TOWARDS NMR SPECTROSCOPIC STUDIES OF THE NTS1-NT COMPLEX

Satita Tapaneeyakorn

Thesis submitted to the Board of the Medical Sciences Division in partial fulfilment of the requirements for the Degree of Doctor of Philosophy at the University of Oxford

St. Hugh's College, Oxford

Trinity Term 2009

Abstract

The tridecapeptide, neurotensin (NT), acts as a neurotransmitter in the central nervous system and peripherally in the gastrointestinal tract. Its receptor, NTS1, belongs to the G protein-coupled receptor family and is a potential target for the treatment of pain, eating disorders, stress, schizophrenia, Parkinson's disease, Alzheimer's disease, and cancer. The six C-terminal amino acids (8-13) of NT are sufficient for binding to this receptor and eliciting the major pharmacological effects of this peptide. The aims of this study are to examine the conformation of the neurotensin free and bound to its receptor using non-perturbing NMR-visible probes (^{13}C, ^{15}N) detected with either solution- or solid-state NMR, and to characterise further this interaction together with other biophysical tools.

^{15}N-Pro10, $^{13}C\alpha$-Tyr11, ^{13}CO-Ile12, and $^{13}C\gamma$-Leu13 enriched NT_{8-13} and $^{13}C\alpha$-Arg9 and ^{13}CO-, $^{13}C\beta$-Pro10 labelled NT_{8-13} were specifically designed and produced by manual Fmoc-solid phase peptide synthesis to permit a determination of the structure of the C-terminal bound ligand and the bound conformation of Pro10 (*cis*/*trans*) of ligand to be made at very high resolution *via* solid-state NMR distance measurements, respectively. ^{13}C solution NMR has been used to determine the assignment for these labelled peptides.

The expression system of NT as a fusion protein in *E. coli* and released by cyanogen bromide cleavage was developed, to allow uniformly (^{13}C, ^{15}N) labelled NT and NT_{8-13} to be produced using two different plasmids. The deuterated peptides were also expressed and will be used for probing the NTS1-NT interactions by solution NMR spectroscopy. The resonance assignments of NT and NT_{8-13} spectra obtained from solution NMR experiments were performed.

Recombinant NTS1 was expressed in *E. coli* and purified with sufficient amount for high-resolution structural studies. Using the uniformly ^{15}N enriched NT in the presence of TEV-cleaved NTS1 (NTS1 alone), specific interactions could not be observed in ^{1}H-^{15}N HSQC spectra as the protein was likely aggregate during experimental time at 25˚C. To facilitate solid-state NMR studies, detergent-purified NTS1 was reconstituted into brain polar lipid liposomes without any specific interactions. The reconstituted NTS1 was incubated with uniformly ^{13}C, ^{15}N enriched NT and further used for solid-state NMR studies. 1D CP MAS spectra of ligand in the BPL and bound ligand were superimposed and chemical shift perturbation at 144 ppm was observed which could arise from side-chain of Arg or Tyr of bound ligand or protein.

Computer modelling was aimed to predict the structure of NTS1-NT_{8-13} complex and rationalise all the experimental data at the nanoscale. The NT_{8-13} ligand probably adopted a U-shape conformation at the putative binding site between EC2 and EC3 loops.

List of publications

Song, C., **Tapaneeyakorn, S.**, Butts, C., Watts, A. and Willis, C. L. (2009) Enantioselective syntheses of α-Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1, 3-$^{13}C_2$]-L-proline and incorporation into the neurotensin receptor 1 ligand, NT_{8-13}. *J Org Chem*, **74**(23), 8980-7.

Tapaneeyakorn, S., Ross, S., Attrill, H. and Watts, A. (2009) Heterologous high yield expression and purification of neurotensin and its functional fragment in *Escherichia coli*. (To be submitted)

List of papers presented

Oral presentation

1. Tapaneeyakorn, S. (2009) High-yield expression and purification of the neurotensin and its functional fragment in *Escherichia coli* for NMR studies. In *the 2nd Samaggi Academic Conference,* February 7-8, 2009, Cambridge, United Kingdom.

2. Tapaneeyakorn, S. (2008) Probing the interaction of neurotensin with its receptor, NTS1, at the nanoscale using solid-state NMR. In *the 4th Workshop of the UK-Japan Bionanotechnology Collaboration,* September 17-18, 2008, Kobe, Japan.

Poster presentation

1. Tapaneeyakorn, S. *et al.* (2008) Structural and orientational information of the neurotensin bound to its receptor, NTS1, by solid-state NMR. In *the 50th Rocky Mountain Conference on Analytical Chemistry,* July 27-31, 2008, Colorado, United States.

2. Tapaneeyakorn, S. *et al.* (2008) Biophysical studies of GPCRs. In *the Department of Biochemistry's Annual Recess,* April 14-15, 2008, Oxford, United Kingdom.

3. Tapaneeyakorn, S. *et al.* (2008) Probing the structure of the receptor bound neurotensin using solid-state NMR. In *the NMRDG meeting – NMR in Structural Biology,* April 3-4, 2008, Cambridge, United Kingdom.

4. Tapaneeyakorn, S. *et al.* (2007) A Structural Study of receptor bound neurotensin by solid-state NMR. In *the 6th Biophysics Congress,* July 14-18, 2007, London, United Kingdom.

5. Tapaneeyakorn, S. *et al.* (2007) Biomolecular solid-state NMR: its application to membrane proteins and drug design. In *the Industry Day – Bionanotechnology IRC,* May 30, 2007, Oxford, United Kingdom.

6. Tapaneeyakorn, S. *et al.* (2006) Probing the interaction of the neurotensin with its receptor, NTS1, by solid-state NMR. In *the NANO2006 Workshop "Perspectives in Nanoscience and Nanotechnology",* September 4-6, 2006, San Sebastian, Spain.

7. Tapaneeyakorn, S. *et al.* (2006) A solid-state NMR study of the neurotensin interaction with its receptor, NTS1. In *the International school on high field NMR spectroscopy for solids and liquids,* May 14-19, 2006, Les Houches, France.

Acknowledgements

The four years of my graduate student life at Oxford University were influenced and touched by many people who supported and inspired me professionally and emotionally in great and important ways.

I would like to express my gratitude to Professor Anthony Watts for his support during my D.Phil. course, for providing resources which have made this project possible and for allowing me to attend scientific conferences to get the pulse of the community and find out what is new and exciting in different subject areas.

Dr Helen Attrill is acknowledged for her valuable discussions. She was working on the optimisation of NTS1B purification providing the possibility to obtain large amounts of active receptor for NMR studies.

I would like to thank Dr Xin Zhao and Dr Chang Gyeom Kim for introducing me to solid-state NMR in my first year and many thanks to Dr Krisztina Varga for her help in setting up all solid-state NMR experiments in my third year. I would also like to thank Dr Phil Williamson for helping setting up some solid-state NMR experiments and for his useful comments. Thank you to Dr Peter Judge for his efforts to set up some solid-state NMR experiments in the last few months.

I am grateful to our collaborators Professor Christine L. Willis and Dr Chuanjun Song for synthesising the non-commercially available labelled amino acids.

I would like to thank Professor Ian Campbell for allowing me to use his solution NMR facilities and many thanks to Dr Tim Claridge, Dr Ioannis Vakonakis, Dr Kate Wegener, and Dr Pengju Jiang for their help in collecting solution NMR data.

I am indebted to Professor Mark Sansom for providing me with the infrastructure to carry out the molecular modelling and Dr Phillip Stansfeld and Assistant Professor Issara Sramala for teaching me homology modelling and MD simulations.

Many thanks to Dr Simon Ross for preliminary data of the expression of neurotensin peptide in the *E. coli* system.

Thank you to Dr Timothy Hadingham for teaching me manual Fmoc-solid phase peptide synthesis and Dr Tracy Lam for teaching me how to use the automatic peptide synthesis machine.

I thank Robin Antrobus for carrying out ESI mass spectrometry of my synthesised samples when the open access machine was out of order.

I thank Dr Peter James Harding for optimisation of reconstitution of NTS1 into lipid bilayers during his D.Phil. course providing me with a good start for SSNMR sample preparation.

Thank you to Liz and Pete who sorted out several things for me during my D.Phil. period.

I am sincerely grateful to the past and present members of the Watts Lab who I have failed to mention personally for their input and I wish you all the best for your future.

I would like to acknowledge the financial support of the Royal Thai Government Scholarship for the D.Phil. funding.

I would like to thank all my friends in Thailand, Germany and England, especially Chai, Chris, Grace, Krisztina, Morakot, Nichanun, Olivia, Parun, Ruxapon, Songpol, and Yaowaluck for always being there for me, listening to me, offering me

support, making me laugh, and cheering me up. Particular thanks to Yaowaluck for making some fantastic Thai food and Grace and Olivia for making delicious cakes. Thank you to Lubica who help me to improve my spoken English and Marcella for her support during my sickness in the Rocky mountain conference.

Last but certainly not least, I would like to thank my family and P' Oat for their deep love, emotional support and always believing in me – I love all of you!

Abbreviations

1D, 2D, 3D	one-dimensional, two-dimensional, three-dimensional
1pda	direct excitation
Å	Ångstrom
AII	angiotensin II
$A_{2A}R$	human A_{2A} adenosine receptor
β_1AR	turkey β_1 adrenergic receptor
β_2AR	human β_2 adrenergic receptor
B_0	static external magnetic field
bRho	bovine rhodopsin
BPL	brain polar lipid extract (porcine)
BSA	bovine serum albumin
cAMP	3'-5'-cyclic adenosine monophosphate
cGMP	guanosine 3',5'-cyclic phosphate
cisNT6	NT_{8-13} homology model containing *cis* conformation at Pro10
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHS	cholesteryl hemisuccinate
CNBr	cyanogen bromide
COSY	correlation spectroscopy
CP	cross-polarisation
CP MAS	cross-polarisation magic angle spinning
CSA	chemical shift anisotropy
CV	column volume
CW	continuous-wave
δ	chemical shift
DARR	dipolar-assisted rotational resonance
DCM	dichloromethane
DDM	dodecyl-β-D-maltoside
DIPEA	*N*,*N*'-diisopropylcarbodiimide
DMF	dimethylformamide

DSS	2,2-dimethyl-2-silapentane-5-sulfonate
DTT	dithiolthreitol
E. coli	*Escherichia coli*
EC1	first extracellular loop
EC2	second extracellular loop
EC3	third extracellular loop
EDT	1,2-ethanedithiol
EDTA	ethylenediaminetetracetic acid
ESI	electrospray ionisation
eCFP	enhanced cyan fluorescent protein
eYFP	enhanced yellow fluorescent protein
Fmoc	9-fluorenylmethoxycarbonyl
γ	gyromagnetic ratio
ΔG_b	binding free energy
GA	genetic algorithm
GB1	immunoglobulin-binding domain of streptococcal protein G
GDP	guanosine diphosphate
GPCR	GTP-binding protein-coupled receptor
GTP	guanosine triphosphate
GST	glutathione S-transferase
hNTS1	human neurotensin receptor 1
HBTU	2-(H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HF	hydrofluoric acid
HOBt	N-hydroxybenzotriazole
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IMAC	immobilised metal affinity chromatography
IP3	inositol-1,4,5-triphosphate
IC1	first intracellular loop
IC2	second intracellular loop
IC3	third intracellular loop

IPTG	isopropyl β-D-1-thiogalactopyranoside
K_i	inhibition constant
LB	Luria-Bertani
LGA	Lamarckian genetic algorithm
MAS	magic-angle spinning
MBP	*E. coli* maltose binding protein
MD	molecular dynamic
MES	2-(*N*-morpholino)ethanesulfonic acid
MWCO	molecular weight cut-off
ω_{Lar}	Larmor frequency
ω_R	rotor spinning frequency
NMR	nuclear magnetic resonance
NT	neurotensin
NT_{8-13}	the six C-terminal residues of NT
NTS	neurotensin receptor
NTS1A	MBP-T43NTS1-TrxA-H10
NTS1B	MBP-TEV-T43NTS1-TEV-TrxA-H10
NTS1C	MBP-TEV-T43NTS1-CFP-TEV-TrxA-H10
NTS1Y	MBP-TEV-T43NTS1-YFP-TEV-TrxA-H10
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PME	Particle Mesh Ewald
PMSF	phenylmethylsulphonyl fluoride
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanol-amine
POPG	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol
ppm	parts per million
REDOR	rotational echo double resonance
rf	radio frequency

RFDR	radiofrequency driven dipolar recoupling
ROESY	rotating-frame overhauser effect spectroscopy
RMSD	root mean square deviation
RMSF	root mean square fluctuation
rNTS1	rat neurotensin receptor 1
rNTS1-B1	rNTS1 homology model containing β structure at EC2 loop
rNTS1-H1	rNTS1 homology model containing helix structure at EC2 loop
rpm	rotations per minute
RR	rotational resonance
sRho	squid rhodopsin
SA	simulated annealing
SDM	site-directed mutagenesis
SDS	sodium dodecyl sulphate
SolNMR	solution-state nuclear magnetic resonance
SPPS	solid phase peptide synthesis
SSNMR	solid-state nuclear magnetic resonance
τ_c^{-1}	tumbling rate
T_2	transversal relaxation time
tBu	*tert*-butyl group
T43NTS1	N-terminally truncated NTS1
t-Boc	*tert*-butoxycarbonyl
transNT6	NT_{8-13} homology model containing *trans* conformation at Pro10
trNOE	transferred nuclear overhauser effects
TEV	tobacco etch virus
TFA	trifluoroacetic acid
TLC	thin layer chromatography
TM	transmembrane
TOCSY	total correlation spectroscopy
TPPM	two-pulse phase modulation
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
TrxA	*E. coli* thioredoxin
xg	x gravity

Contents

Page

Chapter 1

Introduction

1.1 Biological membranes and membrane proteins

The basic functions of the cell membrane are to provide different spatial compartments and to act as highly selective barriers for the exchange of molecules between the different compartments. The cell membrane is composed of a lipid bilayer and various membrane proteins (**Figure 1.1**). Lipid types and protein-lipid ratios depend on the membrane's location and biological function. Individual protein and lipid molecules within membranes diffuse more or less freely from the basal to the lateral surface of the cell or *vice versa* in accordance with the fluid mosaic model (Singer and Nicolson, 1972).

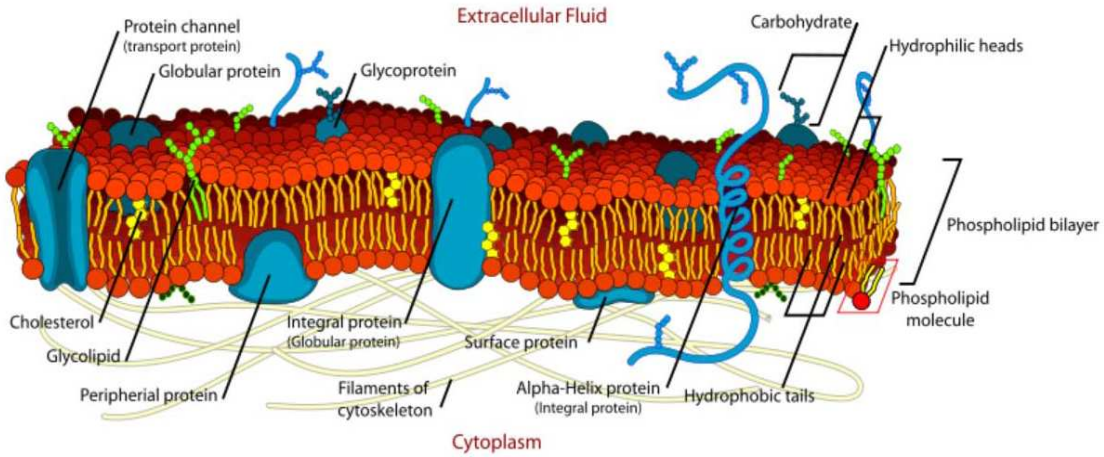


Figure 1.1: Schematic structural components of a typical cell membrane (http://commons.Wikimedia.org/wiki/File:Cell_membrane_detailed_diagram_pl.svgv). This cell membrane figure is based on the fluid mosaic model (Singer and Nicolson, 1972). The cell membrane consists of lipid bilayer and proteins, which are either embedded into the bilayer or peripherally attached.

Phospholipids, the most abundant lipids in eukaryotic cells, are amphiphilic molecules with a hydrophilic head group and usually two lipophilic hydrocarbon chains. In cell membranes, the polar head groups are directed outward and exposed to water, while the hydrophobic tails, consisting of fatty acyl chains are pointed toward the centre of the cell membrane. The head groups of membrane phospholipids can form non-covalent bonds with peripheral membrane-binding proteins and covalent bonds with particular proteins. The cell membrane is freely permeable to water molecules but the hydrophobic part makes the membrane relatively impermeable to ions and polar molecules (Nelson and Cox, 2004). Lipids generally give the cell membrane their flexibility, but cholesterol, a type of lipid, helps stiffen the membrane of vertebrate cells. The hydroxyl group of the cholesterol molecule is positioned at the membrane surface. Cholesterol is a rigid molecule and restricts the motion of the fatty acyl chains of the phospholipid molecules, condensing the lipid molecules and hence causing a decrease of membrane fluidity and permeability to small water-soluble molecules (Ikonen, 2008).

There are two types of proteins in the cell membrane, peripheral and integral membrane proteins (Nelson and Cox, 2004). The integral membrane proteins, firmly embedded in the hydrophobic layer of the membrane, are removable only by agents that disrupt the bilayer structure, such as detergents, organic solvents, or denaturants. Some of these proteins, crossing the bilayer and existing on both sides, form channels or pumps, which control molecular traffic across the membrane. These allow concentration gradients to be maintained between the inside and outside of the cell. The integral membrane proteins are also classified as transporters, receptors, enzymes, structural membrane-anchoring domains, proteins involved in accumulation and

transduction of energy, and proteins responsible for cell adhesion. The peripheral membrane proteins, on the surface of the membrane, are anchored with the membrane, hydrophilic domains of integral proteins and the polar head groups of membrane lipids through various intermolecular interactions, such as electrostatic interactions and hydrogen bonds. They can be released by relatively mild treatments, such as mild non-ionic detergents, high pH and high salt concentrations. The peripheral membrane proteins may act as regulators of membrane-bound enzymes, ion channels and transmembrane receptors, or may limit the mobility of integral membrane proteins by tethering them to intracellular structures. It has been shown that some peripheral membrane proteins regulate cell signalling, electron transport and membrane trafficking through reversible interaction of proteins to biological membranes (Cho and Stahelin, 2005).

Membranes also contain carbohydrates, which are present as part of glycoproteins and glycolipids. The oligosaccharide components of both glycoproteins and glycolipids are generally exposed on the cell's outer surface. The oligosaccharide may be covalently attached to its protein through a glycosidic link to the hydroxyl group of a Ser or Thr residue (*O*-linked), or through an *N*-glycosyl link to the amide nitrogen of an Asn residue (*N*-linked). As the oligosaccharide portions of glycoproteins are often branched, they are rich in information, forming highly specific sites for recognition and high-affinity binding by other proteins. A number of proteins, for example the human protein interferon IFN-β1, produced in two types of tissues have different glycosylation patterns so the oligosaccharide chains may represent a tissue-specific marker (Nelson and Cox, 2004). Glycolipids also serve as markers for cellular recognition by carbohydrate-binding proteins. These components of the

membrane may be protective, insulators, or sites of receptor binding (Nelson and Cox, 2004). A change in glycosylation patterns has been proposed to be a potential biomarker for cancer diagnostic (Dube and Bertozzi, 2005).

Approximately 30-40% of all genes in the human genome are believed to encode membrane proteins (Wallin and von Heijne, 1998). Since membrane proteins are key regulators of cellular function, they account for up to two thirds of known drugable targets, highlighting their critical pharmaceutical importance (Lundstrom, 2007). Structural determination of membrane proteins will aid in structure-based drug design and discovery. However, there are more than 50,000 structures of proteins available in the Protein Data Bank (PDB) (Berman et al., 2003) but three-dimensional structures of only ~191 different membrane proteins (~20 have unique folds) are currently determined at high resolution due to the difficulties with overexpression, extraction, purification, and crystallisation (http://blanco.biomol.uci.edu/Membrane_ Proteins_xtal.html; July, 2009). Moreover, the membrane protein solubility, size and its stability, especially when not membrane-embedded, are major bottlenecks for structural determination by solution-state NMR techniques**.**

1.2 G Protein-coupled receptors

G protein-coupled receptors (GPCRs), the largest group of membrane receptors (~800 in human), are encoded for by 1% of the genome of higher life forms. GPCRs are seven transmembrane helix proteins, connected by three intracellular and three extracellular loops. GPCRs are activated by a wide range of stimuli, including hormones, neurotransmitters, ions, odorants, and photons of light (Lundstrom, 2007). The GPCR superfamily is involved in a variety of biological and pathological

processes such as development and proliferation (Lowes et al., 2002), neurological disorders (Miller et al., 2004), angiogenesis (Parker et al., 2005), metabolic disorders (Rayasam et al., 2007), and immune system and inflammation (Blakeney and Fairlie, 2005). Consequently, they are one of the largest classes of drug targets to which agonists and antagonists are currently focused. Although approximately 50% of pharmaceutical products target GPCRs, only 10% of the GPCRs excluding olfactory receptors are targeted by marketed drugs, indicating the possibility of the remaining 90% of the GPCRs for the treatment of human diseases (Kroeze et al., 2003).

1.2.1 Classification and mechanism of GPCRs

Based on protein sequence similarity, GPCRs are divided in three distinct families, A, B and C (**Figure 1.2**) (Humphrey and Barnard, 1998; Pierce et al., 2002), but some types of classifications divide them into more than three groups, depending on the criteria used. For instance, Fredriksson and colleagues have proposed two more families of GPCRs, frizzled/smoothened-like and adhesion-like families on the basis of phylogenetic criteria (Fredriksson et al., 2003). All receptors in each family share ≥ 20% sequence identity and > 40% similarity within the seven-transmembrane domains. Family A, the largest group of GPCRs, includes GPCRs homologous in sequence to the rhodopsin. Members of family A are activated by widespread inducers, including hormones, neurotransmitters and light. Family B denotes the secretin-like receptors which are mostly regulated by peptide hormones from the glucagon hormone family. Family C is the metabotropic glutamate receptors which bind to neurotransmitter glutamate *via* a Venus-flytrap mechanism (Humphrey and Barnard, 1998; Pierce et al., 2002).

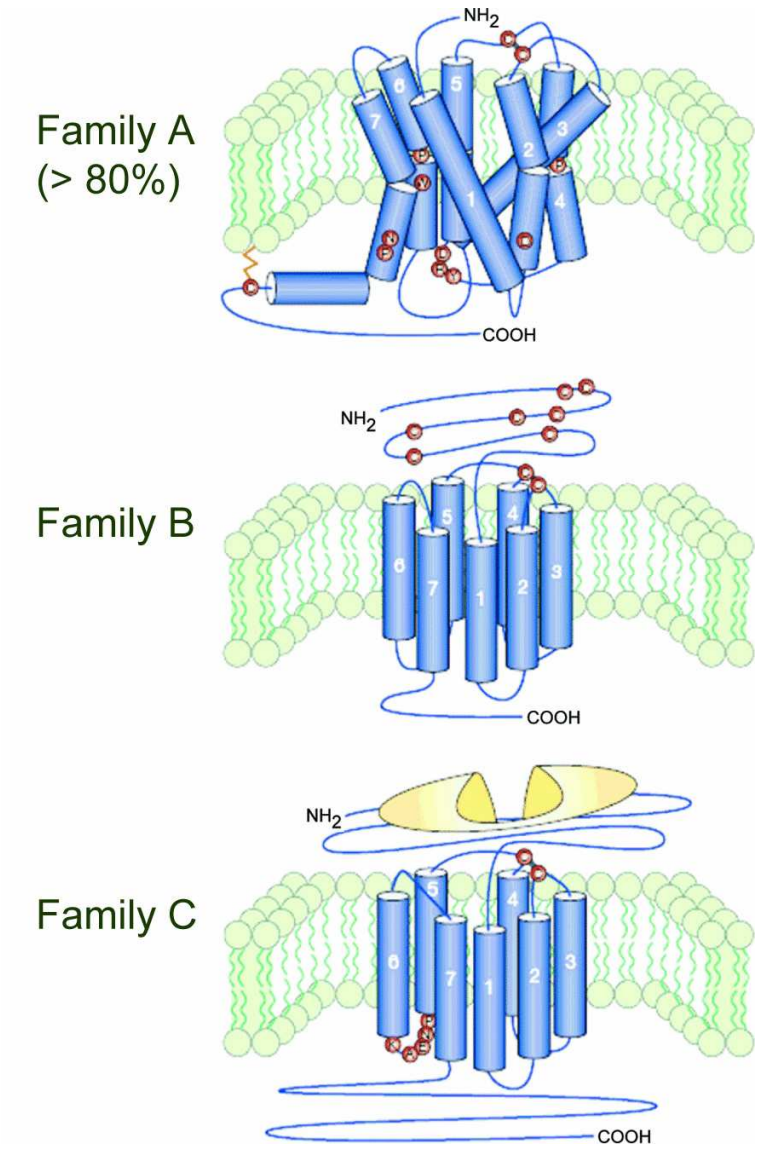


Figure 1.2: Representative structures of the three main GPCR families which have been classified based on their sequences. Some conserved amino acids are indicated in the structures by red circles. Adapted from (Strosberg and Nahmias, 2007).

Family A comprises approximately 180 liganded GPCRs, 110 orphan GPCRs and approximately 350 olfactory GPCRs, thus including more than 80% of total GPCRs, and most of clinically used pharmaceuticals which target GPCRs are directed at this family (Lundstrom, 2007; Tyndall and Sandilya, 2005). Several drugs including antihistamines, cardiovascular drugs, maraviroc, and endothelin are widely used for

the treatment of diseases caused by malfunction of receptors in this family. However, as most of existing GPCR drugs on the market are nonpeptide mimic and non-specific to the target GPCRs, they have been found to have side effects and toxicities. For instance, maraviroc (chemokine receptor-5 antagonist), has been approved for the treatment of drug-resistant patients with R5 HIV-1 since 2007 but it is the potential for some adverse events including cardiovascular side effects and liver toxicity, that limits its use and has initiated debates about its long-term consequences (Fatkenheuer et al., 2005; Telenti, 2009). Cannabinoid-1 receptor blocker rimonabant was approved by the European Union but not the USA in 2006, to be an effective drug for the treatment of obese patients, however this anti-obesity drug causes severe adverse effects, especially of the nervous system, of psychiatric or of gastro-intestinal origin (Curioni and Andre, 2006; Van Gaal et al., 2005) and so it was withdrawn by the manufacturers in 2008.

The general action of GPCRs is shown in **Figure 1.3.** The GPCRs are coupled with heterotrimeric G proteins, composed of a nucleotide-binding α-subunit plus βγ-subunits, although in some cases, the GPCRs may signal through non-G-protein mediated events (Bockaert et al., 2003). Upon agonist binding, an induced receptor changes its conformation which facilitates the exchange of G protein-bound GDP (inactive) for GTP (active) causing dissociation of the heterotrimer. Subsequent effector activation is dependent on the dissociated component parts, α-subunit and βγ subunits. Both G-protein components can themselves play active roles in signal transduction inside cells through the activation or inhibition of effector molecules, such as adenylyl cyclase, phospholipase Cβ, K^+ and Ca^{2+} channels, and phosphoinositide 3-kinase-γ. For instance, the activated Gα subunit can interact with adenylyl cyclase, leading to increased 3'-5'-cyclic adenosine monophosphate (cAMP)

levels, which in turn can induce protein kinase A activation. The protein kinase A is a serine-threonine kinase that phosphorylates specific committed enzymes in the metabolic pathway and also regulates specific gene expression, cellular secretion and membrane permeability (Pierce et al., 2002; Sprang, 1997).

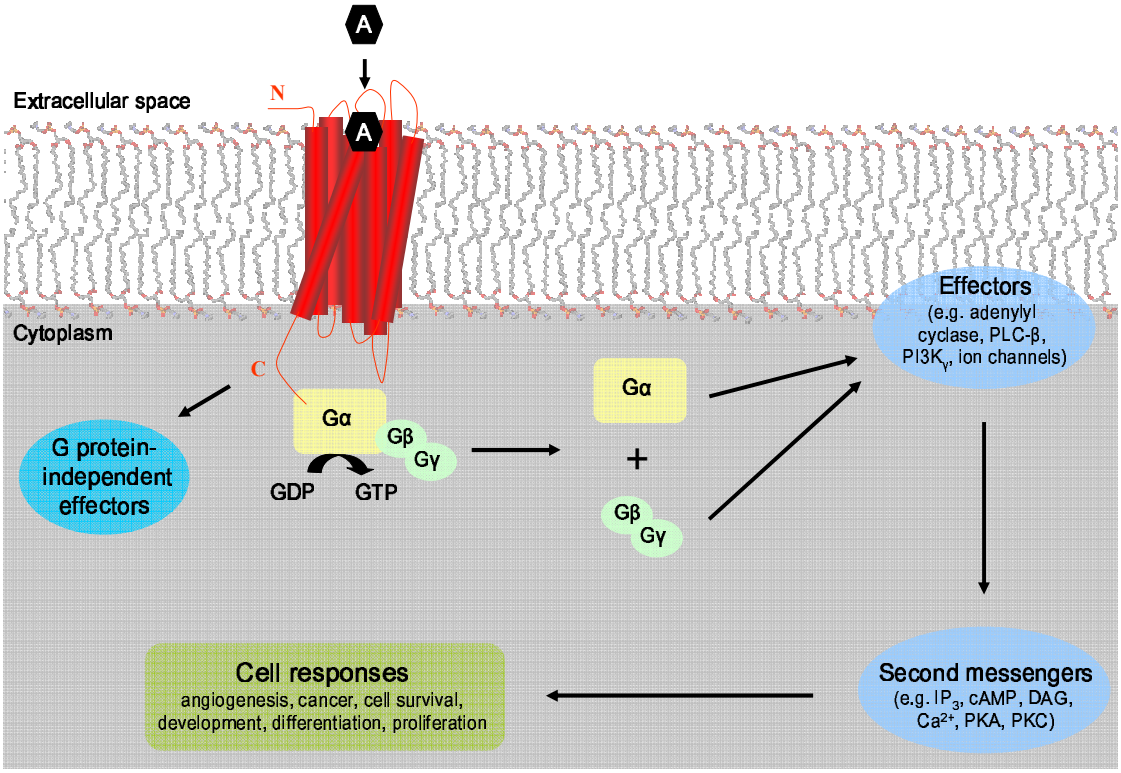


Figure 1.3: Schematic representation of signal transduction cascade within the cell of GPCRs. Agonists (A) can be widespread inducers to activate the GPCRs, depending on the receptor type. They can be peptides (*e.g.* neurotensin, angiotensin, thrombin), amino acids (*e.g.* glutamate), lipids (*e.g.* leukotriene, LPA), biogenic amines (*e.g.* acetylcholine, dopamine), or light. Upon agonist binding, the GPCR changes its conformation which promotes exchange of G-protein-bound GDP for GTP. The activated heterotrimer dissociates into the Gα subunit and the Gβγ subunits, both of which can play independently active roles in signal transduction through the activation or inhibition of effector molecules. This results in second messenger responses, which in turn triggers a cascade of cellular signal transduction events. Hydrolysis of GTP to GDP leads to signal termination and reassociation of the heterotrimer. The G protein-independent effector pathway is also indicated. Abbreviations: cAMP, 3'-5'-cyclic adenosine monophosphate; DAG, diacylglycerol; IP3, inositol-1,4,5-triphosphate; $PI3K_{\gamma}$, phosphoinositide 3-kinase-γ; PKA, protein kinase A; PKC, protein kinase C; PLC, phospholipase C.

There are more than 20 identified α-subunits and 16 of them have been cloned to date. Five isoforms of β subunits and 12 of γ subunits have also been cloned (Pierce et al., 2002). G proteins are generally referred to by their α-subunits, which are divided into four distinct classes: G_s proteins couple to stimulation of adenylyl cyclase; G_i proteins couple to inhibition of adenylyl cyclase and activation of G protein-coupled inwardly rectifying potassium channels; G_q proteins couple to the activation of phospholipase Cβ resulting in the intramembrane hydrolysis of phosphatidylinositol-4,5-bisphosphate to inositol-1,4,5-triphosphate and diacylglycerol; and G_{12} proteins couple to the activation of Rho guanine nucleotide exchange factors. However, it has been shown that most biological responses mediated by GPCRs do not involve a sole biochemical route, but instead result from the integration of intracellular signalling pathways involving several different G proteins, as well as G protein independent pathways (Marinissen and Gutkind, 2001).

G proteins deactivate themselves as a result of their intrinsic GTPase activity, but in some cases G protein-coupled effectors or other molecules can act as GTPase-activating proteins to regulate heterotrimeric G proteins by increasing the rates at which their α-subunits hydrolyse bound GTP and thus return to the inactive state (Ross and Wilkie, 2000).

1.2.2 GPCR structures to date

So far, structural studies of GPCR-ligand interactions are still challenging due to the difficulty of functional GPCR production. To date, only five inactivated GPCR structures, bovine rhodopsin (bRho) (PDB ID: 1U19 and 1F88) (Okada et al., 2004; Palczewski et al., 2000), human β_2 adrenergic receptor (β_2AR) (PDB ID: 2RH1)

(Cherezov et al., 2007), squid rhodopsin (sRho) (PDB ID: 2Z73) (Shimamura et al., 2008), turkey β_1 adrenergic receptor (β_1AR) (PDB ID: 2VT4) (Warne et al., 2008), and human A_{2A} adenosine receptor (A_{2A}R) (PDB ID: 3EML) (Jaakola et al., 2008), have been resolved at high resolution by X-ray crystallography. All resolved structures belong to GPCR class A family and these structures show extensive similarity in their overall structures consisting of seven transmembrane spans, an extra small helix (TM VIII) in the intracellular side and conserved disulfide bridges (see **Figure 1.4**). The structural divergence arising in the extracellular and intracellular regions, are likely to result in their specificity of ligand binding and G-protein coupling. However, the function of these loops is still in debate. The rhodopsin receptors are activated by light incident on the covalently bound retinal, and the three remaining structures are peptide-activated GPCRs, but none of these receptors were crystallised with their native peptide ligands. The structures of β_2AR with its inverse agonist carazolol, and A_{2A}R with its antagonist ZM241385 were crystallised using the T4 lysozyme fusion strategy and lipidic cubic-phase crystallisation methods. Most of the disordered third intracellular (IC3) loop was replaced with lysozyme from T4 bacteriophage to increase the available surface area potential for crystal contacts. In contrast to β_2AR and A_{2A}R, the structure of β_1AR with its antagonist cyanopindolol was crystallised using mutagenesis to improve its thermostability and vapour diffusion method.

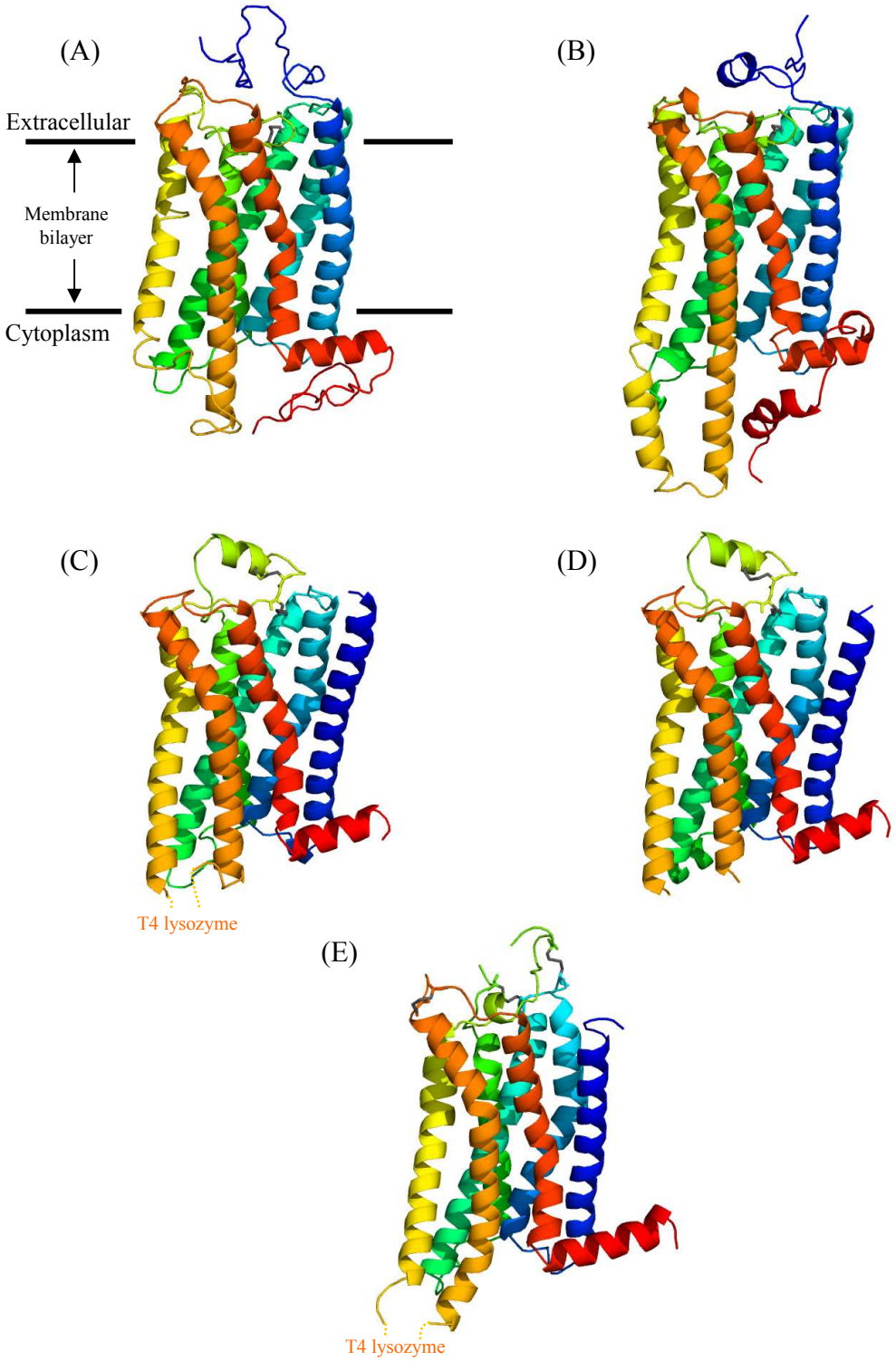


Figure 1.4: Crystal structures of inactivated GPCRs solved to date: bRho (1U19), (A); sRho (2Z73), (B); β_2AR (rRH1), (C); β_1AR (2VT4), (D); and A_{2A}R (3EML), (E). All structures are rendered with the same orientation and colour scheme in PyMOL: TM I (blue); TM II (marine); TM III (cyan); TM IV (green); TM V (yellow); TM VI (orange); TM VII+TM VIII (red). The Cys residues which form disulfide bridges are coloured grey and rendered as sticks. The figure illustrates high overall structure similarity.

Extracellular regions of GPCRs

The extracellular region (EC1, EC2 and EC3 loops) of the βARs is more open than that of rhodopsin (**Figure 1.5A and 1.5B**). There is a short helical segment within EC2 loop, stabilised by limited interactions with EC1 loop and two disulfide bridges, one within the loop and the other fixing the whole loop to the top of TM III. The EC2 loop of βARs forms the top of the ligand-binding pocket, exposed to the extracellular medium, whereas that of rhodopsin is a compact β-strand structure. The EC3 loop of βARs has no interactions with other extracellular loops, while the whole extracellular region of rhodopsin forms a compact folded unit. These observations would suggest that the extracellular region of the β-adrenergic family has evolved to allow access to the ligand binding site (Cherezov et al., 2007; Okada et al., 2004; Warne et al., 2008). The EC2 loop of $A_{2A}R$ (**Figure 1.5C**) lacks the prominent secondary structural elements, such as β-strand and α helix, which were observed in the rhodopsin and βARs, respectively. Instead, the EC2 loop of $A_{2A}R$ is mainly a spatially constrained random coil having two disulfide linkages with EC1 loop and one disulfide bridge with the top of TM III. However, the tip of EC2 loop is highly flexible and not observed in the electron density maps. The small EC3 loop of $A_{2A}R$ (**Figure 1.5C**) contains an intraloop disulfide bridge whose function is unknown. There are also multiple polar and van der Waals interactions among the three extracellular loops. It appears that the extensive disulfide bond network forms a rigid, open structure exposing the ligand-binding cavity to solvent, possibly allowing free access for small molecule ligands (Jaakola et al., 2008).

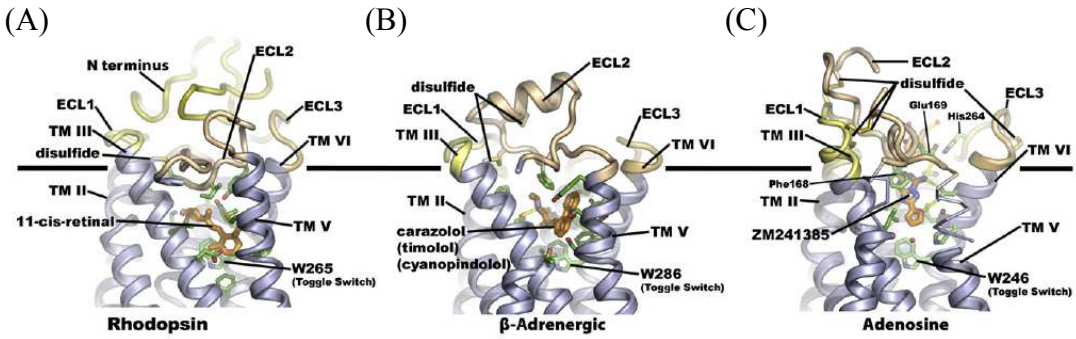


Figure 1.5: Extracellular view and ligand binding pocket of bRho (1U19), (A); β_2AR (2RH1), (B); and A_{2A}R (3EML), (C). Residues interacting with the ligand in each case (orange sticks) are coloured as green sticks. Polar interactions where applicable are shown as yellow lines between the interacting atoms (Hanson and Stevens, 2009).

Ligand binding to GPCRs

Although the helical bundle orientation is similar among the solved class A GPCRs, shifts in the relative positions of the various helices result in root mean square deviations between 0.7 and 2.4 Å depending on which structures are being compared (Hanson and Stevens, 2009). The retinal binding site of rhodopsin is quite similar to the βAR binding pocket which makes most contact with TM III, V and VI, but very different to that of A_{2A}R which shifts to the interface of TM VI and VII and is much higher as almost half of the ligand is entirely exposed to bulk solvent (**Figure 1.5**). The ligand ZM241385 for A_{2A}R also forms extensive interactions with EC2 loop and binds in an extended conformation perpendicular to the plane of plasma membrane (**Figure 1.5**). This suggests that there is no general, family-conserved receptor-binding pocket, so ligand selectivity can be achieved through the pocket in a variety of positions and orientations.

It has been suggested that the binding site of nonpeptide agonists/antagonists to a receptor is different from that of peptide ligands (Marshall, 2001). Nonpeptide

agonists/antagonists are revealed to bind within the transmembrane segments, whereas peptide ligands bind at the interface between the transmembrane sections and the extracellular loops. It is still unclear whether the GPCR-bound conformation of peptide is structurally related to nonpeptide agonists and antagonists. Peptide ligands bound to receptors belonging to the same family may have different structures, indicating receptor selectivity and specificity of the ligands. The conformation of bound ligand is important for rational drug design for the treatment of diseases caused by GPCR malfunctions. Most GPCR-targeted drugs are nonpeptide mimics to date and they are seldom able to mimic the interactions required to induce a full peptide ligand signal with selectivity so patients have several side effects from taking these drugs. There are many attempts to determine receptor-bound and unbound conformations of peptide ligands in order to understand the differences in receptor binding and activation, with obvious consequences for drug discovery.

Intracellular regions of GPCRs

Among GPCRs, IC2 and IC3 loops are believed to have an important role in the binding, selectivity and activation of G proteins. IC2 and IC3 loops are more likely to be important for the strength of the interaction and for specificity, respectively (Burstein et al., 1998a; Burstein et al., 1998b; Wess et al., 1990). Not surprisingly, the solved class A GPCR structures have similar conformations for IC1 loop, a sequence conserved region in all GPCRs, but different conformations for IC2 and IC3 loops (see **Figure 1.6**) (Cherezov et al., 2007; Jaakola et al., 2008; Okada et al., 2004; Shimamura et al., 2008; Warne et al., 2008). The overall structure of βARs are very similar with the exception of the second intracellular (IC2) loop, which forms a

random coil in β_2AR and a short helical conformation in β_1AR (**Figure 1.6B and 1.6C**), although this region is highly conserved among βARs. The helical segment of IC2 loop in β_2AR was not observed because of lattice contacts with adjacent molecules. In β_1AR, the IC2 loop makes different lattice contacts between each of the four molecules and it is therefore likely that the helical conformation found here represents the physiologically relevant structure for all βARs in the inactive conformation. The IC2 loop structure of β_1AR supports the proposed function of this loop as the switch enabling G-protein activation since Tyr149 in IC2 loop is located close enough to form a hydrogen bond with Asp138 of the highly conserved DRY motif at the cytoplasmic end of TM III. The interaction between the E/DRY motif and a glutamate residue at the base of TM VI as an ionic lock was believed to be an important component in stabilising the inactive state of most class A GPCRs but no interaction was observed in both βARs and A_{2A}R (**Figure 1.6**). This suggests that the ionic lock is not an important feature of the inactive state. The IC3 loop has been suggested to be one of the most flexible regions and a major site of interaction with G-proteins so it was modified by either deletion or replacement with an easily-crystallised protein to improve crystal formation for the crystallisation of βARs and A_{2A}R. Therefore, only bRho and sRho structures have native IC3 loops, the most notable different regions (**Figure 1.7**). It ascribes this difference to the extra sequence that squid rhodopsin possesses in this loop region. As the IC3 loop of bRho has a different conformation in a different crystal form, it has been debated that under physiological conditions the G-protein-binding site in the dark state of bRho is dynamically disordered. In contrast to the flexible IC3 loops of βARs and A_{2A}R, there are hydrophilic extensions of TM V and VI into the cytoplasmic medium, which forms

a rigid conformation in IC3 loop region in sRho as shown in **Figure 1.7**. The extension of both helices can be regarded as an essential structural motif for specifying the coupling mode with G_q-type G proteins as the amino acid sequence in this region is well conserved among invertebrate rhodopsins and other G_q-coupled receptors (Horn et al., 2003).

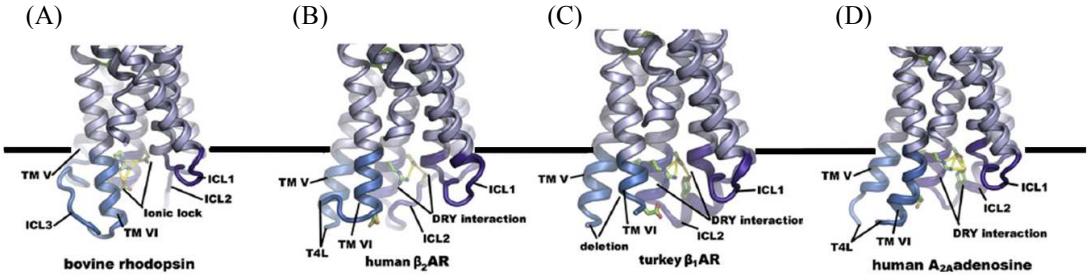


Figure 1.6: Intracellular view of bRho (1U19), (A); β_2AR (2RH1), (B); β_1AR (2VT4), (C); and A_{2A}R (3EML), (D). The interaction between the E/DRY motif and a glutamate residue at the base of TM VI as an ionic lock is only found in the bRho structure. With the exception of bRho, the aspartate residue of the DRY motif of TM III forms a polar hydrogen bond with either a serine or tyrosine residue on IC2 loop. Adapted from (Hanson and Stevens, 2009).

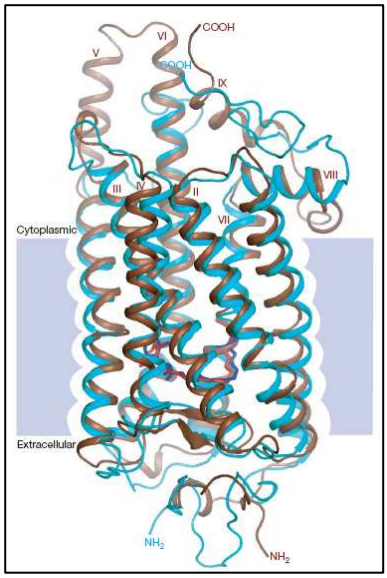


Figure 1.7: Structural comparisons between bRho and sRho. bRho is indicated by cyan and blue; sRho by brown and magenta (Shimamura et al., 2008).

In the past year, although two structures of opsin (unliganded form of rhodopsin) (PDB ID: 3CAP and 3DQB) have been resolved in the presence and absence of transducin peptide (Park et al., 2008; Scheerer et al., 2008), these structures may only be partially activated (Schwartz and Hubbell, 2008). There is therefore still a need to resolve the structures of diffusible-ligand binding GPCRs in both ground and activated states, and any conformationally distinguishable states in between (Jaakola et al., 2005).

1.2.3 Conformation of GPCR-activating peptide ligands

Neurotensin

Neurotensin (NT), a tridecapeptide: E-L-Y-E-N-K-P-R-R-P-Y-I-L, was originally isolated from bovine hypothalamic extracts by Carraway and Leeman (Carraway and Leeman, 1973). It acts as a neuromodulator in the brain and as a local hormone in the gastrointestinal tract. Amino acids 8-13 of NT (NT_{8-13}) are sufficient for biological functions including analgesia, hypothermia, and increased locomotive activity (Goedert, 1989; Kitabgi et al., 1977; Tanaka et al., 1990).

The NMR data indicated that neurotensin in aqueous solution and sodium dodecyl-d_{25} sulphate adopts unstructured conformation (Nieto et al., 1986; Xu and Deber, 1991). Mutagenesis and structure-activity studies combined with modelling techniques (Barroso et al., 2000; Pang et al., 1996) were also utilised to predict the receptor binding site and the conformation of bound NT_{8-13}. The model has shown that the ligand adopts a β-turn conformation, which is consistent with suggestion by other research groups (Sefler et al., 1995) and the hydroxyl group of Tyr11 forms an intra hydrogen bond with the guanidinium proton of Arg9. Nevertheless, the data obtained

from a series of double-quantum filtering 2D correlation experiments using solid-state NMR (SSNMR) were analysed by TALOS program and exhibited an unordered state of the NT_{8-13} in the absence of receptor and a β-strand conformation of the ligand in the presence of neurotensin receptor 1 (**Figure 1.8**) (Luca et al., 2003). Although TALOS program (Cornilescu et al., 1999) is widely used for backbone determination of proteins, the application of this technique to the study of small peptides which generally do not have enough restraints has been unpopular. Even though many attempts have been made to resolve the structure of the bound NT_{8-13} forms using computer modelling, solution-state NMR and SSNMR, its bound conformation is still under dispute.

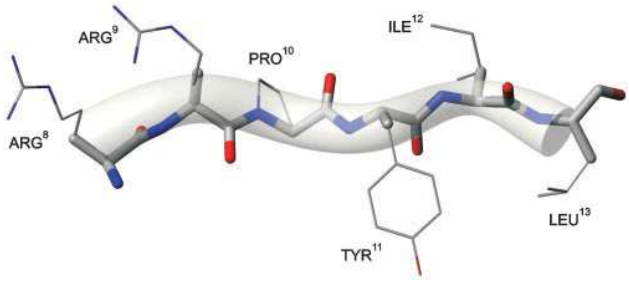


Figure 1.8: Representative backbone conformation of NT_{8-13} bound to its G-protein coupled receptor, NTS1. All side-chains are indicated for reference only, but with no refined conformation (Luca et al., 2003).

Other GPCR-activating peptide ligands

Glucagon, a 29 amino acid peptide, is an important peptide hormone involved in the regulation of glucose level in the blood. It is commercially available for the treatment of hypoglycemia and other clinical applications. The conformation of glucagon in aqueous solution was studied by Boesch and his co-workers using ^1H NMR approach and the result revealed that the prominent structure of monomeric glucagon is a flexible extended random coil form with some ordered structure of the

fragment 22-25 (Boesch et al., 1978). The conformation of glucagon is strongly dependent on surrounding conditions and conformation changes were induced in alkaline, detergent and lipid environments (**Figure 1.9**) (Braun et al., 1983; Epand et al., 1977; Gratzer and Beaven, 1969; Schneider and Edelhoch, 1972). The circular dichroism and infrared spectral studies have shown a predominant β-sheet structure for glucagon (Chou and Fasman, 1975; Gratzer et al., 1967). The molecular conformation of glucagon crystal was also investigated by X-ray crystallography and the structure has shown that it adopts a mainly large α-helical region (residue 6-29) (**Figure 1.9A**) (Sasaki et al., 1975). However, its biological conformation is not known except that it may adopt a helical conformation.

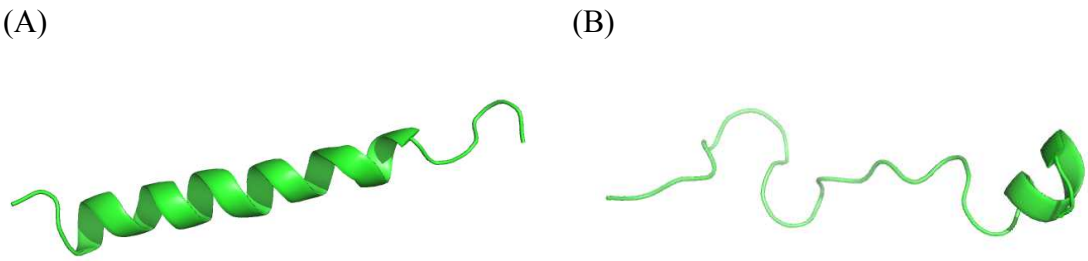


Figure 1.9: Solved conformations of glucagon under different conditions. Crystal structure of glucagon (1GCN), (A). NMR solution structure of glucagon in a lipid-water interphase (1KX6), (B). Rendered in PyMOL.

Angiotensin II (AII), an octapeptide hormone: D-R-V-Y-I-H-P-F, has the major physiological activity in increasing the blood pressure as a part of the renin-angiotensin system. It is the drug target for the treatment of hypertension and heart failure. AII in solution was investigated by various spectroscopic techniques, such as circular dichroism, optically detected magnetic resonance and nuclear magnetic resonance (NMR). The optical rotatory dispersion studies in aqueous solution at 20 °C, pH 6.0 have demonstrated the conformation of AII in random coil (Paiva et al., 1963).

The two-dimensional COSY and ROESY spectra of AII in aqueous solutions (90% H_2O/10% (*v/v*) D_2O) at 10 °C, pH 7.2 have indicated that it has some extended conformation with *cis-trans* isomerisation of the His-Pro peptide bond (Zhou et al., 1991). Using infrared and Raman spectroscopy, the spectra have revealed that this peptide adopts a predominant β structure with a small proportion of unstructured conformation in the dry state and in concentrated aqueous solutions (Fermandjian et al., 1972), however from the circular dichroism studies, a helical structure has been found in AII film, prepared from trifluoroethanol (Fermandjian et al., 1971). Conformational analysis of AII in dimethylsulfoxide and 2,2,2-trifluoroethanol/H_2O was performed using 2D ^{1}H-NMR spectroscopy. The backbone structure of AII is a S-shape conformation possessing a turn-like N-terminus (**Figure 1.10**) (Spyroulias et al., 2003). More recent studies of AII backbone conformation by UV resonance Raman and absorption spectroscopy have shown that this peptide has several conformations in aqueous solution, whereas it adopts mainly an ordered β-turn conformation in dodecylphosphocholine micelles (Cho, 1995). Analogues of AII adopt a γ-turn structure upon binding to the angiotensin receptor 1 (Schmidt et al., 1997). Therefore, it is more likely that the bioactive conformation of AII could be a turn.

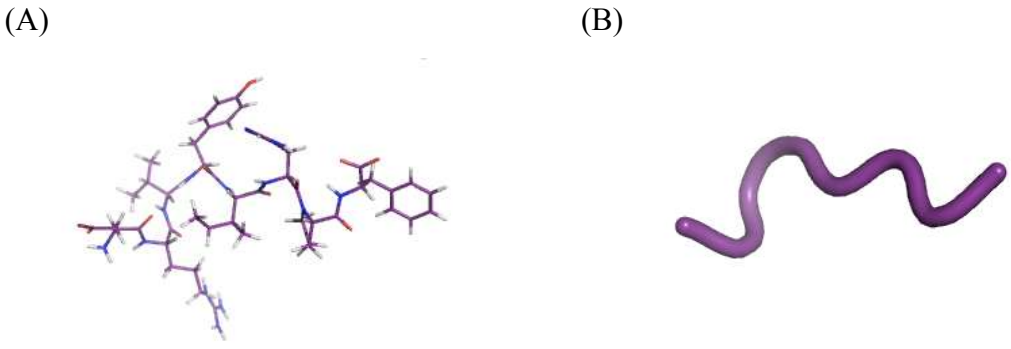


Figure 1.10: NMR solution structure of angiotensin II (1N9V) in dimethylsulfoxide and 2,2,2-trifluoroethanol/H_2O. The structure is rendered as line (A) and ribbon (B) representations in PyMOL.

Overall, there is no structure or more than one conformation of ligands in aqueous solutions but it is believed that the receptor-bound ligand has preferable conformation. Over one hundred agonists and antagonists were reviewed and the evidence suggests that GPCRs tend to recognise ligands with turn conformation (Tyndall et al., 2005).

1.3 Neurotensin receptors and neurotensin

To date, three characterised receptors for NT, termed NTS1 (Tanaka et al., 1990; Vita et al., 1993), NTS2 (Chalon et al., 1996; Mazella et al., 1996) and NTS3 (Mazella et al., 1998), have been identified in mammals. NTS1 and NTS2 share 60% homology and belong to the family of GPCRs containing seven hydrophobic transmembrane (TM) domains. In contrast, NTS3 comprises a single transmembrane domain and belongs to a Vps10p domain receptor family. Most of known effects of NT are mediated through the high-affinity NTS1 receptor.

1.3.1 Neurotensin receptor 1 (NTS1)

The first neurotensin receptor, named rNTS1, to be cloned from rat brain, was shown to consist of a 424 amino acid protein (Tanaka et al., 1990). The human neurotensin receptor (hNTS1) was later cloned from the colonic adenocarcinoma cell line HT29 and the protein sequence contains 418 amino acids which shares 84% amino acid identity and 92% amino acid similarity with rNTS1 (Vita et al., 1993). According to a high number of sequence identity and similar signalling properties among rNTS1 and hNTS1, their ligand binding sites are thought to be very similar.

The binding of [^{3}H]-NT and NT_{8-13} to its receptor, hNTS1, indicates dissociation constants (K_d) of 2.1 nM (±12%) and 0.63 nM (±11%), respectively (Harterich et al., 2008). The binding efficiency of NT is about 3-4 fold less than that of NT_{8-13}. Therefore, most studies of neurotensin focus on the functional fragment of NT, NT_{8-13}, however it is still interesting to understand the importance of the seven N-terminal amino acid of NT (NT_{1-7}), of which the function, if any, is still not known.

NTS1 and ligand binding

The NTS1 receptor has high affinity (K_d ~ nM) for NT, while NTS2 receptor has a very low affinity for the peptide and is selectively recognised by levocabastine, a non-peptide H_1 histamine receptor antagonist (Kitabgi et al., 1987). The NTS1 has cGMP, inositol 1,4,5-triphosphate and Ca^{2+} as second messengers (Mazella et al., 1987). Also, it is sensitive to Na^+ ions and GTP, which decrease its affinity for NT (Martin et al., 1999). Mutagenesis studies have revealed that the IC3 loop is involved in activation of the preferential rNTS1 receptor signalling pathway through G_q coupling, while the first half of the C-terminal domain most likely interacts with G_s to activate adenylyl cyclase and $G_{i/o}$ to abolish adenylyl cyclase activation or activate phospholipase A2 (Kitabgi, 2006). The introduction of single mutation, F358A, in TM VII led to the increase of agonist-independent inositol phosphate production and the constitutive inactivation of cAMP production. The constitutive activity of this mutant indicated that TM VII is involved in the transconformational changes, resulting in coupling of the NTS1 to G_q (Barroso et al., 2002).

Before any crystal structures of GPCRs at high resolution were available, Pang and her group proposed NT_{8-13} binding site comprising eight residues located in the

EC3 loop: Phe331, Ile334, Trp339, Phe342, Phe344, Phe346, Tyr347 and Tyr349, in which all aromatic residues form an aromatic pocket on the surface of the receptor based on homology modelling studies but there were no mutagenesis studies to support this idea (Pang et al., 1996). Later on some aromatic and charge residues in the EC3 loop of rNTS1 have been found to be important for NT binding by site-directed mutagenesis (Barroso et al., 2000; Labbe-Jullie et al., 1998). In addition, Arg327 in TM VI (Botto et al., 1997a; Labbe-Jullie et al., 1998), Asp139 in the EC1 loop (Barroso et al., 2000; Botto et al., 1997a), and an N-terminal extracellular segment near TM I (residues 45-60) (Labbe-Jullie et al., 1995) of rNTS1 are involved in ligand binding (**Figure 1.11**). Competition binding experiments of NT_{8-13} and its analogs containing modified side-chains with rNTS1 using ^{125}I-NT were performed and the data have shown that the Tyr11, side-chain methyl groups of Ile12 and Leu13 have the major role for ligand binding, and Pro10 and the positive charge of Arg8 and Arg9 are much less important. Mutagenesis and binding studies have shown that Tyr11 could have π-π interaction with Trp339, Phe344 and Tyr347 and a hydrogen bond might be formed between the hydroxyl group of Try11 and Try347, which was supported by solid-state NMR data obtained from our group (Williamson et al., 2002). It is also observed that Trp339 might interact with Pro10 in the ligand binding site. Met208 and Phe331 might interact through hydrophobic contacts with the side-chain methyl group of Ile12 and Leu13, respectively. Moreover, Phe331 may form cation-π interactions with the side-chain of Arg9. The functional importance of the C-terminal end of neurotensin was also studied and the results have revealed that the free C-terminus of NT is required for ligand binding by forming an ionic bond with Arg327 (Barroso et al., 2000). In contrast, the free NH_2 of NT is not involved in ligand binding (Lazarus et

al., 1977). Although it has been demonstrated that Arg8 is involved in ligand binding, here it does not seem to have any particular interactions with receptor residues (Barroso et al., 2000). The guanidinium group of Arg8 was proposed to form ionic interaction with Asp139 (Botto et al., 1997a) but it is thought that mutation of Asp139 to Ala might affect the formation of disulfide bridge between Cys142 and Cys225 rather than interacting with the ligand (Barroso et al., 2000). ^{13}C-chemical shift perturbation studies of bound NT_{8-13} to rNTS1 receptor by solid-state NMR (SSNMR) were performed and the results have shown that the carboxy terminus and tyrosine side-chain are important in the interaction of the neurotensin with the receptor-binding site. No perturbations were observed for the side-chains of both Arg8 and Arg9, suggesting little interaction with the agonist-binding site on the rNTS1 (Williamson et al., 2002). These SSNMR data support the above structure-activity studies and mutagenesis and modelling studies. For all structural studies, it will be useful to design labelled positions in neurotensin or neurotensin receptor for precise internuclear distance measurements using solid-state NMR techniques in the future.

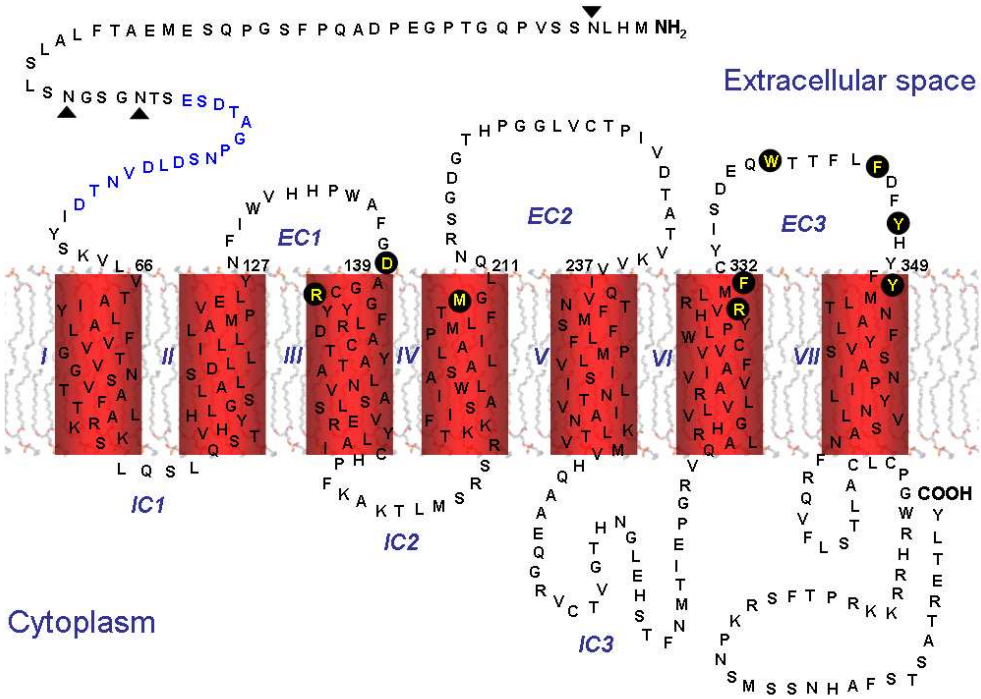


Figure 1.11: Schematic diagram of a rat NTS1 receptor embedded in a cell membrane. Residues involved in the binding of NT (> 5 times decrease of NTS1-binding efficiency after the mutation) are indicated by black circles and the N-terminal segment (residues 45-60), important in the ligand binding is shown in blue. The region of the third intracellular loop (IC3) is essential for G_q activation and coupling to phospholipase C. Putative glycosylation sites are indicated in black triangles.

Neurotensin receptors and antagonist binding

SR 48692, a selective, nonpeptide antagonist of the neurotensin receptor was developed by Sanofi-Aventis (**Figure 1.12**). The SR 48692 inhibits [^{125}I] NT binding in various cell types with K_i values ranging from 1 to 20 nm and the compound has higher affinity for the NTS1 than the NTS2 (Gully et al., 1993). The antagonist binding pocket was proposed based on homology modelling and mutagenesis studies. The model shows that it is located in the upper part of TMs IV, VI and VII and partially overlaps with the last two residues of NT binding site which explains the competitive behaviour towards NT binding. Putative essential residues in antagonist-receptor interactions were also proposed. The adamantine cage of the compound may

interact through hydrophobic contacts with the side-chain of Met208 in TM IV and Phe331 in VI. The carboxylic group forms ionic interaction with the side-chain of Arg327 in TM VI, confirmed in structural-activity studies using SR 48692 analogs. The pyrazole group may form π-π interactions with Tyr351 in TM VII. The chloroquinolinyl and the dimethoxyphenyl moieties may make π-π interactions with Tyr324 and Tyr351, respectively. The dimethoxyphenyl moiety can also form hydrogen bonds with some surrounded aromatic residues (Labbe-Jullie et al., 1998).

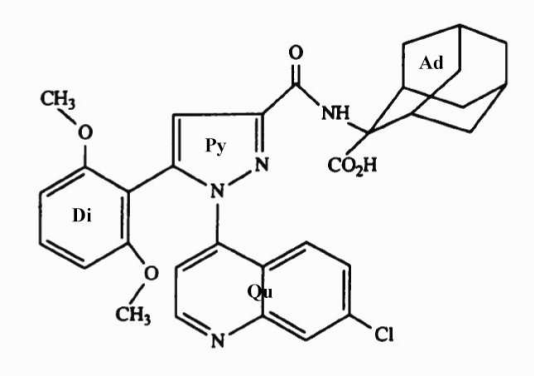


Figure 1.12: Chemical structure of the NTS1 antagonist, SR 48692. Abbreviations: Ad, adamantine; Py, pyrazole; Di, dimethoxyphenyl; Qu, quinolinyl. Adapted from (Gully et al., 1993).

1.3.2 Other neurotensin receptors

i) Neurotensin receptor 2 (NTS2)

The NTS2 receptor was cloned from rat (Chalon et al., 1996), mouse (Mazella et al., 1996) and human (Vita et al., 1998) brain. The rat and mouse NTS2 (416 amino acids) are slightly longer than the human NTS2 (410 amino acids). The rNTS2 shares 43% sequence identity and 64% sequence similarity with rNTS1. The NTS2 receptor has a shorter N-terminal extracellular tail and a longer third intracellular loop than the NTS1 receptor. In addition, it does not have putative N-glycosylation sites in the N-terminal region and the conserved Asp residue in TM II was substituted with Ala or

Gly. This replacement is responsible for the low sensitivity of the NTS2 receptor to Na^+ ions. The NTS2 receptors from rat, mouse or human in COS cells have similar affinity for NT (K_d = 2-10 nM). However, they bind SR 48692 with different affinities (K_d values of 82, 300 and 67 nM for rat, mouse and human NTS2, respectively). As the intracellular Ca^{2+} and IP_3 responses after NT binding can only be observed in the sensitive cell models like the oocyte expression system, but not in human embryonic kidney (HEK) cells, it suggested that the NTS2 is only weakly coupled to phospholipase C (Botto et al., 1998; Mazella et al., 1996). Furthermore, NT may act either as an agonist of the mouse and rat NTS2 or as antagonists of the human NTS2, in *Xenopus* oocytes or Chinese hamster ovary (CHO) cells from signal transduction studies. In contrast to NTS1, SR 48692 acts as a potent agonist for NTS2 in *Xenopus* oocytes and CHO cells (Botto et al., 1997b; Vita et al., 1998; Yamada et al., 1998).

ii) Neurotensin receptor 3 (NTS3)

The 100-kDa neurotensin receptor (NTS3) is the first transmembrane neuropeptide receptor that does not belong to GPCR superfamily. It was first found in an active form in buffer containing non-denaturing detergent CHAPS for solubilisation from newborn human brain. The protein was purified by affinity chromatography and it consists of 833 amino acids, which share 100% sequence identity with the cloned gp95/sortilin. The 100-kDa NTS3-gp95/sortilin is the mature form of the 110-kDa propeptide after the cleavage by the propeptide convertase, furin and contains high-affinity NT binding sites (K_d = 0.3 nM). (Zsurger et al., 1994). The structure of this protein is similar to sorting receptors, such as the mannose 6-phosphate and yeast Vps10p receptors bearing a single transmembrane domain. The physiological

functions of the NTS3 are still not known. It might regulate the biosynthesis and the translocation of GLUT4-containing vesicles, clear ligands from the circulation, or act as a sorting receptor (Lin et al., 1997).

1.4 Homology modelling

Since protein structure determination by X-ray crystallography and NMR spectroscopy is a long and complicated process, if a protein structure is wanted or needed, there is a reliable and booster way to calculate what the structure of that protein probably looks like. Homology (comparative) modelling is a cheap and fast method to approximate the 3D structure of a target protein sequence using an empirical 3D template structure with sufficiently high sequence identity. Built models are useful in driving experimental design that answer biological questions, such as guiding mutagenesis experiments (Vernal et al., 2002; Wu et al., 1999) and hypothesising about structure-function relationships (Bonneau et al., 2004; Bonneau et al., 2002). It is also useful for the rational design of drugs (modelling ligand docking) (Traxler et al., 1997; Vankayalapati et al., 2003). Comparative modelling process can be divided into 4 main stages (**Figure 1.13**): (i) template detection, (ii) sequence alignment, (iii) model generation, and (iv) model validation. The first two steps are very important keys to indicate how good the model is since the better template and alignment you use, the more accurate model you get.

The homologous protein structure can be easily obtained from database search programs like FASTA (Pearson, 1990) or BLAST (Altschul et al., 1990). The chosen homologs must have high sequence similarity and the minimum percentage of sequence identity for safe homology modelling is 25% (Sander and Schneider, 1991).

Clustal is a computer program for multiple sequence alignment between a target sequence and template sequences. Two main variations of this program are ClustalX (Thompson et al., 1997) and ClustalW (Thompson et al., 1994). Insertions (gaps in the template sequence) and deletions (gaps in the model sequence) are sometimes in the sequence, so alignments tend to be wrong around those regions. These problems need to be corrected by hand. The common method is to just remove these areas from the target protein sequence and build the model without the regions. It is also possible to get another template that has the required portion and it can then be used to model the insertion and deletion regions. Alignments could be improved by moving the residues around the gaps which should be placed in loop and turn portions instead of helix and β parts.

Once the alignment is ready, the model can be built by several servers such as SWISS-MODEL (Guex and Peitsch, 1997), WHAT IF (Vriend, 1990), MODELLER (Eswar et al., 2007), and so on. The correctness of a model is dependent on the quality of the sequence alignment, and the accuracy of a model is restricted by the deviation of the used template structures. Since X-ray protein structures always contain errors (Hooft et al., 1996) which might be from poor electron density in the X-ray diffraction map or human errors, the constructed model must also contain some errors. Therefore, the quality of a model needs to be verified. Many evaluated programmes and servers can be used, for example PROCHECK (Laskowski et al., 1993), ProsaII (Sippl, 1993), WHATCHECK (Hooft et al., 1996), and Rampage (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). The correct fold of a model should be checked first and ProsaII program can be used for this purpose. Then, the stereochemistry of the model has to be evaluated. The prediction accuracy is commonly quite high so if the

residues of regular secondary structure of a protein model are located at the right areas in the Ramachandran plot, the model is more likely reliable. If errors are in local regions such as a loop which is far away from binding or active sites, then it is less of a problem because even with good templates, there are several possible loop conformations, and so the loop parts of a protein model are usually the least reliable. A visualisation tool is also important to verify the model. Nowadays, there are many choices for such tools, such as Rasmol (http://www.rasmol.org/), VMD (Humphrey et al., 1996), PyMOL (http://www.pymol.org/), Swiss-PdbViewer (Guex and Peitsch, 1997), and WebLab Viewer (Accelrys Inc.). In addition, energy minimisation and molecular dynamic (MD) simulations of homology models can be performed to detect and remove small errors for model optimisation. However, if large errors are observed, this means the whole process of model building must be iterated.

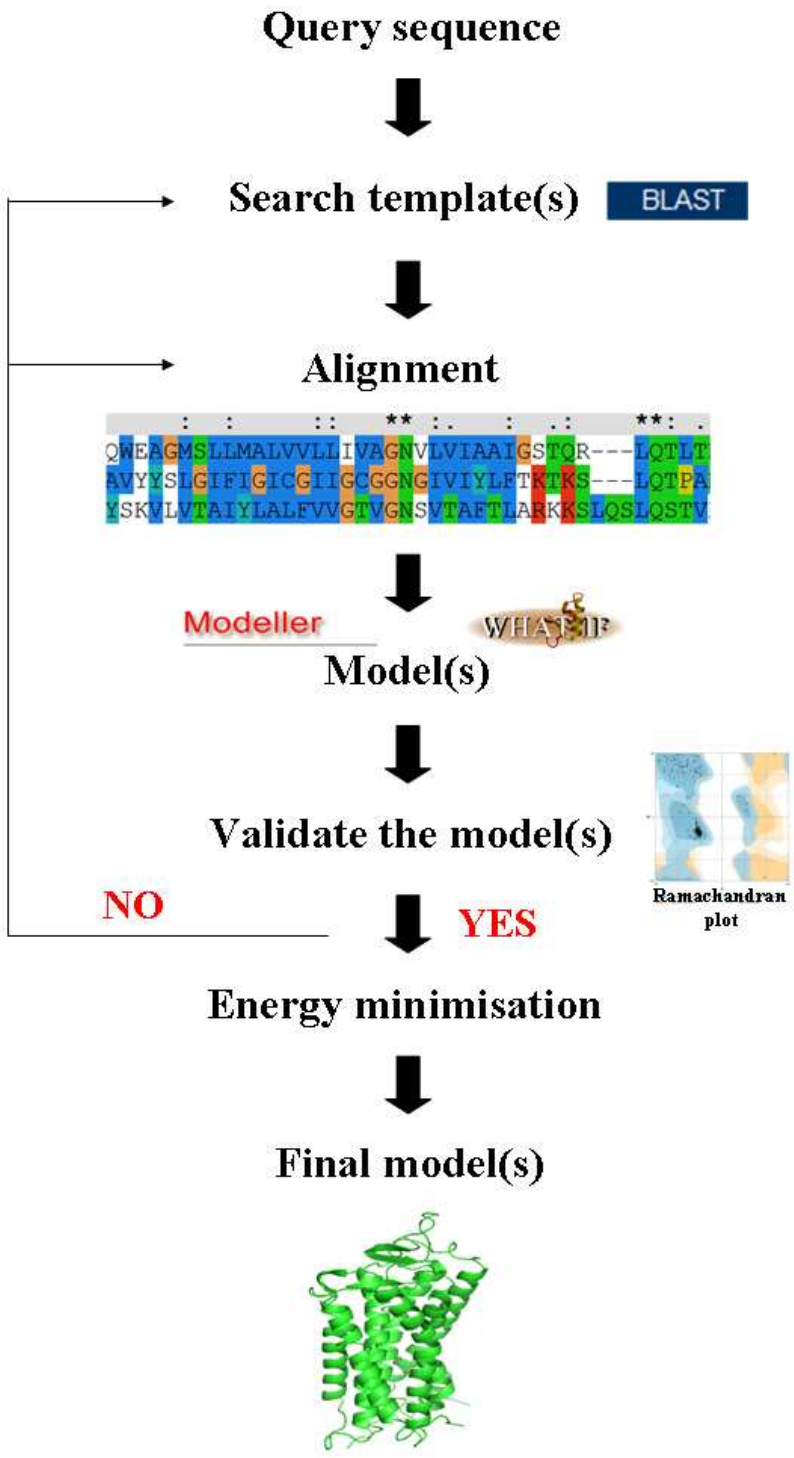


Figure 1.13: A generic flow-chart of homology modelling.

1.5 Docking simulations and molecular dynamic simulations

As mentioned earlier membrane proteins are very important in pharmacology, but their structural and functional information are still very scarce. X-ray crystallography and NMR approaches provide static structures, although membrane proteins do not exist in one conformation, but switch between different structures of their conformational space. The conformational changes can be associated with the protein function. Molecular dynamic (MD) simulations can calculate the time dependent behaviour of a membrane protein. These methods are now regularly used to study the structure, thermodynamics and kinetics of biological molecules such as protein-drug interactions, protein stability and folding (Karplus, 2002). The disadvantages of MD simulations are the limitation of time scales, usually no more than 100 ns, and the size of system. Nowadays, one of the most common used programs to study macromolecule dynamics is Gromacs (Van Der Spoel et al., 2005), which uses classical molecular dynamics theory (Hess et al., 2008). The program usually focuses on two methods, energy minimisation (EM) and molecular dynamics. The goal of EM is to find a steepest descent local energy minimum occupied by the initial conformation and to remove bad van der Waals contacts as the consequence. To simulate the natural motion of biological macromolecules, input files, a solvated structure, minimisation, equilibration must be set up for Gromacs environments. Then, various methods as part of Gromacs for analysing and interpreting a MD simulation trajectory are performed.

Protein-ligand docking is a molecular modelling technique for computer-aided rational drug design. The aim of this technique is to predict the position and orientation of a ligand bound to a protein receptor. It has been widely used in the

virtual screening of large databases of available chemicals in order to select putative drug candidates, but a protein crystal structure is needed to obtain reliable results. There are two common approaches based on molecular dynamics to perform protein-ligand docking. The first approach is the incorporation of ligand and receptor flexibility and their freedom of movement in MD simulations. Docking of different random initial positions of KTX ligand in the extracellular face of KcsA-Kv1.3 embedded in a POPC membrane was performed by Gromacs 3.3 with no bias or artificial driving force. The results revealed spontaneous and stable binding of ligand to the potassium channel (Zachariae et al., 2008). However, the major drawbacks with MD simulations are that it is a time-consuming method and also the system can get trapped in local minima. The second approach is the use of a pre-defined bound ligand in MD simulations, obtained by docking software or experimental data. There are several protein-ligand docking software such as AutoDock (Morris et al., 2009), DOCK (Ewing et al., 2001), GOLD (Jones et al., 1997), GLIDE (Friesner et al., 2004), and so on. Automated docking software can be used to construct a set of protein-ligand configurations, which are then refined in MD simulations. The combined docking-MD approach has been demonstrated for a number of studies such as the binding of organic cations to gramicidin A and the drug-Vpu interactions (Kim et al., 2006; Patra et al., 2007).

Generally, docking programs are composed of two main parts, a search of the configurational and conformational degrees of freedom and the scoring or evaluation function (Brooijmans and Kuntz, 2003). The search algorithm utilises different positions of a flexible ligand as well as torsional degrees of freedom of the ligand. It contains two important components, speed and effectiveness in covering the relevant

conformational space. The scoring function assigns the most favorable scores to the ligand-receptor complex. It must be good enough to differentiate the true binding modes from all the other alternative modes explored and to rank them consequently. The scoring function normally evaluates both the steric and chemical complementarity between the ligand and the receptor.

Molecular docking studies of both ligand and receptor in rigid geometry were successfully performed to determine a potential ligand binding site at the first time. For rigid docking, it means that the search algorithm uses translational and rotational degrees of freedom of the ligand only (without internal degrees of freedom), but this algorithm has a clear limitation (Kuntz et al., 1982). It has been shown that the ligand flexibility in docking simulations gives a reasonable binding geometry when compared to the one observed crystallographically (DesJarlais et al., 1986). The binding site and binding orientation can be affected by protein flexibility (Teague, 2003). The effect of protein conformation on ligand-docking accuracy was examined by comparing the docking of the ligands back to its own co-crystal structures and to the unbound structures. The docking accuracy decreased greatly once the unbound structures were used (Erickson et al., 2004). Many docking algorithms have been developed for a few decades to include ligand and receptor flexibility in docking simulations. However, the receptor flexibility is still a formidable challenge to determine the ligand binding conformations, especially without the knowledge of the binding site. So far, only partially flexibility of the protein receptor is in automated docking packages, for example GOLD providing flexibility of protein polar hydrogens (Jones et al., 1997) and AutoDock including flexibility of protein side-chains (Morris et al., 2009).

1.6 Introduction to NMR of membrane proteins

1.6.1 Principle of NMR

Nuclei consist of neutrons and protons. To be NMR active, the nuclei need to possess a property called the Spin (*I*) (Levitt, 2001). When introduced in an external magnetic field B_0, the nuclear magnetic moments align themselves with the external field into $(2I + 1)$ possible orientations, *e.g.* two orientations (parallel and anti-parallel) if $I = ½$ as in ^{1}H, ^{13}C or ^{15}N. This generates a small magnetic field. The nuclei therefore possess a magnetic moment, μ, which is proportional to their spins, *I*.

$$\mu = \frac{\gamma I h}{2\pi} \tag{1.1}$$

The constant, γ, is called the gyromagnetic ratio and is a fundamental nuclear constant which has a different value for every nucleus. *h* is Planck's constant.

The net effect is the precession of the magnetic moments about B_0. The frequency of this precession is found from the equation:

$$\omega_{Lar} = -\gamma B_0 \tag{1.2}$$

where ω_{Lar} is the Larmor frequency.

Since the ω_{Lar} is directly proportional to the strength of applied magnetic field, increasing the magnetic field of a spectrometer will have many effects: an increase of the Larmor frequency of the various nuclei; an increase of sensitivity by increments from energy gap between spin orientations (the sensitivity varies as the cube of the Larmor frequency); a simple spectrum with higher resolution.

The overall net magnetisation at equilibrium produced from a population difference existing between spin states can be perturbed into a non-equilibrium arrangement through the application of radio frequency pulses. Plotting of time versus energy absorption (NMR signal or free-induction decay, FID) can be generated and Fourier transformation converts this time domain data to the frequency domain data, called the NMR spectrum.

Nuclear spins engage in a number of external and internal interactions giving an NMR spectrum. The energy of the interactions for a spin system (also called the Hamiltonian, H) is given by the sum of the individual interactions (Ernst et al., 1990):

$$H = H_Z + H_{CS} + H_D + H_J + H_Q \tag{1.3}$$

The Zeeman Hamiltonian (H_Z) determines the interaction of the nuclear spin I with the external magnetic field B_0.

$$H_Z = \gamma\hbar B_0 I_Z \tag{1.4}$$

where I_Z is the z-component of the spin operator I and $\hbar$ is Planck's constant divided by 2π.

Each nuclear spin experiences a slightly different field according to its chemical environments and the way that the surrounding electrons are affected by the external magnetic field. This is called chemical shift phenomenon. The Hamiltonian for the chemical shift interaction is

$$H_{CS} = \gamma\hbar(\sigma_{11}\sin^2\theta\cos^2\Phi + \sigma_{22}\sin^2\theta\sin^2\Phi + \sigma_{33}\cos^2\theta)B_0 I_Z \tag{1.5}$$

where θ and Φ describe the angles formed between the principle axis and the laboratory fixed axis. σ_{11}, σ_{22} and σ_{33} are principal tensor elements. These principal

elements are also defined as the shielding anisotropy ($\Delta\sigma$), asymmetry parameter (η) and the isotropic chemical shift (σ_{iso}):

$$\Delta\sigma = \sigma_{33} - \frac{1}{2}(\sigma_{11} + \sigma_{22}) \tag{1.6}$$

$$\eta = \frac{\sigma_{22} - \sigma_{11}}{\sigma_{33} - \sigma_{iso}} \tag{1.7}$$

$$\sigma_{iso} = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) \tag{1.8}$$

The chemical shift Hamiltonian describes the effect of the electronic distribution around the nuclear spin. Since the electronic distribution is not isotropic, the chemical shift interaction depends on the orientation of the nucleus with respect to B_0.

The dipolar Hamiltonian (H_D) describing the through-space, or direct dipolar coupling between two nuclear spins I_I and I_J can be written as:

$$H_D = I_I \, D \, I_J \tag{1.9}$$

where D is the dipolar coupling tensor. In the cases of homonuclear spin interactions only, the dipolar Hamiltonian is

$$H_D = \frac{1}{2}\frac{\gamma_I\gamma_J}{r^3}(1 - 3\cos^2\theta)(3I_{IZ}I_{JZ} - I_II_J) \tag{1.10}$$

and the Hamiltonian describing a heteronuclear dipolar coupling is

$$H_D = \frac{1}{2}\frac{\gamma_I\gamma_J}{r^3}(1 - 3\cos^2\theta)(2I_{IZ}I_{JZ}) \tag{1.11}$$

where θ is the angle between the internuclear vector and the magnetic field. The dipolar coupling has an r^{-3} distance dependence, where r is the internuclear distance which provides the origin of distance restraints in NMR.

The J-coupling Hamiltonian relates to interactions between nuclear spins, corresponding to indirect (electron mediated, through bond) contributions. The J-coupling can be described by:

$$H_J = 2\pi I_I J I_J \quad (1.12)$$

where J is the indirect spin coupling tensor. In SSNMR, J-couplings are much smaller than the dipolar interactions and are often masked by line-broadening or other information.

The quadrupolar Hamiltonian (H_Q) determines the interactions between the nuclear spin $I \geq 1$ and the electric field gradients.

$$H_Q = IQI \quad (1.13)$$

where Q is the quadrupolar tensor.

During the experiment, if the exchange rate of two inter-convertible conformations, A and B of a compound is very fast, both conformations cannot be resolved in the spectrum. The interconversion between A and B is fast on the NMR time scale when it happens at a rate much greater than the difference in frequency between two conformations. In addition, the frequency difference between A and B depends on the Larmor frequency of the nucleus being observed and the strength of the applied field. Therefore, the NMR time scale relies on the particular experiment and the field strength.

In solution-state NMR (solNMR), the molecule rapidly tumbles in three dimensions in solution on an NMR time scale. It has no net average orientational preference with respect to an external magnetic field as the consequence of isotropic

tumbling. Relatively fast tumbling averages dipolar interactions to zero and chemical shifts to isotropic values, resulting in sharp NMR resonances. The Hamiltonian under these conditions may be simplified as:

$$H = H_Z + H_{CS} + H_J \quad (1.14)$$

However, traditional solNMR is of limited use for structural studies on large proteins, membrane proteins and solid materials, because long correlation times or restricted motions result in incomplete averaging of the anisotropic interactions leading to the broadened spectrum.

SSNMR techniques are designed to elucidate structural parameters of slowly tumbling or solid-phase samples at atomic resolution. They are also useful for dynamic studies, and have no theoretical limit of size (Watts, 2002). Unlike solNMR, the resolution and sensitivity of SSNMR are affected by the size and orientation-dependence of the nuclear spin interactions, *i.e.* the chemical shielding and the homonuclear and heteronuclear dipolar spin-spin couplings. These interactions are not averaged to zero, usually generating line broadening for resonance SSNMR spectra of a static sample. Sample orientation or high spinning of sample at the magic-angle spinning (MAS) together with isotopic labelling can improve both resolution and signal-to-noise ratio.

1.6.2 Basic solid-state NMR techniques

i) Magic angle spinning

Magic angle spinning is a technique often used in solid-state NMR for obtaining high resolution spectra (Andrew et al., 1958; Lowe, 1959). The basic idea is

to spin the sample rotor at 54.7° (the magic angle) with respect to the static field B_0 (**Figure 1.14**). A nuclear spin mainly experiences three interactions (dipolar, chemical shift anisotropy (CSA), and quadrupolar for spin ≥ 1) which often lead to very broad and featureless lines. These three interactions in solids are orientation-dependent and can be averaged by MAS. The nuclear dipole-dipole interaction between magnetic moments of nuclei averages to zero. The CSA, a nuclear-electron interaction, averages to a non-zero value. The quadrupolar interaction is only partially averaged by MAS leaving a residual secondary quadrupolar interaction. In liquids, *e.g.* a solution of an organic compound, most of these interactions will average out because of the rapid time-averaged molecular motion that occurs. This orientational averaging in solution is mimicked by MAS of a solid. At the spinning frequency, higher than the amplitude of weak spin interactions, it causes the signal to become narrow centre-bands at the isotropic chemical shift and rotational side-bands appear which are spaced by the MAS frequency. The intensity of spinning side-bands is weak in comparison with that of the central resonance but they can obscure other signals in the spectrum. High spinning frequency can reduce the number and intensity of rotational side-bands and therefore increase the centre-band intensity (**Figure 1.15**). Typical MAS frequencies are 5-20 kHz. The disadvantage of MAS is that the spectrum loses the geometry information due to the suppression of anisotropic interactions.

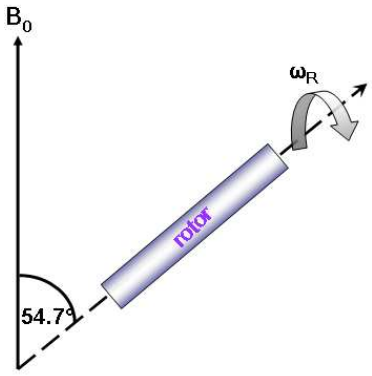


Figure 1.14: Schematic diagram for magic angle spinning (MAS). A rotor is spinning at ω_R at the magic angle, 54.7° with respect to the direction of the external magnetic field B_0.

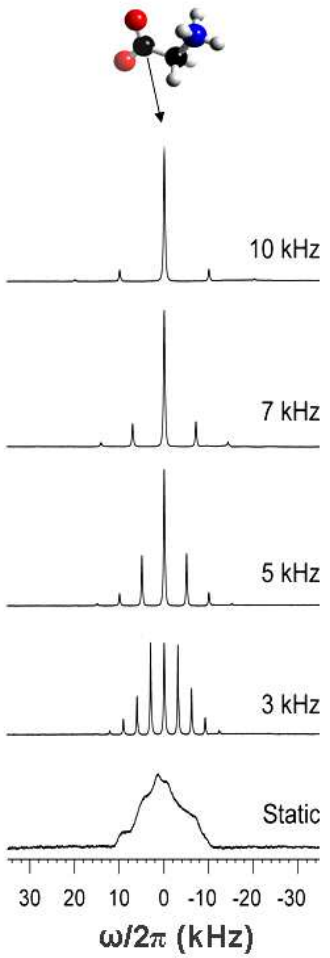


Figure 1.15: Solid-state ^{13}CO NMR spectra of solid glycine. The spectra show broad NMR resonance in a static sample and the effects of magic angle spinning at 3, 5, 7, 10 kHz. MAS collapses the broad lineshapes into sharp centre-bands at the isotropic chemical shifts and rotational side-bands spaced at the spinning frequency. Adapted from (Varga and Watts, 2007).

ii) Oriented samples

For oriented membrane samples, lipid bilayers naturally tend to self assemble and align themselves with flat surfaces. The molecules are aligned in a single specific orientation, rather than the normal random distribution found in powder samples. As a result, it can get rid of much of the resonance broadening. Alignment of membrane proteins can be done by layering the membranes onto glass slides (**Figure 1.16**). Orientational information, determined from dipolar couplings and chemical shift interactions, is dependent on the alignment of the molecule with respect to the direction of the applied magnetic field. Polarisation inversion spin exchange at the magic angle (PISEMA) is a common technique, applying the oriented sample to study the structure, dynamics and topology of aligned molecules such as membrane-associated peptides and proteins. For example the orientation of secondary structure elements by using ^{15}N labelled samples (Marassi and Opella, 2000; Ramamoorthy et al., 1999). However, aligning membrane proteins on glass slides is a difficult step and the resolution is often limited, particularly for large proteins (resonance overlapping). Moreover, the resolution relies on the degree of orientation of the membrane sample (Vosegaard and Nielsen, 2002).

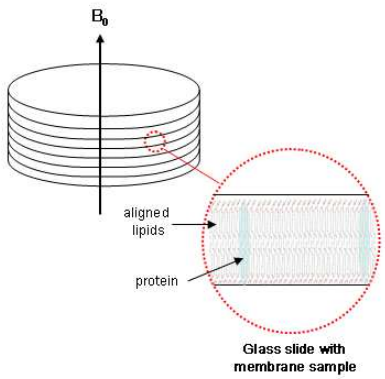


Figure 1.16: Schematic diagram for an oriented sample. The membrane sample is aligned on glass plates and oriented at the same direction as the external magnetic field, B_0.

Magic angle-oriented sample spinning (MAOSS), an experimental approach combining oriented sample with magic angle spinning technique, was proposed (Glaubitz and Watts, 1998). The advantages of this approach are obtaining both orientational information and high resolution at the same time. The method has been employed to study several membrane proteins, such as rhodopsin (Grobner et al., 2000), bacteriorhodopsin (Mason et al., 2004) and M13 coat protein (Glaubitz et al., 2000).

iii) Cross polarisation

Cross polarisation (CP) is a fundamental building block of most solid-state NMR radio frequency (rf) pulse sequences (Pines et al., 1972). CP can be used to enhance the signal of nuclei with low gyromagnetic ratio, S (*e.g.* ^{13}C, ^{15}N) *via* a transfer of magnetisation from a nucleus with high gyromagnetic ratio, I (*e.g.* ^{1}H). The standard CP pulse scheme is illustrated in **Figure 1.17**. In order to establish magnetisation transfer, the rf pulses applied on the two frequency channels must fulfil the Hartmann–Hahn condition (Hartmann and Hahn, 1962).

$$\left| \gamma_I B^I_{rf} \right| = \left| \gamma_S B^S_{rf} \right| \qquad (1.15)$$

where B^I_{rf} and B^S_{rf} are rf fields applied to spin I and S, respectively and γ_I and γ_S are gyromagnetic ratio of spin I and S, respectively. Under MAS, the Hartmann–Hahn matching condition breaks down into a series of narrow matching bands separated by the sample spinning frequency (Stejskal et al., 1977) and it becomes difficult to establish and maintain the efficiency of the magnetisation transfer on an exact match. A number of schemes have been proved to alleviate this problem, such as ramped CP

(Metz et al., 1994), variable amplitude CP (Peersen and Smith, 1993) and amplitude-modulated CP (Hediger et al., 1995).

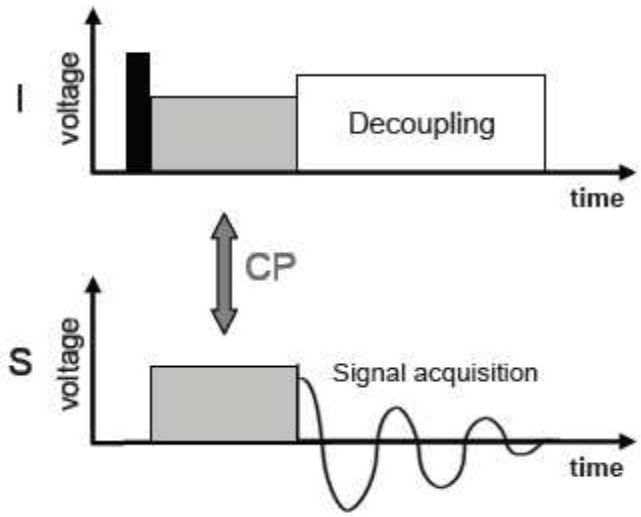


Figure 1.17: Standard CP pulse sequence with decoupling scheme during the signal acquisition.

The combination of the cross polarisation (CP) technique with magic angle spinning (MAS) has become one of the most commonly performed SSNMR experiments. This technique offers high-sensitivity of ^{13}C and ^{15}N in solids by polarisation transfer from abundant proton spins and by eliminating broadening from chemical shift anisotropy and dipolar coupling.

iv) Proton decoupling

The gyromagnetic ratio γ of proton nucleus is the second highest, just under that of tritium, and the proton nucleus has almost 100% natural abundance. Therefore, it is often used as the donor transfer of magnetisation in CP. Generally, it is needed to remove the heteronuclear dipolar coupling between the proton and the other nuclei in

order to increase the resolution of NMR spectra and to isolate spin systems. Several proton decoupling schemes have been introduced. The two most popular ones are the continuous-wave (CW) decoupling and two-pulse phase-modulated (TPPM) decoupling. In the CW decoupling, the spins are irradiated by a strong rf field of typically 50-150 kHz and the decoupling efficiency improves with increasing rf power. In contrast, the TPPM decoupling consists of two pulses with a phase difference between the two pulses. Experimentally, it was found that the TPPM decoupling is very sensitive to the parameter setting of the two pulses, the pulse length and the phase angle, however it is more efficient than the CW decoupling at the same power (Ernst, 2003).

v) Recoupling

Recoupling schemes have been devised to selectively retrieve the information loss of molecular geometry due to the use of MAS. These sequences aim to recouple the CSA, homonuclear or heteronuclear dipolar couplings selectively, thereby allowing the measurement of structurally important geometry parameters without loss of spectral sensitivity and resolution. Under the MAS condition, the anisotropic heteronuclear dipolar coupling becomes time-dependent and leads to a side-band pattern as does CSA so if one imposes an additional time-dependence on the spin system by rf irradiation, then one can selectively interfere with the decoupling process performed by the MAS.

Recoupling experiments can be used to measure the dipolar coupling between atoms. As the dipolar coupling is distance-dependent, this can be used to calculate interatomic distances in isotopically labelled molecules. The examples of recoupling

experiments for homonuclear spin systems are rotational resonance (RR) (Raleigh, 1988) and radiofrequency driven dipolar recoupling (RFDR) (Bennett et al., 1992) experiments. The examples of recoupling experiments for heteronuclear spin systems are dipolar-assisted rotational resonance (DARR) (Takegoshi and Terao, 1999) and rotational echo double resonance (REDOR) (Gullion and Schaefer, 1989) experiments. The DARR method will be explored in detail in Chapter 5. Here, the basics of two common techniques (RR and REDOR) using MAS combined with recoupling methods in order to get distance information are explained.

Rotational resonance

Rotational resonance (RR) is used to measure the homonuclear dipolar coupling between spin ½ pairs such as ^{13}C -^{13}C and ^{1}H-^{1}H pairs (Raleigh et al., 1989). The homonuclear dipole-dipole interactions between two nuclei can be reintroduced selectively through the RR condition, $\Delta = n\omega_R$, where ω_R is the sample spinning frequency, Δ is the difference in the isotropic chemical shifts of the coupled spins, and n is a small integer. This technique has been applied to structural determination of several proteins. For instance, ^{13}C-^{13}C RR technique was employed to investigate the molecular mechanism of transmembrane signalling of the bacterial chemoreceptor upon ligand binding (Isaac et al., 2002). The RR experiments were also used to measure the distance between the positively charged nitrogen and the carbonyl once the uniformly labelled acetylcholine bound to nicotinic acetylcholine receptor. A distance of 5.1 Å was measured indicating a bent conformation of bound acetylcholine (Williamson et al., 2007).

Rotational echo double resonance

Rotational echo double resonance (REDOR) is used to measure the heteronuclear dipolar coupling between spin ½ pairs such as ^{13}C-^{15}N and ^{13}C-^{19}F pairs and also between a spin ½ nucleus and a quadrupolar nucleus such as ^{13}C-^{17}O (Gullion, 1998). This technique depends on attenuation of signal from the observed nucleus (typically ^{13}C) by reapplying its dipolar coupling with heteronuclear spins. In REDOR experiment, for example ^{13}C-^{15}N dipolar coupling, the ^{13}C rotational echoes forming each rotor period after the ^{1}H-^{13}C CP transfer can be prevented from reaching full intensity by applying two ^{15}N π-pulses per rotor period. Changes in spectral intensities of the ^{13}C spectra obtained in the presence and in the absence of ^{15}N π-pulses are a direct measure of ^{13}C-^{15}N coupling constant which allows the internuclear distance to be calculated. The applicability of the method has been demonstrated for a number of rigid biological systems such as ^{13}C-^{19}F dipolar measurement to probe the conformation of the pentafluorophenylmethoxy analogue of TMPIP bound to the gastric H/K-ATPase (Watts et al., 2001) and the determination of ^{13}C-^{15}N distances of intermolecular chains of a pentapeptide of *Bombyx mori* silk fibroin (Kameda et al., 2002).

Although RR and REDOR techniques provide fairly precise distance measurements, there is a limited range of measurable distances (summarised in **Table 1.1**). Therefore, a structural model is important to choose the residue pairs to be measured within limited distances (Kovacs et al., 2007).

Table 1.1: Measurable distances (in Ångstrom) by RR (homonuclear spin systems) and REDOR (heteronuclear spin systems) techniques. Adapted from (Kovacs et al., 2007).

Nuclei	Fluorine	Phosphorus	Carbon	Nitrogen
Fluorine	15.3	11.5	9.8	7.3
Phosphorus		8.7	7.4	5.5
Carbon			6.3	5.3
Nitrogen				3.5

1.6.3 NMR studies of membrane proteins

Solution-state NMR (solNMR) can be used to determine three-dimensional structure and dynamics in solution. A small number of integral membrane protein structures have been studied using solNMR methods because of their stability in detergent micelles, such as KcsA (Baker et al., 2007), DsbB (Zhou et al., 2008), OmpX (Fernandez et al., 2004), Vpu (Park et al., 2003), VDAC-1 (Hiller et al., 2008), and phospholamban (Oxenoid and Chou, 2005). SolNMR also allows measurement of macromolecular motions at near-atomic resolution, but motions on different time scales and amplitudes providing functional information are not well understood. PagP, a bacterial outer membrane phospholipid transferase enzyme, in detergent micelles is an example of functional insight by TROSY-based NMR relaxation measurements. It revealed flexibility in the loops connecting individual β-strands (**Figure 1.18**) (Hwang et al., 2002) and dynamic switching between two states (Hwang et al., 2004). These results were linked to explain the required structural features for substrate recognition and the required conformational changes for enzymatic catalysis. However, it is generally difficult to keep membrane proteins active at high concentration for a period of experimental time in order to gain spectra with high signal-to-noise ratio and resolution. So far, the only partial three-dimensional GPCR structure, successfully

determined by solNMR method is rhodopsin, a stable GPCR expressed at high level (mg per litre or mM). The secondary structures of rhodopsin in the ground (PDB ID: 1JFP) and activated (PDB ID: 1LN6) states were from a set of overlapping peptides spanning the sequence of the protein and similar to transmembrane helices of rhodopsin solved by X-ray crystallography (**Figure 1.19**) (Choi et al., 2002; Yeagle et al., 2001). Some structures of GPCR fragments and domains were also characterised by solNMR techniques, for example the binding domain of human parathyroid hormone receptor (PDB ID: 1BL1) (Pellegrini et al., 1998) and the third intracellular loop of muscarinic acetylcholine receptor m3 (PDB ID: 2CSA) (Iverson et al., 2005).

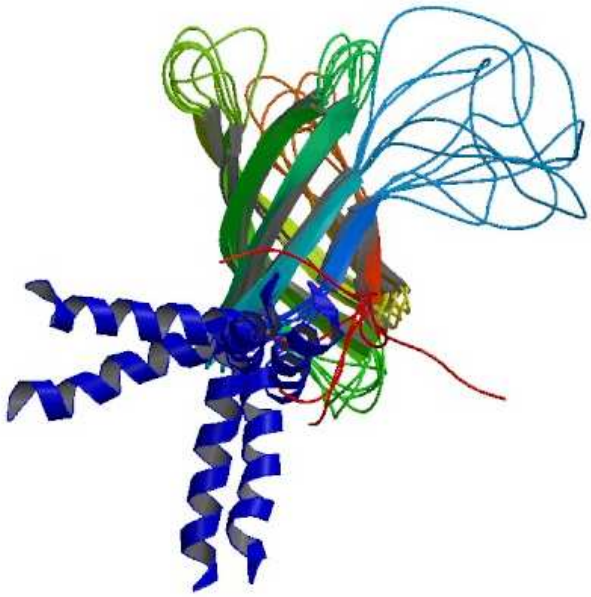


Figure 1.18: Ribbon representation of the solution NMR structure of the outer membrane enzyme PagP in DPC micelles (1MM4). The high flexibility in the extracellular loops is observed from overlay of 10 conformers (Hwang et al., 2002).

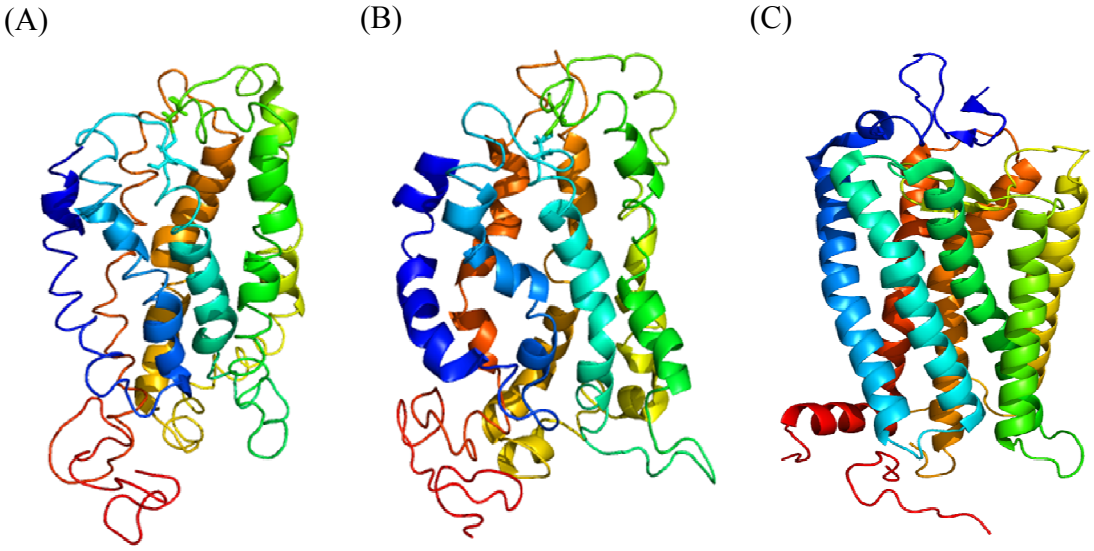


Figure 1.19: Three-dimensional structures of bovine rhodopsin. Solution NMR structures of rhodopsin in the ground (PDB ID: 1JFP), (A) and activated (PDB ID: 1LN6), (B) states. A crystal structure of inactive rhodopsin ground state (1F88), (C). Rendered in PyMOL.

NMR linewidths are normally proportional to the tumbling rate (τ_c^{-1}), so large molecules which have slow tumbling rates ($\tau_c^{-1} \sim 0.75\ kT/\pi r^3\eta$, where r = molecular radius and η = viscosity) will give large linewidths. Therefore, structural determination of the majority of membrane proteins by solNMR spectroscopy is still a significant challenge since the large molecular weight of the protein-lipid or protein detergent complexes cause the molecule to tumble slowly leading to poor resolution spectra. Lysolipids and detergents have been used to isolate membrane proteins in functional forms, to prevent protein aggregation and to increase protein mobility for solution studies, but they often form micelles instead of bilayer structures (Hagen et al., 1979; Krueger-Koplin et al., 2004; Peersen et al., 1990). Solid-state NMR (SSNMR) has been rapidly developed for studying membrane proteins and other non-soluble or large biological molecules since crystals and rapid isotropic tumbling are not necessary. It has the potential to gain accurate internuclear distances and

orientations which give information about molecular structures. SSNMR has been employed to provide structural insight and mechanisms of many membrane proteins, for example rhodopsin (Brown et al., 2009; Grobner et al., 2000), bacteriorhodopsin (Kamihira et al., 2005; Varga et al., 2007; Varga et al., 2008), gramicidins (Nicholson and Cross, 1989) and phage-coat proteins (Schiksnis et al., 1987; Thiriot et al., 2004). Some structures of membrane proteins determined by SSNMR experiments are summarised in **Figure 1.20**.

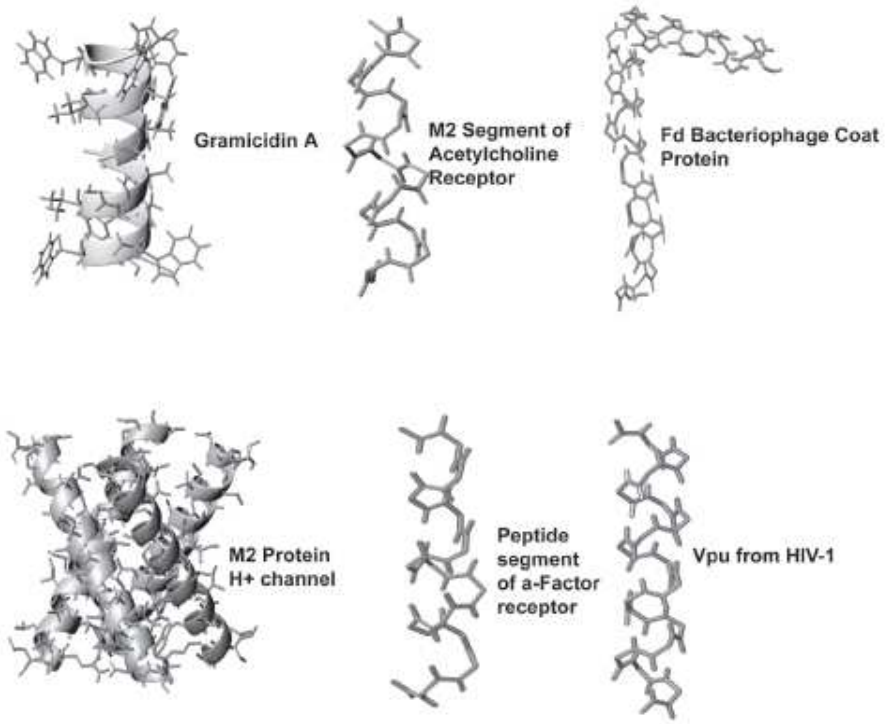


Figure 1.20: Three-dimensional structures of membrane proteins determined by solid-state NMR in oriented bilayers (Grisshammer and Buchanan, 2006).

1.7 Aims of the thesis

The principle aims of this project are to probe the interactions of neurotensin (NT) or truncated neurotensin (NT_{8-13}) with its receptor, rat neurotensin receptor 1 (rat NTS1), a member of GPCR class A family, and to determine the conformation of

bound ligands. Solution-state NMR (solNMR), solid-state NMR (SSNMR), homology modelling, docking simulations, and molecular dynamic (MD) simulations are major techniques which have been used to address the questions of interest in the thesis.

Using high resolution NMR techniques, selective or uniform isotope labelling of a peptide is required to dominate its signal from the large background of an unlabelled receptor protein. Chapter 2 will outline a strategy to specifically label peptide ligands and explain the use of manual Fmoc-solid phase peptide synthesis (SPPS) to synthesise selectively labelled peptides for distance measurement in peptides by SSNMR techniques and to produce unlabelled peptides for other applications. However, producing uniformly labelled peptides by the SPPS method is extremely expensive. Therefore, Chapter 3 will use an *Escherichia coli* expression system for uniform isotope labelling of NT and NT_{8-13} peptides. The expression procedure of the peptide will be optimised in order to maximise the yields.

A large quantity of receptor protein is necessary for structural studies by NMR techniques. The NTS1 protein has been prepared in two forms, detergent-solubilised and membrane-reconstituted samples for solNMR and SSNMR experiments, respectively (Chapter 4).

Unlabelled NT_{8-13} and uniformly ^{13}C, ^{15}N labelled NT have been characterised by 2D and 3D solNMR experiments (Chapter 5). Their resonances have been assigned in order to allow the perturbation shift of peptide bound to its receptor to be observed. A ^{1}H-^{15}N HSQC solNMR method will be employed to probe the resonances shift or broadening of the uniformly ^{15}N labelled NT in the presence of the detergent-solubilised protein. Sample conditions are important to obtain high resolution spectra from SSNMR experiments, so the unlabelled NT_{8-13} peptide in the microcrystalline

form is produced to examine its effects on ^{13}C NMR linewidths. Reconstituted NTS1 sample and uniformly ^{13}C, ^{15}N labelled NT will be used to investigate the protein-ligand interactions by dipolar-assisted rotational resonance (DARR) method, a SSNMR technique.

Computational methods, such as comparative modelling, automated docking and MD simulations, together with experimental data have been used to map the location and conformation of NT_{8-13} peptide at the putative binding site (Chapter 6).

The overall conclusion of the thesis and possible future work to obtain precise structural information of ligand when bound to its receptor are outlined in Chapter 7.

This thesis is completed by four appendices. In appendix A, general methods are described, such as *in vitro* site-directed mutagenesis, DNA transformation by heat shock, SDS-PAGE and Western blot analysis, silver stain, preparation of Ni-IMAC column, and ninhydrin test. The details of buffer compositions used in all experiments in the thesis are shown in appendix B. All crystallisation reagents used to precipitate the unlabelled NT_{8-13} peptide are displayed in appendix C and some solNMR and SSNMR spectra are shown in appendix D.

Chapter 2

Synthesis and characterisation of NT peptides for structural studies

2.1 Introduction

2.1.1 Peptide synthesis

Natural materials, peptides or small proteins, could be difficult to express in bacteria systems. Peptide synthesis can overcome this problem, and two major methods used are solution and solid phase peptide synthesis. The former is less common procedure because of unpredictable solubility characteristics of intermediates even though it is a more practical method for large scale production of well characterised peptides. The latter allows the synthesis of peptide/protein backbone modification, D-proteins, and selectively $^{13}C/^{15}N$ labelled peptides, which are widely used in NMR studies. In solid phase peptide synthesis (SPPS), the peptide is synthesised with the C-terminus covalently bound to an insoluble support. This allows easy separation of the growing peptide from unreacted precipitants and by-products (Merrifield, 1969). Protecting groups are used to mask reactive side-chains and the N-terminus from side-reactions during the synthesis. Common transient protecting groups for the N-terminus of amino acids are *tert*-butoxycarbonyl (*t*-Boc) and 9-fluorenylmethoxycarbonyl (Fmoc). Most laboratories use Fmoc, which is stable in acid but removable by base, for N-terminal protecting group because *t*-Boc requires

strong acid (HF) to cleave the desired peptide from insoluble supports whereas Fmoc needs less harsh precipitants (*e.g.* TFA).

The general scheme for Fmoc-SPPS is shown in **Figure 2.1**. Unlike natural protein synthesis by ribosomes, the synthesis proceeds in a C-terminal to N-terminal direction. The carboxyl group of the first Fmoc amino acid is covalently attached to a solid support resin, which bears a reactive group such as -OH. Then, the Fmoc group is removed by a base (usually piperidine) so that the next Fmoc amino acid can be coupled using *in situ* activation (all precipitants mixed together nearly at once) or a pre-activated species (carboxyl group activated prior to coupling). These steps are repeated until the desired peptide sequence is gained. The peptide-resin is then deprotected and TFA is applied to cleave the bond between the C-terminal amino acid and the support resin. Several scavengers are also included in the cleavage reaction to reduce side-reactions, which occur during the cleavage. For instance, once the *t*-butyl group is present, its carbocation can react with Trp, Tyr and Met to form *t*-butyl derivatives. Therefore, combination of 1,2-ethanedithiol (EDT) and water is necessarily added in the reaction for this case.

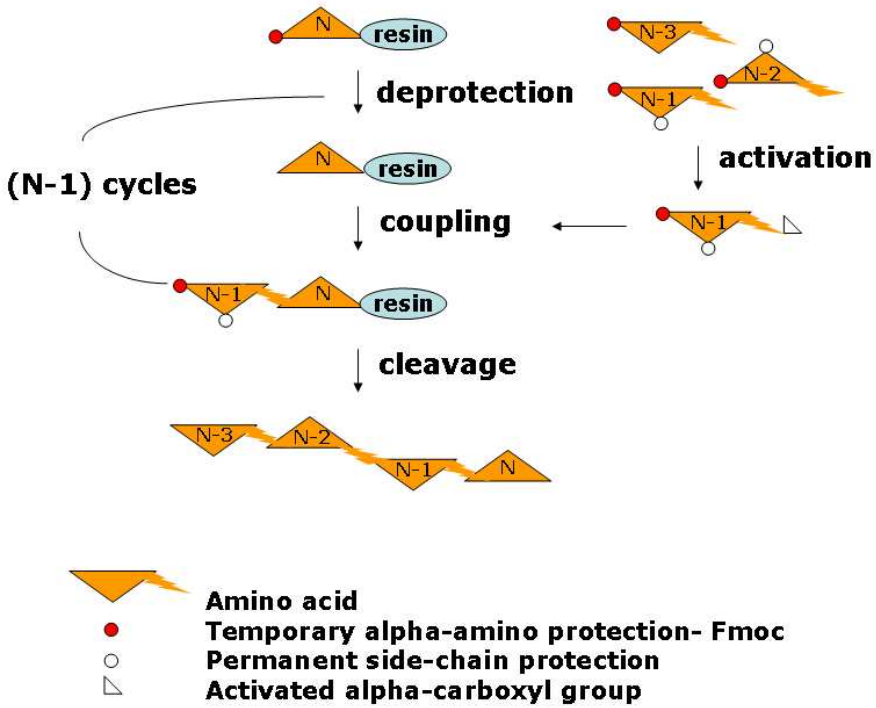


Figure 2.1: General scheme for Fmoc-solid phase peptide synthesis. The first Fmoc-amino acid (N) is covalently bound to a resin and the peptide-resin is then deprotected to provide the coupling of the second Fmoc-amino acid. After deprotection, coupling steps are repeated until the desired peptide is obtained, and the Fmoc group of the last amino acid is removed to release the peptide from the resin by the cleavage mixtures.

2.1.2 Selective isotope labelling

Incorporation of ^{15}N and ^{13}C labelled amino acids into ligands, peptides and proteins is an important method for determining the 3D structures by NMR spectroscopy (Fielding, 2003; Watts, 2002). Labelling aids assignments and increase sensitivity significantly. A peptide can be isotopically labelled in two ways, uniform and selective/specific isotope labelling, depending on the purpose of an experiment. Uniform isotope labelling can give spectra which are too complex, leading to misassignments, or an analysis which may be too complicated through unwanted dipolar coupling from neighbouring spin sites. Selective isotope labelling is commonly

used to aid assignment and structure determination of large proteins, but here it will be used for solving the conformation of a peptide by labelling at the desired spin pairs to facilitate the use of RR and REDOR techniques. These approaches will provide accurate distances leading to the structural model of a peptide at the binding site.

To focus on the site of interest within a peptide ligand, the labelling of the ligand has to be carefully designed as shown in **Figure 2.2**. RR and REDOR techniques require different strategies for ligand labelling. According to the RR condition in section 1.6.2: $\Delta = n\omega_R$, the target spin pairs must have appropriate chemical shift differences which are compatible with the MAS approach. For example, if the chemical shift difference is too high, the experiment will also require high spinning frequency which could exceed the maximum speed of a rotor. Therefore, the integer (n) has to be increased and the signal-to-noise ratio will be decreased as the consequence. In contrast, if the experiment requires too slow spinning frequency, it will be even worse since MAS sequence would not work so a poor resolution spectrum will be obtained. For a REDOR experiment, the spin of the nucleus is required to have as long a T_2 as possible. Carbonyl atoms are preferred because of their relatively long T_2 value. The short T_2 of observed spins causes intensity signal losses leading to the decrease of measure precision and distance limits. Design of selectively isotopically-labelled samples would be the most important step to address the question of interest and match with the technique requirements. Then, the selectively labelled peptide will be obtained by peptide synthesis method using site-specific isotopically-labelled amino acids as starting materials.

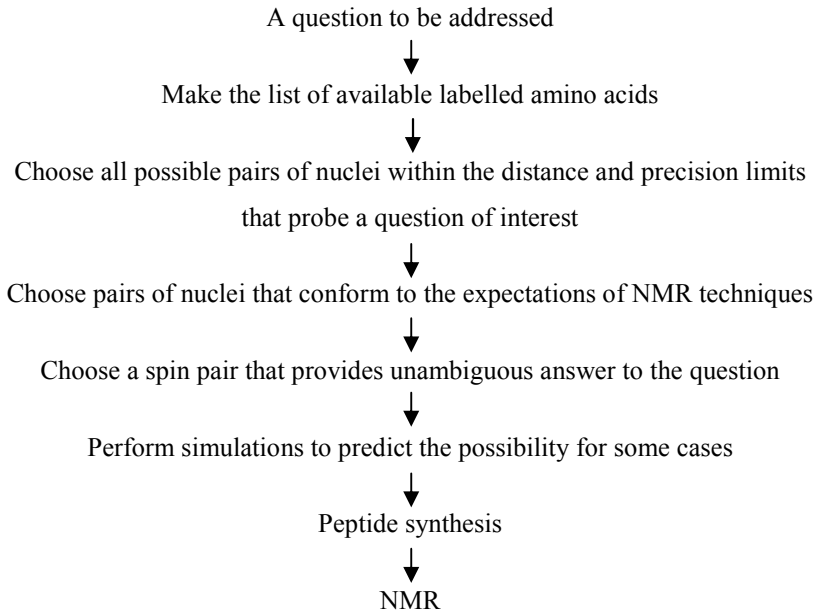


Figure 2.2: Diagram to design a specific-isotopically labelling scheme for NMR distance measurements.

2.1.3 Aims

The aim here is to obtain some site-specific isotopically-labelled NT_{8-13} peptides by manual Fmoc-SPPS method for peptide structural determination when bound to NTS1 using SSNMR. Two labelling schemes (**Figure 2.3**) to facilitate the elucidation of the backbone and some side-chain conformations have been designed, synthesised and purified. Selective labelling at ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 of NT_{8-13} will be used to resolve the conformation of Pro10 of NT_{8-13}. A labelling scheme of ^{15}N-Pro10, ^{13}Cα-Tyr11, ^{13}CO-Ile12, and ^{13}Cγ-Leu13 enriched NT_{8-13} will be employed to determine some backbone and side-chain conformation of the four C-terminal amino acids of NT_{8-13}. The homonuclear and heteronuclear distance measurements will be performed by RR- and REDOR-SSNMR experiments, respectively.

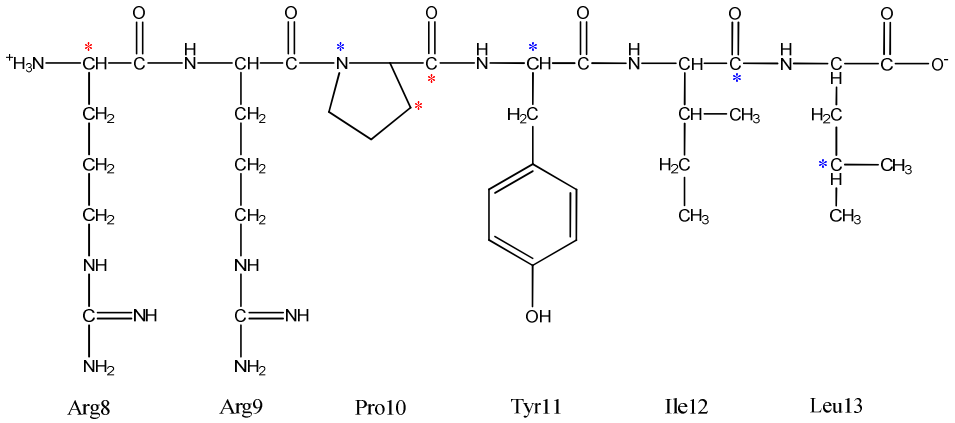


Figure 2.3: Diagram showing the sequence of NT_{8-13} with labelled sites for $^{13}C\alpha$-Arg9 and ^{13}CO-, $^{13}C\beta$-Pro10 labelled NT_{8-13} (red asterisks) and ^{15}N-Pro10, $^{13}C\alpha$-Tyr11, ^{13}CO-Ile12, and $^{13}C\gamma$-Leu13 enriched NT_{8-13} (blue asterisks).

2.2 Materials and methods

2.2.1 Molecular dynamic simulations

The initial coordinates of *cis* and *trans* configurations of Pro10 of NT_{8-13} peptides were generated by DeepView - Swiss-Pdb Viewer (Guex and Peitsch, 1997). The GROMACS 3.3 simulation package (http://www.gromacs.org/) was used for all simulations with a united atom Gromos96 force field (van Gunsteren et al., 1996). Each NT_{8-13} model was solvated with SPC water model (Hermans et al., 1984). Counterions were added to yield an electroneutral system that was energy minimised prior to starting simulations. Periodic boundary conditions were applied to the systems. Long-range electrostatics were calculated using the Particle Mesh Ewald (PME) method (Darden et al., 1993) with a real-space cut-off of 10 Å. For the van der Waals interactions, a cut-off of 10 Å was used. The simulations were performed at a temperature of 300 K using a Berendsen thermostat (Berendsen et al., 1984) with coupling constant of tau T = 0.1 ps. A constant pressure of 1 bar was maintained using

a Berendsen barostat with an isotopic coupling constant of tau P = 1.0 ps and compressibility = $4.5x10^{-5}$ bar^{-1}. The integration time step was 2 fs using LINC method (Hess et al., 1997) to constrain bond lengths. Coordinates were save every 1 ps for analysis. Analysis of all simulations was performed using the GROMACS package. VMD (Humphrey et al., 1996) was used for visualisation.

2.2.2 Synthesis of Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1,3-$^{13}C_2$]-L-proline

Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1,3-$^{13}C_2$]-L-proline were synthesised as previously described (Song et al., 2009). The strategy uses the Williams's oxazinone (Williams, 1992) as a precursor. Isotope labelling of L-arginine at position C-2 was prepared from [2-^{13}C] bromoacetic acid (**Figure 2.4**) whereas that of L-proline at position C-1 and C-3 was prepared from [1-^{13}C] bromoacetic acid and [3-^{13}C] iodide, respectively (**Figure 2.5**).

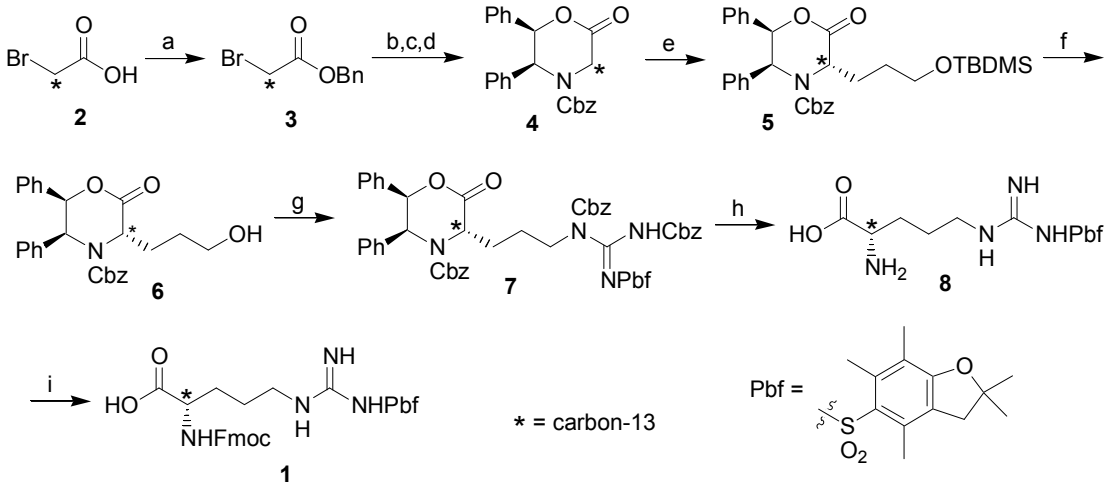


Figure 2.4: Reagents and conditions for making Fmoc-Pbf-[2-^{13}C]-L-arginine: a) BnOH, DCC, DMAP, DCM, r.t., 4 h, 100%; b) (*1R,2S*)-2-amino-1,2-diphenylethanol, Et_3N, THF, r.t., 2.5 h; c) Cbz-Cl, $NaHCO_3$, H_2O/DCM, 0 ^oC, 2.5 h; d) *p*-TsOH.H_2O, toluene, Dean-Stark, reflux, 60% over the 3 steps; e) 3-(*tert*-butyldimethylsilyloxy)-1-iodopropane, LiHMDS, HMPA, THF, -78 ^oC – r.t., 6 h, 63%; f) $KHSO_4$, THF/MeOH/H_2O, r.t., 3 h, 78%; g) **16**, DIAD, PPh_3, 80 ^oC, 16 h; h) H_2, Pd/C, MeOH/THF, 2 days, 18% over the 2 steps; i) Fmoc-ONSu, Na_2CO_3, dioxane/H_2O, 0 ^oC – r.t., 2.5 h, then citric acid, 75%. (Collaboration with University of Bristol)

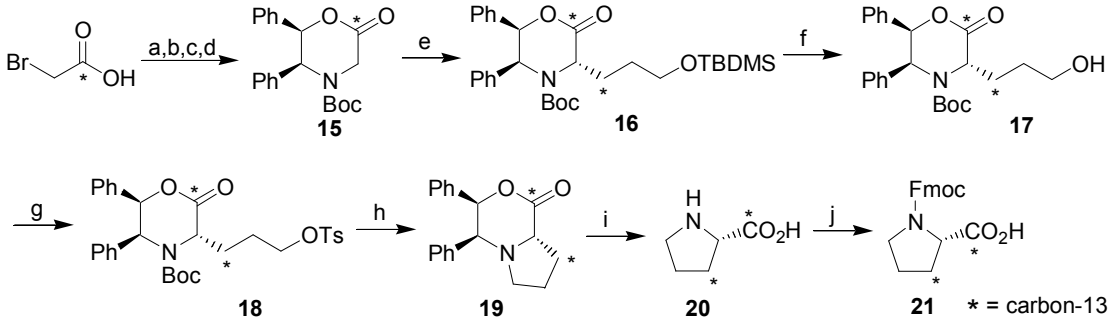


Figure 2.5: Reagents and conditions for making Fmoc-[1,3-$^{13}C_2$]-L-proline: a) BnOH, DCC, DMAP, DCM, 0 ^{o}C, 3 h, then r.t., 1 h, 89%; b) (*1R,2S*)-2-amino-1,2-diphenylethanol, Et_3N, THF, r.t., 2 h; c) Boc_2O, toluene, reflux, 18.5 h; d) *p*-TsOH.H_2O, toluene, Dean-Stark, reflux 1 h, 59% over the 3 steps; e) **14**, LiHMDS, HMPA, THF, -78 ^{o}C - r.t., 3.5 h, 48%; f) TBAF, THF, r.t., 16.5 h, 57%; g) TsCl, DMAP, DCM, r.t., 41 h, 67%; h) TFA, DCM, r.t., 2 h, then $NaHCO_3$ (aq), 63%; i) H_2, Pd/C, THF/EtOH, 21.5 h, 78%; j) Fmoc-Cl, Na_2CO_3, H_2O/dioxane, 0 ^{o}C - r.t., 17 h, 85%. (Collaboration with University of Bristol)

2.2.3 Fmoc protection of [^{15}N]-L-proline, [2-^{13}C]-L-tyrosine, [1-^{13}C]-L-isoleucine, and [4-^{13}C]-L-leucine

Fifty mmol of L-amino acid and 50 mmol of sodium bicarbonate anhydrous were added to a round-bottomed flask containing 15 mL chilled MilliQ water. Fifty mmol of Fmoc-*N*-hydroxysuccinimide ester in 15 mL acetone was then added. The pH of the mixed solution was adjusted to 9-10 with 0.1 M NaOH and stirred overnight at 4 ˚C. The solution was warmed to room temperature and the pH was adjusted to 1-2 with 5 M HCl. Acetone was removed under vacuum and the product was extracted with double volume of chloroform. The chloroform fraction was washed with equal volume of 0.1 M HCl and followed by MilliQ water. The final fraction was dried under vacuum. To crystallise, the product was dissolved in a minimum amount of warm (~40 ˚C) ethylacetate. An equal volume of warm petroleum spirit was added for Fmoc protection of labelled Pro, Ile and Leu and double volume was required for labelled Tyr. The product was allowed to precipitate overnight at 4 ˚C. The

precipitated product was collected by filtration through fine sintered glass. To obtain higher purity, the product was recrystallised. The purity and identity was verified by thin layer chromatography (TLC) using a chloroform:methanol:ammonia solvent system (in a ratio of 65:15:1 (v:v:v)) and electrospray ionisation (ESI) mass spectrometry.

2.2.4 Manual Fmoc-solid phase peptide synthesis

Unlabelled and labelled peptides were manually synthesised in a fume hood. All isotope-labelled amino acid derivatives were purchased from Cambridge Isotope Laboratories, Inc. Novasyn TGA resin was used to synthesise the peptides. The resin was swollen in dichloromethane (DCM) for 20 min, then washed and swollen with dimethylformamide (DMF) for 20 min. Loading of Fmoc amino acid on the resin was accomplished by *N*, *N*'- diisopropylcarbodiimide (DIPEA)-assisted coupling. No excess of all Fmoc-amino acids and 3 times excess of coupling precipitants were used in the peptide coupling steps. The amino acid and coupling precipitants (HOBt/HBTU) were resuspended in a minimum amount of DMF and 200 µl of DIPEA was added to start the reaction. The mixture was added to the resin and the coupling was allowed to proceed for 2 h with agitation. If the coupling was successful, a negative ninhydrin test (Appendix A.6) result was obtained.

Before Fmoc deprotection was carried out, the coupling solution was drained and the resin was washed with excess DMF. The resin was incubated in 20% (*v*/*v*) piperidine in DMF for 40 min with agitation for deprotection. A positive ninhydrin test indicates that the coupling has been successful. Subsequently, further coupling and deprotection cycles were carried out.

After the final deprotection, the resin was washed with excess DMF followed by excess DCM, methanol and diethyl ether. The resin was dried in a desiccator under vacuum overnight. The peptide was cleaved from the resin overnight with the following precipitant mix: 82.5% TFA, 5.0% Phenol, 5.0% MilliQ water, 5.0% thioanisole, and 2.5% EDT (*v*/*v*). The cleavage solution containing the peptide product was collected and the resin was washed 3 times with TFA. Excess TFA was evaporated off under a stream of nitrogen until around 1 mL remained. Cold diethyl ether (40 mL) was added, followed by centrifugation at 8000 xg at 4 ˚C for 10 min. The diethyl ether extraction was then repeated 3 times. The final pellet was dissolved in MilliQ water and then lyophilised.

2.2.5 High performance liquid chromatography (HPLC)

Small scale purification was performed using a C18 Hichrom 150x10 mm semi-preparative HPLC column at a flow rate of 3.5 mL/min and for large scale purification, a C18 Hichrom 150x21.2 mm preparative HPLC column was used at a flow rate of 14 mL/min. Lyophilised peptide was dissolved in MilliQ water containing 0.1% (*v*/*v*) TFA. Peptides were purified using a linear gradient of de-gassed MilliQ water and acetonitrile solvents containing 0.1% (*v*/*v*) TFA. Absorption of the eluted fraction was monitored at 220 nm and 280 nm. The major peak fraction was collected and analysed by ESI mass spectrometry. The acetonitrile in the product fraction was removed under vacuum using a rotary evaporator and the purified peptide was lyophilised. The lyophilised sample was subjected to a C18 Hichrom 150x4.6 mm analytical column at a flow rate of 1 mL/min and a linear gradient of 20% to 45% (*v*/*v*) acetonitrile for purity check.

2.2.6 Resonance assignments for labelled NT_{8-13} in solution

NMR experiments were performed on a Bruker DRX 500 MHz NMR spectrometer in chemistry department's facilities (Oxford). All recorded data were referenced to an external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) standard at 298 K. Approximately 5 mg of each labelled peptide was dissolved in 0.60 mL of H_2O containing 5% D_2O and transferred into a 5-mm NMR tube. All NMR data were processed with nmrPipe (Delaglio et al., 1995) and analysed with the program nmrDraw (Delaglio et al., 1995).

2.3 Results and discussion

Unlabelled and labelled peptides were synthesised with standard solid-phase methodology by manual Fmoc strategy. The synthesis was carried out at a 0.1 mmol scale using no excess amino acid in the fume hood. The progress of synthesis was checked by ninhydrin test for the presence of primary amides. All crude peptides were purified to greater than 95% purity by semi-preparative and/or preparative reversed-phase HPLC using a C18 column using a linear gradient of acetonitrile. The purified peptides were eluted at 26-27% (*v*/*v*) acetonitrile and analysed by ESI mass spectrometry. The purity of the product was checked from peak area in analytical HPLC. **Figure 2.6** is a representative of the purity of purified peptides.

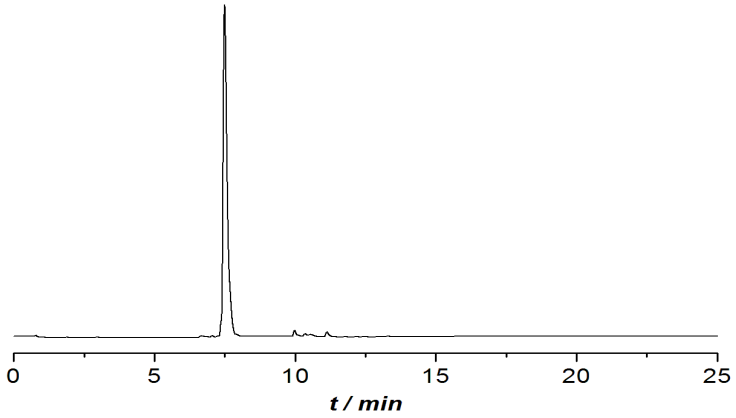


Figure 2.6: Analytical reversed-phase HPLC trace (280 nm) showing the purity of purified NT_{8-13}. The purified peptide was run by solvent A: 0.1% TFA in water; solvent B: 0.1% TFA in acetonitrile; gradient: 20-45% in 23 min. The purified NT_{8-13} peptide was eluted at 26.0% B.

2.3.1 The synthesis of unlabelled peptides

After the synthesis at a 0.1 mmol scale, the yields of crude peptides of NT and NT_{8-13} are about 71% and 84% compared to the theoretical maximum yield, respectively. **Table 2.1** shows the summary of the synthesis of unlabelled NT and NT_{8-13}. The peptides were purified by reverse-phase HPLC (**Figure 2.7**), and ESI mass spectra (**Figure 2.8**) of the purified peptides showed correct molecular weight. The final yields of purified NT and NT_{8-13} were about 36 and 24 mg, respectively.

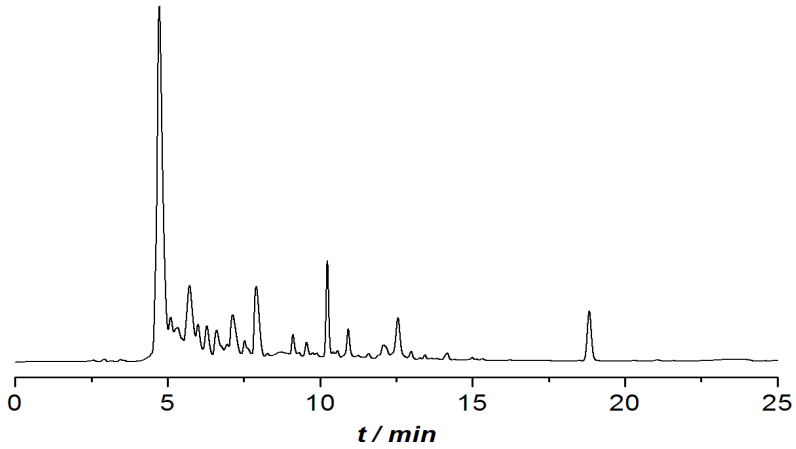
Table 2.1: Summary of the synthesis of unlabelled peptides. See **Figure 2.8** for details.

Sample	Final amount (mg)	Final yield (%)	Expected mass (Da)	ESI mass observed (Da)
Full length NT	36.0	30.0	1689.9	1689.49, 845.67, 564.54
NT_{8-13}	24.0	29.4	816.5	408.92

The unlabelled full length NT and NT_{8-13} were manually made in order to ensure that the manual peptide synthesis protocol was efficient in preparation for

labelled residues. The results showed that excess Fmoc-amino acids are not necessary for the synthesis of neurotensin peptides by manual Fmoc-SPPS method. This will save a lot of cost for the synthesis of labelled peptides. These unlabelled peptides will be used for crystallisation trials, NMR experiments and other applications.

(A)



(B)

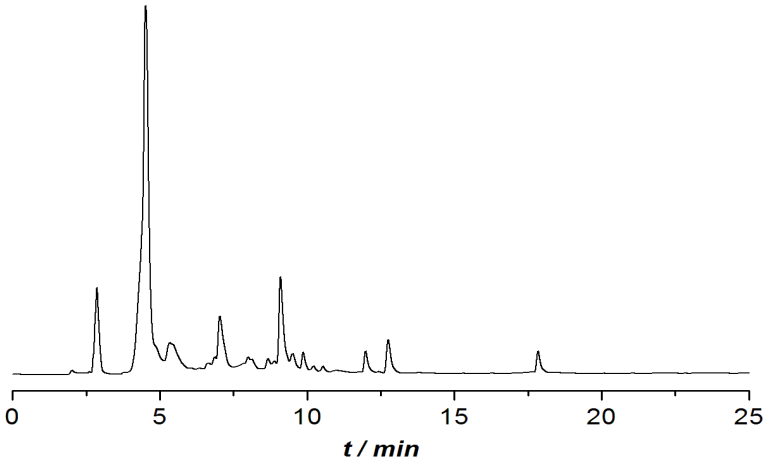
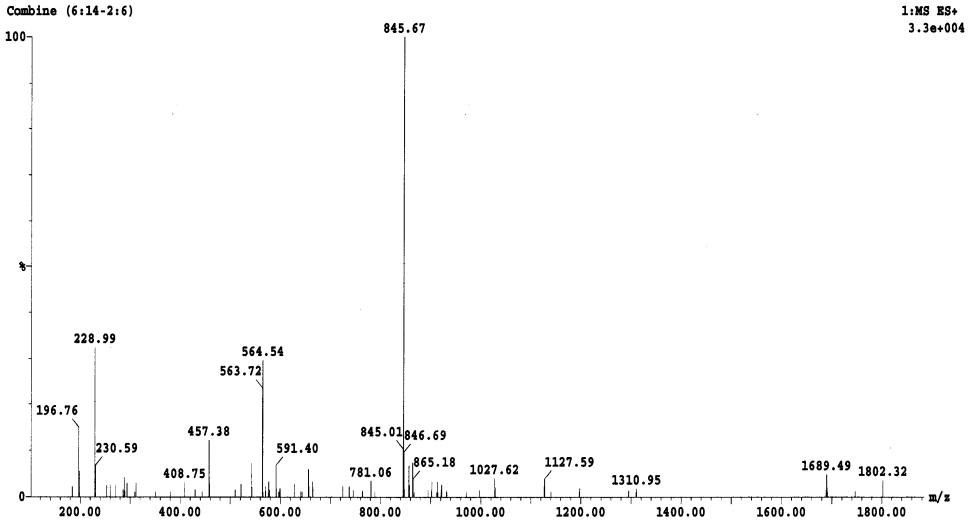


Figure 2.7: Reverse-phase HPLC traces (280 nm) for the full length NT (A) and NT_{8-13} (B) purification. The crude peptide was purified by solvent A: 0.1% TFA in water; solvent B: 0.1% TFA in acetonitrile; gradient: 20-70% in 23 min. The NT and NT_{8-13} peptides were eluted at 25.9% B and 25.4% B, respectively.

(A)



(B)

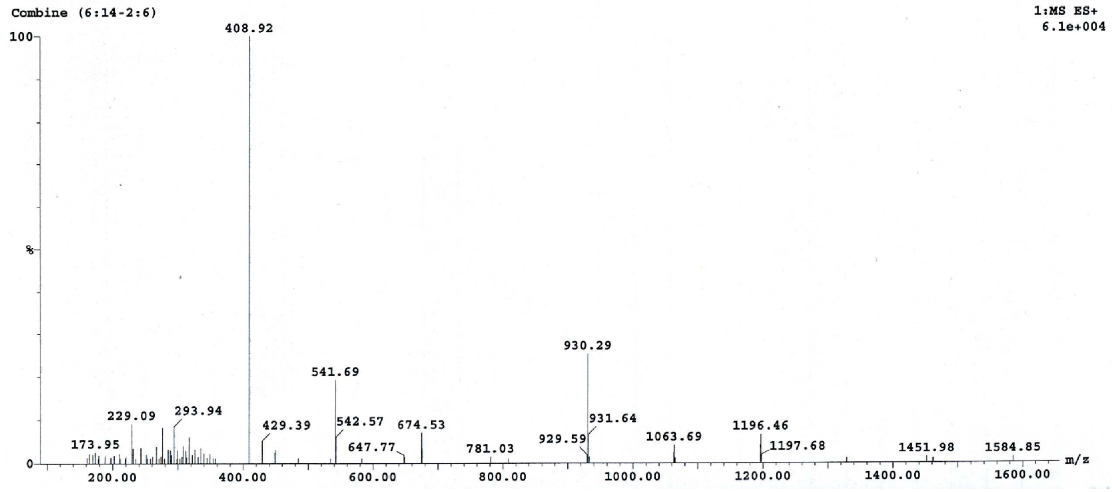


Figure 2.8: Positive-ion mass spectra of purified NT (A) and NT_{8-13} (B). Mass observed of NT is 1689.49 Da $(M+H)^+$, 845.67 Da $(M+2H)^{2+}$ and 564.54 Da $(M+3H)^{3+}$ and theoretical isotope mass number of NT is 1689.9. Mass observed of NT_{8-13} is 408.92 Da $(M+2H)^{2+}$ and theoretical isotope mass number of NT_{8-13} is 816.5.

2.3.2 The synthesis of selectively labelled NT_{8-13}

i) Labelling scheme 1

The hexapeptide, NT_{8-13} is sufficient and required for the binding of NT (Kitabgi et al., 1977), and most of studies focus on this functional fragment. Interestingly, there is a proline residue in the middle of NT_{8-13} sequence. Pro10 is expected to play a crucial role in determining the conformation of bound ligand (Tyndall et al., 2005). Distance measurements using selectively labelled NT_{8-13} are direct methods to answer the first question whether Pro10 is in a *cis* or *trans* conformation. A labelling scheme of NT_{8-13} has to be designed following the scheme in **Figure 2.2**. To address this question, the torsion angles of the backbone of Pro10 need to be determined. The *psi* angle of Arg9 and the *omega* angle between Arg9 and Pro10 are the most important indexes for the conformation of Pro10. The distance between Cα of Agr9 and Cα of Pro10 is the direct method to obtain the *psi* and *omega* angles and differentiates the *cis* and *trans* conformations of Pro10. Nevertheless, the distance between same nuclei which have no difference in chemical shifts cannot be measured experimentally using the RR technique, so an indirect method for measuring the distance between Cα of Agr9 and Cα of Pro10 by labelling at different nuclei is needed. Since there are very limited specifically labelled arginine and proline available commercially, it might be necessary to synthesise those labelled amino acids, even though it is time consuming and expensive. ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 labelled NT_{8-13} (**Figure 2.9**) has been designed to obtain higher sensitivity and to facilitate the distance measurement between Cα of Agr9 and Cα of Pro10 indirectly for the conformational determination of Pro10 by RR-SSNMR experiments. Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1,3-$^{13}C_2$]-L-proline are required for the synthesis of this

labelled peptide by manual Fmoc-SPPS. The first distances of ^{13}Cα Arg9 - ^{13}CO Pro10 which gives the highest chemical shift difference will indicate the conformation of Pro10 (**Table 2.2**). However, the second distance of ^{13}Cα Arg9 - ^{13}Cβ Pro10 will be used to confirm the Pro10 conformation by MD simulations. These two distances are within the limitation of ^{13}C-^{13}C pair (up to 7 Å), yet the precision of the second distance might be less compared to the first one since the spinning frequency is quite low. For instance, 4 kHz is the minimum speed for 6 mm thick wall NMR rotors in order to get a good resolution NMR spectrum. An oriented sample might be able to overcome the problem of low resolution spectra resulted from low spinning frequency.

MD simulations were used to determine whether site-specific labelling at ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 of NT_{8-13} is able to differentiate the *cis* and *trans* Pro10 isomers. Each starting model of NT_{8-13} with *cis* and *trans* configurations was placed in hydrated cell. All-atom simulations were run for 10 ns. Distances from Cα of Arg9 to neighbouring atoms of Pro10 (R9Cα-P10Cβ, R9Cα-P10CO and R9Cα-P10N) were obtained from the last 5 ns of equilibrated simulations. Correlations of Arg9-Pro10 dihedral angle and distances between R9Cα and P10Cα were demonstrated in **Figure 2.10**. The distances between R9Cα and P10Cα were geometrically calculated from the averaged distances of R9Cα-P10Cβ, R9Cα-P10CO and R9Cα-P10N. This labelling scheme has been confirmed on the possibility to distinguish the *cis* and *trans* conformations of Pro10 of NT_{8-13}.

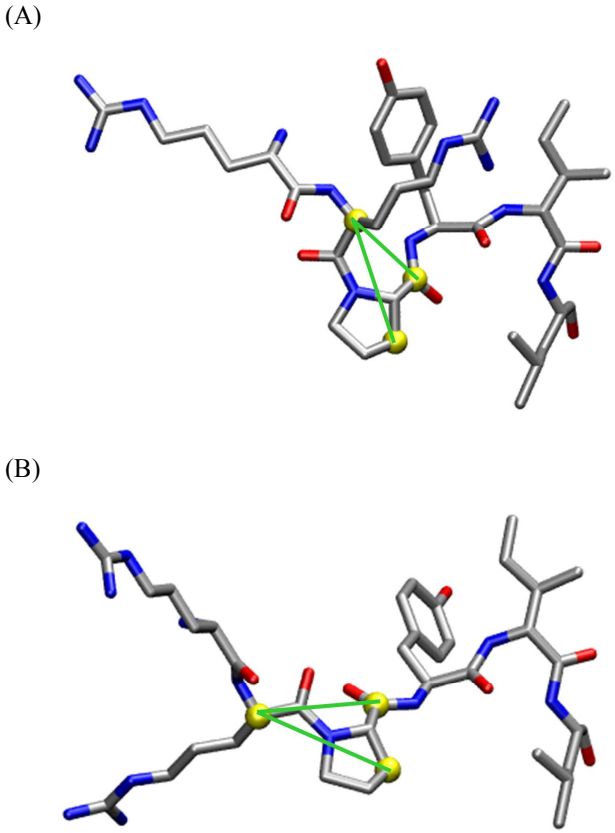


Figure 2.9: Models of *cis*- (A) and *trans*- (B) ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 labelled NT_{8-13} showing the ^{13}C (in yellow spheres) with carbon atoms in grey, nitrogen atoms in blue, and oxygen atoms in red. The green lines show pairs, which will be measured by SSNMR experiments. The models were generated by DeepView - Swiss-Pdb Viewer and rendered in VMD.

Table 2.2: Predicted homonuclear-distance measurement (in Ångstrom) of ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 labelled NT_{8-13}.

No.	Spin pair labelling sites in NT_{8-13} for homonuclear distance measurements	ω_R for RR experiments (Hz)[a]	Predicted inter-spin distances for *cis*/*trans* Pro10 (Å)[b]
1	^{13}Cα Arg9 - ^{13}CO Pro10	14,987	3.2/4.5
2	^{13}Cα Arg9 - ^{13}Cβ Pro10	3,130	4.5/4.8

[a] The spinning frequency differences between spins on a 500 MHz NMR spectrophotometer when the integer (n) = 1.
[b] Predicted distances are measured from the models in **Figure 2.9** in DS ViewerPro 5.0 (Accelrys Inc.).

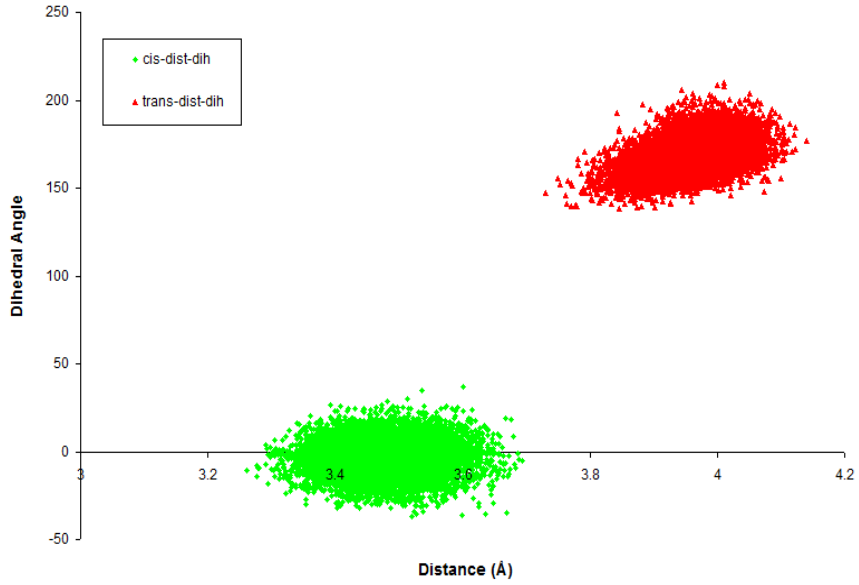
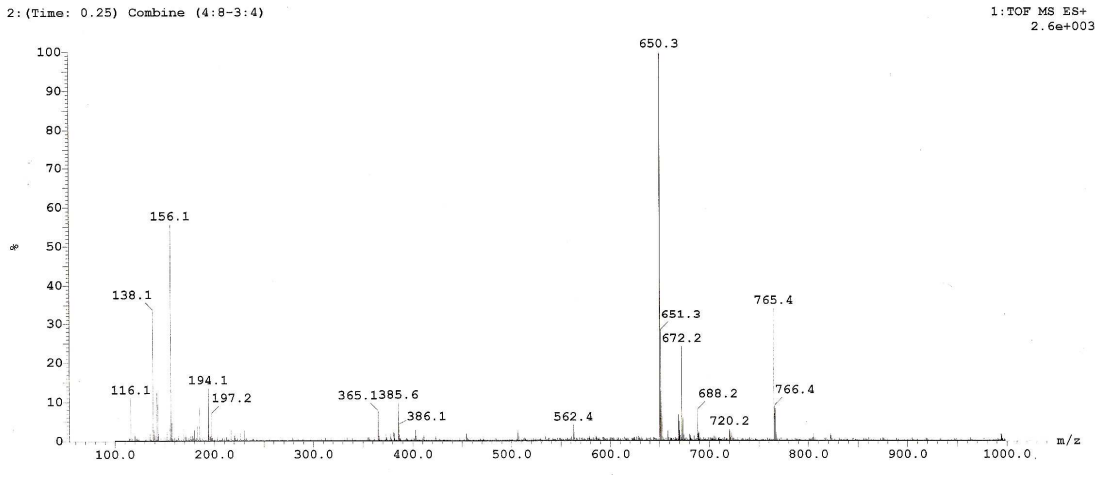


Figure 2.10: Average distances of Cα of Arg9 and Cα of Pro10 (calculated from Cα of Arg9-Cβ of Pro10, Cα of Arg9-CO of Pro10, and Cα of Arg9-N of Pro10 distances) and Arg9-Pro10 dihedral angles to distinguish between *cis* and *trans* Pro10 isomers of NT_{8-13} (GROMACS 3.3).

Synthesis of Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1,3-$^{13}C_2$]-L-proline

The commercially unavailable Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1,3-$^{13}C_2$]-L-proline were synthesised as describe in section 2.2.2. A new strategy was required for the synthesis of both labelled amino acids by using the Williams' oxazinone as the starting step. After the synthesis, the products were confirmed by ESI mass spectrometry. The ESI mass spectra (**Figure 2.11**) show the correct molecular weigh of 650.3 $(M_2\text{-}H)^-$ and 678.5 $(M_2\text{-}H)^-$ for Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1,3-$^{13}C_2$]-L-proline, respectively.

(A)



(B)

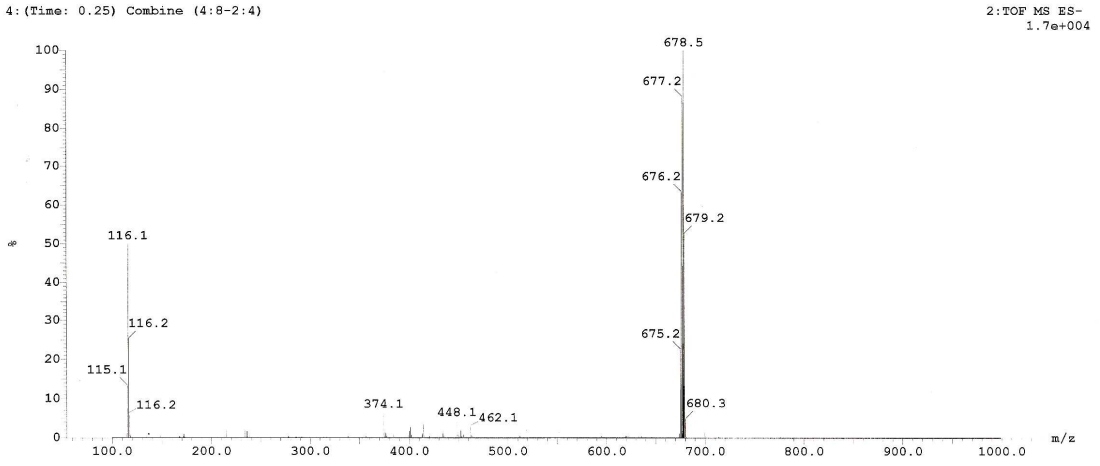


Figure 2.11: Negative-ion mass spectrum of Fmoc-Pbf-[2-^{13}C]-L-arginine. The major peak at *m/z* 650.3 $(M_2\text{-}H)^-$ corresponds to the Fmoc-labelled arginine (A). Negative-ion mass spectrum of Fmoc-[1,3-$^{13}C_2$]-L-proline. The major peak at *m/z* 678.5 $(M_2\text{-}H)^-$ corresponds to the Fmoc-labelled proline (B).

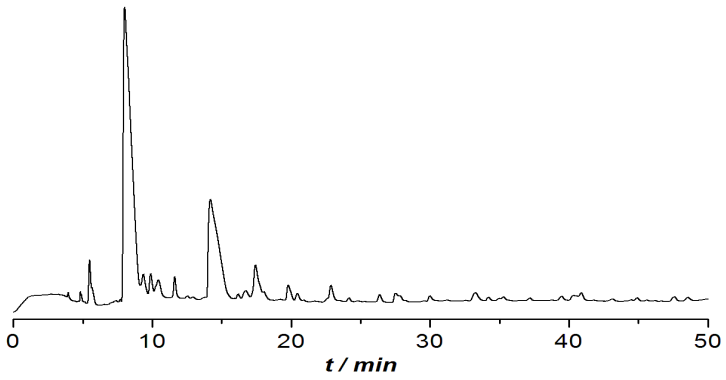
Synthesis of ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 enriched NT_{8-13}

This labelled peptide was designed and synthesised in order to determine the conformation of Pro10 which plays an important role in overall structure of bound NT_{8-13}. Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1,3-$^{13}C_2$]-L-proline, which were specially synthesised for this labelling scheme were used for the production of ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 labelled NT_{8-13}. **Table 2.3** shows the summary of the synthesis of labelling scheme 1. The labelled peptide was purified using a preparative reversed-phase HPLC column (**Figure 2.12A**). The correct molecular weight of the purified peptide was verified by ESI mass spectrometry (**Figure 2.12B**). The purification gave the labelled peptide in 27% yield. The ^{13}C-NMR spectrum of labelled ligand (**Figure 2.13**) was used to authenticate labelled positions.

Table 2.3: Summary of the synthesis of labelled NT_{8-13}. See **Figure 2.12 and 2.17** for details.

Sample	Final amount (mg)	Final yield (%)	Expected mass (Da)	ESI mass observed (Da)
Labelling scheme 1	20.5	27	819.5	820.5, 410.8
Labelling scheme 2	29.0	34	820.5	820.18, 410.79

(A)



(B)

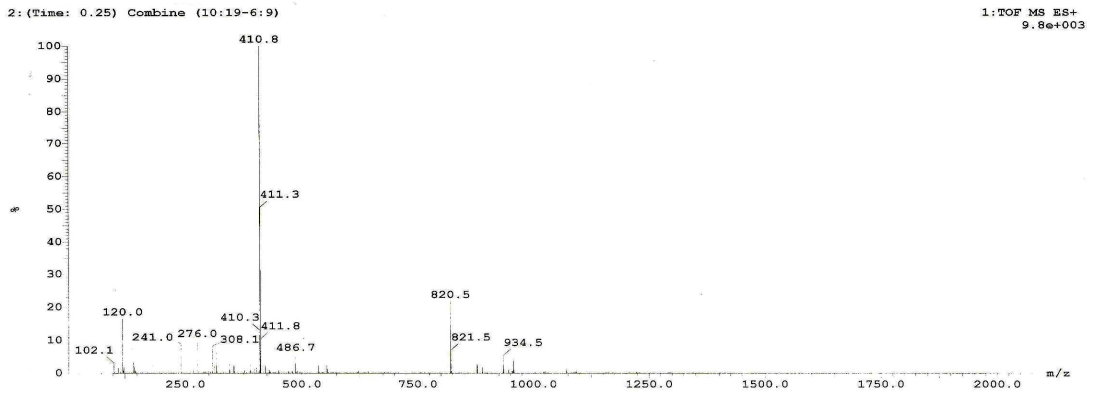


Figure 2.12: Reverse-phase HPLC trace (280 nm) for the ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 labelled NT_{8-13} purification. The crude peptide was purified by solvent A: 0.1% TFA in water; solvent B: 0.1% TFA in acetonitrile; gradient: 20-40% in 45 min. The NT_{8-13} peptide was eluted at 25.4% B (A). Positive-ion mass spectrum of purified ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 labelled NT_{8-13}. The major peaks at *m/z* 410.8 $(M+2H)^{2+}$ and 820.5 $(M+H)^{+}$ correspond to the specific-isotopically labelled NT_{8-13} (B).

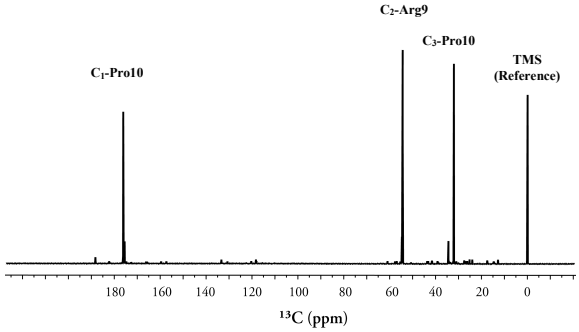


Figure 2.13: 1D ^{13}C-NMR spectrum of labelled ligand: δC_1-Pro10 175.5, 176.2 ppm, δC_2-Arg9 54.5 ppm, δC_3-Pro10 32.1, 34.4 ppm.

ii) Labelling scheme 2

The conformation of the four C-terminal amino acids of the bound hexapeptide is another question of interest. Following the scheme in **Figure 2.2**, based on the availability of labelled amino acids and the requirement of RR and REDOR techniques, a labelling scheme, ^{15}N-Pro10, ^{13}Cα-Tyr11, ^{13}CO-Ile12, and ^{13}Cγ-Leu13 enriched $NT_{8\text{-}13}$, has been proposed to address this question as shown in **Figure 2.14** and the details of predicted distances are shown in **Table 2.4**. Actually, CO of either Pro10 or Tyr11 would be more appropriate for REDOR since it has long T_2, yet the distance between NH of Pro10 and CO of Pro10 does not give any useful information since the angles between them are fixed, and the potential distance (6.08 Å) between NH of Pro10 and CO of Tyr11 exceeds the limitation of REDOR ^{13}C-^{15}N distance measurements (see **Table 1.1** in section 1.6.2).

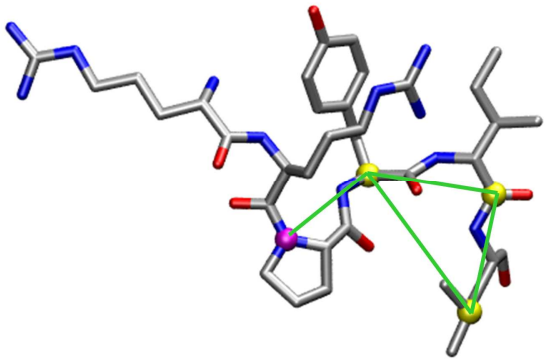


Figure 2.14: Model of the ^{15}N-Pro10, ^{13}C-Tyr11, ^{13}C-Ile12, and ^{13}C-Leu13 labelled $NT_{8\text{-}13}$ showing the ^{13}C (yellow spheres) and the ^{15}N (magenta sphere), with carbon atoms in grey, nitrogen atoms in blue, and oxygen atoms in red. The green lines show pairs, which will be measured by SSNMR experiments. The model was generated by DeepView - Swiss-Pdb Viewer and rendered in VMD.

Table 2.4: Predicted distance measurement (in Ångstrom) of ^{15}N-Pro10, ^{13}Cα-Tyr11, ^{13}CO-Ile12, and ^{13}Cγ-Leu13 labelled $NT_{8\text{-}13}$.

Spin pair labelling sites in $NT_{8\text{-}13}$ for internuclear distance measurements	ω_R for RR experiments (Hz)[a]	Predicted inter-spin distances (Å)[b]
^{13}Cγ Leu13 - ^{13}CO Ile12	18,682	4.46
^{13}Cγ Leu13 - ^{13}Cα Tyr11	3,840	5.97
^{13}CO Ile12 - ^{13}Cα Tyr11	14,782	4.52
^{13}Cα Tyr11 - ^{15}N Pro10	-	5.07

[a] The spinning frequency differences between spins on a 500 MHz NMR spectrophotometer when the integer (n) = 1.
[b] The RR and REDOR techniques will be used to recouple homonuclear and heteronuclear dipolar interactions, respectively. Potential distances are measured from the model in **Figure 2.14** in DS ViewerPro 5.0 (Accelrys Inc.).

The Fmoc protection of labelled amino acids

Since from a commercial provider, N-termini of labelled amino acids are usually not protected, here manual Fmoc protection of labelled amino acids must be performed for Fmoc-SPPS method. Following the synthesis procedure of Fmoc-amino acids as described in section 2.2.3, the nature and purity of products were examined by TLC and compared with standard unlabelled compounds. The R_f values of Fmoc-[^{15}N]-L-proline, Fmoc-[2-^{13}C]-L-tyrosine, Fmoc-[1-^{13}C]-L-isoleucine, and Fmoc-[4-^{13}C]-L-leucine were the same as those of the standard Fmoc-amino acids under the same condition (as shown in **Table 2.5**). The correct molecular weight of all Fmoc-labelled amino acids was confirmed by ESI mass spectrometry (**Figure 2.15**).

Table 2.5: The R_f values (calculated from TLC) of all Fmoc-labelled amino acids. The R_f value of each Fmoc-labelled amino acid matches to the standard commercial compounds. See **Figure 2.15** for details.

Fmoc-labelled amino acids	Yield (%)	R_f values	Expected mass (Da)	ESI mass observed (Da)
^{15}N Proline	91	0.70	338.40	675.41, 337.22
$^{13}C_2$ Tyrosine	75	0.63	404.40	807.38, 403.39
$^{13}C_1$ Isoleucine	92	0.63	354.40	707.47, 353.30
$^{13}C_4$ Leucine	90	0.79	354.40	707.45, 353.25

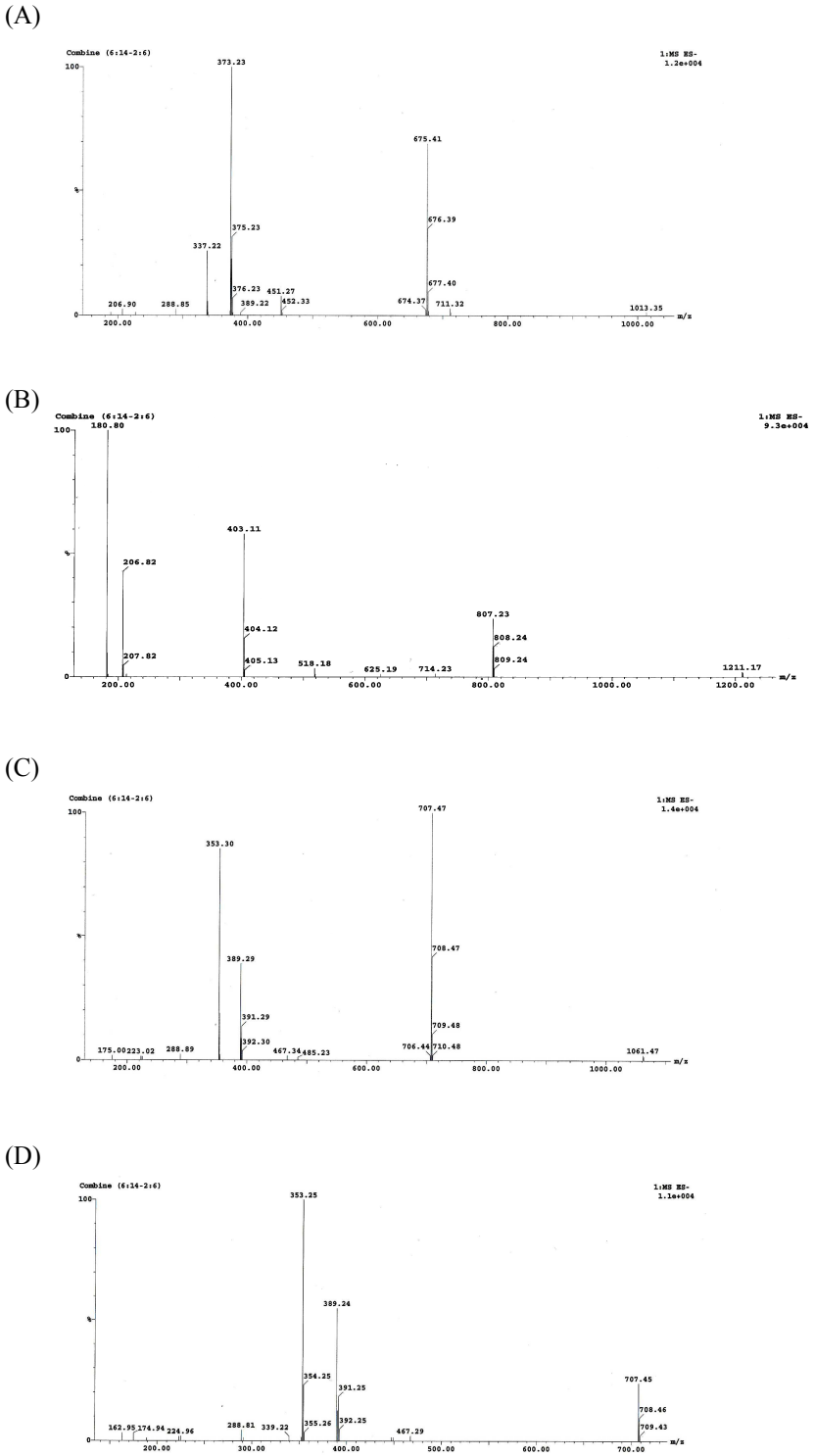
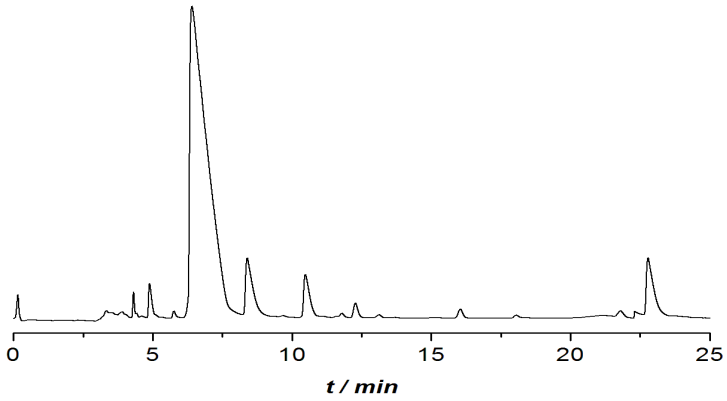


Figure 2.15: Negative-ion mass spectra of Fmoc-labelled amino acids. Two major peaks at *m/z* 337.22 $(M\text{-}H)^-$ and 675.41 $(M_2\text{-}H)^-$ correspond to the Fmoc-[1-^{15}N]-L-proline (A). Two major peaks at *m/z* 403.39 $(M\text{-}H)^-$ and 807.38 $(M_2\text{-}H)^-$ correspond to the Fmoc-[2-^{13}C]-L-tyrosine (B). Two major peaks at *m/z* 353.30 $(M\text{-}2H)^-$ and 707.47 $(M_2\text{-}H)^-$ correspond to the Fmoc-[1-^{13}C]-L-isoleucine (C). The major peaks at *m/z* 353.25 $(M\text{-}H)^-$ and 707.45 $(M_2\text{-}H)^-$ correspond to the Fmoc-[4-^{13}C]-L-leucine (D).

Synthesis of ^{15}N-Pro10, ^{13}Cα-Tyr11, ^{13}CO-Ile12, and ^{13}Cγ-Leu13 enriched NT_{8-13}

^{15}N-Pro10, ^{13}Cα-Tyr11, ^{13}CO-Ile12, and ^{13}Cγ-Leu13 labelled NT_{8-13} was designed based on the commercial availability of labelled amino acids to determine the conformation of the last four residues of NT_{8-13}. For Fmoc-SPPS method, not only the N-terminus but also the side-chain functional group of amino acids are protected to avoid side-reactions occurred during the peptide synthesis. Tyrosine contains a reactive hydroxyl group which is usually protected with *tert*-butyl group (*tBu*). Doubly protection of an amino acid is time consuming and results in a yield loss. To evaluate the need for side-chain protection for Tyr, Fmoc-Tyr-OH and Fmoc-Tyr(*tBu*)-OH were used in the synthesis of NT_{8-13} and compared. The results revealed that the doubly protection of Tyr is not necessary for the synthesis of a short peptide, NT_{8-13} as the yield was not significantly decreased (14% decrease) by using singly protection of tyrosine. **Table 2.3** shows the summary of the synthesis of labelling scheme 2. The labelled NT_{8-13} was purified by a preparative reversed-phase HPLC column (**Figure 2.16A**). The correct molecular weight of purified peptide was verified by ESI mass spectrometry (**Figure 2.16B**). The purification gave the labelled peptide in 34% yield. The ^{13}C- and ^{15}N-NMR spectra of labelled ligand (**Figure 2.17**) were used to confirm labelled positions.

(A)



(B)

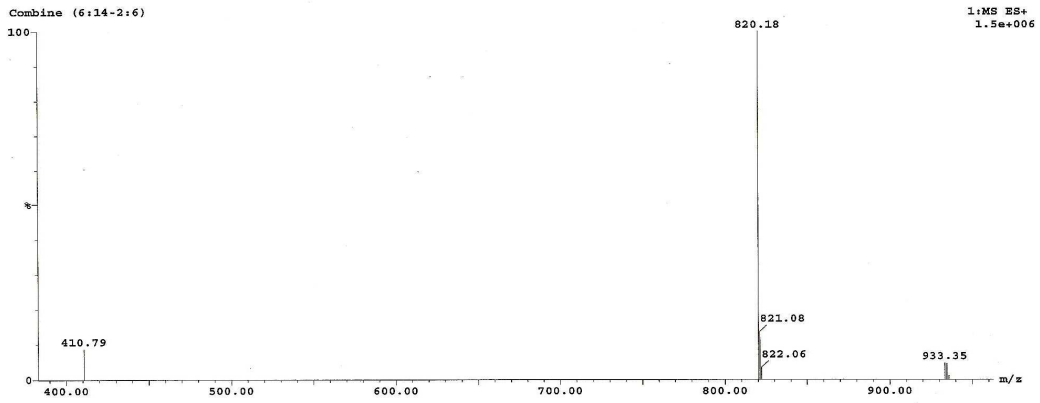


Figure 2.16: Reverse-phase HPLC trace (280 nm) for the ^{15}N-Pro, ^{13}Cα-Tyr, ^{13}CO-Ile, and ^{13}Cγ-Leu labelled NT_{8-13} purification. The crude peptide was purified by solvent A: 0.1% TFA in water; solvent B: 0.1% TFA in acetonitrile; gradient: 20-50% in 23 min. The NT_{8-13} peptide was eluted at 25.7% B (A). Positive-ion mass spectrum of purified ^{15}N-Pro, ^{13}Cα-Tyr, ^{13}CO-Ile, and ^{13}Cγ-Leu labelled NT_{8-13}. The major peaks at *m/z* 410.79 $(M+2H)^{2+}$ and 820.18 $(M+H)^{+}$ correspond to the specific-isotopically labelled NT_{8-13} (B).

(A) (B)

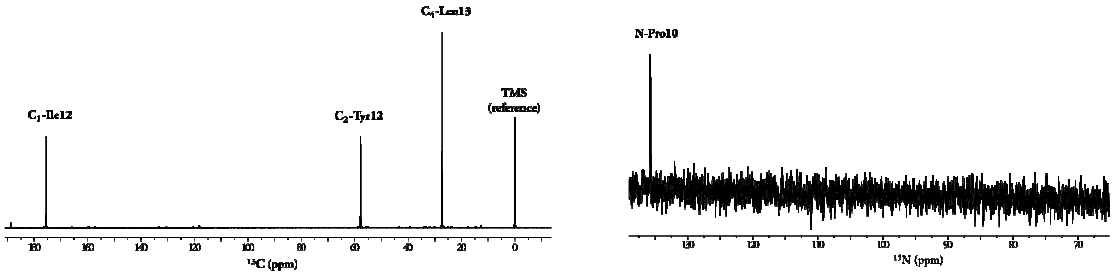


Figure 2.17: 1D ^{13}C- (A) and ^{15}N- (B) NMR spectra of labelled ligand: δC_1-Ile12 175.3 ppm, δC_2-Tyr11 57.98 ppm, δC_4-Leu13 27.5 ppm, and δN-Pro10 135.7 ppm.

2.4 Conclusions

Unlabelled peptides were synthesised and will be used for crystallisation trials, NMR experiments and other applications. ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 enriched NT_{8-13} and ^{15}N-Pro10, ^{13}Cα-Tyr11, ^{13}CO-Ile12, and ^{13}Cγ-Leu13 enriched NT_{8-13} were designed to facilitate the determination of the conformation of the ligand bound to the purified receptor by accurate (to < 0.5 Å precision) distance determinations evaluated from dipolar couplings between site-directed coupled ^{13}C-spins (over < 7 Å range) and ^{13}C-^{15}N spins (over < 6 Å range) using homonuclear recoupling RR-SSNMR and heteronuclear recoupling REDOR-SSNMR methods, respectively (Karlsson and Levitt, 1998; Watts et al., 2004). The labelling scheme of ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 labelled NT_{8-13} has been designed through modelling approaches to permit two pairs of inter-spin distances to be measured. This resolves whether Pro10, which plays a crucial role of the conformation of bound ligand, is in a *cis* or *trans* conformation. The labelled peptide was synthesised with a good yield of 27% recovery. The labelling scheme of ^{15}N-Pro10, ^{13}Cα-Tyr11, ^{13}CO-Ile12, and ^{13}Cγ-Leu13 labelled NT_{8-13} has been designed to determine the structure of the four C-terminal residues of NT in the bound form. 34% recovery of labelled peptide was obtained from manual Fmoc-SPPS method. The labelled positions of each labelling scheme were confirmed by 1D ^{13}C- and ^{15}N- solution NMR experiments.

Chapter 3

Expression, isotope labelling, purification, and characterisation of neurotensin

3.1 Introduction

A uniformly labelled peptide is needed for backbone conformation study, and the entire conformation would require a number of labelled peptides, which ^{13}C or ^{15}N is introduced at the relevant sites. In the production of isotopically labelled peptides, an economically viable scheme is an important factor. Commonly, uniformly labelled peptides are produced by expression using a bacterial recombinant system or solid phase peptide synthesis (SPPS) (Lindhout et al., 2003; Luca et al., 2003). For SPPS, individual labelled amino acids are very expensive, commercially available at a minimum amount of ~100 mg and require N-terminal protection with either *t*-Boc or Fmoc for peptide synthesis. For NT_{8-13} (6-mer peptide) and using no excess of N-terminal protected labelled-amino acids for the SPPS method (see section 2.3.2), the yield was 20-30 mg of purified labelled peptide at the scale of 0.1 mmol reaction-the longer sequence, the lower yield. The length of GPCR activating ligands can vary from 6 to 52 residues. It is therefore more practical to synthesise small peptides but not long sequence polypeptides. For example, the cost of peptide synthesis of U-^{13}C, ^{15}N labelled NT, which is a 13-residue peptide would be at least £5,000 at the scale of 0.1 mmol reaction, whereas using bacterial expression methods would cost less than

£1,000 for the same amount of labelled NT. Actually, solid-state NMR experiments do not require such a high amount of purified labelled peptide. An amount of 5 μmol of labelled peptide (up to £200 using expression methods) would be suitable for NMR studies. Therefore, expression can potentially yield high levels of peptide at relatively low cost, with easy incorporation of isotopic labels. Also, using the expression methods could reduce the use of hazardous chemicals needed in the SPPS method.

3.1.1 Expression of small peptides in *Escherichia coli*

The use of a fusion protein tag can aid peptide stability and solubility following expression (Waugh, 2005). Fusion proteins incorporating various affinity tags and proteolytic cleavage sites can assist in the purification and subsequent peptide fragment production. The expression of a target peptide as inclusion bodies has been sometimes more preferable than as soluble forms (Majerle et al., 2000; Mukhopadhyay, 1996; Suzuki et al., 1998). The expression in inclusion bodies is good for mass production, for example peptide hormones for therapeutic uses since it starts with high purity of peptide before purification. In addition, some peptide hormones are toxic to host cells, which is avoided if expressed peptides as inclusion bodies (Jones et al., 2000). However, peptide recovery can be limited by the methods of solubilisation and refolding. There are several choices of cytosolic soluble protein partners for production of a peptide in *Escherichia coli* (*E. coli*). **Table 3.1** summarises published fusion strategies used for recombinant production of small peptides (< 50 amino acids in length) in *E. coli* in a soluble form. Glutathione S-transferase (GST) (Smith and Johnson, 1988), immunoglobulin-binding domain of streptococcal protein G (GB1 domain) (Lindhout et al., 2003) and thioredoxin (LaVallie et al., 1993) are ones of the

most popular protein partners to enhance the expression levels, solubility and stability of the target peptides. Recovery of the fusion product from the bulk of other bacterial proteins can be done quickly with up to 95% purity by adding polyhistidine tags at the N- or C-terminus of the target peptides (Hochuli et al., 1988). To make a construct, it is frequently difficult to choose the best fusion protein because different protein tags can lead to substantially different expression levels and solubility of the target peptides, and the expression of fusion proteins can be sequence-dependent (Koenig et al., 2003; Waugh, 2005). For instance, the final yield of S-peptide (20 residues) from the construction using GB1 as a fusion protein was about tenfold higher than that obtained using 6 kDa polypeptide plus His-tag (Koenig et al., 2003). Moreover, recombinant production of the GB1-CD4 peptide fusion protein was approximately 11-fold lower than that of the GB1-CMPcc peptide fusion protein (Koenig et al., 2003). Therefore, a significant amount of experimentation is required to find the correct combination of solubilisation tags, affinity tags and expression conditions.

Table 3.1: The summary of the recombinant production of short peptides (< 50 residues) as fusion proteins in *E. coli* in soluble form.

Peptide (# of residues)	Fusion tags	Enzyme or chemical cleavage	Peptide yield (mg L^{-1})		Reference
			Luria-Bertani media	Minimal media	
ACTH(1–24) peptide (24)	*E. coli* thioredoxin + His-tag + linker	Enterokinase	10	6	(Uegaki et al., 1996)
PAK pilin peptide[a] (12)	*E. coli* OmpA + *de novo* polypeptide	CNBr		~2	(Campbell et al., 1997)
Bradykinin (9)	Glutathione S-transferase + linker	Plasma kallikrein		~0.3	(Ottleben et al., 1997)
S-peptide[b] (20)	Polypeptide + His-tag	Thrombin		1.5	(Alexandrescu et al., 1998)
C-peptide[c] (31)	HSA binding region of Streptococcal protein G	Trypsin and Carboxy peptidase B	23		(Jonasson et al., 1998)
Neurotensin (13)	Glutathione S-transferase + linker	CNBr	1.8		(Williamson et al., 2000)
S2 peptide (11)	GB1 domain of Streptococcal protein G + linker	Factor Xa		15	(Koenig et al., 2003)
CMPcc (47)	GB1 domain of Streptococcal protein G + linker	Factor Xa		48	(Koenig et al., 2003)
GLP-1 peptide (30)	CMFH domain + His tag	CNBr	21.5		(Li et al., 2006)
CUE domain[b] (43)	GB1 domain of Streptococcal protein G + His-tag	Thrombin	~18		(Bao et al., 2006)

[a]C-terminal homoserine (lactone) as a result of CNBr cleavage.
[b]Residues Gly-Ser are retained at the N-terminus of the target peptide based on thrombin specificity.
[c]Tandem repeats of the peptide linked to the fusion partner.

3.1.2 Aims

This chapter aims to produce and characterise ^{13}C, ^{15}N labelled NT and NT_{8-13} for subsequent NMR studies. ^{2}H, ^{13}C, ^{15}N labelled NT will be also produced because it has been planned that the interactions between neurotensin and neurotensin receptor 1 can be probed by solution-state NMR techniques. Here, the expression of neurotensin in a bacterial system using the pGEV-1 vector has been proposed for rapid manipulation of labelled sites and economic isotope labelling. Optimisation of the expression of GB1-NT fusion protein in *E. coli* is attempted to improve the yield, which will be useful for further large scale production if needed. Then, the smaller peptide, NT_{8-13}, will be expressed in *E. coli* by introducing the Met cyanogen bromide (CNBr) cleavage site at Pro7 of NT in the pGEV-NT construct.

Uniformly labelled peptides can be used to provide backbone conformations and some side-chain interactions from correlation crosspeaks in NMR spectra. Then, these interactions can be confirmed by distance measurements using either uniformly or selectively isotope-labelled peptides. To determine whether the seven N-terminal amino acid of NT peptide is important for receptor binding, uniformly labelled NT and NT_{8-13} when bound to the receptor can be used in comparative studies.

3.2 Materials and methods

3.2.1 Construction of pGEV-NT

Oligonucleotides (MWG-Biotech) coding for 8xHis for affinity purification, followed by a CNBr cleavage site (Met) and the NT sequence (**Figure 3.1**) were synthesised using optimal codons for *E. coli* (http://www.kazusa.or.jp/codon). The oligonucleotides were designed to provide an overhanging *NheI* site on the 5' end of

the sense strand, and an overhanging *XhoI* site on the 5' end of the antisense strand for cloning into the pGEV-1 vector (Huth et al., 1997). Sense and antisense oligonucleotides were annealed (95 °C; 1 h), followed by cooling (65 °C; 30 min) prior to ligation to the vector. The plasmids were transformed into *E. coli* XL1 Blue (Stratagene) and plated onto LB-agar plus 100 µg/mL ampicillin. Presence of the correct insertion was verified by DNA sequencing. For expression, plasmids were transformed into *E. coli* BL21(DE3) (Novagen).

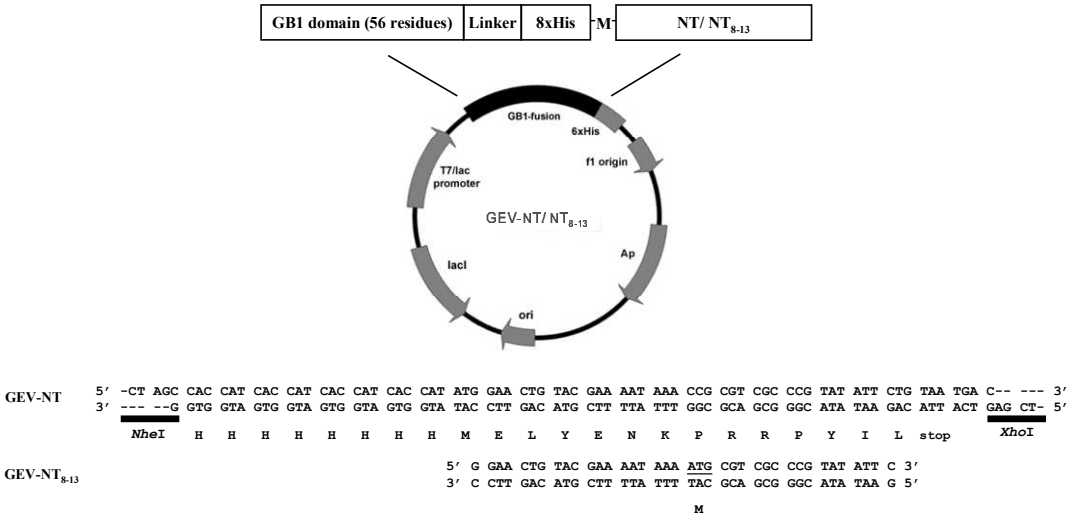


Figure 3.1: Scheme of GEV-NT and GEV-NT_{8-13} (P7M) plasmids. A pGEV-1 plasmid map is shown with a schematic of the insertion. Oligonucleotide sequences used for cloning the full length peptide (NT) and site-directed mutagenesis to give pGEV-NT_{8-13} are given. In the pGEV-1 plasmid, a noncleavable linker and 6xHis tag is encoded (plasmid map). By the inclusion of stop codons downstream of the NT coding sequence, this tag is not included in the translated product. A sequence encoding 8xHis-Met-NT was inserted into this plasmid using *Nhe*I and *Xho*I cloning sites. The inclusion of Met generates a CNBr-cleavable site. By placing stop codons between the C-terminal of NT and the vector-encoded 6xHis tag, a free C-terminus can be generated (introducing a Met site at the 3′ end of an insert is not a viable strategy, as it would yield a C-terminal homoserine upon CNBr cleavage of the product). The GEV-NT_{8-13} vector was constructed by using pGEV-NT as a template for a P7M substitution using site-directed mutagenesis. Underlined letters indicate mutated nucleotide residues.

3.2.2 Construction of pGEV-NT_{8-13}

The recombinant plasmid pGEV-NT was used as a template for the generation of Pro7 to Met (P7M) mutation. The mutant plasmid was designed to provide another CNBr cleavage site to obtain the NT_{8-13} peptide. A pair of complementary mutagenic oligonucleotide primers (MWG-Biotech) (**Figure 3.1**) was synthesised. The mutant was generated by polymerase chain reaction (PCR) using high-fidelity *Pfu* DNA polymerase, using the QuickChange mutagenesis kit (Stratagene) (Appendix A.1). The mutant clones were verified by DNA sequencing.

3.2.3 Optimisation of pGEV-NT expression

All optimisation experiments were conducted in 100 mL conical flasks containing 20 mL LB media plus 100 μg/mL ampicillin. Uninduced cell culture was used as a negative control. *E. coli* cultures after induction (1 OD_{600} ~10^8 cells) were collected by centrifugation at 5000 xg for 10 min. Comparisons of expression levels among different conditions were analysed by loading the same amount of cells on each well of NuPAGE 12% Bis-Tris gel (Biorad).

Screening of OD_{600} of cell culture before induction

For determining the optimum cell density for expression of the fusion peptide, a LB starter culture (2 mL) containing 100 μg/mL ampicillin was inoculated with a single colony of pGEV-NT-transformed BL21(DE3) and grown at 37 °C overnight with 200 rpm shaking. LB media (20 mL, 100 μg/mL ampicillin) in 100 mL conical flasks was inoculated with 0.2 mL starter culture and incubated at 37 °C with 200 rpm

shaking until the cells reached OD_{600} of 0.5, 0.8, 1.0, and 1.2. The expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h.

Screening of concentration of IPTG

The expression of NT peptide was examined in various concentrations of IPTG. The *E. coli* cultures incubated until OD_{600} = 0.8 at 37 °C were induced with IPTG at final concentrations of 0.25, 0.5, 1.0, and 2.0 mM for 4 h.

Screening of IPTG induction periods

The *E. coli* cultures incubated until OD_{600} = 0.8 at 37 °C were induced with 1 mM IPTG for 2, 4, 6, 8, and 10 h.

Screening of induction temperature

The *E. coli* cultures incubated until OD_{600} = 0.8 at 37 °C were induced at 30 °C with 1 mM IPTG for 4, 6, 8, and 10 h.

3.2.4 Expression of GB1 fusion proteins

E. coli cells strain BL21(DE3) harbouring the pGEV-NT and pGEV-NT_{8-13} plasmids were grown at 37 °C in 15 mL M9 minimal media plus 100 µg/mL ampicillin overnight with 200 rpm shaking. Half a litre of M9 minimal media plus yeast nitrogenous base (YNB) in 2 L baffled conical flask was inoculated with 15 mL of overnight culture of the fusion protein construct and incubated at 37 °C with 200 rpm shaking until the cells reached an OD_{600} of 0.8. The cells were then induced with

1 mM IPTG for 4 h. Cells were harvested by centrifugation (6000 xg; 20 min; 4 °C) and the cell pellet was frozen at -20 °C.

3.2.5 Expression and ^{13}C, ^{15}N isotope labelling of GB1 fusion proteins

U-^{15}N labelled NT, U-^{13}C, ^{15}N labelled NT and U-^{13}C, ^{15}N labelled NT_{8-13} were prepared in M9 minimal media plus YNB supplemented with $^{15}NH_4Cl$ and/or [$^{13}C_6$] glucose (Spectra) as the sole nitrogen and carbon sources, respectively. Cells containing each plasmid were grown in 30 mL of media at 37 °C overnight with 200 rpm shaking. One-litre cultures of media were inoculated with 30 mL of overnight culture of the fusion protein construct, grown at 37 °C with shaking until the cells reached an OD_{600} of 0.8, and induced with 1 mM IPTG for 4 h. Cells were harvested by centrifugation (6000 xg; 20 min; 4 °C).

3.2.6 ^{2}H, ^{13}C, ^{15}N isotope labelling of GB1-NT fusion proteins

The expression of deuterated fusion proteins was performed in collaboration with the ILL-EMBL Deuteration Laboratory at the Partnership for Structural Biology in Grenoble, France. To produce the deuterated ^{13}C, ^{15}N GB1-NT, the growth of BL21(DE3) cells containing the pGEV-NT plasmid in deuterated minimal media was optimised. The ampicillin resistance of the plasmid pGEV-NT was changed by transposition to the kanamycin resistance which allows efficient selection of the expression construct under high-cell-density conditions in deuterated minimal media. For ^{2}H, ^{13}C, ^{15}N isotope labelling of GB1-NT, the carbon source (D-glucose [U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-D7, 97-98%]) and the nitrogen source ($^{15}NH_4Cl$) were from Cambridge Isotope Laboratories, Inc. Triple labelled minimum media (200 mL) was

prepared with 1 g of D-glucose for the batch phase of the fermentation. Feeding solution was made by dissolving 5 g of D-glucose in 50 mL of deuterated water.

The batch solution (20 mL) was inoculated with D_2O adapted BL21(DE3) cells containing the pGEV-NT plasmid. This pre-culture was grown overnight at 37 °C in shacking incubator and used to inoculate the fermentor (180 mL). The fermentor was grown in the batch phase for 42 h until OD_{600} ~4.2. The fed batch phase occurred for 17 h with an exponential feeding. When the OD_{600} reached 8.6, the induction phase (15 h, 30 °C) was followed by addition of 1 mM IPTG in 2 mL of D_2O. The fermentation was stopped with an OD_{600} of 7.4. The expression of GB1-NT was checked on SDS-PAGE. Cell paste was collected by centrifugation (6000 xg; 20 min; 4 °C) and frozen at -80 °C.

3.2.7 Purification of GB1 fusion proteins

The frozen cell pellets were thawed in 15-20 mL of cold solubilisation buffer containing 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 5 mM imidazole and 1 mM phenylmethanesulfonyl fluoride (PMSF). The cell suspension was disrupted by using a French Press Cell 3 times at 10,000 psi. To determine the extent of soluble expression, 100 µL of crude lysate was pellet by centrifugation (8000 xg; 15 min; 4 °C). The insoluble fraction was resuspended in the buffer, equal volume to that of the soluble fraction. The total crude lysate, soluble fraction and insoluble fraction were analysed on NuPAGE 12% Bis-Tris gel (Biorad).

Purification was carried out with crude lysate using high speed centrifugation (70000 xg; 90 min; 4 °C). The supernatant fraction was passed through 0.40 and then 0.22 µm filters to remove all insoluble cellular debris. The supernatant was loaded

onto a 5-mL His-Trap column, previously charged with 50 mM $NiSO_4$, and equilibrated in 5 mM imidazole, 500 mM NaCl, and 20 mM Tris-HCl (pH 8.0). The column was washed with 40 mM imidazole, 500 mM NaCl, and 10 mM Tris-HCl (pH 8.0) and the GB1 fusion proteins were eluted with 500 mM imidazole, 500 mM NaCl, and 20 mM Tris-HCl (pH 8.0). The fraction containing the target fusion protein was lyophilised, resuspended in water, desalted using two 5-mL Sephadex G25 superfine columns (10 mM NH_4HCO_3 at pH 8.0 for equilibrium and buffer exchange), and lyophilised.

3.2.8 Cleavage of GB1 fusion proteins and purification of peptides

Neurotensin peptides were cleaved from the GB1 fusion proteins by CNBr. The lyophillised proteins were dissolved in 0.1 M HCl at the final concentration of 5 mg/mL. Solid CNBr was then added to a final molar ratio of 100:1 (peptide:CNBr). The sample was sealed under a nitrogen atmosphere and incubated at room temperature in the dark for 24 h. The reaction was terminated by diluting the sample 10-fold with MilliQ water and lyophilised. Samples were subjected to a preparative C18 reverse-phase HPLC column with a linear gradient of acetonitrile from 20 to 40% (*v/v*) over 45 min. Individual peaks were collected and verified by ESI mass spectrometry. Peaks containing pure peptide were pooled and twice lyophilised to remove all organic solvents. To check purity, analytical HPLC chromatography was performed.

3.2.9 NMR characterisation

^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC experiments were performed on a Bruker 600 MHz NMR spectrometer in NMR facilities of Prof. Iain Campbell (IDC group, Oxford). Approximately 0.1 mg of each labelled peptide was dissolved in 0.60 mL of H_2O containing 5% (*v/v*) D_2O and transferred into a 5-mm NMR tube. HSQC data were acquired with 128 t_1 increments using States-TPPI phase cycling. The recorded data were referenced to an external DSS standard at 298 K. Data in t_1 was linear predicted to 256 points and data points in both dimensions were zero-filled to 1024 for NMR data processing with nmrPipe (Delaglio et al., 1995). The spectra were analysed with Sparky (http://www.cgl.ucsf.edu/home/sparky/).

3.3 Results and Discussion

3.3.1 Preparation of pGEV-NT and pGEV-NT$_{8-13}$ expression vectors

To produce isotopically-labelled NT, using solid phase peptide synthesis would be very expensive. Incorporation of isotopic labels (^{13}C, ^{15}N, ^{2}H) within a peptide using a bacterial recombinant system is an alternative to the chemical synthetic approach. Here, we propose the expression of neurotensin in a bacterial system using the pGEV-1 vector (Huth et al., 1997) encoding the GB1 domain of streptococcal protein G, target peptide and 6xHis. The GB1 domain is small, has a stable fold and high bacterial expression capability and therefore has been widely used as a fusion tag to produce several peptides as soluble proteins at high yield (Gronenborn and Clore, 1996; Koenig et al., 2003; Lindhout et al., 2003). 8xHis-Met-NT was fused inframe with the GB1 domain by the ligation of a synthetic double-stranded oligonucleotide into the *NheI/XhoI* cloning site (**Figure 3.1**). To facilitate removal of tags using CNBr

cleavage, a methionine was introduced N-terminal to the peptide. The Met residue is converted to homoserine as the consequence of CNBr cleavage. Placing stop codons at the 3′ end of NT and introducing the 8xHis tag at the 5′ end of NT instead of using the vector-encoded C-terminal 6xHis tag for affinity purification resulted in a free C-terminus of NT, required for ligand binding (Barroso et al., 2000; Lazarus et al., 1977). The resulting plasmid encodes the fusion protein, GB1-8xHis-Met-NT.

To enable the expression of the 8-13 active sub-fragment of NT in *E. coli*, by extension of the strategy described above, the mutant plasmid (pGEV-NT) was generated by PCR-based directed mutagenesis using the resulting pGEV-NT plasmid as a template and mutagenic primers containing mutated nucleotides. The replacement of Pro7 of NT with Met provides a CNBr cleavage site to yield the pGEV-NT_{8-13} expression construct (**Figure 3.1**).

The presence of the correct insertion for pGEV-NT plasmid and the mutated positions for pGEV-NT_{8-13} plasmid were verified by DNA sequencing. Both plasmids were transformed into *E. coli* BL21(DE3) for expression.

3.3.2 Optimisation of GB1-NT expression

Screening of optimal OD_{600} for induction

The cell cultures were grown at 37 °C until the OD_{600} reached 0.5, 0.8, 1.0, and 1.2. The cultures were induced with 1 mM IPTG for 4 h at 37 °C. Analysis of 0.25 OD_{600} of crude extracts by SDS-PAGE shows induced protein bands migrating between the 6.5- and 14.3-kDa markers, agreeing with the predicted molecular weight of the fusion proteins of 9.2 kDa. The optimal point for induction was at an OD_{600} of 0.8, giving the highest expression level (**Figure 3.2**).

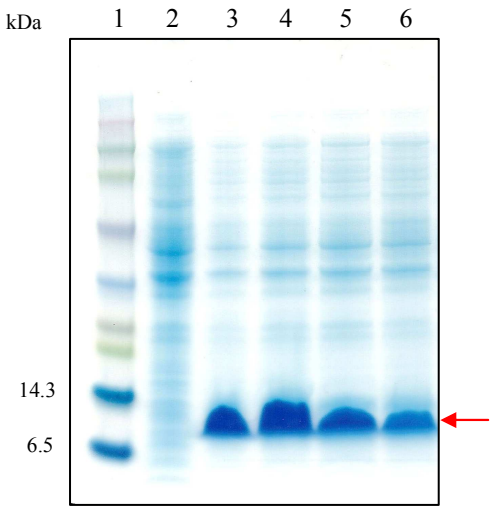


Figure 3.2: SDS-PAGE (Coomassie blue stained 12% NuPAGE gel) analysis of the expression of the GB1-NT fusion protein from *E. coli* transformant at different OD_{600} for 4 h induction by 1 mM IPTG at 37 °C. Lane 1, molecular weight markers 7702S (New England BioLabs); lane 2, supernatant of the cell lysate from non-induced cells; lane 3 to 6, supernatant of the cell lysate from OD_{600} of 0.5, 0.8, 1.0, and 1.2, respectively. The GB1-NT fusion protein migrates as a 9.2-kDa band indicated by a red arrow.

Screening of concentrations of IPTG

The cell cultures were induced by four different IPTG concentrations when they reached an OD_{600} of 0.8. Equal amounts (0.1 OD_{600}) of the crude extracts were analysed by SDS-PAGE. The expression levels of 9.2-kDa GB1-NT at 0.25, 0.5, 1.0, and 2.0 mM of IPTG are similar (**Figure 3.3**). Therefore, the IPTG concentration of 1.0 mM was chosen for further experiments.

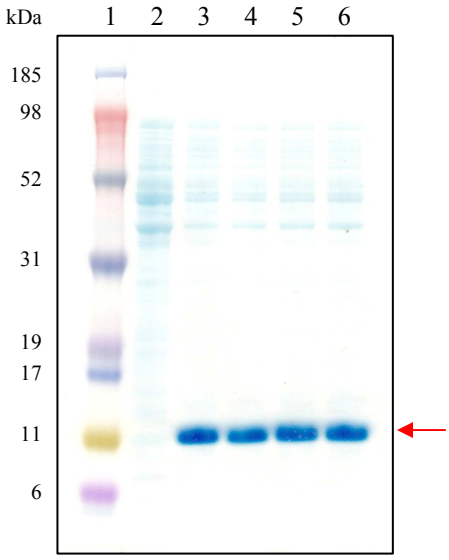


Figure 3.3: SDS-PAGE (Coomassie blue stained 12% NuPAGE gel) analysis of the GB1-NT fusion protein derived from pGEV-NT-harbouring cells with an OD_{600} of 0.8, induced with different IPTG concentrations for 4 h at 37 °C. Lane 1, molecular weight markers MultiMark (Invitrogen); lane 2, supernatant of the cell lysate from non-induced cells; lane 3 to 6, supernatant of the cell lysate from 0.25, 0.5, 1.0, and 2.0 mM IPTG, respectively. A red arrow indicates the expression of the 9.2-kDa GB1-NT fusion protein.

Screening of IPTG induction periods

The *E. coli* culture incubated until an OD_{600} of 0.8 was induced with 1 mM IPTG for 2, 4, 6, 8, and 10 h. Equal amount (0.1 OD_{600}) of crude extracts was analysed on SDS-PAGE. As can be seen in **Figure 3.4**, 4 h and 6 h induction time yield the same expression level of the 9.2-kDa GB1-NT fusion protein, higher than 2 h induction time. The decreased cell viability probably makes the expression level of the fusion protein low for 8 and 10 induction periods. Induction time of 4 h was chosen for further experiments.

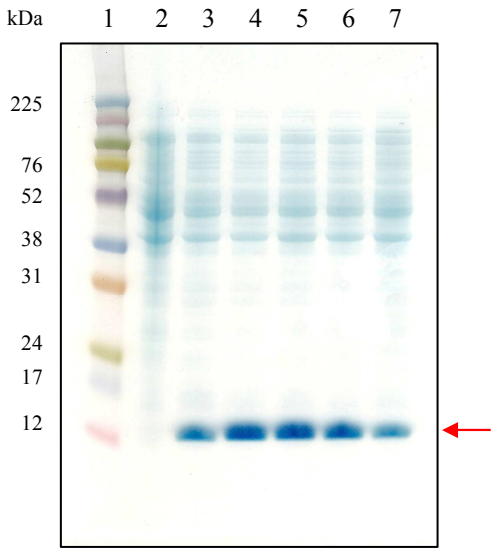


Figure 3.4: SDS-PAGE (Coomassie blue stained 12% NuPAGE gel) analysis of the expression of the GB1-NT fusion protein from *E. coli* transformant with different induction durations at 37 °C. Lane 1, molecular weight markers RPN800E (GE Healthcare); lane 2, supernatant of the cell lysate from non-induced cells; lane 3 to 7, supernatant of the cell lysate from pGEV-NT-harbouring cells induced with 1 mM IPTG for 2, 4, 6, 8, and 10 h, respectively. A red arrow indicates the position of the GB1-NT fusion protein.

Screening of induction temperature

The cell cultures were grown at 37 °C until an OD_{600} of 0.8 was reached and induced with 1 mM IPTG at 30 °C for 4, 6, 8, and 10 h. Equal amount (0.1 OD_{600}) of crude extracts was analysed on SDS-PAGE. The levels of expression at 30 °C for any periods of induction time were similar and lower than the expression level at 37 °C (**Figure 3.5**). The results revealed that decreasing the temperature during induction did not increase the expression level of the 9.2-kDa GB1-NT, so further reduction of induction temperature might not be needed.

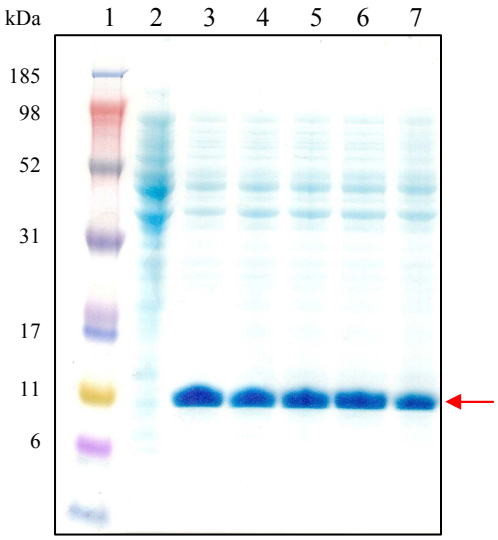


Figure 3.5: SDS-PAGE (Coomassie blue stained 12% NuPAGE gel) analysis of the expression of the GB1-NT fusion protein from *E. coli* transformant at 30 °C and 37 °C. Lane 1, molecular weight markers MultiMark (Invitrogen); lane 2, supernatant of the cell lysate from non-induced cells; lane 3, supernatant of the cell lysate from pGEV-NT-harbouring cells induced with 1 mM IPTG at 37 °C for 4 h; lane 4 to 7, supernatant of the cell lysate from pGEV-NT-harbouring cells induced with 1 mM IPTG at 30 °C for 4, 6, 8, and 10 h, respectively. The 9.2-kDa GB1-NT bands are indicated with a red arrow.

3.3.3 Expression and purification of GB1 fusion proteins

pGEV-NT was transformed into BL21(DE3) cells and expression was induced with 1 mM IPTG for 4 h at 37 °C with shaking at 200 rpm. The cultures were grown in baffled flasks to increase mixing and aeration. From the trials, although the expression level of fusion protein in the baffled flask was similar to that in the non-baffled flask, the amount of cell pellet was 1.5 times higher in the baffled flask (data not shown). Analysis of French Press cell lysate, pellet from cell lysate and soluble fraction by SDS-PAGE shows that most of the fusion proteins are located in the supernatant fraction, suggesting that the GB1-8xHis-Met-NT is soluble in the solution (**Figure 3.6**). The same expression approach was applied for pGEV-NT_{8-13}. The GB1-8xHis-Met-NT(P7M) fusion protein was expressed as soluble protein at high level, similar to GB1-8xHis-Met-NT (**Figure 3.7**). Expression of GB1 fusion proteins in M9 culture was achieved at up to 12.8 mg/10^8 cells, approximately 80% of the yield achieved with LB media (example of GB1-NT in **Figure 3.8**), suggesting that a high expression level of uniformly isotopically labelled peptides using minimal media can be achieved.

The fusion proteins were purified from the soluble cell fraction using Ni^{2+} affinity purification. The fusion proteins were eluted with a step gradient using 500 mM imidazole. The eluant yielded highly pure (> 95 %) fusion proteins determined by densitometry of gel (see **Figure 3.9**).

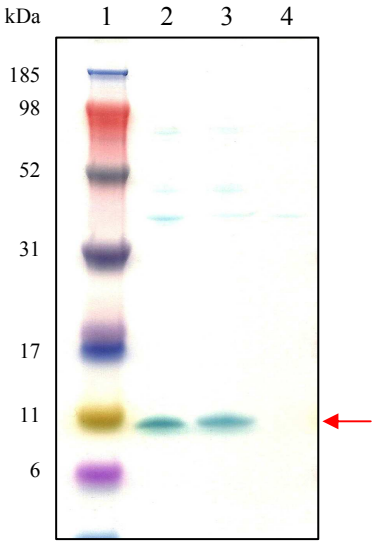


Figure 3.6: SDS-PAGE analysis of the cell lysate of *E. coli* BL21(DE3) transformed with pGEV-NT plasmid from 4 h induction with 1 mM IPTG at 37 °C after French press. 12% NuPAGE electrophoresis was performed on the following samples: lane 1, molecular weight markers MultiMark (Invitrogen); lane 2, whole cell lysate; lane 3, the supernatant fraction; lane 4, pellet from the cell lysate. Samples were visualised using Coomassie Brilliant Blue staining. The 9.2-kDa GB1-NT bands are indicated with a red arrow.

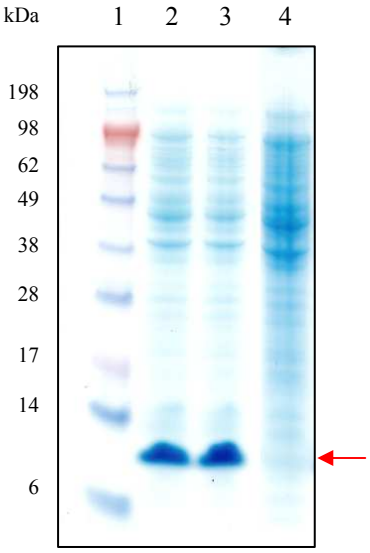


Figure 3.7: SDS-PAGE analysis of expressed fusion proteins. GB1-NT (lane 2) and GB1-NT(P7M) (lane 3) were expressed in BL21(DE3) cells for 4 h with 1 mM IPTG at 37 °C. Soluble fractions were separated on a 12% NuPAGE gel and visualised using Coomassie Brilliant Blue staining. Induced bands indicated by a red arrow migrating below the 14-kDa marker (lane 1 SeeBlue from Invitrogen) are seen. The bacterial proteins from uninduced cells are shown in lane 4. The predicted molecular weight of the expressed proteins is 9.2 kDa.

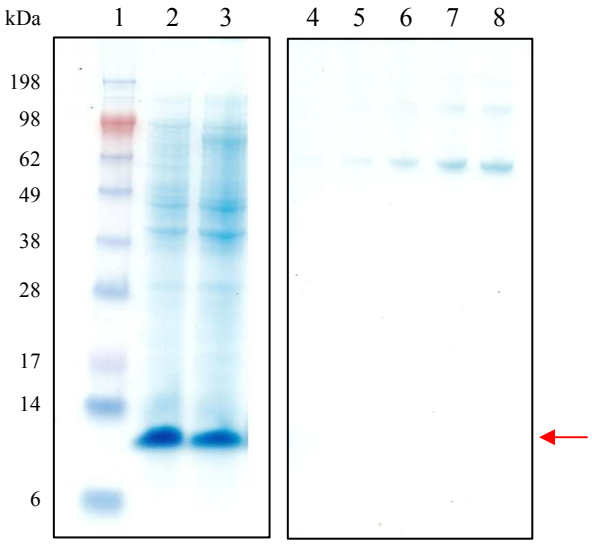


Figure 3.8: SDS-PAGE (Coomassie blue stained 12% NuPAGE gel) analysis of the expression of the GB1-NT fusion protein from *E. coli* transformant induced with 1 mM IPTG for 4 h in different media at 37 °C. Lane 1, molecular weight markers SeeBlue (Invitrogen); lane 2 and 3, supernatant of the cell lysate from pGEV-NT-harbouring cells in LB and M9 minimal media, respectively; lane 4-8, BSA standards 0.025, 0.05, 0.1, 0.20, and 0.25 mg/mL, respectively. Varying amounts of BSA standards were loaded onto the same gel to enable quantification of the amount of produced fusion protein using gel densitometry. A red arrow indicates the position of the GB1-NT fusion protein.

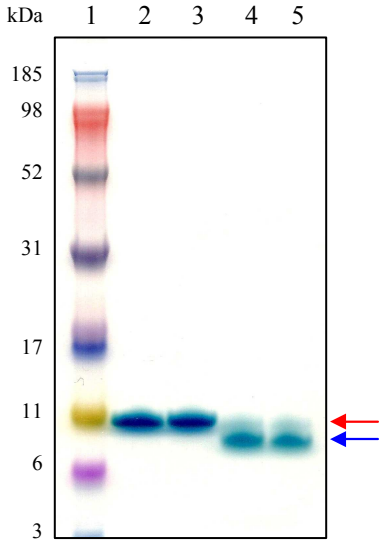


Figure 3.9: SDS-PAGE (Coomassie blue stained 12% NuPAGE gel) analysis of purified and CNBr cleaved GB1-NT and GB1-NT(P7M) from *E. coli* cells containing expressed fusion proteins. Lane 1, molecular weight markers MultiMark (Invitrogen); lane 2 and 3, Ni^{2+} affinity column-purified GB1-NT and GB1-NT(P7M), respectively, indicating the high degree of purification. Following CNBr cleavage, a small increase mobility as a product after the cleavage of GB1-NT (lane 4) and GB1-NT(P7M) (lane 5) indicated by a blue arrow is observed. The efficiency of CNBr cleavage determined by gel densitometry is > 70%. A red arrow indicates the position of the GB1 fusion protein.

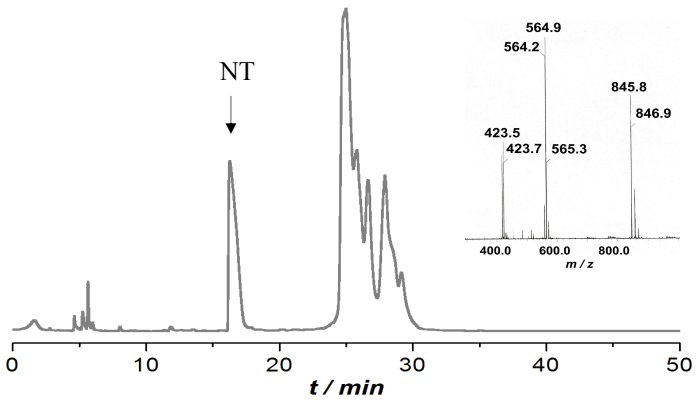
3.3.4 Cleavage of GB1 fusion proteins and purification of peptides

The fusion carriers were removed before using the target peptides. Since enzymatic cleavage (*i.e.* thrombin or Factor Xa) is quite expensive for separating target peptides from their fusion partners, chemical cleavage method is more often used for quantitative cleavage of peptides. Cyanogen bromide (CNBr) is commonly used since the efficiency of the CNBr cleavage at methionine residues is up to 99% (Lindhout et al., 2003) but it is not suitable for the production of peptides containing methionine residues. Fortunately, the neurotensin sequence lacks of internal methionine residues so here CNBr was used in order to give as high yield as possible at the cost-effectiveness.

CNBr cleaves C-terminally to Met residues (Gross, 1967), and can be used to remove the GB1-8xHis-Met and GB1-8xHis-Met-NT_{1-7} fusions from GB1-8xHis-Met-NT and GB1-8xHis-Met-NT(P7M) proteins, yielding the NT and NT_{8-13} peptides, respectively. Cleavage was performed with CNBr under acidic conditions (0.1 M HCl) with > 70% cleavage efficiency (**Figure 3.9**). The cleaved samples were purified by preparative reverse-phase HPLC. The linear gradient of acetonitrile from 20 to 40% (*v/v*) over 45 min separated NT and NT_{8-13} from the fusion tag and uncleaved products (**Figure 3.10**). Both NT peptides were eluted as a single peak at approximately 26.5% (*v/v*) acetonitrile. ESI mass spectrometry was used to analyse peaks from HPLC. The results confirmed that the purified peptides were cleaved at the specific methionine sites, yielding products of 1690 Da and 817 Da for NT and NT_{8-13}, respectively (**Figure 3.10**). The purity of the purified peptides determined from the HPLC chromatogram of a second elution of fractions shown in **Figure 3.10** was > 95%.

The expression of NT was previously performed using GST as a tag protein, yielding 1.8 mg per litre of LB (Williamson et al., 2000). Here, using GB1 as a fusion protein, NT and NT_{8-13}, yields of 9.2 mg (5.4 μmole) and 4.2 mg (5.1 μmole) per litre of M9 are achieved, respectively (**Table 3.2**). This improvement supports the observation of Hammarstrom, *et al.* (2006) (Hammarstrom et al., 2006) for over 45 human proteins, that the GB1+His fusion gives significantly higher yields of protein than the GST+His fusion. To produce a final yield of about 10 mg (5.6 μmole) of U-^{13}C, ^{15}N-labelled NT, suitable for NMR studies, only 4.4 g of ^{13}C-glucose and 1.7 g of $^{15}NH_4Cl$ would be required with a significant cost saving when compared with individual protected labelled amino acids using solid phase peptide synthesis.

(A)



(B)

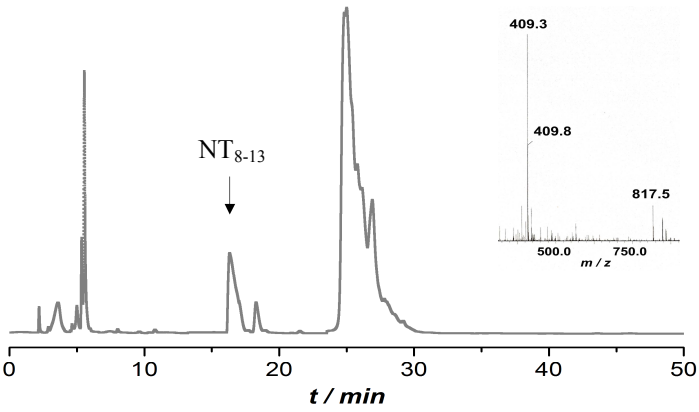


Figure 3.10: Reverse-phase HPLC purification of NT peptides. Reverse-phase HPLC traces (280 nm) and ESI mass spectra (insert) of purified peptides showing the separation of NT (A) and NT_{8-13} (B) from the GB1 tag and uncleaved products. Mass observed of NT: 845.8 Da $(M+2H)^{2+}$, 564.9 Da $(M+3H)^{3+}$ and 423.5 Da $(M+4H)^{4+}$ and mass observed of NT_{8-13}: 817.5 Da (M+H) and 409.3 Da $(M+2H)^{2+}$. Theoretical isotope mass number of NT and NT_{8-13} are 1689.9 and 816.5, respectively.

Table 3.2: Purification of NT and NT_{8-13} from *E. coli* BL21(DE3) harbouring pGEV-NT and pGEV-NT_{8-13} plasmids. The results were derived from 1 L M9 minimal media cultures of *E. coli* expressing each recombinant protein.

Construct	pGEV-NT	pGEV-NT_{8-13}
Cell pellet[a]	4.4 g	4.4 g
GB1 fusion protein[b]	65.8 mg	64.7 mg
Fusion protein purity[c]	> 95%	> 95%
Purified peptide[d]	9.2 mg	4.2 mg
Peptide purity[e]	> 95%	> 95%

[a] The wet weight of cells after 1 mM IPTG induction for 4 h at 37 °C
[b] Fusion protein after Ni^{2+} affinity purification
[c] Purity of fusion protein determined by densitometry of gel shown in **Figure 3.9** (lane 2 and 3)
[d] Target peptide after HPLC purification
[e] Purity of peptide determined by HPLC chromatogram

3.3.5 Production of uniformly labelled peptides

Following optimisation (section 3.3.2) and characterisation (section 3.3.4) of unlabelled peptides, these methods were then used for production of uniformly labelled peptides. For production of U-^{15}N labelled NT, U-^{13}C, ^{15}N labelled NT and U-^{13}C, ^{15}N labelled NT_{8-13}, the expression was performed in M9 minimal media plus YNB supplemented with $^{15}NH_4Cl$ and/or [$^{13}C_6$] glucose as the sole nitrogen and carbon sources, respectively. Positive ESI-MS *m/z* spectra (**Figure 3.11**) of HPLC-purified peptides verify the correct molecular weight of each labelled peptide. The yields of U-^{15}N labelled NT, U-^{13}C, ^{15}N labelled NT and U-^{13}C, ^{15}N labelled NT_{8-13} are 8.7, 9.1, 4.0 mg/L M9, respectively, indicating that incorporation of isotopic labels into the peptides in *E. coli* does not have any effects on expression.

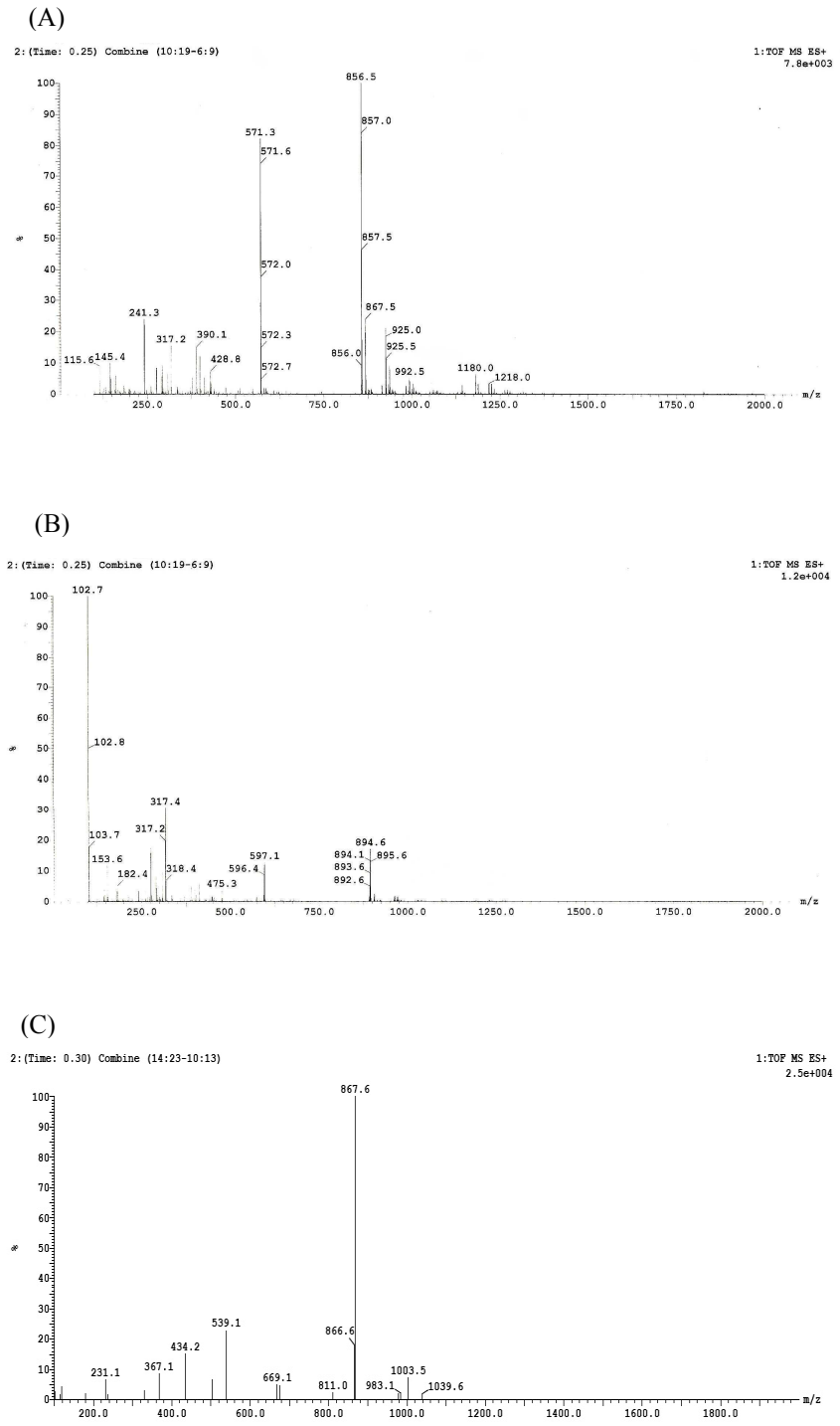


Figure 3.11: Positive-ion mass spectra of uniformly labelled peptides. Two peaks at *m/z* 571.3 $(M+3H)^{3+}$ and 856.5 $(M+2H)^{2+}$ correspond to the U-^{15}N labelled NT (A). Two major peaks at *m/z* 597.1 $(M+3H)^{3+}$ and 894.6 $(M+2H)^{2+}$ correspond to the U-^{13}C, ^{15}N labelled NT (B). One major peak at *m/z* 867.6 $(M+H)^{+}$ corresponds to the U-^{13}C, ^{15}N labelled NT_{8-13} (C).

3.3.6 Production of deuterated labelled peptides

The ampicillin resistance of the plasmid pGEV-NT was changed by transposition to the kanamycin resistance without any loss of neurotensin expression (data not shown). U-^{2}H labelled NT was produced as the consequence of the optimisation of expression of pGEV-NT in deuterated minimal media. Cell paste (4 g) was collected from 300 mL of culture with an OD_{600} ~10. The expression of deuterated GB1-NT was analysed on SDS-PAGE (**Figure 3.12**). Compared to the expression of unlabelled fusion protein, cells express significantly less in deuterated minimal media. The expected molecular weight of U-^{2}H labelled NT is about 1815 Da, and the positive ESI-MS *m/z* spectrum (**Figure 3.13**) of HPLC-purified peptide shows two peaks at *m/z* 595.8 $(M+3H)^{3+}$ and 893.3 $(M+2H)^{2+}$ corresponding to the highly (~98%) ^{2}H-labelled NT. The final yield of purified peptide is about 2.4 mg.

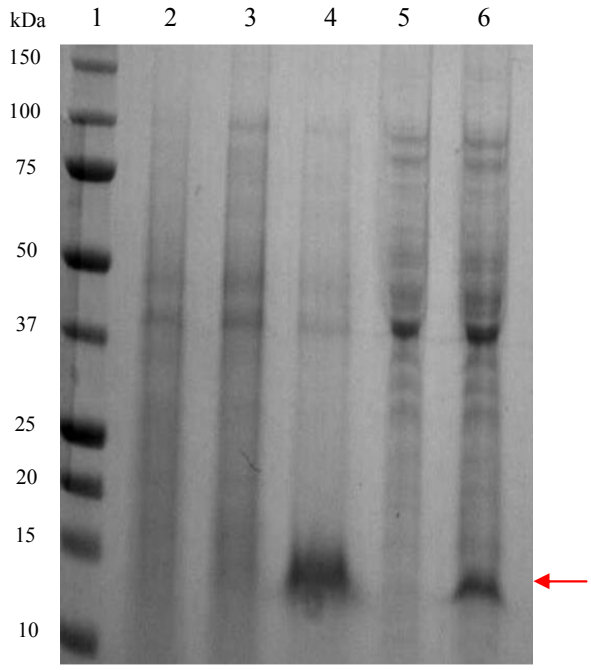


Figure 3.12: SDS-PAGE analysis of the expression of the GB1-NT in non-labelled and deuterated minimal media. Lane 1, molecular weight markers Odyssey (LI-COR Biosciences); lane 2 and 3, supernatant of the cell lysate from non-induced cells in non-labelled minimal media flask; lane 4, supernatant of the cell lysate from 1 mM IPTG induced cells in non-labelled minimal media flask; lane 5, supernatant of the cell lysate from non-induced cells in ^{2}H labelled media fermentor; lane 6, supernatant of the cell lysate from 1 mM IPTG induced cells in ^{2}H labelled media fermentor after 10 h. A red arrow indicates the position of the GB1-NT fusion protein. (Collaboration with the ILL-EMBL Deuteration Laboratory in Grenoble)

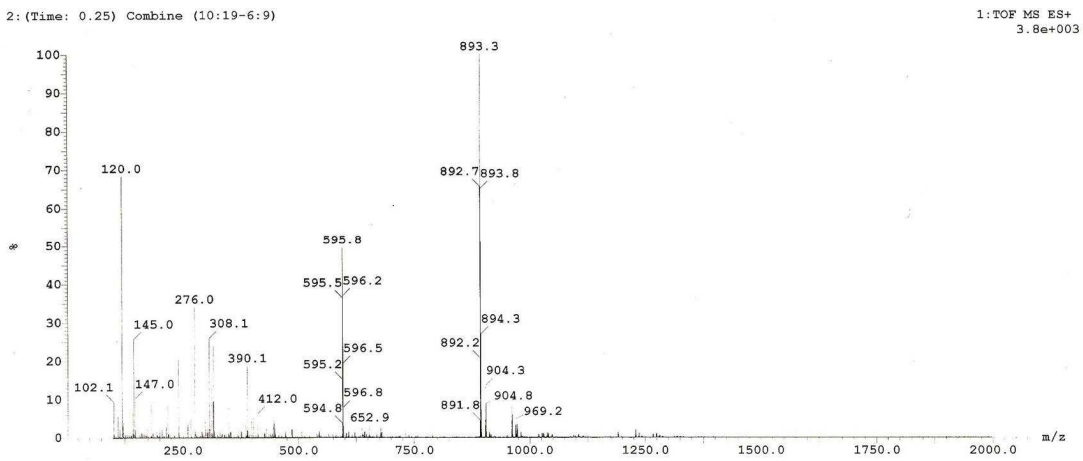


Figure 3.13: Positive-ion mass spectrum of uniformly ^{2}H labelled NT. Two peaks at *m/z* 595.8 $(M+3H)^{3+}$ and 893.3 $(M+2H)^{2+}$ correspond to the highly (~98%) deuterated NT.

U-^{2}H, ^{13}C, ^{15}N labelled NT was produced in triple labelled minimal media. The carbon and the nitrogen sources are D-glucose and $^{15}NH_4Cl$, respectively. Cell paste (3.6 g) was collected from 200 mL of culture with an OD_{600} of 7.4. Analysis of the triple labelled GB1-NT by SDS-PAGE shows the fusion protein with low expression level (**Figure 3.14**). The ESI mass spectrum (**Figure 3.15**) of HPLC-purified peptide verifies the correct molecular weight of ^{2}H (~98%), ^{13}C, ^{15}N labelled NT. The final yield of purified peptide is about 1.4 mg.

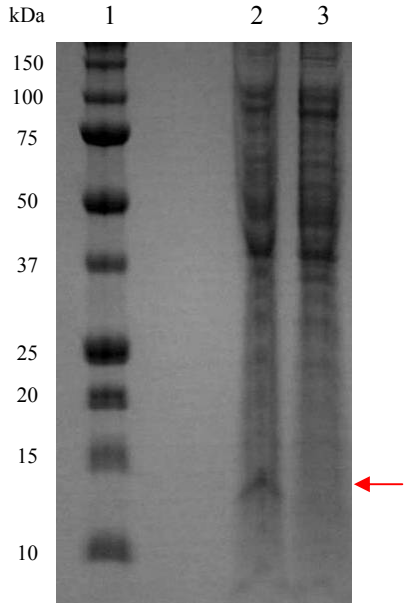


Figure 3.14: SDS-PAGE analysis of expression of the GB1-NT in triple labelled minimal media in a fermentor. Lanes 1, molecular weight markers Odyssey (LI-COR Biosciences); lane 2 and 3, supernatant of the cell lysate from 1 mM IPTG induced cells and non-induced cells after 15 h in ^{2}H, ^{13}C and ^{15}N labelled media at 30 °C, respectively. A red arrow indicates the position of the GB1-NT fusion protein. (Collaboration with the ILL-EMBL Deuteration Laboratory in Grenoble)

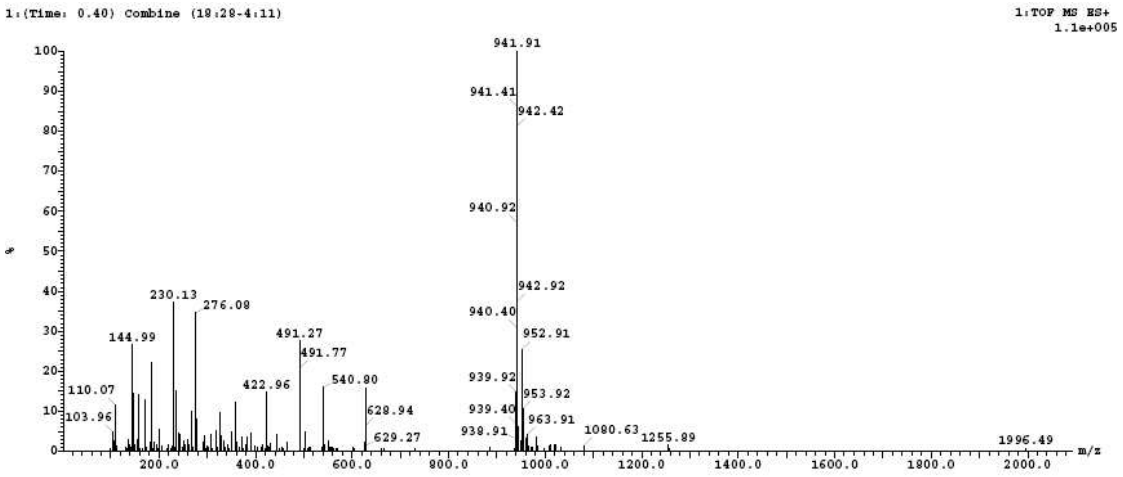


Figure 3.15: Positive-ion mass spectrum of uniformly ^{2}H, ^{13}C, ^{15}N labelled NT. The major peak at *m/z* 941.91 $(M+2H)^{2+}$ corresponds to the ^{2}H (~98%), ^{13}C, ^{15}N labelled NT.

3.3.7 NMR characterisation

The efficiency of labelling of NT and NT_{8-13} is demonstrated in **Figure 3.16 and 3.17**, respectively with ^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC of U-^{13}C, ^{15}N labelled NT and U-^{13}C, ^{15}N labelled NT_{8-13} being displayed. All backbone amide resonances of NT and NT_{8-13} were assigned by comparison with assignments reported in **Table 5.6 and 5.5**, respectively, with the exception of residues Glu1 and Leu2 not observed because of fast exchange with water and residues Asn5 and Arg8 not observed probably because of peptide dynamics (**Figure 3.16A and 3.17A**). NT and NT_{8-13} are unstructured in solution, as evidenced by the limited spectral dispersion in the ^{1}H-^{15}N HSQC NMR spectra. Moreover, two amide proton resonances instead of one resonance were detected for some residues of both peptides, indicating that the peptides are inhomogeneous in the aqueous solution.

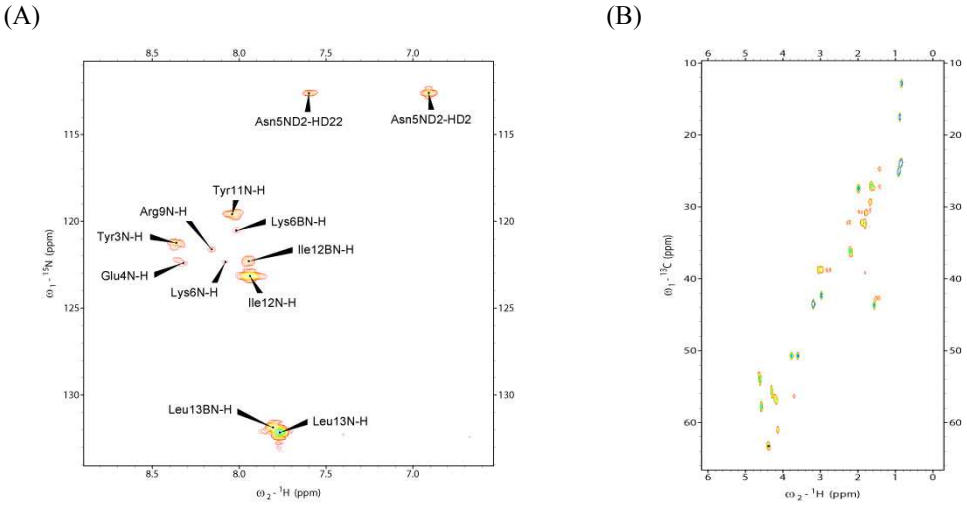


Figure 3.16: ^{1}H-^{15}N HSQC (A) and ^{1}H-^{13}C HSQC (B) spectra of U-^{13}C, ^{15}N labelled NT in aqueous solution (H_2O and 5% (*v/v*) D_2O, pH 6.0). All backbone amides were assigned by comparison with assignments reported in **Table 5.6**. The spectra were recorded from a Bruker 600 MHz NMR spectrometer at 298 K.

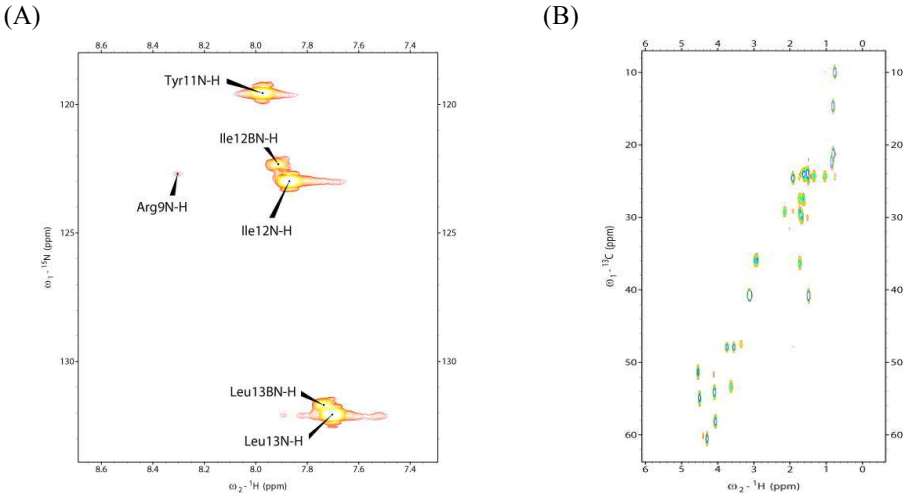


Figure 3.17: ^{1}H-^{15}N HSQC (A) and ^{1}H-^{13}C HSQC (B) spectra of U-^{13}C, ^{15}N labelled $NT_{8\text{-}13}$ in aqueous solution (H_2O and 5% (*v/v*) D_2O, pH 6.0). The backbone amide resonances were completely assigned by comparison with assignments reported in **Table 5.5**. The spectra were recorded from a Bruker 600 MHz NMR spectrometer at 298 K.

3.4 Conclusions

Solid phase peptide synthesis is a technique commonly used in the current production of peptides, however it can be hindered by high cost and relatively low

yields, particularly for the introduction of isotopic labels within the polypeptides. Here, we propose a use of a modified pGEV-1 plasmid to express neurotensin (NT), pGlu^{1}-Leu^{2}-Tyr^{3}-Glu^{4}-Asn^{5}-Lys^{6}-Pro^{7}-Arg^{8}-Arg^{9}-Pro^{10}-Tyr^{11}-Ile^{12}-Leu^{13}-OH as a C-terminal fusion protein with the GB1 domain of streptococcal protein G. The free carboxyl-terminus is important for the function of several peptide hormones, including neurotensin. Therefore, for the pGEV-NT construct, the C-terminal pGEV-encoded 6xHis tag was removed and an N-terminal 8xHis tag was introduced for affinity purification, followed by a methionine for tag removal using CNBr cleavage. Furthermore, this pGEV-NT plasmid was used as a template to include a Pro7 to Met mutation for CNBr cleavage, giving NT_{8-13}, the sub-fragment crucial for the biological activity of this peptide (Barroso et al., 2000; Lazarus et al., 1977). Both constructs produce the target peptides with high yield and low cost for potential use as uniformly isotopically labelled peptides for NMR and other spectroscopic and biophysical studies. In addition, the expression of peptides with a free C-terminus using pGEV-1 plasmid at high yield in solution could be useful for other peptide hormone and ligand expression.

pGEV-NT and pGEV-NT_{8-13} constructs were used to produce uniformly labelled NT and NT_{8-13} in high yield and in a cost effective way, using low cost ^{15}N and/or ^{13}C sources. The efficiency of labelling of neurotensin peptides is checked by ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC solution NMR experiments. Deuterated NT peptide was produced during optimisation of the expression of pGEV-NT construct in deuterated minimal media and triple labelled NT peptide was prepared for future solution NMR studies to investigate the protein-ligand interactions.

Chapter 4

Purification and reconstitution of neurotensin receptor 1

4.1 Introduction

4.1.1 Expression and purification of NTS1

Structural studies require large amounts of active protein and therefore the yield of protein following expression and purification are important (Chelikani et al., 2006; Shibata et al., 2009; Warne et al., 2003). Overexpression of integral membrane proteins has proved extremely difficult. Unlike rhodopsin, which is naturally concentrated and available in high yields in retinal membranes, other GPCRs are endogeneously expressed at low levels and so their purification from natural sources is not practicable. Thus, heterologous expression systems are needed for overexpression of GPCRs in functionally active forms, and once this has been accomplished, the expressed proteins have to be extracted from the host cell membranes and purified (Sarramegna et al., 2006; Sarramegna et al., 2003).

Neurotensin receptor 1 (NTS1) from rat is one of a few examples of GPCRs that have been expressed as a functionally ligand-binding protein at the *E. coli* plasma membrane. The NTS1 fusion protein has been expressed in *E. coli* from pRG/III-hs-MBP-T43NTS1-TrxA-H10 plasmid (termed NTS1A construct) (**Figure 4.1**) (Grisshammer et al., 1999; Tucker and Grisshammer, 1996). The plasmid contains

truncated neurotensin receptor (Δ1-42), which is fused at its N-terminus to a maltose binding protein (MBP) together with its signal sequence in order to target the protein to the periplasm of *E. coli*. The receptor is also fused at its C-terminus to thioredoxin (TrxA), to aid stability, and a deca-his tag to assist in purification. Proteolysis during expression of the NTS1 fusion protein was prevented by removing a protease-sensitive site found at the N-terminus of NTS1. The activity of the NTS1A construct was similar to that of the membrane-bound protein and the native receptor, and high expression levels of functional NTS1A receptor (800 receptors/cell) were obtained (Tucker and Grisshammer, 1996).

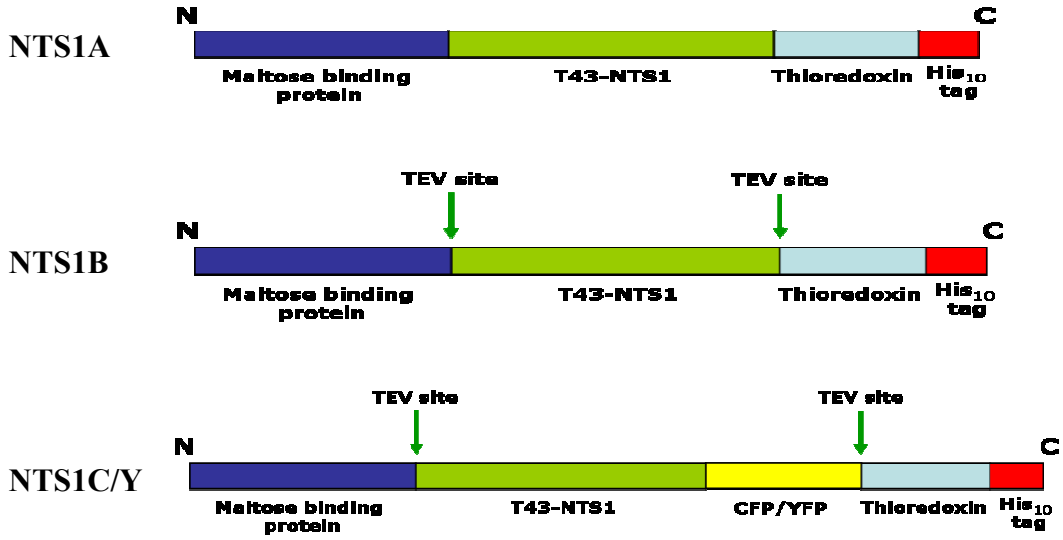


Figure 4.1: Schematic representation of the rat NTS1 constructs. NTS1A consists of maltose binding protein (MBP), an N-terminally truncated form of NTS1 (T43NTS1), thioredoxin and a deca-histidine tag (His_{10} tag). The NTS1B was constructed by including tobacco etch virus (TEV) protease cleavage sites into the NTS1A construct between the MBP and T43NTS1 moieties and the T43NTS1 and thioredoxin moieties to facilitate proteolytic removal of the fusion tags. The eCFP and eYFP proteins were included into the NTS1B construct between the T43NTS1 and C-terminal TEV site moieties to yield the NTS1C and NTS1Y constructs, respectively.

The NTS1B construct was generated by including tobacco etch virus (TEV) protease recognition sites between the MBP and T43NTS1 moieties and the T43NTS1 and TrxA moieties, yielding the NTS1 receptor alone after TEV cleavage and purification (**Figure 4.1**). The presence of the TEV cleavage sites does not influence expression levels (Grisshammer et al., 2005; White et al., 2004). Recently, the optimisation of expression and purification of NTS1B construct were demonstrated by us (Attrill et al., 2009). The plasmid was transformed into different *E. coli* strains *i.e.* DH5*α* and BL21(DE3), C41(DE3), and C43(DE3) cells. Expression of the NTS1B receptor in DH5α cells, which have been traditionally used for its expression at 20 °C for 30 h under the control of the *lac* operator/promoter yields of 0.08 mg of active receptor per litre of culture. Among four different *E. coli* strains, the BL21(DE3) cells gave the highest yield of NTS1B of 0.2 mg per litre of culture, which is an approximate 2.5-fold increase when compared with DH5*α* strain. Furthermore, the presence of POPC or a combination of phospholipids (POPC/POPG/POPE, 3:1:1) was shown to enhance the recovery of functional receptor during the purification process using the optimal concentration of DDM/CHS/CHAPS (Attrill et al., 2009).

Recently, the NTS1B construct was modified to include enhanced cyan fluorescent protein (eCFP) (termed NTS1C, **Figure 4.1**) and enhanced yellow fluorescent protein (eYFP) (termed NTS1Y, **Figure 4.1**) between the T43NTS1 and C-terminal TEV site moieties for multimerisation studies of NTS1 using a fluorescence resonance energy transfer technique. Proteolytic removal of fusion tags yields fluorescent-tagged NTS1 after expression and purification. A saturation radioligand binding assay was used to verify the high affinity of purified fluorescent-

tagged NTS1 for NT which is comparable to that of nonfluorescent receptor (Harding et al., 2009).

4.1.2 Reconstitution of purified NTS1

To purify membrane proteins, it is necessary to solubilise them using detergents, stripping away the membrane lipids which surround its hydrophobic transmembrane regions and keeping the membrane proteins in a non-denaturing environment during further purification (Seddon et al., 2004). Re-introducing the purified protein into a lipid membrane is desirable for structural and biophysical studies, particularly SSNMR experiments, allowing it to be observed in a more native environment which supports function (Watts et al., 2004). Detergent-mediated reconstitution (**Figure 4.2**) is the most successful and frequently used strategy for proteoliposome preparation (Rigaud et al., 1995). The principle of this technique is that the protein is first co-solubilised with an excess of phospholipids in the appropriate detergent in order to form an isotropic solution of lipid-protein-detergent and lipid-detergent micelles. The detergent is then removed from the micellar solution, resulting in the progressive formation of bilayer vesicles with incorporated protein. There are several methods to remove detergents, for example dialysis, gel filtration technique, hydrophobic absorption, and dilution. The hydrophobic absorption method using adsorptive polystyrene beads (Biobeads SM-2) has been shown so far to be efficient for all kinds of detergents (Rigaud et al., 1995) and it was chosen for reconstitution of NTS1 and fluorescent-tagged NTS1 into a model membrane system (Harding et al., 2009; Luca et al., 2003).

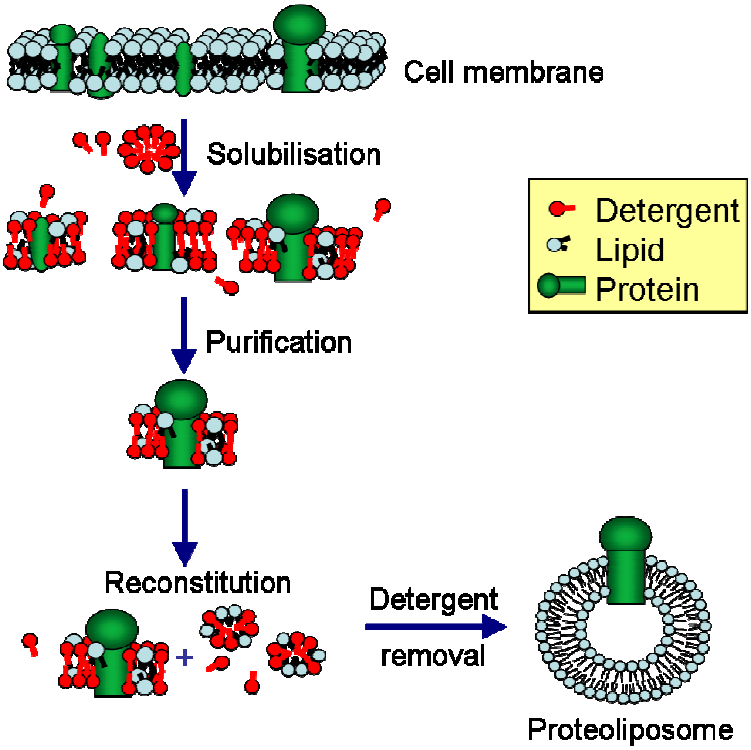


Figure 4.2: Schematic representation of detergent-mediated reconstitution of membrane proteins into liposomes. Starting from a cell membrane, most membrane proteins are isolated with solubilising detergent concentrations. After isolation and purification, the membrane proteins are supplemented with an excess of phospholipids and reconstituted into liposomes with appropriate detergent to form a solution of mixed lipid-protein-detergent and lipid-detergent micelles. Next, the detergent is removed, leading to the formation of closed lipid bilayers with incorporated protein.

4.1.3 Aims

The aim here is to obtain sufficient amounts of NTS1 receptor for NMR studies. Overexpression of NTS1B protein in *E. coli* BL21(DE3) and efficient purification by nickel affinity chromatography followed by gel filtration chromatography allow protein production in large quantities. TEV-cleaved NTS1B in detergent buffer will be used for probing NTS1-NT interactions by solNMR techniques. For SSNMR experiments, in principle the presence of liposomes and fusion tags has no effect on spectral resolution. Therefore, the NTS1B receptor will be reconstituted into liposomes which mimic a native environment.

4.2 Materials and methods

4.2.1 Expression of neurotensin receptor 1

For small scale expression in shake flasks, *E. coli* clone strain BL21(DE3) harbouring pRG/III-hs-MBP-TEV-T43NTS1-TEV-TrxA-H10 plasmid expressing the NTS1B was inoculated into 5 mL of 2xTY media (100 µg/mL ampicillin and 0.2% (*w/v*) glucose) at 37 °C and 200 rpm agitation for 16-20 h. One percentage of the overnight culture was inoculated into fresh 2xTY media and grown at 37 °C until the OD_{600} reached 0.4. The temperature was decreased to 20 °C and protein expression was induced with IPTG at a final concentration of 0.25 mM for 40 h when an OD of 0.6 was reached. *E. coli* cultures were collected by centrifugation at 6000 xg, 4 °C for 30 min, flash frozen in liquid nitrogen, and stored at -80 °C. Large scale expression in a fermentor at 500 litres was done at GSK in Stevenage.

4.2.2 Large scale purification of neurotensin receptor 1

All steps were carried out at 4 °C. Cell pellet (40 g) from large scale protein production was resuspended in 80 mL of 2x solubilisation buffer (100 mM Tris-HCl (pH 7.4), 400 mM NaCl, 60% (*v/v*) glycerol, protease inhibitors (Leupeptin, Pepstatin A, Aprotinin, all 1 µg/mL)) with a hand homogeniser. Lysozyme was added at a concentration of 1 mg/mL, followed by gentle agitation (30 min). Detergent solutions of 5% (*w/v*) CHAPS/1% (*w/v*) CHS (16 mL) and 10% (*w/v*) DDM (16 mL) were added dropwise whilst stirring. MilliQ water was added to give a final volume of 160 mL. DNase (1 U/mL) and $MgCl_2$ (final concentration of 5 mM) were added followed by stirring for 1 h. The resulting suspension was centrifuged at 75000 xg, 4 °C for 90

min and then the supernatant was filtered by passing through 0.45 μm and 0.22 μm filters. Protease inhibitors and imidazole (final concentration of 50 mM) were further added.

An AKTA Basic FPLC purification system (GE Healthcare) with Frac 920 fraction collector was used for column purification steps. The solubilised material was purified by immobilised metal affinity chromatography (IMAC) using a 5 mL HisTrap HP column (GE Healthcare) (see Appendix A.5 for column preparation). The column was equilibrated with 5 column volume (CV) of NiA buffer (50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 30% (*v/v*) glycerol, 0.5% (*w/v*) CHAPS, 0.1% (*w/v*) DDM, 0.1% (*w/v*) CHS, 50 mM imidazole, protease inhibitors) and the sample was loaded into the column with a flow rate of 2.5 mL/min. After loading, the column was washed with 40 CV of NiA buffer to remove non-specifically bound protein, and the NTS1B was eluted using NiB buffer (NiA plus 350 mM imidazole). Eluted fractions were combined and then concentrated to ~3.5 mL using Vivaspin 20, 100,000 MWCO (VWR International) and loaded onto a Superdex 200 XK16/100 gel filtration column pre-equilibrated with GF1 buffer (50 mM Tris-HCl (pH 7.4), 15% (*v/v*) glycerol, 500 mM NaCl, 0.1% (*w/v*) DDM, 0.01% (*w/v*) CHS, 0.05 mg/mL POPC, 1 mM EDTA, protease inhibitors). Proteins were separated at a flow rate of 0.25 mL/min and 2 mL fractions collected. Fractions containing the NTS1B receptor were pooled and concentrated to ~3.5 mL using Vivaspin 20, 100,000 MWCO (VWR International). Purification was analysed by Western blotting using anti-His antibody (GE Healthcare) and radioligand binding assay.

4.2.3 Purification of TEV-cleaved neurotensin receptor 1

The cleavage of fusion tags from NTS1 was carried out by Dr Helen Attrill for solNMR experiments. The IMAC-purified NTS1B from a 60 g pellet weight was treated with TEV protease at a concentration of approximately 4 μg TEV-His6 per μg fusion protein in a total volume of 35 mL cleavage buffer (50 mM Tris-HCl (pH 7.4), 15% (*v/v*) glycerol, 750 mM NaCl, 0.5% (*w/v*) CHAPS, 0.1% (*w/v*) DDM, 0.01% (*w/v*) CHS, 0.05 mg/mL of a combination of POPC/POPG/POPE, 1 mM EDTA, 5 mM DTT), at 4 °C overnight with gentle agitation. The mixed solution was concentrated to 3.5 mL using Vivaspin 20, 100,000 MWCO (VWR International) and loaded onto a Superdex 200 XK16/100 gel filtration column preequilibrated with GF2 buffer (50 mM Tris-HCl (pH 7.4), 50 mM NaCl, 0.1% (*w/v*) DDM, 0.01% (*w/v*) CHS). Proteins were separated at a flow rate of 0.25 mL/min and 2 mL fractions collected. Purification was analysed by SDS-PAGE and radioligand binding assay. Fractions containing the NTS1 receptor were pooled and concentrated to 300 μL using Vivaspin 20, 30,000 MWCO (VWR International). Glycerol (5% (*v/v*)) was added to the sample prior to storage at -80 °C.

4.2.4 Preparation of lipid for protein reconstitution

Liposomes were prepared from porcine brain polar lipids (BPL, Avanti Polar Lipids, Inc.) dissolved in chloroform and methanol in a ratio of 1:1 (*v/v*). The solvent was evaporated off using a stream of nitrogen gas and placed under vacuum desiccator overnight. The resultant film was then resuspended in reconstitution buffer (50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA) to a final concentration of 5 mg/mL and small unilamellar vesicles were formed by bath sonication (three 2 minute bursts).

Unilamellar vesicles with relatively homogenous size were made by 10 cycles of freezing in liquid nitrogen and slow thawing at room temperature, followed by extrusion through 400 nm and 100 nm polycarbonate filters (Whatman) using a mini-extruder (Avanti Polar Lipids). It has been previously proved that 11 passes through each membrane pore size yield unilamellar vesicles of uniform size distribution (MacDonald et al., 1991).

4.2.5 Reconstitution of neurotensin receptor 1

Liposomes (5 mg/mL) were incubated with 5 mM DDM for 3 h at 4 °C to destabilise them and then mixed with purified, detergent-solubilised receptor in a molar ratio of 2000:1 (lipid:protein) and incubated at 4 °C for 1 h with mild shaking. Meanwhile, Bio-beads (Biorad) were washed with methanol for 3 times at 4 °C with mild shaking, to saturate the lipid binding sites on the Bio-beads, and then washed twice with water and reconstitution buffer. For the removal of detergent, 80 mg/mL of saturated Bio-beads was added to the protein/lipid mixture above, and the sample was incubated overnight at 4 °C with gentle agitation. The Bio-beads were filtered off and the reconstituted samples were layered on top of a discontinuous sucrose gradient (5% steps, 0-35% (*w/v*) sucrose in reconstitution buffer) and centrifuged overnight at 100,000 xg, 4°C (Beckman SW28 Ti rotor). Fractions from the gradient were collected and analysed by SDS-PAGE. Fractions containing the receptor were pooled and stored at -80 °C.

4.2.6 Preparation of reconstituted NTS1 pellet for SSNMR experiments

The frozen samples from section 4.2.5 were thawed and diluted with equal volume of reconstitution buffer. The proteoliposomes were harvested by centrifugation at 360,000 xg, 4 °C for 2 h (Beckman Type 70 Ti rotor). The resulting pellets from the ultracentrifugation were then washed and exchanged with four different buffers (50 mM Tris-HCl (pH 7.4), 1 mM EDTA; 50 mM Tris-HCl (pH 7.4), 20 mM KCl, 1 mM EDTA; 50 mM Tris-HCl (pH 7.4), 20 mM KCl, 1 mM EDTA, 5% (*v/v*) glycerol; 50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA) to find the most suitable buffer for structural studies by SSNMR spectrometry. The samples were again centrifuged at 325,000 xg, 4 °C for 2 h (Beckman Type 70.1 Ti rotor). The supernatant fractions were discarded and the pellets were resuspended in their buffers. The radioligand binding assay (section 4.2.8) was used to determine the effect of salt and glycerol on protein stability.

4.2.7 Resolubilisation of reconstituted NTS1

Reconstituted samples were resolubilised with detergents to expose any binding sites which have been occluded after reconstitution. 2x solubilisation buffer (150 µL) and 10% DDM/1% CHS (*w/v*) (50 µL) were added to reconstituted samples (200 µL). The incubation period was 1 h at 4 °C to solubilise reconstituted receptor.

4.2.8 Radioligand binding assay

The amount of active protein in detergent-purified and reconstituted samples was measured using the binding of [^{3}H]-NT (specific activity of 3.33 TBq/mmol, equating to 63,636 cpm/pmol, purchased from PerkinElmer) (Grisshammer et al.,

1993). Binding experiments were performed in the final volume of 60 µL of assay buffer (50 mM Tris-HCl (pH 7.4), 0.1% (*w/v*) DDM, 0.01% (*w/v*) CHS, 1 mM EDTA, 0.1 mg/mL BSA) containing [^{3}H]-NT to a final concentration of 5 nM. Detergents were omitted for assaying the reconstituted sample. The amount of active receptor was calculated assuming one binding-site per molecule, a molecular weight of 101 kDa for the NTS1B and a saturation of 90% of total sites (Fractional occupancy = [Ligand]/[Ligand] + K_d, where receptor K_d was determined as 1.0 ± 0.2 nM for the NTS1B). Non-specific binding was accounted for using competition with excess unlabelled NT (13 µM). The incubation period was 1 h at 4 °C and bound ligands were separated from free ligands by gel filtration using Micro Bio-Spin 30 columns in Tris buffer (Biorad) for detergent solubilised sample and by filtration using Durapore PVDF membranes with a 0.2 µm MWCO (Millipore) for reconstituted samples. Flow-through sample and filter membranes were collected and added to 5 mL of scintillation fluid (Wallac Optiphase HiSafe, PerkinElmer). The amount of [^{3}H]-NT was measured by scintillation counting. The filter membranes were incubated for 6 h and mixed by vortexing prior to scintillation counting. Specific binding was determined by subtracting non-specific cpm from total cpm. The protocol was done in triplicate for each sample.

4.2.9 Quantitation of the total protein

Protein samples were separated by SDS-PAGE and visualised using Coomassie Brilliant Blue staining. Intensities of bands were measured by gel densitometry (ImageJ, NIH). The total protein content of NTS1B was calculated by

using different concentrations of bovine serum albumin (BSA) loaded on the same gel as references.

4.3 Results and discussion

4.3.1 Production of neurotensin receptor 1

The NTS1B construct in *E. coli* strain BL21(DE3) encoded 101 kDa-NTS1B receptor upon IPTG induction as described in section 4.2.1. Several batches of 40 g cell pellet (600 g in total) were solubilised, purified using nickel affinity chromatography. The Ni-column elute was further purified using gel filtration chromatography. **Figure 4.3** demonstrates elution profiles of a representative purification. The NTS1B contains ten histidine residues at the C-terminus, which allows purification by IMAC. With 2 CV of NiB buffer containing 350 mM imidazole, the recombinant protein was eluted with a single peak monitored by the absorbance at 280 nm (**Figure 4.3A**). SDS-PAGE was used to analyse the proteins in the eluted fractions. The expected bands migrating above the 80-kDa marker corresponding to the NTS1B receptor were observed (**Figure 4.3B**). Fractions containing the receptor (2-6) were pooled and loaded on a gel filtration column.

The purification of NTS1B receptor by gel filtration chromatography was monitored at 280 nm (**Figure 4.3C**). The major peak corresponds to the recombinant protein with a high degree of purity, as assayed by SDS-PAGE (**Figure 4.3D**). Gel filtration fractions 11-16 were combined and concentrated. **Figure 4.4** shows the progress of NTS1B purification by SDS-PAGE visualised by Coomassie Brilliant Blue staining. The binding activity of purified NTS1B receptor for NT was confirmed by the binding of [^{3}H]-NT. The details of 15 purifications are shown in **Table 4.1**. A total

of 28.2 nmol of active protein as determined by radioligand binding assay was obtained.

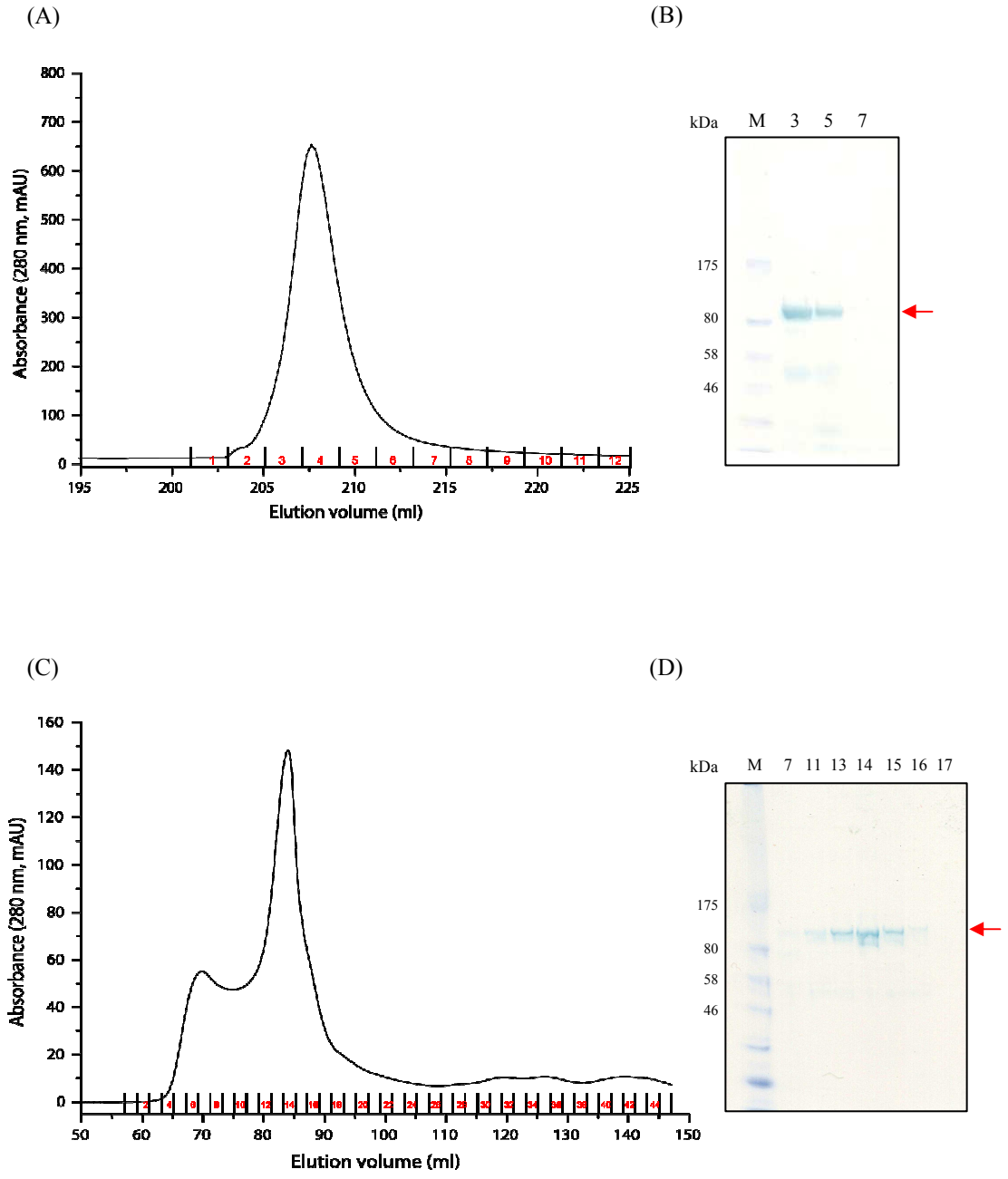


Figure 4.3: Elution profiles (A) and (C) monitored at 280 nm of the purification of recombinant NTS1B by IMAC and gel filtration chromatography, respectively and SDS-PAGE analysis (B) and (D) of the elution fractions. The proteins were separated on 12% SDS-PAGE gels and were visualised by Coomassie Brilliant Blue staining. The molecular masses of the marker proteins P7708S (BioLabs) are indicated (lane M). The lanes indicated in panels (B) and (D) correspond to the respective fraction numbers indicated in the elution pattern (A) and (C), respectively. Red arrows indicate the position of NTS1B receptor.

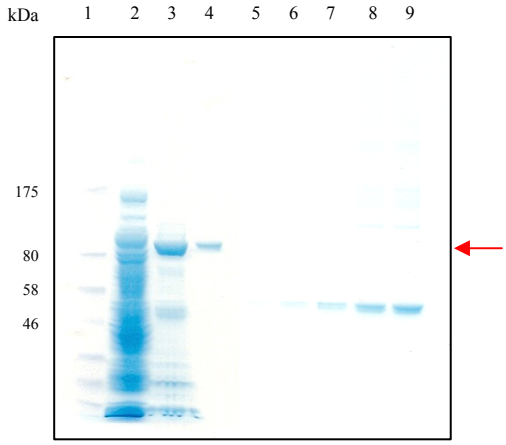


Figure 4.4: SDS-PAGE (Coomassie blue stained 12% SDS-PAGE gel) analysis of the NTS1B (MBP-TEV-T43NTS1-TEV-TrxA-H10) purification from an *E. coli* BL21(DE3) culture. Lane 1, molecular weight markers P7708S (BioLabs); lane 2, supernatant of the cell lysate; lane 3, IMAC elute; lane 4, gel filtration elute; lane 5-9, BSA standards 0.025, 0.05, 0.1, 0.20, and 0.25 mg/mL, respectively. The bands of NTS1B receptor (101 kDa) indicated by a red arrow migrate between the 175-kDa and 80-kDa markers. Varying amounts of BSA standards were loaded onto the same gel to enable quantification of the amount of receptor (by gel densitometry) after gel filtration purification which will be used for further reconstitution step.

Table 4.1: Yields obtained after gel filtration column from 15 purifications (40 g pellet for each purification). The total protein is determined by gel densitometry using BSA standards as references and the active protein is calculated from the radioligand binding assay.

Number	Total protein (nmol)	Active protein (nmol)
1	4.9	1.6
2	6.0	1.9
3	3.6	1.2
4	5.9	1.7
5	5.9	1.9
6	14.2	3.9
7	12.0	3.0
8	6.2	1.7
9	6.5	1.7
10	5.9	1.8
11	4.0	1.2
12	5.1	1.6
13	6.0	1.7
14	6.0	1.8
15	4.7	1.5
Total	96.9	28.2

4.3.2 Production of TEV-cleaved neurotensin receptor 1

Cell pellet (120 g) of *E. coli* strain BL21(DE3) harbouring the NTS1B construct was solubilised with DDM/CHS/CHAPS and the full length protein was purified by nickel affinity chromatography. NTS1 (TEV-cleaved NTS1B) was obtained by incorporating a proteolytic cleavage reaction step after IMAC purification of the protein before subsequent gel filtration purification. After protease treatment, the proteolytic fragments were analysed by SDS-PAGE (**Figure 4.5**). The Coomassie-stained band migrating between the 30-kDa and 46-kDa markers corresponds to the calculated molecular weight of TEV-cleaved NTS1B of 43 kDa. The target product was separated from the unwanted proteins in the mixture with a high degree of purity by gel filtration chromatography. The amount of functional protein detected in the combined gel filtration elutes by radioligand binding assay was 10.5 nmol from 120 g pellet.

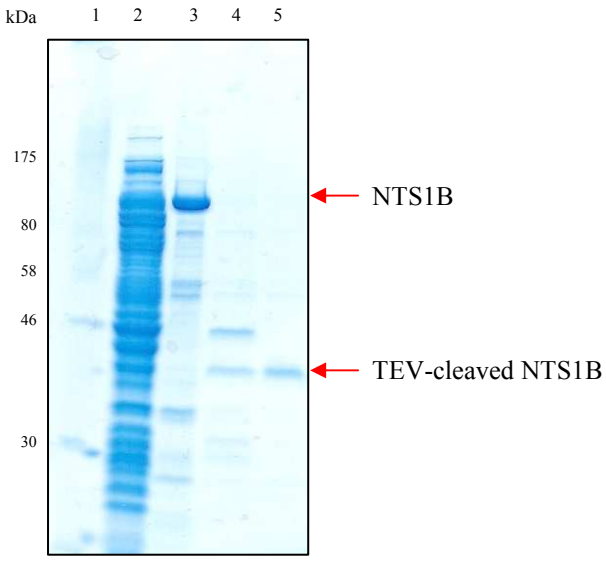


Figure 4.5: SDS-PAGE analysis of the TEV-cleaved NTS1B purification from an *E. coli* BL21(DE3) culture. The products were separated by 12% SDS-PAGE gel and Coomassie Brilliant Blue staining. Lane 1, molecular weight markers P7708S (BioLabs); lane 2, supernatant of the cell lysate; lane 3, IMAC elute; lane 4, the protein mixture after incubation with TEV protease; lane 5, gel filtration elute. The bands of NTS1B receptor (101 kDa) migrating between the 175-kDa and 80-kDa markers and TEV-cleaved NTS1B (43 kDa) migrating between 46-kDa and 30-kDa markers are seen.

4.3.3 Reconstitution of neurotensin receptor 1

Reconstitution of membrane proteins into liposomes yields the protein in a favourable environment for structure-function studies. The detergent purified NTS1B receptor (11 of 15 purifications from section 4.3.1) was reconstituted into BPL with a lipid to protein molar ratio of 2000:1. After detergent removal using Biobeads SM-2 (Biorad), proteoliposomes were isolated from the reconstitution samples using sucrose density gradient ultracentrifugation (**Figure 4.6A**). A band close to the top of the sucrose gradient (0-35%) was lipid and another relatively tight band between sucrose concentrations of 15-20% was a band of reconstituted NTS1B, as analysed by SDS-PAGE (**Figure 4.6B**). The proteoliposomes are present as a single band in the sucrose gradient suggesting that the receptors were incorporated into the BPL and the resulting proteoliposome population was relatively homogenous amongst the liposomes. The fractions from the sucrose gradient which contained NTS1B were combined and the amount of active proteoliposomes was measured by radioligand binding assay. **Table 4.2** gives results of 11 batch reconstitution experiments. A total of 8.64 nmol of active protein was reconstituted into the BPL liposomes (as determined by radioligand binding assay of resolubilised proteoliposomes). However, only half (4.32 nmol) of reconstituted receptor is available for ligand binding because about 50% of the reconstituted receptor was oriented with the binding site on the inner face of the vesicle (as determined by radioligand binding assay of proteoliposomes).

(A) (B)

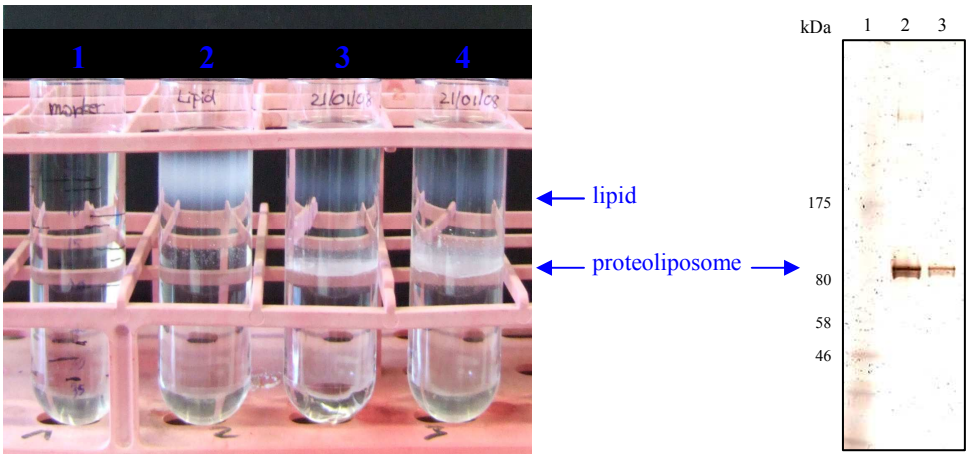


Figure 4.6: Reconstitution of NTS1 into BPL liposomes at a lipid-protein molar ratio of 2000:1. The reconstitution samples were loaded on top of sucrose gradients (0-35%) and centrifuged at 100,000 xg for 16 h (A). Tube 1, sucrose concentration markers; tube 2, BPL alone; tube 3 and 4, reconstitution samples. Free lipid can be seen on the top of tubes. Proteoliposome bands at the centre of tubes 3 and 4 can be seen between sucrose concentrations of 15-20%. Fractions containing the proteoliposomes were analysed by SDS-PAGE and silver staining (B). Lane 1, molecular weight markers P7708S (BioLabs); lane 2 and 3, fractions collected from the proteoliposome bands of tube 3 and 4, respectively. The reconstituted receptor is present in each tube at the location of the proteoliposome band.

Table 4.2: Yields obtained after reconstitution from 11 of 15 preparations in section 4.3.1. The amount of active protein reconstituted into the BPL is calculated from the radioligand binding assay of resolubilised proteoliposomes.

Number	Active reconstituted NTS1B (nmol)
1	0.52
2	0.75
3	0.24
4	0.78
5	1.03
6	1.69
7	1.26
8	0.69
9	0.68
10	0.73
11	0.27
Total	8.64

4.3.4 Preparation of reconstituted NTS1 pellet for SSNMR studies

Potassium ions have slightly more favourable effect on protein stability than do sodium ions (Pikal-Cleland et al., 2000; von Hippel and Schleich, 1969) and the sensitivity of NMR experiments can be reduced by high salt concentrations because the increase of ionic and dielectric conductivity of samples results in dissipation of the radio frequency power and appearance of ring currents in the sample tube (Kelly et al., 1993; Kovacs et al., 2005). Therefore, here attempts were made to prepare NMR sample using KCl buffer instead of NaCl buffer and the concentration of salt was kept as low as possible.

Proteoliposomes were recovered by ultracentrifugation and resuspended in four different buffers to investigate whether salt and glycerol have any effects on protein stability as described in section 4.2.6. **Figure 4.7** shows the binding activity of proteoliposomes in different conditions. After freeze-thawing the samples two times, the protein in a buffer lacking salt (50 mM Tris-HCl (pH 7.4), 1 mM EDTA) lost its binding activity to NT about 50% compared to a reference sample (reconstitution buffer:50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1mM EDTA), whereas the proteins in buffers containing low concentration of salt (50 mM Tris-HCl (pH 7.4), 20 mM KCl, 1mM EDTA) either with or without cryoprotectant (5% (*v/v*) glycerol) gave similar amount of active proteins to the reference after 2 rounds of freeze-thawing. These results have revealed that salt is needed for the stability of NTS1. In addition, glycerol is likely not necessary for protecting the protein from freezing damage. However, for the long-term storage and several freeze-thaws which might be occurred during SSNMR experiments, the low-salt buffer containing 50 mM Tris-HCl (pH 7.4), 20 mM KCl, 1 mM EDTA and 5% (*v/v*) glycerol will be used.

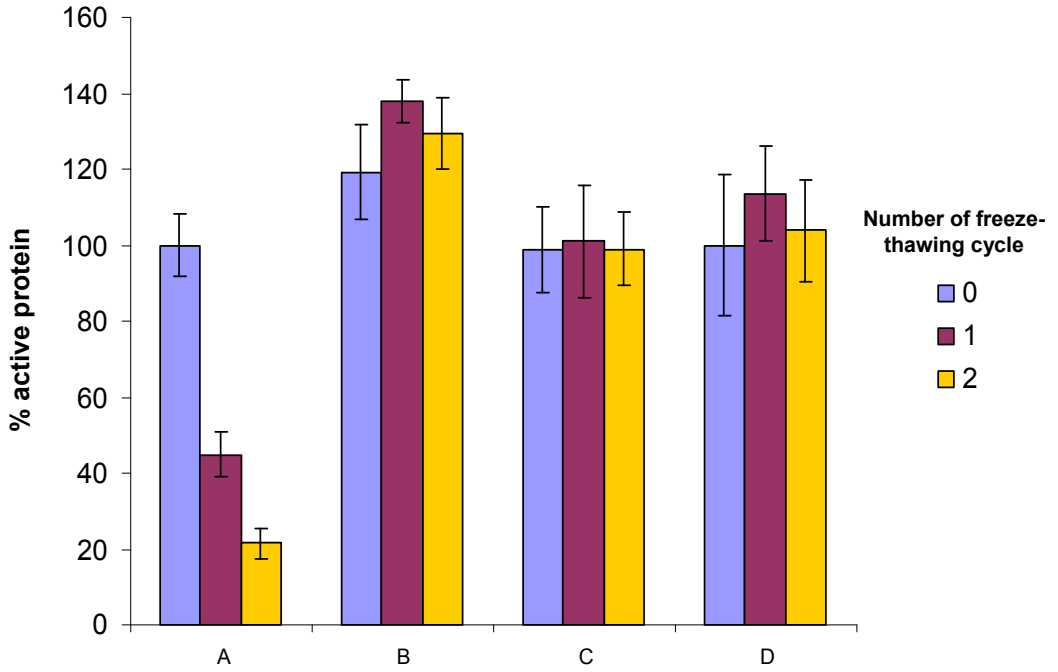


Figure 4.7: Effects of salt and cryoprotectant on protein stability during freeze-thawing. Buffers A: 50 mM Tris-HCl (pH 7.4), 1 mM EDTA; B: 50 mM Tris-HCl (pH 7.4), 20 mM KCl, 1 mM EDTA; C: 50 mM Tris-HCl (pH 7.4), 20 mM KCl, 1 mM EDTA, 5% (*v/v*) glycerol; D: 50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA. The control sample is the protein in reconstitution buffer (buffer D). Error bars represent standard error of the mean from three independent experiments.

4.4 Conclusions

Large amounts of active NTS1B (28.2 nmol) were obtained and some of it was reconstituted into the BPL liposomes for SSNMR studies of NTS1-NT complex (Chapter 5). After reconstitution, the pellet of proteoliposomes was recovered from ultracentrifugation and washed with a low salt buffer containing 50 mM Tris-HCl (pH 7.4), 20 mM KCl, 1 mM EDTA and 5% (*v/v*) glycerol as the protein requires some salt for its stability and some cryoprotectant might be needed for long-term storage and many times of freeze-thawing (**Figure 4.7**).

The Ni-purified NTS1B was incubated with TEV protease and the TEV-cleaved receptor, which lacks the MBP and TrxA-H10 fusions, was obtained. This protein sample will be used for solNMR studies of NTS1-NT interactions (Chapter 5).

Chapter 5

Structural studies of bound neurotensin by NMR spectroscopy

5.1 Introduction

5.1.1 Sample preparation for solid-state NMR studies

As the resolution of a spectrum is important for structural studies, samples are sometimes prepared under different conditions for resolution comparison. A solid-state NMR (SSNMR) sample must be some kind of a rigid form. Examples of sample forms for SSNMR studies include lyophilised powders, bicelle-protein suspensions, microcrystals or nanocrystals precipitated by organic molecules, membrane proteins in membranes, and frozen sample solutions. The spectral linewidth, an indication of spectral resolution, is determined by the structural homogeneity of the sample. Lyophilisation (freeze-drying) of samples is the simplest method of preparing an SSNMR sample and can improve the sample stability in the long-term. The lyophilised or powder samples have been used in several SSNMR experiments, for example the studies of molecular mobility of biological materials (Saitô et al., 1987; Yoshioka et al., 1999) and the structural and dynamic studies of polymeric systems (Brown, 2009). However, lyophilisation may destroy the protein sample and alter the protein secondary structure (Desai et al., 1994). Also, it often provides the sample with structural inhomogeneity as although the protein backbones are correct, the

conformation of side-chains of each residue may be quite variable. The lyophilised sample may give the spectrum with poor resolution since it typically shows 1-2 ppm linewidths and this resolution is not sufficient for most of structural studies (**Figure 5.1A**) (Martin and Zilm, 2003). The sample heterogeneity can sometimes be improved by rehydration as shown in several proteins (Gregory et al., 1993a; Gregory et al., 1993b; Pauli et al., 2000).

A single, large, good quality crystal is required to obtain good diffraction in X-ray crystallography, but the crystallisation trial often end up with the sample as a precipitate or in a nanocrystalline form instead, which may be useless in terms of high resolution X-ray diffraction. This is because X-ray diffraction requires long-rang order in a sample. However, these nanocrystals or precipitates are useful in structure determination by SSNMR spectroscopy since NMR needs only short-range order. It has been demonstrated that microcrystals yielded spectra with narrow linewidths in SSNMR studies of many proteins (see an example in **Figure 5.1A**) (Igumenova et al., 2004; Martin and Zilm, 2003; McDermott et al., 2000; Pauli et al., 2000; Polenova, 2004). In addition, precipitation conditions can affect the SSNMR spectral linewidth as shown in **Figure 5.2** (Igumenova et al., 2004). For structural studies of membrane proteins by SSNMR, freezing the membrane protein solution at very low temperature (typically below -50 °C) is the straightforward method of performing SSNMR experiments as shown in the structure determination of neurotensin peptide when bound to neurotensin receptor (Luca et al., 2003; Williamson et al., 2002). Detergent-solubilised membrane proteins can also be precipitated as microcrystals or reconstituted into lipid bilayers to form proteoliposomes or 2D crystals (Knol et al., 1998; Rigaud et al., 1997). There is no general consensus as to which of four forms

(nanocrystalline, microcrystalline, proteoliposome and 2D crystalline forms) is the best for structural membrane protein studies by SSNMR, however a few cases have suggested that the spectra obtained from proteoliposomes are worse than the others. For instance, the 2D crystals of *E. coli* outer membrane protein G (OmpG) yielded better resolution spectra than the proteoliposome form (Hiller et al., 2005). Also, spectra obtained from nanocrystals of diacylglycerol kinase (DGK) exhibited narrower linewidths than spectra from proteoliposomes (Lorch et al., 2005). Bicelles, bilayer disc with both long and short chain phospholipids, can be used for sample preparation of membrane proteins. The uniformly ^{15}N labelled fd coat protein was studied under several different conditions. The sample in phospholipid bilayers yielded better spectral resolution than the sample in bicelles, yielding better resolved lines than the one in powder form (**Figure 5.1B**) (Marassi and Opella, 1998).

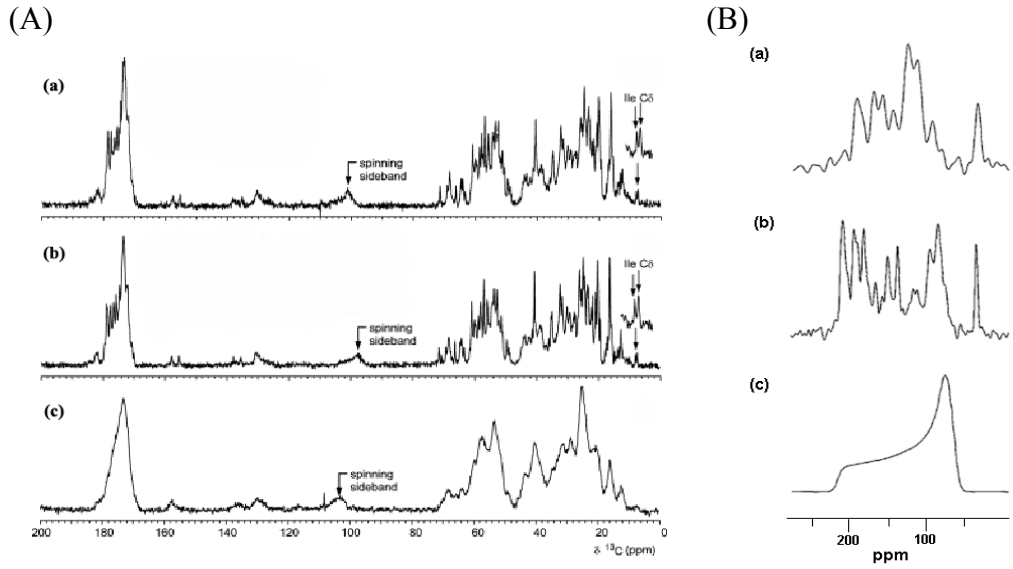


Figure 5.1: Effects of sample conditions on SSNMR spectra. A: the natural abundance ^{13}C spectra of polycrystalline (a) and nanocrystalline (b) ubiquitin, produced from the same crystallisation condition, and lyophilised ubiquitin. The nanocrystalline sample yields well resolved spectrum, practically identical to that of the larger crystals, whereas structural inhomogeneity is observed in the spectrum of lyophilised sample. Adapted from (Martin and Zilm, 2003). B: the spectra of uniformly ^{15}N labelled fd coat protein in oriented bicelles with the addition of lanthanide ions (a), in oriented phospholipid bilayers (b) and in powder form (c). The spectral resolution of the sample in bilayers is better than those of the samples in bicelles and powder condition. Adapted from (Marassi and Opella, 1998).

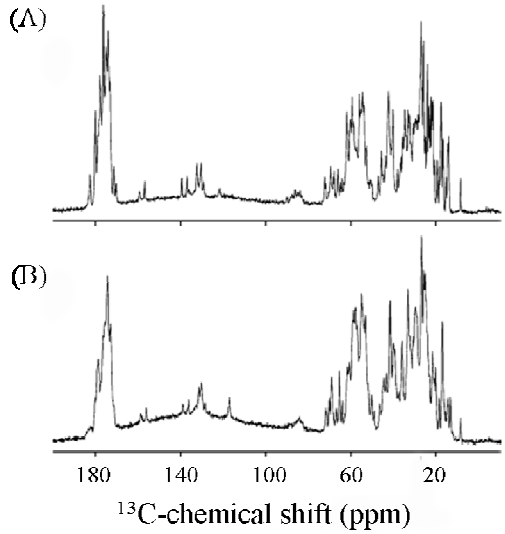


Figure 5.2: Effects of precipitating agents on spectral linewidths. 1D ^{13}C CP MAS spectra of microcrystalline ubiquitin, crystallised from 2-methyl-2,4-pentanediol (A) and PEG 8000 (B) at 400 MHz. Adapted from (Igumenova et al., 2004).

5.1.2 NMR assignments

In order to study the structure and dynamics of proteins by either solution- or solid-state NMR spectroscopy, it is necessary to complete the assignment for a sufficient number of atoms in the sequence. The spectral resonance assignment means correlation of each resonance with associated nuclei in the investigated molecule. The assignment can be from homonuclear and/or heteronuclear spectra depending on available NMR samples, *i.e.* homonuclear 2D spectra for unlabelled proteins, ^{15}N heteronuclear spectra for ^{15}N labelled proteins, or triple resonance spectra for $^{15}N/^{13}C$ doubly labelled proteins. In general, the assignment of proteins can be divided in two main parts: the sequential assignment of residues in the protein sequence and the assignment of the amino acid side-chains. Two types of nuclear interactions are used in NMR spectra for the assignment of the chemical shifts of the nuclei: through-bond interaction (scalar coupling) and through-space interaction (dipolar coupling) (Duer, 2002; Teng, 2005).

In solNMR, multidimensional homonuclear and heteronuclear chemical shift correlation methods, such as correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY) and heteronuclear single quantum correlation (HSQC) use scalar coupling to correlate the spins within a spin system for the resonance assignment of proteins. Only signals of protons which are two or three bonds apart are visible in a COSY spectrum and correlations of all protons of a spin system are detected in the TOCSY experiment. In the HSQC experiment, the magnetisation is transferred from protons to the hetero nucleus (^{13}C or ^{15}N) (Teng, 2005). At MAS in SSNMR experiments, the resonance assignment of proteins in a rigid form is from multidimensional spectra using both through-bond and through-space interactions for coherence transfer (Ader et al., 2007; Brus and Jegorov, 2004; Duer, 2002).

Signals of proton, as well as ^{13}C and ^{15}N are commonly detected for NMR assignments in solNMR experiments. However, proton detection encounters considerable difficulties in SSNMR since strong dipolar couplings between protons give rise to structureless broad lines, even when using sophisticated resolution enhancement techniques combined with fast MAS of the high magnetic field. Since ^{13}C and ^{15}N nuclei have smaller dipolar couplings and larger chemical shift range than ^{1}H nucleus, they give rise to much better resolved spectra. Therefore, they are the nuclei of choice for detection in SSNMR experiments (Duer, 2002).

5.1.3 Investigation of protein-ligand complexes by NMR spectroscopy

Currently, there are three major approaches, X-ray crystallography, solNMR and SSNMR, to study protein-ligand interactions at atomic resolution. X-ray crystallography is an excellent method for the large molecules, but it has some

significant difficulties, *i.e.* getting crystals and good diffraction. NMR-based methods have become more important techniques to study protein-ligand interactions because of the difficulty of ligand-GPCR crystallisation. These NMR approaches can be applied to weak and transient protein-ligand complexes that are difficult to study by other structural methods (Villar et al., 2004). Also, proteins can be studied in a variety of environments, such as different buffer conditions, oriented bilayers, crystals, or bicelles (Sanders and Sonnichsen, 2006; Varga and Watts, 2007). There are several methods in solNMR to determine the conformation of ligand at the receptor interacting site. Studying the conformation of bound ligand to its receptor using NMR techniques, the peptide ligand must be isotopically labelled to avoid the crowding signals from large receptors. Transferred nuclear overhauser effects (trNOE) is a method to observe the signal of the labelled peptide bound to the unlabelled receptor if the dissociation constant (off-rate) of the ligand from the protein is faster than the longitudinal relaxation rate of the magnetisation of the peptide.

The conformation of isotopically labelled peptide antagonists ($cCRF_{30-41}$) of corticotrophin-releasing factor (CRF) receptors was determined by using ROESY and trNOE techniques to study the free and bound ligands, respectively. The analysis of NMR spectra has indicated that the peptide has no conformation in the solution and a helical model while bound to the isolated N-terminal extracellular domain of the CRF receptor (Mesleh et al., 2007). However, if the affinity between a ligand and a protein does not give significant transferred NOEs, this method would not be a suitable approach.

Differential perturbation in linewidths and chemical shifts in ^{15}N-^{1}H heteronuclear correlation NMR spectra is another method to provide the information

of residues of peptide ligand involved in interactions with its protein. This method is used as a ligand screening tool for drug discovery (Shuker et al., 1996). However, it may not be appropriate for structural studies of large, integral membrane or membrane associated protein since molecules in solution have no rapid isotropic motions that would normally lead to an averaging of anisotropic interaction and provide the high resolution spectra from which structural information may be obtained.

Most GPCRs are sensitive to temperature so the experiment must be performed at low temperature but solNMR normally works well at high temperatures, such as room temperature. Additionally, GPCRs are membrane proteins so it would be more interesting to study the ligand in the bound state in the native environment which is surrounded by lipids. Alternatively, SSNMR can overcome these averaging limitations to provide structural and dynamic information on proteins and their ligands (Watts, 2002). Detecting chemical shift changes of labelled ligand upon the binding is one of the common methods in SSNMR studies to observe the interaction around the labelled ligand at the binding site. For instance, perturbation in chemical shifts by 1.6 ppm of $N^{+}(^{13}CH_3)$ labelled acetylcholine bound to its receptor, nicotinic acetylcholine receptor was observed at the methyl groups attached to the quaternary ammonium (**Figure 5.3**). It has been suggested that there are cation-π interactions between aromatic residues of ligand and some residues of receptor (Williamson et al., 1998).

Observed chemical shift changes by SSNMR experiments have indicated that both tyrosine side-chain and the C-terminal end of U-^{13}C, ^{15}N labelled NT_{8-13} interact strongly with the neurotensin receptor upon binding (Williamson et al., 2002). However, it would be even more interesting to confirm those observed interactions

using more precise method, a distance measurement. This promising method can answer more specific questions focusing on sites of interest.

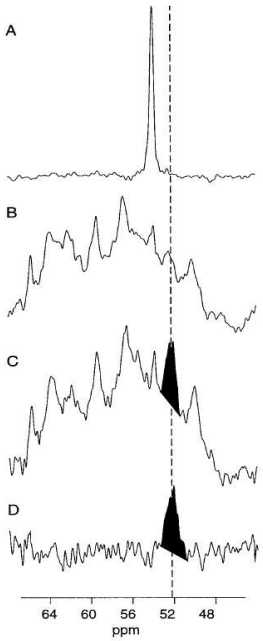


Figure 5.3: Studies of ligand-receptor interaction by CP MAS experiments. ^{13}C spectra of $N^+(^{13}CH_3)$ acetylcholine (A), acetylcholine receptor-rich membranes in the absence of $N^+(^{13}CH_3)$ acetylcholine (B), as in (B) but upon the addition of 40 nmol of $N^+(^{13}CH_3)$ acetylcholine (C), difference spectrum resulting from the subtraction of B from C (D). Black area represents intensity due to bound $N^+(^{13}CH_3)$ acetylcholine, and dotted line indicates the position for the bound $N^+(^{13}CH_3)$ acetylcholine (Williamson et al., 1998).

5.1.4 Aims

As the project focuses on the conformation of neurotensin peptide whilst bound to its G-protein coupled receptor, NTS1, spectral linewidth of the peptide in the absence of receptor is important in resonance assignments by SSNMR experiments, which will be used to compare with the one in the presence of receptor. The aim is to obtain spectra of peptide with narrow lines and hence maximum resolution and signal intensity. The optimisation of peptide preparations should be similar to that of protein samples, so the crystallisation of the NT_{8-13} peptide has been performed.

The NTS1-NT interactions have been studied by NMR spectroscopy. The changes in chemical shifts of peptide when the receptor is present are expected to be observed in either solNMR or SSNMR experiments.

5.2 Materials and methods

5.2.1 Preparation of the precipitated NT_{8-13} peptide

The crystallisation of unlabelled NT_{8-13} was screened by a crystallisation kit containing 50 different solution conditions (Structure Screen 1 - Catalogue Number MD1-01, Appendix C). Two concentrations of peptide solutions (50 mg/mL and 100 mg/mL) were prepared by resuspension of lyophilised samples (from section 2.3.1) in water, followed by sonication for 15 minutes in a sonication bath. The samples were spun at 13,000 xg for 1 minute to remove aggregates. The hanging drop method was used for setting up crystallisation trials as shown in **Figure 5.4**. A thin drop of vacuum grease was applied to the rim of each well of a vapour diffusion plate. Five hundred microlitre of precipitant 1 from the Structure Screen 1 was added into well A1. A 1 μL drop of peptide solution was placed into the centre of a siliconised glass coverslip. The same volume of precipitating agent from the Structure Screen 1 was added to the sample drop and suspended. The coverslide was quickly inverted over well A1 and sealed on the edge to minimise evaporation. The other precipitants from the Structure Screen 1 were continued by the same procedure in remaining wells. The vapour diffusion plate was then incubated at 15 °C. The progress of the crystallisation was checked under a microscope.

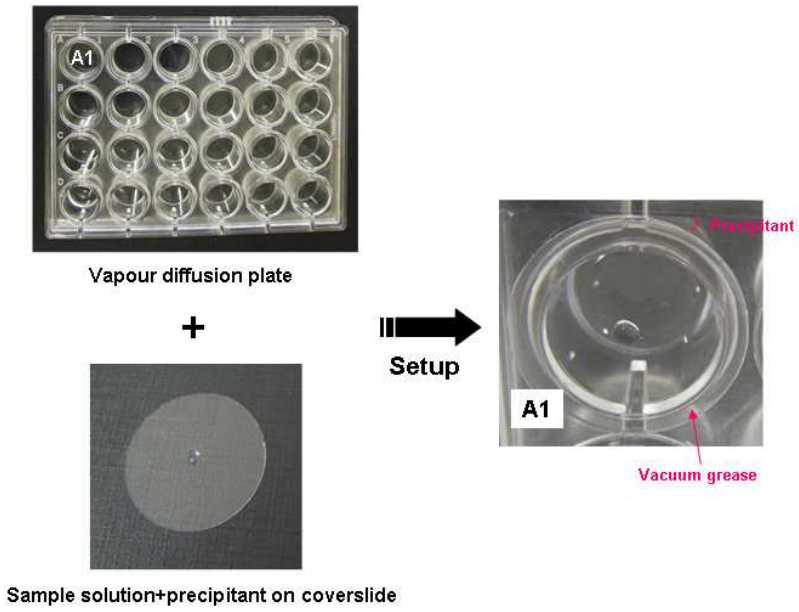


Figure 5.4: Hanging-drop vapour-diffusion technique of sample crystallisation (see text for details).

The chosen condition (precipitant 9) was further optimised in order to obtain a high yield of the precipitate. The crystallisation of 100 mg/mL concentration of sample in the selected crystallisation solution was tested in two different containers (Eppendorf tube and sealed yellow pipette tip) and at two temperatures (4 °C and 15 °C).

The NT_{8-13} peptide was precipitated in the mixture of 20% PEG 4000, 50 mM sodium citrate pH 6.0 and 20% isopropanol (precipitating agent 9) in the sealed yellow pipette tip. The solution was kept at 4 °C for 2 days before the precipitate was separated by centrifugation at 10000 xg; overnight; 4 °C as show in **Figure 5.5A**. The precipitate was resuspended in water and analysed by ESI mass spectrometry. Approximately 1 mg of the precipitate was transferred into a thick wall 4-mm rotor (Varian, USA) by centrifugation at 3000 xg; 2 min; 4 °C (**Figure 5.5B and 5.5C**).

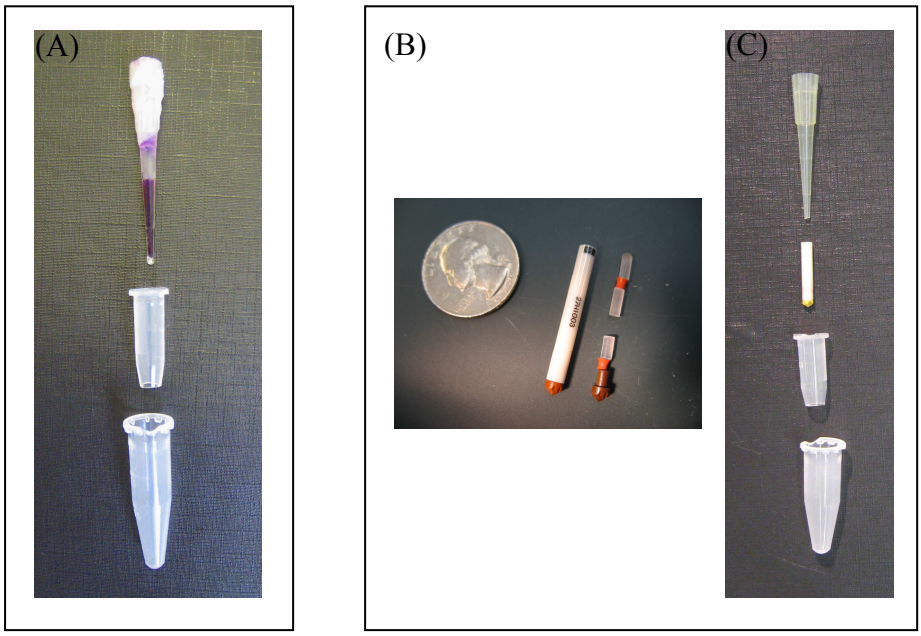


Figure 5.5: Sample packing for solid-state NMR experiments. The precipitate (blue colour) was separated from the crystallisation solution by fast centrifugation (A). The sample was packed in a 4-mm rotor in the following order: driving tip, rubber disc, spacer, sample, spacer, rubber disc, end cap (B). The bottom of the sealed pipette tip containing the sample was cut and the sample was transferred into the centre of the rotor where the homogeneity of the magnetic field is located by slow centrifugation (C).

5.2.2 Preparation of uniformly ^{13}C, ^{15}N labelled NT in the brain polar lipid

The U-^{13}C, ^{15}N labelled NT was prepared in brain polar lipid (BPL) for a negative control, so the BPL was prepared using exactly the same condition as for reconstitution of detergent-purified NTS1B in section 4.2.4. The lipid was destabilised with DDM and the detergent was removed using Bio-beads. The lipid was then loaded into the sucrose-gradient tube and the fraction of BPL was collected after ultracentrifugation. The pellet of BPL was washed once with a low-salt buffer containing 50 mM Tris-HCl (pH 7.4), 20 mM KCl, 1mM EDTA and 5% (*v/v*) glycerol. The solid sample of BPL was transferred into the sealed yellow pipette tip and the labelled peptide was added and mixed. Approximately 100 µL of the

calculated volume of solid BPL with labelled ligand was loaded into a thin wall 4-mm rotor (maximum capacity is 100 μL).

5.2.3 Preparation of uniformly ^{13}C, ^{15}N labelled NT with reconstituted NTS1

From section 4.3.4, in the washing step of the reconstituted NTS1B pellet with low-salt buffer, the U-^{13}C, ^{15}N labelled NT was added before the high speed ultracentrifugation to ensure that there was no excess of unbound ligand in the SSNMR sample. The sample was packed in a thin wall 4-mm rotor, containing ~1.36 nmol of active reconstituted protein, calculated from radioligand binding assays. Therefore, approximately 2.43 μg (~1.36 nmol) of bound labelled ligand was loaded into the rotor.

5.2.4 Resonance assignments for unlabelled NT_{8-13} in D_2O and precipitant

^{1}H-^{13}C HSQC and COSY experiments were performed on a Bruker DRX 500 MHz NMR spectrometer in chemistry department's facilities (Oxford). TOCSY and ROESY experiments were performed on an Oxford Instruments 750 MHz NMR spectrometer with a GE/Omega computer in NMR facilities of Prof. Iain Campbell (IDC group, Oxford). All recorded data were referenced to an external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) standard, whose carbon and proton resonances are defined as 0 ppm at 298 K. Approximately 5 mg of natural abundance peptides was dissolved in 0.60 mL of pure D_2O (99.96-100%), pH 6.0 and the same amount was dissolved in 0.60 mL of precipitating agent 9 prepared with pure D_2O, 5% (*v/v*) H_2O to verify whether the precipitant has any effects on the conformation of the peptide. The samples were loaded into 5-mm NMR tubes. All NMR data were

processed and analysed using nmrPipe (Delaglio et al., 1995) and Sparky (http://www.cgl.ucsf.edu/home/sparky/), respectively.

^{1}H-^{13}C HSQC, COSY, TOCSY, and ROESY data were acquired with 128 t_1 increments using States-TPPI phase cycling. The mixing time for TOCSY and ROESY is 80 and 300 ms, respectively. All four spectra were recorded for the peptide in D_2O and only two spectra of TOCSY and ROESY were collected for the peptide in the precipitant 9. Data in t_1 was linear predicted to 256 points and data points in both dimensions were zero-filled to 1024 for NMR data processing.

5.2.5 Resonance assignments for uniformly ^{13}C, ^{15}N labelled NT in solution

About 0.5 mg of uniformly ^{13}C, ^{15}N labelled NT was dissolved in 0.60 mL of detergent buffer (GF2 buffer with 5% (*v/v*) glycerol), pH 7.4 with 5% (*v/v*) pure D_2O and loaded into a 5-mm NMR tube. U-^{13}C, ^{15}N labelled NT facilitates the backbone and side-chain assignments by a number of double and triple-resonance experiments (Cavanagh et al., 2007). The NMR experiments were performed on an Oxford Instruments 600 MHz NMR spectrometer with a GE/Omega computer (IDC group, Oxford) at 298 K. All recorded data were referenced to an external DSS standard and processed by nmrPipe (Delaglio et al., 1995). The spectra were analysed by Sparky (http://www.cgl.ucsf.edu/home/sparky/).

Here, a combination of 3D CBCANH, CBCA(CO)NH and HNCO was utilised for sequential backbone assignment. The CBCANH spectra correlate the ^{1}H and ^{15}N amide resonances with those of the intra- and inter-residue $^{13}C\alpha$ and $^{13}C\beta$ resonances (**Figure 5.6**) (Grzesiek and Bax, 1992b), whereas the CBCA(CO)NH spectra correlate the ^{1}H and ^{15}N amide resonances of one residue with both $^{13}C\alpha$ and $^{13}C\beta$ resonances of

its preceding residue *via* the intervening ^{13}CO spin (exclusive inter-residue connectivities) (**Figure 5.7**) (Grzesiek and Bax, 1992a). These two experiments provide not only complete sequential assignments but also assignment of side-chain carbons (Cβ), which is useful for the complete assignment of aliphatic resonances. The HNCO spectra correlate the ^{1}H and ^{15}N amide resonances of one residue with the carbonyl (^{13}CO) resonance of the preceding residue (**Figure 5.8**) (Kay et al., 1990). The advantage of using triple resonance experiments is the simplicity and accuracy of assignment since the spectra show a relatively broad spread of the Cα (~45-65 ppm) and Cβ (~25-45 ppm, ~60-70 ppm for threonine and serine) resonances (non-overlapped spectra).

Since the assignments of Cα and Cβ have been obtained by the backbone assignment, the side-chain resonances can be assigned by transferring the magnetisation from side-chain protons to backbone amide spins. Complete assignments were obtained with 3D C-DIPSY-(CO)NH (**Figure 5.9**) (Grzesiek et al., 1993; Logan et al., 1993), 3D HCCH-COSY (**Figure 5.10**) (Ikura et al., 1991), and 2D ^{1}H-^{13}C HSQC (Cavanagh et al., 2007).

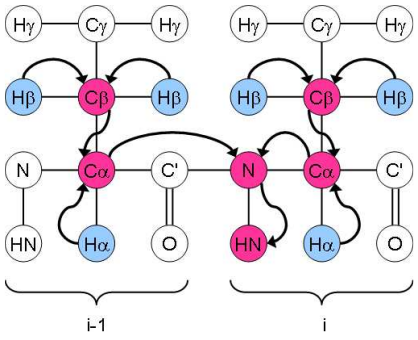


Figure 5.6: Magnetisation transfer pathways (black arrows) for the CBCANH triple-resonance experiment (http://www.protein-nmr.org.uk/spectra.html). The experiment correlates the amide ^{15}N and ^{1}H resonances with $^{13}C\alpha$ and $^{13}C\beta$ of the same and the preceding residue. The blue atoms act as relay nuclei, so their frequencies are not detected. Only frequencies of pink atoms are detected.

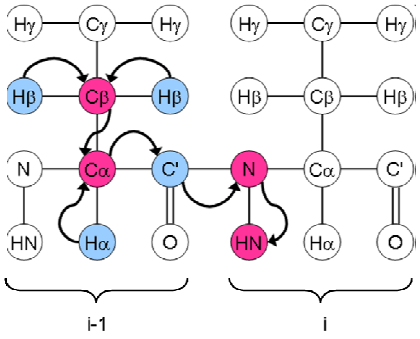


Figure 5.7: Magnetisation transfer pathways (black arrows) for the CBCA(CO)NH triple-resonance experiment (http://www.protein-nmr.org.uk/spectra.html). The experiment correlates the amide ^{15}N and ^{1}H resonances with ^{13}Cα and ^{13}Cβ of the preceding residue. This provides confirmation of the inter-residue correlations in the CBCANH experiment. The blue atoms act as relay nuclei, so their frequencies are not detected. Only frequencies of pink atoms are detected.

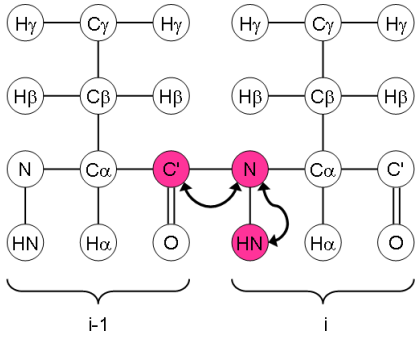


Figure 5.8: Magnetisation transfer pathways (black arrows) for the HNCO triple-resonance experiment (http://www.protein-nmr.org.uk/spectra.html). The experiment correlates sequence backbone connectivities between the ^{15}N-^{1}H pair of one residue with the attached carbonyl (^{13}CO) resonance of the preceding residue. The frequencies of pink atoms are detected.

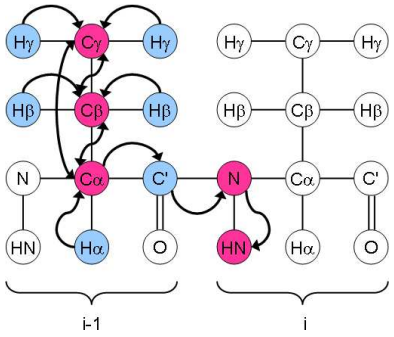


Figure 5.9: Magnetisation transfer pathways (black arrows) for the C-DIPSY-(CO)NH triple-resonance experiment (http://www.protein-nmr.org.uk/spectra.html). The experiment correlates the amide ^{15}N and ^{1}H resonances with ^{13}Cα and all side-chain ^{13}C resonances of the preceding residue. The blue atoms act as relay nuclei, so their frequencies are not detected. Only frequencies of pink atoms are detected.

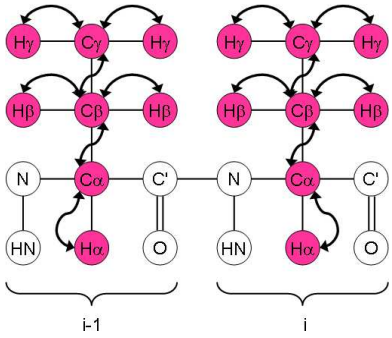


Figure 5.10: Magnetisation transfer pathways (black arrows) for the HCCH-COSY triple-resonance experiment (http://www.protein-nmr.org.uk/spectra.html). The experiment establishes correlations of ^{1}H and ^{13}C resonances within H-C-C-H molecular fragments. Only the hydrogen resonances of the own and neighbouring carbons are visible. The frequencies of pink atoms are detected.

5.2.6 Protein-ligand interactions by ^{1}H-^{15}N HSQC experiments

About 0.5 mg of U-^{15}N labelled NT was dissolved in 0.60 mL of GF2 buffer with 5% (*v/v*) glycerol and 5% (*v/v*) pure D_2O, pH 7.4 and loaded into a 5-mm NMR tube as a negative control. To probe the NTS1-NT interaction, approximately 10.5 nmol of active TEV-cleaved NTS1B in 0.30 mL of GF2 buffer with 5% (*v/v*) glycerol and 5% (*v/v*) pure D_2O, pH 7.4 was incubated with 9.8 nmol of U-^{15}N labelled NT for 1 h and then transferred to a Shigemi NMR microtube, which is suitable for small volume samples with high concentration, leading to measurements with improved signal-to-noise ratios and greatly reduced spectrum acquisition times. The NMR experiments were carried out on an Oxford Instruments 950 MHz NMR spectrometer with a GE/Omega computer (IDC group, Oxford) at 298 K. All recorded data were referenced to an external DSS standard.

^{1}H-^{15}N HSQC spectra were acquired with 128 t_1 increments using States-TPPI phase cycling. Data in t_1 was linear predicted to 256 points and data points in both dimensions were zero-filled to 1024 for NMR data processing by nmrPipe (Delaglio et

al., 1995). The spectra were analysed by Sparky (http://www.cgl.ucsf.edu/home/sparky/).

5.2.7 1D ^{13}C SSNMR experiments

All SSNMR spectra were acquired on an 11.75 T (500 MHz ^{1}H frequency) Varian/Magnex Scientific NMR spectrometer in a 4mm APEX HX probe (for precipitate) and a 4mm APEX HXY probe (for U-^{13}C, ^{15}N labelled NT). Ramped-amplitude cross polarisation (CP) (Metz et al., 1994) was performed with contact time 1 ms and two-pulse phase-modulated decoupling (Bennett et al., 1995) during the acquisition period. Dwell was 10 μs and spectral width was 100 kHz in all experiments. All the spectra were referenced externally to ^{13}C adamantane methylene peak (Morcombe and Zilm, 2003).

i) semi-SSNMR experiments at 278 K

One-dimensional ^{13}C spectra of U-^{13}C, ^{15}N labelled NT in BPL were recorded using a CP MAS pulse sequence (**Figure 5.11**) and a single pulse sequence with proton decoupling (direct excitation or 1pda) (**Figure 5.12**). The spectra were acquired at 278 K and MAS frequency $\omega_R/2\pi$ of 6 kHz. For the CP experiment, the ^{1}H and ^{13}C field strengths were 55 kHz and 33.3 kHz during CP, respectively. TPPM proton decoupling of 71 kHz was applied during the 20.48 ms ^{13}C acquisition time. The pulse delay was 3 s and 12 k scans were collected. For the 1pda experiment, TPPM proton decoupling of 40.2 kHz was applied during the 40.96 ms ^{13}C acquisition time. The pulse delay was 3 s and 2 k scans were collected.

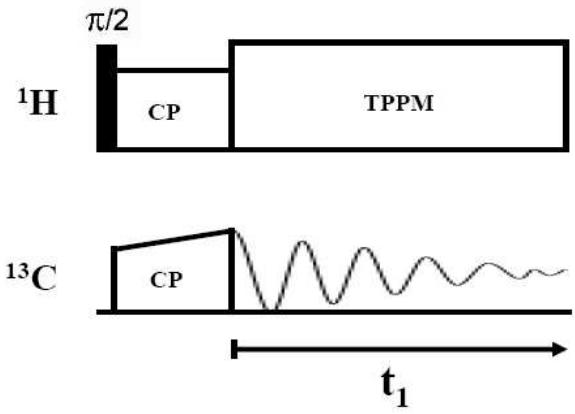


Figure 5.11: CP pulse sequence with two-pulse phase-modulated (TPPM) decoupling scheme during acquisition.

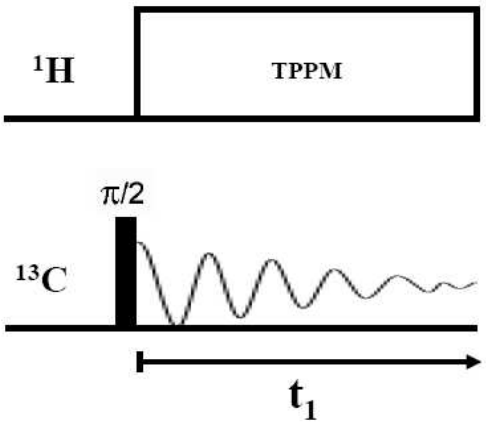


Figure 5.12: Pulse sequence with two-pulse phase-modulated (TPPM) decoupling scheme during acquisition for the direct excitation ^{13}C experiment (1pda).

ii) ^{13}C CP MAS NMR at lower than 273 K

Solid samples of NT_{8-13} precipitate, U-^{13}C, ^{15}N labelled NT in BPL environment and U-^{13}C, ^{15}N labelled NT with reconstituted NTS1B receptor were used to acquire 1D ^{13}C spectra. The pulse sequence for CP MAS experiments is shown in **Figure 5.11**. The ^{13}C CP MAS spectrum of NT_{8-13} precipitate was recorded at 258 K and the sample was spun at a MAS frequency of 8 kHz. The ^{1}H and ^{13}C field strengths

were 68 kHz and 43.8 kHz during CP, respectively. TPPM proton decoupling of 68 kHz was applied during the 20.48 ms ^{13}C acquisition time. The pulse delay was 3 s and 64 k scans were collected.

For the ^{13}C CP MAS spectrum of U-^{13}C, ^{15}N labelled NT in BPL environment, the experiment was carried out at the temperature of 233 K and MAS frequency $\omega_R/2\pi$ of 6 kHz. The ^{1}H and ^{13}C field strengths were 55 kHz and 33.3 kHz during CP, respectively. TPPM proton decoupling of 71 kHz was applied during the 20.48 ms ^{13}C acquisition time. Recycle delays of 3 s and 8192 scans were employed. For the ^{13}C CP MAS spectrum of U-^{13}C, ^{15}N labelled NT with reconstituted NTS1B receptor, the experiment was carried out at the temperature of 233 K and the sample was rotated at a MAS frequency of 6 kHz. The ^{1}H and ^{13}C field strengths were 58 kHz and 34.7 kHz during CP, respectively. TPPM proton decoupling of 75.8 kHz was applied during the 20.48 ms ^{13}C acquisition time. The pulse delay was 3 s and 19704 scans were collected.

5.2.8 2D ^{13}C-^{13}C homonuclear SSNMR experiments

i) ^{13}C-^{13}C COSY semi-SSNMR experiment at 278 K

Carbon correlations of U-^{13}C, ^{15}N labelled NT in BPL were detected from the COSY experiment at 278 K. The COSY pulse sequence (**Figure 5.13**) is the same as in solNMR but the experiment was performed on a 500 MHz Varian/Magnex SSNMR spectrometer (4mm APEX HXY probe) under TPPM proton decoupling during acquisition. Adamantane was used as a chemical shift reference. Dwell was 25 μs and spectral width was 40 kHz. The spectrum was acquired at a spinning frequency $\omega_R/2\pi$ of 6 kHz and proton decoupling field strength was set to 40.2 kHz during acquisition.

There were 1200 points and 2048 points collected in direct and indirect ^{13}C dimensions for 30 ms, respectively. The pulse delay was 3 s and 32 scans were collected.

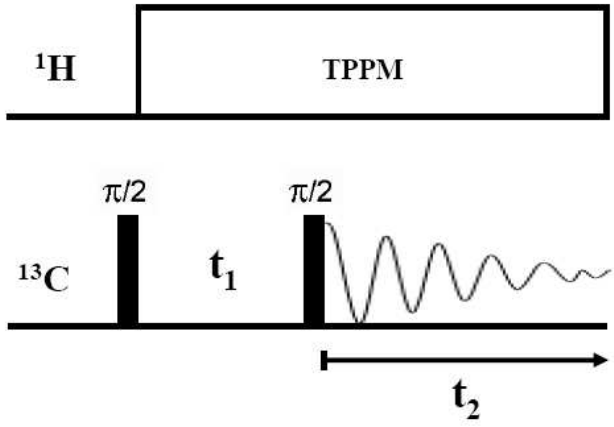


Figure 5.13: Pulse sequence with two-pulse phase-modulated (TPPM) decoupling scheme during acquisition for the COSY experiment.

ii) 2D ^{13}C-^{13}C DARR homonuclear SSNMR experiment

Another CP MAS experiment, dipolar-assisted rotational resonance (DARR), was used to recouple long-range dipolar interactions by a combination of mechanical rotation of the sample and the ^{13}C-^{1}H dipolar interaction (Takegoshi et al., 2001). The pulse sequence for DARR experiment is shown in **Figure 5.14**. Magnetisation is exchanged when a spinning side-band of one spin overlaps with the isotropic resonance or side-band of another. Irradiation of the protons at the rotational resonance condition recouples the ^{13}C-^{1}H dipolar interaction, broadening the lines in the carbon spectrum. After cross polarisation from proton magnetisations, longitudinal magnetisation of ^{13}C spin is obtained by a π/2 radio frequency (rf) pulse. The length of the mixing time determines the number of bond transfers of the magnetisation. In

DARR, reintroduction of ^{13}C-^{1}H dipolar interaction is utilised by two-pulse phase-modulated (TPPM) ^{1}H irradiation with the intensity satisfying the rotary-resonance condition $\omega_1 = n\omega_R$ (n = 1 or 2), where ω_1 is the field strength of the recoupling ^{1}H pulse and ω_R is the rotor spinning frequency. In the end, the magnetisations are converted to observable (-1)-quantum coherences by the last $\pi/2$ rf pulse on the ^{13}C channel.

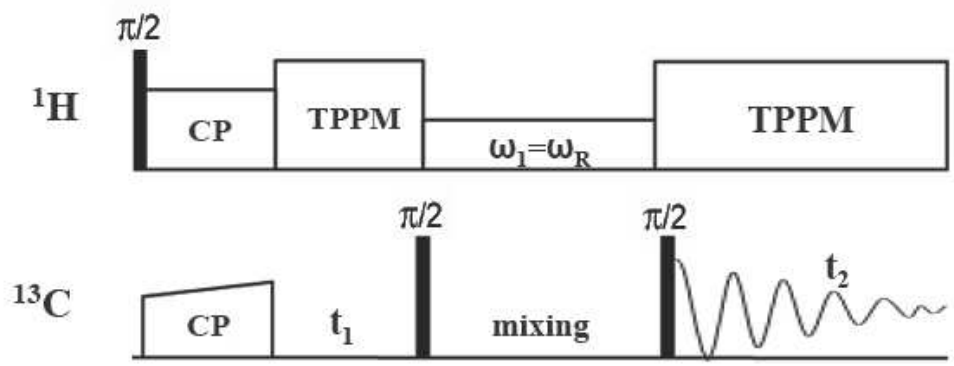


Figure 5.14: Pulse sequence for the DARR experiment. The sequence begins with variable amplitude CP using a ramped pulse on the ^{13}C channel. After the t_1 evolution period, magnetisation is placed along the z-axis with a 90° pulse and mixing occurs longitudinally assisted by a low power ^{1}H 'recoupling' pulse in DARR. During the DARR mixing period, the ^{1}H radio frequency (rf) field strength is set to the $n = 1$ rotational resonance condition. TPPM decoupling is used during acquisition.

U-^{13}C, ^{15}N labelled NT in the presence of reconstituted neurotensin receptor 1 to study protein-ligand interactions was used to acquire a 2D spectrum. The 2D ^{13}C-^{13}C homonuclear DARR spectrum was recorded at 11.75 T, on a 500 MHz (^{1}H frequency) Varian/Magnex Scientific NMR spectrometer (4mm APEX HXY probe) with adamantane as a chemical shift reference. Dwell in direct and indirect ^{13}C dimensions were 10 μs and 13.88 μs, respectively and spectral width was 100 kHz. The rotor was spun pneumatically at 6 kHz at the magic angle. The DARR mixing time was carried out with n = 1 rotary resonance condition for the protons, and a

mixing time of 13 ms in order to identify 2-3 bond transfers. During the whole experiment, the temperature was maintained at 233 K. Cross polarisation was achieved using a 34.7 kHz square ^{1}H pulse at 500.10925 MHz and a 33-56 kHz tangent ramped ^{13}C pulse at 125.76449 MHz for 0.8 ms. There were 1600 points and 32 points collected in direct and indirect ^{13}C dimensions, respectively. The FID was acquired for 16 ms and 2.2 ms in direct and indirect ^{13}C dimensions, respectively. TPPM proton decoupling of 75.8 kHz was applied during acquisition. The recycle delay was 2.5 s, and there were 1 k scans collected for a total of 115 h.

5.2.9 Data processing for SSNMR data

All 1D spectra were processed using sinebell square (cosine shifted) apodisation and zero filled to 32 K points before Fourier transform in Spinsight (Varian). Signal-to-noise and integral measurements were also carried out with Spinsight. The 2D spectrum was processed using sinebell apodisation function, then zero filled to 4 k points in both the direct and indirect dimensions before being Fourier transformed using nmrPipe (Delaglio et al., 1995). The 2D spectrum was analysed by Sparky software (http://www.cgl.ucsf.edu/home/sparky/).

5.3 Results and discussion

5.3.1 Crystallisation of NT_{8-13} peptide for SSNMR experiments

Since the lyophilised sample generally gives broad lines from SSNMR experiments, the attempt was made to crystallise the NT_{8-13} peptide. Two concentrations of peptide NT_{8-13} solution, 50 and 100 mg/mL were tested for the screen crystallisation kit at 15 °C. The precipitants 5, 7, 14, 27-30, and 32 gave

precipitates with both concentrations, while the precipitating agents 8, 9, 25, 34, 35, 42, and 44 only gave precipitates with the concentration of 100 mg/mL. The list of successful precipitants is shown in **Table 5.1**. The precipitants 9, 25 and 42, which gave a reasonable precipitate were chosen for more screening and tested in larger volume in the peptide concentration of 100 mg/mL. The precipitant 9 (20% (*v/v*) PEG 4000, 50 mM sodium citrate pH 6.0 and 20% (*v/v*) isopropanol) was selected as the final condition for crystallisation of NT_{8-13} as it showed large tolerance for changes in the screen and large quantity of the precipitate in the total volume of 80 µL.

Table 5.1: The list of successful precipitants for unlabelled NT_{8-13} from a crystallisation screening kit (Structure Screen 1 - Catalogue Number MD1-01).

Precipitant no.	Mixture solution		
5	None	0.1M Na acetate trihydrate pH 4.6	2.0M Ammonium sulphate
7	0.2M Ammonium acetate	0.1M tri-sodium citrate dihydrate pH 5.6	30% w/v PEG 4000
8	0.2M Ammonium acetate	0.1M tri-sodium citrate dihydrate pH 5.6	30% v/v 2-methyl-2,4-pentanediol
9	None	0.1M tri-Sodium citrate dihydrate pH 5.6	20% w/v 2-propanol, 20% w/v PEG 4000
14	0.2M Ammoniium sulphate	0.1M Na Cacodylate pH 6.5	30% w/v PEG 8000
25	0.2M tri-sodium citrate dihydrate	0.1M Na Hepes pH 7.5	20% v/v 2-propanol
27	None	0.1M Na Hepes pH 7.5	1.5M Lithium sulphate monohydrate
28	None	0.1M Na Hepes pH 7.5 0.8M K dihydrogen phosphate monohyd.	0.8M Na dihydrogen phosphate
29	None	0.1M Na Hepes pH 7.5	1.4M tri-Sodium citrate dihydrate
30	None	0.1M Na Hepes pH 7.5	2% v/v PEG 400, 2.0M Amm sulphate
32	None	0.1M Tris HCl pH 8.5	2.0M Ammoniium sulphate
34	0.2M tri-sodum citrate dihydrate	0.1M Tris HCl pH 8.5	30% v/v PEG 400
35	0.2M Lithium sulphate monohydrate	0.1M Tris HCl pH 8.5	30% w/v PEG 4000
42	0.2M Ammonium sulphate	None	30% w/v PEG 8000
44	None	None	2.0M Ammonium sulphate

The crystallisation of NT_{8-13} solution was initially performed in an Eppendorf tube. The precipitate needs to be transferred from the Eppendorf tube to a sealed yellow pipette tip for rotor packing for SSNMR experiments. It was found that the precipitate was sensitive to resuspension during the transfer, and so the crystallisation was set up in the sealed yellow pipette tip instead. Fortunately, a larger amount of precipitate was found in the pipette tip compared to the Eppendorf tube when starting from the same volume. This was measured by the peptide concentration of the supernatant using absorbance at 280 nm. It might be because of the contact preference as the precipitate on the surface of the pipette tip was observed. The temperature of peptide crystallisation was also optimised. The NT_{8-13} peptide precipitated better (in term of yield) at 4 °C than at 15 °C in the pipette tip. A number of crystallisation setup of unlabelled NT_{8-13} using the precipitating agent 9 was performed in the sealed yellow pipette tip at 4 °C and the correct molecular weight of NT_{8-13} precipitate was verified by ESI mass spectrometry (**Figure 5.15**).

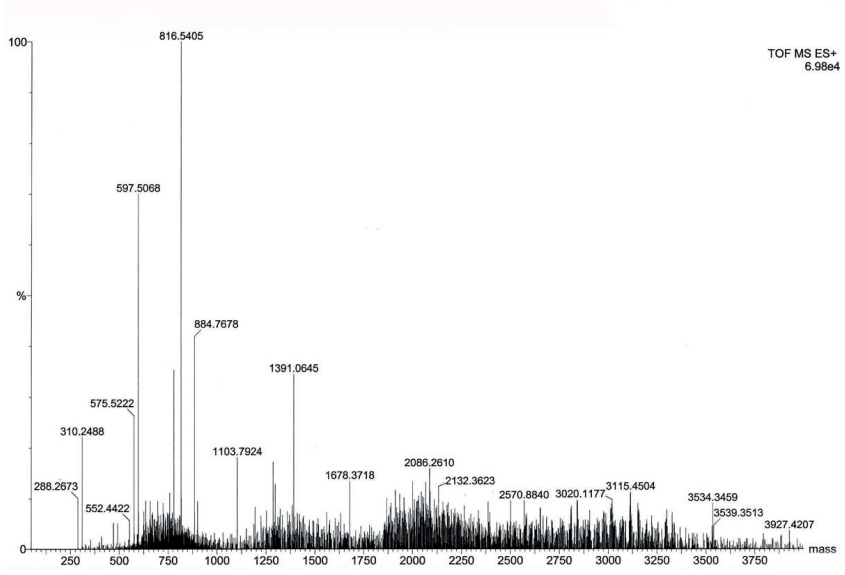


Figure 5.15: Positive-ion mass spectrum of precipitated unlabelled NT_{8-13}. The major peaks at *m/z* 816.5405 $(M+H)^+$ correspond to the precipitated NT_{8-13}.

5.3.2 Resonance assignments for unlabelled NT_{8-13} in D_2O and precipitant

Solution-state NMR techniques were employed to determine whether the crystallisation solution affects or induces the conformation of NT_{8-13} peptide. The 2D NMR spectra for the NT_{8-13} peptide in pure D_2O, pH 6.0 were analysed by the program Sparky. In TOCSY spectrum (Appendix D.1), the mixing time of 80 ms was added to ensure that all or most of the protons involved in each individual spin system are detected. The ROESY spectrum (Appendix D.2) provides complementary information necessary to complete all resonance assignments. The assignment results are summarised in **Table 5.2 and 5.3**. Here, the ^{1}H and ^{13}C chemical shifts of NT_{8-13} were assigned from solNMR experiments for the first time and the ^{13}C assignments correlate with those previously obtained from SSNMR experiments (Williamson et al., 2001) (see **Table 5.4** for the summary of ^{13}C chemical shifts of NT peptides in the free and bound states from SSNMR studies). The ^{1}H assignments are based on a conventional sequential homonuclear assignment protocol for unlabelled peptides. Although there are two repeated residues of Arg (Arg8 and Arg9) in the NT_{8-13} sequence, those can be differentiated and the assignment is achievable since the first and the last residues of the sequence are usually more sensitive to environments than other residues. As can be seen in **Table 5.2**, the chemical shift of Hα of Arg8, the first residue in the sequence, is significantly lower than that of Arg9 and averaged chemical shift for Arg (BMRB database, http://www.bmrb.wisc.edu/). Assignments of Arg9 ^{1}H chemical shifts were confirmed by the strong ROESY crosspeaks with the Hα resonance of the preceding residue Arg8. The ^{13}C chemical shifts were subsequently assigned from the ^{1}H-^{13}C HSQC spectrum (**Figure 5.16**), which has no overlapping crosspeaks. Based on the chemical shift dispersion (Wang and Jardetzky, 2002), the

NT_{8-13} peptide in the absence of its receptor has no preferred conformation, in agreement with other NMR studies (Luca et al., 2003; Nieto et al., 1986).

Table 5.2: ^{1}H-chemical shifts (ppm) for unlabelled NT_{8-13} residues at 298 K in an Oxford Instruments 750 MHz NMR spectrometer (pure D_2O, pH 6.0).

Residue	Hα	Hβ	Hγ	Hδ	Hε
Arg8	3.878	1.809	1.568	2.902	-
Arg9	4.561	1.769	1.642	3.113	-
Pro10 (T)[a]	4.324	2.193 / 1.765	1.940	3.568 / 3.77	-
Pro10 (C)[b]	4.434	2.236 / 1.830	2.028	3.392	-
Tyr11	4.497	2.922	-	7.047	6.754
Ile12	4.042	1.724	1.034 / 1.335 / 0.982	0.891	-
Leu13	4.193	1.531	1.538	0.862 / 0.834	-

[a] (T) = *trans*, [b] (C) = *cis*

Table 5.3: ^{13}C-chemical shifts (ppm) for unlabelled NT_{8-13} residues at 298 K in an Oxford Instruments 750 MHz NMR spectrometer (pure D_2O, pH 6.0).

Residue	Cα	Cβ	Cγ	Cδ	Cε
Arg8	52.66	28.70	24.39	40.57	-
Arg9	51.49	27.35	24.03	40.61	-
Pro10 (T)[a]	60.43	29.14	24.65	47.94	-
Pro10 (C)[b]	60.03	-	-	47.78	-
Tyr11	54.93	36.00	-	130.50	115.50
Ile12	57.85	36.39	24.38 / 14.61	9.98	-
Leu13	52.16	39.73	23.55	22.21 / 21.02	-

[a] (T) = *trans*, [b] (C) = *cis*

Table 5.4: Summary of ^{13}C-chemical shifts (ppm) for neurotensin peptides with and without the receptor, NTS1, from solid-state NMR experiments.

Nuclei	Sample	Arg8	Arg9	Pro10	Tyr11	Ile12	Leu13	Ref.
	NT(9–13)		55.1	66.1	60.1	62.6	56.1	A
	lyophillised NT(8-13)	53.3	52.6	61.1	52.8	57.4	53.4	B
	NT(8–13)	54.95	53.73	62.64	57.05	60.25	55.65	C
		58.1	56.1	64.1	60.1	61.6	58.1	A
	NT(8–13)/NTS–1A	58.6	56.6	64.1	58.1	61.1	57.1	A
Cα	NT(8–13)/NTS–1B	60.1	56.1	62.6	58.6	55.6	56.6	A
	NT(9–13)		31.1	33.1	41.6	40.1	43.6	A
	lyophillised NT(8-13)	28.9	28.9	29.5		36.8	38.7	B
	NT(8–13)	31.14	30.39	31.17	38.57	38.46	43.26	C
		33.6	32.1	34.1	41.1	40.1	44.1	A
	NT(8–13)/NTS–1A	32.6	32.6	34.1	41.1	41.1	43.1	A
Cβ	NT(8–13)/NTS–1B	34.1	32.1	36.6	43.6	40.6	46.1	A
	lyophillised NT(8-13)	24.5	24.5	24.1		14.2	23.7	B
	NT(8–13)	27.09	27.09	26.51		17.21	27.09	C
Cγ1	NT(8–13)/NTS–1A					17.32		C
	lyophillised NT(8-13)					25.6		B
Cγ2	NT(8–13)					26.03		C
	lyophillised NT(8-13)	40.9	40.9	48		11	18-27	B
	NT(8–13)	43.26	43.26	60.73		12.56	23.67	C
Cδ	NT(8–13)/NTS–1A					12.63	23.94	C
	lyophillised NT(8-13)						18-27	B
Cδ	NT(8–13)						24.89	C
Cε	NT(8–13)	159.27	159.27					C
	lyophillised NT(8-13)	169.6		172.3	174.2	172.3	172.5	B
	NT(8–13)	172.3	174.11	175.59	174.92	174.63	180.92	C
CO	NT(8–13)/NTS–1A	173.81					181.35	C
Cζ	lyophillised NT(8-13)	157.6	157.6					B
	lyophillised NT(8-13)				127.9			B
aromatic	NT(8–13)				130.13			C
ring-C1	NT(8–13)/NTS–1A				130			C
	lyophillised NT(8-13)				130.7			B
aromatic	NT(8–13)				133.31			C
ring-C2,6	NT(8–13)/NTS–1A				132.69			C
	lyophillised NT(8-13)				115.7			B
aromatic	NT(8–13)				118.01			C
ring-C3,5	NT(8–13)/NTS–1A				117.86			C
	lyophillised NT(8-13)				155.6			B
aromatic	NT(8–13)				157.07			C
ring-C4	NT(8–13)/NTS–1A				157.27			C

Description:

NT(9-13): free uniformly ^{13}C, ^{15}N labelled NT_{9-13} in detergent buffer
NT(8-13): free uniformly ^{13}C, ^{15}N labelled NT_{8-13} in detergent buffer
NT(8-13)/NTS-1A: bound uniformly ^{13}C, ^{15}N labelled NT_{8-13} with detergent solubilised NTS1
NT(8-13)/NTS-1B: bound uniformly ^{13}C, ^{15}N labelled NT_{8-13} with reconstituted NTS1

References: A (Luca et al., 2003); B (Williamson et al., 2001); C-the values are referenced to DSS adapted from (Williamson et al., 2002).

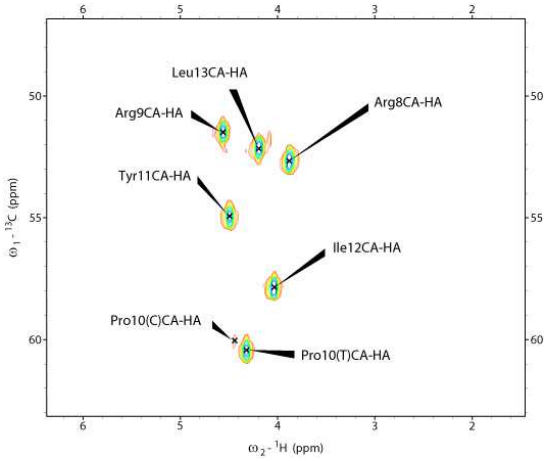


Figure 5.16: ^{1}H-^{13}C HSQC spectrum of NT_{8-13} peptide in aqueous solution (pure D_2O, pH 6.0), ^{1}Hα-^{13}Cα region. The spectrum was recorded on a Bruker 500 MHz NMR spectrometer at 298 K.

There is one extra set of minor signals detected in the TOCSY and ROESY spectra. Since the purity of the sample was checked by HPLC and ESI mass spectrometry, these resonances were ascribed to the presence of two isomers of the Pro10 residue. It has been known that proline in the *trans* conformation is more stable than the *cis* conformation (Wedemeyer et al., 2002). Therefore, the major resonances of Pro10 must correspond to the *trans* form and the minor signals are from the *cis* form. The ratio of *trans*/*cis* conformation calculated from the intensity of signals is 4.2. The presence of *cis*/*trans* isomerisation of Pro10 residue in the NT_{8-13} sequence was already observed in the conformational studies of NT_{8-13} in aqueous solution by ^{1}H NMR (Nieto et al., 1986). In principle, this observation can be confirmed by the chemical shift differences of Cβ and Cγ of the Pro ring (Siemion et al., 1975) but unfortunately, in the ^{1}H-^{13}C HSQC spectrum, crosspeaks of Pro in *cis* form cannot be clearly assigned because of resonance and crosspeak overlap with the exception of Cα (60.03 ppm) and Cδ (47.78 ppm).

The 2D TOCSY and ROESY spectra of NT_{8-13} peptide in the precipitant 9 prepared with pure D_2O, 5% (*v/v*) H_2O showed very small chemical shift changes (< 0.02 ppm) compared to those of the NT_{8-13} peptide in pure D_2O, pH 6.0 suggesting that the NT_{8-13} peptide can be precipitated in precipitant 9 without any effects on the peptide conformation and the peptide has no interaction with the precipitant. Therefore, if a certain conformation of the NT_{8-13} peptide in the presence of the neurotensin receptor was observed, it will be from the binding, not from the chemical environments.

Figure 5.17 shows proton correlations in the fingerprint region (HN-Hα) of TOCSY of the NT_{8-13} peptide in the precipitant 9 prepared with pure D_2O, 5% (*v/v*) H_2O and their assignments are summarised in **Table 5.5**. These correlations were observed in this solvent because of the exchange between the amide protons in peptides and solvent water protons. There is more than one resonance for some amino acids of NT_{8-13} peptide indicating sample heterogeneity in the solvent which was also observed in aqueous solution (see **Figure 3.17** in section 3.3.7). There is no HN-Hα crosspeak of Arg8 due to fast exchange with solvent. Proline has no amide proton, but its Hα and side-chain proton resonances are readily assigned from other regions of the TOCSY spectrum.

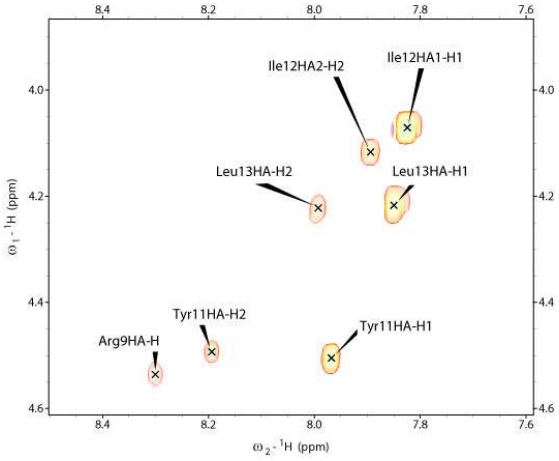


Figure 5.17: Amide proton correlations (HN-Hα) in the TOCSY spectrum of NT_{8-13} peptide in precipitant 9 prepared with pure D_2O, 5% (*v/v*) H_2O suggesting that the sample is not homogeneous. The spectrum was recorded on an Oxford Instruments 750 MHz NMR spectrometer at 298 K.

Table 5.5: ^{1}HN-chemical shifts (ppm) for unlabelled NT_{8-13} residues at 298 K in an Oxford Instruments 750 MHz NMR spectrometer (precipitant 9 prepared with pure D_2O, 5% (*v/v*) H_2O).

Residue	HN
Arg8	-
Arg9	8.320
Pro10 (T)[a]	-
Pro10 (C)[b]	-
Tyr11	7.965 / 8.191
Ile12	7.830 / 7.891
Leu13	7.847 / 7.978

[a] (T) = *trans*, [b] (C) = *cis*

5.3.3 Effects of NT_{8-13} precipitate on ^{13}C SSNMR linewidths

The obtained precipitate was then examined for the effect of crystallisation on ^{13}C SSNMR linewidths. Unfortunately, the SSNMR spectrum has indicated that the precipitate sample is not homogeneous as the resonance broadening was observed. The problem could have been sample packing because the sample could have dehydrated. Although it is unlikely to have been the cause since the sample was precipitated in the

presence of PEG 4000, a few microlitres of sterile water were added into the rotor for rehydration. The spectrum from the rehydrated sample is identical to the previous one (data not shown). From the 1D spectrum (**Figure 5.18**), the very narrow lines are solvent and chemicals from the crystallisation solution and the broad resonances are the NT_{8-13} peptide. Focusing on the chemical shifts of Ile side-chain methyl groups around 10-20 ppm, there are at least four conformations present. These results have revealed that the precipitate from the precipitating agent 9 did not yield high spectral resolution. The choice of precipitant can strongly influence the apparent linewidth as illustrated in **Figure 5.2** (Igumenova et al., 2004), so the spectral linewidth of neurotensin might be improved by different precipitating agents.

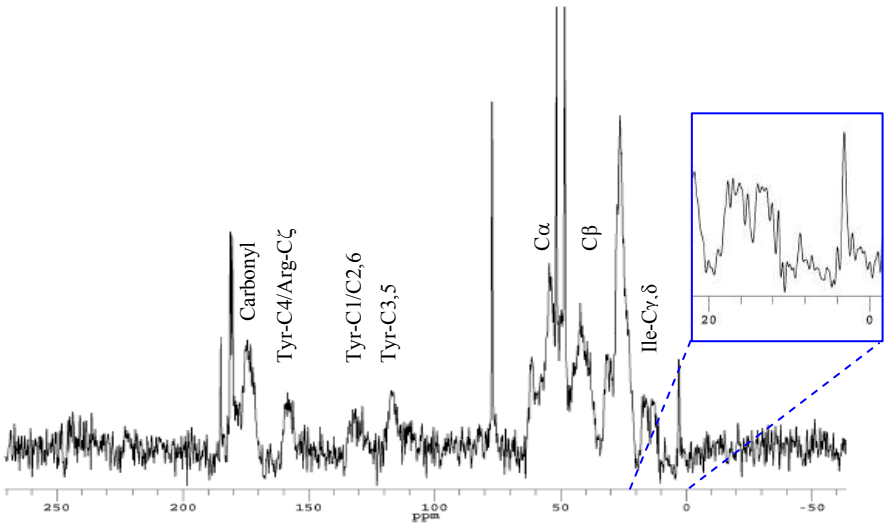


Figure 5.18: 1D ^{13}C CP MAS spectrum of precipitated unlabelled NT_{8-13}. The spectrum was recorded on a Varian 500 MHz SSNMR spectrometer. The region with Ile side-chain methyl carbon (around 10-20 ppm) is shown in larger scale. Conditions: 8 kHz, 258 K, 64 k scans and 20.48 ms acquisition time.

5.3.4 Sequential assignments for uniformly ^{13}C, ^{15}N labelled NT

Spectral resonances must be identified before a protein structure can be determined. The NMR assignment may be limited by overlapping signals leading to the difficulty in finding the exact location of the adjacent peak centres. Also, ambiguities in the assignment of repeated residues such as Leu2-Leu13, Tyr3-Tyr11, Pro7-Pro10, and Arg8-Arg9 in the NT sequence can occur. Therefore, backbone and aliphatic side-chain assignments of U-^{13}C, ^{15}N labelled NT were achieved using a set of double and triple resonance NMR experiments described in section 5.2.5. The chemical shifts of this labelled peptide are listed in **Table 5.6 and 5.7**. These assignments correlate with those previously obtained (Nieto et al., 1986; Xu and Deber, 1991). Two backbone resonances (Cα-Hα and N-HN) were detected for Lys6, Pro7, Arg8, Arg9, Ile12, and Leu13 as shown in either the ^{1}H-^{13}C HSQC (**Figure 5.19**) or the ^{1}H-^{15}N HSQC (**Figure 5.20**) spectrum. The additional minor set of low-level resonances is referred to as xB, where x = residue number. In the ^{1}H-^{13}C HSQC spectrum (**Figure 5.19**), Cα-Hα crosspeaks of Tyr3-Tyr11, Pro7-Pro7B, Ile12-Ile12B are overlapped, but they are differentiated by the 3D CBCANH spectrum (see example in appendix D.3). Cα-Hα minor resonances of Lys6B and Leu13B might be very close to those of Lys6 and Leu13, respectively, and so they cannot be observed. In the ^{1}H-^{15}N HSQC spectrum (**Figure 5.20**), the resonances of Glu1, Leu2 and Asn5 are not detected due to rapid exchange rate with the solvent which limited the assignments of their nitrogen atoms, however signals that correlate with the backbone nitrogen atom of Asn5 are very strong in the CBCANH, CBCACONH and C-DIPSY-(CO)NH. Also, the absence of amide protons of Pro7 and Pro10 limited the assignment of their nitrogen atoms. In contrast to the ^{1}H-^{13}C HSQC spectrum, the crosspeaks of the minor

set are unambiguously assigned in the ^{1}H-^{15}N HSQC spectrum. The detected minor resonance should arise from the *cis* conformation of Pro10 since there are N_i-$C\alpha_{i-1}$-HN_i crosspeaks of Leu13B-Ile12B, Ile12B-Tyr11, Arg9B-Arg8B, Arg8B-Pro7B, and Lys6B-Asn5. This observation supports the existence of a small population of *cis* isomer of Pro10 in neurotensin in aqueous solution as already observed in previous studies by Xu and Deber (Xu and Deber, 1991).

Table 5.6: ^{15}N and ^{1}H-chemical shifts (ppm) for U-^{13}C, ^{15}N labelled NT residues at 298 K in an Oxford Instruments 600 MHz NMR spectrometer (detergent buffer, pH7.4 with 5% (*v/v*) D_2O).

Residue	N	HN	Hα	Hβ	Hγ	Hδ	Others
Glu1	-	-	3.820	1.950 / 2.253	2.258	-	-
Leu2	-	-	4.322	1.551 / 1.454	1.502	0.900 / 0.890	-
Tyr3	121.30	8.368	4.595	3.057 / 2.980	-	-	-
Glu4	122.39	8.323	4.238	1.905 / 1.997	2.183	-	-
Asn5	119.92	8.458	4.663	2.828 / 2.782	-	-	Nδ2 112.75, Hδ21/22 7.609 / 6.929
Lys6	122.51	8.112	4.620	1.823 / 1.708	1.437 / 1.330	1.686	Hε 2.993
Lys6B	120.56	8.008	-	-	-	-	-
Pro7	-	-	4.420	1.861 / 2.296	1.993	3.780 / 3.622	-
Arg8	121.96	8.488	4.305	1.797 / 1.821 / 1.767	1.665	3.209	-
Arg8B	121.85	8.413	4.388	-	-	-	-
Arg9	123.35	8.330	4.619	1.793 / 1.705	1.657	3.135	-
Arg9B	121.62	8.167	4.177	-	-	-	-
Pro10	-	-	4.379	1.815 / 2.225	1.993	3.780 / 3.622	-
Tyr11	119.68	8.052	4.595	3.066 / 2.971	-	-	-
Ile12	123.20	7.961	4.151	1.822	0.896 / 1.121 / 1.419	0.847	-
Ile12B	122.44	7.981	-	-	-	-	-
Leu13	132.19	7.784	4.182	1.579	1.581	0.934 / 0.850	-
Leu13B	131.93	7.827	-	-	-	-	-

Table 5.7: ^{13}C-chemical shifts (ppm) for U-^{13}C, ^{15}N labelled NT residues at 298 K in an Oxford Instruments 600 MHz NMR spectrometer (detergent buffer, pH7.4 with 5% (*v/v*) D_2O).

Residue	Cα	Cβ	Cγ	Cδ	Cε	CO
Glu1	56.05	31.52	35.91	-	-	-
Leu2	55.36	42.62	26.92	24.64 / 24.00	-	176.80
Tyr3	57.75	38.76	-	-	-	175.68
Glu4	56.52	30.63	36.34	-	-	-
Asn5	53.30	38.84	177.20	-	-	174.69
Lys6	54.12	32.70	24.64	29.22	42.23	-
Pro7	63.11	32.21	27.39	50.72	-	176.87
Pro7B	63.17	-	-	-	-	176.50
Arg8	56.06	30.91	27.17	43.44	-	176.20
Arg8B	55.68	31.45	27.08	43.65	-	175.36
Arg9	53.88	30.33	26.82	43.41	-	-
Arg9B	54.23	-	-	-	-	-
Pro10	63.28	32.03	27.34	50.68	-	176.19
Tyr11	57.65	38.49	-	-	-	175.22
Ile12	60.96	38.96	27.11 / 17.45	12.48	-	174.76
Ile12B	60.87	-	-	-	-	-
Leu13	56.89	43.57	27.41	25.14 / 23.73	-	-

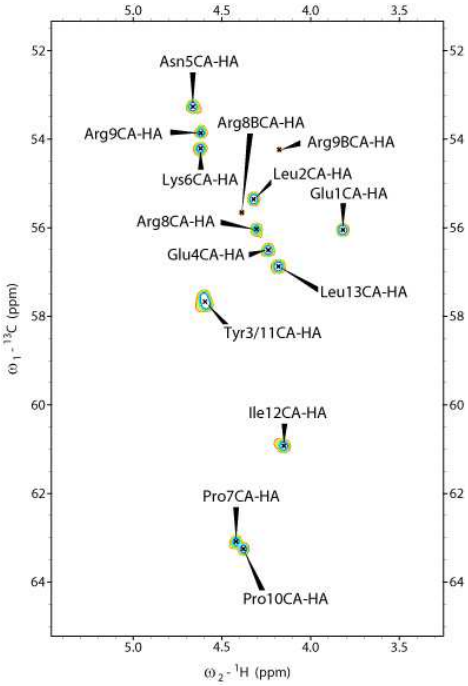


Figure 5.19: ^{1}H-^{13}C HSQC spectrum of U-^{13}C, ^{15}N labelled NT in aqueous solution (detergent buffer, pH 7.4 with 5% (*v/v*) D_2O), ^{1}Hα-^{13}Cα region. The spectrum was recorded on an Oxford Instruments 600 MHz NMR spectrometer at 298 K.

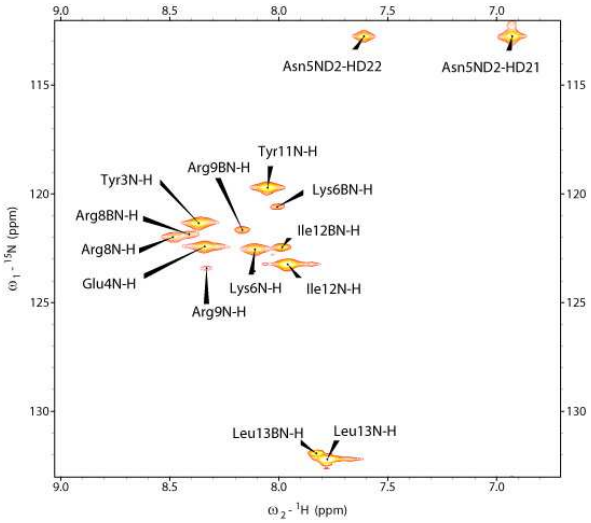


Figure 5.20: ^{1}H-^{15}N HSQC spectrum of U-^{13}C, ^{15}N labelled NT in aqueous solution (detergent buffer, pH 7.4 with 5% (*v/v*) D_2O). The spectrum was recorded on an Oxford Instruments 600 MHz NMR spectrometer at 298 K.

5.3.5 Studies of NTS1-NT interactions by solution-state NMR spectroscopy

Small molecule ligands tumble and diffuse much more rapidly when they are free in solution compared with when they are bound to a protein. Upon the binding, the ligand resonances shift or broaden selectively and these chemical shift changes can be used to identify residues involved in the protein-ligand interaction. If focusing on the protein, the resonance shifts facilitate the location of the ligand to the protein binding site. The resulting chemical shift perturbations can now be used as restraints in data-driven docking algorithms to generate possible structures for the biomolecular complexes. Due to the limitation of producing large amount of labelled NTS1 and large size of receptor for solNMR studies, here the NTS1-NT interaction was aimed at the conformation of receptor-bound ligand by using U-^{15}N labelled NT and unlabelled TEV-cleaved NTS1B.

NTS1-NT interaction surfaces were investigated by ^{1}H-^{15}N HSQC experiments at 298 K. The HSQC spectrum of U-^{15}N labelled NT was first recorded for a negative control (**Figure 5.21** red peaks). The unlabelled TEV-cleaved NTS1B was incubated with the U-^{15}N labelled NT for 1 h before the HSQC experiment was performed to probe the protein-ligand interactions. As shown in **Figure 5.21**, the HSQC spectrum of free ligand (red peaks) was overlaid on the HSQC spectrum of ligand in the presence of the TEV-cleaved NTS1B receptor (blue peaks). Since the neurotensin binds strongly to its receptor (K_d ~nM), broad lines are expected to be observed. Although some resonance shifts were detected when the ligand bound to the receptor, these crosspeaks are as narrow as the ones in the free form. Upon binding, some crosspeaks were missing, but once comparison with the HSQC spectrum of U-^{13}C, ^{15}N labelled NT in **Figure 5.20**, in which those missing crosspeaks were observed, indicates that

the presence of some crosspeaks of the neurotensin is dependent on the dynamics of the ligand. Even though the half life of the receptor is about 20 h at 298 K in the presence of 10% (*v/v*) glycerol, those observed signals from the one with the receptor are unlikely to arise from the binding. The protein might aggregate during the experiment. The stability of protein at room or high temperature is very important for structural studies by solNMR. Recently, Shibata and co-workers successfully produced a thermostable NTS1 mutant (Shibata et al., 2009), which requires further characterisation.

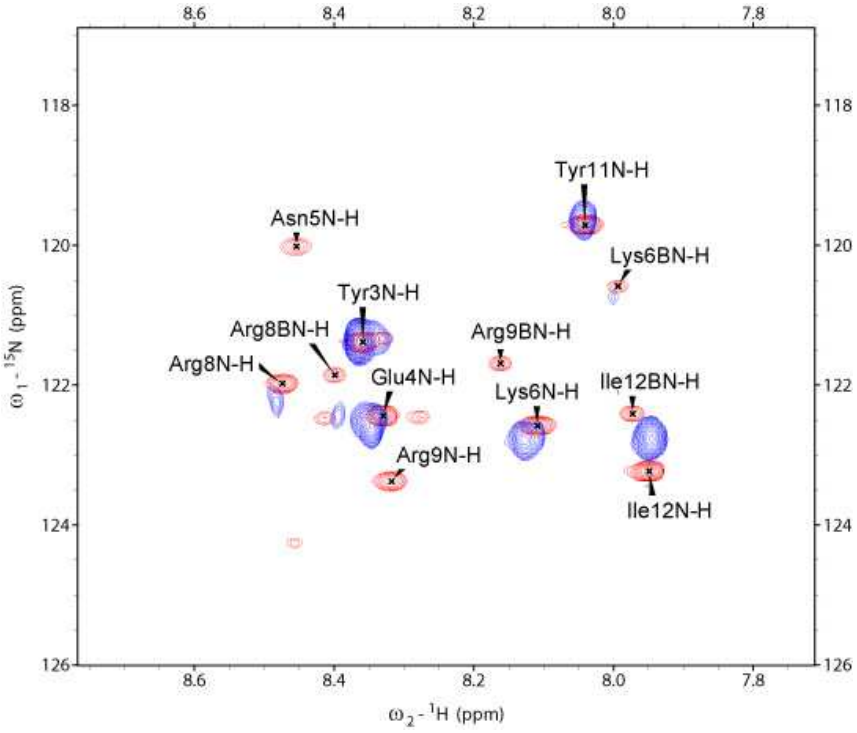


Figure 5.21: Superposition of ^{1}H-^{15}N HSQC spectra of U-^{15}N labelled NT in detergent buffer, pH 7.4 with 5% (*v/v*) D_2O (red peaks) and U-^{15}N labelled NT with unlabelled TEV-cleaved NTS1B in detergent buffer, pH 7.4 with 5% (*v/v*) D_2O (blue peaks). Red peaks were assigned by comparison with assignments reported in **Figure 5.20**. The spectra were recorded on an Oxford Instruments 950 MHz NMR spectrometer at 298 K.

5.3.6 Studies of NTS1-NT interactions by solid-state NMR spectroscopy

To study the protein-ligand interactions, the NTS1B receptor was reconstituted into brain polar lipids (BPL) as in the native form. To ensure that the ligand has no interactions with this lipid, the radioligand binding assay (see section 4.2.8) of the BPL with ^{3}H-NT was performed. There were no differences between the specific and non-specific binding, suggesting that the ligand does not bind to the lipid. The U-^{13}C, ^{15}N labelled NT in the detergent buffer, pH 7.4 (from section 5.3.4) and in the BPL were used as negative controls. The BPL for the labelled peptide in the presence as well as absence of NTS1B receptor was prepared in the same manner to provide identical condition where possible including the concentrations of residual sucrose, detergent, glycerol, and salt.

One-dimensional ^{13}C spectrum of labelled NT in the BPL was first collected by the CP MAS technique, however the sample was not rigid enough to use this method. As can be seen in **Figure 5.22** (black signals), the cross polarisation cannot efficiently enhance the signal of ^{13}C *via* the transfer of magnetisation from ^{1}H. Therefore, the 1pda sequence was used to collect the 1D spectrum of this labelled peptide instead (**Figure 5.22** (red signals)). The experiment was performed similarly to solNMR experiments but with low spinning frequency on the SSNMR magnet. The sample was in a semi-liquid form so here it is called a semi-SSNMR experiment. This 1pda sequence worked better in comparison with the CP MAS technique for this sample. The spectra are dominated by the natural abundance ^{13}C from phospholipid acyl chains between 10 and 45 ppm. The ^{13}Cα ligand resonances are observed between 50 and 70 ppm. The resonances at 65 and 75 ppm arise from natural abundance of glycerol molecules. The chemical shifts arising between 70 and 80 ppm are attributable to the

glycerol backbone of the phospholipids. The resonances between 115 and 135 ppm are derived from the ^{13}C side-chain of Tyr, but it should be noted that peaks at ~130-135 ppm could also arise from the C=C of the lipid chains. The ^{13}Cζ side-chain of Arg and Tyr residues are observed between 155 and 165 ppm. Carbonyl group of the ligand backbones and the lipid gives rise to an envelope appearing between 170 and 180 ppm. Resonances between 180 and 185 ppm arise from the side-chain of Glu. The chemical shifts of this labelled peptide in the BPL were assigned from the ^{13}C-^{13}C COSY spectrum (Appendix D.4). Resonance assignments of the labelled peptide in the BPL are identical to those in the detergent buffer, pH 7.4 (**Figure 5.23**), verifying that there are no interactions between the ligand and the BPL. These assignments will be further used as chemical shift references.

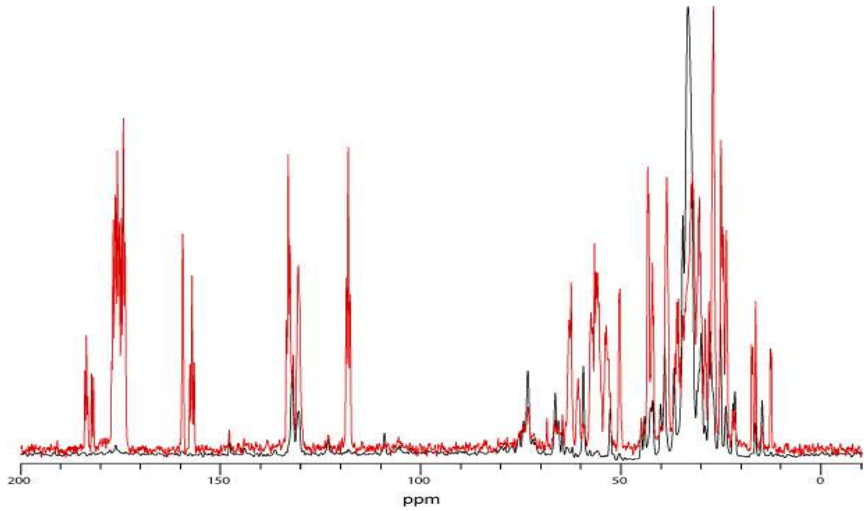


Figure 5.22: Superposition of 1D ^{13}C spectra of U-^{13}C, ^{15}N labelled NT in BPL, pH 7.4. The black signals are collected by the CP MAS technique and the red signals are recorded by the 1pda method. The spectra were recorded on a Varian 500 MHz SSNMR spectrometer at 278 K. Conditions: for 1pda, 6 kHz, 2 k scans and 40.96 ms acquisition time; for CP MAS, 6 kHz, 12 k scans and 20.48 ms acquisition time.

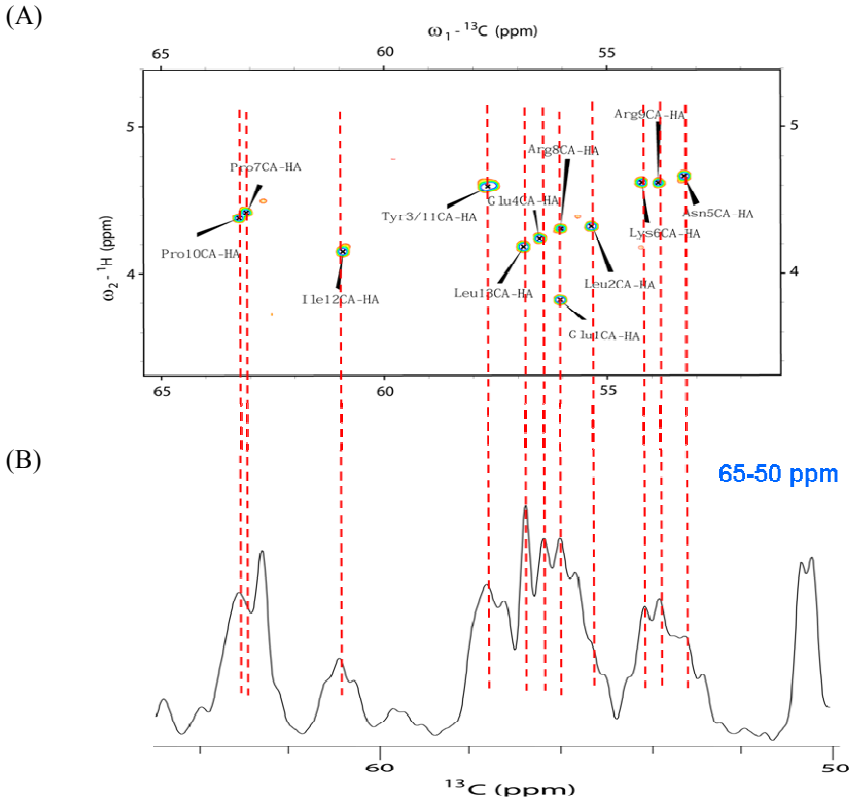


Figure 5.23: ^{13}C spectra of U-^{13}C, ^{15}N labelled NT in detergent buffer (A) and in the BPL (B), pH 7.4. The 2D HSQC spectrum was recorded on an Oxford Instruments 600 MHz NMR spectrometer at 298 K. The 1D ^{13}C spectrum was recorded on a Varian 500 MHz SSNMR spectrometer at 278 K. The dashed lines indicate the same chemical shifts of both spectra.

Chemical shift-perturbation mapping can be used to define the interaction between neurotensin ligand and the NTS1 receptor. Here, the chemical shift changes in ligand upon binding are expected to be observed. Since there is no excess of unbound ligand in the rotor, any chemical shift changes observed are likely to arise from protein-ligand interactions. In **Figure 5.24**, the ^{13}C CP MAS spectra of the labelled peptide with and without reconstituted NTS1B receptor are shown. Broad and overlapped resonances cannot be unambiguously assigned. However, a resonance at 144 ppm is observed in the spectrum of ligand with the reconstituted NTS1B receptor but not in the one without the receptor. This resonance might arise from the $^{13}C\zeta$ side-

chain of Arg and/or Tyr residue according to the binding. After adjusting the signal intensity of the one in the presence of receptor close to the one in the absence of receptor, this resonance is clearly seen. In earlier studies by our group, the chemical shift changes of carboxyl terminus and tyrosine side-chain of U-^{13}C, ^{15}N labelled NT_{8-13} were observed, suggesting the importance of these residues in protein-ligand interactions (Williamson et al., 2002).

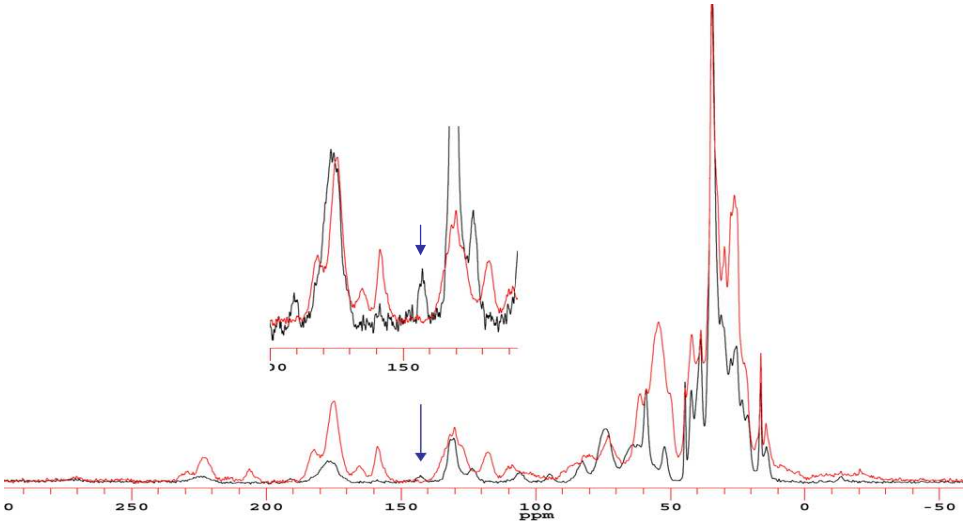


Figure 5.24: Superposition of 1D CP MAS ^{13}C spectra of U-^{13}C, ^{15}N labelled NT in BPL (red signals) and U-^{13}C, ^{15}N labelled NT bound to the NTS1B reconstituted into the BPL (black signals) (bottom). The 110-200 ppm region is shown in same scale and similar intensity for both spectra (top). Blue arrows indicate the observed resonance shift (around 140-150 ppm) of the labelled ligand in the presence of receptor which might arise from Arg and/or Tyr side-chain. The spectra were recorded on a Varian 500 MHz SSNMR spectrometer at 233 K. Conditions: 6 kHz, 20.48 ms acquisition time, 8192 scans for free ligand, and 19704 scans for bound ligand.

A 2D ^{13}C-^{13}C correlation spectrum with ^{13}C-^{13}C dipolar recoupling under magic-angle spinning (MAS) is useful for peak assignments as well as for structure determination. Dipolar-assisted rotational resonance (DARR)is an attractive method for the study of relatively large and complex molecules, because it can handle with the

dipolar truncation effect and thus enables long-range distance measurements. The 2D ^{13}C-^{13}C DARR experiment was performed for U-^{13}C, ^{15}N labelled NT in the presence of reconstituted NTS1B receptor. This experiment is expected to facilitate the resonance assignments of the receptor-bound ligand. However, no crosspeaks are detected because of very broad signals of the diagonal region from 5 k scans at 233 K and too short experimental time (Appendix D.5). To improve the signal intensity, higher protein-lipid molar ratio for reconstitution of NTS1B receptor, corresponding to higher number of bound ligand in a rotor and greater number of total scans can be performed. Furthermore, to confirm that the observed resonance arises from the labelled ligand not the unlabelled receptor, the spectra of reconstituted receptor alone are required.

5.4 Conclusions

To improve the spectral linewidth, the crystallisation trials of unlabelled NT_{8-13} peptide were tried using 50 precipitants from the Structure Screen 1 crystallisation kit instead of using the peptide in the lyophilised form which usually gives the spectrum with low resolution. The NT_{8-13} peptide was precipitated well in the precipitating agent 9 (20% (*v/v*) PEG 4000, 50 mM sodium citrate pH 6.0 and 20% (*v/v*) isopropanol) at 4°C in the sealed yellow pipette tip. This precipitant does not have any effects on the conformation of the NT_{8-13} peptide from solNMR studies of the unlabelled peptide in D_2O and precipitant 9, so it was used to produce NT_{8-13} precipitate. However, the precipitated peptide from precipitating agent 9 did not make any improvement on the ^{13}C spectral resolution.

The doubly labelled peptide, U-^{13}C, ^{15}N labelled NT, was useful for NMR backbone and side-chain assignments. These assignments are important for the investigation of NTS1-NT interactions by either solNMR or SSNMR experiments. SolNMR studies by ^{1}H-^{15}N HSQC experiments of U-^{15}N labelled NT with unlabelled TEV-cleaved NTS1B were not successful as the observed crosspeaks of the ligand were narrow with very small chemical shift changes, instead of broad lines upon the strong receptor binding. These results indicated that the receptor might be aggregating during the experiments and those resonances arise from the ligand in the free form.

The resonance assignments of U-^{13}C, ^{15}N labelled NT with the BPL were derived from the ^{13}C-^{13}C COSY semi-SSNMR experiment at 278 K. The obtained assignments are the same as from the U-^{13}C, ^{15}N labelled NT in the detergent buffer, pH 7.4, suggesting that there are no interactions between the ligand and the BPL in agreement with the results from radioligand binding assays. The NTS1-NT interactions were studied by SSNMR methods. The resonance shift at 144 ppm, probably arising from side-chain of either Arg and/or Tyr residue was observed in the 1D CP MAS spectrum of the labelled ligand with reconstituted NTS1B receptor. To confirm whether this shift is derived from the binding, the ^{13}C-^{13}C DARR experiment needs to be performed with long acquisition time, and the spectra of reconstituted NTS1B alone are further required as a negative control to verify that the observed resonance arises from the labelled ligand and not the natural abundance of the receptor. The amount of bound ligand can be increased which improve the signal-to-noise ratio by using the higher protein to lipid ratio for reconstitution of NTS1B receptor.

Chapter 6

Modelling study of NTS1-NT interactions

6.1 Introduction

6.1.1 NTS1 homology models to date

To date, atomic-resolution structural information is received predominantly from either X-ray crystallography or NMR spectroscopy but the ratio of identified structures and protein sequences is still relatively low. Currently, the number of protein sequences is more than 500,000 whereas only 59939 protein structures have been determined to high resolution (http://www.rcsb.org/pdb/home/home.do; August, 2009). Of these known protein structures, only 200 are unique membrane proteins (White, 2009). Bovine rhodopsin was the first GPCR structure determined by X-ray crystallography in 2000 (Palczewski et al., 2000). It had been more than 7 years until the end of 2007, that the second GPCR structure, β_2 adrenergic receptor, was resolved with several mutations to increase the protein stability (Cherezov et al., 2007). The structures of squid rhodopsin (Murakami and Kouyama, 2008), turkey β_1 adrenergic receptor (Warne et al., 2008) and human A_{2A} adenosine receptor (Jaakola et al., 2008) have been recently obtained by the same technique of X-ray crystallography. Not surprisingly, there are not many GPCR structures in the protein database because of the difficulty in preparing a quantity of GPCR proteins. Because of the lack of high resolution structural data, homology modelling on the basis of a known 3D

homologous protein structure is at present an alternative method to provide structural information (Havel and Snow, 1991; Marti-Renom et al., 2000). It is generally accepted that GPCRs can exist as dimers or as part of larger oligomeric complexes (Milligan, 2004) and our group has recently reported the constitutive dimer form of NTS1 receptor in lipid bilayers using FRET measurements (Harding et al., 2009). However, it has been common to build models of GPCRs as monomers because many studies support monomeric GPCRs as functional units (Chabre and le Maires, 2005; Gurevich and Gurevich, 2008; Whorton et al., 2007) and it is also convenient for molecular modelling studies.

In 1998, there was an attempt to construct a homology model of rNTS1 using a rhodopsin model based on an analysis of approximately 500 alignable sequences and a low resolution structure of frog rhodopsin. This homology model has provided useful information about residues involved in NT binding and the potential binding site (Barroso et al., 2000; Labbe-Jullie et al., 1998). Last year, two homology model of hNTS1 were proposed. The one, built by using bovine rhodopsin (1F88) as a template has shown the importance of EC1 loop in agonist binding through π-stacking clusters between ligand and binding site (Harterich et al., 2008), and the other one, constructed by using bovine rhodopsin (1U19) has revealed the helices involved in homodimerisation (Casciari et al., 2008). For the last decade, a few homology-based structural models of the neurotensin receptor were published but were based on the structure of rhodopsin only available from either electron microscopic or crystallographic studies. Nowadays, 5 inactivated X-ray structures of class A GPCRs are obtainable and this offers the possibility of building comparative models of the neurotensin receptor, which are more reliable and thus enable the computational

exploration of the potential ligand-receptor interactions at the atomic level. This structural approach is also applicable to predict or guide laboratory experiments and propose mechanisms of action.

6.1.2 Introduction to AutoDock

AutoDock (Morris et al., 2009) is an automated docking software for predicting the interaction of ligands with macromolecules such as proteins. The software is composed of three separate programs: AutoTors defines rotatable bonds in the ligand; AutoGrid generates grid maps on the macromolecule (precalculation of atomic affinity potentials) for each atom type present in the ligand; AutoDock reads in the grid maps prepared by AutoGrid and a docking parameter file to perform a docking calculation. AutoDock can use a variety of search algorithms to explore a wide range of conformational states. These include Monte Carlo Simulated Annealing (SA); a Genetic Algorithm (GA); and a hybrid local search GA, also known as the Lamarckian Genetic Algorithm (LGA). However, the SA and GA are not as efficient and reliable in finding the lowest energy of the system as the LGA (Morris et al., 1998). With the release of AutoDock4 in 2009, it now allows incorporation of limited flexibility in the receptor to partially solve the problem arising from protein conformation changes upon binding to ligands. The maximum number of torsions in a protein-ligand complex for AutoDock4 is 32. Also, AutoDockTools (ADT), a graphical user interface, was developed to simplify the design and preparation of docking experiments. In addition, it provides methods for clustering, displaying, and analysing the docking results.

The AutoDock software uses a variation on the AMBER force field with terms empirically determined by linear regression analysis from a set of protein-ligand complexes with known binding constants for energy assessments (Morris et al., 1998). The program uses five free energy terms, dispersion/repulsion (ΔG_{vdW}), hydrogen bonding (ΔG_{hbond}), electrostatic (ΔG_{elec}), torsional (ΔG_{tor}), and desolvation (ΔG_{sol}) free energies, to model the binding free energy (ΔG_b).

$$\Delta G_b = \Delta G_{vdW} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{tor} + \Delta G_{sol} \quad (6.1)$$

AutoDock has proven to be an effective tool capable of fast and precise predictions of bound conformations and binding energies of ligands with macromolecular targets close to experimentally determined structures (Hetenyi and van der Spoel, 2002; Osterberg et al., 2002). For instance, the conformation of inhibitor bound to the homology model of the gastric H^+/K^+-ATPase using AutoDock3 is in good agreement with site-directed mutagenesis data and solid-state NMR data (Kim et al., 2005).

6.1.3 Aims

Structural information is a prerequisite for the rational design of mutagenesis experiments of a protein and can be of great importance for the rational drug design. Since the crystallisation of the neurotensin receptor 1 has not been successful and the large quantity of concentrated and stable protein is really difficult to prepare for NMR or X-ray diffraction studies, the structure of neurotensin receptor 1 at high resolution has not been determined yet. Therefore, here the aim is to use homology modelling approach, as a reliable way to get structural details within a short time, to generate 3D

structure models of rNTS1 as monomers. The method will be based on the high resolution crystal structures of turkey β_1 adrenergic receptor and squid rhodopsin following the general flow-chart of homology modelling (section 1.4). Also, energy minimisation and MD simulations will be performed to remove minor errors and see how stable the models are, respectively. To obtain information about the NT-NTS1 interactions, leading to new drug discovery, ligand docking methods will be employed. Together with useful information of mutagenesis and SSNMR studies, the bound conformation of NT_{8-13} ligand to its receptor, rNTS1 and the binding site will be proposed.

6.2 Materials and methods

6.2.1 Homology modelling of NT_{8-13} ligand

Recently, the crystal structure of the four C-terminal amino acid of neurotensin in the ten-bladed β-propeller domain of NTS3 receptor has been determined at 2 Å resolution (Quistgaard et al., 2009). It was used as a template for the modelling of the six C-terminal residues of NT (NT_{8-13}). *Cis* or *trans* conformation of Pro10 can be the key residue to determine the conformation of bound ligand. Therefore, thirty models were generated with the program MODELLER9v4 for each conformation of Pro10 and the ones with the lowest internal objective function value were selected to be the final models.

6.2.2 Homology modelling of neurotensin receptor 1

i) Template detection for neurotensin receptor 1

The first high resolution structure of GPCR superfamily is bovine rhodopsin (bRho; 1U19, 1F88) (Edwards et al., 2004; Okada et al., 2004; Palczewski et al., 2000), determined by X-ray crystallography. The bRho belongs to class A GPCR family which have very conserved seven transmembrane spans. This observation makes the bRho structure a good template for modelling all other proteins in this family, including neurotensin receptor. The crystal structure of human β_2 adrenergic receptor (β_2AR; 2RH1) had been determined in 2007 at 2.3 Å resolution (Cherezov et al., 2007) and was used together with bRho crystal structure as templates for modelling A_1 and A_{2A} adenosine receptors (Yuzlenko and Kiec-Kononowicz, 2009). In 2008, the crystal structures of squid rhodopsin and β_1 adrenergic receptor (β_1AR) was resolved at 2.5 Å and 2.7 Å resolution, respectively (Murakami and Kouyama, 2008; Warne et al., 2008). In addition, the crystal structure of human A_{2A} adenosine receptor has recently been determined at 2.6 Å (Jaakola et al., 2008) after the rNTS1 homology model had been generated. However, rhodopsin structures are the only complete structures at the moment (**Table 6.1**). The percentage identity between the sequence of interest, rNTS1 and a possible template is high enough to be detected with simple sequence alignment programs like BLAST as shown in **Table 6.1**. The neurotensin receptor 1 and adrenergic receptors are ligand-activated receptors, but rhodopsin has a covalently bound ligand, so adrenergic receptors are preferable as a template for comparative modelling of rNTS1. Since β_1AR has a slightly higher sequence similarity (25% identity) with rNTS1 compared to β_2AR (23% identity), it has been used as a template in this study. However, the crystal structure of β_1AR lacks

the intracellular loop 3 (IC3) region because of its flexibility. Therefore, the squid rhodopsin (25% sequence identity with rNTS1) was chosen as another template for modelling the IC3 portion of rNTS1. Several studies have shown that the transmembrane domains of class A GPCR family are very similar despite little sequence similarity (Lodowski et al., 2009; Vassilatis et al., 2003). Thus, some membrane protein models have been successfully performed with less than 20% sequence identity to their templates (Khalid et al., 2006; Oakhill et al., 2005).

Table 6.1: Possible templates for rNTS1 modelling, obtained from BLAST search. The letters in brackets indicate the PDB IDs of structures.

BLAST-Detected templates	The percentage identity between rNTS1 and BLAST-detected templates	Structure
Turkey β_1 adrenergic receptor (2VT4)	25	lack of IC3 loop
Squid rhodopsin (2Z73)	25	complete
Human A_{2A} adenosine receptor (3EML)	25	incomplete EC2 loop and lack of IC3 loop
Human β_2 adrenergic receptor (2RH1)	23	lack of IC3 loop
Bovine rhodopsin (1U19, 1F88)	23	complete

ii) Sequence alignment

The target sequence, rNTS1 (P20789), was retrieved from the SWISS-PROT database of the ExPASy Molecular Biology Server (O'Donovan et al., 2002). The crystal structures of squid rhodopsin (PDB ID: 2Z73) and β_1 adrenergic receptor (PDB ID: 2VT4) were used as templates for multiple alignment. All sequences were initially aligned with Clustal X2 program (Larkin et al., 2007) using the default values of 10.0 for gap opening and 0.2 for gap extension. The multiple sequence alignment helped to identify conserved residues in this subfamily. The manual alignment was performed to

avoid gap opening within the seven transmembrane helices and give the alignment with the highly conserved residues. The 61 N-terminal amino acids of the first transmembrane domain and 38 C-terminal amino acids of the seventh transmembrane domain of rNTS1, which are large non-conserved parts were removed in order to prevent unwanted bias.

iii) Model generation

In this study, MODELLER9v4 was used to perform homology building of possible three-dimensional structures of rNTS1 by satisfaction of spatial restraints (Eswar et al., 2007; Sali and Blundell, 1993). The inputs to the program are an alignment of a target sequence with known template structures and the structures of template proteins in the PDB format. The MODELLER automatically calculates models containing all non-hydrogen atoms. Here, the sequence alignment was obtained from Clustal X2 together with manual alignment. An ensemble of 30 model structures was generated from each input file by using the PDB structure of β_1AR (2VT4) and squid rhodopsin (2Z73). These models were ranked by the objective function value from MODELLER (calculation in terms of residue-based probability density function and residue-based energy) to help in selecting the best model. Here, two different input files were made in order to generate different homology models at EC2 loop which might affect the ligand binding. Firstly (helix group), the rNTS1 sequence was mainly aligned with 2VT4 sequence and partially aligned with IC3 loop of 2Z73 sequence because of the lack of IC3 loop in the 2VT4 crystal structure. The second input file was to observe whether the structure of EC2 loop would be involved in ligand binding in the further studies. The EC2 loop of 2VT4 has a secondary

structure of helix which is involved in ligand binding whereas that of 2Z73 has two β strands which function as a plug of the retinal-binding pocket (Murakami and Kouyama, 2008; Warne et al., 2008). As the EC2 loop of rNTS1 is long (~26 residues), it could be in either structure or none. Also, some residues in this loop might be involved in ligand binding. Therefore, the second (β group), the target sequence was aligned with 2VT4 structure containing EC2 and IC3 loops of the 2Z73.

iv) Model optimisation and validation

It is necessary to check whether generated models are reliable. There are many alternatives for model validation, such as PROCHECK (Laskowski et al., 1993), WHATCHECK (Hooft et al., 1996), PROVE (Pontius et al., 1996), Verify3D (Luthy et al., 1992), ProsaII (Sippl, 1993), and so on. In this study, ProsaII was used to check the model fold and PROCHECK was used to assess the stereochemistry of models (bond length and angles, dihedral angles, atom-atom overlaps, etc.). Moreover, visualisation tools, such as WebLab ViewerPro (Accelrys Inc.), PyMOL (http://www.pymol.org/) and VMD (Humphrey et al., 1996) were used to evaluate the directions of side-chains of the models whether they are consistent with experimental knowledge such as site-directed mutagenesis (SDM) and ligand binding data.

The top five models of each ensemble ranked by MODELLER's internal objective function value were evaluated and the one with the highest PROCHECK G-factor and the highest residue number in core regions of Ramachandran plot was selected.

The chosen models were then subjected to loop refinements using Modloop program (Fiser and Sali, 2003). As EC3 is a large, non conserved loop which is

important for ligand binding (Barroso et al., 2000; Labbe-Jullie et al., 1998), one hundred loops of a selected model were built completely *de novo* for refinement, without any alignment between the sequences of 2VT4 and rNTS1 using MODELLER9v4. The top five loop models which have side-chains of important residues in the same direction and the highest PROCHECK G-factor were selected for ligand screening experiments using AutoDock4 in order to choose the loop model which agrees with experimentally-attained data. Then, the models containing the selected loop were refined by energy minimisation to obtain the final model. PyMOL was used for the preparation of all figures.

6.2.3 Protein-ligand docking

AutoDock4 (Morris et al., 2009) was used to determine the putative receptor binding site of a ligand. The program uses the Lamarckian Genetic Algorithm, a hybrid algorithm of global and local searches, for docking simulations. Adding polar hydrogens and removing non-polar hydrogens were performed to the model structure and ligand. The resulting file was saved as a .pdbq file which is an altered form of the .pdb file containing charges. The ligand's rotatable bonds were automatically selected in the ligand menu in AutoDockTools. Flexible docking was performed allowing fully flexible ligand and eight torsions to rotate in residues Met208, Trp339 and Tyr347 of rNTS1 protein. AutoGrid was used to pre-calculate atomic affinity potentials for each atom type in the macromolecule. A three-dimensional grid or search area is centred on a region of interest on the model. This grid consists of an array of grid points. Here, a blind dock was performed in AutoDock4 using a grid centred on the top of TM helices and extracellular regions of the model with parameters as following. The xyz

coordinates for the grid centre were set to 37, 48, and 96, respectively. The size of each grid was 0.375 Å and the numbers of points for the grid in xyz directions were set to 126, 126, and 70, respectively. The grid box is shown with the rNTS1 model in **Figure 6.1**. A grid log file was produced. The algorithm used for docking was simulated annealing with default parameters. A docking log file was produced. AutoGrid was then run on the grid log file and subsequently AutoDock was run on the docking log file using parameters as following. For the parameters of the global search algorithm, a population size of 300, a maximum number of energy evaluations of 35 million, a maximum number of generations of 27,000, a rate of gene mutation of 0.02, and a rate of crossover of 0.8 were applied for each experiment. For the local search algorithm of Solis & Wets (Solis and Wets, 1981), a maximum number of iterations of 300 and a probability of performing a local search on an individual of 0.06 were used. Each docking experiment was performed 50 times, yielding 50 docked conformations. Final docked conformations were clustered using a tolerance of 2.0 Å RMSD. Three independent experiments were performed for each protein-ligand complex and the five lowest energy docked structures of each complex were analysed. Docked ligands which agreed with experimental data were selected for further simulation experiments.

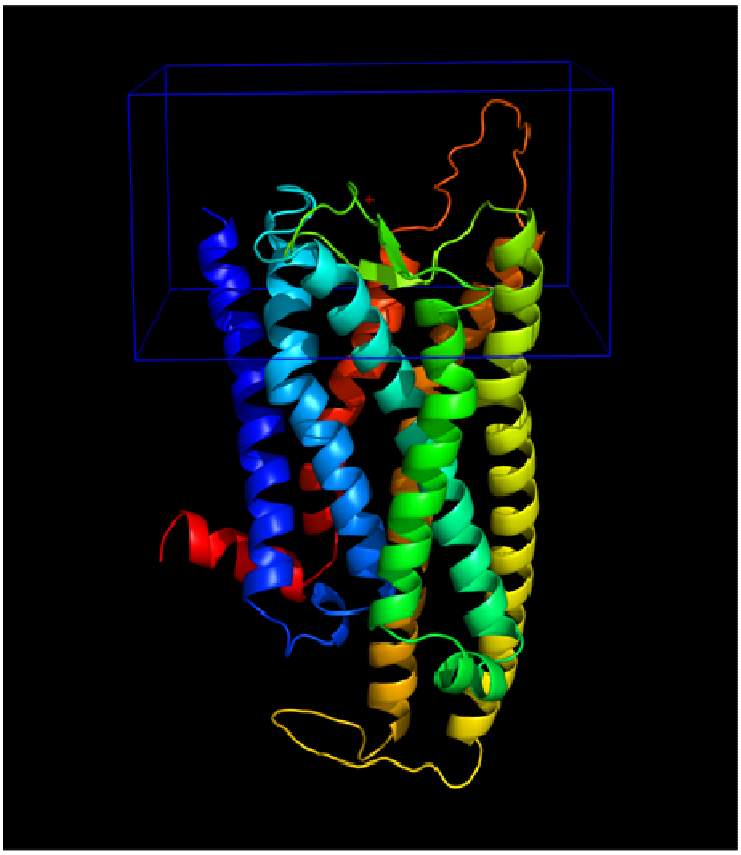


Figure 6.1: Grid box of rNTS1 model for AutoDock-based blind docking. The xyz coordinates for the grid centre were set to 37, 48, and 96. The size of each grid was 0.375 Å and the numbers of points for the grid in xyz directions were set to 126, 126, and 70, respectively.

6.2.4 Molecular dynamic simulations

The GROMACS 3.3 simulation package (http://www.gromacs.org/) was used for all simulations with a united atom Gromos96 force field (van Gunsteren et al., 1996). Each model was embedded in a membrane mimetic POPC288 slab (Tieleman et al., 1999) and then solvated with SPC water model (Hermans et al., 1984). Counterions were added to yield an electroneutral system that was energy minimised prior to starting simulations. Periodic boundary conditions were applied to the systems. Long-range electrostatics were calculated using the Particle Mesh Ewald (PME) method (Darden et al., 1993) with a real-space cut-off of 10 Å. For the van der

Waals interactions, a cut-off of 10 Å was used. The simulations were performed at a temperature of 300 K using a Berendsen thermostat (Berendsen et al., 1984) with coupling constant of tau T = 0.1 ps. A constant pressure of 1 bar was maintained using a Berendsen barostat with an isotopic coupling constant of tau P = 1.0 ps and compressibility = $4.5x10^{-5}$ bar^{-1}. The integration time step was 2 fs using LINC method (Hess et al., 1997) to constrain bond lengths. Coordinates were save every 2 ps for analysis. Analysis of all simulations was performed using the GROMACS package. VMD (Humphrey et al., 1996) and PyMOL (http://www.pymol.org/) were used for visualisation.

6.3 Results and discussion

6.3.1 Homology modelling of NT_{8-13} ligand

The crystal structure of the NT_{10-13} peptide in the ten-bladed β-propeller domain of NTS3 receptor was used as a template for the modelling of NT_{8-13} ligand using MODELLER9v4. Two conformations (*cis* and *trans*) of Pro10 of NT_{8-13} peptide were generated leading to two final models, cisNT6 (dihedral angle of-0.5 degrees) and transNT6 (dihedral angle of-179.9 degrees). Both ligand models were used for general screen of binding site and for the observation of specific conformation of Pro10 in ligand binding.

6.3.2 Homology modelling of neurotensin receptor 1

i) Sequence alignment and model generation

Comparative modelling, widely approved for the prediction of 3D model structure of rhodopsin-like GPCR family, was employed to construct homology models of the rat neurotensin receptor 1. The crystal structures of turkey β_1 adrenergic receptor (2VT4) and squid rhodopsin (2Z73) were used as the templates for the modelling. Although the sequence similarity between templates and target sequence was only 25%, the sequence similarity in TMs was about 52% and the target sequence was aligned very well with the templates giving consensus residues (**Figure 7.2**) as reported in GPCR Data Bank (Horn et al., 2003). The highly conserved motifs of GN in TM I, LxxxD in TM II, C(23x)E/DRY in TM III, W in TM IV, P in TM V, C/SWxP in TM VI, NPxxY in TM VII, and FR in TM VIII are believed to share important functions, and some of them have been studied by mutagenesis. The presence of a well conserved Asp residue in TM2 of GPCRs is involved in the sodium sensitivity of high affinity agonist binding (Ceresa and Limbird, 1994; Kitabgi, 2006). Two different functions of E/DRY motif in TM III within class A GPCRs were proposed. A subgroup of this family uses the E/DRY motif for constitutive activity and agonist affinity, whereas in the other subgroup, the E/DRY motif is involved in regulating receptor conformation and G-protein coupling/recognition (Alewijnse et al., 2000; Rovati et al., 2007). The Trp in TM IV is totally conserved in GPCR class A family and it has been shown to play an essential role in ligand binding for many receptors, including the human CB_2 (Rhee et al., 2000), m3 muscarinic (Wess et al., 1993), 5-HT_{2A} serotonergic (Roth et al., 1997), and human Y1 (Sautel et al., 1996) receptors. The Tyr in the conserved NPxxY motif in TM VII has been found to be involved in

agonist affinity, signal transduction, sequestration, and constitutive activity in different GPCRs (Barak et al., 1995; Gabilondo et al., 1996; Laporte et al., 1996; Rosendorff et al., 2000). A proline-induced kink at conserved positions in TM II, V, VI, and VII are thought to be responsible for structural rearrangements needed for activation of G protein effectors (Yohannan et al., 2004). Moreover, the disulfide bridge between the top of TM III and the EC2 loop, the most conserved aspect of GPCRs, are defined. From crystal structures of GPCRs, this conserved disulfide bridge was shown to be important in the structure stability of the large EC2 loop (Cherezov et al., 2007; Jaakola et al., 2008; Murakami and Kouyama, 2008; Okada et al., 2004; Warne et al., 2008).

The alignment was used for model construction of rNTS1 receptor by MODELLER9v4 as described in section 6.3.2. To determine whether the secondary structure of EC2 loop is involved in ligand binding, rNTS1-B1 and rNTS1-H1 final models were chosen which have β and helix structures at EC2 loop, respectively.

Figure 6.2: Multiple sequence alignment of the rat neurotensin receptor sequence (rNTS1) with the structural sequences of turkey β_1 adrenergic receptor (2VT4) and squid rhodopsin (2Z73). Bold blue colour shows the extremely variable amino termini upstream of the first transmembrane domain and carboxyl termini downstream of the seventh transmembrane domain which are not counted in model generation. Bold red colour shows consensus motifs among GPCR class A family. The sequence identity of rNTS1 with either 2VT4 or 2Z73 is 25%. "*" means that the residues in that column are identical in all sequences in the alignment. ":" means that conserved substitutions have been observed. "." means that semi-conserved substitutions are observed.

ii) Model optimisation and validation

ProsaII

A model will have a correct fold if the correct template is chosen and the alignment between the template and the target sequence is approximately correct. Z-

scores, an important factor calculated by ProsaII4.0 program, indicate the quality of protein structures. The program uses a knowledge-based potential, derived from high-resolution crystal structures, to assess how well a structure fits a sequence. The ProsaII Z-score of a model measures the compatibility between its structure and sequence compared to alternative folds compiled from the PDB. Z-scores of native proteins are dependent on the length of the protein. If the chosen template is correct, the ProsaII Z-scores of the model and the template structures should be comparable. The low Z-score of a model means that the model has a good quality of protein structure. Z_p is a Z-score of protein and Z_1 is a Z-score of fragment of lowest energy found in the protein. **Table 6.2** shows the Z_p and Z_1 values of both templates and model proteins. The Z-scores of the chosen models show good compatibility with those of templates (2VT4 and 2Z73). These indicate that both rNTS1-H1 and rNTS1-B1 are reliable.

Table 6.2: Comparison of Z-scores between templates (PDB ID: 2VT4 and 2Z73) and selected models (rNTS1-H1 and rNTS1-B1).

Protein	Z_p			Z_1		
	Combined energy	Pair energy	Surface energy	Combined energy	Pair energy	Surface energy
Template						
2VT4	-4.07	-5.37	-2.59	-3.51	-3.58	-3.62
2Z73	-5.01	-5.23	-3.33	-3.66	-3.79	-3.41
Model						
rNTS1-H1	-2.73	-4.32	-1.54	-3.77	-4.38	-3.60
rNTS1-B1	-3.22	-4.23	-2.04	-3.77	-4.38	-3.60

PROCHECK

The PROCHECK program was used to verify how normal the geometry of the residues in a target protein structure, once compared with stereochemical parameters derived from high-resolution structures. The output of this program is a series of Ramachandran plots, GLY and PRO Ramachandran plots, Chi_1-Chi_2 plots, main-chain parameters, side-chain parameters, residue properties, etc.

The percentage of residues in the core regions of the Ramachandran plot is one of the best guides to the stereochemical quality of a protein structure. Over 90% of the residues in the core regions were ideally expected. Overall G-factor is another important parameter given by PROCHECK. It is calculated for the properties of torsion angles and covalent geometry. This value is a measure of the overall normality of the structure. If the value is less than -0.5, the protein structure is unusual and if it is less than -1.0, the protein structure is regarded as a highly unusual. A low G-factor of a given residue indicates that the property corresponds to a low-probability conformation. So, for example, residues falling in the unfavourable Chi_1-Chi_2 regions will have a low (or very negative) G-factor. **Table 6.3** shows the summary of the overall G-factors and the percentages of the residues in the Ramachandran plot (**Figure 6.3**) for the selected models, rNTS1-H1 (helix at EC2 loop) and rNTS1-B1 (β-strand at EC2 loop) with those of templates. The final model of each ensemble, constructed by MODELLER has approximately 90% of amino acid residues in the core regions and no residue in disallowed regions. Moreover, their overall G-factor values are very close to zero. The comparable Ramachandran plot characteristic and G-factors indicate the quality of selected models. The final models of rNTS1 are shown in **Figure 6.4**.

Table 6.3: Summary of the overall G-factors and the percentages of the residues in the Ramachandran plot for the templates and final models generated by PROCHECK.

	Template (2VT4)	Template (2Z73)	Model rNTS1-H1	Model rNTS1-B1
Core region	92.4	90.2	88.9	88.2
Allowed region	7.2	9.8	9.4	11.4
Generously allowed region	0.4	0.0	1.7	0.3
Overall G-factor	0.17	0.34	-0.08	-0.09

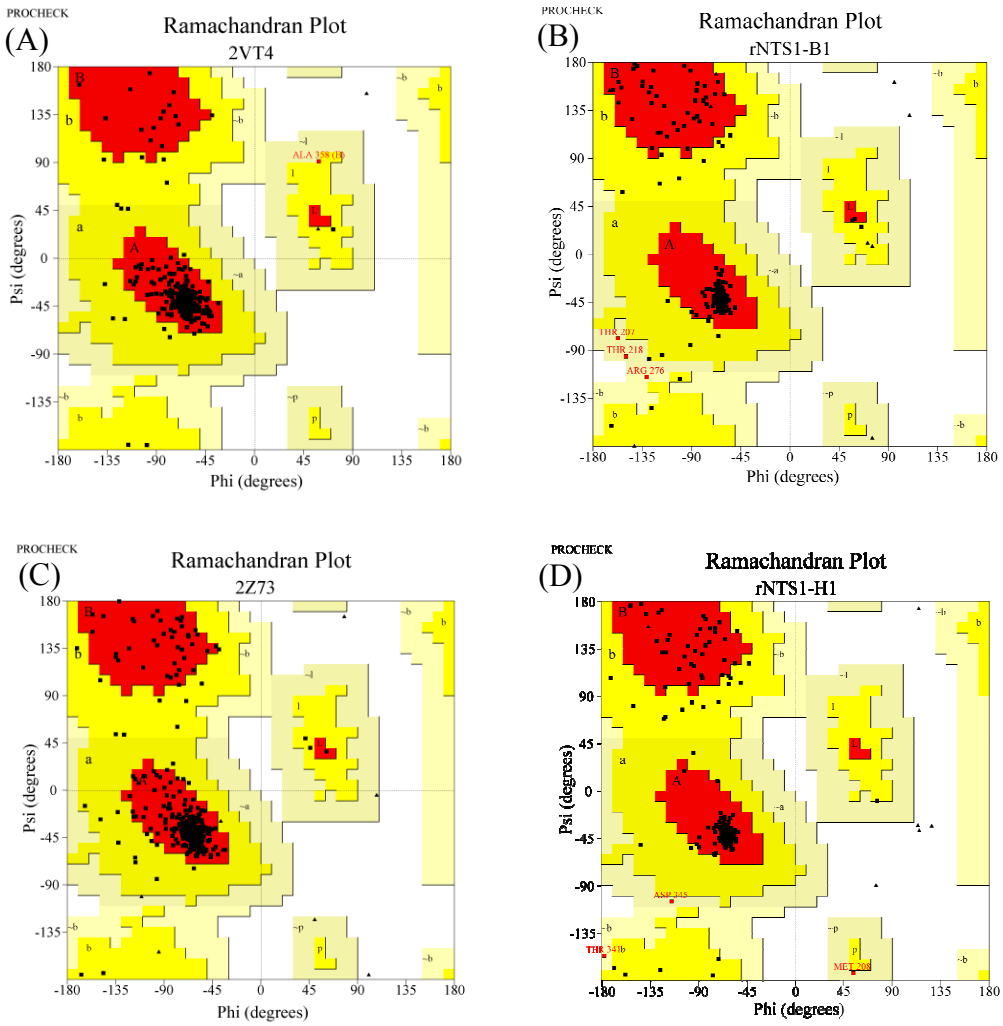


Figure 6.3: Ramachandran plots of templates [2VT4 (A) and 2Z73 (C)] and models [rNTS1-B1 (B), and rNTS1-H1 (D)] created by PROCHECK program.

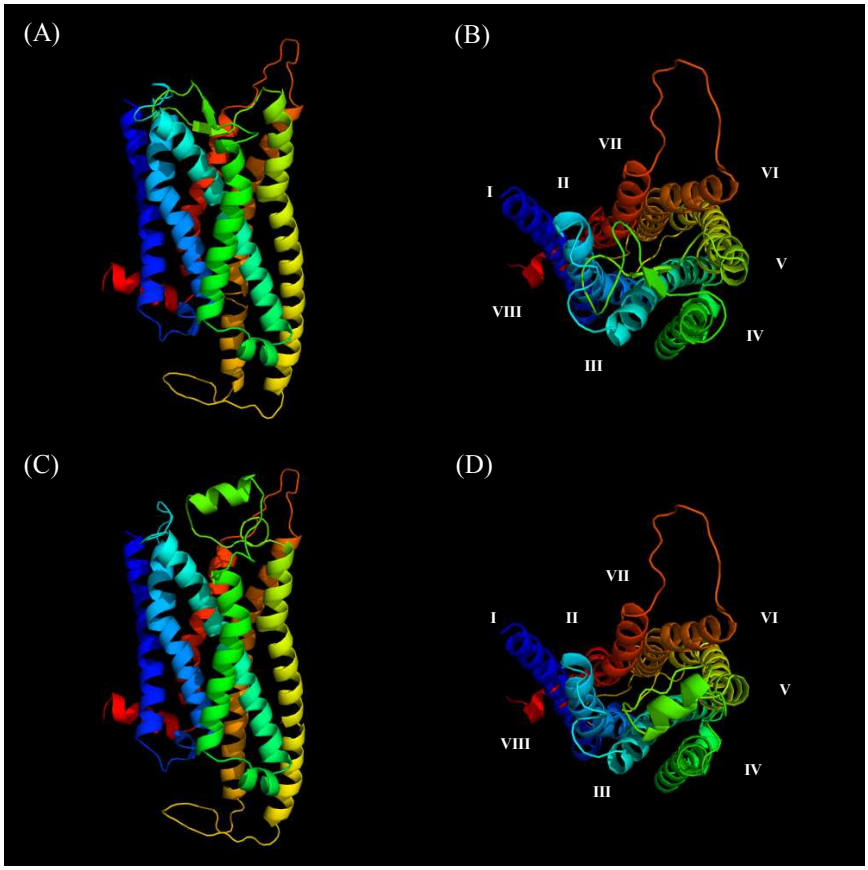


Figure 6.4: Three-dimensional model structures of rNTS1 constructed by MODELLER9v4. The side view of rNTS1-B1 (A) and rNTS1-H1 (C) are shown. The top down view of rNTS1-B1 (B) and rNTS1-H1 (D) are also illustrated. Rendered in PyMOL.

RMSD

It has been known for a long time that the most reliable part of a predicted protein model is the portion it shares with the modelling template, whereas the non-conserved loop parts were a major contributor to model inaccuracy (Bajorath et al., 1993; Harrison et al., 1995). Therefore, the conformation of conserved transmembrane spans should be evaluated as a check of reliability. Root mean square deviation of the positions of α carbons or RMSD is a factor to assess the overall difference in protein backbone structures. A model can be considered to be accurate enough once its RMSD is within the spread of deviations observed for experimental structures showing a

similar sequence identity level as the target and template sequences. The RMSD of backbone atoms in independent determinations of the same protein is up to 0.5 Å (Chothia and Lesk, 1986), whereas that of α carbons of almost 80% of the sequences sharing 50-59% identity with their templates is less than 3 Å from their control structures (Schwede et al., 2000). Here, only one template (2VT4) was used to measure the RMSD value for the models since it is the only template for the modelling of TM regions. The RMSD of TMs between the 2VT4 and the models, calculated from PyMOL, is 0.420 Å for rNTS1-B1 and 0.422 Å for rNTS1-H1. These results suggest that the modelling step has not significantly altered the protein backbone and also confirm the conserved 7TMs among the GPCR class A family.

Visualisation programs for model evaluation

PyMOL, VMD and WebLab ViewerPro were used for visual inspections. Superposition of the models was done over the template 2VT4. The overall conformation except the EC2 and IC3 loops of rNTS1 models is similar to the crystal structure of β_1AR, corresponding to the small number of RMSD. The model validation was carried out through an inspection of the experimental data. All mutagenesis studies of NTS1 receptor with the agonists NT binding presented earlier are summarised in **Table 6.4**. The binding of agonist was absent in the following mutations of Asp139, Arg143, Arg327, and Tyr347 and deletion of residues 45-60. Furthermore, mutations of Met208, Phe331, Trp339, Phe344, and Tyr351 showed a dramatic decrease of agonist affinity (> 5 fold) indicating direct involvement of binding. There are also other residues which might be indirectly involved in agonist binding by displaying only small decrease of agonist binding after mutagenesis. The

viewer program showed that these crucial residues were located on the ligand-accessible surface of the models.

Table 6.4: Summary of the effects of mutation of residues in extracellular loops and the top of transmembrane domains of rat NTS1 and human NTS1 on the affinity for NT, NT_{8-13}, SR48692 (K_d).

Region	Residue	K_d (nM)		Reference
		Neurotensin	SR48692	
	Wild type	0.12±0.02 0.39±0.10 1.7* 0.63 (0.49-0.80)±11%*	2.60±0.20 8.9 (7.5-11)* 2.6±0.2	A, B C D E F
N-terminus	Δ14-28 Δ45-60 E45G-D47G D54G-D56G D60G K64G	NC **ND** 0.45±0.07 0.42±0.04 0.30±0.04 1.20±0.13	NC 3.0±0.3	F F C C C C
TM I	V65A L66A V67A T68A	0.11±0.02 0.12±0.02 0.07±0.01 0.29±0.05	2.21±0.13 2.17±0.16 2.59±0.34 1.89±0.21	A A A A
TM II	D113G	0.39±0.05		C
EC I	E124G F127A* W129A* H131A* P133A* W134A* F136A* D139A D139G	0.43±0.02 0.74 (0.60-0.94)* **ND*** 0.64 (0.51-0.84)* 0.64 (0.39-1.10)* **ND*** 0.46 (0.36-0.64)* **ND** **ND**	 19 (12-29)* 27 (19-38)* 8.4 (6.9-10)* 9.2 (7.1-12)* 23 (12-44)* 33 (17-62)*	C E E E E E E B C
TM III	R143G R143K R143Q R143M Y145A Y146A F147A R149G D150G Y154A T156A	**ND** 0.11±0.01 0.13±0.03 0.09±0.04 0.16±0.03 0.30±0.03 0.14±0.01 0.37±0.01 0.36±0.02 0.26±0.06 0.22±0.05	 1.55±0.29 1.77±0.32 2.75±0.62 4.42±1.12 2.69±0.33 2.25±0.26 2.43±0.61 2.32±0.27	C A A A A A A C C A A
TM IV	M204A F206A	0.20±0.06 0.08±0.02	4.56±0.45 2.36±0.25	A A
EC II	M208A R213G D216G D230G	**1.26±0.26** 0.95±0.09 0.91±0.09 0.32±0.05	**ND**	A C C C

Continued

Region	Residue	K_d (nM)		Reference
		Neurotensin	SR48692	
	Wild type	0.12±0.02 0.39±0.10 1.7* 0.63 (0.49-0.80)±11%*	2.60±0.20 8.9 (7.5-11)* 2.6±0.2	A, B C D E F
TM V	K235A K235Q K235G K235R V236A V237A I238A Q239A T242A F243A	0.18±0.04 0.32±0.11 0.22±0.03 0.18±0.07 0.12±0.01 0.12±0.01 0.13±0.01 0.25±0.09 0.26±0.15 0.13±0.02	2.70±0.52 4.11±0.77 2.21±0.35 4.21±0.66 3.14±0.55 3.04±0.46 4.85±0.63 4.32±1.29 1.42±0.16	A A C A A A A A A A
TM VI	R311G Y324A H325A R327M R327E R327G-R328G R328M I329G* YIS328-330AAA* F331A	0.39±0.10 0.20±0.07 0.10±0.01 **ND** **ND** **ND** 0.28±0.04 0.47±23%* 0.80±36%* **1.29±0.15**	 **ND** 1.84±0.16 **ND** **ND** 5.00±0.80 **ND**	C A A A A C A E E A
EC III	Y333A D336G-E337G W339A Y339F* D340H* F344A D345G F346A Y347A Y347M Y347F H348A	0.20±0.05 1.05±0.05 **1.13±0.20** 2.08* 4.37* **0.69±0.18** 0.84±0.10 0.22±0.02 **ND** **ND** **ND** 0.22±0.02	2.86±0.61 2.63±0.14 2.51±0.35 3.12±0.42 3.99±0.73 4.92±0.95 1.51±0.13 1.71±0.34	A C B D D B C A A B B A
TM VII	Y349A F350A Y351A T354A F358A Y359A	0.45±0.10 0.23±0.03 **0.90±0.27** 0.33±0.11 0.13±0.04 0.40±0.07	3.02±0.77 5.37±1.19 **ND** **13.10±1.40** 8.94±1.90 **ND**	A A A A A A

Description:

Red-bold numbers display the effect of mutation which decreases the ligand binding affinity by more than 5 times.

SDM data are from rat NTS1 except residues with *, which are from human NTS1.

References: A (Labbe-Jullie et al., 1998) using ^{125}I-NT and ^{3}H-SR48692; B (Barroso et al., 2000) using ^{125}I-NT and ^{3}H-SR48692; C (Botto et al., 1997) using ^{125}I-Tyr11-NT; D (Pang et al., 1996) using ^{3}H-NT; E (Harterich et al., 2008) using ^{3}H-NT_{8-13}; F (Labbe-Jullie et al., 1995) using ^{125}I-NT and ^{3}H-SR48692.

Abbreviations: ND not detectable and NC no change.

iii) Summary

On the basis of two sequence alignments differing at the EC2 loop, two alternative homology models (rNTS1-B1 and rNTS1-H1) of the rat neurotensin receptor (rNTS1) were generated using MODELLER9v4. In both cases, the disulfide bridge formed between the EC2 loop and TM III was manually defined. The quality of the models was mainly examined in terms of protein folding and stereochemistry, as calculated by ProsaII and PROCHECK, respectively. Comparing evaluation results of the models with those of templates, the structure of selected models could be regarded as quite realistic for subsequent docking studies.

6.3.3 Docking simulations between NTS1 and NT_{8-13}

Two approaches were used to study binding of NT_{8-13} ligands to the rNTS1 homology models generated by MODELLER. Also, the ligands with Pro10 in either *cis* or *trans* conformation were examined for ligand-receptor binding to observe whether the ligand has a favourable Pro10 conformation in a bound state. The first method was the use of MD simulations by allowing free movement of proteins and ligands with full flexibility in the system. For the second approach, an automated docking program, AutoDock4 was first employed to determine the binding orientations and free energy of binding of protein-ligand complexes. Then, MD simulations were further used to refine those protein-ligand binding conformations obtained from AutoDock4.

i) Equilibration of model structures

The rNTS1 homology models were subjected to MD simulations to observe how long they reach the equilibrium which will help to decide the simulation time for the studies of protein-ligand complex. The protein models were introduced into an equilibrated bilayer of 288 POPC molecules, surrounded by water molecules. The systems had a nonzero total charge (+10) so they were neutralised with 10 chloride ions. The energy of resulting system was minimised and the system was equilibrated with the protein restrained for 500 ps in order to allow relaxation of surrounding lipid and solvent molecules around the protein. After system equilibration, trajectories with a length of 5000 ps were recorded. Analysis of the trajectories (energies, root mean square deviation (RMSD) and root mean square fluctuation (RMSF)) was performed using GROMACS analysis tools. For rNTS1-B1 model, the total system energy dropped to its final value during unrestrained MD simulation after ~4000 ps. The stable potential energy (**Figure 6.5A**) and temperature (data not shown) indicated the equilibrium state of the system. From the Cα-RMSD results (**Figure 6.6A**), the seven transmembrane helices remained within an Cα atom RMSD of ~0.25 nm from the starting structure and get into the equilibrium state after 3000 ps. Considering the whole protein, it rose to an average RMSD value of 0.33 nm within 1000 ps and remained stable at this point for 1200 ps, and climbed gradually to a stable RMSD value of ~0.5 nm after 4000 ps. These results have suggested the high flexibility of protein loops. The last 1000 and 3000 ps of the trajectory were analysed. During the last 3000 ps period, the loop regions showed pronounced movements with a RMSF of Cα atom coordinates of 0.05-0.47 nm, particularly the IC3 loop, while the seven transmembrane helices remained stable in position (RMSF of 0.05-0.07 nm) (see

Figure 6.6B). However, at the equilibrium state (after 4000 ps), the RMSF is in the range of 0.03 and 0.27 nm. Excluding the IC3 loop for RMSD analysis (**Figure 6.6A**), the protein reached equilibrium after 3000 ps with RMSD value of only 0.4 nm. It was not surprising that high flexibility of the IC3 loop was observed because it has been proposed to interact with G-proteins for signal transduction in cells.

For rNTS1-H1, an additional 5000 ps MD simulation time was required for the equilibration of the protein. Analysis by GROMACS tools showed that the potential energy of the system reached a minimum after 8000 ps (**Figure 6.5B**) and the temperature was stable at 300 K during the experiment (data not shown). The seven transmembrane helices reached equilibrium rapidly within 500 ps with RMSD value of 0.2 nm from the starting structure, whereas an average RMSD value of 0.55 nm was detected for the whole protein after 8000 ps (**Figure 6.7A**). Although the IC3 loop was not used for RMSD analysis, it still rose to an average value of 0.45 nm resulting from large fluctuations of several model loops, as can be seen in RMSF analysis of the last 6000 ps of the trajectory (**Figure 6.7B**). However, those loops were refolded and finally adopted a stable conformation as the short range (0.03-0.17 nm) of RMSF of the last 1000 ps of the trajectory was observed. In comparison with rNTS1-B1, the loops of the rNTS1-H1 model are more flexible and reach the equilibrium state more slowly.

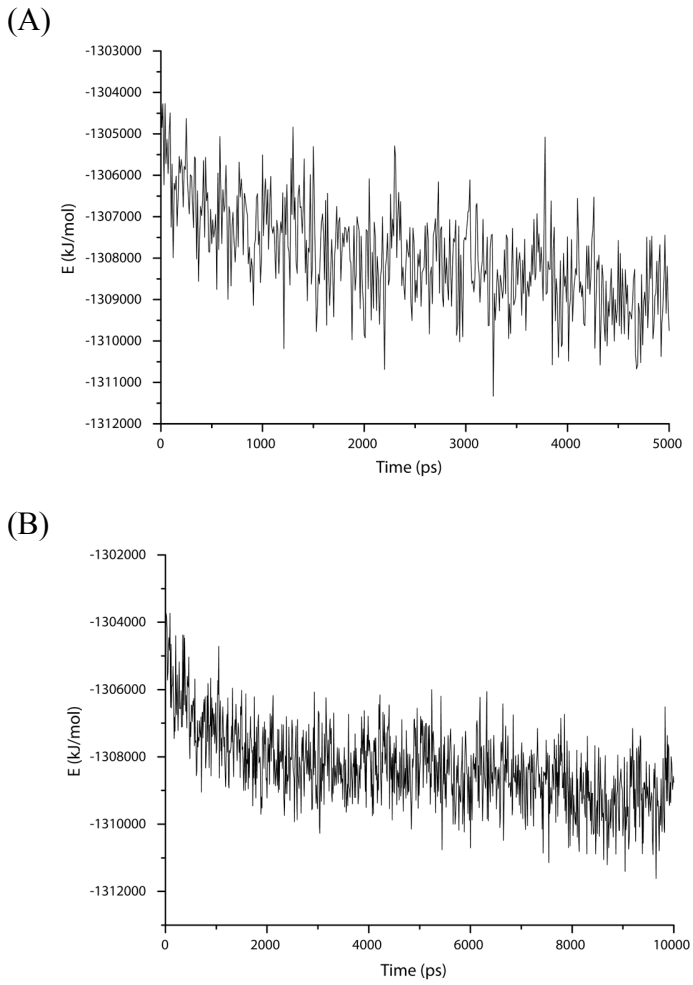


Figure 6.5: Variation in potential energy during the 5000 ps and 10000 ps of MD on the rNTS1-B1 (A) and rNTS1-H1 (B) models, respectively.

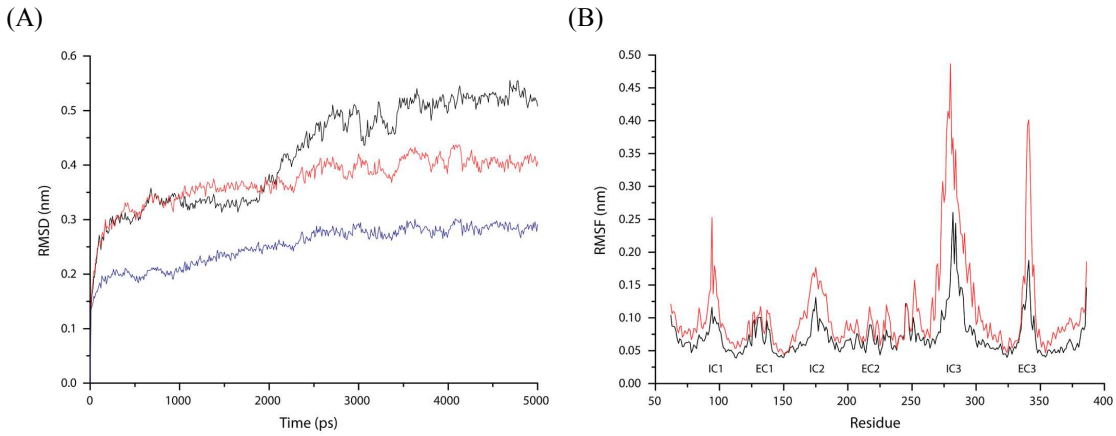


Figure 6.6: Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) of rNTS1-B1 homology model. Cα-RMSD of the transmembrane helices (blue), the model without IC3 loop (red) and the whole model (black) is shown as a function of MD simulation time (A). The Cα-RMSF of the model during last 1000 ps (black) and 3000 ps (red) of MD trajectory was calculated (B).

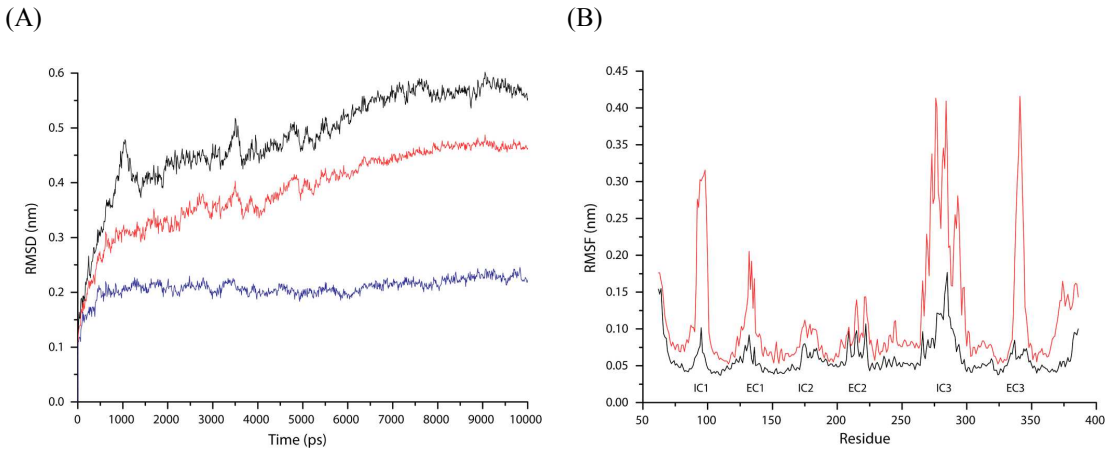


Figure 6.7: RMSD and RMSF of rNTS1-H1 homology model. Cα-RMSD of the transmembrane helices (blue), the model without IC3 loop (red) and the whole model (black) is shown as a function of MD simulation time (A). The Cα-RMSF of the model during last 1000 ps (black) and 6000 ps (red) of MD trajectory was calculated (B).

The time evolution of the potential energies, RMSD and RMSF of the MD trajectory from 2VT4 and 2Z73 crystal structures was computed to analyse the structural stability of the rNTS1 homology models. Both crystal structures reached equilibrium with a Cα-RMSD of ~0.2 nm indicating the high structural stability of backbone structures making them good templates for homology modelling (**Figure 6.8**). Focusing on the secondary structure of EC2 loop, the β structure of 2Z73 gives more compact structure (small fluctuation) compared to the helix structure of 2VT4 which correlates with the EC2-loop structure of rNTS1-H1 and rNTS1-B1 models. The RMSF of coordinates of Cα atoms of both template structures at equilibrium during last 1000 ps (**Figure 6.9**) is comparable with that of rNTS1 models suggesting acceptable receptor models. The model structures at equilibrium will be used to compare with the structures upon ligand binding to observe any conformational changes.

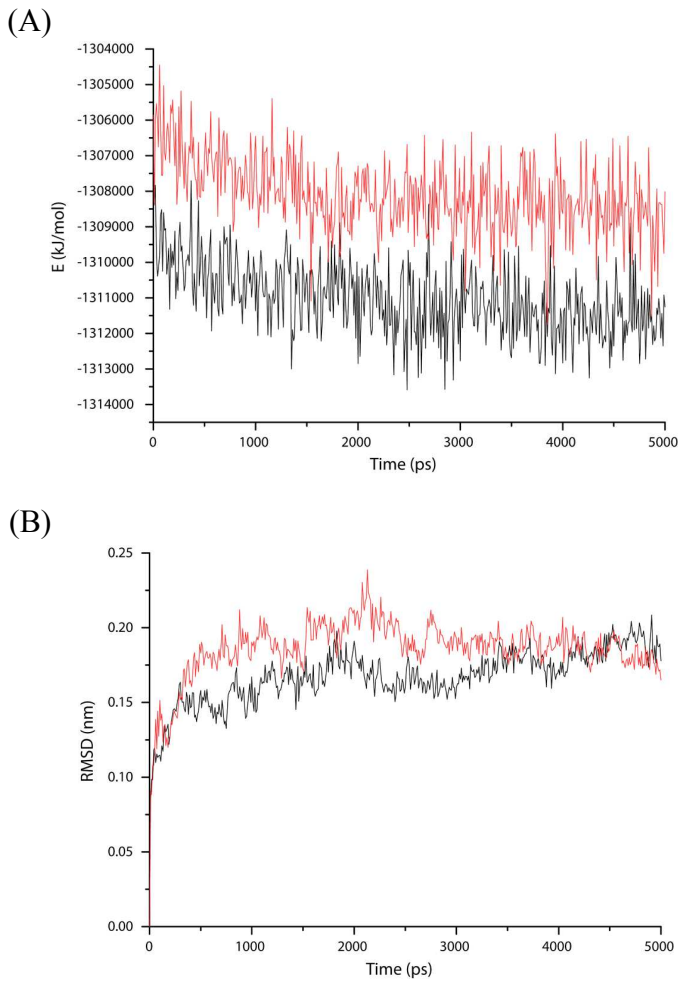


Figure 6.8: Potential energy (A) and Cα-RMSD (B) of 2VT4 (black) and 2Z73 (red) templates for homology modelling of rNTS1 receptor. The values are shown as a function of MD simulation time.

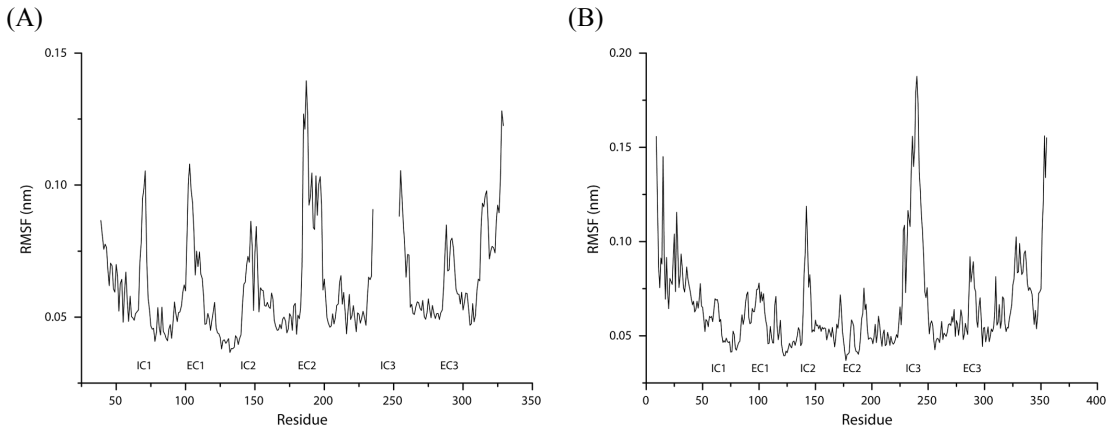


Figure 6.9: Cα-RMSF of 2VT4 (A) and 2Z73 (B) crystal structures during last 1000 ps of MD trajectories.

ii) MD simulations approach

Two ligands (transNT6 and cisNT6) generated from the program MODELLER9v4 were docked to each rNTS1 model by GROMACS 3.3. Therefore, four sets of simulations of protein-ligand complexes *i.e.* rNTS1-B1 and transNT6; rNTS1-B1 and cisNT6; rNTS1-H1 and transNT6; rNTS1-B1 and cisNT6 were performed. After energy minimisation, each complex was submitted to MD simulations using the GROMOS 96 force field.

At the approximately 2.5 nm distance of the ligand from the membrane bilayer, interestingly both *cis*/*trans* ligands moved close to the rNTS1-H1 but not rNTS1-B1 and both bound to the EC3 loop. **Figure 6.10** shows the location of protein and ligands after 10 ns MD simulation time. For the rNTS1-H1 and transNT6 complex, the ligand formed hydrogen bonding interactions to residues Asp336 and Glu337 of the protein. These two residues in the receptor were proposed to be involved in indirect ligand binding because only small decreases of ligand binding after mutagenesis were observed (Botto et al., 1997). This position might be just pre-binding site but unluckily the complex was trapped in local minima which might be solved by longer simulations or manual movement of ligand fragment. Upon binding, the ligand was still in an extended conformation, and no significant conformational change was seen, although the shifts of the top of TM I and II and the intracellular parts of TM V and VI were clearly observed in cryo-EM studies of liganded NTS1 receptor at < 7 Å resolution (Selmi et al., unpublished work). Prolonged MD simulations are probably required for determining significant movement of transmembrane parts.

During relaxing of water and lipid molecules around the protein in the rNTS1-H1 and *cis*NT6 complex, the ligand came closer to the protein at the top of EC2 loop

with a distance of approximately 15 nm from the membrane. At this point, the ligand adopted a U-shape conformation with dihedral angle of Pro10 close to -90 degrees. After that, the ligand moved to the position next to the EC3 loop between three extracellular loops during MD simulations and forms interactions to these three loops. However, no significant conformational change in helices was seen.

Although the docking results of both complexes did not agree well to each other, these experiments showed strong binding between the protein and ligands and high affinity to the EC3 loop for ligands. Note that specific interactions between protein and ligand are not mentioned because side-chain of critical residues from experimental data must be first refined in the models and further distance-restraining MD simulations at the protein binding site may be required. This is a future project.

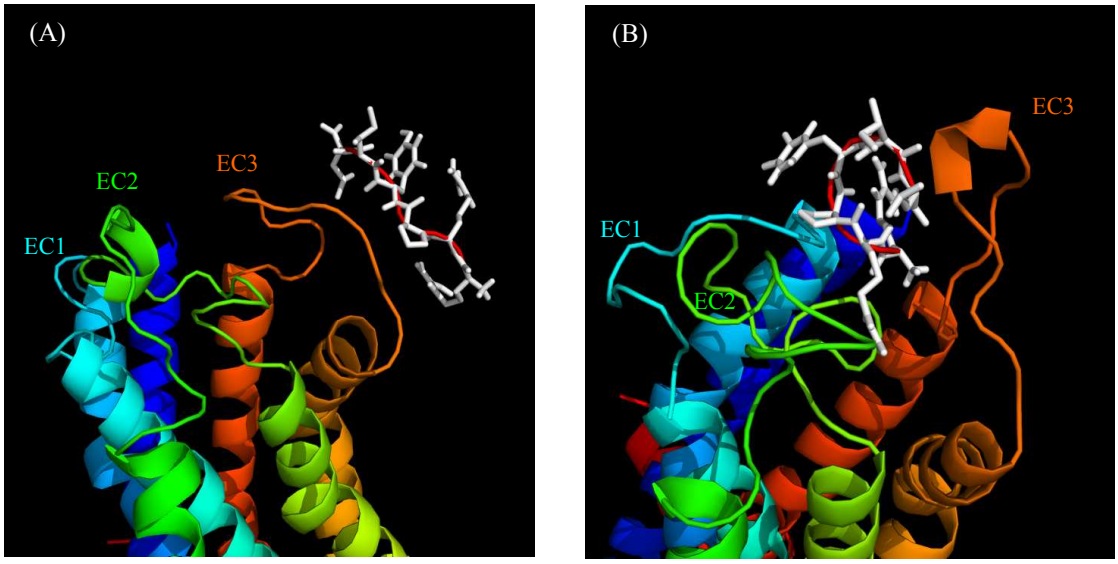


Figure 6.10: The position of transNTS1 (A) and cisNTS1 (B) ligands (white sticks) on rNTS1-H1 model structure after 10 ns simulation time. The backbone conformations of docked ligands are shown as ribbon in red colour. Rendered in PyMOL.

iii) Combination of AutoDock4 and MD simulations approach

The same four sets of protein-ligand complexes as used in MD simulations approach were studied by AutoDock4 and three independent experiments of each set were done. In addition to the fully flexible ligands, side-chains of residues Met208, Trp339 and Tyr347 of rNTS1, which are involved in ligand binding according to the SDM data, were freely rotated during the docking simulations. These three residues providing 8 torsion angles in total were selected to obtain appropriate number of torsions. The total torsion angles of NT_{8-13} are 24 and the allowed number of torsions is 32. One hundred and fifty conformations of bound NT_{8-13} from combination of three independent experiments were analysed and ranked by binding energy. Most of them are located close to the EC3 loop as expected. The position between TMs I, II and VII was also found as the second most-populous location but in disagreement with experimentally ascertained data. The five energetically most favourable docking poses of each complex were selected for further analysis and only docking positions of ligands in agreement with mutational data are summarised in **Table 6.5**.

Table 6.5: The results of the calculated free energy of binding of each complex from docking experiments using AutoDock4 docking program.

Complex	ΔG_b (kcal/mol)	Rank	Range of ΔG_b (kcal/mol)
rNTS1-B1 – transNT6	-10.50	1	from -10.50 to -3.91
	-9.32	4	
	-9.13	5	
rNTS1-B1 – cisNT6	-9.66	3	from -10.33 to -3.06
	-9.42	5	
rNTS1-H1 – transNT6	-8.81	2	from -8.94 to -0.75
	-8.51	3	
	-7.97	4	
	-7.87	5	
rNTS1-H1 – cisNT6	-9.22	1	from -9.22 to -0.79
	-7.53	2	
	-7.52	3	
	-6.61	5	

In general, the conformations of Pro10 (*cis*/*trans*) did not give significant differences in AutoDock results. Although the ligands showed slightly stronger binding to the rNTS1-B1 than rNTS1-H1, it could not be concluded that the secondary structure of the EC2 loop is involved in ligand recognition or binding. **Figure 6.11** and **6.12** show locations of docked ligands in putative binding sites from AutoDock4 studies of the complexes rNTS1-B1 – transNT6, rNTS1-B1 – cisNT6, rNTS1H1 – transNT6, and rNTS1-H1 – cisNT6. Although the bound ligands have several conformations, most of them are in U-shape conformations. Following the results from **Table 6.5**, selected docked structures from each ensemble were used as starting structures for the MD simulations.

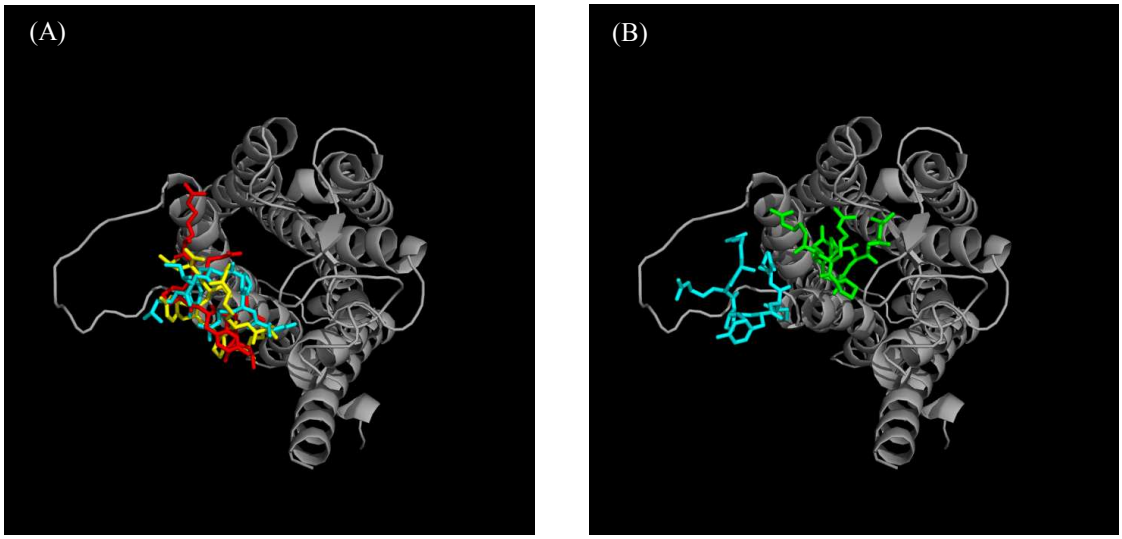


Figure 6.11: The locations of selected transNT6 (A) and cisNT6 (B) ligands in their putative ligand binding sites of the rNTS1-B1 model structure from AutoDock4 studies. The rank orders correspond to the AutoDock data provided in **Table 6.5**. Rank 1 red; rank 3 green; rank 4 yellow; rank 5 cyan. Rendered in PyMOL.

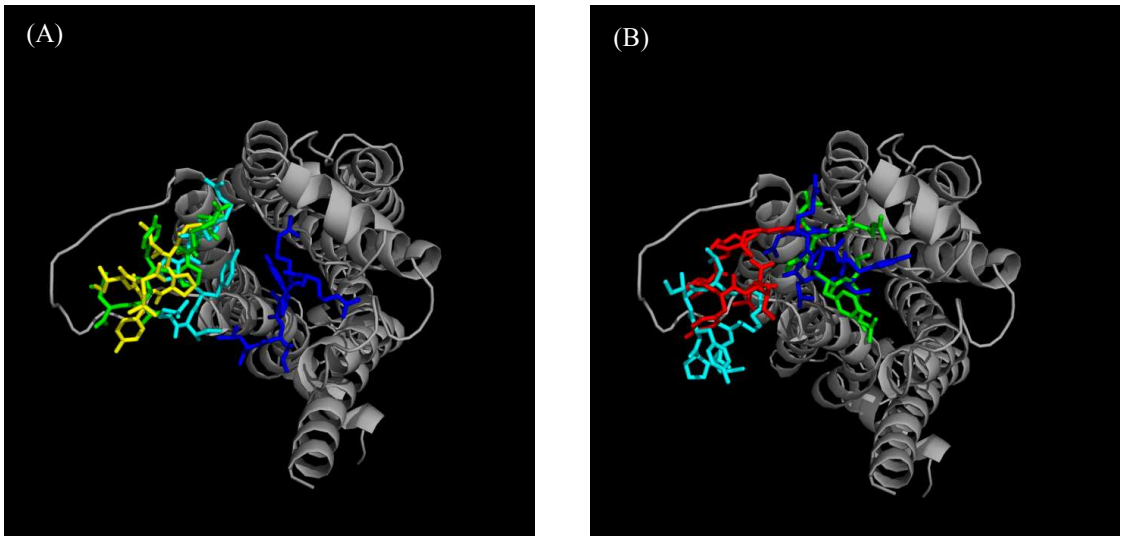


Figure 6.12: The locations of selected transNT6 (A) and cisNT6 (B) ligands in their putative ligand binding sites of rNTS1-H1 model structure from AutoDock4 studies. The rank orders correspond to the AutoDock data provided in **Table 6.5**. Rank 1 red; rank 2 blue; rank 3 green; rank 4 yellow; rank 5 cyan. Rendered in PyMOL.

rNTS1-B1 – transNT6 complex

The pre-defined docked structures from AutoDock4 studies of rNTS1-B1 – transNT6 complexes were subjected to 10 ns MD simulations. Rank 1 and 4 moved slightly to the middle of EC2 and EC3 loops, whereas the location of rank 5 was pretty much the same as before during 10 ns of unrestrained MD simulations. These three docked ligands shared similar conformations with a U-shape structure, as observed in fully-flexible docking studies of rNTS1-H1 – cisNT6 complex. Also, their Pro10-conformations remained stable with dihedral angles close to -180 degrees. Throughout the simulations, structure determining hydrogen bonds were observed. Almost all of hydrogen bonds observed arose from residues which show indirect involvement in ligand binding such as Arg213, Asp216, Asp345, Thr349, and only a few critical residues for ligand binding were found during simulations, for example, the two residues Phe331 and Trp339, and three residues Arg327, Trp339 and Tyr347 for rank

1 and 4, respectively (see **Table 6.4** for SDM data). Even though their positions are slightly different, all docked ligands had interactions with both EC2 and EC3 loops.

rNTS1-B1 – cisNT6 complex

After 10 ns of unrestrained MD simulations of the rNTS1-B1 – cisNT6 complexes, rank 5 which located next to the EC3 loop moved to the middle of EC2 and EC3 loops where rank 3 had been, and their structures were still in a U-shape conformation. The dihedral angles of Pro10 in both docked ligands were close to -90 degrees *i.e.* -91.3 degrees for rank 3 and -88.4 degrees for rank 5. Structure determining hydrogen bonds observed during the simulations were very similar to those of the rNTS1-B1 – transNT6 complex.

rNTS1-H1 – transNT6 complex

For the rNTS1-H1 – transNT6 complexes, rank 2, 4 and 5 but not 3 (extended conformation) had U-shape conformations at the putative binding site on the EC3 loop of rNTS1-H1 model. The *trans* isomer of Pro10 was not changed for all docked ligands. Only rank 2 showed interactions (H-bonds) with all three extracellular loops, while the others did not form any hydrogen bonds to residues in the EC1 loop throughout MD simulations.

rNTS1-H1 – cisNT6 complex

All docked ligands remained where they were during 10 ns of unrestrained MD simulations for the rNTS1-H1 – cisNT6 complexes. Only rank 3 adopted a S-shape conformation upon ligand binding, whereas the other three docked ligands remained as

U-shape conformations. The dihedral angles of Pro10-docked ligands were -90.7, -91.9, -90.3, and -91.7 degrees for rank 1, 2, 3, and 5, respectively. Hydrogen bonds between EC2 and EC3 loop residues and docked ligands were observed during simulations.

Summary

Figure 6.13 shows examples of pre-defined rNTS1-B1 – transNT6 and rNTS1-B1 – cisNT6 complexes after 10 ns MD simulations. In summary, the ligands more likely adopted a U-shape conformation when bound to the receptor models and their binding sites seemed to be between EC2 and EC3 loops which correspond to mutational data (Botto et al., 1997; Labbe-Jullie et al., 1998). Throughout 10 ns simulations time, the docked ligands always had hydrogen bonding interactions with residues in the EC2 and EC3 loops wherever they were. These preliminary MD data have suggested that the EC1 loop might not be involved in ligand binding. None of residues in EC1 loop of rNTS1 receptor except Glu124 was studied by site-directed mutagenesis. According to highly comparable rat and human species of NTS1 receptor (the similarity and identity are 92% and 84%, respectively (Vita et al., 1993)), inclusion of mutagenesis data of the hNTS1 receptor in EC1 loop was considered. Recently, alanine scanning mutagenesis of EC1 loop residues of hNTS1 receptor together with docking studies have suggested that the EC1 loop does not take part in ligand binding, and stabilising of this loop is involved in controlling the shape of the binding site crevice (**Figure 6.14**) (Harterich et al., 2008) which correlates with our observation.

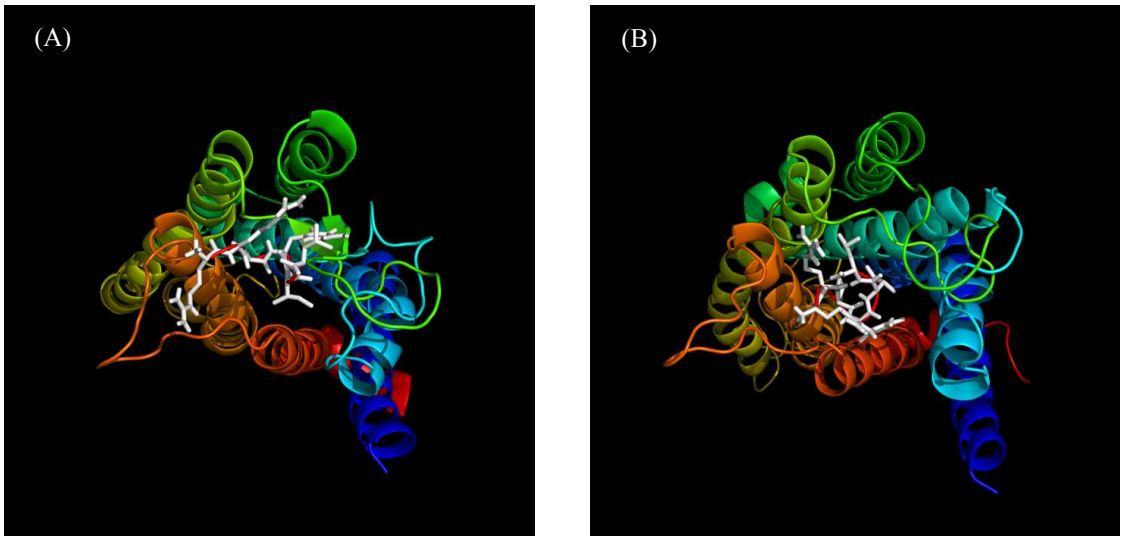


Figure 6.13: Putative binding sites of transNT6 and cisNT6 ligands (white sticks) on the rNTS1-B1 model proposed by the combination of AutoDock4 and MD simulations approach. The backbone conformations of docked ligands are shown as ribbon in red colour. Rendered in PyMOL.

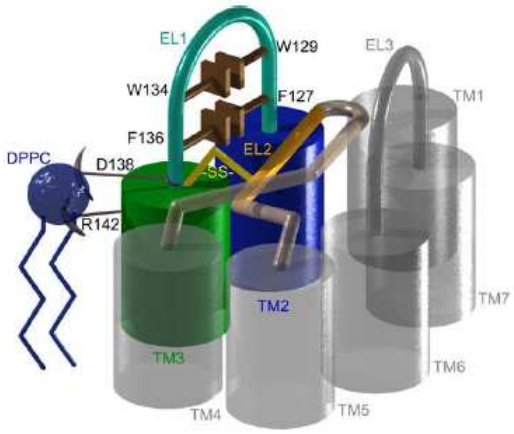


Figure 6.14: hNTS1 model showing the functional role of EC1 loop. The rigid EC1 loop resulted from aromatic π-stacking is proposed to stabilise the extracellular regions. Charge residues (Asp138 and Arg142) anchor TM III to the hydrophilic head group of DPPC molecules, and a disulfide bridge is formed between TM III and EC2 loop (Harterich et al., 2008).

No significant difference of docking results from using rNTS1-B1 and rNTS1-H1 models was observed, but the helix structure of EC2 loop of rNTS1-H1 model disappeared upon transNT6 and cisNT6 binding during simulations, whereas the β

structure of EC2 loop of rNTS1-B1 remained after the binding of transNT6 but not the cisNT6. The *cis* isomer generally exists as a minor population of amino acids including proline. Interestingly, from Autodock4 studies, the Pro10 in *cis* conformation showed high binding affinity to the receptor models at similar level to the *trans* one, and its high affinity still remained after 10 ns of unrestrained MD simulations. However, it has been noted that upon the binding the dihedral angles of Pro10 in *cis* conformation were changed from -0.5 degrees to the angles close to -90 degrees for all docked ligands, whereas that of the transNT6 remained stable at approximately -180 degrees.

6.4 Conclusions

From the mutagenesis and solid-state NMR data, it can be seen that a little is known about the binding site of the rNTS1 receptor. Here, attempts have been made to produce reliable model structures of rNTS1 which can be used to determine structural conformation of the ligand at putative binding sites. Also, it will be useful for rationalising future SSNMR data and designing mutagenesis experiments. To observe whether the structure of EC2 loop is involved in ligand binding, two acceptable rNTS1 homology models, rNTS1-B1 which has β structure at EC2 loop and rNTS1-H1 which has helix structure at EC2 loop, were generated by using the crystal structure of turkey β_1 adrenergic receptor (β_1AR) as the main template. Since the β_1AR structure which has helix structure at EC2 loop lacks the IC3 loop, the crystal structure of squid rhodopsin (sRho) was used for modelling the IC3 loop part to obtain complete model structures and generating β structure at EC2 loop for the rNTS1-B1 model. For ligands, NT_{8-13} models (transNT6 and cisNT6) were constructed by using the crystal

structure of bound NT_{10-13} to the NTS3 receptor and their differences are the conformations of Pro10 (*cis*/*trans*) in the ligand sequence.

Ligands were docked into the rNTS1 models by using two approaches, MD simulations and combination of AutoDock4 and MD simulations. High affinity binding of ligands to the EC3 loop of receptors was observed in both methods. The preliminary studies of the binding site of ligands proposed that the ligands are likely in the U-shape conformations in agreement with the observation that ligands with a turn structure are recognised by GPCRs (Tyndall et al., 2005), and the docked ligands occupy the position between EC2 and EC3 loops as putative binding sites. Although the EC2 loop is probably involved in ligand binding, the importance of the secondary structure of this loop cannot be concluded. The Pro10 of NT_{8-13} ligand may adopt either *cis* or *trans* conformation upon the binding, so direct experimental evidence is still required.

Modelled protein-ligand complexes have to be further optimised by using experimental data to determine specific interactions between proteins and ligands. Antagonists for example, SR49862 can be also used to refine the side-chain geometry of receptor models or test how well these models are suited for antagonist binding. These models might be more suitable for antagonist binding than agonist binding because the models were generated using the β_1AR crystal structure bound with an antagonist (inactive conformation). Distance restraints between ligands and receptors using mutagenesis data will be applied in MD simulations, and ligand-binding free energies after simulations will be calculated.

Chapter 7

Conclusions and future work

So far, there are five crystal structures of inactivated GPCRs, and only the high resolution structures of bovine rhodopsin and squid rhodopsin, which are not ligand-activated GPCRs, are complete. The other three peptide-activated GPCR structures have been determined together with either antagonist or inverse agonist, so at present, no high resolution structure of a GPCR-agonist complex is available. Structural data of an agonist bound to its receptor will be valuable information for structure-based design of therapeutics in drug discovery.

In this thesis, rat NTS1, a member of the class A GPCR family was studied. NTS1 is of particular interest as it has been identified as a potential drug target with critical roles in signalling networks involving medical conditions such as pain, eating behaviour, Schizophrenia, Parkinson's disease, colon cancer, and many others (Cusack et al., 1993; Kitabgi et al., 1989; Tanaka et al., 1990). Several groups have attempted to crystallise the NTS1, but diffraction-quality crystals have not yet been obtained since it is not stable in detergent solution and can adopt numerous potential conformations. The ultimate goals of this project were to investigate the NTS1-NT interactions and to resolve the structure of NT bound to the NTS1 receptor using NMR spectroscopy. In order to study the conformation of the small NT ligand in the presence of the large receptor by NMR techniques, the ligand must be isotopically

labelled to improve the signal-to-noise ratio and sensitivity. Two methods, solid phase peptide synthesis (SPPS) (Chapter 2) and an *E. coli* expression system (Chapter 3), were used to produce a set of labelled NT peptides. The former method was employed to synthesise specifically labelled peptides and the latter one was developed to facilitate uniform isotope labelling at low cost. Although it has been shown that only the six C-terminal amino acids of NT (NT_{8-13}) are sufficient for ligand binding (Kitabgi et al., 1977), both labelled NT and NT_{8-13} were produced to assess the importance if any of the NT_{1-7} fragment.

Two selective labelling schemes were specially designed, and the NT peptides synthesised and purified to solve the conformation of bound NT_{8-13} by RR and/or REDOR SSNMR distance measurements (section 2.3.2; **Figure 2.9 and 2.14**). The first specifically labelled peptide, ^{13}Cα-Arg9 and ^{13}CO-, ^{13}Cβ-Pro10 enriched NT_{8-13}, was designed through MD simulations to determine whether Pro10 in the NT_{8-13} sequence is in *cis* or *trans* isomer, which is expected to play an important role in peptide conformation. This labelled peptide was successfully synthesised by using manually-synthesised Fmoc-Pbf-[2-^{13}C]-L-arginine and Fmoc-[1,3-$^{13}C_2$]-L-proline, which are not commercially available for SPPS. The second selectively labelled peptide, ^{15}N-Pro10, ^{13}Cα-Tyr11, ^{13}CO-Ile12, and ^{13}Cγ-Leu13 enriched NT_{8-13}, was designed based on available labelled amino acids in the market to determine the conformation of the four C-terminal residues of NT_{8-13}. Both of these labelled ligands have been characterised by several methods including ESI mass spectrometry, reverse-phase HPLC and solNMR.

From site-directed mutagenesis studies together with a heuristic method, the phenol group of Tyr11 establishes an intra hydrogen bond with the guanidinium group

of Arg9 when the NT_{8-13} is bound to the NTS1 receptor (Pang et al., 1996). Moreover, ^{13}C-chemical shift changes of Tyr11 side-chain were observed after addition of the receptor in solid-state NMR experiments (Williamson et al., 2002). Even though it was proposed that the side-chain of Tyr11 interacts with the receptor at the binding site, this chemical shift perturbation can result from inter and/or intra interactions. To investigate whether there is an intra interaction between the side-chains of Arg9 and Tyr11, selective labelling at U-^{15}N-Arg9 and phenol-4-^{13}C-Tyr11 of NT_{8-13} is proposed based on commercially-available labelled amino acids (see **Figure 7.1**). Correlation crosspeaks and distance measurements can be obtained by DARR and REDOR SSNMR experiments, respectively, if there is an interaction between these two residues in the bound state of NT_{8-13}.

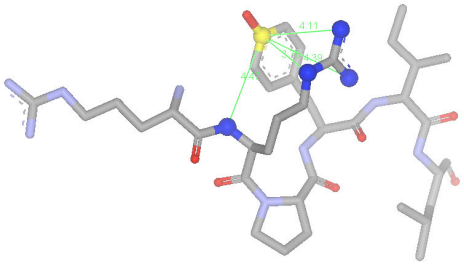


Figure 7.1: Model of U-^{15}N-Arg9 and phenol-4-^{13}C-Tyr11 enriched NT_{8-13} showing the ^{13}C (in yellow spheres) with carbon atoms in grey, nitrogen atoms in blue, and oxygen atoms in red. The green lines and numbers indicate the potential distance (Å) of the labelled amino acids (measured in DS ViewerPro 5.0 (Accelrys Inc.)). The model was generated by DeepView - Swiss-Pdb Viewer and rendered in VMD.

Isotope labelling is an expensive approach, especially for uniform isotope labelling. Expression of proteins or peptides in a bacterial system is an attractive method to produce uniformly labelled samples because the cost of starting materials (^{13}C and ^{15}N sources) is reasonable, while the labelled amino acids are extremely

expensive for the SPPS method. Here, the NT peptide was successfully expressed in *E. coli* with high yield of up to 9.2 mg/L M9 using pGEV-1 plasmid (section 3.3.3 and 3.3.4). Also, presented for the first time is the expression of NT_{8-13} peptide in a bacterial system by using the pGEV-NT plasmid as a template for P7M mutation, yielding the NT_{8-13} 4.2 mg/L M9 after CNBr cleavage and HPLC purification (section 3.3.3 and 3.3.4). Even if these two constructs provide good yields of peptide, it might be possible to obtain higher product yield by expressing peptides as tandem repeats, proposed by Shen (Shen, 1984). Multimers of fusion proteins of neuropeptide substance P and antimicrobial peptide buforin II have improved the peptide yields (Kempe et al., 1985; Lee et al., 1998).

Incorporation of ^{13}C and/or ^{15}N into the NT and NT_{8-13} peptides was performed (section 3.3.5). U-^{15}N-labelled NT, U-^{13}C, ^{15}N-labelled NT and U-^{13}C, ^{15}N-labelled NT_{8-13} were produced in sufficient quantities for structural studies by NMR spectroscopy. Expression of NT peptides in deuterated media was also developed and U-^{2}H labelled NT and U-^{2}H, ^{13}C, ^{15}N labelled NT were produced (section 3.3.6). The efficiency of isotope labelling was verified by ESI mass spectrometry and HSQC solNMR experiments. The ^{15}N-^{1}H HSQC spectra of U-^{13}C, ^{15}N-labelled NT and NT_{8-13} show the sample heterogeneity in aqueous solution (**Figure 3.16 and 3.17**).

Several batches of 40 g NTS1 pellet were used to prepare sufficient quantities of NTS1 receptor for NMR studies (Chapter 4). To probe NTS1-NT interactions by solNMR experiments, the NTS1 receptor was separated from the fusion tags by TEV cleavage in order to reduce the size of the protein sample. For SSNMR studies, the spectral resolution is not relative to the size of a macromolecule or complex, so

detergent purified NTS1B was directly used for lipid bilayer reconstitution providing the receptor in a native environment.

As a lyophilised sample usually gives spectra with poor resolution in SSNMR studies (Martin and Zilm, 2003), the sample in micro/nanocrystalline form is an alternative form to provide narrow spectral linewidths. The unlabelled NT_{8-13} peptide was crystallised by 20% PEG 4000, 50 mM sodium citrate pH 6.0 and 20% isopropanol from a crystallisation kit at 4°C in the sealed yellow pipette tip. Although the peptide was precipitated in this reagent without any effects on the conformation, the NT_{8-13} precipitate did not give a ^{13}C CP MAS spectrum with high resolution. However, other choices of precipitants might give better spectral linewidths.

Chemical shift assignments of NT and NT_{8-13} ligands in solution must be performed prior to NMR studies of the NTS1-NT interactions to allow subsequent comparisons with the resonances obtained when the ligands are bound to the receptor. ^{13}C and ^{1}H chemical shifts of unlabelled NT_{8-13} in pure D_2O from solNMR experiments were unambiguously assigned (section 5.3.2). The *cis*/*trans* isomerisation of Pro10 was observed in the TOCSY and ROESY spectra, with the *cis* conformation existing in a minority of Pro10 signals. However, ^{13}C chemical shifts of the *cis* and *trans* conformations of Pro10 could not be distinguished in the ^{1}H-^{13}C HSQC spectrum according to the crosspeak overlap except for Cα and Cδ (**Figure 5.16**).

U-^{13}C, ^{15}N-labelled NT in detergent buffer was used to facilitate complete backbone and side-chain assignments by triple resonance solNMR experiments (section 5.3.4). The chemical shifts of some nitrogens were not detected because of a rapid amide proton exchange rate with the solvent or dynamics of the peptide. Again,

an additional minor set of low-level resonances corresponding to the presence of *cis* isomer of Pro10 in the NT sequence was observed.

Although chemical shift perturbations in the ^{1}H-^{15}N HSQC spectrum were observed for the U-^{15}N-labelled NT ligand upon binding to the receptor protein in solution at 25 °C, the spectral linewidths of these resonances were not broad as expected (section 5.3.5; **Figure 5.21**). Also, the missing crosspeaks of the labelled ligand in the presence of the receptor were observed in the ^{1}H-^{15}N HSQC spectrum of the same labelled peptide in the absence of receptor as well suggesting dynamic effects of ligand in detergent buffer. These results have revealed that the protein might be aggregating during the experiments and the obtained resonances arising from the ligand in the free form. Therefore, probing the NTS1-NT interactions by solNMR spectroscopy would require a thermostable receptor mutant such as that studied by Shibata and his co-workers (Shibata et al., 2009). A thermostable NTS1 mutant in both the absence and the presence of the peptide agonist NT, suitable for crystallisation was constructed, but its ability to couple to $G_{\alpha q}$ was very poor. Once a functional thermostable NTS1 is available, the U-^{2}H, ^{13}C, ^{15}N labelled NT will be used for structural studies of ligand bound to the receptor.

NTS1-NT interactions were also studied by solid-state NMR experiments (section 5.3.6). ^{13}C-^{13}C COSY spectrum of U-^{13}C, ^{15}N-labelled NT with BPL gives chemical shifts identical to the one of the same labelled peptide but in detergent buffer suggesting that the ligand has no interactions with the lipid used for protein reconstitution. 1D ^{13}C CP MAS spectrum of the U-^{13}C, ^{15}N-labelled NT with BPL referred to as a negative control was compared with the one of the labelled peptide in the presence of reconstituted NTS1B receptor. There is only one resonance (144 ppm)

observed in the spectrum of bound ligand likely arising from the $^{13}C\zeta$ side-chain of Arg and/or Tyr residue. However, a crosspeak of this resonance cannot be observed in the short scan of 2D DARR experiments probably because the number of scans is not high enough. Also, other crosspeaks might be overlapped with the large area of diagonal region. If the resonance at 144 ppm is derived from ligand binding, the proteoliposome alone is needed to verify whether this observed resonance arises from the labelled ligand and is not due to natural abundance signals from the receptor.

To increase SSNMR signals from the bound ligand relating to the number of active receptor, the protein to lipid ratio for protein reconstitution should be adjusted. Using frozen detergent-purified receptor might be an alternative choice even it is not in the native environment. Also, a 6-mm SSNMR rotor (maximum capacity is ~240 μL) could be used to accommodate a larger amount of sample.

Computer modelling was also used to rationalise all available experimental data at the nanoscale level (Chapter 6). rNTS1-B1 and rNTS1-H1 models which have β and helix structures at the EC2 loop, respectively, were generated by using the crystal structure of turkey β_1 adrenergic receptor as the main template. Ligands which have *trans* (transNT6) and *cis* (cisNT6) isomers at Pro10 of $NT_{8\text{-}13}$ were docked onto these receptor models and the results revealed high affinity binding of both ligands to the rat NTS1 models. It has been suggested that the conformation of Pro10 can be either *cis* or *trans*. From our preliminary investigation, the ligands adopted bent structures at the head and tail of the sequence giving a U-shape structure at the putative binding sites between EC2 and EC3 loops. However, more experiments are required to determine whether the secondary structure of the EC2 loop is involved in ligand binding. Also, the receptor models have to be further refined to obtain reliable

protein-ligand complexes using available experimental data. Distance-restraining MD simulations at the protein binding site will be applied.

Here, new strategies for labelling peptides have been developed to resolve the conformation of a clinically important bound peptide at its site of action. Several labelled NT and NT_{8-13} were produced to probe the NTS1-NT interactions and solve the conformation of the bound ligand. The NTS1B receptor was successfully reconstituted at a level suitable for SSNMR studies and will be invaluable in future structural studies. The major challenge will be to find conditions to obtain spectra with high resolution. Although crystal structures with bound peptides may be solved in time, NMR also gives dynamics, and can be performed on a membrane-embedded, fully functional receptor.

References

Ader, C., R. Schneider, K. Seidel, M. Etzkorn, and M. Baldus. 2007. Magic-angle-spinning NMR spectroscopy applied to small molecules and peptides in lipid bilayers. *Biochem Soc Trans*. 35:991-5.

Alewijnse, A.E., H. Timmerman, E.H. Jacobs, M.J. Smit, E. Roovers, S. Cotecchia, and R. Leurs. 2000. The effect of mutations in the DRY motif on the constitutive activity and structural instability of the histamine H(2) receptor. *Mol Pharmacol*. 57:890-8.

Alexandrescu, A.T., K. Rathgeb-Szabo, K. Rumpel, W. Jahnke, T. Schulthess, and R.A. Kammerer. 1998. 15N backbone dynamics of the S-peptide from ribonuclease A in its free and S-protein bound forms: toward a site-specific analysis of entropy changes upon folding. *Protein Sci*. 7:389-402.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403-10.

Andrew, E.R., A. Bradbury, and R.G. Eades. 1958. Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature*. 4650:1659.

Attrill, H., P.J. Harding, E. Smith, S. Ross, and A. Watts. 2009. Improved yield of a ligand-binding GPCR expressed in E. coli for structural studies. *Protein Expression and Purification*. 64:32-38.

Bajorath, J., R. Stenkamp, and A. Aruffo. 1993. Knowledge-based model building of proteins: concepts and examples. *Protein Sci*. 2:1798-810.

Baker, K.A., C. Tzitzilonis, W. Kwiatkowski, S. Choe, and R. Riek. 2007. Conformational dynamics of the KcsA potassium channel governs gating properties. *Nat Struct Mol Biol*. 14:1089-95.

Bao, W.J., Y.G. Gao, Y.G. Chang, T.Y. Zhang, X.J. Lin, X.Z. Yan, and H.Y. Hu. 2006. Highly efficient expression and purification system of small-size protein domains in Escherichia coli for biochemical characterization. *Protein Expr Purif*. 47:599-606.

Barak, L.S., L. Menard, S.S. Ferguson, A.M. Colapietro, and M.G. Caron. 1995. The conserved seven-transmembrane sequence NP(X)2,3Y of the G-protein-coupled receptor superfamily regulates multiple properties of the beta 2-adrenergic receptor. *Biochemistry*. 34:15407-14.

Barroso, S., F. Richard, D. Nicolas-Etheve, P. Kitabgi, and C. Labbe-Jullie. 2002. Constitutive activation of the neurotensin receptor 1 by mutation of Phe(358) in Helix seven. *Br J Pharmacol*. 135:997-1002.

Barroso, S., F. Richard, D. Nicolas-Etheve, J.L. Reversat, J.M. Bernassau, P. Kitabgi, and C. Labbe-Jullie. 2000. Identification of residues involved in neurotensin binding and modeling of the agonist binding site in neurotensin receptor 1. *J Biol Chem*. 275:328-36.

Bennett, A.E., R.G. Griffin, J.H. Ok, and S. Vega. 1992. Chemical shift correlation spectroscopy in rotating solids: Radio frequency-driven dipolar recoupling and longitudinal exchange. *J Chem Phys*. 96:8624-8627.

Bennett, A.E., H.M. Rienstra, M. Auger, K.V. Lakshmi, and R.G. Griffin. 1995. Heteronuclear decoupling in rotating solids. *J Chem Phys*. 103:6951–6957.

Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690.

Berman, H., K. Henrick, and H. Nakamura. 2003. Announcing the worldwide Protein Data Bank. *Nat Struct Biol*. 10:980.

Blakeney, J.S., and D.P. Fairlie. 2005. Nonpeptide ligands that target peptide-activated GPCRs in inflammation. *Curr Med Chem*. 12:3027-42.

Bockaert, J., P. Marin, A. Dumuis, and L. Fagni. 2003. The 'magic tail' of G protein-coupled receptors: an anchorage for functional protein networks. *FEBS Lett*. 546:65-72.

Boesch, C., A. Bundi, M. Oppliger, and K. Wuthrich. 1978. 1H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur J Biochem*. 91:209-14.

Bonneau, R., N.S. Baliga, E.W. Deutsch, P. Shannon, and L. Hood. 2004. Comprehensive de novo structure prediction in a systems-biology context for the archaea Halobacterium sp. NRC-1. *Genome Biol*. 5:R52.

Bonneau, R., I. Ruczinski, J. Tsai, and D. Baker. 2002. Contact order and ab initio protein structure prediction. *Protein Sci*. 11:1937-44.

Botto, J.M., J. Chabry, D. Nouel, M. Paquet, P. Seguela, J.P. Vincent, A. Beaudet, and J. Mazella. 1997a. Identification in the rat neurotensin receptor of amino-acid residues critical for the binding of neurotensin. *Brain Res Mol Brain Res*. 46:311-7.

Botto, J.M., J. Chabry, P. Sarret, J.P. Vincent, and J. Mazella. 1998. Stable expression of the mouse levocabastine-sensitive neurotensin receptor in HEK 293 cell line: binding properties, photoaffinity labeling, and internalization mechanism. *Biochem Biophys Res Commun*. 243:585-90.

Botto, J.M., E. Guillemare, J.P. Vincent, and J. Mazella. 1997b. Effects of SR 48692 on neurotensin-induced calcium-activated chloride currents in the Xenopus oocyte expression system: agonist-like activity on the levocabastine-sensitive neurotensin receptor and absence of antagonist effect on the levocabastine insensitive neurotensin receptor. *Neurosci Lett*. 223:193-6.

Braun, W., G. Wider, K.H. Lee, and K. Wuthrich. 1983. Conformation of glucagon in a lipid-water interphase by 1H nuclear magnetic resonance. *J Mol Biol*. 169:921-48.

Brooijmans, N., and I.D. Kuntz. 2003. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct*. 32:335-73.

Brown, M.F., K. Martinez-Mayorga, K. Nakanishi, G.F. Salgado, and A.V. Struts. 2009. Retinal conformation and dynamics in activation of rhodopsin illuminated by solid-state H NMR spectroscopy. *Photochem Photobiol*. 85:442-53.

Brown, S.P. 2009. Recent Advances in Solid-State MAS NMR Methodology for Probing Structure and Dynamics in Polymeric and Supramolecular Systems. *Macromolecular Rapid Communications*. 30:688-716.

Brus, J., and A. Jegorov. 2004. Through-Bonds and Through-Space Solid-State NMR Correlations at Natural Isotopic Abundance: Signal Assignment and Structural Study of Simvastatin. *J Phys Chem A*. 108:3955–3964.

Burstein, E.S., T.A. Spalding, and M.R. Brann. 1998a. The second intracellular loop of the m5 muscarinic receptor is the switch which enables G-protein coupling. *J Biol Chem*. 273:24322-7.

Burstein, E.S., T.A. Spalding, and M.R. Brann. 1998b. Structure/function relationships of a G-protein coupling pocket formed by the third intracellular loop of the m5 muscarinic receptor. *Biochemistry*. 37:4052-8.

Campbell, A.P., D.L. Bautista, B. Tripet, W.Y. Wong, R.T. Irvin, R.S. Hodges, and B.D. Sykes. 1997. Solution secondary structure of a bacterially expressed peptide from the receptor binding domain of Pseudomonas aeruginosa pili strain PAK: A heteronuclear multidimensional NMR study. *Biochemistry*. 36:12791-801.

Carraway, R., and S.E. Leeman. 1973. The isolation of a new hypotensive peptide, neurotensin, from bovine hypothalami. *J Biol Chem*. 248:6854-61.

Casciari, D., D. Dell'Orco, and F. Fanelli. 2008. Homodimerization of neurotensin 1 receptor involves helices 1, 2, and 4: insights from quaternary structure predictions and dimerization free energy estimations. *J Chem Inf Model*. 48:1669-78.

Cavanagh, J., W.J. Fairbrother, A.G. Palmer III, M. Rance, and N.J. Skelton. 2007. Protein NMR spectroscopy: principles and practice. Boston : Academic Press, Amsterdam. 885 pp.

Ceresa, B.P., and L.E. Limbird. 1994. Mutation of an aspartate residue highly conserved among G-protein-coupled receptors results in nonreciprocal disruption of alpha 2-adrenergic receptor-G-protein interactions. A negative charge at amino acid residue 79 forecasts alpha 2A-adrenergic receptor sensitivity to allosteric modulation by monovalent cations and fully effective receptor/G-protein coupling. *J Biol Chem*. 269:29557-64.

Chabre, M., and M. le Maires. 2005. Monomeric G-protein-coupled receptor as a functional unit. *Biochemistry*. 44:9395-9403.

Chalon, P., N. Vita, M. Kaghad, M. Guillemot, J. Bonnin, B. Delpech, G. Le Fur, P. Ferrara, and D. Caput. 1996. Molecular cloning of a levocabastine-sensitive neurotensin binding site. *FEBS Lett*. 386:91-4.

Chelikani, P., P.J. Reeves, U.L. Rajbhandary, and H.G. Khorana. 2006. The synthesis and high-level expression of a 2-adrenergic receptor gene in a tetracycline-inducible stable mammalian cell line. *Protein Science: A Publication of the Protein Society*. 15:1433.

Cherezov, V., D.M. Rosenbaum, M.A. Hanson, S.G. Rasmussen, F.S. Thian, T.S. Kobilka, H.J. Choi, P. Kuhn, W.I. Weis, B.K. Kobilka, and R.C. Stevens. 2007. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*. 318:1258-65.

Cho, W., and R.V. Stahelin. 2005. Membrane-protein interactions in cell signaling and membrane trafficking. *Annu Rev Biophys Biomol Struct*. 34:119-51.

Choi, G., J. Landin, J.F. Galan, R.R. Birge, A.D. Albert, and P.L. Yeagle. 2002. Structural studies of metarhodopsin II, the activated form of the G-protein coupled receptor, rhodopsin. *Biochemistry*. 41:7318-24.

Chothia, C., and A.M. Lesk. 1986. The relation between the divergence of sequence and structure in proteins. *Embo J*. 5:823-6.

Chou, P.Y., and G.D. Fasman. 1975. The conformation of glucagon: predictions and consequences. *Biochemistry*. 14:2536-41.

Cornilescu, G., F. Delaglio, and A. Bax. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*. 13:289-302.

Curioni, C., and C. Andre. 2006. Rimonabant for overweight or obesity. *Cochrane Database Syst Rev*:CD006162.

Cusack, B., E. Richelson, Y.P. Pang, J. Zaidi, and A.P. Kozikowski. 1993. Pharmacological studies on novel neurotensin mimetics: discovery of a pharmacologically unique agent exhibiting concentration-dependent dual effects as antagonist and agonist. *Mol Pharmacol*. 44:1036-40.

Darden, T., D. York, and L.G. Pedersen. 1993. Particle mesh Ewald—an N.log(N) method for Ewald sums in large systems. *J Chem Phys*. 98:10089–10092.

Delaglio, F., S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A.D. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 6:277–293.

Desai, U.R., J.J. Osterhout, and A.M. Klibanov. 1994. Protein Structure in the Lyophilized State: A Hydrogen Isotope Exchange/NMR Study with Bovine Pancreatic Trypsin Inhibitor. *J Am Chem SOC*. 116:9420-9422.

DesJarlais, R.L., R.P. Sheridan, J.S. Dixon, I.D. Kuntz, and R. Venkataraghavan. 1986. Docking flexible ligands to macromolecular receptors by molecular shape. *J Med Chem*. 29:2149-53.

Dube, D.H., and C.R. Bertozzi. 2005. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov*. 4:477-88.

Duer, M.J. 2002. Solid State NMR Spectroscopy: Principles and Applications. Wiley-Blackwell. 567 pp.

Edwards, P.C., J. Li, M. Burghammer, J.H. McDowell, C. Villa, P.A. Hargrave, and G.F. Schertler. 2004. Crystals of native and modified bovine rhodopsins and their heavy atom derivatives. *J Mol Biol*. 343:1439-50.

Epand, R.M., A.J. Jones, and B. Sayer. 1977. Molecular interactions in the model lipoprotein complex formed between glucagon and dimyristoylglycerophosphocholine. *Biochemistry*. 16:4360-8.

Erickson, J.A., M. Jalaie, D.H. Robertson, R.A. Lewis, and M. Vieth. 2004. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J Med Chem*. 47:45-55.

Ernst, M. 2003. Heteronuclear spin decoupling in solid-state NMR under magic-angle sample spinning. *J Magn Reson*. 162:1-34.

Ernst, R.R., G. Bodenhausen, and A. Wokaun. 1990. Principles of nuclear magnetic resonance in one and two dimensions. Oxford University Press, USA.

Eswar, N., B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, and A. Sali. 2007. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*. Chapter 2:Unit 29.

Ewing, T.J., S. Makino, A.G. Skillman, and I.D. Kuntz. 2001. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*. 15:411-28.

Fatkenheuer, G., A.L. Pozniak, M.A. Johnson, A. Plettenberg, S. Staszewski, A.I. Hoepelman, M.S. Saag, F.D. Goebel, J.K. Rockstroh, B.J. Dezube, T.M. Jenkins, C. Medhurst, J.F. Sullivan, C. Ridgway, S. Abel, I.T. James, M. Youle, and E. van der Ryst. 2005. Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. *Nat Med*. 11:1170-2.

Fermandjian, S., P. Fromageot, A.M. Tistchenko, J.P. Leicknam, and M. Lutz. 1972. Angiotensin II conformations. Infrared and raman studies. *Eur J Biochem*. 28:174-82.

Fermandjian, S., J.L. Morgat, and P. Fromageot. 1971. Studies of angiotensin-II conformations by circular dichroism. *Eur J Biochem*. 24:252-8.

Fernandez, C., C. Hilty, G. Wider, P. Guntert, and K. Wuthrich. 2004. NMR structure of the integral membrane protein OmpX. *J Mol Biol*. 336:1211-21.

Fielding, L. 2003. NMR methods for the determination of protein-ligand dissociation constants. *Curr Top Med Chem*. 3:39-53.

Fiser, A., and A. Sali. 2003. ModLoop: automated modeling of loops in protein structures. *Bioinformatics*. 19:2500-1.

Fredriksson, R., M.C. Lagerstrom, L.G. Lundin, and H.B. Schioth. 2003. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol*. 63:1256-72.

Friesner, R.A., J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, and P.S. Shenkin. 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 47:1739-49.

Gabilondo, A.M., C. Krasel, and M.J. Lohse. 1996. Mutations of Tyr326 in the beta 2-adrenoceptor disrupt multiple receptor functions. *Eur J Pharmacol*. 307:243-50.

Glaubitz, C., G. Grobner, and A. Watts. 2000. Structural and orientational information of the membrane embedded M13 coat protein by (13)C-MAS NMR spectroscopy. *Biochim Biophys Acta*. 1463:151-61.

Glaubitz, C., and A. Watts. 1998. Magic angle-oriented sample spinning (MAOSS): A new approach toward biomembrane studies. *J Magn Reson*. 130:305-16.

Goedert, M. 1989. Radioligand-binding assays for study of neurotensin receptors. *Methods Enzymol*. 168:462-81.

Gratzer, W.B., E. Bailey, and G.H. Beaven. 1967. Conformational states of glucagon. *Biochem Biophys Res Commun*. 28:914-9.

Gratzer, W.B., and G.H. Beaven. 1969. Relation between conformation and association state. A study of the association equilibrium of glucagon. *J Biol Chem*. 244:6675-9.

Gregory, R.B., M. Gangoda, R.K. Gilpin, and W. Su. 1993a. The Influence of Hydration on the Conformation of Bovine Serum-Albumin Studied by Solid-State C-13-Nmr Spectroscopy. *Biopolymers*. 33:1871-1876.

Gregory, R.B., M. Gangoda, R.K. Gilpin, and W. Su. 1993b. The Influence of Hydration on the Conformation of Lysozyme Studied by Solid-State C-13-Nmr Spectroscopy. *Biopolymers*. 33:513-519.

Grisshammer, R., P. Averbeck, and A.K. Sohal. 1999. Improved purification of a rat neurotensin receptor expressed in Escherichia coli. *Biochem Soc Trans*. 27:899-903.

Grisshammer, R., and S.K. Buchanan. 2006. Structural Biology of Membrane Proteins. Royal Society of Chemistry, Cambridge. 420 pp.

Grisshammer, R., R. Duckworth, and R. Henderson. 1993. Expression of a rat neurotensin receptor in Escherichia coli. *Biochem J*. 295 (Pt 2):571-6.

Grisshammer, R., J.F. White, L.B. Trinh, and J. Shiloach. 2005. Large-scale expression and purification of a G-protein-coupled receptor for structure determination -- an overview. *J Struct Funct Genomics*. 6:159-63.

Grobner, G., I.J. Burnett, C. Glaubitz, G. Choi, A.J. Mason, and A. Watts. 2000. Observations of light-induced structural changes of retinal within rhodopsin. *Nature*. 405:810-3.

Gronenborn, A.M., and G.M. Clore. 1996. Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy. *Protein Sci*. 5:174-7.

Gross, E. 1967. [27] The Cyanogen Bromide Reaction. *Methods Enzymol*. 11:238–255.

Grzesiek, S., J. Anglister, and A. Bax. 1993. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic Mixing of 13C Magnetization *J Magn Reson, Ser. B*. 101:114-119.

Grzesiek, S., and A. Bax. 1992a. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc*. 114:6291-6293.

Grzesiek, S., and A. Bax. 1992b. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins *J Magn Reson*. 99:201-207.

Guex, N., and M.C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 18:2714-23.

Gullion, T. 1998. Introduction to rotational-echo, double-resonance NMR. *Concepts Magn Reson*. 10:277 - 289

Gullion, T., and J. Schaefer. 1989. Detection of weak heteronuclear dipolar coupling by rotational-echo double-resonance nuclear magnetic resonance. *Adv. Magn. Reson.* 13:57-83.

Gully, D., M. Canton, R. Boigegrain, F. Jeanjean, J.C. Molimard, M. Poncelet, C. Gueudet, M. Heaulme, R. Leyris, A. Brouard, and et al. 1993. Biochemical and pharmacological profile of a potent and selective nonpeptide antagonist of the neurotensin receptor. *Proc Natl Acad Sci U S A*. 90:65-9.

Gurevich, V.V., and E.V. Gurevich. 2008. GPCR monomers and oligomers: it takes all kinds. *Trends in neurosciences*. 31:74-81.

Hagen, D.S., J.H. Weiner, and B.D. Sykes. 1979. Investigation of solvent accessibility of the fluorotyrosyl residues of M13 coat protein in deoxycholate micelles and phospholipid vesicles. *Biochemistry*. 18:2007-12.

Hammarstrom, M., E.A. Woestenenk, N. Hellgren, T. Hard, and H. Berglund. 2006. Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J Struct Funct Genomics*. 7:1-14.

Hanson, M.A., and R.C. Stevens. 2009. Discovery of new GPCR biology: one receptor structure at a time. *Structure*. 17:8-14.

Harding, P.J., H. Attrill, J. Boehringer, S. Ross, G.H. Wadhams, E. Smith, J.P. Armitage, and A. Watts. 2009. Constitutive Dimerization of the G-Protein Coupled Receptor, Neurotensin Receptor 1, Reconstituted into Phospholipid Bilayers. *Biophysical Journal*. 96:964-973.

Harrison, R.W., D. Chatterjee, and I.T. Weber. 1995. Analysis of six protein structures predicted by comparative modeling techniques. *Proteins*. 23:463-71.

Harterich, S., S. Koschatzky, J. Einsiedel, and P. Gmeiner. 2008. Novel insights into GPCR-peptide interactions: mutations in extracellular loop 1, ligand backbone methylations and molecular modeling of neurotensin receptor 1. *Bioorg Med Chem*. 16:9359-68.

Hartmann, S.R., and E.L. Hahn. 1962. Nuclear double resonance in the rotating frame. *Phys. Rev.* 128:2042.

Havel, T.F., and M.E. Snow. 1991. A new method for building protein conformations from sequence alignments with homologues of known structure. *J Mol Biol*. 217:1-7.

Hediger, S., B.H. Meier, and R.R. Ernst. 1995. Rotor synchronized amplitude modulated nuclear magnetic resonance spin lock sequences for improved cross polarization under fast magic angle sample spinning. *The Journal of Chemical Physics*. 102:4000.

Hermans, J., H.J.C. Berendsen, W.F. van Gunsteren, and J.P.M. Postma. 1984. A consistent empirical potential for water-protein interactions. *Biopolymers*. 23:1513–1518.

Hess, B., H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *J Comp Chem*. 18:1463–1472.

Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comp.* 4:435-447.

Hetenyi, C., and D. van der Spoel. 2002. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci*. 11:1729 - 1737.

Hiller, M., L. Krabben, K.R. Vinothumar, F. Castellani, B.J. van Rossum, W. Kuhlbrandt, and H. Oschkinat. 2005. Solid-State Magic-Angle Spinning NMR of Outer Membrane Protein G from *Escherichia coli*. *ChemBioChem*. 6:1679-1684.

Hiller, S., R.G. Garces, T.J. Malia, V.Y. Orekhov, M. Colombini, and G. Wagner. 2008. Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science*. 321:1206-10.

Hochuli, E., W. Bannwarth, H. Döbeli, R. Gentz, and D. Stüber. 1988. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio/Technology*. 6:1321-1325.

Hooft, R.W., G. Vriend, C. Sander, and E.E. Abola. 1996. Errors in protein structures. *Nature*. 381:272.

Horn, F., E. Bettler, L. Oliveira, F. Campagne, F.E. Cohen, and G. Vriend. 2003. GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res*. 31:294-7.

Humphrey, P.P., and E.A. Barnard. 1998. International Union of Pharmacology. XIX. The IUPHAR receptor code: a proposal for an alphanumeric classification system. *Pharmacol Rev*. 50:271-7.

Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J Mol Graph*. 14:33-8, 27-8.

Huth, J.R., C.A. Bewley, B.M. Jackson, A.G. Hinnebusch, G.M. Clore, and A.M. Gronenborn. 1997. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci*. 6:2359-64.

Hwang, P.M., R.E. Bishop, and L.E. Kay. 2004. The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc Natl Acad Sci U S A*. 101:9618-23.

Hwang, P.M., W.Y. Choy, E.I. Lo, L. Chen, J.D. Forman-Kay, C.R. Raetz, G.G. Prive, R.E. Bishop, and L.E. Kay. 2002. Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc Natl Acad Sci U S A*. 99:13560-5.

Igumenova, T.I., A.E. McDermott, K.W. Zilm, R.W. Martin, E.K. Paulson, and A.J. Wand. 2004. Assignments of carbon NMR resonances for microcrystalline ubiquitin. *J Am Chem Soc*. 126:6720-7.

Ikonen, E. 2008. Cellular cholesterol trafficking and compartmentalization. *Nat Rev Mol Cell Biol*. 9:125-38.

Ikura, M., L.E. Kay, and A. Bax. 1991. Improved three-dimensional 1H-13C-1H correlation spectroscopy of a 13C-labeled protein using constant-time evolution. *J Biomol NMR*. 1:299-304.

Isaac, B., G.J. Gallagher, Y.S. Balazs, and L.K. Thompson. 2002. Site-directed rotational resonance solid-state NMR distance measurements probe structure and mechanism in the transmembrane domain of the serine bacterial chemoreceptor. *Biochemistry*. 41:3025-36.

Iverson, H.A., D. Fox, 3rd, L.S. Nadler, R.E. Klevit, and N.M. Nathanson. 2005. Identification and structural determination of the M(3) muscarinic acetylcholine receptor basolateral sorting signal. *J Biol Chem*. 280:24568-75.

Jaakola, V.P., M.T. Griffith, M.A. Hanson, V. Cherezov, E.Y. Chien, J.R. Lane, A.P. Ijzerman, and R.C. Stevens. 2008. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*. 322:1211-7.

Jaakola, V.P., J. Prilusky, J.L. Sussman, and A. Goldman. 2005. G protein-coupled receptors show unusual patterns of intrinsic unfolding. Vol. 18. Oxford Univ Press. 103-110.

Jonasson, P., P.A. Nygren, B.L. Johansson, J. Wahren, M. Uhlen, and S. Stahl. 1998. Gene fragment polymerization gives increased yields of recombinant human proinsulin C-peptide. *Gene*. 210:203-10.

Jones, D.H., E.H. Ball, S. Sharpe, K.R. Barber, and C.W. Grant. 2000. Expression and membrane assembly of a transmembrane region from Neu. *Biochemistry*. 39:1870-8.

Jones, G., P. Willett, R.C. Glen, A.R. Leach, and R. Taylor. 1997. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*. 267:727-48.

Kameda, T., Y. Nakazawa, J. Kazuhara, T. Yamane, and T. Asakura. 2002. Determination of intermolecular distance for a model peptide of Bombyx mori silk fibroin, GAGAG, with rotational echo double resonance. *Biopolymers*. 64:80-5.

Kamihira, M., T. Vosegaard, A.J. Mason, S.K. Straus, N.C. Nielsen, and A. Watts. 2005. Structural and orientational constraints of bacteriorhodopsin in purple membranes determined by oriented-sample solid-state NMR spectroscopy. *J Struct Biol*. 149:7-16.

Karlsson, T., and M.H. Levitt. 1998. Longitudinal rotational resonance echoes in solid state nuclear magnetic resonance: Investigation of zero quantum spin dynamics. *J. Chem. Phys*. 109:5493-5507.

Karplus, M. 2002. Molecular dynamics simulations of biomolecules. *Acc Chem Res*. 35:321-3.

Kay, L.E., M. Ikura, R. Tschudin, and A. Bax. 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins *J Magn Reson*. 89:496-514.

Kelly, A.E., H.D. Ou, R. Withers, and V. Dפtsch. 1993. Low-conductivity buffers for high-sensitivity NMR measurements. *Appl. Supercond*. 3:2450-2453.

Kempe, T., S.B. Kent, F. Chow, S.M. Peterson, W.I. Sundquist, J.J. L'Italien, D. Harbrecht, D. Plunkett, and W.J. DeLorbe. 1985. Multiple-copy genes: production and modification of monomeric peptides from large multimeric fusion proteins. *Gene*. 39:239-45.

Khalid, S., P.J. Bond, S.S. Deol, and M.S. Sansom. 2006. Modeling and simulations of a bacterial outer membrane protein: OprF from Pseudomonas aeruginosa. *Proteins*. 63:6-15.

Kim, C.G., V. Lemaitre, A. Watts, and W.B. Fischer. 2006. Drug-protein interaction with Vpu from HIV-1: proposing binding sites for amiloride and one of its derivatives. *Anal Bioanal Chem*. 386:2213-7.

Kim, C.G., J.A. Watts, and A. Watts. 2005. Ligand docking in the gastric H+/K+-ATPase: homology modeling of reversible inhibitor binding sites. *J Med Chem*. 48:7145-52.

Kitabgi, P. 2006. Functional domains of the subtype 1 neurotensin receptor (NTS1). *Peptides*. 27:2461-8.

Kitabgi, P., R. Carraway, J. Van Rietschoten, C. Granier, J.L. Morgat, A. Menez, S. Leeman, and P. Freychet. 1977. Neurotensin: specific binding to synaptic membranes from rat brain. *Proc Natl Acad Sci U S A*. 74:1846-50.

Kitabgi, P., D. Herve, J.M. Studler, G. Tramu, W. Rostene, and J.P. Tassin. 1989. [Neurotensin/dopamine interactions]. *Encephale*. 15 Spec No:91-4.

Kitabgi, P., W. Rostene, M. Dussaillant, A. Schotte, P.M. Laduron, and J.P. Vincent. 1987. Two populations of neurotensin binding sites in murine brain: discrimination by the antihistamine levocabastine reveals markedly different radioautographic distribution. *Eur J Pharmacol*. 140:285-93.

Knol, J., K. Sjollema, and B. Poolman. 1998. Detergent-mediated reconstitution of membrane proteins. *Biochemistry*. 37:16410-5.

Koenig, B.W., M. Rogowski, and J.M. Louis. 2003. A rapid method to attain isotope labeled small soluble peptides for NMR studies. *J Biomol NMR*. 26:193-202.

Kovacs, F.A., D.A. Fowler, G.J. Gallagher, and L.K. Thompson. 2007. A practical guide for solid-state NMR distance measurements in proteins. *Concepts in Magnetic Resonance Part A*. 30A:21-39.

Kovacs, H., D. Moskau, and M. Spraul. 2005. Cryogenically cooled probes—a leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 46:131-155.

Kroeze, W.K., D.J. Sheffler, and B.L. Roth. 2003. G-protein-coupled receptors at a glance. *J Cell Sci*. 116:4867-9.

Krueger-Koplin, R.D., P.L. Sorgen, S.T. Krueger-Koplin, I.O. Rivera-Torres, S.M. Cahill, D.B. Hicks, L. Grinius, T.A. Krulwich, and M.E. Girvin. 2004. An evaluation of detergents for NMR structural studies of membrane proteins. *J Biomol NMR*. 28:43-57.

Kuntz, I.D., J.M. Blaney, S.J. Oatley, R. Langridge, and T.E. Ferrin. 1982. A geometric approach to macromolecule-ligand interactions. *J Mol Biol*. 161:269-88.

Labbe-Jullie, C., S. Barroso, D. Nicolas-Eteve, J.L. Reversat, J.M. Botto, J. Mazella, J.M. Bernassau, and P. Kitabgi. 1998. Mutagenesis and modeling of the neurotensin receptor NTR1. Identification of residues that are critical for binding SR 48692, a nonpeptide neurotensin antagonist. *J Biol Chem*. 273:16351-7.

Labbe-Jullie, C., J.M. Botto, M.V. Mas, J. Chabry, J. Mazella, J.P. Vincent, D. Gully, J.P. Maffrand, and P. Kitabgi. 1995. [3H]SR 48692, the first nonpeptide neurotensin antagonist radioligand: characterization of binding properties and evidence for distinct agonist and antagonist binding domains on the rat neurotensin receptor. *Mol Pharmacol*. 47:1050-6.

Laporte, S.A., G. Servant, D.E. Richard, E. Escher, G. Guillemette, and R. Leduc. 1996. The tyrosine within the NPXnY motif of the human angiotensin II type 1 receptor is involved in mediating signal transduction but is not essential for internalization. *Mol Pharmacol*. 49:89-95.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23:2947-8.

Laskowski, R.A., M.W. MacArthur, D.S. Moss, and J.M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26:283-291.

LaVallie, E.R., E.A. DiBlasio, S. Kovacic, K.L. Grant, P.F. Schendel, and J.M. McCoy. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Biotechnology (N Y)*. 11:187-93.

Lazarus, L.H., M.H. Perrin, M.R. Brown, and J.E. Rivier. 1977. Mast cell binding of neurotensin. II. Molecular conformation of neurotensin involved in the stereospecific binding to mast cell receptor sites. *J Biol Chem*. 252:7180-3.

Lee, J.H., I. Minn, C.B. Park, and S.C. Kim. 1998. Acidic peptide-mediated expression of the antimicrobial peptide buforin II as tandem repeats in Escherichia coli. *Protein Expr Purif*. 12:53-60.

Levitt, M.H. 2001. Spin dynamics: basics of muclear magnetic resonance. John Wiley & Sons Ltd, Chichester. 686 pp.

Li, H., C.X. Zhou, and J.Z. Su. 2006. Chemical ligation and cleavage on solid support facilitate recombinant peptide purification. *Protein Expr Purif*. 50:238-46.

Lin, B.Z., P.F. Pilch, and K.V. Kandror. 1997. Sortilin is a major protein component of Glut4-containing vesicles. *J Biol Chem*. 272:24145-7.

Lindhout, D.A., A. Thiessen, D. Schieve, and B.D. Sykes. 2003. High-yield expression of isotopically labeled peptides for use in NMR studies. *Protein Sci*. 12:1786-91.

Lodowski, D.T., T.E. Angel, and K. Palczewski. 2009. Comparative Analysis of GPCR Crystal Structures. *Photochemistry and photobiology*. 85:425.

Logan, T.M., E.T. Olejniczak, R.X. Xu, and S.W. Fesik. 1993. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J Biomol NMR*. 3:225-31.

Lorch, M., S. Fahem, C. Kaiser, I. Weber, A.J. Mason, J.U. Bowie, and C. Glaubitz. 2005. How to Prepare Membrane Proteins for Solid-State NMR: A Case Study on the Alpha-Helical Integral Membrane Protein Diaglycerol Kinase from *E. coli*. *ChemBioChem*. 9:1693-1700.

Lowe, I.J. 1959. Free induction decays of rotating solids. *Physical Review Letters*. 2:285-287.

Lowes, V.L., N.Y. Ip, and Y.H. Wong. 2002. Integration of signals from receptor tyrosine kinases and g protein-coupled receptors. *Neurosignals*. 11:5-19.

Luca, S., J.F. White, A.K. Sohal, D.V. Filippov, J.H. van Boom, R. Grisshammer, and M. Baldus. 2003. The conformation of neurotensin bound to its G protein-coupled receptor. *Proc Natl Acad Sci U S A*. 100:10706-11.

Lundstrom, K. 2007. Structural genomics and drug discovery. *J Cell Mol Med*. 11:224-38.

Luthy, R., J.U. Bowie, and D. Eisenberg. 1992. Assessment of protein models with three-dimensional profiles. *Nature*. 356:83-5.

MacDonald, R.C., R.I. MacDonald, B.P. Menco, K. Takeshita, N.K. Subbarao, and L.R. Hu. 1991. Small-volume extrusion apparatus for preparation of large, unilamellar vesicles. *Biochim Biophys Acta*. 1061:297-303.

Majerle, A., J. Kidri, and R. Jerala. 2000. Production of stable isotope enriched antimicrobial peptides in Escherichia coli: an application to the production of a 15N-enriched fragment of lactoferrin. *Journal of Biomolecular NMR*. 18:145-151.

Marassi, F.M., and S.J. Opella. 1998. NMR structural studies of membrane proteins. *Curr Opin Struct Biol*. 8:640-8.

Marassi, F.M., and S.J. Opella. 2000. A solid-state NMR index of helical membrane protein structure and topology. *J Magn Reson*. 144:150-5.

Marinissen, M.J., and J.S. Gutkind. 2001. G-protein-coupled receptors and signaling networks: emerging paradigms. *Trends Pharmacol Sci*. 22:368-76.

Marshall, G.R. 2001. Peptide interactions with G-protein coupled receptors. *Biopolymers*. 60:246-77.

Marti-Renom, M.A., A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 29:291-325.

Martin, R.W., and K.W. Zilm. 2003. Preparation of protein nanocrystals and their characterization by solid state NMR. *Journal of Magnetic Resonance*. 165:162-174.

Martin, S., J.M. Botto, J.P. Vincent, and J. Mazella. 1999. Pivotal role of an aspartate residue in sodium sensitivity and coupling to G proteins of neurotensin receptors. *Mol Pharmacol*. 55:210-5.

Mason, A.J., S.L. Grage, S.K. Straus, C. Glaubitz, and A. Watts. 2004. Identifying anisotropic constraints in multiply labeled bacteriorhodopsin by 15N MAOSS NMR: a general approach to structural studies of membrane proteins. *Biophys J*. 86:1610-7.

Mazella, J., S. Amar, J.C. Bozou, P. Kitabgi, and J.P. Vincent. 1987. Functional properties and molecular structure of central and peripheral neurotensin receptors. *J Recept Res*. 7:157-65.

Mazella, J., J.M. Botto, E. Guillemare, T. Coppola, P. Sarret, and J.P. Vincent. 1996. Structure, functional expression, and cerebral localization of the levocabastine-sensitive neurotensin/neuromedin N receptor from mouse brain. *J Neurosci*. 16:5613-20.

Mazella, J., N. Zsurger, V. Navarro, J. Chabry, M. Kaghad, D. Caput, P. Ferrara, N. Vita, D. Gully, J.P. Maffrand, and J.P. Vincent. 1998. The 100-kDa neurotensin receptor is gp95/sortilin, a non-G-protein-coupled receptor. *J Biol Chem*. 273:26273-6.

McDermott, A., T. Polenova, A. Bockmann, K.W. Zilm, E.K. Paulsen, R.W. Martin, and G.T. Montelione. 2000. Partial NMR assignments for uniformly (C-13, N-15)-enriched BPTI in the solid state. *Journal of Biomolecular NMR*. 16:209-219.

Merrifield, R.B. 1969. Solid-phase peptide synthesis. *Adv Enzymol Relat Areas Mol Biol*. 32:221-96.

Mesleh, M.F., W.A. Shirley, C.E. Heise, N. Ling, R.A. Maki, and R.P. Laura. 2007. NMR structural characterization of a minimal peptide antagonist bound to the extracellular domain of the corticotropin-releasing factor1 receptor. *J Biol Chem*. 282:6338-46.

Metz, G., X. Wu, and S.O. Smith. 1994. Ramped-amplitude cross polarization in magic-angle-spinning NMR. *J Magn Reson A*. 110:219–227.

Miller, K.J., B.J. Murphy, and M.A. Pelleymounter. 2004. Central G-Protein Coupled Receptors (GPCR)s as molecular targets for the treatment of obesity: assets, liabilities and development status. *Curr Drug Targets CNS Neurol Disord*. 3:357-77.

Milligan, G. 2004. G protein-coupled receptor dimerization: function and ligand pharmacology. *Molecular pharmacology*. 66:1.

Morcombe, C.R., and K.W. Zilm. 2003. Chemical shift referencing in MAS solid state NMR. *J Magn Reson*. 162:479-86.

Morris, G.M., D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, and A.J. Olson. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*. 19:1639-1662.

Morris, G.M., R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, and A.J. Olson. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*.

Mukhopadhyay, A. 1996. Inclusion bodies and purification of proteins in biologically active forms. *Advances in Biochemical Engineering Biotechnology*. 56:61-110.

Murakami, M., and T. Kouyama. 2008. Crystal structure of squid rhodopsin. *Nature*. 453:363-7.

Nelson, D.L., and M.M. Cox. 2004. Lehninger Principles of Biochemistry. W. H. Freeman, New York. 1100 pp.

Nicholson, L.K., and T.A. Cross. 1989. Gramicidin cation channel: an experimental determination of the right-handed helix sense and verification of beta-type hydrogen bonding. *Biochemistry*. 28:9379-85.

Nieto, J.L., M. Rico, J. Santoro, J. Herranz, and F.J. Bermejo. 1986. Assignment and conformation of neurotensin in aqueous solution by 1H NMR. *Int J Pept Protein Res*. 28:315-23.

O'Donovan, C., M.J. Martin, A. Gattiker, E. Gasteiger, A. Bairoch, and R. Apweiler. 2002. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform*. 3:275-84.

Oakhill, J.S., B.J. Sutton, A.R. Gorringe, and R.W. Evans. 2005. Homology modelling of transferrin-binding protein A from Neisseria meningitidis. *Protein Eng Des Sel*. 18:221-8.

Okada, T., M. Sugihara, A.N. Bondar, M. Elstner, P. Entel, and V. Buss. 2004. The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. *J Mol Biol*. 342:571-83.

Osterberg, F., G.M. Morris, M.F. Sanner, A.J. Olson, and D.S. Goodsell. 2002. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins*. 46:34-40.

Ottleben, H., M. Haasemann, R. Ramachandran, M. Gorlach, W. Muller-Esterl, and L.R. Brown. 1997. An NMR study of the interaction of 15N-labelled bradykinin with an antibody mimic of the bradykinin B2 receptor. *Eur J Biochem*. 244:471-8.

Oxenoid, K., and J.J. Chou. 2005. The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc Natl Acad Sci U S A*. 102:10870-5.

Paiva, T.B., A.C. Paiva, and H.A. Scheraga. 1963. The Conformation of Angiotensin Ii in Aqueous Solution. *Biochemistry*. 2:1327-34.

Palczewski, K., T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, and M. Miyano. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*. 289:739-45.

Pang, Y.P., B. Cusack, K. Groshan, and E. Richelson. 1996. Proposed ligand binding site of the transmembrane receptor for neurotensin(8-13). *J Biol Chem*. 271:15060-8.

Park, J.H., P. Scheerer, K.P. Hofmann, H.W. Choe, and O.P. Ernst. 2008. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*. 454:183-7.

Park, S.H., A.A. Mrse, A.A. Nevzorov, M.F. Mesleh, M. Oblatt-Montal, M. Montal, and S.J. Opella. 2003. Three-dimensional structure of the channel-forming trans-membrane domain of virus protein "u" (Vpu) from HIV-1. *J Mol Biol*. 333:409-24.

Parker, S.L., M.S. Parker, R. Sah, and F. Sallee. 2005. Angiogenesis and rhodopsin-like receptors: a role for N-terminal acidic residues? *Biochem Biophys Res Commun*. 335:983-92.

Patra, S.M., T. Bastug, and S. Kuyucak. 2007. Binding of organic cations to gramicidin A channel studied with AutoDock and molecular dynamics simulations. *J Phys Chem B*. 111:11303-11.

Pauli, J., B. van Rossum, H. Forster, H.J.M. de Groot, and H. Oschkinat. 2000. Sample optimization and identification of signal patterns of amino acid side chains in 2D RFDR spectra of the alpha-spectrin SH3 domain. *Journal of Magnetic Resonance*. 143:411-416.

Pearson, W.R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol*. 183:63-98.

Peersen, O.B., E.A. Pratt, H.T. Truong, C. Ho, and G.S. Rule. 1990. Site-specific incorporation of 5-fluorotryptophan as a probe of the structure and function of the membrane-bound D-lactate dehydrogenase of Escherichia coli: a 19F nuclear magnetic resonance study. *Biochemistry*. 29:3256-62.

Peersen, O.B., and S.O. Smith. 1993. Rotational resonance NMR of biological membranes. *Concepts in Magnetic Resonance*. 5:303-303.

Pellegrini, M., A. Bisello, M. Rosenblatt, M. Chorev, and D.F. Mierke. 1998. Binding domain of human parathyroid hormone receptor: from conformation to function. *Biochemistry*. 37:12737-43.

Pierce, K.L., R.T. Premont, and R.J. Lefkowitz. 2002. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol*. 3:639-50.

Pikal-Cleland, K.A., N. Rodriguez-Hornedo, G.L. Amidon, and J.F. Carpenter. 2000. Protein denaturation during freezing and thawing in phosphate buffer systems: monomeric and tetrameric beta-galactosidase. *Arch Biochem Biophys*. 384:398-406.

Pines, A., M.G. Gibby, and J.S. Waugh. 1972. Proton-enhanced nuclear induction spectroscopy. Method for high-resolution NMR of dilute spins in solids. *Journal of Chemistry Physics*. 52:1776-1777.

Polenova, T. 2004. Homonuclear 13C Solid-State MAS Spectroscopy of the Uniformly (13C, 15N)-Enriched E. coli Thioredoxin at 17.6 Tesla. Spin System Identification and Sidechain Resonance Assignments. *45th ENC, talk*.

Pontius, J., J. Richelle, and S.J. Wodak. 1996. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol*. 264:121-36.

Quistgaard, E.M., P. Madsen, M.K. Groftehauge, P. Nissen, C.M. Petersen, and S.S. Thirup. 2009. Ligands bind to Sortilin in the tunnel of a ten-bladed beta-propeller domain. *Nat Struct Mol Biol*. 16:96-8.

Raleigh, D.P., F. Creuzet, S.K. Das Gupta, M.H. Levitt, and R.G. Griffin. 1989. Measurement of Internuclear Distances in Polycrystalline Solids: Rotationally Enhanced Transfer of Nuclear Spin Magnetization. *J.Am.Chem.Soc.* 111:4502.

Ramamoorthy, A., C.H. Wu, and S.J. Opella. 1999. Experimental aspects of multidimensional solid-state NMR correlation spectroscopy. *J Magn Reson*. 140:131-40.

Rayasam, G.V., V.K. Tulasi, J.A. Davis, and V.S. Bansal. 2007. Fatty acid receptors as new therapeutic targets for diabetes. *Expert Opin Ther Targets*. 11:661-71.

Rhee, M.H., I. Nevo, M.L. Bayewitch, O. Zagoory, and Z. Vogel. 2000. Functional role of tryptophan residues in the fourth transmembrane domain of the CB(2) cannabinoid receptor. *J Neurochem*. 75:2485-91.

Rigaud, J.L., G. Mosser, J.J. Lacapere, A. Olofsson, D. Levy, and J.L. Ranck. 1997. Bio-Beads: an efficient strategy for two-dimensional crystallization of membrane proteins. *J Struct Biol*. 118:226-35.

Rigaud, J.L., B. Pitard, and D. Levy. 1995. Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *Biochim Biophys Acta*. 1231:223-46.

Rosendorff, A., B.J. Ebersole, and S.C. Sealfon. 2000. Conserved helix 7 tyrosine functions as an activation relay in the serotonin 5HT(2C) receptor. *Brain Res Mol Brain Res*. 84:90-6.

Ross, E.M., and T.M. Wilkie. 2000. GTPase-activating proteins for heterotrimeric G proteins: regulators of G protein signaling (RGS) and RGS-like proteins. *Annu Rev Biochem*. 69:795-827.

Roth, B.L., M. Shoham, M.S. Choudhary, and N. Khan. 1997. Identification of conserved aromatic residues essential for agonist binding and second messenger production at 5-hydroxytryptamine2A receptors. *Mol Pharmacol*. 52:259-66.

Rovati, G.E., V. Capra, and R.R. Neubig. 2007. The highly conserved DRY motif of class A G protein-coupled receptors: beyond the ground state. *Mol Pharmacol*. 71:959-64.

Saitô, H., R. Tabeta, Y. Yoshioka, C. Hara, T. Kiho, and S. Ukai. 1987. A High-Resolution Solid-State 13C NMR Study of the Secondary Structure of Branched (1→3)-β-D-Glucans from Fungi: Evidence of Two Kinds of Conformers, Curdlan-Type Single-Helix and Laminaran-Type Triple-Helix Forms, as Manifested from the Conformation-Dependent 13C Chemical Shifts. *Bull Chem Soc Jpn*. 60:4267-4272.

Sali, A., and T.L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 234:779-815.

Sander, C., and R. Schneider. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*. 9:56-68.

Sanders, C.R., and F. Sonnichsen. 2006. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem*. 44 Spec No:S24-40.

Sarramegna, V., I. Muller, A. Milon, and F. Talmont. 2006. Recombinant G protein-coupled receptors from expression to renaturation: a challenge towards structure. *Cellular and Molecular Life Sciences (CMLS)*. 63:1149-1164.

Sarramegna, V., F. Talmont, P. Demange, and A. Milon. 2003. Heterologous expression of G-protein-coupled receptors: comparison of expression systems fron the standpoint of large-scale production and purification. *Cell Mol Life Sci*. 60:1529-46.

Sasaki, K., S. Dockerill, D.A. Adamiak, I.J. Tickle, and T. Blundell. 1975. X-ray analysis of glucagon and its relationship to receptor binding. *Nature*. 257:751-7.

Sautel, M., K. Rudolf, H. Wittneben, H. Herzog, R. Martinez, M. Munoz, W. Eberlein, W. Engel, P. Walker, and A.G. Beck-Sickinger. 1996. Neuropeptide Y and the nonpeptide antagonist BIBP 3226 share an overlapping binding site at the human Y1 receptor. *Mol Pharmacol*. 50:285-92.

Scheerer, P., J.H. Park, P.W. Hildebrand, Y.J. Kim, N. Krauss, H.W. Choe, K.P. Hofmann, and O.P. Ernst. 2008. Crystal structure of opsin in its G-protein-interacting conformation. *Nature*. 455:497-502.

Schiksnis, R.A., M.J. Bogusky, P. Tsang, and S.J. Opella. 1987. Structure and dynamics of the Pf1 filamentous bacteriophage coat protein in micelles. *Biochemistry*. 26:1373-81.

Schmidt, B., S. Lindman, W. Tong, G. Lindeberg, A. Gogoll, Z. Lai, M. Thornwall, B. Synnergren, A. Nilsson, C.J. Welch, M. Sohtell, C. Westerlund, F. Nyberg, A. Karlen, and A. Hallberg. 1997. Design, synthesis, and biological activities of four angiotensin II receptor ligands with gamma-turn mimetics replacing amino acid residues 3-5. *J Med Chem*. 40:903-19.

Schneider, A.B., and H. Edelhoch. 1972. Polypeptide hormone interaction. 3. Conformational changes of glucagon bound to lysolecithin. *J Biol Chem*. 247:4992-5.

Schwartz, T.W., and W.L. Hubbell. 2008. Structural biology: A moving story of receptors. *Nature*. 455:473-4.

Schwede, T., A. Diemand, N. Guex, and M.C. Peitsch. 2000. Protein structure computing in the genomic era. *Res Microbiol*. 151:107-12.

Seddon, A.M., P. Curnow, and P.J. Booth. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta*. 1666:105-17.

Sefler, A.M., J.X. He, T.K. Sawyer, K.E. Holub, D.O. Omecinsky, M.D. Reily, V. Thanabal, H.C. Akunne, and W.L. Cody. 1995. Design and structure-activity relationships of C-terminal cyclic neurotensin fragment analogues. *J Med Chem*. 38:249-57.

Shen, S.H. 1984. Multiple joined genes prevent product degradation in Escherichia coli. *Proc Natl Acad Sci U S A*. 81:4627-31.

Shibata, Y., J.F. White, M.J. Serrano-Vega, F. Magnani, A.L. Aloia, R. Grisshammer, and C.G. Tate. 2009. Thermostabilization of the Neurotensin Receptor NTS1. *J Mol Biol*.

Shimamura, T., K. Hiraki, N. Takahashi, T. Hori, H. Ago, K. Masuda, K. Takio, M. Ishiguro, and M. Miyano. 2008. Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. *J Biol Chem*. 283:17753-6.

Shuker, S.B., P.J. Hajduk, R.P. Meadows, and S.W. Fesik. 1996. Discovering high-affinity ligands for proteins: SAR by NMR. *Science*. 274:1531-4.

Siemion, I.Z., T. Wieland, and K.H. Pook. 1975. Influence of the distance of the proline carbonyl from the beta and gamma carbon on the 13C chemical shifts. *Angew Chem Int Ed Engl*. 14:702-3.

Singer, S.J., and G.L. Nicolson. 1972. The fluid mosaic model of the structure of cell membranes. *Science*. 175:720-31.

Sippl, M.J. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins*. 17:355-62.

Smith, D.B., and K.S. Johnson. 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*. 67:31-40.

Solis, F.J., and R.J.-B. Wets. 1981. Minimization by Random Search Techniques. *Mathmatics of Operations Research*. 6:19-30.

Song, C., S. Tapaneeyakorn, A.C. Murphy, C. Butts, A. Watts, and C.L. Willis. 2009. Enantioselective syntheses of alpha-Fmoc-Pbf-[2-(13)C]-L-arginine and Fmoc-[1,3-(13)C2]-L-proline and Incorporation into the neurotensin receptor 1 ligand, NT(8-13). *J Org Chem*. 74:8980-7.

Sprang, S.R. 1997. G protein mechanisms: insights from structural analysis. *Annu Rev Biochem*. 66:639-78.

Spyroulias, G.A., P. Nikolakopoulou, A. Tzakos, I.P. Gerothanassis, V. Magafa, E. Manessi-Zoupa, and P. Cordopatis. 2003. Comparison of the solution structures of angiotensin I & II. Implication for structure-function relationship. *Eur J Biochem*. 270:2163-73.

Stejskal, E.O., J. Schaefer, and J.S. Waugh. 1977. Magic-angle spinning and polarization transfer in proton-enhanced NMR. *J. magn. Reson*. 28:105–112.

Strosberg, A.D., and C. Nahmias. 2007. G-protein-coupled receptor signalling through protein networks. *Biochem Soc Trans*. 35:23-7.

Suzuki, Y., M. Yabuta, and K. Ohsuye. 1998. High-level production of recombinant human parathyroid hormone 1-34. *Applied and Environmental Microbiology*. 64:526-529.

Takegoshi, K., S. Nakamura, and T. Terao. 2001. 13C-1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chemical Physics Letters*. 344:631-637.

Takegoshi, K., and T. Terao. 1999. 13C-1H dipolar recoupling under very fast magic-angle spinning using virtual pulses. *Solid State Nucl Magn Reson*. 13:203-12.

Tanaka, K., M. Masu, and S. Nakanishi. 1990. Structure and functional expression of the cloned rat neurotensin receptor. *Neuron*. 4:847-54.

Teague, S.J. 2003. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov*. 2:527-41.

Telenti, A. 2009. Safety concerns about CCR5 as an antiviral target. *Curr Opin HIV AIDS*. 4:131-5.

Teng, Q. 2005. Structural Biology, Practical NMR Applications. Springer, Berlin. 295 pp.

Thiriot, D.S., A.A. Nevzorov, L. Zagyanskiy, C.H. Wu, and S.J. Opella. 2004. Structure of the coat protein in Pf1 bacteriophage determined by solid-state NMR spectroscopy. *J Mol Biol*. 341:869-79.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 25:4876-82.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673-80.

Tieleman, D.P., H.J. Berendsen, and M.S. Sansom. 1999. An alamethicin channel in a lipid bilayer: molecular dynamics simulations. *Biophys J*. 76:1757-69.

Traxler, P., P. Furet, H. Mett, E. Buchdunger, T. Meyer, and N. Lydon. 1997. Design and synthesis of novel tyrosine kinase inhibitors using a pharmacophore model of the ATP-binding site of the EGF-R. *J Pharm Belg*. 52:88-96.

Tucker, J., and R. Grisshammer. 1996. Purification of a rat neurotensin receptor expressed in Escherichia coli. *Biochem J*. 317 (Pt 3):891-9.

Tyndall, J.D., B. Pfeiffer, G. Abbenante, and D.P. Fairlie. 2005. Over one hundred peptide-activated G protein-coupled receptors recognize ligands with turn structure. *Chem Rev*. 105:793-826.

Tyndall, J.D., and R. Sandilya. 2005. GPCR agonists and antagonists in the clinic. *Med Chem*. 1:405-21.

Uegaki, K., N. Nemoto, M. Shimizu, T. Wada, Y. Kyogoku, and Y. Kobayashi. 1996. 15N labeling method of peptides using a thioredoxin gene fusion expression system: an application to ACTH-(1-24). *FEBS Lett*. 379:47-50.

Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J. Berendsen. 2005. GROMACS: fast, flexible, and free. *J Comput Chem*. 26:1701-18.

Van Gaal, L.F., A.M. Rissanen, A.J. Scheen, O. Ziegler, and S. Rossner. 2005. Effects of the cannabinoid-1 receptor blocker rimonabant on weight reduction and cardiovascular risk factors in overweight patients: 1-year experience from the RIO-Europe study. *Lancet*. 365:1389-97.

van Gunsteren, W.F., P. Kruger, S.R. Billeter, A.E. Mark, A.A. Eising, W.R.P. Scott, P.H. Huneberger, and I.G. Tironi. 1996. Biomolecular simulation: The GROMOS96 manual and user guide Biomos and Hochschulverlag AG an der ETH Zurich, Zurich.

Vankayalapati, H., D.J. Bearss, J.W. Saldanha, R.M. Munoz, S. Rojanala, D.D. Von Hoff, and D. Mahadevan. 2003. Targeting aurora2 kinase in oncogenesis: a structural bioinformatics approach to target validation and rational drug design. *Mol Cancer Ther*. 2:283-94.

Varga, K., L. Aslimovska, I. Parrot, M.T. Dauvergne, M. Haertlein, V.T. Forsyth, and A. Watts. 2007. NMR crystallography: the effect of deuteration on high resolution 13C solid state NMR spectra of a 7-TM protein. *Biochim Biophys Acta*. 1768:3029-35.

Varga, K., L. Aslimovska, and A. Watts. 2008. Advances towards resonance assignments for uniformly-^{13}C, ^{15}N enriched bacteriorhodopsin at 18.8 T in purple membranes. *J Biomol NMR*. 41:1-4.

Varga, K., and A. Watts. 2007. Introduction to solid state NMR and its application to membrane protein-ligand binding studies. In: Biophysical Analysis of Membrane Proteins (ed. E. Pebay-Peroula). Wiley-VCH, Weinheim. 55-90 pp.

Vassilatis, D.K., J.G. Hohmann, H. Zeng, F. Li, J.E. Ranchalis, M.T. Mortrud, A. Brown, S.S. Rodriguez, J.R. Weller, A.C. Wright, J.E. Bergmann, and G.A. Gaitanaris. 2003. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci U S A*. 100:4903-8.

Vernal, J., A. Fiser, A. Sali, M. Muller, J.J. Cazzulo, and C. Nowicki. 2002. Probing the specificity of a trypanosomal aromatic alpha-hydroxy acid dehydrogenase by site-directed mutagenesis. *Biochem Biophys Res Commun*. 293:633-9.

Villar, H.O., J. Yan, and M.R. Hansen. 2004. Using NMR for ligand discovery and optimization. *Curr Opin Chem Biol*. 8:387-91.

Vita, N., P. Laurent, S. Lefort, P. Chalon, X. Dumont, M. Kaghad, D. Gully, G. Le Fur, P. Ferrara, and D. Caput. 1993. Cloning and expression of a complementary DNA encoding a high affinity human neurotensin receptor. *FEBS Lett*. 317:139-42.

Vita, N., F. Oury-Donat, P. Chalon, M. Guillemot, M. Kaghad, A. Bachy, O. Thurneyssen, S. Garcia, C. Poinot-Chazel, P. Casellas, P. Keane, G. Le Fur, J.P. Maffrand, P. Soubrie, D. Caput, and P. Ferrara. 1998. Neurotensin is an antagonist of the human neurotensin NT2 receptor expressed in Chinese hamster ovary cells. *Eur J Pharmacol*. 360:265-72.

von Hippel, P.H., and T. Schleich. 1969. The effects of neutral salts on the structure and conformational stability of macromolecules in solution. In: Structure and Stability of Biological Macromolecules. Marcel Dekker, New York. 694 pp.

Vosegaard, T., and N.C. Nielsen. 2002. Towards high-resolution solid-state NMR on large uniformly 15N- and [13C,15N]-labeled membrane proteins in oriented lipid bilayers. *J Biomol NMR*. 22:225-47.

Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J Mol Graph*. 8:52-6, 29.

Wallin, E., and G. von Heijne. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci*. 7:1029-38.

Wang, Y., and O. Jardetzky. 2002. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci*. 11:852-61.

Warne, T., J. Chirnside, and G.F.X. Schertler. 2003. Expression and purification of truncated, non-glycosylated turkey beta-adrenergic receptors for crystallization. *BBA-Biomembranes*. 1610:133-140.

Warne, T., M.J. Serrano-Vega, J.G. Baker, R. Moukhametzianov, P.C. Edwards, R. Henderson, A.G. Leslie, C.G. Tate, and G.F. Schertler. 2008. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature*. 454:486-91.

Watts, A. 2002. Direct studies of ligand-receptor interactions and ion channel blocking (Review). *Mol Membr Biol*. 19:267-75.

Watts, A., S.K. Straus, S. Grage, M. Kamihira, Y.H. Lam, and Z. Xhao. 2004. Methods in Molecular Biology - Protein NMR Techniques. *Humana Press, New Jersey* 278 (ed. K. Downing):403-474.

Watts, J.A., A. Watts, and D.A. Middleton. 2001. A model of reversible inhibitors in the gastric H+/K+-ATPase binding site determined by rotational echo double resonance NMR. *J Biol Chem*. 276:43197-204.

Waugh, D.S. 2005. Making the most of affinity tags. *TRENDS in Biotechnology*. 23:316-320.

Wedemeyer, W.J., E. Welker, and H.A. Scheraga. 2002. Proline cis-trans isomerization and protein folding. *Biochemistry*. 41:14637-44.

Wess, J., T.I. Bonner, F. Dorje, and M.R. Brann. 1990. Delineation of muscarinic receptor domains conferring selectivity of coupling to guanine nucleotide-binding proteins and second messengers. *Mol Pharmacol*. 38:517-23.

Wess, J., S. Nanavati, Z. Vogel, and R. Maggio. 1993. Functional role of proline and tryptophan residues highly conserved among G protein-coupled receptors studied by mutational analysis of the m3 muscarinic receptor. *Embo J*. 12:331-8.

White, J.F., L.B. Trinh, J. Shiloach, and R. Grisshammer. 2004. Automated large-scale purification of a G protein-coupled receptor for neurotensin. *FEBS Lett*. 564:289-93.

White, S.H. 2009. Biophysical dissection of membrane proteins. *Nature*. 459:344-6.

Whorton, M.R., M.P. Bokoch, S.G.F. Rasmussen, B. Huang, R.N. Zare, B. Kobilka, and R.K. Sunahara. 2007. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proceedings of the National Academy of Sciences*. 104:7682.

Williams, R.M. 1992. Asymmetric Synthesis of a-Amino Acids. *Aldrichimica Acta*. 25:11-25.

Williamson, P.T., S. Bains, C. Chung, R. Cooke, and A. Watts. 2002. Probing the environment of neurotensin whilst bound to the neurotensin receptor by solid state NMR. *FEBS Lett*. 518:111-5.

Williamson, P.T., G. Grobner, P.J. Spooner, K.W. Miller, and A. Watts. 1998. Probing the agonist binding pocket in the nicotinic acetylcholine receptor: a high-resolution solid-state NMR approach. *Biochemistry*. 37:10854-9.

Williamson, P.T., J.F. Roth, T. Haddingham, and A. Watts. 2000. Expression and purification of recombinant neurotensin in Escherichia coli. *Protein Expr Purif*. 19:271-5.

Williamson, P.T., A. Verhoeven, K.W. Miller, B.H. Meier, and A. Watts. 2007. The conformation of acetylcholine at its target site in the membrane-embedded nicotinic acetylcholine receptor. *Proc Natl Acad Sci U S A*. 104:18031-6.

Williamson, P.T.F., S.K. Bains, C. Chung, R. Cooke, B. Meier, and A. Watts. 2001. Characterization and assignment of uniformly labeled NT(8-13) at the agonist binding site of the G-protein coupled neurotensin receptor. *In: Focus on Structural Biology Vol. 1 Perspectives on Solid State NMR in Biology (Eds. S.R. Kühne & H.J.M. de Groot)*:191-202.

Wu, G., A. Fiser, B. ter Kuile, A. Sali, and M. Muller. 1999. Convergent evolution of Trichomonas vaginalis lactate dehydrogenase from malate dehydrogenase. *Proc Natl Acad Sci U S A*. 96:6285-90.

Xu, G.Y., and C.M. Deber. 1991. Conformations of neurotensin in solution and in membrane environments studied by 2-D NMR spectroscopy. *Int J Pept Protein Res*. 37:528-35.

Yamada, M., M. Yamada, A. Lombet, P. Forgez, and W. Rostene. 1998. Distinct functional characteristics of levocabastine sensitive rat neurotensin NT2 receptor expressed in Chinese hamster ovary cells. *Life Sci*. 62:PL 375-80.

Yeagle, P.L., G. Choi, and A.D. Albert. 2001. Studies on the structure of the G-protein-coupled receptor rhodopsin including the putative G-protein binding site in unactivated and activated forms. *Biochemistry*. 40:11932-7.

Yohannan, S., S. Faham, D. Yang, J.P. Whitelegge, and J.U. Bowie. 2004. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc Natl Acad Sci U S A*. 101:959-63.

Yoshioka, S., Y. Aso, S. Kojima, S. Sakurai, T. Fujiwara, and H. Akutsu. 1999. Molecular Mobility of Protein in Lyophilized Formulations Linked to the Molecular Mobility of Polymer Excipients, as Determined by High Resolution 13C Solid-State NMR. *Pharm Res*. 16:1621-1625.

Yuzlenko, O., and K. Kiec-Kononowicz. 2009. Molecular modeling of A1 and A2A adenosine receptors: comparison of rhodopsin- and beta2-adrenergic-based homology models through the docking studies. *J Comput Chem*. 30:14-32.

Zachariae, U., R. Schneider, P. Velisetty, A. Lange, D. Seeliger, S.J. Wacker, Y. Karimi-Nejad, G. Vriend, S. Becker, O. Pongs, M. Baldus, and B.L. de Groot. 2008. The molecular mechanism of toxin-induced conformational changes in a potassium channel: relation to C-type inactivation. *Structure*. 16:747-54.

Zhou, N., G.J. Moore, and H.J. Vogel. 1991. Proton NMR studies of angiotensin II and its analogs in aqueous solution. *J Protein Chem*. 10:333-43.

Zhou, Y., T. Cierpicki, R.H. Jimenez, S.M. Lukasik, J.F. Ellena, D.S. Cafiso, H. Kadokura, J. Beckwith, and J.H. Bushweller. 2008. NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. *Mol Cell*. 31:896-908.

Zsurger, N., J. Mazella, and J.P. Vincent. 1994. Solubilization and purification of a high affinity neurotensin receptor from newborn human brain. *Brain Res*. 639:245-52.

Appendices

Appendix A - General methods

A.1 *In vitro* site-directed mutagenesis

Site-directed mutagenesis was carried out according to the Quick Change™ Site-directed Mutagenesis Kit (Stratagene). The method would make use of *Pfu* DNA polymerase which replicates both strands of DNA plasmid with high fidelity without displacing the mutant oligonucleotide primers (**Figure A.1**).

An automated GeneAmp PCR System 2400 thermal cycler (Perkin-Elmer, CA, USA) was used to perform polymerase chain reactions (PCR). The reaction mixture (50 µl) was composed of

DNA template	50-100 ng
10 mM dNTP mix (2.5 mM each)	1 µL
Forward primer	10 pmole
Reverse primer	10 pmole
10x *Pfu* buffer	5 µL
Pfu DNA polymerase	3 U
Sterile distilled water making final volume to	50 µL

Then the PCR was performed by the following amplification condition, as shown in **Tables A.1**.

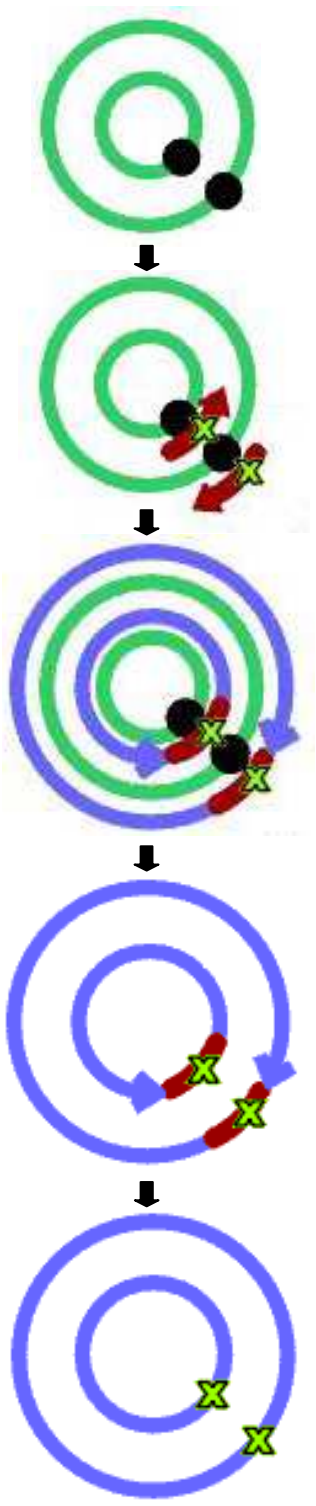


Figure A.1: Overview of the QuickChange site-directed mutagenesis method based on the Quick Change™ Site-directed Mutagenesis Kit (Stratagene).

Table A.1: Temperature cycling parameters for site-directed mutagenesis

Temperature (°C)	Time (min:sec)	Cycles
95	5:00	1
95	0:30	20
60	1:00	
68	15:00	
68	7:00	1
25	HOLD	

*Dpn*I restriction endonuclease (1 µL) was added to the PCR product and incubated at 37 °C for 1 h in order to digest the parental DNA template and leave only mutated nicked plasmid. *Dpn*I restriction endonuclease is specific for methylated and hemimethylated DNA (5′-G^{me}A↓TC-3′). DNA isolated from almost all *E. coli* strains is *dam* methylated and it would be therefore susceptible to *Dpn*I digestion. The *Dpn*I-digested PCR products were subsequently transformed into *E. coli* competent cells.

A.2 DNA transformation by heat shock

An aliquot of 50 µL competent *E. coli* cells was mixed with a 1.5 µL aliquot of 1 µg/µL DNA vector and incubated on ice for 30 min. The cells were made to undergo heat shock at 42 °C for 45 s, and then immediately chilled on ice for additional 5 min. 250 µL of SOC media were added to transformed cells, mixed and incubated at 37 °C for 1 h with shaking at 200 rpm. Culture was plated onto ampicillin-selective LB agar and incubated overnight at 37 °C.

A.3 SDS-PAGE and Western blot analysis.

For sample preparation, protein samples were mixed with 2x loading buffer (200 mM Tris-HCl (pH 8.8), 4% (*w/v*) SDS, 20% (*v/v*) glycerol, 0.02% (*w/v*) bromophenol blue, 5 mM EDTA and 50 mM DTT) in the ratio of 1:1 and heated at 95 °C for 5 min. The heated samples were vigorously vortex-mixed and centrifuged at 12,000 xg for 10 min. The supernatant was loaded onto the well of SDS-12% polyacrylamide gel with Tris-glycine buffer (25 mM Tris-HCl (pH 8.3), 192 mM glycine, and 0.1% SDS) or NuPAGE 12% Bis-Tris gel with MES running buffer (50 mM MES, 50 mM Tris-HCl (pH 7.3), 0.1% (*w/v*) SDS, 1 mM EDTA). Electrophoresis was performed with constant voltage at 120 V for 90 min for the SDS-PAGE gel and at 200 V for 40 min for the NuPAGE gel.

After electrophoresis, protein bands were visualised by soaking the gel in staining solution (50% (*v/v*) methanol, 10% (*v/v*) glacial acetic acid and 0.1% (*v/v*) Coomassie Brilliant Blue R-250 in water) for 1 h. The gel was then soaked in destaining solution (45% (*v/v*) methanol and 10% (*v/v*) glacial acetic acid in water) overnight or until the background was clear.

For Western blots, the gel, blotting pads and nitrocellulose membrane were soaked in Western transfer buffer (12 mM Tris-HCl (pH 8.3) and 96 mM glycine) for 10 min. Transfer onto the nitrocellulose membrane was carried out at a constant current of 26 mA for 1 h. The nitrocellulose membrane was then washed in PBS-tween buffer (0.1% (*v/v*) Tween) containing low fat milk powder for 1 h. The membrane was then transferred into PBS-Tween milk powder buffer containing the primary antibody and incubated for 1 h. Three 5 min washes of the membrane were

carried out using PBS-Tween buffer (0.1% (*v/v*) Tween). The blot was developed using a chemiluminscence substrate and photographic film.

A.4 Silver stain

After gel electrophoresis, proteins were fixed by incubating the gels in fixative enhancer solution (Silver Stain Plus, Biorad) for 20 min with gentle agitation. The solution was then decanted and the gel was rinsed with MilliQ water with mild agitation for 20 min. After that, the gel was stained in staining solution for approximately 15 min or until desired staining intensity. After the desired staining is reached, the gel was placed in 5% (*v/v*) acetic acid to stop the staining reaction for a minimum of 15 min.

A.5 Preparation of Ni-IMAC column

The column was saturated with Ni^{2+} ions, by using a 50 mM $NiSO_4$ solution. According to the information of the manufacturer the capacity of the column is ~15 μmol Ni^{2+} ions per mL of medium. After saturation of the column with Ni^{2+} ions, the column was washed with MilliQ water and equilibrated with NiA buffer.

The column was regenerated by washing with 500 mM EDTA, which removed all Ni^{2+} ions. After washing with MilliQ water the column was reloaded with Ni^{2+} ions and re-used. This re-use had no effect on the reproducibility of the experiments.

A.6 Ninhydrin test

A small number of resin beads were removed from the reaction mixture and gently dried on a hot plate, before being sprayed liberally with ninhydrin solution (0.5

% (*v/v*) ninhydrin in butan-1-ol) and heated again on the hot plate. The solution and beads turn blue-purple in the presence of primary amide group as described in **Figure A.2**.

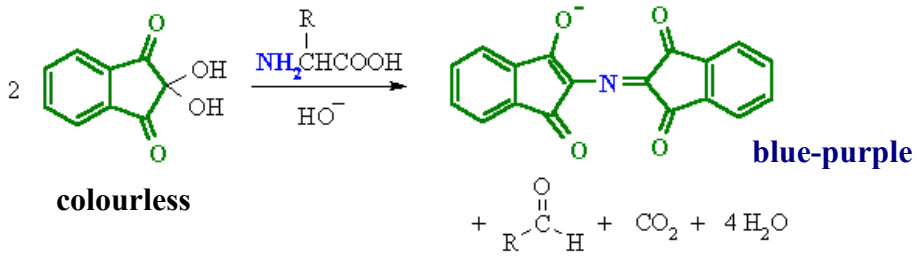


Figure A.2: Ninhydrin reaction (http://www.chem.ucalgary.ca/courses/351/Carey5th/Ch27/ch27-3-3.html).

Appendix B - Buffer compositions

B.1 Growth media

Luria-Bertani (LB)

Per Litre: 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl, adjusted to pH 7.0

Luria-Bertani (LB) agar

Per Litre: 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl, 15 g agar, adjusted to pH 7.0

M9

5x M9 salts: 30 g Na_2HPO_4, 15 g KH_2PO_4, 2.5 g NaCl, 7.5 g NH_4Cl

Per Litre: 1x M9 salts, 1mL of 0.1 M $CaCl_2$, 2 mL of 1 M $MgSO_4$, 20 mL of filtered 20% (*w/v*) glucose, 20 mL of filtered 8.5% (*w/v*) YNB

2xTY

Per Litre: 16 g bacto-tryptone, 10 g yeast extract, 5 g NaCl, adjusted to pH 7.0

SOC

Per Litre: 20 g bacto-tryptone, 5 g yeast extract, 0.5 g NaCl, 2.5 mM KCl, 10 mM MgCl2, 20 mM glucose, adjusted to pH 7.0

B.2 SDS-PAGE analysis

2x loading buffer

200 mM Tris-HCl (pH 8.8), 4% (*w/v*) SDS, 20% (*v/v*) glycerol, 0.02% (*w/v*) bromophenol blue, 5 mM EDTA and 50 mM DTT

Tris-glycine SDS running buffer

25 mM Tris-HCl (pH 8.3), 192 mM Glycine, 0.1 % (*w/v*) SDS

Tris-glycine transfer buffer

12 mM Tris-HCl (pH 8.3), 96 mM Glycine

MES SDS running buffer

50 mM MES, 50 mM Tris-HCl (pH 7.3), 0.1% (*w/v*) SDS, 1 mM EDTA

B.3 Neurotensin purification buffers

Solubilisation buffer

50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 5 mM imidazole, 1 mM PMSF

Equilibration buffer

5 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl (pH 8.0)

Washing buffer

40 mM imidazole, 500 mM NaCl, 10 mM Tris-HCl (pH 8.0)

Elution buffer

500 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl (pH 8.0)

Desalting buffer

10 mM NH_4HCO_3 at pH 8.0

B.4 NTS1 purification buffers

N.B. 'protease inhibitors' consist of Leupeptin, Pepstatin A, Aprotinin, all at a final concentration of 1 μg/mL.

2x solubilisation buffer

100 mM Tris-HCl (pH 7.4), 400 mM NaCl, 60% (*v/v*) glycerol, protease inhibitors

NiA

50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 30% (*v/v*) glycerol, 0.5% (*w/v*) CHAPS, 0.1% (*w/v*) DDM, 0.1% (*w/v*) CHS, 50 mM imidazole, protease inhibitors

NiB

NiA plus 350 mM imidazole

Cleavage Buffer

50 mM Tris-HCl (pH 7.4), 15% (*v/v*) glycerol, 750 mM NaCl, 0.5% (*w/v*) CHAPS, 0.1% (*w/v*) DDM, 0.01% (*w/v*) CHS, 0.05 mg/mL of a combination of POPC/POPG/POPE, 1 mM EDTA, 5 mM DTT

GF1 buffer

50 mM Tris-HCl (pH 7.4), 15% (*v/v*) glycerol, 500 mM NaCl, 0.1% (*w/v*) DDM, 0.01% (*w/v*) CHS, 0.05 mg/mL POPC, 1 mM EDTA, protease inhibitors

GF2 buffer

50 mM Tris-HCl (pH 7.4), 50 mM NaCl, 0.1% (*w/v*) DDM, 0.01% (*w/v*) CHS, 1 mM EDTA, protease inhibitors

Detergent buffer (for NMR experiments)

GF2 buffer with 5% (*v/v*) glycerol

B.5 Reconstitution

Reconstitution buffer

50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA

Sucrose gradient

0-35% (*w/v*) sucrose in 50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA

Low-salt buffer

20 mM KCl, 1 mM EDTA, 5% (*v/v*) glycerol

B.6 Binding assays

Assay buffer (detergent-solubilised samples)

50 mM Tris-HCl (pH 7.4), 0.1% (*w/v*) DDM, 0.01% (*w/v*) CHS, 1 mM EDTA, 0.1 mg/mL BSA

Assay buffer (reconstituted samples)

50 mM Tris-HCl (pH 7.4), 1 mM EDTA, 0.1 mg/mL BSA.

Appendix C - List of a crystallisation screening kit

Table C.1: The list of a crystallisation screening kit (Structure Screen 1 - Catalogue Number MD1-01)

Precipitant no.	Mixture solution		
1	0.02M Calcium chloride dihydrate	0.1M Na Acetate trihydrate pH 4.6	30% v/v 2-methyl-2,4 pentanediol
2	0.2M Ammonium acetate	0.1M Na Acetate trihydrate pH 4.6	30% w/v PEG 4000
3	0.2M Ammonium sulphate	0.1M Na acetate trihydrate pH 4.6	25% w/v PEG 4000
4	None	0.1M Na acetate trihydrate pH 4.6	2.0M Sodium formate
5	None	0.1M Na acetate trihydrate pH 4.6	2.0M Ammonium sulphate
6	None	0.1M Na acetate trihydrate pH 4.6	8% w/v PEG 4000
7	0.2M Ammonium acetate	0.1M tri-sodium citrate dihydrate pH 5.6	30% w/v PEG 4000
8	0.2M Ammonium acetate	0.1M tri-sodium citrate dihydrate pH 5.6	30% v/v 2-methyl-2,4-pentanediol
9	None	0.1M tri-Sodium citrate dihydrate pH 5.6	20% w/v 2-propanol, 20% w/v PEG 4000
10	None	0.1M Na Citrate pH 5.6	1.0M Ammonium dihydrogen phosphate
11	0.2M Calcium chloride dihydrate	0.1M Na acetate trihydrate pH 4.6	20% v/v 2-propanol
12	None	0.1M Na Cacodylate pH 6.5	1.4M Na acetate trihydrate
13	0.2M tri-sodium citrate dihydrate	0.1M Na Cacodylate pH 6.5	30% v/v 2-propanol
14	0.2M Ammoniium sulphate	0.1M Na Cacodylate pH 6.5	30% w/v PEG 8000
15	0.2M Magnesium acetate tetrahydrate	0.1M Na Cacodylate pH 6.5	20% PEG 8000
16	0.2M Magnesium acetate tetrahydrate	0.1M Na Cacodylate pH 6.5	30% v/v 2-methyl-2,4-pentanediol
17	None	0.1M Imidazole pH 6.5	1.0M Sodium acetate trihydrate
18	0.2M Sodium acetate trihydrate	0.1M Na Cacodylate pH 6.5	30% w/v PEG 8000
19	0.2M Zinc acetate dihydrate	0.1M Na Cacodylate pH 6.5	18% w/v PEG 8000
20	0.2M Calcium acetate hydrate	0.1M Na Cacodylate pH 6.5	18% w/v PEG 8000
21	0.2M tri-sodium citrate dihydrate	0.1M Na Hepes pH 7.5	30% v/v 2-methyl-2,4-pentanediol
22	0.2M Magnesium chloride hexahydrate	0.1M Na Hepes pH 7.5	30% v/v 2-propanol
23	0.2M Calcium chloride dihydrate	0.1M Na Hepes pH 7.5	28% v/v PEG 400
24	0.2M Magnesium chloride hexahydrate	0.1M Na Hepes pH 7.5	30% v/v PEG 400

Continued

Precipitant no.	Mixture solution		
25	0.2M tri-sodium citrate dihydrate	0.1M Na Hepes pH 7.5	20% v/v 2-propanol
26	None	0.1M Na Hepes pH 7.5	0.8M K, Na tartrate tetrahydrate
27	None	0.1M Na Hepes pH 7.5	1.5M Lithium sulphate monohydrate
28	None	0.1M Na Hepes pH 7.5 0.8M K dihydrogen phosphate monohyd.	0.8M Na dihydrogen phosphate
29	None	0.1M Na Hepes pH 7.5	1.4M tri-Sodium citrate dihydrate
30	None	0.1M Na Hepes pH 7.5	2% v/v PEG 400, 2.0M Amm sulphate
31	None	0.1M Na Hepes pH 7.5	10% v/v 2-propanol, 20% w/v PEG 4000
32	None	0.1M Tris HCl pH 8.5	2.0M Ammoniium sulphate
33	0.2M Magnesium chloride hexahydrate	0.1M Tris HCl pH 8.5	30% w/v PEG 4000
34	0.2M tri-sodum citrate dihydrate	0.1M Tris HCl pH 8.5	30% v/v PEG 400
35	0.2M Lithium sulphate monohydrate	0.1M Tris HCl pH 8.5	30% w/v PEG 4000
36	0.2M Ammonium acetate	0.1M Tris HCl pH 8.5	30% v/v 2-propanol
37	0.2M Sodium acetate trihydrate	0.1M Tris HCl pH 8.5	30% w/v PEG 4000
38	None	0.1M Tris HCl pH 8.5	8% w/v PEG 8000
39	None	0.1M Tris HCl pH 8.5	2.0M Ammonium dihydrogen phosphate
40	None	None	0.4M K, Na Tartrate tetrahydrate
41	None	None	0.4M Ammonium dihydrogen phosphate
42	0.2M Ammonium sulphate	None	30% w/v PEG 8000
43	0.2M Ammonium sulphate	None	30% w/v PEG 4000
44	None	None	2.0M Ammonium sulphate
45	None	None	4.0M Sodium formate
46	0.05M Potassium dihydrogen phosphate	None	20% w/v PEG 8000
47	None	None	30% w/v PEG 1500
48	None	None	0.2M Magnesium formate
49	1.0M Lithium sulphate monohydrate	None	2% w/v PEG 8000
50	0.5M Lithium sulphate monohydrate	None	15% w/v PEG 8000

Appendix D – Solution- and solid-state NMR spectra

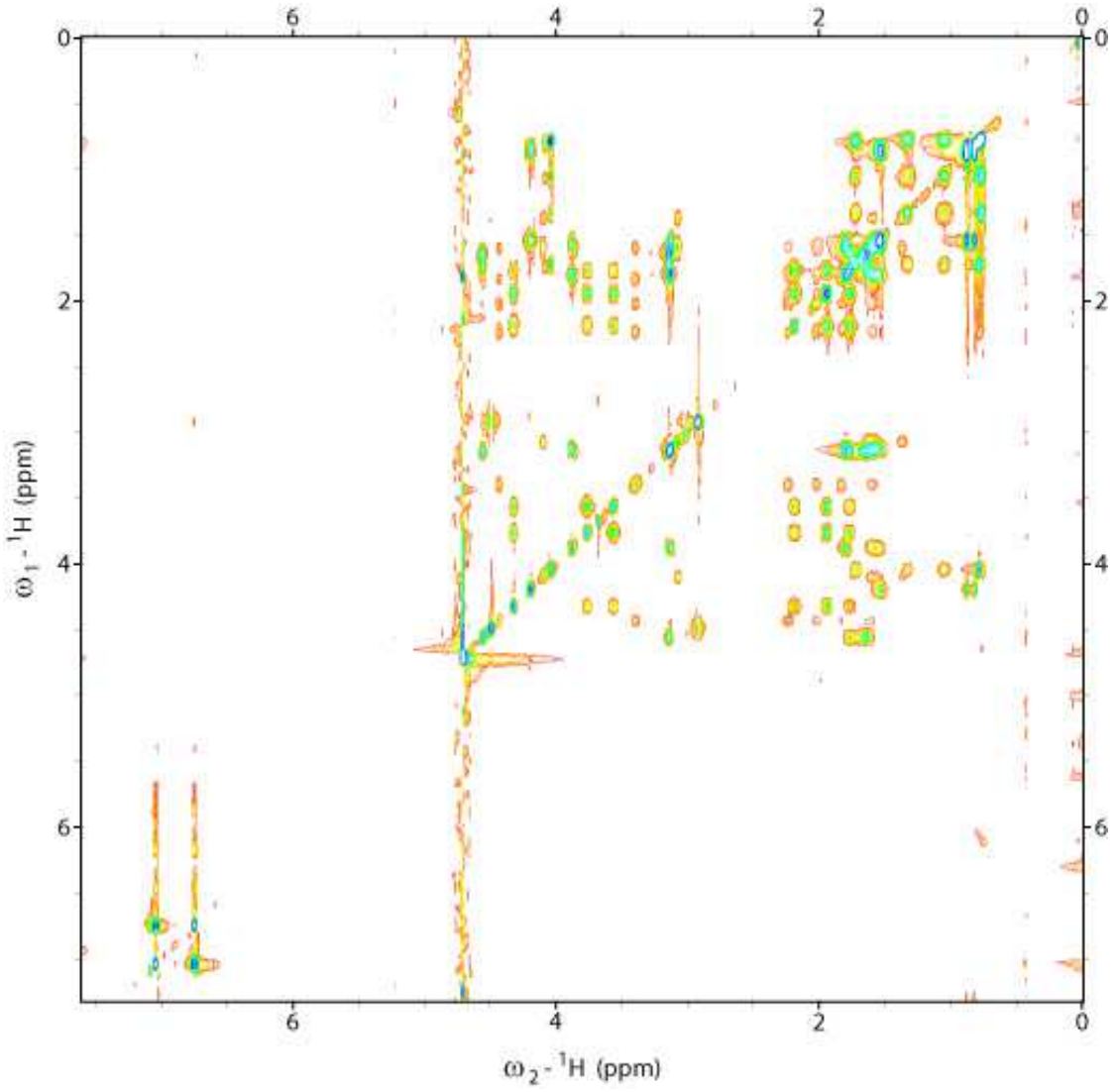


Figure D.1: TOCSY spectrum of $NT_{8\text{-}13}$ peptide in aqueous solution (pure D_2O, pH 6.0). The spectrum was recorded on an Oxford Instruments 750 MHz NMR spectrometer at 298 K.

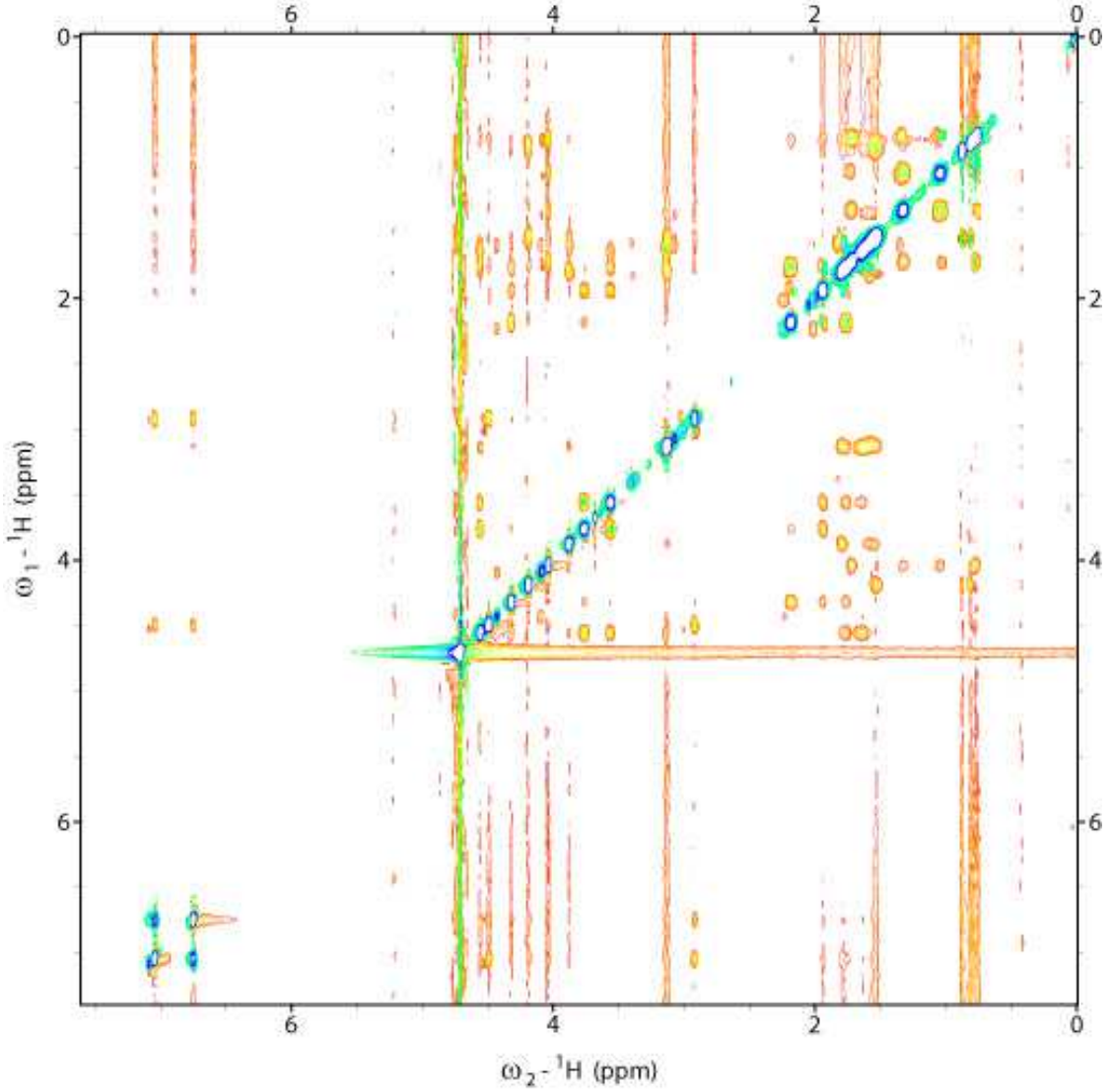


Figure D.2: ROESY spectrum of $NT_{8\text{-}13}$ peptide in aqueous solution (pure D_2O, pH 6.0). The spectrum was recorded on an Oxford Instruments 750 MHz NMR spectrometer at 298 K.

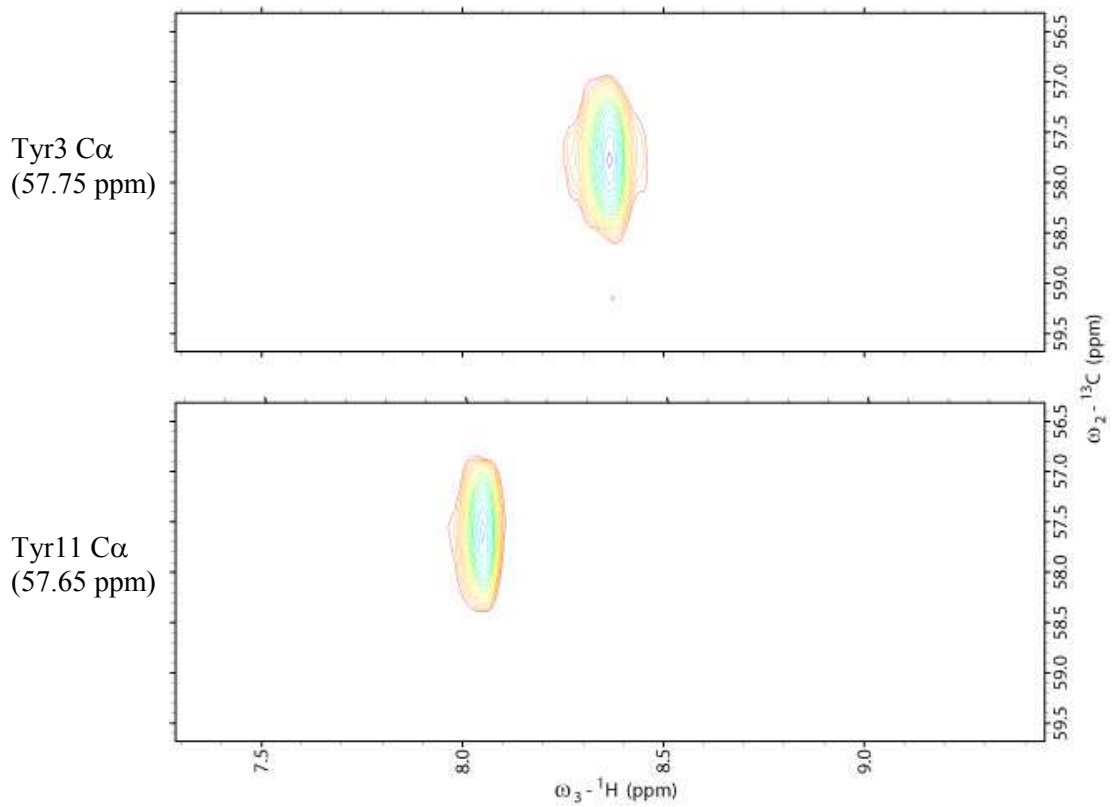


Figure D.3: Selected contour plots of slices taken from the 3D CBCANH spectrum of U-^{13}C, ^{15}N labelled NT in detergent buffer, pH7.4 with 5% (*v/v*) D_2O. The spectrum was recorded on an Oxford Instruments 600 MHz NMR spectrometer at 298 K.

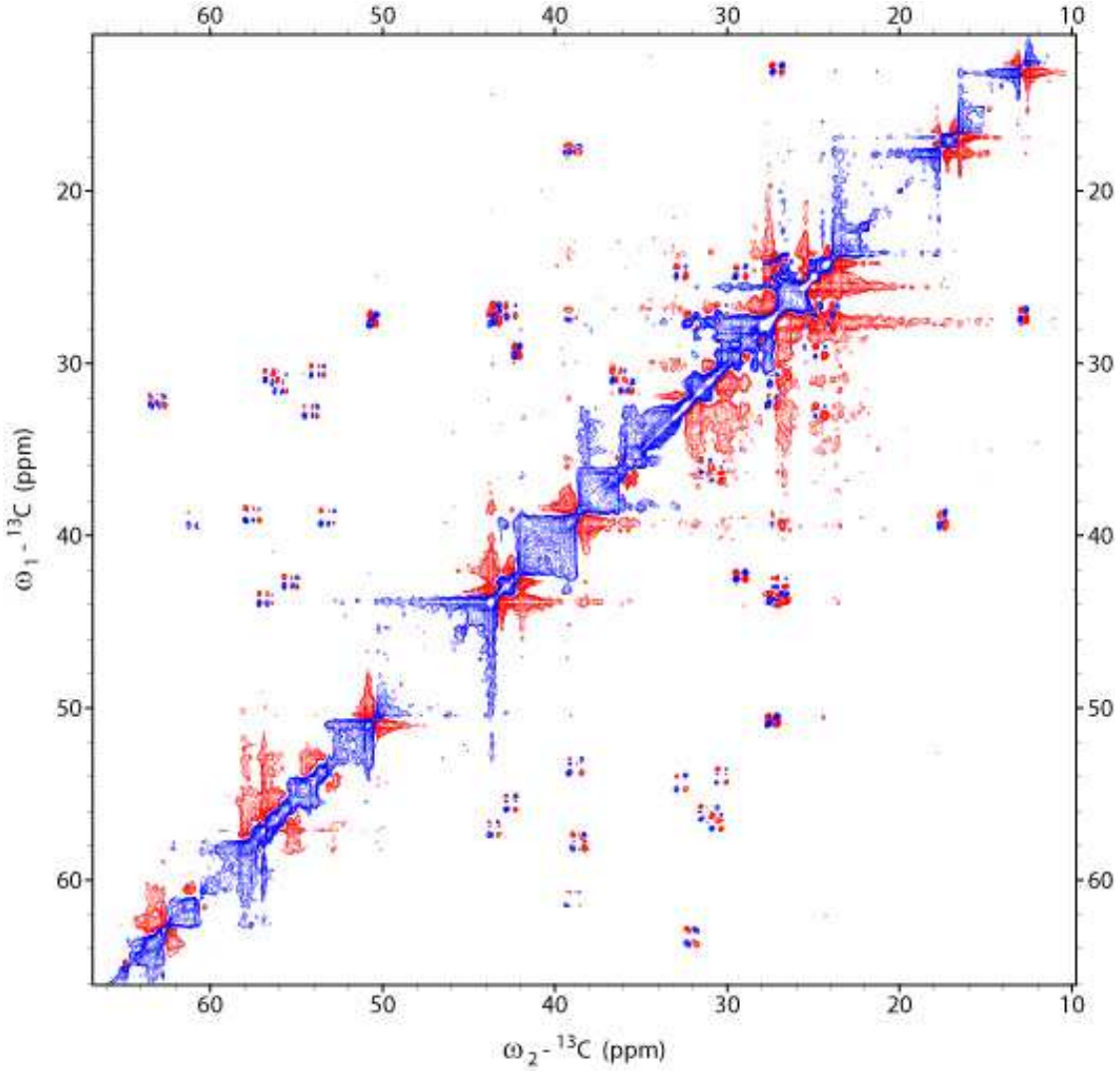


Figure D.4: 2D ^{13}C-^{13}C COSY spectrum of U-^{13}C, ^{15}N labelled NT in BPL, pH 7.4. The spectrum was recorded on a Varian 500 MHz SSNMR spectrometer at 278 K. Conditions: 6 kHz, 32 scans and 20.48 ms acquisition time.

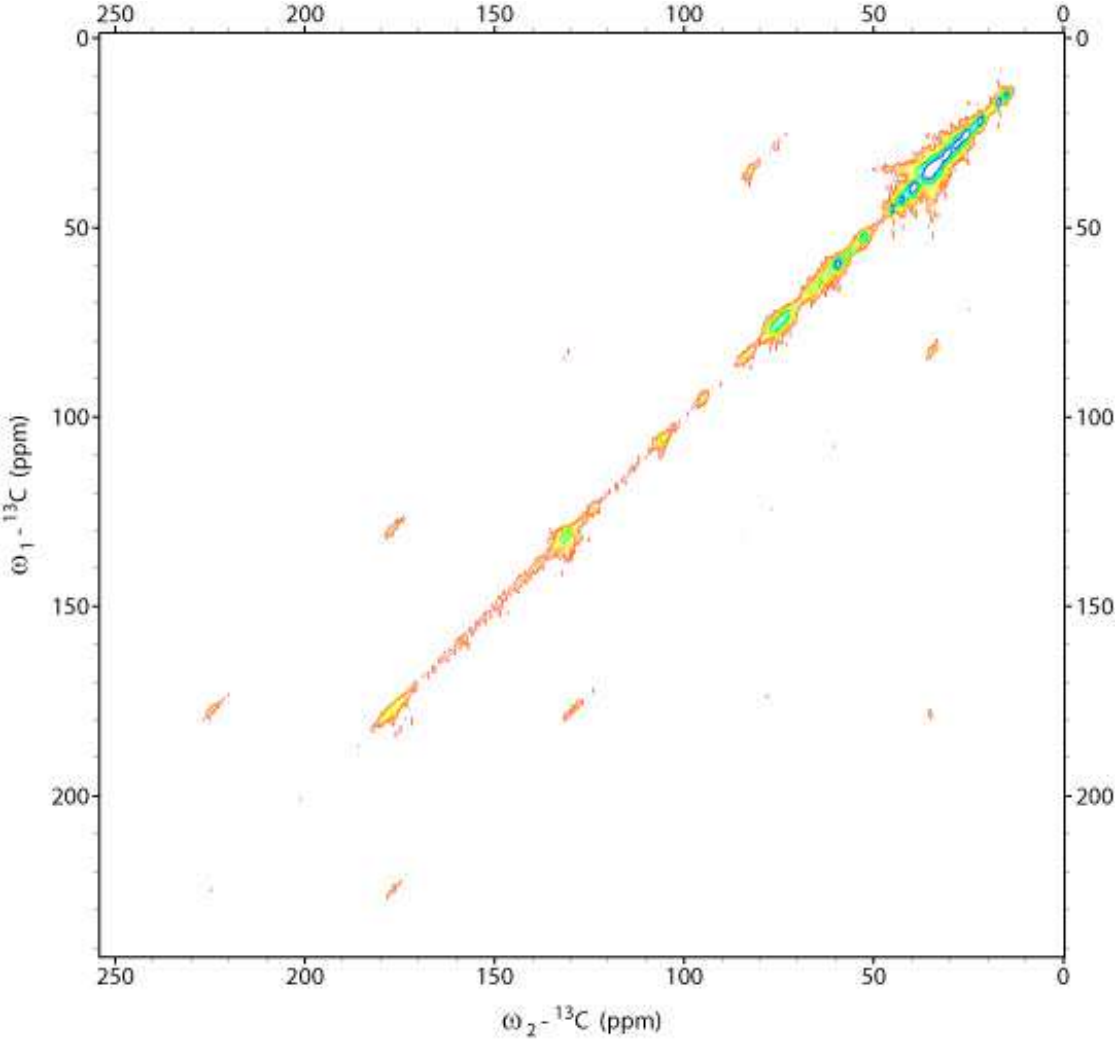


Figure D.5: 2D ^{13}C-^{13}C DARR spectrum of U-^{13}C, ^{15}N labelled NT bound to the NTS1B reconstituted into the BPL, pH 7.4. The spectrum was recorded with 13 ms mixing time, on a Varian 500 MHz SSNMR spectrometer, at 233 K, and with a spinning frequency of 6 kHz.