A Structural Investigation of Squash Aspartic Peptidase Inhibitor (SQAPI) using Nuclear Magnetic Resonance spectroscopy (NMR).

A thesis presented in partial fulfillment of the requirements for the degree of

Master of Science
in
Biochemistry

at Massey University, Palmerston North
New Zealand

Ursula Kate MacAskill

2007

Abstract

Peptidases are enzymes that hydrolyse peptide bonds. This potentially dangerous activity is regulated by post translational modification and peptidase inhibitors. The best characterized of the peptidase inhibitors are the serpins whilst the aspartic peptidase inhibitors are the least characterized. Aspartic peptidase inhibitors are rare with only nine known sources. However, they are of great interest because they play an important part in several human diseases such as metastasis of breast cancer cells, *Candida albicans* infections and HIV.

The aims of this research project were to investigate the structure of Squash Aspartic peptidase inhibitor (SQAPI), using nuclear magnetic resonance spectroscopy (NMR). This required large amounts of relatively pure and isotopically labeled protein, which was achieved by heterologously expressing His-tagged rSQAPI fusion protein in *Escherichia coli* using a rich to minimal media transfer method. The fusion protein was purified with a nickel column and the N-terminal extension containing the His_6-tag was removed by cleavage of the fusion protein with enterokinase followed by nickel column purification.

Preliminary 1 dimensional NMR spectra indicated that SQAPI was folded in solution at pH 3. This was confirmed from the results of a preliminary ^{15}N-edited HSQC. These results combined justified the production of a ^{15}N ^{13}C labeled SQAPI sample for the collection of further NMR spectra. From the spectra produced with double labeled protein the backbone and the side-chain atoms of SQAPI were assigned. The chemical shifts are currently 88.89% complete and have been submitted to the biological magnetic resonance bank (BMRB). A preliminary estimate of the secondary structure of SQAPI has been calculated from the HNHA spectrum suggesting that the SQAPI structure has some similarity to the previously proposed model of the inhibitor's structure. Furthermore, the region corresponding to the putative binding loop on the model of SQAPI was found to be mobile and deuterium exchange experiments indicate that the SQAPI structure is more globular than open.

Acknowledgements

I would like to express my appreciation and gratitude to the following people for their valuable contribution to this work:

My supervisor Dr. Peter Farley for his time, expertise and constant encouragement.

To Dr. Stephen Headey, for his generous investment of time and knowledge.

Quentin, Carole, Carel and Lin from the Mainland Lab. Jolyon, Martin and Dr. Stephen Pascal from the BioNMR Lab. Matt, Simon, Alice and others from the X-lab.

My family and friends for their support and encouragement.

The NZFGW fellowship trust and the J.P. Skipworth Plant Biology scholarship.

Table of Contents

Abbreviations

SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide get electrophoresis
NMR	Nuclear magnetic resonance spectroscopy
SQAPI	Squash aspartic peptidase inhibitor
mM	Milli mole
mL	Milli litre
g	Gram
µg	Micro gram
µl	Micro litre
ppm	Parts per million
1D	One dimensional
MHz	Mega hertz
IPAP	In-phase/Anti-phase
LB	Luria broth
OD_{600}	Optical density (at a wavelength of 600 nano-metres)
IPTG	Isopropyl-b-D-thiogalactopyranoside
Da	Dalton
kDa	Kilo-Dalton
HIV	Human immunodeficiency virus
N	Nitrogen
C	Carbon
DNA	Deoxyribonucleic acid
PI	Peptidase inhibitor
SAP	Secreted aspartic peptidase
PCR	Polymerase chain reaction
HSQC	Hetero-nuclear single quantum correlation
EDTA	Ethylenediaminetetraacetic acid

List of Figures

Technical Note: Due to technical issues the figure legends in chapter four are printed on the pages preceding their respective figures.

List of Tables

Chapter One: Introduction

Peptidase Activity and Regulation

Peptidases, proteinases, proteases and proteolytic enzymes are all terms used to describe enzymes that catalyze the hydrolysis of peptide bonds in a protein. The number of different names for this group of enzymes reflects their importance across a number of different fields, from medicine to molecular biology and biochemistry. The term peptidase will be used to describe these enzymes in this thesis as this is the term used by the MEROPS database (http//merops.sanger.ac.uk). Furthermore, from a biochemist's point of view, "peptidase" best describes the enzymatic activity of cleaving peptide bonds.

Although peptidases were originally only associated with digestion (Dash *et al*. 2003) it has been discovered that they are involved in every aspect of organismal function including catabolism, cell growth, regulation of transcription, blood coagulation and cellular protein turnover (Dash *et al*., 2003; Bochtler *et al*., 1999). In mammals, the role of peptidases extends from processing of peptides for antigen presentation to transport of vesicles across the plasma membrane (Matsui *et al*., 2006). In microbes, some peptidases can bind DNA and act as transcription repressors for the biosynthesis of pyrimidine and cholera toxins (Matsui *et al*., 2006). In plants, peptidases have roles in defense, membrane transport of hormone receptors and meiosis (Matsui *et al*., 2006). In fact 2% of all genes encode a peptidase (Barret *et al*., 2001).

However, unless their activity is regulated peptidases are potentially very dangerous to a cell. Peptidase activity is typically controlled by the cell synthesizing them as inactive zymogens that must undergo cleavage to become active peptidases. In addition, cells are known to synthesize peptidase inhibitors, although the target enzyme for many of these inhibitors is not known. This is largely due to the fact that most inhibitors were discovered by screening with widely available peptidases such as chymotrypsin and pepsin, rather than by being co-purified with their target enzyme. Furthermore, not all inhibitors are synthesized for the purpose of controlling the cells own peptidases. For example, some plants produce peptidase inhibitors as a defense against chewing insects and pathogens (Laing and McManus, 2002)

Classification and Naming of Peptidase Inhibitors

Peptidases are classified in terms of the residue that is present in the active site and takes part in the catalytic activity. Thus there are aspartic, serine, cysteine and metallo peptidases (Barret *et al*., 1999). In general peptidase inhibitors are classified according to the type of peptidase that they inhibit (Laskowski and Kato, 1980). There is some confusion with the naming of inhibitors as they were often named after the species in which they were first discovered. The problem being that orthologs of the same inhibitor are later discovered in a range of other organisms (Laing and McManus, 2002).

In 2004 Rawlings *et al*. constructed a classification system by comparing the amino acid sequences of most known inhibitors. This is a robust system that organizes the inhibitors into 31 families based on similarity between the amino acid sequences. These families are then grouped into 26 clans based on similarities between their tertiary structures. The system has been implemented in the MEROPS peptidase database (http//merops.sanger.ac.uk). In this thesis, however, the inhibitors will be discussed in terms of the catalytic residues of the peptidases that they inhibit as the tertiary structure of many of the aspartic peptidase inhibitors are not known.

The Serine Peptidase Inhibitors

The best characterized super-amily of peptidase inhibitors are the serine peptidase inhibitors (serpins). Serpins are found in plants, animals and some viruses. Little is known about the mechanism of action of plant serpins (Laing and Mcmanus, 2002). However, the general mechanism of action of animal serpins involves a reactive centre loop at the C-terminus of the inhibitor that acts as a bait for the target peptidase. Initially a non-covalent enzyme-inhibitor complex forms, then a peptide bond of the inhibitor is attacked by the active site serine of the peptidase. This attack results in the formation of a covalent bond between the serine and the carboxyl group of the residue in the inhibitor next to the cleaved bond (Gettins *et al*., 1996). The reactive centre loop of the inhibitor then inserts completely into the A β-sheet of the enzyme (Huntington *et al*., 2000). This

induces massive conformational disorder in the peptidase rendering it inactive (Huntington *et al*., 2000, Silverman *et al*., 2001).

The Cysteine Peptidase Inhibitors

Several groups of cysteine peptidase inhibitors are known from both plant and animal sources. The largest class of cysteine peptidase inhibitors are the cystatins which are found in plants and animals (Turk *et al*., 1997). The animal cystatins are split into three groups: type 1 cystatins, type 2 cystatins and the kininogens. The plant cystatins or phytocystatins have the greatest sequence similarity to the type 2 cystatins but have been classified as type 1 on the basis of size and lack of disulfide bonds. It has however, been suggested that phytocystatins should be considered as a completely separate class as phylogenetic analysis places them as a separate major evolutionary tree branch (Turk *et al*., 1997, Margis *et al.,* 1998). The active loops of cystatins bind tightly to the peptidase either side of the active site cysteine thus avoiding cleavage and inactivating the enzyme (Bjork *et al*., 1990).

The Metallopeptidase Inhibitors

Proteinaceous metallopeptidase inhibitors are uncommon but examples are known from plants, animals and microbes (Peanasky *et al*., 1974; Rancour and Ryan, 1968; Oda *et al.,* 1979). A common mechanism of action of metallopeptidase inhibitors is the "Lakowski mechanism", whereby the inhibitor acts as a highly specific but limited substrate of the target enzyme (Oda *et al.*, 1979; Seeram *et al.*, 1997). Structures have been solved for most of the members of this group of inhibitors (http//merops.sanger.ac.uk).

The Aspartic Peptidase Inhibitors

Proteinaceous aspartic peptidase inhibitors are rare with only nine known sources (Laing and McManus, 2002). There are four plant sources; potato cathepsin D inhibitor (Keilova & Tomasek, 1976), wheat inhibitor (Galleschi *et al*., 1993) a squash inhibitor (Christeller *et al*., 1998) and a recently discovered inhibitor from kiwifruit seed (Rassam and Laing, 2006). The other sources are Nematode worm (Abu-Erreish & Peanasky, 1974), sea anemone (Lenaric and Turk, 1999) and yeast (Schu and Wolf, 1991). The remaining two aspartic peptidase inhibitors are small non-proteinaceous microbial peptides that will not be discussed.

The aspartic peptidase inhibitors are the least well characterized peptidase inhibitors. Their mechanisms of action are largely unknown, as is their structure and function *in vivo.* The exception is the yeast inhibitor IA3 which has been well characterized. However, because aspartic peptidases are involved in human disease (examples of such are HIV-1 peptidase in acquired immune deficiency syndrome, cathepsin D in metastasis of breast cancer, β□-secretase in Alzheimer's disease and secreted aspartic peptidases in candidial infection (Dash *et al*., 2003)) their inhibitors are of particular interest.

The Yeast Inhibitor (IA3)

The yeast inhibitor gene encoding IA3 was discovered by oligonucleotide screening of a *Saccharomyces cerevisiae* genomic library. Degenerate oligonucleotides, designed from the amino acid sequence of the inhibitor, were used to probe the cDNA library. The gene codes for a protein of 68 amino acids (Schu & Wolf, 1991).

IA3 inhibits the vacuolar *S. cerevisiae* endopeptidase yscA with sub-nanomolar potency (Phylip *et al.,* 2001). The inhibitor is only effective against yscA; it does not inhibit even close relatives of yscA from other yeast species. IA3 is a substrate, rather than an inhibitor, for all tested aspartic peptidases other than yscA (Phylip *et al.,* 2001).

Although IA3 is 68 amino acids in length, systematic truncation and mutagenesis located the inhibitory activity to the N-terminal half of the protein, in particular residues 2-34 (Phylip *et al*., 2001, Li *et al*., 2000). CD spectra and NMR analysis showed that IA3 is totally unstructured in the absence of yscA (Green *et al*., 2004). The crystal structures of yscA in a complex with either full length IA3 or residues 2-34 were resolved to 2.2 and 1.8 angstroms respectively (Li *et al.*, 2000, Phylip *et al*., 2001). This revealed that the aspartic peptidase yscA folds residues 2-34 of the inhibitor into an almost perfect α helix whereas the rest of the inhibitor remains unstructured. (Li *et al*., 2000, Phylip *et al*., 2001).

The peptidase yscA acts as a template that stabilizes the formation of an α helix that blocks the proteolytic activity of yscA (Li *et al*., 2000). The interactions with the enzyme are between the hydrophobic face of the amphipathic helix and the binding site on the enzyme. There is a hydrophobic cluster of residues at each end of the helix and a leucine in the centre of the helix. The placement of these hydrophobic 'pins' into the hydrophobic 'sockets' of the enzyme active site is crucial for formation of the helix (Phylip *et al*., 2001). The α helix is lodged in the binding cleft of the protein in such a manner that the residues of the inhibitor are no closer than 5 angstroms from the catalytic aspartates of the enzyme, and thus cannot be cleaved (Li *et al.*, 2000).

The Nematode Inhibitors (PI 1-4)

Four inhibitors of pepsin were purified from the body walls of adult *Ascaris lumbricoides* var *suis* (Abu-Erreish & Peanasky, 1973). The inhibitors were found to be active against bovine, porcine and human pepsin between pH 1 and pH 6 and have no significant sequence similarity to any other protein (Abu-Erreish & Peanasky, 1973; Martzen *et al*., 1990).

The four inhibitors have molecular masses ranging between 17,515 and 31,719 Da. Edman sequencing of the primary structure revealed that all of the cysteines in the nematode inhibitors are involved in disulfide bonds (Abu-Erreish & Peanasky, 1973). Ng *et al*. (2000) solved the three dimensional structure of pepsin inhibitor-3 (PI-3) from *Ascaris suum* using X-ray crystallography. PI-3 has a unique fold consisting of two

domains separated by a short hydrophobic region. Each domain comprises an anti parallel β sheet flanked by an α helix.

The structure of the PI-3 and pepsin complex revealed that the inhibitor has a unique mode of action. The N-terminal β strand of PI-3 pairs with a β strand on the active site flap of pepsin. This forms an eight stranded β sheet that spans the two proteins. The N – terminal residues of PI-3 block the area that holds the three residues of the substrate that are upstream of the cleaved peptide bond; thus inactivating the enzyme (Ng *et al*., 2000).

The Potato Inhibitor

The potato inhibitor was the first known proteinaceous cathepsin D inhibitor from plants. Comparison of the amino acid sequence of the potato inhibitor with other inhibitor sequences revealed that it was most likely derived from the Kunitz/soybean trypsin inhibitor family (Mares *et al*., 1989) and probably arose from this family by active site modification (Laing & McManus, 2002).

The potato inhibitor inhibits both the aspartic peptidase cathepsin D and the serine peptidase trypsin but is not effective against pepsin. In 1989, Mares *et al.* discovered that the trypsin inhibitory activity could not be separated from the cathepsin D inhibitory activity by affinity chromatography. Thus it is likely that the protein has two domains with separate activity.

In 1992, Strukelj *et al*. concluded that the potato inhibitor exists as a multigene family. This conclusion was reached by cloning and characterization of three cDNAs of protein inhibitor homologues and the detection of four protein isoforms.

A protein with 84% homology to the potato inhibitor was recently found in tomato (Cater *et al*., 2002). The two inhibitors were found to have greater potency against the yeast peptidase yscA than cathepsin D. This indicates that perhaps the *in vivo* function of these inhibitors is to protect from fungal infection (Cater *et al*., 2002). The structure of these inhibitors has not been solved.

The Wheat Inhibitor

The wheat inhibitor is the least well characterized aspartic peptidase inhibitor. It is reported to have a molecular weight of 58 kDa; which is exceptionally large for a plant inhibitor (Galleschi *et al*., 1993). The nucleic acid sequence is not known nor is the amino acid sequence.

The Sea Anemone Inhibitor (Equistatin)

Equistatin is a 22 kDa protein composed of three thyroglobulin type 1 domains each 65 amino acids long (Lenarcic *et al*., 1997; Lenarcic *et al.*, 1998). It was the first known inhibitor of cathepsin D of animal origin (Lenarcic & Turk, 1999). Equistatin is also known to inhibit the papain-like cysteine peptidases, papain and cathepsins B and L. Equistatin does not inhibit aspartic peptidases other than cathepsin D.

Interestingly cathepsin D and papain are inhibited simultaneously by equistatin. This proves that there are two separate binding sites on the inhibitor (Lenarcic & Turk,1999). In 2003 Galesa *et al*. determined the structural basis of the inhibition of cathepsin D and papain. By heterologously expressing the three separated domains in *Pichia pastoris* the authors were able to characterize the interactions of each domain with the target enzymes. It was discovered that the second domain of equistatin is responsible for the inhibition of cathepsin D. Domain 2 alone inhibited the enzyme at comparable levels to full sized equistatin.

The CD spectra of the interaction between domain 2 and cathepsin D revealed that this equistatin domain undergoes a conformational change when interacting with the enzyme (Galesa *et al*., 2003). This parallels the induced conformational change observed with the yeast aspartic peptidase inhibitor IA3 (Li *et al*., 2000).

The Kiwifruit Inhibitor

The pepsin inhibiting activity of kiwifruit is due to a 20 kDa β−subunit of the 11S globulin-like protein (11s-GLP), legumin (Rassam and Laing, 2006). It appears to inhibit bovine spleen cathepsin D, *Candida albicans* secreted aspartic peptidases (SAPs) 1, 2 and 4, the plant fungus *Glomerella cingulata* SAP, apple seed aspartic peptidase and is a weak inhibitor of kiwifruit seed aspartic peptidase (Rassam and Laing, 2006). Kinetic studies with pepsin show that this inhibitor fits the Michaelis-Menten model for competitive inhibition, rather than the typical tight binding model of peptidase inhibitors. Furthermore, from extrapolation of the kinetics analysis it was estimated that a 900 fold excess of 11S-GLP would be necessary to achieve complete inhibition of pepsin (Rassam and Laing, 2006). Because of this, 11S-GLP would be better classed as a difficult substrate rather than an inhibitor of pepsin.

Squash Aspartic Peptidase Inhibitor (SQAPI)

In 1998 Christeller *et al*. reported the discovery of an aspartic peptidase inhibitor that they had purified from squash (*Cucurbita maxima*) phloem exudate. The authors had noted that squash phloem exudate inhibits pepsin and that this activity was lost when the exudate was heated. A purification procedure involving mild acid treatment, chromatography on trypsin agarose and Sephadex G-75 followed by reverse phase HPLC resulted in the elution of three main peaks that possessed the inhibitory activity. These peaks were shown to contain SQAPI. The molecular mass of HPLC-purified SQAPI was determined by MALDI-TOF MS. The main species had a molecular mass of 10,552 Da. Smaller and larger species of SQAPI were also detected in small amounts (Christeller *et al*., 1998).

SQAPI inhibits several aspartic peptidases. The activity of pepsin and the *Glomerella cingulata* secreted aspartic peptidase are inhibited by SQAPI. Cathepsin D and the *C. albicans* secreted aspartic peptidase 1 and 2 exhibit little inhibition by SQAPI. However *C. albicans* secreted aspartic peptidase 4 showed strong inhibition by SQAPI. The natural target of SQAPI is not yet known nor is its physiological function (Christeller *et al*., 1998; Farley *et al*., 2002).

SQAPI inhibits pepsin completely at a 1:1 molar ratio of pepsin to SQAPI monomer (Farley *et al*., 2002). Surface plasmon resonance analysis showed that the SQAPI:pepsin complex forms slowly but dissociates at a much slower rate than it forms. The same is also true for the one of the *C. albicans* secreted aspartic peptidase isoform that was susceptible to the inhibition by SQAPI. Strong inhibition of the *C. albicans* secreted aspartic peptidase 4 was characterized by slow formation of the secreted aspartic peptidase 4-inhibitor complex but even slower dissociation of the complex. In contrast, weak inhibition of the *G. cingulata* and other *C. albicans* SAPs were characterized by larger dissociation constants (Farley *et al.,* 2002).

SQAPI is an effective inhibitor of pepsin because the pepsin-inhibitor complex is very stable, particularly at low pH. The dissociation rate constant increases with rising pH. Furthermore the association rate between pH5.0-6.0 is lower than at low pH. These factors combined result in a decrease in the efficiency of inhibition at higher pH (Farley *et al*., 2002). SQAPI was classified as a reversible tight binding inhibitor of pepsin (Farley *et al*., 2002)

SQAPI behaves as a dimer of 21,000 Da when subjected to gel filtration chromatography at pH 7, however, SQAPI is active only at low pH (Christeller *et al.,* 1998, Farley *et al.,* 2002). The conditions of gel chromatography may not reflect the *in planta* behaviour of SQAPI. Even though SQAPI inhibits pepsin at a 1:1 ratio of SQAPI monomer to pepsin it is unclear if SQAPI functions as a dimer or a monomer.

In vivo functions of SQAPI

As little is known about the biochemistry of the phloem it is possible that SQAPI functions to regulate the activity of the plants own peptidases (McManus *et al*., 2000). However, SQAPI may protect against chewing insects by inhibiting their digestive peptidases (McManus *et al*., 2000). This is supported by the wide spectrum of inhibition demonstrated by SQAPI (Christeller *et al*., 1998). Furthermore SQAPI is located in the phloem exudate which accumulates at the site of wounds such as those made by chewing insects (Christeller *et al.,* 1998).

The *Cucurbita maxima* SQAPI gene family

Two SQAPI variants were isolated using PCR amplification of cDNA and gDNA. These were called the DVIS and HDVA isoforms according to the amino acids differences between the two isoforms. Subsequently further clones were identified using 3' rapid amplification of cDNA ends establishing that SQAPI exists as a small gene family (Christeller *et al.,* 2006).

The existence of a gene family was confirmed by Southern blotting which indicates that at least four genes were present in the genome of squash (Christeller *et al.,* 2006).

SQAPI Gene Phylogeny

PCR analysis of gDNA from the *Cucurbitales* family with primers from squash sequences revealed SQAPI gene homologs in every member of the order tested (Christeller *et al.,* 2006). This indicates that SQAPI evolved before speciation occurred within the order Curcubitales. It is possible that small SQAPI gene families exist in most of the species tested as in most cases more than one SQAPI gene was isolated (Christeller *et al*., 2006).

Sequence comparisons indicate that SQAPI evolved from the phytocystatins. The SQAPI amino acid sequence has 23-31% identity to phytocystatins and 45-50% similarity. An illustration of the close relationship between SQAPI and the cystatin family is, that the degree of homology and similarity between SQAPI and the *Cucimis sativus* cystatin is similar to that between the apple and pear cystatins (Christeller *et al*., 2006).

Christeller *et al*. (2006) further tested this hypothesis by modelling SQAPI onto the solution structure of the rice phytocystatin, oryzacystatin. The threaded structure is an excellent fit throughout the entire molecule. Areas of SQAPI that have been identified as hyper-variable (which is a characteristic of the binding loops of protease inhibitors) correspond to the binding loops of the oryzacystatin (Christeller *et al*., 2006).

Further support for the model is the position of the solitary tryptophan on a binding loop. The fluorescence of this tryptophan is quenched when an enzyme inhibitor complex forms. Site directed mutagenesis or chemical modification of the tryptophan residue abolishes the activity of the inhibitor, possibly because it can no longer bind the peptidase (Christeller *et al*., 2006).

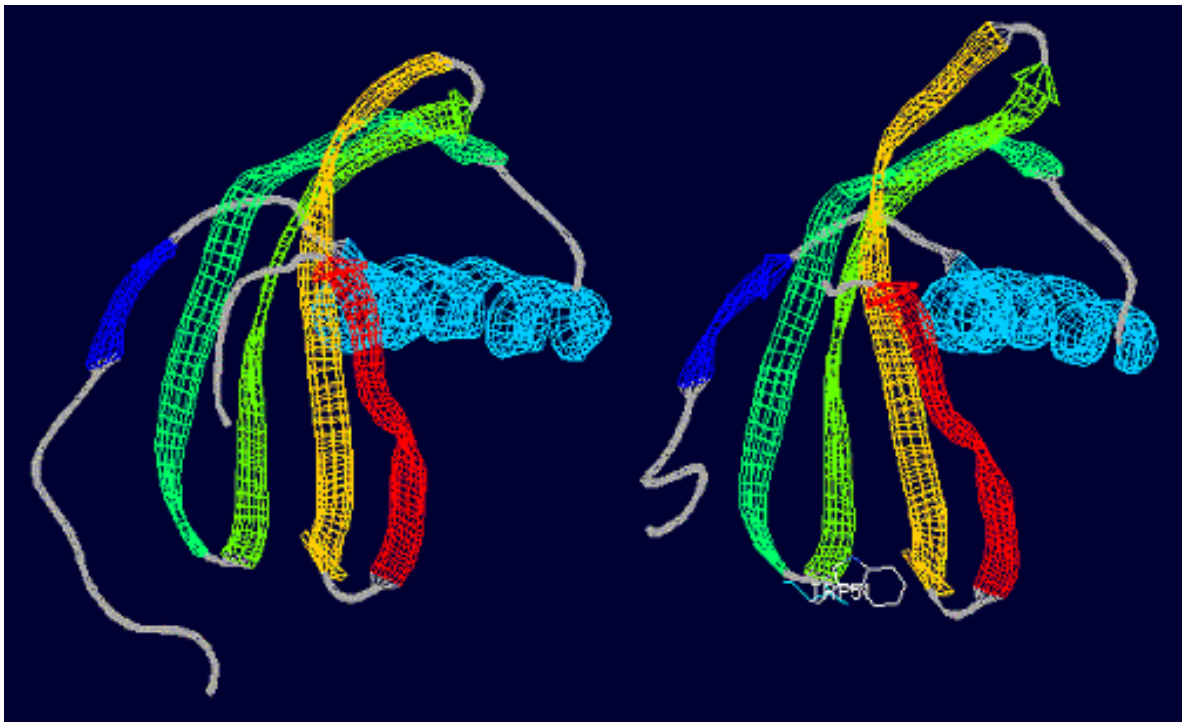


Figure 1. Superimposition of SQAPI onto the rice oryzacystatin structure.

The structure on the left is that of rice cysteine peptidase inhibitor (Nagata et al., 2000). The structure on the right is the threaded fit of SQAPI. The Tryptophan shown on the right is believed to be at the binding site of SQAPI. This loops forms part of the binding site in cystatins. (Fig. 1 is sourced from Christeller et al., 2006)

Protein Structure

Proteins are polymers of 20 different L-amino acids joined by peptide bonds. There are four levels of protein structure. The primary structure is the sequence of amino acids in the polypeptide chain. Secondary structure refers to the regular sub-structures, mainly α-helices and β-sheets that form due to hydrogen bonding of the backbone. The tertiary

structure of a protein refers to the folding together of the secondary structures due to the properties of the side chains (R-groups) of the amino acid residues. Quaternary structure, which not all proteins possess, is formed from several individually folded polypeptides combining to make a protein with multiple subunits.

Structural Determination Techniques

There are two main techniques for the resolution of protein structures, X-ray crystallography and Nuclear Magnetic Resonance spectroscopy (NMR). X-ray crystallography relies on the production of light refracting protein crystals. Protein crystallisation is potentially time consuming and attempts with SQAPI have proved unsuccessful in the past (Elliot, unpublished). Furthermore, structures produced by X-ray crystallography usually provide little information about mobile regions of the protein, such as loops and hinges.

Conventional NMR spectroscopy techniques are limited by the size of proteins that can be resolved, due to spectral complexity and the fact that larger molecules have broader resonances. However, this is not an issue in this case as SQAPI is a 10.5 kDa protein and well below the current size limit of approximately 30 kDa. For these reasons NMR spectroscopy was deemed an appropriate technique for the resolution of the structure.

Stages of Protein Structure Determination

Protein structural determination by NMR has three stages. The first stage is the assignment of backbone resonances to sequential residues. Initially a hetero-nuclear single quantum correlation (HSQC) spectrum is run. This correlates the backbone amide protons and nitrogens. Spectra are then run which correlate resonances associated with backbone amides to backbone carbon resonances. For example, the HNCACB, this is a three dimensional spectrum that correlates the amide proton of a residue (i) with the Cα □and Cβ □of i and the preceding residue, i-1. It is possible to find sequential residues by matching both the $C\alpha^{i}$ and $C\alpha^{i-1}$ resonances and the $C\beta^{i}$ and $C\beta^{i-1}$ resonances. The assignments found on the HNCBCA are confirmed on a separate spectra such as the HNCO (correlates the resonances of the carbonyl atoms of the backbone with the resonances of the backbone amides), HNCA (correlates the Cα resonances with that of the backbone amides) and HN(CO)CBCA (correlates the Cα and Cβ resonances with the resonances of the backbone amides, however the magnetization is transferred in a different manner to the HNCBCA).

The second stage of protein structural determination is the prediction or identification of secondary structural elements in the protein. This is achieved using the chemical shifts, calculation of angle restraints and determination of hydrogen bonds.

The third stage of structural determination is the identification of the tertiary structure of the protein. The principal forms of structural information used are nuclear over-hauser effect (NOE) and residual dipolar coupling (RDC) restraints. NOEs provide short-range restraints and RDCs give both local and more global information about the protein's structure.

NOEs are measured in nuclear over-hauser effect spectroscopy (NOESY) spectra. In a NOESY spectrum magnetization is transferred through space rather than through bonds. A NOESY provides information about residue atoms that are within approximately 5 angstroms of one another which includes but is not limited to atoms of residues that are sequential. The signal intensity in a NOESY rapidly decreases as inter-atomic distance increases, therefore limiting the amount of information on longer distance interactions

provided by the NOESY. These interactions are instead provided by the RDCs. .

The aims of this project were:

1. To determine if SQAPI is structured when free in solution.
2. To obtain a preliminary three dimensional structure of SQAPI in solution using NMR.

Chapter Two: Materials and Methods

Materials

Chemicals

The following chemicals were purchased from BDH (Poole, England): potassium di-hydrogen orthophosphate, di-sodium orthophosphate, sodium dodecyl sulphate, Triton X-100, acrylogel, potassium acetate, sucrose, ammonium chloride, glycerol, imidazole and glycine-HCl. Tetramethylethylenediamine (TEMED), acetic acid, sodium azide and 2-mercaptoethanol were purchased from Riedel-de Haen (Seelze, Germany). Sodium chloride was purchased from Pure Science (Wellington, New Zealand). These chemicals were purchased from Ajax chemicals (Auckland, New Zealand): Tris-HCl, di-potassium orthophosphate, ammonium persulphate, D-glucose, ethylenediaminetetraacetic acid (EDTA) and sodium hydroxide. Chloramphenicol was purchased from Boehringer (Mannheim, Germany) and ampicillin was purchased from USB Corporation (Cleveland, USA). Agar was purchased from Oxoid ltd (Hampshire, England). Thiamine, polyoxyethylene 5-lauryl ether and isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Sigma (Auckland, New Zealand).
The following chemicals were purchased from Merck (Darmstadt, Germany): yeast extract, bacto tryptone, sodium di-hydrogen phosphate, ethanol, chloroform, isoamyl alcohol, isopropanol and hexanol. The ^{15}N enriched ammonium chloride was purchased from spectra stable isotopes (Columbia, USA) and the ^{13}C-D-glucose was purchased from Cambridge Isotope Laboratories (Andover, USA).

Enzymes

Eneterokinase (EKmax) was purchased from Invitrogen (Victoria, Australia). Lysozyme and RNAse were purchased from Roche (Mannheim, Germany).

Other

The Hi-trap chelating Nickel columns were purchased from Pharmacia Biotech / GE Healthcare (Uppsala, Sweden). The centricons were purchased from Amicon

Bioseparations (Billerica, USA). The Viva-spins were purchased from Sartorius Stedim Biotech (Aubagne, France). The EKaway resin was purchased from Invitrogen. The syringe filters and broad range SDS-PAGE standards were purchased from Biorad (California, USA).

Phosphate buffer preparation

The phosphate buffer that was used for collection of NMR samples was prepared by adjusting a 10 mM solution of potassium di-hydrogen phosphate containing EDTA (1mM) to pH 3 with HCl. After sterilisation by autoclave 0.02% w/v sodium azide was added. The buffer was stored in the dark at room temperature.

Vectors

The pRSET/30.2/18 plasmid carrying the His-tagged rSQAPI fusion protein (HDVA isoform) was kindly provided by Dr. Peter Farley. Stocks of this vector were made by transforming electrocompetent *Escherichia coli* DH5α cells with the plasmid. These cells were used to inoculate 200 ml of Luria broth (LB) containing 100 mg/ml ampicillin and grown over night. The plasmid was extracted using either the rapid boil method or the alkaline lysis method.

Plasmid preparation methods

Bulk plasmid preparation (Alkaline lysis method)

One hundred milli-litres of an overnight *E. coli* culture was centrifuged at 5000 rpm to pellet cells. The supernatant was decanted and the cells re-suspended in 5 ml of Glucose/Tris/EDTA (50 mM Glucose, 25 mM Tris – HCl and 10 mM EDTA pH 8) and left at room temperature for 5 minutes. Ten milli-litres of a fresh NaOH/SDS (200 mM NaOH; 1 % SDS solution) was added and the solution incubated on ice for 10 minutes after being mixed by inversion. Seven and a half milli-litres of ice cold solution III (1.875 M potassium acetate, 7.2 % acetic acid) was added and incubated on ice for a further 10 minutes before centrifugation at 10000 rpm and filtration. The DNA was

precipitated from the filtrate by the addition of 31.25 ml of ethanol and incubation at – 20 ^{0}C for at least 30 minutes. After centrifugation the resulting pellet was air dried and re-suspended in TE. The suspension was then incubated at 65 ^{0}C for 20 minutes followed by incubation at 37 ^{0}C for 30 minutes after the addition of RNase. The DNA was ethanol precipitated (as described previously) after two extractions with an equal volume of phenol and one extraction with an equal volume of chloroform:isoamyl (24:1) alcohol solution. The resulting DNA pellet was re-suspended in TE.

Plasmid Preparation (Rapid Boil Method)

The cells only from 1.5 ml of an overnight *E. coli* culture were suspended in 350 µl of STET (8% Sucrose; 50 mM EDTA; 50 mM Tris; 0.5% Triton X-100 pH 8) and 25 µl of lysozyme (10mg/ml freshly prepared). The solution was centrifuged for 10 minutes at 12,000 rpm after boiling for 40 seconds. The precipitated proteins and cellular debris were removed from the sample with a toothpick. An equal volume of isopropanol was added to the supernatant before incubation at – 20 ^{0}C for at least 30 minutes. The DNA pellet that resulted from centrifugation at 12,000 rpm for 15 minutes was washed with 70 % ethanol and air dried, then re-suspended in 50 µl of TE.

Protein sample production methods

Expression of His-tag rSQAPI fusion protein

Electrocompetent *E. coli* BL21 pLysS cells were transformed with the pRSET/30.2/18 expression vector containing the His-tag rSQAPI fusion construct. These cells were streaked onto an LB plate containing ampicillin (100 µg/ml) and chloramphenicol (35 µg/ml) and grown at 37 ^{0}C overnight. A single colony was picked with a sterile toothpick and used to inoculate a seed culture of 20 ml of LB broth containing ampicillin (100 µg/ml) and chloramphenicol (35 µg/ml). The culture was grown at 37 ^{0}C to an OD_{600} of 0.2 - 0.6 and then stored at 4 ^{0}C overnight. The next day the cells were harvested from the seed culture and used to inoculate 400 ml of LB broth containing ampicillin (100 µg/ml) and chloramphenicol (35 µg/ml). This culture was grown at 37 ^{0}C to an OD_{600} of 0.2 - 0.6 and stored overnight at 4 ^{0}C. In the morning the

cells were harvested and split between two flasks containing 400 ml of pre-warmed LB broth containing ampicillin and chloramphenicol. The cultures were grown at 37 ^{0}C until the OD_{600} reached 0.5 and 2 gram of glucose was added to each flask. An hour later the cells were harvested and then shaken at 37 ^{0}C in 300 ml of M9 minimal media in a baffled flask for 30 minutes. The cells only were then transferred to 300 ml fresh M9 minimal media containing 2.4 grams glucose and 0.5 grams of ammonium chloride in a baffled flask. When the OD_{600} began to increase, indicating that growth had recommenced, 23 mg of IPTG was added to the media. The cells were harvested upon cessation of growth as assessed by the stabilizing of the OD_{600} and stored at – 20 ^{0}C (Cai *et al*., 1998).

Optimization of the amount of Glucose required for double labelled His-tag rSQAPI fusion protein production

The cell were prepared as described as above except that the induction media was split into three equal volumes. Each flask contained 1.6 mg/ml of ammonium chloride and either 2 mg/ml, 4mg/ml or 8 mg/ml of glucose. The cells from the three separate cultures were harvested upon cessation of growth and frozen at – 20 ^{0}C overnight.

The His-tag rSQAPI fusion protein from each culture was purified separately as described below. The resulting EDTA fractions that contained the fusion protein were quantified using the Nanodrop (Nanodrop) and the yield calculated for each concentration of glucose.

Purification of His-tag rSQAPI fusion protein

Cells from the induction medium were disrupted using either sonication or the French Press in an appropriate volume of cold Buffer B (20 mM sodium phosphate buffer pH 7.8 containing 500 mM sodium chloride). Particulate matter was removed by centrifugation (10000 X *g* for 20 min) and filtration (Syringe filter 0.45 μm). The supernatant was applied to a nickel affinity column (Pharmecia Biotech; 1 ml or 5 ml HiTrap Chelating) pre-equilibrated with buffer B. The column was washed with 10 bed volumes of buffer B, 20 bed volumes of buffer W (20 mM sodium phosphate buffer pH 6.0 containing 500 mM sodium chloride) and 20 bed volumes of Buffer I (20 mM

sodium phosphate buffer pH 6.0 containing 500 mM sodium chloride and 300 mM imidazole). The column was then eluated with 10 bed volumes of buffer E (20 mM sodium phosphate buffer pH 6.0 containing 500 mM sodium chloride and 50 mM EDTA) and fractions (1 bed volume each) were collected.
The fractions containing rSQAPI fusion protein were indentified by visualisation on an SDS-PAGE gel and buffer exchanged into potassium phosphate buffer (10 mM potassium phosphate buffer; 0.2 % sodium azide; pH 3) and finally concentrated using centrifugal filter devices (Centricon; NMWC 3 kDa membrane Amicon Bioseparations).

Cleavage of His-tag rSQAPI Fusion protein with enterokinase

rSQAPI fusion protein at a concentration of no more than 3 mg/ml in Ekmax reaction buffer (50 mM Tris-HCl, pH 8 containing 1 mM $CaCl_2$ and 0.1% Tween; Invitrogen) was incubated for 16 - 18 hours at 37 ^{0}C with Enterokinase (Ekmax; Invitrogen) at a ratio of 1 unit per 20 μg of His-tag rSQAPI fusion protein.

Removal of uncleaved His-tag rSQAPI and the free His-tag from cleaved rSQAPI fusion protein

The enterokinase digested rSQAPI fusion protein sample was applied to a 1 ml Ni column pre-equilibrated with buffer B. The column was subsequently washed with 10 ml buffer B, 10 ml of buffer W, 10 ml of buffer i (20 mM sodium phosphate buffer, pH 6.0; 30 mM imidazole; 500 mM NaCl), and then eluted with 10ml of buffer I before the nickel was displaced with 10 ml of buffer E. All the eluate from the column was collected in 1.5 ml aliquots and run on SDS-PAGE to determine which fractions contained the cleaved fusion protein.

Protein electrophoresis

Protein samples were prepared for SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) as described by Shagger & Von Jagow (1987). Proteins were separated using 12 % SDS-PAGE Tricine gels constructed with stacking and resolving

solutions. The resolving gel solution contained 1.28 ml glycerol, 3.06 ml Milli-Q water, 4.00 ml of Gel Buffer (3.0 M Tris base pH 8.45; containing 0.3% sodium dodecyl sulfate) 3.60 ml of 40% Acrylamide (Acrylogel 2.6; BDH laboratory supplies) 6.0 μl of TEMED (tetramethylethylendiamine; Riedel de Haen) and 25 μl of ammonium persulfate (300 mg/ml). The stacking gel solution contained 3.85 ml of Milli- Q water, 1.55 ml of Gel Buffer, 0.6 ml of 40% Acrylamide, 7.5 μl of TEMED and 25 μl of 300 mg/ml ammonium persulfate. Gels were routinely run at 50 V for 15 minutes and then at 150 V for one and a quarter hours.

Glycerol stocks

Seventy micro-litres of sterile glycerol was added to 400 μl of an *E. coli* culture that had an OD_{600} of 0.2 – 0.6. The cell suspension was mixed by pippeting and stored at – 80 ^{0}C.

NMR Methods

General NMR methods

All data acquisition and processing was carried out by Dr. Stephen Headey (Massey University, Palmerston North). The NMR experiments were performed in Shigemi tubes with the protein in 95% H_2O/ 5% ^2H_2O containing 10 mM phosphate buffer (pH 3) and 0.02% (w/v) sodium azide unless stated otherwise. Spectra were routinely recorded on a 700 MHz Bruker Avance spectrometer using a Bruker CryoProbe^{TM}. Probe temeperature was regulated with a Bruker CryoPlatform ^{TM}. A one dimensional spectra was collected at 37 ^{0} C with an unlabelled 0.06 mM sample of His-tagged rSQAPI fusion protein in glycine-HCl buffer (10 mM glycine-HCl pH 3 containing 1 mM EDTA and 0.02% (w/v) sodium azide). One dimensional spectra and a NOESY 2D ^2H_2O spectrum were collected with an unlabelled sample of His-tagged rSQAPI fusion protein in phosphate buffer pH 3. A ^{15}N enriched sample of His-tagged rSQAPI fusion protein was used to acquire an initial ^{15} N HSQC at 37^{0}C. A 0.2 mM ^{15}N enriched sample of rSQAPI protein was used to acquire a nitrogen HSQC and a ^{15} N edited NOESY-HSQC (tm 120 ms) at 37^{0}C. A 0.15 mM rSQAPI sample enriched with both

^{13}C and ^{15}N was used to collect a HNCACB, HNCA, HNCO, HN(CA)CO, HN(CO)CACB, HCCH-TOCSY, HBHA(CO)NH, HNHA, ^{13} C edited NOESY-HSQC (tm 120 ms) and a ^{15}N –edited NOESY-HSQC (tm 120 ms) at 37 ^{0} C. These spectra were used for the assignment of the backbone. However, as not all residues were visible, it was necessary to repeat some experiments at the higher temperature of 50 ^{0} C. Spectra collected at the higher temperature included HCCCONH TOCSY (tm 12 ms), HBHA(CO)NH, HNCA, HNCO, HN(CA)CO, HN(CO)CACB and a constant time ^{13}C edited HSQC (13 ms periods) . An impure 0.3 mM sample of rSQAPI enriched with both ^{13}C and ^{15}N was used to collect an HNCO and HCCCONH TOCSY (tm 12 ms or 16 ms) at 50 ^{0}C. A 0.6 mM sample of ^{13} C and ^{15} N enriched rSQAPI was used to acquire isotropic and anisotropic IPAP ^{15} N HSQC, ^{13} C HSQC, ^{15} N HSQC, ^{13} C-edited NOESY-HSQC (120ms tm), ^{15} N-edited NOESY-HSQC (120 ms tm) and a NOESY 2D ^{2} H_2O (tm 100 ms). An additional ^{13} C-edited NOESY was acquired with the pulse frequency focussed on exciting the aromatic residues, however, this was unsuccessful due to noise in the T1. The Standard Bruker Biospin pulse sequences were used for data acquisition. Water suppression was achieved by using WATERGATE. The chemical shifts were referenced according to the method of Wishart *et al.* (1995) where the ^{1}H chemical shift is referenced to the water peak whilst the ^{13}C and ^{15}N chemical shifts were referenced by the $^{13}C/^{1}H$ and $^{15}N/^{1}H$ γ- ratios. NMR data were processed in TOPSPIN version 2.0 (Bruker biospin). The processed spectra were analysed using XEASY software on a silicon graphics O_2 workstation.

Deuterium exchange

A 350 μl ^{15}N labelled sample of SQAPI in 10 mM phosphate buffer was frozen in liquid nitrogen and freeze dried overnight. The lyophilized sample was re-suspended in 250 μl $^{2}H_2O$ and placed in a shigemi tube. Four consecutive HSQC experiments were recorded on the 700 MHz NMR spectrometer at 288 K. The first HSQC ran for 12 minutes, the following two for 48 minutes and the last for 4 hours. These HSQC were overlaid to give information about which protons are involved in H-bonds, as the peaks corresponding to these protons persist in the presence of $^{2}H_2O$ due to a slower exchange with the solvent.

Residual dipolar couplings

Otting media is a mixture of a poly ethylene glycol derivatives and hexanol that forms liquid crystalline phases in solution. Biological macromolecules suspended in this media within a magnetic field become partially aligned due to the liquid crystals (Rukert and Otting, 2000). This results in splitting of protein resonances due to dipole interactions; which can be measured using complementary in-phase and antiphase (IPAP) J-coupled NMR spectra obtained with isotropic and anisotropic solutions of protein (Cordier *et al*., 1999).

The Ottings medium was prepared by the stepwise addition of hexanol in 1 μl aliquots to following mixture, and checking the appearance of the media between two pieces of juxtaposed light-polarizing film:

Polyoxyethylene 5-lauryl ether	27 μl
$^{2}H_2O$	50 μl
Phosphate buffer (3 mM) pH 3.0	450 μl

After the addition of 8 μl of hexanol the media appeared slightly viscous and opalescent between the polarized film. The media was placed in the 700 MHz NMR spectrometer to check the splitting of the $^{2}H_2O$ lock signal which indicated the solvent was not quite aligned. A further 1 μl of hexanol was added and the $^{2}H_2O$ lock signal was observed to be split by 22.5 MHz. A lyophilized sample of $^{13}C^{15}N$ SQAPI was re-suspended in a 250 μl aliquot of the Ottings medium. This sample was returned to the NMR spectrometer and the IPAP were run on the anisotropic solution. An IPAP was previously obtained with the same sample before it was lyophilized (isotropic control).

ϕ angle restraints

Using the method of Kuboniwa *et al.* (1994) the $^{3}J_{H}{}^{N}{}_{H}{}^{\alpha}$ coupling constants were calculated and used to determine the ϕ angle restraints. The calculation of the restraints uses only the information from the HNHA spectrum. In regions of the HNHA spectrum where there is little overlap the intensity of the HA peaks (cross peak: I_x) and their

corresponding HN peaks (diagonal peak: I_D) were recorded using XEASY. From the intensity data the $^3J_{H^N H^\alpha}$ values were calculated using the following formula:

$$I_X/I_D = \tan^2(\pi\ ^3J_{H^N H^\alpha} T)$$

The value obtained is then multiplied by 1.11 to compensate for H^α spin flips during T (The constant time period) (Vuister, 1993).

Chapter Three: Protein Sample Preparation for NMR Analysis

Expression and purification of rSQAPI

Introduction

As NMR requires large amounts of relatively pure and labelled protein, it is necessary to have an efficient expression system to minimize the cost. This is particularly relevant when utilizing ^{13}C-D-glucose as the carbon source. Typically, bacterial expression systems are used when large amounts of protein are required. SQAPI does not have any post-translational modifications such as glycosylation and therefore a bacterial expression system is appropriate. In this study *E. coli* cells carrying an expression vector encoding SQAPI were grown in minimal medium, which decreases the yield of protein as compared with rich medium but allows for selective isotopic labelling of the protein. The recombinant SQAPI was expressed with a N-terminal extension containing six consecutive histidines. This tag allowed for efficient purification of the fusion protein due to the strong affinity of the histidines to Ni-Sepharose.

The aim of the work described in this section was to express and purify single labelled and double labelled rSQAPI for structural determination using NMR.

Expression and purification of His-tagged rSQAPI fusion protein

The His-tagged rSQAPI fusion protein was expressed in *E. coli* BL21 (DE3) pLysS cells containing the plasmid pRSET/30.2/18. The pRSET vector encodes a 43 amino acid leader sequence that, in addition to the His_6 tag, contains an enterokinase cleavage site that can be used to enzymatically remove most of the N-terminal extension. In the pLysS system protein expression is negatively regulated by the lacI repressor and is induced upon the addition of isopropyl-β-D-thiogalactopyranoside (IPTG). It is important to have tight control of protein expression for two main reasons. (i) illicit protein production uses resources that should be used for cell growth and (ii) production of protein before the addition of the isotopic label results in protein samples that are a mixture of labelled and unlabelled protein.

The first stage in expression of the His-tagged rSQAPI fusion protein was the production of a large number of cells carrying the expression plasmid. To achieve this cells were grown in selective rich media in three steps (a 20 ml culture, a 400 ml culture and finally two 400 ml cultures). These cultures were grown to an OD_{600} of no more than 0.5. This was to insure that the cells retained both the pRSET/30.2/18 and the pLysS plasmids. After the final two 400 ml cultures reached an OD_{600} of 0.5, glucose (2 g/flask) was added in order to accelerate growth and increase the number of ribosomes per cell so that they would be more efficient at protein production when they were induced to express SQAPI.

After the growth in rich medium, the cells were transferred to minimal media for half an hour. The purpose of this incubation was to allow time for the cells to use up any residual metabolites derived from the rich medium. This is especially important for maximizing the incorporation of isotopic label into the protein. The cells were then transferred to fresh minimal media that had been supplemented with ammonium chloride and glucose. When producing ^{15}N labelled His-tagged rSQAPI fusion protein ^{15}N enriched ammonium chloride was used. When producing $^{15}N^{13}C$ labelled His-tagged rSQAPI fusion protein both ^{15}N enriched ammonium chloride and ^{13}C-D-glucose were used. SQAPI expression was induced by the addition of IPTG. Growth continued for approximately 2 hours after induction and the cells were harvested at the end of this period of growth.

The cells were re-suspended in 100 ml of ice cold binding buffer and lysed in the French press. Cell debris was removed by centrifugation and the lysate filtered to minimize particulate matter that may damage the column used in the next step. The filtered cell lysate was applied to a charged 5 ml nickel affinity column and the flow through collected. The column was then washed with 10 bed volumes of binding buffer, 10 bed volumes of wash buffer and 10 volumes of imidazole buffer to remove non-specifically bound protein. To elute the His-tagged rSQAPI fusion protein the column was washed with 10 volumes of EDTA buffer, which strips the nickel from the column along with the His-tagged rSQAPI fusion protein. Eluates from all of these steps were analysed by SDS-PAGE as shown in Figure 2. A faint band corresponding to His-tagged rSQAPI fusion protein was seen in the lysate before it was applied to the column

(lane 3). The low level of expression relative to the other proteins is typical of His-tagged rSQAPI expression in this system. Since mg quantities of protein were required for NMR this low yield of protein was overcome by growing cells on a large scale.

His-tagged rSQAPI fusion protein was absent from the unbound fraction (lane 4), the wash buffer fraction (lane 5) and the imidazole wash (lane 6). The His-tagged rSQAPI fusion protein is very tightly bound to the column as the imidazole buffer contains 300 mM imidazole, a concentration of imidazole sufficient to elute many other His-tagged proteins. The fusion protein is so tightly bound to the column it was eluted with the nickel ion using buffered EDTA.

The His-tagged rSQAPI fusion protein is present in lanes 10 through 15. These fractions appeared free of contaminating proteins and were therefore pooled and buffer exchanged in an ultra-filtration cell with a 3 kDa membrane. This method of buffer exchange proved to result in high protein losses, as less than 0.25 mg of His-tagged rSQAPI fusion protein was recovered from about 9 mg of protein from the pooled fractions (Fig. 2). The ultra-filtrator was therefore replaced with centricon concentrators in later sample preparation.

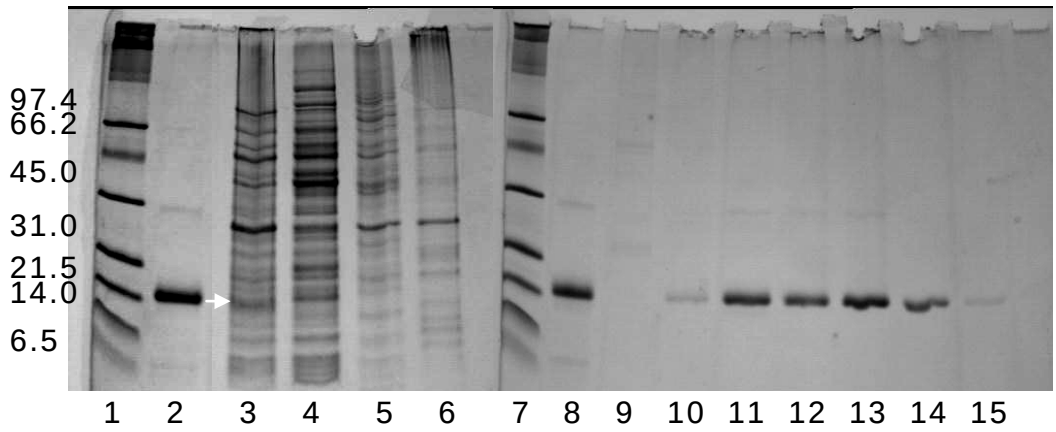


Fig. 2. SDS-PAGE of fractions from Nickel column purification of His-tagged rSQAPI fusion protein.

Lane 1. Molecular weight marker; Lane 2: His-tagged rSQAPI fusion protein standard (15.32 kDa); Lane 3: Crude extract/lysate; Lane 4: unbound fraction; Lane 5: Buffer wash; Lane 6: Imidazole wash; Lane 7: Molecular weight marker; Lane 8: His-tagged rSQAPI fusion protein standard; Lanes 9-15: Buffered EDTA fractions. The His-tagged rSQAPI fusion protein can be clearly seen in lanes 10-15 migrating with the His-tagged rSQAPI fusion protein standard in lanes 2 and 8. The arrow identifies the His-tag rSQAPI fusion protein band in the crude extract

Glucose optimization

The reagents used for production of ^{13}C and ^{15}N enriched proteins are costly, particularly the ^{13}C-D-Glucose. Thus the amount of glucose and the induction time were varied to establish the optimal conditions for protein expression. In this experiment three different glucose concentrations were tested and therefore the cells from a LB culture (800 ml) that contained the appropriate antibiotics were harvested and re-suspended in 300 ml of pre-warmed minimal media which was then divided into three equal parts. Glucose was added after the cells had been incubated for 30 minutes and protein expression was induced after a further 40 minutes.

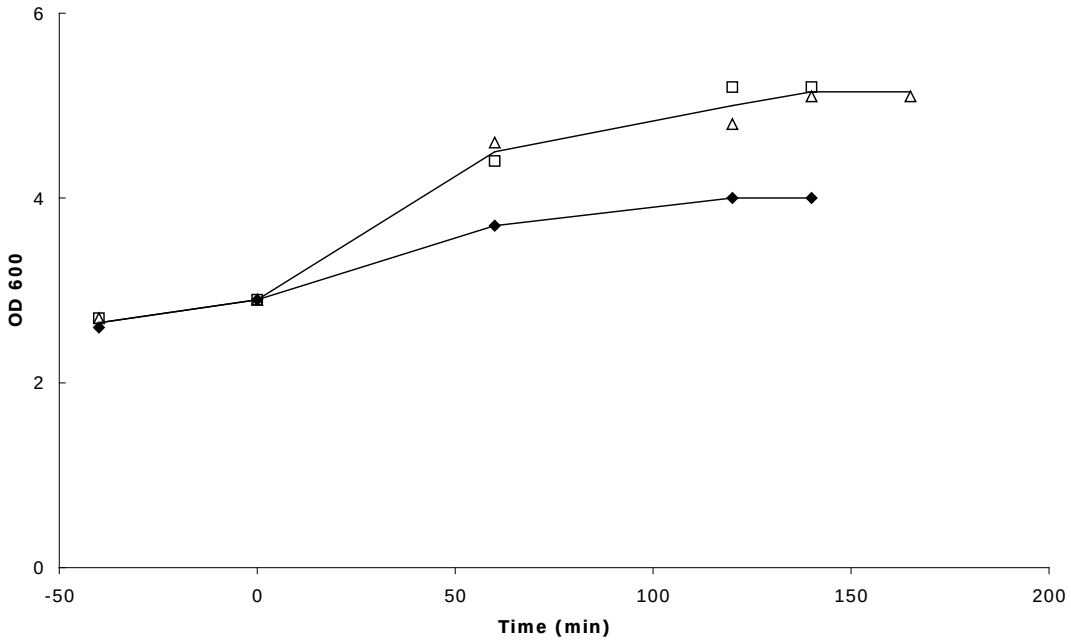


Figure 3. Growth of induced cells in minimal media at three different concentrations of glucose. Black filled diamonds (2 g/l glucose); squares (4 g/l glucose) and triangles (8 g/l glucose). The cells were induced by addition of IPTG at time point zero.

The three cultures all had an OD_{600} of 2.9 when they were induced. The graph clearly shows that the growth of the 2 g/l glucose culture lagged behind that of the 4 g/l and 8 g/l glucose cultures. There was no increase in growth from adding more than 4 g/l glucose to the minimal media as the growth of the 4 g/l and 8 g/l glucose cultures were similar. The growth of all three cultures reached a plateau between 1-2 hours after induction.

The clear cell lysates from the cultures in medium containing three different concentrations of glucose were purified by nickel affinity chromatography. The results of SDS-PAGE analysis of the resulting fractions are shown in figure 3. The His-tagged rSQAPI fusion protein was eluted with the EDTA. The final yield was calculated by quantifying the pooled EDTA fractions for each batch. The 4 g/l glucose culture gave the highest yield of His-tagged rSQAPI fusion protein with 7.2 mg of His-tagged rSQAPI fusion protein per gram of glucose (Table 1). It was therefore decided to use a glucose concentration of 4 g/l for future protein production.

Table 1. Table of yields of His-tagged rSQAPI fusion protein from glucose optimization trials.

Glucose concentration in induction medium (g/l)	Yield of His-tagged rSQAPI fusion protein (mg/g of glucose)
2	3.7
4	7.2
8	1.4

After the cells were lysed in the French press the crude lysate (50 ml) was centrifuged. A small pellet of insoluble matter formed, after decanting the lysate the pellet was re-suspended in 50 ml of buffer. This was run on SDS-PAGE (Fig 3). The His-tagged rSQAPI fusion protein is not visible in the pellet fraction, however it is visible in the crude lysate lane of all three gels. This means that the insoluble pellet was not due to aggregation of the SQAPI protein.

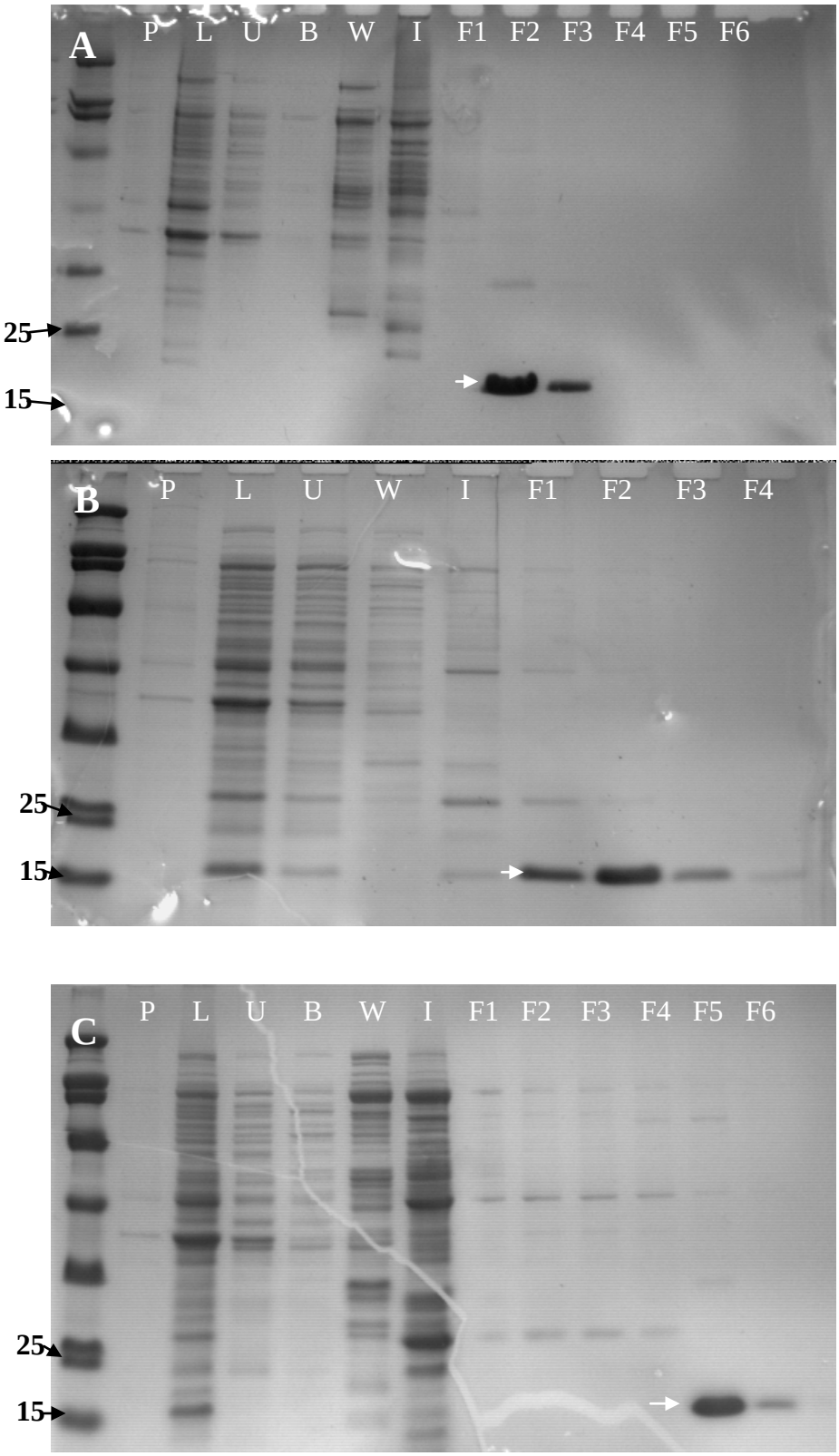


Fig 4. Nickel column purification of the lysate from cells induced in minimal media containing 2 g/l (A), 4 g/l (B) and 8 g/l (C) glucose concentration. In panels A and B the fusion protein appears in the first two or three fractions, whereas in panel C the fusion protein appears later. This was most likely due to the incorrect buffer being applied to the column before the elution buffer was used. The His-tagged rSQAPI protein is indicated with an arrow. Key P: re-suspended pellet; L: Load; U: unbound; B: Buffer B wash; W: Buffer W wash; I: Buffer I wash; F1-F6: Buffered EDTA fractions.

Optimization of induction time

To estimate the optimal induction time for expression of SQAPI, samples (1 ml) were taken from a minimal media culture used for the glucose optimization experiment at various time points for two and half hours after induction. An aliquot (5 μl) was boiled with Laemmli buffer (15 μl) and run on a 12% SDS-PAGE gel (Fig. 4). The expression of His-tagged rSQAPI fusion protein was evident 1 hour after induction (lane 5). The levels of SQAPI stabilized between 1-2 hours after induction. This correlates well with the stabilization of the OD_{600} and the end of the growth phase (Fig. 3).

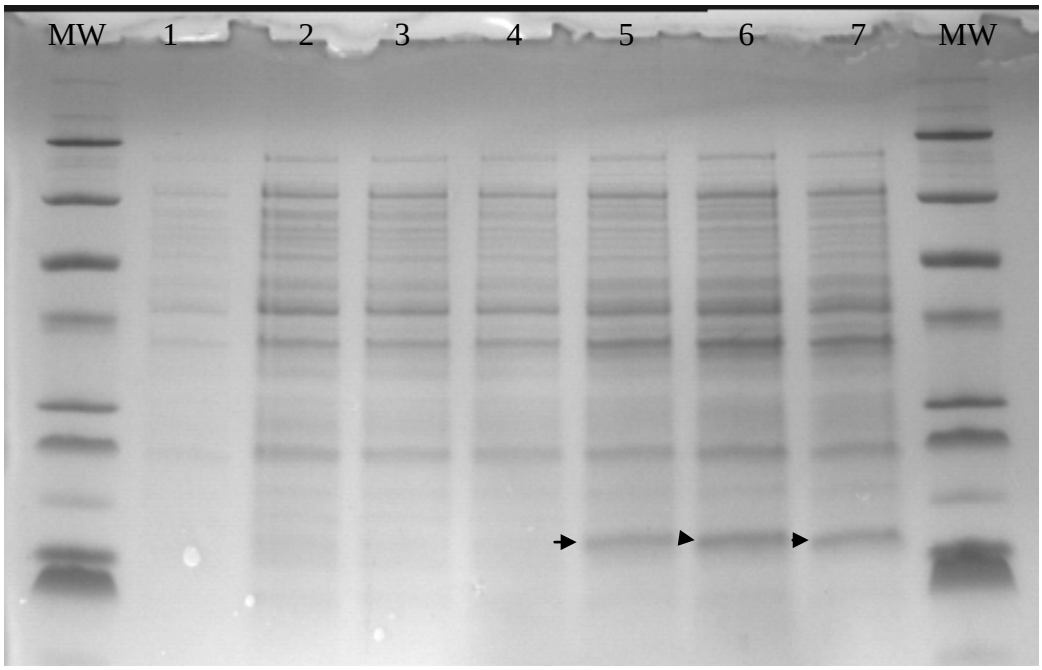


Fig 5. Expression profile of SQAPI in cells incubated in minimal media containing 4 g/l glucose. Samples were first taken after the cells were first added to the minimal medium (lane 1), after 30 minutes of incubation in the minimal media but before the addition of glucose and ammonium chloride (lane 2), 20 minutes after the addition of glucose and ammonium chloride (lane 3), 40 minutes after the addition of glucose and ammonium chloride (lane 4), and then one, two and two and a half hours after the addition of IPTG (lanes 5- 7). His-tagged rSQAPI fusion protein indicated by arrows.

Optimization of enterokinase digestion

The N-terminal extension of the His-tag fusion protein contains an enterokinase cleavage site. Cleavage of this peptide bond would result in 10 extra amino acids from the N-terminal extension remaining on the rSQAPI. However, this would be unlikely to pose a problem for resolution of the structure and removal of most of the N-terminal extension facilitates interpretation of NMR experiments. To find the optimal conditions and amount of enterokinase for the removal of the His-tag from the fusion protein several optimization reactions were run. The conditions to be optimized were the amount of enterokinase and the duration of the incubation. Figure 6 shows the digestion of 20 μg of unlabelled His-tagged rSQAPI fusion protein at 37 ^{0}C for 18 hours with varying amounts of enterokinase. The optimal conditions for the digestion of the fusion protein with enterokinase were judged to be 1 unit of EKmax per 20 μg of protein at 37 ^{o}C for 18 hours. The optimization experiment was repeated at 4 ^{o}C with similar results (data not shown).

In the digest containing 0.01, 0.1 and 1 units of enterokinase two products are visible on the gel. This indicates that the enterokinase has cleaved the protein in an unexpected manner. The larger cleaved species is visible in the lanes with less enterokinase, suggesting that the enzyme has higher affinity for the cleavage site that generate this product than the site which generates the lower molecular weight species. Thus the higher molecular weight species arguably arises from cleavage at the expected site and the lower molecular weight species arises from an unanticipated cleavage site. However, as judged from this gel, it is easier to obtain a homogenous sample of the lower molecular weight species. Since homogeneity of the sample is crucial for NMR analysis all subsequent digests were performed with a ratio of 1 unit enterokinase to 20 μg His-tagged rSQAPI fusion protein.

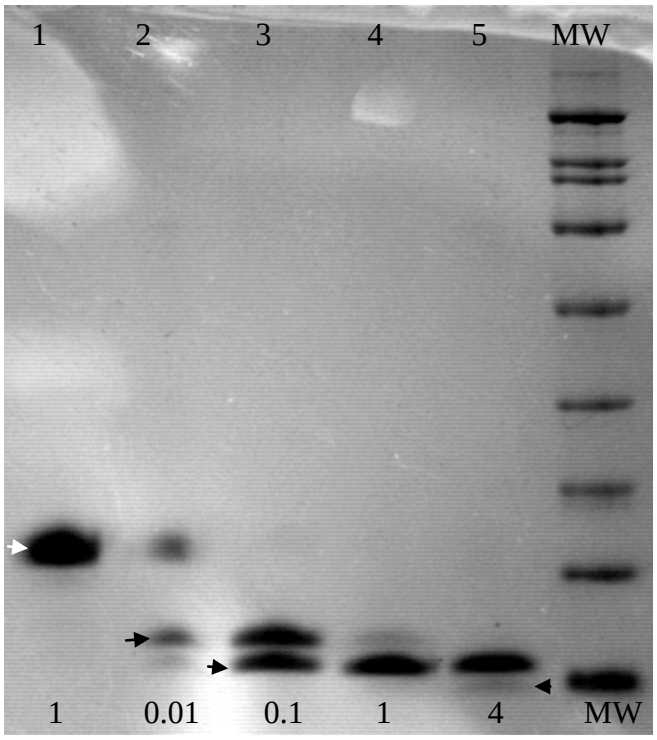


Fig. 6. Optimization of the amount of enterokinase for the digestion of His-tagged rSQAPI fusion protein. Undigested fusion protein (lane 1), digest containing 0.01 units of enterokinase per 20 μg of His-tagged rSQAPI fusion protein (lane 2), a digest containing 0.1 units of enterokinase per 20 μg of His-tagged rSQAPI fusion protein (lane 3), a digest containing 1 unit of enterokinase per 20 μg of fusion protein (lane 4), a digest containing 4 units enterokinase per 20 μg of fusion protein (lane 5), molecular weight marker (lane 6). White arrow indicates His-tagged rSQAPI fusion protein, black arrows indicate cleaved rSQAPI fusion protein.

Removal of enterokinase and cleaved N-terminal extension from the protein sample

It was first thought that the clean up of the cleaved rSQAPI fusion protein would be a two step purification. The first step being the removal of the enterokinase with a commercially available resin (EKaway). Early trials with this resin showed that an unacceptably high amount of the cleaved rSQAPI was lost along with the enterokinase (data not shown). This purification step was therefore deemed unnecessary as the concentration of the un-labelled enzyme in the final protein sample would be low in comparison to the amount of rSQAPI. The second step of the purification would have been to pass the digest through a nickel column (1 ml) to remove the cleaved N-terminal extension. Trials of this purification step revealed that the cleaved rSQAPI fusion protein possessed an intrinsic affinity for the nickel column and required a 300 mM imidazole wash to be eluted from the column. No elution of SQAPI was seen with buffer containing 30 mM imidazole. Fortuitously, the enterokinase showed no affinity for the column and was eluted in the flow-through (Fig. 7). The protocol was therefore simplified to a single step purification which removed both the enterokinase and the N-terminal extension. This purification procedure was not fail safe, however, as some samples contained residual N-terminal extension (as judged from the ^{15}N HSQC) which had to be removed by repeating the nickel column purification. Furthermore, there appears to be some variation between batches of resin, as when a new column was purchased the SQAPI eluted at much lower concentrations of imidazole. This was not investigated further.

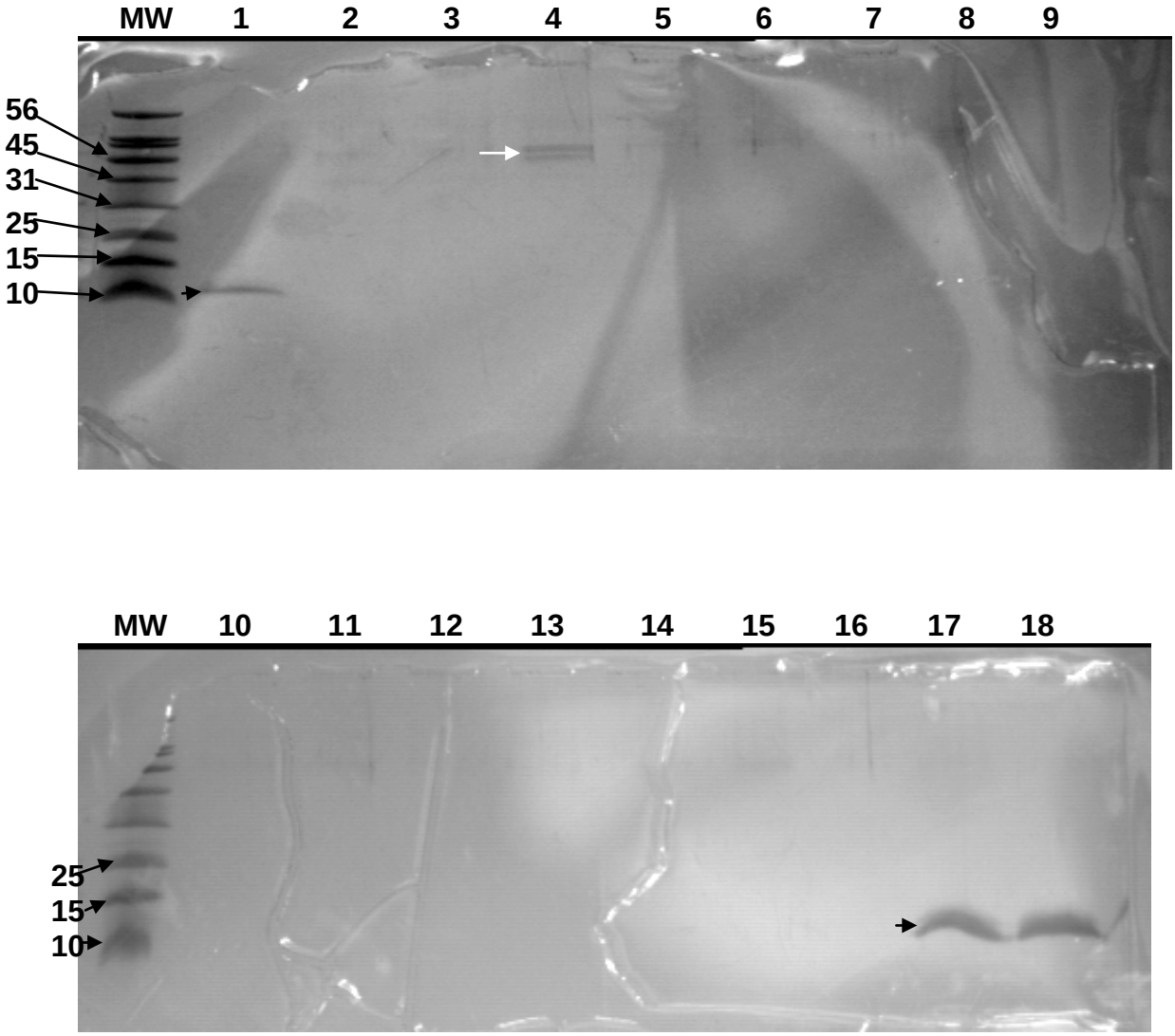


Fig. 7. Removal of the His-tag from the cleaved rSQAPI fusion protein. The unbound fractions (lane 1-8), Buffer B wash (lanes 9 & 10), Buffer W wash (lanes 11-13), Buffer i (containing 30 mM imidazole) wash (lane 14-16), Buffer I (containing 300 mM Imidazole) wash (lanes 17 & 18). Black arrows indicate the cleaved rSQAPI fusion protein, the white arrow indicates the two enterokinase bands (56 kDA and 58 kDa).

Chapter Four: Solution structure of SQAPI at pH3

Introduction

There are at present two techniques for the high resolution structural determination of proteins, X-Ray crystallography and nuclear magnetic resonance spectroscopy (NMR). X-ray crystallography is by far the least labour intensive technique once crystals have been obtained. However, as stated previously, attempts to crystallise SQAPI both by itself and in complex with pepsin have been unsuccessful (Elliot, unpublished). Thus NMR spectroscopy was investigated as an alternative means of structural determination.

Preliminary spectra

Preliminary spectra acquired on unlabelled protein

If a protein is folded in solution it will have a well dispersed ^{1}H one dimensional (1D) spectrum. Preliminary spectra were run on unlabelled solutions of His-tagged rSQAPI fusion protein to examine its suitability for structural determination using NMR. Since the yeast aspartic peptidase inhibitor IA3 is unstructured until in contact with the target aspartic peptidase it was necessary to determine whether or not SQAPI possesses an intrinsically folded structure.

Spectra were collected on an unlabelled sample of 0.2 mM His-tagged rSQAPI fusion protein (in 25 mM glycine-HCl buffer pH 3 containing 0.02% sodium azide and 1 mM EDTA in 95%H_2O/5% $^{2}H_2O$) at 25 ^{o}C. 1D ^{1}H spectra were collected at 700 MHz over 32 scans with a sweep width of 11261 Hz. The spectra revealed that the glycine-HCl buffer was unsuitable for further NMR experiments. The high concentration of glycine resulted in a large peak at 7.65 ppm in the amide region (6-10 ppm) of the 1D spectrum (Fig. 8). All subsequent spectra were acquired using a phosphate buffer. Nevertheless, the well dispersed peaks from 6-10 ppm and 0-4 ppm indicated that SQAPI was folded.

The next two spectra were collected with a 0.06 mM sample of unlabelled His-tagged rSQAPI fusion protein in phosphate buffer (10 mM potassium phosphate buffer pH3;

0.02 % sodium azide; and 1 mM EDTA; in 95% H_2O/5% ^2H_2O) at 25 ^oC. A 1D ^1H spectrum was collected at 700 MHz over 1024 scans with a sweep width of 11261 Hz. A 2D ^1H NOESY spectrum was acquired with 256 increments in the t^1 dimension and 148 scans.

The 1D ^1H spectra of His-tagged rSQAPI fusion proteinin phosphate buffer at pH 3 (Fig. 9) shows well resolved and well dispersed peaks between 0-4 ppm and 6-10 ppm indicating the presence of moderate levels of secondary structure. Although the 2D ^1H NOESY had poor signal to noise ratio, due to the low protein concentration, this spectrum did show well dispersed peaks which indicated the presence of secondary and tertiary structure (data not shown).

As these preliminary spectra run on un-labelled protein suggested that SQAPI was folded when free in solution at pH 3 it was decided that further spectra should be acquired with ^{15}N labelled protein.

Figure 8. A one dimensional ^{1}H spectrum of His-tagged rSQAPI fusion protein in a glycine –HCl buffer at pH 3. The broad peak centred at 7.65 ppm arises from the glycine in the buffer.

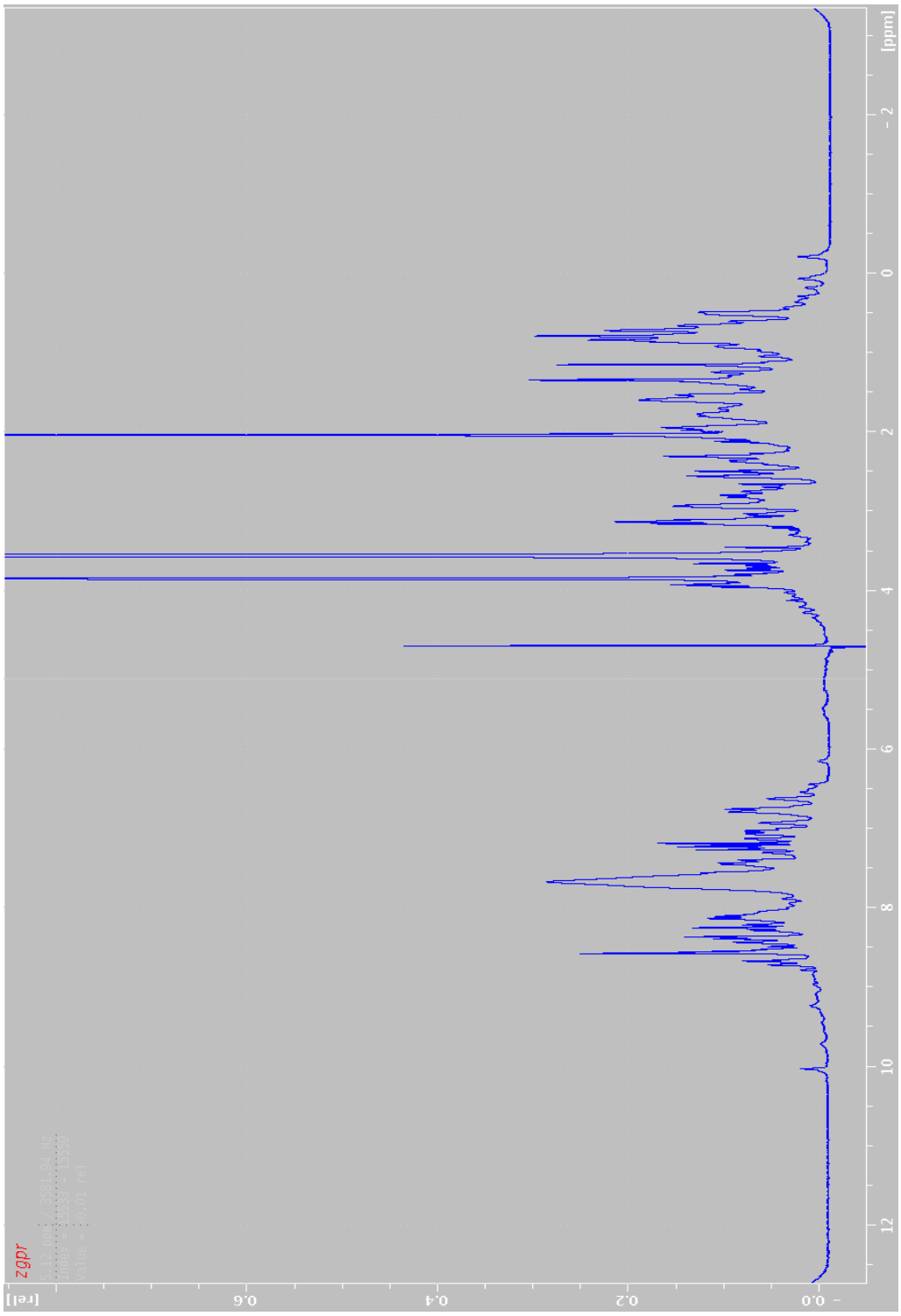
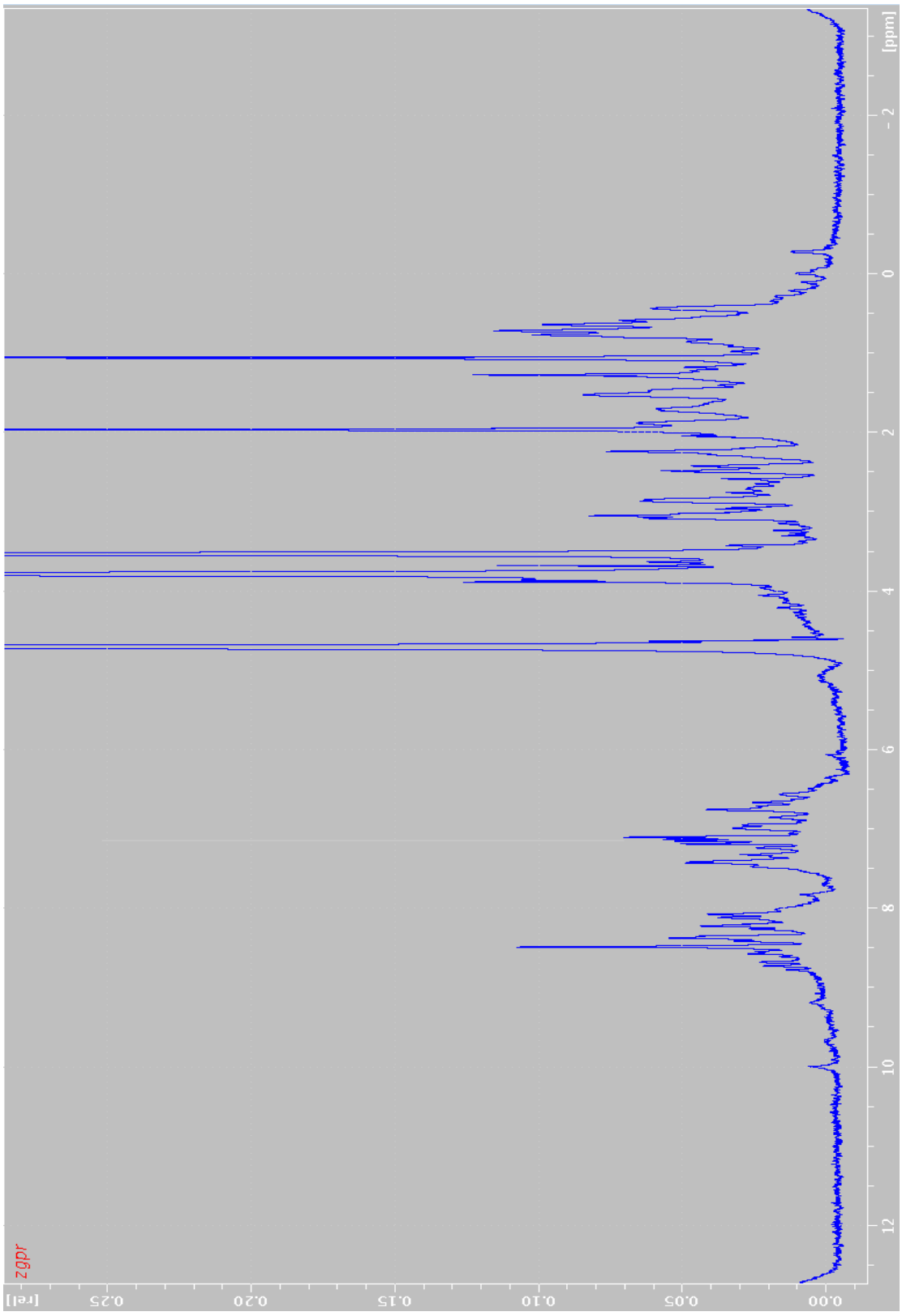

zgpr
[rel]
0.0
0.2
0.4
0.6
12
10
8
6
4
2
0
– 2
[ppm]

Figure 9. A one dimensional ^{1}H spectra of 0.06 mM His-tagged rSQAPI fusion protein in 10 mM phosphate buffer pH 3. The well defined peaks suggest that SQAPI is folded in this buffer and it is suitable for NMR. Data was collected at 37 ^{0}C.


zgpr
[rel]
0.00
0.05
0.10
0.15
0.20
0.25
12
10
8
6
4
2
0
– 2
[ppm]

Preliminary spectra acquired on labelled protein

Preliminary ^{15} N edited HSQC spectra were run on ^{15}N labelled rSQAPI to determine suitable conditions for the rest of the spectra to be run. Only protons bound to a nitrogen atom in the protein produce a peak in this type of spectrum. This includes the side-chains of residues such as tryptophan and histidine, but most importantly the amide protons in the backbone. A protein that has a structure that is amenable to NMR analysis will give an HSQC with well separated peaks.

The ^{15} N edited HSQC for the His-tagged rSQAPI fusion protein (Fig. 10) showed fair dispersion of peaks. However, the cluster of peaks in the centre of the HSQC correspond to regions of the protein that are unstructured (random coil). The overlapping of the peaks makes assignments in this region ambiguous. It was assumed that the N-terminal extension of the His-tagged rSQAPI fusion protein was unstructured as these regions typically are. It was therefore decided to remove the N-terminal extension before we proceed with further NMR experiments. However, the rest of the peaks were well resolved and the spread of the shifts indicated folded protein, consistent with the results obtained with unlabelled protein.

The ^{15} N edited HSQC spectrum of [^{15}N] rSQAPI protein that has had the N-terminal extension removed with enterokinase (Fig 11) shows well dispersed and well resolved peaks. The cluster of peaks in the centre of the HSQC obtained with the fusion protein (Fig. 10) are not present in the HSQC obtained after the N-terminal extension was removed (Fig. 11). This difference is even more obvious when the two spectra are overlaid (Fig. 12). The expected length of the cleaved rSQAPI protein was 106 amino acids, however, six of these are prolines. Prolines are imines, containing no amide proton, and thus do not give rise to peaks in the HSQC. Although this spectrum was later found to be missing signals from some residues, the preliminary count of peaks in this HSQC (99 peaks) was considered to be sufficiently consistent with the expected number of peaks at that time as to proceed with further analysis of the protein. The combined results from the 1D ^{1}H spectra and the ^{15}N HSQC lead to the decision to proceed with the collection of further NMR data at 37^{0}C.

The preliminary spectra obtained with ^{15}N rSQAPI were acquired with the purpose of indicating how easily the structure of SQAPI could be resolved by NMR spectroscopy. The favourable results from these spectra justified the expense of producing ^{15}N ^{13}C labelled protein for collection of further spectra (Table 2).

Fig 10. A ^{15}N hetero-nuclear quantum correlation (HSQC) spectrum of the 0.2 mM His-tagged rSQAPI fusion protein in phosphate buffer. The proton frequency is along the x-axis and the nitrogen frequency is along the y-axis.

Fig 11. A ^{15}N- HSQC ran at 37 ^{0}C on a 0.2 mM rSQAPI sample in phosphate buffer pH 3 after the N-terminal extension had been removed with enterokinase.

Fig 12. An overlay of the HSQCs shown in Fig. 10 and Fig 11. This clearly shows the improvement of the spectra after the N-terminal extension was removed.

Table 2 Spectra acquired for structural determination of SQAPI

Spectra type	Sample	Temperature (^{0}C)	Comments
nitrogen edited HSQC	0.15 mM [13C][15N]	37	
HNCACB	0.4 mM [13C][15N]	37	Impure sample[1]
HNCO	0.4 mM [13C][15N]	37	
HNCO	0.15 mM [13C][15N]	37	
HNCACO	0.15 mM [13C][15N]	37	
HNCA	0.15 mM [13C][15N]	37	
HN(CO)CACB	0.15 mM [13C][15N]	37	
HN(CO)CACB	0.15 mM [13C][15N]	37	
HCCH-TOCSY	0.15 mM [13C][15N]	37	
HCC(CO)NH-TOCSY	0.15 mM [13C][15N]	37	
13C-edited HSQC	0.15 mM [13C][15N]	37	
13C-edited NOESY-HSQC	0.15 mM [13C][15N]	37	poor signal to noise
15N-edited NOESY-HSQC	0.15 mM [13C][15N]	37	
15N-HSQC	0.15 mM [13C][15N]	10	
15N-HSQC	0.15 mM [13C][15N]	25	
15N-HSQC	0.15 mM [13C][15N]	40	
15N-HSQC	0.15 mM [13C][15N]	50	
HNCA	0.15 mM [13C][15N]	50	
HNCO	0.15 mM [13C][15N]	50	
HN(CO)CACB	0.15 mM [13C][15N]	50	
HN(CO)HAHB	0.15 mM [13C][15N]	50	
HCCCONH-TOCSY	0.15 mM [13C][15N]	50	poor signal
HCCH-TOCSY	0.15 mM [13C][15N]	50	
15N NOESY-HSQC	0.6mM [13C][15N]	50	
13C NOESY-HSQC	0.6mM [13C][15N]	50	
IPAP	0.6mM [13C][15N]	25	
IPAP	0.6mM [13C][15N]	25	alignment media
HNHA	0.6mM [13C][15N]	50	

Impure sample[1] The HSQC of this sample showed extra peaks and was purified by nickel column before collecting more spectra.

Assignments

Backbone Assignments

The ^{1}H-^{15}N hetero-nuclear shift correlated peaks were picked on a [^{1}H-^{15}N] HSQC spectrum and this peak list was transferred to the HNCACB. Sequential residues were identified by correlation of the frequencies of the $C^{\alpha\,(i)}$, $C^{\beta\,(i)}$ and $C^{\alpha\,(i-1)}$, $C^{\beta\,(i-1)}$ peaks between strips. The HNCA, HNCO, HN(CO)CACB and HN(CA)CO and in some cases the ^{15}N NOESY-HSQC were used to resolve ambiguities. During the backbone assignments it became apparent that the peaks corresponding to the region between histidine 51 and serine 61 were not visible. This region of the backbone corresponded to the proposed binding loop on the model of SQAPI (Christeller *et al.*, 2006). In some cases mobile regions of proteins may not be visible in NMR spectra. For example the loop may move at a rate of 700 MHz, the same frequency of the spectrometer and therefore be invisible. Two methods were used to attempt to find the missing peaks: (i) an HSQC and HNCO were acquired using a 500 MHz spectrometer and (ii) a series of HSQC spectra were recorded at increasing temperature. The data collected on the 500 MHz spectrometer was of poor quality due to the low protein concentration of the sample. However, the temperature trial proved to be fruitful. A faint peak on the HSQC corresponding to residue 62 increased in intensity when the spectrum was acquired at higher temperatures and further peaks appeared. These peaks disappeared when the temperature was lowered back to 37 ^{0}C and were therefore not due to degradation. This result is shown clearly in an overlay of four ^{15}N edited HSQCs collected at different temperatures (Fig. 13). An HNCO was run at 50 ^{0}C and it was discovered that very weak peaks that had been dismissed as noise on the HNCO at 37 ^{0}C were actually the peaks that corresponded to the missing residues. The result of this trial lead to the acquisition of all spectra at the new temperature of 50 ^{0}C (Table 2). These spectra were used to confirm the assignments already made at 37 ^{o}C and complete the rest of the backbone assignments.

It worth noting at this point that SQAPI inhibits pepsin in activity assays performed at $50^{0}C$ (Laing, unpublished).

The specific identity of the sequential residues was established by entering the chemical shifts from the HNCACB, HNCA and HNCO run at 37 ^{0}C into an online database RESCUE2 (Marin *et al.*, 2004). This program identifies what each residue is likely to be based on the chemical shifts. These possibilities were then matched to the SQAPI sequence.

Fig 13. An overlay of 4 HSQCs run at 10^0C, 25^0C, 40^0C and 50^0C. The HSQC collected at 50^0C shows the greatest improvement in peak shape and intensity. This series of HSQCs were collected on a [^{15}N ^{13}C] double labelled 0.15 mM rSQAPI sample in phosphate buffer.

Vector encoded leader sequence assignments

As mentioned previously the enterokinase was expected to cleave the His-tag ten residues upstream from the beginning of the native SQAPI sequence. The remaining residues of the N-terminal extension, if unstructured would produce intense peaks in the spectra that would be easily assigned. However, no such peaks corresponding to these residues were found. N-terminal backbone assignments ceased at the first native sequence residue, a methionine that looks back to an arginine. The results of the optimization trials for enterokinase cleavage did show evidence for cleavage at more than one site in the fusion protein (Fig. 6). To determine where the enterokinase had cleaved the protein, an aliquot of the ^{15}N rSQAPI sample that was used to generate the HSQC shown in figure 11 was used for N-terminal sequencing. This revealed that 70% of the rSQAPI had been cleaved at the arginine prior to the native SQAPI sequence and the remaining 30 % had been cleaved at the methionine of the native sequence. The N-terminal sequencing results confirmed the suspicion that the enterokinase had not cleaved where it was expected to and why there were no peaks corresponding to the vector encoded leader sequence. Furthermore, the single residue difference in length and the cis-trans isomerisation of the prolines near the N-terminus accounts for the extra peaks in the HSQC.

Side-chain assignments

The assigned ^{1}H-^{15}N hetero-nuclear shifts were then used to assign the H^{α} and H^{β} chemical shifts using the HBHA(CO)NH spectrum. The remaining side chain ^{1}H and ^{13}C frequencies were assigned using both the low temperature and high temperature HCC(CO)NH-TOCSY, HCCH-TOCSY, ^{13}C-HSQC and ^{13}C-NOESY-HSQC spectra where possible. The HCC(CO)NH spectra at both temperatures had very poor signal/noise and thus it was necessary to use the HCCH-TOCSYs for assignments. All of the aromatic residues except histidine 51 were assigned by first finding the Cα and Cβ strips of the residue on the ^{13}C-NOESY-HSQC and then finding the strips that correspond to the side chain carbons that are further from the polypeptide backbone. These assignments were assisted with information from the ^{15}N-NOESY-HSQC. Histidine 51 remains unassigned as it was not visible on the ^{13}C-NOESY-

HSQC. Attempts were made to acquire ^{13}C-NOESY-HSQC specifically centred on the aromatic region but these were unusable due to significant noise in the t_1 dimension.

Chemical shifts

The chemical shifts assigned as described above have been submitted to the Biological Magnetic Resonance Bank (BMRB, Marin *et al*., 2004) with the accession number 15483. The server responded that the chemical shift file contained 665 proton chemical shifts, 405 ^{13}C chemical shifts and 90 ^{15}N chemical shifts. Preliminary analysis with CANDID found the atom assignments to be 88.89% complete. Table 3 shows the atom assignments that were missing. Since the atom assignments must be more than 90% complete to continue with structural calculations further efforts are required to exceed this limit (as described in the future work section). Ideally the chemical shifts should be as complete as possible as incomplete assignments result in incorrect structures. However, once the residual dipolar couplings have been calculated these will be used to refine the NOESY derived structure (see below).

Furthermore, the atom assignments must agree with the two NOESY spectra. The ^{13}C NOESY has 62.15 % agreement with the atom assignments and the ^{15}N NOESY has 58.46 % agreement. This difference is probably due to the peaks being just outside the tolerance range that CANDID was searching and can be corrected manually.

Table 3: Abridged CANDID output of missing assignments

Residue	Amino acid	Missing assignments	Residue	Amino Acid	Missing assignments
			73	TYR	CG CZ CE2 CD2
2	PRO	QG QD			
8	GLU	CD	74	ASN	CG QD2
15	ASN	CG QD2	79	GLU	CG CD
16	ASP	CG	80	LYS	QZ
18	ARG	HE CZ QH1 QH2	83	ASP	CG
			84	ASN	CG QD2
20	LYS	QE QZ	87	LYS	QZ
21	GLU	CD	91	PHE	CG CZ CE2 CD2
24	GLU	CD	95	PHE	CG CZ CE2 CD2
25	PHE	CG CZ CE2 CD2			
28	LYS	QZ			
29	GLN	CD QE2			
30	HIS	CG HD1			
32	GLU	CD			
33	GLN	CD QE2			
34	ASN	CG QD2			
41	ASP	CG			
44	GLN	CD QE2			
46	ILE	QD1			
47	LYS	QZ			
49	ILE	QG1			
51	HIS	CG CD2 HD1 CE1 HD2 HE1			
52	TRP	CG CD2 CE2			
53	ASP	CG			
54	ASN	CG QD2			
55	TYR	CG CZ CE2 CD2			
56	TYR	CG CZ CE2 CD2			
57	ASN	CG QD2			
63	LYS	QZ			
64	HIS	CG HD1			
67	HIS	CG HD1			
68	GLU	CD			
69	PHE	CG CZ CE2 CD2			
71	LYS	QZ			
72	PHE	CZ CE2			

Summary

Requirement 1

89 missing chemical shifts.

Completeness of all atom assignment is 88.89 %.

30 missing proton chemical shifts.

Completeness of proton atom assignment is 92.75 %.

59 missing heavy atom chemical shifts.

Completeness of heavy atom assignment is 84.75 %.

Requirement 2

Chemical shift agreement after automatic calibration for 13C_noesy_hsqc_323_3.3D.16: 62.15 %.

Chemical shift agreement after automatic calibration for 15N-noesy-hsqc_323_3.3D.16: 58.46 %.

ϕ angle restraints

The HNHA is a three dimensional spectrum that correlates the $^{1}H^{N}$, ^{15}N and $^{1}H^{\alpha}$. The three dimensional J-coupling ($^{3}J_{H}{}^{N}{}_{H}{}^{\alpha}$) between these atoms can be calculated from the intensity ratio of the diagonal peak to the cross peak (Kuboniwa *et al*., 1994).

During a constant time period (T) in the HNHA the $^{3}J_{H}{}^{N}{}_{H}{}^{\alpha}$ coupling is allowed to evolve whereby some of the magnetization is transferred to the $^{1}H^{\alpha}$ spin and the rest remains on the $^{1}H^{N}$. The magnetization on the $^{1}H^{\alpha}$ resonates at the $^{1}H^{\alpha}$ frequency and forms the cross peak. The magnetization remaining on the $^{1}H^{N}$ resonates at the $^{1}H^{N}$ frequency and forms the diagonal peak. The magnetization is then transferred back to the $^{1}H^{N}$ and refocused prior to acquisition. The dependence on signal intensity for the diagonal (I_D) and cross peaks (I_X) can be expressed mathematically as $\cos^2(\pi^3 J_{H}{}^{N}{}_{H}{}^{\alpha}\ T)$ and $\sin^2(\pi^3 J_{H}{}^{N}{}_{H}{}^{\alpha}\ T)$ respectively (Kuboniwa *et al*., 1994). Thus it is possible to determine the phi angle restraints from the formula:

$$I_X/I_D = \tan^2(\pi\ ^3 J_{H}{}^{N}{}_{H}{}^{\alpha}\ T)$$

If the calculated $^{3}J_{H}{}^{N}{}_{H}{}^{\alpha}$ is less than 6, the residue is deemed to be in a α-helix. If the value is greater than 8 the residue is classified as straight chain and if the calculated $^{3}J_{H}{}^{N}{}_{H}{}^{\alpha}$ is between 6 and 8 no constraints are placed on the backbone.

The results from the HNHA analysis (Table 4) indicate that residues 16-29 and 42-46 are in an alpha helix. The 16-29 α-helix is reasonably consistent with the model of SQAPI structure by Christeller *et al*. (2006) which predicts residues 18-38 form an α-helix. The 42-46 α-helix and the small amount of straight chain differs from the model.

Table 4. ϕ angle restraints as calculated from the HNHA

Residue	ϕ Angle Restraint	Residue	ϕ Angle Restraint
3	No constraints	52	No constraints
5	No constraints	53	Alpha helical
6	No constraints	54	No constraints
7	Alpha helical	55	Straight chain
9	No constraints	56	No constraints
10	Straight chain	57	Straight chain
11	Alpha helical	58	Straight chain
12	No constraints	59	Alpha helical
13	Straight chain	60	Straight chain
14	Alpha helical	61	No constraints
15	Straight chain	62	No constraints
16	Alpha helical	63	Straight chain
18	Alpha helical	64	No constraints
19	Alpha helical	65	Alpha helical
20	Alpha helical	67	No constraints
21	Alpha helical	68	No constraints
22	Alpha helical	69	Alpha helical
23	Alpha helical	71	No constraints
24	Alpha helical	72	No constraints
26	Alpha helical	73	No constraints
27	Alpha helical	74	No constraints
28	Alpha helical	75	No constraints
29	Alpha helical	76	No constraints
30	No constraints	77	Straight chain
31	Alpha helical	78	Straight chain
33	No constraints	79	Alpha helical
34	Alpha helical	80	Alpha helical
35	No constraints	82	Alpha helical
36	No constraints	83	Straight chain
38	Straight chain	84	Alpha helical
39	Alpha helical	85	No constraints
40	Straight chain	87	No constraints
41	No constraints	88	Alpha helical
42	Alpha helical	89	Straight chain
43	Alpha helical	90	Alpha helical
44	Alpha helical	91	No constraints
46	Alpha helical	94	No constraints
47	Straight chain	95	Alpha helical
48	Alpha helical		
49	No constraints		
51	No constraints		

Deuterium exchange

The rate of exchange between the backbone amides and the solvent was measured by lyophilising 250 µl of a 0.2 mM sample of [^{15}N] rSQAPI and re-hydrating it in $^{2}H_2$0. A series of HSQC experiments were then run. An overlay of these spectra is shown in figure 14 A, along with an overlay with a HSQC acquired before lyophilisation of the sample (Fig. 14 B).

The [^{15} N -^{1} H] peaks corresponding to amide protons that are in slow exchange with the solvent are visible on the HSQC for several minutes after the addition of $^{2}H_2O$.
SQAPI consists of 95 residues, 6 of which are prolines that do not produce a peak on the HSQC. Of the 89 peaks of the HSQC, 64 persist in the presence of $^{2}H_2O$ after 52 minutes. After four hours only four of these peaks have exchanged with the solvent and 60 peaks remain. This would suggest that these residues are shielded from the solvent or are involved in hydrogen bonding.

The results of this analysis also allow a preliminary evaluation of the model of SQAPI proposed by Christeller *et al.* (2006) on the basis of sequence similarity between SQAPI and oryzacystatin.

The results from the solvent exchange experiment indicate the SQAPI structure is more globular than open. Of the 89 HSQC peaks, 60 peaks persist over four hours after re-suspending the lyophilised SQAPI sample in $^{2}H_2O$. Most of the amide protons corresponding to these peaks are involved in the anti-parallel β–sheet and the α –helix in the model.

Fig 14. Panel A: HSQC of lyophilised ^{15}N rSQAPI fusion protein that has been dissolved in $^{2}H_2O$. Panel B: An overlay of the HSQC shown in panel A and the HSQC shown in figure 10.

Residual dipolar couplings

In isotropic solution the dipole-dipole interactions of nuclei within a molecule average to zero due to molecular tumbling. However, if a molecule has a tendency to align in a magnetic field the dipolar term no-longer averages to zero and a small residual dipolar coupling can be observed. This effect can also be exploited in proteins. Tolman *et al*. (1995) demonstrated that the paramagnetic protein cyanometmyoglobulin has a preference for a certain orientation when placed in a magnetic field. This allows the residual dipole-dipole effects to be measured. Furthermore, a weak alignment can be induced with non-magnetic proteins by dissolving them in liquid crystal media that aligns in the magnetic field. In the partially aligned molecule the contribution of the dipole coupling does not average to zero, this is the residual dipolar coupling (RDC). The size of the RDC gives information about the associated bond's orientation with respect to the overall alignment tensor of the entire molecule.

RDCs are very useful for determining the orientation of domains within multi-domain proteins and the relative orientation of components of a complex. However the RDC data collected in this project will be used to refine the NOE data.

There are several types of liquid crystal media such as virus capsids, membrane particles and bicelles (Clore *et al*., 1998, Koenig *et al*., 1999, Bax & Tjandra, 1997). In this experiment (Fig. 15) Otting media was used, which consists of synthetic liquid crystals and the IPAP were collected according the method of Cordier *et al*. (1999) (Ruckert and Otting, 2000).

Figure15. The [in-phase plus anti-phase] (A) and [in-phase minus anti-phase] (B) HSQCs collected on SQAPI both out of alignment media (isotropic) and in alignment media (anisotropic). Note the splitting of the peaks in the nitrogen dimension. In the isotropic solution this splitting represents J, the single bond coupling. In the anisotropic solution the splitting represents J+D, the single bond coupling plus the residual dipolar coupling. The first methionine of the native SQAPI sequence is numbered zero in these spectra.

Discussion

The overall aim of the research described in this thesis is to resolve the tertiary structure of SQAPI using NMR. This has not been achieved within the available time however very good progress has been made towards it. The chemical shifts are at this stage 88.89% complete (the minimum threshold for CANDID is 90%) and the information required for the calculation of the residual dipolar couplings have been obtained. Furthermore, the results obtained so far have provided valuable information about the structure of SQAPI.

Both the ^{1}H 1 dimensional spectra obtained with un-labelled His-tagged rSQAPI fusion protein and the ^{15}N-HSQC obtained with cleaved fusion protein indicated that SQAPI is folded when free in solution at pH 3, unlike the yeast aspartic peptidase inhibitor IA3.

Previous to this study, it was unknown whether SQAPI acted as a dimer or a monomer when inhibiting pepsin. It was known that SQAPI inhibits pepsin at a 1:1 molar ratio but SQAPI behaved as a dimer on gel filtration chromatography at pH 7. The appearance of the peaks on the NMR spectra produced by SQAPI all indicate that it behaves as a monomer at pH 3 (Dr. Stephen Headey, unpublished). Since SQAPI inhibits pepsin a pH 3 this suggests that SQAPI is active in the monomeric form.

The results of this analysis also allow a preliminary evaluation of the model of SQAPI proposed by Christeller *et al.* (2006) on the basis of sequence similarity between SQAPI and oryzacystatin.

A stretch of the backbone from histidine 51 to serine 61 could not be assigned from any of the spectra obtained at 37 ^{0}C. This region corresponds to a proposed binding loop on the model of SQAPI. It is likely that these peaks do not appear on the spectra collected at 37 ^{0}C due to the dynamics of the movement of this region of the protein. The existence of the loop is further supported by the fact that it was possible to assign this region of the backbone from spectra obtained at 50 ^{0}C as the elevated temperature would alter the rate of movement of the loop.

The method of Kuboniwa *et al.* (1999) was used to calculate the ϕ angle restraints of the backbone of SQAPI from the HNHA spectrum. The estimate of the secondary structure of SQAPI generated from this method indicates that residues 16-29 are involved in an α-helix however the model of SQAPI structure has an α-helix from residues 18 – 38. The ϕ–angle restraints indicate that SQAPI has some straight chain secondary structure which does not agree with the model. Although the ϕ–angle restraints are only an estimate they are a more accurate representation of the secondary structure of SQAPI than the model. This is because they are based from data collected directly from the protein itself, rather than from similarity between the sequences of two proteins.

The results from the solvent exchange experiment indicate that the SQAPI structure is more globular than open. Of the 89 HSQC peaks, 60 peaks persist over four hours after re-suspending the lyophilised SQAPI sample in $^{2}H_2O$. Most of the amide protons corresponding to these peaks are involved in the anti-parallel β–sheet and the α –helix in the model.

Future work

Structural calculations in CANDID

Most of the necessary information to start structural calculations has been obtained and this stage has already been started by Dr. Stephen Headey. A structural calculation however, is only as accurate as the information used to generate it. The minimum threshold for atom assignments in CANDID is 90% complete; currently the shifts are 88.89% complete. The table of missing assignments (Table X) shows that one or the other of the δ–carbons (CD1 or CD2) and ε-carbons (CE1 or CE2) of the aromatic ring in the phenylalanines and tyrosines are missing from the chemical shift file. As these pairs of carbons are degenerate it is simply a case of entering the chemical shift of the known carbon into the degenerate carbon position in the file. Furthermore, upon inspection of the hand written record of chemical shifts it was found that a few chemical shifts were known but were not entered into the chemical shifts file. These small corrections were made to the chemical shift file and it has been resubmitted for analysis.

The second obstacle to the structural calculation is the low agreement between the two NOESY spectra and the chemical shifts list. This difference is probably due to the peaks being just outside the tolerance range that CANDID was searching and can be corrected manually.

Residual dipolar couplings

The necessary spectra have been obtained for the calculation of the residual dipolar couplings however it has not been analysed.

Mechanism of action of SQAPI

Once the structure of SQAPI has been solved, it will make information gained from site directed mutagenesis variants more meaningful. When residues in a protein are changed by site directed mutagenesis it is crucial to know what potential effect that change can have on the structure of the protein and take that into account when assessing the effect

of the mutations on the activity. For example, if a residue was mutated and this abrogated the activity of SQAPI this residue could be apart of the active site or could play an important structural role. When the structure of SQAPI is solved, it will allow for specific targeting of residues that are likely to be involved in the active site.

When the structure of SQAPI is solved the next logical step is to determine the structure of SQAPI in complex with pepsin. This will provide information about how SQAPI inhibits its target enzyme. Since attempts to co-crystallise SQAPI with pepsin failed (Elliot, unpublished) it was necessary to first solve the structure of SQAPI by NMR before attempting to solve the SQAPI: pepsin complex.

References

Abu-Erreish, G. M. & R. J. Peanasky (1974). "Pepsin inhibitors from ascaris-lumbricoides – Isolation, purification and some properties." Journal of Biological Chemistry 249(5): 1558-1565.

Barret, A. J., Rawlings, N. D. & O'Brien, E. A. (2001). "The MEROPS database as a protease information system." Journal of Structural Biology 134: 95-102.

Barrett, A. J., Rawlings, N. D., & Woessner, J. F. Eds. (1999). Introduction, in Handbook of proteolytic enzymes, Academic press.

Barret, A. J., Ritnja, A., & Buttle, D.J. (1990). "The amino acid sequence of a novel inhibitor of cathepsin D from potato." FEBS Letters 267: 13-15.

Bjork, I. & Ylinenjarvi, K. (1990). "Equilibrium and kinetic studies of the interaction of chicken cystatin with with four cysteine proteinases." Biochemistry 371: 119-124.

Bochtler, M., Ditzel, L., Groll, M., Hartmann, C., & Huber, R (1999). "The proteasome." Annual Review of Biophysics and Biomolecular Structure 28: 295-+.

Cai, M. L., Huang, Y., Sakaguchi, K., Clore, G. M., Gronenborn, A. M., & Craigie, R. (1998). "An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli." Journal of Biomolecular NMR 11(1): 97-102.

Cater, S. A., Lees, W. E., Hill, J., Brzin, J., Kay, J., & Phylip, L. H. (2002). "Aspartic proteinase inhibitors from tomato and potato are more potent against yeast proteinase A than cathepsin D." Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1596(1): 76-82.

Christeller, J. T., Farley, P. C., Marshall, R. K., Anandan, A., Wright, M. M., Newcomb, R. D., *et al*. (2006). "The squash aspartic proteinase inhibitor SQAPI is widely present in the cucurbitales, comprises a small multigene family, and is a member of the phytocystatin family." Journal of Molecular Evolution 63(6): 747-757.

Christeller, J. T., Farley, P. C., Ramsay, R. J., Sullivan, P. A., & Laing, W. A. (1998). "Purification, characterization and cloning of an aspartic proteinase inhibitor from squash phloem exudate." European Journal of Biochemistry 254(1): 160-167.

Clore, G. M., Starich, M. R., & Gronenborn, A. M. (1998). "Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses." Journal of the American Chemical Society 120(40): 10571-10572.

Cordier, F., Dingley, A. J., & Grzeseik, S. (1999). "A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and and dipolar one-bond J-couplings." Journal of Biomolecular NMR 13(2): 175-180.

Dash, C., Kulkarni, A., Dunn, B., & Rao, M. (2003). "Aspartic peptidase inhibitors: Implications in drug development." Critical Reviews in Biochemistry and Molecular Biology 38(2): 89-119.

Farley, P. C., Christeller, J. T., Sullivan, M. E., Sullivan, P. A., & Laing, W. A. (2002). "Analysis of the interaction between the aspartic peptidase inhibitor SQAPI and aspartic peptidases using surface plasmon resonance." Journal of Molecular Recognition 15(3): 135-144.

Galesa, K., Pain, R., Jongsma, M. A., Turk, V., & Lenarcic, B. (2003). "Structural characterization of thyroglobulin type-1 domains of equistatin." FEBS Letters 539(1-3): 120-124.

Galleschi, L., Friggeri, M., Repiccioli, R., & Come, D. (1993). Aspartic proteinases from wheat: some properties. Proceedings of the Fourth International Workshop of Seeds; Basic and applied aspects of seed biology, Angers, France.

Green, T. B., Ganesh, O., Perry, K., Smith, L., Phylip, L. H., Logan, T. M., *et al*. (2004). "IA(3), an aspartic proteinase inhibitor from Saccharomyces cerevisiae, is intrinsically unstructured in solution." Biochemistry 43(14): 4071-4081.

Huntington, J. A., McCoy, A., Belzar, K. J., Pei, X. Y., Gettins, P. G. W., & Carrell, R. W. (2000). "The conformational activation of antithrombin - A 2.85-angstrom structure of a fluorescein derivative reveals an electrostatic link between the hinge and heparin binding regions." Journal of Biological Chemistry 275(20): 15377-15383.

Keilova, H. & V. Tomasek (1976). "Isolation and some properties of cathepsin-D inhibitor from potatoes." Collection of Czechoslovak Chemical Communications 41(2): 489-497.

Koenig, B. W., Hu, J. S., Ottiger, M., Bose, S., Hendler, R. W., & Bax, A. (1999). "NMR measurement of dipolar couplings in proteins aligned by transient binding to purple membrane fragments." Journal of the American Chemical Society 121(6): 1385-1386.

Kuboniwa, H., Grzesiek, S., Delaglio, F., & Bax, A. (1994). "Measurement of H-N-H-ALPHA J-couplings in calcium-free calmodulin using new 2D and 3D Water-flip-back methods." Journal of Biomolecular NMR 4(6): 871-878.

Laing, W. A. & M. T. McManus (2002). "Proteinase inhibitors." Annual Plant Review 7: 77-119.

Laskowski, M. & I. Kato (1980). "Protein inhibitors of proteinases." Annual Review of Biochemistry 49: 593-626.

Lenarcic, B., Ritonja, A., Strukelj, B., Turk, B., & Turk, V. (1997). "Equistatin, a new inhibitor of cysteine proteinases from Actinia equina, is structurally related to thyroglobulin type-1 domain." Journal of Biological Chemistry 272(21): 13899-13903.

Lenarcic, B. & V. Turk (1999). "Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D." Journal of Biological Chemistry 274(2): 563-566.

Li, M., Phylip, L. H., Lees, W. E., Winther, J. R., Dunn, B. M., Wlodaawer, A., *et al*. (2000). "The aspartic proteinase from Saccharomyces cerevisiae folds its own inhibitor into a helix." Nature Structural Biology 7(2): 113-117.

Mares, M., Meloun, B., Pavlik, M., Kostka, V., & Baudys, M. (1989). "Primary structure of cathepsin-D inhibitor from potatoes and it's structure relationship to soybean trypsin inhibitor family." FEBS Letters 251(1-2): 94-98.

Margis, R., Reis, E. M., & Villeret, V. (1998). "Structural and phylogenetic relationships among plant and animal cystatins." Archives of Biochemistry and Biophysics 359(1): 24-30.

Marin, A., Malliavin, T. E., Nicolas, P., & Delsuc, M. A. (2004). "From NMR chemical shifts to amino acid types: Investigation of the predictive power carried by nuclei." Journal of Biomolecular NMR 30(1): 47-60.

Martzen, M. R., McMullen, B. A., Smith, N. E., Fujikawa, K., & Peanasky, R. J. (1990). "Primary structure of the major pepsin inhibitor from the intestinal parasitic nematode Ascaris-suum." Biochemistry 29(32): 7366-7372.

Matsui, M., Fowler, J. H., & Walling, L. L. (2006). "Leucine aminopeptidases: diversity in structure and function." Biological Chemistry 387(12): 1535-1544.

McManus, M. T., Ryan, S. N., & Laing, W. A. (2000). "Proteinase inhibitors as storage proteins in seeds." Seed Research NZ 12: 3-14.

Nagata, K., Kudo, N., Abe, K., Arai, S., & Tanokura, M. (2000). "Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, Oryza sativa L. japonica." Biochemistry 39(48): 14753-14760.

Ng, K. K. S., Petersen, J. F. W., Cherney, M. M., Garen, C., Zalatoris, J. J., Rao-Naik, C., *et al.* (2000). "Structural basis for the inhibition of porcine pepsin by Ascaris pepsin inhibitor-3." Nature Structural Biology 7(8): 653-657.

Oda, K., Koyama, T., & Murao, S. (1979). "Purification and properties of a proteinaceous metallo-proteinase inhibitor from Streptomyces nigrescens TK-23." Biochimica Biophysica Acta 571: 147-156.

Phylip, L. H., Lees, W. E., Brownsey, B. G., Bur, D., Dunn, B. M., Winther, J. R., *et al.* (2001). "The potency and specificity of the interaction between the IA(3) inhibitor and its target aspartic proteinase from Saccharomyces cerevisiae." Journal of Biological Chemistry 276(3): 2023-2030.

Rancour, J. M. & C. A. Ryan (1968). "Isolation of a carboxypeptidase B inhibitor from potatoes." Archive Biochemica Biophysica 125: 380-383.

Rassam, M. & W. A. Laing (2006). "The interaction of the 11S globulin-like protein of kiwifruit seeds with pepsin." Plant Science 171(6): 663-669.

Rawlings, N. D., Morton, F. R., & Barrett, A. J. (2005). "MEROPS: the peptidase database." FEBS Journal 272: 139-139.

Rawlings, N. D., Tolle, D. P., & Barrett, A. J. (2004). "Evolutionary families of peptidase inhibitors." Biochemical Journal 378: 705-716.

Ruckert, M. & G. Otting (2000). "Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments." Journal of the American Chemical Society 122(32): 7793-7797.

Schu, P. & D. H. Wolf (1991). "The proteinase YSCA-Inhibitor, IA3, Gene - studies of cytoplasmic proteinase- inhibitor deficiency on yeast physiology." FEBS Letters 283(1): 78-84.

Seeram, S. S., Hiraga, K., & Oda, K. (1997). "Resynthesis of reactive site peptide bond and temporary inhibition of Streptomyces metalloproteinase inhibitor." Journal of Biochemistry 122(4): 788-794.

Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G. W., *et al*. (2001). "The serpins are an expanding superfamily of structurally similar but functionally diverse proteins - Evolution, mechanism of inhibition, novel functions, and a revised nomenclature." Journal of Biological Chemistry 276(36): 33293-33296.

Strukelj, B., Pungercar, J., Mesko, P., Barlicmaganja, D., Gubensek, F., Kregar, I., *et al.* (1992). "Characterization of Aspartic proteinase-inhibitors from potato at the gene, cDNA an protein levels" Biological Chemistry Hoppe-Seyler 373(7): 477-482.

Tjandra, N. & A. Bax (1997). "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium." Science 278(5340): 1111-1114.

Tolman, J. R., Flanagan, J. M. & Kennedy, M. A., (1995). Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution, Proceedings of the National Academy of Science. 92: 9279-83.

Turk, B., Turk, V. & Turk, D., (1997). "Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors." FEBS Letters 378(3-4): 141-150.

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., *et al*. (1995). "H-1, C-13 and N-15 Chemical-Shift Referencing in Biomolecular NMR." Journal of Biomolecular NMR 6(2): 135-140.