domesticus* (Q16983) and *Bombyx mori*, plus *Drosophila* gene products belonging to Family-B receptor proteins (CG4395, CG17415, CG8422 and CG12370) (Fig. 4.6a). Alignment of related insect sequences reveals a strong degree of sequence conservation, especially within the transmembrane domains. The *N.lugens* receptor is closely related to the *Bombyx mori* DHR (51 % overall identity, and 69% sequence identity in the transmembrane domains), the *Manduca sexta* DHR (50 % overall sequence identity) and the *Acheta domesticus* DHR (46 % overall sequence identity), plus *Drosophila* gene products CG12370 and CG8422 (with an overall sequence identity of 47 % and 41 %, respectively). However, the putative *N.lugens* diuretic hormone receptor is significantly longer than the closely related insect diuretic hormone receptors from *Bombyx mori*, *Manduca sexta* and *Acheta domesticus*. The longest of the closely related insect diuretic hormone receptors is the *Acheta domesticus* DHR, which is only 441 amino acids, compared to 639 amino acids of the putative *N.lugens* DHR. Although overall sequence similarity is high between the two receptors the *N.lugens* DHR has unusually large extracellular N-terminal domain (260 amino acids), and intracellular C-terminal region (124 amino acids). The N-terminal

(a)



(b)

Figure 4.6 (a and b). **(a)** Amino acid alignment using CLUSTAL V algorithm (Higgins and Sharp, 1989) of the *N.lugens* (BPH) DHR-I (AJ634780) with three other insect DHR sequences and *Drosophila* Family-B receptor proteins. Residues that match the consensus are shaded grey. Full circles indicate conserved cysteine residues. Aligned sequences are the *Drosophila* gene products CG12370, CG8422, CG17415, CG4395 and diuretic hormone receptors from *Manduca sexta* (MANSE, P35464), *Acheta domesticus* (CRICKET, Q16983) and *Bombyx mori* (BOMBYX). **(b)** The phylogenetic relationship of the aligned sequences.

domain of the *Acheta domesticus* DHR is predicted to be 134 amino acids and the C-terminal region 48 amino acids. Therefore, the unusual N-terminal and C-terminal extensions account for the difference in size between the putative *N.lugens* DHR and other insect DHRs. The *Drosophila* gene product CG4395 is the closest match in length to the putative *N.lugens* DHR, but overall sequence similarity between these two proteins is only 16%.

Protein alignment also reveals well-conserved cysteine residues in the N-terminal domain and first and second extracellular loop regions of the majority of aligned proteins (Fig. 4.6a). In the putative *N.lugens* DHR conserved cysteine residues occur within the extracellular N-terminal domain at cys 184, cys 193, cys 207, cys 226 and cys 241; and cys 325 from extracellular loop region 1 and cys 407 from extracellular loop region 2 are also well conserved (Fig. 4.6a). The *N.lugens* diuretic hormone receptor is also distantly related to mammalian calcitonin receptors (30 % overall identity with Norway rat, *Rattus norvegicus* calcitonin receptor, A37430) (Sexton *et al.*, 1993), secretin receptors (21 % overall identity with Norway rat, *Rattus norvegicus* secretin receptor) (Dong *et al.*, 2002); and corticotropin releasing factor receptor (30 % overall identity with house mouse, *Mus musculus* corticotropin releasing factor receptor, NP_034083) (Kishimoto *et al.*, 1995). Furthermore, cysteine residues in these proteins are also highly conserved and occur at the same sequence positions as in the aligned insect receptors (data not shown). This demonstrates structural similarities between the aligned insect receptor proteins and the more distant mammalian receptor proteins. Phylogenetic analysis of the aligned receptor proteins shows distinct grouping of the

N.lugens receptor with insect diuretic hormone receptors from *Manduca sexta, Acheta domesticus* and *Bombyx mori* and *Drosophila* gene products CG12370, CG8422 (Fig. 4.6b). Other members of the *Drosophila* Family B receptors (CG17415 and CG4395) have a limited sequence similarity with the DHR group of proteins and form a distinct sub-group (Fig. 4.6b).

4.2.4 Bacterial and Insect Cell Expression Constructs

Both the N-terminal domain and full-length DHR coding sequences were amplified from *N.lugens* gut specific cDNA. Sequences used for expression constructs matched that of the more abundant DHR-I, AJ634780 (Fig. 4.5a) The primer pair for the full-length *N.lugens* DHR was designed to amplify 1920 bp of coding sequence (amino acid 1-639). The primer pair for the N-terminal domain was designed to amplify the first 780 bp of the coding sequence, corresponding to the predicted extracellular N-terminal domain (amino acids 1-260). Primers were synthesised (Table 2.2) and used as detailed in section 2.8.4 (full-length receptor), and section 2.8.5 (N-terminal domain). Both sets of primers were synthesised and used in PCR on *N.lugens* gut specific DNA. After 30 PCR amplification cycles a specific product corresponding to the full-length DHR coding sequence appeared (Fig. 4.7, lane 3). Also specific amplification of the N-terminal domain (amino acids 1-260) was achieved after 30 PCR cycles (Fig. 4.8, lane 1). All products were cloned and sequenced and shown to match the consensus exactly.

4.2.5 Variable cDNA Sequences Encoding N.lugens DHR

PCR amplification of the full-length DHR sequence using DHR CDS fwd and rev primers (Table 2.2) resulted in the co-amplification of two sequences of different length.

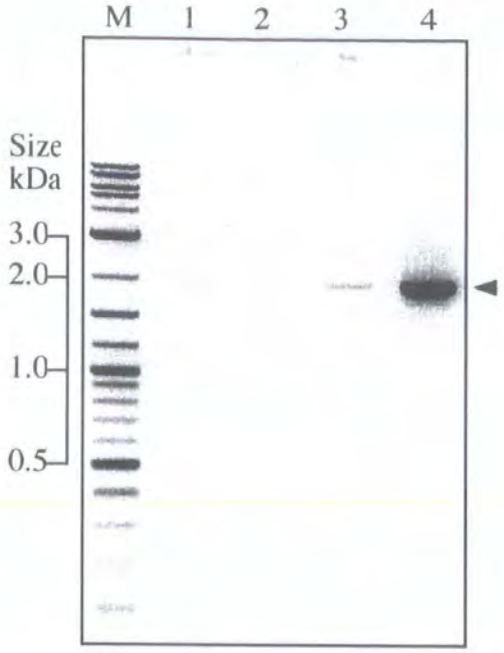


Figure 4.7. The putative *N.lugens* DHR coding sequence (CDS) was amplified from cDNA by PCR using a proof reading polymerase (Clontech). PCR errors were avoided by using a minimal number of cycles. Reaction products were removed after 20 cycles (lane 1), 25 cycles (lane 2), 30 cycles (lane 3) and 35 cycles (lane 4). Products were separated on a 1.1% agarose/TAE gel, and DNAs were stained with ethidium bromide (0.5μg/ml). Reaction products were cloned after 30 amplification cycles. Arrow indicates PCR reaction product corresponding to the putative DHR CDS (1920 bp).

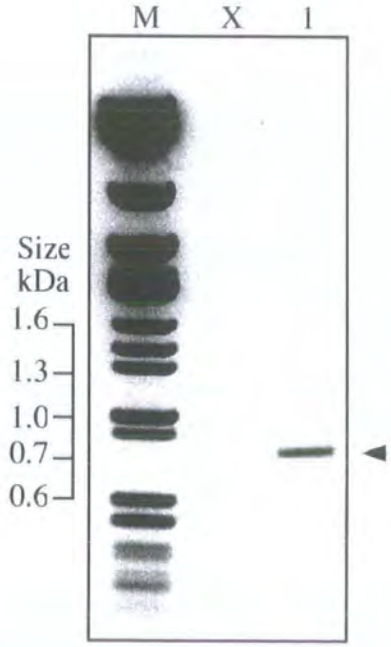


Figure 4.8. Amplification of the putative *N.lugens* N-terminal domain of the *N.lugens* DHR. The N-terminal domain was predicted by hydrophobicity analysis (Fig. 4.5b) and amplified by PCR using proof-reading polymerase and a minimal number of cycles. Amplification product is marked with an arrow. NOTE: X denotes an empty lane.

However, on the gel the long and short isoforms were not resolved and therefore appear as a single amplification product (Fig. 4.7). Reaction products from the amplification reaction were cloned and analysed by restriction with *Nde* I and *Xho* I (Fig. 4.9); of the 4 clones analysed 3 appear to be the long isoform (Fig. 4.9, lanes 1, 2 and 4), while a single clone represents the short isoform (Fig. 4.9, lane 3). Both the long and short isoform were completely sequenced on both strands of DNA. The long isoform (1920 bp) matched the *N.lugens* DHR-I (AJ634780) consensus sequence exactly (Fig. 4.5a). While the short isoform (1599 bp) matches the consensus sequence until glu 529, after which the sequence diverges and ends prematurely (Fig. 4.10a & b), after glu 529 the predicted sequence then continues EEEA* (where * represents the stop codon).

The origin of the short isoform was investigated. DNA from the variable region was analysed to elucidate whether long and short versions were encoded by different genes, or if they resulted from alternative splicing of a single genomic copy. It is known that all genes from this family are encoded on multiple exons, and alternative splicing throughout this family is common. A primer 5' to the variable sequence (DHR genomic, Table 2.2) was used in conjunction with DHR CDS rev primer (Table 2.2) in PCR amplification from genomic DNA (Fig. 4.11). Amplification resulted in a strong product of approx. 1600 bp, this genomic product was cloned and sequenced. Sequence results demonstrate that the location of the intron matches that for the long isoform (Fig. 4.12). As the short isoform was not represented by a genomic copy it was deemed to be a PCR artefact. Therefore, the long isoform matching the DHR-I sequence (AJ634780), was expressed as a recombinant protein in Sf21 insect cells (Section 4.2.8).

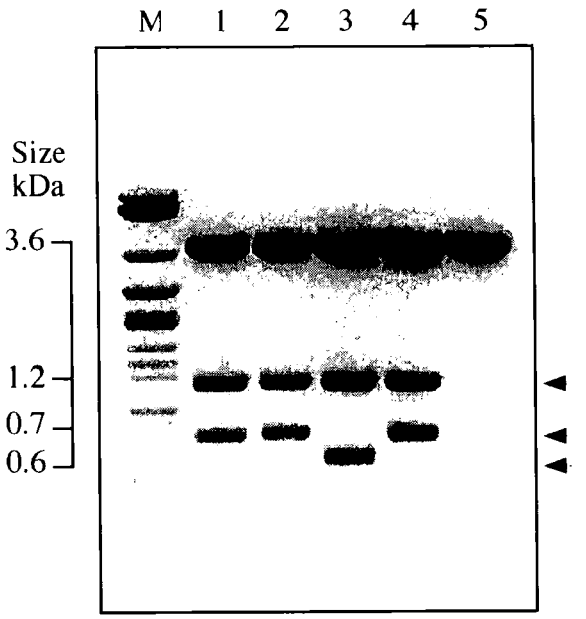


Figure 4.9. Amplification of the *N.lugens* DHR coding sequence, the majority of clones are the 'long' isoform (lanes 1,2 and 4). However a short version is also apparent (Lane 3). The long and short isoforms were analysed by DNA sequencing (Fig. 4.10 a and b).

(a)

```
GTAGCACTGTTCTACTGTTTCCTGAATTCAGAAGTGCAGCACACGCTCAGTTCGCACTTTGAACGCTGGAAG
 V  A  L  F  Y  C  F  L  N  S  E  V  Q  H  T  L  S  S  H  F  E  R  W  K

GAGGCGAGGCAGATAGGAGGAGGAGGTGGTGGTGGGAGGAGGTATACCTATAGCAAGGACTGGTCCCCC
 E  A  R  Q  I  G  G  G  G  G  G  G  G  R  R  Y  T  Y  S  K  D  W  S  P

AATACTAGAACAGAAAGTATCCGCCTCTGCAGCACTAACCAGGTGGGAGGAGGAGGTGGTGGAGGCGGAGGA
 N  T  R  T  E  S  I  R  L  C  S  T  N  Q  V  G  G  G  G  G  G  G  G  G

GGAGGAGGCATGAGTTACAAGAAGCGCGAATCAACAGTTAGTGAGACAACCACAATGACTCTGGTTGGCGGT
 G  G  G  M  S  Y  K  K  R  E  S  T  V  S  E  T  T  T  M  T  L  V  G  G

GGTGTGGGGTCGAATCGGGTGAGCAACGGGTCTATGGGACCCACTATGGGGTCTCTCAGACCCCCACAGGGG
 G  V  G  S  N  R  V  S  N  G  S  M  G  P  T  M  G  S  L  R  P  P  Q  G

TCCACCAGTCCCTATTTGGAGGTGGCGCCTCTCGAGAATGCCGTGTGA
 S  T  S  P  Y  L  E  V  A  P  L  E  N  A  V  .>
```

(b)

```
GTAGCACTGTTCTACTGTTTCCTGAATTCAGAAGTGCAGCACACGCTCAGTTCGCACTTTGAACGCTGGAAG
 V  A  L  F  Y  C  F  L  N  S  E  V  Q  H  T  L  S  S  H  F  E  R  W  K

GAGGAGGAGGCATGAGTTACAAGA
 E  E  E  A  .  V  T  R
```

Figure 4.10 (a&b). DNA sequence of PCR amplification products corresponding to the long isoform (a) and the short isoform (b) as amplified from *N.lugens* gut cDNA using DHR CDS fwd and rev primers (Table 2.2). Only the variable region of the sequence is detailed, sequence data for the rest of the sequence matches the consensus exactly (Fig. 4.5a).

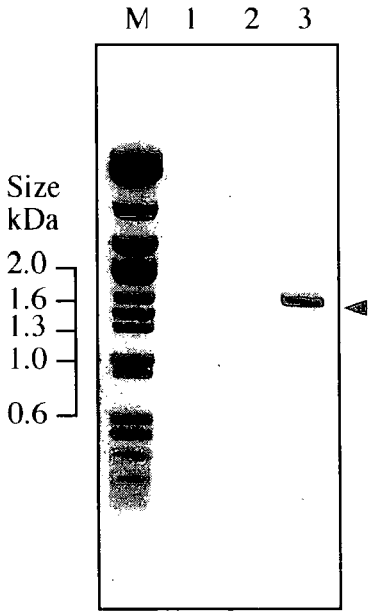


Figure 4.11. Amplification of a partial *N.lugens* DHR sequence from genomic DNA. Primers were designed corresponding to regions flanking the variable sequence (Fig. 4.10 a and b). Amplification product (approx. 1600 bp) was cloned and sequenced (Fig. 4.12).

```
GTAGCACTGTTCTACTGTTTCCTGAATTCAGAAGTGCAGCACACGCTCAGTTCGCACTTTGAACGCTGGAAG
 V  A  L  F  Y  C  F  L  N  S  E  V  Q  H  T  L  S  S  H  F  E  R  W  K

GAGGCGAGGCAGATAGGAGGAGGAGGTGGTGGTGGTGGGAGGAGGTATACCTATAGCAAGGACTGGTCCCCC
 E  A  R  Q  I  G  G  G  G  G  G  G  G  R  R  Y  T  Y  S  K  D  W  S  P

AATACTAGAACAGAAAGTATCCGGTAGGTACTAAAAACAAAAACCAAAAATCTGGTGTGGCGCACTCACACA
 N  T  R  T  E  S  I  R

ACTTTCCTTGCCGTTATGAAAATTTATCACCTGACGCTAGTGTTCACGCTCAACTCAAGTCTTCTATTCAAG
GATCTGAGCCAGCTGGTGACAGGACAATAACGCTGGAGACACACATGAGGTCTGCTATCTCTTCATAGTGAA
TCATTTAATAGATTCAAGTTGCCAACAAATTGCAATTGAATAATCTTATTTTCTCGAACATCGAGCTTATTT
CCAATTTATGTCAAAATGTTACTGAACATTAATTGTAGAGAATTTTATGCTCAATCTTCTCGACTCGGATTT
TTTTGGTTATTTATTTTTTTTTTTTATTTATTTCATTTATTTGCAATCAAATGAAACTAATGTAATAGCATT
GGTAGATAA.....
.....UNSEQUENCED INTRON.....
.....
.....CAACTTTCTCAACTACACAAAAAGTACATTCAAGTTTTTGAATAAAAG
CTATTGAAACTAATACTCACGTGTCTACGCACGCTTTCGGATTTTTACGAACTGTAGTGTCGTCAAAAATGA
TCCAAAAATGATCGATTCTATCATTATCAAATTCTGTTTTGTTCTCTGTTCTCTATCCTCTTCTAATAATTC
TCTTTTGTTCTCTCTGTCCGCAGCCTCTGCAGCACTAACCAGGTGGGAGGAGGAGGTGGTGGAGGCGGAGGA
                       L  C  S  T  N  Q  V  G  G  G  G  G  G  G  G  G

GGAGGAGGCATGAGTTACAAGAAGCGCGAATCAACAGTTAGTGAGACAACCACAATGACTCTGGTTGGCGGT
 G  G  G  M  S  Y  K  K  R  E  S  T  V  S  E  T  T  T  M  T  L  V  G  G

GGTGTGGGGTCGAATCGGGTGAGCAACGGGTCTATGGGACCCACTATGGGGTCTCTCAGACCCCCACAGGGG
 G  V  G  S  N  R  V  S  N  G  S  M  G  P  T  M  G  S  L  R  P  P  Q  G

TCCACCAGTCCCTATTTGGAGGTGGCGCCTCTCGAGAATGCCGTGTGA
 S  T  S  P  Y  L  E  V  A  P  L  E  N  A  V  .>
```

Figure 4.12. DNA sequence of PCR product amplified from *N.lugens* genomic DNA using primers flanking the variable region from the long and short DHR isoforms (Fig. 4.11). Nucleotides corresponding to the coding sequence are shown in bold type with amino acid translation. Intronic regions are shaded grey.

4.2.6 Production of anti-DHR antibody by means of expression of recombinant DHR

The small size of *N.lugens* and low abundance of hormone receptor makes purification of these proteins in sufficient quantity for characterisation and antibody raising difficult. Therefore a recombinant protein expression system was used to produce a fragment of the DHR for the purpose of antibody production (Section 2.8.6). A cDNA fragment encoding the predicted N-terminal extracellular domain (amino acids 1-260) was cloned into a bacterial expression vector, in frame with a C-terminal poly-His fusion tag. The construct was transformed into an *E.coli* host strain for expression, which was induced in cells grown to mid-log phase. The accumulation of recombinant protein after induction of expression was monitored by SDS-PAGE and Coomassie blue staining (Fig. 4.13a). The recombinant protein accumulated in cells to a level of about 15 mg/l. Cell lysis and fractionation showed the DHR N-terminal domain was present in the insoluble fraction (Fig. 4.13a, lane 4). On denaturing SDS-PAGE gels the recombinant protein migrates with a mobility similar to the 36 kDa band of the molecular weight marker, however the predicted M_r of the N-terminal domain is only 28.1 kDa. Therefore, M_r of the purified recombinant protein was checked by SELDI-TOF mass spectrometry, and shown to coincide with the predicted molecular weight (Fig. 4.13b). The anomalous migration of the recombinant polypeptide on SDS-PAGE may be due to the amino acid composition and a reduced capacity to bind SDS detergent. As the protein was expressed in the insoluble fraction the DHR N-terminal domain was purified under denaturing conditions (Section 2.8.6). Cells were lysed by sonication in a buffer containing 6 M urea, and the poly-His tagged recombinant protein was purified by metal affinity chromatography (Section 2.8.6). The affinity column was washed to remove

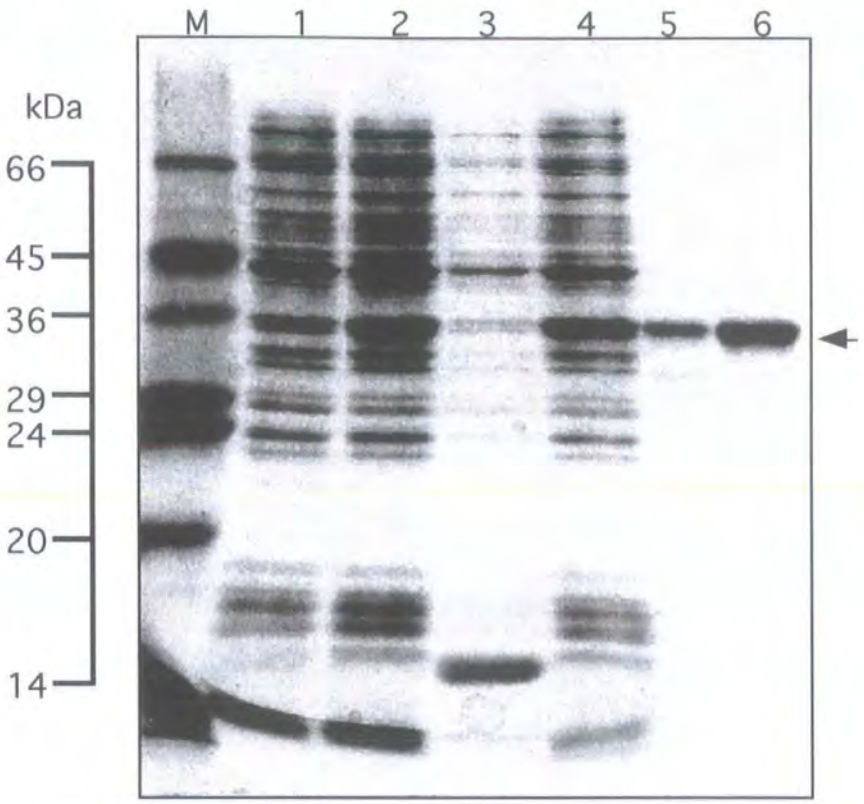


Figure 4.13(a). SDS-PAGE analysis of recombinant DHR protein expressed in BL21 (DE3) STAR (Invitrogen) and purification under denaturing conditions. Protein samples were prepared for SDS-PAGE analysis and resolved on a 12% polyacrylamide gel, proteins were visualised by staining with Coomassie blue. Uninduced cells were grown to an OD λ 600 nm (lane 1) and induced with IPTG (0.1 mM). Following induction cells were grown for a further 3 hours (lane 2). A sample of induced cells was split into soluble (lane 3) and insoluble (lane 4) fractions. Recombinant DHR receptor binding domain was purified under denaturing conditions, as described in section 2.8.6 (lane 5) and refolded by dialysis into 10 mM Tris-HCl (pH 7.5) and finally water (lane 6). SDS-7 marker (lane M) was run along with protein samples to estimate protein molecular weight, each band corresponds to approximately 3 μg protein. The molecular weight of purified recombinant protein was confirmed by SELDI-TOF mass spectrometry (Fig. 4.13b).

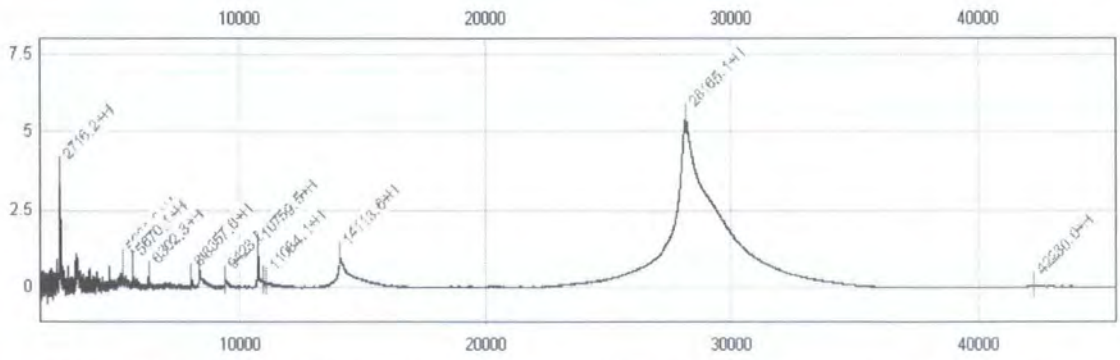


Figure 4.13(b). SELDI-TOF mass spectrometry (MS) analysis of purified recombinant DHR hormone binding domain (Figure 1, lane 6). 5 μg of recombinant protein was applied to an individual spot of the SELDI protein chip, WCX2 (weak cation exchange, Ciphergen). Samples were analysed by SELDI-TOF MS and the purified recombinant protein was confirmed as M_r 28165 Da.

non-specifically bound proteins and the recombinant protein was eluted, and diluted to a concentration of 50 μg/ml (Fig. 4.13a, lane 5). The protein was refolded by dialysis, which resulted in very little protein precipitation (Fig. 4.13a, lane 6). The soluble protein was concentrated by freeze-drying and resuspended at a concentration of 5 mg/ml. Purified recombinant protein was used in *N.lugens* artificial diet feeding trials (Section 4.2.7) and for production of antibodies in rabbits (Section 4.2.9).

4.2.7 Feeding trials with recombinant DHR N-terminal domain

The toxicity of the DHR N-terminal domain towards first instar *N.lugens* was assessed by artificial diet feeding trials. As the N-terminal domain is known to be important for hormone binding, it was anticipated that if recombinant protein could enter the insect haemolymph (and retain its hormone binding properties) it could cause a disruption to insect water homeostasis. Therefore, first instar *N.lugens* were fed purified recombinant N-terminal DHR protein at a concentration of 10 mg/ml (equivalent to 1 % (w/v)) in liquid artificial diet. Insect survival was monitored daily (Fig. 4.14). Insects feeding on artificial diet and artificial diet containing purified recombinant N-terminal DHR, or purified albumin showed similar survival over an 11 day period. Therefore, feeding trials demonstrate that ingestion of purified recombinant *N.lugens* DHR N-terminal domain had no significant toxic effect. As diuretic hormone is known to be present in the haemolymph micro-injection of the recombinant N-terminal hormone-binding domain was also tried. Microinjection experiments resulted in a high insect mortality rate, as a result of the injection procedure. Therefore survival data for microinjected *N.lugens* could not be generated.

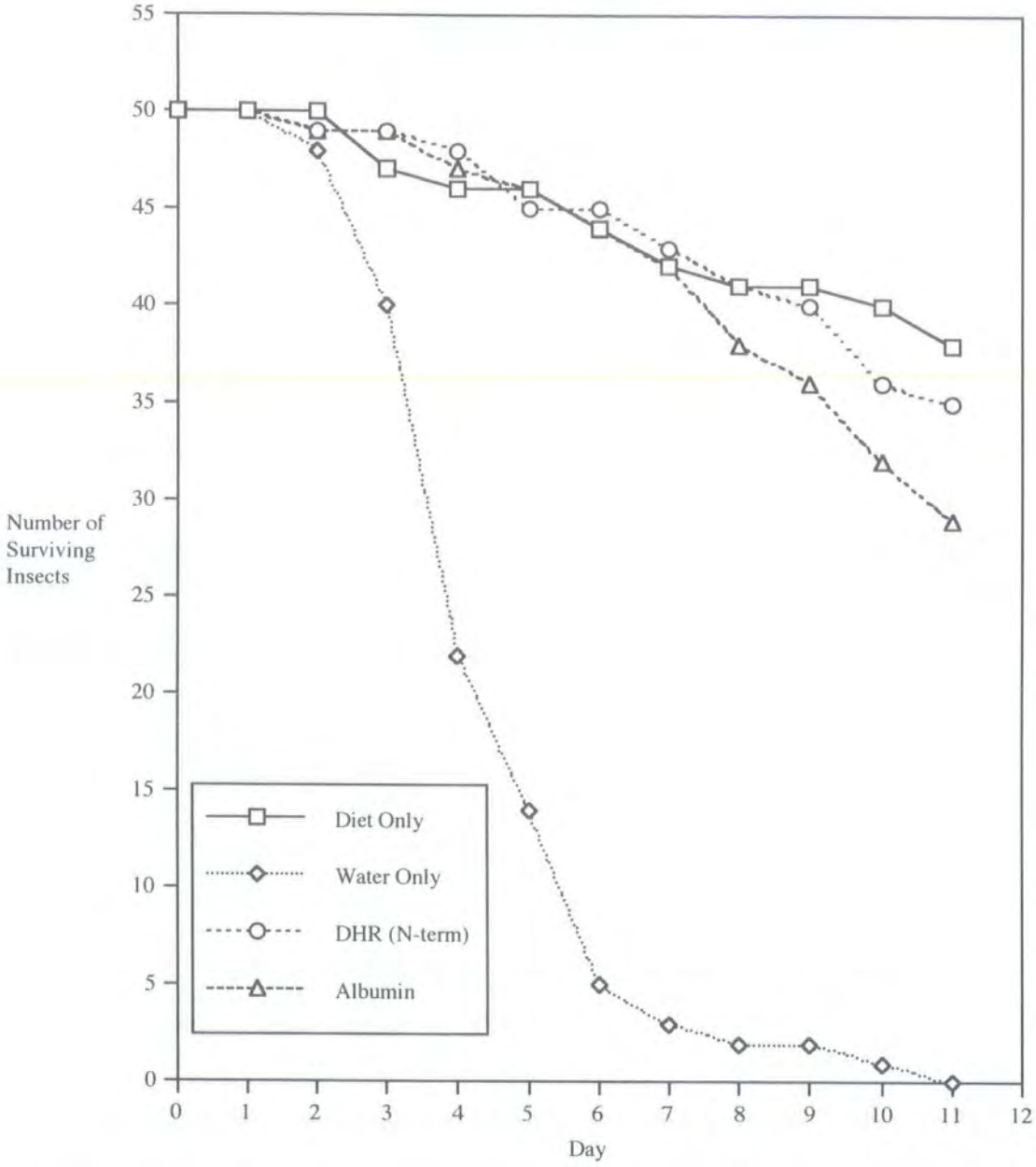


Figure 4.14. Survival of first instar *N.lugens* feeding on diet with purified recombinant N-terminal DHR at a concentration of 10 mg/ml (equivalent to 1% w/v). Control experiments were diet only, diet containing albumin (at a concentration of 10 mg/ml), and no diet. Insect survival was recorded daily. Diet was changed on alternate days to prevent bacterial contamination of diet.

4.2.8 Insect Cell Expression of the Full-Length N.lugens DHR

Attempts were made to functionally express the full-length *N.lugens* DHR in Sf21 insect cells from *Spodoptera frugiperda.* Functional expression of the DHR would aid identification of the peptide hormone using an orphan receptor screening approach. A functionally expressed receptor and the hormone ligand would facilitate *in vitro* screening for potential protein agonists and antagonists. Therefore, confluent Sf21 cells, with a viability >95% were transfected with the pIZ/DHR expression construct, using a lipid mediated protocol (Invitrogen). Cells were also transfected with a positive control, which consisted of CAT (chloramphenicol acetyl-transferase) in the pIZ insect expression vector (Invitrogen). Transient expression of the proteins was monitored at 2 and 4 days post transfection. Harvested cells were split into soluble and insoluble fractions and proteins were separated by SDS-PAGE and stained with Coomassie Brilliant Blue (Fig. 4.15a). Similar gels were run and transferred to nitrocellulose and probed with anti-V5 antibody in western blot experiments (Fig. 4.15b). The V5 epitope is a 10 amino acid fusion tag that is fused to the carboxy terminus of the mature protein that allows detection using the anti-V5 epitope antibody.

Analysis of the western blot shows transient expression of the positive control (CAT), a protein that migrates at approx. 24 kDa on gel (Fig. 4.15b lane, 4 & 6). As expected the positive control is expressed in the soluble fraction. The expression appears strongest at 2 days post transfection (Fig. 4.15b, lane 4) and is significantly reduced at 4 days post transfection (Fig. 4.15b, lane 6). The western blot also demonstrates transient expression of the *N.lugens* DHR in the soluble protein fraction. Partitioning of the membrane

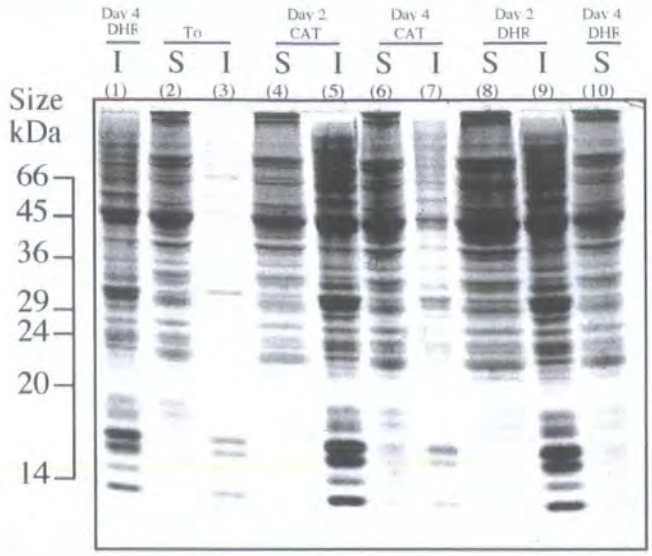


Figure 4.15(a). SDS-PAGE analysis of soluble and insoluble protein extracts from Sf21 insect cell lines expressing DHR (full-length) and a positive control CAT (chloramphenicol acetyl-transferase). Cells were harvested on day 2 and 4 post transfection and prepared for SDS-PAGE on a 12% gel. 10 µg of total cell protein (TCP) was loaded in each lane. Lane loadings are indicated above the appropriate lane (S-soluble fraction and I-insoluble fraction). A cell extract was also taken from non-expressing Sf21 cells at time = 0 (T0). Recombinant protein expression was monitored by transferring a similar gel to nitrocellulose and probing with anti-V5 antiserum (Fig. 4.15b).

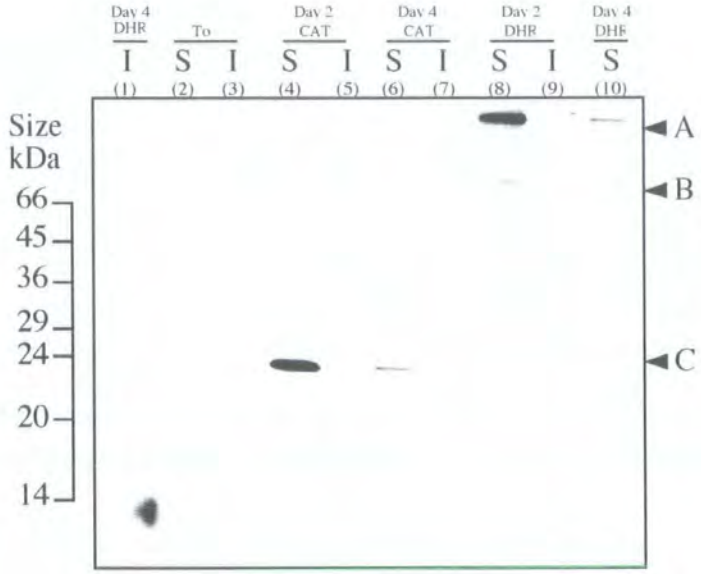


Figure 4.15(b). A similar gel to above was transferred to a nitrocellulose membrane and used in a western blot experiment to test for recombinant protein expression in the Sf21 insect cells. The anti-V5 antiserum (Invitrogen) was used as the primary antibody (dilution 1:10,000) and binding was visualised by using a HRP-labelled secondary antibody and ECL detection. A-immunoreactive material which failed to enter the resolving gel B-immunoreactive material corresponding to the predicted molecular mass of the full length *N.lugens* DHR and C-immunoreactive material corresponding to the appropriate molecular weight of the positive control CAT.

protein to the soluble fraction was expected due to the presence of 1% (v/v) Nonidet-P40 detergent in the protein extraction buffer. A faint band of approx. 69 kDa is observed after 2 days post transfection (Fig. 4.15b, lane 8), which corresponds well with the predicted size of the mature *N.lugens* DHR protein. However, expression is lost at 4 days post transfection (Fig. 4.15b, lane 10). Also, immuo-reactive material that did not enter the resolving gel is also present at 2 days post transfection, and is significantly reduced after 4 days post transfection in the soluble fraction of cells transfected with the DHR pIZ expression construct. It is likely that this immuno-reactive material is due to the transient expression of the *N.lugens* DHR, as background on the western blot is low, and similar bands are not observed in the protein extracts from non-transfected cells (Fig. 4.15b, lane 2 and 3). Attempts were made to generate a cell line that stably expressed the full-length DHR. At 48 hours post transfection, recombinant cells were selected with zeocin, and upon regrowth of the selected Sf21 cells protein expression was analysed by western blot. Although transformants demonstrated antibiotic resistance expression of the recombinant protein could not be detected (data not shown). Similar procedures were followed for cell lines transiently expressing CAT, however, stable cell lines were successfully generated (data not shown). Transfected cell lines were analysed by PCR to try and determine a reason for failure of protein expression (Fig. 4.16). Analysis of transfected cells before and after antibiotic selection demonstrated the stable incorporation of the construct into the insect cell genome. The DHR coding sequence could be amplified from genomic DNA from insect cells before and after antibiotic selection (Fig. 4.16 lane 10 and 11, respectively), but not from untransfected cells (Fig. 4.16, lane 9). Similar results were obtained for the positive

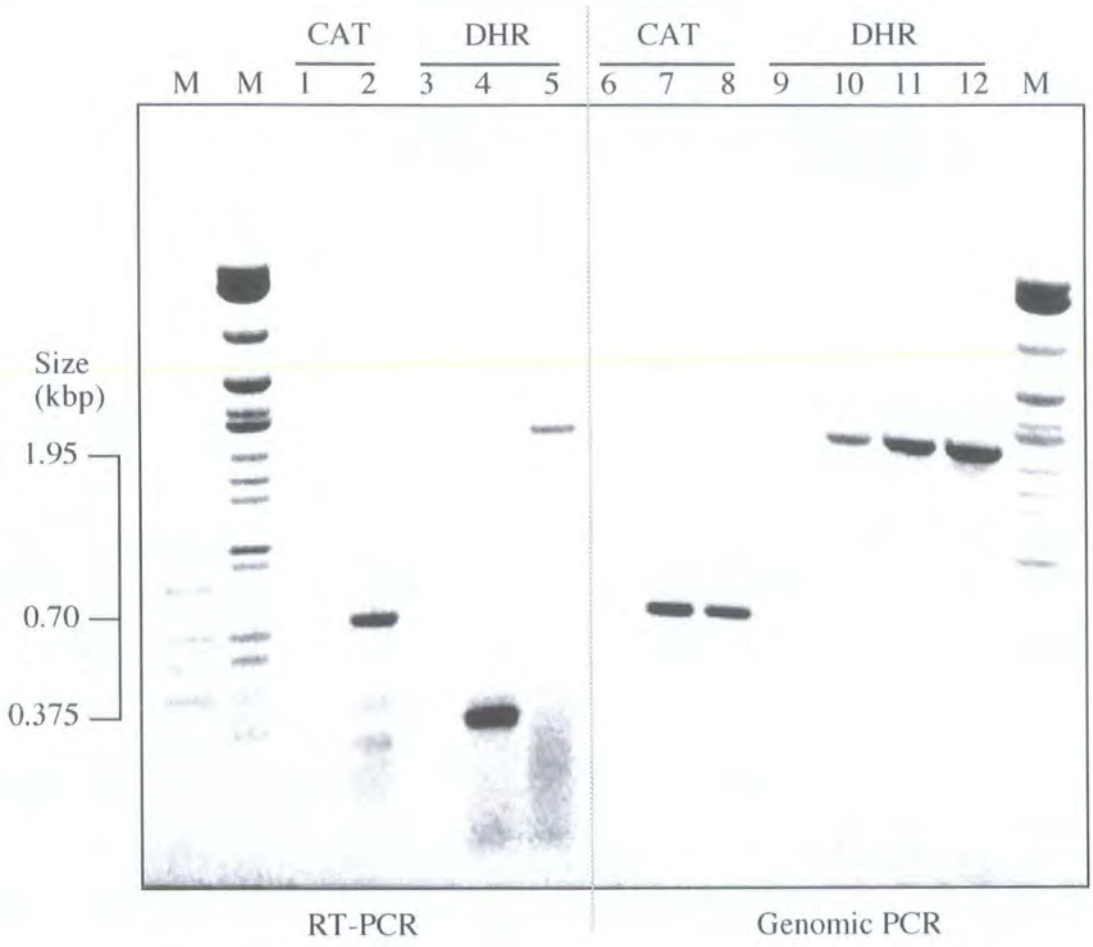


Figure 4.16. Expression analysis of *N.lugens* DHR in a transformed SF21 insect cell line, using genomic PCR and RT-PCR with primers specific to DHR. A positively expressing control CAT was also analysed, again using specific primers. 1,2 (CAT RT-PCR): (1) -ve control, (2) polyclonal expressing cell line; 3,4,5 (DHR RT-PCR): (3) -ve control, (4) 96 hours after transformation with antibiotic selection, (5) immediately after transformation without selection; 6,7,8 (CAT Genomic PCR): (6) -ve control, (7) 96 hours after transformation with antibiotic selection, (8) untransformed plasmid DNA; 9,10,11,12 (DHR Genomic PCR): (9)-ve control, (10) 96 hours after transformation with antibiotic selection, (11) immediately after transformation without selection, (12) untransformed plasmid DNA.

control, CAT. After transfection the CAT coding sequence could be amplified from insect cell genomic DNA, before and after antibiotic selection (Fig. 4.16, lane 7), but not from untransfected cells (Fig. 4.16, lane 6). The same set of primers were used for analysis of the message in RT-PCR experiments on RNA extracted from transfected insect cells. Message for the positive control CAT accumulated as expected, after transfection (Fig. 4.16, lane 2), resulting in the expression of recombinant CAT. However, message corresponding to the *N.lugens* DHR was observed after transfection (Fig. 4.16, lane 5). But upon antibiotic selection the message became significantly truncated (from approx. 2 kbp to 400 bp) and was therefore unable to code for the DHR (Fig. 4.16, lane 4).

4.2.9 Anti-DHR Antibodies

Antibodies were raised in rabbits against purified recombinant N-terminal domain protein (amino acids 1-260) and against a synthetic peptide corresponding to amino acids 238 - 260 of the extracellular N-terminal domain. IgG fractions from independent rabbits were purified by chromatography on immobilised protein G, and both antibodies were concentrated to 10 mg/ml. The purified IgG fractions were tested for specific binding to their 'targets' by western blotting. The putative anti-recombinant DHR antibody and anti-peptide DHR antibody were reacted with purified recombinant DHR N-terminal domain. As expected, antibodies showed binding, with both anti-recombinant DHR antibody and anti-peptide DHR antibody generating a strong and specific signal when used at 1:10,000 dilution, and 1:2,000 dilution, respectively (Fig. 4.17, lane 1). The antibody was also used in western blots on protein extracts from *N.lugens* gut tissues. Both antibodies bound specifically to common polypeptides of

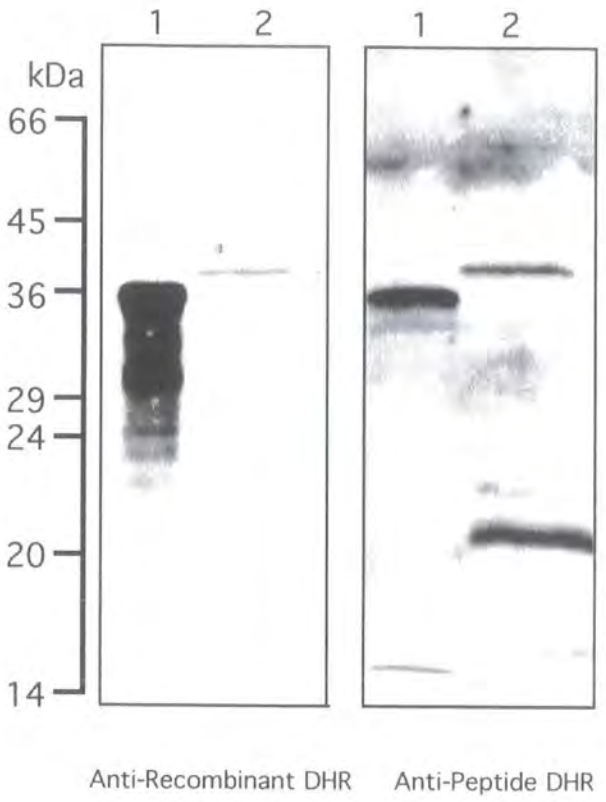


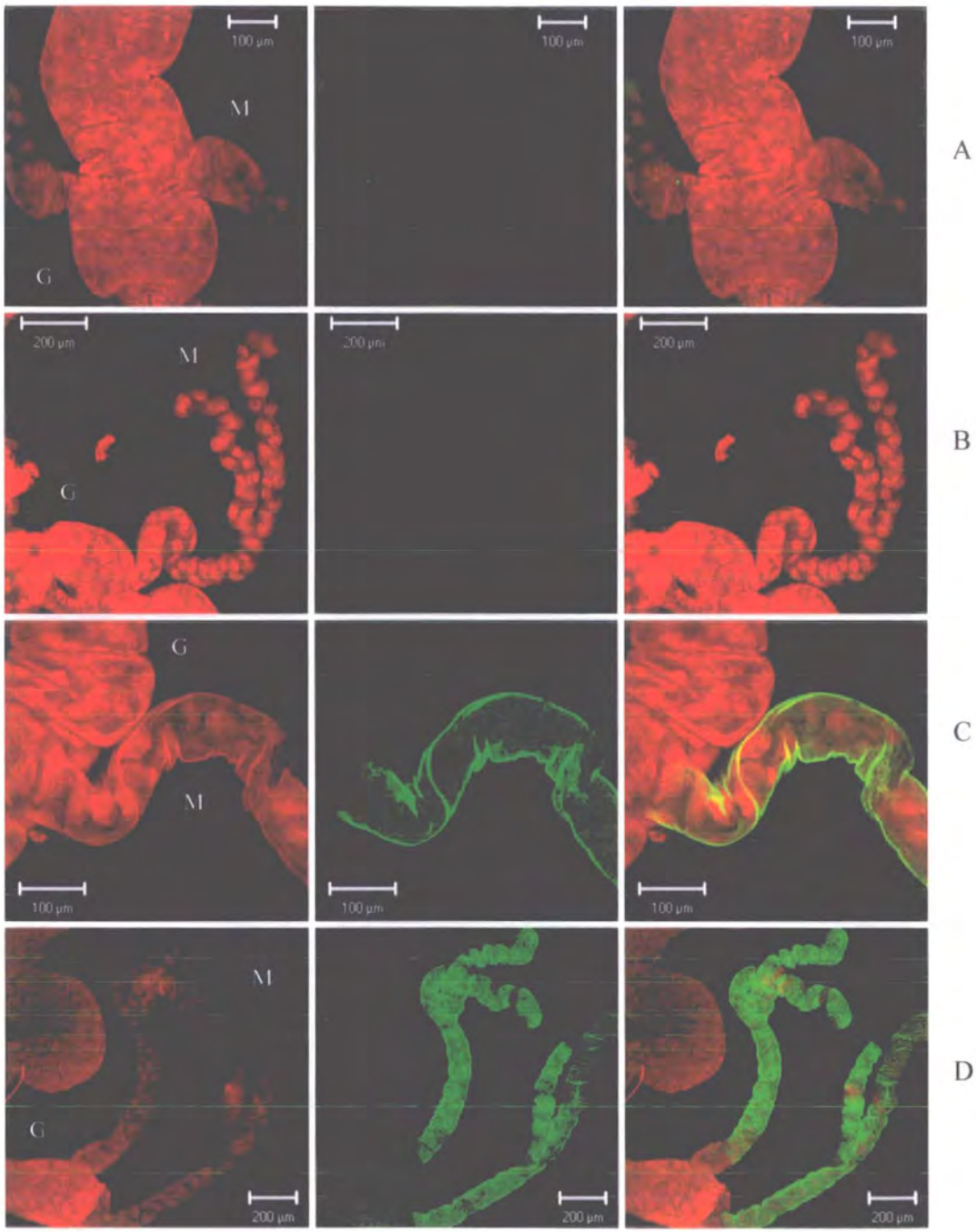
Figure 4.17. Western blot analysis of proteins extracted from *N.lugens* gut tissue and probed with anti-peptide DHR antibody (anti-peptide Ab) and anti-recombinant DHR antibody (anti-recombinant Ab). After electrophoresis the proteins were transferred onto nitrocellulose by electroblotting (0.15 mA, 90 minutes). Proteins were detected by either anti-peptide antibody (Rabbit 1, 1 : 2,000) or anti-recombinant DHR antibody (Rabbit 2, 1 : 10,000). As a positive control 14 ng of recombinant DHR (hormone binding domain) was run along side *N.lugens* gut proteins (lane 1). Primary antibody binding was visualised by using anti-rabbit HRP conjugate and ECL detection, the above film was after exposing photo-sensitive (Fuji-RX) film for 30 seconds.

approx. 40 kDa and at approx. 20 kDa (Fig. 4.17, lane 2). No background non-specific binding to gut proteins was observed on western blots even at low serum dilutions.

4.2.10 Immunolocalisation of DHR in the N.lugens Gut

The interaction of anti-DHR antibody with proteins in the gut of *N.lugens* was further investigated by immunolocalisation (Fig. 4.18). Dissected insect gut tissues were incubated with purified anti-recombinant DHR IgG, and then a fluorescent-labelled secondary antibody. Gut tissue incubated with either primary antibody only or fluorescent secondary antibody only were used as controls. Fluorescence due to bound secondary antibody was visualised using the confocal microscope. Gut tissue incubated with primary antibody only has no green fluorescence, eliminating the possibility of tissue autofluorescence (Fig. 4.18, row A). In contrast gut tissue incubated with both primary and secondary antibody showed strong fluorescence, which was specifically located at the Malpighian tubules. No fluorescence was observed in foregut, mid-gut or hindgut tissue, and a very clear boundary between fluorescence on tubule tissue, and no fluorescence in gut tissue was observed (Fig. 4.18, rows C and D). The gut tissue incubated with fluorescent secondary antibody only showed no significant fluorescence confirming that the strong fluorescence observed is due to anti-recombinant DHR antibodies specifically binding DHR rather than non-specific binding of the secondary antibody (Fig. 4.18, row B). The binding of anti-DHR IgG observed in this experiment suggests that the 'target' of the antibody is on the extracellular surface of the Malpighian tubules, and therefore exposed to the haemolymph *in vivo*.

Figure 4.18. Immunolocalisation of DHR in *N.lugens* gut tissue. Dissected gut tissues were incubated with primary anti-DHR antibody and fluorescent secondary antibody and observed under the confocal microscope (rows C and D). Green fluorescence represents antibody binding; red fluorescence is background counterstaining of the tissue using ethidium bromide. Similar experiments using either primary (row A) or secondary (row B) antibodies independently provided a control for auto-fluorescence and non-specific binding of the fluorescent secondary antibody. Images from the red and green channel were recorded independently and digitally overlaid to produce a final image. M-Malpighian tubule and G-midgut.


100 µm
M
G
A
200 µm
B
C
D
RED
GREEN
RED & GREEN MERGE

4.2.11 Feeding trials with anti-DHR antibody

Feeding trials were set up to investigate whether ingested anti-recombinant DHR antibodies were detrimental to insect survival, and if orally administered antibodies were capable of interacting with the *N.lugens* DHR *in vivo*. Insects were fed anti-DHR IgG at a concentration of 0.1 mg/ml (equivalent to 0.01% w/v) in liquid artificial diet, for up to 8 days. Controls were diet with no antibody, diet containing goat anti-rabbit IgG, and no diet. All treatments except the no diet control showed similar survival, and thus anti-recombinant DHR antibodies administered via liquid artificial diet had no significant toxic effect (Fig. 4.19). During the feeding trial honeydew samples from feeding insects were collected, and after the feeding trial haemolymph samples were taken from surviving insects. Samples were analysed by western blot (Fig. 4.20). Anti-recombinant DHR was detected in the honeydew of feeding insects at a similar level to the artificial diet, showing that the insects had ingested the antibody and that it had passed through the gut (Fig. 4.20, lanes 1 and 2). However, no antibody was detected in the haemolymph fraction (Fig. 4.20, lane 3); suggesting that it had not crossed the gut wall. Insects fed anti-DHR antibody followed by fluorescent labelled secondary antibody showed no fluorescence in the gut or Malpighian tubules after a chase period with diet (result not presented) suggesting that the anti-recombinant DHR antibody had not bound to gut tissue.

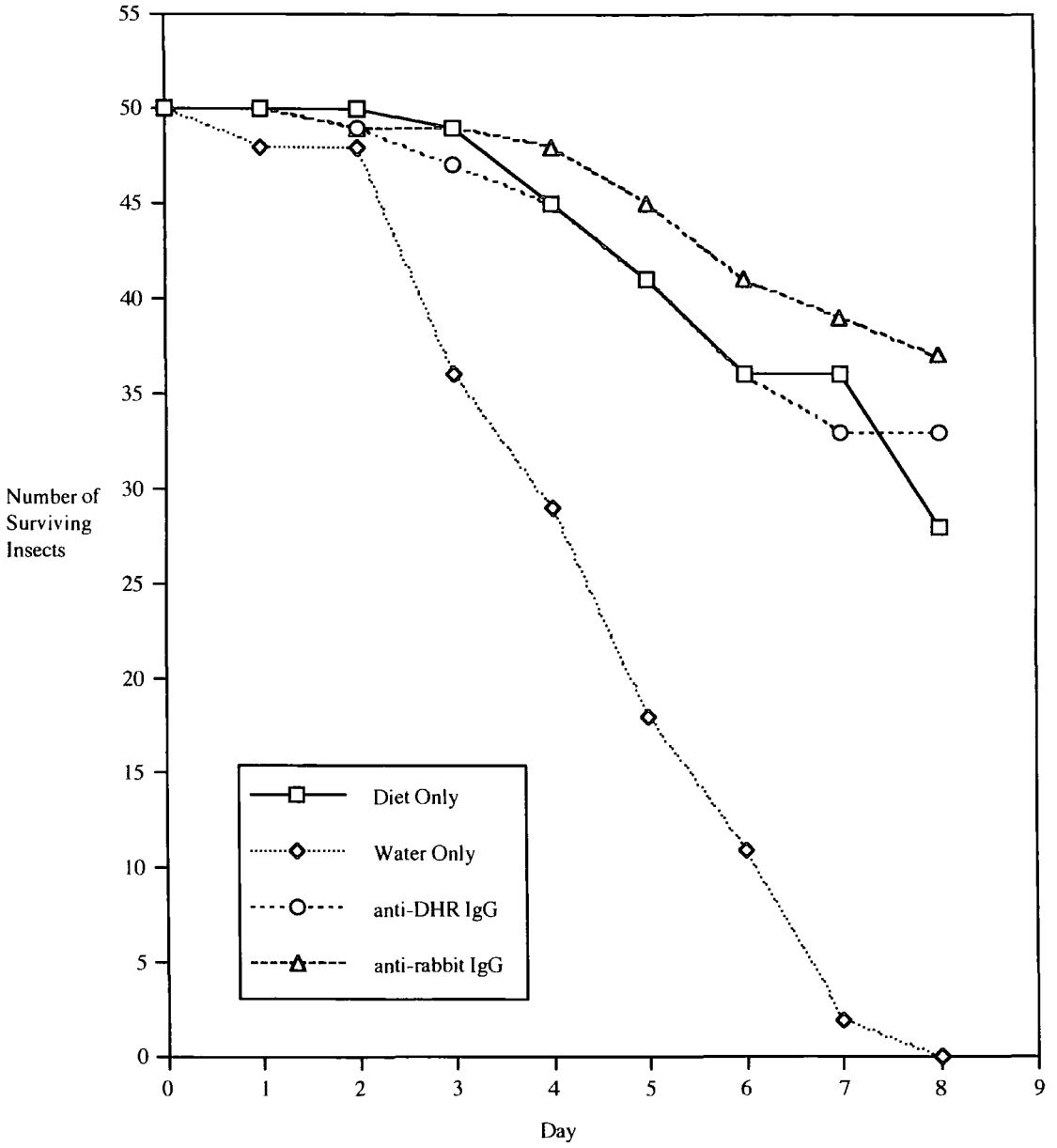


Figure 4.19. Survival of first instar *N.lugens* feeding on diet with anti-recombinant antibody incorporated at a concentration of 0.1 mg/ml (equivalent to 0.01% w/v). Control experiments of diet with no antibody, diet containing goat anti-rabbit IgG, and no diet were set up and run in parallel. Insect survival was recorded daily. Diet was changed on alternate days to prevent bacterial contamination of diet.

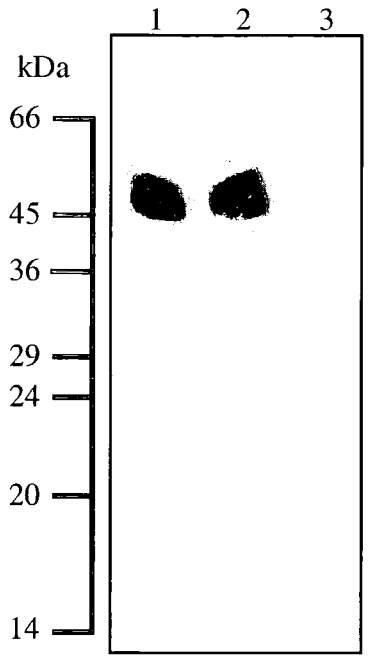


Figure 4.20. Western blot analysis of honeydew (lane 2) and haemolymph (lane 3) from *N.lugens* feeding from liquid artificial diet containing anti-DHR antibody (0.1 mg/ml, lane 1). Hoppers were fed liquid artificial spiked with anti-DHR antibody diet for 48 hours and then honeydew and haemolymph samples were analysed by western blot. Samples were resolved by SDS-PAGE and proteins were transferred onto nitrocellulose by electroblotting. Anti-DHR antibodies were visualised by using anti-rabbit HRP conjugate and ECL detection, the above film was after exposing photo-sensitive (Fuji-RX) film for 30 seconds.

4.3 Discussion

4.3.1 Cloning N.lugens *DHR*

The degenerate primers that were designed against highly conserved residues in TM 3 and TM 6 of aligned diuretic hormone receptor proteins were successful in amplifying the *N.lugens* homologue. A similar degenerate primer approach has been already used for the retrieval of the *Bombyx mori* DHR cDNA (Ha *et al.*, 2000) and the *Acheta domesticus* DHR cDNA (Reagan, 1996). However, the primers were based solely on the *Manduca sexta* coding sequence. As the set of degenerate DHR primers designed here (Section 3.2.3) reflects codon usage from a wide variety of insects it is likely that the primers could be used on other insect species for amplification of an internal fragment of the DHR. Amplification resulted in a single product, which upon sequencing was shown to have sequence similarity to other DHR proteins. With the amplification of an internal fragment of the *N.lugens* DHR two approaches were amenable for the retrieval of the full-length coding sequence. These were library screening and RACE PCR. Due to the anticipated rarity of the message the RACE approach was preferred over library screening, as the PCR step would enrich the DHR cDNA population and facilitate cloning. Therefore, using the partial sequence data gene specific primers were designed for use in RACE experiments. The full-length receptor coding sequence plus untranslated regions was retrieved by 5' and 3' RACE PCR, sequences were aligned on a computer to produce a contiguous sequence.

4.3.2 N.lugens *DHR Sequence Analysis*

Sequence analysis demonstrates that the *N.lugens* DHR is a member of the calcitonin/secretin/CRF subfamily of G-protein coupled receptors (GPCR), or family-B peptide receptors. The first insect receptor identified as a member of this family was a diuretic hormone receptor (DHR) from *Manduca sexta* (Reagan, 1994) and now DHRs from *Acheta domesticus* (Reagan, 1996) and *Bombyx mori* (Ha *et al.*, 2000) have been described. This family of receptors is characterised by long N-terminal domains with five well-conserved cysteine residues. The high degree of conservation among these key residues suggests that they have an essential role for either structural of functional integrity of these receptor proteins. Presumably these cysteine residues form a network of disulphide bridges, which may be important in stabilisation of the large extracellular N-terminal domain. Cysteine residues are also conserved in the second and third extracellular loop regions, these putative disulphide forming cysteines are the only residues well conserved across all families of GPCRs.

The *Drosophila* genome (http://flybase.bio.indiana.edu/) contains five receptor proteins with similarity to vertebrate corticotropin releasing factor and calcitonin receptors (family-B receptor proteins). These include gene products, CG8422, CG12370, CG17415, CG4395 and CG13758, although it is unclear whether CG13758 is a member of group I, II or III (Brody and Cravchik, 2000; Hewes and Taghert, 2001). Two receptors from this group, encoded by CG8422 and CG12370 are related to insect diuretic hormone receptors. The high degree of sequence identity (57%) between CG8422 and CG12370 is suggestive of gene duplication, and therefore a redundancy in

function (Hewes and Taghert, 2001). Multiple DHR sequences have not been reported for any other insect species. Results reported here suggest that *N.lugens* has two DHR cDNAs (DHR-I and DHR-II) (Fig. 4.5a), although they are more similar to each other than CG8422 and CG12370 in *Drosophila*. Sequence analysis between *Drosophila* Family-B peptide receptors and DHR proteins from *Manduca, Bombyx, Acheta* and *N.lugens* reveal strong sequence conservation between receptors, with sequence identity up to 95% between any two receptors (Fig. 4.6a). The *N.lugens* DHR has greatest sequence similarity to the *Bombyx mori* DHR with an overall sequence similarity of 51%. Greatest sequence conservation between receptors occurred within the transmembrane domains, in particular TM 3 and TM 6. The large N-terminal domain is a characteristic feature of this protein family, where it is thought to be important in hormone binding, although sequence conservation within this domain is relatively low. The *N.lugens* N-terminal domain (amino acid 1-260) contains the 5 well-conserved cysteine residues, but it is significantly longer than any of the other reported DHR sequences. The *Drosophila* gene product CG4395 has an extracellular N-terminal domain of 196 amino acids, the closest match in length, however phylogenetic analysis based on the primary structure of *Drosophila* Family-B receptors and other insect DHRs (Fig. 2) reveals that CG4395 is most similar to CG17415, and forms a group distinct from the DHR-type receptors. *Drosophila* gene product CG4395 has limited sequence identity (16% sequence similarity) with the *N.lugens* DHR. Therefore, based on sequence identity, the cDNA cloned from *N.lugens* has been classified as a putative diuretic hormone receptor, although its hormone ligand remains unidentified at present. Every receptor in this family is encoded on multiple exons, and several of these genes

are alternatively spliced to yield functionally distinct products. However, there was no evidence to suggest that the *N.lugens* DHR is alternatively spliced, which is also the case for other reported insect DHRs.

4.3.3 Ligand Binding and Receptor Activation

The putative *N.lugens* diuretic hormone receptor is a member of the G protein-coupled receptor (GPCR) superfamily. GPCRs have been named based on their ability to recruit and regulate the activity of intracellular heterotrimeric G proteins. GPCRs interact with heterotrimeric G proteins and initiate guanine-nucleotide exchange. The activated receptor induces a conformational change in the associated G protein α-subunit leading to a release of GDP followed by binding of GTP (Bourne *et al.*, 1991). Subsequently, the GTP-bound form of the α-subunit dissociates from the receptor as well as from the stable βγ-dimer. Both the GTP-bound α-subunit and the released βγ-dimer can modulate several cellular signalling pathways. These include, among others, stimulation or inhibition of adenylate cyclase and activation of phospholipases, as well as the regulation of potassium and calcium channel activity (Hamm, 1998). Analysis of functionally expressed insect diuretic hormone receptors demonstrates that receptor activation leads to stimulation of an intracellular adenylate cyclase, which results in an intracellular increase in cAMP (Reagan, 1994; Reagan, 1995a; Reagan, 1996). Diuretic hormone binding its receptor is proposed to result in the activation of a Na^+-K^+-$2Cl^-$ cotransporter, which transports ions from the haemolymph to the lumen of the Malpighian tubule (Reagan, 1995b; Reagan, 1995c).

The complexity of GPCR signalling has been further underlined by data indicating that GPCRs may not act solely via heterotrimeric G-proteins. Following hormone-binding desensitisation of some receptors occurs and the receptor is no longer capable of responding to hormone. This process is mediated by two protein families: G protein-coupled receptor kinases (GRKs) and arrestins. GRKs specifically bind to the agonist-G protein coupled receptor complex, promoting receptor phosphorylation, which in turn leads to arrestin binding. Arrestin binding inhibits the receptor and G protein interaction, leading to desensitization. Furthermore, many GPRs are then removed from the plasma membrane via clathrin-mediated endocytosis. Therefore, the receptor is no longer capable to respond to hormone (Krupnick and Benovic, 1998). As numerous phosphorylation sites were predicted at the intracellular C-terminal region of the putative *N.lugens* diuretic hormone receptor, it is possible that unconventional signalling pathways exist for the putative *N.lugens* diuretic hormone receptor.

Numerous studies have been carried out to identify domains involved in ligand binding to various subclasses of GPCRs. Only recently have we have gained insight into ligand-binding domains in peptide receptors, which are of interest due to the discovery of small molecule non-peptide ligands that can act with high potency at peptide receptors as antagonists and agonists. The ligand binding sites for peptide ligands in family B receptors involve the extracellular domains. The large amino terminus that characterises family B receptors seems to play a key role for most ligands, including secretin (Holtmann *et al.*, 1996), calcitonin (Stroop *et al.*, 1996), and CRF (Liaw *et al.*, 1997; Dautzenberg *et al.*, 1998). However, the amino terminus is not sufficient for binding

these ligands and additional interactions are found in the extracellular loops (Holtmann *et al.*, 1996; Stroop *et al.*, 1996; Liaw *et al.*, 1997; Dautzenberg *et al.*, 1998). Due to the structural similarities between these receptors and related insect receptors it is likely that loop regions are also important for the binding of peptide hormones in related insect hormone receptors.

4.3.4 N.lugens *Feeding Trial with Purified Recombinant DHR N-terminal domain*

Oral ingestion of purified recombinant DHR N-terminal domain exhibited no toxic effect when fed to first instar *N.lugens* over an 11 day artificial diet feeding trial (Fig. 4.14). As insect diuretic hormone is known to be present in the haemolymph (Coast, 1996; Coast, 1998) it is likely that orally delivered DHR N-terminal domain is incapable of interaction with its corresponding hormone. Therefore, injection of recombinant DHR N-terminal domain into *N.lugens* haemolymph was attempted, however, as a consequence of the high insect mortality rate (due to the microinjection procedure) survival data could not be generated. If this microinjection experiment was repeated an insect species more amenable to experimental manipulation would be chosen, such as *Manduca sexta.* Diuretic hormone (Kataoka *et al.*, 1989) and the receptor (Ha *et al.*, 2000) have been well characterised in this insect, making such an approach a viable option for validation of this receptor protein as a potential target. Also, the lectin *Galanthus nivalis* agglutinin (GNA) has been shown to have the capability to enter the insect haemolymph, and furthermore to act as a carrier protein for the delivery of small peptides fused to the GNA protein (Fitches *et al.*, 2002). As the *N.lugens* DHR N-terminal domain is situated on the haemolymph side of the gut, delivery of protein antagonists or agonists may be accomplished through fusions with GNA.

4.3.5 N.lugens *DHR Coding Sequence Expression*

Functional expression of receptor proteins in insect cell lines is well documented (McCarroll and King, 1997; Towers and Sattelle, 2002), however, a cell line stably expressing *N.lugens* DHR could not be generated. In both the CAT and DHR transfection experiments protein expression was observed at 2 days post transfection, and was significantly reduced at 4 days post transfection. This reduction in transient expression is expected, as non-expressing cells continue to grow and dilute the recombinant protein (McCarroll and King, 1997; Pfeifer *et al.*, 1997; Hegedus *et al.*, 1998; Towers and Sattelle, 2002). In order to select stable cell lines transformants were selected with zeocin. For the positive control this resulted in a stable cell line expressing CAT protein. When selection was applied to the DHR transformants protein expression was lost, even though there were resistant insect cells capable of regrowth. Therefore, the accumulation of mRNA corresponding to CAT and DHR was analysed by PCR before antibiotic selection and post antibiotic selection. It was apparent, for unknown reasons, that upon selection the DHR transformants were truncating the DHR message, therefore, the cells were unable to express functional DHR.

4.3.6 Anti-DHR Antibody Characterisation

Antibodies raised against recombinant DHR (N-terminal domain) and against the DHR peptide (amino acid 237-260) bound specifically to polypeptides from *N.lugens* gut tissue (Fig. 4.19). The anti-peptide antibody bound to polypeptides of M_r 41,000, 22,000 and 20,000, with the latter giving the strongest signal. The anti-recombinant DHR antibodies also bound to the 41,000 M_r and 20,000 M_r polypeptides, although binding to the latter was weak. In addition they bound to a polypeptide of approximately M_r

40,000. Since both antibodies bound to common polypeptides in the *N.lugens* gut it was concluded that the 41,000 M_r band must represent DHR. Cross reactivity with a protein corresponding to the full length DHR (M_r 69 kDa) was not observed for either antibody. A possible reason for this discrepancy is cleavage of the protein by proteinases present or released during preparation of *N.lugens* gut proteins, resulting in the liberation of a fragment containing the N-terminal domain, which would be similar in size on SDS-PAGE to the recombinant protein encoded by the N-terminal domain expression construct. The slight increase in M_r could be accounted for through glycosylation of the N-terminal domain *in vivo*. The *N.lugens* N-terminal DHR hormone-binding domain has 4 potential N-glycosylation sites matching the N-X-S/T consensus at Asn 7, Asn 221, Asn 229 and Asn 234 (Fig. 4.5a). Multiple glycosylation sites are found in the N-terminal domain of other reported insect DHR sequences, and glycosylation may be important for expression and function of hormone receptor proteins (Russo *et al.*, 1991).

4.3.7 Immunolocalisation of N.lugens *DHR in Dissected Gut Tissue*

Immunolocalisation experiments using anti-recombinant DHR antibody gave a positive signal upon immunofluorescence microscopy of dissected *N.lugens* gut tissue, under nonpermeabilizing conditions. Although proteins may be partially denatured by the mild fixation procedure, the results indicate that the N-terminal domain of the DHR is exposed on the cell surface and is therefore accessible to antibody in dissected gut tissue *in vitro* or haemolymph borne diuretic hormone peptides *in vivo*. The *N.lugens* DHR is expressed in a tissue specific manor and is present only in the Malpighian tubules of dissected gut tissue. The extracellular N-terminal domain is exposed on the cell surface where it is thought to interact with haemolymph borne diuretic hormones. Diuretic

hormones from insects have significant sequence homology to the vertebrate CRF-like peptides. The interaction between CRF-like diuretic peptide and diuretic hormone receptor has been studied in *Acheta domesticus*, using isolated Malpighian tubules (Coast and Kay, 1994) and functionally expressed receptor (Reagan, 1996). Insect CRF-like diuretic peptides are synthesised in the neurosecretory cells of the pars intercerebralis and are released into the haemolymph via the corpora cardiaca (Coast, 1996). This leads to the prediction that the N-terminal domain of the *N.lugens* DHR is exposed to the haemolymph, where it interacts with haemolymph borne CRF-like diuretic peptides, in agreement with the conclusions presented here. The specific localisation to Malpighian tubules is in agreement with the conclusion that this cDNA does encode a DHR, because the tubules are specifically concerned with water balance.

4.3.8 anti-DHR Feeding Trials

The appearance of functionally active ingested antibodies in the haemolymph has been reported in the whitefly, *Bemisia tabaci* (Morin *et al.*, 1999) and in the European corn borer, *Ostrinia nubilalis* (BenYakir and Shochat, 1996). Therefore, antibodies were delivered to *N.lugens* in liquid artificial diet feeding trials to determine whether an interaction between ingested antibody and DHR would occur, and if such an interaction would be detrimental to insect survival. Antibodies delivered to insects via artificial diet feeding trials had no significant effect on insect survival (Fig. 4.21). This is due to either of the following possibilities, anti-recombinant DHR antibody is interacting with the receptor but has no effect on insect survival; or, no interaction is occurring between the antibody and receptor. To investigate this further the fate of ingested antibodies was monitored in insects feeding on liquid diet containing anti-DHR IgG (Fig. 4.21).

Analysis of the honeydew of *N.lugens* feeding on anti-DHR antibodies proves that both ingestion and excretion of the IgG protein occurs, and the antibody remains intact due to the low level of protease activity in the gut of the *N.lugens* (Foissac *et al.*, 2002). However, there was no detectable IgG in the haemolymph of *N.lugens*. Therefore the DHR N-terminal domain is most likely inaccessible to ingested antibodies in the *N.lugens*. An effective delivery system needs to be developed so an interaction between a potential agonist or antagonist and DHR can be established. A baculoviral delivery system has been used with the lepidopteran, *Manduca sexta* (tobacco hornworm). A baculovirus that has been genetically modified to express *Manduca sexta* diuretic hormone causes a strong alteration in *M.sexta* larval fluid homeostasis, resulting in an increased insect mortality rate (Maeda, 1989). These results demonstrate that a modulation of fluid homeostasis has a significant effect on insect survival, and therefore the DHR still provides an attractive target for the development of novel insecticides. The discovery of an *N.lugens* DHR presents a target for the development of novel insecticidal proteins, which may prove useful in controlling this insect species and limiting the damage caused to rice crop.

CHAPTER 5

THE *N.LUGENS* ALLATOSTATIN RECEPTOR

5.1 Introduction

Using a similar methodology to that described in the previous chapter degenerate primers were used to retrieve a putative allatostatin receptor (ALSTR) from *N.lugens* gut specific cDNA. Based on similarity, the amplified sequence has been classified as a putative allatostatin receptor (ALSTR), although its hormone ligand remains unidentified at present. The ALSTR plays a central role in the regulation of moulting in insects (Section 1.10.6), therefore this receptor has been classified as a potential 'target' for the development of insecticidal proteins. The cDNA cloning of such a receptor is fundamental for the future design of potential insecticidal proteins, which may be used as part of a control strategy when expressed as a transgene in target crops.

5.2 Results

5.2.1 Isolation of a cDNA encoding a ALSTR from the gut of N.lugens

Degenerate primers, and nested degenerate primers corresponding to conserved residues from transmembrane helices three and seven of aligned insect ALSTRs and related *Drosophila* family-A peptide receptors were synthesised (Section 3.2.4). Primers were first optimised on *Drosophila* ALSTR-1 cDNA (DAR-1) cDNA (CG2872), which was a kind gift from Dr Hans-Jürgen Kreienkamp, University of Hamburg. Both pairs of forward and reverse degenerate primers were used in all possible combinations on the

Drosophila template (Fig. 5.1a). All amplification reactions produced specific products of the predicted size (approx. 500 bp), therefore the degenerate primers were deemed effective in the amplification of the *Drosophila* allatostatin receptor-1 clone (Fig. 5.1b, lanes D1, D2, D3 and D4).

PCR reaction conditions that were optimised on the *Drosophila* cDNA were used on *N.lugens* gut specific cDNA, using the primers in the same combination (Fig. 5.1a and b). Again, all amplification reactions produced products of the expected size (Fig. 5.1b, lanes N1, N2, N3 and N4). The amplification products were deemed specific as reactions with no primers (Fig. 5.1b, lane C1), and single primers (Fig. 5.1b, lanes C2, C3, C4) did not result in amplification. The longest *N.lugens* PCR product was a result of amplification between ALSTR degenerate (fwd) and ALSTR degenerate (rev) (Fig. 5.1b N4), therefore this amplification product was purified from the gel, cloned and sequenced. Comparison to proteins in the GenBank database revealed the amplified product had greatest similarity to insect allatostatin receptor proteins, in particular the allatostatin receptor from the American cockroach, *Periplaneta americana*, AF336364 (Auerswald *et al.*, 2001). Gene specific primers were designed from regions of the putative ALSTR fragment and used in rapid amplification of cDNA ends (RACE) PCR for retrieval of the 5' and 3' end of the gene (Table 2.3). Based on estimates from the similar *Periplaneta americana* cDNA the 5' RACE product was predicted to be 1008 bp in length, and the 3' RACE product 333 bp; plus additional untranslated regions (UTR) at either end.

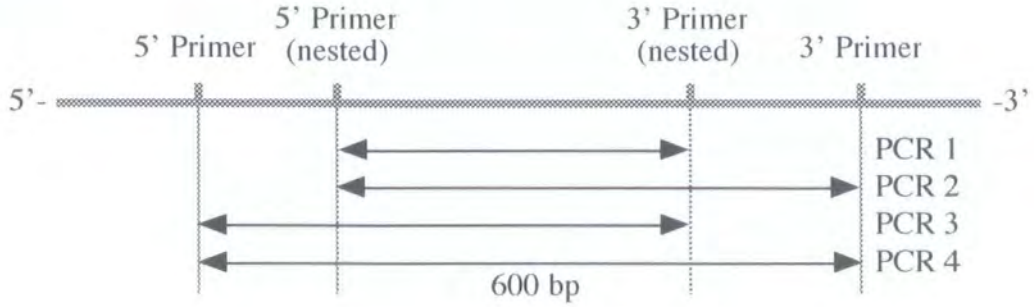


Figure 5.1(a). Location of degenerate allatostatin receptor (ALSTR) primers within the *Drosophila* ALSTR gene, CG2872. PCR conditions were optimised on *Drosophila* cDNA and then applied to *N.lugens* insect cDNA (Fig. 5.1b). Primer combinations used in the amplification reaction are shown above (PCR 1-4).

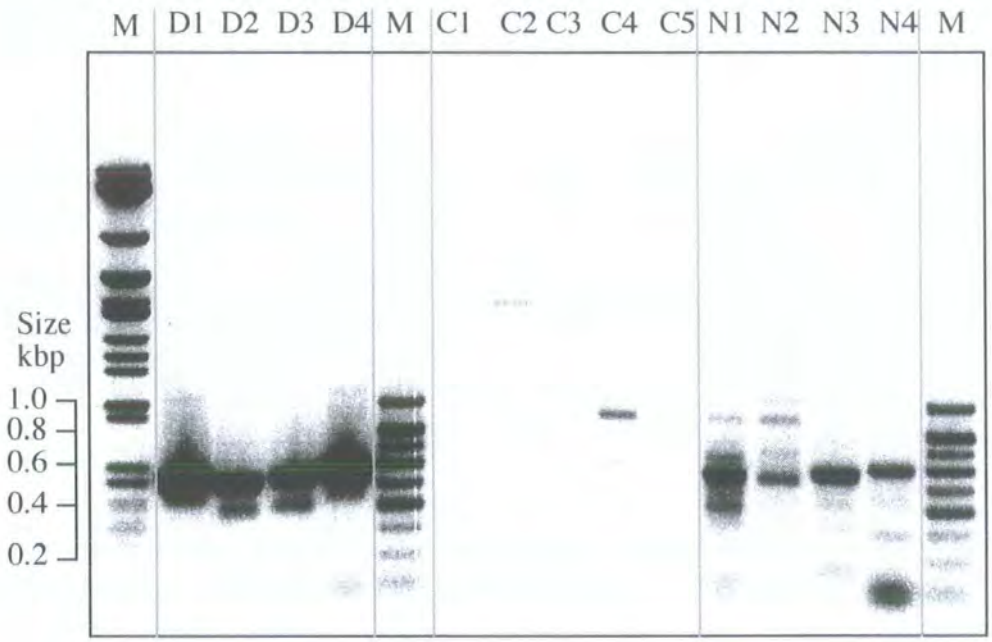


Figure 5.1(b). PCR amplification of ALSTR fragment from a *Drosophila* cDNA (CG2872) and total BPH cDNA using a degenerate primer set. Primers were used in combinations shown in Fig. 5.1(a) (PCR 1-4) on ***D****rosophila* cDNA template (**D**1-4) and on ***N****.lugens* first strand cDNA (**N**1-4). To check the PCR amplification was specific 5 negative control reactions were included: C1, no primers; C2, 3' nested primer only; C3, 3' primer only; C4, 5' nested primer only; C5, 5' primer only. The 600bp band from the B4 reaction was purified from the gel, cloned into pCR2.1 and sequenced.

Primers were used in RACE reactions on prepared *N.lugens* cDNA, according to the manufacturers protocol (Clontech). The 5' RACE reaction yielded a strong product of approx. 1100 bp (Fig. 5.2, lane 4). Similar products did not appear in any of the control PCR reactions using single primers (Fig. 5.2, lane 1 and 2), therefore the reaction product was deemed specific and was cloned and sequenced. The initial 3' RACE reaction resulted in non-specific amplification of the target cDNA (Fig. 5.3, lane 1). Therefore, an aliquot of the reaction was diluted and reamplified using a nested GSP (which was the reverse and complement of the 5' RACE primer). Reamplification of the initial 3' RACE PCR produced strong products at approx. 350 bp and 500 bp (Fig. 5.3, lane 2). The two reaction products appeared to be a result of specific amplification, as products of similar sizes were absent from the single primer control reactions (Fig. 5.3, lane 3, 4, 5 and 6). Therefore the 350 bp and 500 bp reaction products were purified from the gel, cloned and sequenced. 5' and 3' RACE sequences were aligned on a computer to produce a contiguous sequence. BLAST searches of GenBank showed this protein had greatest sequence similarity (69 %) to the ALSTR from *Periplaneta americana* (Auerswald *et al.*, 2001).

Based on RACE sequence data, full-length ALSTR cDNAs amplified from *N.lugens* gut specific cDNA were of 1600 bp, 1510 bp, 1427 bp and 1367 bp. In all cases the coding sequence was identical, however the 3' untranslated region differed in length and sequence. In total 4 clones with unique 3' UTR regions were discovered (Fig. 5.4a; a, b, c and d). For each of the 4 independent clones the longest open reading frame (ORF) consists of 1152 bp of identical sequence, which encodes a protein of 384 amino acids

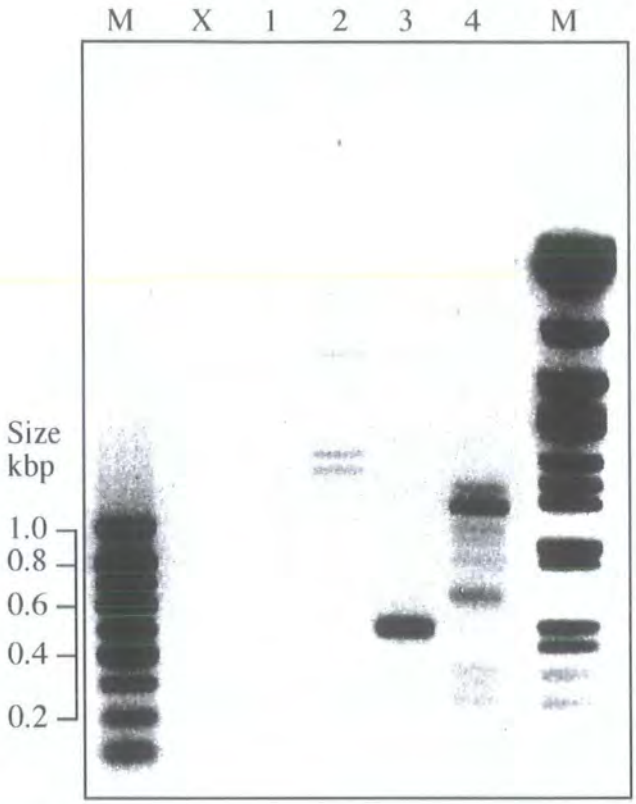


Figure 5.2. 5' RACE (rapid amplification of cDNA ends) on *N.lugens* cDNA using an ALSTR gene specific primer (GSP) and a universal primer (UP) situated at the extreme 5' end of the *N.lugens* cDNA. The GSP corresponded to position 951-977 of the *N.lugens* ALSTR cDNA (ALSTR 5' RACE, Table 2.3) and the universal primer common to all full-length cDNAs. Amplification between ALSTR 5' RACE and the 5' universal primer produced a product of 1.2 kbp (Lane 1), amplification of an internal 500 bp fragment confirmed the integrity of the cDNA (Lane 2). Primer pairs were checked for miss-priming by incorporation of a single primer in a similar PCR reaction, GSP only (Lane 3) or universal primer only (Lane 4). The 1.2 kb amplification product was cut from the gel, purified and TA-cloned into the TOPO vector (Invitrogen) Note: X denotes an empty lane.

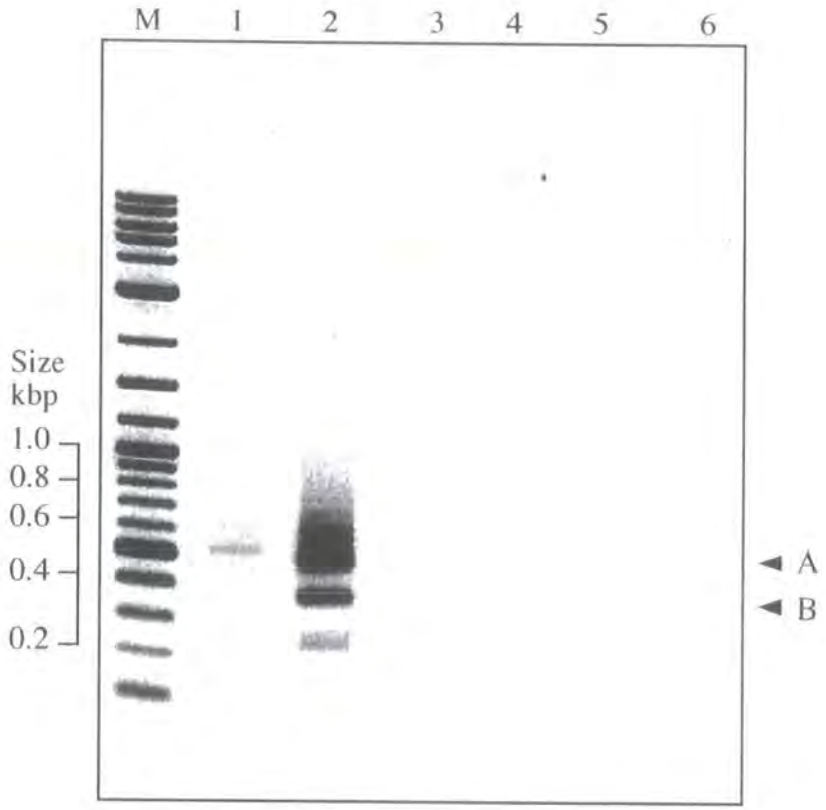


Figure 5.3. 3' rapid amplification of cDNA ends (RACE) on *N.lugens* cDNA using an ALSTR gene specific primer (GSP) and a universal primer (UP) situated at the extreme 3' end of the *N.lugens* cDNA. The GSP corresponded to nucleotide position 907-935 of the *N.lugens* ALSTR cDNA (ALSTR 3' RACE, Table 2.3) and the universal primer common to all full-length cDNAs. cDNAs were initially amplified by PCR using ALSTR 3' RACE and UP (Lane 1), following amplification an aliquot of the reaction was removed and re-amplified using nested primers. The nested GSP corresponded to nucleotide position 952-977 of the *N.lugens* ALSTR cDNA (ALSTR 3' RACE nested) while the nested-UP (NUP) was common to all full-length cDNAs. Reamplification using ALSTR 3' RACE nested and NUP resulted in products of approximately 450 bp (A) and 350 bp (B). Amplification products A and B were cut from the gel cloned into pCR2.1 (Invitrogen) and sequenced. To check the specificity of the amplification reaction control reactions were run in parallel: GSP2 only (lane 3); NUP only (lane 4); no primers (lane 5); GSP2 and NUP without DNA template (lane 6). DNA products were run on a 1.1% agarose gel in TAE buffer, DNA was stained with ethidium bromide (0.5μg/ml).

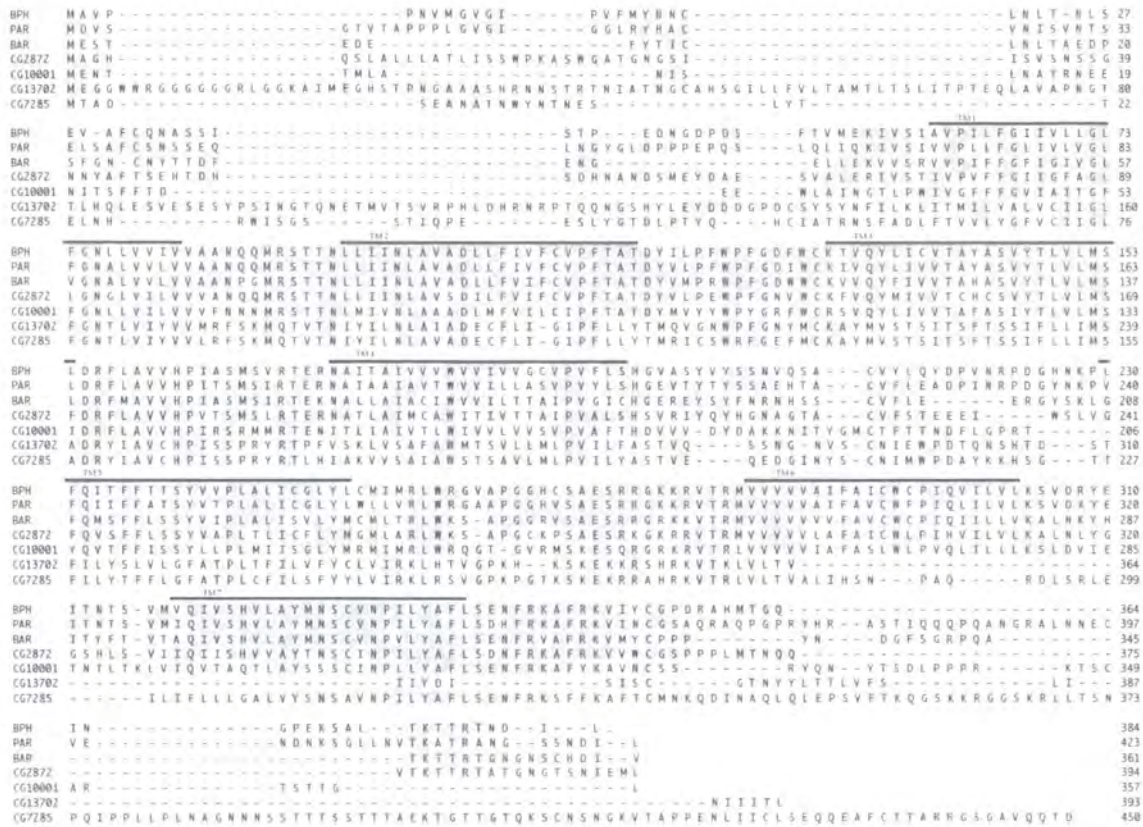
with a predicted M_r of 42.4 kDa. The 5' untranslated region (UTR) was 69 bp (Fig. 5.4a), and the 3' UTR was either 379 bp (Fig. 5.4a, a), 289 bp (Fig. 5.4a, b), 206 bp (Fig. 5.4a, c), or 146 bp (Fig. 5.4a, d). Analysis of the 3' UTR sequence reveals that although each of the sequences are different all share some sequence similarity, and either diverge at different points, or include premature poly-A sequences. Regions of sequence similarity within the 3' UTR region is highlighted in fig. 5.4a.

The *N.lugens* ALSTR has 4 potential glycosylation sites (matching the N-X-S/T consensus), 3 of which occur in the extracellular N-terminal region. Glycosylation sites are situated at asn 22 asn 25, asn 34 and asn 313 (Fig. 5.4a). Based on hydrophobicity analysis (Krogh *et al.*, 2001) the protein is predicted to have an extracellular N-terminus (amino acid 1-59), seven transmembrane helices connected by short peptide loops (33 amino acids or less) and an intracellular C-terminal region (Fig. 5.4b), all of which are characteristic features of G-protein coupled receptors. Phosphorylation sites are also a common feature of GPCR proteins, therefore serine and threonine sites were predicted using NetPhos, version 2.0 software programme (http://www.cbs.dtu.dk/services/NetPhos/) (Blom *et al.*, 1999). Potential serine phosphorylation sites were predicted at ser 167, ser 271; and potential threonine phosphorylation sites at thr 170 and thr 279 (Fig. 5.4a). Analysis of the protein using SignalP, version 2.0 software programme (http://www.cbs.dtu.dk/services/SignalP-2.0/) (Nielsen *et al.*, 1997) did not reveal a conventional signal sequence for RER membrane translocation.

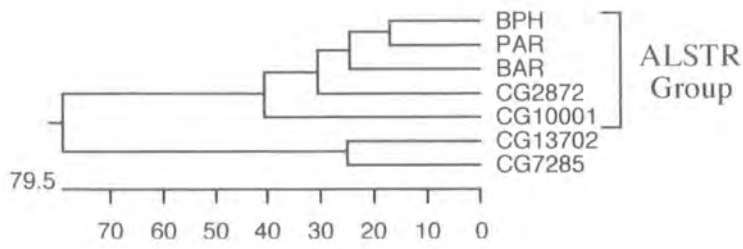
```
                                                                        -69

ATGGCGGTACCCCCGAACGTAATGGGAGTCGGCATCCCCGTCTTCATGTACAACAACTGTCTCAACCTCACCAACCTG +78
<M  A  V  P  P  N  V  M  G  V  G  I  P  V  F  M  Y  N  N  C  L  N  L  T  N  L

TCTGAGGTCGCCTTCTGTCAGAACGCCTCCTCGATCTCCACCCCCGAGGACAACGGAGACCCCGACAGTTTCACGGTG +156
 S  E  V  A  F  C  Q  N  A  S  S  I  S  T  P  E  D  N  G  D  P  D  S  F  T  V

ATGGAGAAGATCGTATCCATAGCAGTTCCCATACTGTTTGGAATCATCGTGCTCCTCGGTCTCTTCGGGAATCTGCTG +234
 M  E  K  I  V  S  I  A  V  P  I  L  F  G  I  I  V  L  L  G  L  F  G  N  L  L
                     TM1

GTTGTTATTGTGGTAGCAGCTAACCAACAGATGCGATCTACAACCAACCTACTGATTATCAATCTGGCAGTGGCTGAT +312
 V  V  I  V  V  A  A  N  Q  Q  M  R  S  T  T  N  L  L  I  I  N  L  A  V  A  D
                                                    TM2

TTACTGTTTATTGTGTTCTGTGTACCGTTCACGGCCACTGATTATATACTGCCGTTCTGGCCTTTTGGTGACTTCTGG +390
 L  L  F  I  V  F  C  V  P  F  T  A  T  D  Y  I  L  P  F  W  P  F  G  D  F  W

TGTAAGACTGTGCAGTACCTGATCTGTGTGACTGCCTACGCCAGTGTCTACACTCTGGTGCTCATGTCCTTGGACAGG +468
 C  K  T  V  Q  Y  L  I  C  V  T  A  Y  A  S  V  Y  T  L  V  L  M  S  L  D  R
          TM3

TTCCTGGCGGTGGTGCACCCKATCGCGTCGATGTCGGTTCGCACGGAACGCAACGCGATCACGGCGATCGTGGTGGTG +546
 F  L  A  V  V  H  P  I  A  S  M  S  V  R  T  E  R  N  A  I  T  A  I  V  V  V
                                 *     ~              TM4

TGGGTGGTGATCGTGGTCGGCTGCGTGCCCGTCTTCCTGTCGCACGGCGTCGCCTCCTATGTCTACTCGAGCAACGTG +624
 W  V  V  I  V  V  G  C  V  P  V  F  L  S  H  G  V  A  S  Y  V  Y  S  S  N  V

CAAAGTGCCTGCGTCTACCTACAGTACGACCCCGTCAACCGGCCCGACGGCCATAATAAGCCGCTCTTCCAGATCACG +702
 Q  S  A  C  V  Y  L  Q  Y  D  P  V  N  R  P  D  G  H  N  K  P  L  F  Q  I  T
                                                                    TM5

TTCTTCACGACATCGTACGTGGTTCCGCTGGCGCTGATCTGCGGCCTGTACCTGTGCATGATCATGCGTCTGTGGCGC +780
 F  F  T  T  S  Y  V  V  P  L  A  L  I  C  G  L  Y  L  C  M  I  M  R  L  W  R

GGCGTGGCGCCCGGCGGCCATTGCTCGGCCGAGAGCCGACGTGGCAAGAAGCGCGTCACCAGGATGGTTGTCGTCGTC +858
 G  V  A  P  G  G  H  C  S  A  E  S  R  R  G  K  K  R  V  T  R  M  V  V  V  V
                              *                 ~                    TM6

GTCGCAATATTCGCCATCTGCTGGTGTCCCATACAGGTGATCCTGGTGCTGAAGAGTGTGGACCGTTACGAGATCACC +936
 V  A  I  F  A  I  C  W  C  P  I  Q  V  I  L  V  L  K  S  V  D  R  Y  E  I  T

AACACATCAGTGATGGTGCAAATCGTGAGTCACGTGTTGGCCTACATGAACTCATGCGTCAACCCCATCCTCTACGCT +1014
 N  T  S  V  M  V  Q  I  V  S  H  V  L  A  Y  M  N  S  C  V  N  P  I  L  Y  A
                          TM7

TTCCTCTCGGAAAACTTCAGGAAAGCTTTCCGCAAGGTGATCTACTGTGGACCGGACCGCGCCCATATGACTGGACAA +1092
 F  L  S  E  N  F  R  K  A  F  R  K  V  I  Y  C  G  P  D  R  A  H  M  T  G  Q

ATTAATGGGCCCGAAAAGTCGGCGCTCACAAAAACCACTCGCACAAACGATATCTTGTGA                  +1152
 I  N  G  P  E  K  S  A  L  T  K  T  T  R  T  N  D  I  L  .>

(a)
ATGGCGGTACCCCCGAACGTAATGGGAGTCGGCATCCCCGTCTTCATGTACAACAACTGTCTCAACCTCACCAACCTG +1230
TCTGAGGTCGCCGGAATCAAGGTCCAAATTACCATGAAAGTTGCAATCAAGATGGCTGACAGAAAGAGATGGAGCTTC +1308
AACACCTTGGACACATTATTTGCTTTCCTTCATTAAAATATTTCAATCGAAAAATTGTCAAGACAACCATAAGGTTAA +1386
TAGTTATTGTTTAGCTGAACATAATTTTGTTTCAGGCAATATCTTAAGGATTGCCGTAATTGTATAATTGAAAAACTA +1464
GAAGTATTGTTGAAATTGTCAAGTAAATCTACTCTGATTTAAAAAAAAAAAAAAAAAAAAAAAAAA             +1531

(b)
GGAATCAAGGTCCAAATTACCATGAAAGTTGCAATCAAGATGGCTGACAGAAAGAGATGGAGCTTCAACACCTTGGAC +1230
ACATTATTTGCTTTCCTTCATTAAAATATTTCAATCGAAAAATTGTCAAGACAACCATAAGGTTAATAGTTATTGTTT +1308
AGCTGAACATAATTTTGTTTCAGGCAATATCTTAAGGATTGCCGTAATTGTATAATTGAAAAACTAGAAGTATTGTTG +1386
AAATTGTCAAGTAAATCTACTCTGATTTAAAAAAAAAAAAAAAAAAAAAAAAAAA                       +1441

(c)
GGAATCAAGATTAAACTCCTTTTGGGCGCACCATTCAATGTGAGATGAGATGAGACGACGCGCCCTATTTAAAAGATA +1230
CGAATACGGATGAACAGAGTGAGCAATCGCCGTTTTATTTTTACTGTATATACATGGGTGAGCTAATAAATAACGGGG +1308
TTGCTTTTTTTTGTTTCATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA                            +1358

(d)
GGAATCAAGGTCCAAATTACCATGAAAGTTGCAATCAAGATGGCTGACAGAAAGAGATGGAGCTTCAACCCCTTGGAC +1230
ACATTTTTTGCTTTCCTTCATTAAAATATTTCAATTGAAAAAAAAAAAAAAAAAAAAAAAAAAAA             +1298
```

The *N.lugens* ALSTR sequence was aligned with other allatostatin receptors from *Bombyx mori* (AF254742), *Periplaneta americana* (AF336364), plus related *Drosophila* gene products belonging to Family-A receptor proteins (CG2872, CG10001, CG13702 and CG7285) (Fig. 5.5a). Alignment with other insect sequences reveals the strong degree of sequence conservation throughout the protein, however the sequence conservation is not represented in the predicted extracellular N-terminal region (Fig. 5.5a). Among this family of receptor proteins the transmembrane regions and first two cytoplasmic loops show greatest sequence conservation, this pattern of sequence conservation can be seen in the aligned receptors (Fig. 5.5a).

Alignment shows that the *N.lugens* receptor is strongly related to the *Periplaneta americana* allatostatin receptor, (86 % sequence identity in the transmembrane regions, and 69 % overall identity), the *Bombyx mori* allatostatin receptor allatostatin (73 % sequence identity in the transmembrane region, and 57 % overall identity), plus the *Drosophila* ALSTR proteins DAR-1 and DAR-2 (with an overall sequence identity of 51% and 45% respectively). The *N.lugens* allatostatin receptor is also related to the mammalian galanin receptors (35 % overall identity with the house mouse, *Mus musculus* galanin receptor type-2, AAC95468 (Pang *et al.*, 1998)), somatostatin receptors (31 % with the human somatostatin receptor, AAA20828 (Panetta *et al.*, 1994)), and opioid receptor (30 % with the pig, *Sus scrofa*, μ opioid receptors Q95247 (Pampusch *et al.*, 1998)). Furthermore, the first and second potential glycosylation sites (filled circles in Fig. 5.5a) of *N.lugens* allatostatin receptor occur at the same positions



(a)



(b)

Figure 5.5 (a & b). **(a)** Amino acid alignment using the CLUSTAL V algorithm of the putative *N.lugens* ALSTR (Fig. 5.4a) with two other insect ALSTR sequences and related *Drosophila* Family-A receptor proteins. Residues that match the consensus are shaded grey. Aligned sequences are the *Drosophila* gene products CG2872, CG10001, CG13702, and CG7285, and ALSTRs from *Bombyx mori* (BAR, AF254742) and *Periplaneta Americana* (PAR, AF336364) **(b)** The phylogenetic relationship of aligned sequences.

in AF336364, AF254742 and CG13702 (data not shown). These results reveal structural similarities between related receptor proteins.

Also, a conserved acidic-Arg-aromatic triplet is present in the N-terminal extremity of the third cytoplasmic loop (Attwood *et al.*, 1991) and could be implicated in the interaction with G-proteins. In the case of the *N.lugens* ALSTR the triplet DRF (amino acid 155 – 157) is well conserved. Phylogenetic analysis (Fig. 5.5b) of the aligned receptor proteins shows a distinct grouping of the insect allatostatin receptors and predicted *Drosophila* allatostatin receptors (CG2872 and CG10001). Other members of the related *Drosophila* family-A peptide receptor proteins have a limited similarity with the ALSTR group of proteins and form a distinct group away from the more closely related ALSTRs (Fig. 5.6b). The ALSTR group (BPH, PAR, BAR, CG2872 and CG10001) all have DRF amino acids as the conserved triplet at the N-terminal extremity of the third cytoplasmic loop, whereas the more distant *Drosophila* gene products (CG13702 and CG7285) have a DRY amino acid triplet. In all cases though, this triplet matches the acidic-Arg-aromatic consensus, which is well conserved throughout Family-A peptide receptor proteins.

5.2.2 Bacterial Expression Constructs

Due to the absence of an overlapping region in RACE products a full-length clone could not be assembled by restriction enzyme digestion and ligation. Therefore, a full-length was assembled by PCR using primers at the extreme 5' and 3' end of the open reading frame. The primer pair was designed to amplify 1152 bp of coding sequence (amino acid 1 – 384). Restriction enzyme sites were incorporated on either end of the primer to

allow directional cloning of the coding sequence into the pIZ insect cell expression vector (Table 2.3). To reduce the chance of introducing PCR errors aliquots of the PCR cloning reaction were removed at cycle 20 and then at every five cycles after. PCR products after 30 cycles (Fig. 5.6, lane 3) were sequenced on both strands of DNA. An error free clone was restricted with *Eco*R V and *Not* I and ligated into the pIZ insect expression vector. In a similar cloning reaction a primer pair for the N-terminal extracellular region was designed to amplify the first 177 bp of the coding sequence (amino acid 1 - 59). Restriction enzyme sites were incorporated at either end to allow the directional cloning of the coding sequence into the pET24(a) expression vector (Table 2.3). The forward primer incorporated a *Nde* I site, while the reverse primer had a *Xho* I site. The N-terminal region was amplified using a minimal number of PCR cycles, previously determined as 30 (Fig. 5.7). An amplification product with the correct sequence was restricted with *Nde* I and *Xho* I and ligated into the pET24(a) bacterial expression vector. The expression construct pET24/ALSTR(1-59) was sequenced over the 3' end to check that the coding sequence was in frame with the poly-His tag and stop codon of the vector.

5.2.3 Production of anti-ALSTR antibodies

The production of antigenic material for anti-ALSTR antibodies used two approaches, namely the expression and purification of the predicted N-terminal region (amino acids 1-59), and antibodies raised against synthetic peptides from extracellular loop regions of the *N.lugens* ALSTR.

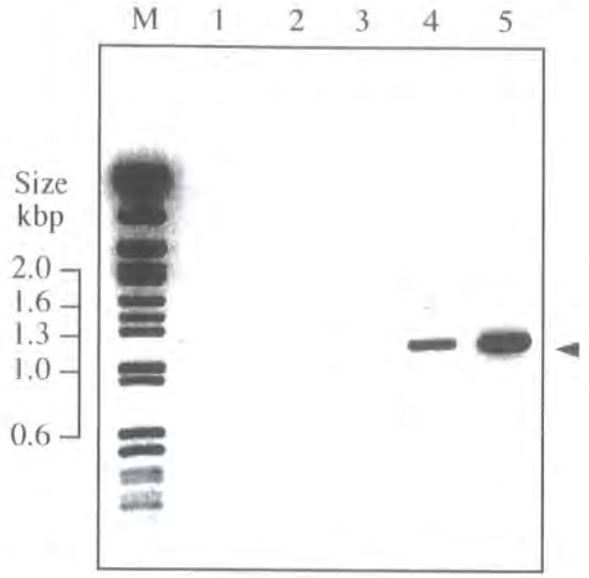


Figure 5.6. The *N.lugens* ALSTR coding sequence (CDS) was amplified from cDNA by PCR using a proof reading polymerase (Clontech). PCR errors were avoided by using a minimal number of cycles. Reaction products were removed after 20 cycles (lane 1), 25 cycles (lane 2), 30 cycles (lane 3), 35 cycles (lane 4) and 40 cycles (lane 5). Products were separated on a 1.1% agarose/TAE gel, and DNAs were stained with ethidium bromide (0.5μg/ml). Reaction products were cloned after 30 amplification cycles (lane 3). ALSTR CDS was sequenced on both strands and a correct reading frame was established. (A) PCR reaction product corresponding to the ALSTR CDS (1152 bp).

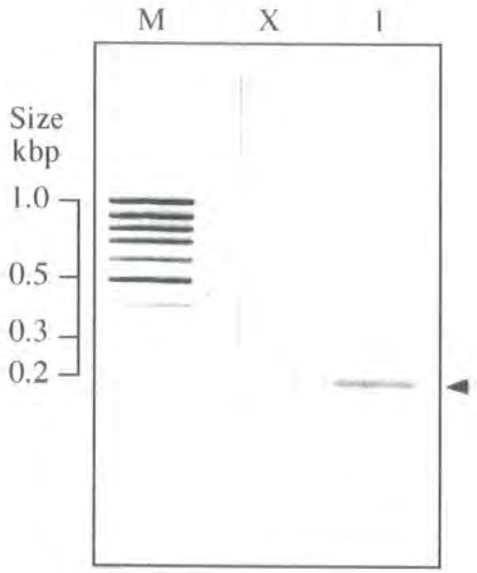


Figure 5.7. The N-terminal region of the *N.lugens* ALSTR was cloned by PCR (lane 1) using a proof reading polymerase (Clontech) and a minimal number of cycles, sequence errors were check by sequencing. The PCR product marked with an arrow was gel purified (Qiagen) and ligated into pET24a bacterial expression vector. X denotes an empty lane.

The pET24/ALSTR(1-59) consrtuct was transformed into a *E.coli* expression host and recombinant protein expression was induced in cells grown to mid-log phase. The accumulation of recombinant protein after induction was monitored by SDS-PAGE and Coomassie blue staining (Fig. 5.8a). After cell lysis and fractionation into soluble and insoluble fractions the accumulation of recombinant protein was not evident on the Coomassie stained gel. Therefore, a similar gel was run and transferred onto nitrocellulose membrane. Transferred proteins were probed with INDIA His probe-HRP (Pierce) in a western blot experiment (Fig. 5.8b). Western blot showed that the recombinant protein accumulated in the insoluble fraction. However, the immunoreactive material was significantly smaller than the predicted molecular weight of 6.3 kDa for the N-terminal 59 amino acids. Analysis of the Coomassie blue stained SDS gel shows accumulation of a product after induction of expression in the insoluble fraction (Fig. 5.8a, lane 4). However, as this product appears as a diffuse band on the gel, and it was significantly smaller than the predicted Mr it was concluded that the expression of the N-terminal ALSTR resulted in degradation of the recombinant peptide. Due to our inability to express recombinant N-terminal ALSTR, antibodies were raised against synthetic peptides corresponding to predicted extracellular regions of the *N.lugens* ALSTR. Regions of the protein that were chosen for antibody production are shown in Table 5.1.

5.2.4 Characterisation of anti-ALSTR Antibodies

Antibodies raised against synthetic peptides were first purified by affinity to immobilised protein G. The purified antibodies were concentrated to 10 mg/ml. Antibodies were used in western blot experiments on protein extracts from *N.lugens.*

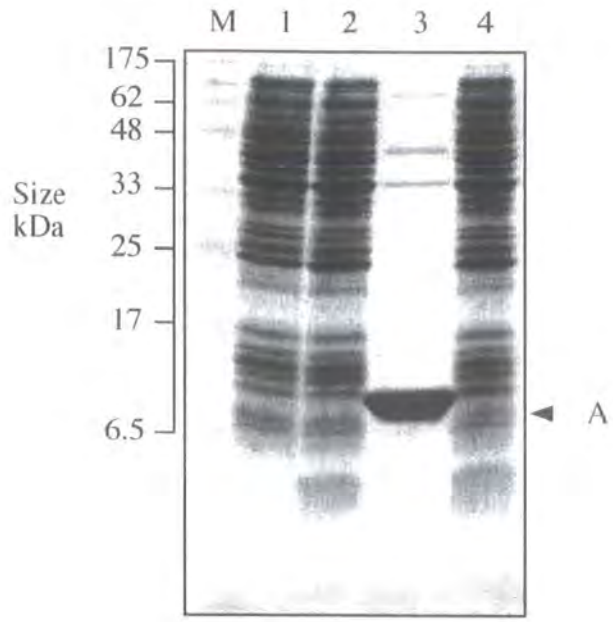


Figure 5.8a. SDS-PAGE analysis of recombinant *N.lugens* ALSTR N-terminal domain expressed in RosettaBlue (DE3). Protein samples were prepared for SDS-PAGE analysis and resolved on a 15 % acrylamide gel, proteins were visualised by staining and destaining with Coomassie blue. Uninduced cells were grown to an OD λ 600 nm (lane 1) and induced with IPTG (0.1mM). Following induction cells were grown for a further 3 hours (lane 2). Induced cells were treated with lysozyme (A) and split into soluble (lane 3) and insoluble (lane 4) fractions.

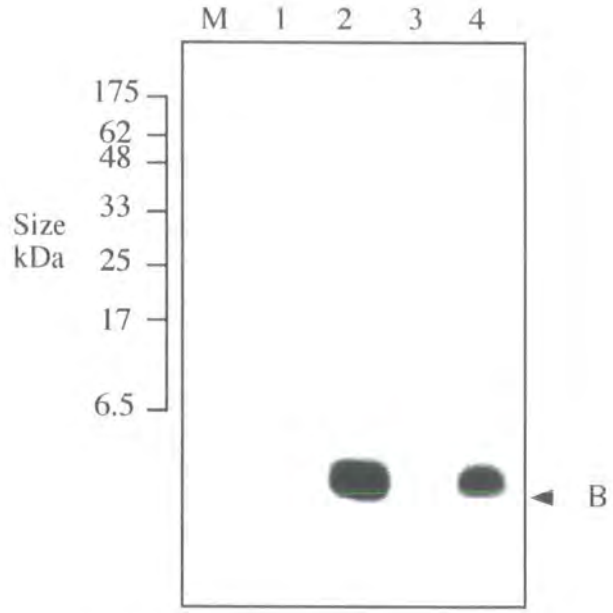


Figure 5.8b. A similar gel was transferred to a nitrocellulose membrane that was used in western blot to test for expression of *N.lugens* ALSTR N-terminal domain. Blots were probed with INDIA His probe-HRP (Pierce) and binding of the primary antibody was visualised directly using ECL detection.

Region	From	To	Amino Acid Sequence
Outside	1	59	MAVPPNVMGVGIPVFMYNNCLNLTNLSEVAFCQNASS**ISTPEDNGDPDSFT**VMEKIVSI
TM	60	82	AVPILFGIIVLLGLFGNLLVVIV
Inside	83	94	VAANQQMRSTTN
TM	95	117	LLIINLAVADLLFIVFCVPFTAT
Outside	118	131	DYILPFWPFGDFWC
TM	132	154	KTVQYLICVTAYASVYTLVLMSL
Inside	155	173	DRFLAVVHPIASMSVRTER
TM	174	196	NAITAIVVVWVVIVVGCVPVFLS
Outside	197	229	HGVASYVYSSNVQSACVYLQ**YDPVNRPDGHNKP**
TM	230	252	LFQITFFTTSYVVPLALICGLYL
Inside	253	282	CMIMRLWRGVAPGGHCSAESRRGKKRVTRM
TM	283	303	VVVVVAIFAICWCPIQVILVL
Outside	304	317	**KSVDRYEITNTSVM**
TM	318	340	**VQIVSHVLA**YMNSCVNPILYAFL
Inside	341	383	SENFRKAFRKVIYCGPDRAHMTGQINGPEKSALTKTTRTNDIL

Table 5.1. Regions of the *N.lugens* ALSTR, as predicted by TMHMM analysis (Fig. 5.4b). Regions shown in bold type correspond to synthetic peptides that were synthesised for the production of polyclonal antibodies. Regions predicted as 'outside' are in contact with insect haemolymph and the regions predicted as 'inside' are cytoplasmic.

However, the non-specific binding of antibodies was unacceptable, and furthermore did not bind to proteins to the predicted size of the *N.lugens* allatostatin receptor. Therefore, further characterisation of these antibodies was not possible.

5.3 Discussion

5.3.1 Cloning N.lugens ALSTR

The degenerate primers that were designed against highly conserved residues in TM 3 and TM 7 of aligned allatostatin receptor proteins were successful in amplifying a *Drosophila* positive control (DAR-1, CG2872), and a similar protein from *N.lugens* gut specific cDNA. It is therefore likely that this primer set may be used for amplification of ALSTRs from other insect species. Long stretches of conserved amino acids allowed nested primers to be designed, which are useful to increase the specificity of the PCR. However, due to the specificity of the initial PCR a nested amplification approach was not required. After successful amplification of a partial sequence the full-length CDS was retrieved by 5' and 3' RACE PCR. The full-length receptor coding sequence plus untranslated regions was aligned on a computer to produce a contiguous sequence.

5.3.2 N.lugens *ALSTR Sequence Analysis*

Sequence analysis demonstrates that the cloned *N.lugens* receptor is a member of the somatostatin/galanin/opioid receptor subfamily of G-protein coupled receptors (GPCRs), or family-A peptide receptors. The first insect receptor to be cloned with structural similarities to proteins from this family was the allatostatin receptor from *Drosophila* (DAR-1, CG2872) (Birgul *et al.*, 1999). Since this initial description allatostatin receptors *Bombyx mori* (Secher *et al.*, 2001), *Periplaneta americana*

(Auerswald *et al.*, 2001) have been characterised, and a second ALSTR from *Drosophila* has been cloned (DAR-2) (Lenz *et al.*, 2000), and functionally characterised (Larsen *et al.*, 2001; Lenz *et al.*, 2001). Proteins belonging to this receptor family (family A) have an extracellular N-terminal region, which is often glycosylated, 7 transmembrane regions and an intracellular C-terminal region, which is often phosphorylated. All of these features are present in the putative ALSTR receptor amplified from *N.lugens* gut specific cDNA.

As this protein is a member of the GPCR superfamily it possesses the well-conserved cysteine residues on TM 2 and TM 3, which are conserved across all classes of GPCR (Probst *et al.*, 1992). However, these are the only residues that are well conserved across different classes of GPCRs. Within the Family-A subgroup the only conserved motif is the DRY/F triplet, which is located at the cytoplasmic side after TM 3. Strict conservation of this triplet among members of this family suggests that is may be important for functional integrity of the receptor protein. It has been proposed that this motif may be important for the interaction with intracellular G proteins.

The *Drosophila* genome contains four receptor proteins with similarity to vertebrate somatostatin/galanin/opioid receptor subfamily. These include gene products CG7285, CG13702, CG10001 and CG2872. Two receptors from this group have been functionally characterised and classified as ALSTRs (DAR-1, CG2872 and DAR-2, CG10001). *Drosophila* is the only insect to date, with two well-characterised ALSTRs. However, the general presence of two or more allatostatin receptors in an insect would

agree well with previous pharamacological experiments in cockroaches, where some A-type allatostatins are potent inhibitors of juvenile hormone (JH) biosynthesis in the *corpora allata*, whereas they were relatively ineffective as inhibitors of hindgut muscle contraction, while other A-type allatostatins have the reverse properties (Pratt *et al.*, 1997). The presence of two allatostatin receptors in cockroaches is also suggested by direct binding studies (Bowser and Tobe, 2000). With these data in mind it is likely that *N.lugens* has more that one ALSTR; and as in *Drosophila* Type-A, Type-B and Type-C allatostatins. Although, only through functional expression of the receptor can the bioactive ligand be identified for the *N.lugens* ALSTR. Physiological experiments would have to be used to determine which of the allatostatins (type-A, B or C) exert a true allatostatic effect (inhibition of JH synthesis) in *N.lugens*. Nevertheless, identification of a putative ALSTR in *N.lugens* is the first stage for the characterisation of a potential 'target' and design of specific insecticidal proteins.

The putative *N.lugens* ALSTR has greatest sequence similarity to the *P.americana* ALSTR with an overall sequence identity of 69%. Greatest sequence conservation occurs between and within the transmembrane regions. However, there is little sequence conservation within the predicted N-terminal region. Although the N-terminal region is significantly smaller than that of family-B receptors it is thought to be important in binding of peptide hormones. Glycosylation of the N-terminal domain may also be important for ligand binding, the fact that glycosylation sites have been conserved in related receptors suggest the significance of these sites. Therefore, based on sequence

identity, the cDNA cloned from *N.lugens* gut cDNA has been classified as a putative allatostatin receptor, although its hormone ligand remains unidentified at present.

5.3.3 Alternative ALSTR 3' Ends

The coding region for each of the independent clones that was sequenced is identical, however variation is seen within the 3' UTRs (Fig. 5.4a; a, b, c, d). It is possible that sequence diversity among the different UTRs may be generated by several different mechanisms. 3' UTRs from (a) and (b) are identical apart from an additional 90 bp sequence at the start of the UTR region (a). Staggeringly this additional 90 bp region is a 100% nucleotide match with the initial 90 bp from the coding region. However, the biological significance of this is unknown, as is the mechanism which generated this diversity. Although it is speculative to suggest it may be a result of alternate RNA splicing. The 3' UTR sequence represented by (c) matches that of the (b) UTR for the first 8 bp but then the sequence diverges, although speculative this diversity may be a result of alternative splicing within this region. The sequence represented by (d) matches that of (b) for the first 98 bp, however the poly-A site is added prematurely, this may be due to a different location of the poly-A addition sequence in the unprocessed RNA.

In summary, the diversity observed in the 3' UTR may be due to either alternative RNA splicing, different poly-A addition sites or PCR artefacts. Further work is needed to elucidate the mechanism responsible for generating this diversity and the significance, if any, of these alternative 3' UTRs.

5.3.4 Recombinant Protein Expression and Anti-ALSTR Antibodies

The longest extracellular region was chosen for protein expression, which was the extracellular N-terminal region (amino acids 1 - 59). Upon induction of expression hosts transformed with the ALSTR-N-terminal region expression construct the recombinant protein accumulated in the insoluble fraction. However, expression of the recombinant protein did not produce a crisp band and it looked as if protein degradation had taken place. Due to the small N-terminal region (amino acids 1-59) it made protein expression and purification of the 6.3 kDa peptide difficult. Therefore, this region should have been expressed as a fusion protein, linked to a carrier protein like BSA. Expression of such a fusion would make expression and purification simpler, as well as reducing the chances of protein degradation. Due to the inadequate quality of the expressed peptide synthetic peptides from regions of the *N.lugens* ALSTR were synthesised, linked to a BSA carrier and used to raise antibodies in rabbits. However, due to the lack of specificity of the antibodies generated, and failure of binding to the *N.lugens* ALSTR target further characterisation was not possible.

CHAPTER 6.

GENERAL DISCUSSION

Due to the significant reduction in rice yield caused by the insect pest *Nilaparvata lugens* it is of great importance to generate effective control methods to limit the damage caused (section 1.4). Control methods must be effective, specific and have little environmental impact. Plant genetic modification (GM) provides an appropriate solution to address these needs. Using GM approaches it is possible to engineer crop plants to have increased endogenous resistance to insect pest species, allowing effective and specific control with limited environmental impact. The success of crops expressing *Bacillus thuringiensis* toxin (*Bt* toxin) stands testament to the potential benefits offered by this new technology. However, as no *Bt* toxin has been described for *N.lugens* novel approaches must be considered for control of this pest, and other Homopteran pest species.

Traditionally, the search for novel insecticidal proteins has involved purifying extracts from plants and toxins from bacteria, and screening against the chosen pest species in artificial diet feeding trials. Using these methods the lectin from *Galanthus nivalis* (GNA) has been discovered as an insecticidal protein, as has the *Bt* toxin. An advantage of this approach is that insecticidal proteins can be identified with no prior knowledge of the pest biology. However, such an approach is not without significant drawbacks. For example, large numbers of candidate compounds need to be screened in order to identify

insecticides. Therefore, such an approach is both time-consuming and expensive. Also, compounds identified as insecticidal towards a pest species are sometimes toxic to a broad range on non-target insect species, and therefore have an unacceptable environmental impact.

This thesis aimed to design novel insecticidal proteins with prior knowledge of the 'target' protein. It was anticipated that a modulation in function of such a target would be detrimental to insect survival. This experimental design can be viewed as the reciprocal of traditional screening methods. However, for such an approach to be effective knowledge of insect biology is necessary for the identification of potential targets, and to this extent is reliant on and limited by the current knowledge available. Further developments in insect cell signalling and homeostasis will certainly yield many more potential targets for insecticidal protein design. Therefore, using knowledge of host biology 'targets' have been identified and then potential insecticidal proteins have been designed to interact and cause a modification in function. This approach is possible by utilising the sequences deposited in GenBank databases (which is growing at an exponential rate), as well as an increasing understanding of regulation in insects. Also fuelling this type of research is the genome sequences from *Drosophila* and *Anopheles*, plus the huge number of ESTs for the phloem-feeding pea aphid, *Acyrthosiphon pisum.* With candidate targets identified specific protein agonists and antagonists can be designed. However, due to the lack of sequence data for *N.lugens* the first stage in such a stratergy is the identification of potential targets. This project has been successful in identifying potential targets proteins. Due to the central role played in the coordination

of crucial insect processes both the insect diuretic hormone receptor (DHR) and allatostatin receptor (ALSTR), members of the G-protein coupled receptor (GPCR) family, were selected as suitable targets. Also, the low abundance of such proteins further strengthens their role as a target. Due to their low abundance and central role in homeostasis it can be anticipated that a modulation in function may be achieved by very small amounts of antagonist or agonist, which is likely to have a dramatic effect on host biology.

This project utilised two approaches for identification of potential targets. These were a mass sequencing approach (Section 3.1) and amplification of selected targets using degenerate primers (Section 3.1). The random screening approach was not effective in GPCR target identification, even though large numbers of independent clones were sequenced. However, it can be anticipated that if a greater number of clones were sequenced potential target GPCR sequences should occur. Therefore, the major drawback with this approach is the significant cost and time spent in generating a large EST database. Also, GPCR proteins identified using this approach would still need biological characterisation to establish if they would provide effective targets. The diuretic hormone receptor (DHR) and allatostatin receptor (ALSTR) were identified as potential *N.lugens* targets. Therefore, these sequences were amplified using degenerate primers and completed by RACE PCR. Out of the 7680 EST sequences neither of these receptor proteins was represented. This highlights the failings of the mass sequencing approach and low abundance of these receptor proteins.

With targets identified protein antagonists and agonists can be designed. This thesis also describes the production of antibodies that recognise the N-terminal domain of the DHR. When incorporated into artificial diet the antibodies are ingested by the insect and excreted in the honeydew. As expected very little proteolysis of the antibody occurs, due to the low abundance of proteases in the *N.lugens* gut. Therefore, it can be anticipated that oral ingestion of protein antagonists and agonists by Homopteran insect pests will serve as an effective delivery mechanism as little protein degradation takes place. Although ingestion of antibodies took place they were ineffective in crossing the gut wall and entering the haemolymph, and therefore could not interact with the target DHR. For such an approach to be effective a delivery system needs to be developed.

Although ingestion of anti-DHR antibodies did not exert a toxic effect this is probably due to the lack of an interaction between the antibody and the receptor protein. Therefore, the DHR still remains a valid target for insecticidal proteins. However, if this work was continued I would utilise another insect model species to validate targets, microinjection of the protein antagonists should experimentally resolve the problem of no interaction with receptor proteins. An insect species amenable to this kind of manipulation is *Manduca sexta*, this insect also has a functionally characterised diuretic hormone and receptor protein. This approach would be effective in validating potential insecticidal targets. Once validated as insecticidal further research would be required to elucidate an effective delivery system. One such delivery system may be the creation of fusion proteins. Upon ingestion the lectin GNA is known to cross the gut wall, small

peptides fused to this 'carrier' are also transported into the haemolymph. Therefore, creation of agonist/antagonist-GNA fusions may serve as an effective delivery system.

In conclusion, the initial stages required for the design of insecticidal proteins has been achieved with the isolation of potential targets from *N.lugens* gut specific cDNA. Both of these targets are members of the GPCR superfamily and have been classified as a putative DHR and ALSTR. Because the roles played by these receptors is crucial for insect water homeostasis (DHR) and control of moulting (ALSTR) it is likely that they will potential targets across all orders of insect pest.

REFERENCES.

Abo, M. E.,Sy, A. A., 1998. Rice virus diseases: Epidemiology and management strategies. Journal of Sustainable Agriculture 11, 113-134.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H. C., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Miklos, G. L. G., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, A., de Pablos, B., Delcher, A., Deng, Z. M., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F. C., Gorrell, J. H., Gu, Z. P., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, D. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z. X., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z. W., Lasko, P., Lei, Y. D., Levitsky, A. A., Li, J. Y., Li, Z. Y., Liang, Y., Lin, X. Y., Liu, X. J., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D. C., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E.,

Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G. G., Zhao, Q., Zheng, L. S., Zheng, X. Q. H., Zhong, F. N., Zhong, W. Y., Zhou, X. J., Zhu, S. P., Zhu, X. H., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M.,Venter, J. C., 2000. The genome sequence of Drosophila melanogaster. Science 287, 2185-2195.

Ahn, Y. J., Kwon, M., Park, H. M.,Han, C. K. (1997). Potent insecticidal activity of Ginkgo biloba derived trilactone terpenes against Nilaparvata lugens. Phytochemicals for Pest Control. **658:** 90-105.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W.,Lipman, D. J., 1990. Basic Local Alignment Search Tool. Journal of Molecular Biology 215, 403-410.

Altstein, M., 2001. Insect neuropeptide antagonists. Biopolymers 60, 460-473.

Ashford, D. A., Smith, W. A.,Douglas, A. E., 2000. Living on a high sugar diet: the fate of sucrose ingested by a phloem-feeding insect, the pea aphid Acyrthosiphon pisum. Journal of Insect Physiology 46, 335-341.

Attwood, T. K., Eliopoulos, E. E.,Findlay, J. B. C., 1991. Multiple Sequence Alignment of Protein Families Showing Low Sequence Homology - a Methodological Approach Using Database Pattern-Matching Discriminators for G-Protein-Linked Receptors. Gene 98, 153-159.

Audsley, N., Kay, I., Hayes, T. K.,Coast, G. M., 1995. Cross-Reactivity Studies of Crf-Related Peptides on Insect Malpighian Tubules. Comparative Biochemistry and Physiology a-Physiology 110, 87-93.

Auerswald, L., Birgul, N., Gade, G., Kreienkamp, H. J.,Richter, D., 2001. Structural, functional, and evolutionary characterization of novel members of the allatostatin receptor family from insects. Biochemical and Biophysical Research Communications 282, 904-909.

Bargmann, C. I., 1998. Neurobiology of the Caenorhabditis elegans genome. Science 282, 2028-2033.

Bargmann, C. I.,Kaplan, J. M., 1998. Signal transduction in the Caenorhabditis elegans nervous system. Annual Review of Neuroscience 21, 279-308.

Baum, T. J., Hiatt, A., Parrott, W. A., Pratt, L. H.,Hussey, R. S., 1996. Expression in tobacco of a functional monoclonal antibody specific to stylet secretions of the root-knot nematode. Molecular Plant-Microbe Interactions 9, 382-387.

Bendena, W. G., Garside, C. S., Yu, C. G.,Tobe, S. S. (1997). Allatostatins: Diversity in structure and function of an insect neuropeptide family. Neuropeptides in Development and Aging. **814:** 53-66.

Bennett, R. N.,Wallsgrove, R. M., 1994. Secondary Metabolites in Plant Defense-Mechanisms. New Phytologist 127, 617-633.

BenYakir, D.,Shochat, C., 1996. The fate of immunoglobulin G fed to larvae of Ostrinia nubilalis. Entomologia Experimentalis Et Applicata 81, 1-5.

Beyenbach, K. W., 1995. Mechanism and Regulation of Electrolyte Transport in Malpighian Tubules. Journal of Insect Physiology 41, 197-207.

Birgul, N., Weise, C., Kreienkamp, H. J.,Richter, D., 1999. Reverse physiology in Drosophila: identification of a novel allatostatin-like neuropeptide and its cognate

receptor structurally related to the mammalian somatostatin/galanin/opioid receptor family. Embo Journal 18, 5892-5900.

Blackburn, M. B., Kingan, T. G., Bodnar, W., Shabanowitz, J., Hunt, D. F., Kempe, T., Wagner, R. M., Raina, A. K., Schnee, M. E.,Ma, M. C., 1991. Isolation and Identification of a New Diuretic Peptide from the Tobacco Hornworm, Manduca-Sexta. Biochemical and Biophysical Research Communications 181, 927-932.

Blom, N., Gammeltoft, S.,Brunak, S., 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology 294, 1351-1362.

Bourne, H. R., Sanders, D. A.,McCormick, F., 1991. The Gtpase Superfamily - Conserved Structure and Molecular Mechanism. Nature 349, 117-127.

Bowser, P. R. F.,Tobe, S. S., 2000. Partial characterization of a putative allatostatin receptor in the midgut of the cockroach Diploptera punctata. General and Comparative Endocrinology 119, 1-10.

Bradford, M. M., 1976. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. Analytical Biochemistry 72, 248-254.

Brody, T.,Cravchik, A., 2000. Drosophila melanogaster G protein-coupled receptors. Journal of Cell Biology 150, F83-F88.

Broeck, J. V., 2001. Insect G protein-coupled receptors and signal transduction. Archives of Insect Biochemistry and Physiology 48, 1-12.

Brown, M. R., Crim, J. W., Arata, R. C., Cai, H. N., Chun, C.,Shen, P., 1999. Identification of a Drosophila brain-gut peptide related to the neuropeptide Y family. Peptides 20, 1035-1042.

Brownlee, D. J. A.,Fairweather, I., 1999. Exploring the neurotransmitter labyrinth in nematodes. Trends in Neurosciences 22, 16-24.

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M.,Hayashizaki, Y., 2000. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Research 10, 1617-1630.

Chen, W. L.,Sun, C. N., 1994. Purification and Characterization of Carboxylesterases of a Rice Brown Planthopper, Nilaparvata-Lugens Stal. Insect Biochemistry and Molecular Biology 24, 347-355.

Choi, M. Y., Fuerst, E. J., Rafaeli, A.,Jurenka, R., 2003. Identification of a G protein-coupled receptor for pheromone biosynthesis activating neuropeptide from pheromone glands of the moth Helicoverpa zea. Proceedings of the National Academy of Sciences of the United States of America 100, 9721-9726.

Clottens, F. L., 1994. Isolation and characterisation of a diuretic peptide common to the house fly and stable fly. Peptides 15, 971-979.

Clyne, P. J., Warr, C. G.,Carlson, J. R., 2000. Candidate taste receptors in Drosophila. Science 287, 1830-1834.

Coast, G. M., 1996. Neuropeptides implicated in the control of diuresis in insects. Peptides 17, 327-336.

Coast, G. M., 1998. Insect diuretic peptides: Structures, evolution and actions. American Zoologist 38, 442-449.

Coast, G. M., Holman, G. M.,Nachman, R. J., 1990. The Diuretic Activity of a Series of Cephalomyotropic Neuropeptides, the Achetakinins, on Isolated Malpighian Tubules of the House Cricket, Acheta-Domesticus. Journal of Insect Physiology 36, 481-488.

Coast, G. M.,Kay, I., 1994. The Effects of Acheta Diuretic Peptide on Isolated Malpighian Tubules from the House Cricket Acheta-Domesticus. Journal of Experimental Biology 187, 225-243.

Cottrell, G. A., 1997. The first peptide-gated ion channel. Journal of Experimental Biology 200, 2377-2386.

Cox, K. J. A., Tensen, C. P., VanderSchors, R. C., Li, K. W., vanHeerikhuizen, H., Vreugdenhil, E., Geraerts, W. P. M.,Burke, J. F., 1997. Cloning, characterization, and expression of a G-protein- coupled receptor from Lymnaea stagnalis and identification of a leucokinin-like peptide, PSFHSWSamide, as its endogenous ligand. Journal of Neuroscience 17, 1197-1205.

Dai, S. M.,Sun, C. N., 1984. Pyrethroid Resistance and Synergism in Nilaparvata-Lugens Stal (Homoptera, Delphacidae) in Taiwan. Journal of Economic Entomology 77, 891-897.

Dautzenberg, F. M., Wille, S., Lohmann, R.,Spiess, J., 1998. Mapping of the ligand-selective domain of the Xenopus laevis corticotropin-releasing factor receptor 1: Implications for the ligand-binding site. Proceedings of the National Academy of Sciences of the United States of America 95, 4941-4946.

De Jaeger, G., De Wilde, C., Eeckhout, D., Fiers, E.,Depicker, A., 2000. The plantibody approach: expression of antibody genes in plants to modulate plant metabolism or to obtain pathogen resistance. Plant Molecular Biology 43, 419-428.

Dixon, R. A. F., Kobilka, B. K., Strader, D. J., Benovic, J. L., Dohlman, H. G., Frielle, T., Bolanowski, M. A., Bennett, C. D., Rands, E., Diehl, R. E., Mumford, R. A., Slater, E. E., Sigal, I. S., Caron, M. G., Lefkowitz, R. J.,Strader, C. D., 1986. Cloning of the Gene and Cdna for Mammalian Beta-Adrenergic- Receptor and Homology with Rhodopsin. Nature 321, 75-79.

Dong, M. Q., Zang, M. W., Pinon, D. I., Li, Z. J., Lybrand, T. P.,Miller, L. J., 2002. Interaction among four residues distributed through the secretin pharmacophore and a focused region of the secretin receptor amino terminus. Molecular Endocrinology 16, 2490-2501.

Du, J. P., Foissac, X., Carss, A., Gatehouse, A. M. R.,Gatehouse, J. A., 2000. Ferritin acts as the most abundant binding protein for snowdrop lectin in the midgut of rice brown planthoppers (Nilaparvata lugens). Insect Biochemistry and Molecular Biology 30, 297-305.

Egerod, K., Reynisson, E., Hauser, F., Williamson, M., Cazzamali, G.,Grimmelikhuijzen, C. J. P., 2003. Molecular identification of the first insect proctolin receptor. Biochemical and Biophysical Research Communications 306, 437-442.

Elekes, K., Hernadi, L., Muren, J. E.,Nassel, D. R., 1994. Peptidergic Neurons in the Snail Helix-Pomatia - Distribution of Neurons in the Central and Peripheral Nervous Systems That React with an Antibody Raised to the Insect Neuropeptide, Leucokinin-I. Journal of Comparative Neurology 341, 257-272.

Fecker, L. F., Koenig, R.,Obermeier, C., 1997. Nicotiana benthamiana plants expressing beet necrotic yellow vein virus (BNYVV) coat protein-specific scFv are partially protected against the establishment of the virus in the early stages of infection and its pathogenic effects in the late stages of infection. Archives of Virology 142, 1857-1863.

Feng, G. P., Reale, V., Chatwin, H., Kennedy, K., Venard, R., Ericsson, C., Yu, K., Evans, P. D.,Hall, L. M., 2003. Functional characterization of a neuropeptide F-like receptor from Drosophila melanogaster. European Journal of Neuroscience 18, 227-238.

Fitches, E., Audsley, N., Gatehouse, J. A.,Edwards, J. P., 2002. Fusion proteins containing neuropeptides as novel insect contol agents: snowdrop lectin delivers fused allatostatin to insect haemolymph following oral ingestion. Insect Biochemistry and Molecular Biology 32, 1653-1661.

Foissac, X., Edwards, M. G., Du, J. P., Gatehouse, A. M. R.,Gatehouse, J. A., 2002. Putative protein digestion in a sap-sucking homopteran plant pest (rice brown plant hopper; Nilaparvata lugens: Delphacidae) - identification of trypsin-like and cathepsin B-like proteases. Insect Biochemistry and Molecular Biology 32, 967-978.

Foissac, X., Loc, N. T., Christou, P., Gatehouse, A. M. R.,Gatehouse, J. A., 2000. Resistance to green leafhopper (Nephotettix virescens) and brown planthopper (Nilaparvata lugens) in transgenic rice expressing snowdrop lectin (Galanthus nivalis agglutinin; GNA). Journal of Insect Physiology 46, 573-583.

Fourney, R. M., Miyakoshi, J., Day, R. S.,Paterson, M. C., 1988. Northern blotting: efficient RNA staining and transfer. Focus 10, 5-7.

Furuya, K., Milchak, R. J., Schegg, K. M., Zhang, J. R., Tobe, S. S., Coast, G. M.,Schooley, D. A., 2000. Cockroach diuretic hormones: Characterization of a calcitonin- like peptide in insects. Proceedings of the National Academy of Sciences of the United States of America 97, 6469-6474.

Fuse, M., Zhang, J. R., Partridge, E., Nachman, R. J., Orchard, I., Bendena, W. G.,Tobe, S. S., 1999. Effects of an allatostatin and a myosuppressin on midgut carbohydrate enzyme activity in the cockroach Diploptera punctata. Peptides 20, 1285-1293.

Garczynski, S. F., Brown, M. R., Shen, P., Murray, T. F.,Crim, J. W., 2002. Characterization of a functional neuropeptide F receptor from Drosophila melanogaster. Peptides 23, 773-780.

Gatehouse, A. M. R., Down, R. E., Gatehouse, J. A.,Christou, P. (2000). Improved strategies for aphid resistant transgenic crops. Bcpc Conference: Pests & Diseases 2000, Vols 1-3, Proceedings. **1-3:** 593-598.

Gatehouse, J. A., 2002. Plant resistance towards insect herbivores: a dynamic interaction. New Phytologist 156, 145-169.

Gerald, C., Walker, M. W., Criscione, L., Gustafson, E. L., BatzlHartmann, C., Smith, K. E., Vaysse, P., Durkin, M. M., Laz, T. M., Linemeyer, D. L., Schaffhauser, A. O., Whitebread, S., Hofbauer, K. G., Taber, R. I., Branchek, T. A.,Weinshank, R. L., 1996. A receptor subtype involved in neuropeptide-Y-induced food intake. Nature 382, 168-171.

Gether, U., 2000. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. Endocrine Reviews 21, 90-113.

Gullan, P. J.,Cranston, P. S. (1998). The Insects: An Outline of Entomology. London, Chapman & Hall: 400-432.

Ha, S. D., Kataoka, H., Suzuki, A., Kim, B. J., Kim, H. J., Hwang, S. H.,Kong, J. Y., 2000. Cloning and sequence analysis of cDNA for diuretic hormone receptor from the Bombyx mori. Molecules and Cells 10, 13-17.

Halton, D. W., Shaw, C., Maule, A. G.,Smart, D. (1994). Regulatory Peptides in Helminth-Parasites. Advances in Parasitology, Vol 34. **34:** 163-227.

Hamm, H. E., 1998. The many faces of G protein signaling. Journal of Biological Chemistry 273, 669-672.

Hanahan, D. (1985). DNA Cloning. London, UK, IRL Press Ltd.

Hasui, H.,Ozaki, K., 1984. Electrophoretic Esterase Patterns in the Brown Planthopper, Nilaparvata-Lugens Stal (Hemiptera, Delphacidae) Which Developed Resistance to Insecticides. Applied Entomology and Zoology 19, 52-58.

Hazelwood, R. L., 1993. The Pancreatic-Polypeptide (Pp-Fold) Family - Gastrointestinal, Vascular, and Feeding Behavioral-Implications. Proceedings of the Society for Experimental Biology and Medicine 202, 44-63.

Hegedus, D. D., Pfeifer, T. A., Hendry, J., Theilmann, D. A.,Grigliatti, T. A., 1998. A series of broad host range shuttle vectors for constitutive and inducible expression of heterologous proteins in insect cell lines. Gene 207, 241-249.

Hetru, C., Li, K. W., Bulet, P., Lagueux, M.,Hoffmann, J. A., 1991. Isolation and Structural Characterization of an Insulin-Related Molecule, a Predominant Neuropeptide from Locusta-Migratoria. European Journal of Biochemistry 201, 495-499.

Hewes, R. S.,Taghert, P. H., 2001. Neuropeptides and neuropeptide receptors in the Drosophila melanogaster genome. Genome Research 11, 1126-1142.

Hibino, H., 1996. Biology and epidemiology of rice viruses. Annual Review of Phytopathology 34, 249-274.

Higgins, D. G.,Sharp, P. M., 1989. Fast and Sensitive Multiple Sequence Alignments on a Microcomputer. Computer Applications in the Biosciences 5, 151-153.

Holman, G. M., Nachman, R. J.,Wright, M. (1990). A stratergy for the isolation and structural characterisation of certain insect myotropic peptides that modify the spontaneous contractions of the isolated cockroach hindgut. Chromotography and isolation of insect hormones and pheromones. A. R. McCaffery and I. D. Wilson, New York: Plenum: 195-204.

Holmes, S. P., He, H., Chen, A. C., Ivie, G. W.,Pietrantonio, P. V., 2000. Cloning and transcriptional expression of a leucokinin-like peptide receptor from the Southern cattle tick, Boophilus microplus (Acari : Ixodidae). Insect Molecular Biology 9, 457-465.

Holt, J., Chancellor, T. C. B., Reynolds, D. R.,Tiongco, E. R., 1996. Risk assessment for rice planthopper and tungro disease outbreaks. Crop Protection 15, 359-368.

Holtmann, M. H., Ganguli, S., Hadac, E. M., Dolu, V.,Miller, L. J., 1996. Multiple extracellular loop domains contribute critical determinants for agonist binding and activation of the secretin receptor. Journal of Biological Chemistry 271, 14944-14949.

Hoon, M. A., Adler, E., Lindemeier, J., Battey, J. F., Ryba, N. J. P.,Zuker, C. S., 1999. Putative mammalian taste receptors: A class of taste-specific GPCRs with distinct topographic selectivity. Cell 96, 541-551.

Howard, A. D., McAllister, G., Feighner, S. D., Liu, Q. Y., Nargund, R. P., Van der Ploeg, L. H. T.,Patchett, A. A., 2001. Orphan G-protein-coupled receptors and natural ligand discovery. Trends in Pharmacological Sciences 22, 132-140.

Hu, Y. H., Bloomquist, B. T., Cornfield, L. J., DeCarr, L. B., FloresRiveros, J. R., Friedman, L., Jiang, P. L., LewisHiggins, L., Sadlowski, Y., Schaefer, J., Velazquez, N.,McCaleb, M. L., 1996. Identification of a novel hypothalamic neuropeptide Y receptor associated with feeding behavior. Journal of Biological Chemistry 271, 26315-26319.

Imai, K., Konno, T., Nakazawa, Y., Komiya, T., Isobe, M., Koga, K., Goto, T., Yaginuma, T., Sakakibara, K., Hasegawa, K.,Yamashita, O., 1991. Isolation and Structure of Diapause Hormone of the Silkworm, Bombyx-Mori. Proceedings of the Japan Academy Series B-Physical and Biological Sciences 67, 98-101.

Iversen, A., Cazzamali, G., Williamson, M., Hauser, F.,Grimmelikhuijzen, C. J. P., 2002. Molecular identification of the first insect ecdysis triggering hormone receptors. Biochemical and Biophysical Research Communications 299, 924-931.

Karnik, S. S., Gogonea, C., Patil, S., Saad, Y.,Takezako, T., 2003. Activation of G-protein-coupled receptors: a common molecular mechanism. Trends in Endocrinology and Metabolism 14, 431-437.

Kataoka, H., Troetschler, R. G., Li, J. P., Kramer, S. J., Carney, R. L.,Schooley, D. A., 1989. Isolation and Identification of a Diuretic Hormone from the Tobacco Hornworm, Manduca-Sexta. Proceedings of the National Academy of Sciences of the United States of America 86, 2976-2980.

Kay, I., Coast, G. M., Cusinato, O., Wheeler, C. H., Totty, N. F.,Goldsworthy, G. J., 1991a. Isolation and Characterization of a Diuretic Peptide from Acheta-Domesticus - Evidence for a Family of Insect Diuretic Peptides. Biological Chemistry Hoppe-Seyler 372, 505-512.

Kay, I., Patel, M., Coast, G. M., Totty, N. F., Mallet, A. I.,Goldsworthy, G. J., 1992. Isolation, Characterization and Biological-Activity of a Crf- Related Diuretic Peptide from Periplaneta-Americana L. Regulatory Peptides 42, 111-122.

Kay, I., Wheeler, C. H., Coast, G. M., Totty, N. F., Cusinato, O., Patel, M.,Goldsworthy, G. J., 1991b. Characterization of a Diuretic Peptide from Locusta-Migratoria. Biological Chemistry Hoppe-Seyler 372, 929-934.

Kishimoto, T., Pearse, R. V., Lin, C. J. R.,Rosenfeld, M. G., 1995. A Sauvagine Corticotropin-Releasing Factor-Receptor Expressed in Heart and Skeletal-Muscle. Proceedings of the National Academy of Sciences of the United States of America 92, 1108-1112.

Kitamura, A., Nagasawa, H., Kataoka, H., Inoue, T., Matsumoto, S., Ando, T.,Suzuki, A., 1989. Amino-Acid Sequence of Pheromone-Biosynthesis-Activating Neuropeptide (Pban) of the Silkworm, Bombyx-Mori. Biochemical and Biophysical Research Communications 163, 520-526.

Kolakowski, L. F., 1994. Gcrdb - a G-Protein-Coupled Receptor Database. Receptors & Channels 2, 1-7.

Kolhekar, A. S., Roberts, M. S., Jiang, N., Johnson, R. C., Mains, R. E., Eipper, B. A.,Taghert, P. H., 1997. Neuropeptide amidation in Drosophila: Separate genes encode the two enzymes catalyzing amidation. Journal of Neuroscience 17, 1363-1376.

Konopinska, D., 1997. Insect neuropetide proctolin and its analogues - An overview of the present literature. Journal of Peptide Research 49, 457-466.

Koroleva, O. A., Davies, A., Deeken, R., Thorpe, M. R., Tomos, A. D.,Hedrich, R., 2000. Identification of a new glucosinolate-rich cell type in Arabidopsis flower stalk. Plant Physiology 124, 599-608.

Kozak, M., 1990. Downstream Secondary Structure Facilitates Recognition of Initiator Codons by Eukaryotic Ribosomes. Proceedings of the National Academy of Sciences of the United States of America 87, 8301-8305.

Kramer, S. J., Toschi, A., Miller, C. A., Kataoka, H., Quistad, G. B., Li, J. P., Carney, R. L.,Schooley, D. A., 1991. Identification of an Allatostatin from the Tobacco Hornworm

Manduca-Sexta. Proceedings of the National Academy of Sciences of the United States of America 88, 9458-9462.

Krogh, A., Larsson, B., von Heijne, G.,Sonnhammer, E. L. L., 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology 305, 567-580.

Krupnick, J. G.,Benovic, J. L., 1998. The role of receptor kinases and arrestins in G protein-coupled receptor regulation. Annual Review of Pharmacology and Toxicology 38, 289-319.

Larsen, M. J., Burton, K. J., Zantello, M. R., Smith, V. G., Lowery, D. L.,Kubiak, T. M., 2001. Type A allatostatins from Drosophila melanogaster and Diplotera puncata activate two Drosophila allatostatin receptors, DAR-1 and DAR-2, expressed in CHO cells. Biochemical and Biophysical Research Communications 286, 895-901.

Lee, J. D.,Park, J. S. (1977). Biology and control of the brown planthopper *Nilaparvata lugens*, in Korea. The Rice Brown Planthopper. Taiwan, ASPAC: 199-213.

Lee, S. I., Lee, S. H., Koo, J. C., Chun, H. J., Lim, C. O., Mun, J. H., Song, Y. H.,Cho, M. J., 1999. Soybean Kunitz trypsin inhibitor (SKTI) confers resistance to the brown planthopper (Nilaparvata lugens Stal) in transgenic rice. Molecular Breeding 5, 1-9.

Lefkowitz, R. J., 1998. G protein-coupled receptors III. New roles for receptor kinases and beta-arrestins in receptor signaling and desensitization. Journal of Biological Chemistry 273, 18677-18680.

Lehane, M. J. (1996). Digestion and fate of the vertebrate bloodmeal in insects. The Immunology of Host-Ectoparasitic Arthropod Relationships. Wallingford, UK, CAB International: 131-149.

Lenz, C., Sondergaard, L.,Grimmelikhuijzen, C. J. P., 2000a. Molecular cloning and genomic organization of a novel receptor from Drosophila melanogaster structurally related to mammalian galanin receptors. Biochemical and Biophysical Research Communications 269, 91-96.

Lenz, C., Williamson, M.,Grimmelikhuijzen, C. J. P., 2000b. Molecular cloning and genomic organization of a second probable allatostatin receptor from Drosophila melanogaster. Biochemical and Biophysical Research Communications 273, 571-577.

Lenz, C., Williamson, M.,Grimmelikhuijzen, C. J. P., 2000c. Molecular cloning and genomic organization of an allatostatin preprohormone from Drosophila melanogaster. Biochemical and Biophysical Research Communications 273, 1126-1131.

Lenz, C., Williamson, M., Hansen, G. N.,Grimmelikhuijzen, C. J. P., 2001. Identification of four Drosophila allatostatins as the cognate ligands for the Drosophila orphan receptor DAR-2. Biochemical and Biophysical Research Communications 286, 1117-1122.

Liaw, C. W., Grigoriadis, D. E., Lorang, M. T., DeSouza, E. B.,Maki, R. A., 1997. Localization of agonist- and antagonist-binding domains of human corticotropin-releasing factor receptors. Molecular Endocrinology 11, 2048-2053.

Libert, F., Parmentier, M., Lefort, A., Dinsart, C., Vansande, J., Maenhaut, C., Simons, M. J., Dumont, J. E.,Vassart, G., 1989. Selective Amplification and Cloning of 4 New Members of the G- Protein Coupled Receptor Family. Science 244, 569-572.

Lorenz, M. W., Kellner, R.,Hoffmann, K. H., 1995. A Family of Neuropeptides That Inhibit Juvenile-Hormone Biosynthesis in the Cricket, Gryllus-Bimaculatus. Journal of Biological Chemistry 270, 21103-21108.

Maeda, S., 1989. Increased Insecticidal Effect by a Recombinant Baculovirus Carrying a Synthetic Diuretic Hormone Gene. Biochemical and Biophysical Research Communications 165, 1177-1183.

Maqbool, S. B., Riazuddin, S., Loc, N. T., Gatehouse, A. M. R., Gatehouse, J. A.,Christou, P., 2001. Expression of multiple insecticidal genes confers broad resistance against a range of different rice pests. Molecular Breeding 7, 85-93.

Matteson, P. C., 2000. Insect pest management in tropical Asian irrigated rice. Annual Review of Entomology 45, 549-574.

McCarroll, L.,King, L. A., 1997. Stable insect cell cultures for recombinant protein production. Current Opinion in Biotechnology 8, 590-594.

Meola, S. M., Clottens, F. L., Coast, G. M.,Holman, G. M., 1994. Localization of Leucokinin-Viii in the Cockroach, Leucophaea- Maderae, Using an Antiserum Directed against an Achetakinin-I Analog. Neurochemical Research 19, 805-814.

Miles, P. W., 1999. Aphid saliva. Biological Reviews of the Cambridge Philosophical Society 74, 41-85.

Milligan, G., 2002. Strategies to identify ligands for orphan G-protein-coupled receptors. Biochemical Society Transactions 30, 789-793.

Mitsuhashi, J., 1974. Methods for rearing leafhoppers and planthoppers on artificial diets. Review of Plant Protection Research 7, 57-67.

Moran, P. J.,Thompson, G. A., 2001. Molecular responses to aphid feeding in Arabidopsis in relation to plant defense pathways. Plant Physiology 125, 1074-1085.

Morin, S., Ghanim, M., Zeidan, M., Czosnek, H., Verbeek, M.,van den Heuvel, J., 1999. A GroEL homologue from endosymbiotic bacteria of the whitefly Bemisia tabaci is implicated in the circulative transmission of tomato yellow leaf curl virus. Virology 256, 75-84.

Nakamura, Y., Gojobori, T.,Ikemura, T., 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Research 28, 292-292.

Nassel, D. R., 2002. Neuropeptides in the nervous system of Drosophila and other insects: multiple roles as neuromodulators and neurohormones. Progress in Neurobiology 68, 1-84.

Nielsen, H., Engelbrecht, J., Brunak, S.,vonHeijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10, 1-6.

Noronha, K. F., Lange, A. B.,Osborne, R. H., 1997. Proctolin analogues and a proctolin metabolite as antagonists of the peptide proctolin. Peptides 18, 67-72.

O'Donnell, M. J.,Spring, J. H., 2000. Modes of control of insect Malpighian tubules: synergism, antagonism, cooperation and autonomous regulation. Journal of Insect Physiology 46, 107-117.

Osbourn, A. E., 1996. Preformed antimicrobial compounds and plant defense against fungal attack. Plant Cell 8, 1821-1831.

Pampusch, M. S., Osinski, M. A., Brown, D. R.,Murtaugh, M. P., 1998. The porcine mu opioid receptor: molecular cloning and mRNA distribution in lymphoid tissues. Journal of Neuroimmunology 90, 192-198.

Panetta, R., Greenwood, M. T., Warszynska, A., Demchyshyn, L. L., Day, R., Niznik, H. B., Srikant, C. B.,Patel, Y. C., 1994. Molecular-Cloning, Functional-Characterization, and Chromosomal Localization of a Human Somatostatin Receptor (Somatostatin Receptor Type-5) with Preferential Affinity for Somatostatin-28. Molecular Pharmacology 45, 417-427.

Pang, L., Hashemi, T., Lee, H. J. J., Maguire, M., Graziano, M. P., Bayne, M., Hawes, B., Wong, G.,Wang, S. K., 1998. The mouse GalR2 galanin receptor: Genomic organization, cDNA cloning, and functional characterization. Journal of Neurochemistry 71, 2252-2259.

Park, Y., Filippov, V., Gill, S. S.,Adams, M. E., 2002. Deletion of the ecdysis-triggering hormone gene leads to lethal ecdysis deficiency. Development 129, 493-503.

Park, Y., Kim, Y. J., Dupriez, V.,Adams, M. E., 2003. Two subtypes of ecdysis-triggering hormone receptor in Drosophila melanogaster. Journal of Biological Chemistry 278, 17710-17715.

Peeters, K., De Wilde, C., De Jaeger, G., Angenon, G.,Depicker, A., 2001. Production of antibodies and antibody fragments in plants. Vaccine 19, 2756-2761.

Peterson, C. J., Tsao, R.,Coats, J. R., 1998. Glucosinolate aglucones and analogues: Insecticidal properties and a QSAR. Pesticide Science 54, 35-42.

Pfeifer, T. A., Hegedus, D. D., Grigliatti, T. A.,Theilmann, D. A., 1997. Baculovirus immediate-early promoter-mediated expression of the Zeocin(TM) resistance gene for use as a dominant selectable marker in Dipteran and Lepidopteran insect cell lines. Gene 188, 183-190.

Pierce, K. L., Premont, R. T.,Lefkowitz, R. J., 2002. Seven-transmembrane receptors. Nature Reviews Molecular Cell Biology 3, 639-650.

Pingali, P. L.,Gerpacio, R. V., 1997. Living with reduced insecticide use for tropical rice in Asia. Food Policy 22, 107-118.

Plant, A., Stieber, F., Scherkenbeck, J., Losel, P.,Dyker, H., 2001. Syntheses of analogues of the insect neuropeptide proctolin containing an oxazole ring as an amide bond replacement. Organic Letters 3, 3427-3430.

Powell, K. S. (1993). Antimetabolic effect of plant proteins on homopteran insect pests. Biological Sciences. Durham, Durham University: 191.

Powell, K. S., Gatehouse, A. M. R., Hilder, V. A.,Gatehouse, J. A., 1995. Antifeedant Effects of Plant-Lectins and an Enzyme on the Adult Stage of the Rice Brown Planthopper, Nilaparvata-Lugens. Entomologia Experimentalis Et Applicata 75, 51-59.

Powell, K. S., Spence, J., Bharathi, M., Gatehouse, J. A.,Gatehouse, A. M. R., 1998. Immunohistochemical and developmental studies to elucidate the mechanism of action of the snowdrop lectin on the rice brown planthopper, Nilaparvata lugens (Stal). Journal of Insect Physiology 44, 529-539.

Pratt, G. E., Farnsworth, D. E., Siegel, N. R., Fok, K. F.,Feyereisen, R., 1989. Identification of an Allatostatin from Adult Diploptera- Punctata. Biochemical and Biophysical Research Communications 163, 1243-1247.

Pratt, G. E., Unnithan, G. C., Fok, K. F., Siegel, N. R.,Feyereisen, R., 1997. Structure-activity studies reveal two allatostatin receptor types in corpora allata of Diploptera punctata. Journal of Insect Physiology 43, 627-634.

Probst, W. C., Snyder, L. A., Schuster, D. I., Brosius, J.,Sealfon, S. C., 1992. Sequence Alignment of the G-Protein Coupled Receptor Superfamily. DNA and Cell Biology 11, 1-20.

Rafaeli, A. (2002). Neuroendocrine control of pheromone biosynthesis in moths. International Review of Cytology - a Survey of Cell Biology, Vol 213. 213: 49-91.

Raina, A. K., Jaffe, H., Kempe, T. G., Keim, P., Blacher, R. W., Fales, H. M., Riley, C. T., Klun, J. A., Ridgway, R. L.,Hayes, D. K., 1989. Identification of a Neuropeptide Hormone That Regulates Sex- Pheromone Production in Female Moths. Science 244, 796-798.

Rao, K. V., Rathore, K. S., Hodges, T. K., Fu, X., Stoger, E., Sudhakar, D., Williams, S., Christou, P., Bharathi, M., Bown, D. P., Powell, K. S., Spence, J., Gatehouse, A. M. R.,Gatehouse, J. A., 1998. Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper. Plant Journal 15, 469-477.

Reagan, J. D., 1994. Expression Cloning of an Insect Diuretic Hormone-Receptor - a Member of the Calcitonin Secretin Receptor Family. Journal of Biological Chemistry 269, 9-12.

Reagan, J. D., 1995a. Functional Expression of a Diuretic Hormone-Receptor in Baculovirus-Infected Insect Cells - Evidence Suggesting That the N-Terminal Region of Diuretic Hormone Is Associated with Receptor Activation. Insect Biochemistry and Molecular Biology 25, 535-539.

Reagan, J. D., 1995b. Molecular-Cloning of a Putative Na+-K+-2cl(-) Cotransporter from the Malpighian Tubules of the Tobacco Hornworm, Manduca- Sexta. Insect Biochemistry and Molecular Biology 25, 875-880.

Reagan, J. D., 1995c. Molecular-Cloning of Insect Diuretic Hormone Receptors and Na+- K+-2cl-Cotransporters. Abstracts of Papers of the American Chemical Society 210, 59-AGRO.

Reagan, J. D., 1996. Molecular cloning and functional expression of a diuretic hormone receptor from the house cricket, Acheta domesticus. Insect Biochemistry and Molecular Biology 26, 1-6.

Rosenkilde, C., Cazzamali, G., Williamson, M., Hauser, F., Sondergaard, L., DeLotto, R.,Grimmelikhuijzen, C. J. P., 2003. Molecular cloning, functional expression, and gene silencing of two Drosophila receptors for the Drosophila neuropeptide pyrokinin-2. Biochemical and Biophysical Research Communications 309, 485-494.

Rosso, M. N., Schouten, A., Roosien, J., BorstVrenssen, T., Hussey, R. S., Gommers, F. J., Bakker, J., Schots, A.,Abad, P., 1996. Expression and functional characterization of a single chain FV antibody directed against secretions involved in plant nematode infection process. Biochemical and Biophysical Research Communications 220, 255-263.

Rubia, E. G.,Heong, K. L., 1989. Vertical distribution of two hopper species on rice plants. International Rice Research Newsletter 14, 31-32.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Miklos, G. L. G., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., Cherry, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Birney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Chervitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S. B., Gong, F. C., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J. Y., Li, Z. Y., Hynes, R. O., Jones, S. J. M., Kuehl, P. M., Lemaitre, B., Littleton, J. T., Morrison, D. K., Mungall, C., O'Farrell, P. H., Pickeral, O. K., Shue, C., Vosshall, L. B., Zhang, J., Zhao, Q., Zheng, X. Q. H., Zhong, F., Zhong, W. Y., Gibbs, R., Venter, J. C., Adams, M. D.,Lewis, S., 2000. Comparative genomics of the eukaryotes. Science 287, 2204-2215.

Russo, D., Chazenbalk, G. D., Nagayama, Y., Wadsworth, H. L.,Rapoport, B., 1991. Site-Directed Mutagenesis of the Human Thyrotropin Receptor - Role of Asparagine-

Linked Oligosaccharides in the Expression of a Functional Receptor. Molecular Endocrinology 5, 29-33.

Sakata, K.,Miyata, T., 1994. Correlation of Esterase Isozymes to Malathion Resistance in the Small Brown Planthopper (Homoptera, Delphacidae). Journal of Economic Entomology 87, 326-333.

Sambrook, J.,Russell, D. W. (2001). Molecular cloning a laboratory manual, Cold Spring Harbour Laboratory Press.

Scherkenbeck, J., Plant, A., Stieber, F., Losel, P.,Dyker, H., 2002. Syntheses of depsipeptide analogues of the insect neuropeptide proctolin. Bioorganic & Medicinal Chemistry Letters 12, 1625-1628.

Schmid, A.,Becherer, C., 1996. Leucokinin-like immunoreactive neurones in the central nervous system of the spider Cupiennius salei. Cell and Tissue Research 284, 143-152.

Schoofs, L., Holman, G. M., Proost, P., Vandamme, J., Hayes, T. K.,Deloof, A., 1992. Locustakinin, a Novel Myotropic Peptide from Locusta- Migratoria, Isolation, Primary Structure and Synthesis. Regulatory Peptides 37, 49-57.

Secher, T., Lenz, C., Cazzamali, G., Sorensen, G., Williamson, M., Hansen, G. N., Svane, P.,Grimmelikhuijzen, C. J. P., 2001. Molecular cloning of a functional allatostatin gut/brain receptor and an allatostatin preprohormone from the silkworm Bombyx mori. Journal of Biological Chemistry 276, 47052-47060.

Sehnal, F.,Zitnan, D. (1996). Midgut endocrine cells. Biology of the Insect Midgut. M. J. Lehane and P. F. Billingsley. London, Chapman & Hall: 55-85.

Sexton, P. M., Houssami, S., Hilton, J. M., Okeeffe, L. M., Center, R. J., Gillespie, M. T., Darcy, P.,Findlay, D. M., 1993. Identification of Brain Isoforms of the Rat Calcitonin Receptor. Molecular Endocrinology 7, 815-821.

Small, G. J.,Hemingway, J., 2000. Molecular characterization of the amplified carboxylesterase gene associated with organophosphorus insecticide resistance in the brown planthopper, Nilaparvata lugens. Insect Molecular Biology 9, 647-653.

Smart, D., Johnston, C. F., Shaw, C., Halton, D. W.,Buchanan, K. D., 1993. Use of Specific Antisera for the Localization and Quantitation of Leucokinin Immunoreactivity in the Nematode, Ascaris-Suum. Comparative Biochemistry and Physiology C-Pharmacology Toxicology & Endocrinology 106, 517-522.

Starratt, A. N.,Brown, B. E., 1975. Stucture of the pentapeptide proctolin, a proposed neurotransmitter in insects. Life Sciences 17, 1253-1256.

Storm, J. E., Rozman, K. K.,Doull, J., 2000. Occupational exposure limits for 30 organophosphate pesticides based on inhibition of red blood cell acetylcholinesterase. Toxicology 150, 1-29.

Stout, M. J., Fidantsef, A. L., Duffey, S. S.,Bostock, R. M., 1999. Signal interactions in pathogen and insect attack: systemic plant-mediated interactions between pathogens and herbivores of the tomato, Lycopersicon esculentum. Physiological and Molecular Plant Pathology 54, 115-130.

Stroop, S. D., Nakamuta, H., Kuestner, R. E., Moore, E. E.,Epand, R. M., 1996. Determinants for calcitonin analog interaction with the calcitonin receptor N-terminus and transmembrane-loop regions. Endocrinology 137, 4752-4756.

Tanaka, K., 1999. Quantitative genetic analysis of biotypes of the brown planthopper Nilaparvata lugens: heritability of virulence to resistant rice varieties. Entomologia Experimentalis Et Applicata 90, 279-287.

Tavladoraki, P., Benvenuto, E., Trinca, S., Demartinis, D., Cattaneo, A.,Galeffi, P., 1993. Transgenic Plants Expressing a Functional Single-Chain Fv- Antibody Are Specifically Protected from Virus Attack. Nature 366, 469-472.

Teal, P. E. A.,Nachman, R. J., 2002. A brominated-fluorene insect neuropeptide analog exhibits pyrokinin/PBAN-specific toxicity for adult females of the tobacco budworm moth. Peptides 23, 801-806.

Tensen, C. P., Cox, K. J. A., Burke, J. F., Leurs, R., van der Schors, R. C., Geraerts, W. P. M., Vreugdenhil, E.,van Heerikhuizen, H., 1998. Molecular cloning and characterization of an invertebrate homologue of a neuropeptide Y receptor. European Journal of Neuroscience 10, 3409-3416.

Towers, P. R.,Sattelle, D. B., 2002. A Drosophila melanogaster cell line (S2) facilitates post- genome functional analysis of receptors and ion channels. Bioessays 24, 1066-1073.

Truman, J. W., Ewer, J.,Ball, E. E., 1996. Dynamics of cyclic GMP levels in identified neurones during ecdysis behaviour in the locust Locusta migratoria. Journal of Experimental Biology 199, 749-758.

Ulrich, C. D., Holtmann, M.,Miller, L. J., 1998. Secretin and vasoactive intestinal peptide receptors: Members of a unique family of G protein-coupled receptors. Gastroenterology 114, 382-397.

Unson, C. G., Cypess, A. M., Wu, C. R., Goldsmith, P. K., Merrifield, R. B.,Sakmar, T. P., 1996. Antibodies against specific extracellular epitopes of the glucagon receptor

block glucagon binding. Proceedings of the National Academy of Sciences of the United States of America 93, 310-315.

Van Damme, E. J. M., Allen, A. K.,Peumans, W. J., 1987. Isolation and characterisation of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. Febs Letters 215, 140-144.

VandenBroeck, J., Schoofs, L.,DeLoof, A., 1997. Insect neuropeptides and their receptors - New leads for medical and agricultural applications. Trends in Endocrinology and Metabolism 8, 321-326.

Vaughan, M., 1998. G protein-coupled receptors minireview series. Journal of Biological Chemistry 273, 17297-17297.

Veenstra, J. A., 1999. Isolation and identification of three RFamide-immunoreactive peptides from the mosquito Aedes aegypti. Peptides 20, 31-38.

Vontas, J. G., Small, G. J.,Hemingway, J., 2001. Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in Nilaparvata lugens. Biochemical Journal 357, 65-72.

Vontas, J. G., Small, G. J., Nikou, D. C., Ranson, H.,Hemingway, J., 2002. Purification, molecular cloning and heterologous expression of a glutathione S-transferase involved in insecticide resistance from the rice brown planthopper, Nilaparvata lugens. Biochemical Journal 362, 329-337.

Voss, A., Niersbach, M., Hain, R., Hirsch, H. J., Liao, Y. C., Kreuzaler, F.,Fischer, R., 1995. Reduced Virus Infectivity in Nicotiana-Tabacum Secreting a Tmv- Specific Full-Size Antibody. Molecular Breeding 1, 39-50.

Warburton, H., Palis, F. G.,Pingali, P. L. (1995). Impact of Pesticides on Farmer Health and the Rice Environment. Los Bafios, Laguna, Philippines, Academic Publishers, Norwell, MA and International Rice Research Institute.

Watanabe, T.,Kitagawa, H., 2000. Photosynthesis and translocation of assimilates in rice plants following phloem feeding by the planthopper Nilaparvata lugens (Homoptera : Delphacidae). Journal of Economic Entomology 93, 1192-1198.

Wettstein, J. G., Earley, B.,Junien, J. L., 1995. Central-Nervous-System Pharmacology of Neuropeptide-Y. Pharmacology & Therapeutics 65, 397-414.

Williamson, M., Lenz, C., Winther, A. M. E., Nassel, D. R.,Grimmelikhuijzen, C. J. P., 2001a. Molecular cloning, genomic organization, and expression of a C- type (Manduca sexta-type) allatostatin preprohormone from Drosophila melanogaster. Biochemical and Biophysical Research Communications 282, 124-130.

Williamson, M., Lenz, C., Winther, M. E., Nassel, D. R.,Grimmelikhuijzen, C. J. P., 2001b. Molecular cloning, genomic organization, and expression of a B- type (cricket-type) allatostatin preprohormone from Drosophila melanogaster. Biochemical and Biophysical Research Communications 281, 544-550.

Woodhead, A. P., Stay, B., Seidel, S. L., Khan, M. A.,Tobe, S. S., 1989. Primary Structure of 4 Allatostatins - Neuropeptide Inhibitors of Juvenile-Hormone Synthesis. Proceedings of the National Academy of Sciences of the United States of America 86, 5997-6001.

Yu, C. G., Stay, B., Ding, Q., Bendena, W. G.,Tobe, S. S., 1995. Immunochemical Identification and Expression of Allatostatins in the Gut of Diploptera-Punctata. Journal of Insect Physiology 41, 1035-1043.

Zimmermann, S., Schillberg, S., Liao, Y. C.,Fisher, R., 1998. Intracellular expression of TMV-specific single-chain Fv fragments leads to improved virus resistance in Nicotiana tabacum. Molecular Breeding 4, 369-379.

Zitnan, D., Sauman, I.,Sehnal, F., 1993. Peptidergic Innervation and Endocrine-Cells of Insect Midgut. Archives of Insect Biochemistry and Physiology 22, 113-132.

Zlotkin, E., 1999. The insect voltage-gated sodium channel as target of insecticides. Annual Review of Entomology 44, 429-455.