Modelling and expression of the extracellular domain of the human guanylyl cyclase C receptor

A prelude to the study of the interaction between the GC-C receptor and the heat-stable enterotoxin from enterotoxigenic Escherichia coli

Marie-Josée Porcheron

Thesis submitted in partial fullfilment of the requirements for the degree of Master of Science

Department of Molecular Biology,
University of Bergen, Norway

Bergen, September 2011

Acknowledgements

This work was carried out during the period of November 2009 to September 2011 at the Department of Molecular Biology, University of Bergen.

I would like, as a first, to give my most sincere thanks to my supervisor, Dr. Pål Puntervoll, as well as my co-supervisors Pr. Rein Aasland and Dr. Nathalie Reuter, for their guidance through all the steps of this work. I really appreciated your encouragements and your ability to keep me in (relative) focus.

I am also very grateful to the members of the EntVac project I had the chance to work with, especially Arne Michael Taxt and Dr. Yuleima Diaz for their daily counselling, but also Thomas A. Aloysius and Stian Henriksen for their tips and discussion.
My thanks also to all the members of Lab 4 for always being ready to help, from directions to protocols, even after the fifth time.
Another thanks is in order for Marielle Ryste Hauge from the Administration team and Carol Issalene from Lab 3 for their help with the difficult administrative paperwork.
I would like to thank Pr. Johan Lillehaug for his interest.
All of you, as well as the other members of the department, I thank for the marvellous work atmosphere at MBI.

I acknowledge the Computing Biology Unit from the Uni Computing department for the use of their machines.

Special thanks to my family and friends.

Bergen, 14th of September 2011
Marie-Josée Porcheron

Abstract

The human guanylyl cyclase C receptor is the target for the heat-stable enterotoxin (STa) from enterotoxigenic *Escherichia coli*, which is responsible for more than 200 million episodes of diarrhea and 300,000 deaths per year in developing countries. The STa toxin is currently a candidate for the generation of a toxoid vaccine, and the determination of the receptor-ligand interaction would provide invaluable information for its design. In this study, we have prepared a set of three-dimensional models for the extracellular, ligand-binding domain of the human GC-C receptor (GCC-ECD), based on homology with the homologous natriuretic peptide receptors (NPRs). The modelled GCC-ECD monomer was similar to previously published models, and the models for the dimer enabled us to identify residues potentially involved in the oligomerization of the receptor, as well as the receptor-ligand interaction. Those residues are located within two regions of the GCC-ECD, from Ser75 to Ser127 and from Glu175 to Arg218. Previously published studies have shown that point mutations in the first region have an effect on ligand-binding, but the second region has not been investigated at all. Two residues that had been previously proposed as the ligand-binding residues were located within the hinge region between the two sub-domains of the GCC-ECD models. Additional candidate template structures were also obtained through threading, all belonging to the Type 1 periplasmic binding fold superfamily. Finally, we have taken the first step towards the setup of *in vitro* interaction studies by cloning the pro-sequences for the endogenous ligands of the GC-C receptor, guanylin and uroguanylin. A fragment of the GCC-ECD was also cloned, and it was successfully expressed in *E. coli*. Those results provide a basis for further interaction studies, both experimentally and using bioinformatics.

Abbreviations

BLAST	Basic Local Alignment Search Tool
BLASTp	protein BLAST
CDD	Conserved Domain Database
PSI-BLAST	Position Specific Iterative -BLAST
ETEC	enterotoxigenic *Escherichia coli*
ECD	Extracellular domain
HMM	Hidden Markov Model
GC	Guanylyl cyclase receptor
GC-C	Guanylate Cyclase C receptor
GST	Glutathione-S-transferase
NCBI	National Center for Biotechnology information
GCC-ECD	extracellular domain of the GC-C receptor
NPR-(A,B,C)	Natriuretic Peptide Receptor (A, B, or C).
NPR(A,B,C)-ECD	extracellular domain of the NPR-A, B or C receptor
NPRs	Natriuretic Peptide receptors
MAFFT	Multiple Alignment Fast Fourier Transform
PCR	Polymerase chain Reaction
PDB	Protein Data Bank
PBPD1	Type 1 periplasmic binding fold superfamily
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	SDS-Polyacrylamide Gel Electrophoresis
STa	Heat-stable enterotoxin
STh	STa toxin produced by human strains of ETEC
STp	STa toxin produced by porcine strains of ETEC

Contents

1 Introduction

1.1 Context of the study

1.1.1 Enterotoxigenic *Escherichia coli* mediates diarrhea through several toxins

Diarrheal diseases account for more than 3 million deaths per year amongst young children in the developing countries, with Enterotoxigenic *Escherichia coli* (ETEC) being the most commonly isolated pathogen (World Heath Organization, 2006). It is also the main cause of traveler's diarrhea (Navaneethan and Giannella, 2008; Okoh and Osode, 2008).

ETEC strains are a type of *E. coli* secreting toxins in the host's intestine, causing increased fluid excretion leading to diarrhea. They express colonization factors allowing their attachment to the epithelium in the small intestine where they release different exotoxins, the heat-stable (STa) and/or the heat-labile (LT) enterotoxins chief amongst them (Sack et al., 1975). The LT toxin is a 84 kDa, hexameric protein very similar to the cholera toxin (Spangler, 1992). The STa toxins are small peptides secreted by the pathogen, and are characterized by their resistance to the effects of high temperature (Sack, 1975). The STa toxin secreted by human ETEC strains, commonly known as STh, is a 19 amino-acid long peptide which is currently one of the targets for the development of a vaccine against ETEC-induced diarrhea (Aimoto Saburo et al., 1982; Walker et al., 2007). The STa toxins were shown to mediate increased fluid excretion via an augmentation of intracellular cGMP, and a membrane receptor, named the heat-stable enterotoxin receptor (STaR) was identified in the beginning of the 80s (Field et al., 1978; Frantz et al., 1984). It was found later that the guanylyl cyclase activity due to STa was located within the receptor itself, and it was renamed to guanylyl cyclase receptor C (GC-C, Schulz et al., 1990; de Sauvage et al., 1991). Evidence of another, GC-C independent pathway, exist in kidney epithelial cells (Sindićе et al., 2002; Carrithers et al., 2004).

1.1.2 The STa toxins are small, highly structured peptides

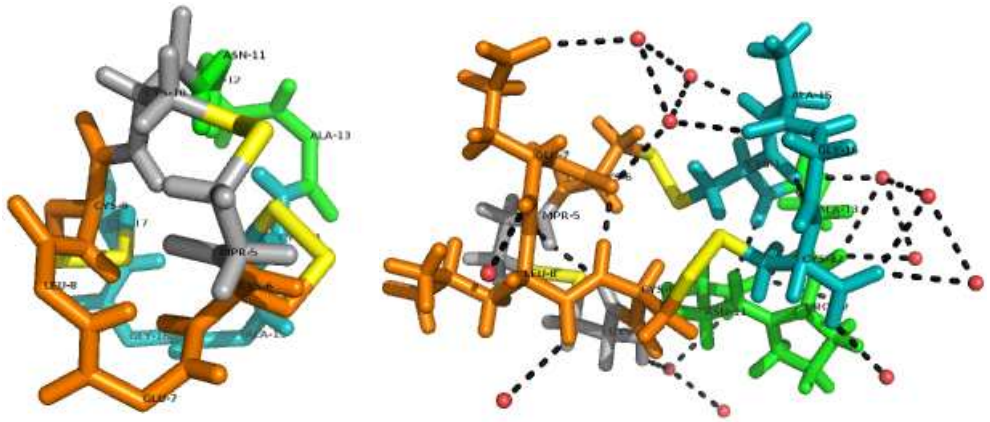
STa toxins are expressed as precusors

The STa toxins are encoded by three different *estA* alleles, with *estA1* coding for the STa secreted by the porcine strains of ETEC , named STp, and the others for STh (Guzman-Verduzco and Kupersztoch, 1989). All alleles have a 72 residue open reading frame and both toxins are synthetized as a pre-pro-precursor (Okamoto and Takahara, 1990; Rasheed et al., 1990). While the 19 amino-acid long pre-sequence is cleaved off after the initiation of translation, allowing translocation of the pro-precursor to the periplasm of the cell, the location for the cleavage of the pro-sequence remains unclear (Yamanaka et al., 1997; Yang et al., 1992). The sequences of the mature toxins are NSSNYCCELCCNPACTGCY for STh and NTFYCCELCCNPACAGCY for STp. The 14 C-terminal residues form the toxic domain of the STa peptides, its small size making it non-immunogenic and thus a difficult candidate for the generation of a toxoid[1] vaccine (Yoshimura Shoko et al., 1985). It contains three disulfide bridges required for biological activity, involving cysteines 5-10, 6-14 and 9-17 of STp and 6-11,7-15 and 10-18 of STh (Gariépy et al., 1987; Shimonishi et al., 1987). Their formation is supposed to occur inside the periplasmic space and involve the disulfide bond formation protein A (DsbA), although it has been suggested to happen outside the cell in a DsbA-independent fashion (Yamanaka et al., 1994; Batisson and Der Vartanian, 2000).

STa toxins form a spiral maintained by disulfide bonds

The reference structure for the STa toxins is the crystal structure of a synthetic analog of STp (PDB entry 1ETN), Mpr^5-STp(5-17), although the structure of the toxic domain of STa has been studied earlier by NMR spectroscopy (Ozaki et al., 1991; Gariépy et al., 1986). The analog is there described as a right-hand spiral composed of three β-turns, held together by the disulfide bonds mentionned earlier (Figure 1.1a). The segments composing the two first β-turns along the sequence form a cleft into which three water molecules are present, connecting Ala-15 to Cys-6 and Glu-7. Another is buried between the second and the third β-turn,

[1] toxin whose toxicity has been weakened but which retains its immunogenicity



(a) **(b)**

Figure 1.1: Structure of STp analog Mpr^5-STp(5-17) (Ozaki et al., 1991; PDB id 1ETN).(a) Stick representation of the right-hand spiral. The spiral is formed by a succession of three β-turns along the sequence. The residues forming them are colored in orange (Cys-6 to Cys-9), green (Asn-11 to Cys-14), and teal (Cys-14 to Cys-17). The disulfide bonds holding the structure are represented in yellow. **(b)** Stick representation of the STp analog including solvent and hydrogen bonds. The water molecules surrounding the analog are represented by red spheres and the hydrogen bonds by black dashes. This figure was generated using PyMol.

connecting Pro-12 to Cys-14.

More recently, the STp(5-17) fragment has been crystallized, showing the same global fold as the analog, even though the structural elements are described differently (Sato and Shimonishi, 2004). The structure of STh(6-18), which has been determined by NMR, confirms the crystallographic data obtained for the STp monomer (Matecko et al., 2009).

The STa toxins are bacterial enterotoxins similar to mammalian guanylin

The STa toxins produced by ETEC belong to a larger family of heat-stable enterotoxins produced by other pathogens. The first members of this family were purified in 1983 from *Klebsiella pneumoniae* and *Yersinia enterocolitica*, for which an additional one was found later (Klipstein et al., 1983; Takao et al., 1983; Yoshino et al., 1995). The toxins produced by different strains of *Vibrio cholerae* were purified over following years (Takao et al., 1985; Arita et al., 1986; Takeda et al., 1991). Similar toxins were purified from *Citrobacter freundii* and the enteroaggregative *Escherichia coli* (Guarino et al., 1987; Savarino et al., 1991). The STa from the

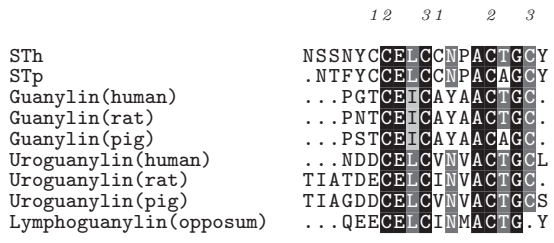


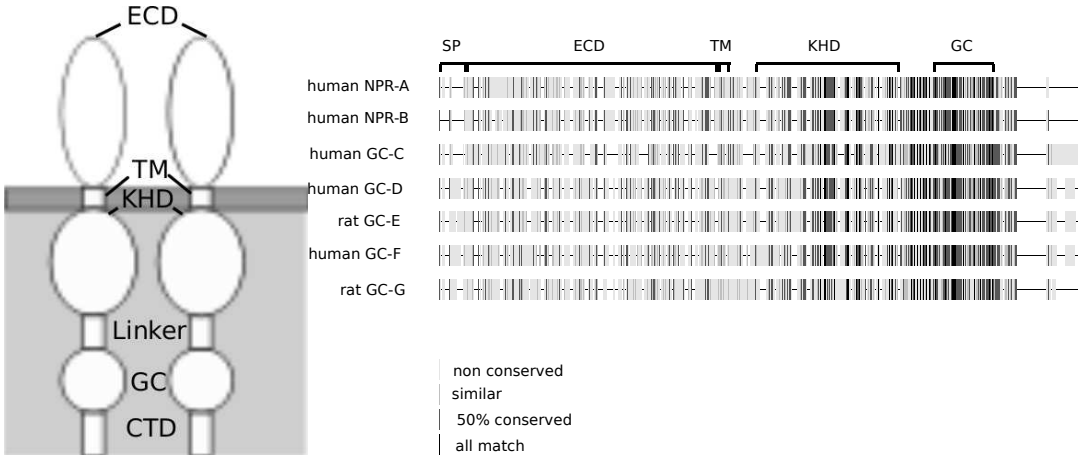
Figure 1.2: Alignment of STa, guanylin, and uroguanylin peptides. The UniProt sequences corresponding to the mature STa toxins were aligned with the endogenous guanylin and uroguanylin from human, pig, and rat (mouse sequences being identical to rat sequences). The cysteines are labelled 1, 2 or 3, according to the disulfide bond they form (disulfide bond 1 is only present in STa toxins). This figure was generated using the TeXshade package for latex.

latter, named EAST1, is thought to have a mechanism of action similar to that of the STa toxins (Savarino et al., 1993).

Interestingly, the mammalian receptor for STa has three other endogenous ligands in the intestine and kidney, guanylin, uroguanylin, and lymphoguanylin (Figure 1.2; Currie et al., 1992; Hamra et al., 1993; Forte et al., 1999). Together, these peptides form the guanylin peptide family.

As for STa, the endogenous guanylin peptides are small peptides expressed as pre-pro precursors. Guanylin is 15 amino acids long, and its precursor contains 115 residues (Wiegand et al., 1992a,b). It is organized into a pre- signal peptide of 19 residues, a pro- sequence, and the sequence for the mature guanylin at its C-terminus (Schulz et al., 1992; de Sauvage et al., 1992). The mature peptide contains 4 cysteine residues that are organized in two disulfide bonds between the cysteine pairs 4-12 and 7-15 (Cuthbert et al., 1994; Nokihara et al., 1997). The precursor for uroguanylin has a length of 112 residues, and the mature sequence is 16 amino-acids long (Hill et al., 1995; Li et al., 1997; Miyazato et al., 1996). It contains the four cysteines conserved with guanylin, forming the same disulfide bonds. The precursor for lymphoguanylin is 109 amino acids long, and the mature peptide consist of the 15 C-terminal residues (Forte et al., 1999). The C-terminal cysteine that was present in both guanylin and uroguanylin is replaced by a tyrosine in lymphoguanylin, and thus lymphoguanylin possess only one disulfide bond.

The structure of guanylin fragments of various sizes has been studied by NMR, revealing the existence of two topological forms termed A and B (PDB entries 1GNA and 1GNB, Skelton et al., 1994). The A-form has a fold highly similar to



(a) **(b)**

Figure 1.3: The Guanylyl cyclase receptor family. (a) Cartoon representation of the GC-C receptor. The membrane is shown in dark grey and the cytoplasm in light gray. (b) Multiple sequence alignment of mammalian guanylyl cyclase receptors. The UniProt sequence for the human GC-C was used as query to perform a BLASTp search against the UniProtKB/Swiss-Prot databases, and the 32 hits showing an E-value under 1e-50 were aligned using the MAFFT alignment tool. Although all hits were used for conservation calculations, only the human sequences are shown (with the exception of the second sequence for the human GC-B receptor, which is also hidden), or when non existing, the rat sequence. ***SP**: signal peptide, **ECD**: extracellular domain, **TM**: transmembrane helix, **KHD**: kinase homology domain, **GC**: guanylate cyclase catalytic domain, **CTD**: C-terminal domain.* The alignment figure was generated using the Texshade package for latex.

that of STa, whereas the B-form is described as an assembly of three turns in a left-handed spiral (in opposition to a right-handed spiral, which is the fold adopted by the A-form and STa). The presence of topological isomers was also determined for uroguanylin, and their structures solved by NMR (PDB entries 1UYA and 1UYB, Marx et al., 1998). For both peptides, only the A-form is active.

1.2 The Guanylyl Cyclase C receptor and its interaction with STa

The GC-C receptor is a member of the guanylyl-cyclase coupled receptors family (GCs), which counts to this day 6 other members (Figure 1.3b). The guanylyl cyclases A and B are receptors for the natriurectic peptides and are thus also known as the natriuretic peptides receptors A and B (NPR-A and NPR-B; Chinkers et al.,

1989; Chang et al., 1989). Three of the family members (GC-D, GC-E, and GC-F) are orphans receptors involved in the sensory system (Yang et al., 1995; Fülle et al., 1995). The last one is the murine renal guanylyl cyclase GC-G (Kuhn et al., 2004). All GC receptors are single-pass transmembrane proteins, with their extracellular domain (ECD) responsible for ligand-binding. The intracellular domain consists of a kinase homology domain (KHD) that is attached to the catalytic domain through a linker region. Some of the GCs contain a C-terminal domain (CTD). Despite this common organization, the sequence idenity between the GCs is low: local pairwise alignments between the full-length human GC-C sequence and the other human GC receptors show that the human receptor shares less than 35% of its sequence with NPR-A and B, and that only the intracellular domains of GC-D, E and F are similar, with about 45% sequence identity.

The human GC-C receptor is coded by the *gucy2c* gene, located on chromosome 12, and its open reading frame corresponds to a 1073 amino-acid long polypeptide for the human sequence (Mann et al., 1996). Transcription is regulated by the hepatocyte nuclear factor-4 (HNF-4), the homebox protein CDX2, and the Protein kinase C (Swenson et al., 1999; Park et al., 2000; Di Guglielmo et al., 2001; Roy et al., 2001). The mature receptor, with a theoritical molecular mass around 121 kDa, is expressed as N-glycosylated forms of 130 and 145 kDa, the latter being the active form found on the plasma membrane (Vaandrager et al., 1993; Ghanekar et al., 2004). Expression is localized to the brush border of epithelial cells in the small intestine as well as the crypts of the colon (de Jonge, 1975; Swenson et al., 1996).

1.2.1 cGMP mediated GC-C signalling leads to fluid secretion and cell proliferation in the intestine

The GC-C receptor catalyzes the synthesis of the cyclic guanosine monophosphate (cGMP), thus increasing its intracellular concentration and triggering several signalling cascades (for a review, see Basu et al., 2010). The main target of GC-C signalling is the cystic fibrosis transmembrane conductance receptor (CFTR), a chloride ion channel member of the ATP-binding cassette (ABC) transporter family. Activation of CFTR is achieved through several pathways illustrated in Figure

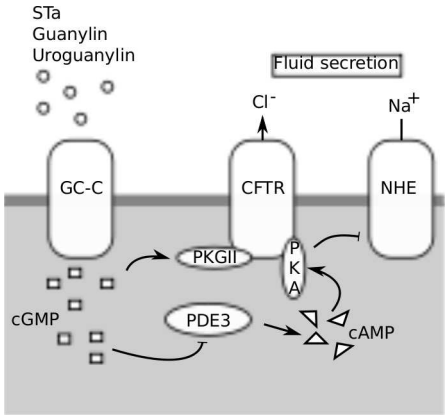


Figure 1.4: Fluid secretion mediated by the GC-C receptor. Synthesis of cGMP by the GC-C receptor upon its activation by STa, guanylin, or uroguanylin triggers several signalling cascades leading to fluid secretion (see text for details). ***STa:*** *heat-stable enterotoxin,* ***GC-C:*** *guanylyl cyclase C receptor,* ***cGMP:*** *cyclic guanosine monophosphate,* ***PKGII:*** *cGMP-dependent protein kinase II,* ***PDE3:*** *phosphodiesterase 3,* ***cAMP:*** *cyclic adenosine monophosphate,* ***PKA:*** *protein kinase A,* ***CFTR:*** *cystic fibrosis transmembrane conductance receptor,* ***NHE:*** *Na^+/H^+-exchanger.*

1.4. The increase of intracellular cGMP levels activates the cGMP-dependent protein kinase II (PKGII), which is responsible for the phosphorylation of CFTR (Markert et al., 1995; Vaandrager et al., 1997). Cyclic GMP also inhibits the Phosphodiesterase 3 (PDE3), resulting in the accumulation of cAMP inside the cell and the activation of protein kinase A (PKA). The activated PKA is able to activate CFTR but also to inhibit the Na^+/H^+-exchanger (NHE), thus preventing the uptake of Na^+ (Cheng et al., 1991).

Activation of the GC-C receptor has also an effect on cell proliferation, by the means of prolonging the cell cycle and via the activation of cyclic nucleotide-gated channels, leading to an anti-proliferating effect (Pitari et al., 2001, 2003).

1.2.2 GC-C binds STa through its extracellular domain

The GC-C receptor binds the STa toxins and the guanylin peptides via its extracellular domain, which can be expressed independently of the rest of the receptor (Nandi et al., 1996; Hasegawa et al., 1999c). It is a 407 amino acid long polypeptide (residues 23 to 430 of the full-length receptor) containing 8 cysteines residues conserved amongst the species (Hasegawa and Shimonishi, 2005). Those are organized into 4 disulfide bonds, between the cysteine pairs 7-94, 72-77, 101-128, and 179-226, respectively (numbering from the first amino acid of the GCC-ECD). It also contains 10 potential N-glycosylation sites, 7 of which conserved amongst the species (Ghanekar et al., 2004). Glycosylation is not required for ligand binding in itself, but it is essential for proper folding and activation of the receptor, in par-

ticular the conserved Asn172 and Asn379 sites (Hasegawa et al., 1999b; Ghanekar et al., 2004).

Separate expression of the extracellular and intracellular domains of GC-C indicates that it forms a dimer in the abscence of ligand and a trimer in its presence, even though the unit responsible for ligand binding is the dimer (Hasegawa et al., 1999c; Vijayachandra et al., 2000). The trimer had been previously observed for the full-length receptor (Vaandrager et al., 1994). The interaction between the receptor and its ligands has not been solved yet, but, following photoaffinity labeling and mutagenesis studies, it has been proposed that the binding sequence for the STa toxins is the segment spanning residues 387 to 393 ("SPTFTWK" for the human GC-C) along the sequence, near the C-terminus of the domain (Hasegawa et al., 1999a). Earlier mutagenesis studies, also on the pig GC-C, had proposed the Arg136 and Asp347 as the ligand-binding residues and the same C-terminal region as important for the conformation of the receptor (Wada et al., 1996). However, the Asp347 is not conserved with the human sequence, for which there is an Asn residue at that position.

GCC-ECD has a fold similar to that of the NPRs

The only member of the GC family for which three-dimensional structures are available is the NPR-A receptor, and two homology models for the ECD of the GC-C receptor have been presented based on it (van den Akker et al., 2000; Ogawa et al., 2004; Hasegawa and Shimonishi, 2005; Lauber et al., 2009). However, the NPR family counts another member that is not a guanylyl cyclase: the NPR clearance receptor, or NPR-C, for which several structures have also been published (He Xl et al., 2001; He et al., 2006). This receptor is a protein G coupled receptor that binds all natriuretic peptides, and its ECD shares about 20% of its sequence with that of GC-C. The available crystal structures for the ligand-bound extracellular domains of the NPR-A and C receptors reveal that, even though the sequence homology between them is low (less than 36%), their structures are remarkably similar (Figure 1.5a, He Xl et al., 2001; Ogawa et al., 2004; He et al., 2006). Each monomer is organized into two highly structured sub-domains, each of them centered around a β-sheet that is covered on each side by α-helices. The

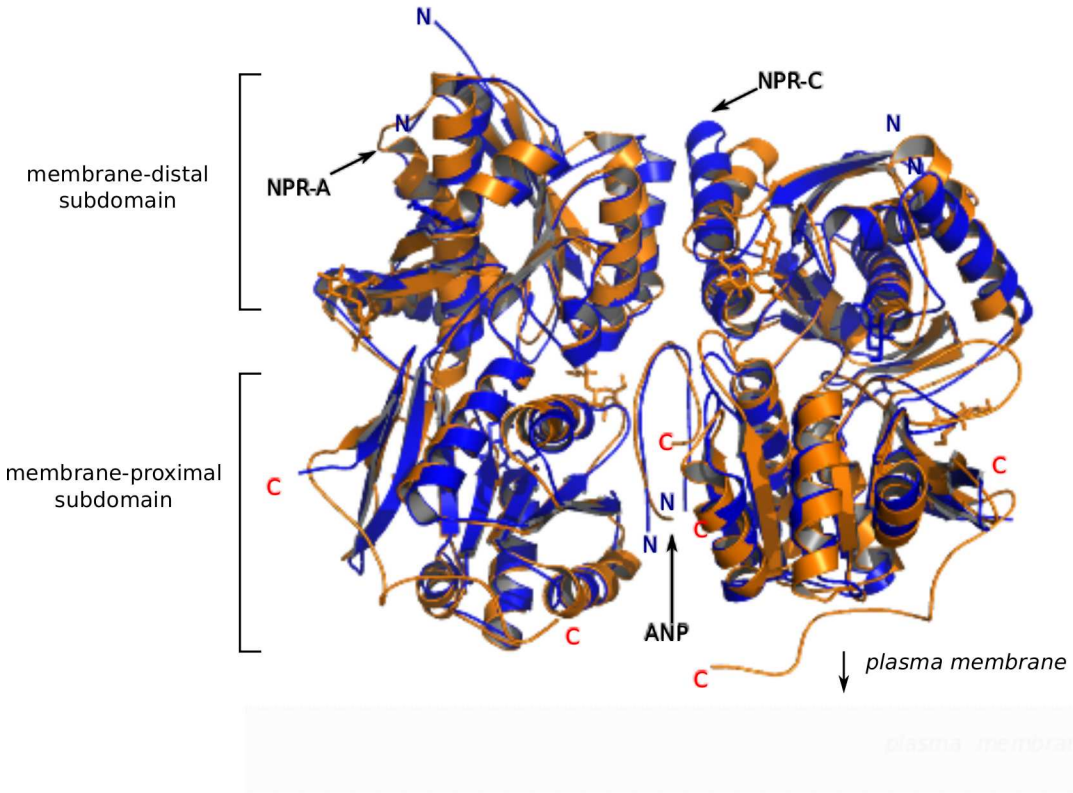


Figure 1.5: Structure of the ligand-bound extracellular domain of the natriuretic peptide receptors (NPRs). (a) Superposition of the crystal structures for the extracellular domains of the NPR-A receptor (in orange) bound to ANP (PDB entry 1T34) and the NPR-C receptor (in blue) bound to ANP, BNP, and CNP (PDB entries 1YK0, 1YK1, and 1JDP, respectively). **(b)** Secondary structure organization for the extracellular domains of the NPR-A and C receptors, according to their published crystal structures. The sequences are not aligned.

organization of the secondary structure elements along the sequence and within the structure is identical for both receptors, with the exception on an extra helix located on the outside of the membrane-proximal domain of the NPR-C receptor (Figure 1.5b). The sub-domains are interconnected by three cross-overs, and the expression of the putative membrane-proximal sub-domain of the GC-C receptor, as well as even even smaller portion of it, suggests that its ECD shares the same type of organization (Hidaka et al., 2002; Lauber et al., 2009). Interaction between the monomers is mediated by the membrane distal sub-domain, via the interaction of two helices located of the membrane-distal domain of each monomer (h4 and h6 along the sequence), forming a 2x2 helix bundle.

Several structures exists for each of the NPR-A and NPR-C receptors, corresponding to their unliganded and ligand-bound forms (Figure 1.6). The conformational change that occurs, upon ligand binding, within the NPR-A receptor, involves the relative position of each monmer, but the intramolecular structure remains mostly unchanged (Figure 1.6; Ogawa et al., 2004). On the contrary, the NPR-C monomers adopt different conformations when bound to a ligand: the angle formed between the helices h2 and h10, which illustrates that of the membrane-distal and -proximal subdomains, is augmented by more than 10°, which brings the ligand binding regions of the receptor that are located within the membrane-proximal domain closer (He Xl et al., 2001). The segment between b19 to h20 (from Leu279 to Pro285), which links the two subdomains, is described as a spring that is stretched upon ligand binding. This fragment interacts strongly, in the unliganded form of the receptor, with the N-linked glycan at Asn248, and the interaction is broken unpon ligand binding. This site is aligned in the sequence alignment with the Asn306 site of NPR-A, but their localization on the structures is, although near, different. The properties observed for Asn248 of NPR-C are not similary observed in NPR-A, for which the N-glycosylated residues are not involved in ligand-binding (Miyagi et al., 2000). However, the NPR-A glycosylation sites are conserved between the species, and it has been prosposed that they have a role in the proper folding of the receptor. The Asn13 and Asn180 sites are conserved with the NPR-B receptor. The Asn41 site of the NPR-C receptor is located within the missing segment in the structures.

The disulfide bonds of each receptor are situated at the exact same location,

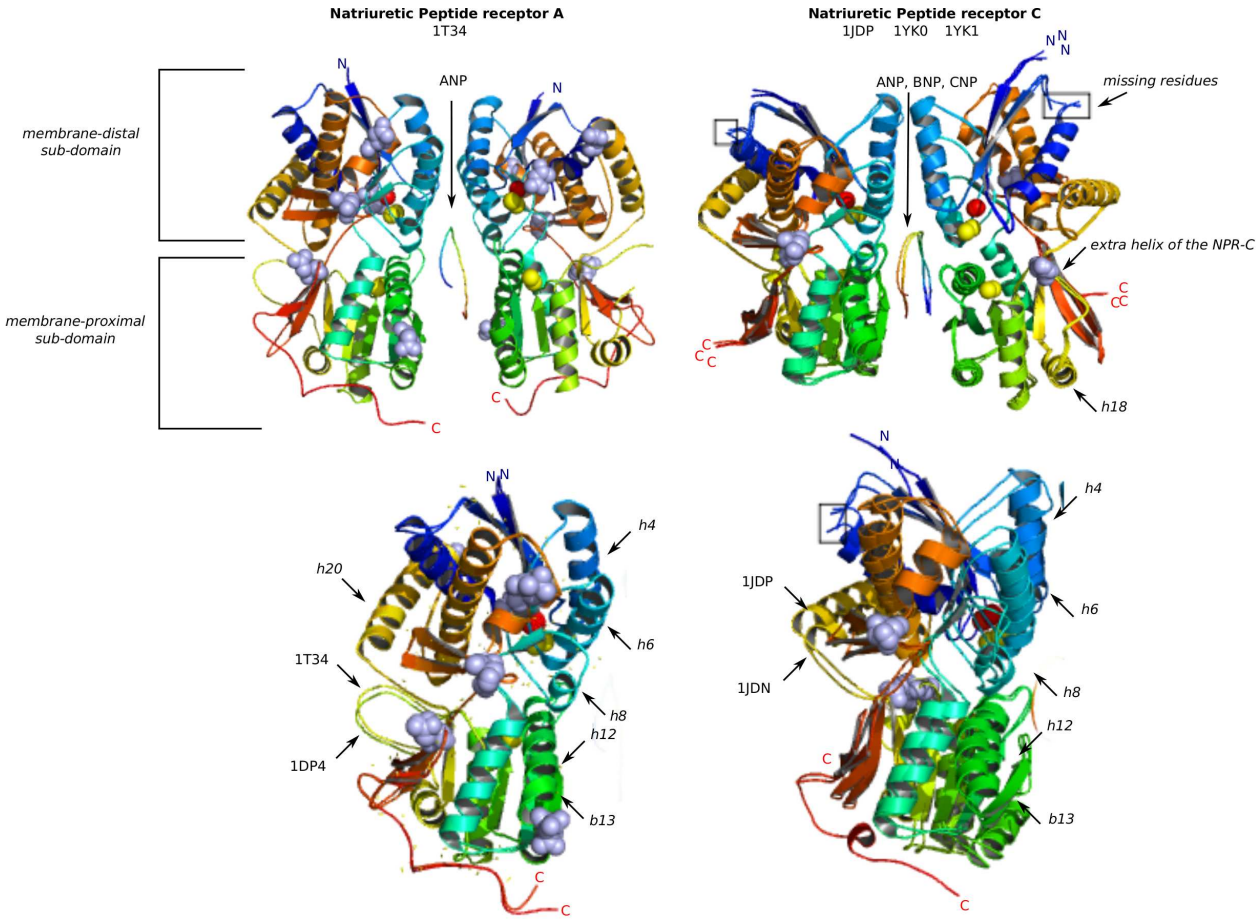


Figure 1.6: Crystal structures for the NPR-A and NPR-C receptors. The different structures for the natriuretic receptors (NPR) A and C are shown in cartoon representation, each chain colored as a rainbow from N-terminus to C-terminus. *Up:* Structures for the ligand-bound NPR-A (PDB entry 1T34) and NPR-C (PDB entries 1JDP, 1YK0, and 1YK1) receptors. Since they are very similar, the structures for the bound forms of the NPR-C are superposed *Down:* Structures for the unliganded NPR-A and NPR-C receptors (PDB entries 1DP4 and 1JDN, monomers). They are superposed with the corresponding ligand-bound form (PDB entries 1T34 and 1JDP, respectively) for easier vizualisation. The sulphur atomes from the cysteine residues are shown as yellow spheres, chlorides as green ones, and the potential N-glycosylation sites as gray ones. Secondary structure elements of interest (see text) are indicated***ANP:*** *Atrial Natriuretic peptide,* ***BNP:*** *Brain Natriuretic peptide,* ***CNP:*** *Natriuretic peptide type C.* Figure generated using PyMol.

behind the ligand-binding helices h6 and h12 which they seem to lock the position of. The structures also reveal a bound chloride ion, located in the viccinity of the disulfide bond from the membrane-distal domain, that has been shown to be necessary for receptor activity (van den Akker et al., 2000). This bond is the one that, within the GC-C receptor, is separated into two different disulfide bridges (Hasegawa and Shimonishi, 2005).

The GC-C receptor seems to bind ligands in a different fashion from that of the NPRs

The ligand binding site of the NPR receptors is located between the monomers, where different subsets of amino-acids from each monomer (sites I and II, involving mostly helix h8 and the region from b11 to h14, see Figure 1.5b) bind a different part of the natriuretic peptides (He et al., 2006). This data is in contradiction with the hypothesis for the ligand-binding site of GC-C according to which the ligand-binding sequence of the GC-C receptor involves C-terminal residues, the latter being further supported by the observed lignad-binding capabilities of the GCC-ECD fragments (Hasegawa et al., 1999a; Hidaka et al., 2002; Lauber et al., 2009). The interaction between the binding sequence and the ligand is proposed to mimick the interaction that takes place between guanylin and its prosequence, which form a β-hairpin (Lauber et al., 2003). Titration of the complex between STa and the proximal domain of GC-C (miniGC-C) indicates a 1:1 stoichiometry, which is also in contracticton with that of the NPR receptors which bind one molecule of ligand per dimer (Lauber et al., 2009; He et al., 2006).

STa and the guanylin peptides bind to GC-C with different affinities

The STa toxins possess high affinity for the GC-C receptor, with values for the dissociation constant between 0.4 x 10^{-11} M and 2.2 x 10^{-9} M, and the presence of several affinity sites has been reported (Wada et al., 1994; Deshmane et al., 1995). Values for the dissociation constant for the extracellular domain of GC-C range between 4.0 x 10^{-10} M and 7.3 x 10^{-8} M, and the miniGC-C shows an affinity somewhat 10-fold weaker than that of the full receptor, with K_D values between 4.5 x 10^{-9} M and 7.2 x 10^{-9} M (Hasegawa et al., 1999c; Lauber et al., 2009).

Guanylin and uroguanylin, which compete with STa, were shown to inhibit the binding of radio-labeled STa in a similar pH dependent fashion: at pH 5, guanylin has a K_i of 10^{-7} M, against 10^{-9} for uroguanylin. The effect is reversed at pH 8, with K_i values of 10^{-9} and 10^{-8} (Hamra et al., 1997).

The binding of ligands to the extracellular domain of the GC-C receptor is thought to induce a conformational change within the receptor leading to the activation of the catalytic domain, but the nature of this change is unknown. Data obtained for the extracellular domains of the NPR-A and NPR-C receptors show that the dimer undergoes either a twist motion, or a translation of its membrane-proximal domains (He Xl et al., 2001; Ogawa et al., 2004). A recent study on the juxtamembrane region of the GC-A receptor suggests that relative orientation is more crucial than proximity, although the rotation mechanism they propose is different from the one inferred by the crystallographic data (Parat et al., 2010).

2 Aims of the Study

One current strategy to counter diarrhea induced by enterotoxigenic *Escherichia coli* (ETEC) involves the development of a toxoid vaccine based on its heat-stable enterotoxin (STa), as is it a key virulent factor (Taxt et al., 2010). The STa toxin binds to the guanylyl cyclase C receptor (GC-C), and detailed knowledge on this interaction would be a great asset for the design of the vaccine. In order to study this interaction, it was chosen to use an experimental approach combined with bioinformatics.

The primary aim was to establish an *in vitro* system to study the interaction. The subaims were:

1. to clone and to express, in *Escherichia coli*, the GCC-ECD and its endogenous ligands guanylin and uroguanylin.

2. to conduct pilot binding experiments such as GST-pulldown assays or surface plasmmon resonance spectroscopy experiments (Biacore).

The main bioinformatical aim was to investigate the residues of the GCC-ECD that may be involved in ligand binding, but also in the oligomerization of the receptor.

1. For this puropse, a goal was to generate a homology model for the extracellular domain of the GC-C receptor (GCC-ECD). This included the aim to construct a high quality multiple sequence alignment and the indentification of different template structures.

2. The development procedure for the vaccine involves work with several model organisms, such as the mouse or the pig. In order to assess the suitability of those organisms for this purpose, an aim was to use the obtained homology models for the GCC-ECD , as well as the sequence alignments, to evaluate, amongst those organism, potential differences in ligand-binding.

3 Materials

3.1 Software

3.1.1 Databases and database search programs

Protein sequences were obtained from the UniProt and UniRef90 protein sequence databases, the latter containing clustered sets of sequences sharing at least 90% identity (Magrane and Consortium, 2011; Suzek et al., 2007). Structures were obtained from the Protein Data Bank (Berman, 2000; www.pdb.org).

Database searches for sequence similarity were carried out using the BLAST and HMMER program suites (Altschul et al., 1990; www.hmmer.org). BLAST, for "Basic local alignment search tool", detects the sequences segments of a database that produce alignments of high statistical significance with the query. In this study, protein sequences were compared by using the BLASTp program, either from the NCBI website (for National Center for Biotechnology Information, www.ncbi.nlm.nih.gov), via the command-line version of the program suite, `blastall`, or using its more sensitive version, PSI-BLAST (for Position-Specific Iterated). The latter uses the results of an initial BLASTp search to build a position-specific scoring matrix (PSSM), or profile, using the multiple alignment of the returned sequences. Profiles are statistical descriptions of a multiple alignment or even one sequence which gives, for each column of the alignment, the propensity of the amino acid that is most represented. The profile is used as query in the next iteration of the search, thus giving the possibility to find more distantly related homologs to the initial query.

The HMMER package provides a group of programs that also makes use of profiles for sequence similarity searches (Eddy, 1998; www.hmmer.org). In this case, the probabilistic model used to construct the profiles is the Hidden Markov Model, and the profile can be used to search sequences databases as well as profile HMM databases, in a non-iterative or iterative fashion (Krogh et al., 1994). HMMER 2.3.2 was used in this study, for which the `hmmbuild` program is used to

build the profiles HMM. Parameters for the profile are calculated separately by the `hmmcalibrate` program, and the database search is performed by the `hmmsearch` program.

Database search for fold recognition, was performed using the pGenThreader program, from the PSIPRED web-server (Lobley et al., 2009). This method is based on the comparison of PSSM profiles between the target sequence and template structures. The profile for the target is obtained through PSI-BLAST, after 8 iterations.

3.1.2 Sequence alignment programs

Several multiple sequence aligment tools were used in this study. The Multiple Alignment Fast Fourier Transform (MAFFT) tool was used a the default multiple sequence alignment program (Katoh et al., 2005). In MAFFT, amino-acid sequences are converted into sequences of vectors, which describe each residue in terms of volume and polarity (Katoh, 2002). The similarity between such sequences is represented by the correlation between them, and the discrete fourier transform (corresponding to the fast fourier transform algorithm) is used to simplify its expression.

Structure-based multiple sequence alignment was performed using the EXPRESSO server, where a set of sequences is submitted to the server, which assigns, via a BLASTp search against the PDB database, structural templates to the sequences whenever possible (Armougom et al., 2006).

Sequence-sequence and sequence-structure alignments were also performed using the diverse alignment commands of the MODELLER program, which 9.9 version was used in this study (Sali and Blundell, 1993).

3.1.3 Other programs related to sequence alignments

Highly similar sequences provide very similar sequence information, and can thus be considered redundant. Removal of this redundancy from a set of sequences was performed by using the CD-HIT program from its web-interface (Li and Godzik, 2006; Huang et al., 2010).

Another useful information, when constructing sequence alignments, is to observe the phylogenetic distribution of the sequences composing the alignment. Phylogenetic trees based on sequence aligments were obtained using the MrBayes 3 program, which uses the Bayes probabilistic theorem to infer phylogeny (Ronquist and Huelsenbeck, 2003; Huelsenbeck and Ronquist, 2001).

3.1.4 Secondary structure prediction: PSIPRED

The PSIPRED secondary structure prediction method was used to predict the secondary structure of the extracellular domain of the GC-C receptor (Jones, 1999). This method uses, as it improves the prediction, sequence information provided by sequences related to the query, and more specifically the information provided by PSI-BLAST (Zvelebil, 1987; Altschul et al., 1997). Neural networks are used to process the information.

3.1.5 Homology modelling using MODELLER

Homology modelling was performed using MODELLER v9.9, which generates a three-dimensional model for a protein, given spacial restraints (Sali and Blundell, 1993). One of those restraints is the experimentally determined structure for an homologous protein (template), but additional data from other sources can be used as constraints. More precisely, it is the alignment between the template structure(s) and the target sequence that is used as input to the progam, the output being one or several models. Several alignment tools are included in MODELLER in order to align the target with the template, but also perform the structure-structure alignment of several templates, which can then be used togheter for doing multi-template modelling.

3.1.6 Model evaluation

PROCHECK calculates, from the coordinates of a structure, its stereochmichal parameters: the (phi,psi) angles, peptide bond planarity, bond lengths, bond angles, hydrogen-bond geometry, and side-chain conformations. The values for these

parameters can then be compared with that of known proteins structures (Morris et al., 1992).

3.2 Biological and chemical materials

3.2.1 Bacterial strains and DNA material

Table 3.1: *Escherichia coli* strains

Strain	Genotype
TOP10 One shot	F– *mcr*A Δ(*mrr* – *hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G
Origami B	F- *omp*T *hsd*S_B(r_B^- m_B^-) *gal dcm lacY*1 *ahp*C (DE3) *gor*522:: Tn10 *trx*B (Kan^R, Tet^R)

Two *Escherichia coli* strains were used in study, the TOP10 One Shot cells from Invitrogen and the Origami B cells from Novagen, for which the genotypes are presented in Table 3.1. The TOP10 cells were used for plasmid preparation and the Origami cells for protein expression.
The template DNAs used in this study, i.e. the human sequences for pro-guanylin, pro-uroguanylin, and the GC-C receptor, were obtained as recombinant pCR4-TOPO plasmids from Invitrogen.

Fragments of interest were cloned into the pSXG expression vector, a mutated version of the pGEX-2TK vector for which the polylinker was replaced with that of the pGAD424 vector (Figure 3.1; (Ragvin et al., 2004)). The pSXG vector enables the construction of glutathione-S-transferase fusion proteins which are inducible by IPTG: the multiple-cloning site is placed in 3' of the gene coding for the GST, which expression is directed by the IPTG-inducible tac-promoter. To further ensure that the construct will be expressed only upon induction, the vector contains the gene coding for the LacI repressor of the Lac operon.

The different primers used in this study were obtained from Sigma (see Table 3.2), and the nucleotides (dNTPs) were purchased from TaKaRa. The GeneRuler DNA ladder was obtained from Fermentas.

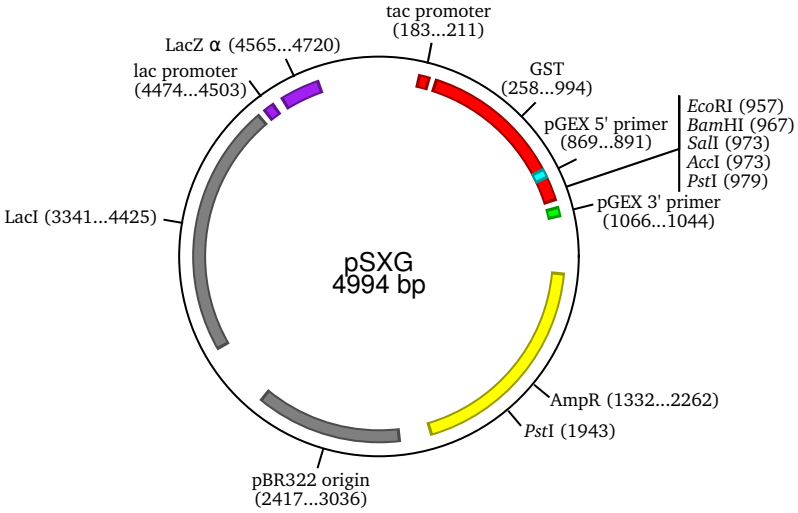


Figure 3.1: Graphical map for the pSXG vector. The pSXG vector is an *E. coli* expression vector for the inducible expression of glutathione-S-transferase fusion proteins using the Lac operon. The tac-promoter is induced by the addtion of IPTG in the culture medium, and the presence of the gene coding for the repressor of the Lac operon, LacI, prevents the expression of the trget protein in the absence of IPTG. For selection purposes, the vector contains the genes coding for resistance to ampicilin and the α-segment of LacZ.

Table 3.2: Primers

Primer name	Sequence	Restriction Site
Pro-guanylin fwd *	5'-GCCTTGGCAGAATTCGTCACCGTGCAG-3'	*Eco*RI
Pro-guanylin rev *	5'-TGGGCCCATGGATCCTTAGCATCCGGT-3'	*Bam*HI
Pro-uroguanylin fwd *	5'-GCAGAGCACAGAATTCGTCTACATCCAGTACC-3'	*Eco*RI
Pro-uroguanylin rev *	5'-TGGGCGGATCCTACCCAGGGCTATCTCA-3'	*Bam*HI
GCC-ECD fwd	5'-TGGCTGTCCGGATCCTCCCAGGTGAGTCAGAAC-3'	*Bam*HI
GCC-ECD rev/ miniGC-C D (rev)	5'-GGAGGAGGATCCTTACTGAGGGCCCCGGCCTGTAATATC-3'	*Bam*HI
miniGC-C A (fwd)	5'-GGAGGAGAATTCTCTCCAGCTAGAAAGTTGATGTACTTC-3'	*Eco*RI
miniGC-C B (rev)	5'-CACCATGGTGTCTCCAGGAGCCAGCGTCAGAACAAGGACATTTTTCATATAGTC-3'	None
miniGC-C C (fwd)	5'-CTGACGCTGGCTCCTGGAGACACCATGGTGCTTCTGTATACCTCTGTG-3'	None
pGEX 5' (fwd)	5'-GGGCTGGCAAGCCACGTTTGGTG-3'	None
pGEX 3' (rev)	5'-CCGGGAGCTGCATGTGTCAGAGG-3'	None

* Primers designed by Arne M. Taxt

3.2.2 Proteins

The enzymes used in this study, i.e. the Taq DNA polymerase (ExTaq), *Eco*RI and *Bam*HI endonucleases, calf intestine alkaline phosphatase, and T4 DNA ligase, as well as their corresponding buffers, were purchased from TaKaRa.

The anti-GST polyclonal rabbit antibody was obtained from Sigma and the anti-rabbit, HRP-coupled antibody from GE Healthcare.

Bovine serum albumine (BSA) was obtained from TaKaRa. The PageRuler protein molecular weight marker was obtained from Fermentas.

3.2.3 Common chemicals and Solutions

Table 3.3: Buffers, culture media, and other solutions

Name	Composition	pH
Phosphate buffered saline (PBS)	137 mM NaCl; 2.7 mM KCl; 4.3 mM Na_2HPO_4; 1.47 mM KH_2PO_4	7.4
PBS-T	PBS with 0.05 % (v/v) Tween 20	7.4
Ethylene-diamine-tetra-acetate (EDTA)	500 mM EDTA	8.0
Tris-acetate-EDTA (TAE)	40 mM Tris; 20 mM acetic acid; 1 mM EDTA	
Electrophoresis buffer	25 mM Tris-HCl pH 8.5; 1 % (w/v) SDS	
Transfert buffer	3.03 g/l Tris; 14.4 g/l glycine; 20 % (v/v) methanol	
LB medium	Tryptone 10 g/l; Yeast extract 5 g/l; NaCl 10 g/l; dH_20	
LB-agar	Tryptone 10 g/l; Yeast extract 5 g/l; NaCl 10 g/l; Agar 15 g/l; dH_20	

Most of the chemicals used in this study were obtained from either Merck or Sigma, with the execption of Agarose, which was purchased from Invitrogen, Sodium Docecyl Sulfate (SDS) from Fluka, Trizma-base from Prolabo, and the 6x loading buffer for DNA, which was provided by TaKaRa. Ethanol and Isopropanol were obtained from Arcus. Buffers and their composition are described in Table 3.3.

4 Methods

4.1 Construction of sequence alignments

The critical determinant for the quality of a three-dimensional model built by homology is the sequence alignment between the target sequence to be modelled and the structure(s) that will serve as template for the modelling. When the template and the target possess a very high sequence identity, they are easy to align and thus no other sequence information is needed. However, when the sequence identity is low, it is necessary to include in the alignment other sequences that will provide additional information, leading to a better alignment of the target and template sequences. The additional sequences, which are homologous to the target, should not be too similar to each other so as to avoid redundancy. It is therefore necessary to gather sequences that are homologous to the target but that are as diverse from each other as possible, referred to as the remote homologs to the target.

4.1.1 Gathering of remote homologs

The sequence corresponding to the extracellular domain of the GC-C receptor (UniProt accession number P25092, residues 24 to 430) was used as query for a preliminary BLASTp search against the UniProtKB/SwissProt database from the UniProt web-server, using an E-value threshold of 1. A PSI-BLAST search was also carried out from the NCBI web-site against the SwissProt database, using the same query sequence (GCC-ECD) and the default parameters.

The gathering of remote homologs for the GCC-ECD was performed according to the strategy presented in Figure 4.1. The sequence corresponding to full-length human GC-C receptor was used as query to gather homologous protein sequences using the BLASTp alignment tool (Altschul et al., 1997). The UniRef90 database, which contains representatives for sequences groups sharing above 90% sequence identity, was chosen for this search (Suzek et al., 2007). It was also chosen to

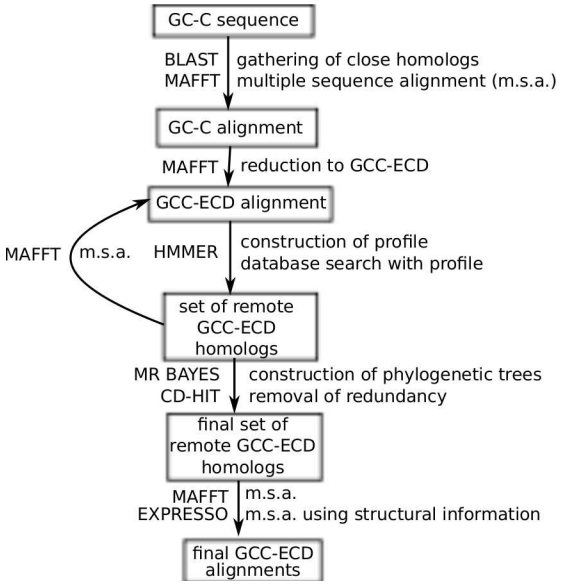


Figure 4.1: Strategy for the construction of a high quality sequence alignment. From the sequence of the full length GC-C receptor, close homologs were gathered in order to build an accurate profile for its extracellular domain (GCC-ECD). This profile was used to gather sub-sequences homologous to the GCC-ECD, which were thereafter aligned. Sequences containing important deletions were removed from the alignment, which was used to build another profile. This procedure was repeated until no more homologs of interest were gathered this way. Sequences redundant above 90% were removed from the set of gathered homologs, provided that they were not associated to a structure, belonging to one of the chosen representative organisms, or a member of an under-represented phylogenetic group. This final set of sequences was aligned using two different alignment programs, giving the final alignments.

limit the search to the Coelomata taxonomic group, as to exclude plants and fungi. Version 2.2.20 of the `blastall` program was used, with default values for all parameters and an E-value threshold of 0.0. Sequences obtained were aligned using the MAFFT multiple alignment program (Katoh et al., 2005). Sequences containing either prominent deletions or dissimilarities within the region of the alignment corresponding to the extracellular domain of the GC-C receptor (GCC-ECD) were removed from the alignment. The alignment was then edited so as to contains only the portion corresponding to the GCC-ECD and the sub-sequences aligned to it.

This alignment, now a GCC-ECD alignment, was used to build a Hidden Markov Model (HMM) profile using the `hmmbuild` program of the HMMER 2.3.2 package (Eddy, 1998). After calibration (`hmmcalibrate`), the profile was used to perform a database search within the UniRef90-Coelomata database, with the `hmmsearch` program, using an E-value threshold of 1e-100. From the search results, the subsequences corresponding to the portion of the GCC-ECD with 10 additional residues in the N- and C-terminal were harvested using a script provided by Dr. Pål Puntervoll. As described above, the obtained sequences were aligned using MAFFT and sequences were removed according to the same criteria as for the

initail alignment. This new alignement was used to build a new profile, and the procedure was repeated until no more new sequences were harvested. The last HMM profile for which new sequences were gathered from the UniRef90 database, limited to the Coelomata taxonomic group, was used to search the corresponding portion of the UniProt database. This was done in order to obtain all the sequences that could be harvested with this profile, not just the representatives of identity clusters.

4.1.2 Determination of the final set of sequences

In order to obtain a highly informative but non-redundant sequence alignment, several analyses were performed on the sequence set obtained through the profile HMM search procedure: sequences for which structures are available were determined, redundancy above 90% was removed from the set, and phylogenetic trees were built for both the original set and the clustered one. Removal of redundancy above 90% was performed using the CD-HIT program (Huang et al., 2010). The phylogenetic trees were built using the version 3.1.1 of the MrBayes program (Ronquist and Huelsenbeck, 2003).

The information gathered by those analyses was used to remove redundant sequences from the original set of homologs for which no structures were associated and that did not belong to an underepresented phylognetic cluster or an organism of interest.

4.1.3 Multiple sequence alignments

The final, non-redundant, set of sequences obtained as described above was aligned using the MAFFT and EXPRESSO (3D-Coffee) alignment tools (See *Materials*, Section 3.1.2; Katoh et al., 2005; Armougom et al., 2006). The latter takes into account the structural information for the sequences to which structures are associated, in this case the sequences corresponding to the rat NPR-A and the human NPR-C. The sequences for the extracellular domains of the NPR-A and NPR-C receptors were also aligned with that of the GC-C receptor using the alignment programs included in MODELLER.

4.2 Homology Modelling

The structures for the NPR-A and NPR-C receptors were used to model the extracellular domain of the GC-C receptor (GCC-ECD). The structure corresponding to the unbound NPR-A receptor (PDB entry 1DP4), contains a dimer which organizzation is due to crystal packing, so only the monomer was modelled based on it (Ogawa et al., 2004). The structures for the bound forms of the NPR-C receptor (PDB codes 1JDP, 1YK0, and 1YK1) were used as a group (multi-template modelling). For the structures containing dimers (with the exception of 1DP4), modelling was performed on both chains simultaneously (without symmetry constraints) and separately. Disulfide bonds of the GCC-ECD, as described by Hasegawa and Shimonishi (2005), were added as constraints for all models. The template(s) and the target were aligned by the MAFFT tool, using the final set of sequences gathered by HMM profile search. The evaluation of the models was done by submitting them to the PDBsum database, where different analyses were performed, including the evaluation of their stereochemical parameters (Laskowski, 2001).

4.3 Cloning of the GCC-ECD, pro-guanylin and pro-uroguanylin

4.3.1 Preparation of inserts by site-directed mutagenesis

The DNAs encoding pro-guanylin, pro-uroguanylin and the complete GC-C receptor were used as templates to perform site-directed mutagenesis via Polymerase Chain Reaction (PCR), under the following conditions:

Reaction mix		cycle conditions	
DNA template	100-200 ng	initial denaturation	94 °C, 30 sec.
10x ExTaq buffer	5μl	denaturation	94 °C, 10 sec.
dNTPs	2.5 mM	annealing	55 °C, 10 sec.
Primers	0.2 μM	extension	72 °C, 30 sec.
ExTaq DNA polymerase	0.025 U/μl	final extension	72 °C, 1 min.
dH_2O	up to 50 μl	number of cycles:	25

Pro-guanylin and Pro-uroguanylin

The DNA fragments coding for the human pro-guanylin and pro-uroguanylin were amplified using the primers mentionned in Table 3.2, introducing an *Eco*RI restriction site in 5' of the coding sequence and a *Bam*HI site in 3'. After the PCR, the reaction mixtures were subjected to agarose gel electrophoresis on a 1% agarose gel containing 3μg/ml of ethidium bromide (EtBr) in TAE buffer, for the purposes of analysis and purification. Each mixture was loaded as two samples of 40μl and 5μl, alongside 500 ng of 100 bp DNA ladder. The gel was run at 80V for 80 minutes, and the bands corresponding to the expected PCR products (315 bp for pro-guanylin and 302 bp for pro-uroguanylin) for the 40μl samples were purified from it using the QIAquick gel extraction kit and according to the manufacturer's instructions. The purified PCR products were digested by the *Eco*RI and *Bam*HI endonucleases, according to the following reaction mixture:

Reagent	concentration/volume
PCR product	20 ng/μl
Buffer 10H	5.5 μl
Bovine serum albumine (BSA)	0.2μg/μl
*Eco*RI	2 U/μl
*Bam*HI	3 U/μl
dH_2O	up to 50 μl

The digestion was carried out for 1h at 37 °C, and the enzymes were inactivated at 80 °C for 20 minutes. The digested inserts were stored at -2 0°C.

GCC-ECD

The extracellular domain of the GC-C receptor (residues 24 to 430) was amplified by PCR according to the condtions presented above. The primer couple used for the reaction introduced *Bam*HI restriction sites on each side of the receptor sequence, along with a stop-codon for the reverse primer. The PCR amplified fragment was, as in the case of pro-guanylin and pro-uroguanylin, subjected to agarose gel electrophoresis and gel extraction, using the same conditions. The purified PCR product, which has an expected size of 1251 bp, was digested by *Bam*HI according to the digestion reaction presented for pro-guanylin and pro-uroguanylin .

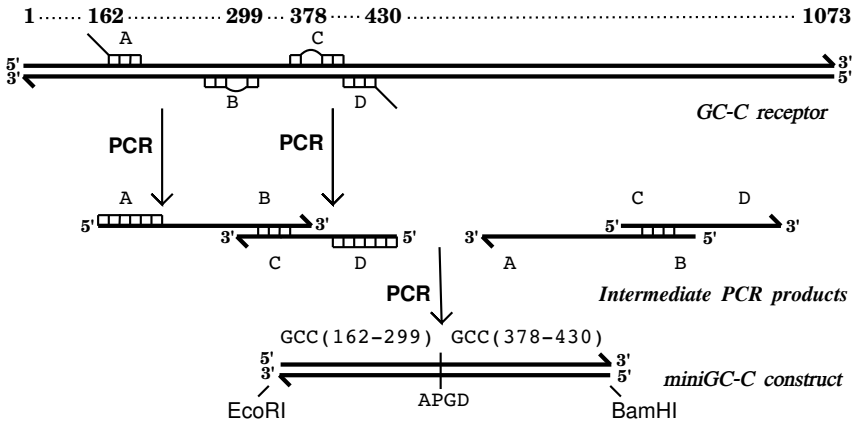


Figure 4.2: Synthesis of the miniGC-C insert. From the coding sequence for the GC-C receptor, 2 fragments are amplified by PCR using the primers couples A/B and C/D, respectively. Those fragments are used in an additionnal PCR reaction using primers A and D but also each other, leading to the miniGC-C construct.

miniGC-C

In order to reproduce the miniGC-C fragment obtained by Lauber et al., two portions of the GC-C receptor, from residues 142 to 299 and 378 to 430, were amplified by PCR (Lauber et al., 2009). The first fragment was amplified using primers A and B, and the second using primers C and D (see Table 3.2). Each PCR product (with expected sizes of 444bp for the the first fragment and 189bp for the second) was purified by gel extraction Primer A introduces a *Eco*RI restriction site in 5' of the first fragment, and primer B introduces the Ala-Pro-Gly-Asp coding sequence in 3'. Primer C introduces the same sequence but in 5' of the second fragment, and primer D a *Bam*HI restriction site in 3'. Due to the common Ala-Pro-Gly-Asp coding sequence, the PCR products for the fragments are complementary, which allows their combination during a third PCR reaction (Figure 4.2). The final PCR product was then purified and double-digested as was done for the guanylin and uroguanylin inserts.

4.3.2 Cloning into the pSXG vector

The empty pSXG vector was digested by either *Bam*HI or *Eco*RI/*Bam*HI, according to the conditions presented for the inserts (20 ng/μl of DNA, 2 U/μl for *Eco*RI, 3 U/μl for *Bam*HI, total reaction volume of 50 μl). Dephosphorylation was then

carried out on 44 μl of the reaction mixture, using 0.4 U/μl of calf intestine alkaline phosphatase (CiAP) in its associated reaction buffer, for a total reaction volume of 50 μl. The rest was kept for analysis on agarose gel. Incubation conditions for the dephosphorylation were 1h at 37 °C, and the enzyme was inactivated for 15 minutes at 75 °C. The dephosphorylated plasmid was purified by phenol/chloroform extraction. The insert and dephosphorylated vector were ligated overnight at 16 °C using 0.5 U/μl of T4-DNA ligase.

4.3.3 Transformation, plasmid purification, and analysis

The ligated products were transformed into TOP 10 One shot cells as described by the manufacturer (Invitrogen), but using LB medium instead of SOC medium. The transformed cells were spread on LB-agar plates containing 100 μg/ml ampicilin and grown overnight at 37 °C. Single colonies were used to inoculate 5ml of LB medium (containing 100 μg/ml ampicilin) and the cultures were grown overnight at 37 °C, 250 rpm. The plasmids were purified using the QIAspin mini-prep kit, according to the manufacturer's instructions. The presence of the inserts with the expected size was assessed by restriction digestion analysis of the pSXG constructs using the *Eco*RI and *Bam*HI endonucleases for the pro-guanylin, pro-uroguanylin, and miniGCC constructs, and *Bam*HI for the GST-GCCECD. The reaction mix and incubation conditions were identical to what was done for the digestion of the empty pSXG. The pSXG constructs were also subjected to sequencing which was performed by the Sequencing Facility of the Department of Molecular Biology (MBI), University of Bergen. The sequencing reaction was prepared according to the instructions from the facility, which uses the Big-Dye version 3.1 DNA sequencing kit from Applied Biosystems:

Reaction mix	
DNA template	[a]
Sequencing buffer	1 μl
Big-Dye v3.1	1 μl
Primer[b]	3,2 pmol
dH_2O	up to 50μl

[a] the amount of DNA depends on the size of the template

[b] the pGEX 5' and 3' were used in two sequencing reactions for each template

cycle conditions	
initial denaturation	96 °C, 5 min.
cycle phase 1	96 °C, 10 sec.
cycle phase 2	50 °C, 5 sec.
cycle phase 3	60 °C, 4 min.
number of cycles:	25

4.4 Expression of GST-tagged miniGC-C

The pSXG-miniGCC construct, as well as the empty pSXG vector, were transformed into Origami cells, spread on LB-agar plates containing 100 μg/ml ampicillin, 15 μg/ml kanamycin and 12.5 μg/ml tetracyclin, and grown at 37 °C for at least 24 hours. Precultures of 5 ml of LB-medium (also containing 100 μg/ml ampicillin, 15 μg/ml kanamycin and 12.5 μg/ml tetracyclin) were made from a single colony and grown overnight at 37 °C, 250 rpm. The precultures were used to inoculate 50 ml of LB medium (containing the same anitbiotics at the same concentrations) at an OD_{600nm} of 0.1. Cultures were grown at 37 °C, 250 rpm to an OD_{600nm} of 0.8. Induction of protein expression was carried by addition of IPTG to a final concentration of 100 μM, and the cultures were placed at 30 °C for protein expression. After 6 hours of incubation, cells were harvested by centrifugation at 5000 xg, 20 min, 4 °C and resuspended in 1 ml of PBS-T per 100 mg wet pellet.

In order to verify the expression of GST and GST-miniGCC, SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western Blot analyses were carried out (Shapiro et al., 1967; Towbin et al., 1979). Cell samples of 1.5 ml were lysed either by sonication or using French Press, and 250 μl of each lysed sample were clarified by centrifugation at 13 000 rpm for 5 min, at room temperature. The pellet was resuspended in 250 μl of PBS. Samples from whole cell samples as well as supernatant sample, for both non-induced and induced cultures transformed with either pSXG-miniGCC or the empty vector were subjected to SDS-PAGE analysis on two identical 12% polyacrylamide gels. The gels were either Coomassie stained or used for western blot analysis using an anti-GST antibody from rabbit as primary antibody and a horseradish peroxidase (HRP)-conjugated anti-rabbit antibody as secondary one. Detection of the secondary antibody was carried out using the ECL Western Blotting detection kit from GE Healthcare.

5 Results

5.1 Modelling of the GC-C receptor

One step towards the determination of the interaction between the GC-C receptor and its ligands is the knowledge of the structure for the receptor itself. In the absence of an experimentally solved structure, the construction of a three-dimensional model for the ligand-binding domain of the receptor may provide valuable information. Two homology models have been published for the extracellular domain of the GC-C receptor (GCC-ECD), and several *in vitro* experiments have been carried out based on the acquired data (Hasegawa and Shimonishi, 2005; Lauber et al., 2009). The structural models were built based on the crystal structure of the unliganded NPR-A receptor, which is also a GC receptor (see *Introduction*, 1.2). However, one other structure exists for the bound form of the receptor, and several ones for the NPR-C receptor, which is related to the NPR-A receptor and possess a highly similar structure (Ogawa et al., 2004; He Xl et al., 2001; He et al., 2006). No dimeric model of GCC-ECD has been published, even though evidence of its presence has been presented and the structures for the bound NPR receptors all contain dimers (Vaandrager et al., 1994; Hasegawa et al., 1999c; Vijayachandra et al., 2000).

5.1.1 Gathering of GC-C homologs

Preliminary sequence analysis

An initial BLASTp search using the human sequence for the extracellular domain of the GC-C receptor was performed as described in *Methods*, section 4.1.1 (Table 5.1). The search was carried out against the manually annotated UniProtKB/SwissProt database and yielded nine sequences with an E-value below 1. The identified sequences were six GC-C receptors, 2 NPR-C receptors, and one sequence corresponding to the centrosomal protein CEP57L1. Note that sequence identity drops abruptly from 70% to 20% with no sequences having intermediate

Table 5.1: BLASTp search result using the sequence for the human GCC-ECD

E-value	Identity	Entry name	Protein name	Organism	Accesion number
0.0	100.0%	GUC2C_HUMAN	GC-C receptor	Homo sapiens	P25092
0.0	77%	GUC2C_PIG	GC-C receptor	Sus Scrofa	P55204
$1.0x10^{-174}$	71%	GUC2C_RAT	GC-C receptor	Rattus norvegicus	P23897
$1.0x10^{-173}$	72%	GUC2C_CAVPO	GC-C receptor	Cavia porcellus	P70106
$1.0x10^{-172}$	71%	GUC2C_MOUSE	GC-C receptor (isoform 2)	Mus musculus	Q3UWA6-2
$1.0x10^{-172}$	71%	GUC2C_MOUSE	GC-C receptor	Mus musculus	Q3UWA6
$9.0x10^{-3}$	21%	ANPRC_HUMAN	NPR-C receptor (isoform 2)	Homo sapiens	P17342-2
$9.0x10^{-3}$	21%	ANPRC_HUMAN	NPR-C receptor	Homo sapiens	P17342
$6.7x10^{-1}$	25%	CE57L_RAT	Centrosomal protein CEP57L1	Rattus norvegicus	Q6AXZ4

values, which makes it difficult to assess whether the non-GC-C receptor sequences are homologous to the GC-C sequences, especially considering their poor statistical values. In addition, no NPR-A receptor was found, even though it is, when the full-length receptor is considered, the most similar protein to GC-C in terms of sequence identity (data not shown).

Considering those results, it was chosen to perform the same search, but using PSI-BLAST (see *Materials*, 3.1.1). More than 100 sequences were gathered at the fourth iteration, suggesting that a profile-based search method is, in the case of GCC-ECD, the right approach to gather remote homologs. However, the user control over the procedure is limited, since PSI-BLAST does not provide the user with the ability to modify, at each iteration, the multiple sequence alignment nor the profile. It was therefore chosen, in order to build a sequence alignment as good as possible, to use another profile-based method allowing that kind of control.

Database search using HMM profiles

Table 5.2: Gathering of GCC-ECD homologs by HMM profile search

Profile Nr.	E-value cutoff	Sequences gathered	Sequences aligned
-	0.0	15	9
1*	1e-100	14	11
2*	1e-100	18	15
3*	1e-10	49	41
4*	1e-100	47	47
4**	1e-100	93	88

* Search against the UniRef90 database
** Search against the UniProt database

The chosen method, HMMER, is based on the use of Hidden Markov Model

(HMM) profiles (see *Materials*, 3.1.1), which were used in place of the query sequence to perform sequence similarity searches (See *Methods*, 4.1.1). From the sequence of the full-length receptor, close homologs were gathered by a BLASTp search against the UniRef90 database limited to the Coelomata taxonomic group, which yielded 15 sequences. A multiple sequence alignment was constructed from those homologs, which was edited so as to contain only sequences highly similar to the GCC-ECD, but also only the portion of the alignment corresponding to the length of GCC-ECD. This first "GCC-ECD alignment", which contained nine sequences, was the starting point for the iterative procedure that was carried out to harvest remote homologs to the GCC-ECD (Table 5.2). Briefly, a HMM profile was built from the alignment and used to perform a database search, and, from the search results, the sub-sequences corresponding to the length of the GCC-ECD were fetched and aligned. Sequences containing important deletions were removed from the alignment, which was against reduced to the exact length of the GCC-ECD, leading to a new "GCC-alignment".

Four HMM profile searches were conducted in total, leading to a set of 47 sequences (Table 5.2). However, since the searches were performed against the UniRef90 database, the set of sequences only contains cluster representatives of sequences that are identical above 90%. In order to harvest all corresponding sequences, the last HMM profile was used to search the UniProt database, again limited to the Coelomata taxonomic group. Ninety-three sequences were thus gathered, 88 of which were kept in the subsequent "GCC-ECD alignment".

Analysis of the gathered sequences

The set of 88 sequences gathered by the iterative HMM profile search contains a lot of highly similar sequences that provide the very similar information, and thus can be removed without loss of information. The CD-HIT program was used to remove the redundancy above 90% from the set. This new set was compared to the original one in order to identify the redundant sequences: phylogenetic trees were built for both sets, and the sequences removed by the CD-HIT program were located in the tree corresponding to the original set of sequences. Sequences that were (i) removed by the CD-HIT program, (ii) not a human sequence, and (iii)

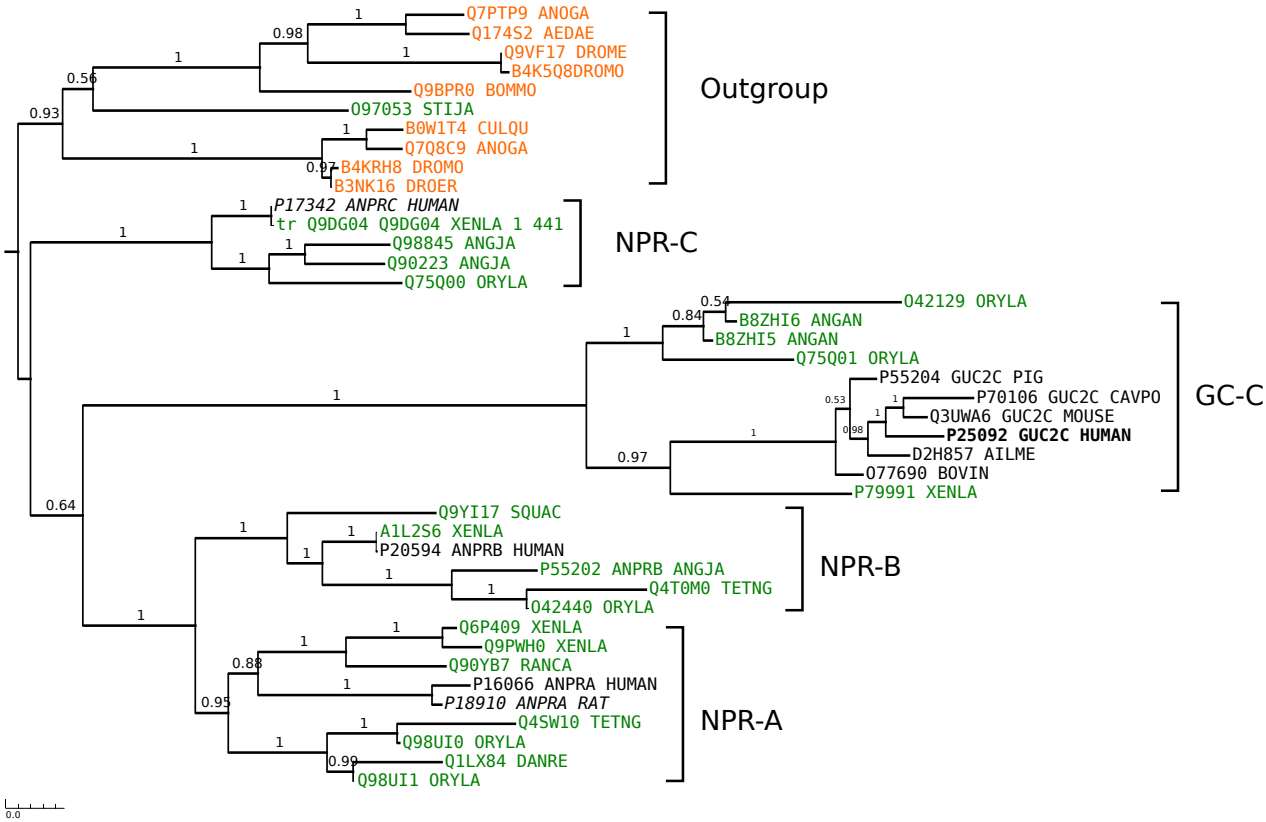


Figure 5.1: Phylogenetic distribution of homologous sequences to the GCC-ECD. The phylogenetic tree was built using the MrBayes program, using the set of 41 sequences obtained by profile HMM search, after removal of redundancy above 90%. Insects (in orange) were chosen as an outgroup. Sequences belonging to mammals are represented in black, and those belonging to fishes or amphibians in green. A bold font was used for the sequence of the human GCC-ECD, and a italic one for the sequences for which structures are available. Clusters are indicated the by name of the sequence they represent.

not associated with an experimentally determined structure were removed from the original set of sequences, leading to the final set.

The result was a set of 41 sequences, for which the corresponding phylogenetic tree is presented in Figure 5.1. The sequences are separated in four clusters corresponding to either orthologs of the human GC-C, or belonging to NPR receptors, which strongly suggests that they are indeed homologs. Interestingly, the tree suggests that the sequences for the NPR-B receptor, as well as the sequences for the NPR-A receptor, are less distant from the GC-C sequences than the NPR-C sequences are, which the opposite of what their level of sequence identity would

suggest. It is, however, not a surprising result considering that both the NPR-A and NPR-B receptors are guanylate cyclase receptors, whereas the NPR-C receptor is a protein G-coupled receptor.

The sequences for the GC-C receptor are more distant to any other cluster of sequences than they are from each other. This organisation also separates the two structures in different clusters. A human sequence, which was chosen as representative organism for the mammals, is present in each group. It is the only mammalian sequence in all groups apart from the GC-C cluster, the presence of the rat NPR-A sequence being due to its association with a structure. This reflects the high sequence identity between mammalian sequences for the NPRs, and the fact that, even within closely related organisms, the sequences for the GC-C receptor are still different, the sequence identity between them being around 70% only.

5.1.2 Alignments of GCC-ECD with its homologs

In order to evaluate the effect on both the sequence information provided by the gathered homologs to the GC-C receptor and the structural information provided by the structures of the NPR-A and C receptors, three alignments programs were used. The full set of sequences was aligned by the MAFFT and EXPRESSO multiple sequence alignment tools, the latter taking into acount the structural information provided by the structures associated to the NPR-C and rat NPR-A sequences. For the third alignment, which was performed using the alignment tools from the MODELLER program, only the sequences for the GC-C, NPR-C, and rat NPR-A, that is our target protein and the sequences for which structures are associated, were considered. In this case, a structure-structure alignment was done on the NPR sequences, and the GC-C was thereafter aligned to them in a structure-sequence alignment.

The three alignments are presented in parallel in Figure 5.1, with, in the case of the alignments performed on the full set of sequences, only the GC-C, NPR-C, and rat NPR-A sequences represented (the full alignment that was done using MAFFT in presented as appendix). All alignments present the secondary structure elements and for the NPR-A and C receptors (NPRs) well aligned with each other, but also with the predicted structural elements for the GCC-ECD, which are

```
                                10        20        30        40        50        60        70        80        90       100
mafft|NPRA_RAT      LLLLRGGHASDLTVAVVLPLTNTSYPWSWARVGPAVELALARVKARPD----LLPGWTVRMVLGSSENAA-GVCSDTAAPLAAVDLKWE---HSPAVFLGPGCVYSAAPVGRFTAHWRVPLLTAG  108
mafft|NPRC_HUMAN    RQEREALPPQKIEVLVLLPQDDS-YLFSLTRVRPAIEYALRSVEGNGTGRRLLPPGTRFQVAYED------SDCGN-RALFSLVDRVAAARGAKPDLILGPVCEYAAAPVARLASHWDLPMLSAG  113
mafft|GCC_HUMAN     SQVSQNCHNGSYEISVLMMGNSA-FAEPLKNLEDAVNEGLEIVRGRLQ-NAGLNVTVNATFMYSDGLIHNSGDCRSSTCEGLDLLRKISNAQRMGCVLIGPSCTYSTFQMY-LDTELSYPMISAG  122

expresso|NPRA_RAT   LLLLRGGHASDLTVAVVLPLTNT-SYPWSWARVGPAVELALARVKARPD---LL-PGWTV--RMVLGSSE---NAAGVCSDTAAPLAAVD---LKWE-HSPAVFLGPGCVYSAAPVGRFTAHWRV  102
expresso|NPRC_HUMAN RQEREALPPQKIEVLVLLPQD--DSYLFSLTRVRPAIEYALRSVEGNGTGRRLLPPGTRF--Q--VA------YEDSDCGNR-ALFSLVDRVAAARG-AKPDLILGPVCEYAAAPVARLASHWDL  107
expresso|GCC_HUMAN  SQVSQNCHNGSYEISVLMMG-N-SAFAEPLKNLEDAVNEGLEIVRGRL----QN-AGLNVTVNATFMYSDGLIHNSGDCRSSTCEGLDLL-RKISNAQRMGCVLIGPSCTYSTFQMY-LDTELSY  116

modeller|NPRA_RAT   --------SDLTVAVVLPLTNTSYPWSWARVGPAVELALARVKA-------RPDLLP-GWTVRMVLGSSENA---AGVCSDTAAPLAAVDLKWE---HSPAVFLGPGCVYSAAPVGRFTAHWRVP  103
modeller|NPRC_HUMAN P-------QKIEVLVLLP-QDDSYLFSLTRVRPAIEYALRSVEGNGTGRRL---LPP-GTRFQVAYEDSD---------CGNRALFSLVDRVAAARGAKPDLILGPVCEYAAAPVARLASHWDLP  108
modeller|GCC_Human  SQVSQNCHNGSYEISVLMMGNSAFAEPLKNLEDAVNEGLEIVRG-------RLQNAGLNVTVNATFMYSDGLIHNSGDCRSSTCEGLDLLRKISNAQRMGCVLIGPSCTYSTFQMYLDTEL-SYP  117

                                110       120       130       140       150       160       170       180       190       200       210       220
mafft|NPRA_RAT      APALGIGVK-DEYALTTRTGPSHVKLGDFVTAL-------HRRLGWEHQALVLYADRLGDDRPCFFIVEGLYMRVRERLNITVNHQEFVEGDPDHYPKLLRAVRRKGRVIYICSSPDAFRNLMLL  225
mafft|NPRC_HUMAN    ALAAGFQHKDSEYSHLTRVAPAYAKMGEMMLAL-------FRHHHWSRAALVYSDD--KLERNCYFTLEGVHEVFQEE-GLHTSIYSFDETKDLDLEDIVRNIQASERVVIMCASSDTIRSIMLV  228
mafft|GCC_HUMAN     S----FGLSCDYKETLTRLMSPARKLMYFLVNFWKTNDLPFKTYSWST-SYVYKNG--TETEDCFWYLNALEASVSYF-SHELGFKVVLRQ-DKEFQDILMDHNRKSNVIIMCGGPEFLYKLK--  236

expresso|NPRA_RAT   PLLTAGAPALGIGVKD-EYALTTRTGPSHVKLGDFVTAL-------HRRLGWEHQALVLYADRL-GDDRPCFFIVEGIYMRVRER-LNITVNHQEFVE-G-DPDHYPKLLRAVRRKGRVIYICSS  215
expresso|NPRC_HUMAN PMLSAGALAAGFQHKDSEYSHLTRVAPAYAKMGEMMLAL-------FRHHHWSRAA-LVYSDDKL--ERNCYFTLEGVHEVFQEEGLH--TSIYSFDETK--DLDLEDIVRNIQASERVVIMCAS  218
expresso|GCC_HUMAN  PMISAGSFGLSCD----YKETLTRLMSPARKLMYFLVNFWKTNDLPFKTYSWST-S-YVYKNG--TETEDCFWYLNALEASVSYF--SHELGFKVVL---RQDKEFQDILMDHNRKSNVIIMCGG  228

modeller|NPRA_RAT   LLTAGAPALGIGVKD-EYALTTRTGPSHVKLGDFVTALHRRLGWEHQALVLYADRLGDDRPCFFIVEGLYMRVRERLNITVNHQEFVEGDPDHYPKLLRAVRRKGRVIYICSSPDAFRNLMLLAL  227
modeller|NPRC_HUMAN MLSAGALAAGFQHKDSEYSHLTRVAPAYAKMGEMMLALFRHHHW-SRAALVYSDDK-LERNCYFTLEGVHEVFQEE-GLHTSIYSFDETKDLDLEDIVRNIQASERVVIMCASSDTIRSIMLVAH  230
modeller|GCC_Human  MISAGSFGLSCDYKETLTRLMSPARKLMYFLVNFWKTNDLPFKTYSWSTSYVYKNGTETEDCFWYLNALEASVSY--FSHELGFKVVLRQDKEFQDILMDHNRKSNVIIMCGGPE----FLYKLK  236

                                230       240       250       260       270       280       290       300       310       320       330       340       350
mafft|NPRA_RAT      ALNAGLTGEDYVFFHLDVFGQSLKSAQGLVPQKPWERGDGQDRSARQAFQAAKIITYKEPDNPEYLEFLKQLKLLADKKFNFTVEDGLKNIIPASFHDGLLLYVQAVTETLAQGGTVTDGENITQ  350
mafft|NPRC_HUMAN    AHRHGMTSGDYAFFNIELFNSSSYG------DGSWKRGDKHDFEAKQAYSSLQTVTLLRTVKPEFEKFSMEVKSSVEKQGLN--MEDYVNMFVEGFHDAILLYVLALHEVLRAGYSKKDGGKIIQ  345
mafft|GCC_HUMAN     GDRAV--AEDIVIILVDLFNDQYLE----------------DNVTAPDYMKNVLVLTLSPGNSLLNSSFSRN-------------LSPTKRDFALAYLNGILLFGHMLKIFLENG-ENITTPKFAH  330

expresso|NPRA_RAT   PDAFRNLMLLALNAGLTGEDYVFFHLDVFGQSLKSAQGLVPQKPWERG-DGQDRSARQAFQAAKIITYKEPDNPEYLEFLKQLKLLADKKFNFTV-EDGLKNIIPASFHDGLLLYVQAVTETLAQ  338
expresso|NPRC_HUMAN SDTIRSIMLVAHRHGMTSGDYAFFNIELFNSSSYGD------GSWKRGD-KHDFEAKQAYSSLQTVTLLRTVKPEFEKFSMEVKSSVEK-QGL-NM-EDYVNMFVEGFHDAILLYVLALHEVLRA  333
expresso|GCC_HUMAN  PEFLYKLKGD----RAVAEDIVIILVDLFNDQYLED----------------NVTAPDYMKNVLVLTLSPGNSLLNSSFS-------------R-NLSPTKRDFALAYLNGILLFGHMLKIFLEN  319

modeller|NPRA_RAT   NAGLTGEDYVFFHLDV--FGQSLKSAQGLVPQKPWERGDGQDRSARQAFQAAKIITYKEPDNPEYLEFLKQLKLLADKKFNFTVEDGLKNIIPASFHDGLLLYVQAVTETLAQGGTVTDGENITQ  350
modeller|NPRC_HUMAN RHGMTSGDYAFFNIELFNSSSYGD--------GSWKRGDKHDFEAKQAYSSLQTVTLLRTVKPEFEKFSMEVKSSVEK--QGLNMEDYVNMFVEGFHDAILLYVLALHEVLRAGYSKKDGGKIIQ  345
modeller|GCC_Human  GDRAVAEDIVIILVDLFNDQ-------------YLE----DNVTAPDYMKNVLVLTLSPGNSLLNSSFSRNLS-------------PTKRDFALAYLNGILLFGHMLKIFLENGENIT-TPKFAH  330

                                360       370       380       390       400       410       420       430
mafft|NPRA_RAT      RMWNRSFQGVTGYLKIDRNGDRDTDFSLWDM-DPETGAFRVVLNYNGTSQELMAV-SEHKLYWPLG------------YPPPDVPKC-GFDNEDPACNQDH--------------  436
mafft|NPRC_HUMAN    QTWNRTFEGIAGQVSIDANGDRYGDFSVIAMTDVEAGTQEVIGDYFGKEGRFEMR-PNVKYPWGPLKLRIDENRIVEHTNSSPCKSSGGLEESAVTGIVVG--------------  445
mafft|GCC_HUMAN     AFRNLTFEGYDGPVTLDDWGDVDSTMVLLYT-SVDTKKYKVLLTYDTHVNKTYPVDMSPTFTWKNS------------KLPNDITGR-GPQILMIAVFTLT--------------  417

expresso|NPRA_RAT   GG--TVTDGENITQRMWNRSFQGVTGYLKIDRNGDRDTDFSLWDMD-PETGAFRVVLNYNGTSQEL-MAVSEHKLYWPLGY------------P-PPDVPKCGFDNEDPACNQDH  436
expresso|NPRC_HUMAN GYSK--KDGGKIIQQTWNRTFEGIAGQVSIDANGDRYGDFSVIAMTDVEAGTQEVIGDYFGKEGRF-EMRPNVKYPWGPLKLRIDENRIVEHTNSSPCKSSGGLEESAVTGIVVG  445
expresso|GCC_HUMAN  G-E-N-ITTPKFAHAFRNLTFEGYDGPVTLDDWGDVDSTMVLLYT-SVDTKKYKVLLTYDTHVNKTYPVDMSPTFTWKNSK------------L-PNDITGRGPQILMIAVFTLT  417

modeller|NPRA_RAT   RMWNRSFQGVTGYLKIDRNGDRDTDFSLWDM-DPETGAFRVVLNYNGTSQELMAVS-EHKL-YWPLGYPPPDVPKCGF----------------------------------  425
modeller|NPRC_HUMAN QTWNRTFEGIAGQVSIDANGDRYGDFSVIAMTDVEAGTQEVIGDYFGKEGRFEMRP-NVKYPWGPLKLRIDENR----------------------------------  418
modeller|GCC_Human  AFRNLTFEGYDGPVTLDDWGDVDSTMVLLYTSV-DTKKYKVLLTYDTHVNKTYPVDMSPTFTWKNSKLPNDITGRGPQ----------------------------------  407
```

Figure 5.1: Multiple alignment of GCC-ECD with NPR-A and NPR-C. The "mafft" and "expesso" alignments are the portion of the alignment of the set of 41 sequences obtained by profile HMM search (empty columns, which correspond to insertions from other sequences, were removed for an easier vizualisation). The "modeller" alignment was generated by aligning the GCC-ECD with the previously aligned structures 1DP4 and 1JDN, which correspond to the unliganded NPR-A and NPR-C receptors. Residues belonging to secondary structure elements are shown in red (α-helices) or green (β-strands). N-glycosylation sites are highlighted in blue, and cysteines in yellow, and ligand-binding residues in gray. In the case of GCC-ECD, the secondary structure was predicted by PSIPRED, the N-glycosylation sites by NetNglyC 1.0, and the ligand binding sequence is the one inferred from photoaffinity labeling studies (Wada et al., 1996; Hasegawa et al., 1999a). The numbering of each sequence corresponds to their respective extracellular domains.

described later on see section 5.1.3. The residues of the NPRS that are involved in ligand binding, as well as the conserved cysteines, are also aligned in most cases. Two regions are most different for each alignment, located, in terms of secondary structure elements,(i) between the first and second α-helices (residues 45 to 92 of GCC-ECD), and (ii) between the fourth and seventh β-strands (residues 120 and 195 of GCC-ECD, see Figure 5.1). These regions are also both located just before the ligand-binding sequences of the NPRs, and two NPR ligand-binding fragments, located on the fourth α-helix and the sixth β-strand and α-helix (residues 126 to 131 and 170 to 188 of GCC-ECD, approximately) is within the second region.

The first "region of uncertainity" appears best aligned in the "mafft" alignment, with the least gaps and conserved cysteines aligned. The "expresso" alignment contains more gaps, especially within the second β-strand, and the "modeller" alignment does not present the conserved cysteines as aligned. It is to be noted that this region contains the two additionnal cysteines of the GC-C receptor. The second region contains also more gaps for the "mafft" and "expresso" alignments, but they are very similar to each other, and all three alignments present the ligand-binding residues of the NPRs aligned.

In summary, the "modeller" alignment, for which the only sequence information is that of the templates and the target, presents, as expected, less gaps than the other alignments, but conserved residues are not always aligned. For the "mafft" alignment, which contains additional sequence information but no structure information, conserved residues as well as the sequence fragments corresponding to secondary structure elements are aligned. In fact, the secondary structure

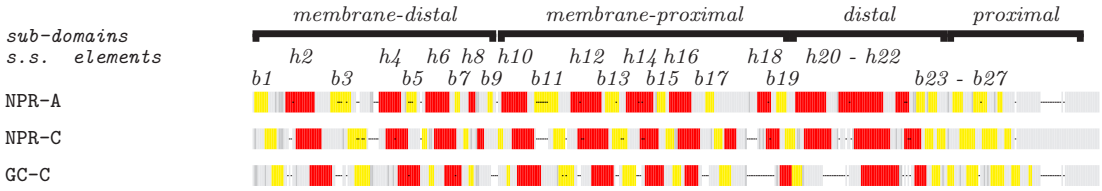


Figure 5.2: Secondary structure for the GC-C, NPR-A, and NPR-C extracellular domains. The secondary structure elements for the NPR-A and NPR-C ECDs are represented along their sequences according to the data from their three-dimensional structures. The secondary structure elements for the GCC-ECD have been predicted by the PSIPRED program. α-helices are represented in red and β-strands in yellow. Numbering of the secondary structure elements is done with NPR-A as reference.

elements are better aligned with MAFFT than they are with EXPRESSO, which is suprising considering that the "expresso" alignment includes structural information. It was therefore chosen to use the "mafft" alignment for the modelling.

5.1.3 Comparison of the secondary structures for the GC-C, NPR-A, and NPR-C ECDs

The secondary structure for the extracellular domain of the GC-C receptor (GCC-ECD) was predicted using PSIPRED, which relies on the results from a PSI-BLAST database search, and is therefore based on the secondary structure from other proteins. As expected, the organization of secondary structure elements for the GCC-ECD is globally the same as for the NPRs, with an alternance of α-helices and β-strands along most of the sequence, with the C-teminus exclusively composed of β-strands (Figure 5.2). Not predicted are the α-helices h8 and h20 from the NPRs, and the extra α-helix of NPRC-ECD, located just before h18. The helix h8 corresponds to the ligand-binding fragment of the NPRs that is between residues 111 and 115 of NPR-A, located within the second region of uncertainity in the sequence alignments (see section 5.1.2). The extra helix of NPRC-ECD (residues 267 to 271) and the helix h20 (residues 286 to 305 of NPR-A) are located, on the sequence alignment, within the gapped region appearing after the last conserved cysteine. The other binding regions for the NPRs correspond to helices h6 and h12, and the β-strand b13. Helices h4 and h6 are involved in the

Table 5.3: Structures obtained via threading

PDB ID	Resolution (Å)	P-value	UniProt ID	Protein description
1JDP	2.00	7e-05	P17342	CNP-bound NPRC-ECD
3JPW	2.80	1e-04	Q00960	Glutamate NMDA receptor subunit NR2B
3OM0	1.40	3e-04	Q63273	Ionotropic Glutamate Receptor Kainate 5
3LOP	1.55	4e-04	B5RX19	Substrate-binding periplasmic protein (Pbp) from Ralstonia solanacearum, engineered
3H6G	2.70	4e-04	P42260	Ionotropic Glutamate Receptor Kainate 2
3HUT	1.93	6e-04	Q2RQC5	Branched-chain amino acid ABC transporter from Rhodospirillum rubrum, putative
3OLZ	2.75	6e-04	D3ZDH2	Ionotropic Glutamate Receptor Kainate 3
3H5L	1.70	7e-04	Q5LQF6	Branched-chain amino acid ABC transporter from Silicibacter pomeroyi, putative

dimerization interface of the NPR receptors.

Identification of remote structures by fold recognition

With the purpose of exploring the possibilty for the extracellular domain of the GC-C receptor to have a fold different to that of the NPR receptors, its sequence was submitted as query for a database search by fold recognition, using the pGenThreader tool of the PSIPRED web-server (Bryson et al., 2005). Eight structures with a p-value below 1e-03 were obtained, with the structure for the NPR-C receptor bound to the natriuretic peptide C (CNP; PDB entry 1JDP) scoring highest, with a p-value of 7e-05 (Table 5.3). This result further supports the hypothesis according to which the GCC-ECD has a structure very similar to that of the NPR receptors.

Four structures corresponding to different ionotropic glutamate receptors (iGluRs), which mediate excitatory synaptic neurotransmission in the central nervous system, were also harvested (PDB entries 3JPW, 3OM0, 3H6G, and 3OLZ) (Karakas et al., 2009; Kumar et al., 2009; Kumar and Mayer, 2010). They describe their regulatory extracellular amino-terminal domains (ATD), which are located before their ligand-binding domains in terms of sequence. As for the GC-C and the NPR receptors, the fold adopted by those sub-domains is recognized as belonging to the Type 1 periplasmic binding fold superfamily (PBPD1), suggesting that they may be remote homologs to the GC-C receptor (Marchler-Bauer et al., 2011). The remaining three structures, which remained to be published, correspond to putative members of the ABC transporter family (PDB entrys 3HUT and 3H5L) and to an

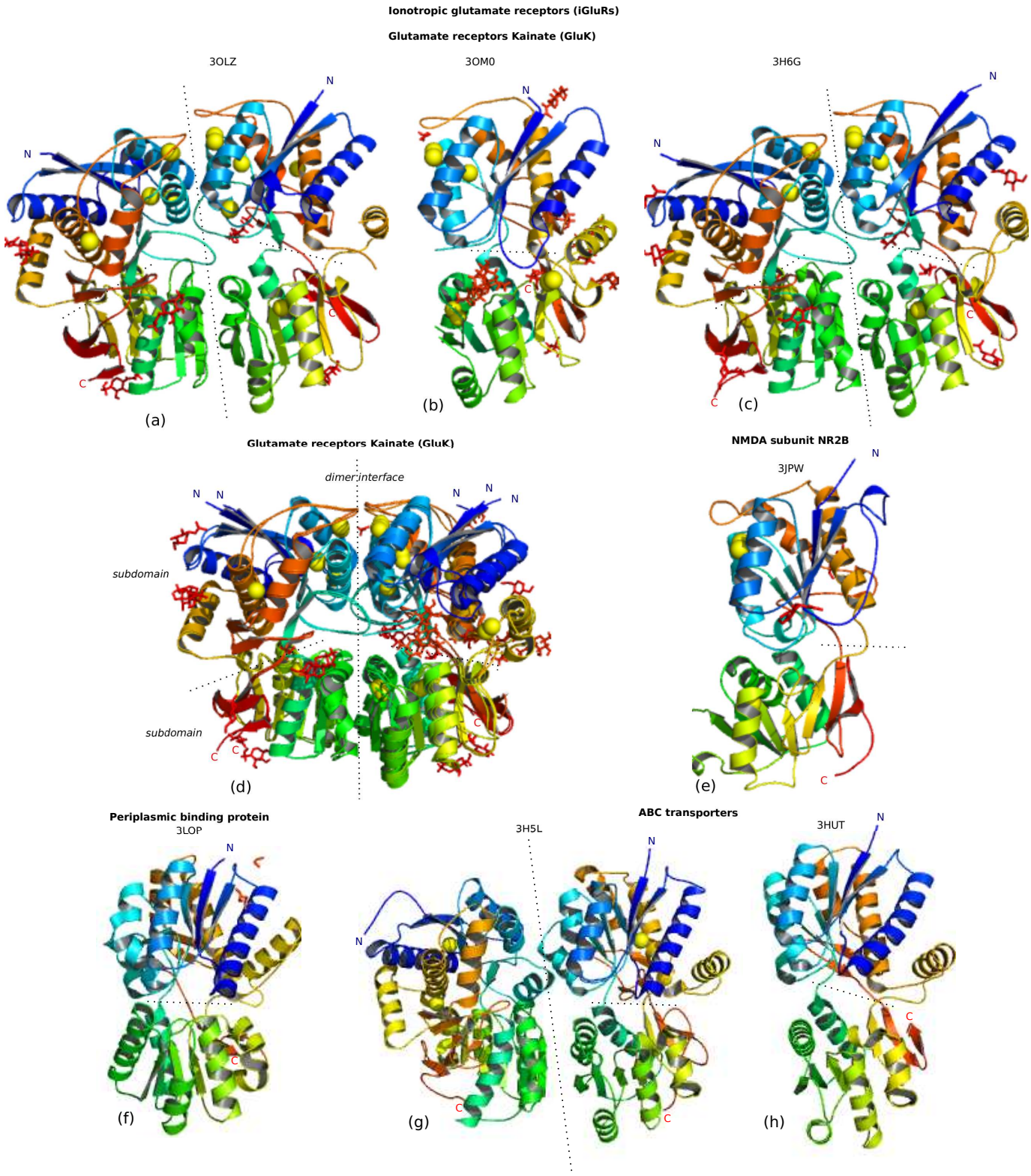


Figure 5.3: Structures identified by threading. Cartoon representation of the different structures obtained by threading using the sequence for the GCC-ECD as query. Each chain is colored as a rainbow from N-terminus to C-terminus. (a) Ionotropic Glutamate Receptor Kainate (GluK) 3 (PDB entry 3OLZ). (b) GluK 5 (PDB entry 3OM0). (c) (PDB entry 3H6G). (d) Superposition of the structures for the GluK receptors. (e) Glutamate NMDA receptor subunit NR2B (PDB entry 3JPW). (f) Engineered substrate-binding periplasmic protein from Ralstonia solanacearum(PDB entry 3LOP). (g)Putative branched-chain amino acid ABC transporter from Silicibacter pomeroyi (PDB entry 3H5L). (h) Putative branched-chain amino acid ABC transporter from Rhodospirillum rubrum (PDB entry 3HUT).

Table 5.4: NPR-A and NPR-C structures

PDB entry	Resolution (Å)	Oligomerization	UniProt ID	Protein description
1DP4	2.00	monomer*	P18910	NPRA-ECD
1T34	2.95	dimer	P18910	ANP-bound NPRA-ECD
1JDN	2.90	monomer	P17342	NPRC-ECD
1JDP	2.00	dimer	P17342	CNP-bound NPRC-ECD
1YK0	2.40	dimer	P17342	ANP-bound NPRC-ECD
1YK1	2.90	dimer	P17342	BNP-bound NPRC-ECD

* The structure presents a dimer which organization is due to crystal packing, therefore only the monomer was used as template (see van den Akker et al., 2000; Ogawa et al., 2004).

engineered substrate-binding periplasmic protein (pbp, PDB entry 3LOP), all of then also members of the PBPD1 superfamily.

All the structures share the fold with 2 subdomains linked by three cross-overs, with each subdomain a β-sheet surrounded by α-helices (Figure 5.3). However, the relative orientation of subdomains and secondary structure elements is different, as well as their lengths, giving an idea of the variability of the fold. As for the NPR receptors, several of the threaded structures contain disulfide bonds, some of them having a greater number of cysteines than the NPRs or even the GC-C receptor. The N-glycosylation sites belonging to some of the structures are mostly located within the hinge region between the two subdomains.

The structure for the unliganded NPRA-ECD (PDB entry 1DP4) was also harvested, but with a p-value of 1e-03 (data not shown), which reflects what was obtained by the initial BLASTp sequence similarity search.

5.1.4 Homology Modelling based on the natriuretic peptide receptors

NPR template structures

As mentionned earlier, six structures are available for the NPR receptors (Table 5.4). The two structures corresponding to the extracellular domain of the NPR-A receptor (NPRA-ECD) describe its unliganded (PDB entry 1DP4) and bound (PDB entry 1T34) forms, respectively (van den Akker et al., 2000; Ogawa et al., 2004). The extracelllular domain of the NPR-C receptor (NPRC-ECD) is described by four structures, one for the unliganded form (PDB entry 1JDN) and three

Table 5.5: PROCHECK analysis of the GCC-ECD models (main chain parameters).

Parameter	NPRA-based models				NPRC-based models				Comparison values	
	unbound	ligand-bound			unbound	ligand-bound			Typical	Bandwidth
	(monomer)	(dimer)	ch. A	ch. B	(monomer)	(dimer)	ch. A	ch. B		
Ramachandran[a]	79.9	84.8	84.3	83.8	82.1	80.2	81.0	84.3	83.8	10.0
Planarity[b]	5.2	4.2	6.7	4.5	5.0	4.5	4.6	4.1	6.0	3.0
Bad contacts[c]	8.6	7.1	6.9	4.9	7.1	6.1	4.2	6.6	4.2	10.0
Cα distortion[d]	2.0	2.0	2.9	2.1	1.9	1.8	1.8	1.7	3.1	1.6
H-bond energy[e]	0.9	0.8	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.2

[a] Percentage of residues in the most favored regions of the Ramachandran plot.
[b] Standard deviation of the ω torsion angle, gives a measure of the planarity of the peptide bond.
[c] Number of bad contacts per 100 residues, with a distance of closest approach less than or equal to 2.6$\AA$.
[d] Standard deviation of the ζ torsion angle, gives a measure of the tetrahedral distortion of the Cα.
[e] Standard deviation of the hydrogen bond energies for main-chain hydrogen bonds.

for the bound one (PDB entries 1JDP, 1YK0, and 1YK1), corresponding to the complexes between NPR-C and the natriuretic peptides A, B, and C (He Xl et al., 2001; He et al., 2006).

General features of the GCC-ECD models

Homology models for the GCC-ECD monomer were built based on the structures for the unliganded NPR-A and NPR-C receptors (PDB entries 1DP4 and 1JDN), but also using the "dimer structures" for the ligand-bound receptors (the 3 structures corresponding to the NPR-C receptor bound to its various ligands were used as a group), modelling each chain of the dimer separately. Dimer models were built based on the same "dimer structures". The "mafft" sequence alignment (see Figure 5.1) was used as basis for the modelling, and the experimentally determined disulfide bonds for GCC-ECD were added as constraints for all models (Hasegawa and Shimonishi, 2005).

Analysis of the modelled structures was carried out by submitting them to the PDBsum structure database, thus generating several analyses concerning the features of the models but also their stereochemical quality by PROCHECK, for which the main chain parameters are presented in Table 5.5 (Laskowski, 2001; Morris et al., 1992). All models are inside the observed values for known protein structures for all parameters, although the percentage of residues with a good value for the (ϕ,ψ) torsion angles (Ramachandran) is below to 90%, and that the number of bad contacts is rather high. However, this kind of result is expected considering that the models were not refined.

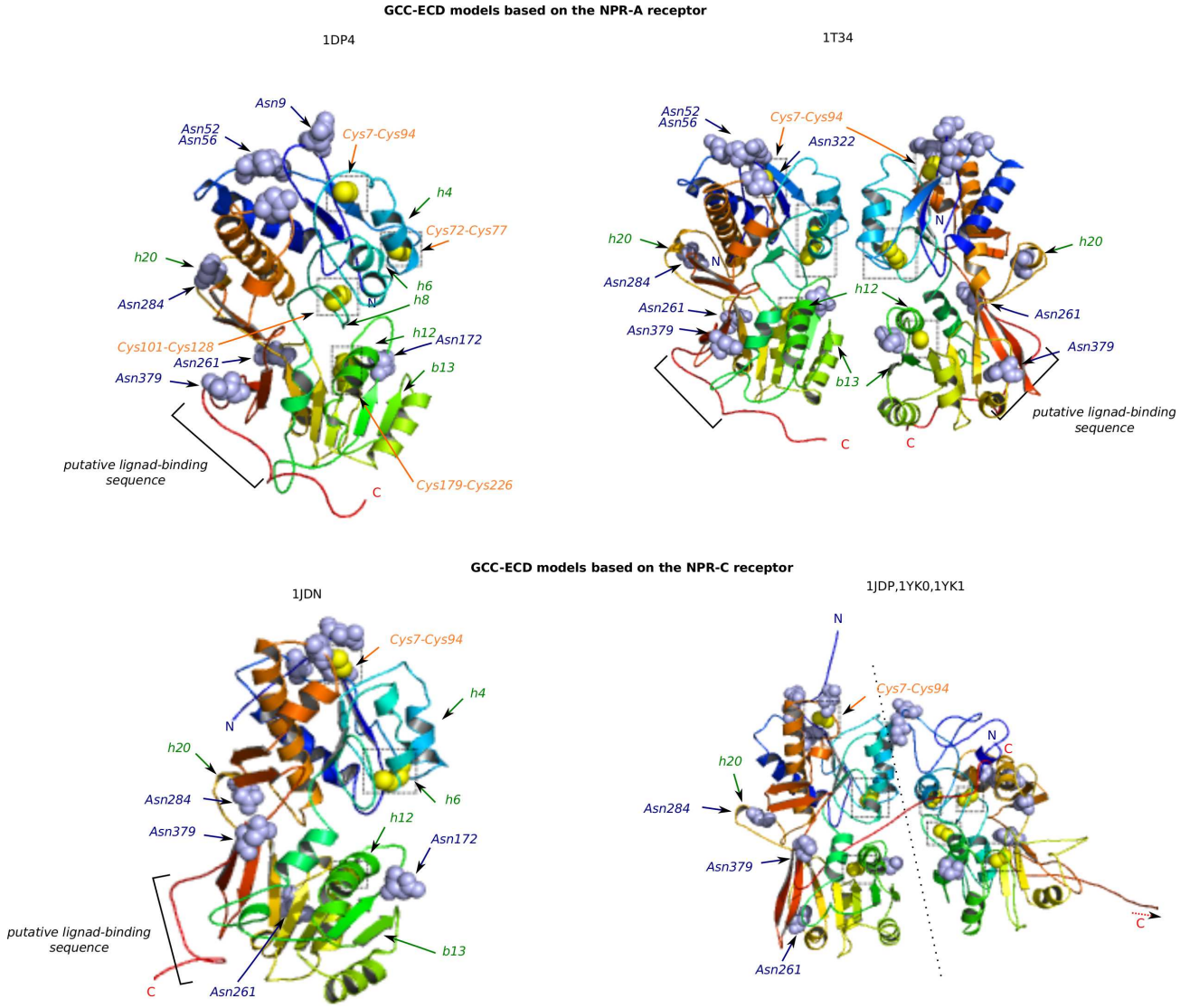


Figure 5.4: Homology models of the GCC-ECD. Each model is shown as a cartoon representation, each chain colored as a rainbow from N-terminus to C-terminus. The PDB code(s) of the corresponding template(s) is(are) indicated above the models. Several features are represented on the models, such as the disulfide bonds (sulphur atomes shown as yellow spheres) and potential N-glycosylation sites (atomes represented as gray spheres) of the GC-C receptor. The secondary structure elements of interest are indicated by their number according to the sequence of the NPR-A receptor. The putative ligand-binding sequence, as described by Hasegawa et. al, is indicated by a bracket.

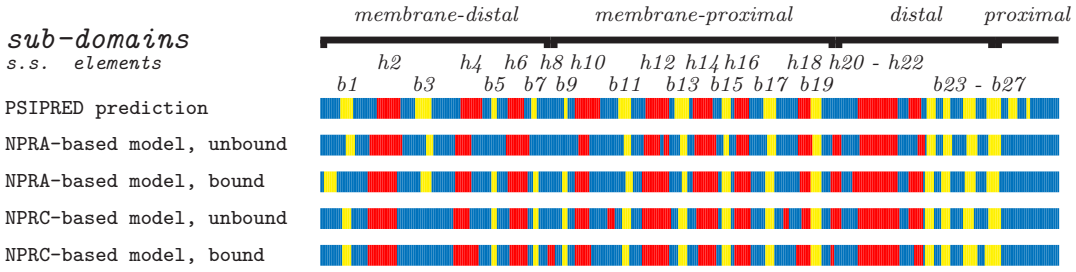


Figure 5.5: Secondary structure of the GC-C models. The organization of the secondary structure elements along the sequence of the GCC-ECD is represented for the different models, as well as the PSIPRED secondary structure prediction. α-helices are shown as in red and β-strands in yellow.

The structures of the models for the extracellular domain of the GC-C receptor (GCC-ECD) are presented in Figure 5.4, with the exception of the monomer models based on the ligand-bound structures for the NPR receptors, which were identical to the chains from the dimer models. All models show, as expected, the same overall structure, although the two chains from the model based on the ligand-bound NPR-C receptor are totally different from each other. The most structured chain is similar to the model obtained based on the unliganded NPR-C and to the NPR-C templates, so the unstructured chain is most likely an artefact.

The N- and C-terminal ends also adopt different conformations depending on the model, aberrant in several cases. The N-terminal region corresponds to the portion of the GCC-ECD sequence (before the first β-strand) that is not present within the structure files for the NPR templates (see the "modeller" alignment in Figure 5.1). The C-terminal region is, within the templates, either not defined or present as a coil under the membrane-proximal domain, which makes it difficult for the modelling.

The secondary structure of the extracellular domain of the GC-C receptor presented in the models is globally in aggreement with the secondary structure prediction, although the length of the secondary structure elements varies from model to model (Figure 5.5). The helices that were not predicted for the GC-C receptor are present in the models in a more or less defined fashion, suggesting that program tried to fit the corresponding GC-C sequence onto an helix structure even though it was not optimal.

Disulfide bonds of GCC-ECD

The disulfides bonds of the extracellular domain of the GC-C receptor, which were experimentally determined, were added as a constraint for the modelling (Hasegawa and Shimonishi, 2005). The disulfide bond of the membrane-proximal domain (Cys179-Cys226), which is conserved with that of the NPR receptors, is also located behind the helix h12 on all models (Figure 5.4). The disulfide bond of the NPR receptors which is located in the membrane-distal domain (behind helix h6), is split into two bonds in the GCC-ECD (Cys72-Cys77 and Cys7-Cys94, see Figure 5.1). The Cys7-Cys94 bond is located near the top of the structures, and seems to be responsible for the folding of the N-terminal portion of the chain back into the structures for the models based on the NPR-A receptor. The Cys72-77 bond takes the place of the NPR disulfide bond, and is positionned, in all models but the one based on the unliganded NPR-A receptor, at the exact same position as the Cys101-Cys128 bond.

N-glycosylation sites of GCC-ECD

All models present the potential glycosylation sites for the GC-C receptor on the outside of their respective structures, apart from Asn261 that is buried for the models based on the NPR-C receptor (Figure 5.4). However, the sequence identity between models and templates is not high enough to describe with precision the orientation of residues along the polypeptide chain, the side chains of the Asn residues being shown only for visualization purposes. The sites for the Asn9, Asn52, Asn56, and Asn322 are situated on the top of the structures, often very close to each other. The Asn172 site is at the NPR-like dimer interface within the membrane-proximal domain, with the Asn261 and Asn379 on the other side. The Asn284 site is located between the two subdomains, within the helix h20 that was not predicted for the GCC-ECD.

Residues involved in dimer interface and ligand binding

The residues of the NPR receptors that are involved in either the dimer interface or the ligand binding belong to the sequence stretches that go from the α-helices h4 to h8 for the membrane-distal domain, and from the β-strand b11 to helix h14

Table 5.6: Residues of GCC-ECD potentially involved in interactions

secondary structure	NPRA-based models		NPRC-based models	
	unbound	bound	unbound	bound
h4	Tyr76, Leu80, Leu83	Ser75, Gly79, Leu82	Ser75, Gly79, Arg84	Gly79, Leu82, Lys85
h6	Ser104, Glu107, Asp111, Leu114	Gln107, Leu110, Glu113	Gln107, Leu110, Glu113	Gln107, Leu110, Glu113
"h8"*	Leu126 to Lys131	Ser127 to Lys131	Leu126 to Lys131	Ser127
h12	Glu175, Trp181, Ala185	Phe180, Trp181, Asn184 Ala188	Glu175, Phe180, Trp181, Asn184, Ala188, Tyr191	
b13	Lys200, Val201, Val202	Lys200, Val201, Val202	Phe199, Lys200, Val201	Phe199, Lys200, Val201
h14	Phe209, Ile212, His216	Lys207, Asp210, Met213, Arg218	Asp210, Ile212, Asp214, Arg218, Lys 219	Asp210, Asp214, Arg218

* The α-helix h8 is not predicted for the GCC-ECD, and is not present on all models.

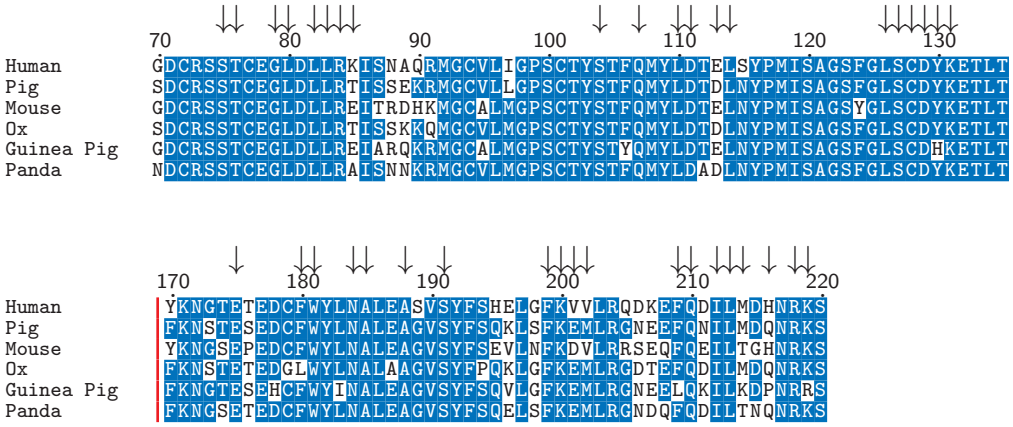


Figure 5.6: Putative interface residues of GCC-ECD. The residues located at the NPR-like interface for the extracellular domain of the GC-C receptor have been identified on the GCC-ECD models (see text, section 5.1.4). This alignment shows the conservation, amongst the GC-C orthologs, of the sequence fragments corresponding to those residues. Residues conserved in more than half the sequences are shaded in blue, and the residues identified on the models are indicated by arrows. The separation between the two regions is shown by a red line

for the membrane proximal domain (Figure 5.1). For the extracellular domain of the GC-C receptor, this corresponds, according to the sequence alignments, to the regions from Cys72 to Leu134 and from Trp164 to His216. On the GCC-ECD models, the residues facing the outside of the molecule on the NPR-like interface can be identified (Table 5.6). They are more or less the same ones for all models, and are highly conserved amongst the GC-C orthologs (Figure 5.6).

The PDBsum analysis on the model for the extracellular domain of the GC-C receptor (GCC-ECD) based on the bound NPR-A reports the putative interactions that occur between the two monomers. Those interactions involve the residues Ile66, Arg73, Ser75, Glu78, Leu82, Leu83, Leu110, Glu113, Tyr130 and Lys131 (data not shown).

5.2 Cloning and expression of the GC-C receptor and its endogenous ligands

In order to study, by biochemical means, the interaction between the GC-C receptor and its ligands, it was chosen to develop an *in vitro* system that would complement the cell-based and sukling mouse assays already in place. For this purpose, the extracellular domain of the GC-C receptor (GCC-ECD), as well as the pro-sequences for its endogenous ligands guanylin and guanylin, were cloned into the pSXG vector. In addition, a small fragment of the GCC-ECD, named miniGCC and corresponding to its putative membrane-proximal sub-domain, was also cloned into the pSXG vector and expressed in *Escherichia coli*.

5.2.1 Construction of the pSXG vectors

The cDNAs for human guanylin, uroguanylin, GCC-ECD and miniGC-C were cloned into the pSXG vector to form the pSXG-guanylin, pSXG-uroguanylin, pSXG-GCCECD and pSXG-miniGCC constructs (Figure 5.7). Each fragment was amplified from a PCR4-TOPO vector containing the sequence for the prohormone and the full-length GC-C receptor, and primers introducing restriction sites in 5' and 3' of the insert sequences were used (see *Materials*, section 3.2.1). In the case of the miniGC-C insert, several amplifications were necessary: one for

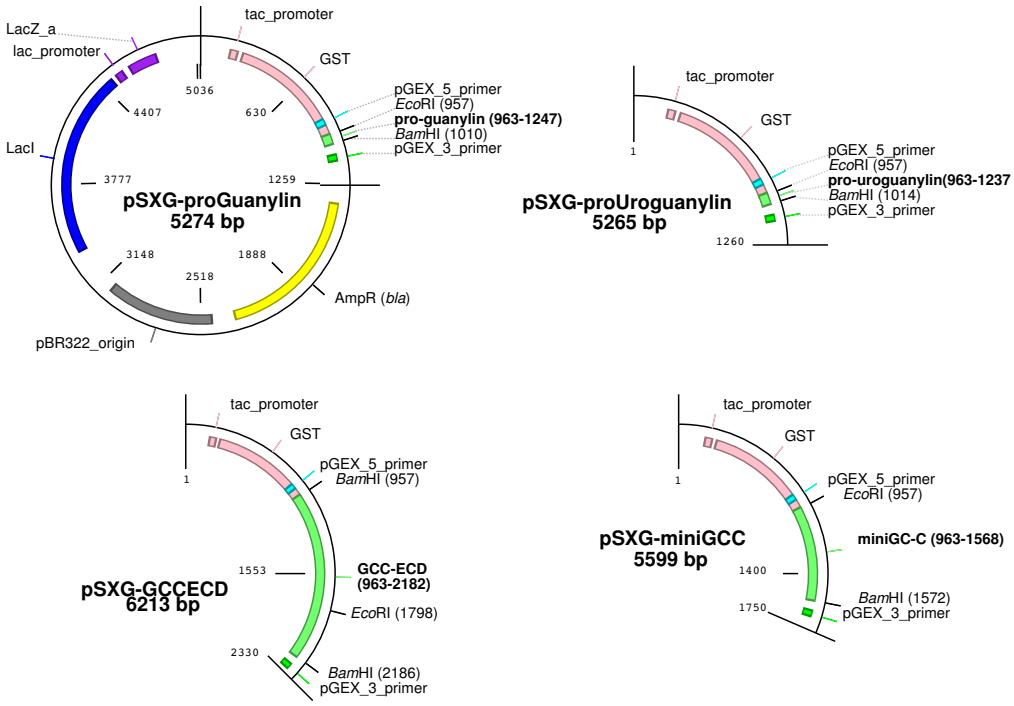


Figure 5.7: Graphical map for the pSXG constructs. The map of the full vector is shown for the pSXG-proGuanylin construct, and the portion containing the insert for the others (pSXG-proUroguanylin, pSXG-GCC-ECD, and pSXGminiGCC). in the case of guanylin and uroguanylin, it is the pro-peptides that have been cloned. The GCC-ECD inserts corresponds to the region of the GC-C receptor that codes for its extracellular domain. The miniGCC insert codes for the putative membrane-proximal domain of the GCC-ECD, according to the design by Lauber et. al (Lauber et al., 2009).

each of the sequence fragments from the GC-C receptor, and another to join them (see Figure 4.2 in Section 4.3.1). The GCC-ECD was cloned in collaboration with Dr Yuleima Diaz.

The inserts were retrieved from their respective clones by conducting a restriction analysis on the pSXG constructs (data not shown). The size of the bands corresponded to the expected sizes of the inserts, i.e. 290 bp for pro-guanylin, 280 bp for pro-uroguanylin, 1229 bp for GCC-ECD and 615 bp for miniGC-C. The clones were submitted to sequencing, which revealed, in the case of GCC-ECD, a frameshift caused by two missing bases after the introduced *Bam*HI restriction site (data not shown). Otherwise all sequences were confirmed, and the pro-guanylin and pro-uroguanylin peptides have been successfully expressed by Arne M. Taxt (personal communication). Pro-guanylin has also been purified using its GST-tag, but difficulties are currently met for the purification of pro-uroguanylin.

5.2.2 Pilot expression of miniGCC

The putative membrane-proximal sub-domain of the extracellular domain of the GC-C receptor (miniGC-C) was expressed as a Glutathione-S-transferase fusion protein in an *E.coli* strain possessing an oxidative cytoplasm, in order to allow the formation of disulfide bonds. Analysis on SDS-PAGE revealed the over-expression, upon induction by ITPG, of a 43kDa protein which was identified as the GST-miniGCC fusion by western blot analysis (Figure 5.8). The presence of GST and GST-miniGCC for non-induced cells indicates a leakage of the pSXG vector. The use of French Press as a lysis method augnmented the yield of protein in the supernatant compared to sonication.

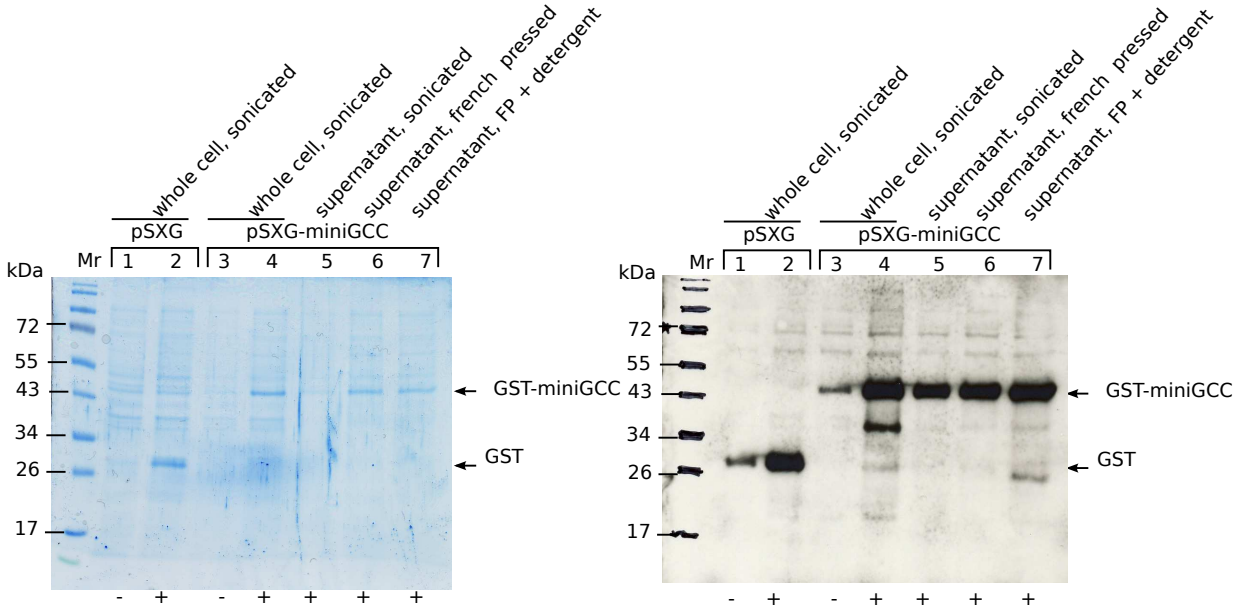


Figure 5.8: Pilot expression of miniGC-C. Expression was carried out as described in the *Methods* section. (a) SDS-PAGE analysis for the expression of miniGCC. (b) Western blot analysis for the expression of miniGCC, using anti-GST antibodies. Lanes 1 and 2: pSXG, lanes 3-7: miniGCC. The lanes marked with a minus sign represent the non-induced cultures, whereas those with a plus sign represnt cultures induced with 100μM IPTG. The nature of the samples (i.e. whole cell lysate or supernatant), as well as the lysis method is indicated above the lanes.

6 Discussion

6.1 Homology modelling of the GCC-ECD

The characterization of the interaction between the guanylyl cyclase C receptor (GC-C) and its ligands, the endogenous guanylin and uroguanylin peptides, as well as the heat-stable enterotoxin (STa) from the enterotoxigenic *Escherichia coli* (ETEC) would be an asset for the design of a toxoid vaccine against the latter. In the abscence of a crystal structure for the receptor, homology modelling of its extracellular domain (GCC-ECD), which is responsible for ligand-binding, has been previously performed based on the structure for another guanylyl cyclase receptor, the natriuretic peptide receptor A (NPR-A) (Hasegawa and Shimonishi, 2005; Lauber et al., 2009). The NPR-A receptor is the protein for which a structure is available that possesses the highest sequence identity to the GC-C receptor (data not shown). Ligand-binding studies using a fragment of the GCC-ECD, which design was motivated by the NPR-A based homology model, supports the hypothesis according to which the GCC-ECD has a fold similar to that of the NPR-A receptor (Lauber et al., 2009). However, the published models only present the monomeric form of the GCC-ECD, although it has been shown that the ligand-binding unit is a dimer (Vijayachandra et al., 2000). In addition, when considering the sequence for the GCC-ECD alone, it appears that it more similar to the the natriuretic peptide receptor C (NPR-C, see *Results*, Table 5.1). The NPR-C receptor, which is a protein G-coupled receptor, is homologous to the NPR-A receptor, and structures are available for its extracellular domain, both for the ligand-bound and unbound forms of the receptor (He Xl et al., 2001; He et al., 2006).

In this study, we have built homology models for the GCC-ECD using all the structures that are available for the NPR-A and NPR-C receptors as templates (6 in total, 2 for the NPR-A and 4 for the NPR-C). In order to achieve the highest possible quality for these models, special care was taken when building the sequence alignment to be used as basis for the modelling: 41 sequences for remote homologs to the GCC-ECD were gathered using itrative building of and searching with

Hidden Markov Model profiles, and two different multiple sequence alignments built from this set were compared to the default structure-sequence alignment used by MODELLER. As excepted, the additional sequence information provided by the GCC-ECD homologs improved the alignment between the extracellular domains of the GC-C, NPR-A, and NPR-C receptors. The two alignments built from the set differed in that one of them was containing structrural information (in addition to the sequence information provided by the set of GCC-ECD homologs). However, the other one performed best and was therefore used for the modelling procedure. Another alternative could have been to manually construct a fourth alignment from the information provided by all three sequence alignments, as well as the information from the analysis of the template structures (see *Results*, 5.1.4).

The obtained models all shared the same global structure and were within what is observed for protein in terms of stereochemical parameters (see *Results*, 5.1.4). The N-terminal ends, for which the GCC-ECD sequence was not aligned with the templates, showed aberrant conformations. The same was observed for the C-terminal ends, its corresponding portion in the template structure being either missing or having a coil structure. However, these regions are not critical for the rest of the models, and can be omitted, as is the case with the automaticcally generated models from the ModBase model database (Pieper et al., 2011). The potential N-glycosylation sites for the GCC-ECD were well located on the outside of the model structures. The unique disulfide bonds of GCC-ECD were not as well placed, and resulted in aberrant conformations (such as the folding of the N-terminal end into the structure due to the Cys7-Cys94 bond) or clashes (the Cys72-Cys77 and Cys101-Cys128 situated at the same location). The relative position of those disulfide bridges, even if it is not well modeled, seems to indicate that they maintain the structure of the membrane-distal subdomain in the abscence of the chloride ion that is bound to the NPR receptors, but, unfortunately, there is no data available concerning whether chloride is necessary for the activity of the GC-C receptor.

Another uncertain region of the models is the portion corresponding to the secondary structure elements that are present in the NPR receptors but were not predicted for the GCC-ECD. These elements are present in some of the models but not all, although it is from the conformation of the polypeptide chain that it was

attempted to model them as such. These results, along with the one concerning the terminal ends and the disulfide bonds, reflect the difficulty to model by homology the regions of the target that are most likely structurally different form that of the template. The homology procedure tries to fit to the template as closely as possible, thus sometimes creating clashes and aberrant conformations within the model(s). In order to address these issues, the refinement of the models is necessary.

6.2 Identification of remotely related structures by threading

Considering the low sequence identity between the extracellular domains of the GC-C, NPR-A and NPR-B receptors, it was chosen to perform a database search based on fold recognition rather than sequence similarity (see *Results*, section 5.1.3). The obtained structures all belonged to the Type 1 Periplasmic Binding fold superfamily (PBPD1), which also contains the GC-C and NPR receptors (domain accession number cl10011). This result suggests that more putative homologs to the GCC-ECD than were picked up in the HMM profile search may exist, and that the alignment derived from the gathered set of sequences could have been extended. Indeed, the Conserved Domain Database lists the hierarchy of the superfamily, along with sequence clusters that could have been used as a starting point for the building of the sequence alignment (Marchler-Bauer et al., 2011). In addition, some of the threaded structures contain more cysteines than the NPR receptors, located in the same region as that of the GCC-ECD models. This suggest that such structures might be better to use as templates for the modelling of this region.

6.3 Hypotheses for dimer interaction and ligand-binding

Experimental data based on photo-affinity labelling studies of the STa toxin suggest that the ligand-binding sequence for the GC-C receptor as the ECD fragment

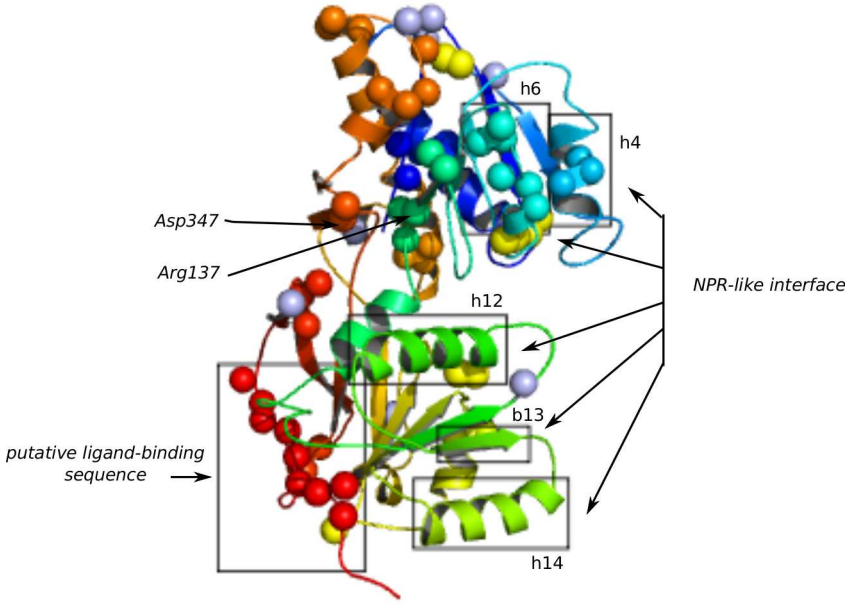


Figure 6.1: Localization of the mutated residues of the GCC-ECD. the residues of the extracellular domain of the GC-C receptor (GCC-ECD) that were submitted to site-directed mutagenesis (see Wada et al., 1996 and Hasegawa et al., 1999a) are shown on the model for the GCC-ECD based on the structure for the ligand-bound NPR-A receptor (PDB entry 1T34). The Cαs of the mutated residues are shown as spheres which color follows that of the main chain. Sulphur atoms from the cysteine residues are shown as yellow spheres and the Cαs of potential N-glycosylation sites as gray ones. The Asn121 site is colored in gray even though it has been subjected to mutagenesis.

between the 387 and 393 residues, at the C-terminal end of the domain (Hasegawa et al., 1999a). However, this region of the GCC-ECD models is not well defined, and they are thus not very well suited to address this issue. On the other hand, the dimer models can be used to predict residues that may be involved in the oligomerization of the receptor but also the ligand-binding, assuming that the GC-C receptors binds its ligands in the same fashion of that of the NPR-A and NPR-C receptors.

Following this hypothesis, a set of residues located at the NPR-like interface of GCC-ECD models that may be involved in either the interaction between GCC-ECD monomer or ligand-binding have been proposed (see *Results*, 5.1.4). Those residues are located, within the GCC-ECD sequence, in the segments from Ser75 to Ser127 and from Glu175 to Arg218. Within the first segment, nine residues

have been subjected previously to site-directed mutagenesis: three (residues 78 to 80) located in h4 and 6 (residues 107 to 109 and 111 to 113) in h6 (Figure 6.1; Wada et al., 1996). Those mutations consisted in the alaline substitution of the polar residues, and lead to various results depending on the mutation: the EGL(78-80)AAV mutation, which was used in combination with the VS(3,4)AG mutation, resulted in an important reduction for the binding of the STa toxin, whereas the QMY(107-109)AIS and DT(111,112)AA mutations had no effect, and the TD(112,113)GA mutation leaded to increased binding of STa. This results suggest that this region is indeed related to the binding of ligand, but most likely in a indirect fashion, as would be the case if this region was involved in the dimer interface rather than ligand binding. Interestingly, no mutations have been performed on residues belonging to the second segment (Glu175 to Arg218, α-helix h12 and β-strand b13), even though it is corresponding to the main ligand-binding region for the NPR receptors. The only mutation in the vicinity, ET(230,231)AA, has a moderate effect on STa binding.

The residues that were proposed as the ligand binding residues by this study, Arg136 and Asp347, are located, on the GCC-ECD model, at the hinge region between the two subdomains, which may explain their importance. The mutation of the residue fragment Arg296-Phe298, which is also located at the hinge region, leads to loss of binding affinity, although not complete. Two other regions were submitted to site-directed mutagenesis, corresponding to the top of the structure (residues 321 to 326) and the end of the domain. In the first case, those mutations had no effect at all, which would seem logical. The mutations at the C-terminal end of the domain, which all resulted in important to complete loss of binding, cannot be related to the models since this region is not well defined.

In summary, the GCC-ECD models fit well with the mutational data, which supports the hypothesis according to which the extracellular domain of the GC-C receptor not only has a fold very similar to the NPR receptors, but also interacts with its ligand in the same fashion, although the other hypothesis for ligand binding could not be investigated.

6.4 Cloning and expression of guanylin, uroguanylin, GCC-ECD and miniGCC

As mentioned earlier, the development of a toxoid vaccine against STa would benefit from the characterization of the interaction between the GC-C receptor and its ligands. To achieve this purpose, the expression of the different protagonists (GCC-ECD, guanylin, uroguanylin and STa) of the interaction would provide a system to study the interaction *in vitro*. It was chosen to express them as gluthatione-S-transferase fusion proteins in *Escherichia coli*, as was done previously for the GCC-ECD (Nandi et al., 1996).

The cloned pro-guanylin and pro-uroguanylin peptides have both been successfully expressed in *Escherichia coli* and purified using their GST-tag, but it has not been established of yet whether they are functional. The extracellular domain of the GC-C receptor has to be cloned again due to the presence of a frameshift occuring between the GST and GCC-ECD fragment of the fusion protein, but it seems only a matter of time before it is, as the miniGCC, expressed in *Escherichia coli*.

6.5 Diversity of the model organisms used for the development of the vaccine against STa

One strategy chosen to identify toxoid candidates for an STa vaccine, pursued by the EntVac consortium, is to screen a library of all possible single amino acid mutants of STa for effects on toxicity and antigenicity. In the process of the development of the toxoid vaccine against STa, the pig and the mouse are used as model organisms (Taxt et al., 2010). In particular, the mouse is used at an early level in suckling mouse assays to assess the ability of toxoid candidates to induce diarrhea. However, putative ligand-binding sequence located at the C-terminal end of the extracellular domain of the GC-C receptor is not strictly conserved between the three organisms (SPTFTWK for the human, SPTFIWK for the pig, and NPNFIWK for the mouse). In this regard, STa toxoid canditates may have a different effect depending on the organism, which could lead either to overlook

good candidates or to a considerable waste of time and ressources pursuing bad ones, in addition to the unecessary sacrifice of animals. On the other hand , the potential ligand-binding residues identified in this study are almost all strictly identical (30 out of 37) between not only human, pig, and mouse, but also including other mammalian species (see *Results*, 5.1.4), suggesting that mouse and pig are relevant model organisms for assessing toxoids aimed for human vaccine usage.

7 Future Perspectives

The homology models built in this study allowed us to propose a set of residues from the GC-C receptor that may be involved in its oligomerization and/or its interaction with the guanylin peptides, making them good candidates for site-directed mutagenesis.

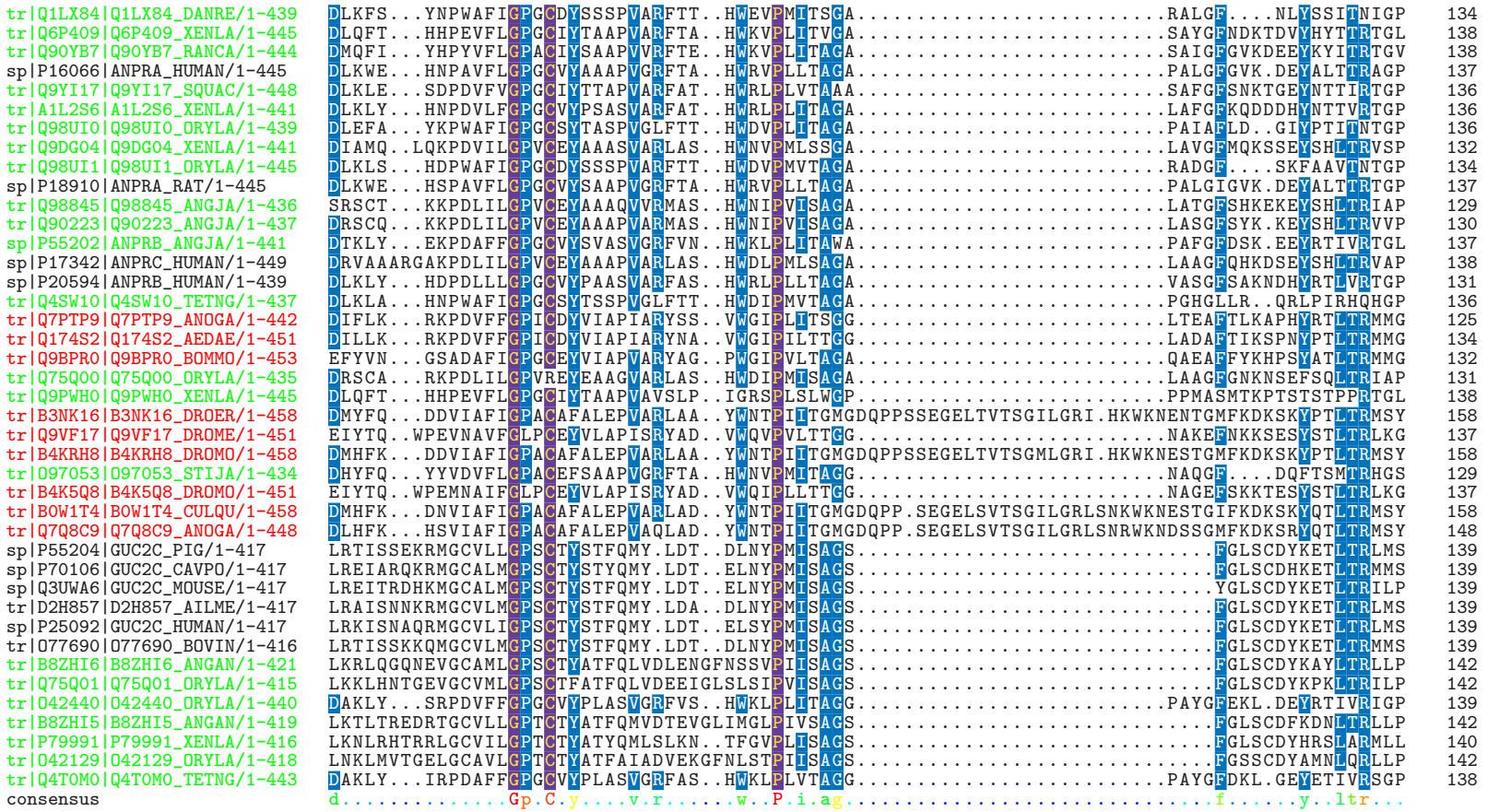
The expression of the extracellular domain of the GC-C receptor, the miniGCC and the guanylin peptides as GST-fusion proteins would provide a complete *in vitro* system to study the interaction between the GC-C receptor and its ligands. The range of possible experiments is wide, from qualitative GST pull-down assays to quantitative binding studies using surface plasmon resonance. In addition to interaction studies, the issue of the oligomerization of the receptor, which remains unclear, could be investigated. It is therefore of high interest to continue the current cloning and expression attempt.

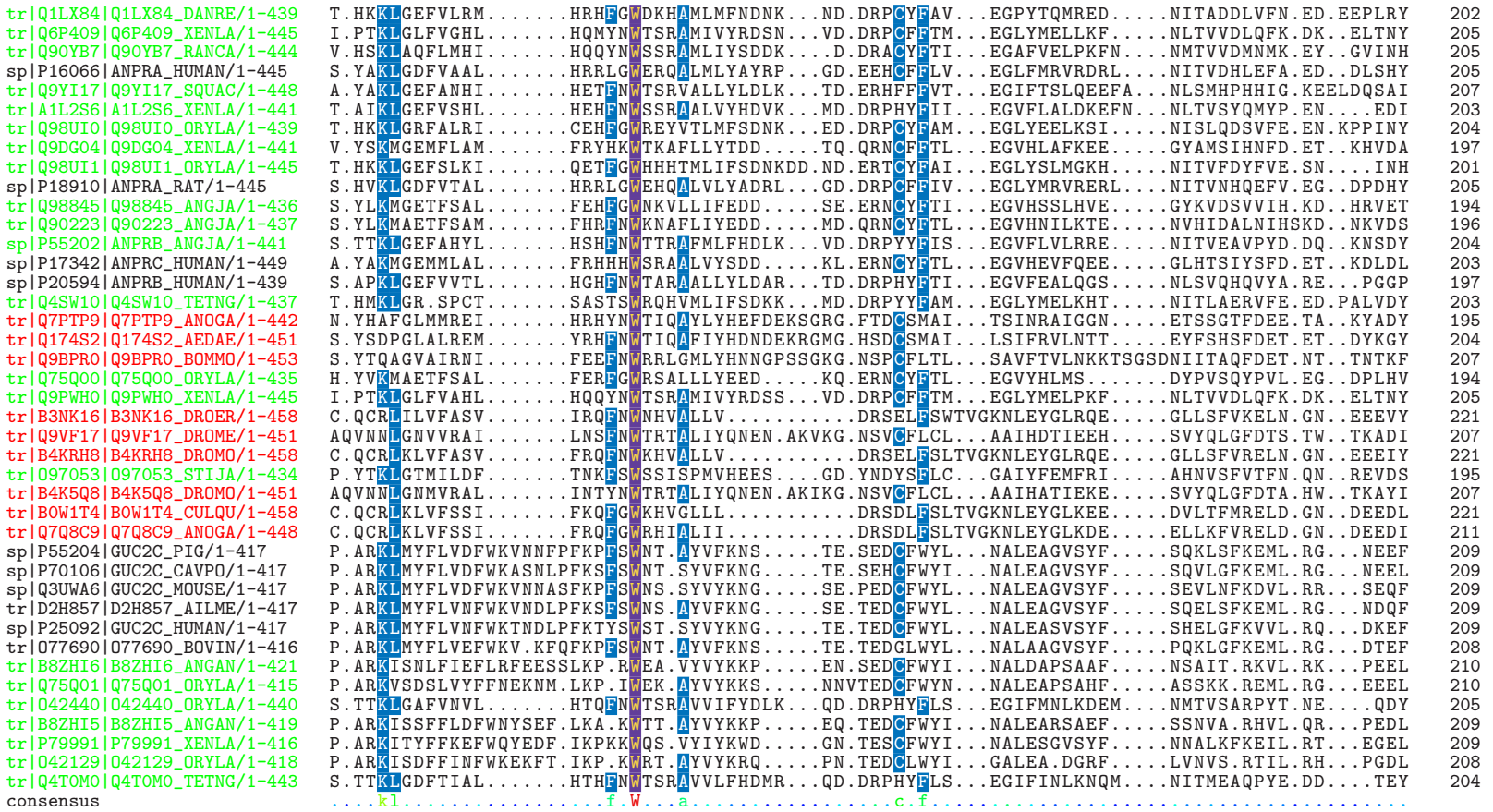
We have also seen that structures other than the ones corresponding to the NPR-A and NPR-C receptors could be used as templates for the modelling og the GCC-ECD, porvided that a high quality sequence alignment is built. Those new models might be able to describe the regions of the GCC-ECD that were not well modeled when based on the NPRs, but also provide us with another alternative for the prediction of residues of interest for the experimental studies. In addition, both groups of models (based on the NPRs and the threaded structures) could be refined in the hope of achieving a quality high enough to carry out molecular dynamics simulations and maybe even docking experiments, with the prior use of ligand-binding site prediction programs. Those models could then be used in combination with the experimental data obtained from mutagenesis and binding studies, and together, form a double edged, self-enhancing studying tool.

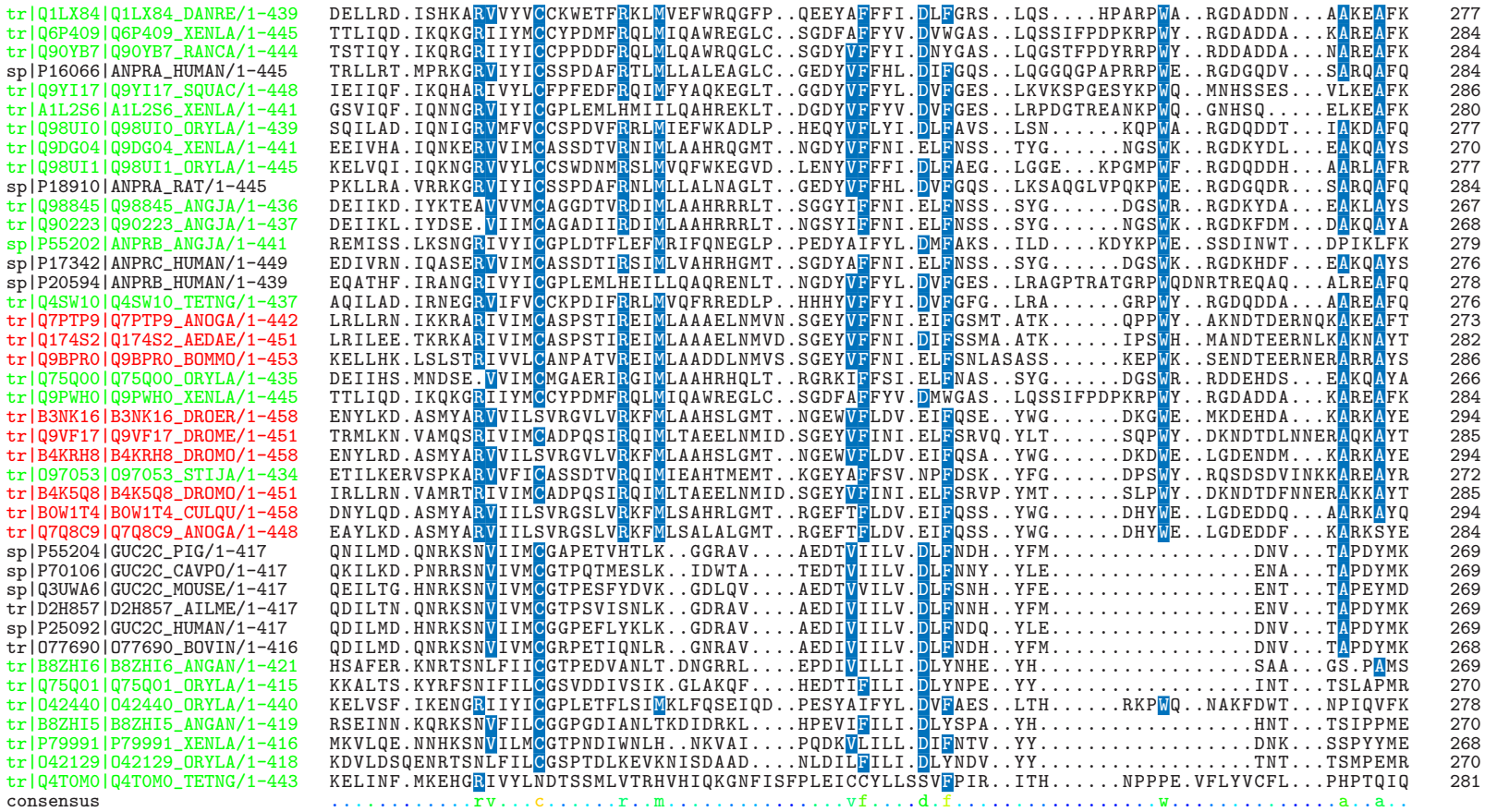
Appendix

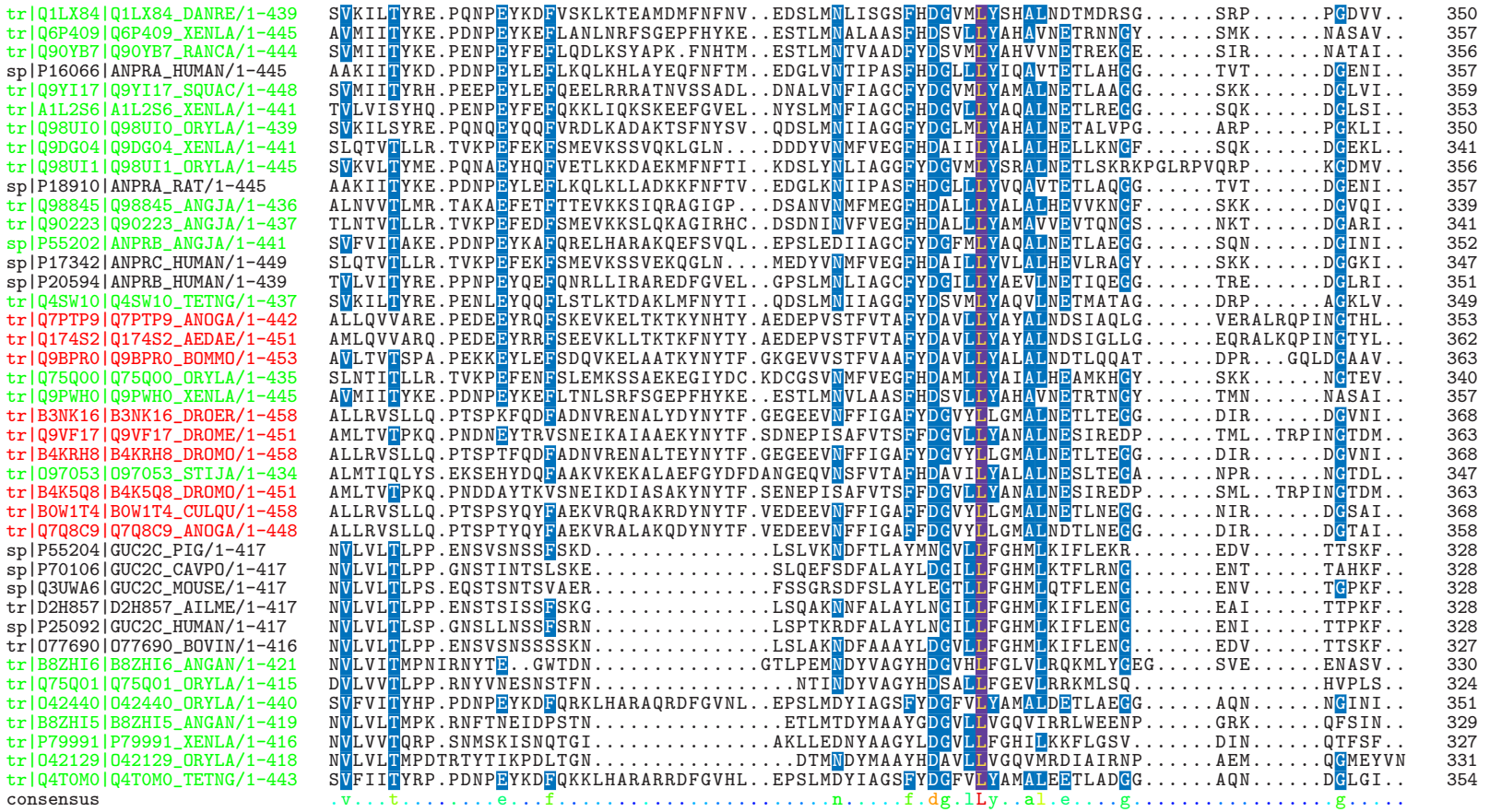
Appendix 1: Multiple alignment of the GCC-ECD remote homologs. The set of 41 sequences obtained by profile HMM search was aligned using the Multiple Alignment Fast Fourirer Transform (MAFFT) alignment tool. The residues are shaded according to their similarity: *from 80% conserved:* yellow letters on dark blue shading, *between 50% and 80% conserved:* white letters on blue shading. The consensus sequence is shown at the bottom, using a color scale from blue to red ("cold-hot"). similar residues (above 50% conserved) are shown in llower case letters and residues conserved above 80% in upper case letters. The residues are numbered according to their full-length sequences in the UniProt database. Figure generated with the texshade package for latex.

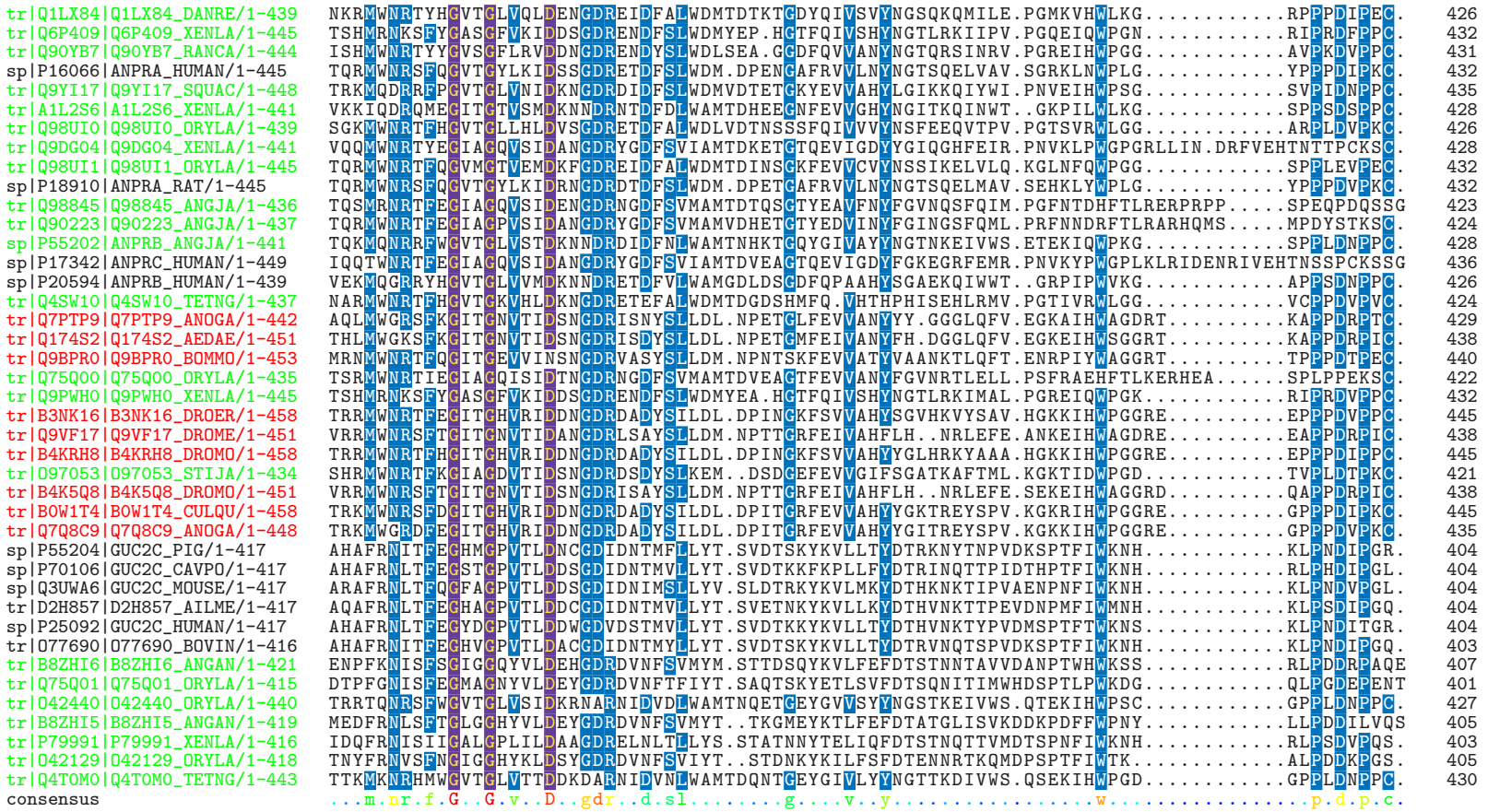

tr|Q1LX84|Q1LX84_DANRE/1-439 Q.SPSAHASQKNITLAVILPLHNT..EYPWAWPRVGPALYWALEKVNSDPN....LLAGYH..LQLVFNSSENKE..GLCSDSVAPLVAV 79
tr|Q6P409|Q6P409_XENLA/1-445 G.VAVGDDGVNNLTMAVLLPKTNR..VYPWAWPRVGPAIQLAIDRINNDPS....LLPDLH..VQTFFGNSEDKD..GVCSDSTAPVVAV 79
tr|Q90YB7|Q90YB7_RANCA/1-444 V.VCHCSQEAQNLTIAVLLPKTST..SYPWALPRVGPAIQMAIDRVNADRD....LLPDFH..LWAVYGDTEDKH..KRCSESAAPVTAV 79
sp|P16066|ANPRA_HUMAN/1-445 L.LLLRGSHAGNLTVAVVLPLANT..SYPWSWARVGPAVELALAQVKARPD....LLPGWT..VRTVLGSSENAL..GVCSDTAAPLAAV 79
tr|Q9YI17|Q9YI17_SQUAC/1-448 G.TGRSQSPPETINIAIVLPH.NT..KYAWAWPRVGPAIQMAIERINNDGD....LLKDYV..LTYKYKSSEEN...GGCADSLAPLHAV 77
tr|A1L2S6|A1L2S6_XENLA/1-441 R.SHQNVTSNHTLTLAVVLPESNI..RYAWSWPRVAPALRMAVDRAQELQ.....LLSGYQ..VKWVFLTSELN...GACSEYVAPLNAV 77
tr|Q98UI0|Q98UI0_ORYLA/1-439 A.HGHHERPRQNITLAIILPETNT..AYPWAWPRIGPALERAIIKINSDPN....LLPNHH..LTYVFKSSENSN..GICSESVAPLVAV 79
tr|Q9DG04|Q9DG04_XENLA/1-441 L.AKSNPMDEDTVNMLVLLPKDNS...YMFSMDRVKPAIDHALSSIQENQT....LLPGVH..FNVIYND.......SDCGN.QALFSLI 72
tr|Q98UI1|Q98UI1_ORYLA/1-445 S.SENTTDDLQEVTLAAILPLTNT..DYPWAWPRVAPALYRAVDSVNSDPH....LLPGLK..LQLVHGSSENRE..GFCSDSAAPLVAV 79
sp|P18910|ANPRA_RAT/1-445 L.LLLRGGHASDLTVAVVLPLTNT..SYPWSWARVGPAVELALARVKARPD....LLPGWT..VRMVLGSSENAA..GVCSDTAAPLAAV 79
tr|Q98845|Q98845_ANGJA/1-436 L.PVRTSALNEEIEVLVLLPKNNS...YIFSMPRVRPAIEYAKIRLSAD.....LYPGLN..FTVHYDN.......SDCGN.EALFSLV 70
tr|Q90223|Q90223_ANGJA/1-437 L.PVNESASTEDIDVLILLPKNNS...YHFSISMVAPAIDYAQKKMKSVNG....LYSGLN..FNFHYEN.......SNCGD.EALFRLV 72
sp|P55202|ANPRB_ANGJA/1-441 M.ARCRTEIGKNITVVVMLPDNHL..KYSFAFPRVFPAIRMAHDDIQKKGK....LLRGYT..INLLNHSTESQG..AGCSESQAQIMAV 79
sp|P17342|ANPRC_HUMAN/1-449 R.QEREALPPQKIEVLVLLPQDDS...YLFSLTRVRPAIEYALRSVEGNGTGRRLLPPGTR..FQVAYED.......SDCGN.RALFSLV 76
sp|P20594|ANPRB_HUMAN/1-439 A.GGVRPPGARNLTLAVVLPEHNL..SYAWAWPRVGPAVALAVEA..........LGRALP..VDLRFVSSELE...GACSEYLAPLSAV 72
tr|Q4SW10|Q4SW10_TETNG/1-437 D.QPGAERERPNITLAVILPQSNT..EYPWAWPRIGPAIDRAVRTVNANAT....LLPDHH..LTYAFKSSEDKA..GICSELAASLMAV 79
tr|Q7PTP9|Q7PTP9_ANOGA/1-442 ..........YVKFAILLPKKPSKNRDIRILSTVLPVIEMATRVVTAPGG....LLQNLR..IEIDYRD.......TQCSSTYGALGAF 66
tr|Q174S2|Q174S2_AEDAE/1-451 S.ASEDSYHEPHIKFAILLPEH.GRSRDSRILSTVRPVIEMATNLVTGPNG....VLHNLK..IEIDYRD.......TQCSSTYGALGAF 75
tr|Q9BPR0|Q9BPR0_BOMMO/1-453 R.PERHQFHRKIVKLGVLLPADPN...QVFSLVKVLPILEMAIPAVTKQDG....PLPGWK..ILVDYRD.......TLCSSVEGPLAAF 73
tr|Q75Q00|Q75Q00_ORYLA/1-435 L.DPVLSGRTEDIDVLVFLPQNNS...FLFSSARVAPALRYAQRRLQAGEG....NFSGFH..FNLHFQS.......SDSPN.EALFALV 72
tr|Q9PWH0|Q9PWH0_XENLA/1-445 G.VSVGDDGVNNLTMAVLLPKTNR..VYPWAWPRVGPAIQLAIDRINNDPS....LLPDLH..VQTFLGNSEDKD..GVCSDSTAPVVAV 79
tr|B3NK16|B3NK16_DROER/1-458 E.VGEMGSTMRVYNVGVLMASHLD...SPFDLERCGPAVDLALDEINKV......FLKPHN..ITLLKKKGSY....PSCSGARAPGLAA 74
tr|Q9VF17|Q9VF17_DROME/1-451 P.DHRRLGARRQLVFVALLPSVESDNKNDCIMPKVLPVLELAIRHVQRMG.....FVGGSHFDIQLISRD.......TFCSSKYGPIGFF 77
tr|B4KRH8|B4KRH8_DROMO/1-458 L.NERDLSRMRVYNVGVLMASHLD...SPFDLERCGPAVDLALDQINKR......FLSPHN..IRLVKKKASY....PSCSGAKAPGLAA 74
tr|O97053|O97053_STIJA/1-434 R.ELRIGLMLPKFPLVSLLPETTQRPMYPFFLQMVQPAVEIALQEVKAT......TLPFHQ..VSVVSND.......TLCNVNTAQIVVV 74
tr|B4K5Q8|B4K5Q8_DROMO/1-451 Q.QHDGPVQRRQLVFVALLPSVESDNKNDCIMPKVLPVLELAIGHVQRMG.....FVGGVQIDITLISRD.......TFCSSTYGPLGFF 77
tr|B0W1T4|B0W1T4_CULQU/1-458 P.EEDYGGNYTTYNVGVLMASHLD...SPFDLERCGPAVDLALVFINEF......LMAHHR..IKLLKVQQSY....ASCSGAKSPGLAA 74
tr|Q7Q8C9|Q7Q8C9_ANOGA/1-448 ..........VYHVGVLMASHLD...SPFDLERCGPAIDLALELVNQS......LMKVHN..VRLSKVQRSY....ATCSGSKSPGLAA 64
sp|P55204|GUC2C_PIG/1-417 S.SVSQNCHNGSYEISVLMMNNSA...FPESLDNLKAVVNEGVNIVRQRLL.EAGLTVTVN..ATFVYSEGVIYKS.SDCRSSTCEGLDL 82
sp|P70106|GUC2C_CAVPO/1-417 S.QISQNCHNGSYEITVLMMNNYA...FQESLESLKTAVNKGLDIVKQRLQ.EAALYVTVN..ATFIHSDGLIHKS.GDCRSSTCEGLDL 82
sp|Q3UWA6|GUC2C_MOUSE/1-417 S.QVRQNCRNGSYEISVLMMDNSA...YKEPMQNLREAVEEGLDIVRKRLR.EADLNVTVN..ATFIYSDGLIHKS.GDCRSSTCEGLDL 82
tr|D2H857|D2H857_AILME/1-417 S.QISRNCNNGSYEISVLMMNNSA...FPESLDNLKEAVNEGVEIVRQRLL.NAGINVTVN..VTFIYSDSVIYKS.NDCRSSTCEGLDL 82
sp|P25092|GUC2C_HUMAN/1-417 S.QVSQNCHNGSYEISVLMMGNSA...FAEPLKNLEDAVNEGLEIVRGRLQ.NAGLNVTVN..ATFMYSDGLIHNS.GDCRSSTCEGLDL 82
tr|O77690|O77690_BOVIN/1-416 S.HVSRNCQDGSYEISVLMMNNSA...FPESLDSLEEVVKEGVKIVSQRLL.KAGLNVTVN..ATFIYSEGVIYKS.SDCRSSTCEGLDL 82
tr|B8ZHI6|B8ZHI6_ANGAN/1-421 S.LMVRDCLKSAYVLNVVLLEDDV...SEWSLKFVKAGVERAIAIENQRNA.EEGLNFKLT..ANYCGFNTSSYRR.RGCGSSTCEGVEI 82
tr|Q75QO1|Q75QO1_ORYLA/1-415 MLDDCLES.NPRYTMNVVLLEDNT...YEWSRPFVQEAVEGAIKKDAEENR.KAGLNFTLT..ANYNWFNTNLYNR.QGCGSSTCEGVAI 82
tr|O42440|O42440_ORYLA/1-440 C.FCLLPGCRSNITAAVMLPDNYH..KYPWALPRVFPALLMAQEDLHTKHK....LLLGHT..ITILNYSTENPAAPGSCAESRAQVVVV 81
tr|B8ZHI5|B8ZHI5_ANGAN/1-419 S.PSANACPQGEDIINVVLLDDNV...SQWSLDFVKNAVNEAIIHDNELNV.AAGVGFNMT..ASYDGYKTNQYQR.KGCGSSTCEGVEL 82
tr|P79991|P79991_XENLA/1-416 D.LLEANCMSGSLTMNVIMLNDSM...TEWNIKAVQEAVSIGMHVVTKDLE.REGIKVTIN..ADFQTFNTDLYAT.PGCVSSGCEGVEK 82
tr|O42129|O42129_ORYLA/1-418 CVQDTGQCMDG.ITVNVILLEDEE...SPWSLKYVGGQILEAIEKDAAINA.EEGMEFNLT..VNFEGFNTTLYRQ.RGCITSACEGAEK 82
tr|Q4TOM0|Q4TOM0_TETNG/1-443 T.CCLLPGCRGNITVAVMLPDNHH..KYPWALPRVFPAILMAHEDLQSKHG....LLLGRS..INIWNYSTEDPTA.GSCAESRAQVVAV 80
consensus .....v.lp.....v.pa...a.....l.....C.....


tr|Q1LX84|Q1LX84_DANRE/1-439 DLKFS...YNPWAFIGPGCDYSSSPVARFTT..HWEVPMITSGA.....RALGF....NLYSSITNIGP 134
tr|Q6P4O9|Q6P4O9_XENLA/1-445 DLQFT...HHPEVFLGPGCIYTAAPVARFTA..HWKVPLITVGA.....SAYGFNDKTDVYHYTTRTGL 138
tr|Q90YB7|Q90YB7_RANCA/1-444 DMQFI...YHPYVFLGPACIYSAAPVVRFTE..HWKVPLITAGA.....SAIGFGVKDEEYKYITRTGV 138
sp|P16066|ANPRA_HUMAN/1-445 DLKWE...HNPAVFLGPGCVYAAAPVGRFTA..HWRVPLLTAGA.....PALGFGVK.DEYALTTRAGP 137
tr|Q9YI17|Q9YI17_SQUAC/1-448 DLKLE...SDPDVFVGPGCIYTTAPVARFAT..HWRLPLVTAAA.....SAFGFSNKTGEYNTTIRTGP 136
tr|A1L2S6|A1L2S6_XENLA/1-441 DLKLY...HNPDVLFGPGCVYPSASVARFAT..HWRLPLITAGA.....LAFGFKQDDDHYNTTVRTGP 136
tr|Q98UI0|Q98UI0_ORYLA/1-439 DLEFA...YKPWAFIGPGCSYTASPVGLFTT..HWDVPLITAGA.....PAIAFLD..GIYPTITNTGP 136
tr|Q9DG04|Q9DG04_XENLA/1-441 DIAMQ..LQKPDVILGPVCEYAAASVARLAS..HWNVPMLSSGA.....LAVGFMQKSSEYSHLTRVSP 132
tr|Q98UI1|Q98UI1_ORYLA/1-445 DLKLS...HDPWAFIGPGCDYSSSPVARFTT..HWDVPMVTAGA.....RADGF....SKFAAVTNTGP 134
sp|P18910|ANPRA_RAT/1-445 DLKWE...HSPAVFLGPGCVYSAAPVGRFTA..HWRVPLLTAGA.....PALGIGVK.DEYALTTRTGP 137
tr|Q98845|Q98845_ANGJA/1-436 SRSCT...KKPDLILGPVCEYAAAQVVRMAS..HWNIPVISAGA.....LATGFSHKEKEYSHLTRIAP 129
tr|Q90223|Q90223_ANGJA/1-437 DRSCQ...KKPDLILGPVCEYAAAPVARMAS..HWNIPVISAGA.....LASGFSYK.KEYSHLTRVVP 130
sp|P55202|ANPRB_ANGJA/1-441 DTKLY...EKPDAFFGPGCVYSVASVGRFVN..HWKLPLITAWA.....PAFGFDSK.EEYRTIVRTGL 137
sp|P17342|ANPRC_HUMAN/1-449 DRVAAARGAKPDLILGPVCEYAAAPVARLAS..HWDLPMLSAGA.....LAAGFQHKDSEYSHLTRVAP 138
sp|P20594|ANPRB_HUMAN/1-439 DLKLY...HDPDLLLGPGCVYPAASVARFAS..HWRLPLLTAGA.....VASGFSAKNDHYRTLVRTGP 131
tr|Q4SW10|Q4SW10_TETNG/1-437 DLKLA...HNPWAFIGPGCSYTSSPVGLFTT..HWDIPMVTAGA.....PGHGLLR..QRLPIRHQHGP 136
tr|Q7PTP9|Q7PTP9_ANOGA/1-442 DIFLK...RKPDVFFGPICDYVIAPIARYSS..VWGIPLITSGG.....LTEAFTLKAPHYRTLTRMMG 125
tr|Q174S2|Q174S2_AEDAE/1-451 DILLK...RKPDVFFGPICDYVIAPIARYNA..VWGIPILTTGG.....LADAFTIKSPNYPTLTRMMG 134
tr|Q9BPR0|Q9BPR0_BOMMO/1-453 EFYVN...GSADAFIGPGCEYVIAPVARYAG..PWGIPVLTAGA.....QAEAFFYKHPSYATLTRMMG 132
tr|Q75Q00|Q75Q00_ORYLA/1-435 DRSCA...RKPDLILGPVREYEAAGVARLAS..HWDIPMISAGA.....LAAGFGNKNSEFSQLTRIAP 131
tr|Q9PWH0|Q9PWH0_XENLA/1-445 DLQFT...HHPEVFLGPGCIYTAAPVAVSLP..IGRSPLSLWGP.....PPMASMTKPTSTSTPPRTGL 138
tr|B3NK16|B3NK16_DROER/1-458 DMYFQ...DDVIAFIGPACAFALEPVARLAA..YWNTPIITGMGDQPPSSEGELTVTSGILGRI.HKWKNENTGMFKDKSKYPTLTRMSY 158
tr|Q9VF17|Q9VF17_DROME/1-451 EIYTQ..WPEVNAVFGLPCEYVLAPISRYAD..VWQVPVLTTGG.....NAKEFNKKSESYSTLTRLKG 137
tr|B4KRH8|B4KRH8_DROMO/1-458 DMHFK...DDVIAFIGPACAFALEPVARLAA..YWNTPIITGMGDQPPSSEGELTVTSGMLGRI.HKWKNESTGMFKDKSKYPTLTRMSY 158
tr|O97053|O97053_STIJA/1-434 DHYFQ...YYVDVFLGPACEFSAAPVGRFTA..HWNVPMITAGG.....NAQGF....DQFTSMTRHGS 129
tr|B4K5Q8|B4K5Q8_DROMO/1-451 EIYTQ..WPEMNAIFGLPCEYVLAPISRYAD..VWQIPLLTTGG.....NAGEFSKKTESYSTLTRLKG 137
tr|B0W1T4|B0W1T4_CULQU/1-458 DMHFK...DNVIAFIGPACAFALEPVARLAD..YWNTPIITGMGDQPP.SEGELSVTSGILGRLSNKWKNESTGIFKDKSKYQTLTRMSY 158
tr|Q7Q8C9|Q7Q8C9_ANOGA/1-448 DLHFK...HSVIAFIGPACAFALEPVAQLAD..YWNTPIITGMGDQPP.SEGELSVTSGILGRLSNRWKNDSSGMFKDKSRYQTLTRMSY 148
sp|P55204|GUC2C_PIG/1-417 LRTISSEKRMGCVLLGPSCTYSTFQMY.LDT..DLNYPMISAGS.....FGLSCDYKETLTRLMS 139
sp|P70106|GUC2C_CAVPO/1-417 LREIARQKRMGCALMGPSCTYSTYQMY.LDT..ELNYPMISAGS.....FGLSCDHKETLTRMMS 139
sp|Q3UWA6|GUC2C_MOUSE/1-417 LREITRDHKMGCALMGPSCTYSTFQMY.LDT..ELNYPMISAGS.....YGLSCDYKETLTRILP 139
tr|D2H857|D2H857_AILME/1-417 LRAISNNKRMGCVLMGPSCTYSTFQMY.LDA..DLNYPMISAGS.....FGLSCDYKETLTRLMS 139
sp|P25092|GUC2C_HUMAN/1-417 LRKISNAQRMGCVLIGPSCTYSTFQMY.LDT..ELSYPMISAGS.....FGLSCDYKETLTRLMS 139
tr|O77690|O77690_BOVIN/1-416 LRTISSKKQMGCVLMGPSCTYSTFQMY.LDT..DLNYPMISAGS.....FGLSCDYKETLTRMMS 139
tr|B8ZHI6|B8ZHI6_ANGAN/1-421 LKRLQGQNEVGCAMLGPSCTYATFQLVDLENGFNSSVPIISAGS.....FGLSCDYKAYLTRLLP 142
tr|Q75QO1|Q75QO1_ORYLA/1-415 LKKLHNTGEVGCVMLGPSCTFATFQLVDEEIGLSLSIPVISAGS.....FGLSCDYKPKLTRILP 142
tr|O42440|O42440_ORYLA/1-440 DAKLY...SRPDVFFGPGCVYPLASVGRFVS..HWKLPLITAGG.....PAYGFEKL.DEYRTIVRIGP 139
tr|B8ZHI5|B8ZHI5_ANGAN/1-419 LKTLTREDRTGCVLLGPTCTYATFQMVDTEVGLIMGLPIVSAGS.....FGLSCDFKDNLTRLLP 142
tr|P79991|P79991_XENLA/1-416 LKNLRHTRRLGCVILGPTCTYATYQMLSLKN..TFGVPLISAGS.....FGLSCDYHRSLARMLL 140
tr|O42129|O42129_ORYLA/1-418 LNKLMVTGELGCAVLGPTCTYATFAIADVEKGFNLSTPIISAGS.....FGSSCDYAMNLQRLLP 142
tr|Q4TOM0|Q4TOM0_TETNG/1-443 DAKLY...IRPDAFFGPGCVYPLASVGRFAS..HWKLPLVTAGG.....PAYGFDKL.GEYETIVRSGP 138
consensus d.....Gp.C.y....v.r.....w..P.i.ag.....f.....y..ltr...


tr|Q1LX84|Q1LX84_DANRE/1-439 T.HKKLGEFVLRM.....HRHFGWDKHAMLMFNDNK...ND.DRPCYFAV...EGPYTQMRED.....NITADDLVFN.ED.EEPLRY 202
tr|Q6P409|Q6P409_XENLA/1-445 I.PTKLGLFVGHL.....HQMYNWTSRAMIVYRDSN...VD.DRPCFFTM...EGLYMELLKF.....NLTVVDLQFK.DK..ELTNY 205
tr|Q90YB7|Q90YB7_RANCA/1-444 V.HSKLAQFLMHI.....HQQYNWSSRAMLIYSDDK....D.DRACYFTI...EGAFVELPKFN....NMTVVDMNMK.EY..GVINH 205
sp|P16066|ANPRA_HUMAN/1-445 S.YAKLGDFVAAL.....HRRLGWERQALMLYAYRP...GD.EEHCFFLV...EGLFMRVRDRL....NITVDHLEFA.ED..DLSHY 205
tr|Q9YI17|Q9YI17_SQUAC/1-448 A.YAKLGEFANHI.....HETFNWTSRVALLYLDLK...TD.ERHFFFVT...EGIFTSLQEEFA...NLSMHPHHIG.KEELDQSAI 207
tr|A1L2S6|A1L2S6_XENLA/1-441 T.AIKLGEFVSHL.....HEHFNWSSRAALVYHDVK...MD.DRPHYFII...EGVFLALDKEFN...NLTVSYQMYP.EN....EDI 203
tr|Q98UI0|Q98UI0_ORYLA/1-439 T.HKKLGRFALRI.....CEHFGWREYVTLMFSDNK...ED.DRPCYFAM...EGLYEELKSI.....NISLQDSVFE.EN.KPPINY 204
tr|Q9DG04|Q9DG04_XENLA/1-441 V.YSKMGEMFLAM.....FRYHKWTKAFLLYTDD.....TQ.QRNCFFTL...EGVHLAFKEE.....GYAMSIHNFD.ET..KHVDA 197
tr|Q98UI1|Q98UI1_ORYLA/1-445 T.HKKLGEFSLKI.....QETFGWHHHTMLIFSDNKDD.ND.ERTCYFAI...EGLYSLMGKH.....NITVFDYFVE.SN....INH 201
sp|P18910|ANPRA_RAT/1-445 S.HVKLGDFVTAL.....HRRLGWEHQALVLYADRL...GD.DRPCFFIV...EGLYMRVRERL....NITVNHQEFV.EG..DPDHY 205
tr|Q98845|Q98845_ANGJA/1-436 S.YLKMGETFSAL.....FEHFGWNKVLLIFEDD.....SE.ERNCYFTI...EGVHSSLHVE.....GYKVDSVVIH.KD..HRVET 194
tr|Q90223|Q90223_ANGJA/1-437 S.YLKMAETFSAM.....FHRFNWKNAFLIYEDD.....MD.QRNCYFTL...EGVHNILKTE.....NVHIDALNIHSKD..NKVDS 196
sp|P55202|ANPRB_ANGJA/1-441 S.TTKLGEFAHYL.....HSHFNWTTRAFMLFHDLK...VD.DRPYYFIS...EGVFLVLRRE.....NITVEAVPYD.DQ..KNSDY 204
sp|P17342|ANPRC_HUMAN/1-449 A.YAKMGEMMLAL.....FRHHHWSRAALVYSDD.....KL.ERNCYFTL...EGVHEVFQEE.....GLHTSIYSFD.ET..KDLDL 203
sp|P20594|ANPRB_HUMAN/1-439 S.APKLGEFVVTL.....HGHFNWTARAALLYLDAR...TD.DRPHYFTI...EGVFEALQGS.....NLSVQHQVYA.RE...PGGP 197
tr|Q4SW10|Q4SW10_TETNG/1-437 T.HMKLGR.SPCT.....SASTSWRQHVMLIFSDKK...MD.DRPYYFAM...EGLYMELKHT.....NITLAERVFE.ED.PALVDY 203
tr|Q7PTP9|Q7PTP9_ANOGA/1-442 N.YHAFGLMMREI.....HRHYNWTIQAYLYHEFDEKSGRG.FTDCSMAI...TSINRAIGGN.....ETSSGTFDEE.TA..KYADY 195
tr|Q174S2|Q174S2_AEDAE/1-451 S.YSDPGLALREM.....YRHFNWTIQAFIYHDNDEKRGMG.HSDCSMAI...LSIFRVLNTT.....EYFSHSFDET.ET..DYKGY 204
tr|Q9BPR0|Q9BPR0_BOMMO/1-453 S.YTQAGVAIRNI.....FEEFNWRRLGMLYHNNGPSSGKG.NSPCFLTL...SAVFTVLNKKTSGSDNIITAQFDET.NT..TNTKF 207
tr|Q75QO0|Q75QO0_ORYLA/1-435 H.YVKMAETFSAL.....FERFGWRSALLLYEED.....KQ.ERNCYFTL...EGVYHLMS.....DYPVSQYPVL.EG..DPLHV 194
tr|Q9PWH0|Q9PWH0_XENLA/1-445 I.PTKLGLFVAHL.....HQQYNWTSRAMIVYRDSS...VD.DRPCFFTM...EGLYMELPKF.....NLTVVDLQFK.DK..ELTNY 205
tr|B3NK16|B3NK16_DROER/1-458 C.QCRLILVFASV.....IRQFNWNHVALLV..........DRSELFSWTVGKNLEYGLRQE.....GLLSFVKELN.GN..EEEVY 221
tr|Q9VF17|Q9VF17_DROME/1-451 AQVNNLGNVVRAI.....LNSFNWTRTALIYQNEN.AKVKG.NSVCFLCL...AAIHDTIEEH.....SVYQLGFDTS.TW..TKADI 207
tr|B4KRH8|B4KRH8_DROMO/1-458 C.QCRLKLVFASV.....FRQFNWKHVALLV..........DRSELFSLTVGKNLEYGLRQE.....GLLSFVRELN.GN..EEEIY 221
tr|O97053|O97053_STIJA/1-434 P.YTKLGTMILDF.....TNKFSWSSISPMVHEES....GD.YNDYSFLC...GAIYFEMFRI.....AHNVSFVTFN.QN..REVDS 195
tr|B4K5Q8|B4K5Q8_DROMO/1-451 AQVNNLGNMVRAL.....INTYNWTRTALIYQNEN.AKIKG.NSVCFLCL...AAIHATIEKE.....SVYQLGFDTA.HW..TKAYI 207
tr|B0W1T4|B0W1T4_CULQU/1-458 C.QCRLKLVFSSI.....FKQFGWKHVGLLL..........DRSDLFSLTVGKNLEYGLKEE.....DVLTFMRELD.GN..DEEDL 221
tr|Q7Q8C9|Q7Q8C9_ANOGA/1-448 C.QCRLKLVFSSI.....FRQFGWRHIALII..........DRSDLFSLTVGKNLEYGLKDE.....ELLKFVRELD.GN..DEEDI 211
sp|P55204|GUC2C_PIG/1-417 P.ARKLMYFLVDFWKVNNFPFKPFSWNT.AYVFKNS.....TE.SEDCFWYL...NALEAGVSYF.....SQKLSFKEML.RG...NEEF 209
sp|P70106|GUC2C_CAVPO/1-417 P.ARKLMYFLVDFWKASNLPFKSFSWNT.SYVFKNG.....TE.SEHCFWYI...NALEAGVSYF.....SQVLGFKEML.RG...NEEL 209
sp|Q3UWA6|GUC2C_MOUSE/1-417 P.ARKLMYFLVDFWKVNNASFKPFSWNS.SYVYKNG.....SE.PEDCFWYL...NALEAGVSYF.....SEVLNFKDVL.RR...SEQF 209
tr|D2H857|D2H857_AILME/1-417 P.ARKLMYFLVNFWKVNDLPFKSFSWNS.AYVFKNG.....SE.TEDCFWYL...NALEAGVSYF.....SQELSFKEML.RG...NDQF 209
sp|P25092|GUC2C_HUMAN/1-417 P.ARKLMYFLVNFWKTNDLPFKTYSWST.SYVYKNG.....TE.TEDCFWYL...NALEASVSYF.....SHELGFKVVL.RQ...DKEF 209
tr|O77690|O77690_BOVIN/1-416 P.ARKLMYFLVEFWKV.KFQFKPFSWNT.AYVFKNS.....TE.TEDGLWYL...NALAAGVSYF.....PQKLGFKEML.RG...DTEF 208
tr|B8ZHI6|B8ZHI6_ANGAN/1-421 P.ARKISNLFIEFLRFEESSLKP.RWEA.VYVYKKP.....EN.SEDCFWYI...NALDAPSAAF.....NSAIT.RKVL.RK...PEEL 210
tr|Q75QO1|Q75QO1_ORYLA/1-415 P.ARKVSDSLVYFFNEKNM.LKP.IWEK.AYVYKKS.....NNVTEDCFWYN...NALEAPSAHF.....ASSKK.REML.RG...EEEL 210
tr|O42440|O42440_ORYLA/1-440 S.TTKLGAFVNVL.....HTQFNWTSRAVVIFYDLK...QD.DRPHYFLS...EGIFMNLKDEM....NMTVSARPYT.NE....QDY 205
tr|B8ZHI5|B8ZHI5_ANGAN/1-419 P.ARKISSFFLDFWNYSEF.LKA.KWTT.AYVYKKP.....EQ.TEDCFWYI...NALEARSAEF.....SSNVA.RHVL.QR...PEDL 209
tr|P79991|P79991_XENLA/1-416 P.ARKITYFFKEFWQYEDF.IKPKKWQS.VYIYKWD.....GN.TESCFWYI...NALESGVSYF.....NNALKFKEIL.RT...EGEL 209
tr|O42129|O42129_ORYLA/1-418 P.ARKISDFFINFWKEKFT.IKP.KWRT.AYVYKRQ.....PN.TEDCLWYI...GALEA.DGRF.....LVNVS.RTIL.RH...PGDL 208
tr|Q4TOM0|Q4TOM0_TETNG/1-443 S.TTKLGDFTIAL.....HTHFNWTSRAVVLFHDMR...QD.DRPHYFLS...EGIFINLKNQM....NITMEAQPYE.DD....TEY 204
consensus ....kl.....f.W...a.....c.f.....


tr|Q1LX84|Q1LX84_DANRE/1-439 DELLRD.ISHKARVVYVCCKWETFRKLMVEFWRQGFP..QEEYAFFFI.DLFGRS..LQS....HPARPWA..RGDADDN...AAKEAFK 277
tr|Q6P409|Q6P409_XENLA/1-445 TTLIQD.IKQKGRIIYMCCYPDMFRQLMIQAWREGLC..SGDFAFFYV.DVWGAS..LQSSIFPDPKRPWY..RGDADDA...KAREAFK 284
tr|Q90YB7|Q90YB7_RANCA/1-444 TSTIQY.IKQRGRIIYICCPPDDFRQLMLQAWRQGLC..SGDYVFFYI.DNYGAS..LQGSTFPDYRRPWY..RDDADDA...NAREAFK 284
sp|P16066|ANPRA_HUMAN/1-445 TRLLRT.MPRKGRVIYICSSPDAFRTLMLLALEAGLC..GEDYVFFHL.DIFGQS..LQGGQGPAPRRPWE..RGDGQDV...SARQAFQ 284
tr|Q9YI17|Q9YI17_SQUAC/1-448 IEIIQF.IKQHARIVYLCFPFEDFRQIMFYAQKEGLT..GGDYVFFYL.DVFGES..LKVKSPGESYKPWQ..MNHSSES...VLKEAFK 286
tr|A1L2S6|A1L2S6_XENLA/1-441 GSVIQF.IQNNGRVIYICGPLEMLHMIILQAHREKLT..DGDYVFFYV.DVFGES..LRPDGTREANKPWQ..GNHSQ.....ELKEAFK 280
tr|Q98UI0|Q98UI0_ORYLA/1-439 SQILAD.IQNIGRVMFVCCSPDVFRRLMIEFWKADLP..HEQYVFLYI.DLFAVS..LSN......KQPWA..RGDQDDT...IAKDAFQ 277
tr|Q9DG04|Q9DG04_XENLA/1-441 EEIVHA.IQNKERVVIMCASSDTVRNIMLAAHRQGMT..NGDYVFFNI.ELFNSS..TYG......NGSWK..RGDKYDL...EAKQAYS 270
tr|Q98UI1|Q98UI1_ORYLA/1-445 KELVQI.IQKNGRVVYLCCSWDNMRSLMVQFWKEGVD..LENYVFFFI.DLFAEG..LGGE...KPGMPWF..RGDQDDH...AARLAFR 277
sp|P18910|ANPRA_RAT/1-445 PKLLRA.VRRKGRVIYICSSPDAFRNLMLLALNAGLT..GEDYVFFHL.DVFGQS..LKSAQGLVPQKPWE..RGDGQDR...SARQAFQ 284
tr|Q98845|Q98845_ANGJA/1-436 DEIIKD.IYKTEAVVVMCAGGDTVRDIMLAAHRRRLT..SGGYIFFNI.ELFNSS..SYG......DGSWR..RGDKYDA...EAKLAYS 267
tr|Q90223|Q90223_ANGJA/1-437 DEIIKL.IYDSE.VIIMCAGADIIRDIMLAAHRRRLT..NGSYIFFNI.ELFNSS..SYG......NGSWK..RGDKFDM...DAKQAYA 268
sp|P55202|ANPRB_ANGJA/1-441 REMISS.LKSNGRIVYICGPLDTFLEFMRIFQNEGLP..PEDYAIFYL.DMFAKS..ILD....KDYKPWE..SSDINWT...DPIKLFK 279
sp|P17342|ANPRC_HUMAN/1-449 EDIVRN.IQASERVVIMCASSDTIRSIMLVAHRHGMT..SGDYAFFNI.ELFNSS..SYG......DGSWK..RGDKHDF...EAKQAYS 276
sp|P20594|ANPRB_HUMAN/1-439 EQATHF.IRANGRIVYICGPLEMLHEILLQAQRENLT..NGDYVFFYL.DVFGES..LRAGPTRATGRPWQDNRTREQAQ...ALREAFQ 278
tr|Q4SW10|Q4SW10_TETNG/1-437 AQILAD.IRNEGRVIFVCCKPDIFRRLMVQFRREDLP..HHHYVFFYI.DVFGFG..LRA......GRPWY..RGDQDDA...AAREAFQ 276
tr|Q7PTP9|Q7PTP9_ANOGA/1-442 LRLLRN.IKKRARIVIMCASPSTIREIMLAAAELNMVN.SGEYVFFNI.EIFGSMT.ATK......QPPWY..AKNDTDERNQKAKEAFT 273
tr|Q174S2|Q174S2_AEDAE/1-451 LRILEE.TKRKARIVIMCASPSTIREIMLAAAELNMVD.SGEYVFFNI.DIFSSMA.ATK......IPSWH..MANDTEERNLKAKNAYT 282
tr|Q9BPR0|Q9BPR0_BOMMO/1-453 KELLHK.LSLSTRIVVLCANPATVREIMLAADDLNMVS.SGEYVFFNI.ELFSNLASASS......KEPWK..SENDTEERNERARRAYS 286
tr|Q75Q00|Q75Q00_ORYLA/1-435 DEIIHS.MNDSE.VVIMCMGAERIRGIMLAAHRHQLT..RGRKIFFSI.ELFNAS..SYG......DGSWR..RDDEHDS...EAKQAYA 266
tr|Q9PWH0|Q9PWH0_XENLA/1-445 TTLIQD.IKQKGRIIYMCCYPDMFRQLMIQAWREGLC..SGDFAFFYV.DMWGAS..LQSSIFPDPKRPWY..RGDADDA...KAREAFK 284
tr|B3NK16|B3NK16_DROER/1-458 ENYLKD.ASMYARVVILSVRGVLVRKFMLAAHSLGMT..NGEWVFLDV.EIFQSE..YWG......DKGWE..MKDEHDA...KARKAYE 294
tr|Q9VF17|Q9VF17_DROME/1-451 TRMLKN.VAMQSRIVIMCADPQSIRQIMLTAEELNMID.SGEYVFINI.ELFSRVQ.YLT......SQPWY..DKNDTDLNNERAQKAYT 285
tr|B4KRH8|B4KRH8_DROMO/1-458 ENYLRD.ASMYARVVILSVRGVLVRKFMLAAHSLGMT..NGEWVFLDV.EIFQSA..YWG......DKDWE..LGDENDM...KARKAYE 294
tr|O97053|O97053_STIJA/1-434 ETILKERVSPKARVVFICASSDTVRQIMIEAHTMEMT..KGEYAFFSV.NPFDSK..YFG......DPSWY..RQSDSDVINKKAREAYR 272
tr|B4K5Q8|B4K5Q8_DROMO/1-451 IRLLRN.VAMRTRIVIMCADPQSIRQIMLTAEELNMID.SGEYVFINI.ELFSRVP.YMT......SLPWY..DKNDTDFNNERAKKAYT 285
tr|B0W1T4|B0W1T4_CULQU/1-458 DNYLQD.ASMYARVIILSVRGSLVRKFMLSAHRLGMT..RGEFTFLDV.EIFQSS..YWG......DHYWE..LGDEDDQ...AARKAYQ 294
tr|Q7Q8C9|Q7Q8C9_ANOGA/1-448 EAYLKD.ASMYARVIILSVRGSLVRKFMLSALALGMT..RGEFTFLDV.EIFQSS..YWG......DHYWE..LGDEDDF...KARKSYE 284
sp|P55204|GUC2C_PIG/1-417 QNILMD.QNRKSNVIIMCGAPETVHTLK..GGRAV....AEDTVIILV.DLFNDH..YFM.................DNV...TAPDYMK 269
sp|P70106|GUC2C_CAVPO/1-417 QKILKD.PNRRSNVIVMCGTPQTMESLK..IDWTA....TEDTVIILV.DLFNNY..YLE.................ENA...TAPDYMK 269
sp|Q3UWA6|GUC2C_MOUSE/1-417 QEILTG.HNRKSNVIVMCGTPESFYDVK..GDLQV....AEDTVVILV.DLFSNH..YFE.................ENT...TAPEYMD 269
tr|D2H857|D2H857_AILME/1-417 QDILTN.QNRKSNVIVMCGTPSVISNLK..GDRAV....AEDIVIILV.DLFNNH..YFM.................ENV...TAPDYMK 269
sp|P25092|GUC2C_HUMAN/1-417 QDILMD.HNRKSNVIIMCGGPEFLYKLK..GDRAV....AEDIVIILV.DLFNDQ..YLE.................DNV...TAPDYMK 269
tr|O77690|O77690_BOVIN/1-416 QDILMD.QNRKSNVIVMCGRPETIQNLR..GNRAV....AEDIVIILV.DLFNDH..YFM.................DNV...TAPDYMK 268
tr|B8ZHI6|B8ZHI6_ANGAN/1-421 HSAFER.KNRTSNLFIICGTPEDVANLT.DNGRRL....EPDIVILLI.DLYNHE..YH..................SAA...GS.PAMS 269
tr|Q75QO1|Q75QO1_ORYLA/1-415 KKALTS.KYRFSNIFILCGSVDDIVSIK.GLAKQF....HEDTIFILI.DLYNPE..YY..................INT...TSLAPMR 270
tr|O42440|O42440_ORYLA/1-440 KELVSF.IKENGRIIYICGPLETFLSIMKLFQSEIQD..PESYAIFYL.DVFAES..LTH......RKPWQ..NAKFDWT...NPIQVFK 278
tr|B8ZHI5|B8ZHI5_ANGAN/1-419 RSEINN.KQRKSNVFILCGGPGDIANLTKDIDRKL....HPEVIFILI.DLYSPA..YH..................HNT...TSIPPME 270
tr|P79991|P79991_XENLA/1-416 MKVLQE.NNHKSNVILMCGTPNDIWNLH..NKVAI....PQDKVLILL.DIFNTV..YY..................DNK...SSPYYME 268
tr|O42129|O42129_ORYLA/1-418 KDVLDSQENRTSNLFILCGSPTDLKEVKNISDAAD....NLDILFILI.DLYNDV..YY..................TNT...TSMPEMR 270
tr|Q4TOM0|Q4TOM0_TETNG/1-443 KELINF.MKEHGRIVYLNDTSSMLVTRHVHIQKGNFISFPLEICCYLLSSVFPIR..ITH......NPPPE.VFLYVCFL...PHPTQIQ 281
consensus .....rv...c.....r..m.............vf....d.f.................w.............a..a..


tr|Q1LX84|Q1LX84_DANRE/1-439 SVKILTYRE.PQNPEYKDFVSKLKTEAMDMFNFNV..EDSLMNLISGSFHDGVMLYSHALNDTMDRSG......SRP......PGDVV.. 350
tr|Q6P409|Q6P409_XENLA/1-445 AVMIITYKE.PDNPEYKEFLANLNRFSGEPFHYKE..ESTLMNALAASFHDSVLLYAHAVNETRNNGY......SMK......NASAV.. 357
tr|Q90YB7|Q90YB7_RANCA/1-444 SVMIITYKE.PENPEYFEFLQDLKSYAPK.FNHTM..ESTLMNTVAADFYDSVMLYAHVVNETREKGE......SIR......NATAI.. 356
sp|P16066|ANPRA_HUMAN/1-445 AAKIITYKD.PDNPEYLEFLKQLKHLAYEQFNFTM..EDGLVNTIPASFHDGLLLYIQAVTETLAHGG......TVT......DGENI.. 357
tr|Q9YI17|Q9YI17_SQUAC/1-448 SVMIITYRH.PEEPEYLEFQEELRRRATNVSSADL..DNALVNFIAGCFYDGVMLYAMALNETLAAGG......SKK......DGLVI.. 359
tr|A1L2S6|A1L2S6_XENLA/1-441 TVLVISYHQ.PENPEYFEFQKKLIQKSKEEFGVEL..NYSLMNFIAGCFHDGVLLYAQALNETLREGG......SQK......DGLSI.. 353
tr|Q98UIO|Q98UIO_ORYLA/1-439 SVKILSYRE.PQNQEYQQFVRDLKADAKTSFNYSV..QDSLMNIIAGGFYDGLMLYAHALNETALVPG......ARP......PGKLI.. 350
tr|Q9DG04|Q9DG04_XENLA/1-441 SLQTVTLLR.TVKPEFEKFSMEVKSSVQKLGLN....DDDYVNMFVEGFHDAIILYALALHELLKNGF......SQK......DGEKL.. 341
tr|Q98UI1|Q98UI1_ORYLA/1-445 SVKVLTYME.PQNAEYHQFVETLKKDAEKMFNFTI..KDSLYNLIAGGFYDGVMLYSRALNETLSKRKPGLRPVQRP......KGDMV.. 356
sp|P18910|ANPRA_RAT/1-445 AAKIITYKE.PDNPEYLEFLKQLKLLADKKFNFTV..EDGLKNIIPASFHDGLLLYVQAVTETLAQGG......TVT......DGENI.. 357
tr|Q98845|Q98845_ANGJA/1-436 ALNVVTLMR.TAKAEFETFTTEVKKSIQRAGIGP...DSANVNMFMEGFHDALLLYALALHEVVKNGF......SKK......DGVQI.. 339
tr|Q90223|Q90223_ANGJA/1-437 TLNTVTLLR.TVKPEFEDFSMEVKKSLQKAGIRHC..DSDNINVFVEGFHDALLLYAMAVVEVTQNGS......NKT......DGARI.. 341
sp|P55202|ANPRB_ANGJA/1-441 SVFVITAKE.PDNPEYKAFQRELHARAKQEFSVQL..EPSLEDIIAGCFYDGFMLYAQALNETLAEGG......SQN......DGINI.. 352
sp|P17342|ANPRC_HUMAN/1-449 SLQTVTLLR.TVKPEFEKFSMEVKSSVEKQGLN....MEDYVNMFVEGFHDAILLYVLALHEVLRAGY......SKK......DGGKI.. 347
sp|P20594|ANPRB_HUMAN/1-439 TVLVITYRE.PPNPEYQEFQNRLLIRAREDFGVEL..GPSLMNLIAGCFYDGILLYAEVLNETIQEGG......TRE......DGLRI.. 351
tr|Q4SW10|Q4SW10_TETNG/1-437 SVKILTYRE.PENLEYQQFLSTLKTDAKLMFNYTI..QDSLMNIIAGGFYDSVMLYAQVLNETMATAG......DRP......AGKLV.. 349
tr|Q7PTP9|Q7PTP9_ANOGA/1-442 ALLQVVARE.PEDEEYRQFSKEVKELTKTKYNHTY.AEDEPVSTFVTAFYDAVLLYAYALNDSIAQLG......VERALRQPINGTHL.. 353
tr|Q174S2|Q174S2_AEDAE/1-451 AMLQVVARQ.PEDEEYRRFSEEVKLLTKTKFNYTY.AEDEPVSTFVTAFYDAVLLYAYALNDSIGLLG......EQRALKQPINGTYL.. 362
tr|Q9BPR0|Q9BPR0_BOMMO/1-453 AVLTVTSPA.PEKKEYLEFSDQVKELAATKYNYTF.GKGEVVSTFVAAFYDAVLLYALALNDTLQQAT......DPR...GQLDGAAV.. 363
tr|Q75QO0|Q75QO0_ORYLA/1-435 SLNTITLLR.TVKPEFENFSLEMKSSAEKEGIYDC.KDCGSVNMFVEGFHDAMLLYAIALHEAMKHGY......SKK......NGTEV.. 340
tr|Q9PWH0|Q9PWH0_XENLA/1-445 AVMIITYKE.PDNPEYKEFLTNLSRFSGEPFHYKE..ESTLMNVLAASFHDSVLLYAHAVNETRTNGY......TMN......NASAI.. 357
tr|B3NK16|B3NK16_DROER/1-458 ALLRVSLLQ.PTSPKFQDFADNVRENALYDYNYTF.GEGEEVNFFIGAFYDGVYLLGMALNETLTEGG......DIR......DGVNI.. 368
tr|Q9VF17|Q9VF17_DROME/1-451 AMLTVTPKQ.PNDNEYTRVSNEIKAIAAEKYNYTF.SDNEPISAFVTSFFDGVLLYANALNESIREDP......TML..TRPINGTDM.. 363
tr|B4KRH8|B4KRH8_DROMO/1-458 ALLRVSLLQ.PTSPTFQDFADNVRENALTEYNYTF.GEGEEVNFFIGAFYDGVYLLGMALNETLTEGG......DIR......DGVNI.. 368
tr|O97053|O97053_STIJA/1-434 ALMTIQLYS.EKSEHYDQFAAKVKEKALAEFGYDFDANGEQVNSFVTAFHDAVILYALALNESLTEGA......NPR......NGTDL.. 347
tr|B4K5Q8|B4K5Q8_DROMO/1-451 AMLTVTPKQ.PNDDAYTKVSNEIKDIASAKYNYTF.SENEPISAFVTSFFDGVLLYANALNESIREDP......SML..TRPINGTDM.. 363
tr|B0W1T4|B0W1T4_CULQU/1-458 ALLRVSLLQ.PTSPSYQYFAEKVRQRAKRDYNYTF.VEDEEVNFFIGAFFDGVYLLGMALNETLNEGG......NIR......DGSAI.. 368
tr|Q7Q8C9|Q7Q8C9_ANOGA/1-448 ALLRVSLLQ.PTSPTYQYFAEKVRALAKQDYNYTF.VEDEEVNFFIGAFFDGVYLLGMALNDTLNEGG......DIR......DGTAI.. 358
sp|P55204|GUC2C_PIG/1-417 NVLVLTLPP.ENSVSNSSFSKD..............LSLVKNDFTLAYMNGVLLFGHMLKIFLEKR.......EDV......TTSKF.. 328
sp|P70106|GUC2C_CAVPO/1-417 NVLVLTLPP.GNSTINTSLSKE..............SLQEFSDFALAYLDGILLFGHMLKTFLRNG.......ENT......TAHKF.. 328
sp|Q3UWA6|GUC2C_MOUSE/1-417 NVLVLTLPS.EQSTSNTSVAER..............FSSGRSDFSLAYLEGTLLFGHMLQTFLENG.......ENV......TGPKF.. 328
tr|D2H857|D2H857_AILME/1-417 NVLVLTLPP.ENSTSISSFSKG..............LSQAKNNFALAYLNGILLFGHMLKIFLENG.......EAI......TTPKF.. 328
sp|P25092|GUC2C_HUMAN/1-417 NVLVLTLSP.GNSLLNSSFSRN..............LSPTKRDFALAYLNGILLFGHMLKIFLENG.......ENI......TTPKF.. 328
tr|O77690|O77690_BOVIN/1-416 NVLVLTLPP.ENSVSNSSSSKN..............LSLAKNDFAAAYLDGVLLFGHMLKIFLENG.......EDV......TTSKF.. 327
tr|B8ZHI6|B8ZHI6_ANGAN/1-421 NVLVITMPNIRNYTE..GWTDN..............GTLPEMNDYVAGYHDGVHLFGLVLRQKMLYGEG.....SVE......ENASV.. 330
tr|Q75QO1|Q75QO1_ORYLA/1-415 DVLVVTLPP.RNYVNESNSTFN..................NTINDYVAGYHDSALLFGEVLRRKMLSQ.................HVPLS.. 324
tr|O42440|O42440_ORYLA/1-440 SVFVITYHP.PDNPEYKDFQRKLHARAQRDFGVNL..EPSLMDYIAGSFYDGFVLYAMALDETLAEGG......AQN......NGINI.. 351
tr|B8ZHI5|B8ZHI5_ANGAN/1-419 NVLVLTMPK.RNFTNEIDPSTN..............ETLMTDYMAAYGDGVLLVGQVIRRLWEENP......GRK......QFSIN.. 329
tr|P79991|P79991_XENLA/1-416 NVLVVTQRP.SNMSKISNQTGI..............AKLLEDNYAAGYLDGVLLFGHILKKFLGSV.......DIN......QTFSF.. 327
tr|O42129|O42129_ORYLA/1-418 NVLVLTMPDTRTYTIKPDLTGN..............DTMNDYMAAYHDAVLLVGQVMRDIAIRNP......AEM......QGMEYVN 331
tr|Q4TOM0|Q4TOM0_TETNG/1-443 SVFIITYRP.PDNPEYKDFQKKLHARARRDFGVHL..EPSLMDYIAGSFYDGFVLYAMALEETLADGG......AQN......DGLGI.. 354
consensus .v...t.......e...f.....................n.....f.dg.lLy..al.e....g.................g.....


tr|Q1LX84|Q1LX84_DANRE/1-439 NKRMWNRTYHGVTGLVQLDENGDREIDFALWDMTDTKTGDYQIVSVYNGSQKQMILE.PGMKVHWLKG.....RPPPDIPEC. 426
tr|Q6P409|Q6P409_XENLA/1-445 TSHMRNKSFYGASGFVKIDDSGDRENDFSLWDMYEP.HGTFQIVSHYNGTLRKIIPV.PGQEIQWPGN.....RIPRDFPPC. 432
tr|Q90YB7|Q90YB7_RANCA/1-444 ISHMWNRTYYGVSGFLRVDDNGDRENDYSLWDLSEA.GGDFQVVANYNGTQRSINRV.PGREIHWPGG.....AVPKDVPPC. 431
sp|P16066|ANPRA_HUMAN/1-445 TQRMWNRSFQGVTGYLKIDSSGDRETDFSLWDM.DPENGAFRVVLNYNGTSQELVAV.SGRKLNWPLG.....YPPPDIPKC. 432
tr|Q9YI17|Q9YI17_SQUAC/1-448 TRKMQDRRFPGVTGLVNIDKNGDRDIDFSLWDMVDTETGKYEVVAHYLGIKKQIYWI.PNVEIHWPSG.....SVPIDNPPC. 435
tr|A1L2S6|A1L2S6_XENLA/1-441 VKKIQDRQMEGITGTVSMDKNNDRNTDFDLWAMTDHEEGNFEVVGHYNGITKQINWT..GKPILWLKG.....SPPSDSPPC. 428
tr|Q98UI0|Q98UI0_ORYLA/1-439 SGKMWNRTFHGVTGLLHLDVSGDRETDFALWDLVDTNSSSFQIVVVYNSFEEQVTPV.PGTSVRWLGG.....ARPLDVPKC. 426
tr|Q9DG04|Q9DG04_XENLA/1-441 VQQMWNRTYEGIAGQVSIDANGDRYGDFSVIAMTDKETGTQEVIGDYYGIQGHFEIR.PNVKLPWGPGRLLIN.DRFVEHTNTTPCKSC. 428
tr|Q98UI1|Q98UI1_ORYLA/1-445 TQRMWNRTFQGVMGTVEMDKFGDREIDFALWDMTDINSGKFEVVCVYNSSIKELVLQ.KGLNFQWPGG.....SPPLEVPEC. 432
sp|P18910|ANPRA_RAT/1-445 TQRMWNRSFQGVTGYLKIDRNGDRDTDFSLWDM.DPETGAFRVVLNYNGTSQELMAV.SEHKLYWPLG.....YPPPDVPKC. 432
tr|Q98845|Q98845_ANGJA/1-436 TQSMRNRTFEGIAGQVSIDENGDRNGDFSVMAMTDTQSGTYEAVFNYFGVNQSFQIM.PGFNTDHFTLRERPRPP.....SPEQPDQSSG 423
tr|Q90223|Q90223_ANGJA/1-437 TQRMWNRTFEGIAGPVSIDANGDRYGDFSVMAMVDHETGTYEDVINYFGINGSFQML.PRFNNDRFTLRARHQMS.....MPDYSTKSC. 424
sp|P55202|ANPRB_ANGJA/1-441 TQKMQNRRFWGVTGLVSTDKNNDRDIDFNLWAMTNHKTGQYGIVAYYNGTNKEIVWS.ETEKIQWPKG.....SPPLDNPPC. 428
sp|P17342|ANPRC_HUMAN/1-449 IQQTWNRTFEGIAGQVSIDANGDRYGDFSVIAMTDVEAGTQEVIGDYFGKEGRFEMR.PNVKYPWGPLKLRIDENRIVEHTNSSPCKSSG 436
sp|P20594|ANPRB_HUMAN/1-439 VEKMQGRRYHGVTGLVVMDKNNDRETDFVLWAMGDLDSGDFQPAAHYSGAEKQIWWT..GRPIPWVKG.....APPSDNPPC. 426
tr|Q4SW10|Q4SW10_TETNG/1-437 NARMWNRTFHGVTGKVHLDKNGDRETEFALWDMTDGDSHMFQ.VHTHPHISEHLRMV.PGTIVRWLGG.....VCPPDVPVC. 424
tr|Q7PTP9|Q7PTP9_ANOGA/1-442 AQLMWGRSFKGITGNVTIDSNGDRISNYSLLDL.NPETGLFEVVANYYY.GGGLQFV.EGKAIHWAGDRT.....KAPPDRPTC. 429
tr|Q174S2|Q174S2_AEDAE/1-451 THLMWGKSFKGITGNVTIDSNGDRISDYSLLDL.NPETGMFEIVANYFH.DGGLQFV.EGKEIHWSGGRT.....KAPPDRPIC. 438
tr|Q9BPR0|Q9BPR0_BOMMO/1-453 MRNMWNRTFQGITGEVVINSNGDRVASYSLLDM.NPNTSKFEVVATYVAANKTLQFT.ENRPIYWAGGRT.....TPPPDTPEC. 440
tr|Q75Q00|Q75Q00_ORYLA/1-435 TSRMWNRTIEGIAGQISIDTNGDRNGDFSVMAMTDVEAGTFEVVANYFGVNRTLELL.PSFRAEHFTLKERHEA.....SPLPPEKSC. 422
tr|Q9PWH0|Q9PWH0_XENLA/1-445 TSHMRNKSFYGASGFVKIDDSGDRENDFSLWDMYEA.HGTFQIVSHYNGTLRKIMAL.PGREIQWPGK.....RIPRDVPPC. 432
tr|B3NK16|B3NK16_DROER/1-458 TRRMWNRTFEGITGHVRIDDNGDRDADYSILDL.DPINGKFSVVAHYSGVHKVYSAV.HGKKIHWPGGRE.....EPPPDVPPC. 445
tr|Q9VF17|Q9VF17_DROME/1-451 VRRMWNRSFTGITGNVTIDANGDRLSAYSLLDM.NPTTGRFEIVAHFLH..NRLEFE.ANKEIHWAGDRE.....EAPPDRPIC. 438
tr|B4KRH8|B4KRH8_DROMO/1-458 TRRMWNRTFHGITGHVRIDDNGDRDADYSILDL.DPINGKFSVVAHYYGLHRKYAAA.HGKKIHWPGGRE.....EPPPDIPPC. 445
tr|O97053|O97053_STIJA/1-434 SHRMWNRTFKGIAGDVTIDSNGDRDSDYSLKEM..DSDGEFEVVGIFSGATKAFTML.KGKTIDWPGD.....TVPLDTPKC. 421
tr|B4K5Q8|B4K5Q8_DROMO/1-451 VRRMWNRSFTGITGNVTIDSNGDRISAYSLLDM.NPTTGRFEIVAHFLH..NRLEFE.SEKEIHWAGGRD.....QAPPDRPIC. 438
tr|B0W1T4|B0W1T4_CULQU/1-458 TRKMWNRSFDGITGHVRIDDNGDRDADYSILDL.DPITGRFEVVAHYYGKTREYSPV.KGKRIHWPGGRE.....GPPPDIPKC. 445
tr|Q7Q8C9|Q7Q8C9_ANOGA/1-448 TRKMWGRDFEGITGHVRIDDNGDRDADYSILDL.DPITGRFEVVAHYYGITREYSPV.KGKKIHWPGGRE.....GPPPDVPKC. 435
sp|P55204|GUC2C_PIG/1-417 AHAFRNITFEGHMGPVTLDNCGDIDNTMFLLYT.SVDTSKYKVLLTYDTRKNYTNPVDKSPTFIWKNH.....KLPNDIPGR. 404
sp|P70106|GUC2C_CAVPO/1-417 AHAFRNLTFEGSTGPVTLDDSGDIDNTMVLLYT.SVDTKKFKPLLFYDTRINQTTPIDTHPTFIWKNH.....RLPHDIPGL. 404
sp|Q3UWA6|GUC2C_MOUSE/1-417 ARAFRNLTFQGFAGPVTLDDSGDIDNIMSLLYV.SLDTRKYKVLMKYDTHKNKTIPVAENPNFIWKNH.....KLPNDVPGL. 404
tr|D2H857|D2H857_AILME/1-417 AQAFRNLTFEGHAGPVTLDDCGDIDNTMVLLYT.SVETNKYKVLLKYDTHVNKTTPEVDNPMFIWMNH.....KLPSDIPGQ. 404
sp|P25092|GUC2C_HUMAN/1-417 AHAFRNLTFEGYDGPVTLDDWGDVDSTMVLLYT.SVDTKKYKVLLTYDTHVNKTYPVDMSPTFTWKNS.....KLPNDITGR. 404
tr|O77690|O77690_BOVIN/1-416 AHAFRNITFEGHVGPVTLDACGDIDNTMYLLYT.SVDTSKYKVLLTYDTRVNQTSPVDKSPTFIWKNH.....KLPNDIPGQ. 403
tr|B8ZHI6|B8ZHI6_ANGAN/1-421 ENPFKNISFSGIGGQYVLDEHGDRDVNFSVMYM.STTDSQYKVLFEFDTSTNNTAVVDANPTWHWKSS.....RLPDDRPAQE 407
tr|Q75Q01|Q75Q01_ORYLA/1-415 DTPFGNISFEGMAGNYVLDEYGDRDVNFTFIYT.SAQTSKYETLSVFDTSQNITIMWHDSPTLPWKDG.....QLPGDEPENT 401
tr|O42440|O42440_ORYLA/1-440 TRRTQNRSFWGVTGLVSIDKRNARNIDVDLWAMTNQETGEYGVVSYYNGSTKEIVWS.QTEKIHWPSC.....GPPLDNPPC. 427
tr|B8ZHI5|B8ZHI5_ANGAN/1-419 MEDFRNLSFTGLGGHYVLDEYGDRDVNFSVMYT..TKGMEYKTLFEFDTATGLISVKDDKPDFFWPNY.....LLPDDILVQS 405
tr|P79991|P79991_XENLA/1-416 IDQFRNISIIGALGPLILDAAGDRELNLTLLYS.STATNNYTELIQFDTSTNQTTVMDTSPNFIWKNH.....RLPSDVPQS. 403
tr|O42129|O42129_ORYLA/1-418 TNYFRNVSFNGIGGHYKLDSYGDRDVNFSVIYT..STDNKYKILFSFDTENNRTKQMDPSPTFIWTK.....ALPDDKPGS. 405
tr|Q4TOM0|Q4TOM0_TETNG/1-443 TTKMKNRHMWGVTGLVTTDDKDARNIDVNLWAMTDQNTGEYGIVLYYNGTTKDIVWS.QSEKIHWPGD.....GPPLDNPPC. 430
consensus ...m.nr.f.G..G.v..D..gdr..d.sl.....g....v..y.....w.....p.d.p.c.

```
tr|Q1LX84|Q1LX84_DANRE/1-439   ..GFKNDNPACLAKT   439
tr|Q6P409|Q6P409_XENLA/1-445   ..GFDHSNPECKKSS   445
tr|Q90YB7|Q90YB7_RANCA/1-444   ..GFDNSNPECMKTS   444
sp|P16066|ANPRA_HUMAN/1-445    ..GFDNEDPACNQDH   445
tr|Q9YI17|Q9YI17_SQUAC/1-448   ..VFETDIASCNQAT   448
tr|A1L2S6|A1L2S6_XENLA/1-441   ..VFNADDPSCLKTT   441
tr|Q98UI0|Q98UI0_ORYLA/1-439   ..GFKNDNPACLTKT   439
tr|Q9DG04|Q9DG04_XENLA/1-441   ..GLGESAVTGIVVG   441
tr|Q98UI1|Q98UI1_ORYLA/1-445   ..GFKNDNPACLTST   445
sp|P18910|ANPRA_RAT/1-445      ..GFDNEDPACNQDH   445
tr|Q98845|Q98845_ANGJA/1-436   ..GLGVSAVTGIIVG   436
tr|Q90223|Q90223_ANGJA/1-437   ..GLGVSAVTGITFG   437
sp|P55202|ANPRB_ANGJA/1-441    ..VFSMDEPFCNEDQ   441
sp|P17342|ANPRC_HUMAN/1-449    ..GLEESAVTGIVVG   449
sp|P20594|ANPRB_HUMAN/1-439    ..AFDLDDPSCDKTP   439
tr|Q4SW10|Q4SW10_TETNG/1-437   ..GFKNDNPACLTSE   437
tr|Q7PTP9|Q7PTP9_ANOGA/1-442   ..GFDGSLCPDNSLP   442
tr|Q174S2|Q174S2_AEDAE/1-451   ..GFDGSLCPDKSLP   451
tr|Q9BPR0|Q9BPR0_BOMMO/1-453   ..GFDGSLCPDNSLP   453
tr|Q75Q00|Q75Q00_ORYLA/1-435   ..GLGVSALTGVIVG   435
tr|Q9PWH0|Q9PWH0_XENLA/1-445   ..GFDQSNPECKKST   445
tr|B3NK16|B3NK16_DROER/1-458   ..GFLGNSTDCLGNF   458
tr|Q9VF17|Q9VF17_DROME/1-451   ..GYDGALCPDNSLP   451
tr|B4KRH8|B4KRH8_DROMO/1-458   ..GFLGNAPDCHGNE   458
tr|O97053|O97053_STIJA/1-434   ..GFNGDKCIVDVNN   434
tr|B4K5Q8|B4K5Q8_DROMO/1-451   ..GYDGSLCPDNSLP   451
tr|B0W1T4|B0W1T4_CULQU/1-458   ..GFMGNSPACQRSE   458
tr|Q7Q8C9|Q7Q8C9_ANOGA/1-448   ..GFLGTSPACQGND   448
sp|P55204|GUC2C_PIG/1-417      ..GPQILMIAVFTLT   417
sp|P70106|GUC2C_CAVPO/1-417    ..GPHILLIAVCTLA   417
sp|Q3UWA6|GUC2C_MOUSE/1-417    ..GPQILMIAVFTLT   417
tr|D2H857|D2H857_AILME/1-417   ..GPQALLIAVFTLT   417
sp|P25092|GUC2C_HUMAN/1-417    ..GPQILMIAVFTLT   417
tr|O77690|O77690_BOVIN/1-416   ..GPQMLMIAVFTLA   416
tr|B8ZHI6|B8ZHI6_ANGAN/1-421   QVLLATQDIIVIVL.   421
tr|Q75Q01|Q75Q01_ORYLA/1-415   E.DLSTQDIIVIVLG   415
tr|O42440|O42440_ORYLA/1-440   ..VFSTDDPSCNDGL   440
tr|B8ZHI5|B8ZHI5_ANGAN/1-419   V.HLQIHNIIVIVLA   419
tr|P79991|P79991_XENLA/1-416   ..GPHILTIAVFTLI   416
tr|O42129|O42129_ORYLA/1-418   ..ELETQDIIVVVLG   418
tr|Q4TOM0|Q4TOM0_TETNG/1-443   ..VFSSDDPSCNDVT   443
consensus                      ..g...........
```

X non conserved
X ≥ 50% conserved
X ≥ 80% conserved

Bibliography

Aimoto Saburo, Takao T, Shimonishi Yasutsugu, Hara Saburo, Takeda T, et al. (1982) Amino-acid sequence of a heat-stable enterotoxin produced by human enterotoxigenic Escherichia coli. European journal of biochemistry / FEBS 129: 257--63.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215: 403--10.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25: 3389--402.

Arita M, Takeda T, Honda T, Miwatani T (1986) Purification and characterization of Vibrio cholerae non-O1 heat-stable enterotoxin. Infection and immunity 52: 45--9.

Armougom F, Moretti S, Poirot O, Audic S, Dumas P, et al. (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic acids research 34: W604--8.

Basu N, Arshad N, Visweswariah SS (2010) Receptor guanylyl cyclase C (GC-C): regulation and signal transduction. Molecular and cellular biochemistry 334: 67--80.

Batisson I, Der Vartanian M (2000) Contribution of defined amino acid residues to the immunogenicity of recombinant Escherichia coli heat-stable enterotoxin fusion proteins. FEMS microbiology letters 192: 223--9.

Berman HM (2000) The Protein Data Bank. Nucleic Acids Research 28: 235--242.

Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, et al. (2005) Protein structure prediction servers at University College London. Nucleic acids research 33: W36--8.

Carrithers SL, Ott CE, Hill MJ, Johnson BR, Cai W, et al. (2004) Guanylin and uroguanylin induce natriuresis in mice lacking guanylyl cyclase-C receptor. Kidney international 65: 40--53.

Chang MS, Lowe DG, Lewis M, Hellmiss R, Chen E, et al. (1989) Differential activation by atrial and brain natriuretic peptides of two different receptor guanylate cyclases. Nature 341: 68--72.

Cheng SH, Rich DP, Marshall J, Gregory RJ, Welsh MJ, et al. (1991) Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. Cell 66: 1027--36.

Chinkers M, Garbers DL, Chang MS, Lowe DG, Chin HM, et al. (1989) A membrane form of guanylate cyclase is an atrial natriuretic peptide receptor. Nature 338: 78--83.

Currie MG, Fok KF, Kato J, Moore RJ, Hamra FK, et al. (1992) Guanylin: an endogenous activator of intestinal guanylate cyclase. Proceedings of the National Academy of Sciences of the United States of America 89: 947--51.

Cuthbert AW, Hickman ME, MacVinish LJ, Evans MJ, Colledge WH, et al. (1994) Chloride secretion in response to guanylin in colonic epithelial from normal and transgenic cystic fibrosis mice. British journal of pharmacology 112: 31--6.

de Jonge H (1975) Properties of guanylate cyclase and levels of cyclic GMP in rat small intestinal villous and crypt cells. FEBS Letters 55: 143--152.

de Sauvage FJ, Camerato TR, Goeddel DV (1991) Primary structure and functional expression of the human receptor for Escherichia coli heat-stable enterotoxin. The Journal of biological chemistry 266: 17912--8.

de Sauvage FJ, Keshav S, Kuang WJ, Gillett N, Henzel W, et al. (1992) Precursor structure, expression, and tissue distribution of human guanylin. Proceedings of the National Academy of Sciences of the United States of America 89: 9089--93.

Deshmane SP, Carrithers SL, Parkinson SJ, Crupper SS, Robertson DC, et al. (1995) Rat guanylyl cyclase C expressed in COS-7 cells exhibits multiple affinities for Escherichia coli heat-stable enterotoxin. Biochemistry 34: 9095--102.

Di Guglielmo MD, Park J, Schulz S, Waldman SA (2001) Nucleotide requirements for CDX2 binding to the cis promoter element mediating intestine-specific expression of guanylyl cyclase C. FEBS letters 507: 128--32.

Eddy SR (1998) Profile hidden Markov models. Bioinformatics (Oxford, England) 14: 755--63.

Field M, Graf LH, Laird WJ, Smith PL (1978) Heat-stable enterotoxin of Escherichia coli: in vitro effects on guanylate cyclase activity, cyclic GMP concentration, and ion transport in small intestine. Proceedings of the National Academy of Sciences of the United States of America 75: 2800--4.

Forte LR, Eber SL, Fan X, London RM, Wang Y, et al. (1999) Lymphoguanylin: cloning and characterization of a unique member of the guanylin peptide family. Endocrinology 140: 1800--6.

Frantz JC, Jaso-Friedman L, Robertson DC (1984) Binding of Escherichia coli heat-stable enterotoxin to rat intestinal cells and brush border membranes. Infection and immunity 43: 622--30.

Fülle HJ, Vassar R, Foster DC, Yang RB, Axel R, et al. (1995) A receptor guanylyl cyclase expressed specifically in olfactory sensory neurons. Proceedings of the National Academy of Sciences of the United States of America 92: 3571--5.

Gariépy J, Judd AK, Schoolnik GK (1987) Importance of disulfide bridges in the structure and activity of Escherichia coli enterotoxin ST1b. Proceedings of the National Academy of Sciences of the United States of America 84: 8907--11.

Gariépy J, Lane a, Frayman F, Wilbur D, Robien W, et al. (1986) Structure of the toxic domain of the Escherichia coli heat-stable enterotoxin ST I. Biochemistry 25: 7854--66.

Ghanekar Y, Chandrashaker A, Tatu U, Visweswariah SS (2004) Glycosylation of the receptor guanylate cyclase C: role in ligand binding and catalytic activity. The Biochemical journal 379: 653--63.

Guarino A, Cohen M, Thompson M, Dharmsathaphorn K, Giannella R (1987) T84 cell receptor binding and guanyl cyclase activation by Escherichia coli heat-stable toxin. The American journal of physiology 253: G775--80.

Guzman-Verduzco LM, Kupersztoch YM (1989) Rectification of two Escherichia coli heat-stable enterotoxin allele sequences and lack of biological effect of changing the carboxy-terminal tyrosine to histidine. Infection and immunity 57: 645--8.

Hamra FK, Eber SL, Chin DT, Currie MG, Forte LR (1997) Regulation of intestinal uroguanylin/guanylin receptor-mediated responses by mucosal acidity. Proceedings of the National Academy of Sciences of the United States of America 94: 2705--10.

Hamra FK, Forte LR, Eber SL, Pidhorodeckyj NV, Krause WJ, et al. (1993) Uroguanylin: structure and activity of a second endogenous peptide that stimulates intestinal guanylate cyclase. Proceedings of the National Academy of Sciences of the United States of America 90: 10464--8.

Hasegawa M, Hidaka Y, Matsumoto Y, Sanni T, Shimonishi Y (1999a) Determination of the binding site on the extracellular domain of guanylyl cyclase C to heat-stable enterotoxin. The Journal of biological chemistry 274: 31713--8.

Hasegawa M, Hidaka Y, Wada A, Hirayama T, Shimonishi Y (1999b) The relevance of N-linked glycosylation to the binding of a ligand to guanylate cyclase C. European journal of biochemistry / FEBS 263: 338--46.

Hasegawa M, Kawano Y, Matsumoto Y, Hidaka Y, Fujii J, et al. (1999c) Expression and characterization of the extracellular domain of guanylyl cyclase C from a baculovirus and Sf21 insect cells. Protein expression and purification 15: 271--81.

Hasegawa M, Shimonishi Y (2005) Recognition and signal transduction mechanism of Escherichia coli heat-stable enterotoxin and its receptor, guanylate cyclase C. The journal of peptide research : official journal of the American Peptide Society 65: 261--71.

He Xl, Dukkipati A, Garcia KC (2006) Structural determinants of natriuretic peptide receptor specificity and degeneracy. Journal of molecular biology 361: 698--714.

He Xl, Chow Dc, Martick MM, Garcia KC (2001) Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone. Science (New York, NY) 293: 1657--62.

Hidaka Y, Matsumoto Y, Shimonishi Y (2002) The micro domain responsible for ligand-binding of guanylyl cyclase C. FEBS letters 526: 58--62.

Hill O, Cetin Y, Cieslak A, Mägert HJ, Forssmann WG (1995) A new human guanylate cyclase-activating peptide (GCAP-II, uroguanylin): precursor cDNA and colonic expression. Biochimica et biophysica acta 1253: 146--9.

Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics (Oxford, England) 26: 680--2.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics (Oxford, England) 17: 754--5.

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology 292: 195--202.

Karakas E, Simorowski N, Furukawa H (2009) Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit. The EMBO journal 28: 3910--20.

Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30: 3059--3066.

Katoh K, Kuma Ki, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic acids research 33: 511--8.

Klipstein FA, Engert RF, Houghten RA (1983) Protection in rabbits immunized with a vaccine of Escherichia coli heat-stable toxin cross-linked to the heat-labile toxin B subunit. Infection and immunity 40: 888--93.

Krogh A, Brown M, Mian IS, Sjölander K, Haussler D (1994) Hidden Markov models in computational biology. Applications to protein modeling. Journal of molecular biology 235: 1501--31.

Kuhn M, Ng CKD, Su YH, Kilić A, Mitko D, et al. (2004) Identification of an orphan guanylate cyclase receptor selectively expressed in mouse testis. The Biochemical journal 379: 385--93.

Kumar J, Mayer ML (2010) Crystal structures of the glutamate receptor ion channel GluK3 and GluK5 amino-terminal domains. Journal of molecular biology 404: 680--96.

Kumar J, Schuck P, Jin R, Mayer ML (2009) The N-terminal domain of GluR6-subtype glutamate receptor ion channels. Nature structural & molecular biology 16: 631--8.

Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. Nucleic acids research 29: 221--2.

Lauber T, Neudecker P, Rösch P, Marx UC (2003) Solution structure of human proguanylin: the role of a hormone prosequence. The Journal of biological chemistry 278: 24118--24.

Lauber T, Tidten N, Matecko I, Zeeb M, Rösch P, et al. (2009) Design and characterization of a soluble fragment of the extracellular ligand-binding domain of the peptide hormone receptor guanylyl cyclase-C. Protein engineering, design & selection : PEDS 22: 1--7.

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics (Oxford, England) 22: 1658--9.

Li Z, Perkins AG, Peters MF, Campa MJ, Goy MF (1997) Purification, cDNA sequence, and tissue distribution of rat uroguanylin. Regulatory peptides 68: 45--56.

Lobley A, Sadowski MI, Jones DT (2009) pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. Bioinformatics (Oxford, England) 25: 1761--7.

Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. Database 2011: bar009--bar009.

Mann Ea, Swenson ES, Copeland NG, Gilbert DJ, Jenkins Na, et al. (1996) Localization of the guanylyl cyclase C gene to mouse chromosome 6 and human chromosome 12p12. Genomics 34: 265--7.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic acids research 39: D225--9.

Markert T, Vaandrager AB, Gambaryan S, Pöhler D, Häusler C, et al. (1995) Endogenous expression of type II cGMP-dependent protein kinase mRNA and protein in rat intestine. Implications for cystic fibrosis transmembrane conductance regulator. The Journal of clinical investigation 96: 822--30.

Marx UC, Klodt J, Meyer M, Gerlach H, Rösch P, et al. (1998) One peptide, two topologies: structure and interconversion dynamics of human uroguanylin isomers. The journal of peptide research : official journal of the American Peptide Society 52: 229--40.

Matecko I, Burmann BM, Schweimer K, Kalbacher H, Einsiedel J, et al. (2009) Structural Characterisation of the E. coli Heat Stable Enterotoxin STh. The Open Spectroscopy Journal 2: 34--39.

Miyagi M, Zhang X, Misono KS (2000) Glycosylation sites in the atrial natriuretic peptide receptor: oligosaccharide structures are not required for hormone binding. European journal of biochemistry / FEBS 267: 5758--68.

Miyazato M, Nakazato M, Yamaguchi H, Date Y, Kojima M, et al. (1996) Cloning and characterization of a cDNA encoding a precursor for human uroguanylin. Biochemical and biophysical research communications 219: 644--8.

Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins 12: 345--64.

Nandi A, Mathew R, Visweswariah SS (1996) Expression of the extracellular domain of the human heat-stable enterotoxin receptor in Escherichia coli and generation of neutralizing antibodies. Protein expression and purification 8: 151--9.

Navaneethan U, Giannella RA (2008) Mechanisms of infectious diarrhea. Nature clinical practice Gastroenterology & hepatology 5: 637--47.

Nokihara K, Wray V, Ando E, Naruse S, Hayakawa T (1997) Synthesis, solution structure, binding activity, and cGMP activation of human guanylin and its disulfide isomer. Regulatory peptides 70: 111--20.

Ogawa H, Qiu Y, Ogata CM, Misono KS (2004) Crystal structure of hormone-bound atrial natriuretic peptide receptor extracellular domain: rotation mechanism for transmembrane signal transduction. The Journal of biological chemistry 279: 28625--31.

Okamoto K, Takahara M (1990) Synthesis of Escherichia coli heat-stable enterotoxin STp as a pre-pro form and role of the pro sequence in secretion. Journal of bacteriology 172: 5260--5.

Okoh AI, Osode AN (2008) Enterotoxigenic Escherichia coli (ETEC): a recurring decimal in infants' and travelers' diarrhea. Reviews on environmental health 23: 135--48.

Organization WH (2006) Weekly epidemiological record Relevé épidémiologique hebdomadaire pp. 97--104.

Ozaki H, Sato T, Kubota H, Hata Y, Katsube Y, et al. (1991) Molecular structure of the toxin domain of heat-stable enterotoxin produced by a pathogenic strain of Escherichia coli. A putative binding site for a binding protein on rat intestinal epithelial cell membranes. The Journal of biological chemistry 266: 5934--41.

Parat M, Blanchet J, De Léan A (2010) Role of juxtamembrane and transmembrane domains in the mechanism of natriuretic peptide receptor A activation. Biochemistry 49: 4601--10.

Park J, Schulz S, Waldman SA (2000) Intestine-specific activity of the human guanylyl cyclase C promoter is regulated by Cdx2. Gastroenterology 119: 89--96.

Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, et al. (2011) ModBase, a database of annotated comparative protein structure models, and associated resources. Nucleic acids research 39: D465--74.

Pitari GM, Di Guglielmo MD, Park J, Schulz S, Waldman SA (2001) Guanylyl cyclase C agonists regulate progression through the cell cycle of human colon carcinoma cells. Proceedings of the National Academy of Sciences of the United States of America 98: 7846--51.

Pitari GM, Zingman LV, Hodgson DM, Alekseev AE, Kazerounian S, et al. (2003) Bacterial enterotoxins are associated with resistance to colon cancer. Proceedings of the National Academy of Sciences of the United States of America 100: 2695--9.

Ragvin A, Valvatne Hv, Erdal S, Arskog V, Tufteland KR, et al. (2004) Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. Journal of molecular biology 337: 773--88.

Rasheed JK, Guzmán-Verduzco LM, Kupersztoch YM (1990) Two precursors of the heat-stable enterotoxin of Escherichia coli: evidence of extracellular processing. Molecular microbiology 4: 265--73.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics (Oxford, England) 19: 1572--4.

Roy N, Guruprasad MR, Kondaiah P, Mann EA, Giannella RA, et al. (2001) Protein kinase C regulates transcription of the human guanylate cyclase C gene. European journal of biochemistry / FEBS 268: 2160--71.

Sack DA, Merson MH, Wells JG, Sack RB, Morris GK (1975) Diarrhoea associated with heat-stable enterotoxin-producing strains of Escherichia coli. Lancet 2: 239--41.

Sack RB (1975) Human diarrheal disease caused by enterotoxigenic Escherichia coli. Annual review of microbiology 29: 333--53.

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology 234: 779--815.

Sato T, Shimonishi Y (2004) Structural features of Escherichia coli heat-stable enterotoxin that activates membrane-associated guanylyl cyclase. The journal of peptide research : official journal of the American Peptide Society 63: 200--6.

Savarino SJ, Fasano A, Robertson DC, Levine MM (1991) Enteroaggregative Escherichia coli elaborate a heat-stable enterotoxin demonstrable in an in vitro rabbit intestinal model. The Journal of clinical investigation 87: 1450--5.

Savarino SJ, Fasano A, Watson J, Martin BM, Levine MM, et al. (1993) Enteroaggregative Escherichia coli heat-stable enterotoxin 1 represents another subfamily of E. coli heat-stable toxin. Proceedings of the National Academy of Sciences of the United States of America 90: 3093--7.

Schulz S, Chrisman TD, Garbers DL (1992) Cloning and expression of guanylin. Its existence in various mammalian tissues. The Journal of biological chemistry 267: 16019--21.

Schulz S, Green CK, Yuen PS, Garbers DL (1990) Guanylyl cyclase is a heat-stable enterotoxin receptor. Cell 63: 941--8.

Shapiro AL, Viñuela E, Maizel JV (1967) Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochemical and biophysical research communications 28: 815--20.

Shimonishi Y, Hidaka Y, Koizumi M, Hane M, Aimoto S, et al. (1987) Mode of disulfide bond formation of a heat-stable enterotoxin (STh) produced by a human strain of enterotoxigenic Escherichia coli. FEBS letters 215: 165--70.

Sindićе A, Başoglu C, Cerçi A, Hirsch JR, Potthast R, et al. (2002) Guanylin, uroguanylin, and heat-stable euterotoxin activate guanylate cyclase C and/or a pertussis toxin-sensitive G protein in human proximal tubule cells. The Journal of biological chemistry 277: 17758--64.

Skelton NJ, Garcia KC, Goeddel DV, Quan C, Burnier JP (1994) Determination of the solution structure of the peptide hormone guanylin: observation of a novel form of topological stereoisomerism. Biochemistry 33: 13581--92.

Spangler BD (1992) Structure and function of cholera toxin and the related Escherichia coli heat-labile enterotoxin. Microbiological reviews 56: 622--47.

Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics (Oxford, England) 23: 1282--8.

Swenson ES, Mann EA, Jump ML, Giannella RA (1999) Hepatocyte nuclear factor-4 regulates intestinal expression of the guanylin/heat-stable toxin receptor. Am J Physiol Gastrointest Liver Physiol 276: G728--736.

Swenson ES, Mann EA, Jump ML, Witte DP, Giannella RA (1996) The Guanylin / STa Receptor Is Expressed in Crypts and Apical Epithelium throughout the Mouse Intestine 1 and brain natriuretic peptide receptors , GC-A and GC-B (1). GC-C catalyzes the formation of a profuse , potentially life threatening diarrhea (2 . Biochemical and Biophysical Research Communications 1014: 1009--1014.

Takao T, Hitouji T, Aimoto S, Shimonishi Y, Hara S, et al. (1983) Amino acid sequence of a heat-stable enterotoxin isolated from enterotoxigenic Escherichia coli strain 18D. FEBS letters 152: 1--5.

Takao T, Shimonishi Y, Kobayashi M, Nishimura O, Arita M, et al. (1985) Amino acid sequence of heat-stable enterotoxin produced by Vibrio cholerae non-01. FEBS letters 193: 250--4.

Takeda T, Peina Y, Ogawa A, Dohi S, Abe H, et al. (1991) Detection of heat-stable enterotoxin in a cholera toxin gene-positive strain of Vibrio cholerae O1. FEMS microbiology letters 64: 23--7.

Taxt A, Aasland R, Sommerfelt H, Nataro J, Puntervoll Pl (2010) Heat-stable enterotoxin of enterotoxigenic Escherichia coli as a vaccine target. Infection and immunity 78: 1824--31.

Towbin H, Staehelin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proceedings of the National Academy of Sciences of the United States of America 76: 4350--4.

Vaandrager aB, Schulz S, De Jonge HR, Garbers DL (1993) Guanylyl cyclase C is an N-linked glycoprotein receptor that accounts for multiple heat-stable enterotoxin-binding proteins in the intestine. The Journal of biological chemistry 268: 2174--9.

Vaandrager AB, Tilly BC, Smolenski A, Schneider-Rasp S, Bot AG, et al. (1997) cGMP stimulation of cystic fibrosis transmembrane conductance regulator Cl-channels co-expressed with cGMP-dependent protein kinase type II but not type Ibeta. The Journal of biological chemistry 272: 4195--200.

Vaandrager aB, van der Wiel E, Hom ML, Luthjens LH, de Jonge HR (1994) Heat-stable enterotoxin receptor/guanylyl cyclase C is an oligomer consisting of functionally distinct subunits, which are non-covalently linked in the intestine. The Journal of biological chemistry 269: 16409--15.

van den Akker F, Zhang X, Miyagi M, Huo X, Misono KS, et al. (2000) Structure of the dimerized hormone-binding domain of a guanylyl-cyclase-coupled receptor. Nature 406: 101--4.

Vijayachandra K, Guruprasad M, Bhandari R, Manjunath UH, Somesh BP, et al. (2000) Biochemical characterization of the intracellular domain of the human guanylyl cyclase C receptor provides evidence for a catalytically active homotrimer. Biochemistry 39: 16075--83.

Wada A, Hiayama T, Kitao S, Fujisawa J, Hidaka Y, et al. (1994) Pig Intestinal Membrane-Bound Receptor (Guanylyl Cyclase) for Heat-Stable Enterotoxin : cDNA Cloning , Functional Expression , and Characterization. Microbiol Immmunol 38: 535--541.

Wada A, Hirayama T, Kitaura H, Fujisawa J, Hasegawa M, et al. (1996) Identification of ligand recognition sites in heat-stable enterotoxin receptor, membrane-associated guanylyl cyclase C by site-directed mutational analysis. Infect Immun 64: 5144--5150.

Walker RI, Steele D, Aguado T (2007) Analysis of strategies to successfully vaccinate infants in developing countries against enterotoxigenic E. coli (ETEC) disease. Vaccine 25: 2545--66.

Wiegand RC, Kato J, Huang MD, Fok KF, Kachur JF, et al. (1992a) Human guanylin: cDNA isolation, structure, and activity. FEBS letters 311: 150--4.

Wiegand RC, Katoz J, Curriez MG (1992b) Rat guanylin cDNA: Characterization of the precursor of an endogenous activator of intestinal guanylate cyclase.

Yamanaka H, Kameyama M, Baba T, Fujii Y, Okamoto K (1994) Maturation pathway of Escherichia coli heat-stable enterotoxin I: requirement of DsbA for disulfide bond formation. Journal of bacteriology 176: 2906--13.

Yamanaka H, Nomura T, Fujii Y, Okamoto K (1997) Extracellular secretion of Escherichia coli heat-stable enterotoxin I across the outer membrane. Journal of bacteriology 179: 3383--90.

Yang RB, Foster DC, Garbers DL, Fülle HJ (1995) Two membrane forms of guanylyl cyclase found in the eye. Proceedings of the National Academy of Sciences of the United States of America 92: 602--6.

Yang Y, Gao Z, Guzmán-Verduzco LM, Tachias K, Kupersztoch YM (1992) Secretion of the STA3 heat-stable enterotoxin of Escherichia coli: extracellular delivery of Pro-STA is accomplished by either Pro or STA. Molecular microbiology 6: 3521--9.

Yoshimura Shoko, Ikemura Haruo, Watanabe Hiroyuki, Aimoto Saburo, Hara Saburo, et al. (1985) Essential structure for full enterotoxigenic activity of heat-stable enterotoxin produced by enterotoxigenic Escherichia coli. FEBS Letters 181: 138--142.

Yoshino K, Takao T, Huang X, Murata H, Nakao H, et al. (1995) Characterization of a highly toxic, large molecular size heat-stable enterotoxin produced by a clinical isolate of Yersinia enterocolitica. FEBS letters 362: 319--22.

Zvelebil M (1987) Prediction of protein secondary structure and active sites using the alignment of homologous sequences. Journal of Molecular Biology 195: 957--961.