 Discovery and Characterization of Guanylate Cyclase Receptors in Manduca sexta

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biochemistry

by

Chong Tang

David A. Schooley, Dissertation Advisor

May 2012
We recommend that the dissertation prepared under our supervision by

David A. Schooley

entitled

Discovery and Characterization of Guanylate Cyclase Receptors in Manduca sexta

be accepted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

David A. Schooley, Advisor

David Shintani, Committee Member

Gary J. Blomquist, Committee Member

Claus Tittiger, Committee Member

Kathleen Keef, Committee Member

, Graduate School Representative

Marsha H. Read, Ph. D., Dean, Graduate School
ACKNOWLEDGEMENTS

Foremost, thanks go to my family. It is their unconditional love and support that make me who I am today. They have given me my intelligence, health, love and appreciation for humanity.

I thank Dr. David Schooley, for his patient guidance throughout my 5 years of graduate life in America. Without a doubt, I couldn’t have been able to accomplish the task of higher learning if not for his encouragement, wisdom, humor, ski, and Macintosh training.
ABSTRACT

Natriuretic peptide receptors are very important in mammals for controlling water homeostasis in the body. It is possible that during evolution, these receptors played a key role in the process, by which animals moved from the sea to the land. Until now, there has been little information on most natriuretic peptide receptors in insects. We use Manduca sexta as our research model.

We have now found five guanylate cyclase receptors in this organism similar to the ANP receptor, which are found in different tissues of Manduca. We have complete four mRNA sequences for these large (>1,000 AA) proteins. The one incomplete receptor is very rare in most tissues. An ELISA for human ANP revealed a cross reacting material with hydrophobicity appropriate for ANP in Manduca sexta larvae. We are now pursuing the identity of the hormones which target these receptors.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>ix</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Part I. Guanylate Cyclase Receptors</td>
<td>1</td>
</tr>
<tr>
<td>Receptor Biochemistry</td>
<td>1</td>
</tr>
<tr>
<td>Biochemistry of GC-A (Human guanylate cyclase receptor)</td>
<td>7</td>
</tr>
<tr>
<td>Soluble Guanylate Cyclases</td>
<td>9</td>
</tr>
<tr>
<td>Eclosion Hormone Receptor</td>
<td>12</td>
</tr>
<tr>
<td>Other Guanylate Cyclase Receptors</td>
<td>13</td>
</tr>
<tr>
<td>Part II. Insect water homeostasis</td>
<td>14</td>
</tr>
<tr>
<td>Diuretic peptides</td>
<td>20</td>
</tr>
<tr>
<td>Antidiuretic peptides</td>
<td>27</td>
</tr>
<tr>
<td>Part III. Peptides activating cGMP elevation</td>
<td>28</td>
</tr>
<tr>
<td>Physiology of ANP and BNP</td>
<td>28</td>
</tr>
<tr>
<td>Insect antidiuretic factors :(ADF)</td>
<td>32</td>
</tr>
<tr>
<td>Ion Transport Peptide: ITP</td>
<td>33</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>40</td>
</tr>
</tbody>
</table>
Part I. Bioinformatic analyses

Part II. Searching for the conserved guanylate cyclase domain

Part III. Different methods for RACE-PCR to complete the end sequences

Part IV. Cloning methods for long sequences

Part V. A mutation in NPR5

Part VI. qPCR to detect gene abundance

Part VII. Receptor characterization

RESULTS

DISCUSSION
Sequence analyses............................................................................................................. 100
A new RACE PCR method................................................................................................... 100
Detecting alternative splicing isoforms ........................................................................... 101
Tissue localization............................................................................................................. 104
NPR6.............................................................................................................................. 107
REFERENCE..................................................................................................................110
APPENDIX....................................................................................................................126
LIST OF FIGURES

Figure 1. Receptor classifications................................................................. 2
Figure 2. The structure of a guanylate cyclase receptor; ........................................ 4
Figure 3. A model for GPCRs............................................................ 6
Figure 4. Alignment of human GC-A (P16066.1) and human soluble guanylate cyclase
(CAA75738.1). .............................................................................................. 11
Figure 5. The isolated Malpighian tubule of Aedes aegypti under control conditions.... 16
Figure 6. Mechanism of transepithelial NaCl and KCl secretion in Malpighian tubules of
the yellow fever mosquito Aedes aegypti under control conditions...................... 18
Figure 7. Mechanism of action of mosquito natriuretic peptide (MNP).................. 23
Figure 8. Mechanism of action of kinins. ...................................................... 26
Figure 9. The primary structures of ANP, BNP and CNP. ............................... 30
Figure 10. Proposed model for control of ion transport across desert locust ileum. ...... 36
Figure 11. Alignment of receptors homologous to Rattus norvegicus GC-A in different
species. ........................................................................................................ 42
Figure 12. Protein alignment of guanylate cyclases from different species. .............. 50
Figure 13. Degenerate primers designed for nested PCR..................................... 54
Figure 14A. The general structure of the receptor guanylate cyclases. ..................... 56
Figure 14B. The conserved partial sequence AFFGPEG is located in the ECD....... 56
Figure 15. The mechanism of 3′ RACE. .......................................................... 61
Figure 16. Procedures for 5′ RACE PCR........................................................ 64
Figure 17. Mechanism of 5′ and 3′ RACE-PCR............................................. 72
Figure 18. Experimental design and steps involved in bridging PCR (BPCR)............ 74
Figure 19. The mechanism of In-Fusion technology from Clontech.......................... 76
Figure 20. An alternative splicing site in NPR5. .................................................... 84
Figure 21. Flow chart of the Phusion site-directed mutagenesis kit protocol.............. 87
Figure 22. Results of nested degenerate PCR from cDNA........................................ 95
Figure 23. Different PCR systems compatible with TdT buffer................................. 96
Figure 24. PCR efficiency of the NPR5 3' fragment was tested in different PCR buffers .
................................................................................................................................. 97
Figure 25. Quantitative PCR of cDNA from 4 tissues and from whole body extracts.... 98
Figure 26. Preliminary hormone dose-response curve for NPR5............................... 99
Figure 27. Partial gene sequence of NPR5 which explains the alternative splicing...... 103
LIST OF TABLE

Table 1. Primary structures of selected insect neuropeptides. ........................................... 19
Table 2. Orthologous families of insect rGCase. .................................................................. 43
Table 3. TRIzol composition ................................................................................................. 44
Table 4. Standard mixed base abbreviations with their degeneracy. ................................. 49
Table 5. Primers used to detect the 5’ end of the receptor guanylate cyclases ................. 55
Table 6. Primers used to analyze the 3’ end of other guanylate cyclase receptors ........... 60
Table 7. Primer list for 5’ race PCR to detect the 5’ region of guanylate cyclase receptors. ................................................................................................................................. 68
Table 8. Optimized conditions for PCR. .............................................................................. 78
Table 9. Primers used for fragment cloning. ........................................................................ 80
Table 10. Primers used for qPCR. ...................................................................................... 89
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Amino acid</td>
</tr>
<tr>
<td>ADF</td>
<td>Antidiuretic factor</td>
</tr>
<tr>
<td>CAP</td>
<td>Cardioacceleratory peptide</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CHH</td>
<td>Crustacean hyperglycemic hormone</td>
</tr>
<tr>
<td>CRF</td>
<td>Corticotropin-releasing factor</td>
</tr>
<tr>
<td>CTSH</td>
<td>Chloride transport stimulating hormone</td>
</tr>
<tr>
<td>dATP</td>
<td>Deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>Deoxycytidine triphosphate</td>
</tr>
<tr>
<td>DH</td>
<td>Diuretic hormone</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>ECD</td>
<td>Extracellular domain</td>
</tr>
<tr>
<td>EG</td>
<td>Epitracheal glands</td>
</tr>
<tr>
<td>EGT</td>
<td>Ecdysteroid UDP glucosyltransferase</td>
</tr>
<tr>
<td>EH</td>
<td>Eclosion hormone</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>ETH</td>
<td>Ecdysis-triggering hormone</td>
</tr>
<tr>
<td>ETHR</td>
<td>ETH receptor</td>
</tr>
<tr>
<td>GC-A, B, C, D, E, F</td>
<td>Guanylate cyclase receptor A, B, C, D, E, F</td>
</tr>
<tr>
<td>GCD</td>
<td>Guanylate cyclase domain</td>
</tr>
<tr>
<td>GPCR</td>
<td>G-protein coupled receptor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>GSP</td>
<td>Gene specific primer</td>
</tr>
<tr>
<td>HEK</td>
<td>Human embryonic kidney cell</td>
</tr>
<tr>
<td>ICD</td>
<td>Intracellular domain</td>
</tr>
<tr>
<td>IP3</td>
<td>Inositol (1, 4, 5)-triphosphate</td>
</tr>
<tr>
<td>ITP</td>
<td>Ion transport peptide</td>
</tr>
<tr>
<td>KHD</td>
<td>Kinase homology domain</td>
</tr>
<tr>
<td>MIH</td>
<td>Molt inhibiting hormone</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MT</td>
<td>Malpighian tubules</td>
</tr>
<tr>
<td>NMUR</td>
<td>Neuromedin U receptor</td>
</tr>
<tr>
<td>NO</td>
<td>Nitric oxide</td>
</tr>
<tr>
<td>NP</td>
<td>Natriuretic peptide</td>
</tr>
<tr>
<td>NPR</td>
<td>Natriuretic peptide receptor</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDE</td>
<td>cGMP selective phosphodiesterase</td>
</tr>
<tr>
<td>PKG</td>
<td>cGMP dependent protein kinase</td>
</tr>
<tr>
<td>RACE PCR</td>
<td>Rapid Amplification of cDNA ends</td>
</tr>
<tr>
<td>rGC</td>
<td>Guanylate cyclase receptor</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>sGC</td>
<td>Soluble guanylate cyclase</td>
</tr>
<tr>
<td>TdT</td>
<td>Terminal deoxynucleotidyl transferase</td>
</tr>
<tr>
<td>TMB</td>
<td>3,3,5,5-tetramethylbenzidine</td>
</tr>
<tr>
<td>TMD</td>
<td>Transmembrane domain</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>UAP</td>
<td>Universal amplification primer</td>
</tr>
<tr>
<td>URTP</td>
<td>Universal reverse transcription primer</td>
</tr>
<tr>
<td>VIH</td>
<td>Vitellogenesis inhibiting hormone</td>
</tr>
</tbody>
</table>
INTRODUCTION

Part I. Guanylate Cyclase Receptors

Receptor Biochemistry

Cells can react to environmental changes by transduction of extracellular signals, to produce intracellular responses. Membrane-impermeable signal molecules are recognized by receptors, which are localized on the plasma membrane of the cell. There are mainly three different types of receptors: (A) the receptor may comprise a ligand-gated ion channel such as the nicotinic receptor for acetylcholine; (B) the receptor may be a ligand-regulated transmembrane enzyme, such as the tyrosine kinase receptors; (C) the receptor may not of itself possess either intrinsic enzymatic or transporter activity, but may, upon being activated by its ligand, interact specifically and directly with other membrane-localized signal adaptor molecules or effectors. To date the third type is best represented by the family of G-protein-coupled receptors, like the ones for angiotensin II and adrenaline (Figure 1).
Figure 1. Receptor classifications. (A) Oligomeric ion channel- the receptor may comprise a ligand-gated ion channel. (B) Receptor tyrosine kinase or receptor guanylate cyclase- the receptor may be a ligand-regulated transmembrane enzyme. (C) G-protein-coupled receptor- the receptor may not of itself possess either intrinsic enzymatic or transporter activity (Voet and Voet, 2004).
The activity of a variety of hormones and neurotransmitters (e.g. acetylcholine) depends on a ligand-specific regulation of ion conductance. The nicotinic receptor for acetylcholine can be taken as a prototype for this category of receptor systems. The binding of acetylcholine changes the conductance properties of the receptor channel, thereby causing depolarization and tissue activation (Lodish, 2000).

Another different group of cell surface receptors share a single transmembrane domain in all members. Receptor protein kinases are a typical family in this group (Lodish, 2000). Extracellular signal molecules, like epidermal growth factor, platelet derived growth factor, and insulin bind to receptor protein kinases. These receptors possess one or two intracellular protein kinase regions. Binding of the ligand to a receptor protein kinase may lead to receptor dimerization, intermolecular autophosphorylation, and receptor protein kinase activation. Guanylate cyclase receptors are another family with a single transmembrane domain, which our group is focusing on. Guanylate cyclase receptors serve as receptors that produce cyclic GMP (cGMP) from GTP in response to ligand binding. In general, guanylate cyclase receptors have a variable extracellular domain (ECD), a single transmembrane domain (TMD), an intracellular protein kinase homology domain (KHD) and intracellular guanylate cyclase catalytic domain (GCD).

Binding of ANP to its dimeric receptor has been shown to cause it to rotate (Ogawa et al., 2004b), causing guanylate cyclase activation (Fig. 2). The known ligands for mammalian transmembrane guanylate cyclases fall into two families: (1) the natriuretic peptides, such as atrial natriuretic peptide (ANP), and brain natriuretic peptide (BNP); and (2) guanylin, uroguanylin, and the Eschericia coli heat stable enterotoxins.
Figure 2. The structure of a guanylate cyclase receptor; NPRA. ECD- extracellular domain, TMD- a single transmembrane domain, KHD- an intracellular protein kinase like domain, and GCD- intracellular guanylate cyclase catalytic domain.
G-protein coupled receptors (GPCRs) act by perhaps the most commonly used mechanism for generating intracellular signals. The structural features shared by all members of the GPCR superfamily include seven membrane-spanning domains, a putative extracellular ligand binding domain and an intracellular domain responsible for interaction with G-proteins or other intracellular signaling proteins. Upon activation by ligands, GPCRs can mediate a variety of intracellular responses via G proteins (e.g. regulating ion channels, inducing hormone secretion, or affecting enzyme activity or gene expression). Upon activation, the G protein triggers the exchange of GDP for GTP on its α-subunit, which leads to dissociation of the GTP-bound subunit, which then promotes the stimulation of downstream effectors, such as adenylate cyclase or phospholipase C (Figure 3).
Figure 3. A model for GPCRs. After binding of adrenaline (hormone), the G protein triggers the exchange of GDP for GTP on its α-subunit, which leads to dissociation of the GTP-bound subunit thereby promoting the interaction with adenylate/adenylyl cyclase, producing the second messenger cAMP (Voet and Voet, 2004).
Biochemistry of GC-A (Human guanylate cyclase receptor)

There has been very little research on insect guanylate cyclase receptors. We can only explore insect guanylate cyclase receptors based on research on human guanylate cyclase receptors (rGC). The atrial natriuretic peptide (ANP) receptor (ANPR) is the most studied member of the receptor guanylate cyclase family. The natriuretic peptide receptors, and other rGCs, share an overall topology. Human and rat GC-A mRNA is highly expressed in the kidney, adrenal gland, ileum, aorta, and adipose and lung tissue. The GC-A precursor protein has a signal peptide in the N-terminus, which determines the protein location. The receptor is a homodimer of a single-span trans-membrane polypeptide. The general topology of the GC receptor family consists of four distinct domains, an extracellular ligand-binding domain of approximately 450 amino acids (ECD), a 22- to 25- residue single hydrophobic membrane-spanning domain, and an intracellular domain (ICD) of approximately 500-700 amino acids, consisting of a 250-amino-acid kinase homology domain (KHD), a roughly 40-residue coiled-coil dimerization domain, and an approximately 250-amino-acid carboxyl-terminal guanylate cyclase catalytic domain (GCD) (Figure 1).

GC-A is a complex molecule that needs to be post-translationally modified, including glycosylation, phosphorylation, and disulfide bond formation. As such, modifications are most accurately performed by mammalian cells; they have become the major host for GC-A production. A Japanese group reported that expression in insect cells can give the active ECD of GC-C after changing the secretion signal peptide. In my experiments, the EGT sequence, a strong secretion signal peptide in insects (Hasegawa et
al., 1999), was connected to GC-A lacking the native signal peptide. However, I did not detect any activity upon expression in Sf9 cells.

The ECD of rat GC-A includes three intramolecular disulfide bonds between Cys60-Cys86, Cys164-Cys215, and Cys423-Cys432. Asn-13, Asn-180, Asn-306, Asn-347, and Asn-395 are glycosylated in the ECD. Glycosylation is essential for correct folding, processing and transport to the cell membrane. The crystal structure of the glycosylated, unliganded dimerized ECD of rat GC-A was determined at 2.0 Å resolution by Dr. Misono's group (Ogawa et al., 2003). The monomer structure displays a type I periplasmic binding protein fold and consists of two interconnected subdomains, each containing a central sheet flanked by helices. Although these authors originally proposed that the dimer adopts a tail-to-tail V-shape, the second crystal structure of the ECD bound to ANP revealed that the receptor forms a head-to-head, A-shape dimer with a stoichiometry of one molecule of ANP per two receptor molecules (Misono et al., 2005a; Ogawa et al., 2004a). These two receptors are activated by the rotation caused by ANP binding, which differs from the dimerization manner of receptor protein kinases (Misono et al., 2011).

The KHD of GC-A is more closely related to protein tyrosine kinase receptors than protein serine/threonine kinases. It has also been demonstrated that the KHD of GC-A serves as an important mediator in transducing a ligand-induced signal to activate the GCD (Misono et al., 2011). The ICD expressed from CHO cells (Chinese hamster ovary) has no any detectable GC activity (Kuhn, 2003). This may suggest that ATP serves as an intracellular allosteric regulator of the KHD for the activation of GC-A (Misono et al., 2005b).
The single transmembrane domain ANP receptors contain a single guanylate cyclase catalytic active site per molecule, however, based on homology modeling data with an adenylate cyclase structure, two polypeptide chains seem to be required to activate GC-A guanylate cyclase activity (Liu et al., 1997b). Collectively, the signaling mechanism of the single transmembrane receptors is not well understood. The reason for this void is that no intact receptor structure has been reported for any single transmembrane receptor to date.

**Soluble Guanylate Cyclases**

Guanylate cyclases have two forms: the receptor form, which is a single transmembrane domain protein, and the soluble form. Soluble guanylate cyclases (sGC) work as receptors for the signaling agent nitric oxide (NO). sGC activity is regulated by NO, GTP, ATP and allosteric activator. Soluble guanylate cyclases are heterodimeric hemoproteins, and are important for several physiological functions. Binding of NO to a prosthetic heme group causes marked activation of the soluble guanylate cyclase. There are isoforms in animals including α1, α2, β1 or β2 subunits (Liu et al., 1997b). They are generally only active as heterodimers; active forms include α1β1 and α2β2 heterodimers. *Manduca* has a different β3 sGC, which is active as a homodimer, but inactive as a heterodimer (Morton and Anderson, 2003). A further complication is the existence of splice variants of sGC forms. Until now no crystal structure of the entire receptor has been solved. The sGC structure has been modeled based on an adenylate cyclase crystal structure (Liu et al., 1997a).

The guanylate cyclase catalytic domains are conserved between soluble guanylate cyclases and guanylate cyclase receptors (Figure 4). We performed an alignment of
human GC-A and GC-As in other species using VectorNTI. We found three partial sequences (SDIVGFIA, DVYKVETIGD and MPRYCLFG) which are more conserved than other regions. We believe these three domains are critical for catalytic functions to generate cGMP from GTP. The amino acids marked as bold characters (Gly, Arg, & Cys) determine the substrate specificity (Liu et al., 1997a). Mutations in these three amino acids can change the substrate specificity (Liu et al., 1997a).
Figure 4. Alignment of human GC-A (P16066.1) and human soluble guanylate cyclase (CAA75738.1). The domains SDIVGFT\(_A\) (D1), D\(\text{VYKVE}T\)GD (D2) and MP\(\text{RYCLFG}\) (D3) are conserved between the two different types of guanylate cyclase. Underlines in the abbreviation mean the nucleotide is not conserved.
Eclosion Hormone Receptor

Eclosion hormone (EH) is an insect neuropeptide involved in emergence of the insect from the shed cuticle during ecdysis; it was identified at the protein level in *Manduca sexta* (Kataoka et al., 1987) and *Bombyx mori* (Nagasawa et al., 1985). In *Manduca*, EH is produced by ventromedial cells in the brain and is released into the hemolymph (Hewes and Truman, 1991). In response to circulating EH, massive release of ecdysis-triggering hormone (ETH) occurs from Inka cells which triggers initiation of emergence from the shed exoskeleton (Ewer et al., 1997; Zitnan et al., 1996) (Hewes and Truman, 1991). Several papers have shown that EH can cause cGMP elevation in target cells, such as Inka cells and certain brain cells (Ewer et al., 1997).

In 2009, a Taiwanese group reported an EH receptor from *Bactrocera dorsalis*, BdmGC-1, a guanylate cyclase receptor (Chang et al., 2009). It exhibits all of the characteristics of a typical guanylate cyclase receptor. BdmGC-1 is most similar to NPR5 (one of our *Manduca sexta* rGCs). We located homologous proteins in different species by BLAST search and alignment. We believe their conclusions are incorrect. They also claim that another alternatively spliced variant BdmGC-1B encodes an isoform of this enzyme (Chang et al., 2009). The B isoform has the same conserved domains and putative N-glycosylation sites found in BdmGC-1, but possesses an additional 46 amino acid insertion in the extracellular domain and lacks the C-terminal tail of BdmGC-1. BdmGC-1B has some native guanylate cyclase activity, which should cause the accumulation of cGMP in the cell. There is no sequenced genome for this species. In our experiments, we did not observe such a variant. On the other hand, we found a stop codon between the kinase domain and the guanylate cyclase domain in an isoform. Some putative guanylate
cyclase receptors found by BLAST analysis of insect genomic data also contain this stop codon, for example, in silkworm genomic data and preliminary Manduca genomic data. I will discuss this more in the next chapter.

**Other Guanylate Cyclase Receptors**

Animals contain six guanylate cyclase receptors (Lucas et al., 2000), as well as soluble guanylate cyclases. The rGC are GC-A (also named GC-A or NPR-1 activated by ANP and BNP), GC-B (also named GC-B activated by CNP), GC-C (activated by heat-stable enterotoxin, guanylin, and uroguanylin), GC-D (olfactory guanylate cyclase), GC-E and GC-F (the latter being the retinal guanylate cyclases involved in visual signal transduction). NPR-C is a clearance receptor without a guanylate cyclase domain. GC-C is an intestinal rGC, strongly activated by a heat stable enterotoxin from various bacteria. Production of cGMP in intestinal cells leads to loss of NaCl and water. GC-C is also stimulated by the peptides guanylin and uroguanylin, produced primarily by intestinal mucosa cells. Recent evidence shows uroguanylin also acts on the kidney (Yuge et al., 2003).

While GC-D, -E, and -F are all rGCases and possess the rGCase juxtamembrane motif, they are all orphan receptors with no known ligand. It seems unlikely that ligands are required for the function of the olfactory GC-D and the receptors found in the retina, GC-E and GC-F. The function of GC-D, and another novel receptor GC-G (Schulz et al., 1998), is unknown.

The genome of *Caenorhabditis elegans* contains about ten receptor GCs, and even more soluble GC. A BLAST search of the *Drosophila* genome with the guanylate cyclase conserved catalytic domain shows five homologous rGC proteins. The molecular weights
span from 1000 to 1500 amino acids which is the consistent with other rGC molecular weights. A similar result is found in a BLAST search of the Tribolium genome data. It appears that there are five guanylate cyclase receptors in insects.

**Part II. Insect water homeostasis**

Although the kidney is the primary organ for regulating salt and water balance, it is not the exclusive regulator of the extracellular fluid compartment. Amphibians use the gills and the skin in addition to the kidney, and reptiles and some birds have nasal salt glands to assist the kidney in salt and water balance. Beetles and butterflies possess the astonishing ability to absorb water from the air in the crytonephric complex (Beyenbach, 2003). The primary organs of salt and water balance for insects are generally Malpighian tubules (MTs), salivary glands, midgut and hindgut; it should be noted in the latter context that in vertebrates the intestines can in certain conditions (in the presence of bacterial endotoxins) cause massive fluid transport (Marshall, 2005).

Water and ion balance in insects is tightly controlled, but the basic mechanisms differ from vertebrate animals. There is little or no blood pressure in the insect circulatory system, so urine formation is by active transport, rather than by filtration, which is the mechanism used by the human kidney (Pannabecker et al., 1993). Malpighian tubules (MTs) are blind-end tubes which generally empty into an ampulla which is located at the junction between the midgut and hindgut and secrete KCl or NaCl rich urine. This fluid is essentially isoosmotic to hemolymph, so water and solutes follow by passive diffusion. At the same time, active transport is used for uric acid and toxic metabolites. The driving force for secretion is a vacuolar-type proton pump; a V-H⁺-ATPase located in the apical membrane of MT principal cells. It acts in parallel with secondary active antiporters for
Na\(^+\)/H\(^+\) and K\(^+\)/H\(^+\) exchange. Blood-feeding insects largely secrete NaCl into the tubule lumen, and phytophagous insects secrete largely KCl. Selective reabsorption in the ileum and rectum of essential metabolites, including ions and water, determines the composition of the final excreta. Figure 5 illustrates the basic transepithelial transport system under control conditions in the Malpighian tubule of the yellow fever mosquito. There are two pathways into the tubule lumen: a transcellular pathway through principal and stellate cells and a paracellular pathway between these cells. The transcellular pathway involves solute entry across the basolateral membrane, passing through the cell interior and exiting across the apical membrane into the tubule lumen. In the paracellular pathway, the solute penetrates into the lumen through the junction between the epithelial cells.
Figure 5. The isolated Malpighian tubule of *Aedes aegypti* under control conditions (Aneshansley et al., 1989). To move $K^+$ from 3.4 mM in the peritubular bath to 91 mM in the tubule lumen requires a driving force (chemical potential) of 87.1 mV, calculated as $E^0 = 61 \text{ mV} \log(91/3.4)$. Adding to this the lumen-positive voltage of 52.6 mV (electrical potential), against which $K^+$ is moved, yields a total electrochemical potential (139.7 mV) needed to transport $K^+$ into the tubule lumen. Similar calculations for $Na^+$ yield an electrochemical potential of 40.2 mV against which this cation is secreted. To move $Cl^-$ in the peritubular bath to 161 mmol $Cl^-$ in the tubule lumen requires the small driving force of 0.5 mV. However, the transepithelial voltage is lumen-positive (52.6 mV), pulling $Cl^-$ into the tubule lumen. Thus, $Cl^-$ moves into the tubule lumen down an electrochemical potential (passive transport) of 52.1 mV from 158 mmol.
The mechanism by which solutes pass through the membrane is complex. Here we discuss some modes of action. The active transporter driving all processes is an ATP-consuming proton pump located in the apical plasma membrane of the brush border membrane (Figure 6). These V-type ATPases are found in all species, but are also of special importance in plants. The pump consists of two components: the V$_1$ complex capable of hydrolysis of ATP to supply energy, and the V$_0$ complex which serves to pump protons across the membrane. Protons secreted into the tubule lumen are largely returned into the cell in exchange for Na$^+$ or K$^+$. The lumen positive potential vs. the hemolymph is sufficient to energize Cl$^-$ passage from hemolymph to lumen, which is the mechanism of transepithelial Cl$^-$ secretion.

For hemolymph homeostasis, insects must regulate their water and ion balance. The insect normally needs to conserve water rather than to expel it. Diuretic hormones are defined as peptides that increase water loss by the insect whole body, by increasing the tubule secretion. Antidiuretic peptides are those that cause the conservation of water by the whole insect, either by inhibiting tubule secretion or by promoting the reabsorption of ions and water in the hindgut. Table 1 summarizes the structures and functions of some neuropeptides. Some antidiuretic peptides are considered to be candidates to stimulate guanylate cyclase receptors, hormones activating cGMP production.
Figure 6. Mechanism of transepithelial NaCl and KCl secretion in Malpighian tubules of the yellow fever mosquito *Aedes aegypti* under control conditions (Aneshansley et al., 1989). K⁺ enters principal cells from outside via K⁺ channels located in the basolateral membrane. Na⁺ enters *via* cotransport with K⁺ and Cl⁻. Na⁺ and K⁺ are extruded from the cell across the apical membrane through specific transport proteins.
<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locmi-DH</td>
<td>MGMGPSLSIVNPMDVLQRRLLEIARRRLRDAEEQIKANKDFLQQI-amide</td>
</tr>
<tr>
<td>Locmi-K</td>
<td>AFSSWG-amide</td>
</tr>
<tr>
<td>Manse-CAP\textsubscript{2b}</td>
<td>pELYAFPRV-amide</td>
</tr>
<tr>
<td>Tenmo-ADF\textsubscript{a}</td>
<td>VVNTPGHAVSYHVY</td>
</tr>
<tr>
<td>Tenmo-ADF\textsubscript{b}</td>
<td>YDDGSYKPHIYGF</td>
</tr>
<tr>
<td>Locmi-neuroparsin</td>
<td>NPIRSCEGANCVVDLTRCEYGDVTDFFGRKVCAKGPGDKCGGPYELHGKCGVGMDCRCGLCSLHNLQCFFFEGLPSSC</td>
</tr>
<tr>
<td>Schgr-ITP</td>
<td>SFFDIQCKGVYDKSIFARLDRICEDCYNLFREPQLHSLCRSDCFKSPYFKGCLQALLLIDEEEEKFNQMVEIL-amide</td>
</tr>
<tr>
<td>Manse-EH</td>
<td>NPAIATGYDPMEMICNCAQCKKMLGAWFEGPLCAESCIKFKGKLIPECEDFASIAPFLNLK</td>
</tr>
</tbody>
</table>
Diuretic peptides

The corticotropin-releasing factor-related diuretic hormones

The corticotropin-releasing factor related diuretic hormone family has over 30 identified members (Schooley et al., 2012), which increase urine secretion by MTs isolated in vitro. These peptides have sequence homology to the vertebrate CRF superfamily, which includes urotensin, urocortin, and sauvagine. CRF-related DHs have been indentified in many species, for example, crickets, cockroaches, termites, tenebrionid beetles, sphingid moths, and many species of flies (Schooley et al., 2012).

Some species have two forms of CRF-related DH, a short form (30-37 amino acids) and a long form (41-54 amino acids). It may be the rule that most species contain two forms of DH, which appear to be paralogs. The gene for the CRF-related DH (Drome-DH44) in D. melanogaster has been reported (Cabrero et al., 2002). Its effect is to stimulate the secretion of MT by elevating cAMP in the principal cells. Using a cAMP bioassay, other DHs have been identified. While attempting to isolate CRF-like DH from the cockroach *Diploptera punctata*, Furuya *et al.* (2000) isolated a 46 residue CRF-like DH, together with a shorter peptide (31 amino acids), which they termed Dippu-DH31. This peptide has some sequence similarity to calcitonin, and is quite different from CRF-related DHs.

These peptides are called calcitonin-like DHs.

The hormonal status of Locmi-DH in the locust is well established in tissues and in the circulation (Coast et al., 1993). Manse-DH and the shorter Manse-DPII are also active at physiological doses in increasing excretory water loss in larval and adult *Manduca sexta* (Furuya et al., 2000; Jeffs, 1993). Some CRF-related DH receptors are identified. All of them are G protein coupled receptors (GPCRs) associated with
adenylate cyclase activity. GPCRs are seven-transmembrane domain proteins, as mentioned in the last chapter. The DH ligands have high affinity for these receptors, their binding elevates cAMP concentration. cAMP appears to mediate the activity of all CRF-related DHs (Schooley et al., 2012) in the mosquitoes, *Culex salinarius* (Clark et al., 1998) and *Anopheles gambiae* (Coast et al., 2005). The mode of action of the CRF-related DHs appears to be relatively uniform, and there are varying degrees of interspecific bioactivity. The CRF-related DHs activate GPCRs in primary cells of the MT, elevating cAMP levels in these cells, although in crickets (Coast and Kay, 1994) and the mosquitoes (Clark et al., 1998; Coast et al., 2005) they may also elevate intracellular Ca$^{2+}$. Figure 7 illustrates the general mode of action of the CRF family. The action of cAMP in the primary cells is to increase Na$^+$ conductance across the basolateral membrane. Na$^+$ can enter the cell from hemolymph *via* either Na$^+$ channels, or *via* the Na$^+$/K$^+$/2Cl$^-$ cotransporter, thus increasing intracellular Na$^+$ concentration. This presents the basal Na$^+$/H$^+$ antiporter with a higher concentration of substrate, thereby increasing cation transport, with water and other solutes (Cl$^-$) following passively through specific transport proteins, many in the junction between cells. Whether cAMP leads to stimulation of proton extrusion across the apical membrane is unknown.

In the tsetse fly, cAMP appears to transiently increase MT cell volume by increasing basolateral permeability to both Na$^+$ and Cl$^-$, with cell volume returning to normal as the transcellular movement of fluid increases (Hagger et al., 1997). Ultrastructural studies show that cAMP also affects mitochondrial location and action. It is possible that cAMP also enhances the activity of the apical H$^+$-ATPase by increasing ATP availability, further stimulating cation movement across the apical membrane. There is
recent electrophysiological evidence that cAMP can stimulate the H\textsuperscript+-ATPase directly in

*D. melanogaster*, although this has not yet been demonstrated in other species

(Wieczorek et al., 2009).

Beyenbach has a theory, as does Dow, that increased concentrations of ATP increase the activity of the V-ATPase (Harvey, 2009). However, ATP concentrations are buffered in the cell and rarely change by even 10% (Voet and Voet, 2004). There is a metabolite, arginine phosphate, which in insects serves the same role as creatine phosphate in mammals, providing an ATP buffer.

The inactivation of CRF-related DHs has been studied with Manse-DH, which is proteolytically cleaved between residues Arg30 and Ala31, as well as between Leu29 and Arg30 by a presumed metalloprotease (Li et al., 1997). The products of this proteolytic cleavage are unlikely to be active because the separate N- and C-terminal halves of Locmi-DH have no effect on tubule secretion (Montuenga et al., 1996).
Figure 7. Mechanism of action of mosquito natriuretic peptide (MNP). MNP is a calcitonin-like DH (Coast et al., 2005). In Malpighian tubules of the yellow fever mosquito *Aedes aegypti*, MNP selectively increases the rates of transepithelial Na⁺ secretion by increasing the Na⁺ conductance of the basolateral membrane and by activating the Na⁺/K⁺/2Cl⁻ cotransporter. The second messenger of MNP and CRF-like diuretic peptides is cAMP. Whether cAMP also stimulates proton extrusion across the apical membrane is unknown.
The (myo)kinins

Another large family of diuretic peptides are the insect (myo)kinins. These were originally isolated from extracts of whole heads of Leucophaea maderae on the basis of stimulation of the cockroach hindgut in vitro, and eight members of this family occur in this cockroach. There are currently more than 20 kinins, 8 from L. maderae (leucokinins), 5 from A. domesticus (achetakinins), 1 from L. migratoria (locustakinin), 3 from Culex (culekinin depolarizing peptides), 3 from A. aegypti (aedeskinins or Aedes leucokinins), and 3 from Helicoverpa (helicokinins) (Schooley et al., 2012). These peptides are small, 6-13 residues, and have a highly conserved COOH-terminal pentapeptide sequence, FX₁X₂WGamide, where X₁ is S, H, N, or Y, and X₂ is S or P. In addition to their myotropic action, the kinins may be involved in the release of digestive enzymes into the gut. Their most important action, however, appears to be in their diuretic action on the MT. The kinin receptor of D. melanogaster has recently been cloned (Hauser et al., 2006). When it is expressed heterologously in cell lines, it binds to the conspecific kinin, Drome-K. The receptor is localized mainly in the secondary stellate cells of the MTs, as shown by immunohistochemistry; it appears to mediate kinin activity. In another dipteran species, A. aegypti, a receptor which binds Aedes kinins has been identified in Malpighian tubules (Pietrantonio et al., 2005).

The mode of action of kinins has been studied intensively in MT preparations from D. melanogaster and A. aegypti. Figure 8 demonstrates the mechanism of action of the leucokinins, which are typical kinins. The receptor for kinins is a G-protein coupled receptor. Stimulation of the G-protein is thought to activate phospholipase C and to generate inositol (1, 4, 5)-trisphosphate (IP₃) and diacylglycerol. IP₃ goes on to release
intracellular Ca$^{+2}$ from stores in the ER. The subsequent rise in cytoplasmic Ca$^{+2}$ concentration activates a Ca$^{+2}$ channel in the basolateral membrane (Radford et al., 2002). Extracellular Ca$^{+2}$ entering the cell makes the epithelium “leaky” which aids in water transport (Radford et al., 2002). How Ca$^{+2}$ increases junction conductance or permeability is currently an active field of research. Stellate cells may well mediate transepithelial Cl$^-\$ secretion. Then fluid and Cl$^-\$ can flow through the junction between cells, or through cells.

Structure activity studies on kinin analogs have tested their diuretic effects on cricket MTs in vitro. The conserved C-terminal pentapeptide of the Achdo-Ks and Leuma-Ks is the minimum sequence required to achieve the same activity and potency as the parent compounds (Nachman et al., 1995). The aromatic side chains of the invariant Phe1 and Trp4 residues of the pentapeptide interact with the receptor. When either of these residues is mutated to Ala, the resulting analog is inactive. The minimum requirement for biological activity is the C-terminal dipeptide Trp4-Gly5, extended N-terminal Phe. The amide is not entirely necessary for activity, because an ester analog of AFFWG-acid has improved potency (Nachman et al., 1995).

Degradation of kinins is relatively well researched. A membrane-bound enzyme found in MTs is similar to the vertebrate metalloprotease neprilysin. It cleaves Helze-K-II between Pro5 and Trp6, hence destroying the conserved pentapeptide sequence - the minimum requirement for biological activity (Bland et al., 2008).
Figure 8. Mechanism of action of kinins. Leucokinin belongs to the kinin family of insect diuretic peptides. Leucokinin increases the rate of both transepithelial NaCl and KCl secretion by increasing the paracellular Cl⁻ conductance via intra- and extracellular Ca⁺². Septate junction proteins (claudins, neurexins, etc. (Beyenbach et al., 2010)) are hypothesized to form channel-like extracellular structures with selectivity and variable conductance. Research has shown a non-selective stimulation of transepithelial NaCl and KCl secretion (Yu and Beyenbach, 2001). G, heterotrimeric G protein; PLC, phospholipase C; IP₃, inositol(1,4,5)-trisphosphate.
**Antidiuretic peptides**

Functionally, there are two types of antidiuretic peptides: those that inhibit MT secretion and those that stimulate fluid reabsorption by the ileum and or rectum. Two antidiuretic peptides which act on the MTs are Tenmo-ADFa (Eigenheer et al., 2002), and Tenmo-ADFb (Eigenheer et al., 2003), which will be discussed in more detail in the next chapter. The other antidiuretic peptides which act on reabsorption by the ileum and rectum are known as neuroparsins (Fournier and Girardie, 1988), which influence the rectum, and ion-transport peptide (ITP) (Audsley et al., 1992b), which stimulates ion and water reabsorption from ileum.

The neuroparsins were isolated from the neurosecretory parts of the migratory locust corpus cardiacum. The three neuroparsins are structurally related monomers containing intramolecular disulfide bridges (Hietter et al., 1991). The longer neuroparsins are 83 amino acids and the shortest one is 78 amino acids. They are able to increase water absorption by everted rectal sacs of the locust (Fournier and Girardie, 1988). Their mode of action is proposed to be an elevation of intracellular $\text{Ca}^{2+}$ by stimulating the inositol phosphate cascade. The exact details of ion transport have not been well studied. Neuroparsins expressed using a baculovirus system have positive effects on fluid transport across the rectal epithelium (Girardie et al., 1997). However, they have no effect on ion transport across the rectal epithelium, which is required for fluid uptake. This raises concerns about the antidiuretic activity of neuroparsins.

Another antidiuretic peptide, chloride transport stimulating hormone (CTSH), was isolated from the corpora cardiaca of the desert locust (Audsley et al., 1992a). It is able to increase the $\text{Cl}^-$-dependent short-circuit current across ilea or recta mounted as flat sheets
in Ussing-type chambers. Its mode of action is to stimulate chloride transport into the rectum via cAMP (CTSH acts on the rectum, ITP only on the ileum).

The well characterized antidiuretic peptide ion transport peptide (ITP) will be discussed in the next chapter as a candidate to stimulate a guanylate cyclase receptor.

**Part III. Peptides activating cGMP elevation**

**Physiology of ANP and BNP**

Studies on human or rat ANP and BNP provide a basis for exploring similar hormones in insects. Water retaining and excreting mechanisms play an important role in the evolution process from fish to tetrapods. Natriuretic peptides promote the water excreting process, which is also rather conserved between fish and tetrapods. Natriuretic peptides are a family of structurally related but genetically distinct hormones that regulate blood volume and blood pressure. This peptide family in mammals includes atrial natriuretic peptide (ANP) and brain natriuretic peptide (BNP). Their receptor GC-A, is a member of the guanylate cyclase receptor family, which causes smooth muscle relaxation, blood volume decrease, and sodium excretion by binding ANP or BNP. ANP and BNP elicit their physiological response through the synthesis of cGMP, a classic intracellular second messenger that was originally identified in rat urine in 1963 (Ashman et al., 1963). There are three known cGMP regulated/dependent proteins: cGMP-dependent protein kinase (PKG), cGMP selective phosphodiesterase (PDE), and cyclic nucleotide-gated ion channels. These proteins mediate natriuretic activity via the cGMP signal cascade. Counterbalancing with the water and sodium excretion hormones, renin, aldosterone, and angiotensin stimulate sodium and water retention (Beyenbach, 1993). Therefore, each of these hormonal systems plays a key role in the maintenance of body fluid homeostasis.
ANP is a 28-residue peptide containing a 17-residue peptide ring structure formed by an intramolecular disulfide bond (Figure 1). Structure-function studies have shown that the disulfide bond, several conserved residues in the ring structure and the C-terminal sequence Asn-Ser-Phe-Arg are essential for its biological activity (Qiu et al., 2004).

A homologous peptide, B-type natriuretic peptide (BNP), is produced mainly by the heart ventricle. BNP has similar natriuretic activity to ANP, both of which target the receptor GC-A. C-type natriuretic peptide (CNP) has weak natriuretic and vasodilatory activities. CNP activates the receptor GC-B, which is mainly localized in the brain and nervous systems. It mediates neuronal control of circulation and salt and fluid intake. ANP, BNP, and CNP share common structural features (Figure 9).

**Insect ecdysis involved hormone-EH**

The first insect neuropeptide known to elevate cGMP as second messenger is eclosion hormone (EH), a large peptide with three disulfide bonds which appears essential in the ecdysis process of all insect species studied (Truman et al., 1981).

Molting in insects consists of a sequence of events including: the synthesis and secretion of a new cuticle by the epidermis, the subsequent shedding of the old cuticle (ecdysis), and the process of eclosion, a sequence of behaviors required for the insect to emerge from the shed cuticle. These events are triggered by the regulated release of several peptides, eclosion hormone (EH) and ecdysis-triggering hormone (ETH). They work together on a number of targets to cause eclosion. The behavioral sequence which leads to ecdysis consists of two defined stages: the preecdysis sequence (serves to loosen the
Figure 9. The primary structures of ANP, BNP and CNP. Two Cys residues in each peptide are disulfide-bonded. Conserved residues are shaded.
attachment between the old and new cuticle) and the ecdysis behavior (consists of rhythmic anterior-moving peristaltic waves which culminates in shedding of the cuticle). Ecdysis behaviors are driven by a pattern of neural activity in the abdominal CNS which can be triggered by EH.

A previously unknown endocrine system, the epitracheal glands (EGs) was discovered in *Manduca* (Zitnan et al., 1996). The EGs consist of two distinct cell types, a single Inka cell and glandular cells of unknown function. The Inka cell was shown to contain ETH, which was detected in the hemolymph just after the onset of preecdysis in larvae and pharate pupae. The injection of synthetic ETH induced characteristic preecdysis and ecdysis behavior in larvae and pharate pupae (Horodyski, 1996). This demonstrates that ETH acts directly on the nervous system to trigger preecdysis and ecdysis behavior, the requirement of the tracheal system for action of EH.

It is now recognized that the requirement of the tracheal system for EH action on the CNS *in vitro* is due to the presence of segmentally distributed endocrine Inka cells attached to the tracheal tubes near each spiracle. Blood-borne EH is reported to act on a newly described receptor guanylate cyclase in Inka cells to induce the release of PETH and ETH by cGMP elevation (Zitnan et al., 2003). We described this EH guanylate cyclase receptor in Chapter 1, EH receptor. Further studies have established that EH action on the Inka cell provides a mechanism for the complete depletion of Inka cell contents necessary for the transition from preecdysis to ecdysis behavior. ETH then acts on its receptor (ETHR-A) in VM neurons in the brain to elicit further EH release. Blood-borne EH acts on the Inka cell to cause cGMP production and PETH/ETH depletion in a positive feedback loop (Zitnan and Adams, 2012).
The action of EH to stimulate cGMP production was first studied in *Manduca*, and was thought to involve a soluble GCase (Ewer et al., 1994). Two lines of evidence have been presented previously to explain the mechanism of the increase in cGMP level mediated by EH. Morton and Giunta (1992) suggested that EH stimulated an increase in cGMP level through a pathway involving the intracellular release of arachidonic acid. They later reported that EH signaling involves the activation of phospholipase C (Morton and Simpson, 2002), and that either Ca\(^{2+}\) ions or free fatty acids cause the activation of a novel soluble guanylate cyclase which is not NO sensitive. The receptor was believed to be a GPCR or a receptor Tyr kinase (Morton and Simpson, 2002). In *Bombyx*, Shibanaka et al. (1993) proposed a pathway leading from production of inositol trisphosphate through calcium activation of nitric oxide synthase, production of nitric oxide, and activation of a soluble guanylate cyclase. However, very recently a Taiwanese group (Shibanaka et al., 1993) has shown in the Oriental fruit fly *Bactrocera dorsalis*, that the EH receptor is a receptor guanylate cyclase in Inka cells, as mentioned in the EH receptor chapter.

**Insect antidiuretic factors: (ADF)**

Our group identified two antidiuretic peptides isolated from *Tenebrio molitor* pupal heads based on their ability to elevate cGMP in Malpighian tubules. One of these peptides, Tenmo-ADFa, is exceptionally potent (EC\(_{50}\) ~ 10 fM) (Eigenheer et al., 2002), and is also a potent antidiuretic in *Aedes aegypti* (Massaro et al., 2004), whereas Tenmo-ADFb is much less potent in *Tenebrio*, with an EC\(_{50}\) of 240 pM (Eigenheer et al., 2003).
Eigenheer showed that nitric oxide donors used to stimulate soluble guanylate cyclases cannot stimulate cGMP levels in *Tenebrio* Malpighian tubules in the presence of NO synthase inhibitors, nor can they block the effects of crude preparations of ADFs (Eigenheer et al., 2002); two NO synthase inhibitors were tested at three different concentrations.

These data suggest that these peptides act *via* an integral membrane guanylate cyclase receptor. A study of the second antidiuretic peptide, Tenmo-ADFb, in the cricket *Acheta domestica* showed it to have diuretic, rather than antidiuretic, properties on Malpighian tubules (Coast et al., 2007). Exogenous cGMP was shown to have a diuretic effect, rather than antidiuretic, also in *Manduca* (Skaer et al., 2002), as well as a diuretic effect in *Drosophila* as mentioned above. Thus, the effect of cGMP in Malpighian tubules differs between different orders.

Because both diuretic and antidiuretic hormones are known from *T. molitor*, it was possible to study their interaction in a conspecific assay. Tenmo-DHs act *via* cAMP to increase MT secretion, whereas Tenmo-ADFs act through cGMP to reduce it. The stimulatory effect of Tenmo-DH37 first applied, is reversed by Tenmo-ADFa subsequently added (Wiehart et al., 2002). The likely mode of action is that a cAMP-specific phosphodiesterase is activated by cGMP (Quinlan and O’Donnell, 1998), leading to the rapid degradation of cAMP that mediates diuretic hormone activity.

**Ion Transport Peptide: ITP**

Neuropeptides and endocrine peptides are circulating signaling molecules involved in many physiological, behavioral, and developmental events. Ion transport peptide (ITP) in locusts was identified as having antidiuretic activity in the ileum (Coast
et al., 2002). Synthetic ITP has similar activity to that of native ITP, stimulating Cl\(^-\) transport 10-fold, Na\(^+\) transport ~2-fold, K\(^+\) permeability ~3-fold, and isoosmotic fluid absorption ~4-fold (King et al., 1999). ITP is released from endocrine organs known as the corpora cardiaca; it then stimulates the ileum to transport Cl\(^-\) ion from the lumen to hemolymph, thus forming an electrochemical gradient which drives water resorption.

ITP is a 72 amino-acid peptide that has ~40% sequence identity with crustacean hyperglycemic hormone (CHH), including identical positions of the six cysteine residues and the connectivity of the resulting three disulfide bridges (King et al., 1999). A structure-function study showed that in the six N-terminal residues (SFFDIQ), substitutions of Ala for residues 1, 4, 5 or 6 do not abolish the biological activity (Zhao et al., 2005). However, Phe mutations at residues 2 or 3 are deleterious to biological function (Zhao et al., 2005). At the C terminus (MVEIL), the Leu-amide is essential for biological activity (Zhao et al., 2005). Breaking any disulfide bond causes an activity loss of ~90% (Zhao et al., 2005).

The mode of action of ITP is proposed as shown in Figure 10 by Phillips (Phillips et al., 1998). ITP was reported to start a signal cascade by elevating cAMP. cAMP causes some anion channels to open in the basolateral membrane. Chloride enters these epithelial cells by an active transport mechanism driven by a Cl\(^-\) ATPase. This transport mechanism has been most heavily studied in *Aplysia californica* (Gerencser, 1996). This transport mechanism provides the driving force for passive reabsorption of K\(^+\) via ion channels with different properties in the apical and basolateral membranes. Na\(^+\) reabsorption is less important because of its lower concentration. Protein kinase A, after stimulation by cAMP, activates the apical Cl\(^-\) pump and opens K\(^+\) and Cl\(^-\) channels in the
apical and basolateral membranes. Based on these data, exogenous cAMP should mimic the biological functions of ITP. Exogenous cAMP achieves most of the actions attributed to Schgr-ITP, but it has the opposite effect on proton secretion of Schgr-ITP (Coast et al., 2002). Electrophysiological studies also showed that there is no change in ileal basolateral conductance after cAMP addition to the ileum. Therefore, cAMP may not be the second messenger for Schgr-ITP; another second messenger is anticipated for the inhibition of acid secretion.
Figure 10. Proposed model for control of ion transport across desert locust ileum. Ion transport peptide (ITP) is postulated to act via cAMP at the apical membrane to increase both $K^+$ and $Na^+$ conductances and to stimulate the $Cl^-$ pump directly. ITP must act via another second messenger to inhibit apical acid secretion.
**Insect diuretic peptide: CAP\textsubscript{2b}**

Another insect peptide which acts via cGMP is CAP\textsubscript{2b}. This peptide was isolated based on its ability to stimulate the heartbeat rate of *M. sexta* (Cheung et al., 1992). CAP\textsubscript{2b} is a diuretic in *Drosophila melanogaster*, in which species it elevates cGMP in the Malpighian tubule through the intermediacy of nitric oxide (Davies et al., 1997). The peptide was proposed to act via phospholipase C in principal cells, producing IP\textsubscript{3}, and causing the release of Ca\textsuperscript{+2} from internal stores that stimulate entry of external Ca\textsuperscript{+2}. This in turn activates a cascade consisting of Ca\textsuperscript{+2}-calmodulin-dependent nitric oxide synthase and, subsequently, guanylate cyclase and the V-ATPase. A G-protein coupled receptor has been cloned from *D. melanogaster* that, when expressed in frog oocytes, responds with low EC\textsubscript{50} values to CAP\textsubscript{2b} (Park et al., 2002). This peptide is somewhat homologous to the vertebrate peptide neuromedin U, and its receptor is homologous to the vertebrate neuromedin U receptor. The neuromedin U receptors are two G-protein coupled receptors which bind the neuropeptide hormones neuromedin U and neuromedin S. There are two subtypes of neuromedin U receptor, each encoded by a separate gene (*NMUR1*, *NMUR2*). Neuromedin U is an agonist at both the NMU1 and NMU2 subtypes, while neuromedin S is selective for NMU2, and is a more potent agonist at this subtype than neuromedin U (Mitchell et al., 2009). CAP\textsubscript{2b} gives rise to an elevation of intracellular Ca\textsuperscript{+2} through plasma membrane channels, which in turn stimulates a soluble NO synthase (Rosay et al., 1997). This short lived, reactive gas binds to a heme group in a soluble GC, giving rise to cGMP which increases diuresis by the tubules of this species. Interestingly, Quinlan *et al.* (1997) showed CAP\textsubscript{2b}, and cGMP, to have an antidiuretic effect in *Rhodnius prolixus*. In this species the actions of cAMP and cGMP are antagonistic; it is
likely that cGMP stimulates a cAMP-selective phosphodiesterase (Quinlan and O'Donnell, 1998).

**Crustacean hyperglycemic hormone-CHH**

Studies on crustaceans have contributed substantially to concepts on neurosecretion and neuroendocrine regulation. Abramowitz et al. (1944) first reported a hyperglycemic response to injection of crude eyestalk extract in crustaceans (Kleinholz and Keller, 1973). The hormone responsible for this hyperglycemic effect was characterized as a polypeptide with a molecular weight of about 8-9000 Da belonging to a family of neuropeptides (crustacean hyperglycemic hormone (CHH), molt inhibiting hormone (MIH), and vitellogenesis inhibiting hormone (VIH)). There are more than 40 genes reported in different species, for example shrimp, crayfish, lobster, etc. (Fanjul-Moles, 2006). CHH is an important hormone both for development and during the life cycle of the animal. Different CHH isoforms have multifunctional roles in decapods; they play a central role in carbohydrate metabolism, but one (molt inhibiting hormone, MIH) also exerts an effect on the molt and reproduction. CHH has been reported to have osmoregulatory functions in several species of crustaceans (Chung and Webster, 2006; Serrano et al., 2003; Spanings-Pierrot et al., 2000).

Crustacean hyperglycemic hormone (CHH), depending on species, has 72 to 73 amino acids, and shares 40% similarity with ITP in sequence (Zitnan and Adams, 2012). The invariant peptide backbone always consists of six cysteines, two arginines, one aspartic acid, and a phenylalanine residue in identical positions. The six cysteines form three disulfide bridges, and the bond connectivity between individual Cys residues are
identical in all the peptides (Mettulio et al., 2004). This connectivity is the same in ITP (King et al., 1999).

In the lobster, the crustacean hyperglycemic hormone (CHH) has been shown to act on target tissues to elevate cGMP level via a rGCase (Goy, 1990). More recently, Lee et al. (2007) have shown CHH and MIH to act via cGMP in the crab. Nagai et al. (2009) showed CHH also acts via cGMP in the prawn Marsupenaeus japonicus. The insect homologue of CHH is the ion transport peptide (Audsley et al., 1992b), which acts on the ileum of Schistocerca gregaria to stimulate active transport of Cl− ion, causing reabsorption of NaCl and water (Audsley et al., 1992b; Meredith et al., 1996). This process is similar to the homeostatic mechanism of the human kidney. It is not clear whether cGMP or cAMP is the bona fide second messenger of this peptide (Dr. Neil Audsley, personal communication). From the high similarity to CHH, it seems likely that cGMP may be the second messenger.
MATERIALS AND METHODS

Part I. Bioinformatic analyses

The insect guanylate cyclases should be very different from those in mammals. Figure 11 shows a protein sequence alignment of receptors from *Rattus norvegicus* (AAI28743), *Bactrocera dorsalis* (ACQ90240), *Apis mellifera* (XP_001121839), *Aedes aegypti* (XP_00165228), and *Tribolium castaneum* (XP_970405). The sequence of rat GC-A was downloaded from NCBI, the other sequences were located by BLAST analysis of genome data in NCBI. Based on Clustal W alignments, the guanylate cyclase domain is the most conserved domain between different species. This domain is essential for guanylate cyclase activity, and is highly similar to the soluble guanylate cyclases. In this domain, nearly all species share three identical partial sequences: IYFSDIVGFT, DVYKVETIGDAYMVVSGLP, and MPRYCLFGDTVNTASRMES. We used the conserved sequence MPRYCLFGDTVNTASRMES to search for possible guanylate cyclase receptors in the *Bombyx* genome (http://silkworm.genomics.org.cn/), *Tribolium* genome (http://www.hgsc.bcm.tmc.edu/), and *Drosophila* genome (http://www.ncbi.nlm.nih.gov/). Because soluble guanylate cyclases also have this conserved sequence, the BLAST results include some soluble guanylate cyclase (sGCs) and all receptor guanylate cyclases (rGCs). The receptor guanylate cyclases are at least 1000 amino acids in length, which is much larger than soluble guanylate cyclases. We classified the receptor guanylate cyclases in the BLAST search by the protein size. In Table 2 (Page 43), we list all possible receptor guanylate cyclases in *Bombyx, Tribolium*, and *Drosophila*. We found five different guanylate cyclases in the insect genomes; we call these in Manduca NPR1, NPR5, NPR6, NPR7 and NPR8.
Part II. Searching for the conserved guanylate cyclase domain:

RNA preparation

Because receptor guanylate cyclases are rare transcripts in organisms, we needed to prepare RNA without its degradation. Insect tissues are similar to plant tissues. Many labs have good success with the QIAGEN RNeasy plant mini kit (MacKenzie et al., 1997). This kit gives nearly pure RNA without DNase digestion. A medium concentration (around 3M) of guanidine hydrochloride is used to extract RNA without significant genomic DNA contamination (Wu et al., 2002). However, RNase can still be active in a medium concentration of guanidine hydrochloride. Second, our PCR detection technology is very sensitive; even a little genomic DNA contamination will degrade the results. With this kit, genomic contamination impedes our detection of rare transcripts by PCR. Third, we needed to prepare RNA from the fat body for qPCR studies. The high lipid content in this tissue usually affects RNA binding. I chose stronger denaturing reagents to inactivate RNase, and then performed the DNase digestion in tubes. TRIzol (Invitrogen) is a good choice; it can also dissolve the lipid in the organic phase. The polysaccharides and proteoglycans in insect tissue co-precipitate with RNA (Chomczynski and Mackey, 1995). The RNA pellet is very difficult to dissolve. Here I used a salt precipitation method (step 8) to pellet the RNA without other contamination. TRIzol composition is shown in Table 3. The protocol is given in the following numbered steps.
Table 2. Orthologous families of insect rGCase. For *Manduca sexta*, we show the size of the protein cloned.

<table>
<thead>
<tr>
<th>Manduca sexta</th>
<th>Bombyx BGIBMGA</th>
<th>Tribolium NCBI</th>
<th>Drosophila NCBI</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPR1</td>
<td>1102AA</td>
<td>006891-TA</td>
<td>XP_971364</td>
</tr>
<tr>
<td>NPR2</td>
<td>soluble</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR3</td>
<td>soluble</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR5</td>
<td>1092AA</td>
<td>006799-TA</td>
<td>XP_970405.1</td>
</tr>
<tr>
<td>NPR6</td>
<td>500AA</td>
<td>003981-TA&amp;3982-TA</td>
<td>XP_975601.2</td>
</tr>
<tr>
<td>NPR7</td>
<td>1022AA</td>
<td>None found</td>
<td>XP_001809661.1</td>
</tr>
<tr>
<td>NPR8</td>
<td>1197AA</td>
<td>SP Q9BPR0</td>
<td>NP_001167546.1</td>
</tr>
</tbody>
</table>

*Table 2. Orthologous families of insect rGCase. For *Manduca sexta*, we show the size of the protein cloned.*
Table 3. TRIzol composition. The weights given are based on 1L of solution.

<table>
<thead>
<tr>
<th>Composition</th>
<th>Final concentration</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guanidinium thiocyanate</td>
<td>0.8 M</td>
<td>118.16 g</td>
</tr>
<tr>
<td>Ammonium thiocyanate</td>
<td>0.4 M</td>
<td>76.12 g</td>
</tr>
<tr>
<td>Sodium acetate (pH 5)</td>
<td>0.1 M</td>
<td>33.4 ml of 3 M stock</td>
</tr>
<tr>
<td>Adjust pH to 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycerol</td>
<td>5%</td>
<td>50 ml</td>
</tr>
<tr>
<td>Phenol</td>
<td>38%</td>
<td>380 ml</td>
</tr>
</tbody>
</table>
1. After homogenization, incubate the homogenized sample for 5 minutes at room temperature to permit complete dissociation of the nucleoprotein complex.

2. Add 0.2 mL of chloroform per 1 mL of TRIzol

3. Shake tube vigorously by hand for 15 seconds.

4. Incubate for 3 minutes at room temperature.

5. Centrifuge the sample at 12,000 x g for 15 minutes at 4°C. Note: The mixture separates into a lower red phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. RNA remains exclusively in the aqueous phase. The upper aqueous phase is ~50% of the total volume.

6. Remove the aqueous phase of the sample by angling the tube and pipetting the solution out. Avoid drawing any of the interphase or organic layer into the pipette when removing the aqueous phase.

7. Transfer the aqueous phase to a fresh tube. Precipitate the RNA from the aqueous phase by mixing with isopropyl alcohol and high salt (0.8 M sodium citrate and 1.2 M NaCl). The following modification of the RNA precipitation (step 3) removes these contaminating compounds from the isolated RNA. Add to the aqueous phase 0.25 ml of isopropanol followed by 0.25 ml of a high salt precipitation solution per 1 ml of TRIzol reagent used for the homogenization. Mix the resulting solution, centrifuge and proceed with the isolation as described in the protocol. The modified precipitation effectively precipitates RNA while maintaining polysaccharides and proteoglycans in a soluble form.

8. Incubate samples at room temperature for 10 minutes and centrifuge at no more than 12,000 x g for 10 minutes at 2 to 8°C. The RNA precipitate, often invisible before centrifugation, forms a gel-like pellet on the side and bottom of the tube.
9.  Remove the supernatant. Wash the RNA pellet once with 75% ethanol, adding at least 1 ml of 75% ethanol per 1 ml of TRIzol Reagent used for the initial homogenization. Mix the sample by vortexing and centrifuge at no more than 7,500 x g for 5 minutes at 2 to 8 °C.

10.  Dissolve the pellet in RNase free water after drying. Measure the RNA concentration by measuring OD260.

11.  Promega RNase-free DNase is used to digest the genomic DNA.

- RNA in water 8 µl
- RNase-Free DNase 10X reaction buffer 1 µl
- RNase-Free DNase 1 µg/µg RNA

Nuclease-free water to a final volume of 10 µl

12.  TRIzol is used again to decontaminate DNase reaction (steps 1-11).

**Reverse transcription**

Because guanylate cyclase receptors are of low abundance in insects, general reverse transcriptases do not work efficiently with these rare transcripts. I tried different reverse transcriptases: NEB MMLV reverse transcriptase, Roche reverse transcriptase AMV, Roche Transcriptor, Clontech SMART reverse transcriptase, and Clontech SMARTscribe reverse transcriptase (data are not shown). Based on my data, of all these reverse transcriptases, Clontech SMARTscribe, which can amplify rare transcripts, is the most efficient for the receptor guanylate cyclases. We used SMARTScribe Reverse Transcriptase (Clontech) to perform reverse transcription as shown below:

1.  Add 1 µl oligo dT (12 µM) to RNA (2 µg). Add the RNase-free water to a final volume of 5 µl.
2. Heat the mixture to 72°C for 3 min to destroy secondary structure.

3. After cooling down to room temperature for 1 min, add: 2 µl 5 X first strand, 1 µl dNTP Mix, 1 µl 20 mM DTT, 1 µl SMARTscribe reverse transcriptase (100 units/µl).

4. Incubate at 42°C for 90 mins.

5. 75°C for 15 mins to deactivate reverse transcriptase.

**Degenerate PCR**

Based on the conserved sequence in the guanylate cyclase domain, we designed degenerate primers to target our guanylate cyclase receptors. Both receptor guanylate cyclases and soluble guanylate cyclases have the same conserved sequence in the guanylate cyclase domain, resulting in guanylate cyclase activity. If we used the guanylate cyclase domain to design degenerate primers, we would also clone soluble guanylate cyclase sequences. Receptor guanylate cyclases can be characterized by gene size in 3' RACE PCR. Soluble guanylate cyclase genes have shorter 3' ends (less than 500 bp) than receptor guanylate cyclases. We also did BLAST analyses of our sequences in the silkworm database. *Manduca* and the silkworm usually have high similarity in the protein sequence. This helps greatly to differentiate guanylate cyclase receptors.

Generally, degenerate primers have much lower efficiency than specific primers. A degenerate primer with lower degeneracy is better than a degenerate primer with high degeneracy. We use the DVYKVETI partial sequence as an example to explain this statement. We count the degeneracy as in Table 4. The DVYKVETI degenerate primer (GAYGTNTAYAARGTNGARACNAT) has 2 X 4 X 2 X 2 X 4 X 2 X 4=1024 degeneracy. That means this degenerate primer is a mix of 1024 different primers. If DVYKVETI degenerate primers (shown below) are used to target the template
GACGTATAAGGTGAGAGCAT (the underlined bases are degenerate sites), only a few primers can successfully target the template in these 1024 primers. Second, thousands of more abundant genes interfere with the PCR efficiency in the cDNA pool. Degenerate primers are very difficult for targeting our genes in the cDNA pool.

The first degenerate primers used were designed based on the conserved sequence DVYKVETI and MPRYCLFGD because of their lower degeneracy (Figure 12). However, we did not get good results based on these primers.

DVYKVETI degenerate sense primer

5' GAYGTNTAYAARGTNGARACNAT 3'

MPRYCLFGD antisense degenerate primer

5' RTCNCCRAANARRCARTANCKNGGCAT 3'
<table>
<thead>
<tr>
<th>Character</th>
<th>Mixed base</th>
<th>Degeneracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>A,G</td>
<td>2</td>
</tr>
<tr>
<td>Y</td>
<td>C,T</td>
<td>2</td>
</tr>
<tr>
<td>M</td>
<td>A,C</td>
<td>2</td>
</tr>
<tr>
<td>K</td>
<td>G,T</td>
<td>2</td>
</tr>
<tr>
<td>S</td>
<td>C,G</td>
<td>2</td>
</tr>
<tr>
<td>W</td>
<td>A,T</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>A,C,T</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>C,G,T</td>
<td>3</td>
</tr>
<tr>
<td>V</td>
<td>A,C,G</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>A,G,T</td>
<td>3</td>
</tr>
<tr>
<td>N</td>
<td>A,T,C,G</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 4. Standard mixed base abbreviations with their degeneracy.
Figure 12. Protein alignment of guanylate cyclases from different species. The red arrows (degenerate primers) target the conserved domains DVYKVETI, and MPRYCLFGD.
**Nested PCR using degenerate primers**

The efficiency of degenerate primers is very low, because of sequence degeneracy and cDNA background interference. Nested PCR helps to reduce the cDNA background interference (Moser et al., 2012). We designed two sets of degenerate primers (Figure 13). Set 1 is designed based on the partial sequences DIVGFT and YCLFGD. Set 2 is designed based on DVYKVETI and MPRYCLFGD. At the same time we improved the degenerate primers by using inosine to substitute N for (A, T, G, C) in the degenerate primers.

Set 1: DIVGFT-sense (12 µM): 5' GAYATHGTIGGITYAC 3' (degeneracy - 12 fold)
Set 1: YCLFGD-antisense (12 µM): 5' GTRTCICRAAIARRCARTA 3' (degeneracy – 32 fold)
Set 2: DVYKVETI-sense (12 µM): 5' GAYGTITAYAARGTIGARACIAT 3' (degeneracy – 16 fold)
Set 2: MPRYCLFGD-antisense (12 µM): 5' CCRAAIARRCARTAICKIGGCAT 3' (degeneracy-32 fold)

PCR is performed using GoTaq® Hot Start Colorless Master Mix. The MJ Research Thermal Cycler is used to set up a touch-down PCR program (95°C 2 mins, 35 cycles of [95°C 30 s, 55°C-45°C 30 s, 72°C 1 min] then 72°C 10 mins)

For 1\textsuperscript{st} round PCR, 20 µl of reaction volume included:
2X Hot Start Master Mix, 10 µl; Set 1: DIVGFT-sense, 1 µl; Set 1: YCLFGD-anti, 1 µl; cDNA (by reverse transcription), 1 µl; water, 7 µl.

For 2\textsuperscript{nd} round PCR, 20 µl of reaction volume included:
2X Hot Start Master Mix, 10 µl; Set 2: DVYKVETI-sense, 1 µl; Set 2: MPRYCLFGD-anti, 1 µl; 1st round PCR as template (diluted with 180 µl water), 1 µl; water, 7 µl.

The PCR result is shown below. The PCR product is ligated into pGEM®-T Easy Vector. This PCR product contained 7 different guanylate cyclase fragments. We had to pick out many colonies to include all the different sequences. A white/blue screen was used to pick out 50 positive colonies. Plasmids were prepared using a QIAGEN mini prep and sequenced by the Nevada Genomics Center. The sequences are shown in the appendix.

**Locate sequences using degenerate PCR**

The fragments we got are in the guanylate cyclase domain, which is located around 2.5 kb from the 5’ end (Figure 14A). RACE PCR experiments can be conducted to study the unknown parts of the 5’ end (2.5 kb) and of the 3’ end (0.5 kb). The unknown sequence of the 5’ end is too large for 5’ RACE PCR. Generally, RACE PCR efficiency is even lower than for degenerate PCR. First, one primer in a RACE PCR is a universal primer, which can target all cDNAs. Second, MMLV reverse transcriptase may not transcribe the full length of the gene because of secondary structure in the sequence. I tried hundreds of times to perform 5’ RACE PCR directly without any successes.

Some receptor guanylate cyclases exist in both the *Tribolium* genomic data and the silkworm genomic data. *Manduca* NPR1 corresponds to 006891-TA in the silkworm database and XP_971364 in the *Tribolium* genomic data (Figure 14B). We performed an alignment of these two sequences. We believed that the sequence conserved between *Tribolium* and *Bombyx* should also be conserved in *Manduca*. A new degenerate primer was designed based on the conserved sequence AFFGPEG in the ECD domain. PCR and sequencing conditions were the same as in the last step. The 5 base pairs from the 3’ end
(underlined sequences) are very critical for PCR efficiency. Usually one mismatch in the 3' end base pairs can terminate the PCR. Two primers were used to make sure of an exact match in the 3' end. We got a 2.4 kb fragment as shown below. This is a great improvement for the 5' end unknown sequence (Figure 14B).

NPR1: AFFGPEG-sense-A/G (two primers)

5' GCN TTY TTY GGN CCI GA A/G GG 3'

NPR1: GCD-antisense

5' TTAATGTTTCATTAGGTC 3'

A similar method was employed for NPR5, NPR7, and NPR8. NPR6 transcripts are of too low abundance for degenerate PCR. The sequence of primers is listed below (Table 5).
Figure 13. Degenerate primers designed for nested PCR. DIVGFT and YCLFGD degenerate primers (blue) are used to perform the first round PCR. DVYKVETI and MPRYCLFGD degenerate primers (red) are used to do the second round PCR.
<table>
<thead>
<tr>
<th>Primers</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPR5: AETYG-sense 'T/C</td>
<td>2.7 kb</td>
</tr>
<tr>
<td>5' GCI GAR ACI TAT/C GG</td>
<td></td>
</tr>
<tr>
<td>NPR5: GCD-antisense</td>
<td></td>
</tr>
<tr>
<td>5' CAGCAGCAGGC 3'</td>
<td></td>
</tr>
<tr>
<td>NPR7: NGDAEG-A/G</td>
<td>1.8 kb</td>
</tr>
<tr>
<td>AAY GGI GAY GCI GAA/G GG</td>
<td></td>
</tr>
<tr>
<td>NPR7: GCD-antisense</td>
<td></td>
</tr>
<tr>
<td>5' GGAGTCGTGAAGCA 3'</td>
<td></td>
</tr>
<tr>
<td>NPR8: QDILEN-sense-A/G</td>
<td>1.2 kb</td>
</tr>
<tr>
<td>5' CAR GAY ATY TTI GAA/G AA 3'</td>
<td></td>
</tr>
<tr>
<td>NPR8: GCD-antisense</td>
<td></td>
</tr>
<tr>
<td>5' CTTCAACTGATCCA 3'</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Primers used to detect the 5' end of the receptor guanylate cyclases. The primers were used to detect the 5' end of the receptor guanylate cyclases NPR5, NPR7, and NPR8. Anti-sense primers are specific primers, which are designed based on the known guanylate cyclase domain sequence. The sense primers are degenerate primers, which are designed based on a conserved region in the 5' end.
Figure 14A. The general structure of the receptor guanylate cyclases. The red fragment is the guanylate cyclase domain, whose sequence is already known from the step “degenerate PCR”. The unknown regions are 0.5 kb fragments in the 3’ end and 2.5 kb fragments in the 5’ end.

Figure 14B. The conserved partial sequence AFFGPEG is located in the ECD. The specific primer NPR1-GCD-antisense is designed based on the known region. NPR1-ADDGPEG-sense is the degenerate primer. The final product is a 2.4 kb PCR fragment.
Part III. Different methods for RACE-PCR to complete the end sequences

3′ RACE PCR

Rapid amplification of cDNA ends (RACE) is a variant of RT-PCR that amplifies unknown cDNA sequences corresponding to the 3′- or 5′-end of the RNA. Numerous variations of the original protocols have been published (Delarue et al., 1996; Hodin and Wistow, 1993; Jain et al., 1992; Shen, 1999). Two general RACE strategies exist: one amplifies 5′ cDNA ends (5′ RACE) and the other captures 3′ cDNA end sequences (3′ RACE). In either strategy, the first step involves the conversion of RNA to single-stranded cDNA using a reverse transcriptase. For subsequent amplification, two PCR primers are designed to flank the unknown sequence. One PCR primer is complementary to known sequences within the gene, and a second primer is complementary to an “anchor” site (anchor primer). The anchor site may be present naturally, such as the poly (A) tail of most mRNAs, or can be added *in vitro* after completion of the reverse transcription step. The anchor primer also can carry adaptor sequences, such as restriction enzyme recognition sites, to facilitate cloning of the amplified product. Amplification using these two PCR primers results in a product that spans the unknown 5′ or 3′ cDNA sequence, and sequencing this product will reveal the unknown sequence. The information obtained from partial cDNA sequences then can be used to assemble the full-length cDNA sequence (Frohman et al., 1988).

The 3′-RACE procedure uses a modified oligo (dT) adaptor (3′-RACE primer) as the reverse transcription primer (Figure 15). This oligo(dT) primer/adaptor is comprised of an oligo(dT) sequence, which anneals to the poly(A) tail of the mRNA, and an adaptor sequence at the 5′ end (5′ CAAGCAGTGGTATCAACGCAGAGTAC 3′). A single G, C
or A residue at the 3’ end ensures that cDNA synthesis is initiated only when the primer/adaptor anneals immediately adjacent to the junction between the poly (A) tail and 3’ end of the mRNA. This adaptor sequence is used as the anchor primer in the subsequent amplifications along with a primer complementary to known sequences within the gene. Because the guanylate cyclase receptors are low abundance genes, usually a nested PCR is necessary to enhance the sensitivity.

3’-RACE primer (3’-CDS; 12 µM):

5’ CAAGCAGTGGTATCAACGCAGAGTAC(T)_{30}VNC 3’

(N = A, C, G, or T; V = A, G, or C)

3’ RACE reverse transcription is performed using the 3’ RACE CDS primer (the same protocol as in reverse transcription). The adaptor sequence (5’ CAAGCAGTGGTATCAACGCAGAGTAC 3’) is used to design the following universal primers to flank the unknown 3’ end region:

Universal primer 1: 5’ CTAA TACGACTCACTA TAGGGCAA 3’

Nested universal primer 2: 5’ AAGCAGTGGTATCAACGCAGAGT 3’

NPR1: GCD-specific primer 1

NPR1: GCD-specific primer 2

PCR was performed using GoTaq® Hot Start Colorless Master Mix. The MJ Research thermal cycler was used for the touch-down PCR program (95°C 2 mins, 35 cycles of [95°C 30s, 60°C-55°C 30 s, 72°C 1 min], then 72°C 10 mins)

For 1st round PCR with a 20 µl reaction volume:

2X Hot Start Master Mix, 10 µl; universal primer 1, 0.5 µl; NPR1: GCD-specific primer 1, 0.5 µl; 3’ RACE cDNA (by reverse transcription), 1 µl; water, 7 µl.
For 2nd round PCR in a 20 µl reaction volume: 2X Hot Start Master Mix, 10 µl; nested universal primer 2, 1 µl; NPR1: GCD-specific primer 2, 1 µl; 1st round PCR (diluted with 180 µl water), 1 µl; water, 7 µl.

The other guanylate cyclase 3’ end is analyzed with a similar method. The primers used are listed as below (Table 6).
<table>
<thead>
<tr>
<th>Sense primer</th>
<th>Anti-sense primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPR1: 3r-3014</td>
<td>universal primer 1</td>
</tr>
<tr>
<td>5' AGTAGTGGCTGGTGTTGGGACTTAC 3'</td>
<td>nested universal primer 2</td>
</tr>
<tr>
<td>NPR1: 3r-3042</td>
<td></td>
</tr>
<tr>
<td>5' ATGCCTCGCTATTTTGTGGGAC 3'</td>
<td></td>
</tr>
<tr>
<td>NPR7: 3r-3020</td>
<td>universal primer 1</td>
</tr>
<tr>
<td>5' GCAACGACACACATACCTGCTTTGAAGA 3'</td>
<td>nested universal primer 2</td>
</tr>
<tr>
<td>NPR7: 3r-3114</td>
<td></td>
</tr>
<tr>
<td>5' ATGGTGTTAGTTGGAGAGGA TACGA 3'</td>
<td></td>
</tr>
<tr>
<td>NPR8: 3r-3408</td>
<td>universal primer 1</td>
</tr>
<tr>
<td>5' GGAGAAAAAGTAGACCCCAACGCTCAGA 3'</td>
<td>nested universal primer 2</td>
</tr>
<tr>
<td>NPR8: 3r-3346</td>
<td></td>
</tr>
<tr>
<td>5' TGGAGTGTAGGGGAGAGGTCGCTATGAAG 3'</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. The primers used to analyze the 3' end of other guanylate cyclase receptors NPR5, NPR7, and NPR8. All the anti-sense primers are the same: universal primer 1: 5' CTAATACGACTCTAGTTAGGGCAA 3'; nested universal primer 2: 5'AAGCAGTGGTATCAACGCAGAGT 3'. The sense primers are specific primers based on the guanylate cyclase receptor sequences.
Figure 15. The mechanism of 3' RACE. In 3' RACE PCR reverse transcription, the anchor primer is based on the mRNA poly (A) tail as the priming site. The sense primer is located in a known region and the anti primer is the anchor primer. PCR can amplify an unknown region between a known region and the poly (A) tail.
5' RACE PCR

In 5' RACE (Figure 16), the first-strand cDNA synthesis reaction is primed using an oligonucleotide complementary to a known sequence in the gene. After removing the RNA template, an anchor site oligo(C) at the 3'-end of the single-stranded cDNA is created using terminal deoxynucleotidyl transferase (TdT), which adds a nucleotide tail. A typical amplification reaction follows using an anchor primer complementary to the newly added tail and another primer complementary to a known sequence within the gene.

The 5' RACE system is intended for synthesis of first strand cDNA, purification of first strand products, homopolymeric tailing, and preparation of target cDNA for subsequent amplification.

First strand cDNA is synthesized from total or poly (A) RNA using a gene-specific primer (GSP1) using 1 µl SMARTscribe reverse transcriptase. After first strand cDNA synthesis, the original mRNA template is removed by treatment with the RNase mix (mixture of RNase H, which is specific for RNA: DNA heteroduplex molecules and RNase T1). Unincorporated dNTPs, GSP1, and proteins are separated from cDNA using a silica column. A homopolymeric tail is then added to the 3'-end of the cDNA using TdT and dCTP or dATP.

Since the tailing reaction is performed in a PCR-incompatible buffer (containing CoCl₂), the entire contents of the reaction may not be directly amplified by PCR without intermediate organic extractions, ethanol precipitations, or dilutions. We made a PCR buffer (GC buffer), which reduces the negative impact of TdT buffer on PCR. PCR amplification is accomplished using Roche FastStart Taq DNA polymerase, a nested, gene-specific primer (GSP2, designed by the user) that anneals to a site located within the
cDNA molecule, and a novel deoxyinosine-containing anchor primer provided with the system. If the gene is not successfully detected, we can carefully perform a second round PCR using another nested gene-specific primer (GSP3) and the universal amplification primer (UAP), which anneals to part of the anchor primer. Following amplification, 5′ RACE products can be cloned into a pGEM-T vector for subsequent characterization procedures, which may include sequencing.
Figure 16. Procedures for 5' RACE PCR. 5' RACE PCR uses the terminal transferase to add oligo(C) adaptor to the end after specific primer dependent reverse transcription. The detection primers in the following PCR are oligo(G) and another gene specific primer.
Gene specific cDNA synthesis primer: Gene specific primer (GSP) (Shown in Table 7)

1. Add 1 µl GSP (12 µM) into total RNA (2 µg). Add RNase-free water to a final volume of 5 µl.
2. Heat the mixture to 72°C for 3 min to destroy the secondary structure.
3. After cooling down to room temperature for 1 min, add: 2 µl 5X buffer, 1 µl dNTP Mix, 1 µl 20 mM DTT, 1 µl SMARTscribe reverse transcriptase (100 units/µl).
4. Incubate at 42°C for 90 mins.
5. 75°C for 15 mins to deactivate reverse transcriptase.
6. Add 0.5 µl of RNase H (New England Biolab) and 0.5 µl RNase T1 (Epicentre), mix gently but thoroughly, and incubate for 60 min at 37°C.
7. Collect the reaction by brief centrifugation and place on ice. The procedure may be stopped at this point and the reactions stored at -20°C.

Column purification of cDNA is done using a QIAGEN PCR purification kit using the factory manual. Then the cDNA is eluted with 45 µl water.

**Tailing procedure**

We used terminal transferase to add an oligo (C) tail or oligo (A) tail to the 3’ end of the gene specific cDNA. Each tail has different advantages and disadvantages. In the oligo (A) tailing method, the poly T adaptor’s annealing temperature is very low (around 40°C for a 20 dT oligo). For the oligo (C) tailing method, the oligo (G) adaptor can anneal with the tail at a higher annealing temperature in PCR. The sensitivity and specificity is better than the oligo A tail method. However, the oligo (G) adaptor very
easily causes false priming in PCR. To solve this issue, we substitute some dGs in the poly G adaptor with inosine, which has a higher preference for dC than other nucleotides (details are given in the following PCR protocol).

1. Add the following components to each tube and mix gently:
   10X terminal transferase reaction buffer (NEB), 5 µl; 2.5 mM CoCl\(_2\), 5 µl; purified gene specific cDNA, 38 µl; 10 mM dCTP or dATP, 1 µl.
2. Incubate for 2 to 3 min at 94°C. Chill 1 min on ice. Collect the contents of the tube by brief centrifugation and place on ice.
3. Add 1 TdT, mix gently, and incubate for 30 min at 37°C.

**PCR of dC-tailed cDNA**

Normally, homopolymer primers have melting temperatures that are either higher [poly (dG)•poly (dC)] or lower [poly (dA)•poly (dT)] than a typical GSP. They also can have poor specificity that can lead to mispriming at internal sequences. To minimize these problems, our anchor primers were designed with the selective placement of deoxyinosine residues in the poly (dG) portion. This design eliminates the need to use the mixtures of anchor and adapter primers described in the original method.

Deoxyinosine has the capacity to base-pair with all four bases; however, it does so with varying affinities. The order of stabilities for the different combinations, from greatest to least stable, reported by Martin *et al.* (2006) are as follows: I:C > I:A > I:T > I:G. I:C pairs were found to be slightly less stable than A:T pairs. The selective placement of deoxyinosine residues in the 3’ region of the anchor primer maintains low stability on the primer 3’-end (\(\Delta G = -8.2\) kCal/mol) and creates a melting temperature (\(T_{m}\)) for the 16-base anchor region (66.6°C) which is comparable to that of a typical 20-mer primer with
50% GC content. This maximizes specific priming from the oligo-dC tail and minimizes priming at internal C-rich regions of the cDNA.

The Universal Amplification Primer (UAP) is used to reamplify primary 5' RACE PCR products in applications such as nested PCR or enrichment of RACE products for cloning.

5' RACE abridged anchor primer (AAP): 5' GGC CAC GCG TCG ACT AGT ACG GGI IGG GII GGG IIG3'

5' RACE universal amplification primer (UAP): 5'GGC CAC GCG TCG ACT AGT ACG3'

NPR1 Gene specific primer 1 (GSP1): in Table 7

NPR1 Gene specific primer 2 (GSP2): in Table 7

Usually 5' RACE PCR can only amplify fragments smaller than 500 bp. Multiple 5' RACE PCRs had to be performed for NPR1 in order to sequence the 5' end of NPR1. Below are listed each gene specific anti-sense primer for each 5' RACE PCR. The sequences are in the appendix.
| NPR1   | npr1 5r-GSP 5' AGGGCCGATCCATTGTCAGCTCTTCT 3' |
|        | npr1 5r-GSP1 5' GCCAGAAATAGTGAGACCCTGACGGTC 3' |
|        | npr1 5r-GSP2 5' AGGTACCCAATAGTGTTGTACTTC 3' |
| NPR7   | npr7 5r-GSP 5' TCTGTGTCATCGCCCTTAGAGCCAA 3' |
|        | npr7 5r-GSP1 5' ACCCCACGGTGATTATCTTTCTTTTACTG 3' |
|        | npr7 5r-GSP2 5' CCGCCACAATCGGCATTATAAGGTCAATA 3' |
| NPR8   | npr8 5r-GSP 5' GCGAGTCGTGTCATTACGTCGCCGATACG 3' |
|        | npr8 5r-GSP1 5' CGGCTCCGTATTGTCATTCTTTTCATCA 3' |
|        | npr8 5r-GSP2 5' AGAGCCCTGGGAAGTGGAGAAATCGAA 3' |
| NPR5   | npr5 5r-GSP2 5' ATGCAGACACGTTGGCTGCCACAA 3' |
|        | npr5 5r-GSP1 5' CACACGTCTCCTGGTGATGAA 3' |
|        | npr5 5r-GSP 5' CAGAAGTAGCTATCATGGCGAGTTGAAA 3' |

Table 7. Primer list for 5' race PCR to detect the 5' region of guanylate cyclase receptors. GSPs were used in the gene specific reverse transcription. GSP1 and 5' RACE bridged anchor primers (AAP) were used for PCR detection. GSP2P and 5' RACE universal amplification primers (UAP) were used for nested PCR.
5' and 3' RACE PCR

5' RACE PCR and 3' RACE PCR are generally very complex methods. If linear mRNA or cDNA can be transformed into the circular form, both the 3' end and the 5' end can be sequenced in one step (Figure 17). mRNA has a 5' end cap motif, which can inhibit ligation. When RNA 5´-polyphosphatase is used to convert 5´-triphosphorylated RNA to 5´-monophosphorylated RNA, the phosphatase buffer is not compatible with ligation conditions. Also, single stranded DNA is difficult to ligate. Recently, the product CircLigase™ II ssDNA ligase from Epicentre makes single strand cDNA ligation possible. CircLigase™ II ssDNA ligase is a thermostable enzyme that catalyzes intramolecular ligation (i.e. circularization) of ssDNA templates having a 5´-phosphate and a 3´-hydroxyl group (Polidoros et al., 2006). In contrast to T4 DNA ligase, which ligates DNA ends that are annealed adjacent to each other on a complementary DNA sequence, CircLigase II ssDNA ligase ligates ends of ssDNA in the absence of a complementary sequence. The enzyme is therefore useful for making circular ssDNA molecules from linear ssDNA. Circular ssDNA molecules can be used as substrates for rolling-circle replication or rolling-circle transcription. However, not all the ssDNA can be ligated efficiently. Based on my studies of the Epicentre research, on the 3' end, ligation efficiency is greatest if the 3' end nucleotide is a T, and T is greater than or equal to A ≧ G ≧ C (a "C" on the 3' end will not ligate with any decent efficiency). At the 5' end, the ligation efficiency is greatest if the 5' end nucleotide is a G, and G > A ≧ T > C. We placed a guanine at the 5' end of the oligo dT reverse transcription primer to enhance the ligation in the following steps. However, we could not determine the 3' end sequence
of the cDNA after transcription. This method is not stable and still in the exploratory stage.

**Reverse transcription**

Here we use another reverse transcriptase, MonsterScript reverse transcriptase from Epicentre. MonsterScript Reverse Transcriptase is a thermostable reverse transcriptase that completely lacks RNase H activity. The enzyme's lack of RNase H activity contributes to its ability to make longer cDNAs and more complete full-length libraries of first-strand cDNA molecules compared to other reverse transcriptases.

**Universal reverse transcription primer (URTP):**

phosphorylated GGTTTTTTTTTTTTTTTTT

1. Add 1 µl URTP to total RNA (1 µg). Add RNase-free water to a final volume of 10 µl.

2. Heat the mixture to 65°C for 1 min to destroy the secondary structure.

3. After cooling down to room temperature for 1 min, then add:

RNase-Free Water, 4 µl; MonsterScript 5X reaction buffer, 4 µl; dNTP, 1 µl; MonsterScript Reverse Transcriptase, 1 µl.

4. Incubate at 42°C for 10 mins, 65°C 40 mins.

5. 90°C for 15 mins to deactivate reverse transcriptase.

6. Add 0.5 µl of RNase H (New England Biolab) and 0.5 µl RNase T1 (Epicentre), mix gently but thoroughly, and incubate for 60 min at 37°C.

The primers may negatively impact the circularization. Column purification of cDNA is done using the QIAGEN PCR purification kit using the factory manual. Then the cDNA is eluted with the 45 µl water.
Circularization

1. Combine the following reaction components:
   Single-stranded DNA, 12 µl; CircLigase II 10X reaction buffer, 2 µl; 50 mM MnCl₂, 1 µl; 5M betaine, 4 µl; CircLigaseII ssDNA Ligase, 1 µl; total reaction volume, 20 µl.

2. Incubate the reaction at 60°C overnight.

PCR detection

Because making large fragments is difficult using PCR, we utilized LongAmp Taq DNA Polymerase (New England Biolab), which can generate 30 kb fragments in end point PCR. We only succeeded with NPR8.
Figure 17. Mechanism of 5′ and 3′ RACE-PCR.

1. Synthesize 1st strand cDNA by reverse transcription from target mRNA using 5′-end-phosphorylated RT primer
2. Degradation of RNA in DNA-RNA hybrids by treatment with RNase H
3. Circularization of single-stranded cDNA with DNA Ligase
4. DNA amplification using nested PCR
Part IV. Cloning methods for long sequences

After all the sequences are known, we planned to clone the receptor guanylate cyclase into our expression vector. The expression vector we use is pcDNA5FRT, which is a powerful expression vector from Invitrogen. The problem is how to amplify the low abundance GC-rich 4kb gene fragments from cDNA, given the error rates in PCR. I tried many different commercial polymerases: LongAmp (NEB), Phusion (NEB), LA Taq (Takara), Gotaq (Promega), Phire (NEB), etc. All of them failed for amplifying the guanylate cyclase receptors. To solve this problem, I amplified smaller fragments first using optimized PCR conditions, and then I bridged the fragments in vitro (Figure 18).

The advantages of this method are:
1. Smaller fragments are easily amplified with the polymerase with optimal conditions.
2. Smaller fragments have lower error rates in PCR.

Bridging the fragments is the most difficult part of this experiment. Generally, we have three different methods to bridge the gene fragments. The first method is dependent on the restriction endonucleases and ligase. However, this method is limited by the location of the restriction endonuclease sites in the gene sequence. The second method is based on PCR (Figure 19). If the two fragments have overlapping regions (20-50 bp), the two fragments can hybridize to another in the annealing step to generate a full length fragment. The full length fragment can be amplified using two primers, one at each end. This method is only efficient for bridging two fragments, not for bridging three fragments.
Figure 18. Experimental design and steps involved in bridging PCR (BPCR). Gene fragments A and B have a junction region C. At specific temperatures, the fragments anneal together to form a template molecule, thus triggering the amplification cascade.
The third method uses recent In-Fusion technology from Clontech (Figure 19). In-Fusion cloning kits are designed for fast, directional cloning of one or more fragments of DNA into any vector. The cornerstone of In-Fusion Cloning technology is Clontech’s proprietary In-Fusion enzyme, which fuses DNA fragments, e.g., PCR-generated sequences and linearized vectors, efficiently and precisely, by recognizing a 15 bp overlap at their ends. This 15 bp overlap can be engineered by designing primers for amplification of the desired sequences. In-Fusion HD Kits offer increased cloning efficiency, especially for longer fragments, short oligonucleotides, and multiple fragments. This method is good for bridging three fragments together. We designed an 8-15 bp overlap (overlapping the neighboring gene fragment) on each primer end. The three PCR amplified fragments now have 8-15 bp end sequences which overlap each other. The In-Fusion enzyme can bridge the gene together.
Figure 19. The mechanism of In-Fusion technology from Clontech. The engineered gene fragments have 15 bp overlapping sequences on each end. The plasmid must be linearized using a restriction endonuclease to reduce the negative background. The In-Fusion enzyme recognizes the 15 bp overlapping sequence and inserts the gene into the vector.
**Amplifying small fragments**

Generally, most fragments can be amplified using the Gotaq Hot Start Master Mix. However, some fragments have very strong secondary structure, for which PCR conditions must be optimized. I developed a very good system for optimizing PCR conditions. These ten buffers cover many possible buffer combinations in PCR, which allows amplification of any complex template. I utilized Roche High Fidelity Faststart Taq. FastStart Taq DNA Polymerase has four-fold higher accuracy, allowing amplification of fragments up to 5 kb. It has an error rate less than 1 mutation per 1 kb of amplified fragment.
<table>
<thead>
<tr>
<th>Number</th>
<th>MgCl$_2$ (25 mM)</th>
<th>GC solution</th>
<th>DMSO (100%)</th>
<th>dNTP (10mM)</th>
<th>Buffer</th>
<th>Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48</td>
<td>40</td>
<td>20</td>
<td>80</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>40</td>
<td>20</td>
<td>80 with 1.8 mM MgCl$_2$</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>40</td>
<td>20</td>
<td>80</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>80</td>
<td>20</td>
<td>80</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>80 with 1.8 mM MgCl$_2$</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>48</td>
<td>0</td>
<td>20</td>
<td>80</td>
<td>252</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>64</td>
<td>0</td>
<td>20</td>
<td>80</td>
<td>236</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>48</td>
<td>160</td>
<td>20</td>
<td>80</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>160</td>
<td>20</td>
<td>80 with 1.8 mM MgCl$_2$</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>64</td>
<td>160</td>
<td>20</td>
<td>80</td>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Optimized conditions for PCR. All the figures are based on µl per 200 µl total volume. 5 x GC Buffer: 2.5 M MMNO (N-Methylmorpholine N-oxide) with 2.5 M proline, or 2 M MMNO with 2 M proline. The final working concentration is 0.4 - 0.5 M.
I list all the primers used to amplify the fragment in Table 9. The fragments were cloned into pGEM-T vector. A white-blue screen is performed to pick out the positive colonies. The plasmids are prepared using the Qiagen plasmid Mini Prep. The Nevada Genomics Center did sequencing reactions using the M13F and M13R primers.
<table>
<thead>
<tr>
<th>NPR1: fragment1-sense</th>
<th>GGAGACATATGTGGCGCGGC GCCGC</th>
<th>NPR1: fragment1-antisense</th>
<th>ATTTGCGTGACGATCTGCTCTCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPR1: fragment2- sense</td>
<td>ACGAAATATTTCCACACCAACG GCA</td>
<td>NPR1: fragment2-antisense</td>
<td>GGGGAAAGTCTCTTTTATTTCAACAATTGACG GTAAAC</td>
</tr>
<tr>
<td>NPR1: fragment2-antisense</td>
<td>GGGAAGTCTCTTTTATTTCAACAATTGACGTAAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR7: fragment1- sense</td>
<td>GGAGACATATGGCTGCAAAGCTCAGTAAAGAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR7: fragment1-antisense</td>
<td>A TCTTA T</td>
<td>CACAGTTGA TTCCGTTCAAC</td>
<td></td>
</tr>
<tr>
<td>NPR7: fragment2- sense</td>
<td>AT AAGGATTTGGTTTACCAACAAGAAATGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR7: fragment2-antisense</td>
<td>GGGGAAAGTCTCTCGCAGAGGATACCTACT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR5: fragment1- sense</td>
<td>GGAGACATATGGAAATTTATAAGAGGGATCACAG TAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR5: fragment1-antisense</td>
<td>ATGCTTCCGCACTCGCTTCAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR5: fragment2- sense</td>
<td>AAGGCACTCATATAGATATAATAGAGAAATAAAAA A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR5: fragment2-antisense</td>
<td>GGGGAAAGTCTCTATAATAGGACACGACACA CTAGT GTACTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR8: fragment1- sense</td>
<td>GGAGACATATGTTAAAAGGAGGTCGTGAGTGCGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR8: fragment1-antisense</td>
<td>ACGTGGATCTTCAGCGCCTCGCAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR8: fragment2- sense</td>
<td>AGATCCACGTGTCTCCTGACA CCAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR8: fragment2-antisense</td>
<td>GGGGAAAGTTCATTTAGGTTTTCACCTTTAGATA A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9. Primers used for fragment cloning. The sense primer and antisense primer are used to fish out the fragments from cDNA. Each NPR is bridged using fragment 1 and fragment 2.
Bridging the fragments into an expression vector

Preparation of a Linearized Vector by Restriction Digestion

To achieve a successful In-Fusion reaction, I must first generate a linearized vector. The linearized vector can be generated using restriction enzymes (single or double digests) or by PCR. Due to differences in cutting efficiencies, different restriction enzymes will generate different amounts of background. Generally speaking, digestion with two enzymes is better than with any single enzyme. Efficiency of digestion will always be better if the restriction sites are as far apart as possible. In addition, increasing the enzyme digestion time and the digestion reaction volume will reduce the background.

1. The pcDNA5FRT plasmid is prepared using the QIAGEN plasmid mini prep. The NheI and NotI restriction sites in the multiple cloning sites of this plasmid are utilized to linearize the vector. The following are components in the digestion buffer:
   10x Buffer 4 (NEB), 5 μl; 100 μg/ml BSA (NEB), 0.5 μl; NheI (NEB), 1 μl; NotI (NEB), 1 μl; water (Promega), 37.5 μl; pcDNA5/FRT 200 ng/μl (Invitrogen), 5 μl. Incubation at 37°C from 3 hours to overnight; the latter can increase linearization and reduce background.

2. After digestion, purify the linearized vector using the QIAGEN PCR purification kit.

3. Check the background of the vector by transforming 10 ng of the linearized and purified vector into competent cells.

4. Prepare the PCR fragments

   The In-Fusion method is not affected by the presence or absence of A-overhangs, so one can use any thermostable DNA polymerase for amplification, including
proofreading enzymes. I used Roche High Fidelity Faststart Taq with the same working buffer as the last step (see table 5). When PCR cycling was complete, I analyzed the PCR product by agarose gel electrophoresis to confirm the identity of the DNA fragment. The dNTP, primer and salt concentration decreases the In-Fusion enzyme efficiency. The PCR products are purified with the QIAGEN Gel Purification Kit.

5. Set up for the In-Fusion cloning reaction:

5X In-Fusion HD enzyme premix, 2 µl; linearized vector, 4 µl (150 ng); 3 purified PCR fragments, 4 µl (90 ng); total volume, 10 µl.

Note: The ratio of the vector and fragments is calculated using the In-Fusion molar ratio calculator (http://bioinfo.clontech.com/infusion/molarRatio.do).

Incubate the reaction for 15 min at 50ºC, then place on ice. After transformation, colony PCR is utilized to screen the positive colonies. The positive colonies were then sequenced by the Nevada Genomics Center.
Part V. A mutation in NPR5

When I got the sequence for NPR5, I found that all of the four positive colonies had the same mutation. The mutation (GTAG) is located between the kinase domain and the guanylate cyclase domain (Figure 20). TAG is used as the stop codon in the genetic code, which may cause protein translation to stop without including the guanylate cyclase domain. This mutated guanylate cyclase without guanylate cyclase activity, may function as the clearance receptor in insect systems. NPR-C (the clearance receptor in mammals) binds the unneeded natriuretic peptides (ANP, BNP, and CNP), and then degrades them by endocytosis (Yap et al., 2004).
Figure 20. An alternative splicing site in NPR5. The mutation is shown as GTAG, which is located between the kinase domain and the guanylate cyclase domain. TAG is the stop codon.
To characterize the guanylate cyclase activity of the receptor, NPR5 was mutated to knock out the stop codon sequence GTAG. I used PCR based mutation technology, which linearized the plasmid using two back-to-back primers (Figure 21). The linearized plasmid was treated with the Quick Blunting Kit (New England Biolabs) and then recircularized with T4 DNA ligase (New England Biolabs). DpnI cleaves only when its recognition site is methylated. DNA purified from a dam$^+$ strain will be a substrate for DpnI, which is utilized to destroy the original plasmid template. The identity of the mutated plasmid was confirmed by sequencing by the Nevada Genomics Center.

PCR is performed using Phusion High-Fidelity DNA Polymerase which gives robust performance and can be used in all PCR applications. Its unique structure, a novel *Pyrococcus*-like enzyme fused with a processivity-enhancing domain, increases fidelity and speed. Phusion DNA polymerase is an ideal choice for cloning and can be used for long or difficult amplicons. It has an error rate 50-fold lower than that of *Taq* DNA polymerase and 6-fold lower than that of *Pyrococcus furiosus* DNA polymerase. Phusion is the most accurate thermostable polymerase available. Phusion DNA polymerase possesses 5´→ 3´ polymerase activity, 3´→ 5´ exonuclease activity, and will generate blunt-ended products. PCR is done using an MJ research thermal cycler (98°C 1 min, 30 cycles 98°C 15 sec 72°C 3 mins, 72°C 5 mins):

- primer M-sense, 2.5 µl;
- primer M-anti, 2.5 µl;
- NPR5 plasmid, 1 µl;
- 5x Phusion HF buffer, 10 µl;
- Phusion hot start DNA polymerase 0.5 µl;
- water 33.5 µl.

When PCR cycling was complete, I analyzed the PCR product by agarose gel electrophoresis to confirm the identity of the DNA fragment. The PCR product was purified using a QIAGEN PCR purification kit and eluted with 50 µl water. The quick
blunting kit (NEB) is used to convert DNA with incompatible 5´or 3´overhangs to
5´phosphorylated, blunt-ended DNA, which is appropriate for down-stream self-ligation:
blunting buffer (10X), 2 µl; blunting enzyme mix, 0.5 µl; deoxynucleotide solution
mix (1 mM), 1 µl; eluted DNA, 17.5 µl.

The DNA is recircularized using T4 DNA ligase:
10x ligation buffer, 2 µl; T4 DNA ligase, 1 µl; phosphorylated DNA, 5 µl; water, 13 µl.
Incubate at room temperature overnight. 1 µl DpnI is added and incubated for another
hour. The final product is transformed into TOP10 competent cells. Usually this method
will generate 95% positive colonies. 4 different colonies were sent to the Nevada
Genomics Center to confirm the sequence of the plasmid.
Figure 21. Flow chart of the Phusion site-directed mutagenesis kit protocol. Any target plasmid can be used as a starting material for creating point mutations, deletions or insertions. Phosphorylated primers are designed to introduce the desired mutation (A, B, C1 or C2) so that they first anneal back to back to the plasmid. In the PCR reaction, Phusion DNA Polymerase extends the primers and amplifies the plasmid with the mutation. The mutated plasmid is then circularized by quick ligation and transformed into bacteria.
Part VI. qPCR to detect gene abundance

To study gene abundance, we had to prepare total RNA from head, Malpighian tubule, midgut, hingut, fat body, trachea, ileum. Using high quality total RNA for qPCR is not necessary. I used NucleoSpin RNA II (Clontech) to prepare genomic DNA-free total RNA. Experiments were conducted using the standard protocol.

The cDNA is synthesized using the Roche Transcriptor reverse transcriptase from total RNA as a template. Clontech Smarter Scribe MMLV, which can amplify rare transcripts with PCR bias, is not appropriate for real-time qPCR. Roche Transcriptor reverse transcriptase system:

Total RNA (1-2 µg), 3 µl; random primer (N15), 1 µl; water, 9 µl; final volume to 13 µl.

Incubate at 65°C for 10 minutes, and then place the tubes on ice. This step ensures denaturation of RNA secondary structure. Add the following components:

Transcriptor RT reaction buffer, 5X, 4 µl; Protector RNase Inhibitor, 0.5 µl; dNTP, 2 µl; Transcriptor Reverse Transcriptase, 0.5 µl.

Incubate at 55°C for 30 minutes. Inactivate the reverse transcriptase by heating to 85°C for 5 minutes.

We performed quantitative PCR (95°C 2 mins, 30 cycles 95°C 30 seconds, 65°C 1 min) with an ABI Prism 7500 with specific probes; SYBR Green PCR master mix (Applied Biosystems) was used to detect mRNA in different tissues (ileum, tubule, head, trachea, and whole body). The specific primers are listed:
<table>
<thead>
<tr>
<th>Gene</th>
<th>Product size</th>
<th>Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPR1: qPCR600-sense</td>
<td>600 bp</td>
<td>66°C</td>
</tr>
<tr>
<td>5' AGTCTCTGTGGGTTGCTG 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR1: qPCR600-anti</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' CTCGTCTTTCTGGCTTCCTTT 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR5: qPCR600-sense</td>
<td>600 bp</td>
<td>66°C</td>
</tr>
<tr>
<td>5' CATGCTGGCAGCCTGAACTA 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR5: qPCR600-anti</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' GCCATGGCAGCCTGGGACTTT 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR7: qPCR-sense600</td>
<td>600 bp</td>
<td>66°C</td>
</tr>
<tr>
<td>5' CCCCCCGTGGCAGCCACTCTAC 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR7: qPCR-anti600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' TCTGGCAGGCACTGCCCACAT 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR8: qPCR-sense800</td>
<td>800 bp</td>
<td>66°C</td>
</tr>
<tr>
<td>5' TCTGTGCTGCTGACGGAATCTGC 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR8: qPCR-anti800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' AGGTCCACCACCTCGCGCTG 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR6: qPCR600-sense</td>
<td>600 bp</td>
<td>66°C</td>
</tr>
<tr>
<td>5' GAGTTGCTGCGGGGTGCTG 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPR6: qPCR600-anti</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' CGATTGCTGCTGAAGTGACTG 3'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Primers used for qPCR.
Gene abundance in different molting stages

Samples were collected to study gene variability in different stages as follows: -2 (two days before molting), -1 (the day before molting), 0 (molting), 1 (1st day after molting), 2 (2nd day after molting). The cDNA is synthesized using Transcriptor reverse transcriptase (Roche) from total RNA, which is prepared using a NucleoSpin RNA II cartridge (Clontech).

Part VII. Receptor characterization

The sequenced positive colonies are pre-cultured in 5 ml of terrific broth overnight (37°C, 150 rpm shaker). 100 µl of the pre-culture solution is added to 100 ml of terrific broth overnight (37°C, 150 rpm shaker). The plasmid is prepared using the PureYield Plasmid Maxiprep System (Promega) or the QIAGEN Plasmid Maxi Prep as a standard protocol. The plasmid is quantified using the NanoDrop 8000. CHO cells (Chinese hamster ovary) are plated into 96-well plates the day before transfection. I used 0.5 µg plasmid per well. The transfection is performed using jetPEI or Lipofectamine 2000 as a standard protocol. Generally, jetPEI gives better results than Lipofectamine 2000.

The different hormones were prepared in different dilutions from $10^{-7}$ M to $10^{-14}$ M in F12 cell culture media, which is used for CHO cell cultures. Then the hormone dilutions are added into wells using a multiple-channel pipettor at the same time. After 30 minutes incubation at 37°C, the cells are lysed with 100 µl of 0.1 N HCl. The acid is neutralized with 200 µl of 0.1 N NaHCO₃.

We used a competitive binding assay to analyze the concentration of cGMP (within a range of 100 pmol/ml to 0.5 pmol/ml). Sheep anti-cGMP serum (Abcam) is
used to coat Greiner Microlon 600 plates using universal coating buffer (Cell Technology). The cyclic GMP in samples is analyzed by competitive binding with cGMP-HRP (Genscript). The protocol follows:

EIA buffer: PBS with 0.1% BSA, pH 7.4
Washing buffer: PBS-T with detergent such as 0.05% (v/v) Tween 20

1. Coat 96 well high binding plates with 100 µl of cGMP antibody per well at a dilution of 1:2000. Incubate at 4°C overnight on an orbital shaker.
2. Remove unbound antibody by washing with 200 µl washing buffer (3 times) and pat dry on paper towels.
3. Blocking: add 200 µl EIA buffer and incubate for 2 h at room temperature.
4. Discard solution, wash with 200 µL washing buffer (3 times), and pat dry on paper towels.
5. Prepare HRP-cGMP dilution (1:10000) by diluting the serum with EIA buffer.
6. Add 90 µl HRP-cGMP samples to wells, run in triplicate.
7. Add 5-10 µl testing sample to the wells.
8. Incubate at room temperature for 4 h or overnight.
9. Discard the solution and wash wells thoroughly and rinse immediately with washing buffer (5 times), followed by distilled water (2 times).
10. Add 100 µl TMB (KPL) and develop at room temperature for 10-30 min.
11. Read plate at 630 nm (blue color) in a micro-plate reader.
12. Add 50 µl 1 N HCl to each well to stop the color development and read immediately at 450 nm (yellow color).
The receptor is weakly expressed in CHO cells. COS-7 cells and HEK-293 cells failed to express functional guanylate cyclase receptors.

**Genomic DNA mapping**

Based on *Manduca sexta* base (http://agripestbase.org/manduca/?q=blast), we located different NPRs in assembly 1.0 scaffoldDB (June 2011).
RESULTS

Nested PCR using degenerate primers

The efficiency of degenerate primers is very low, because of sequence degeneracy and cDNA background interference. Nested PCR helps to reduce the cDNA background interference (Moser et al., 2012). We designed two sets of degenerate primers (Figure 14). Set 1 is designed based on the partial sequences DIVGFT and YCLFGD. Set 2 is designed based on DVYKVETI and MPRYCLFGD. At the same time we improved the degenerate primers by using inosine to substitute N for (A, T, G, C) in the degenerate primers.

The PCR result is shown below (Figure 22). The PCR product is ligated into pGEM®-T Easy Vector. This PCR product contained 7 different guanylate cyclase fragments. We had to pick out many colonies to include all the different sequences. A white/blue screen was used to pick out 50 positive colonies. Plasmids were prepared using a QIAGEN mini prep and sequenced by the Nevada Genomics Center. The sequences are shown in the appendix.

PCR of dC-tailed cDNA

The components of the tailing reaction buffer have a negative impact on the PCR. Generally, companies use a large amount of terminal transferase in their buffer, which is compatible with PCR. Here I describe a new method. I studied buffer capability and PCR efficiency by comparing ten different PCR buffers and Taq polymerases (Figure 23). The different PCR buffers were mixed with the tailing buffer. PCR efficiency is checked by amplifying a 3 kb fragment. The results are shown in Figure 18. The Roche Faststart Taq GC buffer mixed with tailing buffer can enhance PCR efficiency.
Amplifying small fragments of NPRs

In my hands, the NPR5 3' fragment was the most difficult fragment to amplify from cDNA. Only the ninth condition was suitable for this fragment (Figure 24).

Gene abundance in tissues

The gene abundance in tissues is plotted by Prism 5 (Figure 25). *Manduca sexta* actin (L13764) and *Manduca sexta* ribosomal protein L7 (GU084319) are used as the reference genes.

Receptor characterization

The receptor NPR5 was expressed in the CHO cell. ITP, ADFb, EH were used to stimulate the cell to produce cGMP. The dose-response cGMP concentrations were plotted by Prism 5 (Figure 26).

Genomic DNA mapping

NPR1 genomic DNA maps to scaffold00097. NPR5 genomic DNA is also mapped to scaffold00097. NPR8 genomic DNA maps to scaffold00044. NPR7 genomic DNA maps to scaffold00063. NPR6 genomic DNA maps to scaffold00198. There is not yet a graphical map of the draft *Manduca* genome.
Figure 22. Results of nested degenerate PCR from cDNA. DVYKVETI-sense: 5’
GAYGTITAYAARGTIGARACIAT 3’, MPRYCLFGD-antisense: 5’
CCRAAIARRCARTAICKIGGCA T 3’ are used for this nested degenerate PCR. The PCR
product size is 300 bp. The PCR templates were cDNAs from different tissues of
*Manduca sexta* (a-midgut, b-hindgut, c-Malpighian tube, and d-head). The ladder is the
NEB 100 bp ladder (e). The lowest band is 200bp.
Figure 23. Different PCR systems compatible with TdT buffer. Polymerases were checked in different PCR buffers mixed with the tailing buffer. The PCR efficiency was checked by amplifying a 1 kb fragment. (1) Roche Faststart Taq system; (2) NEB LongAmp PCR system; (3) Roche GC-rich PCR system; (4) Clontech Advantage PCR system; (5) Phire PCR system; (5) Phusion HF PCR system; (6) Phusion GC PCR system; (7) Invitrogen Taq system; (8) Invitrogen HiFi PCR system; (9) The DNA ladder is the NEB 100 bp ladder.
Figure 24. PCR efficiency of the NPR5 3’ fragment was tested in different PCR buffers (Table 8). The NPR5 3’ fragment was the most difficult fragment to amplify from cDNA that I used. Only the ninth condition ((10 ) the right side of the ladder) tested was suitable for this fragment. The ladder is the NEB 1kb ladder (9). (1) PCR in buffer 1; (2) PCR in buffer 2; (3) PCR in buffer 3; (4) PCR in buffer 4; (5) PCR in buffer 5; (6) PCR in buffer 6; (7) PCR in buffer 7; (8) PCR in buffer 8; (10) PCR in buffer 9; (11) PCR in buffer 10.
Figure 25. Quantitative PCR of cDNA from 6 tissues and from whole body extracts. All tissues were dissected from the same *Manduca sexta*. Three different *Manduca* were used in this experiment. The tissue RNA was prepared using NucleoSpin RNA II. The RNase-free DNase is used to reduce the genomic DNA contamination. The abundance of transcripts is based on ribosome L7 (down) being 1.00.
Figure 26. Preliminary hormone dose-response curve for NPR5. The receptor NPR5 is expressed in CHO cell and HEK cell. CHO cells proved to not be useful for these studies. These data are only duplicates; no standard deviations are possible.
DISCUSSION

Sequence analyses

Based on multiple sequence alignments of NPR1, NPR5, NPR6, NPR7 and NPR8 with their orthologs in other species, we judged the completeness of the sequence analysis. NPR7 is about 100 amino acids shorter than the other 4 NPR sequences. The final decision of whether a sequence is complete or not is to analyze it for a signal peptide using the signalP server (http://www.cbs.dtu.dk/services/SignalP/). From this analysis, we confirmed that NPR1 and NPR5 are complete at the 3' and 5' end. Because a draft Manduca genome sequence was completed recently, I analyzed the NPR7 sequence by searching the genome sequence. I also designed some NPR7 5' end primers to check the 5' end sequence, which is designed based on the available NPR7 genomic sequence. The NPR8 5' end sequence was recently completed based on the Manduca transcript library sequences.

A new RACE PCR method

5' and 3' RACE PCR is not efficient for most genes. The key enzyme for this new method is CircLigase (Epicentre), which is sequence dependent. Based on my studies of research by Epicentre scientists, at the 3' end, ligation efficiency is greatest if the 3' end nucleotide is a T, and T ≧ A >> G >>> C (a "C" on the 3' end will not ligate with efficiently). At the 5' end, the ligation efficiency is greatest if the 5' end nucleotide is a G, and G > A ≧ T > C. The single-strand circularization ligation loses competence if the 3' end nucleotide is G. However, generally the mature mRNA has a 5' cap structure, which is found on the 5' end of an mRNA molecule and consists of a guanine nucleotide connected to the mRNA via an unusual 5' to 5' triphosphate linkage. This would create a
C at the 3' end of the cDNA, which will not ligate efficiently. The only good method is to utilize MMLV rather than other high fidelity reverse transcriptases. MMLV has terminal transferase activity, transferring an unknown nucleotide base to the 3' end of the cDNA, which makes sequence dependent ligation possible. Generally, the nucleotide base added is C or A.

**Detecting alternative splicing isoforms**

We observed that NPR5 has a mutation between the kinase domain and the guanylate cyclase domain. We used the *Manduca* genomic sequence to analyze how this mutation happened. Below is the genomic sequence of the region covering the exons where the mutation occurs. The exons are shown as italic characters and the intron is shown in normal black characters (Figure 27). As a general rule, the intron donor and acceptor consensus sites are GT and (C/T) AG, respectively. The NPR5 intron contains a possible alternative intron donor, GTAG (bold characters), immediately next to another potential splice donor, GT (underlined). In a biological system, alternative splicing may occur, including or excluding the GTAG. I think likely both spliced forms occur *in vivo*. The next question is whether this alternative splicing has biological meaning, such as the use of alternative splicing as a control mechanism for certain functions. This cannot be answered at the moment. There are many genes that have unusual alternative splicing. A large percentage of this type of alternative splicing results from biological mistakes, which fuels evolution. The splicing machinery is not a perfect system that has difficulty in recognizing the exact splicing donor.

One way of testing this hypothesis is comparing this region of gene sequence from multiple taxa (other insect species). However, I did not find the intron "GGTAG"
followed by a "GT" in a similar region of *Tribolium* genomic DNA. In the closely related silkworm, there is no intron in this region. It may not be conserved in other taxa. This stop codon may not have a biological role. A remote possibility is that the protein product of the wrong transcript (including the GTAG) may function as a dominant negative regulator of the functional receptor. The other possibility is that the splicing form without guanylate cyclase activity may function as a clearance receptor, which is similar to NPR-C (clearance receptor) in mammalian animals. The reason for the alternative splicing remains as a long-term question.

Not only does NPR5 contain alternative splicing forms, there are also isoforms in other NPRs. There are 3 alternative splicing isoforms in the 3’ end of NPR1, and 2 different alternative splicing isoforms in the 5’ end of NPR1. Two different splicing isoforms exist in the 3’ end of NPR8. I did not observe any isoforms in NPR7; this gene cannot be found in the draft genome. We presume that the NPRs with the longest ORF (open reading frame) have all expected functions of guanylate cyclase receptors. Based on this assumption, the longest isoforms are expressed in those cells which are hormone targets. We next need to conduct experiments on receptor characterization. The short isoforms may function as dominant negative regulators in the dimeric formation of functional receptor. However, Chang *et al.* (2009) found that the short isoform of the eclosion hormone receptor (NPR5) has native guanylate cyclase activity without hormone stimulation. The short isoforms may have an unknown regulatory function, possibly generating some cGMP before ligand stimulation.
Figure 27. Partial gene sequence of NPR5 which explains the alternative splicing. This is a partial genomic sequence in the location where the stop codon mutation occurs. The italic characters are the exons. The black characters are the introns. The bold characters and underlined characters are the alternative splicing sites.
Tissue localization

We analyzed data by qPCR using two reference genes - *Manduca sexta* actin (L13764) *Manduca sexta* ribosomal protein L7 (GU084319). The Q-PCR data indicate that, NPR1 is rich in Malpighian tubule, NPR5 is rich in trachea, NPR8 is rich in the hindgut and NPR6 is rich in the fatbody (data not shown). Based on analysis for the actin reference gene, all the genes abound in the fatbody. The actin gene expression is usually used as the reference gene in mammalian animal researches, however, the actin expression level significantly varies in different tissues of Manduca (http://silkworm.swu.edu.cn/microarray/Bmarray.php?search=yes&range=sw22671).

Chang et al. (2009) characterized one guanylate cyclase receptor as the eclosion hormone receptor in *Bactrocera*, which is highly expressed in Inka cells on the tracheae. NPR5 is the *Manduca* orthologue of the *Bactrocera* EH receptor according to Clustal W. We expected to find the EH receptor in trachea; the apparent expression of four additional rGCase in this tissue is an interesting enigma. Only limited immunohistochemical data were performed on the *Bactrocera* EH receptor, concentrating on the tracheal junctions. Our Q-PCR data show all four receptors to be highly expressed in the trachea, whereas NPR1 is most highly expressed in Malpighian tubule. We initially expected NPR1 to be the Tenmo-ADFb receptor, and NPR5 to be the EH receptor.

Very recent results attempting to characterize the NPR5 receptor show that it responds similarly to Schgr-ITP and Tenmo-ADFb (Figure 26). In contrast, the response to Manse-EH is barely significant. This is remarkable as Clustal W alignments show that NPR5 is the *Manduca* ortholog of the *Bactrocera* EH receptor reported by Chang et al. (2009). Further, the peptides tested were not the *Manduca* orthologs of ITP and ADFb, as
such samples are not currently available. *Schistocerca gregaria* ITP, the form tested on
the receptor, has high activity in *Schistocerca* and *Periplaneta americana*, but not on
other species (Meredith et al., 1996). No sequence is known for a *Manduca* ortholog of
Tenmo-ADFb. Thus, the EC\textsubscript{50} values seen in this graph for ITP and ADFb are likely
considerably higher than the EC\textsubscript{50} values for their *Manduca* orthologs, which are likely in
the sub-nanoMolar range.

From these surprising findings, we hypothesize based on Clustal W alignments
that NPR1 is the ADFa receptor of *Manduca sexta*, which is located in Malpighian tubule.
It will be quite interesting to determine immunohistochemical (IHC) localization of
NPR5 and NPR1. ISH is more likely to be able to read truncated isoforms, as it is
normally used for a much larger portion of the protein than IHC; the only good antigen in
NPR5 is upstream of the extracellular domain.

NPR5 and NPR8 are moderately abundant in tracheae, which is the respiratory
system in *Manduca*. The tracheae have recently been shown to "breath" very quickly;
perhaps being filled with molting fluid may cause hypoxia (Greenlee et al., 2009;
Westneat et al., 2003). The preecdysis stage occurs just prior to exuviation, and is
characterized by separation of the old exoskeleton from the underlying epidermal layer
(Zitnan and Adams, 2012). The molting fluid may be partly reabsorbed through the
tracheae, and energy reserves are mobilized by the midgut. Is a pulse of ITP-like
immunoreactivity observed at the molt (Drexler et al., 2007) important in clearing the
trachea of molting fluid they take in? Does the lethal effect of deleting the ITP gene in
*Tribolium* ("100% cumulative mortality at the time of eclosion or shortly afterward"
(Begum et al., 2009)) result from inability to clear the airways of fluid? It is possible that
NPR7 also responds strongly to ITP, and perhaps Tenmo-ADFb. Levels of ITP-like immunoreactivity rise in *Manduca* blood when molting fluid is being reabsorbed; most molting fluid is re-ingested into the foregut. This large intake of fluid may require action of ITP for reabsorption in the tracheae, as mentioned above. There are 5 Cys in the ECD of NPR5. Again, ISH data will be invaluable in localizing tissues which express this receptor.

The extracellular domains of the rGC are the least conserved elements. However, analysis of NPR1, -5, -7, and NPR8, and their *Bombyx*, *Tribolium*, and *Drosophila* homologues, shows each to contain the characteristic juxtamembrane motif found just before the TMD in all animal rGCase sequences, with the exception of GC-C (Misono et al., 2011). This domain is crucial for signaling: mutation of certain residues abolishes activity, while mutation of others causes constitutive activity. *Manduca* NPR1, and its *Tribolium* and *Drosophila* orthologues, have 12 Cys residues in the ECD; the similar *Bombyx* rGCase has 11 Cys. This will give a far more cross linked structure than the corresponding ECD of the ANP receptor which has 6 (rat) or 7 (human) Cys residues. The NPR1 receptor is expressed in Malpighian tubule examined. We hypothesize that NPR1 is the likely receptor for ADFa based on ClusterW alignments.

ClustalW alignments of rGCase for *Bombyx*, *Drosophila*, *Tribolium*, and *Manduca* show that NPR 8 is most similar to the rat ANP receptor, which is highly expressed in trachea. The *Bombyx* ortholog of NPR8, SP accession Q9BPR0, is a rGCase which has six Cys in the ECD and is known to align most closely to GC-A, the vertebrate ANP receptor (Tanoue et al., 2001). This receptor was reported to be most highly expressed in the antennae and antennal lobes of the brain, but the referenced paper
includes a Figure showing that this rGCase is expressed in all 8 tissues tested, including the tubules; these data are based on IHC staining of the receptor. ADFb is an antidiuretic hormone identified by our group (Eigenheer et al., 2003).

**NPR6**

We performed a ClustalW alignment of NPR6 orthologs in *Bombyx* (SP Q9BPR0), *Tribolium* (NP_001167546.1), and *Drosophila* (CG31183). NPR1, NPR7, and NPR8 alignments were done using a similar method. We found that NPR6 is the most conserved family member of the guanylate cyclase receptors. NPR6 must have an important function because of this high conservation. Full length NPR6 has not been cloned in any insect. qPCR results show that there is no NPR6 in the head, midgut, hindgut, or rectum in *Manduca sexta* (Figure 25). Based on microarray data in the silkworm genomic database and fruitfly genomic database (flybase), NPR6 may be moderately abundant in testis (adult), fat body (larvae), and hemocytes (larvae). I detected low levels of NPR6 in the fat body and hemocyte mRNA from *Manduca* (data are not shown). Testes are important in insect reproduction. However, we found no published data on the relationship between cGMP and insect reproduction. In the future, I hope to study any function of the guanylate cyclase receptor NPR6 in insect reproduction. Some processes in insect reproduction may rely on cGMP signaling.

Fat body and hemocytes are important in insect immunity (Aggarwal and Silverman, 2008; Leclerc and Reichhart, 2004). The strong functional similarities between the insect immune response and the vertebrate immune response suggests a possible common lineage between these systems. Upon examining the rooted phylogeny
of several immune genes from both vertebrates and arthropods, immune systems have arisen independently in vertebrates and arthropods (Hughes et al., 1989). Nevertheless, there is evidence to support the notion of conservation in function of the vertebrate immune system and the insect immune system at the functional level.

Several proteins involved in both the vertebrate and invertebrate immune responses appear to be conserved. Lysozyme and hemolin are the most common proteins in both immune systems, sharing a conserved sequence (Sun et al., 1990).

The vertebrate immune system is based on arrangement of many modulators to provide the inducible and innate immune response. Neurohormones can influence the behavior of vertebrate immunocytes via Ca\(^{2+}\) concentration, cAMP and inositol trisphosphate (Hallam et al., 1984). I propose an insect immune system model here. It is possible that when pathogens infect the insect, they stimulate the insect to secrete neurohormones. Some of these neurohormones may target guanylate cyclase receptors to produce the second messenger cyclic GMP. The cGMP may initiate the signal cascade of immune reaction.

Insect fat body and hemocytes are known to mediate the immune response in insects (Marmaras and Lampropoulou, 2009). NPR6 has the most conserved sequence between different insect species of any Manduca rGCase. On the other hand, NPR6 is an ultra-low abundance gene in insects. We believe that NPR6 will be related to some important common functions shared by all insects. From micro-array data in the silkworm database and fruitfly database, NPR6 homologous genes are expressed in the hemocytes and fat body, which are the major sites to secrete antimicrobial peptides (Hoffmann, 2003). The larval fat body is a major site of intermediatory metabolism of
insects and has a function analogous to that of the vertebrate liver. It consists of thin strings, one or two cells thick, or small nodules suspended in the hemocel and distributed throughout the insect body (Roma et al., 2010). The fat body is a target tissue for many insect hormones and is also a site of response to microbial infection (Jiang et al., 2010). NPR6 is the receptor likely to be activated by a neurohormone to simulate the fat body immune response. Characterized immune genes, in fat body, are induced by microbial infection and encode antimicrobial peptides which are then released into the hemolymph (Zhu et al., 2003). In addition, Lepidopteran fat body releases several other proteins, such as pattern recognition proteinases: prophenoloxidase activating proteinase and a serine proteinase inhibitor from the serpin family (Jiang et al., 2010).
REFERENCE


Nachman, R.J., Coast, G.M., Holman, G.M., and Beier, R.C. (1995). Diuretic activity of
C-terminal group analogues of the insect kinins in *Acheta domesticus*. Peptides 16, 809-
813.

messenger of crustacean hyperglycemic hormone signaling pathway in the kuruma prawn

Eclosion hormone of the silkworm *Bombyx mori*: Purification and determination of the
N-terminal amino acid sequence. Insect Biochem. 15, 573-578.

thermodynamic analysis of 3’ double-nucleotide overhangs neighboring Watson-Crick

moned-bound atrial natriuretic peptide receptor extracellular domain: rotation mechanism

moned-bound atrial natriuretic peptide receptor extracellular domain: Rotation mechanism

and preliminary X-ray analysis of the atrial natriuretic peptide (ANP) receptor extracellular
domain complex with ANP: use of ammonium sulfate as the cryosalt. Acta Crystallogr


APPENDIX

NPR1

acctctgcaagaaccagtgcttaaatcactgagcccatacacttcaactctaactctttttttttttttttttttggcacaagtgaagttggtagttgatggctgaagtttgaagattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
agccagaagacgagccgtcgtgcgggttcaactacgagaagtgtccgctgcagatcagtgcctggtggtgctgagctgatatacgctgtatcgaacgacgaataaatccacacgagaagacagctgaagacgacgacgactgagcaggagatagaaggactgctgtggaagatagacccgaatgaggtcgtcggctatcagggcagcgggcttgtggtcgctagtaagctgagctt
gagccggagtcattcgattcggtcactatttacttcagcgatatcgttggtttcaccgctatgtcggccgagagcactccgctgcaagtggtcaacttccttaacgacttgtacacgctgttcgatagaattataaggggctacgacgtgtacaaagtggagacggcgacgcatacatggttgtttcgggcttaccgattcgaaataacgaccgccatgttggcgagattgcatccatggcagctgaa
cctactgaacgcggcgaagagtcacaaaatagcgcatagacctaatgaaacattaaaattaagaataggtattcacactggcgca
gtagtggctggtggtgggacttacaatgcctcgctattgtttttggggacacggtgaacaccgcgtctagaatggaagtcgacgggggagccgctgcgtaattcatactatcgccgaactgcaagcaagcactcgacaaactagaaggctacattgtcgacgacgagggactataccgatcaaagggaagggacagcttcagacttattggttagtcagcgccaccgctgctggaccgcgagtcggc
ggtgtgaacacgcgcgcgcacgcgcaccagagagatgtacattagagattatgtgtatatga
NPR5

tgagtatacaacagtcagcgaatcagaacgaagcggctcaaaacgcgtacaaccgcggtcgcgggaaacctacgtaatggcg
gcaaaaaattgtgaatatagtaaatgaatatcatacatataccagttttgtggaacagctaaatatcaatgtaagttctaaatgttgcga
acatggaaatttttaagaggatcacatgaaaaatacatattttgtggtttatatctgtttatctagtctatcatactgcgcgtcagcgcga
atttactttaggatatctgacttgatcacaacgacactagctatcagcagctatccgactgatcgcgaatcagcgcgtcgcggcagctacac
agaaaggaaggtggtatctcagagctgagctgtttggacaccaagtctgctagctactgctgtgctggtggggagctgctgttgcaacaacgctgt
gcgaagttgttcgctggcgcgecgacgcggtgtgtgtgctgtggtgcaacactggtttctgcaacgtgctggcgcgtggtgtgtgtgtgtgtgtgtgt
acgcaattagtttacgtggatgagctggtgggatctgctgggatagctgctgatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg
gacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg
caccacccacgggtacacctcttcggcacaagaaacgaaacacctgacataaaactgtcatcaacggatgttctttatatacgga
aagggcaggtgccccgcataccctctctcgagagccgatataatgtatatttgctacatctgtagttttatagcgcgcgtatgctgt
cgatgtgtcgtatttgtgcttttatgacatcgaagcagagccgccaagccgcaagccgacacctgctgagctgcgggtgcgcgcgc
ctcagtgatttacgtcagttattataagagacgacggttcgatagaaattaatatataaatatatctttaccaaatcagaaaaaa
aaaaaaaaaaaaaaaa

NPR7

gcgggcacaatattaadaaagaaacgacacagagattttgtctatgcataataataactacccataaagatgaattagactattatgata
ttttaacgcgacaagaacaacatcgagagttgtcatacgcagagcatggaatgtttctagaaagaagttctgcggtatattgttcgaacgc
aacagcgaagcagacaactttctcacaagcgaagcagataaactcaagcgcgtacaccacctatcgcgaagataattgtcgtgtagacgacatc
gagctgagtggatggagtagtttgtatgaccttataaatgccgtattttgcggtatagtaaccataaatccttttcataaatcataatgacgtgctagc
gtctcgcagcatcgtgcagcagataaatcagcgtgtggttattgtgacatcgcgactcatcggtataattttattttattttattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
NPR6

agaga
tgac
ttga
tggata
gtgc
ttgat
tcct
catt
tttt
tttaat
taattt
tggtggt
ttgat
tttttt
ttat
taat
tgcag
tgcttt
tttttt
tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
cagggcggttatgaggggagacgctggcgggagccagacagtacgcgttcgggagcccccacgtggtacgcttacgaagctct
actccgagttctctgtgaagctgccactgggctgatagatcagacttttctgtgataactgtccgaaagctgagctgagttaaactaca
tatacagttctgtgatgagagaggttaaaattttctcagggcgcgttctataggcggtgtatgttactgggcaggtggttaaaga
gagctgagcagacagtacgagacgccacgtggggcagaggtctgctgatactgctcggacaagctcagcattaca
actacacggttctctgatggggaagaggtaaaattttctcagggcgcgttctataggcggtgtatgttactgggcaggtggttaaaga
gagctgagcagacagtacgagacgccacgtggggcagaggtctgctgatactgctcggacaagctcagcattaca
tccaccattataactacaactgtgtcataataaagagatatgtgacacgtgtatgtgagttgttgcccgagagtcacccgctgctccgaggtgttt
ggggttgggagagaattcagatgcgctggccctggccagactactggagtggctggggtgctggtgacgtgacgctgagacggcagtacgagcttcgatacattacgctgctgatacttctgatggggaagaggtaaaattttctcagggcgcgttctataggcggtgtatgttactgggcaggtggttaaaga
gagctgagcagacagtacgagacgccacgtggggcagaggtctgctgatactgctcggacaagctcagcattaca
accacgagcagatagacgccgctgatacttataactaatagattatcagctggatattcttttttctgatgtgaagggggagtgcggcgcgtgttatttactgggcatggcgttgaatgagacgctgacagaaggcgaggacatcagagacgccacatcaccagggcgagctgctgatactgctcggacaagctcagcattaca
acgctagcagacagcagagactgtgatactatcagctgctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg
aagaattactgtatcagctgatacttctgagggggaagaggccgtccg