STRUCTURAL CHARACTERIZATION OF THE N-TERMINAL REGION OF THE SACCHAROMYCES CEREVISIAE G-PROTEIN COUPLED RECEPTOR, STE2P

A Thesis Submitted to the College of

Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of Masters of Science

in the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

Stephanie Kendall

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised my thesis work, or in their absence, by Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and the University of Saskatchewan in any scholarly use which may be made of any materials in my thesis.

Request for permission to copy or make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Biochemistry

University of Saskatchewan

Saskatoon, Saskatchewan, S7N 5E5

ABSTRACT

G-protein coupled receptors (GPCRs) form a superfamily of cell surface receptors with in excess of 2000 genes identified across taxa (Pierce *et al.*, 2002). They are integral membrane proteins that are comprised of seven hydrophobic helical segments which form a transmembrane spanning bundle. Ste2p and Ste3p are *Saccharomyces cerevisiae* GPCRs that are the α-factor and a-factor pheromone receptors, respectively (Bardwell, 2005). Ste2p in particular has served as an excellent model for studying the mechanisms of action of GPCRs. Recent results suggest that the extracellular N-terminus of the Ste2p receptor is involved in modulating cell wall degradation and membrane juxtaposition during yeast mating potentially by mediating an intercellular interaction with Ste3p (Shi *et al.*, 2009a).

The goals of this project were to obtain purified mg quantities of a soluble version of a portion of the N-terminus of Ste2p and to acquire structural information about this region by performing biophysical analysis on the soluble N-terminal Ste2p fragments. Initially, a synthetically produced KKK-Ste2p(14-43)-KKK peptide yielded circular dichroism (CD) spectra that indicated peptide secondary structure similar to what has been predicted *in silico*. Preliminary nuclear magnetic resonance (NMR) experiments with this peptide were also promising, yielding a correlation spectroscopy (COSY) spectrum that allowed for limited amino acid assignments. A recombinant version of the N-terminus of Ste2p, including residues 2-48 with terminal lysine residues, was expressed as a fusion protein in *E. coli* and the SK peptide was liberated by cyanogen bromide cleavage. Dynamic light scattering (DLS) analyses of the SK peptide showed it was aggregated when dissolved in water, but soluble in a trifluoroethanolamine/water (TFE/H_2O) mixture. CD analysis of the SK peptide in TFE/H_2O indicated that it contained more β-strand and less α-helix than the KKK-Ste2p(14-43)-KKK peptide. NMR analysis was performed on both an unlabelled and ^{15}N-labelled SK peptide, yielding unusable spectra with very broad bands, most likely arising from aggregation at high concentrations. In conclusion, a high yield recombinant expression and purification system has been developed, yielding a Ste2p N-terminal peptide fragment that demonstrates some expected structural features. High resolution structural information may be obtained upon further optimization of the solvent system.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor, Dr. Michèle Loewen, for all of her encouragement and support throughout my Master's project. Thanks to her endless suggestions, advice, good ideas and good company during this time. I appreciate everything she has done for me and would not have made it through my program without her help. I would like to thank the members of my committee, Dr. Oleg Dmitriev, Dr. Ramji Khandelwal, Dr. Stan Moore, Dr. Bill Roesler and Dr. Veronica Campanucci. I would especially like to thank Dr. Dmitriev for his help and advice with my NMR experiments and assignments. I would like to thank all of the members of the Loewen lab including Dr. Olesya Kharenko, Dr. Chunhua Shi, and Pooja Choudary. A special thank you to Olesya for all of her help throughout my Master's project, she made working in the lab enjoyable. I would like to thank Dang Van and Michelle Forgeron for their help with my NMR experiments and Jason Maley for his help with my biophysical experiments at the SSSC. I would like to thank my family for their constant encouragement and guidance.

Dedication

I would like to dedicate my thesis to my husband and son.
Without their love and support I could never have completed this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2xYT	2x Yeast extract and tryptone
ACN	Acetonitrile
cAMP	3'-5'-Cyclic adenosine monophosphate
β_2AR	β_2- adrenergic receptor
BRET	Bioluminescence resonance energy transfer
CD	Circular dichroism
CNBr	Cyanogen bromide
COSY	Correlation spectroscopy
DLS	Dynamic light scattering
DMPC	Dimyristol phosphatidylcholine
EL	Extracellular loop
Fab	Fragment antigen binding
FID	Free induction decay
FRET	Fluorescent resonance energy transfer
GABA	γ-aminobutyric acid
GAPs	GTPase-activating proteins
GPCRs	G-protein coupled receptors
HMBC	Heteronuclear multiple bond coherence
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
IL	Intracellular loop
IPTG	isopropyl β-D-thiogalactopyranoside
KSI	Ketosteroid isomerase
MAPK	Mitogen activated protein kinase
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
PCR	Polymerase chain reaction
ROESY	Rotating frame Overhauser effect spectroscopy
SDS	Sodium dodecyl sulphate

TFA	Trifluoroacetic acid
TFE	Trifluoroethanolamine
TM	Transmembrane
7-TM	Seven transmembrane proteins
TOCSY	Total correlation spectroscopy

1.0 LITERATURE REVIEW

1.1 Introduction

G-protein coupled receptors (GPCRs) are membrane proteins called seven transmembrane proteins (7 TM) (Pierce *et al.*, 2002). The 7 TM proteins are integral membrane proteins that have seven hydrophobic segments predicted to form transmembrane spanning helices. The transmembrane domains are connected by intracellular and extracellular loops and the N-terminus is extracellular, while the C-terminus is intracellular. These receptors mediate extracellular signals (proton gradients, fungal infection, light, hormones, calcium etc.) into intracellular responses. In the case of the GPCRs this signalling is mediated through downstream G-protein activation. Activation of the G-protein subunits leads to further activation of signalling cascades. To date, a limited number of 3D structures of GPCRs have been resolved using either X-ray crystallography or NMR. GPCR structural studies are important because GPCRs represent an important target for therapeutic drug design with approximately 30% of current drugs targeting GPCRs.

The yeast, *Saccharomyces cerevisiae,* is an important model system for studying the molecular and cell biology of higher eukaryotes. Ste2p and Ste3p are *S. cerevisiae* GPCRs that are the α-factor and a-factor pheromone receptors, respectively (Bardwell, 2005). These receptors are expressed on opposite cell types and it is well documented that activation of these receptors by pheromones leads to G-protein mediated signalling for initiation of mating in yeast. However it was also shown, using tandem affinity purification and proteoliposome fusion assays, that activated forms of these receptors interact with each other directly to promote membrane juxtaposition (Shi *et al.*, 2007). I*n vivo* analysis indicated that when a yeast strain with the Ste2p receptor deleted, had an N-terminally truncated Ste2p receptor expressed, this lead to arrest prior to cell wall degradation. *In vitro* characterization of this truncation indicated loss of the interaction between Ste2p and Ste3p as well as loss of the receptor's ability to promote membrane juxtaposition. More recently it was found that the residues Pro 15, Ile 24 and Ile 29 in the N-terminal region of Ste2p are important in mediating mating efficiency (Shi *et al.*, 2009a). These results suggested that the N-terminus of the Ste2p receptor is involved in cell wall degradation and membrane juxtaposition during yeast mating potentially by mediating the interaction with Ste3p. NMR analyses of the N-terminus of Ste2p would give insight into

the mechanisms associated with how the GPCR modulates events in later steps of mating including cell wall degradation and membrane fusion.

1.2 G Protein Coupled Receptors (GPCRs)

GPCRs, also known as 7 TM receptors, are integral membrane proteins that have seven hydrophobic helical segments predicted to form transmembrane spanning helices (Pierce *et al.*, 2002). The transmembrane domains are connected by intracellular and extracellular loops and the N-terminus is extracellular, while the C-terminus is intracellular. These receptors mediate a wide array of extracellular signals, such as proton gradients, fungal infection, light, hormones, calcium, odorants, peptide pheromones, etc. into intracellular responses (Bockaert and Pin, 1999). The GPCR superfamily is a large and diverse family that has been found in the genomes of animals, plants and fungi. In the human genome alone there are approximately 800 GPCRs (Bhasin and Raghava, 2004; Fredriksson *et al.*, 2003). The GPCRs are an important target family for therapeutic drug design and ~30% of all marketed prescription drugs target GPCRs (Jacoby *et al.*, 2006).

All GPCRs have a similar secondary structure which is characterized by seven transmembrane spanning α-helices connected by alternating intracellular and extracellular loops. However, even though they all have very similar structural features, there is no significant sequence similarity across the GPCR superfamily (Probst *et al.*, 1992). Therefore, classification of the GPCRs into distinct families has been complicated. GPCRs have been classified using a number of different techniques, such as sequence similarity, phylogenetic criteria, structural and ligand binding criteria, amino acid composition, and many others (Bockaert and Pin, 1999; Chou and Elrod, 2002; Fredriksson *et al.*, 2003; Kolakowski Jr., 1994).

Based on sequence similarity classification, the GPCRs have been separated into six different families, labelled A-F (Horn *et al.*, 2003; Kolakowski Jr., 1994). The Class A receptors, or the rhodopsin-like family, is by far the largest family of GPCRs and encompasses ~80% of all GPCRs. This family includes receptors for biogenic amines, hormone proteins, odorants, peptides and lipids (Cheng *et al.*, 2005). This class therefore represents a very important group of GPCRs that are involved in many pathological conditions such as Parkinson's disease, schizophrenia and other mood disorders. Class B receptors, or the

Secretin-like family, is the next largest family of GPCRs. These receptors bind large peptides such as secretin, glucagon, vasoactive intestinal peptide, parathyroid hormone, calcitonin and more (Cardoso *et al.*, 2006). The N-terminus of these receptors, which is important for ligand binding, varies from between 60-80 residues in length and contains conserved cysteine bridges (Fredriksson *et al.*, 2003). The Class C receptors, or the Metabotropic glutamate/pheromone family, are a smaller group of GPCRs. This family includes metabotropic glutamate receptors, γ-aminobutyric acid (GABA) receptors, calcium-sensing receptors, mammalian pheromone receptors and taste receptors (Fredriksson *et al.*, 2003). The majority of the receptors in this family have a very large N-terminal extracellular domain and many have conserved cysteine residues which may be involved in ligand binding (Pin *et al.*, 2003). The Class D receptors include the fungal pheromone receptors, Ste2p and Ste3p. This class of receptors will be discussed in detail in Section 1.2. The Class E receptors, or the cyclic adenosine monophosphate (cAMP) receptors, are a very small group of GPCRs that are related to the cAMP receptors from the species, *D. discoideum*, also known as slime mold (Raisley *et al.*, 2004). The final class of receptors are the Class F receptors or the Frizzled/Smoothened family. The frizzled receptors bind to specific ligands from a family of lipoglycoproteins called WNT. These WNT proteins bind to the cysteine-rich extracellular terminus of the frizzled receptors (Schulte and Bryja, 2007). The smoothened receptors are involved in mediating signalling of the hedgehog pathway (Wang *et al.*, 2009).

In general, GPCR signalling in the cell is accomplished through the interaction with heterotrimeric guanine nucleotide-binding proteins, or G-proteins. The heterotrimer consists of Gα/Gβ and Gγ subunits which are membrane bound and in close proximity to the intracellular surface of GPCRs. Upon activation of the GPCR, due to ligand binding, the receptor undergoes a conformational change which allows for interaction between the intracellular surface of the receptor and the G proteins (Altenbach *et al.*, 2008; Scheerer *et al.*, 2008). This interaction results in activation of the G protein by inducing a conformational change in the G protein and exchanging GDP for GTP (Lambright *et al.*, 1994). Once the GDP is exchanged for GTP, the heterotrimeric G-protein dissociates from the receptor and the Gα subunit dissociates from the Gβγ subunit. The Gα and Gβγ subunits go on to further activate or inhibit effector proteins, such as tyrosine kinases, phosphodiesterase, phosphoinositide-3-kinase, mitogen activated

protein kinases (MAPK), etc (Kristiansen, 2004). The G-protein subunits stay activated until the GTP is hydrolyzed to GDP or inactivated by GTPase-activating proteins (GAPs).

1.2.1 GPCR Structural Analysis

As was mentioned previously, GPCRs all contain seven hydrophobic transmembrane-spanning domains. These domains are linked together by intracellular and extracellular loops. The N-terminal region is extracellular, while the C-terminal tail is intracellular. Increased knowledge of the three-dimensional structure of GPCRs is very important for understanding their mechanisms of signalling and for use in designing new drugs. Currently, the two main tools being used for structural analyses of GPCRs are NMR spectroscopy and X-ray crystallography. X-ray crystallography uses X-rays to bombard protein crystals in order to determine their overall three-dimensional structure. The first high resolution X-ray crystal structure of GPCRs was that of rhodopsin (Palczewski *et al.*, 2000). The structure was solved with the co-factor 11-*cis*-retinal, a derivative of vitamin A, which held the TM region of rhodopsin in the inactive state. This was the first structure to confirm the 7 TM spanning helices, intradiscal and cytoplasmic loop domains as well as the conformation of the side-chains. Since then a number of other rhodopsin structures have emerged which have allowed for further refinement of the original structure (Li *et al.*, 2004a; Okada *et al.*, 2002; Okada *et al.*, 2004; Salom *et al.*, 2006; Teller *et al.*, 2001). These structures have revealed that rhodopsin has a β-hairpin conformation in the second intradiscal loop, which is believed to be involved in sheltering retinal in the closed binding pocket. This feature has been seen in all crystal forms of rhodopsin but not for other GCPRs. The original crystal structure also confirmed the disulfide bridge that connects the intradiscal side of transmembrane 3 (TM3) and intradiscal loop 2. This is a common feature of GPCRs. The N-terminus of Rhodopsin is 34 residues long and was found to contain five distorted β-strands which form a compact unit and overlays the other intradiscal loops. For a number of years this was the only GPCR structure that was available.

The next GPCR structure that was solved was that of the human β_2- adrenergic receptor (β_2AR). Initially, two structures were solved for β_2AR, the first of which was a fusion protein where the third intracellular loop (IL3) was replaced with T4 lysozyme and this fusion protein was bound to the partial inverse agonist, carazolol (Cherezov *et al.*, 2007). Previous attempts to crystallize the wild-type β_2AR were unsuccessful and this was thought to be because the IL3

conformation was highly flexible which would interfere in crystal-lattice formation. Therefore, in the first crystal structure the IL3 was replaced with a soluble T4 lysozyme which resulted in the ability to produce crystals (Rosenbaum *et al.*, 2007). The other structure was bound to carazolol as well as to a antigen binding fragment (Fab) which recognizes the IL3 and provides conformational stability allowing for crystallization (Rasmussen *et al.*, 2007). Both of these structures were similar indicating that the addition of T4 lysozyme or the binding of Fab had little effect on the overall native structure. These structures demonstrate a major difference to that of rhodopsin in the structure of extracellular loop 2 (EL2). In the rhodopsin structure, intradiscal or EL2 forms a β-strand and closes off access to the binding pocket from the extracellular side, however in the β_2AR structures the EL2 is partially α-helical and the binding pocket is left open so that it is easily accessible to ligands. The N-terminus (residues 1-28) is disordered and not visible in the structure. The next structure that was solved was that of β_2AR again fused to T4 lysozyme, but containing a stabilizing point mutation, E122W, which was solved in complex with the inverse agonist timolol (Hanson *et al.*, 2008). This structure suggested a cholesterol binding pocket located between TM2, TM3 and TM4. Then the turkey β_1AR structure was solved, which contained multiple stabilizing mutations and was solved in complex with the strong antagonist cyanopindolol (Warne *et al.*, 2008). The human and turkey receptors had significant sequence similarity and it was therefore no surprise that the structures were very similar. However, the major difference was in a portion of the EL2, in the turkey structure this region was α-helical while in the human structure it was random coil. The N-terminus of the turkey receptor was truncated in order to allow for purification and therefore information on the structure of the N-terminus was unavailable.

The fourth GPCR structure to be solved was that of A_{2A} adenosine receptor in complex with the antagonist ZM241385 (Jaakola *et al.*, 2008). The adenosine structure demonstrated similarities to both the β_2-AR and rhodopsin structures; however it also had a few differences such as a shifted ligand binding site. As well, the EL2 of the adenosine receptor was significantly different from that of either the rhodopsin or β_2AR structures. However, similar to the β_2AR structure, it did leave the binding pocket open for ligand binding. No structural information was provided on the short N-terminus. Another rhodopsin structure that was solved was that of light-activated opsin in complex with a synthetic peptide corresponding to the C-terminal end of the Gα subunit transducin (Scheerer *et al.*, 2008). This structure

demonstrated differences from the inactive rhodopsin structure in the arrangement of the seven transmembrane bundles. In addition, in the activated rhodopsin structure, the N-terminus contained only two antiparallel β-strands, while the inactivated structure contained five β-strands. One of the most recent structures to be solved was that of CXCR4 chemokine receptor (Wu *et al.*, 2010). The structure was solved in complex with a small molecule antagonist, IT1t, and with a cyclic peptide inhibitor, CVX15. These antagonists were chosen based on ligand solubility, binding affinity and induced protein thermostability. This structure also contained the insertion of T4 lysozyme into IL3. The N-terminal residues of this structure did not have interpretable density and were presumed to be disordered. Another recent structure to be solved was that of the human dopamine D3 receptor (D3R), which again contained the insertion of T4 lysozyme into IL3 (Chien *et al.*, 2010). The EL2 is much shorter in the D3R structure when compared to the structure of the βAR's. Overall, these six crystal structures highlight many of the common structural characteristics between GPCRs but also indicate that each GPCR has its own interesting structural features. Historically, the reason for the limited number of crystal structures is that using X-ray crystallography for GPCR structure determination can sometimes be very complicated. First, purification of large amounts of GPCRs can be quite difficult. Also, many GPCRs are very difficult to crystallize because they require detergents or other solvents to solubilise them which can interfere in the crystallization. Finally, while the recent successes described above highlight a bright future, that is based on modifications and ligand associations for crystallizing more GPCRs, structural information about N-terminal regions remains elusive.

NMR takes advantage of the ability of nuclei to absorb electromagnetic energy and radiate this energy out at a specific frequency. This frequency is dependent on the nuclei's environment, which gives information about the structure of the protein. NMR theory will be discussed in more detail in Section 1.3.3. To date there has been no complete 3D structure of a GPCR solved using NMR. This is because recombinant expression of GPCR proteins is quite difficult for cell expression systems. As well the need for a large amount of pure, isotopically labelled protein can be quite expensive to produce. On top of all this, in order to perform NMR on a protein it has to be soluble, which often requires the use of detergents or lipid micelles. Therefore, the combination of these factors as well as the limited thermal stability and short lifetime of GPCRs has led to the inability to determine a full structure using NMR. However,

the first full structure of a 7TM protein using NMR was recently obtained on the photoaxis receptor sensory rhodopsin II (pSRII) from *Natronomonas pharaonis* (Gautier *et al.*, 2010). This structure was obtained with good resolution and is almost identical to the structure obtained by X-ray crystallography. Therefore, recombinant expression systems are improving in order to be able to produce large quantities of higher molecular weight proteins. However, it is still very difficult to recombinantly express the larger full length GPCRs, so all current NMR structural data has been obtained from fragments of GPCRs (Chung *et al.*, 2002; Conner *et al.*, 2008; Werner *et al.*, 2008; Zou *et al.*, 2008). Focusing on data related to the structurally enigmatic N-terminal region, a fragment of the human Y4 GPCR containing the N-terminus, TM1, TM2 and EL1 was solved using NMR (Zou *et al.*, 2008). The N-terminus (residues 1-40) was found to contain a short helix (residues 5-10), an unstructured portion (residues 11-30) and a terminal helix (residues 31-40) which connected the N-terminus to TM1. A terminal helix (residues 39-47) was also found using NMR structural analysis of the yeast GPCR, Ste2p (Neumoin *et al.*, 2009). This demonstrates that NMR can be applied to protein fragments in order to solve structural questions for proteins that are very difficult to crystallize. Another important advantage of NMR is that dynamic information about the protein can be obtained, whereas in X-ray crystallography only static information is available. Thus, receptor-ligand interactions can be studied using NMR (Tapaneeyakorn *et al.*, 2010).

1.3 *Saccharomyces cerevisiae*

Saccharomyces cerevisiae is a species of budding yeast, which is also known as baker's yeast. *S. cerevisiae* has been used as a model system for molecular and cell biology research in eukaryotes. It is an extremely useful model system because the molecular bases of recombination, replication, cell division, and metabolism are generally conserved between yeast and higher eukaryotes, such as mammals. The life cycle of *S. cerevisiae* consists of the yeast having either two genomes (diploid) or one genome (haploid), and during the life cycle the yeast alternates between these two forms using the processes of mating and germination.

1.3.1 *S. cerevisiae* Mating Overview

The yeast *S. cerevisiae* contains two types of cells for use in mating, the mating-type a cell and the mating-type α cell, also known as the *MAT*a and *MAT*α cells, respectively. *MAT*a and *MAT*α are haploid cells, and during mating two opposite cell types fuse to form a *MAT*a/*MAT*α diploid. Mating in *S. cerevisiae* is initiated by the release of peptide pheromones which recognize receptors on the opposite haploid yeast cell (Figure 1.1) [reviewed in (Dohlman, 2002; Dohlman *et al.*, 1991; Elion, 2000; Konopka and Fields, 1992)]. *MAT*α cells secrete α-factor pheromone, a tridecapeptide with the peptide sequence WHWLQLKPGQPMY, which binds to the GPCR, Ste2p, on *MAT*a cells. *MAT*a cells secrete a-factor, a dodecapeptide pheromone with a farnesyl lipid group attached. The a-factor peptide sequence is YIIKGVFWDPAC and the peptide binds to the GPCR, Ste3p, on *MAT*α cells. The mating pheromones are essential for initiation of mating. Cells that cannot produce these pheromones or lack the required GPCR cannot undergo mating.

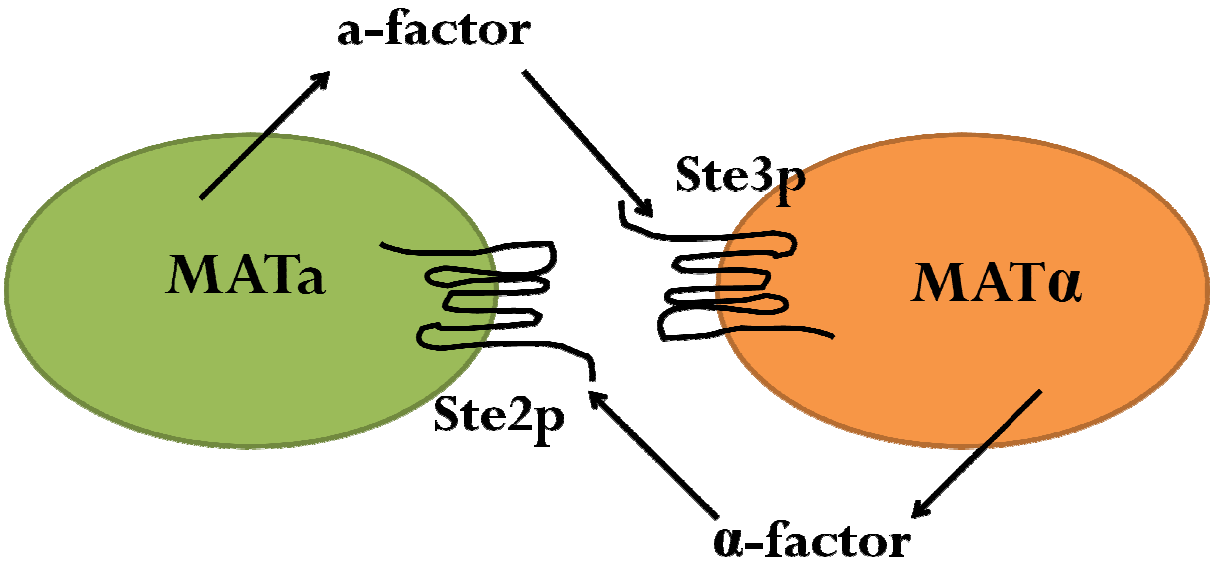

Figure 1.1: Mating of the Yeast *Saccharomyces cerevisiae*.
Mating in *S.cerevisiae* involves two haploid cells, MATa and MATα, which secrete pheromones, a-factor and α-factor, respectively. The pheromones recognize the receptor on the opposite mating type cell, which initiates mating.

Once a yeast cell is stimulated by pheromone secreted from the opposite cell type, this results in activation of the cell's pheromone receptors and the interacting G proteins. The G protein complex that interacts with Ste2p and Ste3p contains Gpa1 (Gα), Ste4p (Gβ) and Ste18p (Gγ). When the G protein is activated, the Gβγ subunits are released from Gα and the Gβγ subunit goes on to bind to three different effectors; the Ste5p complex, the Ste20p protein kinase, and the Far1/Cdc24 complex (Figure 1.2) [reviewed in (Bardwell, 2005)]. The Ste5p complex involves Ste5p as a binding platform (scaffold) for other proteins. Ste5p binds to both Ste4p (Gβ) and Ste11p protein kinase allowing for Ste11p to be closer to the plasma membrane (Pryciak and Huntress, 1998). Ste20p is a member of the p21-activated protein kinase (PAK) family (Lim *et al.*, 1996). The C-terminal end of Ste20p is bound to Gβγ and Ste20p is activated once it binds to the Rho-like G protein, Cdc42 (Ash *et al.*, 2003). This brings Ste20p close to the plasma membrane and to Ste11p. Ste20p phosphorylates and activates Ste11p which initiates the MAP kinase cascade. The Far1/Cdc24 complex involves the N-terminal end of Far1 bound to Gβγ and the C-terminal end bound to Cdc24 (Butty *et al.*, 1998). Cdc24 is a guanine exchange factor for Cdc42. Therefore, this complex is involved in Cdc42 activation by exchanging GDP for GTP on Cdc42. This allows for binding of Cdc42 to Ste20p and further activation of Ste20p. Once the MAP kinase cascade has been initiated this results in a number of changes in the cell. The cells 1) synthesize cell surface molecules necessary for fusion with their mating partners, 2) arrest in G1 phase of the growth cycle in order to synchronize the cell cycles of the two mating partners prior to cell fusion, 3) form mating projections that are involved in the fusion process, and 4) increase expression of the genes involved in sexual conjugation (Dohlman *et al.*, 1991; Herskowitz, 1995; Marsh *et al.*, 1991).

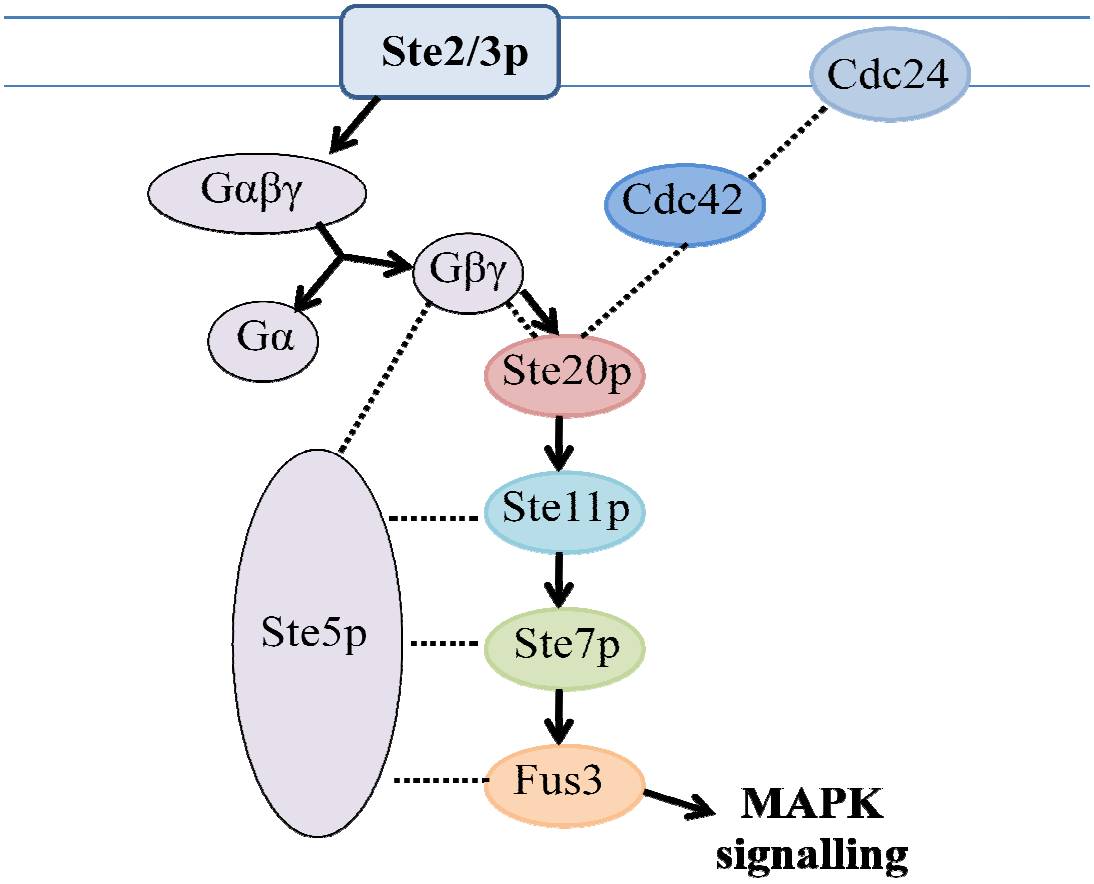

Figure 1.2: The Yeast Mating Pheromone Response Pathway.
The yeast mating pathway is initiated by activation of the receptors Ste2p or Ste3p in response to pheromones. This activates the G protein complex which dissociates into its subunits, Gα and Gβγ. The Gβγ subunit then goes on to interact with other protein complexes, such as the Ste5p complex, the Ste20p protein kinase, and the Far1/Cdc24 complex, which initiates the MAPK signalling cascade.

Physical fusion of MATa cells and MATα cells to each other is initiated by agglutination, which involves binding of glycoproteins on the surface of the opposite cells (Figure 1.3) (reviewed in (Chen *et al.*, 2007). In order for the plasma membranes of the cells to fuse, their cell walls must first be remodelled. The yeast cells assemble a unifying wall surrounding the junction and degrade the cell wall at the contact site, forming a conjugation bridge. After the cell wall is remodelled, osmotic gradients across the two plasma membranes allow for one cell to extend a finger of membrane bound cytoplasm into the other cell. Immediately following the membrane contact, the plasma membranes fuse together. Once the

plasma membranes of the cells have fused, the nuclei migrate into the bridge and subsequently fuse to produce a diploid nucleus. This process is termed karyogamy. All of these processes require involvement of cytoskeletal components, such as microtubules and actin structures, and the cell wall (Madden and Snyder, 1998). After the yeast cell fusion is completed, the final product is a peanut-shaped zygote with a diploid nucleus. The mating is completed by terminating the pheromone response, through down-regulation of Ste4p.

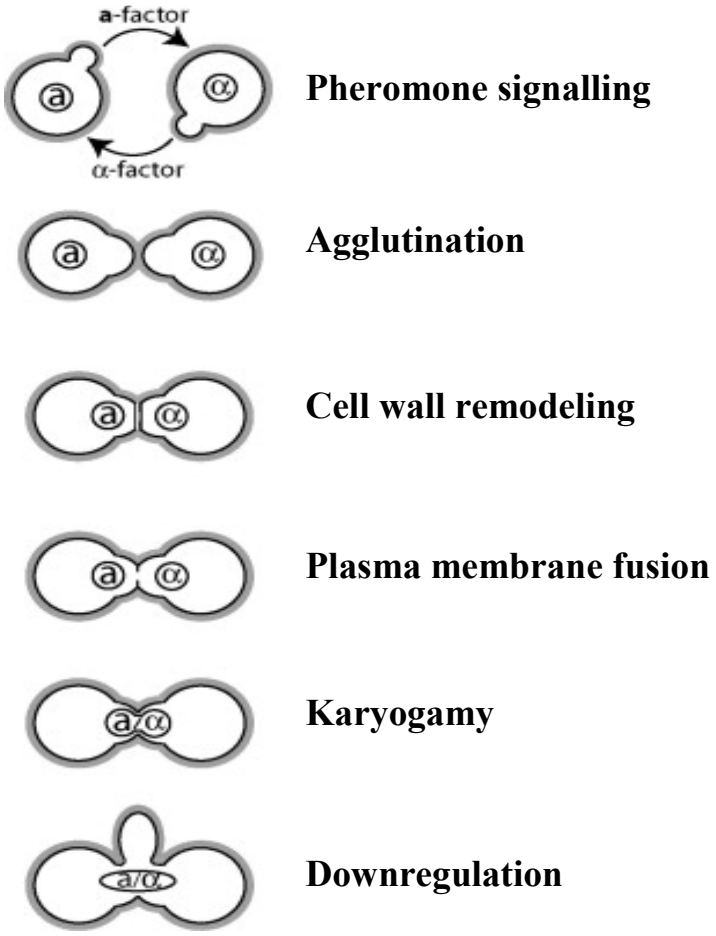

Figure 1.3: The Stages of Yeast Mating and Cellular Fusion.
After mating is initiated by pheromone signalling, the cells form projections towards each other which interact during the agglutination stage through cell surface glycoproteins. Next, the cell wall are remodelled which allows for the formation of a conjugation bridge and leads to plasma membrane fusion. During karyogamy the nuclei fuse to produce a diploid nucleus. After fusion is complete the pheromone response is terminated through down-regulation of Ste4p (Gβ subunit). This image has been modified from (Chen *et al.*, 2007).

1.3.2 Ste2p Role in Mating

As was mentioned previously, Ste2p is the mating pheromone GPCR for *MAT*a cells of yeast. When α-factor is secreted from *MAT*α cells, it interacts with Ste2p and initiates mating with the opposite yeast cells, as described above. However, recent studies suggest additional roles beyond modulation of signal transduction, leading to G1 arrest. In particular, proteoliposome fusion assays demonstrated that activated Ste2p and Ste3p promote lipid mixing and membrane juxtaposition through an interaction between Ste2p and Ste3p (Figure 1.4) *(Shi et al.*, 2007). The interaction between Ste2p and Ste3p was confirmed using tandem affinity purification, surface plasmon resonance and cellular pull-down assay experiments. Mutational analyses highlighted a role for the N-terminal region of Ste2p in mediating these effects. In particular two mutants, one with the first 20 amino acids removed (Ste2pΔN20) and a second with the first 30 residues removed (Ste2pΔ30N), were expressed in a yeast strain where the wild type Ste2p receptor was deleted. The Ste2pΔ30N receptor resulted in reduced signalling and reduced cell cycle arrest (Shi *et al.*, 2007). However, when the Ste2pΔN20 receptor was expressed, the result was a strain that exhibited wild-type signalling levels and phenotypic characteristics, but when mated with *MAT*α resulted in very low levels of fused zygotes (Figure 1.5). Electron microscopy experiments showed mating arrest at a prezygotic stage after cell contact. Tandem affinity purification and proteoliposome fusion assays confirmed that the interaction between Ste2p and Ste3p was eliminated as was the receptor's ability to promote membrane juxtaposition when the first 20 amino acids of Ste2p were deleted. Therefore, the prezygotic arrest that was observed when the Ste2pΔN20 mutant was expressed was possibly due to not only the mediation of cell wall erosion but the loss of interaction between Ste2p and Ste3p. Mutagenic screening of the N-terminus of Ste2p was completed and it was found that mutations at residues Pro15, Ile24 and Ile29 resulted in wild-type levels of cell cycle arrest and a significant loss in mating efficiency, similar to what was seen for the Ste2pΔN20 mutant (Figure 1.6) (Shi *et al.*, 2009a). Hence, these three residues in the N-terminus play an important role in mediating the mating efficiency of Ste2p, possibly through some form of interaction with either/both activated Ste3p and ligand and/or some other component.

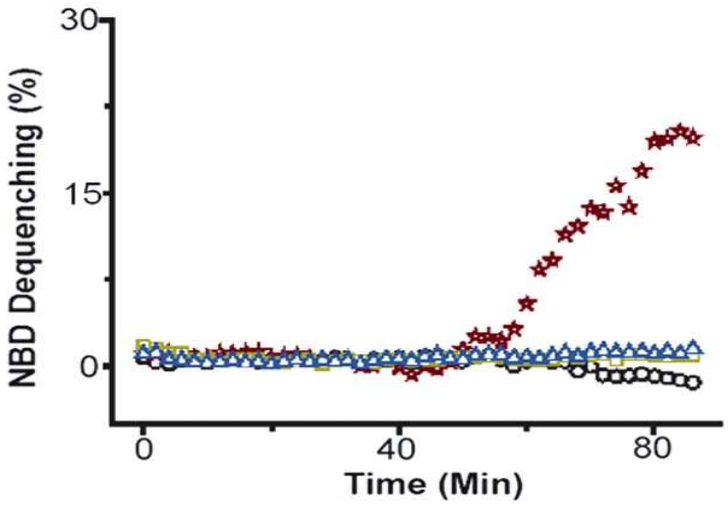

Figure 1.4: Proteoliposome Fusion Assay of Ste2p and Ste3p.
Fluorescently-labelled liposomes containing Ste2p-$(His)_8$ were mixed with non-labelled liposomes containing Ste3p-$(His)_8$ in the presence (red star) or absence (black hexagon) of a- and α-factor. To ensure that the lipid mixing was not due to the pheromones alone, fusion was carried out in the presence (blue triangle) or absence (green square) of a- and α-factor with both Ste2p-$(His)_8$ and Ste3p-$(His)_8$ absent. This image is taken from (Shi *et al.*, 2007).

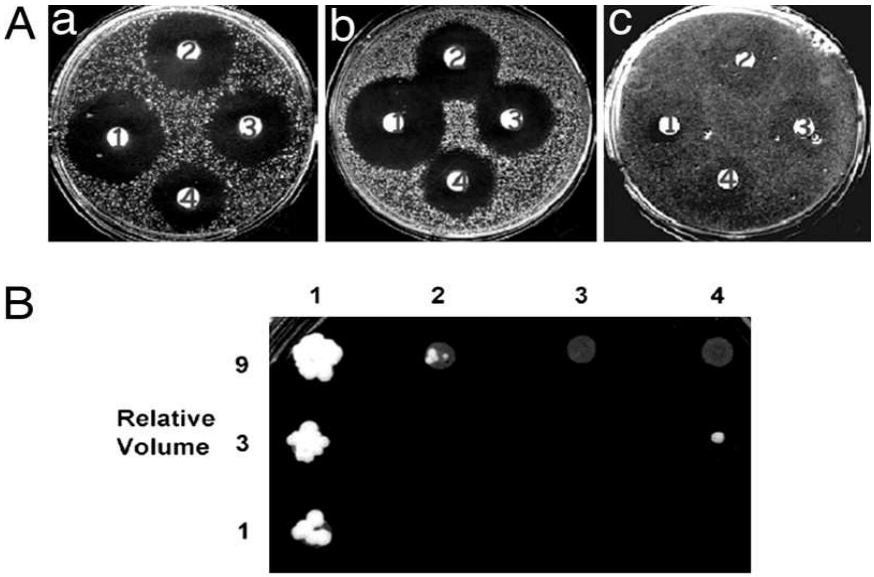

Figure 1.5: Signalling and Mating Efficiency of Ste2p Mutants.
A. Halo assay to determine α–factor stimulated signalling in a receptor null MATa yeast strain overexpressing (a) Ste2p-GFP, (b) Ste2pΔ20N-GFP, (c) Ste2pΔ30N-GFP. Decreasing concentrations of α–factor were spotted on positions 1-4.
B. MATα yeast was mated with receptor null MATa yeast strains overexpressing (lane 1) Ste2p-GFP, (lane 2) Ste2pΔ20N-GFP, (lane 3) Ste2pΔ30N-GFP, and (lane 4) Ste4p.
These images are taken from (Shi *et al.*, 2007).

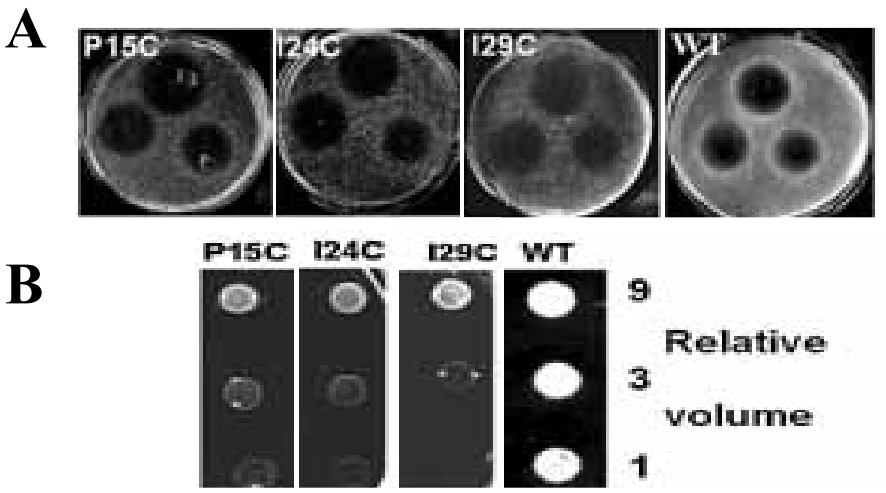

Figure 1.6: Signalling and Mating Efficiency of Ste2p Point Mutants.
A. Halo assay to determine α–factor stimulated signalling in a receptor null MATa yeast strain overexpressing Ste2p-P15C, Ste2p-I24C, Ste2p-I29C, and WT Ste2p. Decreasing concentrations of α–factor were spotted on the plate starting from the upper position and moving counter-clockwise.
B. MATα yeast was mated with receptor null MATa yeast strains overexpressing Ste2p-P15C, Ste2p-I24C, Ste2pI29C and WT Ste2p.
These images are modified from (Shi *et al.*, 2007).

1.3.3 Ste2p Oligomerization

A large number of studies have shown that GPCRs can interact to form dimers or higher-order oligomers (Milligan, 2004; Milligan, 2009; Terrillon and Bouvier, 2004). GPCRs can undergo homodimerization (He *et al.*, 2002; Salahpour *et al.*, 2004) or heterodimerization (Jordan and Devi, 1999; Marshall *et al.*, 1999; Nelson *et al.*, 2001). The ability of GPCRs to dimerize or oligomerize has been thought to play a role in receptor functions such as receptor biogenesis, plasma membrane targeting, signal transduction and down-regulation (Bulenger *et al.*, 2005; Wang and Dohlman, 2004). Ste2p has been shown to dimerize using fluorescent resonance energy transfer (FRET) or bioluminescence resonance energy transfer (BRET) (Gehret *et al.*, 2006; Overton *et al.*, 2005; Raicu *et al.*, 2005). Ste2p has also been shown to oligomerize *in vitro* using atomic force microscopy and DLS (Shi *et al.*, 2009b). It has been suggested that a GXXXG motif within TM1 of Ste2p may mediate the homo-oligomerization interaction (Overton *et al.*, 2003).

1.3.4 Ste2p Structural Studies

The 3D crystal structure of Ste2p to date is unknown. However, through the use of CD and NMR analysis, portions of the Ste2p receptor structure have been illuminated. In addition computer modelling has been used to produce a three-dimensional model of Ste2p (Eilers *et al.*, 2005). The Ste2p model was compared to a three-dimensional structure of rhodopsin and even though these receptors have no sequence homology, they displayed conserved microdomains which suggested similarities between the divergent receptors. This information has given us some understanding into the mechamism of action of Ste2p, although the N-terminus was not included in the model.

To date, very little research has been done on the structure of the N-terminus of Ste2p. The only structural information that is known was obtained from a recent NMR analysis of a recombinant peptide containing residues 31 to 110 of Ste2p. This peptide encompassed a portion of the N-terminus, transmembrane helix 1 (TM1), the first intracellular loop (IL1), transmembrane helix 2 (TM2) and a portion of EL1. NMR analysis on this peptide found that residues 39-47 of the N-terminus of Ste2p were helical (Neumoin *et al.*, 2009). These residues were predicted to interact with phospholipid groups, providing some indication about the possible function of the N-terminus of Ste2p. Some NMR has been performed on the N-terminal domains of other GPCRs (Ravindran *et al.*, 2009; Ulfers *et al.*, 2002; Zou *et al.*, 2008), however because there is no sequence homology, we cannot say anything about the Ste2p N-terminus from these.

Much more structural information has been obtained on the transmembrane domains of Ste2p. The transmembrane domains of GPCRs are predicted to form α-helical structures because it is energetically favourable for a peptide to have this conformation in the non-aqueous interior of the lipid bilayer (Engelman *et al.*, 1986). Biophysical experiments, such as CD, demonstrated that the synthetic transmembrane fragments of all putative 7 TM domains from Ste2p tended to form α-helical structures when in 100% TFE (Xie *et al.*, 2000). However, biophysical analysis of the 7 TM domains in sodium dodecyl sulphate (SDS) or dimyristol phosphatidylcholine (DMPC) yielded evidence of significant structural diversity. Most of the transmembrane domains were helical, but TM3 and transmembrane helix 6 (TM6) could form β-structures. Specifically, TM6 was highly helical when in TFE and formed a stable β-sheet when in DMPC or SDS, demonstrating high conformational flexibility. NMR analysis of a

synthetic TM6 peptide in TFE/H_2O (4:1) showed that the peptide was α-helical but that the Pro258 residue disturbs the structure (Naider *et al.*, 2001). Mutagenesis of Pro258 to Leu results in a constitutively active receptor with 45% of wild type signalling (Konopka *et al.*, 1996; Sommers *et al.*, 2000). Therefore, the conformational changes that TM6 undergoes may be significant for receptor activation as well as receptor oligomerization. This could be the case for transmembrane helix 7 (TM7) as well, which was found to contain an α-helical structure which is disrupted by Pro290 which produces a large kink in the helix (Estephan *et al.*, 2005; Neumoin *et al.*, 2007). However, NMR analysis on TM1 and TM2 showed that they contained very stable helical structures as compared to the flexible TM6 and TM7 (Neumoin *et al.*, 2009). Therefore, structural analysis of the transmembrane domains has demonstrated that they are not all rigid α-helical structures and instead can display some conformational flexibility which may play a role in receptor activation.

As with the N-terminus of Ste2p, the C-terminal end has had very little structural analysis performed on it. Using NMR analysis of a recombinant Ste2p peptide encompassing a portion of the third extracellular loop (EL3), TM7 and 40 residues of the C-terminus, the C-terminus of Ste2p was found to be α-helical in both DPC micelles and TFE/water. However, in the DPC micelles, the C-terminus was unstructured after Gln328 while in TFE/water the helix extended up until Ala336 (Estephan *et al.*, 2005; Neumoin *et al.*, 2007). However, this may be due to the fact that TFE is a known helix inducer.

1.4 Techniques

1.4.1 Circular Dichroism

Dichroism is the property of a material to exhibit some absorption when light in different polarization states is travelling through it. Usually light is linearly polarized, meaning that the electric field component of electromagnetic radiation is oscillating in a single direction, usually in a transverse wave (Figure 1.7A). However, when light is circularly polarized, the electric field component rotates while maintaining constant amplitude, essentially producing a helix through space (Figure 1.7B). When the electric field component is rotating counter-clockwise, this is called left circularly polarized light, and when it is rotating clockwise, it is

called right circularly polarized light. When a chiral molecule is subjected to circularly polarized light, the absorption of light circularly polarized in the right-handed direction is different from the absorption of light circularly polarized in the left-handed direction, and therefore the material is said to demonstrate circular dichroism (CD) (Figure 1.7C) (Abu-Shumays and Duffield, 1966; Hennessey and Johnson, 1981; Oakley *et al.*, 2006). The difference in absorbance (ΔA) between the right and left circularly polarized light can be plotted versus the wavelength of light to produce a CD spectrum.

Traditionally, CD has been reported in terms of its ellipticity (θ). The ellipticity of a material refers to the fact that when the circularly polarized light passes through the material and a differential change in absorbance is observed, the magnitudes of the circularly polarized light are not the same anymore and therefore the polarized light is no longer circular but more elliptical. While the term ellipticity is dated, it is still used out of tradition. It is very easy to calculate the ellipticity from the change in absorbance (ΔA).

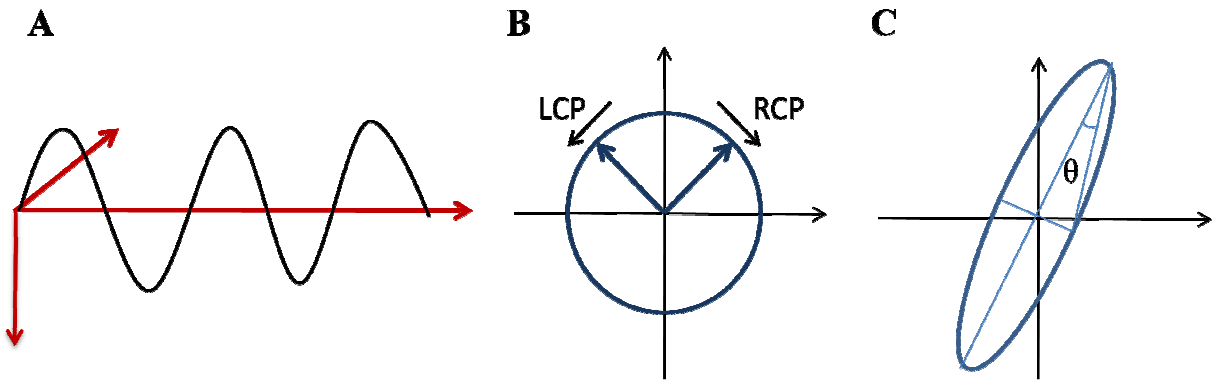

Figure 1.7: Circular Dichroism Spectroscopy.
A. Linearly polarized light travels in a single direction like a transverse wave.
B. Circularly polarized light is made up of two components, right circularly polarized (RCP) and left circularly polarized (LCP), of which the electric field components rotate through space at the same amplitude.
C. When circularly polarized light travels through a chiral molecule the right and left circularly polarized light have different absorbances which causes the overall polarization to become elliptical. This ellipticity can be measured using circular dichroism spectroscopy.

$$\theta = \Delta A \cdot 32.982 \quad (1.1)$$

Where θ is the ellipticity and ΔA is the change in absorbance between right and left circularly polarized light. The value 32.982 accounts for conversion from electric field vector quantities to the angle of ellipticity.

Ellipticity is usually expressed in millidegrees (mDeg). Typically, a CD spectrum consists of a plot of the Molar ellipticity [θ] versus the wavelength of light. The [θ] can be calculated from the ellipticity and takes into account the molar concentration of the sample as well as the pathlength of the cell. Therefore using Beer Lamberts law, the [θ] can be calculated by:

$$[\theta] = 100 \cdot \theta / C \cdot l \quad (1.2)$$

Where C is the concentration of the sample, l is the pathlength of the cell and the factor of 100 is used for converting the units.

Therefore, the final result is [θ] with units of $deg \cdot cm^2 \cdot dmol^{-1}$. For proteins and peptides it is more convenient to express CD in terms of mean residue ellipticity $[\theta]_{MR}$. $[\theta]_{MR}$ reports the molar ellipticity for individual protein residues instead of the whole protein molecule. In order to calculate $[\theta]_{MR}$ the number of residues needs to be known.

$$[\theta]_{MR} = [\theta]/n \quad (1.3)$$

Where $[\theta]_{MR}$ is the mean residue ellipticity, [θ] is the Molar ellipticity and n is the number of amino acid residues.

By displaying the CD spectrum in terms of $[\theta]_{MR}$, this allows for the CD of proteins or peptides with very different molecular weights to be compared.

CD is a very useful tool for structural analysis of proteins because information about the secondary structure can be obtained from the CD spectrum. Secondary structure can be determined by CD spectroscopy in the far UV range (180-250 nm). At the far UV wavelengths,

the chromophore that is observed is the peptide bond (Kelly *et al.*, 2005). Certain properties are indicative of specific secondary structures (Brahms and Brahms, 1980). For example, a CD spectrum with a "w" shape and troughs (also referred to as bands or lines) at around 208 and 222 nm are indicative of α-helical structure, while a CD spectrum with a "v" shape and a trough around 217-220 nm indicates β-sheet secondary structure. A protein that contains random coil will have a characteristic spectrum with a positive band at 212 nm and a trough at 195 nm. The approximate amount of each secondary structure can be determined by analysing the CD spectrum using computational analysis (Bohm *et al.*, 1992). CD spectroscopy can give information about the amount of different secondary structures in a protein but not the specific residues that participate in that portion of secondary structure. This is because the CD spectrum reflects the secondary structure of the entire protein or peptide.

CD spectroscopy is important in protein structural studies for more reasons than just secondary structure determination. CD can be used to compare the structure of the protein when it is in different solvents. This is helpful in determining a suitable solvent system for future structural experiments. In addition, CD can be useful in determining the stability of the protein under stress, such as a change in temperature, pH, or the addition of buffers, salts or denaturants (Kelly and Price, 1997; Sajedi *et al.*, 2004; Sajedi *et al.*, 2007). CD spectroscopy can also be a valuable tool for analysing the conformational changes a protein undergoes during protein-protein or protein-ligand interactions (Boxer *et al.*, 2004). Overall, CD spectroscopy is a very useful technique for analysing the structural characteristics of proteins.

1.4.2 Dynamic Light Scattering

Dynamic light scattering (DLS) is a useful technique in protein structural characterization because it gives information about the size of biomolecules in solution (reviewed in (Li *et al.*, 2004b; Murphy and Tsai, 2006; Nobbmann *et al.*, 2007)). It can therefore determine if a protein is aggregated in solution. The principle of DLS is based on the fact that when light is passed through a solution of biomolecules, part of the light will be scattered. The fluctuations in the scattered light intensity due to Brownian motion can be measured using DLS. An important characteristic of the particles in DLS is the diffusion coefficient (D). The diffusion coefficient of the particles is inversely proportional to the decay time of the light scattering fluctuations. The decay time can be obtained from the time-

dependant correlation function of the scattered light. Using the Stokes-Einstein equation, the hydrodynamic radius of the particles in the solution can be obtained from the diffusion coefficient.

$$R_h = k_B T / 6\pi\eta D \quad (1.4)$$

Where R_h is the hydrodynamic radius with the units nm, k_B is the Boltzmann constant, T is the temperature, η is the solvent viscosity and D is the diffusion coefficient.

An assumption made when using the Stokes-Einstein equation is that the particles are spherical and non-interacting. In reality, however, the biomolecules are not spherical, tumbling and solvated. Therefore, this calculation results in the apparent radius of the dynamic hydrated/solvated biomolecule, also known at the hydrodynamic radius. Since the calculation is an approximation of the radius, DLS is not useful as a measurement of the exact molecular mass. Instead it is used to distinguish molecules that vary in size by a large amount, such as native proteins versus large aggregates. It is difficult to separate small oligomers such as monomers, dimers and trimers from each other and they will often be lumped into the same peak.

1.4.3 Nuclear Magnetic Resonance

NMR spectroscopy is a very important tool for gathering structural information about proteins (reviewed in (Cavanagh *et al.*, 1996; Roberts, 1993; Silverstein *et al.*, 2005; Wüthrich, 1986)). NMR spectroscopy is based on the magnetic properties of nuclei. All nuclei have a charge and on some nuclei the charge spins on the nuclear axis. This produces a quantum mechanical property called angular momentum. The angular momentum is determined using the quantum spin number (I) of the nuclei. The quantum spin number is based on the atomic mass and atomic number of a nucleus (i.e. relative number of protons and neutrons). For example, ^{1}H has an odd atomic number (1 proton) so therefore I = ½, while ^{2}H has an even atomic mass (1 proton + 1 neutron = 2) and an odd atomic number (1 proton) resulting in I=1. Nuclei with a non-zero spin possess a nuclear magnetic moment (μ), which is proportional to the angular momentum using the gyromagnetic ratio (γ). The gyromagnetic ratio is a constant

for each nucleus. The most widely used nuclei in protein NMR spectroscopy are ^{1}H, ^{13}C, and ^{15}N because they have $I = ½$ and ^{2}H, which has $I = 1$. When the nuclei are placed in a magnetic field, the spin ½ nuclei can occupy two energy levels known as the Zeeman levels (Figure 1.8). The difference in population between the two energy levels is given by the Boltzmann distribution, which determines that there is an excess of population in the lower energy state. With the application of an external radiofrequency pulse in a magnetic field, radiation is introduced and energy is absorbed by the nuclei, raising it to the higher energy state. When the nucleus is in the external magnetic field, the nuclear spin will precess around the magnetic field (Figure 1.9). The precession frequency is called the Larmor frequency. As the nuclei transition between the two energy states, the spin orientation is reversed. However, due to the population excess in the lower level, as the nuclei transition to the higher level there is absorption of energy, which produces a signal. As the system returns to thermodynamic equilibrium, relaxation occurs in the form of spin-lattice (longitudinal) relaxation and spin-spin (transverse) relaxation. Spin-lattice relaxation results from the transfer of energy from the excited protons to the surrounding lattice, or nuclei that are vibrating and rotating. Spin-spin relaxation results from interactions between neighbouring nuclei which results in the lifetime of nuclei in the excited state to decrease. The decay of transverse magnetization, or spin-spin relaxation, is detected in the receiver and recorded as the corresponding free induction decay (FID) (Figure 1.10). The FID is Fourier transformed which produces the NMR spectrum.

Information about the protein structure can be interpreted from NMR spectra. The chemical shift (δ) of the nuclei is the variation of the frequency of the nucleus dependent on its environment, which is relative to a reference compound. Each chemically non-equivalent nucleus will have a different chemical shift based on the surrounding environment. For proton NMR spectra, each resonance line represents a single proton (or group of protons) in the protein. The chemical shifts of the nuclei can give information about the environment of that nucleus. Therefore, the chemical shifts of specific nuclei can be assigned from the spectra. For a larger molecule such as a protein, assignments of individual nuclei can become extremely difficult due to the number of nuclei in the protein. It is easier to use two-dimensional (2D) NMR experiments for protein structure assignments. 2D experiments allow for calculation of the distances (through space and through bond) between nuclei and therefore enable overall sequential and conformational assignments for the protein.

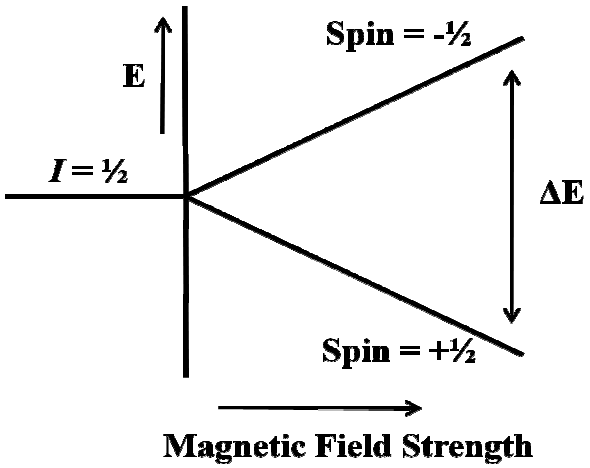

Figure 1.8: Zeeman Levels for a Spin ½ Nucleus in a Magnetic Field.
When a nucleus has a spin number (I) of ½, it can occupy two energy levels, -½ or +½ in a magnetic field. When the nucleus is subjected to radiofrequency radiation, it can transition between the two levels. Therefore, if it was in the lower energy level it will transition to the higher energy level and vice versa. The difference in population between the two levels is given by the Boltzmann distribution.

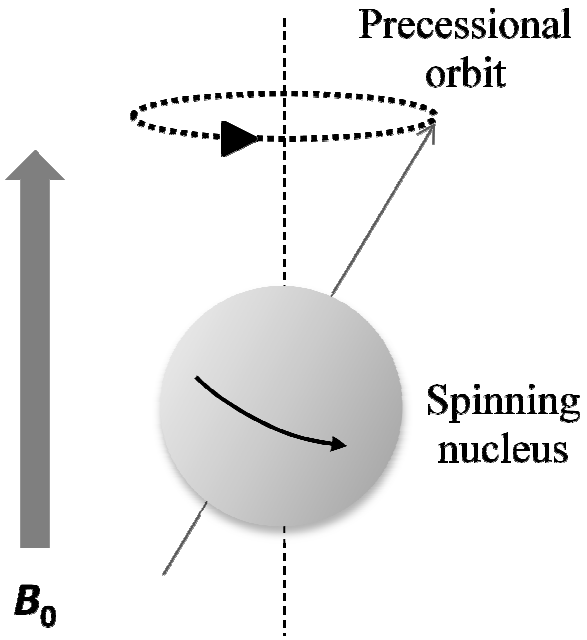

Figure 1.9: Precessing Nucleus in a Magnetic Field.
In the external magnetic field of magnitude, B_0, the nucleus precesses around the magnetic field.

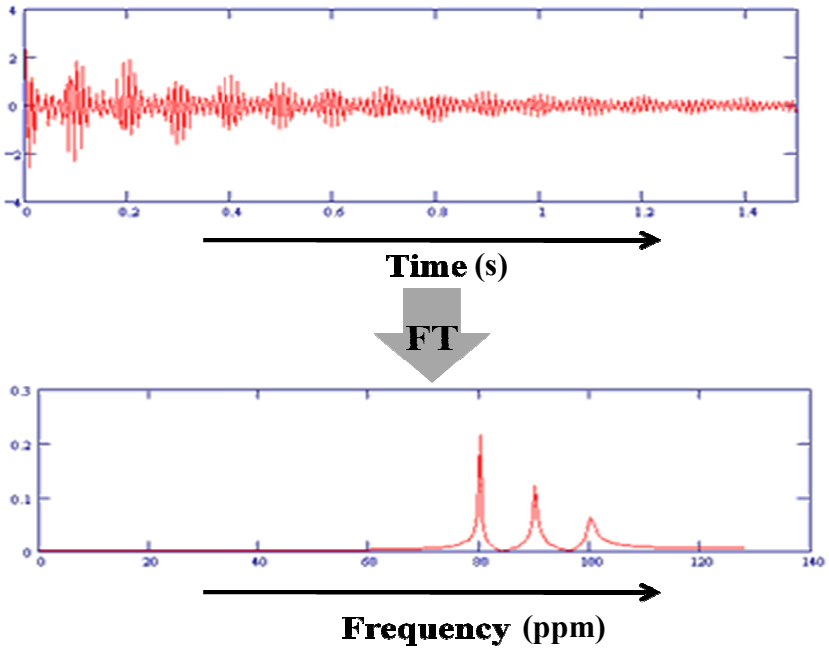

Figure 1.10: Free Induction Decay and the Corresponding Fourier Transformed Spectrum.
Initially the transverse magnetization decay is recorded as the free induction decay (FID) over a period of time (top). After Fourier Transformation of the FID the spectrum is produced (bottom). This figure is modified from (Iannone, 1999).

For protein NMR, the majority of structural information comes from multidimensional homonuclear or heteronuclear experiments. In a one-dimensional (1D) NMR experiment, the FID is recorded as a function of one time variable and Fourier-transformed to produce a spectrum that is a function of one frequency variable, while in a 2D NMR experiment, the signal is recorded as a function of two time variables, t_1 and t_2. The t_1 time variable is an evolution time that increases in increments throughout the experiment, and the t_2 variable is when the signal is recorded. The total sequence of events is called a pulse sequence (Figure 1.11). Thus, a pulse sequence is repeated with increasing t_1 increments and the FID is recorded as a function of t_2 for each value of t_1. The FID's are Fourier-transformed in both dimensions to produce a spectrum which contains two frequency variables, f_1 and f_2. The 2D spectra are represented by a topographical map, where one of the axes is f_1 and the other is f_2. Some common 2D-NMR experiments that are used in protein NMR are correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY), which measure magnetization transfer

through bond between the same type of nuclei (homonuclear). Heteronuclear single quantum coherence (HSQC) or heteronuclear multiple bond coherence (HMBC) measure the magnetization transfer through bond between different types of nuclei (heteronuclear). As well, nuclear Overhauser effect spectroscopy (NOESY) and rotating frame Overhauser effect spectroscopy (ROESY) NMR experiments measure magnetization transfer through space. For example, a very common 2D experiment that is used is COSY. 2D COSY experiments allow for observation of the amide protons ($^1H^N$) and their interactions with alpha carbon protons ($^1H^\alpha$). The $^1H^N$-$^1H^\alpha$ crosspeaks provide the fingerprint of the amino acid sequence of the peptide. There is a known number of expected crosspeaks in the $^1H^N$-$^1H^\alpha$ region which can be determined using the following equation:

$$N - P + G - 1 = \text{crosspeaks in } {}^1H^N\text{-}{}^1H^\alpha \text{ region} \quad (1.5)$$

Where N is the total number of residues, P is the number of proline residues and G is the number of glycine residues in the peptide.

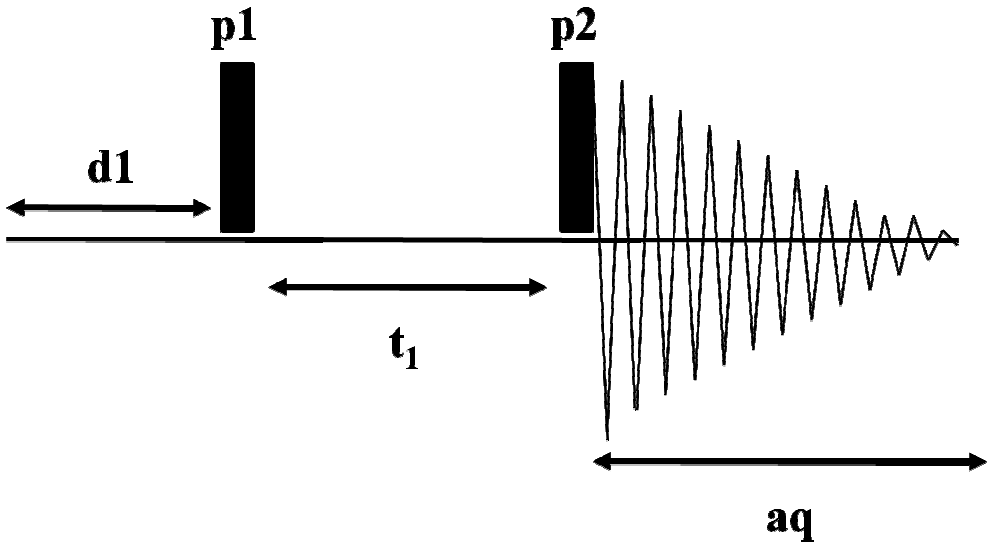

Figure 1.11: 2D COSY NMR Pulse Sequence.
Pulse sequence for a standard 2D COSY NMR experiment. The relaxation delay (d1) is followed by a pulse (p1) which is a variable radiofrequency pulse dependent on the experiment. Following the pulse is a time delay, t_1, which increases in increments over the course of the experiment. Following another pulse, p2, the t_2 magnetization signals are detected by the receiver during the acquisition time (aq).

In order to perform heteronuclear NMR experiments on protein samples, the protein must first be isotopically labelled. The reason behind this is that the most abundant naturally-occurring isotopes for carbon and nitrogen, ^{12}C and ^{14}N respectively, have nuclei properties that make them unfavorable for NMR. In the case of ^{12}C nuclei, they have an even atomic mass and an even atomic number which results in $I = 0$. Therefore, they have no spin and are NMR inactive. However, ^{14}N nuclei have an even atomic mass number and an odd atomic number which results in $I = 1$, which causes unfavorable relaxation properties during NMR experiments. In contrast the, relatively rare isotopes ^{13}C and ^{15}N can be used in heteronuclear NMR experiments. However, in order to detect these isotopes, isotopic labelling of a protein must be carried out to enrich for their presence allowing detection at relatively low protein concentrations. This can be achieved by producing recombinant proteins in media enriched with the required isotope, or an isotopically labelled protein can be chemically synthesized. Uniform isotopic labelling is an important tool because it enables/facilitates assignment procedures.

A disadvantage of protein NMR is that there are size constraints. Larger proteins have slower tumbling rates and shorter NMR signal relaxation times, which leads to reduced sensitivity and broad linewidths. The other problem is the increase in complexity of the NMR spectra due to more NMR-active nuclei in larger proteins and therefore more interactions between the nuclei. This results in more signals that overlap in the limited range of frequencies. In order to combat the problem of large proteins, a number of NMR techniques are available. The first is to label proteins with deuterium, which will take the place of exchangeable protons, allowing for them to not be observed in the proton spectrum. The result is simplified spectra, increased relaxation time and increased spectral resolution (Gardner and Kay, 1998; Goto and Kay, 2000). The second technique that has been useful in NMR spectroscopy of large proteins is transverse relaxation-optimized spectroscopy (TROSY) (Pervushin *et al.*, 1997). TROSY NMR experiments suppress the faster transverse relaxation signals that result from the slow tumbling of high molecular weight proteins and selectively observe the most slowly relaxing component, which results in sharper lines for amide and aromatic groups. Another disadvantage of NMR is the requirement of a large amount of pure protein, which is especially difficult for membrane proteins. The reason for this is that the

signal is averaged a number of times in an experiment and the signal-to-noise ratio is proportional to the concentration of the sample and the square root of the number of scans. Therefore, a sample with lower protein concentration will require a larger number of scans which may result in an impractical length of time for obtaining a satisfactory spectrum. The goal is to provide a protein sample with a high concentration that is still soluble. This problem has been overcome by the use of fusion proteins for recombinant expression in bacteria, as well as the use of detergents in NMR experiments for solubilising membrane proteins (Cook *et al.*, 2010; Gautier *et al.*, 2010; Sanders and Landis, 1995).

However, there are many advantages to using NMR spectroscopy for protein structure determination. One advantage is that because the protein sample is in solution, the difficult issue of trying to crystallize the protein is avoided. Another advantage of NMR is that dynamic information about the protein can also be obtained from solution NMR spectroscopy (Mittermaier and Kay, 2006; Wand, 2001). This can provide information about protein movements on a broad range of timescales. In addition, NMR can be a useful tool for characterizing protein-ligand interactions (Bokoch *et al.*, 2010; Stark and Powers, 2007). Using a technique called chemical shift perturbation, information about the ligand binding interface and conformation changes can be obtained.

1.5 Objectives

Binding of α-factor pheromone to the G-protein coupled receptor, Ste2p, initiates signal transduction events that lead to mating of the yeast *Saccharomyces cerevisiae*. It has been shown that Ste2p and Ste3p interact with each other to promote cell wall degradation and possible membrane juxtaposition (Shi *et al.*, 2007). Experiments performed with a ΔSte2p yeast strain and an N-terminally truncated Ste2p over-expressed resulted in arrest prior to cell wall degradation, a loss of interaction between the Ste2p and Ste3p, and a loss of the receptor's ability to promote membrane juxtaposition *in vitro*. Specifically, mutagenesis studies have shown that residues Pro15, Ile24 and Ile29 from the N-terminus of Ste2p are important for mediating the mating efficiency (Shi *et al.*, 2009a). These results suggest that the N-terminus of Ste2p is involved in cell wall degradation and possibly membrane juxtaposition during yeast mating, potentially by mediating the interaction with Ste3p. However, the mechanism by

which the N-terminus is modulating these effects remains unknown. Therefore, the overall goal of the work presented in this thesis was to perform structural analyses to gain insight into the mechanism of action of the N-terminal region of Ste2p.

The specific objectives of this thesis were:

1. To obtain, by recombinant or synthetic methods, purified mg quantities of a soluble, purified version of a portion of the N-terminus of Ste2p.
2. To obtain structural information about the N-terminal region of Ste2p by performing biophysical analyses, such as CD, DLS and NMR, on the obtained N-terminal Ste2p fragments.

2.0 MATERIALS AND METHODS

2.1 Reagents

The names of the reagents and their suppliers are listed in Table 2.1. The addresses of the suppliers are provided in Table 2.2.

Table 2.1: List of Reagents and Suppliers.

Reagent:	Supplier:
Acetic acid - glacial	Fisher Scientific
Acetonitrile	Fisher Scientific
Ammonium molybdate	Sigma Aldrich
Ammonium persulfate	Sigma Aldrich
Ammonium sulfate	EMD
Ammonium $^{15}N_2$ sulphate (^{15}N, 98%+)	Cambridge Isotopes
Ampicillin	Sigma Aldrich
Bacto-Agar	BD
Bacto-Tryptone	BD
Bacto-Yeast Extract	BD
Boric acid	EMD
Bromophenol blue	Sigma Aldrich
Calcium chloride	Sigma Aldrich
Calcium sulfate	VWR
Chloroform	Fisher Scientific
Cobalt chloride	Sigma Aldrich
Coomassie Brilliant Blue G 250	Sigma Aldrich
Cyanogen bromide	Sigma Aldrich
Deuterium oxide (D, 99.9%)	Cambridge Isotopes
Dimethyl sulfoxide	Fisher Scientific
Formic Acid (88%)	Fisher Scientific
Glucose	Sigma Aldrich
Glycerol	Sigma Aldrich
Glycine	Fisher Scientific

Guanidine hydrochloride	Sigma Aldrich
Imidazole	Sigma Aldrich
Iron(II) sulfate	Fisher Scientific
Isopropyl β-D-thiogalactopyranoside (IPTG)	BioShop
Magnesium sulfate	Fisher Scientific
Manganese chloride	Sigma Aldrich
β-mercaptoethanol	Sigma Aldrich
Methanol	Fisher Scientific
Nickel-NTA resin	Qiagen
n-octyl-β-D-glucopyranoside (OG)	Anatrace
One Shot TOP10F' Competent Cells	Invitrogen
1-palmitoyl-2-hydroxy-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (LPPG)	Avanti Polar Lipids
pET31b+	Novagen
1,4-Piperazinediethanesulfonic acid	Sigma Aldrich
Potassium chloride	Fisher Scientific
Potassium phosphate monobasic	EMD
Primers	IDT
Qiagen mini-prep kit	Qiagen
QuikChange Site-Directed Mutagenesis Kit	Stratagene
SeeBlue® Plus2 Pre-stained Standard	Invitrogen
Sodium phosphate dibasic	EMD
Sodium phosphate monobasic	EMD
2,2,2-trifluoroethanol (TFE)	Sigma Aldrich
2,2,2-trifluoroethanol-d2 (TFE-d2) (D, 98%)	Cambridge Isotopes
N,N,N',N'-tetramethylethylenediamine	Sigma Aldrich
Trifluoracetic acid (TFA)	Sigma Aldrich
Zinc chloride	Sigma Aldrich

Table 2.2: List of Names and Addresses of Suppliers.

Supplier:	Address:
Anatrace	Maumee, Ohio, USA
Avanti Polar Lipids	Alabaster, Alabama, USA
BD	Mississauga, Ontario, Canada
BioShop	Burlington, Ontario, Canada
Cambridge Isotopes	Andover, Massachusetts, USA
EMD	Gibbstown, New Jersey, USA
Fisher Scientific	Ottawa, Ontario, Canada
IDT	Coralville, Iowa, USA
Invitrogen	Carlsbad, California, USA
Novagen	Gibbstown, New Jersey, USA
Qiagen	Mississauga, Ontario, Canada
Sigma-Aldrich	Oakville, Ontario, Canada
Stratagene	La Jolla, California, USA
VWR	Mississauga, Ontario, Canada

2.2 Synthetic KKK-Ste2p(14-43)-KKK Peptide

Initially, residues 14-43 of Ste2p were chemically synthesized by CanPeptide, Inc. (Pointe-Claire, Quebec). However, CanPeptide Inc. was unable to purify this product by HPLC. Therefore, three lysine residues were added respectively to the N and C-termini of the peptide to increase solubility as was successfully carried out previous (Xie *et al.*, 2000). This produced the synthetic 36 residue KKK-Ste2p(14-43)-KKK peptide. This peptide contained an N-terminal acetylation and a C-terminal amidation. Canpeptide Inc. was able to synthesize and purify this peptide by HPLC. The final peptide sequence was KKKDPTYNPGQSTINYTSIYGNGSTITFDELQGKKK.

2.3 Bacterial Strains and Media Preparations

E. coli BL21-CodonPlus (DE3)-RIL strain (Stratagene, La Jolla, CA) was used for plasmid expression and induction of fusion proteins. One Shot® TOP10F´ competent cells were used as the host for propagation of mutagenized plasmid DNA.

2x Yeast Extract and Tryptone (2xYT) plates containing ampicillin were used to propagate both *E. coli* strains. 2xYT plates contain 16 g/L Bacto-Tryptone, 10 g/L Bacto-Yeast Extract, 5 g/L NaCl and 15 g/L Bacto-Agar in 1 L of distilled water (dH_2O). The media was sterilized by autoclaving for 20 mins at 15 psi. Once the media was cooled to ~50°C, a final concentration of 100 μg/mL of sterile ampicillin was added. The 2xYT media was poured into Petri dishes, left to solidify and stored at 4°C.

Liquid 2xYT media containing ampicillin was used for small and large scale expression of plasmid DNA encoding fusion proteins transformed into *E. coli*. 2xYT media consisted of 16 g/L Bacto-Tryptone, 10 g/L Bacto-Yeast Extract, and 5 g/L NaCl in 1 L of dH_2O. The media was sterilized by autoclaving for 20 mins at 15 psi. Once the media cooled, a final concentration of 100 μg/mL of sterile ampicillin was added. The 2xYT media was kept at 4°C for long term storage.

M9 minimal media containing ampicillin was used for larger scale expression of plasmid DNA encoding fusion proteins transformed into *E. coli* to produce ^{15}N-labelled peptide. M9 minimal media consisted of 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4 and 0.5 g/L NaCl in 1 L of dH_2O. After mixing of the media, it was sterilized by autoclaving for 20 mins at 15 psi. After the media cooled, sterile 1 μM $FeSO_4$, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 0.2% (w/v) glucose, 0.1% (w/v) $(NH_4)_2SO_4$ or ^{15}N-$(NH_4)_2SO_4$, 3 nM ammonium molybdate, 0.4 μM H_3BO_3, 30 nM $CoCl_2$, 10 nM $CaSO_4$, 80 nM $MnCl_2$, and 10 nM $ZnCl_2$ were added. In addition, a final concentration of 100 μg/mL of sterile ampicillin was added and the M9 minimal medium was stored at 4°C.

Super Optimal Broth (SOB) media was used for production of BL21-CodonPlus (DE3)-RIL competent cells. SOB media contained 20 g/L Bacto-Tryptone, 5 g/L Bacto-Yeast Extract, 0.5 g/L NaCl, and 10 mM KCl in 1 L of dH_2O. The SOB media was autoclaved for 20 mins at 15 psi to sterilize and allowed to cool. A final concentration of sterile 10 mM $MgCl_2$ and 10 mM $MgSO_4$ was added to the SOB media. The SOB media was kept at 4°C for long term storage.

2.4 Molecular Cloning

2.4.1 Preparation of Competent BL21-CodonPlus (DE3)-RIL *E. coli* Cells

Competent cells were prepared according to Inoue *et al*. Essentially, five mL of 2xYT medium were inoculated with a single colony of BL21-CodonPlus (DE3)-RIL *E. coli*. The cultures were incubated overnight at 37°C with shaking. The overnight cultures were used to inoculate 100 mL of complete SOB medium, 1:50. These cultures were incubated at 37°C with shaking until the $OD_{600} \approx 0.4\text{-}0.5$. The cultures were transferred to 50 mL Falcon tubes and kept on ice for 10 min. They were then centrifuged at 2500x g for 10 min at 4°C. The supernatant was discarded and the pellet was gently re-suspended in 32 mL of ice cold transformation buffer and kept on ice for 10 minutes. The transformation buffer contained 10 mM 1,4-piperazinediethanesulfonic acid, 15 mM $CaCl_2 \cdot 2H_2O$ and 250 mM KCl. It was adjusted to pH 6.7 using KOH pellets. After adjusting the pH, 55 mM $MnCl_2 \cdot 4H_2O$ was added and the buffer was filter sterilized and stored at 4°C. The re-suspended cells were centrifuged again at 2500x g for 10 min at 4°C. The supernatant was discarded and the cells were gently re-suspended in 8 mL ice cold transformation buffer. Dimethyl sulfoxide was added drop by drop, while continually swirling, to a final concentration of 7%. The competent cells were kept on ice for 10 min and aliquoted into sterile 1.5 mL microcentrifuge tubes and stored at -80°C.

2.4.2 Transformation of Competent Bacterial Cells

E. coli BL21-CodonPlus (DE3)-RIL (100 µL) or One Shot TOP10F' (50 µL) competent cells were thawed on ice. To the competent cells, 10 pg of the required plasmid or 5 µL of mutagenesis reactions were added and allowed to incubate on ice for 30 min. The transformation reaction was then heat-shocked at 42°C for 30 sec and then incubated on ice for another 10 min. To allow for recovery of the cells, 100 µL of 2xYT medium was added to the transformation reaction. The reaction was then incubated for one hour with shaking at 37°C. The transformation reaction was spread onto 2xYT plates with ampicillin and incubated overnight at 37°C.

2.4.3 Molecular Cloning of Ste2p N-terminal Fragments into pET31b+

Ste2p N-terminal fusion proteins were all expressed from the pET31b+ vector. The pET31b+ vector is useful for high-level expression of peptides. The vector contains an ampicillin resistance gene for selection and places the target gene under the control of a T7 promoter. The vector also encodes the ketosteroid isomerase (KSI) gene at the 5' end of the multiple cloning region and six histidine amino acids (HIS tag) at the 3' end. The pET31b+/KSI-Ste2p(1-71)-HIS and pET31b+/KSI-Ste2p(1-48)-M-HIS constructs were made previously in our lab by E. Lukiwski and M. Loewen, respectively. Briefly, these constructs were made by amplifying residues coding for 1-71 or 1-48 of Ste2p from the pTTSH8Q1/Ste2p construct using polymerase chain reaction (PCR). The PCR amplification introduced a 5' *AlwN*I restriction site and a 3' *Xho*I restriction site. The *AlwN*I-Ste2p(1-71)-*Xho*I PCR product was purified using agarose gel electrophoresis. Both the PCR product and the pET31b+ vector were digested using *AlwN*I and *Xho*I restriction enzymes. Calf intestinal phosphatase was applied to the vector and both digested vector and fragment were purified by agarose gel electrophoresis. They were then ligated using DNA ligase to produce a gene encoding a fusion protein that contained the KSI peptide at the N-terminus, residues 1-71 or 1-48-M of Ste2p and a C-terminal HIS tag. These resulting fusion protein products were designated KSI-Ste2p(1-71)-HIS and KSI-Ste2p(1-48)-M-HIS fusion protein. The final gene constructs were confirmed by DNA sequencing.

2.4.4 Site Directed Mutagenesis

Mutagenesis primers were designed to introduce three lysine residues between residue 1 (methionine) and residue 2 (serine) of the Ste2p N-terminal fragment, included in the pET31b+/KSI-Ste2p(1-48)-M-HIS construct. Site-directed mutagenesis was carried out using the forward primer 5'-CCAGATGCTGATG<u>AAGAAGAAG</u>TCTGATGCGGCTC-3' as well as the complementary reverse primer. All primers were designed according to the QuickChange Site-Directed Mutagenesis Kit guidelines. The melting temperature (T_m) was calculated using the equation in the guidelines and found to be 79.3°C. Another set of mutagenesis primers were designed to introduce three lysine residues between residue 48 (threonine) and the following methionine residue of the Ste2p N-terminal fragment included in the construct pET31b+/KSI-Ste2p(1-48)-M-HIS. The mutagenesis was carried out using the forward primer

5’-GTTTAGTTAACAGTACTAAGAAGAAGATGCTCGAGCACCACC-3’ and the complementary reverse primer. The T_m was found to be 79.4°C. The mutagenesis was carried out using the QuikChange Site-Directed Mutagenesis Kit, following the manufacturer’s protocol. Essentially, each mutagenesis reaction consisted of (in a total reaction volume of 50 µL): 1x reaction buffer (10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl, pH 8.8, 2 mM $MgSO_4$, 0.1% (v/v) Triton X-100, 0.1 mg/mL of nuclease-free bovine serum albumin), 0.1 mM dNTPs, 125 ng each of forward and reverse primers, 70 ng of purified construct and 2.5 U of *PfuTurbo* DNA polymerase. The mutagenesis thermal cycling conditions were: (1) initial denaturation step at 95°C for 30 sec; (2) denaturation step at 95°C for 30 sec; (3) annealing step at 55°C for 1 min; (4) extension step at 68°C for 1 min/kb of plasmid length; (5) repeat 17 cycles from step (2) to step (4). The mutagenesis reaction was transformed into One Shot TOP10F´ Competent Cells, following the manufacturer’s protocol. Mutagenesis reactions were performed consecutively to incorporate the lysines into the two locations yielding a final single construct, pET31b+/KSI-M-KKK-Ste2p(2-48)-KKK-M-HIS. The associated final fusion protein product arising from this construct was termed ‘KSI-SK-HIS fusion protein (where SK stands for ‘KKK-Ste2p(2-48)-KKK-M’ and/or Stephanie Kendall). Single transformed colonies were grown overnight in 2xYT medium and DNA was extracted from the cells using a Qiagen mini-prep kit. Constructs were confirmed by sequencing.

2.5 Recombinant Protein Expression

2.5.1 Fusion Protein Expression Using 2xYT or M9 Minimal Media

To start the expression of the fusion protein of interest, a starter culture of 3 mL 2xYT medium or M9 minimal medium with 100 µg/mL ampicillin was inoculated with a single colony of BL21-CodonPlus (DE3)-RIL *E. coli* transformed with the appropriate pET31b+ based construct from a fresh plate. The starter culture was incubated for ~16 hours at 37°C with shaking until reaching a cell density of $OD_{600} \approx 0.9\text{-}1.0$. A 4 L Erlenmeyer flask containing 800 mL of either 2xYT or M9 minimal medium was inoculated 1:400 or 1:50, respectively, with the starter culture. The cultures were incubated at 37°C with continuous shaking until an $OD_{600} \approx 0.4\text{-}0.5$ was reached. When the cultures reached this cell density, 0.2

mM isopropyl β-D-thiogalactopyranoside (IPTG) was added to induce the production of the fusion protein. The induced cell culture was incubated at 15°C overnight with shaking.

2.5.2 Optimization of Expression Induction Time

In order to find the optimal time for induction of fusion protein expression, a time course experiment was set up. Cultures were grown and induced as described in section 2.5.1 with the induction time varied and 1 mL aliquots were taken from the culture prior to induction at 1 hour, 2 hours, 3 hours, 4 hours and 5 hours post-induction. The 1 mL aliquots were centrifuged at 16,000x g for 2 minutes. The supernatant was aspirated and the cell pellet was re-suspended in 100 µL of 50 mM Tris, pH 8.0. The suspension was sonicated with 100 W output for 2 x 20 sec. After sonication, 100 µL of 2x SDS loading buffer (5% (v/v) β-mercaptoethanol, 100 mM Tris HCl pH 6.8, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol) was added to the suspension. These samples were the total cell fraction. In order to separate the soluble and insoluble fractions, a 1 mL aliquot of cell culture was centrifuged, re-suspended in 100 µL of 50 mM Tris, pH 8.0 and sonicated on ice at 100 W output for 6 x 20 sec pulses. After sonication, the sample was centrifuged for 2 min at 16,000x g. The supernatant was kept as the soluble fraction and the cell pellet was re-suspended in 100 µL of 50 mM Tris, pH 8.0. To both the soluble and insoluble fractions, 100 µL of 2x SDS loading buffer was added. All samples were boiled at 100°C for 5 min and run on a 15% SDS-polyacrylamide gel using polyacrylamide gel electrophoresis (PAGE) as described below in section 2.7.1.

2.5.3 Optimization of IPTG Concentration

In order to optimize the IPTG concentration for maximum protein production, an IPTG concentration dependent experiment was run. The initial experiment was set up the same as the previous induction time experiment, except that the induction time was maintained for five hours at 37°C and the concentration of IPTG was varied. The following IPTG concentrations were used; no IPTG, 0.2 mM, 0.4 mM, 0.6 mM, 0.8 mM and 1 mM IPTG. Again, 1 mL aliquots were removed and separated into total cell, soluble and insoluble fractions as described above in section 2.5.2. These samples were all analysed using SDS-PAGE.

2.6 Fusion Protein Purification

2.6.1 Ni-NTA Affinity Enrichment

After expression of the fusion protein in either 2xYT or M9 minimal medium with ampicillin, the cell cultures were harvested by centrifugation at 5,500x g for 15 mins at 4°C. After centrifugation, the cell pellet was re-suspended in 15 mL of 50 mM Tris pH 8.0 and sonicated on ice with a 100 W output for 6 x 20 sec pulses. After sonication the sample was centrifuged for 15 mins at 13,500x g at 4°C. The cell pellet was re-suspended in 8 mL guanidine hydrochloride binding buffer (6 M guanidine hydrochloride, 5 mM imidazole, 500 mM NaCl, 20 mM Tris, pH 8.0) and loaded onto 4 mL of Ni-NTA resin. The Ni-NTA resin was prepared according to the manufacturer's protocol and equilibrated with 10 column volumes of binding buffer. The fusion protein was bound to the resin by shaking at 4°C for 1 hour. The column was washed with 2 column volumes of binding buffer, then with another 2 column volumes of binding buffer with 20 mM imidazole and finally with 2 column volumes of binding buffer with 50 mM imidazole. The fusion protein was eluted with 2 column volumes of binding buffer with 1 M imidazole. The final eluted fraction was dialyzed using a dialysis membrane (Spectrum Laboratories, Inc., Rancho Dominguez, CA) with a molecular weight cut off of 3500 Da. The sample was dialyzed in 4 L of dH_2O pH 7 for 4 hrs. This step was repeated a total of 4 times and the resulting solution was lyophilized. Total protein weight was estimated based on dry weight of lyophilized protein.

2.6.2 Cyanogen Bromide Cleavage

Cyanogen Bromide (CNBr) was used to remove the KSI and His tags from the KSI-Ste2p(1-71)-HIS and KSI-SK-HIS fusion proteins, yielding either Ste2p (2-54) peptide (SDAAPSLSNLFYDPTYNPGQSTINYTSIYGNGSTITFDELQGLVNSTVTQAIM) or SK peptide
(KKKSDAAPSLSNLFYDPTYNPGQSTINYTSIYGNGSTITFDELQGLVNSTKKKM)
respectively. The lyophilized fusion protein and CNBr were dissolved separately in 70% (v/v) trifluoroacetic acid (TFA) in water. The reaction was initiated by mixing equal volumes of the two solutions together. The final concentration of protein and CNBr in the reaction was ~20

mg/mL and ~0.5 M, respectively. The cleavage reaction was carried out in the dark, at room temperature for either 4.5 hours or 24 hours. The reaction was stopped by diluting it three-fold with water and speed-vacuuming to produce a dry pellet. The pellet was re-suspended in 50 µL of H_2O/acetonitrile (ACN) (50:50) and speed-vacuumed. This step was repeated three times to ensure that all of the TFA and CNBr were removed from the sample. The digests were analyzed using SDS-PAGE.

Cyanogen bromide cleavage of the KSI-Ste2p(1-71)-HIS fusion protein was tested using a number of different solvents. The cleavage reaction was carried out as described in the previous paragraph. However, both the fusion protein and CNBr were dissolved in one of three solvents; 70% (v/v) TFA, 70% (v/v) formic acid or 6 M guanidine hydrochloride. The reactions were all carried out in the dark at room temperature for 24 hours. All reactions were stopped by diluting them three-fold with water and drying them down with the speed-vaccuum. The reactions were all washed with H_2O/ACN (50:50) and dried down using the speed-vacuum, as previously described.

2.6.3 High Performance Liquid Chromatography (HPLC)

Purification of the peptide of interest from the other cleavage reaction fragments was performed on a Hewlett-Packard/Agilent 1100 series instrument with a quaternary gradient solvent system similar to that previously reported (Zheng *et al.*, 2006). The quaternary gradient solvent system can be used with up to four solvents for gradient analysis. The lyophilized CNBr cleavage reaction was re-suspended in 2,2,2-trifluoroethanol (TFE), incubated at 37°C for 20 min, and then an equal volume of H_2O was added. The final concentration of the protein was ~10 mg/mL. The lyophilized peptides solubilized to produce a clear solution. The sample was loaded onto a SUPELCOSIL LC-18, 5 µm column (Supelco, Bellefonte, PA, USA) (25 cm x 10 mm ID with 2 cm x 4 mm ID guard cartridge) for semi-preparative purification. The solvent gradient system consisted of solvent A (acetonitrile with 0.1% (v/v) TFA) and solvent B (water with 0.1% (v/v) TFA). After injection, solvent A was increased from 5% to 55% over 25 min, then reduced to 5% solvent A over 10 min (to recondition the column). The flow rate was maintained at 1 mL/min throughout the purification, and elution of the peptide was monitored using a diode array detector (DAD), from 200 nm to 280 nm. The peak fractions were collected, pooled and lyophilized.

2.7 Protein Analyses

2.7.1 SDS-Polyacrylamide Gel Electrophoresis

Protein samples were mixed with 1x SDS gel-loading buffer (2.5% (v/v) β-mercaptoethanol, 50 mM Tris-HCl, pH 6.8, 2% (w/v) SDS, 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol) and boiled at 100°C for 5 min. The majority of protein samples were resolved on a Tris-glycine SDS-polyacrylamide gel (15% (v/v) Acrylamide/Bis (37.5:1), 375 mM Tris, pH 8.8, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulfate, 0.04% (v/v) N,N,N',N'-tetramethylethylenediamine) with a 5% stacking gel (5% (v/v) Acrylamide/Bis (37.5:1), 125 mM Tris, pH 6.8, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulfate, 0.1% (v/v) N,N,N',N-tetramethylethylenediamine). The SeeBlue® Plus2 Pre-stained Standard and 15-20 µL of denatured protein samples were loaded onto the SDS gel. Electrophoresis was carried out using the Bio-Rad Mini Protean Tetra cell (Bio-Rad Laboratories, Hercules, CA) in running buffer composed of 25 mM Tris-HCl, 192 mM glycine and 0.1% (w/v) SDS. A constant voltage of 100 V was applied until the dye band was through the stacking gel. The voltage was then increased to 200 V until the dye band ran off the gel. Protein bands were visualized by soaking the gel in Coomassie staining solution (45% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.25% (w/v) Coomassie Brilliant Blue G 250) and then de-staining with 30% (v/v) methanol and 10% (v/v) glacial acetic acid.

2.7.1.1 SDS-PAGE of CNBr Cleavage Products

The dried CNBr cleavage reactions were re-suspended in 2x Tris-tricine sample buffer (100 mM Tris-Cl, pH 6.8, 24% (w/v) glycerol, 8% (w/v) SDS, 0.2 M dithiothreitol) and boiled at 100°C for 5 mins. The samples were resolved on a 15% SDS Tris-tricine polyacrylamide gel (15% (v/v) Acrylamide/Bis (37.5:1), 1 M Tris-Cl, pH 8.45, 0.1% (w/v) SDS, 1% (w/v) glycerol, 0.1% (w/v) ammonium persulfate, 0.04% (v/v) N,N,N',N'-tetramethylethylenediamine) with a 4% stacking gel (4% (v/v) Acrylamide/Bis (37.5:1), 1 M Tris-Cl, pH 8.45, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulfate, 0.1% (v/v) N,N,N',N-tetramethylethylenediamine). The SeeBlue® Plus2 Pre-stained Standard and 15-20 µL of

denatured CNBr cleavage protein samples were loaded onto the Tris-tricine SDS gel. Electrophoresis was carried out using the Bio-Rad Mini Protean Tetra cell (Bio-Rad Laboratories, Hercules, CA) with a cathode buffer (100 mM Tris base, 100 mM tricine, 0.1% (w/v) SDS) and anode buffer (200 mM Tris, pH 8.9). A constant voltage of 100 V was applied until the dye band was through the stacking gel. Then the voltage was increased to 200 V until the dye band ran off the gel. Protein bands were visualized using a silver stain protocol. The gels were placed in a fixing solution (32% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.01% (v/v) formaldehyde) for at least two hours with shaking. The gels were washed once with 30% (v/v) methanol and twice with distilled water. The gels were incubated with a solution containing 0.02% (w/v) sodium thiosulfate in water for 1 minute. After rinsing the gels with water they were soaked in the silver stain (0.2% (w/v) $AgNO_3$, 0.025% (v/v) formaldehyde, 0.02% (w/v) citric acid) for 20 minutes. They gels were washed with water three times and then developed until good resolution using developing solution (6% (w/v) sodium carbonate, 0.02% (v/v) formaldehyde, 0.0004% (w/v) sodium thiosulfate). The development was stopped using a stop solution containing 20% (v/v) methanol and 0.2% (w/v) citric acid.

2.7.2 Matrix Assisted Laser Desorption Ionization – Time of Flight (MALDI-TOF) Mass Spectrometry

Mass spectrometric analysis was performed in the PBI/NRC Mass Spectroscopy facility on an AB SCIEX 4800 tof-tof instrument (AB SCIEX, Concord, ON, CA) equipped with an Nd:YAG laser, 355 nm wavelength 3-7ns pulse width 200Hz firing rate. Four hundred laser shots were accumulated and averaged with the laser power set at 4463 arbitrary units. The mass spectrometer was operated in positive ion linear mode scanning from m/z 2000 to m/z 30000 using the instrument default calibration. The samples were used as submitted (see below) and mixed on the MALDI plate (0.5μL) sample with CHCA matrix (0.5 μL).

The cyanogen bromide cleavage reaction was re-suspended in ACN/H_2O (50:50) to a final concentration of 1 mg/mL. The sample was solubilised to produce a clear liquid, and then sent for MALDI-TOF MS analysis.

The spectra were compared to cyanogen bromide theoretical peptide digest of the protein with two search engines (1) auto digest simulation within Biolynx (Waters) and (2) MS-Digest within Protein Prospector (UCSF Mass spectrometry Facility).

The lyophilized samples from the HPLC peaks were dissolved in ACN/H_2O (50:50) with a final protein concentration of 1 mg/mL and sent for MALDI-TOF MS analysis. The ACN/H_2O sample solubilised to produce a clear liquid.

2.7.3 Dynamic Light Scattering (DLS)

The dried SK peptide was re-suspended in four different solvents: (i) 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0, (ii) TFE/H_2O (1:1) (+0.1% (v/v) TFA), (iii) chloroform:methanol:water ($CHCl_3$:CH_3OH:H_2O) (4:4:1), or (iv) 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 and 0.1% (w/v) n-octyl-β-D-glucopyranoside (OG). The final concentration of all samples was 0.2 mM. The solutions were each filtered using a 0.2 µm nanopore filter (Millipore, Billerica, MA, USA) and 12 µL of each was loaded into quartz cuvettes. The DLS measurements were performed using a DynaPro-MS800 instrument (Wyatt Technology, Santa Barbara, CA, USA) running Dynamics software at 25°C. The laser was at a wavelength of 830 nm for obtaining light-scattering measurements at a 90° angle. Auto-correlated light intensity data was converted to the hydrodynamic radius (Rh) using the Stokes-Einstein equation using the Dynamics software.

2.7.4 Circular Dichroism (CD)

2.7.4.1 Circular Dichroism Temperature Study of Synthetic KKK-Ste2p(14-43)-KKK

To prepare the H_2O sample, the chemically-synthesized KKK-Ste2p(14-43)-KKK peptide was dissolved in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 to a final concentration of 0.1 mg/mL. The TFE/H_2O sample was prepared by dissolving the KKK-Ste2p(14-43)-KKK peptide in 100% TFE and then an equal volume of dH_2O was added to make a ratio of 1:1 TFE:H_2O. The final concentration of the sample was 0.05 mg/mL. TFA was added to a final concentration of 0.1%. The LPPG sample was prepared by dissolving the KKK-Ste2p(14-43)-KKK peptide and 1-palmitoyl-2-hydroxy-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (LPPG), separately, in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0. The solutions were mixed together to produce a final concentration of protein and LPPG of 0.1 mg/mL and 2 mg/mL, respectively. The circular dichroism (CD) spectra of the peptides were recorded on a PiStar-180 spectrometer (Applied Photophysics, Leatherhead, Surrey, UK) running PiStar software. The spectrometer

was equipped with a water bath for temperature control. Approximately 20 µL of each sample was used in a quartz cuvette with a path length of 0.1 cm. All spectra were the average of 8 scans between 260 and 190 nm or 185 nm at an interval of 0.5 nm and a slit width of 6 nm. The spectra were obtained at 20°C, 40°C, 60°C and 80°C. CD spectra on blanks, corresponding to the different solutions without protein dissolved, were collected with the same parameters and subtracted from the spectra containing the protein. The CD intensities were expressed as mean residue ellipticities [(deg·cm^2)/(dmol·residue)]. Spectra were smoothed over five points to reduce the signal-to-noise ratio. The spectra were also normalized to zero at the 250-260 nm region. Secondary structure analysis of the circular dichroism data was performed using the CDNN CD spectra deconvolution software (Applied Photophysics).

2.7.4.2 Circular Dichroism of Recombinant SK peptide

The SK peptide was dissolved in 100% TFE and then an equal volume of dH_2O was added to make a ratio of 1:1. The final concentration of the sample was 1 mg/mL. TFA was added to a final concentration of 0.1%. The sample was filtered with a 0.2 µm nanopore filter. The circular dichroism spectrum was run on the sample the same as was mentioned for the samples in the previous paragraph with a few minor changes. Approximately 200 µL of sample was used in a quartz cuvette with a path length of 1 cm and the spectrum was obtained at room temperature.

2.7.5 Nuclear Magnetic Resonance (NMR)

2.7.5.1 NMR of Synthetic KKK-Ste2p(14-43)-KKK

Initial NMR experiments were conducted using the chemically-synthesized KKK-Ste2p(14-43)-KKK (8.7 mg of lyophilized peptide) dissolved in 90% H_2O/10% deuterium oxide (D_2O) which produced a clear solution with a final peptide concentration of ~ 4 mM. 1D-proton (^{1}H) with water suppression by excitation sculpting (number of scans=64), phase sensitive double-quantum filtered 2D ^{1}H-^{1}H correlated spectroscopy (COSY) with water suppression by excitation sculpting (number of scans=96), phase sensitive 2D nuclear Overhauser effect spectroscopy (NOESY) with water suppression by excitation sculpting (number of scans=112) and phase sensitive 2D total correlated spectroscopy (TOCSY) with

water suppression by excitation sculpting (number of scans=112) experiments were run on the sample. The 2D ^{1}H-^{1}H COSY pulse program was double quantum filtered and phase sensitive, and used 512 t_1 increments. The mixing time for the NOESY was optimized to 0.2 sec. The mixing time was optimized to 0.035 sec and the relaxation delay was optimized to 3 sec for the TOCSY experiments. The water signal was suppressed using excitation sculpting with gradients for all experiments. The 90 degree pulse was optimized prior to running all experiments. All NMR spectra were recorded at 25 °C on Bruker Avance 500 MHz NMR spectrometer (Bruker BioSpin Corporation, Billerica, MA, USA) running TopSpin 2.1 software and equipped with a 5 mm TXI cryoprobe. All ^{1}H dimensions were referenced to the solvent. Assignment of the 2D ^{1}H-^{1}H COSY spectrum peaks was performed using the processing software, Felix (Molecular Simulations Inc.).

2.7.5.2 NMR of Recombinant SK Peptide

NMR experiments were also performed on the SK peptide (12 mg), which was re-suspended in 50% TFE-d2/40% H_2O/10% D_2O (0.1% TFA) to a final peptide concentration of ~4 mM. 1D-^{1}H and 2D ^{1}H-^{1}H COSY experiments were run on the sample as described in section 2.7.5.1 except that the time domain parameter was optimized to 256 for the SK peptide.

NMR experiments were performed on the ^{15}N-SK peptide (9.4 mg), which was re-suspended in 50% TFE-d2/40% H_2O/10% D_2O (0.1% TFA) to a final peptide concentration of ~3 mM. 1D-^{1}H with water suppression by excitation sculpting (number of scans=64) and 2D ^{1}H-^{15}N phase sensitive heteronuclear single quantum coherence (HSQC) using via double inept transfer, number of scans=96, F1=512, F2=2048) experiments were run on the sample.

3.0 RESULTS

3.1 Recombinant Ste2p(2-54) Peptide

3.1.1 The Recombinant Expression and Purification System for Production of KSI-Ste2p(1-71)-HIS Fusion Protein

A requirement for NMR studies is the production of milligram quantities of isotopically labelled peptides. Toward this, an expression system was previously designed to over-express a fusion protein containing the N-terminal region of Ste2p, residues 1-71, in *E. coli*. The pET31b+ vector was selected for cloning of the Ste2p peptide because it is optimal for production of peptides and small proteins (Figure 3.1). The pET31b+ vector contains the T7 promoter (Rosenberg *et al.*, 1987; Studier and Moffatt, 1986). Therefore, the gene of interest is cloned into the pET31b+ vector and is under the control of bacteriophage T7 transcriptional machinery, where T7 polymerase is expressed in the bacterial cell. The pET31b+ vector contains a coding sequence for the KSI peptide which is upstream of the unique cloning site, where Ste2p(1-71) was inserted. The KSI peptide is a useful tool in expression of hydrophobic peptides because it forces the fusion protein into inclusion bodies, from which the hydrophobic fusion protein can be solubilised and purified (Kuliopulos *et al.*, 1987; Kuliopulos and Walsh, 1994). In addition the vector contains a coding region for six histidine residues downstream of where the Ste2p(1-71) coding fragment was inserted. This adds a HIS tag on the fusion protein which allows for affinity purification using Ni-NTA resin. Thus, over-expression of this construct in *E. coli* produced the KSI-Ste2p(1-71)-HIS fusion protein which contained Ste2p(1-71) with an N-terminal KSI tag, which promoted inclusion body formation, and a C-terminal HIS tag which allowed for purification using a Ni-NTA column (Figure 3.2).

An initial expression was performed in BL21-CodonPlus (DE3)-RIL cells. These cells were chosen, as they were engineered to specifically express selected tRNAs for optimized arginine, isoleucine and leucine codon usage and thus maximized recombinant protein yield.The total cell extract, insoluble and soluble fractions were analyzed by SDS-PAGE (Figure 3.3). This experiment was performed in order to verify that the fusion protein was being expressed into inclusion bodies. The SDS-PAGE gel confirmed that the majority of the

KSI-Ste2p(1-71)-HIS fusion protein was in the insoluble fraction of the total cell extract (Figure 3.3, lane 3).

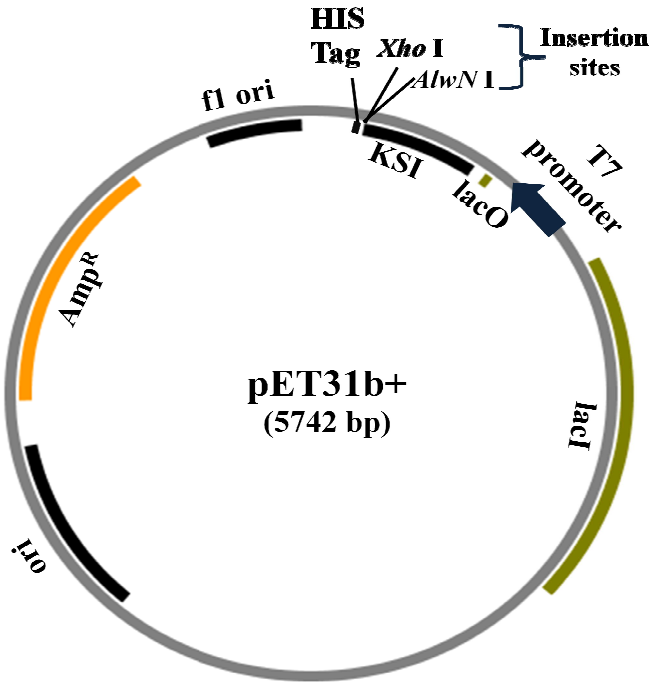

Figure 3.1: A Schematic Representation of the *E. coli* Expression Vector pET31b+.
The pET31b+ vector contains a coding region for ketosteroid isomerase (KSI) which is upstream of where the gene of interest is cloned. The vector also encodes six histidine residues (HIS tag) which are downstream of the gene of interest. The pET31b+ vector encodes a gene for ampicillin resistance (Amp^R) as well as the regulatory Z, Y and A gene segments of the lac operon (lacO) and a gene encoding the lac repressor (lacI) transcription factor. Under normal conditions, lacI binds to lacO preventing transcription from the T7 promoter. However, binding of IPTG to lacI releases it from lacO, thereby permitting the T7 polymerase to bind to the T7 promoter and subsequent transcription to proceed. The plasmid also contains an origin of plasmid replication in *E. coli* (ori) site and an origin of replication derived from filamentous phage (f1 ori).

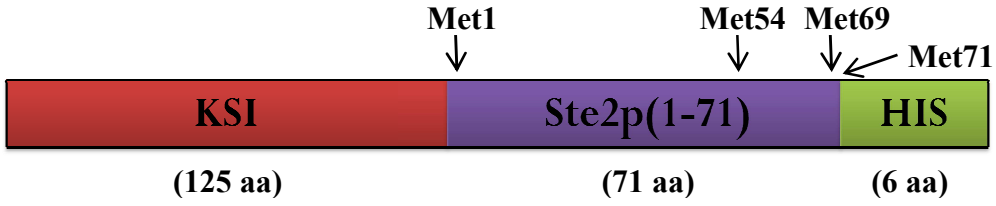

Figure 3.2: A Schematic Representation of the KSI-Ste2p(1-71)-HIS Fusion Protein.
The KSI-Ste2p(1-71)-HIS fusion protein contains an 125 amino acid N-terminal ketosteroid isomerase (KSI) tag. The Ste2p(1-71) portion of the protein contains four Met residues at Met1, Met54, Met69, and Met71. Cleavage after these residues by CNBr liberates the peptide, Ste2p(2-54), from the fusion protein. The six His residues (HIS tag) at the C-terminus are used for purification of the fusion protein using Ni-NTA resin.

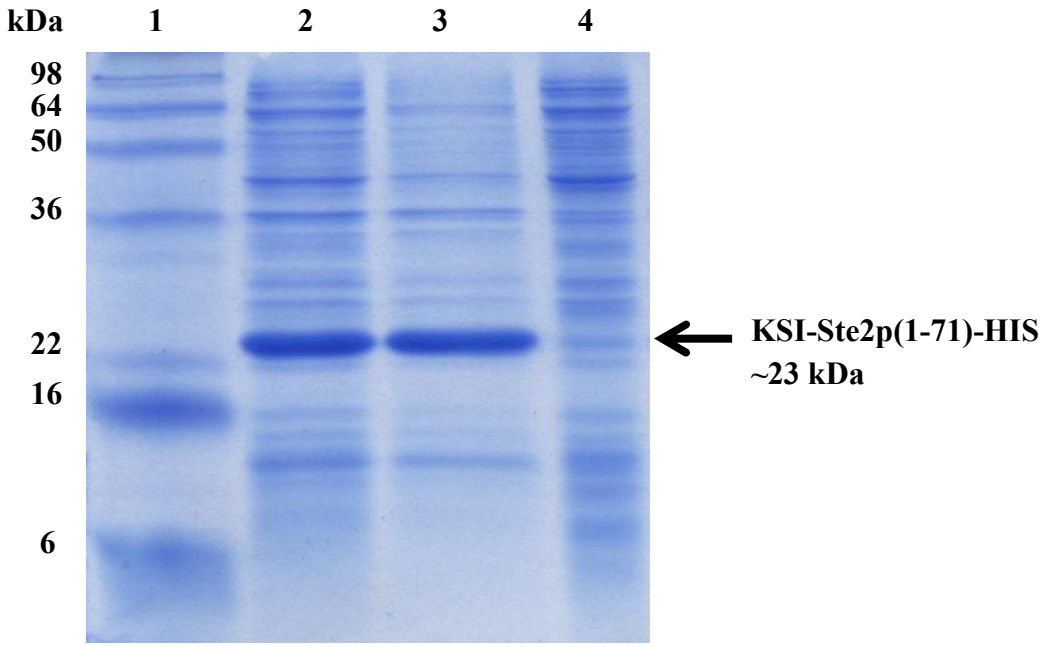

Figure 3.3: Over-expression of KSI-Ste2p(1-71)-HIS.
After over-expression of the KSI-Ste2p(1-71)-HIS fusion protein in BL21-CodonPlus (DE3)-RIL *E. coli*, the total cell extract as well as the insoluble and soluble fractions were analyzed for fusion protein expression levels using SDS-PAGE. The expected mass of the KSI-Ste2p(1-71)-HIS fusion protein is ~23 kDa. The lanes were loaded as follows: (1) protein marker; (2) total cell extract; (3) insoluble fraction of total cell extract; (4) soluble fraction of total cell extract. All samples were induced for 5 hours.

3.1.2 Optimization of KSI-Ste2p(1-71)-HIS Fusion Protein Expression in *E. coli*

In order to optimize the expression of the KSI-Ste2p(1-71)-HIS fusion protein, the protein expression was induced using different concentrations of IPTG in BL21-CodonPlus (DE3)-RIL *E. coli*. The concentrations varied from no IPTG, up to 1 mM IPTG. After 5 hours of induction, a sample of each culture was separated into total cell, insoluble and soluble fractions. Since the KSI tag forces the fusion protein to become insoluble, only the insoluble fraction of each culture was examined using SDS-PAGE (Figure 3.4). It was observed that induction of the KSI-Ste2p(1-71)-HIS fusion protein with 0.2 or 0.4 mM IPTG led to good expression levels of the fusion protein (Figure 3.4, Lane 5 & 6). The addition of higher concentrations of IPTG showed similar levels to that of the 0.2 mM induction. Therefore, the cost-effective conclusion is to use 0.2 mM IPTG for the induction during large scale expression and purification of the KSI-Ste2p(1-71)-HIS fusion protein.

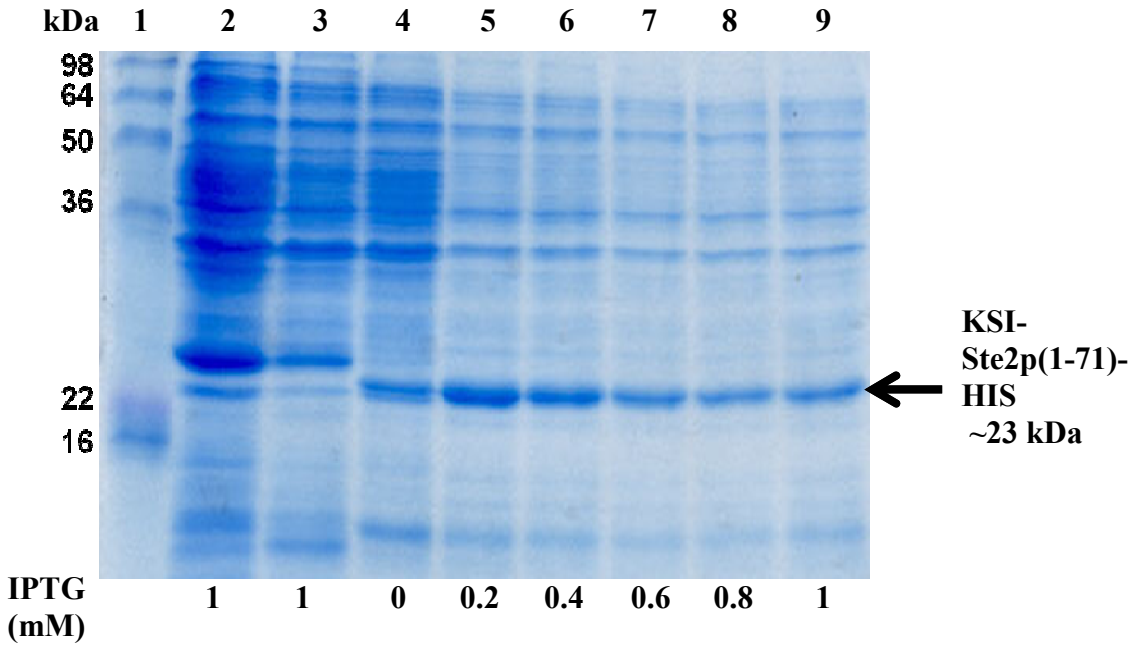

Figure 3.4: Effect of [IPTG] on KSI-Ste2p(1-71)-HIS Over-expression.
After over-expression of the KSI-Ste2p(1-71)-HIS fusion protein in BL21-CodonPlus (DE3)-RIL *E. coli*, the insoluble fraction from the total cell lysates were observed for fusion protein expression levels using SDS-PAGE. The concentration of IPTG that was used to induce the protein expression was varied from 0 - 1 mM IPTG. The expected mass of the KSI-Ste2p(1-71)-HIS fusion protein is ~23 kDa. The lanes were loaded as follows: (1) protein marker; (2) BL21-CodonPlus (DE3)-RIL cells containing no plasmid; (3) empty pET31b+ vector; (4-9) pET31b+/KSI-Ste2p(1-71)-HIS construct with increasing concentration of IPTG. (15% Tris-glycine SDS-PAGE of insoluble fraction of crude cell extract; All samples were induced for 5 hours).

Another aspect that was optimized for the expression of the KSI-Ste2p(1-71)-HIS fusion protein was the induction time after the addition of IPTG. The induction times varied from zero up to 5 hours. The IPTG concentration was kept constant at 0.4 mM. At different time intervals a sample of the cultures was removed and separated into the total cell, insoluble and soluble fractions and the insoluble fractions were analysed using SDS-PAGE (Figure 3.5). The expression of the fusion protein increased until the highest yield was observed after the 5 hour induction (Figure 3.5, Lane 9). However, out of convenience the KSI-Ste2p(1-71)-HIS fusion protein was left to induce overnight at 15°C after expression was initiated by the addition of IPTG.

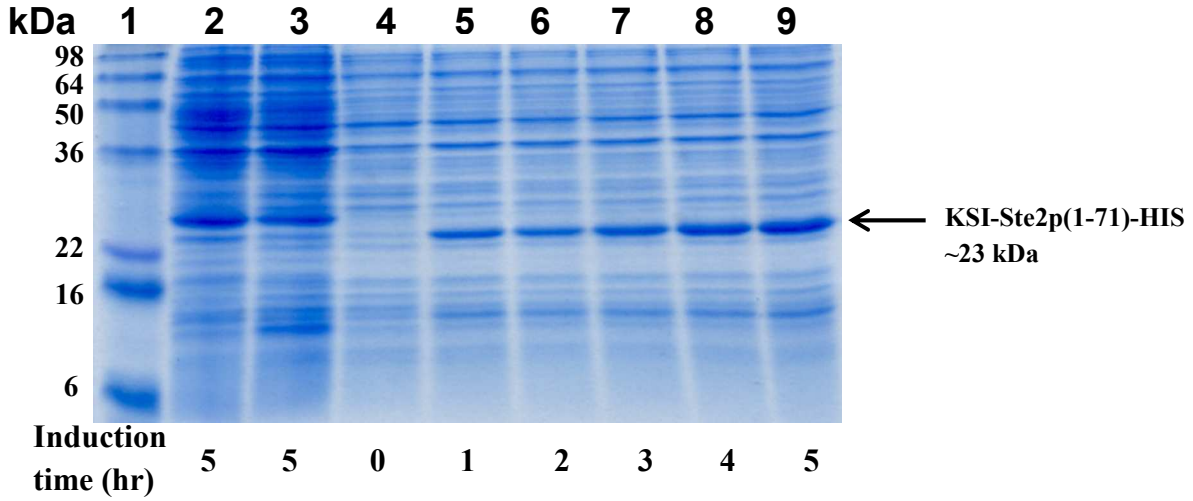

Figure 3.5: Effect of Induction Time on KSI-Ste2p(1-71)-HIS Over-expression.
After over-expression of the KSI-Ste2p(1-71)-HIS fusion protein in BL21-CodonPlus (DE3)-RIL *E. coli*, the insoluble fractions from the total cell lysates were analyzed for fusion protein expression levels using SDS-PAGE. The induction time for protein expression was varied from no induction up to 5 hours induction. The expected mass of the KSI-Ste2p(1-71)-HIS fusion protein is ~23 kDa. The lanes were loaded as follows: (1) protein marker; (2) BL21-CodonPlus (DE3)-RIL cells containing no plasmid; (3) empty pET31b(+) vector; (4-9) pET31b+/KSI-Ste2p(1-71)-HIS construct with increasing induction time. (15% Tris-glycine SDS-PAGE of crude cell extract; All samples were induced with 0.4 mM IPTG)

3.1.3 Purification

After expression of the KSI-Ste2p(1-71)-HIS fusion protein was optimized, a large-scale expression of the fusion protein was carried out in 2xYT medium and purified using Ni-NTA resin. Aliquots were removed from the culture after induction was completed and separated into total cell, insoluble and soluble fractions. These fractions as well as aliquots after each step during purification were analysed using SDS-PAGE (Figure 3.6). As expected, the majority of the fusion protein was in the insoluble fraction (Figure 3.6, Lane 3) with little to none of the fusion protein being expressed in the soluble fraction (Figure 3.6, Lane 4). Some of the fusion protein was lost in the flow through after application to the affinity resin, as well as in the washes, but most of the contaminating proteins were removed (Figure 3.6, Lanes 5-8).

Following the washes, KSI-Ste2p(1-71)-HIS fusion protein was eluted with binding buffer containing 1 mM imidazole and observed on the gel at the expected molecular weight of 23 kDa (Figure 3.6, Lane 9). After elution, the fusion protein was dialyzed to remove the imidazole and salts. The fusion protein was found to be at a high concentration as well as highly pure. After the previously mentioned optimization of the fusion protein expressions levels and purification, the total yield was ~70 mg/liter of culture of pure KSI-Ste2p(1-71)-HIS fusion protein based on the weight of the lyophilized protein.

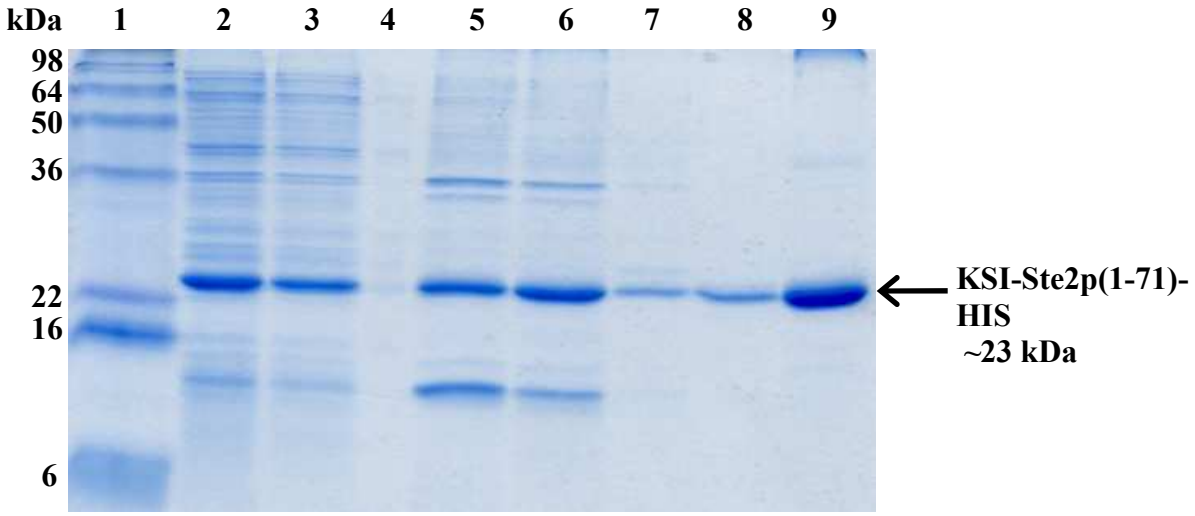

Figure 3.6: Ni-NTA Purification of KSI-Ste2p(1-71)-HIS.
SDS-PAGE was used to analyze the cell extracts of BL21-CodonPlus (DE3)-RIL *E. coli* after large-scale expression to observe the levels of KSI-Ste2p(1-71)-HIS fusion protein in the total cell, insoluble and soluble fractions. SDS-PAGE was also used to observe the purity and loss of the fusion protein during the purification using Ni-NTA resin. The lanes were loaded as follows: (1) protein marker; (2) total cell extract; (3) insoluble fraction of total cell extract; (4) soluble fraction of total cell extract; (5) flow through fraction of Ni-NTA purification; (6) fraction of binding buffer wash; (7) fraction of binding buffer + 20 mM imidazole wash; (8) fraction of binding buffer + 50 mM imidazole wash; (9) elution fraction from Ni-NTA column using binding buffer with 1 M imidazole.

3.1.4 Cyanogen Bromide Cleavage

CNBr is known to cleave proteins specifically at methionine residues by means of cleavage at its C-terminal carboxyl group (Figure 3.7) (Gross and Witkop, 1962). The KSI-Ste2p(1-71)-HIS fusion protein contains four methionine residues at positions Met1, Met54, Met69, and Met71 of the Ste2p(1-71) portion. Therefore, cleavage of the KSI-Ste2p(1-71)-HIS fusion protein will release the peptide of interest, Ste2p(2-54), from the KSI and His tags. Initially, the cyanogen bromide cleavage reaction was carried out for 4.5 hours, as was previously described in the literature (Estephan *et al.*, 2005). To examine if the cleavage reaction was being carried out completely, purified KSI-Ste2p(1-71)-HIS that had not been subjected to cleavage, as well as a zero time CNBr cleavage control and a 4.5 hour CNBr reaction were examined using SDS-PAGE (Figure 3.8). After a 4.5 hour CNBr cleavage reaction the KSI-Ste2p(1-71)-HIS fusion protein band had almost disappeared while a new band appeared at ~14 kDa, corresponding to the expected KSI fragment, as well a band at ~6 kDa appeared, which corresponds to the expected peptide, Ste2p(2-54) (Figure 3.8, Lane 4). After the 4.5 hour cleavage, there also appeared to be a number of other bands which may be due to incomplete cleavage of the fusion protein, or degradation of the other peptide fragments. It has been shown previously that a longer CNBr cleavage reaction time of 24 hours results in completely digested protein (Arevalo *et al.*, 2003; Rodríguez *et al.*, 2003). Therefore, in order to ensure complete digestion of the peptide, all subsequent CNBr cleavage reactions were carried out for 24 hours.

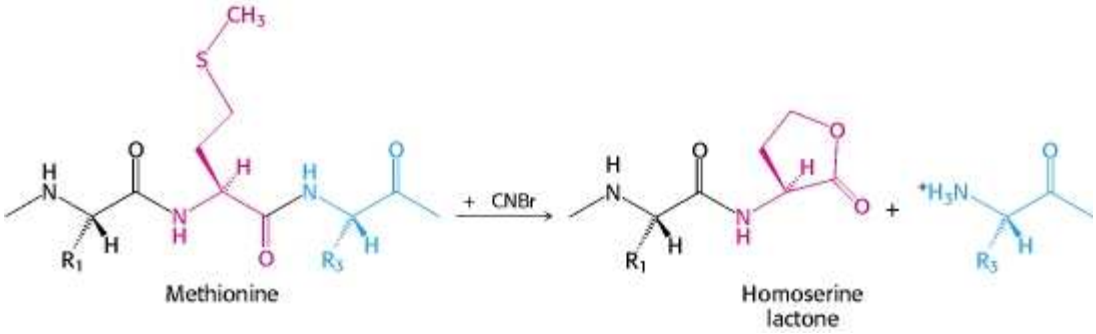

Figure 3.7: Cyanogen Bromide Cleavage Reaction.
Cyanogen bromide cleaves at the C-terminal end of Methionine residues by attacking the sulphur of the Methionine residue. This results in the formation of homoserine lactone at the C-terminal end of the peptide. The sulphur replaces the bromide ion in the cyanogen bromide forming methyl thiocyanate. The figure was taken from (Berg *et al.*, 2002).

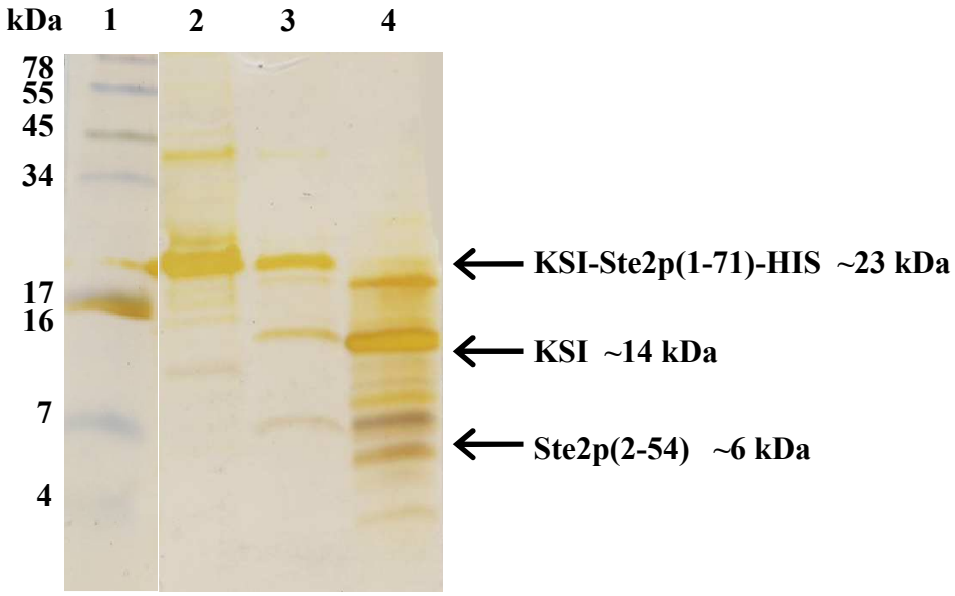

Figure 3.8: CNBr Cleavage of KSI-Ste2p(1-71)-HIS.
Cyanogen bromide was used to cleave the KSI-Ste2p(1-71)-HIS fusion protein at the C-terminal end of Met residues in order to release the Ste2p(2-54) peptide. The cyanogen bromide reaction was incubated for 4.5 hours in order to determine if the cleavage reaction was being carried out to completion during this time. The gel was loaded as follows: (1) protein marker; (2) purified KSI-Ste2p(1-71)-HIS fusion protein; (3) KSI-Ste2p(1-71)-HIS fusion protein CNBr cleavage zero time control; (4) KSI-Ste2p(1-71)-HIS fusion protein 4.5 hour CNBr cleavage. (15% Tris-Tricine SDS-PAGE of evaporated CNBr cleavage products, stained with silver stain.)

3.1.5 HPLC Purification

After cleavage of the KSI-Ste2p(1-71)-HIS fusion protein to produce the peptide Ste2p(2-54), the resulting digested sample was subjected to HPLC separation to purify this desired fragment away from the other cleavage products. The CNBr reaction fragments (~10 mg protein) in TFE/H2O (1:1), were loaded onto the preparative C18 reverse-phase column, which separates peptides based on their hydrophobicity. A gradient system of acetonitrile/water/TFA was used to separate the fragments similar to that used previously (Zheng *et al.*, 2006). After separation on the column, the chromatogram observed at 220 nm

had four peaks, at ~7.5 min, ~9.5 min, ~10.5 min and ~13.5 min respectively (Figure 3.9). The peak fractions from multiple runs were collected, dried down and re-suspended and loaded onto a 15% Tris-tricine gel and silver stained. The resulting gel showed no bands (data not shown). Subsequently, the dried samples were re-suspended in ACN/H_2O (1:1) and sent for MALDI-TOF MS analysis. None of the peak samples contained any of the expected peptide masses after CNBr cleavage (data not shown). The HPLC purification of the Ste2p(2-54) peptide was repeated numerous times and many aspects of the purification were reassessed, but the Ste2p(2-54) peptide was never recovered after the HPLC step. Every time the CNBr cleavage products were loaded onto the column the Ste2p(2-54), KSI and His tag peptides all disappeared.

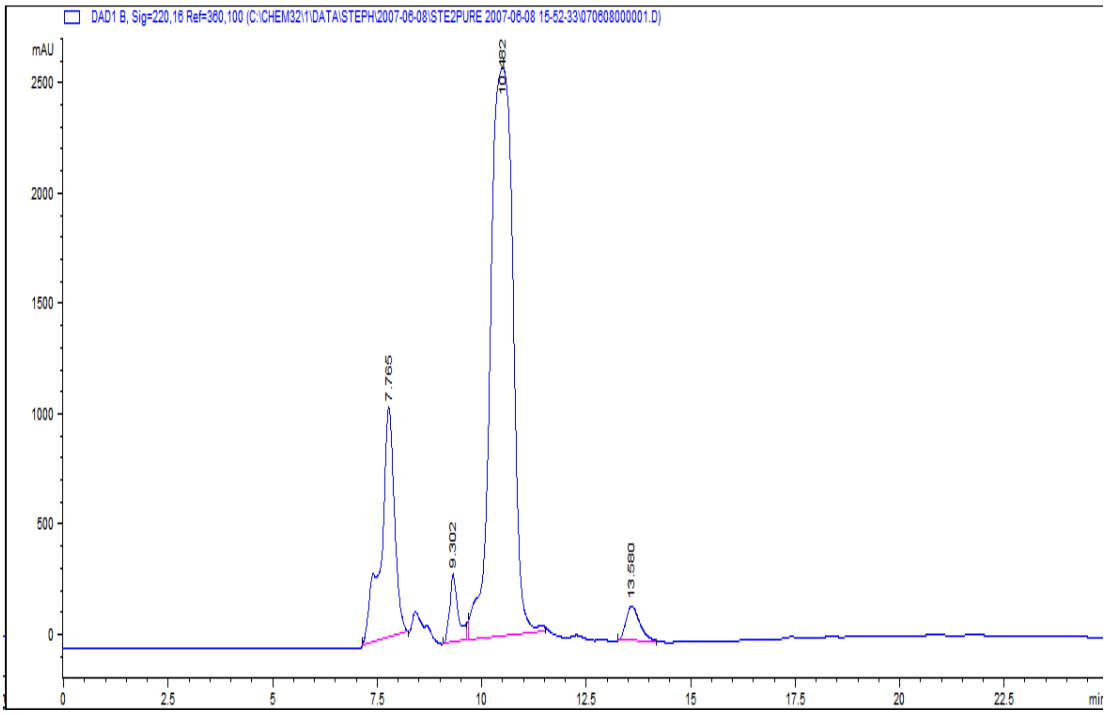

Figure 3.9: HPLC Chromatogram of KSI-Ste2p(1-71)-HIS Fusion Protein CNBr Cleavage Products.
HPLC was performed after CNBr cleavage in order to purify the Ste2p(2-54) peptide from the other cleavage fragments. The chromatogram at 220 nm displays the purification of the CNBr reaction products after a 24 hour digest.

3.2 Synthetic Ste2p(14-43) and KKK-Ste2p(14-43)-KKK Peptides

After purification of the recombinant Ste2p(2-54) peptide proved to be unfeasible, CanPeptide Inc. was contracted to produce a synthetic peptide containing residues 14-43 of Ste2p. The rationale behind choosing residues 14-43 of Ste2p was because this produced a smaller peptide, which would hopefully be easy and inexpensive to synthesize and which encompassed the residues of interest, Pro15, Ile24 and Ile29. However, after synthesizing the Ste2p(14-43) peptide, CanPeptide Inc. was also unable to purify the peptide by HPLC. One possible factor contributing to these difficulties is the hydrophobic nature of the peptide. A review of the literature indicated that the addition of three lysine residues to both the N and C-termini of a peptide can decrease peptide hydrophobicity sufficiently to allow for purification of hydrophobic peptides by HPLC (Xie *et al.*, 2000). Therefore, CanPeptide Inc. was asked to synthesize the Ste2p(14-43) peptide again with the addition of three lysine residues to both ends of the peptide. The addition of the lysine residues allowed them to synthesize the KKK-Ste2p(14-43)-KKK peptide and purify it by HPLC. This peptide was used to assess suitable solvent systems for biophysical analyses and for initial structural analysis by CD and NMR.

3.2.1 Circular Dichroism Temperature Study of the Synthetic KKK-Ste2p(14-43)-KKK Peptide

Circular dichroism analysis of the chemically synthesized KKK-Ste2p(14-43)-KKK peptide was carried out in order to determine the secondary structure of the peptide as well as the stability of the protein at different temperatures. Toward this, the solubility of the peptide was initially assessed in a number of solvents. In total, three solvents were tested, (1) 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0, (2) TFE/H_2O (1:1) with 0.1% TFA, and (3) LPPG (200 mM) in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0. The reason behind choosing three solvent systems was because CanPeptide Inc. had demonstrated that the peptide of interest dissolved in H_2O to produce a clear solution. However, it was unknown if the peptide solubilised completely. Therefore, it was decided to use TFE/H_2O as another solvent because TFE/H_2O is a very common solvent used in protein structural studies and aids in solubilising hydrophobic peptides. In addition, another solvent system that was used was the detergent LPPG in H_2O.

This solvent was chosen because it had been previously shown to produce higher quality NMR spectra and longer sample lifetimes when compared to other traditional NMR solvents (Krueger-Koplin *et al.*, 2004). The samples were prepared following previously described methods (Estephan *et al.*, 2005; Neumoin *et al.*, 2009); the peptide was dissolved in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 and produced a clear solution. However, when the peptide was dissolved in 100% TFE, the protein would not dissolve and produced a cloudy solution. But, upon addition of H_2O the peptide dissolved and the solution became clear. The peptide also produced a clear solution when it was dissolved in the LPPG with 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0.

The peptide samples were run on an π*-180 instrument. All samples were assessed at 20°C, 40°C, 60°C and 80°C. The ellipticity (θ) of the samples were converted to mean residue ellipticity ([θ]) and plotted against the wavelength.

The CD spectra (Figure 3.10 (A-C)) showed that the KKK-Ste2p(14-43)-KKK peptide was somewhat helical and that the helicity of the peptide was increased when in TFE. The helicity of the peptide was determined by the presence of troughs at 208 and 222 nm. The spectra of peptide in the H_2O and LPPG solutions looked similar. The H_2O and TFE samples appeared to denature at 60°C, while the LPPG sample was already denaturing at 40°C. The ellipticity and wavelength were imported into the CDNN CD spectra deconvolution software for structural analysis. The predicted secondary structure obtained from the software for the KKK-Ste2p(14-43)-KKK peptide in H_2O and LPPG were similar at 20°C with ~ 42% β-strand and 9% α-helix and ~46% β-strand and 9% α-helix, respectively (Table 3.1). While the predicted secondary structure in the TFE sample was ~26% β-strand and 13% α-helix. This is consistent with the known tendency of TFE to increase helical content (Sonnichsen *et al.*, 1992).

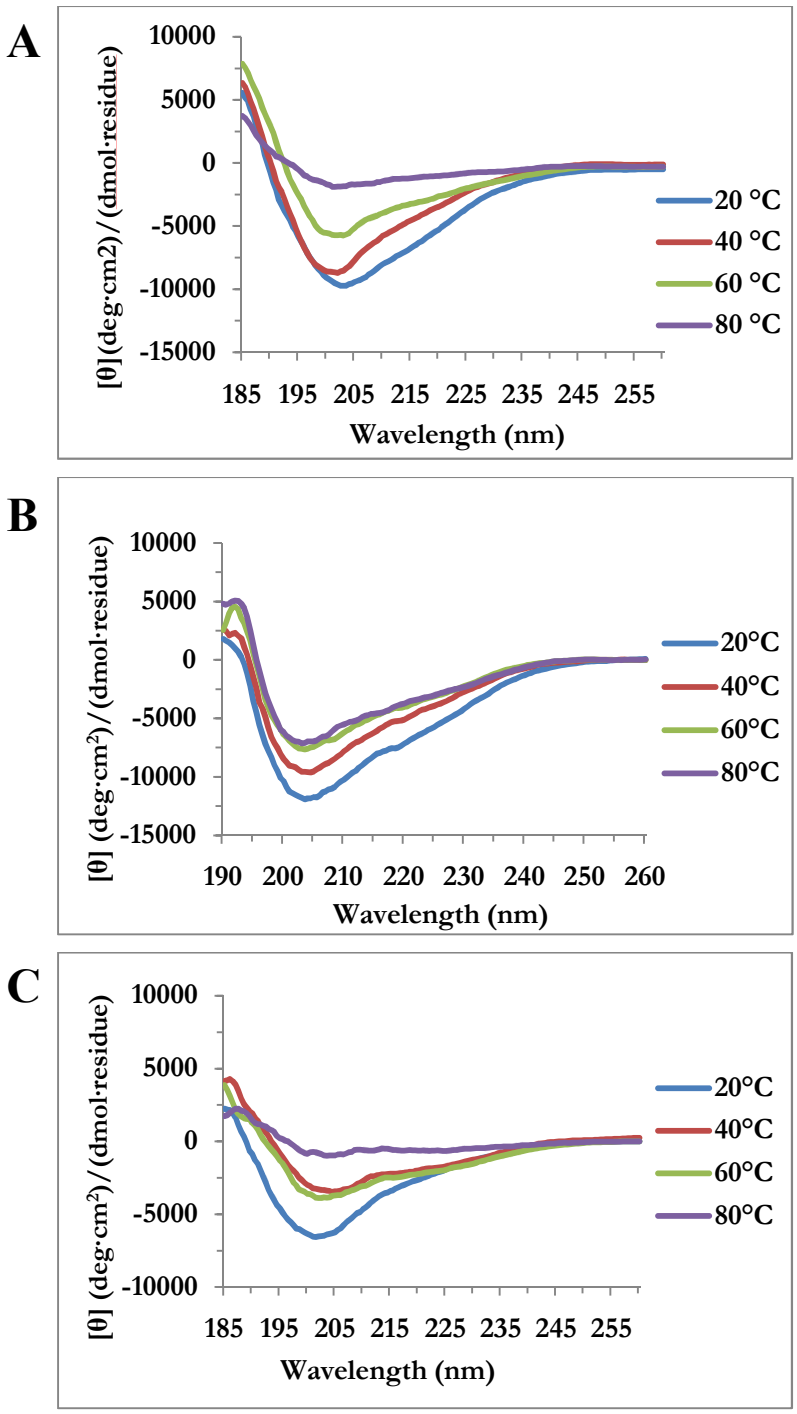

Figure 3.10: CD Spectra of KKK-Ste2p(14-43)-KKK Peptide.
The CD spectra of the KKK-Ste2p(14-43)-KKK peptide was obtained in order to determine the secondary structure of the peptide in different solvents. The spectra were run at increasing temperatures to assess denaturation of the peptide. A. The CD spectra of KKK-Ste2p(14-43)-KKK peptide in H_2O with a peptide concentration of 0.1 mg/mL. B. The CD spectra of KKK-Ste2p(14-43)-KKK peptide in TFE/H_2O (1:1) with 0.1% TFA with a peptide concentration of 0.05 mg/mL. C. The CD spectra of KKK-Ste2p(14-43)-KKK peptide in LPPG with 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 with a peptide concentration of 0.1 mg/mL.

Table 3.1: Predicted Secondary Structure of Synthetic KKK-Ste2p(14-43)-KKK Peptide Using CDNN CD Spectra Deconvolution Software.

KKK-Ste2p(14-43)-KKK in H_2O 185-260 nm				
	20°C	**40°C**	**60°C**	**80°C**
Helix	9.10%	8.60%	8.70%	8.40%
Antiparallel	38.50%	42.40%	54.50%	54.80%
Parallel	3.60%	3.60%	3.90%	3.90%
Beta-Turn	23.80%	23.40%	21.00%	19.30%
Rndm. Coil	31.80%	31.00%	26.40%	26.70%
Total Sum	106.70%	108.90%	114.50%	113.10%
KKK-Ste2p(14-43)-KKK in TFE/H_2O 190-260 nm				
	20°C	**40°C**	**60°C**	**80°C**
Helix	12.50%	10.20%	10.10%	10.30%
Antiparallel	22.60%	41.00%	45.60%	49.00%
Parallel	3.70%	4.10%	4.30%	4.50%
Beta-Turn	27.10%	22.80%	21.20%	20.50%
Rndm. Coil	37.10%	29.70%	27.70%	26.20%
Total Sum	103.00%	107.70%	108.90%	110.70%
KKK-Ste2p(14-43)-KKK in LPPG/H_2O 185-260 nm				
	20°C	**40°C**	**60°C**	**80°C**
Helix	8.50%	8.80%	8.90%	8.50%
Antiparallel	42.20%	53.10%	49.80%	54.10%
Parallel	3.60%	3.80%	3.80%	3.90%
Beta-Turn	22.30%	20.40%	20.90%	19.10%
Rndm. Coil	31.20%	26.60%	27.60%	26.80%
Total Sum	107.90%	112.70%	111.00%	112.40%

3.2.2 Nuclear Magnetic Resonance Spectroscopic Analysis of the Chemically Synthesized KKK-Ste2p(14-43)-KKK Peptide.

Based on the observed solubility and stability of the peptide in H_2O, the chemically-synthesized KKK-Ste2p(14-43)-KKK peptide was initially dissolved in D_2O to a final concentration of ~1 mM for NMR analysis. Initially, basic 2D TOCSY and NOESY experiments were carried out to assess the sample quality and spectral resolution for through bond and through space interactions. The preliminary 2D TOCSY and NOESY results were promising (data not shown). Therefore, the 2D TOCSY (Figure 3.11) and NOESY (Figure 3.12) experiments were repeated with a more concentrated sample, 3.5 mM KKK-Ste2p(14-43)-KKK peptide in D_2O. The 3.5 mM NMR experiments clearly show interactions between amino acid protons. For example, Figure 3.13 shows an expansion of the 2D-TOCSY from the 3.5 mM KKK-Ste2p(14-43)-KKK peptide sample demonstrating the interaction between the 2,6 H's and 3,5 H's of the tyrosine residues in the peptide (residues 7, 16 and 20 in the peptide, representing Y17, Y26 and Y30 in Ste2p). The expected chemical shift values for the 2,6 H's and the 3,5 H's of Tyr are 7.15 ppm and 6.86 ppm, respectively (Wüthrich, 1986). There are two sets of doublets representing the Tyr protons, which are close to the expected proton chemical shift, in the ^1H spectrum which display interacting cross-peaks in the 2D-TOCSY spectrum. For reference the sequence of the KKK-Ste2p(14-43)-KKK peptide is KKKDPTYNPGQSTINYTSIYGNGSTITFDELQGKKK.

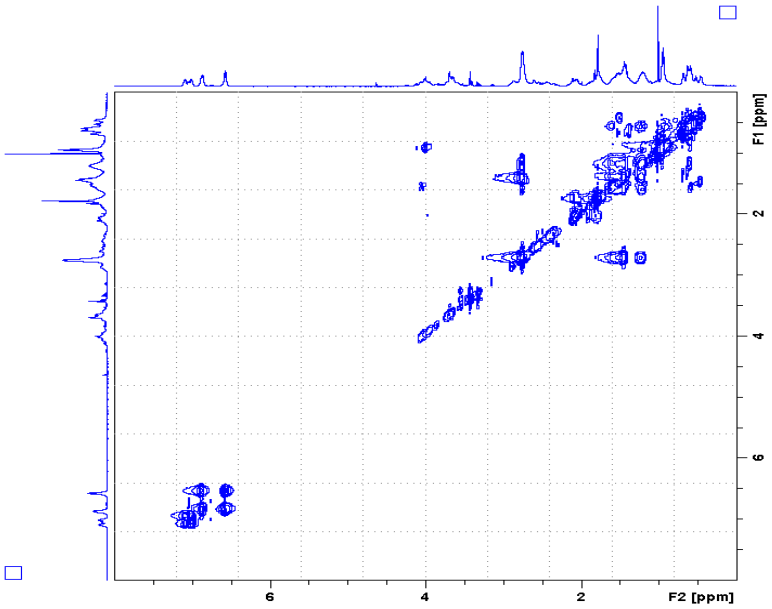

Figure 3.11: 2D ^{1}H-^{1}H TOCSY Spectrum of KKK-Ste2p(14-43)-KKK Peptide.
The TOCSY spectrum of the 3.5 mM KKK-Ste2p(14-43)-KKK peptide in D_2O was obtained in order to observe the through bond interactions between proton residues. The sample was dissolved in 100% D_2O so that the only protons that would be observed in the spectrum are those that are attached to carbon. The spectrum was obtained at 25°C.

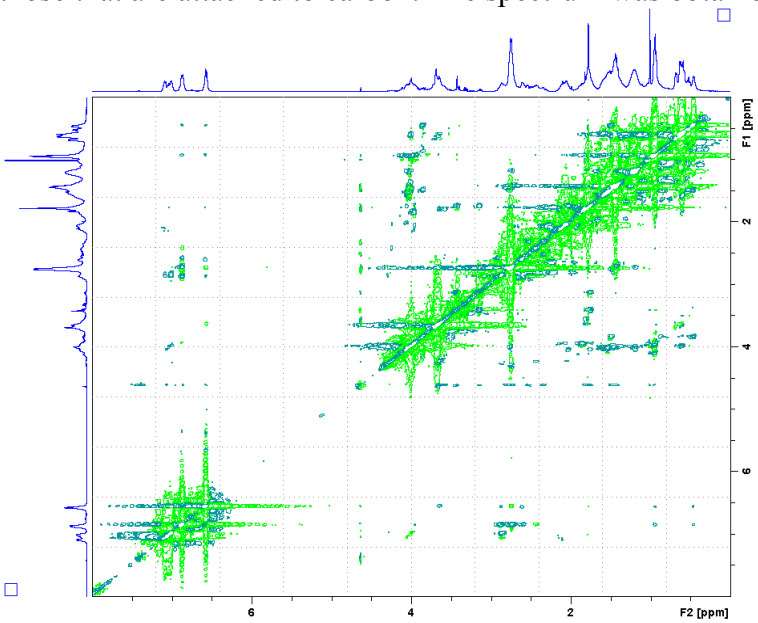

Figure 3.12: 2D ^{1}H-^{1}H NOESY Spectrum of KKK-Ste2p(14-43)-KKK Peptide.
The NOESY spectrum of the 3.5 mM KKK-Ste2p(14-43)-KKK peptide in D_2O was obtained in order to observe through space interactions between protons. The sample was dissolved in 100% D_2O in order to only observe the protons attached to carbon residues. The spectrum was obtained at 25°C.

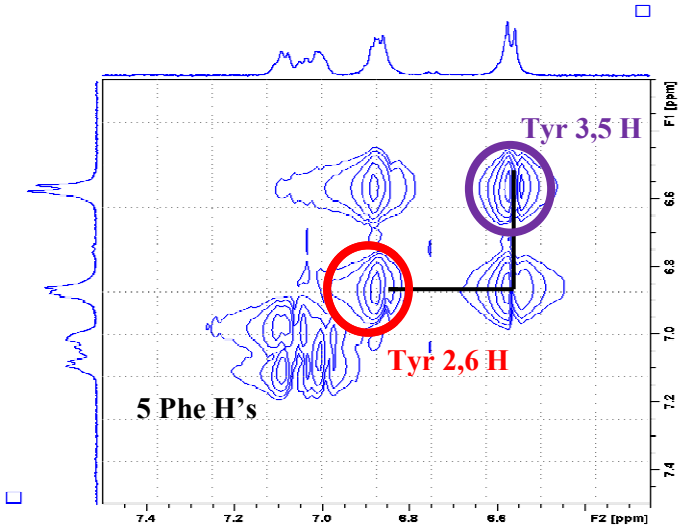

Figure 3.13: Expansion of 2D ^{1}H-^{1}H TOCSY Spectrum of KKK-Ste2p(14-43)-KKK Peptide.
The previous TOCSY spectrum was expanded between 6.25 and 7.5 ppm in order to observe the interactions between the protons of the Tyr residues.

Analysis of the TOCSY and NOESY spectra obtained in D_2O showed a number of cross peaks representing interactions between carbon associated protons in the proteins suggesting some structural information might be obtained for the peptide by NMR analyses. Toward a more complete assignment of residues, the KKK-Ste2p(14-43)-KKK peptide was re-suspended in 90% H_2O/10% D_2O to a final concentration of 4 mM. This was performed to enable observation of the amide protons ($^{1}H^{N}$) and their interactions with alpha carbon protons ($^{1}H^{\alpha}$) using 2D COSY experiments. The reason the $^{1}H^{N}$-$^{1}H^{\alpha}$ interaction is important is because it demonstrates the fingerprint of the amino acid sequence of the peptide. In addition, there is a known number of expected crosspeaks in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region. If the number of actual crosspeaks does not equal the expected number of crosspeaks, it can give information about the homogeneity of the protein. The KKK-Ste2p(14-43)-KKK peptide contains 36 amino acid residues, and using equation 1.5, the total number of expected crosspeaks in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region for the SK peptide is 37.

The 4 mM KKK-Ste2p(14-43)-KKK sample was subjected to NMR analysis. A 2D ^{1}H-^{1}H COSY spectrum was run on the sample (Figure 3.14). The reason the COSY experiment

was chosen (instead of TOCSY) was because it displays interactions through bonds for three bond lengths or less. Therefore, it is an ideal experiment for showing the crosspeak interactions in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region. This particular COSY experiment is double quantum filtered meaning that it only selects those systems with at least two spins, and filters out non-coupled systems such as methyl singlets, therefore simplifying the spectrum. This results in a cleaner spectrum which allows for easier interpretation.

The COSY showed ~30 crosspeaks in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region (Figure 3.15), however, they were not resolved enough to be able to find all of the expected 37 crosspeaks.

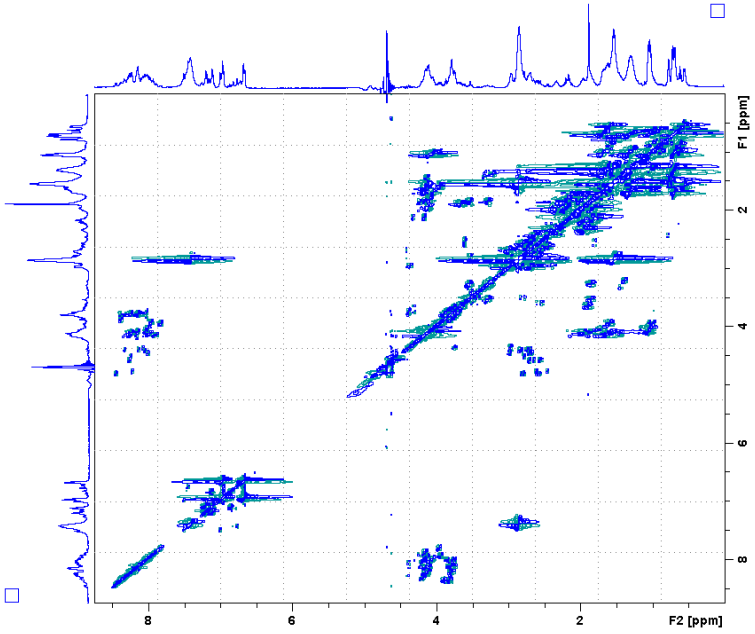

Figure 3.14: 2D ^{1}H-^{1}H COSY Spectrum of KKK-Ste2p(14-43)-KKK Peptide in 90% H_2O/10% D_2O at 25°C.
The COSY spectrum of the 4 mM KKK-Ste2p(14-43)-KKK peptide in 90% H_2O/10% D_2O was obtained in order to observe the through bond proton interactions. More specifically, to see the interaction between the amide protons($^{1}H^{N}$) and alpha carbon protons($^{1}H^{\alpha}$) of the protein. The spectrum was obtained at 25°C.

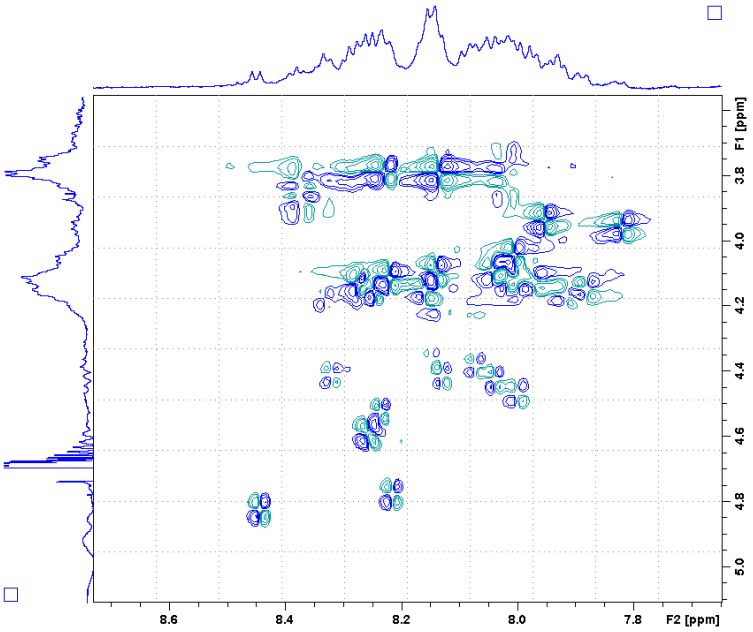

Figure 3.15: Expansion of 2D ^{1}H-^{1}H COSY Spectrum of KKK-Ste2p(14-43)-KKK Peptide.
The previous COSY spectrum was expanded between 3.6 and 5.1 ppm in the F1 and 7.6 and 8.8 ppm in the F2 in order to observe the crosspeaks in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region. The KKK-Ste2p(14-43)-KKK peptide has 36 amino acid residues, which corresponds with 37 $^{1}H^{N}$-$^{1}H^{\alpha}$ crosspeaks.

The spin systems corresponding to the six lysine and five threonine amino acids were approximately assigned using the program Felix (Figure 3.16 and 3.17). Generally, the chemical shifts of the protons were close to predicted values, with variations in the range of 0.1-0.4 ppm, with αH and NH protons shifted more than others (Tables 3.2 and 3.3). The side chain shifts of the six lysine residues were very easy to distinguish because they all produced exactly the same chemical shift which resulted in very large cross peaks. The ε-amino-H of the lysine residues are usually not observed in NMR experiments that use water as a solvent because of rapid exchange with the water. However, they are observed in the COSY spectrum (Figure 3.16) which indicates that they are possibly modified, which is very common with lysine amino groups (Creighton, 1984). In contrast the lysine backbone proton shifts, while not identical, were sufficiently overlapped that identification of individual cross peaks was not possible. Similarly all proton shifts arising from the five threonine residues were not identical, but again so closely overlapping, it was not possible to distinguish individual cross peaks in this spectrum. An example of the overlap is the crosspeak interaction between the γH and the βH of the five threonine residues (Figure 3.17 inset).

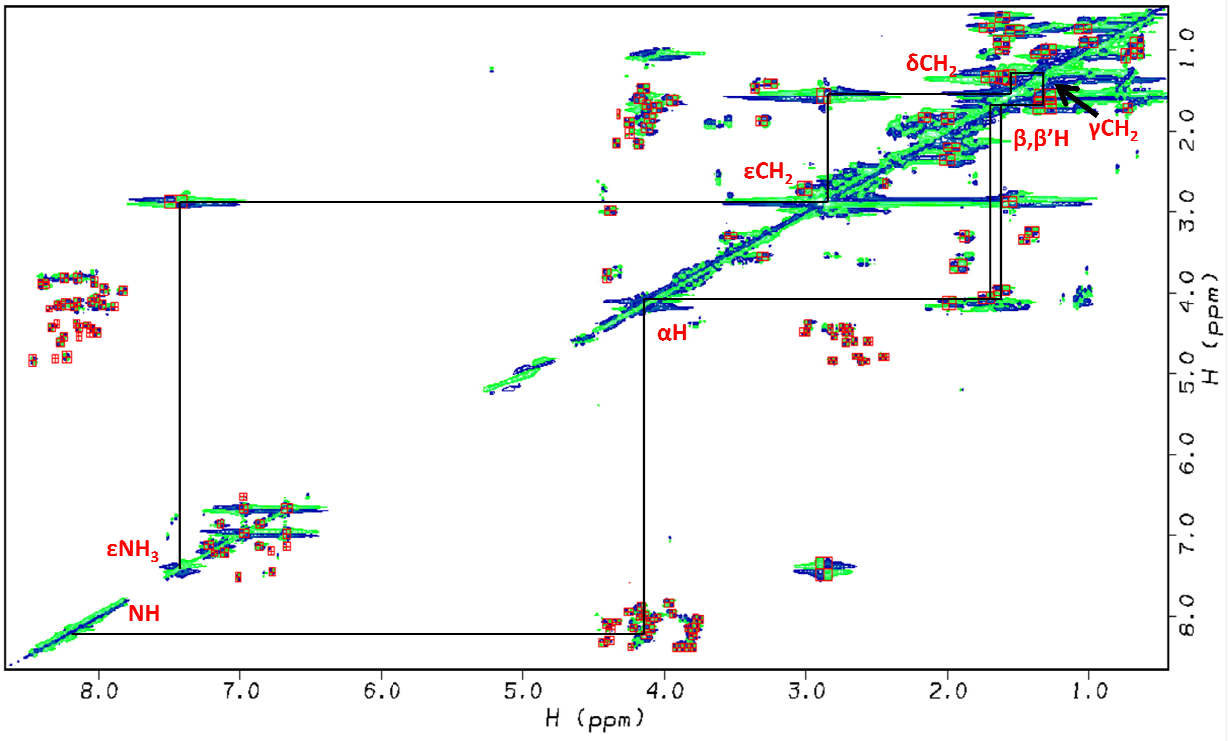

Figure 3.16: Connectivity of Protons from Lysine Residues from 2D ^{1}H-^{1}H COSY Spectrum.
Assignment of the proton connections of lysine residues from the 2D ^{1}H-^{1}H COSY spectrum of 4 mM KKK-Ste2p(14-43)-KKK peptide in 90 % H_2O/ 10% D_2O at 25°C.

Table 3.2: Proton Observed Chemical Shift and Predicted Chemical Shift of the Lysine Amino Acids from KKK-Ste2p(14-43)-KKK.

Proton	Observed Chemical Shift (ppm)	Predicted Chemical Shift (ppm)
εNH_n	7.46	7.5
εCH_2	2.87	3.0
δCH_2	1.56	1.7
γCH_2	1.34	1.4
βCH_2	1.61,1.72	1.8,1.9
αH	3.97-4.07	4.4
NH	8.15-8.29	8.4

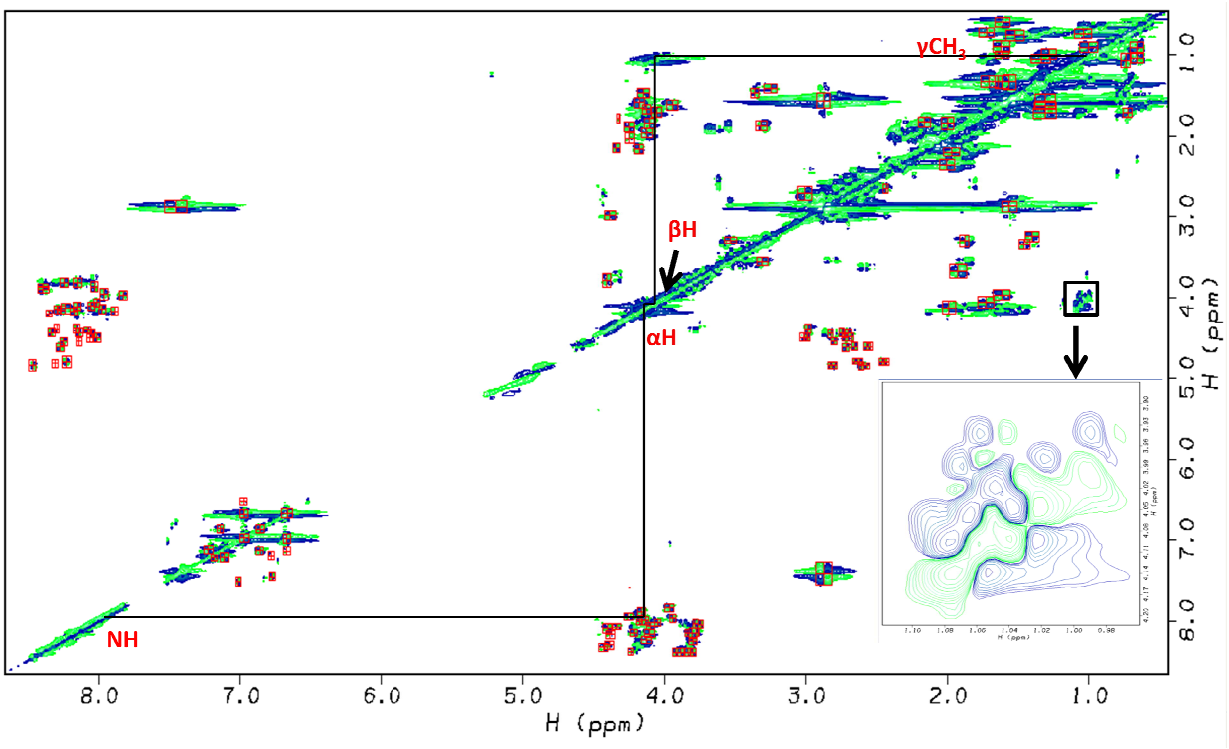

Figure 3.17: Connectivity of Protons from Threonine Residues from 2D ^{1}H-^{1}H COSY Spectrum.
Assignment of the proton connections of threonine residues from the 2D ^{1}H-^{1}H COSY spectrum of 4 mM KKK-Ste2p(14-43)-KKK peptide in 90 % H_2O/ 10% D_2O at 25°C. Inset is expansion of βH-γH crosspeak region.

Table 3.3: Proton Observed Chemical Shift and Predicted Chemical Shift of the Threonine Amino Acids from KKK-Ste2p(14-43)-KKK.

Proton	Observed Chemical Shift (ppm)	Predicted Chemical Shift (ppm)
γCH_3	0.99-1.07	1.2
βH	3.96-4.10	4.2
αH	4.06-4.15	4.4
NH	7.98-8.05	8.2

The NH-αH crosspeak interactions for both the lysine and threonine residues were overlapped so that individual residues were unable to be assigned (Figure 3.18). Use of a higher field magnet for future NMR experiments may be useful to separate these residues.

However, even though a few assignments were able to be made, the proton spectrum in general contained fewer peaks then expected with a lot of broad peaks which indicates that the peptide was possibly aggregated.

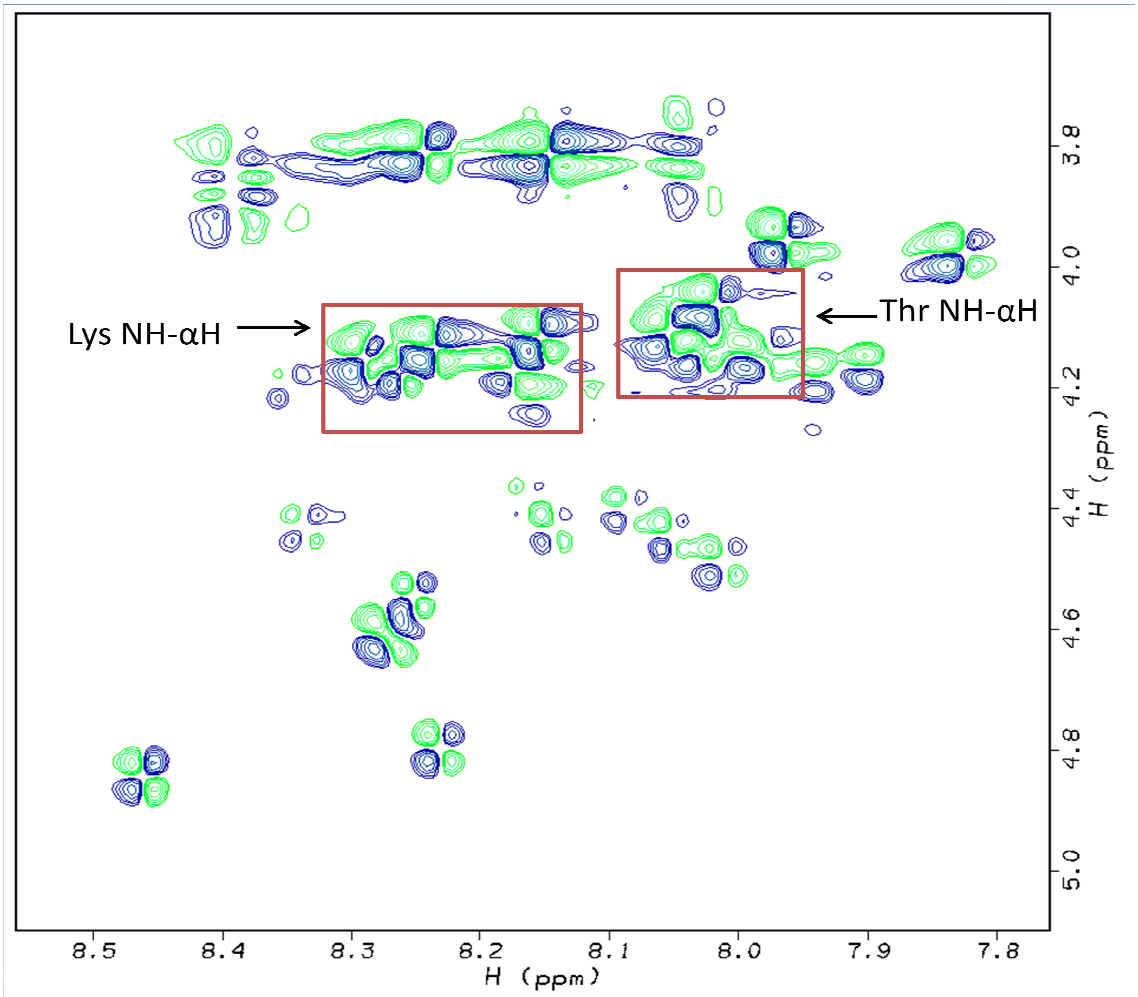

Figure 3.18: Expansion of 2D ^{1}H-^{1}H COSY Spectrum for Identification of Proton Crosspeaks.
The COSY spectrum was expanded between 3.6 and 5.1 ppm in the F1 and 7.8 and 8.5 ppm in the F2 in order to observe the crosspeaks in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region. Assignment of lysine and threonine crosspeaks was performed using Felix.

3.3 KKK-Ste2p(2-48)-KKK-M (SK Peptide)

3.3.1 Molecular Cloning and Mutagenesis of pET31b+/KSI-Ste2p(1-48)-HIS

In light of the success achieved with the synthetic peptide containing the terminal Lys residues, it was decided to shift the focus back to a recombinant Ste2p peptide that could be isotopically labelled to enable heteronuclear NMR experiments for more efficient structural elucidation. Therefore, lysine codons were introduced into a pET31b+/KSI-Ste2p(1-48)-M-HIS construct (already available in the laboratory), which would hopefully allow for the purification of the final recombinant peptide by HPLC. The reason the pET31b+/KSI-Ste2p(1-48)-M-HIS construct was chosen for mutagenesis over the longer Ste2p(1-71) construct that was used initially was to eliminate the portion of the TM1 domain that was included in the later construct, which was most likely increasing the hydrophobicity of the protein and contributing to the purification problems. Therefore, the recombinant studies were continued on a shorter portion containing only the N-terminal residues 1-48. Mutagenesis of the construct was performed to insert three lysine residues between the first methionine residue of the Ste2p peptide and the second residue, serine. Insertion mutagenesis was completed in two sequential steps, with the construct mutated at the 'N-terminal' region first. Upon confirmation of successful insertion at the first site, this mutated construct was prepared and used as the template for the second round of mutagenesis to insert three lysines at the 'C-Terminal' site. After the mutagenesis was completed, the final construct was confirmed by sequencing and named pET31b+/KSI-M-KKK-Ste2p(2-48)-KKK-M-KSI.

3.3.2 Expression and Purification of the SK Fusion Protein

The pET31b+/KSI-M-KKK-Ste2p(2-48)-M-KKK-HIS construct was transformed into BL21-CodonPlus (DE3)-RIL *E. coli* and produced the KSI-M-KKK-Ste2p(2-48)-KKK-M-HIS fusion protein, which from this point forward shall be referred to as the SK fusion protein. A large scale expression was carried out to assess production of the SK fusion protein. The SK fusion protein contains the M-KKK-Ste2p(2-48)-KKK-M peptide with an N-terminal KSI tag and a C-terminal HIS tag theoretically yielding a fusion protein of ~22 kDa (Figure 3.19). After protein expression was induced by IPTG, an aliquot of the culture was removed and separated into the total cell, insoluble and soluble fractions. The remaining culture was

prepared and applied to a Ni-NTA column as described in the methods. The fusion protein was eluted using binding buffer with 1 M imidazole, dialyzed and lyophilized. All samples were analysed using SDS-PAGE (Figure 3.20). The majority of the fusion protein was in the insoluble fraction of the cell lysate (Figure 3.20, lane 3) and the SK fusion protein purified to near homogeneity (Figure 3.20, lane 9). The yield of the final purified, lyophilized SK-fusion peptide was found to be ~70 mg of per litre of culture based on dry protein weight.

Figure 3.19: The SK Fusion Protein.
The SK fusion protein contains an N-terminal KSI tag which forced the fusion protein to become insoluble. The M-KKK-Ste2p(2-48)-KKK-M portion of the fusion protein contains two Met residues at Met1 and a terminal Met residue. The Met residues allowed for cleavage of the fusion protein by CNBr in order to liberate the peptide, KKK-Ste2p(2-48)-KKK-M, also known as the SK peptide from the SK fusion protein. The HIS tag was used for purification of the fusion protein by Ni-NTA resin. The predicted molecular weight of the SK fusion protein is ~22 kDa.

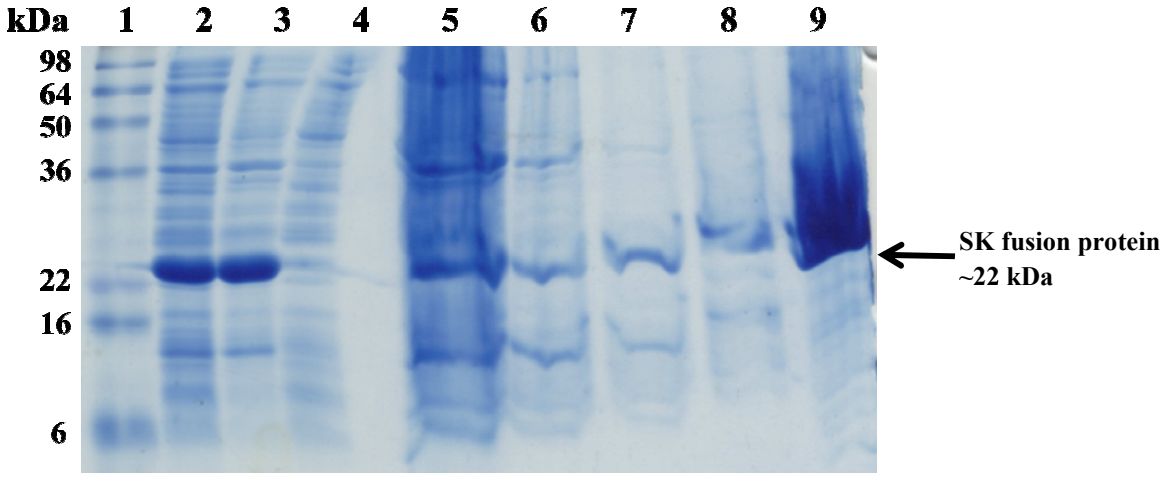

Figure 3.20: Ni-NTA Purification of SK Fusion Protein.
SDS-PAGE was used to analyze the cell extracts of BL21-CodonPlus (DE3)-RIL *E. coli* after large-scale expression to observe the levels of SK fusion protein in the total cell, insoluble and soluble fractions. SDS-PAGE was also used to observe the purity and loss of the SK fusion protein during the purification using Ni-NTA resin. The lanes were loaded as follows: (1) protein marker; (2) total cell extract; (3) insoluble fraction of total cell extract; (4) soluble fraction of total cell extract; (5) flow through fraction of Ni-NTA purification; (6) fraction of binding buffer wash; (7) fraction of binding buffer with 20 mM imidazole wash; (8) fraction of binding buffer with 50 mM imidazole wash; (9) elution fraction from Ni-NTA column.

3.3.3 Cyanogen Bromide Cleavage of SK Fusion Protein

Cyanogen bromide cleavage was performed to digest the SK fusion protein at the C-terminal end of methionine residues. This reaction should produce the peptide of interest KKK-Ste2p(2-48)-KKK-M, which was named the SK peptide. Following CNBr cleavage in 70% TFA, the reaction was terminated and resulting products washed three times by consecutive drying by speed-vaccum and re-suspension in ACN/H_2O (1:1) to make sure that all traces off TFA and CNBr were removed. The dried protein yield was ~70 mg per litre of culture, indicating very little loss of the protein during this step.

3.3.4 HPLC Purification of SK Peptide

HPLC purification was used to isolate the SK peptide from the KSI and HIS tag digested fragments. The lyophilized protein from the CNBr cleavage was dissolved in TFE to produce a clear solution. An equal volume of H_2O was added and no precipitant formed. The final protein concentration was ~10 mg/mL. The sample was injected onto the C18 column in 100 µL volumes and yielded the chromatogram in Figure 3.21 upon elution using an ACN/H_2O gradient system. Fractions for peaks at ~16 min, ~21.5 min, ~25 min, ~31.5 min and ~33 min were collect respectively. The respective peaks were pooled from multiple HPLC runs and lyophilized.

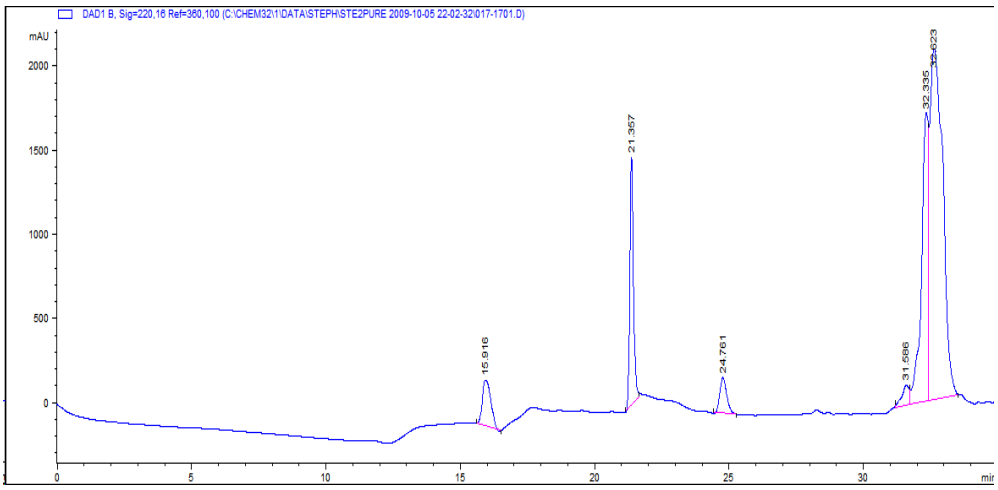

Figure 3.21: HPLC Chromatogram of SK Fusion Protein CNBr Cleavage Products. HPLC was performed in order to purify the SK peptide from the other CNBr cleavage fragments. The chromatogram is at 220 nm. (Reverse-phase HPLC using preparative C18 column with ACN/H_2O + 0.1% TFA solvent system)

3.3.5 Mass Spectrometry of SK Peptide Purification

After drying down the HPLC peak samples, there was ~11.5 mg of the 33 min peak, but only ~0.5 mg of the 16, 21.5, 25 or 31.5 min peaks. Each of these was re-suspended in ACN/H_2O at a final concentration of 1 mg/mL and sent for MALDI-TOF MS analysis. The MALDI-TOF spectrum of the 33 min peak contained two prominent peaks (Figure 3.22). The first peak at 5885 m.u. corresponded to the SK peptide, while the second peak at 6142 m.u. corresponded to the LM-KKK-Ste2p(2-48)-KKK-M peptide (LM-SK-peptide; Table 3.2). The only difference between these two peptides was that the larger peptide had one missed cleavage and contained two extra amino acids, one leucine and one methionine. The ratio of these peptides in the dried fraction was approximately 70% SK peptide and 30% LM-SK peptide based on the MS peak profiles. The final dried peptide weight was 11.5 mg, which taking molecular weight differences into consideration, represented a 38% overall yield of SK peptide from purified SK fusion protein.

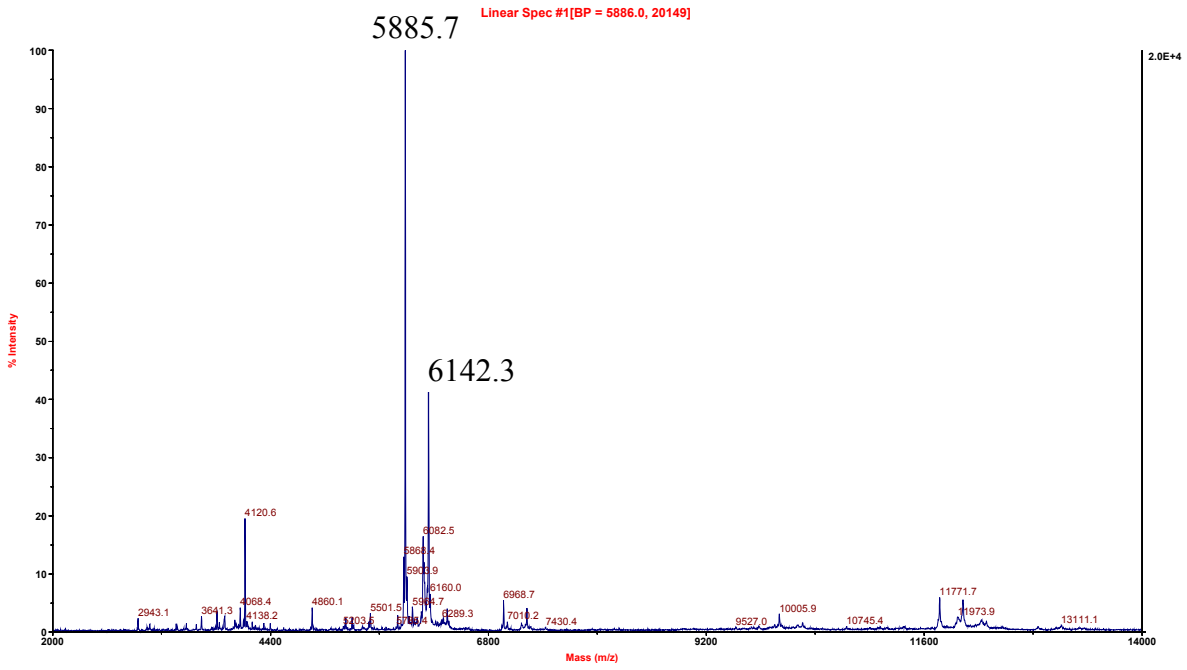

Figure 3.22: MALDI-TOF MS of 33 min Peak from HPLC Purification.
After drying down the 33 min peak from the HPLC purification in Figure 3.21, the sample was dissolved in ACN/H_2O and sent for MS analysis. MALDI-TOF MS was performed on the 33 min peak in order to determine the mass of the peptide that was purified off the column.

Table 3.4: Predicted Mass of Peptides after Cyanogen Bromide Cleavage of SK Fusion Protein.

Peptide	Mass	Sequence	Missed cleavage	Modifications
SK	**5888.56**	KKKSDAAPSLSNLFYDPTYNPGQSTINYT SIYGNGSTITFDELQGLVNSTKKKM(Hsl)	**0**	
LM- SK	**6129.08**	LMKKKSDAAPSLSNLFYDPTYNPGQSTI NYTSIYGNGSTITFDELQGLVNSTKKKM(Hsl)	**1**	
LM- SK	**6145.08**	LMKKKSDAAPSLSNLFYDPTYNPGQSTI NYTSIYGNGSTITFDELQGLVNSTKKKM(Hsl)	**1**	**1 oxidation**
KSI	**13416.9**	HTPEHITAVVQRFVAALNAGDLDGIVAL FADDATVEDPVGSEPRSGTAAIREFYAN SLKLPLAVELTQEVRAVANEAAFAFTVSF EYQGRKTVVAPIDHFRFNGAGKVVSIRA LFGEKNIHACQM(Hsl)	**0**	
His tag	**1083.49**	LEHHHHHH	**0**	

3.3.6 Dynamic Light Scattering of the Recombinant SK Peptide

Based on the NMR results from the synthetic KKK-Ste2p(14-43)-KKK peptide, it was possible that the SK peptide might be aggregated. Therefore, DLS was carried out in order to determine if the purified SK peptide was prone to aggregation. The dried SK peptide from the 33 min HPLC peak was re-suspended in either TFE/H_2O, $CHCl_3$:CH_3OH:H_2O, 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 or 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 and OG detergent to a final concentration of 0.2 mM. OG was used during DLS instead of the previously used LPPG because it was determined that a deuterated form of LPPG was no longer commercially available. If in future NMR experiments detergents were required then a deuterated form would be needed. Therefore, OG was selected to be used for DLS because a deuterated form was available for future use. The TFE/H_2O and $CHCl_3$:CH_3OH:H_2O samples solubilised to produce a clear liquid. However, both of the H_2O samples had a large amount of precipitate.

All of the samples were filtered through a 0.2 µm filter and analysed using DLS. The TFE/H_2O sample showed no aggregates (Figure 3.23), with the majority of the sample having a radius of 1 nm, corresponding to a molecular weight close to the SK peptide. The $CHCl_3$:CH_3OH:H_2O sample was unable to be analyzed by DLS due to the solvent background being too high. The 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 sample showed considerable aggregation (Figure 3.24), with most of the protein having a radius in the highest range. The 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 and OG sample yielded a radius of ~0.45 nm (Figure 3.25), which corresponds to the buffer, meaning that there was little to no peptide in the sample. Therefore, the SK peptide was most likely aggregated and filtered out of the sample in this last condition.

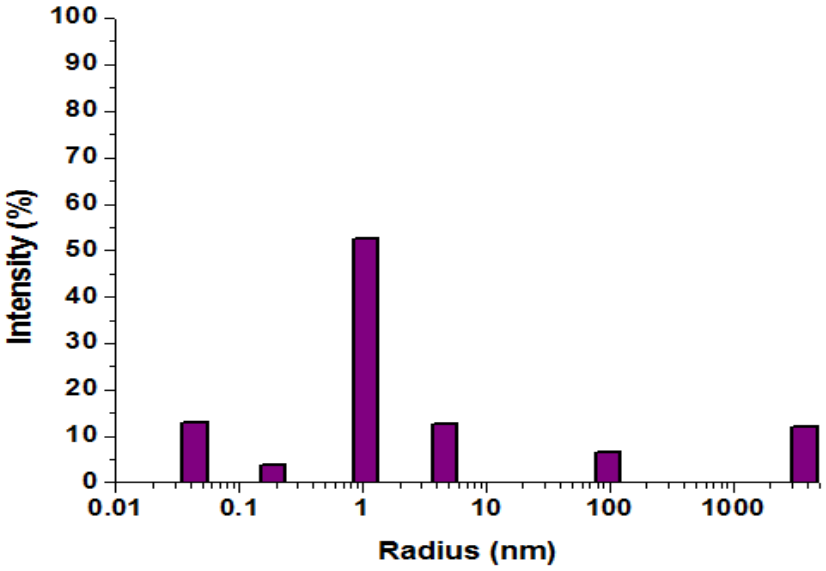

Figure 3.23: DLS of the SK Peptide in TFE/H_2O (1:1) (+0.1% TFA).
DLS was performed on the SK peptide in order to determine if the peptide was aggregated in TFE/H_2O with TFA. The hydrodynamic radius of the sample was measured and this is an indicator of the approximate molecular weight of the molecules in the solution.

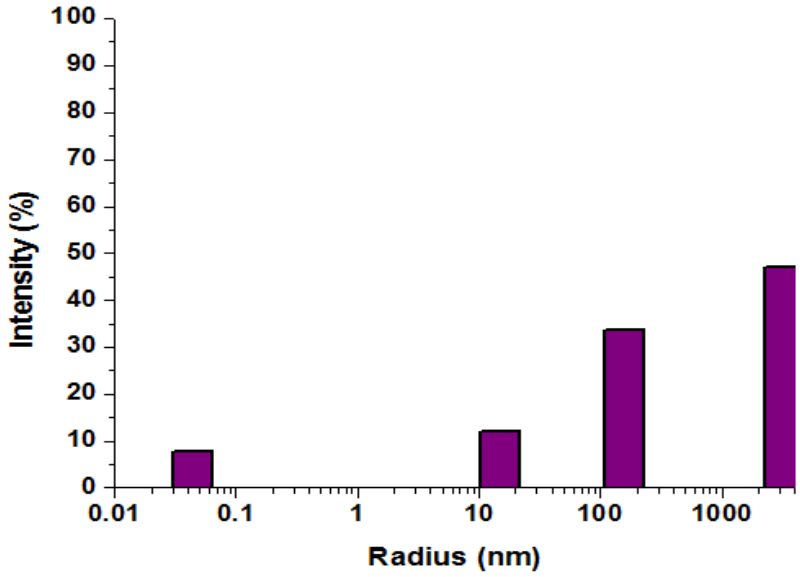

Figure 3.24: DLS of the SK Peptide in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0.
DLS was performed on the SK peptide in order to determine if the peptide was aggregated in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0. The hydrodynamic radius of the sample was measured and this is an indicator of the approximate molecular weight of the molecules in the solution.

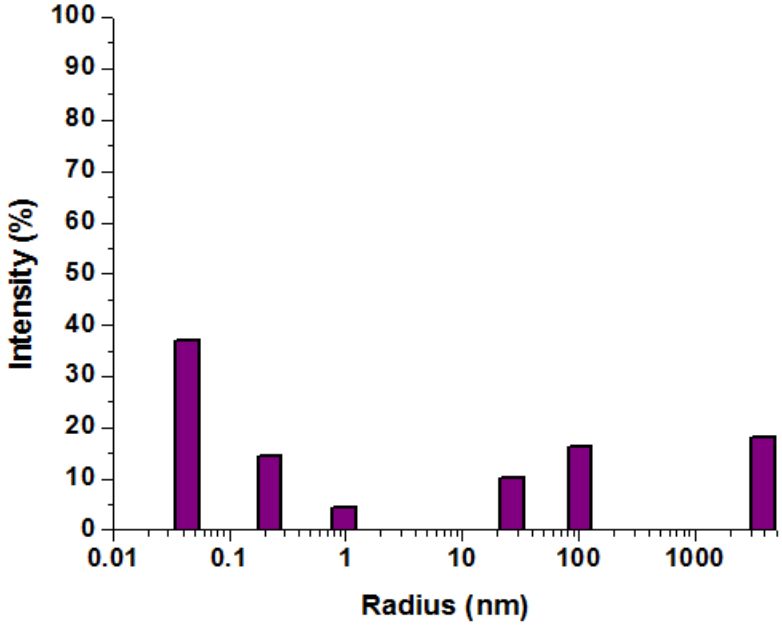

Figure 3.25: DLS of the SK Peptide in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 and 0.1% OG.
DLS was performed on the SK peptide in order to determine if the peptide was aggregated in 10 mM Na_2HPO_4/NaH_2PO_4 pH 7.0 and OG. The hydrodynamic radius of the sample was measured and this is an indicator of the approximate molecular weight of the molecules in the solution.

3.3.7 Circular Dichroism of the Recombinant SK peptide

Circular dichroism was performed on the SK peptide in order to assess the secondary structure of the peptide in different solvents. Initially, the same four solvents were used as in the DLS experiments. The samples were prepared following previously described methods (Estephan *et al.*, 2005). However, after filtering the samples, no CD signal was observed for either of the H_2O samples suggesting there was no protein in solution after filtering. This correlates with DLS results that suggested that they were most likely both aggregated and therefore the peptide was filtered out of the samples. Also, the $CHCl_3$:CH_3OH:H_2O sample was not able to be analyzed using CD because the chloride in the solvent absorbs in the low UV region making it impossible to see the CD spectrum of the SK peptide in this region. Therefore, the only sample that CD was performed on was the SK peptide dissolved in TFE/H_2O (+0.1% TFA). The sample was run on the π*-180 instrument. The ellipticity (θ) of the samples were converted to mean residue ellipticity ($[\theta]_{MR}$) and plotted against wavelength (Figure 3.26). The spectrum shows that the peptide was able to be solubilised in the TFE/H_2O and the overall shape of the spectrum is similar to that of the synthetic KKK-Ste2p(14-43)-KKK peptide. The CD spectrum in figure 3.26 displays peaks at 208 and 222 nm which indicates the presence of α–helical structure, which was similar to what was seen for the synthetic peptide in TFE/H_2O in figure 3.10B.

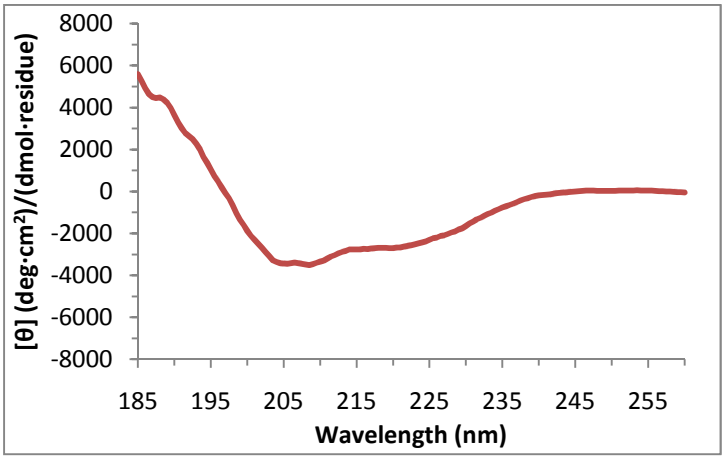

Figure 3.26: CD Spectrum of the SK Peptide in TFE/H_2O (0.1% TFA).
The CD spectrum of the SK peptide was acquired in order to determine the secondary structure of the peptide in TFE/H_2O with TFA. The spectrum was obtained at 25°C and the SK peptide concentration was ~0.2 mM.

The ellipticity and wavelength were imported into the CDNN CD spectra deconvolution software for structural analysis. The predicted secondary structure obtained from the software for the SK peptide in TFE/H_2O (1:1) (+0.1% TFA) was ~52% β-strand and 6% α-helix with a predicted beta-turn (Table 3.3). When compared to the predicted structure from CD analysis of the KKK-Ste2p(14-43)-KKK in TFE/H_2O, which was ~ 26% β-strand and 13% α-helix with an expected beta-turn, it indicated that a larger portion of the longer recombinant SK peptide was β-strand with relatively less α-helix.

Table 3.5: Predicted Secondary Structure of SK Peptide in TFE/H_2O (1:1) (+0.1% TFA) Using CDNN CD Spectra Deconvolution Software.

	190-260 nm
Helix	**6.30%**
Antiparallel	**48.80%**
Parallel	**3.70%**
Beta-Turn	**16.70%**
Rndm. Coil	**29.00%**
Total Sum	**104.50%**

3.3.8 Nuclear Magnetic Resonance of the Recombinant SK Peptide

Based on the DLS and CD data, TFE was chosen as a suitable solvent for NMR analysis of the SK peptide. The SK peptide was re-suspended in 50% TFE-d2/40% H_2O/10% D_2O (0.1% TFA) to a final concentration of 4 mM. This solvent yielded a clear solution of the peptide for NMR analysis. The SK peptide has 54 amino acid residues which theoretically should yield 54 crosspeaks in the $^1H^N$-$^1H^\alpha$ region. A 1D-^{1}H spectrum was run on the SK peptide sample (Figure 3.27). The spectrum displayed a number of sharp peaks; however the peaks were broad and lacking in number for the size of the peptide. A typical 1D-^{1}H spectrum for a peptide of approximately the same size would contain a larger number of sharp peaks, such as the 1D-^{1}H spectrum of the 47-residue peptide of Factor IX (Figure 3.28). None the less, a 2D ^{1}H-^{1}H COSY spectrum was run on the sample (Figure 3.29). The COSY showed almost no crosspeaks in the $^1H^N$-$^1H^\alpha$ region (around 7 or 8 ppm), which suggests that the

peptide may be aggregated at 4 mM concentration. The DLS and CD experiments were performed at a concentration of 0.2 mM and in theory this is the only variable between the experiments. Therefore, the high concentration may be contributing to the aggregation.

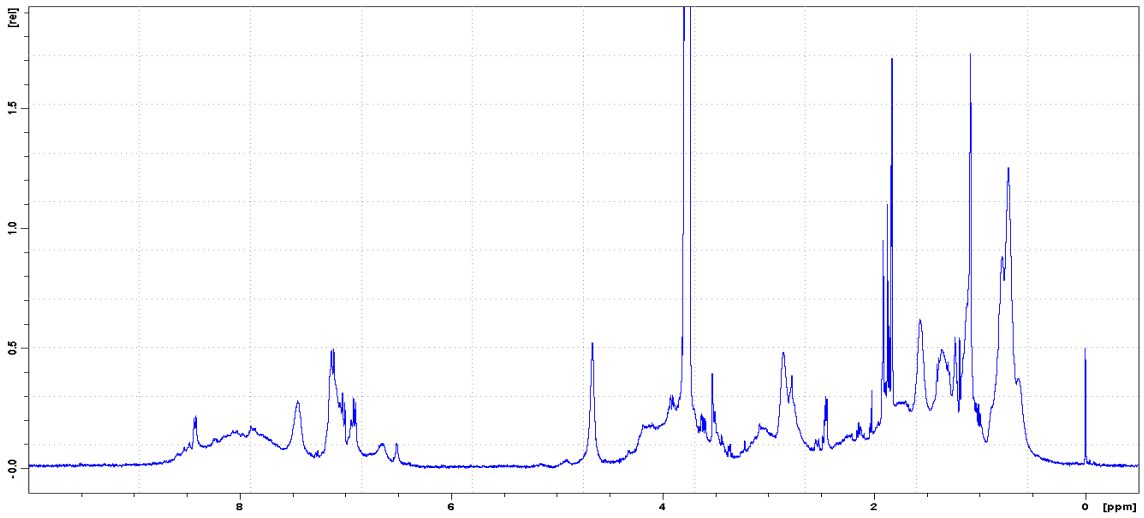

Figure 3.27: 1D ^{1}H Spectrum of SK Peptide.
The 1D proton spectrum of the 4 mM SK peptide was obtained in 50% TFE-d2/40% H_2O/10% D_2O in order to observe the total number of proton peaks for the peptide. The spectrum was obtained at 25°C.

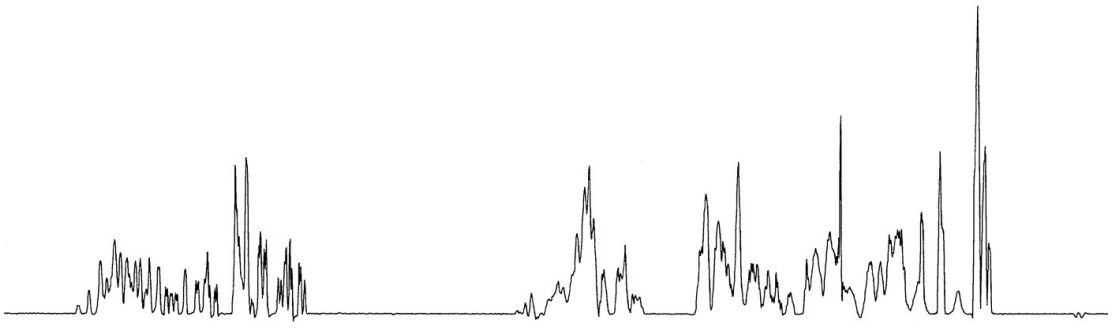

Figure 3.28: 1D-^{1}H spectrum of Factor IX residues 1-47.
The 1D proton spectrum of residues 1-47 of Factor IX was obtained using a peptide concentration of 2.1 mM in H_2O, pH 5.2 at 25°C. This spectrum was taken from (Freedman *et al.*, 1995).

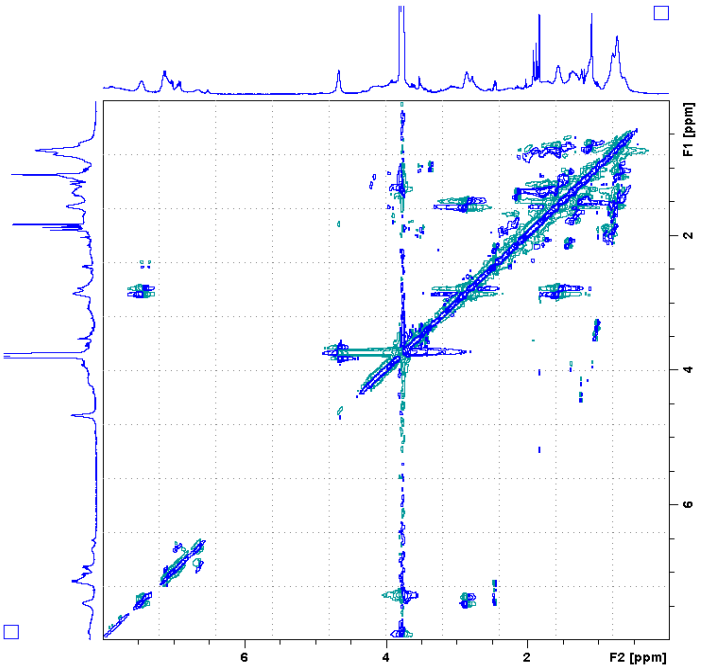

Figure 3.29: 2D ^{1}H-^{1}H COSY Spectrum of SK Peptide.
The COSY spectrum for the 4 mM SK peptide in 50% TFE-d2/40% H_2O/10% D_2O was acquired in order to observe the through bond proton interactions in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region. The spectrum was obtained at 25°C.

3.4 SK Peptide in Enriched ^{15}N M9 Minimal Medium

3.4.1 Purification of SK Fusion Protein from M9 Minimal Medium

In order to enable NMR structural analyses at lower protein concentrations, the pET31b+/KSI-M-KKK-Ste2p(2-48)-M-KKK-HIS construct was transformed into BL21-CodonPlus (DE3)-RIL *E. coli* and a large scale expression in M9 minimal medium supplemented with ^{15}N-labelled ammonium sulphate was performed in order to produce an ^{15}N-labelled SK fusion protein. Labelling with ^{15}N allows 2D heteronuclear ^{15}N-^{1}H NMR experiments to be performed, which facilitiates structural resolution from lower concentration protein samples. After expression of the SK fusion protein in the ^{15}N-labelled M9 minimal medium, the fusion protein was purified as in Section 2.6.1. The total dried ^{15}N-SK fusion protein amount was ~47 mg/L of M9 minimal medium. The ^{15}N-SK fusion protein was cleaved using cyanogen bromide as in Section 2.6.2 and dried down using the speed-vacuum.

3.4.2 HPLC Purification of ^{15}N-labelled SK peptide

HPLC purification was used to isolate the ^{15}N-SK peptide in the ^{15}N-SK fusion protein from other digested fragments. The lyophilized protein from the CNBr cleavage was dissolved in TFE to produce a clear solution. An equal volume of H_2O was added and no precipitant formed. The final protein concentration was ~10 mg/mL. The sample was injected onto the hydrophobic reverse phase C18 column in 100 µL volumes and yielded the chromatogram in Figure 3.30 upon elution using an ACN/H_2O gradient system. Fractions for peaks at ~13 min, ~16 min, ~20 min, ~27 min and ~35 min were collect respectively. The respective peaks were pooled from multiple HPLC runs and lyophilized.

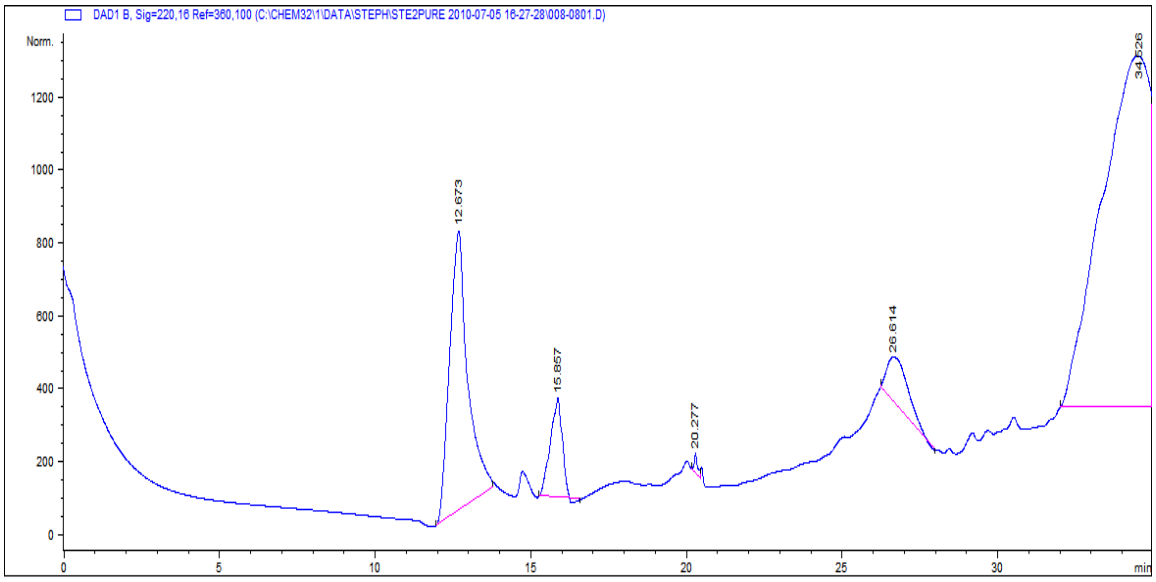

Figure 3.30: HPLC Chromatogram of ^{15}N-labelled SK Fusion Protein CNBr Cleavage Products.
HPLC was performed in order to purify the ^{15}N-SK peptide from the other CNBr cleavage fragments. The chromatogram is at 220 nm.

3.4.3 Mass Spectrometry of ^{15}N-labelled SK Peptide Purification

After drying down the HPLC peak samples, there was no protein from the 13, 16, 20, and 27 min peaks. However, the 35 min peak contained 9.5 mg of protein. The 35 min peak was re-suspended in ACN/H_2O at a final concentration of 1 mg/mL and sent for MALDI-TOF MS analysis. The MALDI-TOF spectrum of the 35 min peak contained a number of prominent peaks in the range from 3000-4000 mass units (Figure 3.31). This range is considerably lower than the expected mass of ^{15}N-labelled SK peptide, which is ~5950 mass units. There are a few peaks around 5950-6027 m.u. but they are very small and not the prominent peptides in the sample. Therefore, it is possible that this sample had degraded prior to MS analysis.

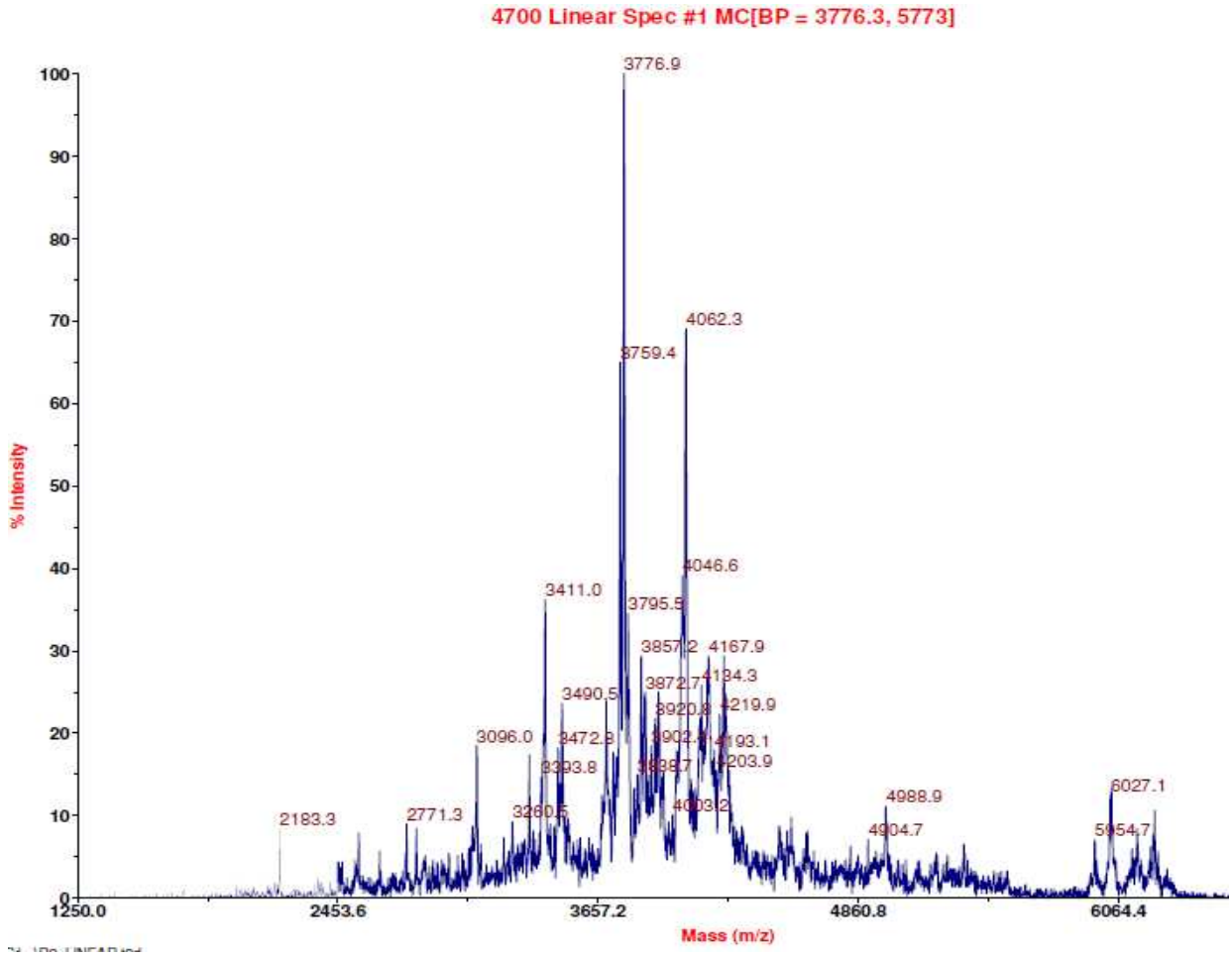

Figure 3.31: MALDI-TOF MS of 35 min Peak from HPLC Purification.
After drying down the 35 min peak from the HPLC purification in Figure 3.30, the sample was dissolved in ACN/H_2O and sent for MS analysis. MALDI-TOF MS was performed on the 35 min peak in order to determine the mass of the peptide that was purified off the column.

3.4.4 Nuclear Magnetic Resonance of ^{15}N-labelled SK Peptide

The ^{15}N-labelled SK peptide from the 35 min HPLC peak was re-suspended in 50% TFE-d2/40% H_2O/10% D_2O (0.1% TFA) to a final peptide concentration of ~3 mM. A 2D ^{1}H-^{15}N HSQC was run on the sample in order to observe interactions between the nitrogens and their attached protons (Figure 3.32). While the sample produced a number of clear interactions, there was considerable overlap for the most part, suggesting aggregation or degradation. This correlates with the observed degradation identified by previous MS analysis.

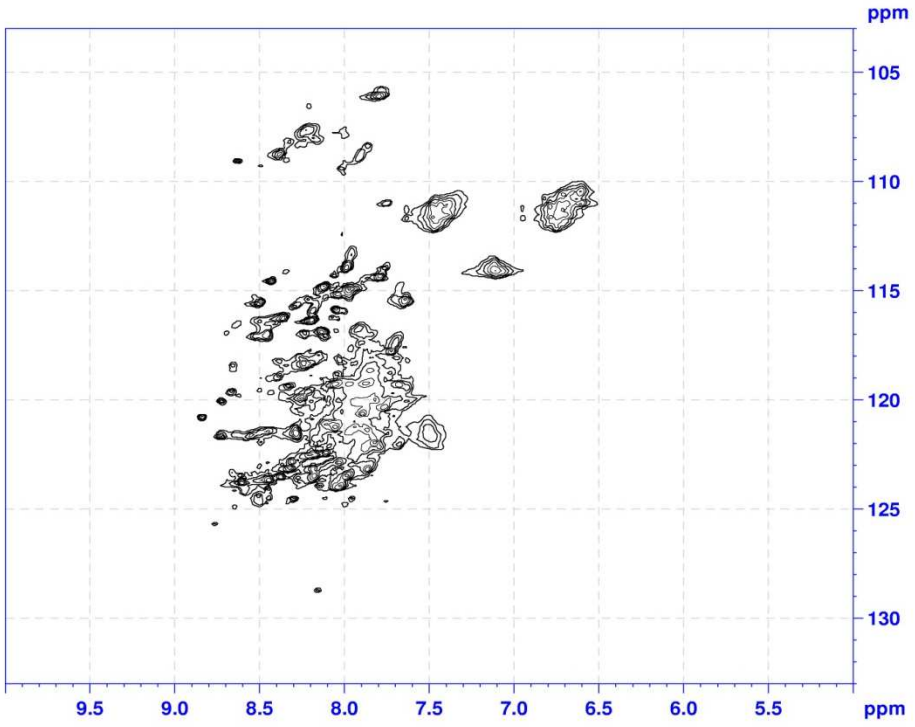

Figure 3.32: 2D ^{1}H-^{15}N HSQC Spectrum of ^{15}N-labelled SK Peptide.
The ^{1}H-^{15}N HSQC spectrum was obtained using 3 mM ^{15}N-labelled SK peptide in 50% TFE-d2/40% H_2O/10% D_2O in order to observe all proton and nitrogen interactions. The spectrum was obtained at 25°C.

4.0 DISCUSSION

The goals of this project were to obtain purified mg quantities of a soluble version of a portion of the N-terminus of Ste2p and to acquire structural information about this region by performing biophysical analysis on the soluble N-terminal Ste2p fragments. Initially, residues 2-54 of Ste2p were purified as part of a fusion protein and cleavage of this protein produced Ste2p(2-54) peptide. Unfortunately, the Ste2p(2-54) peptide was unable to be purified using HPLC. Therefore, a peptide containing residues 14-43 was synthesized but again was unable to be purified using HPLC. However, the addition of three lysine residues to both ends of the peptide allowed for synthesis and purification of the KKK-Ste2p(14-43)-KKK peptide. This peptide was subjected to biophysical analysis which yielded CD spectra that indicated peptide secondary structure and initial NMR experiments that were promising. Following this, a shorter recombinant version of the N-terminus of Ste2p, containing residues 2-48 with the additional terminal lysine residues, was successfully expressed as a fusion protein in *E. coli*. The fusion protein was effectively cleaved in order to liberate the Ste2p N-terminal peptide fragment, SK peptide. The SK peptide was purified, in mg quantities, from the other cleavage products using HPLC and these results were confirmed using MS analysis. DLS analysis determined that the SK peptide was aggregated when dissolved in H_2O but not aggregated in TFE/H_2O. CD analysis of the SK peptide when it was dissolved in TFE/H_2O showed that it demonstrated secondary structure. NMR analysis of both the unlabelled and ^{15}N-labelled peptide was performed on the SK peptide.

4.1 Ste2p(2-54) Peptide Purification

Originally, a longer N-terminal fragment of Ste2p was going to be used for structural analysis based on naturally occurring methionines as cleavage sites and the concept of 'the longer the peptide, the more biologically relevant the resulting structure'. After purification of the fusion protein containing residues 1-71 from Ste2p, the fusion protein was cleaved using cyanogen bromide and attempts to separate the fragments were made using HPLC. However, after many experiments were performed, the Ste2p(2-54) peptide fragment continually was unable to be eluted off the HPLC column. This issue was very difficult to resolve and

approximately a year was focused on trying to determine why the Ste2p(2-54) peptide was lost during HPLC purification. A number of factors were tested to determine if they affected the HPLC purification. A few of these were changing the HPLC gradient solvents, as well as CNBr cleavage solvents and reaction length. Also a number of solvents were tested to solubilise the protein before loading onto the HPLC column. However, none of these changes resulted in the successful purification of the peptide by HPLC. This may be due to the fact that the Ste2p(2-54) peptide contained residues 50 - 54 which are predicted to be part of transmembrane 1 of Ste2p. Of these residues, Ala52, Ile53 and Met54 are considered hydrophobic at pH 7.0 with Thr50 and Gln51 being considered neutral (Monera *et al.*, 1995). A hydrophobicity plot of the Ste2p(2-54) peptide confirmed that both ends of the peptide contained amino acid residues that were quite hydrophobic (Figure 4.1). Therefore the peptide was possibly too hydrophobic to be purified using a reverse-phase C18 column.

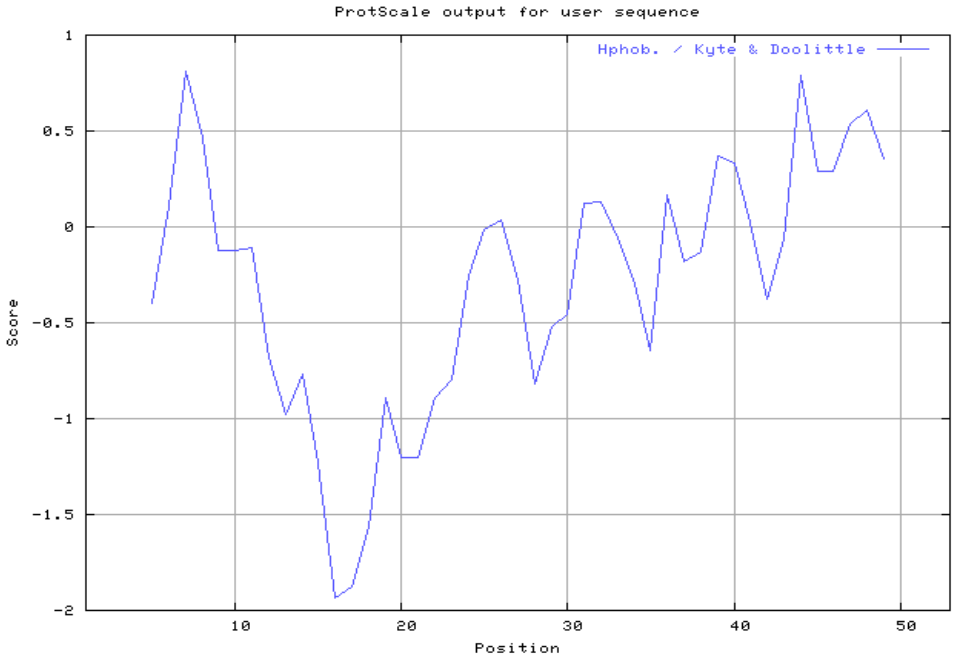

Figure 4.1: Hydrophobicity Plot of the Ste2p(2-54) Peptide.
The hydrophobicity of the Ste2p(2-54) peptide was analyzed using ProtScale and based on the Kyte and Doolittle Hydropathicity model (Kyte and Doolittle, 1982). Areas above 0 are considered to be hydrophobic and areas below 0 are considered to be hydrophilic.

4.2 Synthetic Ste2p(14-43) Structural Analysis

After initial experiments with the Ste2p(1-71) peptide were unsuccessful, it was decided to ask CanPeptide Inc. to synthesize a shorter version of the Ste2p peptide for initial structural analysis. Residues 14-43 from Ste2p were chosen to be synthesized because it encompassed a portion of the N-terminal end of Ste2p which has been previously shown to have helical structure (Neumoin *et al.*, 2009), as well as residues Pro15, Ile24 and Ile29 where previous mutagenesis of these residues lead to significant loss in mating efficiency while maintaining wild-type levels of signalling (Shi *et al.*, 2009a). A shorter version of the peptide was also more cost-effective to produce and easier to synthesize.

CanPeptide Inc. was able to synthesize the Ste2p(14-43) peptide, however, when attempting to purify the synthesized peptide using HPLC, the peptide continually disappeared. These results were similar to what had happened with the Ste2p(2-54) peptide. A hydrophobicity plot of the Ste2p(14-43) peptide demonstrated that portions of the C-terminal end of the peptide were hydrophobic (Figure 4.2). This led us to believe that both of the peptides were too hydrophobic to be purified using the HPLC. It has been previously shown that the addition of lysine residues to both termini of a peptide allowed for the production of a less hydrophobic peptide (Xie *et al.*, 2000). Despite its hydrophobic chain of four methylene groups, lysine has an amino group with a pI of ~10. Thus at pH's less than 10 the amino group is ionized producing a very polar charged amino acid that can interact with aqueous solution yielding a much more soluble overall protein. At the same time In this previous study, after the addition of six lysine residues, the peptide was able to be purified using HPLC and used for structural analysis. It was found that the addition of the lysine residues to both ends of the peptide had no significant effect on the helix formation of the residues in the peptide. However, more research needs to be done to explore the effect of the lysine residues on other domains. Therefore, CanPeptide Inc. was asked to synthesize the KKK-Ste2p(14-43)-KKK peptide which would have three lysine residues before residue 14 and three lysine residues after residue 43. After synthesizing the KKK-Ste2p(14-43)-KKK peptide CanPeptide Inc. was able to purify the peptide using HPLC. Therefore, the addition of the lysine residues resulted in a less hydrophobic protein that could be purified, as predicted by the hydrophobicity plot of the KKK-Ste2p(14-43)-KKK peptide (Figure 4.3).

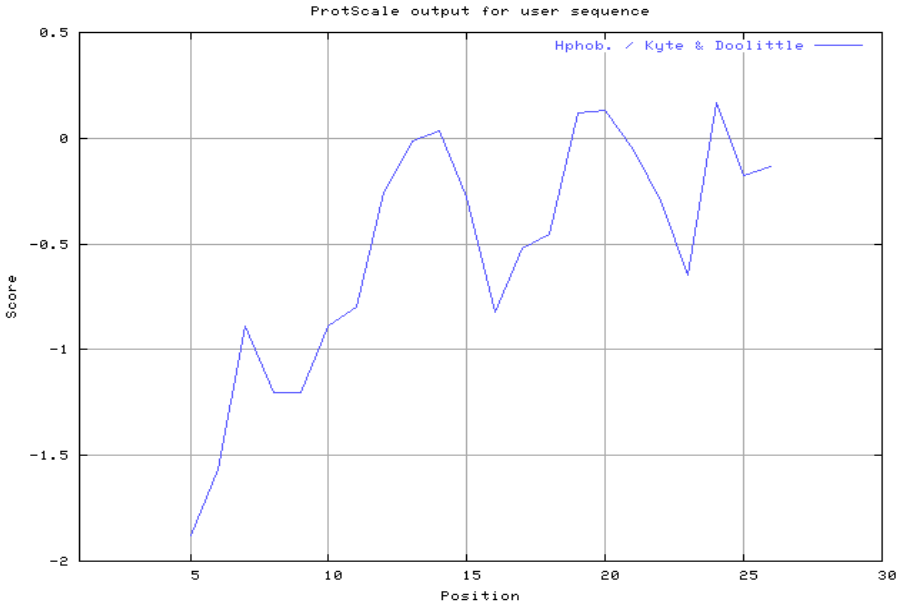

Figure 4.2: Hydrophobicity Plot of the Ste2p(14-43) Peptide.
The hydrophobicity of the Ste2p(14-43) peptide was analyzed using ProtScale and based on the Kyte and Doolittle Hydropathicity model (Kyte and Doolittle, 1982). Areas above 0 are considered to be hydrophobic and areas below 0 are considered to be hydrophilic.

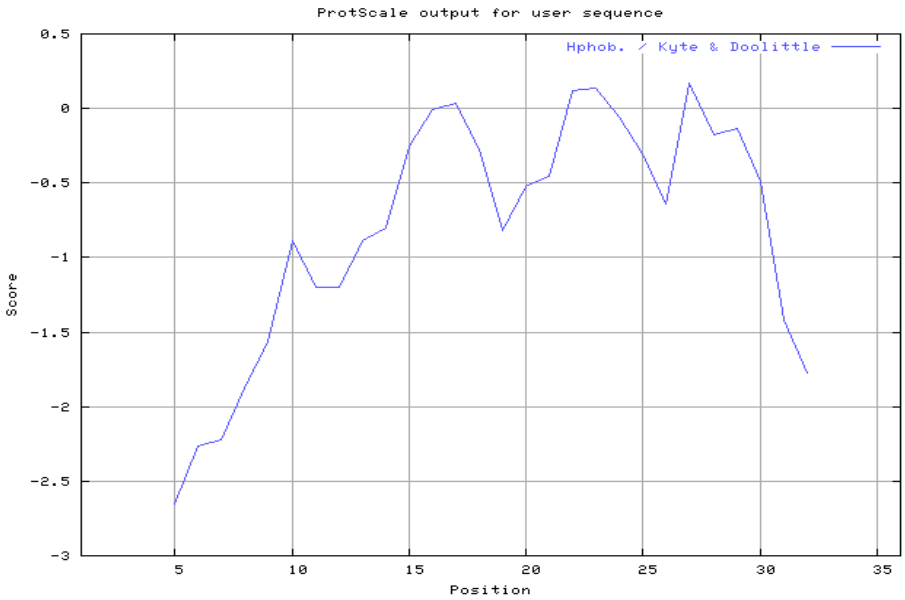

Figure 4.3: Hydrophobicity Plot of the KKK-Ste2p(14-43)-KKK Peptide.
The hydrophobicity of the KKK-Ste2p(14-43)-KKK peptide was analyzed using ProtScale and based on the Kyte and Doolittle Hydropathicity model (Kyte and Doolittle, 1982). Areas above 0 are considered to be hydrophobic and areas below 0 are considered to be hydrophilic.

Initial structural analysis was performed on the KKK-Ste2p(14-43)-KKK peptide in order to determine the solubility of the peptide in various solutions toward developing optimal conditions for NMR experiments. Circular dichroism analysis of the KKK-Ste2p(14-43)-KKK peptide indicated secondary structure when dissolved in three different solvents, H_2O, TFE/H_2O and H_2O with the detergent LPPG. The peptide was also denatured at 60°C when in H_2O and TFE and at 40°C when in H_2O/LPPG. The secondary structure was predicted to be ~ 42% β-strand and 9% α-helix and ~46% β-strand and 9% α-helix, for H_2O and LPPG, respectively, while the predicted secondary structure in the TFE sample was ~26% β-strand and 13% α-helix. These percentages are based on computational analysis and cannot be taken as accurate numbers. Computational analysis is limited by the fact that structure content is estimated based on averaging given sets of control/known samples. However, while the absolute values cannot be taken as accurate, relative changes between similar samples and between different conditions are meaningful for general comparisons. The fact that there were higher amounts of α-helix in the TFE/H_2O sample is not surprising due the fact that TFE is known to induce helical content (Sonnichsen *et al.*, 1992). Overall, the predicted amounts of ~42% β-strand and 9% α-helix were consistent with a previously predicted β-strand loop β-strand fold (Shi *et al.*, 2009a) with the C-terminal residues (39-43) taking on an α-helical structure (Neumoin *et al.*, 2009). As well the KKK-Ste2p(14-43)-KKK peptide contained a number of amino acids that have been shown to prefer β-strand conformation, such as Thr16, Tyr17, Gln21, Thr23, Ile24, Tyr26, Thr27, Ile29, Tyr30, Thr35, Ile36, Thr37, Phe38, and Gln 42 (Chou and Fasman, 1978). Therefore, it is not surprising that a large portion of the synthetic peptide was predicted to form β-strand.

Initial NMR analysis was performed on the KKK-Ste2p(14-43)-KKK peptide. Preliminary 2D –TOCSY and NOESY spectra were obtained from the sample and assignments of Phe and Tyr residues were able to be made from the TOCSY spectrum. However, based on the size of the peptide, the number of crosspeaks that were observed in the spectra was quite a bit less than expected. Therefore, a 2D- COSY spectrum was run on the sample in order to observe all of the backbone interactions. The KKK-Ste2p(14-43)-KKK peptide contains 36 amino acid residues which correlates to 37 expected crosspeaks in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region of the 2D-COSY spectrum. The 2D-COSY spectrum was promising and displayed ~30 crosspeaks in

the $^{1}H^{N}$-$^{1}H^{\alpha}$ region. Assignment of the proton spin systems of the six lysine and five threonine residues were made (Figure 3.16 and 3.17). The observed proton shifts were similar to the predicted proton shifts for most of the protons (Table 3.2 and 3.3). However, the predicted proton chemical shifts are obtained from NMR of very small peptides and real proton chemical shifts show a range of variations due to the sequence and structure of the peptide. There was a larger chemical shift in the amide and alpha protons which may be due to their interactions with the protons of neighbouring amino acids. Resonances arising from the side chains of all six lysines were virtually identical, in the spectrum indicting very little difference in their chemical environment within the peptide structure. This is expected based on their grouped arrangement of 3 consecutive lysines at each of the N and C-termini. The ε-amino-protons were observed in the spectrum which is uncommon due to the rapid exchange of the protons when in water. However, it is very common for lysine amino residues to participate in a multitude of reactions which results in the amino residues being modified (Creighton, 1984). Therefore, most likely the amino protons have undergone a reaction which resulted in their modification. The lack of clear separation for lysine backbone proton shifts and $^{1}H^{N}$-$^{1}H^{\alpha}$ crosspeaks is also likely at least in part representative of the similar chemical environment of each lysine. The overlapped $^{1}H^{N}$-$^{1}H^{\alpha}$ crosspeaks of the six lysines likely accounts for some (if not five) of the missing $^{1}H^{N}$-$^{1}H^{\alpha}$ crosspeaks. In contrast, the five threonines are dispersed throughout the peptide (Thr 16, Thr 23, Thr 27, Thr 35 and Thr 37) and thus shift separation, particularly in the backbone region would be expected. While generally greater separation was observed for the lysine shifts compared to the threonine shifts, the resonances were still sufficiently overlapped that individual proton peaks could not be identified (Figure 3.18). Overlap of the five threonine $^{1}H^{N}$-$^{1}H^{\alpha}$ crosspeaks could also account for some (if not 4) of the missing $^{1}H^{N}$-$^{1}H^{\alpha}$ crosspeaks. While this overlap may still be attributable to a similar chemical environment for the threonines, the problem is likely compounded by the generally broad nature of all the peaks in the spectrum, likely representing some degree of aggregation in the peptide sample. In conclusion, some initial assignments were able to be made and more assignments are definitely possible. However, overall the lack of resolution arising from broad peaks in the ^{1}H spectrum were an indication that the KKK-Ste2p(14-43)-KKK peptide was possibly aggregated, an issue that higher field magnets would not address. The aggregation state of this peptide remains to be assessed by DLS.

4.3 SK Peptide Purification

After promising initial structural analyses of the KKK-Ste2p(14-43)-KKK peptide, the focus was shifted back to producing a recombinant N-terminal Ste2p peptide. The major reason for this was to be able to label the protein with ^{15}N by expressing the protein in isotopically enriched minimal medium. Isotopic labelling of recombinant proteins is essential for structural elucidation using heteronuclear (^{15}N-^{1}H) NMR experiments. The use of ^{15}N-labelling results in the ability to observe specific nuclear interactions, such as the interactions between backbone nitrogens and interacting protons. In addition, labelling the peptide with ^{15}N allows for the sample to contain a lower concentration of peptide, such as 0.1-0.3 mM (Cohen *et al.*, 2008; Su *et al.*, 2008). Working towards this goal a construct was produced that when expressed would contain residues 1-48 of Ste2p again with the KSI and His tags. This construct was subjected to two rounds of mutagenesis to introduce three lysine residues between Met 1 and 2 of Ste2p(1-48) and between 48 and Met49. The construct was successfully expressed in *E. coli* and the resulting protein purified. After cyanogen bromide cleavage of the fusion protein the expected peptide was the KKK-Ste2p(2-48)-KKK-M peptide. Purification of this peptide using HPLC was successful and confirmed using MS. Therefore, the addition of the lysine residues proved to decrease the hydrophobicity enough to allow for the peptide to be purified. A hydrophobicity plot of the SK peptide (Figure 4.4) verified that the peptide is predicted to be less hydrophobic then the Ste2p(2-54) peptide (Figure 4.1). However, using MS analysis it was shown that the peptide was not completely cleaved. The resulting purified product was a combination of ~70% SK peptide as well as ~30% LM-SK peptide, which is the SK peptide with an additional Leucine and Methionine on the N-terminus. Even though the SK peptide was not completely 100% pure it was decided to continue with structural analysis of the peptide. The reasoning behind this is that the only difference between the two peptides is two amino acid residues. Therefore, the majority of the nuclei from the amino acids will have identical chemical shifts in any obtained NMR spectra. The only differences will be due to the nuclei from the L and M residues as well as the nuclei shifts of the N-terminal lysine residues of the LM-SK peptide may be affected, which could actually help assignment of N-terminal

versus C-terminal lysines. However, since the abundance of LM-SK peptide is less than half of the SK peptide the resulting spectra will indicate which shifts are due to the LM-SK peptide and which are due to the SK peptide based on the magnitude of the peaks. The final amount of peptide that was obtained was 11.5 mg from each initial litre of bacterial culture, which filled the requirement for the amount of pure protein needed for NMR. Therefore, it was decided that structural analysis of the SK peptide could feasibly be attempted.

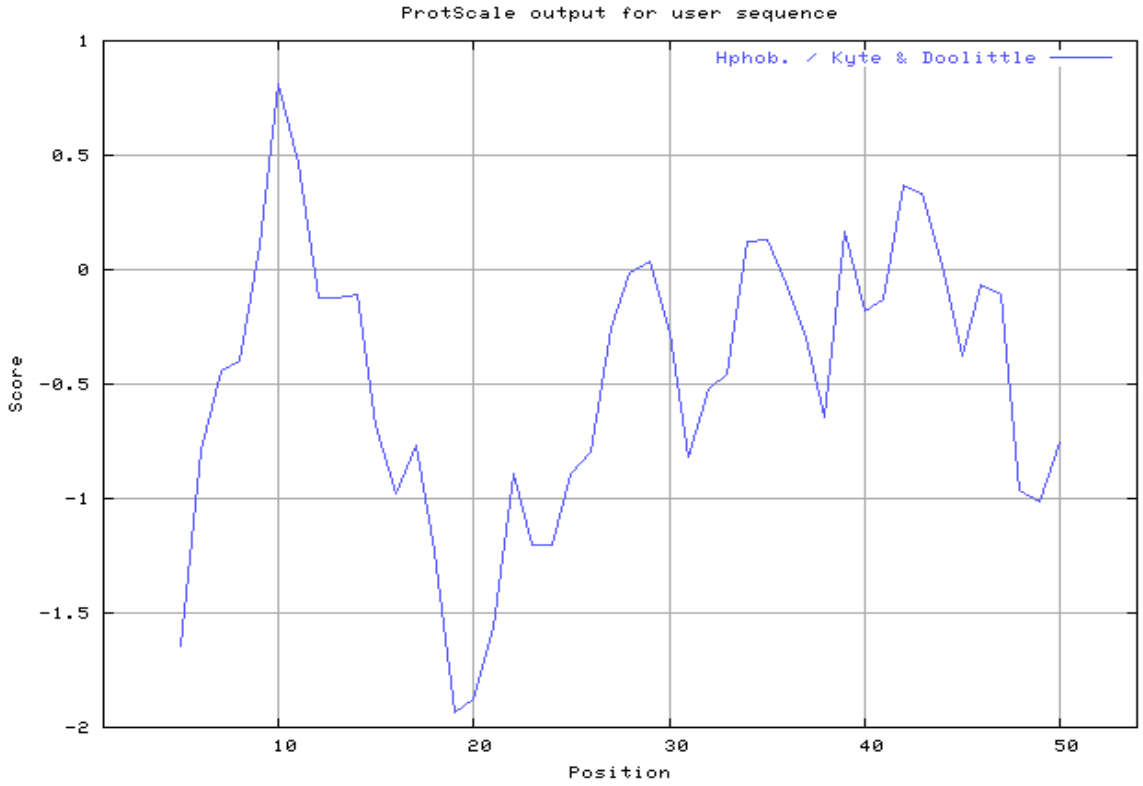

Figure 4.4: Hydrophobicity Plot of the SK Peptide.
The hydrophobicity of the SK peptide was analyzed using ProtScale and based on the Kyte and Doolittle Hydropathicity model (Kyte and Doolittle, 1982). Areas above 0 are considered to be hydrophobic and areas below 0 are considered to be hydrophilic.

4.4 SK Peptide Structural Analysis

Previous broad peaks in the NMR spectra of the KKK-Ste2p(14-43)-KKK peptide had implied that the peptide was possibly aggregated. Therefore, dynamic light scattering was performed on the SK peptide in order to determine the aggregation state of the peptide in a number of solutions. The SK peptide displayed aggregation when dissolved in H_2O or H_2O with LPPG detergent. However, it did not display aggregation when dissolved in TFE/H_2O. It is not unexpected that TFE/H2O would solubilise the peptide; it is a useful solution for dissolving proteins and has been very useful for structural analysis of hydrophobic peptides or proteins (Sonnichsen *et al.*, 1992; Tyukhtenko *et al.*, 2009). Therefore, CD analysis was performed on the SK peptide in TFE/H_2O. The CD spectrum displayed secondary structure and was predicted to have ~52% β-strand and 6% α-helix. When compared to the predicted secondary structure of KKK-Ste2p(14-43)-KKK in TFE/H_2O which was ~ 26% β-strand and 13% α-helix with an expected beta-turn, it indicates that a larger portion of the longer recombinant SK peptide is β-strand with relatively less α-helix. It was shown previously that residues 39-47 of Ste2p form an α-helix (Neumoin *et al.*, 2009), therefore, the higher β-strand content must be due to the structure of the N-terminal residues of the SK peptide.

NMR analysis of a 4 mM SK peptide sample in TFE/H_2O/D_2O resulted in a proton spectrum that contained many sharp peaks. However, there were still fewer peaks than would be expected from a peptide with 54 amino acids. In addition, there were a few very broad peaks especially in the 8 ppm region. This spectrum indicated that the peptide was actually aggregated. However, DLS analysis had shown that the peptide was not aggregated when dissolved in TFE/H_2O. The reason for this may be that the peptide concentration was too high. When the sample was analysed using DLS it was at a concentration of 0.2 mM where as the peptide concentration used during NMR was 4 mM. Therefore, the increase in concentration may have resulted in aggregation of the peptide. Another indication that the peptide was aggregated was that when the 2D-COSY was run on the 4 mM SK peptide sample it displayed very few crosspeaks in the $^{1}H^{N}$-$^{1}H^{\alpha}$ region, when there was expected to be 54 crosspeaks. Therefore, it would be better to decrease the concentration and confirm using DLS that the peptide is not aggregated before continuing with NMR analysis of the peptide. NMR experiments have been previously run on isotopically labelled peptide samples of 3-7 kDa at concentrations of 0.2 - 1 mM and yielded good results (Nielsen *et al.*, 1994; Seebahn *et al.*,

2011). So, if the SK peptide solubilised without aggregation at a lower concentration, there is a good chance the structure could be solved.

The SK peptide was also expressed, cleaved and purified after being isotopically labelled with ^{15}N. This was to allow 2D-heteronuclear (^{1}H-^{15}N) experiments to be performed on the peptide. Confirmation of the presence of the ^{15}N-SK peptide was attempted using MS after purifying the peptide using HPLC. However, no peaks were found in the MS spectrum to indicate that the ^{15}N-SK peptide was present in the sample. There were a large number of peaks from 3000-4000 m.u. which indicated that possibly the ^{15}N-SK peptide had degraded. However, a 2D ^{1}H-^{15}N HSQC was run on the sample at 3 mM to observe all ^{15}N-labelled nitrogens and their attached protons. Many ^{1}H-^{15}N interactions were observed on the 2D-HSQC spectrum indicating that the peptide had definitely been labelled with ^{15}N. However, since the peptide was degraded it was difficult to produce a clearly resolved spectrum. The reason the peptide degraded is unknown but in future experiments MALDI-TOF-MS can be run on the sample after cyanogen bromide cleavage to ensure the peptide has not degraded prior to HPLC purification.

4.5 Conclusions

In conclusion, a portion of the N-terminus of Ste2p was successfully expressed and purified from *E. coli* as a fusion protein, KSI-M-KKK-Ste2p(2-48)-KKK-M-HIS. The fusion protein was cleaved after Met residues using cyanogen bromide cleavage in order to liberate the SK peptide. The SK peptide was successfully purified using HPLC to mg quantities which were required for structural analysis. Biophysical analysis of the SK peptide demonstrated that the peptide was not aggregated at lower concentrations and that there was secondary structure observed, which correlated with previous predictions. NMR analysis on the SK peptide allowed for initial amino acid assignments but displayed very few crosspeaks from the backbone amino acids. However, the successful isotopic labelling of the fusion protein with ^{15}N provides a possible avenue to elucidation of the structure of the N-terminus in the future.

4.6 Future Directions

Future experiments on this project would include trying to alleviate the aggregation problems of the SK peptide. One solution would be to perform a screen of a number of different detergents in order to determine an optimal solution to dissolve the SK peptide. This would require a detergent that solubilised the SK peptide at high concentrations while keeping the detergent concentration low so as not to interfere during NMR. In addition, the detergent would have to be commercially available in a deuterated form for future NMR studies. Also, experimenting with the solubilisation conditions could be helpful, such as increasing the temperature or using sonication. Performing DLS on the peptide at the concentration intended to be used for NMR might be very helpful as well. Another solvent system that could be used would be chloroform:methanol:water. However, it is difficult to assess the aggregation state of the peptide using this solvent. If however the SK peptide was unable to be solubilised then a synthetic ^{15}N-labelled peptide could be produced and studied at lower concentrations. In addition, if the aggregation problem is resolved, future NMR studies could be performed using higher power magnets to increase the resolution of the NMR spectra.

Future experiments that could be performed would be mutagenesis to produce N-terminal fusion proteins containing I24A, I24C, I29A, I24C, P15C and then structural analysis to compare mutated SK peptide structures to wild type SK peptide structure. This would give some insight into why the Ste2p receptor lost its mating efficiency after these residues were mutated. Another interesting experiment that could be performed would be tandem affinity purification using the wild type or N-terminally truncated Ste3p and the SK peptide to determine if there is an interaction between the SK peptide and Ste3p.

5.0 REFERENCES

Abu-Shumays, A., and Duffield, J.J. (1966). Circular dichroism theory and instrumentation. Anal Chem *38*, A29-A58.

Altenbach, C., Kusnetzow, A.K., Ernst, O.P., Hofmann, K.P., and Hubbell, W.L. (2008). High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. Proc Natl Acad Sci USA *105*, 7439-7444.

Arevalo, E., Estephan, R., Madeo, J., Arshava, B., Dumont, M., Becker, J.M., and Naider, F. (2003). Biosynthesis and biophysical analysis of domains of a yeast G protein-coupled receptor. Biopolymers *71*, 516-531.

Ash, J., Wu, C., Larocque, R., Jamal, M., Stevens, W., Osborne, M., Thomas, D.Y., and Whiteway, M. (2003). Genetic Analysis of the Interface Between Cdc42p and the CRIB Domain of Ste20p in Saccharomyces cerevisiae. Genetics *163*, 9-20.

Bardwell, L. (2005). A walk-through of the yeast mating pheromone response pathway. Peptides *26*, 339-350.

Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). Biochemistry, 5th edition (New York, W H Freeman).

Bhasin, M., and Raghava, G.P.S. (2004). GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors. Nucl Acids Res *32*, W383-389.

Bockaert, J., and Pin, J.P. (1999). Molecular tinkering of G protein-coupled receptors: an evolutionary success. EMBO J *18*, 1723-1729.

Bohm, G., Muhr, R., and Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng *5*, 191-195.

Bokoch, M.P., Zou, Y., Rasmussen, S.G.F., Liu, C.W., Nygaard, R., Rosenbaum, D.M., Fung, J.J., Choi, H.-J., Thian, F.S., Kobilka, T.S*., et al.* (2010). Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. Nature *463*, 108-112.

Boxer, D.H., Zhang, H., Gourley, D.G., Hunter, W.N., Kelly, S.M., and Price, N.C. (2004). Sensing of remote oxyanion binding at the DNA binding domain of the molybdate-dependent transcriptional regulator, ModE. Org Biomol Chem *2*, 2829-2837.

Brahms, S., and Brahms, J. (1980). Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. J Mol Biol *138*, 149-178.

Bulenger, S., Marullo, S., and Bouvier, M. (2005). Emerging role of homo- and heterodimerization in G-protein-coupled receptor biosynthesis and maturation. Trends Pharmacol Sci *26*, 131-137.

Butty, A.-C., Pryciak, P.M., Huang, L.S., Herskowitz, I., and Peter, M. (1998). The Role of Far1p in Linking the Heterotrimeric G Protein to Polarity Establishment Proteins During Yeast Mating. Science *282*, 1511-1516.

Cardoso, J., Pinto, V., Vieira, F., Clark, M., and Power, D. (2006). Evolution of secretin family GPCR members in the metazoa. BMC Evol Biol *6*, 108.

Cavanagh, J., Fairbrother, W.J., III, A.G.P., and Skeleton, N.J. (1996). Protein NMR spectroscopy: principles and practice (San Diego, CA, Academic Press).

Chen, E.H., Grote, E., Mohler, W., and Vignery, A. (2007). Cell-cell fusion. FEBS Letters *581*, 2181-2193.

Cheng, B.Y.M., Carbonell, J.G., and Klein-Seetharaman, J. (2005). Protein classification based on text document classification techniques. Proteins: Struct, Funct, Bioinf *58*, 955-970.

Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G.F., Thian, F.S., Kobilka, T.S., Choi, H.-J., Kuhn, P., Weis, W.I., Kobilka, B.K.*, et al.* (2007). High-Resolution Crystal Structure of an Engineered Human β_2-Adrenergic G Protein Coupled Receptor. Science *318*, 1258-1265.

Chien, E.Y.T., Liu, W., Zhao, Q., Katritch, V., Won Han, G., Hanson, M.A., Shi, L., Newman, A.H., Javitch, J.A., Cherezov, V.*, et al.* (2010). Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. Science *330*, 1091-1095.

Chou, K.C., and Elrod, D.W. (2002). Bioinformatical Analysis of G-Protein-Coupled Receptors. J Proteome Res *1*, 429-433.

Chou, P.Y., and Fasman, G.D. (1978). Empirical Predictions of Protein Conformation. Annu Rev Biochem *47*, 251-276.

Chung, D.A., Zuiderweg, E.R.P., Fowler, C.B., Soyer, O.S., Mosberg, H.I., and Neubig, R.R. (2002). NMR Structure of the Second Intracellular Loop of the α2A Adrenergic Receptor: Evidence for a Novel Cytoplasmic Helix. Biochemistry *41*, 3596-3604.

Cohen, L.S., Arshava, B., Estephan, R., Englander, J., Kim, H., Hauser, M., Zerbe, O., Ceruso, M., Becker, J.M., and Naider, F. (2008). Expression and biophysical analysis of two double-transmembrane domain-containing fragments from a yeast G protein-coupled receptor. Peptide Sci *90*, 117-130.

Conner, M., Hicks, M.R., Dafforn, T., Knowles, T.J., Ludwig, C., Staddon, S., Overduin, M., Gunther, U.L., Thome, J., Wheatley, M.*, et al.* (2008). Functional and Biophysical Analysis of the C-Terminus of the CGRP-Receptor; a Family B GPCR. Biochemistry *47*, 8434-8444.

Cook, G., Stefer, S., and Opella, S.J. (2010). Expression and purification of the membrane protein p7 from HCV. Biopolymers *96*, 32-40.

Creighton, T.E. (1984). Proteins: Structures and Molecular Properties (New York, NY, W.H Freeman and Company).

Dohlman, H.G. (2002). G Proteins and Pheromone Signaling. Annu Rev Physiol *64*, 129-152.

Dohlman, H.G., Thorner, J., Caron, M.G., and Lefkowitz, R.J. (1991). Model Systems for the Study of Seven-Transmembrane-Segment Receptors. Annu Rev Biochem *60*, 653-688.

Eilers, M., Hornak, V., Smith, S.O., and Konopka, J.B. (2005). Comparison of Class A and D G Protein-Coupled Receptors: Common Features in Structure and Activation. Biochemistry *44*, 8959-8975.

Elion, E.A. (2000). Pheromone response, mating and cell biology. Curr Opin Microbiol *3*, 573-581.

Engelman, D.M., Steitz, T.A., and Goldman, A. (1986). Identifying Nonpolar Transbilayer Helices in Amino Acid Sequences of Membrane Proteins. Annu Rev Biophys Biophys Chem *15*, 321-353.

Estephan, R., Englander, J., Arshava, B., Samples, K.L., Becker, J.M., and Naider, F. (2005). Biosynthesis and NMR Analysis of a 73-Residue Domain of a Saccharomyces cerevisiae G Protein-Coupled Receptor. Biochemistry *44*, 11795-11810.

Fredriksson, R., Lagerström, M.C., Lundin, L.G., and Schiöth, H.B. (2003). The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. Mol Pharmacol *63*, 1256-1272.

Freedman, S.J., Furie, B.C., Furie, B., and Baleja, J.D. (1995). Structure of the Metal-free γ-Carboxyglutamic Acid-rich Membrane Binding Region of Factor IX by Two-dimensional NMR Spectroscopy. J Biol Chem *270*, 7980-7987.

Gardner, K.H., and Kay, L.E. (1998). The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct *27*, 357-406.

Gautier, A., Mott, H.R., Bostock, M.J., Kirkpatrick, J.P., and Nietlispach, D. (2010). Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. Nat Struct Mol Biol *17*, 768-774.

Gehret, A.U., Bajaj, A., Naider, F., and Dumont, M.E. (2006). Oligomerization of the Yeast α-Factor Receptor. J Biol Chem *281*, 20698-20714.

Goto, N.K., and Kay, L.E. (2000). New developments in isotope labeling strategies for protein solution NMR spectroscopy. Curr Opin Struct Biol *10*, 585-592.

Gross, E., and Witkop, B. (1962). Nonenzymatic Cleavage of Peptide Bonds: The Methionine Residues in Bovine Pancreatic Ribonuclease. J Biol Chem *237*, 1856-1860.

Hanson, M.A., Cherezov, V., Griffith, M.T., Roth, C.B., Jaakola, V.P., Chien, E.Y., Velasquez, J., Kuhn, P., and Stevens, R.C. (2008). A Specific Cholesterol Binding Site Is Established by the 2.8 Å Structure of the Human β2-Adrenergic Receptor. Structure *16*, 897-905.

He, L., Fong, J., von Zastrow, M., and Whistler, J.L. (2002). Regulation of Opioid Receptor Trafficking and Morphine Tolerance by Receptor Oligomerization. Cell *108*, 271-282.

Hennessey, J.P., and Johnson, W.C. (1981). Information content in the circular dichroism of proteins. Biochemistry *20*, 1085-1094.

Herskowitz, I. (1995). MAP kinase pathways in yeast: For mating and more. Cell *80*, 187-197.

Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F.E., and Vriend, G. (2003). GPCRDB information system for G protein-coupled receptors. Nucl Acids Res *31*, 294-297.

Iannone, M. (1999). Exploring Fourier Transform Techniques with Mathcad. J Chem Educ *76*, 286.

Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene *96*, 23-28.

Jaakola, V.-P., Griffith, M.T., Hanson, M.A., Cherezov, V., Chien, E.Y.T., Lane, J.R., IJzerman, A.P., and Stevens, R.C. (2008). The 2.6 Angstrom Crystal Structure of a Human A_{2A} Adenosine Receptor Bound to an Antagonist. Science *322*, 1211-1217.

Jacoby, E., Bouhelal, R., Gerspacher, M., and Seuwen, K. (2006). The 7 TM G-Protein-Coupled Receptor Target Family13. Chem Med Chem *1*, 760-782.

Jordan, B.A., and Devi, L.A. (1999). G-protein-coupled receptor heterodimerization modulates receptor function. Nature *399*, 697-700.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. Biochim Biophys Acta *1751*, 119-139.

Kelly, S.M., and Price, N.C. (1997). The application of circular dichroism to studies of protein folding and unfolding. Biochim Biophys Acta *1338*, 161-185.

Kolakowski Jr., L.F. (1994). GCRDb: A G-Protein-Coupled Receptor Database. Receptor Channel *2*, 1-7.

Konopka, J.B., and Fields, S. (1992). The pheromone signal pathway in *Saccharomyces cerevisiae*. Antonie Van Leeuwenhoek *62*, 95-108.

Konopka, J.B., Margarit, S.M., and Dube, P. (1996). Mutation of Pro-258 in transmembrane domain 6 constitutively activates the G protein-coupled alpha-factor receptor. Proc Natl Acad Sci USA *93*, 6764-6769.

Kristiansen, K. (2004). Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. Pharmacol Therapeut *103*, 21-80.

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

Kuliopulos, A., Shortle, D., and Talalay, P. (1987). Isolation and sequencing of the gene encoding delta 5-3-ketosteroid isomerase of *Pseudomonas testosteroni*: overexpression of the protein. Proc Nat Acad Sci USA *84*, 8893-8897.

Kuliopulos, A., and Walsh, C.T. (1994). Production, Purification, and Cleavage of Tandem Repeats of Recombinant Peptides. J Am Chem Soc *116*, 4599-4607.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J Mol Biol *157*, 105-132.

Lambright, D.G., Noel, J.P., Hamm, H.E., and Sigler, P.B. (1994). Structural determinants for activation of the α-subunit of a heterotrimeric G protein. Nature *369*, 621-628.

Li, J., Edwards, P.C., Burghammer, M., Villa, C., and Schertler, G.F.X. (2004a). Structure of Bovine Rhodopsin in a Trigonal Crystal Form. J Mol Biol *343*, 1409-1438.

Li, S., Xing, D., and Li, J. (2004b). Dynamic Light Scattering Application to Study Protein Interactions in Electrolyte Solutions. J Biol Phys *30*, 313-324.

Lim, L., Manser, E., Leung, T., and Hall, C. (1996). Regulation of phosphorylation pathways by p21 GTPases. Eur J Biochem *242*, 171-185.

Madden, K., and Snyder, M. (1998). Cell Polarity and Morphogenesis in Budding Yeast. Annu Rev Microbiol *52*, 687-744.

Marsh, L., Neiman, A.M., and Herskowitz, I. (1991). Signal Transduction During Pheromone Response in Yeast. Annu Rev Cell Biol *7*, 699-728.

Marshall, F.H., Jones, K.A., Kaupmann, K., and Bettler, B. (1999). $GABA_B$ receptors - the first 7TM heterodimers. Trends Pharmacol Sci *20*, 396-399.

Milligan, G. (2004). G Protein-Coupled Receptor Dimerization: Function and Ligand Pharmacology. Mol Pharmacol *66*, 1-7.

Milligan, G. (2009). G protein-coupled receptor hetero-dimerization: contribution to pharmacology and function. Br J Pharmacol *158*, 5-14.

Mittermaier, A., and Kay, L.E. (2006). New Tools Provide New Insights in NMR Studies of Protein Dynamics. Science *312*, 224-228.

Monera, O.D., Sereda, T.J., Zhou, N.E., Kay, C.M., and Hodges, R.S. (1995). Relationship of sidechain hydrophobicity and α-helical propensity on the stability of the single-stranded amphipathic α-helix. J Peptide Sci *1*, 319-329.

Murphy, R.M., and Tsai, A.M. (2006). Misbehaving Proteins: Protein (Mis)Folding, Aggregation, and Stability (New York, NY, Springer).

Naider, F., Arshava, B., Ding, F.X., Arevalo, E., and Becker, J.M. (2001). Peptide fragments as models to study the structure of a G-protein coupled receptor: The α-factor receptor of *Saccharomyces cerevisiae*. Biopolymers *60*, 334-350.

Nelson, G., Hoon, M.A., Chandrashekar, J., Zhang, Y., Ryba, N.J.P., and Zuker, C.S. (2001). Mammalian Sweet Taste Receptors. Cell *106*, 381-390.

Neumoin, A., Arshava, B., Becker, J., Zerbe, O., and Naider, F. (2007). NMR Studies in Dodecylphosphocholine of a Fragment Containing the Seventh Transmembrane Helix of a G-Protein-Coupled Receptor from Saccharomyces cerevisiae. Biophys J *93*, 467-482.

Neumoin, A., Cohen, L.S., Arshava, B., Tantry, S., Becker, J.M., Zerbe, O., and Naider, F. (2009). Structure of a Double Transmembrane Fragment of a G-Protein-Coupled Receptor in Micelles. Biophys J *96*, 3187-3196.

Nielsen, K.J., Heath, R.L., Anderson, M.A., and Craik, D.J. (1994). The Three-dimensional Solution Structure by ^{1}H NMR of a 6-Kda Proteinase Inhibitor Isolated from the Stigma of Nicotiana alata. J Mol Biol *242*, 231-243.

Nobbmann, U., Connah, M., Fish, B., Varley, P., Gee, C., Mulot, S., Chen, J., Zhou, L., Lu, Y., Sheng, F.*, et al.* (2007). Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. Biotechnol Genet Eng Rev *24*, 117-128.

Oakley, M.T., Bulheller, B.M., and Hirst, J.D. (2006). First-principles calculations of protein circular dichroism in the far-ultraviolet and beyond. Chirality *18*, 340-347.

Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E.M., and Shichida, Y. (2002). Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. Proc Natl Acad Sci USA *99*, 5982-5987.

Okada, T., Sugihara, M., Bondar, A.-N., Elstner, M., Entel, P., and Buss, V. (2004). The Retinal Conformation and its Environment in Rhodopsin in Light of a New 2.2 A Crystal Structure. J Mol Biol *342*, 571-583.

Overton, M.C., Chinault, S.L., and Blumer, K.J. (2003). Oligomerization, Biogenesis, and Signaling Is Promoted by a Glycophorin A-like Dimerization Motif in Transmembrane Domain 1 of a Yeast G Protein-coupled Receptor. J Biol Chem *278*, 49369-49377.

Overton, M.C., Chinault, S.L., and Blumer, K.J. (2005). Oligomerization of G-Protein-Coupled Receptors: Lessons from the Yeast Saccharomyces cerevisiae. Eukaryotic Cell *4*, 1963-1970.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Trong, I.L., Teller, D.C., Okada, T., Stenkamp, R.E*., et al.* (2000). Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. Science *289*, 739-745.

Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA *94*, 12366-12371.

Pierce, K.L., Premont, R.T., and Lefkowitz, R.J. (2002). Seven-transmembrane receptors. Nat Rev Mol Cell Biol *3*, 639-650.

Pin, J.-P., Galvez, T., and Prézeau, L. (2003). Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. Pharmacol Ther *98*, 325-354.

Probst, W.C., Snyder, L.A., Schuster, D.I., Brosius, J., and Sealfon, S.C. (1992). Sequence alignment of the G-protein coupled receptor superfamily. DNA Cell Biol *11*, 1-20.

Pryciak, P.M., and Huntress, F.A. (1998). Membrane recruitment of the kinase cascade scaffold protein Ste5 by the Gβγ complex underlies activation of the yeast pheromone response pathway. Genes Dev *12*, 2684-2697.

Raicu, V., Jansma, D.B., Miller, R.J.D., and Friesen, J.D. (2005). Protein interaction quantified *in vivo* by spectrally resolved fluorescence resonance energy transfer. Biochem J *385*, 265-277.

Raisley, B., Zhang, M., Hereld, D., and Hadwiger, J.A. (2004). A cAMP receptor-like G protein-coupled receptor with roles in growth regulation and development. Dev Biol *265*, 433-445.

Rasmussen, S.G.F., Choi, H.-J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R.P., Sanishvili, R., Fischetti, R.F*., et al.* (2007). Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. Nature *450*, 383-387.

Ravindran, A., Joseph, P.R.B., and Rajarathnam, K. (2009). Structural Basis for Differential Binding of the Interleukin-8 Monomer and Dimer to the CXCR1 N-Domain: Role of Coupled Interactions and Dynamics. Biochemistry *48*, 8795-8805.

Roberts, G.C.K., ed. (1993). NMR of macromolecules: a practical approach (New York, NY, Oxford University Press).

Rodríguez, J.C., Wong, L., and Jennings, P.A. (2003). The solvent in CNBr cleavage reactions determines the fragmentation efficiency of ketosteroid isomerase fusion proteins used in the production of recombinant peptides. Protein Express Purif *28*, 224-231.

Rosenbaum, D.M., Cherezov, V., Hanson, M.A., Rasmussen, S.G.F., Thian, F.S., Kobilka, T.S., Choi, H.-J., Yao, X.-J., Weis, W.I., Stevens, R.C*., et al.* (2007). GPCR Engineering Yields High-Resolution Structural Insights into β_2-Adrenergic Receptor Function. Science *318*, 1266-1273.

Rosenberg, A.H., Lade, B.N., Dao-shan, C., Lin, S.-W., Dunn, J.J., and Studier, F.W. (1987). Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene *56*, 125-135.

Sajedi, R., Naderi-Manesh, H., Khajeh, K., Ranjbar, B., Ghaemi, N., and Naderi-Manesh, M. (2004). Purification, characterization, and structural investigation of a new moderately thermophilic and partially calcium-independent extracellular α-amylase from *Bacillus* sp. TM1. App Biochem Biotechnol *119*, 41-50.

Sajedi, R.H., Taghdir, M., Naderi-Manesh, H., Khajeh, K., and Ranjbar, B. (2007). Nucleotide sequence, structural investigation and homology modeling studies of a Ca^{2+} independant α-amylase with acidic pH-profile. J Biochem Mol Biol *40*, 315-324.

Salahpour, A., Angers, S., Mercier, J.-F., Lagacé, M., Marullo, S., and Bouvier, M. (2004). Homodimerization of the β2-Adrenergic Receptor as a Prerequisite for Cell Surface Targeting. J Biol Chem *279*, 33390-33397.

Salom, D., Lodowski, D.T., Stenkamp, R.E., Trong, I.L., Golczak, M., Jastrzebska, B., Harris, T., Ballesteros, J.A., and Palczewski, K. (2006). Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. P Natl Acad Sci USA *103*, 16123-16128.

Sanders, C.R., and Landis, G.C. (1995). Reconstitution of Membrane Proteins into Lipid-Rich Bilayered Mixed Micelles for NMR Studies. Biochemistry *34*, 4030-4040.

Scheerer, P., Park, J.H., Hildebrand, P.W., Kim, Y.J., Krausz, N., Choe, H.-W., Hofmann, K.P., and Ernst, O.P. (2008). Crystal structure of opsin in its G-protein-interacting conformation. Nature *455*, 497-502.

Schulte, G., and Bryja, V. (2007). The Frizzled family of unconventional G-protein-coupled receptors. Trends Pharmacol Sci *28*, 518-525.

Seebahn, A., Dinkel, H., Mohrlüder, J., Hartmann, R., Vogel, N., Becker, C.-M., Sticht, H., and Enz, R. (2011). Structural characterization of intracellular C-terminal domains of group III metabotropic glutamate receptors. FEBS Letters *585*, 511-516.

Shi, C., Kaminskyj, S., Caldwell, S., and Loewen, M.C. (2007). A role for a complex between activated G protein-coupled receptors in yeast cellular mating. Proc Natl Acad Sci USA *104*, 5395-5400.

Shi, C., Kendall, S.C., Grote, E., Kaminskyj, S., and Loewen, M.C. (2009a). N-terminal residues of the yeast pheromone receptor, Ste2p, mediate mating events independently of G1-arrest signaling. J Cell Biochem *107*, 630-638.

Shi, C., Paige, M.F., Maley, J., and Loewen, M.C. (2009b). *In vitro* characterization of ligand-induced oligomerization of the S. cerevisiae G-protein coupled receptor, Ste2p. Biochim Biophys Acta-Gen Subjects *1790*, 1-7.

Silverstein, R.M., Webster, F.X., and Kiemle, D., eds. (2005). Spectrometric identification of organic compounds, 7th edn (Hoboken, NJ, John Wiley & Sons, Inc.).

Sommers, C.M., Martin, N.P., Akal-Strader, A., Becker, J.M., Naider, F., and Dumont, M.E. (2000). A Limited Spectrum of Mutations Causes Constitutive Activation of the Yeast α-Factor Receptor. Biochemistry *39*, 6898-6909.

Sonnichsen, F.D., Van Eyk, J.E., Hodges, R.S., and Sykes, B.D. (1992). Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. Biochemistry *31*, 8790-8798.

Stark, J., and Powers, R. (2007). Rapid Protein−Ligand Costructures Using Chemical Shift Perturbations. J Am Chem Soc *130*, 535-545.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol *189*, 113-130.

Su, Z., Sukdeo, N., and Honek, J.F. (2008). $^{15}N-^{1}H$ HSQC NMR Evidence for Distinct Specificity of Two Active Sites in *Escherichia coli* Glyoxalase I. Biochemistry *47*, 13232-13241.

Tapaneeyakorn, S., Goddard, A.D., Oates, J., Willis, C.L., and Watts, A. (2011). Solution- and solid-state NMR studies of GPCRs and their ligands. Biochim Biophys Acta-Biomembranes *1808*, 1462-1475.

Teller, D.C., Okada, T., Behnke, C.A., Palczewski, K., and Stenkamp, R.E. (2001). Advances in Determination of a High-Resolution Three-Dimensional Structure of Rhodopsin, a Model of G-Protein-Coupled Receptors (GPCRs). Biochemistry *40*, 7761-7772.

Terrillon, S., and Bouvier, M. (2004). Roles of G-protein-coupled receptor dimerization. EMBO Rep *5*, 30-34.

Tyukhtenko, S., Tiburu, E.K., Deshmukh, L., Vinogradova, O., Janero, D.R., and Makriyannis, A. (2009). NMR solution structure of human cannabinoid receptor-1 helix 7/8 peptide: Candidate electrostatic interactions and microdomain formation. Biochem Biophys Res Commun *390*, 441-446.

Ulfers, A.L., Piserchio, A., and Mierke, D.F. (2002). Extracellular domains of the neurokinin-1 receptor: Structural characterization and interactions with substance P. Peptide Sci *66*, 339-349.

Wand, A.J. (2001). Dynamic activation of protein function: A view emerging from NMR spectroscopy. Nat Struct Biol *8*, 926-931.

Wang, Y., and Dohlman, H.G. (2004). Pheromone Signaling Mechanisms in Yeast: A Prototypical Sex Machine. Science *306*, 1508-1509.

Wang, Y., Zhou, Z., Walsh, C.T., and McMahon, A.P. (2009). Selective translocation of intracellular Smoothened to the primary cilium in response to Hedgehog pathway modulation. Proc Natl Acad Sci USA *106*, 2623-2628.

Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G.W., Tate, C.G., and Schertler, G.F.X. (2008). Structure of a β1-adrenergic G-protein-coupled receptor. Nature *454*, 486-491.

Werner, K., Richter, C., Klein-Seetharaman, J., and Schwalbe, H. (2008). Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. J Biomol NMR *40*, 49-53.

Wu, B., Chien, E.Y.T., Mol, C.D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F.C*., et al.* (2010). Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. Science *330*, 1066-1071.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids (Hoboken, NJ, John Wiley & Sons).

Xie, H., Ding, F.-X., Schreiber, D., Eng, G., Liu, S.-f., Arshava, B., Arevalo, E., Becker, J.M., and Naider, F. (2000). Synthesis and Biophysical Analysis of Transmembrane Domains of a Saccharomyces cerevisiae G Protein-Coupled Receptor. Biochemistry *39*, 15462-15474.

Zheng, H., Zhao, J., Sheng, W., and Xie, X.-Q. (2006). A transmembrane helix-bundle from G-protein coupled receptor CB2: Biosynthesis, purification, and NMR characterization. Biopolymers *83*, 46-61.

Zou, C., Naider, F., and Zerbe, O. (2008). Biosynthesis and NMR-studies of a double transmembrane domain from the Y4 receptor, a human GPCR. J Biomol NMR *42*, 257-269.