Structure, Function and Interaction of MIC2/M2AP complex from *Toxoplasma gondii*

Bing Liu

Imperial College London

2005-2009

Supervisor: Prof. S.J. Matthews

Submitted (2009) in partial fulfillment of the requirement for the degree of Doctor of Philosophy

“Time is a bird forever on the wing.”

T.W. Robertson

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Steve Matthews, who continuously sparking me with new ideas and providing me endless supports. I would also like to thanks my PhD advisor, Dr. Jon Nield, Prof Xiaodong Zhang and Dr. Ernesto Cota for their advices throughout my project.

I'd like to thank all the people from Steve Matthews Lab, for their valuable advice and joyful company. Especially, Pete Simpson, whose NMR knowledge as well as jokes are seems to be endless, deserves the most great thanks. Dr. Savvas Saouros, Dr. Kovilen Sawmynaden and Dr. Tharin Blumenscein shared their knowledge within the project. Jan Marchant, the software expert of the group, dedicated his time helped during the final stage of the structure calculations. And Yi Yang, the other PhD student of the same year, always reminds me the real tastes of home food in London restaurants.

I would also thank my parents for their love and support over the years, who always teach me how to deal with relationships and "try one more time". I'd like to thank my younger sister Jia, who always reminds me the happiest family times.

At last, I would like to thank University of Science and Technology of China, Singaporean Ministry of Education and Chinese Ministry of Education for the funding in the early stage of my studies.

Abstract

Microneme proteins (MICs) belong to a protein family that is actively involved in the invasion of host cells by apicomplexan parasites. Among these proteins, MIC2 is a member of the Thrombospondin-related anonymous (TRAP) family of adhesive proteins, which are highly conserved within the apicomplexans. The other microneme protein M2AP, known to facilitate the transport of MIC2 through the secretory network, is also essential to the virulence of the parasite. Most importantly, MIC2 and M2AP form a multimeric functional complex that presents in the cell invasive stages and serves as the prototype of other TRAP family members.

This thesis describes the NMR assignments and two solution structure calculations of M2AP and a Thrombospondin type-1 repeats (TSR-1) domain. M2AP is composed exclusively of beta strands and, despite no sequence similarity, demonstrates clear structural similarity to the galectin-like domain from MIC1. And TSR-1 repeats from MIC2 demonstrate two typical TSR-like-fold domains. Interactional studies reveal that the M2AP binds both the first and the last pair of TSR-1 repeats of MIC2. The results on structure–function relationship of MIC2 and M2AP reinforce the critical role of TRAP protein in the successful invasion of cells by *T. gondii* and reveal potential therapeutic targets that may be used to control toxoplasmosis.

Contents

Abbreviations

APBS	Adaptive Poisson-Boltzmann Solver
ARIA	Ambiguous restraints for iterative assignment
ATP	Adenosine triphoshpate
BLAST	Basic Local Alignment Search Tool
BMRB	Biological Magnetic Resonance Data Bank
C terminus	Carboxyl terminus
CNS	Crystallography & NMR system
CSI	Chemical shift index
CRD	Carbohydrate recognition domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ER	Endoplasmic reticulum
FPLC	Fast protein liquid chromatography
Gal-3	Galectin 3
GPI	Glycoposhpatidylinsositol
HSQC	Heteronuclear single quantum correlation
Hz	Hertz
IPTG	Isopropyl-β-D-thiogalactopyranoside

LB	Luria Bertani broth
M2AP	Microneme protein 2 associated protein
MAR	Microneme adhesive repeat
MLEV	Malcolm Levitt's CPD sequence
MIC2	Microneme protein 2
MPP	Microneme protein protease
MPP1	Microneme protein protease 1
MPP2	Microneme protein protease 2
N terminus	Amine terminus
NOE	Nuclear Overhauser effect
NOESY	Nuclear overhauser effect spectroscopy
NMR	Nuclear magnetic resonance
ORF	Open reading frame
PALES	Prediction of AlignmEnt from Structure
PCR	Polymerase chain reaction
PDB	Protein data bank
ppm	Parts per million
RDC	Residual dipolar coupling
RNA	Ribonucleic acid
RMSD	Root mean square deviation
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TALOS	Torsion angle likelikhood obtained from shift and sequence similarity

TBE	Tris/Borate/EDTA buffer
TE	Tris/EDTA Buffer
TMS	Tetramethylsilane
TMD	Transmembrane domain
TOCSY	TOtal Correlation SpectroscopY
TROSY	Transverse Relaxation Optimized Spectroscopy
TRAP	Thrombospondin-related adhesive protein
TSP	3 - trimetylsilyl propionate
TSP-1	Thrombospondin -1 protein
TSR	Thrombospondin-like repeat
TSR1	Thrombospondin-like type 1 repeat
UV	Ultra violet

Chapter 1 Biological Introduction

1.1 The parasite - Toxoplasma gondii

T. gondii is one of the most common obligate intracellular parasites that belong to the phylum Apicomplexa and it is the only known member of species genus *Toxoplasma* (Sabin and Olitsky 1937). Its name derives from its morphology, as the 'toxon' meaning arch or bow is after its crescent shape of the active tachyzoite form of the organism. Its close relatives, i.e., *Plasmodium, Eimeria,* and *Cryptosporidium,* are responsible for the more famous diseases including Malaria in humans and Coccidiosis in poultry.

At certain stages of its life, the parasite is able to cause encephalitis (inflammation of the brain) and neurologic diseases that affect the heart, liver, and eyes, causing fatality. The life cycle of this unicellular protozoan can be divided into two phases – the sexual cycle and asexual cycle (See figure 1.1). The sexual cycle only happens in members of Felidae family (wild and domestic cats) (Comstock and Ganley 1973). Thus the cats serve as the primary hosts for this parasite. The parasite is able to infect most warm-blooded animals, which implies the asexual cycle can take place in any infected animals including the human (Carruthers 2002).

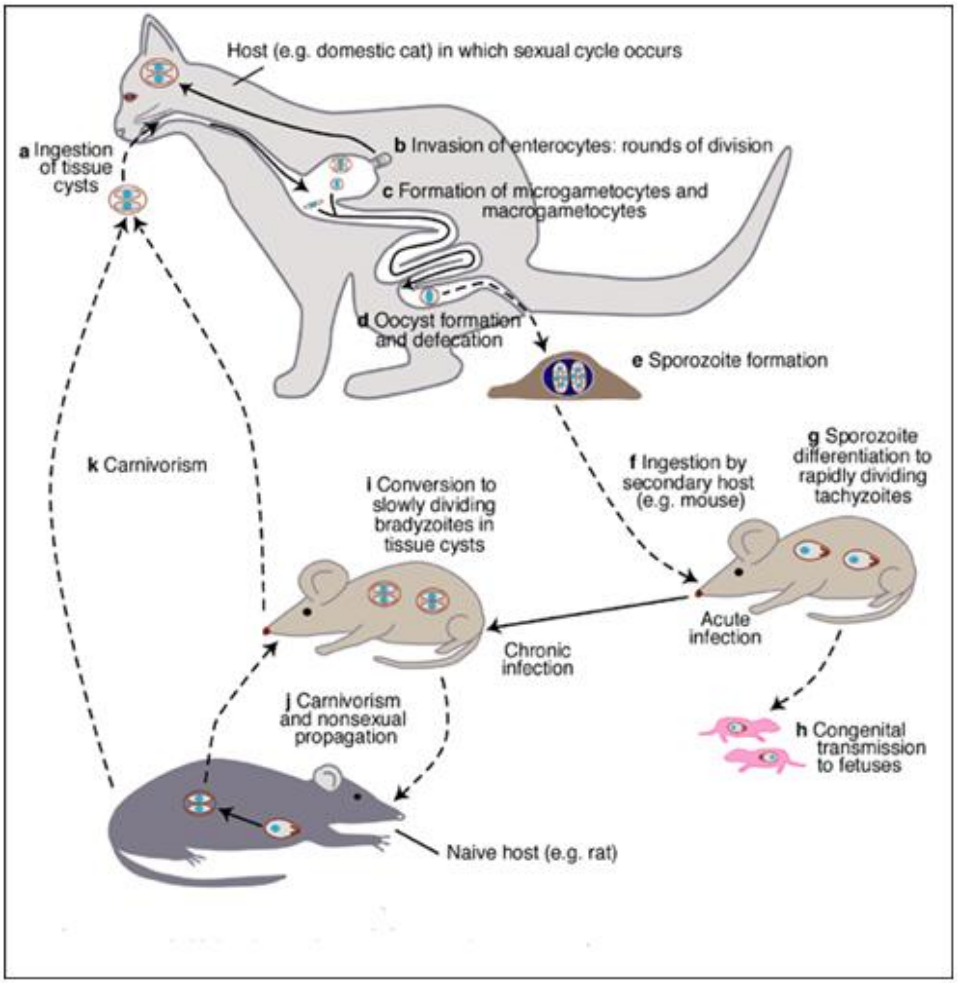


Figure 1.1 The *T. gondii* life cycle**:** The sexual cycle takes place exclusively in the intestinal enterocytes of many members of the cat family (*Felidae*); the asexual cycle takes places in most warm-blooded animals including humans. Figure adapted from (Coppens and Joiner 2001).

The sexual cycle takes place exclusively in the intestinal enterocytes of many members of the cat family (Torda 2001). After ingestion of any type of tissue cysts by the cat, the parasite invades the enterocytes then undergoes rounds of division. The parasites will then begin to differentiate into two different kinds of gametocytes: microgametocytes (male) and macrogametocytes (female). The gametocytes fuse to form a zygote with a thick-walled structure or 'oocyst' that is shed into the

environment with the cat's faeces. The oocyst is capable of meiosis, producing highly infectious 'sporozoites' that are resistant to environmental damage and may persist for years in a moist environment. The sporozoites differentiate into the rapidly dividing 'tachyzoite' form after ingestion by any secondary hosts such as mice or humans. This would cause and sustain an acute or chronic infection. During the acute infection, congenital transmission to the developing fetus can occur. In many hosts, a chronic phase of the disease ensues, as the tachyzoite changes into a slowly dividing form known as the 'bradyzoite'. Latent bradyzoite tissue cysts persist for the life of the host, re-emerging occasionally, but do not produce overt disease in healthy individuals. Carnivorous ingestion of tissue cysts can lead to the infection of a native host, allowing for an indefinite nonsexual propagation of *T. gondii*.

T. gondii consists of three clonal lineages designated type I, II, and III, which differ in virulence and epidemiological pattern of occurrence (Howe et al. 1997). Most strains isolated from patients with AIDS are type II. Type I and II strains have been recorded in patients with congenital disease, whereas strains isolated from animals are mostly genotype III. Strain-specific antibodies could allow typing of *T. gondii* strains with serum from a patient.

1.2 Toxoplasmosis

As described previously, the parasite causes *Toxoplasmosis* by infecting most warm-blooded animals but the primary host is the felid family. Animals are infected by eating infected meat, by contact with cat feces, or by transmission from mother to fetus, where cats have been shown as a major reservoir of this infection.

1.2.1 Toxoplasmosis in humans

Toxoplasmosis in human beings is often minor and self-limiting in many cases but can have serious effects on a fetus whose mother first contracts the disease during pregnancy or on an immunocompromised human (e.g. AIDS patient) (Holliman 1988). An estimated 30% of the total world population carries the infection. The infection rates vary from country to country. While France has prevalence of up to 80% of its population, South Korea has only 4.3% (Novotna et al. 2005). Many factors, from climate in the region to eating habits of the culture, determine the rates.

Healthy individuals could be infected by ingestion of contaminated food or water with oocysts shed by cats, or undercooked meat with tissue cysts (See figure 1.2). Once the parasite has been ingested, it actively invades intestinal epithelial cells (using gliding motility, see section 1.4) or it gets phagocytosed by them. The molecular characterization, importance and function of several proteins, especially microneme proteins, from different parasite organelles have been reported during this event (Achbarou et al. 1991). These molecules, together with immunodominant tachyzoite surface antigen SAG1, are the most promising vaccine candidates (Dlugonska 2008).

The infection soon spreads to all organs, including brain, eye, heart, muscle, placenta and fetus. During the early stage of the infection, patients only suffer mild flu-like symptoms or no illness. After the first few weeks the infection rarely causes any symptoms in otherwise healthy adults. It may become more serious, potentially causing fatality in people with a weakened immune system, such as those infected with HIV or who are pregnant. Brain inflammation, encephalitis and other

neurologic diseases are among the most common causes of death in these cases (Montoya and Liesenfeld 2004).

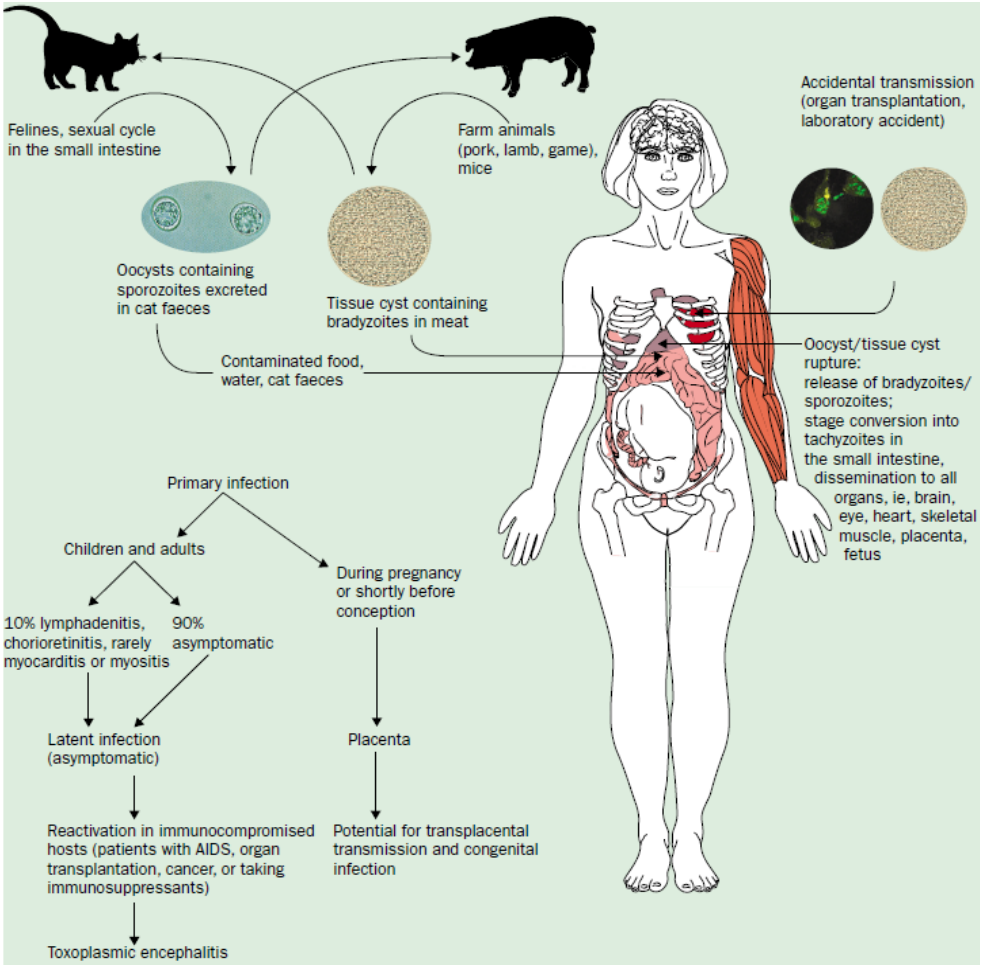


Figure 1.2 Toxoplasmosis in human: Life cycle and clinical manifestations. Picture adapted from (Montoya and Liesenfeld 2004).

Congenital toxoplasmosis is a special but common form in which a fetus is infected via the placenta (Heath and Zuelzer 1944). The consequences can be unpredictable and life threatening. It causes severe syndromes, including mental retardation, blindness, and epilepsy in infancy or much later in life. Women infected with *T. gondii* before conception rarely transmit the parasite to their fetus, but those who

become acutely infected or have reactivation of T. *gondii* during pregnancy (because of immunosuppression) can transmit the organism transplacentally.

1.2.2 Lafferty's hypothesis

Apart from the physical damage of infection, Lafferty found a significant correlation between high levels of the parasite and high levels of neuroticism, suggested by the fact that *T. gondii* is largely a brain parasite (Lafferty 2006). There was a positive but weak correlation between Toxoplasma and levels of uncertainty avoidances and masculine sex roles (Novotna et al. 2005). However, excluding the non-western countries, like China, South Korea, Japan, Turkey, and Indonesia, the correlations of both personality measurements with Toxoplasma got much stronger. Lafferty postulated why the French are more neurotic (*T. gondii* infection rate 88%) and Australians are not (28%). He and many other researchers concluded that *T. gondii*, whilst being just one of many factors, does influence the human culture.

1.3 Current Diagnosis, Treatment and Vaccination

Detection of *Toxoplasma gondii* in human blood samples may be achieved by using the polymerase chain reaction (PCR). Inactive cysts may exist in a host which would evade detection.

Since the invasion mechanism of *T. gondii* is highly similar to that of *Plasmodium* species, anti-malaria drugs (like Quinin and Pyrimethamine) combined with antibiotics are among the most commonly used. For in latent Toxoplasmosis, the antibiotics atovaquone and clindamycin are used in HIV and organ transplanted patients (Almond et al. 2000). Pyrimethamine-sulfadiazine combinations are typically avoided during pregnancy, because of the risk of harm to both mother and baby.

Congenital toxoplasmosis acquired during pregnancy can be treated with the antimicrobial spiramycin, although it is still considered an experimental drug in some parts of the world. The long term effects upon the unborn child cannot be reversed by drug treatment, but subsequent use of pyrimethamine may help to reduce the effects of chorioretinitis in later life.

There is currently no vaccine available for humans. In sheep, toxoplasmosis can be controlled by vaccination with a live, attenuated vaccine (Buxton and Innes 1995). However, since such a vaccine has practical disadvantages and is not acceptable for use in humans, various strategies to develop an effective subunit vaccine have been explored. The most promising vaccination targets include Microneme protein 2 (MIC2) and immunodominant tachyzoite surface antigen (SAG1) (Huynh and Carruthers 2006).

1.4 Host-cell invasion, gliding motility and microneme proteins

T. gondii actively invades a wide range of cells within its mammalian host by a mechanism called gliding motility (See figure 1.3), which is distinct from phagocytosis (Dobrowolski and Sibley 1997). Gliding locomotion begins with the apical release of adhesins and then their redistribution toward the posterior pole of the parasite, which involves the parasite actomyosin system. This motility is powered by an actin-myosin motor, in which the actin is anchored to the host cell by various microneme protein complexes via aldolase, and myosin is anchored to the parasite's inner membrane complex. Movement is achieved by myosin 'walking' along actin filaments. The anchoring of myosin to the inner membrane complex results in the translocation of actin (from anterior to posterior) and the coupled cell-surface proteins. Adhesin-receptor complexes are therefore pushed 'backwards' propelling

the parasite forward (*i.e.* relative to the host-cell surface) and are subsequently proteolytically removed (Baum 2006). This active form of invasion confers *T. gondii* the ability to cross the non-permissive biological barriers including the blood-brain barrier, the intestine and the placenta, where it exhibits its most severe virulence.

Host-cell invasion is a highly regulated process and represents a critical stage in the life-cycle of *T. gondii,* as failure to enter the cell renders the parasite to permanent attack from the host's immune response. Consequently, host-cell invasion occurs quickly and utilises this unique form of locomotion.

Interestingly, gliding motility is a phylum specific mechanism unique to Apicomplexa. It confers the parasite the ability of moving across the solid substrate and penetrate biological barrier. And it is also distinct from gliding mechanisms found in some bacteria which relies upon cell surface appendages, such type IV pili (Sibley 2004). This is particularly evident during host-cell invasion and contrasts the relatively passive mechanisms employed by bacteria (*e.g.* induced phagocytosis) and viruses (*e.g.* clathrin-coated pits) to facilitate cell entry (*i.e.* which use existing endocytic pathways) (Carruthers 2002).

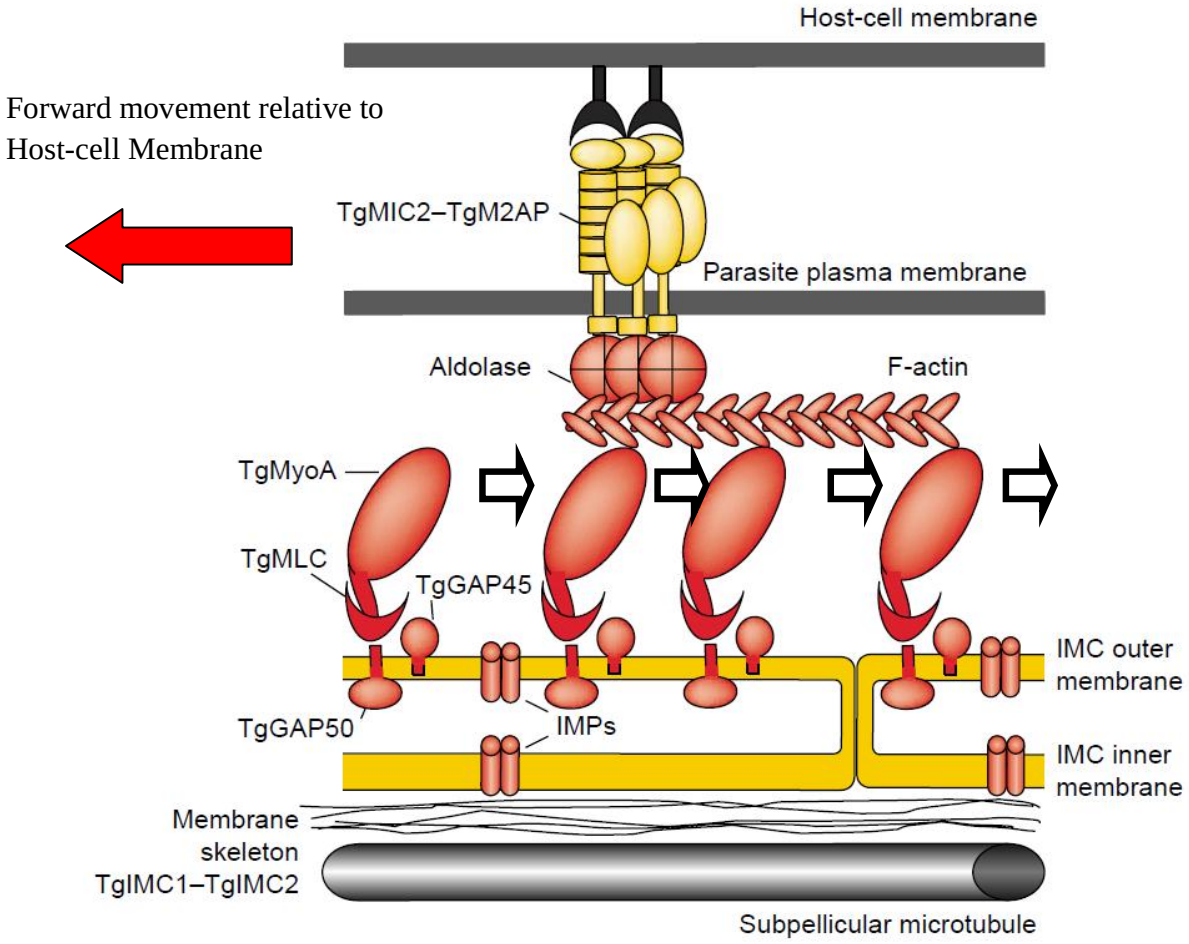


Figure1.3 The molecular machine (MIC2/M2AP complex) that promotes gliding motility in *Toxoplasma gondii.* It comprises the motor complex [myosin A (TgMyoA), the myosin light chain (TgMLC), TgGAP45 and TgGAP50], which is anchored in the outer membrane of the IMC and connected via filamentous actin (F-actin) and aldolase to the hexameric microneme protein complex (MIC2/M2AP), which, in turn, interacts with host-cell receptors. Propulsion is achieved by myosin 'walking' along actin filaments. Figure adapter from (Keeley and Soldati 2004)

1.4.1 Microneme proteins

Three sets of different secretory organelles: the micronemes, rhoptries and dense granules release their contents during the parasite invasion (Striepen et al. 2001).

Microneme proteins released from the micronemes play a dominant role in the early phase of the invasion process by discharging adhesins capable of interacting with host cell receptors. These proteins serve as adhesins in the gliding motility machinery.

Eleven proteins have been confirmed in the micronemes and at least nine additional putative microneme proteins have been identified based on their secretion properties or homology to other known microneme proteins. Among these proteins, half of them were predicted to be type I membrane proteins with a single membrane spanning sequence located within the C terminus. Surprisingly, all the previously identified parasite surface proteins (Surface Antigens - SAG) are tethered to the parasite surface by glycosylphosphatidylinositol (GPI) anchors without any transmembrane domains. This implies that the difference between the anchoring of SAG and MIC proteins has evolved for differential protein targeting and surface regulation (Carruthers 2002).

Among these microneme proteins, a conserved protein is required for both migration and host cell invasion, known as MIC2 and often regarded as the only TRAP (Thrombospondin-related adhesive protein) protein in *T. gondii* (see section 1.7). MIC2 and its partner protein M2AP were found to interact with a conserved host cell receptor and playing central roles in parasite survival and invasion. While deletion of the MIC2 gene led the death of the parasite, even reduction of expression level of MIC2 dramatically decreased the virulence of the parasite (see section 1.7) (Brossier and David Sibley 2005).

Apart from the transmembrane TRAP homologous protein MIC2 and its partner M2AP, other microneme proteins also play an important role during the host cell invasion of Toxoplasmosis (Cerede et al. 2005). Many of them have been assigned roles in invasion and virulence. While individual gene knockout of MIC1 or MIC3

only slightly decreased invasion in mice, the double gene disruption led to severely impaired virulence and conferred the immunity against subsequent challenge (Cerede et al. 2005). These findings support the idea that the parasite has evolved multiple ligand-receptor interactions and these protein proteins play synergistical roles to ensure the invasion in different type of cells.

1.5 Architectural insight into microneme proteins

Microneme proteins show surprising domain conservation across the apicomplexans genera, some of which even show functional complementary across species. For example, TSR type 1 repeats were found in *Cryptosporidium, Eimeria, Neospora, Plasmodium* and *Toxoplasma* species (figure 1.4). This implies not only the invasion machinery is well conserved within phylum apicomplexans, but also the structures at molecular level (Rabenau et al. 2001). For some of these proteins, the domain localization has not been confirmed at the ultrastructural level.

T. gondii microneme proteins contain structural domains that are morphologically and functionally conserved at the molecular level. They include: (i) chitin-binding-like domains; (ii) epidermal growth factor (EGF) like domains; (iii) apple domains; (iv) type I repeat of Thrombospondin (TSR-1); (v) the integrin-like I-domains; and (vi) β-sheet domain (See figure 1.4). These motifs are present in one or multiple copies, so that every microneme protein is structurally unique by different combination of these domains. Some of these domains present homology with known adhesive domains from higher eukaryote proteins, *e.g.* TSR-1 resembles human TSP (thrombospondin) domain.

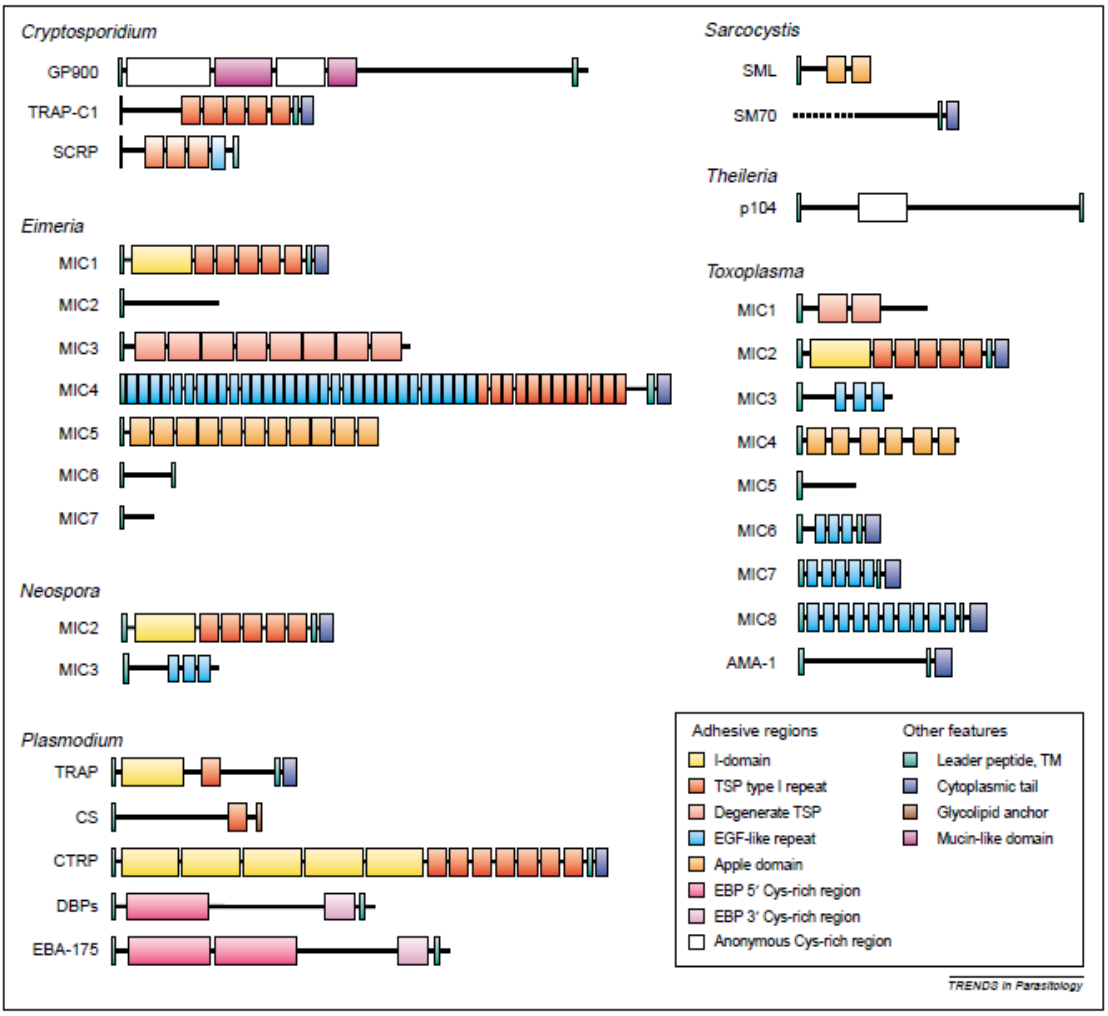


Figure 1.4 Schematic illustration of the structure similarity of microneme proteins in apicomplexans parasites. The domains are coloured differently. Figure adapted from (Rabenau et al. 2001).

1.6 Three microneme protein complexes

Microneme proteins were found to be of great importance during invasion. However, their precise role and the level of functional redundancy remain to be evaluated. Remarkably, most microneme proteins are proteolytically cleaved during biogenesis and post-exocytosis *(e.g.* MIC2 Integrin-like domain) ((Dowse and Soldati 2004). 2004). This feature provides the molecular basis to dissect these proteins or protein complexes.

It is becoming more apparent that many microneme proteins are found in stable adhesive complexes. Three complexes have been identified to be the key adhesive complexes that bind to host cells cooperatively (Cerede et al. 2005).

First, MIC6 partnered with MIC1 and MIC4 form one of the three complexes that work synergistically in host cell binding. MIC6 acts as an escorter protein, whereas MIC1 and MIC4 fulfill the role of adhesins (See figure 1.5). The MIC6 protein contains a transmembrane sequence, which functions to anchor the adhesive complex to the membrane. It is also required for correct targeting of the adhesive complex to the micronemes since genetic disruption of MIC6 caused mis-targeting of MIC1 and MIC6 to the default signal pathway. The NMR structures and crystal structure of domains from this complex are available, *e.g.* MAR, MIC1CT of the MIC1 and EGF domain of MIC6 (Blumenschein et al. 2007; Saouros et al. 2005; Sawmynaden et al. 2008).

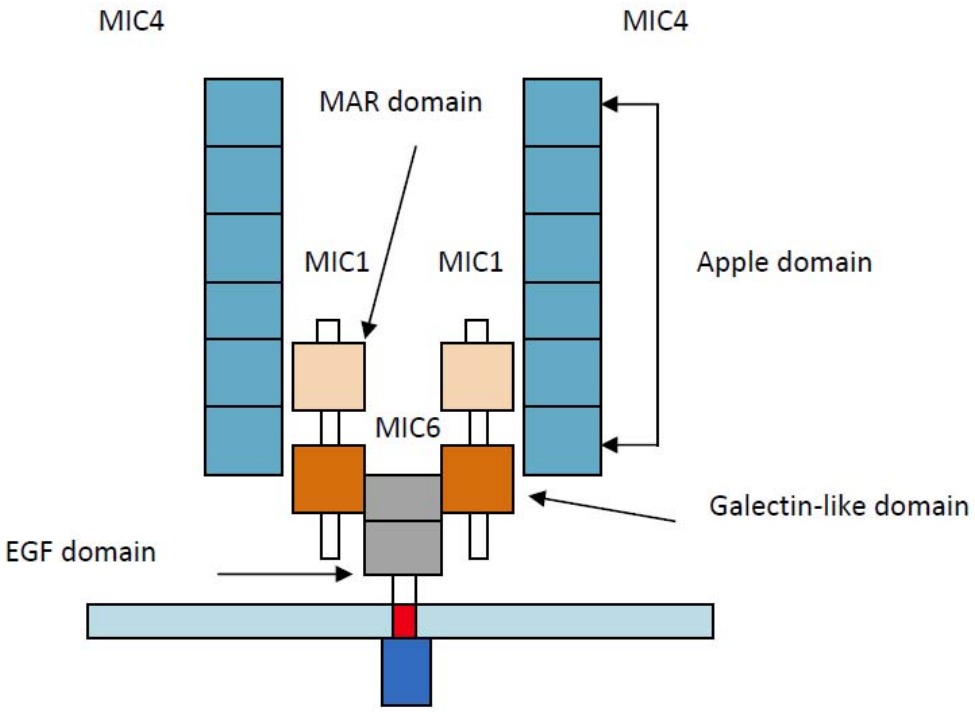


Figure 1.5 A schematic representation of a model for the MIC4–1–6 complex. Microneme adhesive repeat (MAR) domain and galectin-like domain of MIC1, epidermal growth factor-like (EGF) domains of MIC6 and six apple domains of MIC4 are coloured differently: blue for apple domain, pink for MAR domain, orange for Galectin-like domain, grey for EGF domain, red for transmembrane domain and dark blue for cytoplasmic tail.

MIC3 and MIC8 also form an adhesive complex that is released onto the parasite surface during attachment and invasion. MIC3 has an N terminal chitin-binding-like (CBL) domain, which acts a receptor binding site of mammalian cells. It forms a dimer (See figure 1.6) in order to acquire these binding properties and is connected to parasite by an interaction with a MIC8 monomer. The binding site, CBL domain, still remains poorly characterized (Garcia-Reguet et al. 2000).

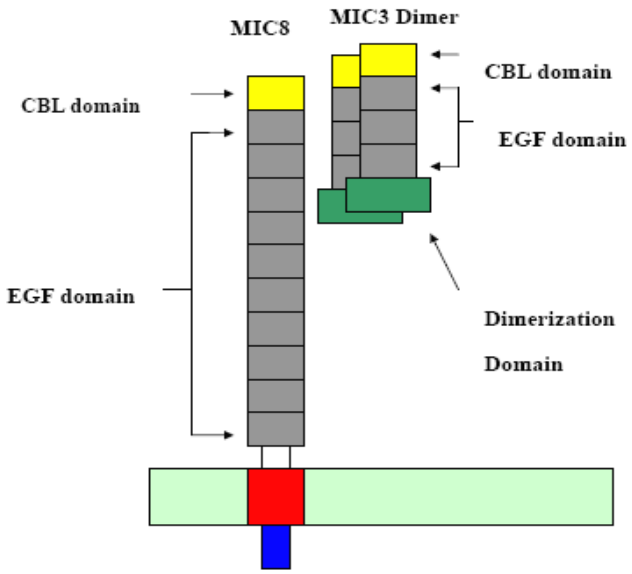


Figure 1.6 MIC8 partners with MIC3 dimer. MIC3 possesses a prodomain that is removed, presumably in the Golgi. MIC8 anchors the adhesive complex to the membrane. CBL domain is coloured yellow, EGF domain is pale, dimerization domain is green, transmembrane domain is red and cytoplasmic tail is blue. The specific contacts between them are still not known. No structural information is yet available.

The third and arguably most important protein complex consists of MIC2 and MIC2 associated protein - M2AP (see figure 1.7). MIC2 and M2AP form a hexameric hetero-complex (Jewett and Sibley 2004b), in which MIC2 provides the membrane anchor for the complex, and binds to the accessory protein M2AP. And the complex formation is essential for intracellular parasite survival. MIC2 is the only TRAP family protein that is expressed at all the invasive stages of the parasite. M2AP is known to facilitate the transport of MIC2 through the secretory network. Mutant parasites that do not express M2AP have a defect in parasite invasion presumably the consequence of insufficient MIC2 transported to the cell surface but accumulated in the parasite endoplasmic reticulum and Golgi (Huynh and Carruthers 2006). The MIC2-M2AP complex is highly conserved within apicomplexans and serves as a model to

understand the first steps of invasion by this phylum. Its presence at all stages of the life cycle suggests that it has a critical role in the successful invasion of cells and reveals potential therapeutic target that may be used to control toxoplasmosis.

This complex is the primary target of the project and will be discussed in detail in the later section.

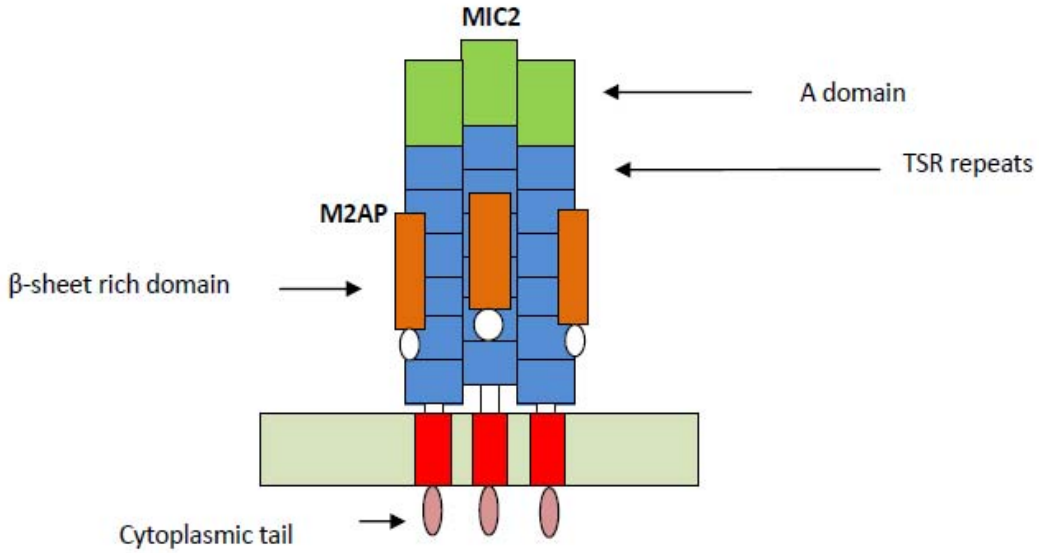


Figure 1.7 Schematic representation of the proposed architecture and interactions of MIC2/M2AP relationship: the extracellular domains of MIC2 participate in the formation of MIC2-M2AP complexes. MIC2 consists of A domain (green), TSR-1 (blue) and cytoplasmic tail (grey). TSR-1 forms an extended stalk on the integrin A domain, which is optimally positioned for interaction with host receptors. No structure related to the complex is yet published.

1.7 MIC2-M2AP complex

As the most important adhesive complex for the parasite, MIC2 and M2AP are reported to form a hexameric complex containing three molecules of each protein. MIC2 has demonstrated fundamental roles in cell gliding motility and invasion, and classified as the only "TRAP" protein in *T. gondii.* Originally discovered in *Plasmodium* (Rogers et al. 1992), TRAP proteins are characterized by having one or more copies of an Integrin-like domain, one or more copies of a type 1 thrombospondin-like repeat (TSR-1), transmembrane domain and cytoplasmic tail. It is now regarded as a popular malaria vaccine. M2AP function is also crucial to these processes, as disruption of M2AP results in secretory impairment of MIC2, leading to impaired invasion. Overall, MIC2 and M2AP complex is one of the most popular TRAP protein models under study.

Although MIC2 is essential for parasite survival and gene knockout of MIC2 is lethal for *T. gondii*, conditional gene knockout confirmed that MIC2 is a major determinant of invasion and virulence of the parasite. Reduced MIC2 expression resulted in mistrafficking of M2AP, defective host-cell recognition and invasion, the loss of helical gliding motility and an inability of *T. gondii* to induce a lethal infection. An attenuated inflammatory immune response was produced by infecting mice with MIC2-deficient parasites, indicating a promising vaccination leading to long-term protective immunity (Huynh and Carruthers 2006).

1.7.1 The Integrin-like A domain of MIC2

The integrin-like A domain is usually found in the α chain of integrins, such as the cartilage matrix protein, factor B and collagen VI (Takagi and Springer 2002). The

domain is located in the N terminal region of MIC2 possesses a cation binding site, termed the metal ion-dependent adhesion site (MIDAS), which contains five conserved residues, Asp-X-Ser-X-Ser, for cation binding. Other ligands that interact with Integrin A domain can be independent of the MIDAS motif, such as in TRAP binding to Glycosaminoglycans. In addition, some studies demonstrated that the binding of MIC2 is not dependent on the MIDAS site. For example, the recombinant A domain of MIC2 interacts with heparin in a MIDAS site-independent manner (Brossier and David Sibley 2005).

MIC2 is also processed by cleavage of a pro-peptide upstream of the A domain, which removes the first 70 amino acids of the protein. This process occurs while MIC2 is at the surface of the parasite. The protease responsible for this cleavage is termed MPP2 (microneme protein protease2). However, the processing of MIC2 by MPP2 is not required for the release of the protein into the supernatant. Recent data suggest this interaction facilitates migration across polarized epithelial cells.

1.7.2 The thrombospondin-like repeats (TSR) of MIC2

TSR repeats are involved in the interaction with many ligands, including GAGs, CD36 or HIV gp120. The crystal structure of TSRs from human TSP-1 consists of three anti-parallel strands that contain conserved residues involved in their proper folding and function (Tan et al. 2002) (see figure 1.8). It is postulated that TSR-like repeats may participate in the interactions with M2AP and host cell receptors although there is currently no experimental evidence to demonstrate this directly.

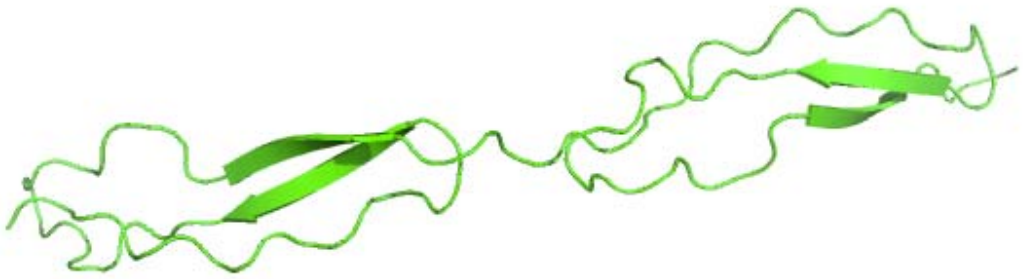


Figure 1.8 Crystal structure of human TSP-1 domain: two subdomains consist of three anti-parallel strands. PDB ID: 1LSL.

1.7.3 The transmembrane domain (TMD) of MIC2

Like most of the microneme proteins from apicomplexans, TMD of MIC2 contains a large percentage of small hydrophobic residues. Studies show that the extracellular part of MIC2 needs to be released for effective invasion to occur and the cleavage happens within the TMD of the molecule (Opitz et al. 2002). Mutation of residue Lysine 692, eleven amino acids upstream of the TMD, prevents cleavage by microneme protein protease 1 (MPP-1). MPP-1 is a protease that is named after its activity. As a result, the parasite retains the mutated MIC2 at its surface and is severely impaired in its capicity to invade the cell (Opitz et al. 2002). These studies suggested that the Lysine 692 involved in the recognition by the protease and the cleavage happens inside the TMD.

1.7.4 The cytoplasmic tail of MIC2

The cytoplasmic tail of MIC2 interacts with the F-actin-binding protein aldolase from the cell cytoskeleton, thus permitting the parasite to move and invade (see figure 1.7). The interaction requires Trp767 at the carboxy terminal extremity of the protein. The cytoplasmic tail of *Plasmodium falciparum* TRAP also contains a Tryptophan residue that is required for the interaction with *Plasmodium* aldolase. This implies that this Tryptophan in the carboxy terminal of the protein is probably conserved within the phylum apicomplexa (Buscaglia et al. 2003).

The C terminus of MIC2 is essential for the parasite survival. Mutation of the TRAP C terminus from *Plasmodium falciparum* prevents the mobility of the parasite. Hence, the C domain plays a critical role in linking to the cytoskeleton and a common mechanism of interaction of adhesins with the cytoskeleton that coordinates invasion is potentially shared in apicomplexans.

1.7.5 M2AP

M2AP is a member of a specific protein family expressed by coccidian parasites including *Toxoplasma gondii*, *Neospora caninum* and *Eimeria tenella*. It is a soluble protein that is secreted by the parasite into its environment during the invasion. M2AP complexes with MIC2 within 15 minutes of synthesis and remains associated with MIC2 in the micronemes, on the parasite surface during invasion and in the culture medium after release from the parasite plasma membrane. Like MIC2, it also possesses a propeptide, but this is removed in the Golgi (Zhou et al. 2004).

Its function and structure remain largely unknown. The phylogenic conservation and association with a key adhesive protein (MIC2) suggest that M2AP is a fundamental

component of the *T. gondii* invasion machinery. Secondary structure alignment by BLAST search shows no homologues outside class Coccidiae, suggesting its novel structure and uniqueness among *T. gondii* microneme proteins. Absence of M2AP homologues in Plasmodium species probably implies the undetermined function of this protein (Jewett and Sibley 2004a).

1.8 Aims of this project and scope of the Thesis

Microneme proteins are both therapeutically (e.g. drug development) and phylogenically (model parasite) interesting. The aim of this project focused on the structure determination of microneme proteins and *in vitro* interactions thereafter. Given that the most parts of the MIC1-4-6 complex structure have been recently solved, studying the other two complexes is important to extend the current understanding of the *T. gondii* invasion machinery. In particular, the MIC2-M2AP complex is central to *T. gondii* infection, and a promising target for vaccines (and potential drug development).

The project initially started with the MIC3-8 complex and later moved to the MIC2-M2AP complex after the cDNA became available. This thesis will focus mainly on the expression, purification and structure determination of putative domains of the MIC2-M2AP complex since no working construct was obtained for the MIC3-8 complex. Calculation of the solution structures of two regions from the complex by NMR is the main focus of the thesis and perhaps more importantly their interactions. The results are discussed with regards to the current state of structural understanding of microneme complexes from *T. gondii*.

Using NMR spectroscopy as a primary tool, this project thus aims to elucidate the structural and functional properties of MIC2/M2AP complex with the following objectives:

1. To provide a structural insight into the extracellular region of M2AP by investigating the predicted β sheet rich domain.
2. To provide a structural insight into the extracellular region of MIC2 by investigating the relevant and presumed TSR domain.
3. To provide a functional insight into microneme protein assembly by examining the nature of the MIC2 – M2AP interaction.
4. To contribute to our overall understanding of how microneme proteins facilitate host-cell invasion and lead to problematic diseases such as toxoplasmosis.

Chapter 2 NMR methods for analysis of protein structure and interaction

NMR spectroscopy, x-ray crystallography and cryo-electron microscopy [1] are the only three techniques capable of solving protein or other macromolecules structures at atomic resolution. NMR has additional advantages in access the solution state and interactions of a molecule, reaction kinetics and molecular dynamics (Montelione et al. 2000). Thus it has become ubiquitous in such divergent fields as chemistry, biology, medicine and forensic science. In protein science, NMR is intensively used to determine the structures of proteins and protein complexes (Cavanagh 2007). A brief interpretation of the basis of NMR is essential as it is being used throughout these structural determination and functional studies (Keeler 2005).

2.1 NMR inactive, NMR active and NMR useful nuclei

The physical basis of NMR spectroscopy can be described by the quantum mechanics of nuclear spin angular momentum. The physical origin of NMR active nuclei is that both protons and neutrons have a property termed spin. Whether a nucleus has net spin arises from whether they pair up or not, which has a direct relation relationship with the atomic mass number (see below).

(i) Nuclei with both even mass number and atomic number, thus a zero spin quantum number (Section 2.2), are NMR inactive.

(ii) Nuclei other than (i) possessing none zero spin quantum number are NMR active.

[1] Cryo-electron microscopy (Cryo-EM) in some favorable cases

a. Nuclei with spin quantum number ½, such as the proton which has two possible spin states, are NMR useful.

b. Nuclei with spin quantum number greater than ½ processing electric quadrupole moments, which have more than two possible spin states, are not commonly used in solution state NMR.

2.2 Nuclear Spin Theory

The spin of a nucleus can be quantized by the nuclear spin number **I** and the magnetic spin quantum number **m**. The magnetic moment **m** of a positively charged nucleus has 2I+1 possible orientations that are degenerate in the absence of an external magnetic field. The proton, with I = ½, has two orientations thus two energy states while in an external magnetic field. (See figure 2.1) In protein NMR, most of the nuclei measured are of I = ½ (i.e., ^{1}H, ^{13}C and ^{15}N) whereas I with a larger value will introduce further multiplicity bringing complexity into the spectrum.

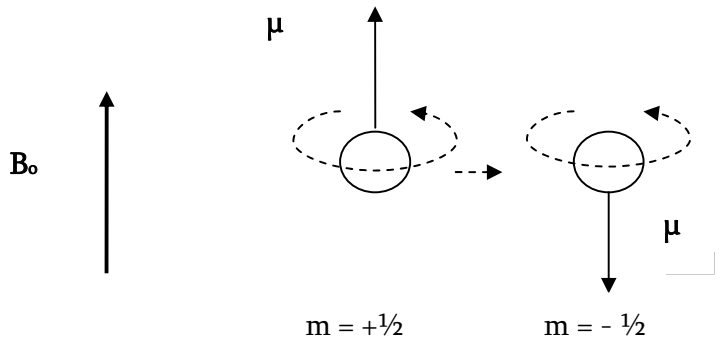


Figure 2.1 Two possible orientations of the magnetic moment (μ) of a proton possessing spin in an external magnetic field (B_0).

2.3 Larmor Precession and Boltzmann Distribution

When a magnetic moment is placed in a magnetic field it will tend to align with the field. Classically, a magnetic moment can be visualized as arising from a current loop and the influence towards alignment can be described as the torque on the moment exerted by the magnetic field. The torque of the magnetic moment causes it to precess about the Bo axis at the frequency termed the Larmor frequency (See figure 2.2). This is governed by the strength of the field and the properties of the nucleus.

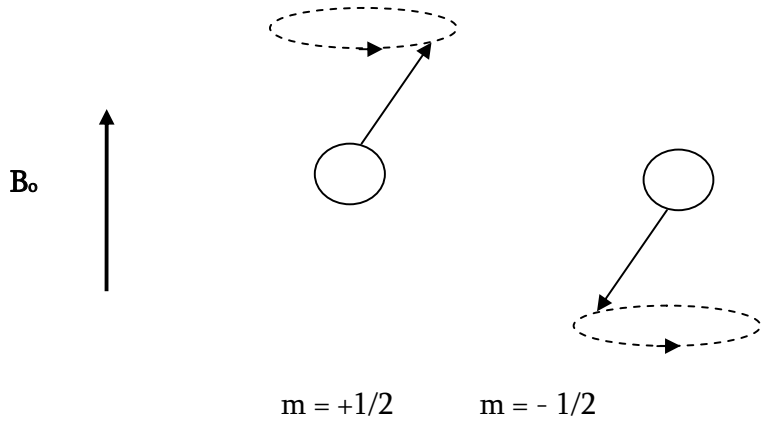


Figure 2.2 Precession of the magnetic moment in each of the two possible spin states of an I=1/2 nucleus in external magnetic field B_o. m = +1/2 aligns with the external field thus it is at the lower energy state. And m = - ½ (anti-parallel with Bo) is at the higher energy state.

Net absorption of radiation requires more particles in the lower energy state than in the higher one to occur. For any system of energy levels at *thermal equilibrium,* there will be more particles in the lower state(s) than in the upper state(s) (see figure 2.3). The distribution of particles is described by Boltzmann distribution (see equation 2.1). This is very important that the lower energy level will contain slightly more nuclei than the higher level under temperature that is relevant for solution NMR. Thus, it is

possible to excite these nuclei into the higher level with electromagnetic radiation. For NMR, this corresponds to the electromagnetic radiation in the radio frequency range (MHz).

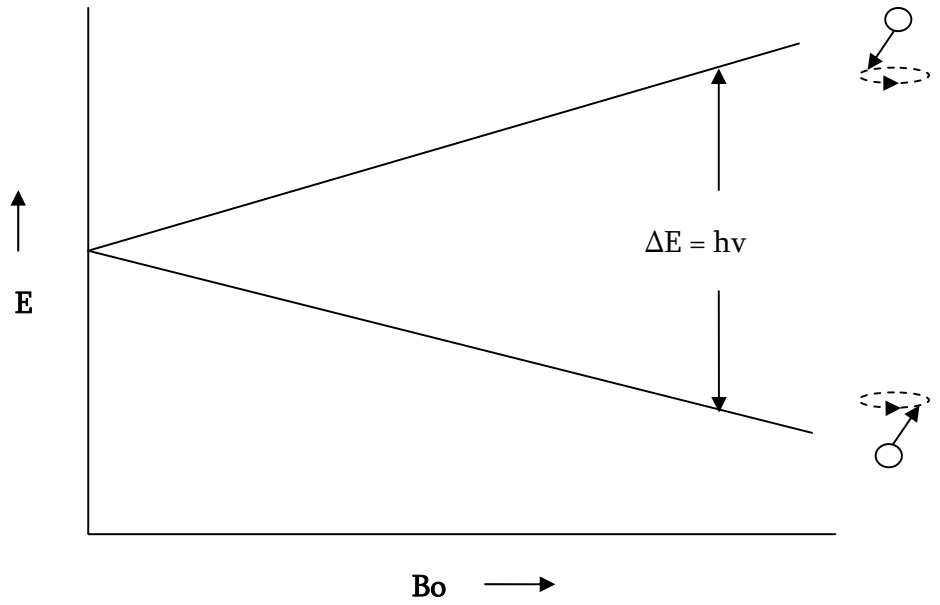


Figure 2.3 Relative energy ΔE of proton spin states in an external magnetic field Bo. Transitions between the two states constitute resonance and can occur if electromagnetic radiation of the appropriate frequency and energy is absorbed.

$$\frac{P_{(m=-1/2)}}{P_{(m=1/2)}} = e^{-\Delta E/kT} = \exp\left(\frac{-\Delta E}{kT}\right) \qquad \text{-- Equation 2.1}$$

Where k is the Boltzmann constant, T is temperature and ΔE is the energy difference possessed by two different states and P is the population.

2.4 Relaxation Processes and Pulses

The re-equilibration of a collection of nuclei after being immersed in an external field is not infinitely fast. The rate of establishing the new equilibrium is governed by spin-lattice (or longitudinal) relaxation time, T_1.

$$\frac{P_{eq} - P_t}{P_{eq} - P_0} = \exp\left(-\frac{t}{T_1}\right) \quad \text{-- Equation 2.2}$$

An oscillating magnetic field B_1 can also interact with an NMR active molecule, which is weaker than Bo and of the same frequency as the precession of nuclear magnetic moments. Establishing the new equilibrium in this magnetic field is normally much shorter than T_1. The re-orientation of individual spins caused by local magnetic fields also influences the transverse components of affected magnetic moments and therefore can drive the bulk transverse magnetisation vector towards zero. This process is called spin-spin (or transverse) relaxation and the rate is controlled by spin-spin relaxation time T_2.

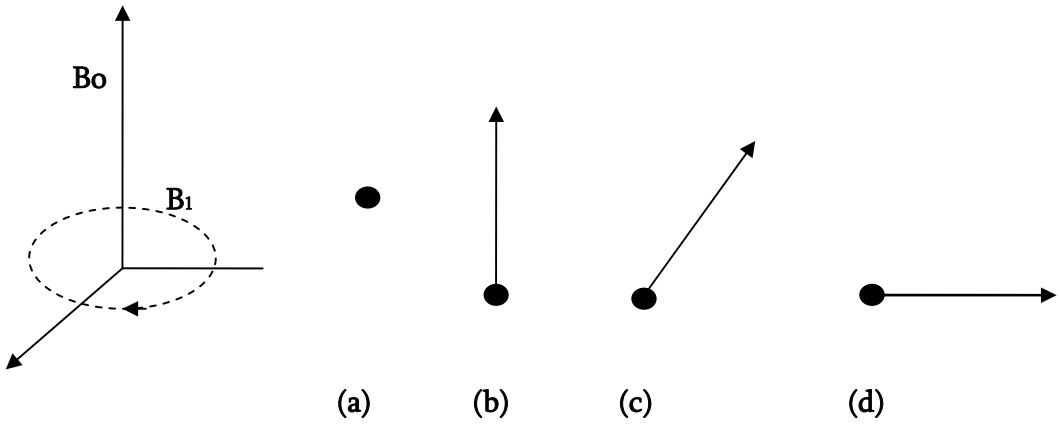


Figure 2.4 Depiction of Net magnetic Moments. (a) Bo and B_1 are off; (b) Bo is on but B_1 is off; (c) Bo and B_1 are both on; and (d) the flip angle of M reaches 90°. T_1 is the relaxation time from state (a) to (b). And T_2 is the relaxation time from (c) or (d) to (a). B_1 can be regarded as a pulse.

A suitable pulse can move a magnetization vector from its equilibrium position (See figure 2.4). The bulk net magnetic vector of the nuclei is along z axis before the pulse is applied. After the pulse is applied, the magnetisation rotates in the x, y plane at the Larmor Frequency of the material under study. Then the induced field in on the z-axis becomes manageable and the transmitted RF signal becomes a fixed electric field vector (E) in the x,y plane relative to the rotating frame. The angle is determined by the power and duration of the irradiation by B_1. When the angle rotated by the magnetic vector is 90°, it is termed a 90° pulse (See figure 2.5).

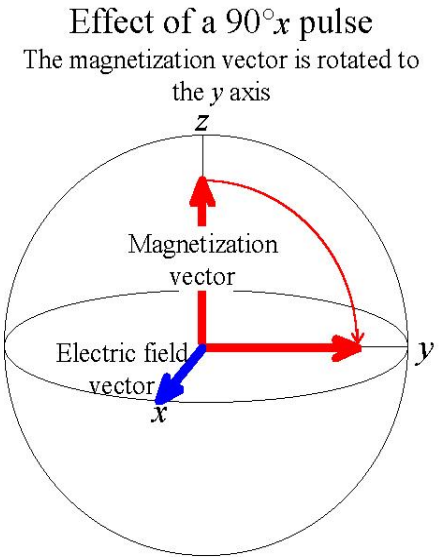


Figure 2.5 Effect of a 90° pulse: the red arrows stand for the magnetization vector and the blue arrow stands for the electric field vector.

2.5 Chemical shift

Each atomic nucleus has a unique pattern of energy levels and resonance frequencies within the same magnetic field. The total magnetic field experienced by a nucleus includes the local magnetic field B_{local} induced by currents of electrons in the molecular orbitals and external field. Even for the same type of nucleus, the electron distribution can vary according to the local environment, i.e. binding partners, bond lengths, torsion angles, etc. The variation in their unique nuclear magnetic resonance frequencies of the same nucleus is conventionally described by the chemical shift (See Equation 2.3). B_{local}, which affects the total magnetic field experienced by nucleus, reflects information on the chemical environment.

$$\text{Chemical shift, } \delta = 10^6 (\nu - \nu_{ref}) / \nu_{ref} \quad \text{-- Equation 2.3}$$

The chemical shift is measured in parts per million (ppm) relative to a reference frequency. The reference compound usually used is tetramethylsilane (TMS), which is defined as 0 ppm. In solution NMR, water soluble derivatives TSP and DSS are commonly used in aqueous buffers. Such δ is measured on a relative scale regardless of the field strength of the magnetic field and this provides a consistent measure across different spectrometers.

2.6 The spin – spin interactions (J coupling)

J coupling (also called indirect scalar coupling) arises from the interaction between two nuclei through intervening chemical bonds (see equation 2.4). It will cause a splitting of signals which may reveal structural information, e.g., this coupling provides detailed insight into the connectivity of atoms in a molecule. If a proton has a single spin coupled neighbor, it will appear as a doublet rather than a singlet in the 1D spectrum (See figure 2.6). If it has two identical partners, it will appear as a triplet. The number of multiplets and the intensity of each peak follow Pascal's Triangle.

The distance of the lines in a multiplet (in Hz) is given by the coupling constant J which is independent of the magnetic field strength. Normally, couplings over one (^{1}J) two (^{2}J, geminal coupling) and three bonds (^{3}J, vicinal coupling) are observed. Vicinal couplings depend on the torsion angle between the coupling protons, which is very important in protein NMR. This is described by the famous Karplus equation, which allows dihedral angle information to be acquired from scalar couplings.

$$J(phi) = A * \cos^2(phi - 60) - B * \cos(phi - 60) + C \qquad \text{-- Equation: 2.4}$$

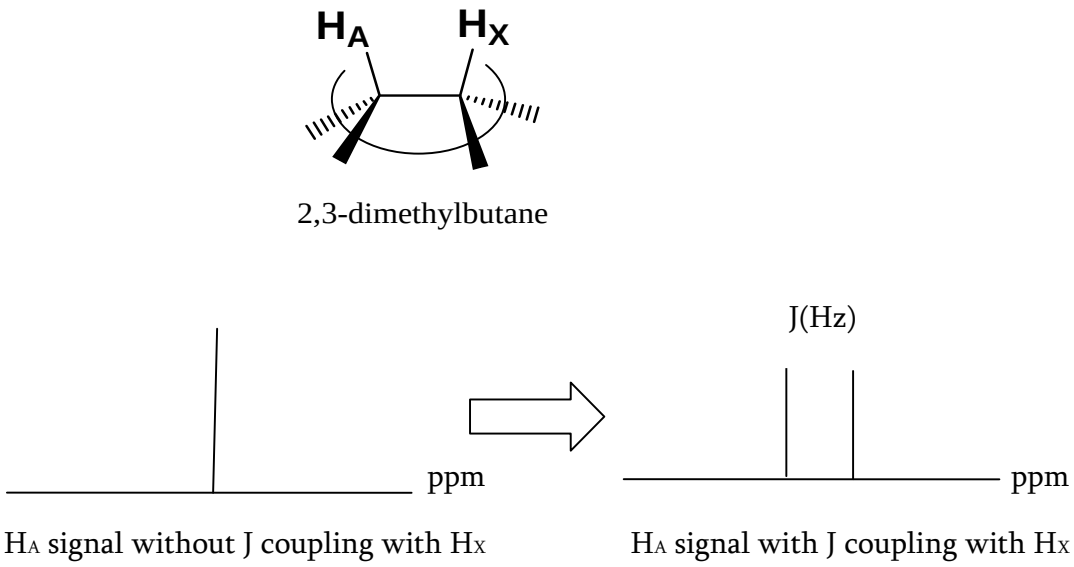


Figure 2.6 1D spectra illustrating a sample case of J coupling between H_A and H_X in 2, 3 – dimethylbutane

2.7 Dipolar Couplings

Dipolar coupling can occur between two spin ½ nuclei that are close in space. It is different from J coupling that occurs via the electrons in the covalent bond. Dipolar coupling occurs as a direct result of the interaction of the magnetic fields of two proximal nuclei through space. Specifically, if nucleus A is positioned within the magnetic field of nucleus B, this can add to or substract from the total magnetic field experienced by nucleus A depending on the spin state of B (see figure 2.7). This is known as dipolar coupling. For molecules tumbling isotropically in solution, the tumbling is such that the dipolar coupling is averaged to zero and the splitting is not directly observed. However, residual dipolar coupling can be measured by putting the molecule into liquid crystalline medium (see section 2.9). Although the time-averaged coupling for nuclei in a tumbling molecule is zero, the fluctuating magnectic fields felt by a nucleus provide a mechanism for interaction, *e.g.* cross-relaxation.

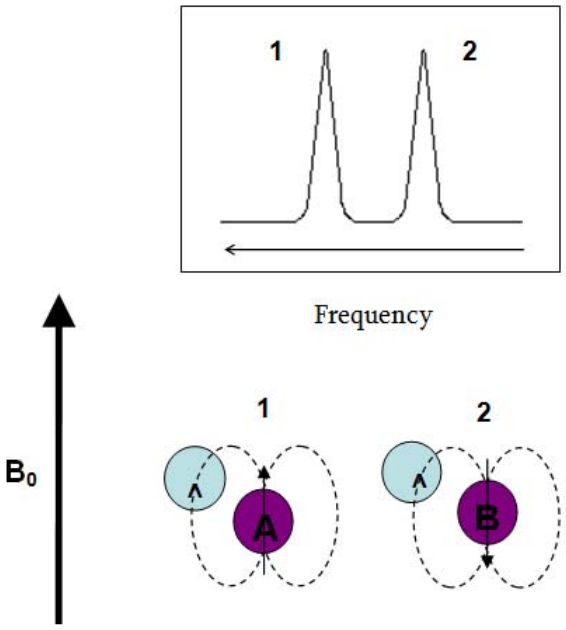


Figure 2.7: Representation of dipolar coupling between two nuclei, A and B. Nucleus B is shown in two different spin states. The magnetic field of B (dotted lines) can add to or subtract from the B_0 field felt by nucleus A. The inset shows the spectral peaks corresponding to nucleus A when nucleus B is in the spin state shown in 1 (aligned with B_0) and 2 (aligned against B_0).

2.8 The Nuclear Overhauser Effect (NOE)

The NOE is the enhancement of intensity of an NMR signal generated by one nucleus when it is near another non-equivalent nucleus whose spin state population is perturbed from equilibrium, *e.g.* by irradiation. When one nucleus is saturated, cross relaxation takes place via dipole-coupled spins, which leads to the modification of signal intensity. NOE is of great importance to NMR spectroscopy, particularly with respect to structure determination. As NOEs arise due to dipolar couplings between nuclei (*i.e.* through-space) and are independent of through-bond scalar couplings. The

presence (or absence) of an NOE can be used to determine the relative proximity of surrounding nuclei and provide key information regarding structural geometry.

2.8.1 The cause of NOE

Considering two spins at thermal equilibrium (*e.g.* I and S), the nuclear Overhauser effect can be defined as the change in the intensity of an "observed" spin I (Interesting) upon saturation of another spin S (Source).

$$f_I\{S\} = (I - I_0) / I_0 \qquad \text{-- Equation 2.5}$$

The orientation of each spin is distributed according to the Boltzmann distribution such that there is a slight excess of nuclei in the lowest energy level (*i.e.* $\alpha\alpha$ orientation). This established population difference can be perturbed from its equilibrium position by application of a RF pulse to the S spin (*i.e.* at the Larmor frequency) causing saturation of the $\alpha \rightarrow \beta$ transitions. It is the cross relaxation (W_{0IS} and W_{2IS}) that generates the NOE. The intensity of signal I is increased by double quantum relaxation W_{2IS} (*i.e.* increased population difference) and decreased by zero quantum relaxation W_{0IS} (*i.e.* recovered population difference). Single quantum relaxation W_{1I} attenuates the NOE (See figure 2.8). The formal equation describe NOE (Solomon equation) is shown as below:

$$f_I\{S\} = \left(\frac{\gamma_S}{\gamma_I}\right)\frac{\left(W_{2IS} - W_{0IS}\right)}{\left(W_{2IS} + 2W_{1IS} + W_{0IS}\right)} = \left(\frac{\gamma_S}{\gamma_I}\right)\frac{\sigma_{IS}}{\rho_{IS}} \qquad \text{-- Equation 2.6}$$

Where $W_{IS} \propto \gamma_I^2\gamma_s^2\frac{3\tau_c}{r^6}$ $W_2 \propto \gamma_I^2\gamma_s^2\frac{12\tau_c}{r^6}$ $W_0 \propto \gamma_I^2\gamma_s^2\frac{2\tau_c}{r^6}$

The term $W_{2IS} - W_{0IS}$ ($= \sigma_{IS}$) describes the contribution of cross relaxation to the relaxation rate of spin I. The denominator $W_{2IS} + 2\,W_{1IS} + W_{0IS}$ ($= \rho_{IS}$) corresponds to the total relaxation rate of spin I due to spin S.

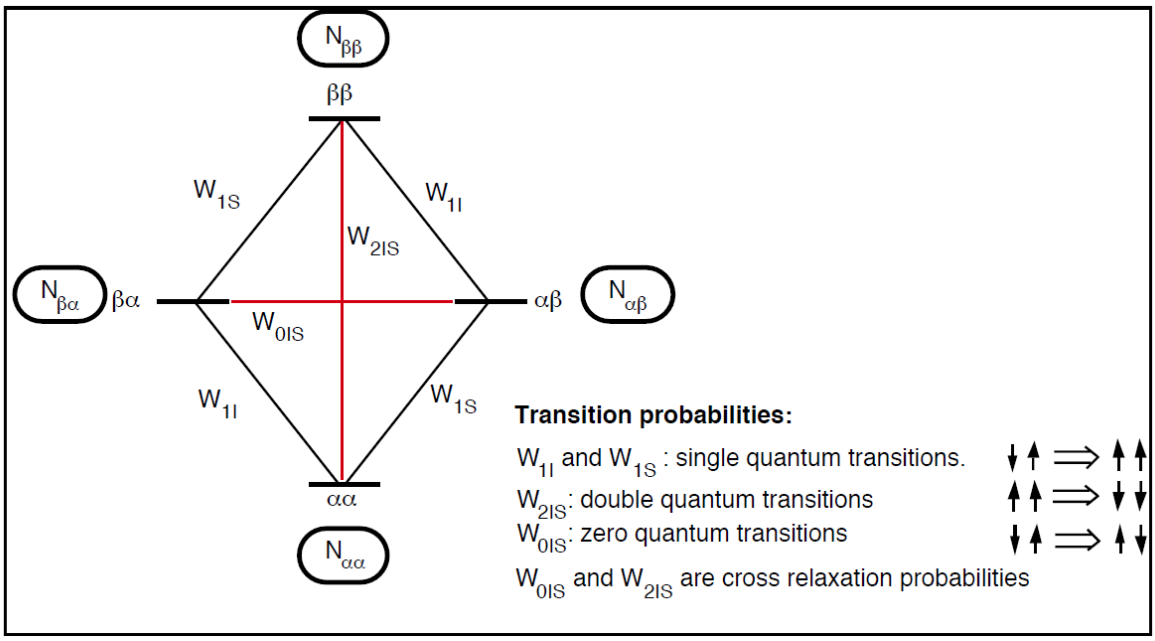


Figure 2.8: A simplified analysis of the NOE phenomenon: Transition probabilities under saturation. Figure adapted from Keeler (2005).

2.8.2 Correlation time

To induce the spin transition, a molecule should tumble with proper frequency to provide a suitable fluctuating field. The dynamics of such movements are expressed by the correlation time τ_c. This defines the average time required for the molecule to rotate through an angle of one radian. A very rough estimation of correlation time can be made according to following formula:

$$\tau_c \approx M_r \times 10^{-12} s$$ Equation 2.7

Where M is the Molecular mass

Molecules which tumble rapidly (small molecules) in solution are likely to favour the higher energy W_{2IS} process and hence exhibit positive NOEs. Whilst those tumble slowly (large molecules) will favour the W_{0IS} process and thus display negative NOEs (see figure 2.9).

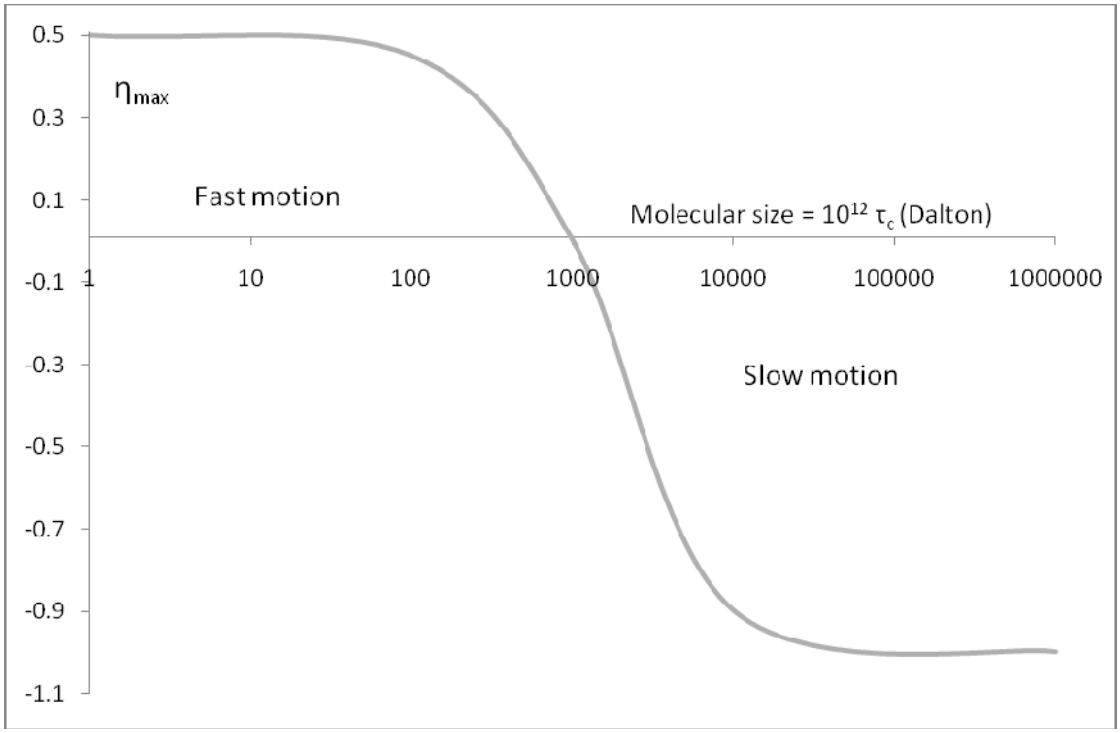


Figure 2.9 The variation in the maximum theoretical NOE in a two-spin system as a function of the molecular size (related to the correlation time τ_c). The region of the fast motion is the extreme narrowing limit and that of the slow motion is the spin-diffusion limit.

2.8.3 NOE and mixing time

The rate at which the system changes itself is proportional to the internuclear distance between the spins I and S. Therefore, the NOE has a distance-dependence such that the change in observed signal intensity for spin I is proportional to r^{-6}. Long-range NOEs are detectable to a maximum distance of 0.4 – 0.5 nm and hence can provide experimental distance restraints for macromolecular structure determination. Generally, strong NOE intensities indicate the distances between protons are less than 0.25 nm, while weak intensities indicate the distances are between 0.37 nm to 0.5 nm.

The NOE is transferred between protons during the "mixing time" in the NOESY experiment. The mixing time may be near 400 millisecond for small molecules (< 1kDa) and near 100 millisecond for large molecules (> 10 kDa). Longer mixing times allow the NOE to enhance in intensity. However, mixing times that are too long can cause spin diffusion to another partner, leading to inaccurate NOEs.

2.9 Residual Dipolar Coupling (RDC)

The residual dipolar coupling between two spins in a protein occurs if the molecules in solution exhibit a partial alignment leading to an incomplete averaging of spatially anisotropic dipolar couplings.

RDC measurement can provide information on the global folding of the protein. As opposed to traditional NOE based NMR structure determination, RDCs provide long distance structural information (see section 2.9.2). It also provides information about the dynamics in molecules on time scales slower than nanoseconds. Thus

measurement can be very important for structural calculation and dynamics studies, especially for large proteins (larger than 25 kDa) (Clore, 1998).

For two dipole-coupled nuclei, A and B, the observable dipolar coupling in solution, D_{AB}, can be expressed as:

$$D_{AB}(\theta,\phi) = A_a^{AB}\left\{(3\cos^2\theta - 1) + \frac{3}{2}R(\sin^2\theta\cos 2\phi)\right\} \qquad \text{Equation 2.8}$$

$A^{AB}{}_a$ and R are the axial and rhombic components, respectively, of the molecular alignment tensor, Aa, in the principal coordinate frame. θ is the angle between the internuclear vector and the z axis of the alignment tensor, Φ is the angle between the projection of the internuclear vector onto the x-y plane and the x axis. To use RDCs in any type of structure refinement, good estimates for Aa and R in Equation 2.8 must be available. There are several methods for determining Aa and R, and the choice of which one to use is depends in part on whether a reasonably accurate structure is available prior to the refinement. For example, the alignment tensor can be calculated from the known structures using PALSE software (Zweckstetter, 2008).

2.9.1 Measurement of RDC

RDC measurement in solution consists of two steps: molecule alignment and NMR experiments.

For a diamagnetic protein and moderate magnetic field, molecules have little preference in orientation. Thus the average dipolar couplings tend to be zero. To obtain RDCs in solution, a co-solute is needed that causes a partial alignment and a net non-zero average value without causing severe coupling interactions and distorted

spectra. For example, filamentous Pf1 bacteriophages (large anisotropic magnetic susceptibility) are commonly used to achieve tailored partial alignment.

Methods based on the separation of peaks centers (splitting) can be used to measure coupling constants between nuclei (See figure 2.10). For example, the dipolar coupling between ^{1}H and attached ^{15}N causes splitting in ^{1}H, ^{15}N HSQC in addition to the J coupling effect.

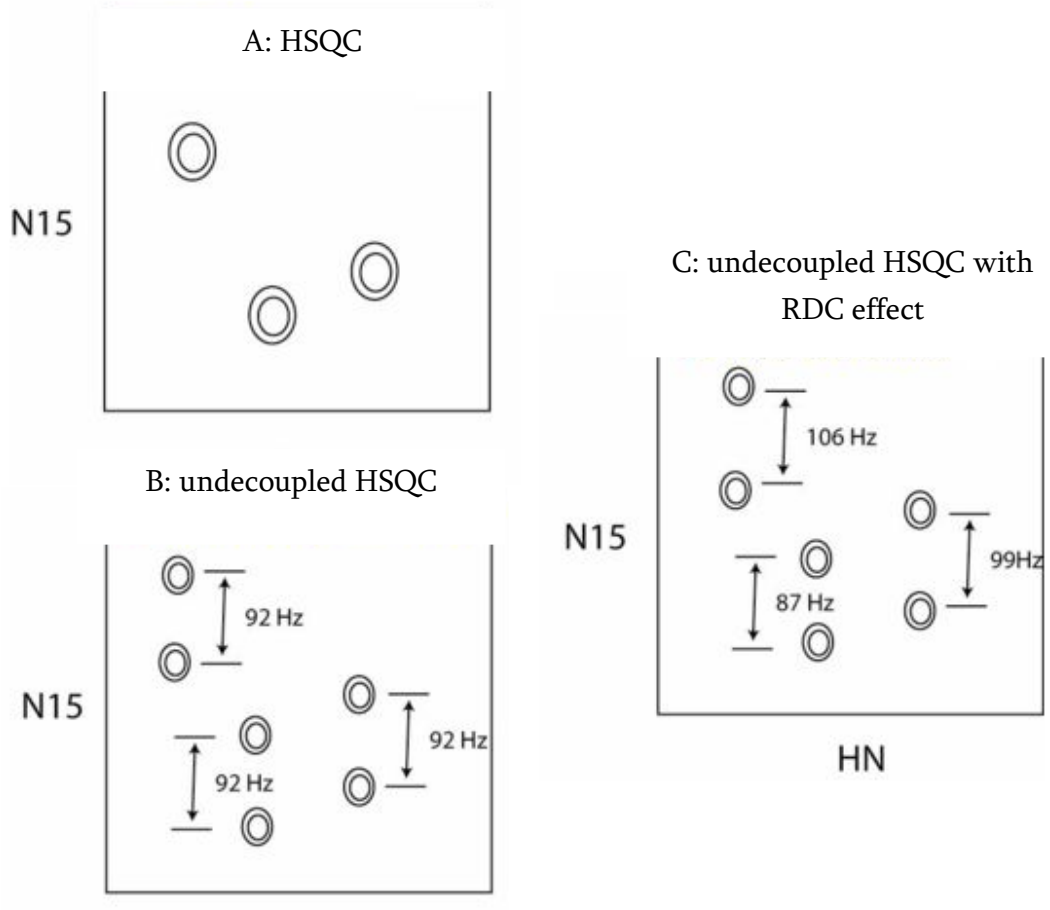


Figure 2.10 RDC effect on HSQC spectrum. Spectrum A: HSQC recorded in a standard manner with ^{1}H, ^{15}N decoupled in the indirect dimension; Spectrum B: J-splitting, Spectrum C: J-splitting and D-splitting (modified from (Lipsitz and Tjandra 2004))

2.9.2 RDC sand NOEs as restraints for structure calculation

Most structure calculations by NMR are based on NOEs between different types of protons in the protein. NOEs provide distance restraints, which depend on the inverted sixth power of the distance between the nuclei, r^{-6}. However, it only provides structural information about the nearest neighborhood. RDCs provide additional restraints information over NOEs (Kessler, 2005):

1. RDCs provide relative orientations of internuclear vectors. These couplings are not subjected to any limitations regarding to the distance of these vectors.
2. ^{1}H - ^{1}H RDCs between nuclei with a distance larger than 5-6 Å can be detected. This is because the RDC is proportional to r^{-3} whereas NOE is proportional to r^{-6} (See equation 2.9).

$$D_{IS} = \frac{\hbar \gamma_I \gamma_S}{4\pi r_{IS}^3}[1 - 3\cos^2\theta] \qquad \text{Equation 2.9}$$

3. In large molecules (>25kDa) it is often impossible to record NOEs due to deuteration. RDCs experiments have no problem at this situation.

2.10 One dimensional NMR spectroscopy (1D NMR)

One dimensional NMR spectroscopy includes all regular nuclei, e.g.^{1}H and two types of experiment are typically used: non J-decoupled and J decoupled. Measurement of 1D NMR can be divided into 3 stages.

- Stage one: the sample reaches equilibrium
- Stage two: a RF signal is transmitted for sufficient time and power to rotate the magnetic vector from z axis to x, y plane.

- Stage three: The signal that evolves due to precession of the magnetization vector is measured after the pulse. The vector returns to equilibrium on the z-axis. The fluctrating voltage induced in the receiver coil by the precessing magnetization is called free induction decay (FID). The receiver coil is placed in such an orientation that only precession in the x-y plane, i.e., transverse magnetization will be detected.

The FID, which contains time domain information, is usually transformed in to the frequency domain by Fourier Transform for further analysis (See figure 2.11). A well featured 1D NMR spectrum of a recombinant protein is useful to characterize the folding features of the protein (*e.g.* folded state, self-association, concentration and some dynamics).

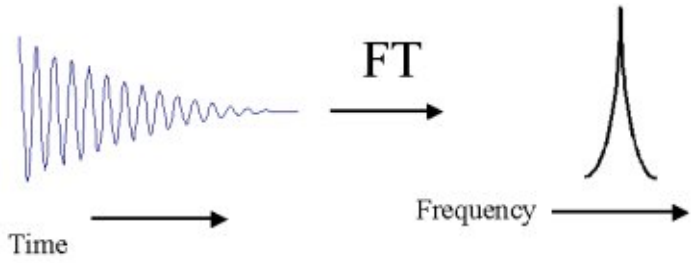


Figure 2.11 Fourier transform: from time dependent to frequency dependent.

2.11 Two Dimensional NMR spectroscopy

The chemical shift of a nucleus is an indication of its electronic and magnetic environments within the molecule; scalar coupling constants imply the through-bond proximity between nuclei; NOE spectra can show the stereochemical (through space) information for intramolecular interactions. However, when we move to more

complicated molecules, the standard 1D spectrum can suffer severe signal overlap and line broadening, which render the spectral parameters hard to analyze. Multidimensional experiments introduce additional frequency domains that spread data and alleviate those problems.

2.11.1 A basic two dimensional experiment

A 1D spectrum is a plot of intensity *vs.* frequency; while in 2D experiments intensity is plotted as a function of two frequencies, usually called F1 and F2. In 2D NMR, spectroscopic data are collected as a function of two time scales, t_1 (evolution) and t_2 (detection) (see figure 2.12). The resulting data set is then processed by two Fourier transformations converting the time domains to two frequencies.

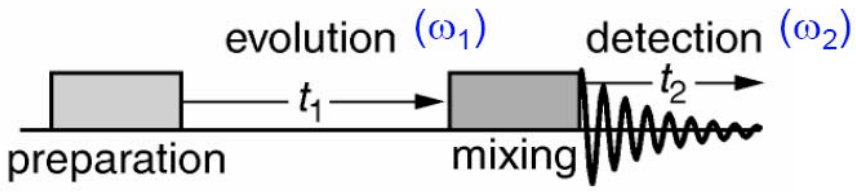


Figure 2.12 A classical 2D NMR experiment can be divided into four phases: preparation, evolution, mixing and detection (adapted from (Keeler, 2005)).

In the first period, called the preparation time, the sample is excited by one or more pulses. The resulting magnetization is allowed to evolve for the first time period, t_1, under its NMR frequency. Another period follows, called the mixing time, which consists of further pulses, where the magnetization may be transferred onto another

nucleus, *e.g.* via NOE or scalar coupling. After the mixing period the signal of the second nucleus is recorded as a function of the second time variable, t_2. The experiment is repeated with t_1 incremented, to build a series of FIDs at multiple values of t_1 (see figure 2.13). The exact nature of the preparation and mixing periods determines the information found in the spectrum, e.g., COSY, NOESY and TOCSY spectra (see below).

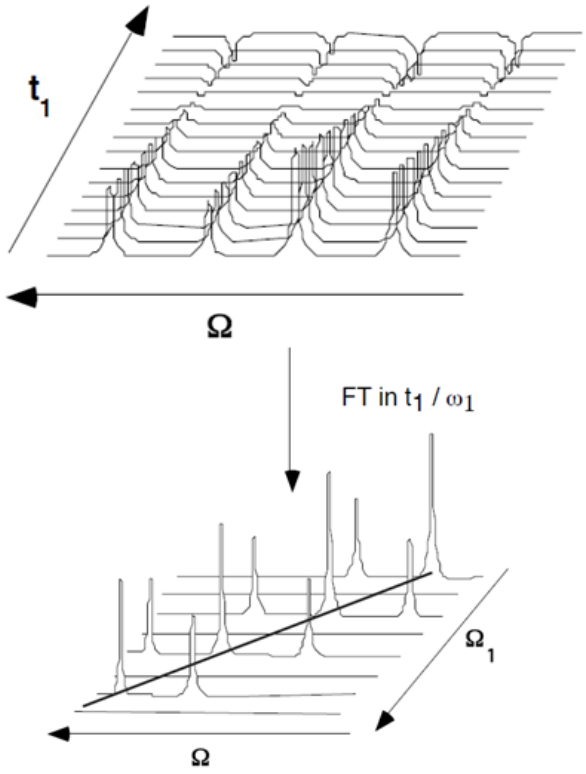


Figure 2.13 2D spectrum builds up: Starting with t1=0, the evolution time t1 is incremented by Δt1. A series of spectra is generated. Note that there is an interferogram ("FID") in the orthogonal direction (t1). FT of t_1/ω_1 generates the indirect dimension Ω1.

2.11.2 COrrelation SpectroscopY (COSY) and Total Correlation SpectroscopY (TOCSY)

Two-dimensional COSY (COrrelation SpectroscopY) experiments are capable of determining the connectivity of a molecule by identifying nuclei that are scalar coupled. In the COSY experiment, magnetization is transferred by J coupling during the mixing period. In ^{1}H ^{1}H COSY spectra, protons that are more than three chemical bonds apart give no cross peaks because the J coupling constants are normally close to 0. Therefore, only signals of protons which are two or three bonds apart are visible in a COSY spectrum. The cross peaks between HN and Hα protons of a protein are of special importance because the *phi* torsion angle of the protein backbone can be derived from the ^{3}J coupling constant between them via Karplus relation.

In most respects, TOCSY is similar to the COSY. The exception is that instead of showing correlations between nuclei within the 3 bonds of each other, the TOCSY can correlate all nuclei in a spin system.

2.11.3 Hereonuclear Single Quantum Coherence (HSQC)

In this special type of COSY experiment, the magnetization is transferred directly to a covalently attached heteronucleus (normally ^{15}N or ^{13}C). The spectrum contains a peak for each unique proton attached to the heteronucleus being considered. Thus, if the chemical shift of a specific proton is known, the chemical shift of the coupled heteronucleus can be determined, and vice versa.

The ^{15}N HSQC experiment is probably the most frequently recorded 2D experiment in protein NMR. The HSQC experiment can be performed either using the natural abundance of the ^{15}N isotope, or more commonly using isotopically labeled protein.

The latter can be recorded at much lower concentrations, but requires recombinant expression of the protein.

Each residue of the protein (except proline) has an amide proton attached to nitrogen in the peptide bond. If the protein is folded, the peaks are usually well dispersed, and most of the individual peaks can be distinguished. The HSQC experiment is thus very useful in screening candidates for structure determination by NMR. The number of peaks in the spectrum should match the number of residues in the protein (though side chains with nitrogen-bound protons will add additional peaks).

2.11.4 Nuclear Overhauser Effect SpectroscopY (NOESY)

^{1}H, ^{1}H Nuclear Overhauser Effect SpectroscopY is a method to identify spins undergoing cross-relaxation. In proton-NMR, direct dipolar couplings provide the primary means of cross-relaxation, and so spins undergoing cross-relaxation are those close to one another in space rather than through chemical bonds (see section 2.8). Thus, the cross peaks of a NOESY spectrum indicates which protons are close in space and their intensities are related with their distance in space. It is different from COSY, which relies on J-coupling to provide spin-spin correlation and forms the basis for tertiary structure determination by NMR spectroscopy (Neuhaus and Williamson 1989). 1H NOESY spectra also has 3D derivations that add ^{13}C and ^{15}N dimensions in protein applications.

2.12 Three dimensional NMR spectroscopy and protein structure determination

A three dimensional NMR experiment (see figure 2.14) can easily be constructed from a two dimensional one by inserting an additional indirect evolution time and a second mixing period between the first mixing period and the direct data acquisition. Each of the different indirect time periods (t_1, t_2) is incremented separately.

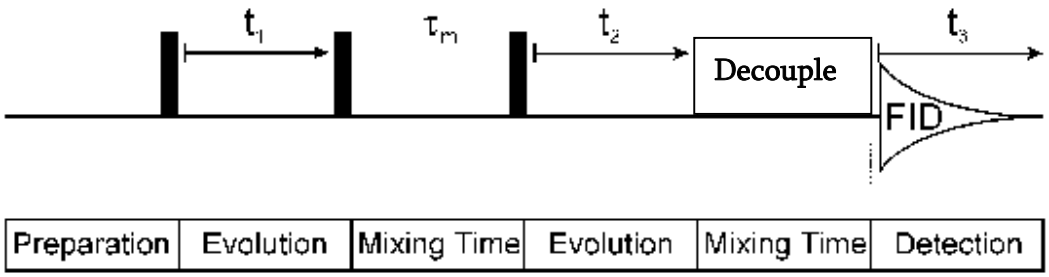


Figure 2.14 A typical pulse sequence for three dimensional NMR (one channel). MLEV (Malcolm Levitt's CPD sequence) is composite pulse decoupling scheme.

3D experiments are more commonly performed on uniformly labeled (mostly ^{15}N and ^{13}C) proteins and are displayed in three frequency domains. Combination of different 3D experiments provides the basis for solving protein structures. The general strategy for determining protein structure and interaction studies can be divided into four steps (Gardner and Kay 1998).

1. Routine 1D NMR to verify that the protein is folded and stable at the experimental conditions for the desired set of spectra. .
2. 2D 15N HSQC to identify flexible polypeptide from the folded domain.

3. Backbone 3D NMR experiment to assign backbone resonances and reveal the secondary structure of the protein. The assigned HSQC spectrum can be used in chemical shift mapping for an interaction between the ligand and protein.

4. Sidechain 3D NMR and NOESY experiments to assign sidechain resonances, and using ^{1}H-^{1}H NOEs, dihedral angles and RDCs thus calculate the structure.

2.13 Backbone assignment experiments

Assignment of backbone atoms and linkages with their sequential neighbours is usually the first step of structure assignment. The experiments discussed below allow us to assign most of the HA, HB, CA, CB and HN resonances of residues of the protein and group them into spin systems.

2.13.1 HNCACB and CBCA(CO)NH

These experiments are used to determine the C_{α} and $C_{ß}$ frequencies of pairs of contiguous amino acids in the protein sequence.

In the HNCACB, the amide proton is pulsed and magnetization is passed to the atoms covalently bonded with it. The magnetization is first transferred to the attached ^{15}N and then further transferred to the $^{13}C_{\alpha}$ and $^{13}C_{ß}$ of the current and preceding residue. After the resonances of $^{13}C_{\alpha}$ and $^{13}C_{ß}$ have been recorded, the magnetization is transferred back to amide proton following the same track (See figure 2.15A). In the CBCA(CO)NH, the magnetization is passed from the alkyl protons to $^{13}C_{\alpha}$ and $^{13}C_{ß}$ atoms of its own residue and then goes backwards to carbonyl and amide of the following residue (See figure 2.15B). Combined with the HNCACB, peaks of residues appearing in HSQC can be assigned.

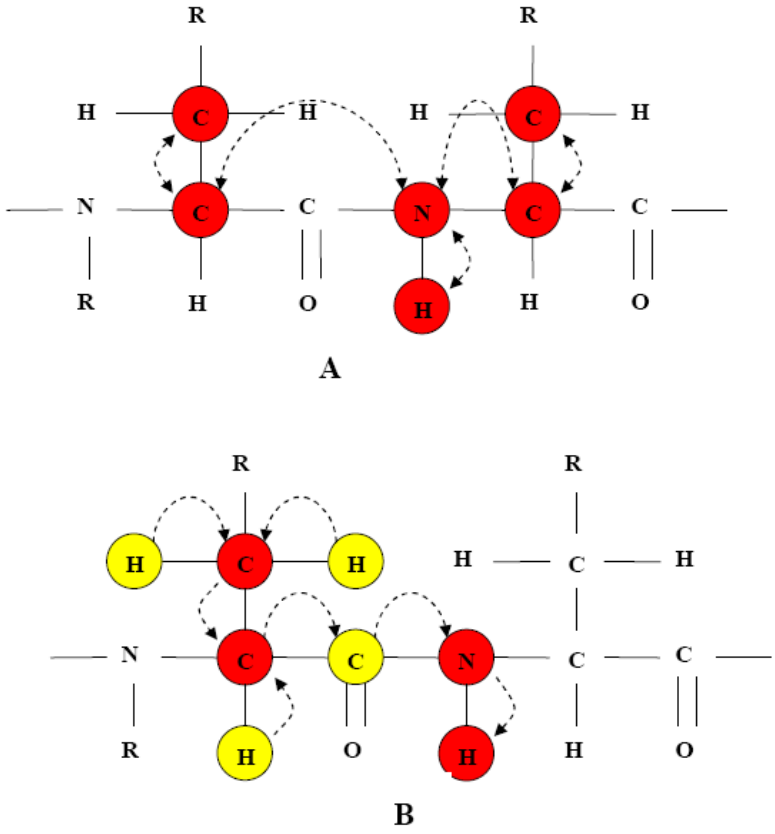


Figure 2.15 The magnetization transfer in the HNCACB and CBCA(CO)NH. A: sequential magnetization transfer through bonds in the HNCACB. B: sequential magnetization transfer through bonds in the CBCA(CO)NH. Both yellow and red atoms are involved in the magnetization transfer whereas only frequencis of the red atoms are recorded.

2.13.2 HN(CA)CO and HNCO

This pair (See figure 2.16) works as supplementary experiments to the HNCACB and CBCA(CO)NH and provides the CO frequencies of the residues. In the HN(CA)CO, the amide proton is pulsed and the magnetization is transferred via the $^{13}C_{\alpha}$ to the carbonyl from both preceding and following residues. In the HNCO, the

magnetization from the amide proton is only transferred to ^{13}C of the preceding carbonyl group.

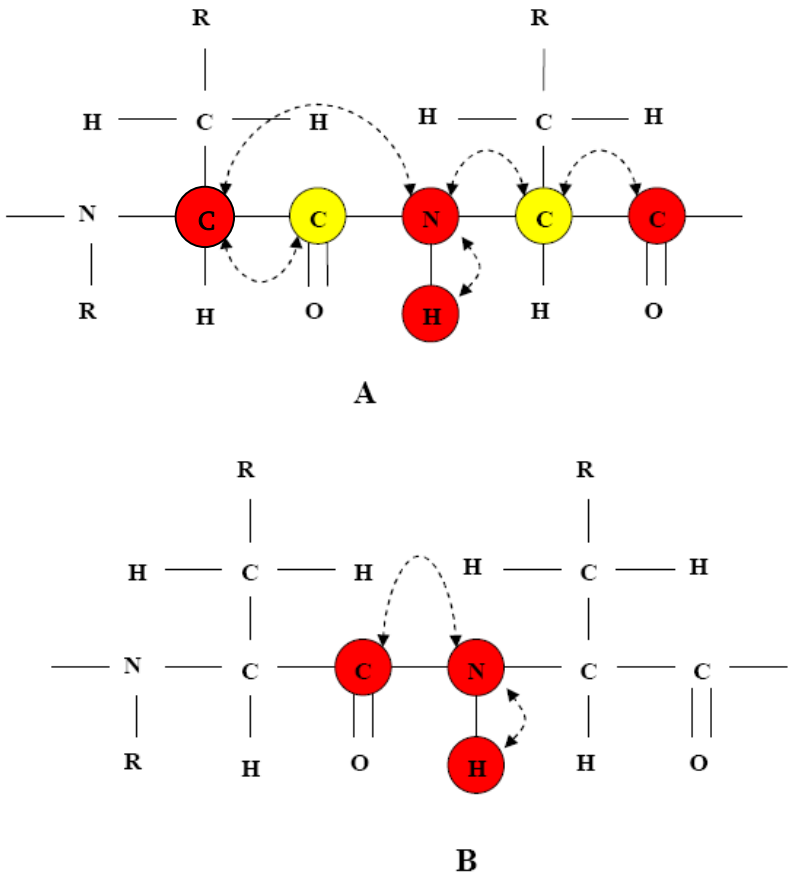


Figure 2.16 The magnetization transfer in the HN(CA)CO and HNCO. A: sequential magnetization transfer through bonds in the HN(CA)CO. B: sequential magnetization transfer through bonds in the HNCO. Both yellow and red atoms are involved in the magnetization transfer whereas only the frequencies of the red atoms are recorded.

2.12.3 HBHA(CBCACO)NH

The magnetization transfer of the HBHA(CBCACO)NH is similar to the CBCA(CO)NH but the H_α/ H_β proton signals are recorded instead of the $^{13}C_\alpha/C_\beta$. This allows assignment of the H_α and H_β frequencies in the sidechain (See figure 2.17).

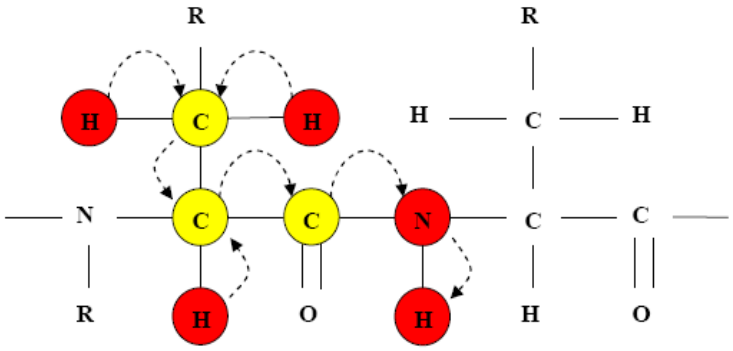


Figure 2.17 Magnetization transfer in the HBHA(CBCACO)NH. Both yellow and red atoms are involved in the magnetization transfer whereas only resonances of the red atoms are recorded.

2.12.4 H(C)CH-TOCSY

The 3D H(C)CH-TOCSY experiment is specifically designed to correlate side-chain aliphatic proton and ^{13}C resonances via J(CH) and J(CC) coupling. The experiment provides nearly complete assignments of all aliphatic ^{1}H and ^{13}C resonances (See figure 2.18). Commonly, in some resonances of the long aliphatic side chains (as Lys or Arg), there is substantial overlap and more than one variant of the experiment (e.g., H(C)CH and (H)CCH-TOCSY) is required.

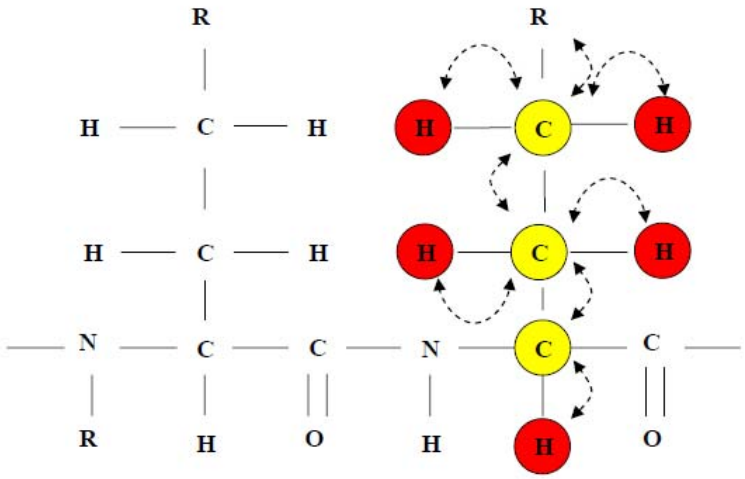


Figure 2.18 Magnetization transfer in the H(C)CH-TOCSY. Both yellow and red atoms are involved in the magnetization transfer whereas only resonances of the red atoms are recorded.

2.13 Resonance assignment and MARS

Backbone resonance assignment is a prerequisite for structure determination of proteins by NMR. In the assignment process, the chemical shifts of ^{1}HN, ^{15}N, ^{13}Ca ^{13}Cb and ^{13}CO of residue I and ^{13}Ca, ^{13}Cb and ^{13}CO of residue i-1 obtained from the previously described experiments are assembled together according to the sequence. This procedure can be done manually or automatically by using software like MARS (robust automatic backbone assignment of proteins).

MARS assembles all chemical shifts into arrays called pseudoresidues (PRs), and each of PR is associated with a single ^{1}HN, ^{15}N root (Jung and Zweckstetter 2004). It comprises three steps: (1) Identification of the amino acid type in PRs based on chemical shifts; (2) Matching and linking sequential PRs into segments; (3) Mapping

PR segment onto the primary sequence. MARS can also accept combinations of automated and manual assignments.

2.14 NMR-based Structure Calculation

NMR data do not uniquely define a protein structure because all the restraints are ranges of allowed distances, angles etc. rather than unique values. This reflects the fact that the experimental data contains uncertainties in measurement and interpretation.

In fact, proteins are dynamic molecules in solution, so a view of protein structure as a single set of atomic coordinates may be unrealistic. Thus NMR methods do not calculate a single structure but rather repeat the structure calculation process many times to generate an ensemble of structures. These methods are designed to explore the regions of conformational space that satisfy the experimentally derived restraints. NMR restraints such as NOE distance restraints and hydrogen bond restraints often explicitly include the positions of hydrogen atoms. By contrast, X-ray structures do not generally include hydrogen atoms in atomic coordinate files, because the heavy atoms dominate the diffraction pattern and the hydrogen atoms are not explicitly seen.

The methods include distance geometry, restrained molecular dynamics, simulated annealing and other hybrid methods. In distance geometry, one uses the NOE-derived distance restraints to generate a distance matrix, which is then used as a guide in calculating a structure. These structures are able to produce the correct overall fold but usually have poor local geometry (e.g. improper bond angles, distances). Hence distance geometry must be combined with some extensive energy minimization method to generate physically reasonable structures.

Restrained molecular dynamics involves computing the potential energy V with respect to the atomic coordinates. Usually this is defined as the sum of a number of terms:

$V_{total} = V_{bond} + V_{angle} + V_{dihedr} + V_{vdW} + V_{coulomb} + V_{NMR}$ Equation 2.8

These terms terms are energy terms corresponding to such forces as van der Waals and electrostatic repulsions and attractions, cost of deforming bond lengths and angles. The NMR restraints are incorporated into the V_{NMR} term, which is a "pseudoenergy" or "pseudopotential" term included to represent the cost of violating the distance and dihedral restraints measured from NMR data.

Simulated annealing is essentially a special implementation of rMD and uses similar potentials but employs raising the temperature of the system and then slow cooling in order not to get trapped in local energy minima. It is an efficient method for locating the global minimum of the target function. Otherwise, it may be extremely complicated due to the large number of parameters contributing to the energy landscape.

Programs such as ARIA (Linge et al. 2003a) incorporated in CNS (Brunger, 1998) as part of the software for the calculation of distance geometry and simulated annealing), with automatic routines for iterative assignment of ambiguous NOE restraints, start the calculation with some sets of unambiguous NOEs and calculate an ensemble in an iterative approach. In general, these softwares do the structure calculation of a protein in four steps:

1. Generation of a random protein starting structure.
2. Perform on this random structure an *in vacuo* iterative simulated annealing based on experimental restraints.

3. Water refinement of the calculated structures (not with CYANA (Güntert, 1997), but with CNS protocols / ARIA)
4. Validation of the calculated structures.

2.15 Summary

NMR spectroscopy is an extremely useful and powerful tool which can be used to study protein structure and dynamics. The NMR phenomenon arises due to a physical property of the nucleus termed 'spin'. The associated magnetic moment becomes quantised in the presence of an external magnetic field, such that nuclei exhibit a precessional motion (*i.e.* about the field axis) and essentially behave as small 'bar magnets'. Surrounding electrons give rise to characteristic chemical shifts, which in turn can be correlated (*i.e.* between nuclei of different types) and detected (as free induction decay) using suitable NMR experiments (*e.g.* triple resonance). Such chemical shift values are ultimately attributed to specific nuclei (*i.e.* comprising the protein backbone or amino acid sidechains) during the process of 'resonance assignment' and which forms the basic pre-requisite for structure determination.

However, a major drawback of macromolecular NMR is its size limitation caused by two technical barriers (Yu 1999). First, larger molecules have slower tumbling rates and shorter NMR signal relaxation times. This reduces the sensitivity of the complicated pulse sequences that often use long delays for the necessary coherence transfer steps. The increased molecular weight also introduces more complexity to a given spectrum, simply because there are more NMR-active nuclei and, therefore, more interactions among them. The current size limit of NMR for uniformly labeled protein is around 35 kDa (Clore and Gronenborn 1998). Structures of 82kDa have been determined by NMR (Malate synthase G, Kay et al.). However, recent specific

methyl labeling schemes and transverse relaxation optimized spectroscopy (TROSY) has extended the molecular size range to 1000kDa for studying the solution properties and interactions of large molecular assemblies (Isaacson et al. 2007; Sprangers and Kay 2007).

Chapter 3 Material and Methods

This chapter describes the details of methods used in the cloning and recombinant protein expression from MIC2-M2AP and MIC3-8 complexes for NMR spectroscopy (Saouros et al. 2007). As the conditions of different recombinant proteins may vary, the specific conditions can be found in Chapter 4 and 5.

3.1 Construct design

Construct design was based on the primary structure alignment and secondary structure prediction for different domains (see individual chapter)(Brossier and David Sibley 2005) (Gasteiger et al. 2003). A series of constructs were designed to produce recombinant protein matched to putative domains from the two microneme protein complexes, including M2AP β-sheet rich domain, A domain and Thrombospondin-like (TSR) domain from MIC2 (see individual chapter).

3.2 Primers, PCR and PCR purification

Forward and reverse primers were designed for Ligation Independent Cloning (LIC - Merck) and were purchased from MWG Biotech (Germany). The LIC is able to clone PCR products without need for the restriction enzyme digestion or ligation reactions, which is achieved by adding a short 5' overhang to the primer. The enzymes used include Pfu (Qiagen), Taq (Novagen) and KOD HiFi (Novagen) DNA polymerase.

cDNAs (MIC2, M2AP, MIC3 and MIC8) were obtained from our collaborator Prof. Dominique Soldati (Univ. of Geneva, formerly at Imperial College) in pGEM vectors. 100 µl PCR reactions were performed by a Mastercycler Gradient (Eppendorf). The PCR products were purified by a Gel Extraction Purification kit (Qiagen) to remove the excess dNTPs and the original pGEM vector. A complete list of primers used is shown in Appendix 4.

3.3 Plasmid and Plasmid cloning

The pET vector (Merck) series is widely used throughout the experiment. The pET-32 Xa/LIC vector (See figure 3.1) fused with the 109aa Trx-Tag thioredoxin protein, His-Tag and S-Tag sequences is used for rapid cloning. Thioredoxin (trx) is 17.5kDa protein that promotes disulphide bridge formation, which is essential for the folding of Cysteine rich microneme proteins. The vector also provides a factor Xa cleavage site and a thrombin cleavage site for complete removal of the fusion partner, i.e., thioredoxin, His-tag or S-Tag. pET-32 variations with an Enterokinase site or thioredoxin deletions were also used in some cases.

Cloning of PCR fragments was conducted using a ligation Independent cloning (LIC) kit (Merck). It has the advantage that it does not require the conventional use of restriction enzymes or ligases, but instead utilizes the 3' → 5' exonuclease activity of T4 DNA polymerase.

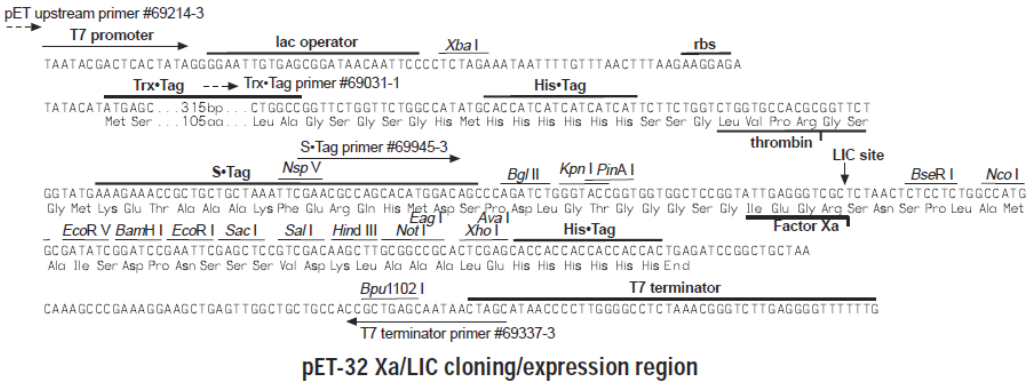


Figure 3.1 Schematic representation of pET-32 Xa/LIC cloning and expression regions (Qiagen)

3.4 Plasmid Sequencing

The plasmids were purified by Miniprep Kit (Qiagen) and sequenced by Advance Biotechnology Centre (ABC), Imperial College London using T7 terminator reverse primer, and other specific primers. Later samples were sent to Cogenics (UK) after the closure of ABC.

3.5 *E. coli* strains

Several strains were used commonly in the laboratory. The XL1-Blue (Novagen) competent cells were used primarily for plasmid cloning purpose. The BL21 (DE3) cells were used for initial expression tests, which has the advantage of no antibiotic resistance. The Origami strain was employed to express all the MIC protein constructs. Origami (DE3) greatly promotes the formation of the disulphide bridge formation by mutations in the thioredoxin reductase (trxB) and glutathione reductase (gor) genes (Prinz et al. 1997). It has the designation (DE3) which is lysogenic for a λ prophage that contains an IPTG inducible T7 RNA polymerase. Its dual anti-biotic resistance (Tetracyclin and Kanamycin) provides high selectivity for the host cells.

3.6 Protein Expression and Purification

For isotopic enrichment of proteins with ^{15}N and ^{13}C, a minimal medium containing ^{15}N-labelled ammonium chloride (0.07%) and ^{13}C-labelled glucose (0.2%) was used for bacterial culture (See section 3.9). Otherwise, Lysogeny broth (LB) medium was used for unlabeled sample. A 50 ml sub-culture in unlabelled minimal media was grown

overnight to facilitate adaptation of the bacteria and isotopically labelled compounds added upon scale-up to a 1 L preparation.

Cells were grown to an optical density of 0.6-0.7 (at 600 nm) before protein expression was induced by addition of 1000 μM isopropyl -β-thiogalactoside (IPTG). Induction temperatures were either 30 °C or 18 °C and cells were cultured for a further 4 hours (minimum) to overnight. Cell pellets were harvested by centrifugation at 4000 rpm (Sorvall), flash frozen in liquid nitrogen and stored at -20 °C until required.

Cells were lysed by French Press (SLM Instruments) and lysates were collected after centrifuged at 1500 rpm (SS-34 rotor, Sorvall) for 30 minutes. EDTA-free protease inhibitor cocktail (Roche) to prevent protein degradation and Benzonase to digest nucleic acids were used in buffers for all chromatography procedures.

The protein was purified by binding 6X histidine-tagged MIC proteins onto a Ni-NTA Agarose (Qiagen) column. The samples were added to the equilibrated column after washing with binding buffer (20mM Tris-HCl, pH 7.9, 5 mM Imidazole, 0.5 M NaCl), then subsequent washing with wash buffer (20 mM Tris-HCl, pH 7.9, 40 mM Imidazole, 0.5 M NaCl) and elution buffer (20 mM Tris-HCl, pH 7.9, 250 mM Imidazole, 0.5 M NaCl) yielded purified protein. Final samples were analysed by SDS-PAGE.

3.7 Digestion of Fusion Protein and Separation

Before digestion, the protein sample was buffer exchanged into Factor Xa buffer (20mM Tis-HCl pH8.0, 500mM NaCl, 2mM $CaCl_2$) or thrombin buffer (20mM Tis-

HCl pH8.4, 150mM NaCl, 2.5 mM $CaCl_2$) for cleavage. The membrane used for dialysis was Spectra/Por® dialysis membrane (Spectrum).

The MIC recombinant fusion proteins were quantified by Bradford assay or direct absorbance assay at 260 and 280 nm. A relevant amount of enzyme (factor Xa) was added to the protein sample and incubated at room temperature. The reaction can be stopped irreversibly by 1mM Pheny Methyl Sulfonyl Fluoride (PMSF), a serine protease inhibitor. A further affinity chromatography step (discussed above) enables retention of the Histidine-tagged fusion protein, but allows MIC proteins to pass through.

3.8 Size exclusion chromatography

Preparative gel filtration was frequently conducted to remove low level contaminants (*i.e.* residual thioredoxin) and analytical gel filtration performed for diagnostic purposes or functional studies. Gel filtration was performed using an AKTA™ design System (Amersham Biosciences) in combination with a Superdex™ 75 or Superdex™ 200 columns. Samples were concentrated to a volume of 100 – 500 µl (see following section for details) and contained at least 150 mM salt (to prevent non-specific charge-based interactions with the resin). Samples were injected onto the column at a flow rate of 1 ml min^{-1} and UV absorbance at 280 nm recorded.

3.9 Protein concentration

Protein samples were concentrated (and buffer exchanged) using a 20 ml Vivaspin (Vivascience) (4500 rpm) at 4 °C. A typical NMR buffer comprised of 20 mM phosphate (pH 6.8 – 7.0) and contained 50 – 150 mM NaCl. Samples were concentrated to a final volume of 300 – 500 µl and protease inhibitors added to prevent proteolytic degradation.

3.10 Co-expression of microneme protein complexes

The LIC duet Trx-Tag Ek Adaptor was used to co-express MIC2 and M2AP. It provides a means to co-express two open reading frames (ORFs) in bacteria from a single plasmid (See figure 3.2).

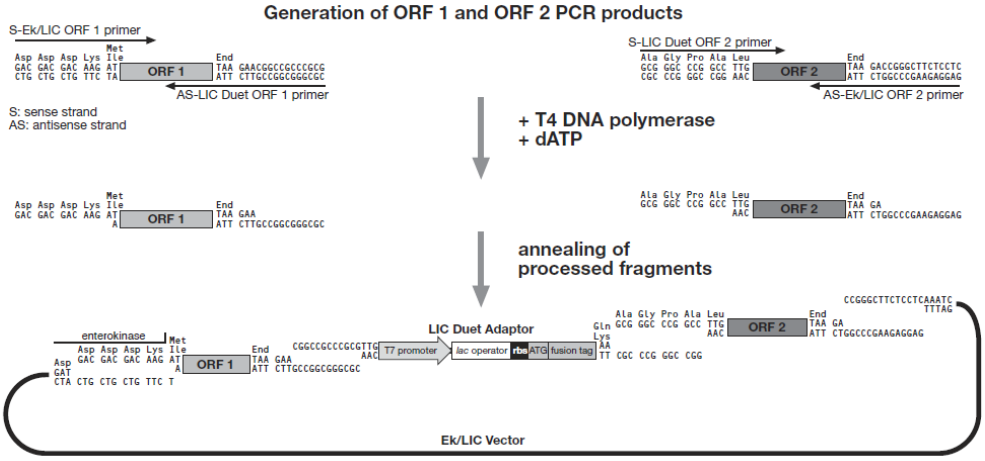


Figure 3.2 Cloning process of LIC duet Adaptors. After ORF1 and ORF2 are amplified with primers that include the indicated 5'-LIC extension, the PCR inserts are treated with LIC-qualified T4 DNA Polymerase (+dATP), annealed to the LIC Duet Adaptor and Ek/LIC vector, and transformed into competent cells. (Merck BioSceinces)

3.11 Isotopic labeling of recombinant proteins

The recombinant proteins were expressed in Minimal media with $^{15}NH_4Cl$ is the only nitrogen source and ^{13}C glucose is the only carbon source (see appendix A2). Minimal media contains the essential ingredients for bacterial species to grow, namely no more nutrition beyond core sources of carbon and nitrogen. With the isotopic labeled carbon and nitrogen sources only available, all the bacterial products will be labeled.

3.12 NMR experiments

Nuclear Magnetic Resonance (NMR) spectra were used throughout the project to determine the foldedness of purified proteins. The Cross Faculty NMR Centre (Imperial College) provides 500MHz and 800MHz (Bruker) NMR Spectrometers equipped with cryoprobes for functional studies of proteins. Occasionally, experiments were recorded in Oxford University on 950MHz spectrometer (Oxford instruments). All experiments for M2AP were running at 308K for optimal spectral properties. And most of the remaining experiments were operated at 303K unless indicated.

3.13 NMR Sample preparation

The protein samples were concentrated and buffer exchanged in Vivaspin (Vivascience) concentrators of molecular weight cut off 5kDa at 4500 rpm. The samples from the digestion were diluted with 20ml NMR buffer and concentrated down to 1ml at 4 ^{o}C. The process was repeated twice to bring down ratio of digestion

buffer to NMR buffer from 1 to 1:399. 10% D_2O was added to the sample for the frequency lock of the NMR spectrometer. A suitable amount of Protease Inhibitor Cocktail (Merck Biosciences) solution with sodium azide (0.1 %) was added to prevent the enzyme activity and microbial growth.

3.14 Backbone and sidechain assignments

All 3D experiments were analyzed by NMRview as described in the introduction. The backbone assignment was carried out first by manual assignment and later using the MARS automatic assignment software (Jung and Zweckstetter 2004), which was powered by scripts from Jan Merchant (Imperial College), using HNCACB, CBCA(CO)NH, HN(CA)CO AND HNCO as raw data.

The chemical shifts lists picked from ^{13}C, ^{15}N TOCSY and HBHA(CBCACO)NH data along with peak information from ^{13}C, ^{15}N NOESY spectra were processed and passed to ARIA (Ambiguous Restraints for Iterative Assignment, software from Institut Pasteur) for structure calculation.

3.15 NMR structure calculation

After side-chain assignment, NOE assignment was completed using the ARIA (ambiguous restraints for iterative assignment) protocol. NMR structures were calculated using ARIA version 2.2 (Linge et al. 2003a) in which CNS is employed for the simulate annealing steps. In each round of structure calculations ARIA, the lowest energy structures are selected and subsequently used to calibrate NOESY spectra (based upon the average distances obtained). After each round, ambiguity is reduced based on the previous round of structure calculation. At the end of the

iterations, 20 lowest energy structures were water refined (A short molecular dynamics trajectory in a thin layer of water that refines the final structure ensemble).

Chapter 4 Identification and structural determination of M2AP

4.1 Preliminary work on MIC3/MIC8 complex

Prior to the project on MIC2/M2AP complex, the MIC3-8 complex was first investigated. MIC3 and MIC8 also form an adhesive complex that is released onto the parasite surface during attachment and invasion. MIC3 has an N terminal chitin-binding-like (CBL) domain, which acts a receptor binding site of mammalian cells. It forms a dimer in order to acquire binding properties and is bound to the parasite by MIC8 monomer (Cerede et al. 2002) (see figure 1.6). In the MIC3-8 complex, MIC3 protein is an extracellular protein with a theoretical molecular weight of 37.9 kDa comprising 359 amino acids. It contains one N-teriminal lectin-like or Chitin binding like domain (CBL), 5 putative epidermal growth factor-like (EGF) motifs and a C-terminal dimerization domain. Similar to MIC3, the transmembrane protein MIC8 has one CBL domain and more overlapping EGF domains and a cytoplasmic tail.

As there is no structural information published yet, the design of suitable constructs *in vitro* for these domains was the primary objective for this complex.

4.1.1 Construct design for MIC3/MIC8

Although an in-depth analysis of MIC3 had been carried out by Cerede group (Cerede, 2006), the boundaries between each domain remain unclear. As there is not much information available about this complex, the strategies used to design constructs were based on the published analysis and supplemented with further alignments. A list of constructs produced can be found in table 4.1.

Construct	Protein	Residue No.	Notes
CBL	MIC3	067-144	CBL motif
CBL-2	MIC3	067-146	CBL motif with one downstream Cysteine
CBL-E	MIC3	067-189	CBL motif with one EGF domain
10CBL	MIC3	077-144	CBL motif without one upstream Cysteine
MIC3C	MIC3	237-359	MIC3 C terminal dimerization domain
MIC3	MIC3	067-359	Mature MIC3 full length
MIC8	MIC8	019-298	Extracellular part

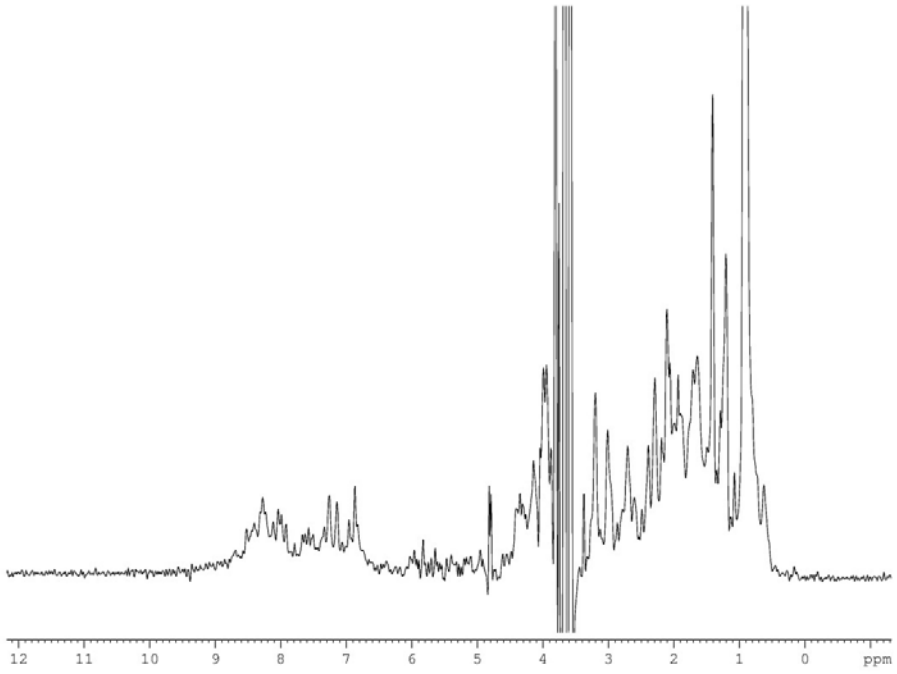
Table 4.1 Constructs tested for the MIC3/MIC8 complex. Residue numbers from the original protein sequence has been indicated.

4.1.2 Constructs expression, purification

All the constructs were successfully expressed under optimized conditions and recombinant proteins with thioredoxin fusion exhibited high solubility. The thioredoxin tag was subsequently removed by proteolysis using factor Xa (described in chapter 3.6).

4.1.3 1D NMR

After successfully cleavage of the fusion protein, the recombinant constructs were assessed by 1D NMR. All the samples were found to be unfolded as no dispersion resonances were observed. 1D spectrum of CBL as an example of the constructs is shown below (see figure 4.1).



^{1}H (ppm)

Figure 4.1 1D (^{1}H) spectrum of unfolded CBL. No amide chemical shifts below 9 ppm, few $C_{\alpha}H$-H shifts in 5 ppm region and structured methyl groups below 0.5 ppm region. (Temperature 303 K, pH 5.7, phosphate buffer)

4.1.4 Conclusions

Producing soluble and folded construct became a bottle neck for this project, and therefore the project was paused before new expression system was introduced. Recently *Pichia* expression system has showed some promise, as a folded recombinant domain of MIC3 has been obtained. Further studies will be continued in the future.

4.2 MIC2-M2AP study

Within a short time after synthesis, M2AP is assembled into a complex with MIC2. The structural and functional characteristics of either component remain largely

unknown. From structural alignments, M2AP is predicted to have a short putative pro-domain, which will be removed in Golgi or nascent micronemes, a β-sheet rich domain and a C-terminal random coil domain (figure 4.2). As the matured protein contains 284 residues, the predicted β-sheet rich domain consisting of around 200 residues is the major part of the protein. The mode of interaction between M2AP and MIC2 remains largely ill-defined, however, the MIC2 interaction site is likely to be located within the C-terminal β-sheet rich domain of M2AP (see section 4.3). The following sections describe the structural characterization of M2AP, especially the structure determination of the β-sheet rich domain.

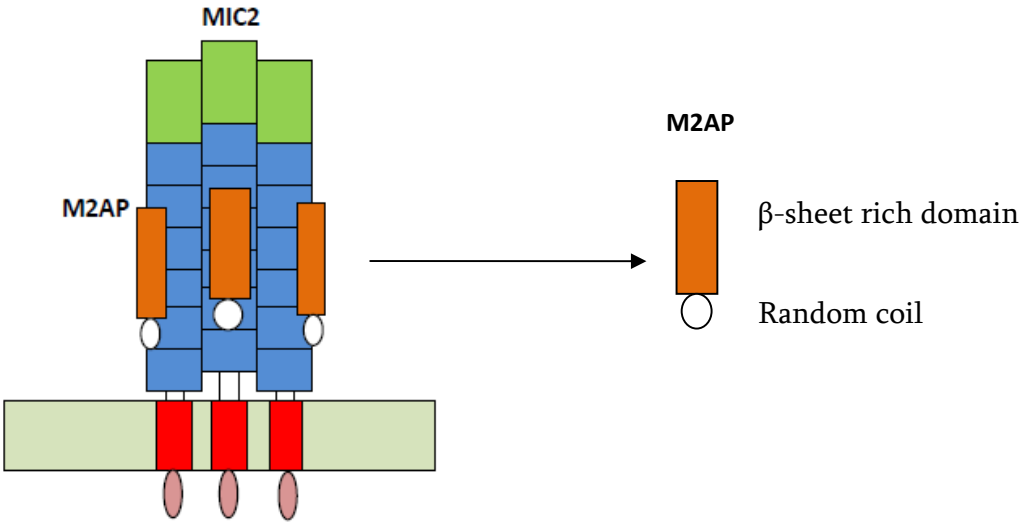


Figure 4.2 Schematic representation of the proposed architecture and interactions of MIC2/M2AP: The extracellular domains of MIC2 participate in the formation of MIC2-M2AP complexes. M2AP contains a β-sheet rich domain and a C-terminal random coil domain.

4.3 Primary sequence alignment, Secondary structure prediction and Construct design

Basic Local Alignment Search Tool (BLAST) found that M2AP from *Neospora caninum*, has a homologous β-sheet rich domain in M2AP from *T. gondii*, but does not contain a C terminal random coil region. The result of this primary sequence alignment helped to define the boundary of the β-sheet rich domain. It also implies the importance of this domain, which probably includes binding to MIC2 to facilitate the complex formation.

Secondary structure predictions were performed by using the Jpred (Cuff et al. 1998) server from the University of Dundee. Jpred uses a consensus based approach to define regions of alpha helix, beta strand and random coil. The results provide a basis to design recombinant protein constructs. Continuous stretches of β-strands/sheets were found in M2AP and a clear boundary between structured and unstructured region could be found (See figure 4.3).

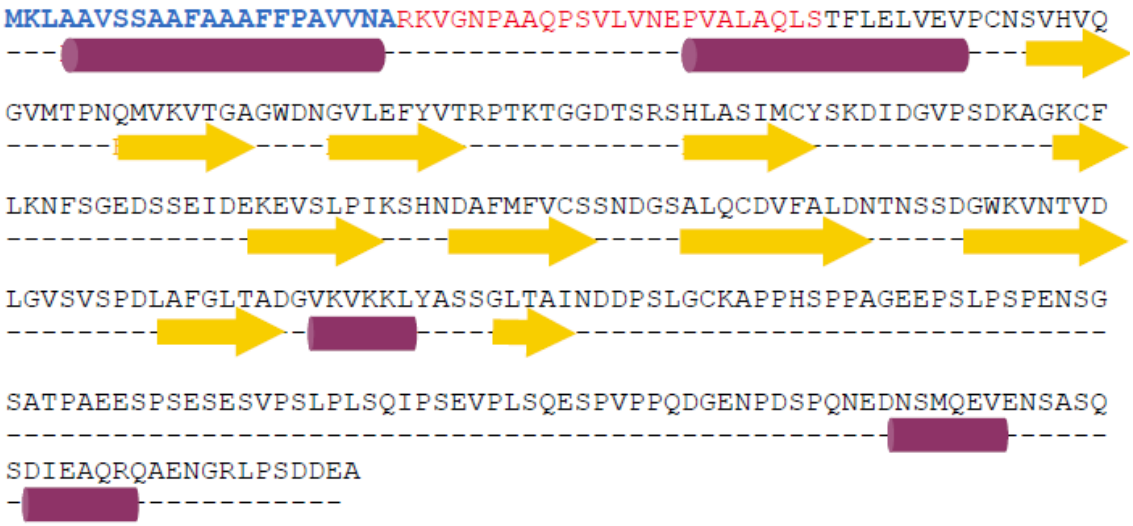


Figure 4.3 JPred secondary structure prediction of M2AP. Note that the putative signal peptide is in blue and the predicted pro-domain is colored red. The exact boundary of the proposed pro-domain at the N-terminus remains unknown. Purple tubes indicate a possible helix and yellow arrows indicate potential strands.

Since there is no information available regarding the M2AP structure, nor the homologous protein, strategies employed to design constructs were dependent on the primary sequence alignment and secondary structure prediction.

Based on these results, the β-sheet rich domain was predicted to end at residue 230 from the M2AP precursor. Three possible starting points of the domain were proposed based on the prediction and the molecular properties of residues around the predicted site. Thus five constructs M2APF (23-330), M2APN (23-228), 38M2APN (38-228), 47M2APN (47-228) and 53M2APN (53-228) were made in order to define the putative the β-sheet rich domain (see table 4.2).

Construct Name	Starting residue	Ending residue	Number of Residues
M2APF	23	330	308
M2APN	23	228	206
47M2APN	47	228	182
38M2APN	38	228	190
53M2APN	53	228	175

Table 4.2 Constructs for M2AP β-sheet rich domain with starting and ending residue numbers

4.4 Constructs Expression, Purification and One Dimensional NMR

The PCR and cloning of the constructs were successful achieved using the methods described in section 3.5 and 3.6 and verified by DNA sequencing.

4.4.1 Initial construct - M2APF

Since no information on M2APF structure exists, an initial full length construct was tested first in order to assess the folding of the recombinant protein. Using the vector system and refined *E. coli* strain, the yield of M2APF (23 -330) was approximately 15 – 20 mg per litre in unlabeled Lysogeny Broth (LB) medium. The recombinant protein expressed as a thioredoxin fusion was found to be very soluble, however the protein degraded very rapidly at room temperature. In an attempt to confirm the expected identity of the protein, a test trypsin digestion of the protein for mass spectrometry was undertaken. However, meaningful data were not obtained (data not shown).

A fresh sample of the recombinant protein was then digested by the enzyme Factor Xa to cleave its 6xHis tag and thioredoxin fusion partner from the N-terminus. The digestion was successful and cleaved protein remained soluble and stable in solution

(figure 4.4). The identity of the sample was confirmed using mass spectrometry (M. Panico, Imperial College London).

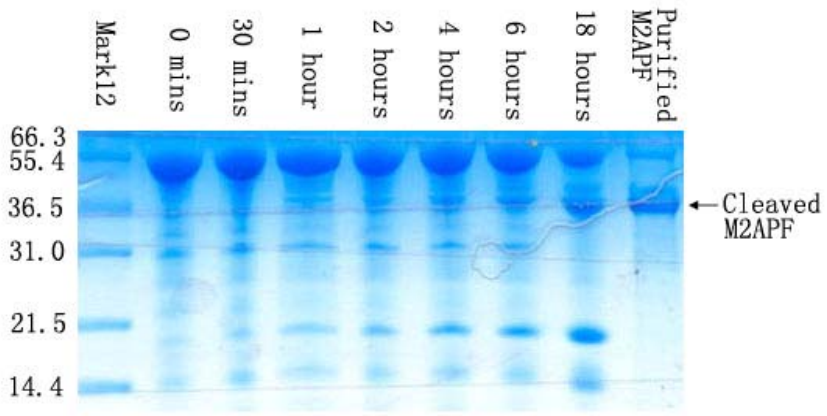


Figure 4.4 SDS-PAGE (Denaturing, Reducing) of M2AP fusion protein cleaved by factor Xa. The fusion partner which possesses a 6xHis tag was removed by passing the cleaved mixture back across the Ni-NTA column. The molecular marker Mark12 is in the first lane. The time on each lane indicates the digestion time after factor Xa added.

The structure of this construct was assessed by 1D (^{1}H) NMR. Appearance of peaks below 0.5 ppm indicates that a folded domain is likely present in the protein. Methyl groups resonate in this region have a characteristic chemical shift about 0.9 ppm. In a folded polypeptide these resonances are shifted due to electronic shielding or deshielding by their environment, in particular due to local aromatic rings and any peak found significantly lower than 0.9 ppm is indicative of a highly structured protein. At 0.8 – 1ppm region, a very sharp and intense peak can be observed, which

indicates the presence of the methyl groups belonging to unstructured or flexible parts of the protein (See figure 4.5). However, a number of peaks in the ring-current shifted methyl region and amide region greater than 9 ppm indicated the presence of a folded domain. Also, peaks around 5 to 6 ppm are indicative of Hα resonances in β sheets, an apparent side-chain Tryptophan NH is also observable around 10.2 ppm.

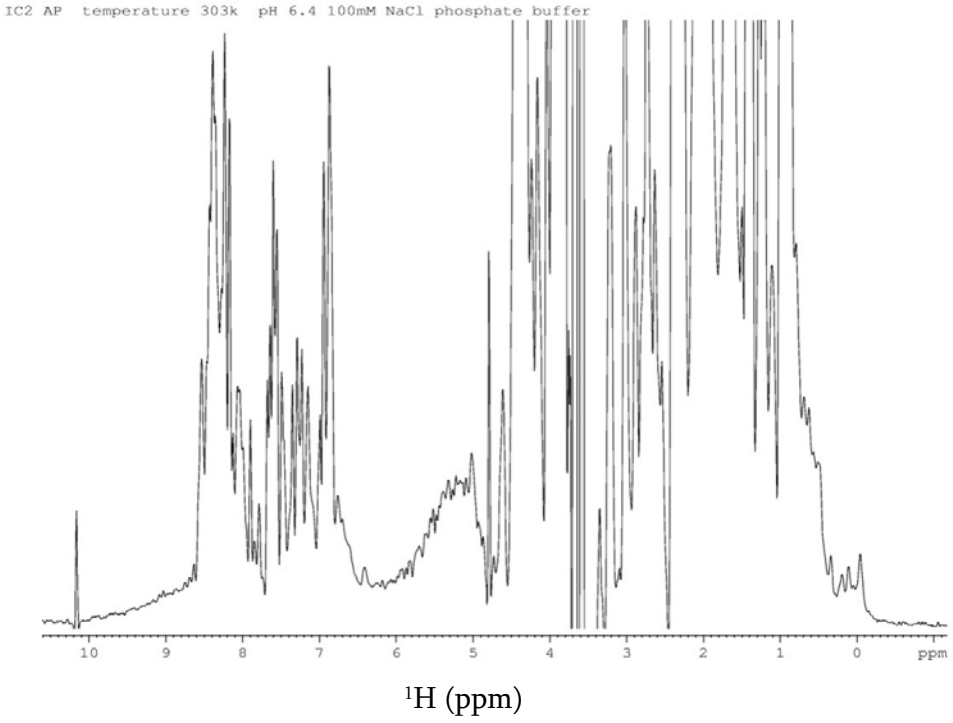


Figure 4.5 1D (^{1}H) spectrum of folded M2APF. High field shifted methyl peaks due to packing with the hydrophobic core are visible in the 0 ppm region. (Temperature 303 K, pH 6.4, phosphate buffer)

4.4.2 Second construct - M2APN

Compared with the full length M2AP construct, M2APN excludes the predicted random coil region, which spans from residue 229 until the C terminus. Using the same protocol as the M2AP construct, this recombinant protein was soluble and did

not degrade for at least three days at 4 °C. The cleavage of its fusion partner was successful and the cleaved protein remained soluble in solution (See figure 4.6).

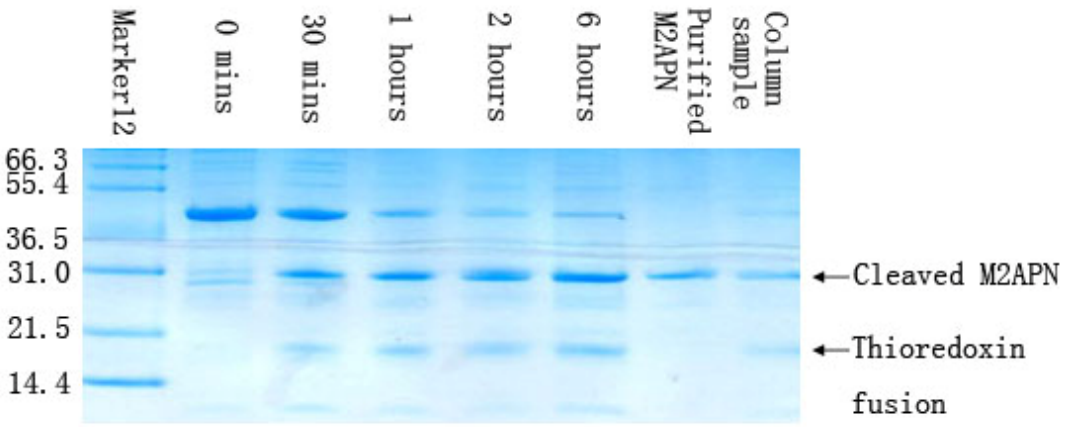


Figure 4.6 SDS-PAGE (Denaturing, Reducing) of M2APN fusion protein under cleavage by factor Xa. The time on each lane indicates the digestion time after factor Xa was added. Column sample was the protein samples which bound to the resin column after second purification.

The folding of this construct was again assessed by 1D (^{1}H) NMR (figure 4.7). The structured region has not been disrupted, but is visibly sharper due to the smaller molecular weight, which means the omitted regions are very likely be to unstructured. From the secondary structure prediction and the chemical characteristics of the amino acids, it was predicted that the C terminal boundary was around residue 228, ending before two prolines. The NMR spectrum is consistent with this prediction, displaying an improved methyl region with few unstructured

peaks around 0.9 ppm and clearly structured amide protons.

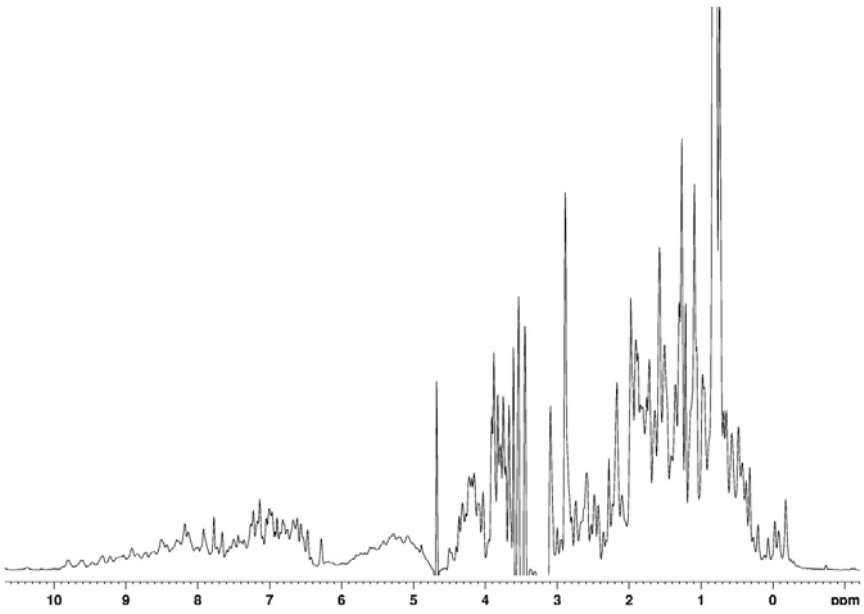


Figure 4.7 1D (^{1}H) spectrum of folded M2APN (23-228). Signals in the 6 to 9 ppm region are greatly improved compared with Figure 4.4. (Temperature 303 K, pH 6.4, phosphate buffer)

4.4.3 Final constructs - 38M2APN, 47M2APN and 53M2APN

After defining the C terminal boundary of β-sheet rich domain, three constructs were designed to locate its N-terminus and their 1D spectra are shown below (See figure 4.8).

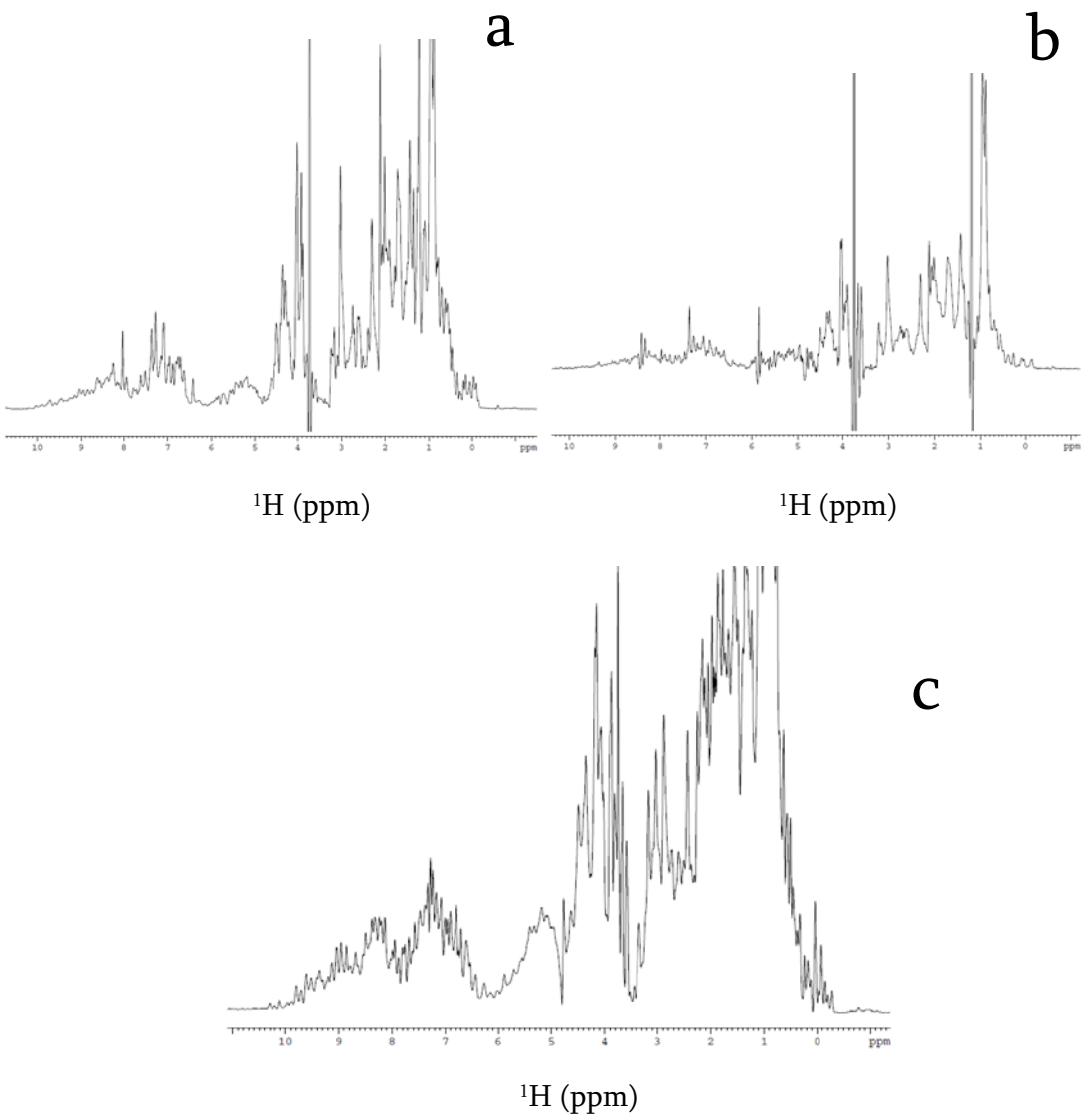


Figure 4.8 1D spectra of M2APN constructs: (a) 1D (^{1}H) spectrum of 38M2APN (38-228), (b) 1D (^{1}H) spectrum of 53M2APN (53-228) and (c) 1D (^{1}H) spectrum of 47M2APN (47-228). (Temperature 303 K, pH 6.4, phosphate buffer)

38M2APN is 11 amino acids shorter at N terminus and the NMR spectrum is unaffected by this deletion (see figure 4.8a). An even shorter construct 47M2APN showed a dramatic improvement, with the methyl region showing improved

linewidths at folded chemical shifts and fewer signals from unstructured polypeptide (see figure 4.8c). Further deletion from N terminus displayed clear structural disturbances, presumably due to destabilization of the protein core (see figure 4.8b). 47M2APN was chosen as the best construct and thus was deemed feasible for further NMR studies.

4.5 2D (1H – 15N) HSQC

A ^{15}N labeled sample of 47m2apn was made for further NMR characterization using the method described in section 3.9.

The spectrum showed a good dispersion of resonances with little overlap and the signals between 8.3-8.5 ppm corresponding to unstructured amides is relatively clear (See figure 4.9). Crowded and overlapped peaks in this region usually mean there are dynamic and unstructured regions present. This spectrum demonstrated that 47M2APN is a suitable construct for further NMR structural analysis.

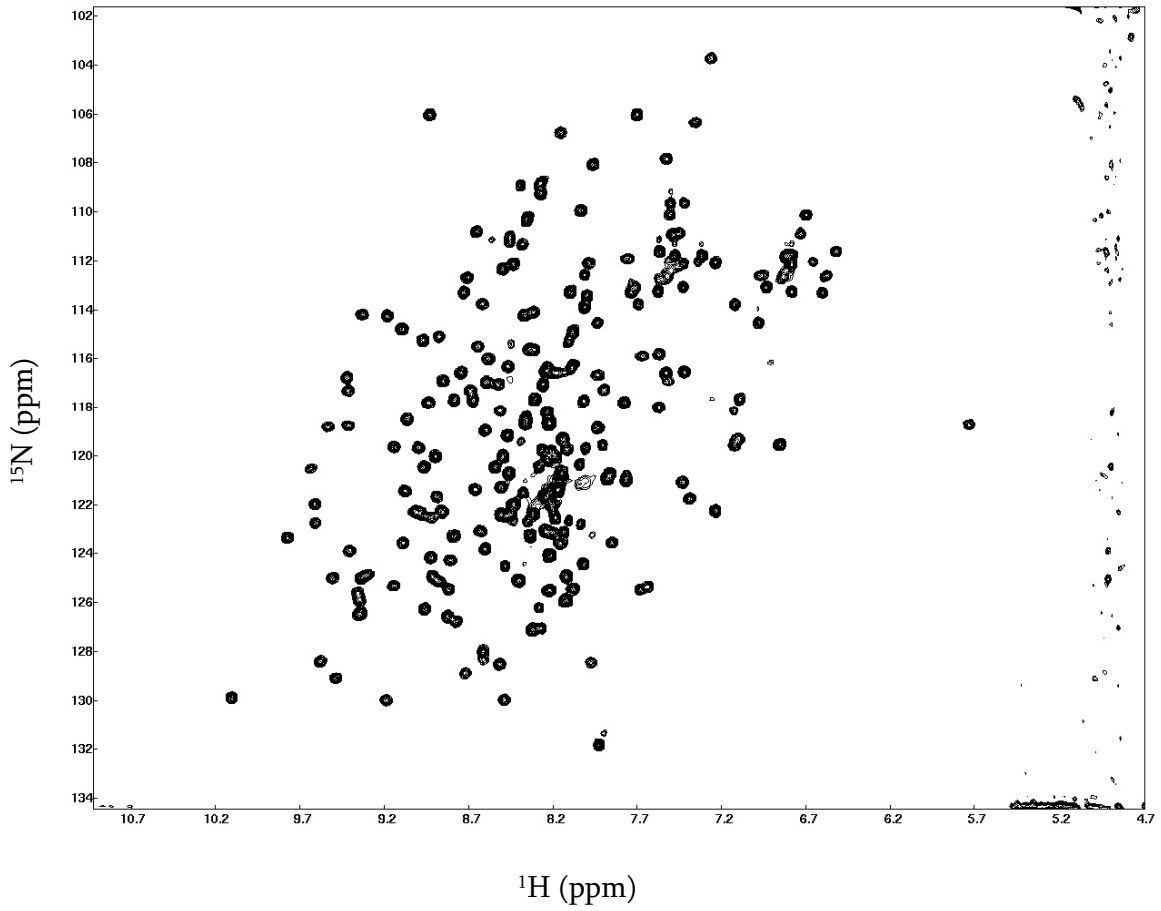


Figure 4.9 2D (^{1}H – ^{15}N) HSQC spectrum of 47M2APN. Buffer conditions: pH 5.9, 100 mM Sodium Phosphate and 50 mM NaCl; Temperature: 308 K.

4.6 3D NMR experiments and Sequential backbone assignments

A ^{15}N, ^{13}C labeled sample of 47M2APN was made for further NMR characterization using the method described in section 3.9. A list of experimental parameters is shown (See table 4.2). However, the sample suffered from aggregation after a long data acquisition (half life around 3 days). A few fresh samples were re-prepared after the old samples proved to be unusable.

Backbone resonances for HN, N, Cα, Cβ and CO were acquired using 2D ($^{1}H/^{15}N$) HSQC and 3D HNCACB, CBCA(CO)NH, HNCO and HN(CA)CO experiments. Hα and Hβ resonances were obtained using an HBHA(CBCACO)NH experiment and side chain assignments acquired using HCCH-TOCSY and (H)CC(CO)NH-TOCSY experiments. Distance restraints were determined by NOESY experiments.

Experiment	Nucleus	Real points	Spectral width
HSQC	^{1}H	1024	6250
	^{15}N	128	1666.67
HNCACB	^{1}H	512	6250
	^{15}N	41	1666.67
	^{13}C	55	8333.33
CBCA(CO)NH	^{1}H	512	6250
	^{15}N	41	1666.67
	^{13}C	24	8333.33
HNCO	^{1}H	512	6250
	^{15}N	40	1666.67
	^{13}C	55	1666.67
HN(CA)CO	^{1}H	1024	6250
	^{15}N	39	1666.67
	^{13}C	55	1666.67
HBHA(CO)NH	^{1}H	512	6250
	^{15}N	41	1666.67
	^{1}H (indirect)	64	3001.2
HC(C)H – TOCSY	^{1}H	512	6250
	^{13}C	32	4401.4
	^{1}H (indirect)	128	4752.85
(H)CCH – TOCSY	^{13}C	128	8810.572
	^{13}C (*J*-coupled)	30	4402.377
	^{1}H	512	6250
NOESY – HSQC (800MHz)	^{1}H	1024	10000
	^{15}N	40	2666.67
	^{1}H (indirect)	138	10000
NOESY – HSQC (800MHz)	^{1}H	1024	10000
	^{13}C	26	14084.507
	^{1}H (indirect)	140	10000
NOESY – HSQC (950MHz)	^{1}H	512	10000
	^{13}C	50	7936.51
	^{1}H (indirect)	109	9090.91

Table 4.2 Experiments and acquisition parameters: A total of 11 experiments were recorded for structure determination of 47M2APN. Acquisition parameters shown include the number of digitised points (used to represent the FID) and spectral width. The mixing time used for the ^{15}N and ^{13}C NOESY-HSQC was 100 ms. The last experiment was recorded at Oxford University (Christina Redfield).

Four 3D-NMR experiments, CBCA(CO)NH, HNCACB, HNCO and HN(CA)CO, were used to sequentially assign the backbone atoms of the protein (See section 2.10). The

HN(CO)CACB and HNCACB are displayed in strips according to the ^{1}H and ^{15}N chemical shifts of the amide resonances in HSQC spectrum (See figure 4.10).

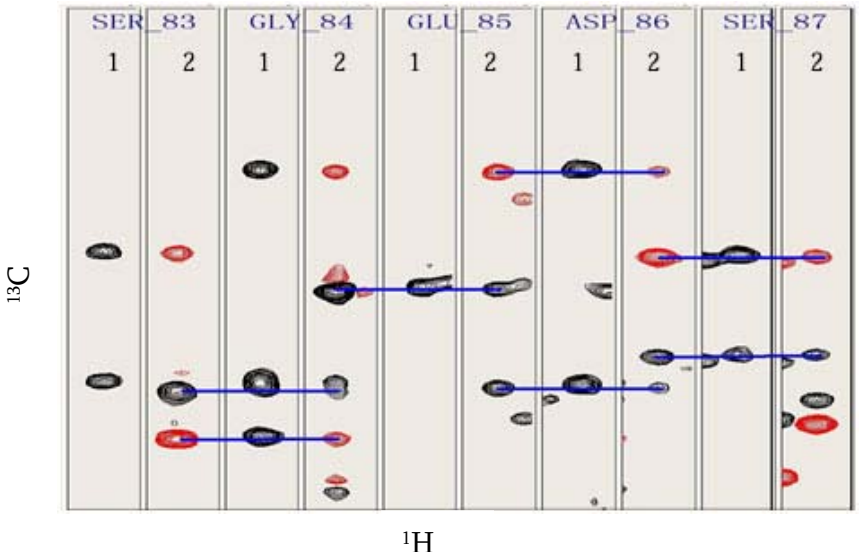


Figure 4.10 Sequential backbone atom assignments: the CBCA(CO)NH strips denoted by '1' and the HNCACB strips denoted by '2'. The linkages between three strips indicate inter and intra residue resonances.

In the HNCACB strips, $^{13}C\alpha$ (black peaks) and $^{13}C\beta$ (red peaks) resonances from both i and i-1 residues are present (peaks from the i residue are generally more intense). In the CBCA(CO)NH strips, only $^{13}C\alpha$ and $^{13}C\beta$ resonances (black spots) from i-1 residues are present. The HN(CA)CO and HNCO experiments provide supplementary information and the ^{13}C chemical shifts for the carbonyl groups (See figure 4.11). All the spectra were processed using NMRview and backbone assignment facilitated by the use of the automated assignment program MARS (around 65% of the residues were assigned by MARS) (Jung and Zweckstetter 2004). The assignments for rest of the residues were achieved manually.

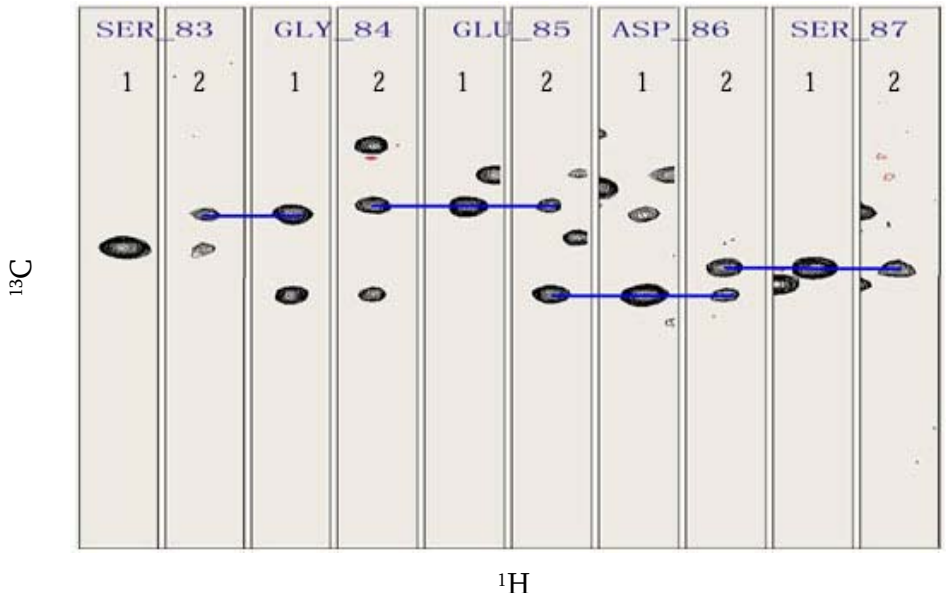


Figure 4.11 Sequential backbone atom assignment: HNCO strips denoted by '1' and HN(CA)CO strips denoted by '2'. The linkages between three strips indicate inter and intra residue resonances.

4.7 Secondary structure prediction and resonance assignments

All spectra were processed by NMRView (Johnson 2004) and backbone assignment facilitated by the use of MARS (Young-Sang and Zweckstetter 2004).

Apart from the first threonine and two histidines, 172 out of 175 possible backbone assignments were made corresponding to 98% of residues (7 prolines excluded) (See figure 4.12).

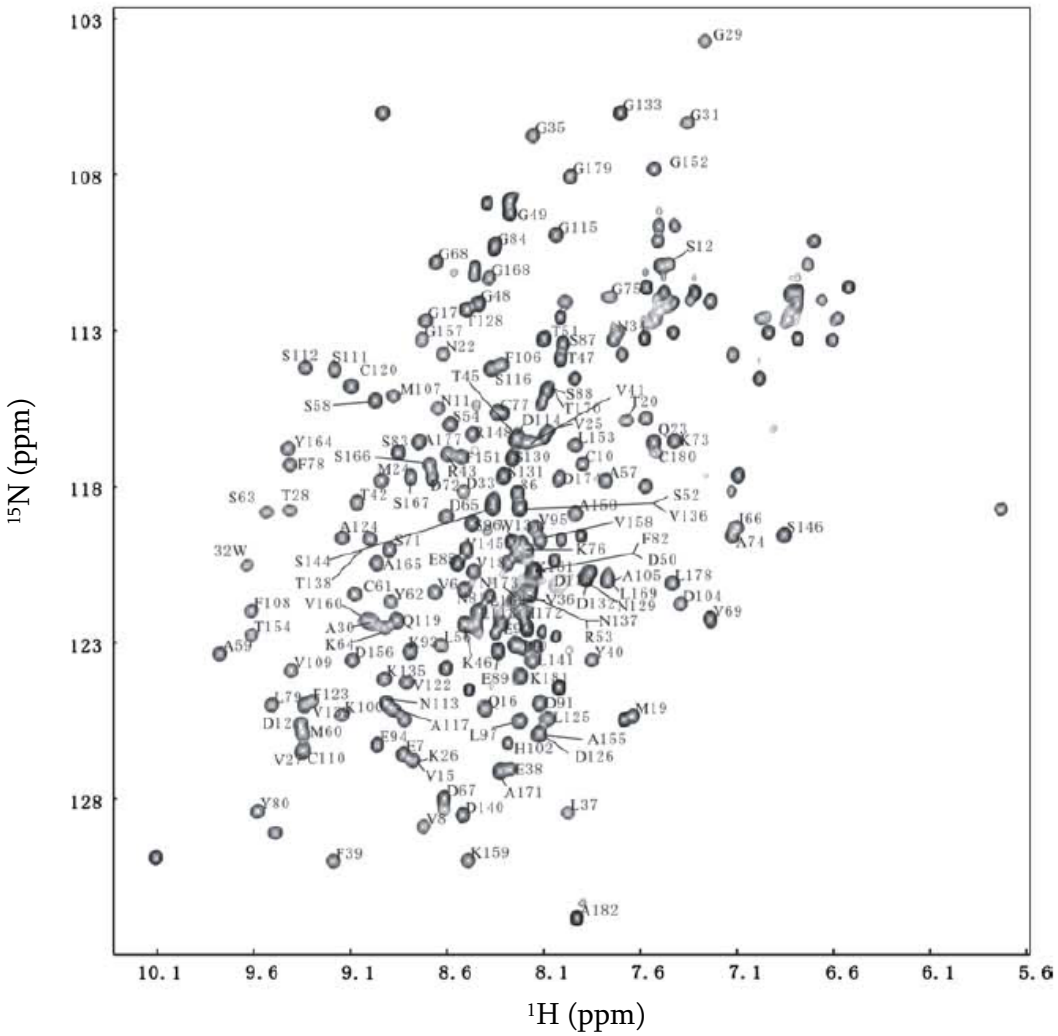


Figure 4.12 2D ($^{1}H/^{15}N$) HSQC spectrum of 47M2APN showing all backbone amide assignments. A total of 172 assignments (out of a possible 175) were made. 7 prolines have no amide signals and are thus not been assigned.

Chemical shift index (CSI), which requires assignment of backbone Hα, Cα, Cβ and CO resonances, is widely accepted to provide accurate prediction of secondary structure (Wishart and Sykes 1994). It is a quick and robust way of deriving secondary structure information from chemical shift data. The chemical shifts of HA, CA, CB and CO can be compared to their random coil values and then assigned with an index of -1 (blue lollipop, beta stand), 0 (grey lollipop, coil) and +1 (red lollipop, alpha

helix). Secondary structure is then predicted based upon the 'clustering' of values in the primary sequence. In the CSI prediction of 47M2APN, 16 β-strands and one three-turn helix were found, indicating a rich β-strand rich structure (See figure 4.13).

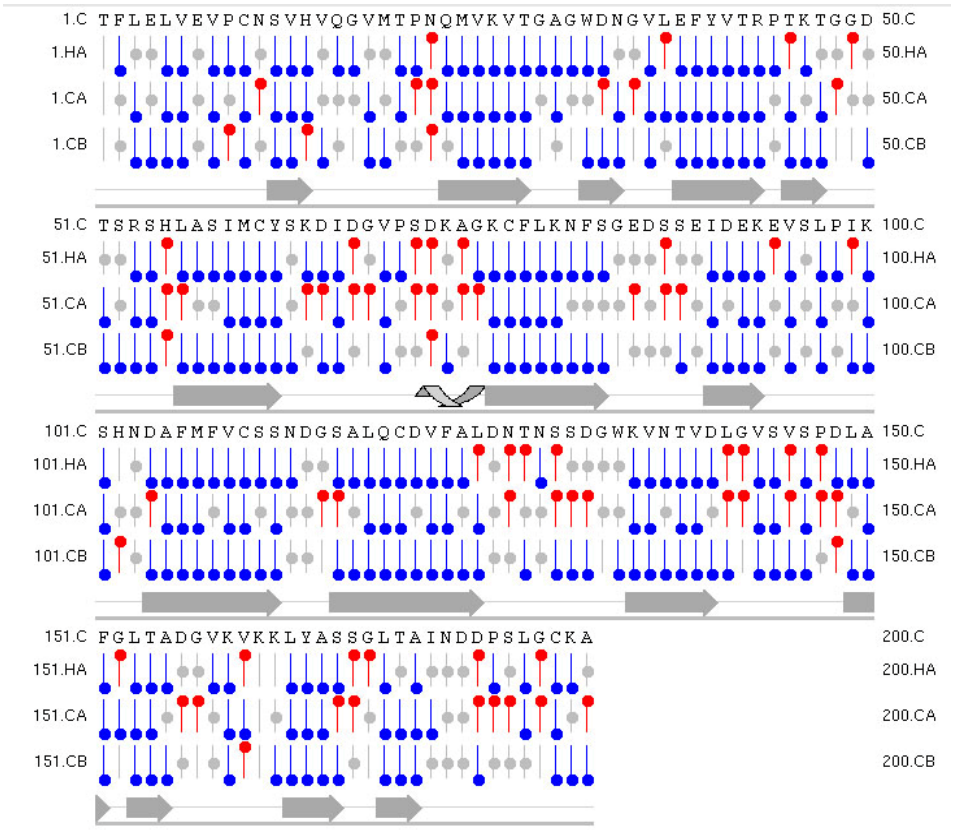


Figure 4.13 Graphical representation of the calculated chemical shift index (CSI) for 47M2APN. The primary sequence is shown along the top, below which is the CSI for each carbon atom or hydrogen per residue: -1 (blue lollipop), 0 (grey lollipop) and +1 (red lollipop). Contiguous stretches correlate to the secondary structure elements (depicted in cartoon format) beneath the CSI prediction.

4.8 Assignments of side-chain chemical shifts

The assignments for the majority of side-chain ^{1}H and ^{13}C resonances were obtained from the combined use of ^{13}C-(H)C(CCO)NH-TOCSY and ^{13}C-HCCH-TOCSY experiments, starting with knowledge of the backbone ^{1}HN, ^{15}N, ^{13}C, ^{13}CO and ^{1}H resonance assignments. HBHA(CBCACO)NH provided the chemical shifts for most Hα and Hβ. Assignments for the aromatic side-chains were based on intraresidue H – H NOEs . For some aliphatic residues, the ^{15}N-NOESY-HSQC and ^{13}C-HSQC-NOESY spectra were required for assignment if poor quality peaks were found in the TOCSY spectra. 93% of the side-chain nuclei could be assigned. An example of how to use NMRview and in-house scripts (Marchant et al. 2008) to assign side-chain chemical shifts when ^{13}C-(H)C(CCO)NH-TOCSY and ^{13}C-HCCH-TOCSY do not work is shown (See figure 4.14). The data was deposited in Biological Magnetic Resonance Bank (BMRB entry 15961).

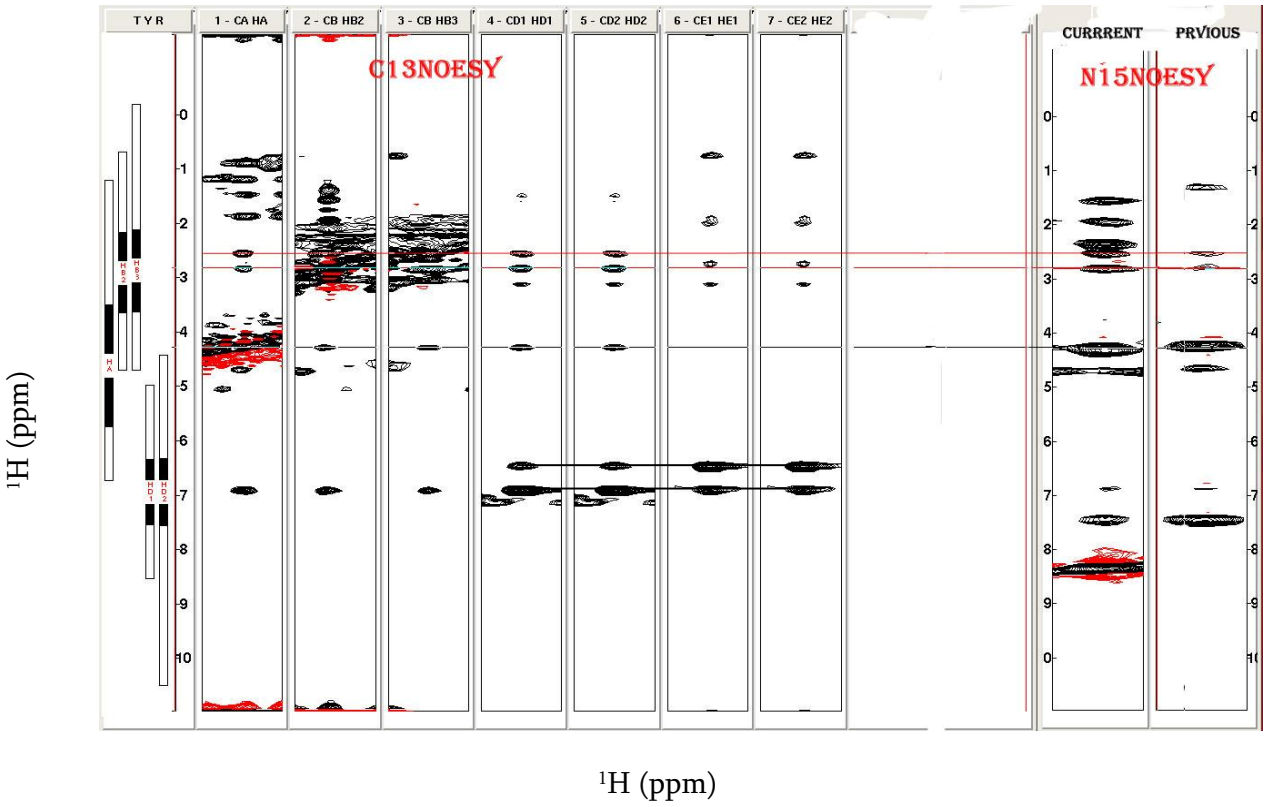


Figure 4.14 Sidechain assignments for Tyrosine 20 spin system using NOESY spectra when ^{13}C-(H)C(CCO)NH-TOCSY and ^{13}C-HCCH-TOCSY do not work. The spectra were displayed using NMRView and in-house scripts, and the lines link the same protons in different spectra. From left to right: 1. the average shifts of relevant atoms; 2. ^{13}C - NOESY planes; 3. ^{15}N - NOESY planes.

4.9 Dihedral angles and TALOS

Dihedral angles are very important restraints for structural calculation of proteins. Backbone dihedral angles are readily obtained from chemical shift information using the TALOS program (Cornilescu et al. 1999). TALOS uses the resonance assignments of atoms to predict the phi and psi dihedral angles within the protein backbone, and hence also predict secondary structure. TALOS searches for the best 10 matches and

provides consistent ϕ and ψ angles (across all sequences) as predictive restraints (in the form of a mean and standard deviation). An updated version, TALOS +, was also used to refine the structure (Shen et al. 2009).

4.10 Residual Dipolar coupling (RDC)

Unlike NOEs (See chapter 4.10), RDCs provide orientation restraints rather than short distance restraints, thus they provide long distance information for the structural calculation. ^{1}H-^{15}N RDCs were measured by alignment in phage. Measurement of D_2O splitting was used to validate the crystalline state of the media. 0.2 mM 47M2APN solution was added to a concentrated solution of the media in stages to dilute the media as much as possible without losing alignment. The final concentration of crystalline media must be low to prevent the protein from interacting with it. It must also be low enough that the protein does not align so much that RDCs other than those between directly bonded atoms can be observed.

Alignment tensor parameters were generated by MODULE (Dosset et al. 2001) based on calculated structures and a recorded RDC data (a set of 134 H-N couplings). These parameters were not included in the initial structure calculation but used to the verify the structure quality.

4.11 Automated NOE assignment and ARIA

NOEs provide distance restraints for atoms that are in proximity to each other through space. A distance can be derived from the assigned NOE data so that three dimensional structures can be calculated in conjunction with other restraints. Unlike J couplings which are observed only when the atoms are bonded to same or

neighboring atoms, all atoms that are in proximity to each other give a NOE (See chapter 2.8).

Assignment of NOEs was made by an automatic approach. Chemical shifts and dihedral angles generated by TALOS+ for 47M2APN were used in ARIA(version 2.2) (Rieping et al. 2007) for structure calculation. Peak lists for ^{15}N and ^{13}C-NOESY spectra were picked prior to the ARIA run and initially contained 2103 and 4302 peaks respectively. Proton tolerances for automated NOE assignment were set to 0.04 ppm and 0.05 ppm (for ^{15}N and ^{13}C-NOESY respectively) and heteronuclear tolerances set to 0.55 ppm (^{15}N), 1.1 ppm (^{13}C – NOESY recorded on 800 MHz) and 0.9 ppm (^{13}C – NOESY recorded on 950 MHz). Three disulphide bridges were assigned from NOESY data between the following cysteine residues 10 -180, 61 -77 and 110 -120.

Other parameters including CNS input parameters were left as default and the calculation implemented as a restrained molecular dynamics (torsion angle) protocol. In each round of NOE assignment performed by ARIA (between iterations 1 and 7) a total of 20 structures were calculated. Six of the lowest energy structures where subsequently used to refine restraints in the following iteration. In the final round of NOE assignment (iteration 8) a total of 40 structures were calculated. The 10 lowest energy structures were subjected to solvent refinement in explicit water (Linge et al. 2003b) and formed the final structure ensemble.

4.12 Solution structure

The 10 lowest energy structures generated after 8 ARIA iterations were water refined in ARIA (see figure 4.15, 4.16 and 4.17). The initial structural calculation did not produce meaningful structure, largely due to poor quality of the initial ^{13}C-NOESY spectrum recorded on 800MHz. The problem was solved by replacing it with a better

resolved ^{13}C-NOESY spectrum recorded on 950MHz. This β-sheet rich structure contains 12 anti-parallel stands and two short helices (see figure 4.16). Further analysis of the structure can be found in chapter 6.

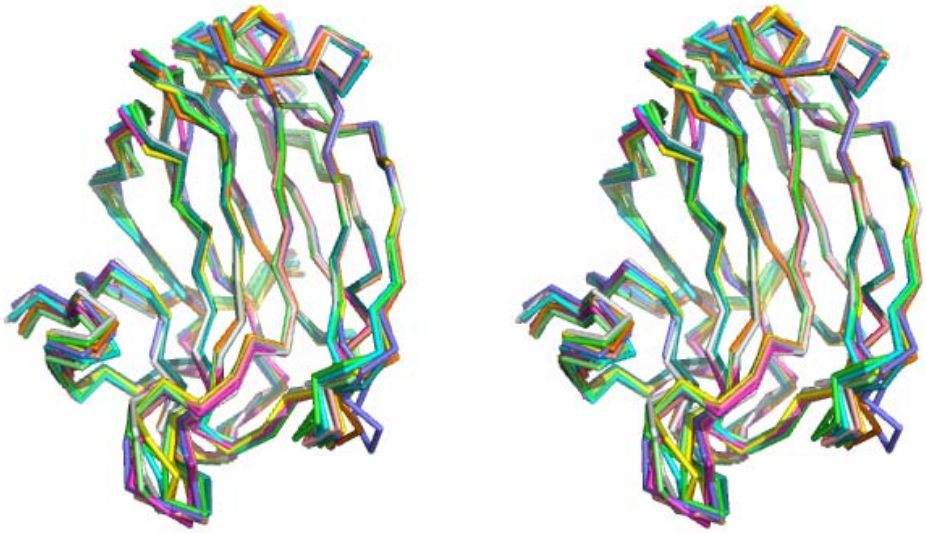


Figure 4.15 – Stereoview of the calculated structure ensemble for M2AP β-sheet rich domain. The structure ensemble is displayed by Pymol (DeLano 2004). It comprises the 10 lowest energy structures and shows a good convergence of the obtained conformers. It is dominated by β-strand secondary structures possessing a Galectin-like fold.

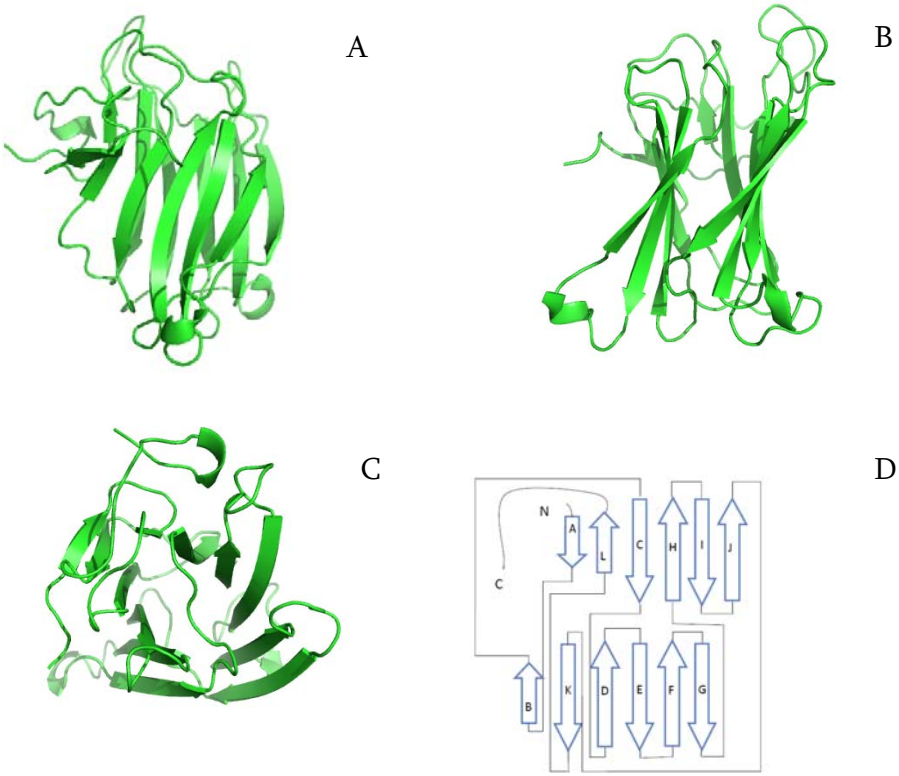


Figure 4.16 Cartoon representation and Topology of 47M2APN solution structure. (A) Front view of the structure, (B) Side view (180° rotation of A), (C) Bottom view (90° rotation of A) and (D) Topology of the structure: 12 β-strands. The images were generated by Pymol.

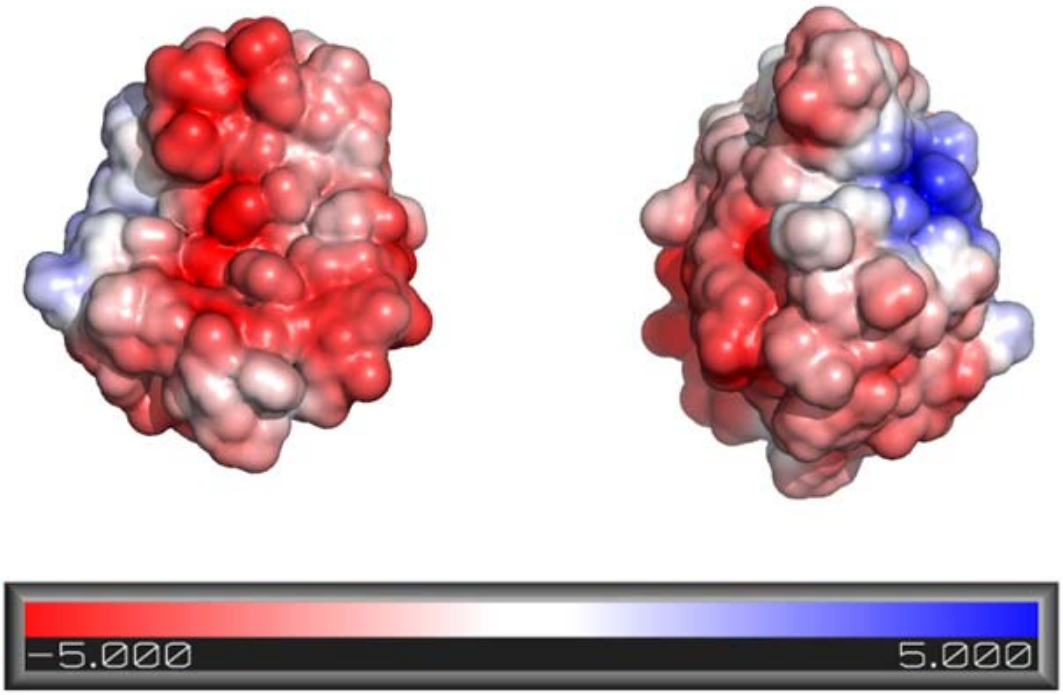


Figure 4.17 Electrostatic surface potential of 47M2AP. The electrostatic surface is calculated by APBS (Baker, 2001), and rendered in blue and red to illustrate electrostatically positive and negative regions, respectively. The second figure depicts a 180° rotation of the surface. The images were generated by Pymol.

4.13 Structure quality

Assessment of structural quality is an important part of the structure determination process. Such validation is required to ensure that the calculated structure represents a viable three-dimensional model. In the case of an ARIA structure, because of the automated nature of the technique, there is an added problem of uncertainty of the NOE assignments. The ensemble was submitted to Protein Structure Validation Suite (PSVS) for quality analysis (Bhattacharya et al. 2007). PSVS integrates multiple known structure quality analysis software including Procheck, Verify3D and ProsaII.

The output includes idealized structural geometry of backbone and side-chain dihedral angles. RMSD values in structural geometry (e.g. superimposition of the calculated ensemble) are presented (See Table 4.3).

The Ramachandran plot is another important indicator of structural quality. The plot shows the combinations of the phi and psi backbone angles in a given protein structure as dots, which are plotted on top of a reference distribution map that contains the regions where the phi/psi combinations are found in good quality reference structures. It shows that 85.4% are in the most favoured regions and only 1.3% adopting a disallowed conformation.

	M2AP
Number of experimental restraints	3937
Total NOE-derived	3659
Ambiguous	1257
Unambiguous	2402
Intraresidue	896
Sequential	528
Medium-range ($\|i-j\| \leq 4$)	146
Long-range ($\|i-j\| > 4$)	832
Talos (ϕ/ψ)	144
RDC	134
RMSD from experimental restraints	
Distance (Å)	0.0415 ± 0.08
Dihedral angle (deg.)	0.20± 0.03
RMSD from idealized covalent geometry	
Bonds (Å)	0.0048 ± 0.0002
Angles (deg.)	0.70 ±0.03
Energies (kcal mol −1)	
Ebond	-63.8 ± 5
Eangle	373.8 ± 30
Evdw	-1059 ±38
Coordinate RMSD (Å)	
Backbone atoms in secondary structure	0.4354 ± 0.06
Heavy atoms in secondary structure[a]	0.82 ± 0.09
Ramachandran plot	
Residues in most favoured regions (%)	85.4%
Residues in allowed regions (%)	12.1%
Residues in disallowed regions (%)	1.3%

Table 4.3 Statistics associated with the experimentally-derived restraints and conformer geometries are shown. All statistics obtained from the appropriate ARIA output files and PSVS outputs.

4.14 Conclusion

The solution structure of M2AP was successfully determined by NMR. This 19.4 kDa protein was verified by multiple validation tools and proved to be a reliable structure for further studies. It provides a new platform that can lead to a better understanding of its structure function relationship, extend the knowledge of this protein family and invasion mechanism of the parasite.

A complete discussion and comparison for this protein can be found in chapter 6.

Chapter 5 Domain scan of MIC2 and structural determination of TSR12 domain

5.1 Introduction

As one of the most promising therapeutic targets, MIC2 drew the most attention among microneme proteins in *T. gondii* (Dautu et al. 2007). It has long been proposed to play the central role in the parasite gliding motility and host cell invasion, however, its precise mechanism of action has not been fully established. Although MIC2 contains several conserved extracellular domains that are involved in receptor binding and an intracellular domain involved in actin-myosin motor, neither the structures nor interaction mechanism of these domains has been established. As a member of the conserved TRAP family, it may serve as a valuable model for understanding the functions of this family. Unfortunately, structural information is really limited within the family, apart from the structure of one single TSR fragment (49 residues) from *Plasmodium falciparum* (Tossavainen et al. 2006). More importantly, its interactions with other parts of the protein and other proteins remain largely unknown.

This chapter describes the structural characterization of MIC2, especially the A domain and TSR domain (See figure 5.1). The determination and structural calculation of TSR_12 domain was the result gained from all the attempts towards a MIC2 structure.

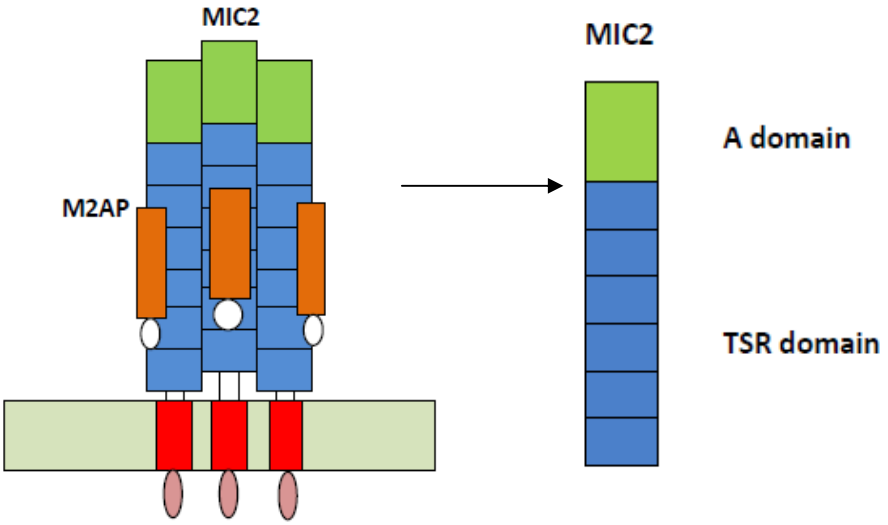


Figure 5.1 Schematic representation of the proposed architecture and interactions of MIC2/M2AP: The extracellular domains of MIC2 participate in the formation of MIC2-M2AP complexes. The extracellular part of MIC2 includes Integrin like A domain (green) and thromspondin-like type 1 repeats (blue).

5.2 Integrin-like A domain (A domain)

As described in section 1.7.1, the precise function and interaction mechanism of the A domain remain is still unknown, even though multiple ligands have been found to bind to the motif (Whittaker and Hynes 2002). Although this highly conserved domain appears in many Apicomplexan members, no structure has been reported.

This section describes the efforts to express soluble, folded recombinant protein for structural analysis and interaction studies. The strategy was to test multiple constructs based on structure alignments and predictions.

5.2.1 Primary sequence alignment, Secondary structure prediction and Construct design

Like M2AP, the MIC2 also contains pro-peptide upstream of the A domain. 70 amino acids in the N terminus are trimmed by a protease called MPP2 (Microneme protein protease 2)(Carruthers et al. 2000). Secondary structure prediction shows good structure conservation with continuous α helices and β strands throughout this domain (See figure 5.2). The constructs design was based on the known pro-peptide and potential disulphide bridges. In total, 12 constructs were made to produce folded recombinant A domain (See Table 5.3).

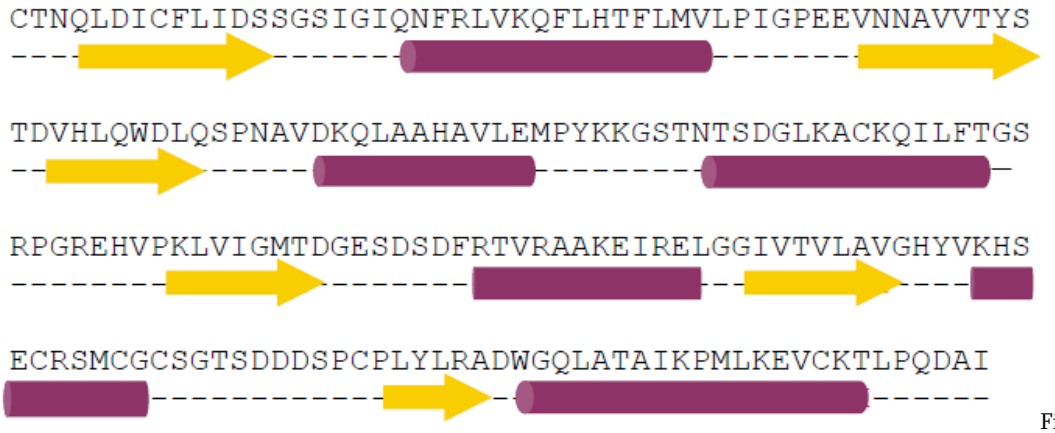


Figure 5.2 JPred secondary structure prediction of MIC2 A domain. Purple tubes indicate possible helices and yellow arrows indicate potential strands.

The strategies to design the constructs were primarily dependent on the primary sequence alignment and secondary structure prediction, especially cysteines that could be involved in disulphide bridge formation. Since the role of the pro-domain in protein folding remains unclear, the first two constructs were designed to include this

domain. However, most of the constructs begin around residue 70 – the putative end point for the pro-domain.

Construct Name	Starting residue	End residue	Number of Residues
MIC2N	27	270	244
MIC2E	27	702	676
67MIC2_4C	67	228	162
67MIC2_5C	67	232	166
67MIC2_6C	67	245	179
67MIC2_7C	67	265	199
67MIC2_8C	67	270	204
73MIC2_4C	73	230	158
73MIC2_5C	73	232	160
73MIC2_6C	73	265	193
27MIC2_7C	27	265	239
MIC2A	73	270	198

Table 5.3 Twelve Constructs tested for A domain. Both starting and end residues from pre-matured MIC2 are listed.

5.2.2 Construct Cloning, Expression and Purification

The PCR and cloning of the constructs were successful using the methods described in section 3.5 and 3.6 and verified by plasmid sequencing. Generally speaking, all A domain constructs had high yield using the protocol described by the previous chapter. An example of a purified recombinant protein (MIC2N) is shown (figure 5.4). The yields for all the constructs were around 15 mg per litre of LB culture.

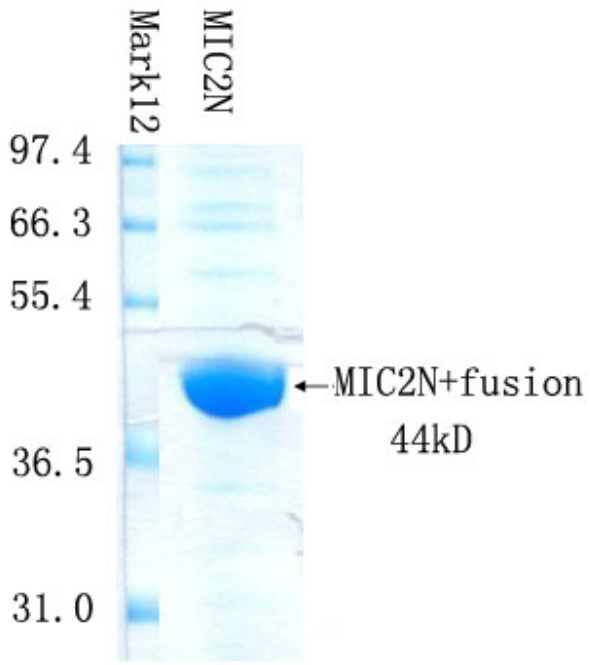


Figure 5.4 SDS-PAGE (Denaturing, Reducing) of MIC2N fusion protein purified under denaturing conditions. Note that the protein includes a 19 kD fusion partner. The lane on the left shows the molecular marker.

Recombinant proteins with thioredoxin fusion exhibited high solubility. The thioredoxin tag was removed by factor Xa (described in chapter 3.6). As the factor Xa cleavage of microneme proteins has been shown in the previous chapter, the cleavage will not be discussed here.

5.2.3 Constructs for A domain were unfolded or partially unfolded

1D (^{1}H) NMR spectra of these recombinant proteins showed that most of the recombinant proteins were unfolded or partially folded. For the construct MIC2N, although a few upfield and downfield proton shifts can be found, the majority of the signals lie in the unfolded regions (See figure 5.5). Line broadening in the aliphatic

region (0.5 ppm – 2 ppm) suggests a degree of self-association/aggregation is present within the sample.

Similar spectra for all the constructs listed were also obtained. Therefore, the recombinant A domain appeared to lack of significant amount of observable NMR peaks indicative of tertiary structure, which is possibly due to misfolding or lack of post translational modification in the bacterial expression system.

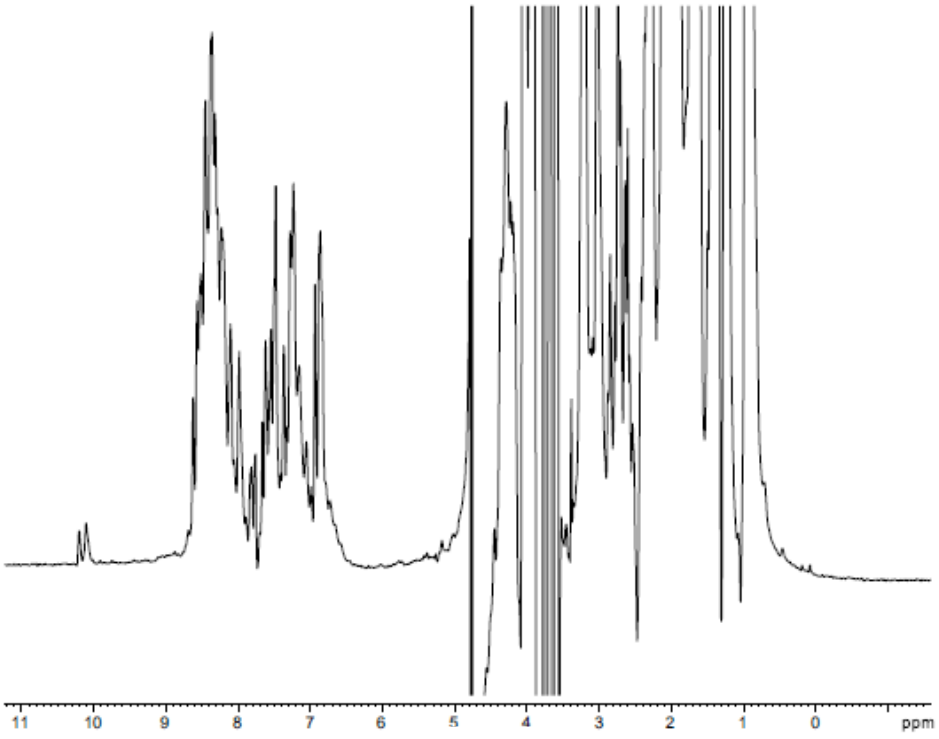


Figure 5.5 1D (^{1}H) NMR spectrum of the recombinant protein MIC2N. Absence of significant signals above 9 ppm (from hydrogen bonding) and below 0.5 ppm (from ring-current effects) is characteristic of an unstructured protein. However, there might be a small folded region within the protein. Spectra recorded at 500 MHz and 303 K.

5.2.4 Summary for A domain

All the constructs for the A domain have exhibited unfolded or partially folded spectra. Purification of the protein from inclusion bodies and refolding using different protocol showed no success for the domain (data not shown). This may imply that this maybe a consensus result under the expression system used. Due to the intrinsic drawback of bacterial expression system, this eukaryotic protein might need a eukaryotic host in order to obtain structurally and functionally characterize this domain.

5.3 TSR-1 domain

MIC2 contains 6 putative TSR type 1 repeats (TSR-1) (See figure 5.1). TSR is a small (around 60 amino acids) domain that is normally found in extracellular proteins or transmembrane proteins that are involved in multiple roles*, e.g.*, immunity, cell adhesion, and neuronal development. The crystal structure from human thrombospondin-1 (TSP-1) has been solved (Tan et al. 2002). The domain has an elongated structure that contains an anti-parallel three-stranded β sheet. The NMR structure of the TSR-1 from *Plasmodium falciparum* indicates that the fold is very similar to TSP-1 and F-spondin TSRs and contains a heparin binding site (Tossavainen et al. 2006). In *T. gondii*, TSR-1 repeats may participate in the interaction with M2AP, the formation of the complex, or interaction with host receptors. Therefore, the structure of TSR from *Toxoplasma gondii* would help establish these interactions, especially after the M2AP structure has been solved (See chapter 4).

5.3.1 Primary sequence alignment and domain annotation of TSR-1

In *T. gondii*, only the first two TSR-1 repeats in MIC2 are well-conserved (See figure 5.6), and they are similar to class E TSR-1 repeats found in TRAP, Spondin and F-spondin.

(a)

TSR repeats	Starting Residue	End Residue	Cysteines
TSR1	271	339	6
TSR2	340	402	6
TSR3	405	469	6
TSR4	470	531	6
TSR5	532	598	4
TSR6	598	648	6

(b)

Figure 5.6 (a) SMART domain detection of the TSR-1 repeats (Letunic et al. 2006). The pink indicates low complexity region. (TSR-1 repeats is also called TSP-1 in mamaliam cell)(b) TSR repeats in TSR domain of MIC2. The boundary and number of cysteines are included.

5.3.2 Construct Design

The strategies to design the constructs were primarily dependent on the conserved TSR domains, especially cysteines that could be involved in disulphide bridge

formation. The constructs include single, double, triple and full TSR repeats (see table 5.7). As individual TSR might have different roles in the complex and interact with one other, constructs containing single and multiple repeats were all tested.

Construct Name	Starting residue	Ending residue	Total amino acids
MIC2T	271	702	430
482MIC2	482	702	220
TSR1	271	339	69
TSR2	340	402	63
TSR3	405	469	65
TSR4	470	531	62
TSR5	532	598	67
TSR6	598	648	51
TSR12	271	402	132
TSR23	340	469	130
TSR34	405	531	127
TSR45	470	598	129
TSR56	532	648	117
TSR	271	648	378

Table 5.7 Fourteen Constructs tested for TSR-1 domain. Starting and ending residues from pre-matured MIC2 are listed.

5.3.4 Constructs Cloning, Expression and Purification

The PCR and cloning of the constructs were successful using the methods described in section 3.5 and 3.6 and verified by plasmid sequencing. Using the vector system and *E. coli* strain described in chapter 3, the yields of single TSR repeats were approximately 15 – 20 mg per litre and multiple repeats were around 10 -15 mg per litre in unlabeled Lysogeny Broth medium. The recombinant protein with a thioredoxin fusion was found to be very soluble. An example of a purified recombinant protein (TSR12) is shown (See figure 5.8) and the yields for all the constructs were 15 mg per litre LB culture.

After removal of the thioredoxin tag by factor Xa (as described in chapter 3.6), the protein remained soluble.

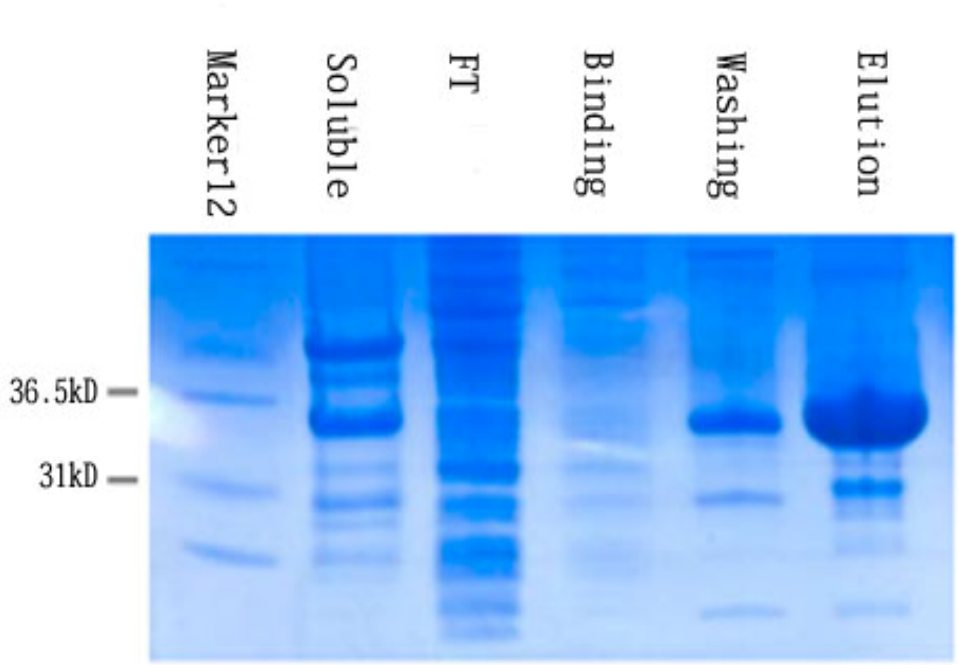


Figure 5.8 SDS-PAGE (Denaturing, Reducing) of TSR12 fusion protein purified under native condition. Note that the protein includes a 19 kD fusion partner. (Soluble = soluble protein, FT = 1st column flow-through, Binding = binding buffer wash, Washing = washing buffer wash and Elution = elution buffer wash).

5.3.5 Unfolded, partially folded and folded contructs found in recombinant TSR proteins

None of the TSR recombinant proteins containing single or triple copies of TSR motifs were found to be folded by 1D (^{1}H) NMR. During the cleavage by factor Xa, these proteins were very unstable and some of the thioredoxin fusion protein proved

difficult to be removed completely from the NMR sample, even after size exclusive chromatography. Apart from the peaks from unfolded protein, the spectra show high similarity to the 1D (^{1}H) of thioredoxin alone, indicating that the thioredoxin was the only folded species in the NMR sample (See figure 5.9).

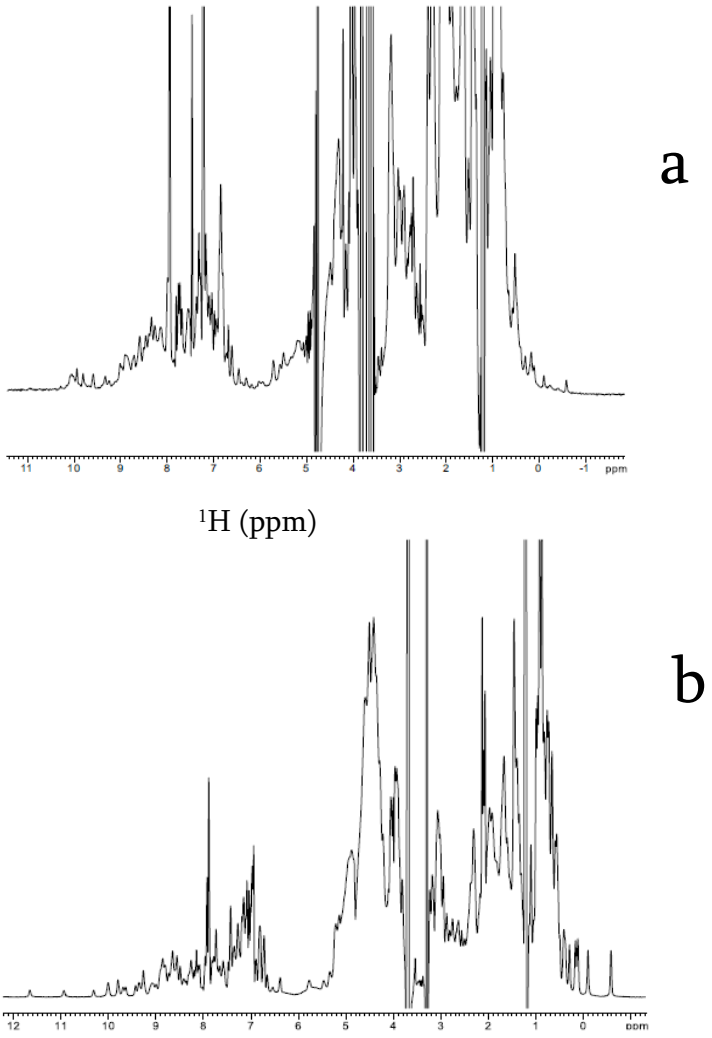


Figure 5.9 (a) 1D (^{1}H) NMR spectrum of the recombinant protein TSR1 (b) 1D (^{1}H) NMR spectrum of thioredoxin. Spectra recorded at 500 MHz and 303 K.

For recombinant proteins with multiple TSR repeats (TSR12, TSR23, TSR34, TSR45 and TSR56), most were found to be unfolded or partially folded except for TSR12, which appeared to be fully folded (See figure 5.10). 1D (^{1}H) spectrum of the isolated TSR12 motif exhibited excellent spectral properties, including good signal-to-noise ratio, and sharp line-widths with minimal peak overlap, therefore TSR12 was a potential candidate for structure determination.

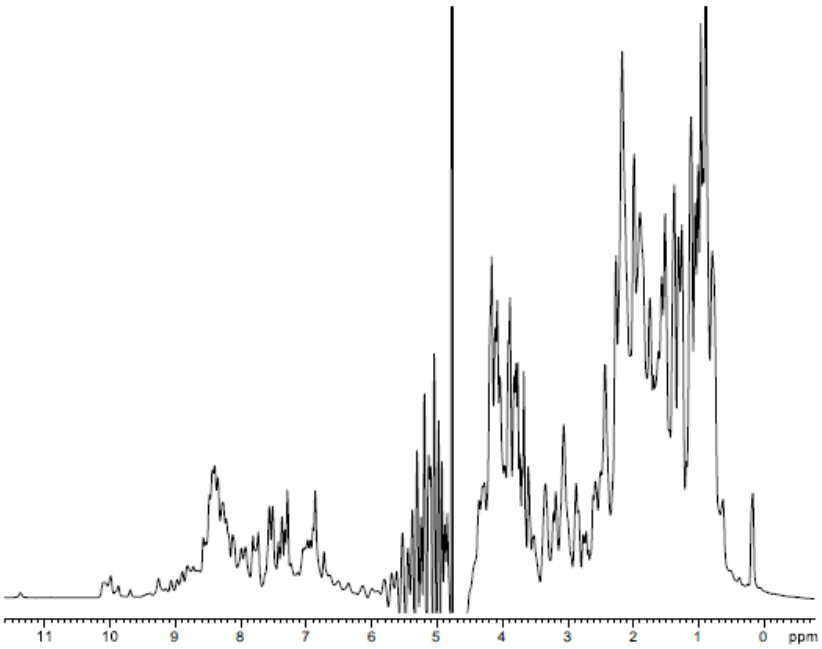


Figure 5.10 1D (^{1}H) NMR spectrum of the recombinant protein TSR12. Spectrum recorded at 500 MHz and 303 K.

5.4 HSQC spectrum of TSR12

A ^{15}N labeled sample of TSR12 was expressed in minimal medium for further NMR characterization using the method described in section 3.9.

($^{1}H/^{15}N$) HSQC spectra of the isolated TSR12 motif exhibited promising spectral properties (See figure 5.11). Although some peak overlap was present in the centre of the spectrum, this might be a result of a small flexible linkage existing in the middle of the two TSR domains and at the flexible N terminus of the protein (incomplete cleavage by factor Xa).

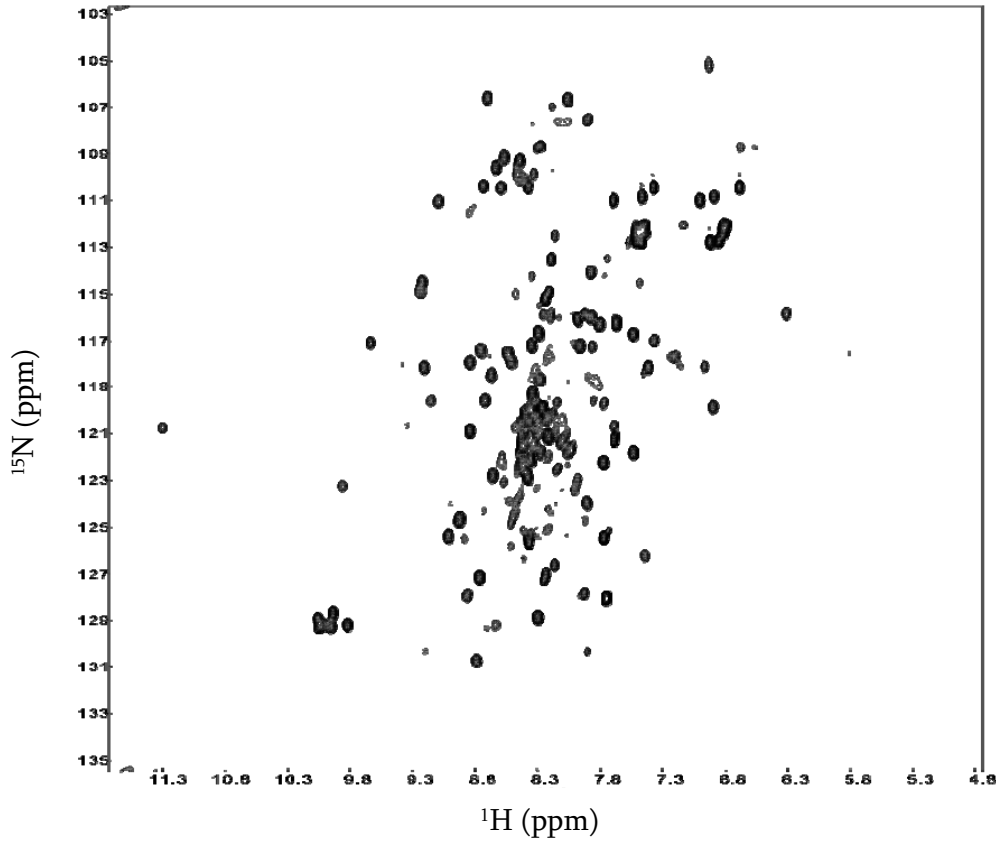


Figure 5.11 2D (^{1}H – ^{15}N) HSQC spectrum of the recombinant protein TSR12. Spectrum recorded at 500 MHz and 303 K.

5.5 Structure determination of TSR12 by NMR

A ^{15}N, ^{13}C labeled sample of TSR12 was successfully expressed in minimal medium for further NMR characterization using the method described in section 3.9.

5.5.1 NMR experiments and parameters

Backbone resonances for HN, N, Cα, Cβ and CO were acquired by CBCA(CO)NH, HNCACB, HNCO, and HN(CA)CO. Hα and Hβ resonances were obtained using an HBHA(CBCACO)NH experiment and side-chain assignments acquired using HCCH-

TOCSY and an (H)CC(CO)NH-TOCSY experiments as described in previous chapter. All spectra were referenced in the direct (^{1}H) dimension according to the resonance frequency of the water peak and in the indirect (*i.e.* heteronuclear) dimension using frequency ratios. A list of experimental parameters is shown (See table 5.13)

Experiment	Nucleus	Real points	Spectral width (Hz)
HSQC	^{1}H	1024	6250
	^{15}N	128	1666.67
HNCACB	^{1}H	512	6250
	^{15}N	40	1666.67
	^{13}C	55	8333.33
CBCA(CO)NH	^{1}H	512	6250
	^{15}N	40	1666.67
	^{13}C	48	8333.33
HNCO	^{1}H	1024	10000
	^{15}N	45	2666.67
	^{13}C	45	2665.96
HN(CA)CO	^{1}H	1024	10000
	^{15}N	63	2666.67
	^{13}C	55	2665.96
HBHA(CO)NH	^{1}H	1024	10000
	^{15}N	55	2666.67
	^{1}H (indirect)	64	4803.07
(H)CC(CO)NH – TOCSY	^{1}H	512	6250
	^{15}N	66	1666.67
	^{13}C	40	11947.43
HC(C)H – TOCSY	^{1}H	512	6250
	^{13}C	34	4401.4
	^{1}H (indirect)	128	4752.85
(H)CCH – TOCSY	^{13}C	118	14084.51
	^{13}C (*J*-coupled)	30	7042.25
	^{1}H	1024	10000
NOESY – HSQC	^{1}H	1024	10000
	^{15}N	41	2666.67
	^{1}H (indirect)	142	10000
NOESY – HSQC	^{1}H	1024	6250
	^{13}C	45	8802
	^{1}H (indirect)	140	6250

Table 5.13 Experiments and acquisition parameters: A total of 13 experiments were recorded for structure determination of TSR12. Acquisition parameters shown include the number of digitised points (used to represent the FID) and spectral width. The mixing time used for the ^{15}N and ^{13}C NOESY-HSQC was 100 ms.

In conjunction with the ^{15}N-^{1}H HSQC, these experiments were used to sequentially assign the backbone atoms of the protein (See section 2.10). All the spectra were

processed using NMRview and backbone assignment facilitated by the use of the automated assignment program MARS (Jung and Zweckstetter 2004). How to use these spectra to achieve the backbone assignment has already been discussed by previous chapter on M2AP protein (See chapter 4.5). The method for assigning TSR12 backbone is similar to M2AP, thus will not be discussed here.

5.5.2 Resonance assignment and CSI prediction

As the TSR12 has 13 prolines, a total of 113 out of a possible 119 backbone assignments were successfully made which is corresponding to 85 % of the residues (See figure 5.14). All corresponding amino acid side-chain assignments were considered complete (including Hδ and Hε for both tyrosine residues) and all cysteine Cβ resonances were characteristic of an oxidised form.

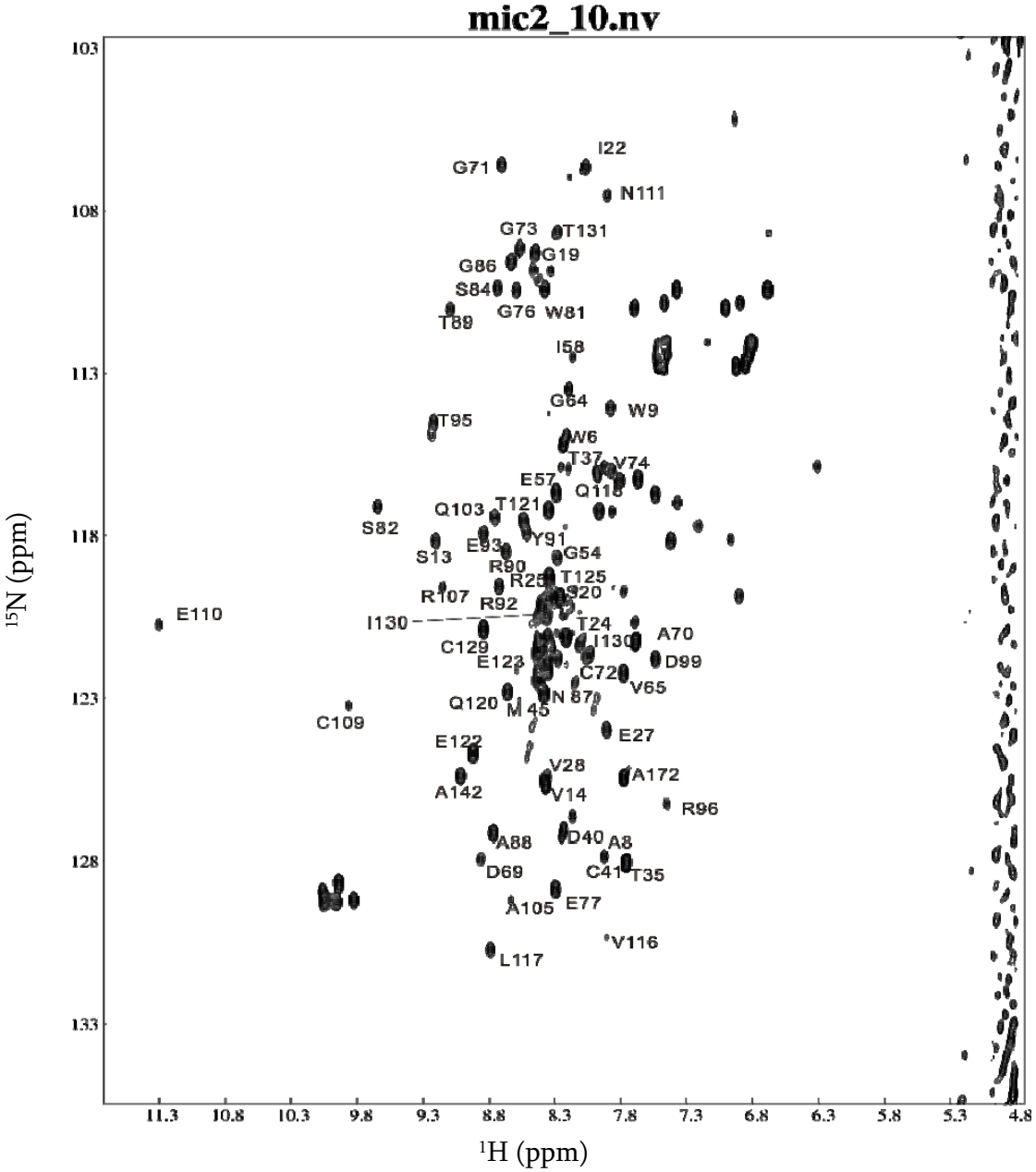


Figure 5.14 2D ($^{1}H/^{15}N$) HSQC spectrum of TSR12 showing all backbone amide assignments. A total of 113 assignments (out of a possible 119) were made. 13 prolines have no amide signal and are thus not been assigned.

Particularly interesting resonances included the amide proton from glutamic acid 110, which was observed at 11.3 ppm. This was significantly different from the average chemical shift (8.34 ± 0.6 ppm) deposited into the Biological Magnetic Resonance Data Bank (BMRB) (Markley et al. 2008) and therefore such a resonance may indicate deshielding by aromatic ring.

The chemical shift index prediction shows that N terminal domains lack significant secondary structure (see figure 5.15). This may be due to too many prolines in the N terminus resulting in many gaps in the assignments. It is also possible that there may be little regular secondary structure.

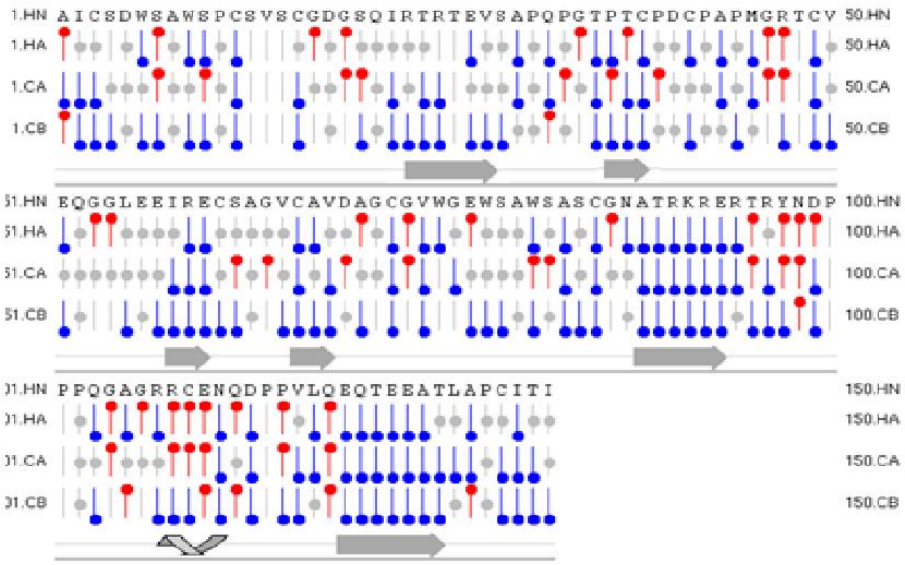


Figure 5.15 Graphical representation of the calculated chemical shift index of TSR12. The primary sequence is shown along the top, below which is the CSI for each carbon atom per residue: -1 (blue lollipop), 0 (grey lollipop) and +1 (red lollipop) (see main text for details). Contiguous stretches correlate to the secondary structure elements (depicted in cartoon format) beneath the CSI prediction.

5.6 Assignment of side-chain chemical shifts

The assignments for the majority of side-chain ^{1}H and ^{13}C resonances were obtained from the combined use of ^{13}C-(H)C(CCO)NH-TOCSY and ^{13}C-HCCH-TOCSY experiments, starting with knowledge of the backbone ^{1}HN, ^{15}N, ^{13}C, ^{13}CO and ^{1}H resonance assignments. However, NOESY spectra were using to assign the side-chain resonances when the two TOCSY experiments had insufficient data (see figure 5.16). Using the similar approach describe in chapter 4, 80% of the side-chain nuclei could be assigned.

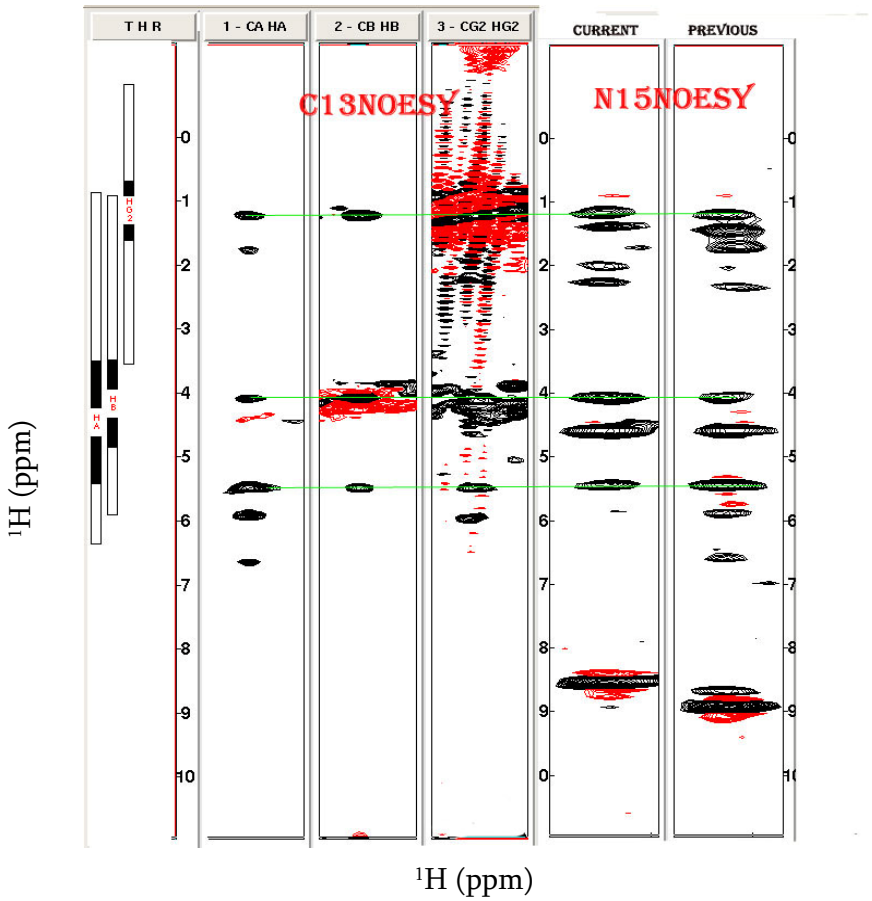


Figure 5.16 Side-chain assignments for Threonine 125 using NOESY spectra when ^{13}C-(H)C(CCO)NH-TOCSY and ^{13}C-HCCH-TOCSY did not work. The spectra were displayed using NMRView and in-house scripts. From left to right: 1. the average shifts of relevant atoms; 2. ^{13}C - NOESY planes; 3 ^{15}N - NOESY planes.

5.7 Dihedral angle restraints: TALOS

Backbone dihedral angles of TSR12 were readily obtained from chemical shift information using the TALOS+ program.

5.8 NOE and Structure calculation

Assignment of NOEs was using ARIA2.2 in an automatic approach (similar to the previous chapter). Chemical shifts, dihedral angles generated by TALOS+ for TSR12 was inputted into ARIA(version 2.2) (Rieping et al. 2007) for structure calculation. Peak lists for ^{15}N and ^{13}C-NOESY spectra were picked prior to the ARIA run and initially contained 1103 and 2302 peaks respectively. Proton tolerances for automated NOE assignment were set to 0.04 ppm and 0.05 ppm (for ^{15}N and ^{13}C-NOESY respectively) and heteronuclear tolerances set to 0.55 ppm (^{15}N) and 1.1 ppm (^{13}C – NOESY recorded on 800 MHz). Six disulphide bridges were left unassigned in the initial runs.

Other parameters including CNS input parameters were left as default and implemented as a restrained molecular dynamics (torsion angle) protocol.

5.9 Structure calculation

The structure was calculated by ARIA 2.2. Although it is still at preliminary stage, the data shows that the structure resembles the TSR fold of TRAP protein from *Plasmodium falciparum* (See figure 5.17). The initial calculated structure contains three anti-parallel β strands in one putative TSR domain. The cross strand NOEs can be found to support the anti-parallel β strands structure. The domain has an elongated structure stabilized by an array of tryptophan and arginine residues as well as disulfide bonds. The fold is very similar to those of thrombospondin type-1 (TSP-1) and F-spondin TSRs. However, the structure calculation has not been finalized and the quality still needs verification. No further discussion will be included in this thesis.

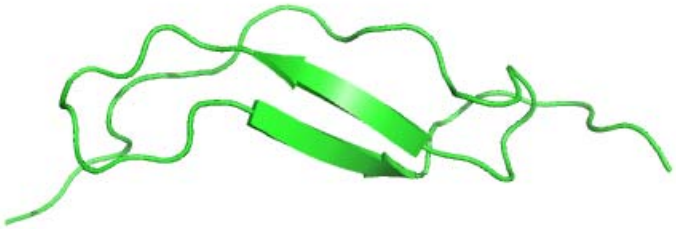


Figure 5.17 TSR structure from *P. Falciparum* (Tossavainen et al. 2006). The structure possesses one thrombospondin-like repeat – a loose packed three beta strands motif. Image generated by Pymol. (PDB ID: 2BBX)

5.10 Summary

Structures of TSR domains from thrombospondin-1 (TSP-1; Tan et al. 2002) and F-spondin (PDB codes 1SZL and 1VEX) have been solved. These show that a TSR domain has an elongated structure consisting of an antiparallel three-stranded β-sheet. The domain core is formed by a stacked array of side chains of conserved tryptophans, arginines, and cysteines. The initial calculated structure of TSR12 is very similar to two solved TSR domains.

Because of time limitations of the project, the structure of TSR12 needs further refinement and quality analysis. However, the structure itself is of significant meaning to the MIC2 (i.*e.* TRAP) architecture and interaction mechanism. It provides a possible interacting partner for TRAP proteins. Further work (*e.g.* structure refinements) is already lined up for the future (see chapter 8.7).

Chapter 6 Structure homology and Interaction Study

6.1 Introduction

After determining the structure of M2AP and obtaining the full backbone assignment of TSR-1 domain (MIC2) from MIC2/M2AP complex, this chapter describes some post-structural analysis and interaction studies. The proteins were compared with known structures in the database. The homologous proteins were found to be both inside and outside the Phylum Apicomplexa. The homology shared by *T. gondii* microneme proteins sheds light on structural and functional conservation in the parasite, while the differences imply the parasite is involved in various mechanisms for invasion.

NMR titration assays were conducted between M2AP and MIC2 domains in order to establish the interaction site. As MIC2 and M2AP have been suggested to form a hexamer, establishment of the interaction sites would provide information to build a model for the complex.

6.2 Carbohydrate binding properties

Carbohydrate-binding properties have been described in microneme proteins in various studies. They typically play an important role in biological recognition phenomena involving cells and proteins. For example, the MAR domain from MIC1 binds sialylated oligosaccharides to permit attachment to the cells of the host organism during infection (Blumenschein et al., 2007). A domain of MIC2 was found to interact with multiple ligands including collagens, laminin and fibronectin (See chapter 1.7.1). Multimerization of MIC2 is required for binding to heparin and human cells (Harper et al., 2004).

As M2AP is a unique protein related to the TRAP family, little information was available at the start of this study. To assess the carbohydrate-binding properties of M2AP, microarray analyses (Campanero-Rhodes et al., 2006) were carried out using the fusion proteins 47m2apn (His-tagged), and lipid-linked oligosaccharide probes (Feizi and Chai, 2004). With 4200 oligosaccharide probes, this microarray encompassed a panel representing diverse mammalian glycan sequences and their analogues, as well as some derived from fungal and bacterial polysaccharides (Palma et al., 2006). Interestingly, M2AP does not show significant affinity to any sugars (Data not shown), implying that M2AP might only be involved in pure protein-protein interaction with MIC2.

MIC2 contains six TSR-1 motifs. These repeats were shown to be involved in the interaction with many ligands including GAGs, CD36, or HIV gp12 (See chapter 1.7.2). Although there is no published evidence, TSR-1 motifs were thought to participate in the interaction with M2AP (Brossier and David Sibley, 2005). No carbohydrate microarray has been carried out so far for TSR-1 motifs.

6.3 M2AP structure homology

Protein BLAST alignment of M2AP shows no putative conserved domains. Only M2AP from *Neospora caninum* shows a high similarity to its counterpart in *Toxoplasma gondii* in terms of primary structure (with a score 206). The sequence alignment algorithm - ClustalW was used to identify the conserved regions (See figure 6.1). The resolved structured motif (highlighted in red) is highly conserved between these two proteins. However, M2AP from *Toxoplasma gondii* has a longer C terminal region indicating undetermined function which might be unique to *T. gondii*.

```
Neospora        MVNLAAVSGAAFAVACFPAVVSARKAVGTAARHTSLVNEPVALAQLSSVLELVTVPCDAV 60
Toxoplasma      -MKLAAVSSAAFAAAFFPAVVNARKVGNPAAQPSVLVNEPVALAQLSTFLELVEVPCNSV 59
                 ::*****.****.* *****.***. ..**: : ***********:.**** ***::*
Neospora        HVQGTMGANDLVKITGAGWGSGALQFQVSHPVKTGGDKSRSQLASVTCYSSDATDAPEGK 120
Toxoplasma      HVQGVMTPNQMVKVTGAGWDNGVLEFYVTRPTKTGGDTSRSHLASIMCYSKDIDGVPSDK 119
                ****.* .*::**:*****..*.*:* *::*.*****.***:***: ***.*  ..*..*
Neospora        AGKCFLTKPAEDGSSDEPQETEVSLPITSHNNAFMFVCSASTAPGVQCDVYAFDNVAN-N 179
Toxoplasma      AGKCFLKNFSGEDSS-EIDEKEVSLPIKSHNDAFMFVCSSNDGSALQCDVFALDNTNSSD 178
                ******.: : :.** * :*.******.***:*******:. ...:****:*:**. . :
Neospora        GWSVKTLDLGLSTSTDLGFSVHAEALNTMKAYASSSLTVARHDISLGCKN---------- 229
Toxoplasma      GWKVNTVDLGVSVSPDLAFGLTADGVKVKKLYASSGLTAINDDPSLGCKAPPHSPPAGEE 238
                **.*:*:***:*.*.**.*.: *:.::. * ****.**. ..* *****
Neospora        ------------TPVEEETT---------------------------------PQGEE 242
Toxoplasma      PSLPSPENSGSATPAEESPSESESVPSLPLSQIPSEVPLSQESPVPPQDGENPDSPQNED 298
                           **.**..:                                    **.*:
Neospora        A------------------------------ 243
Toxoplasma      NSMQEVENSASQSDIEAQRQAENGRLPSDDEA 330
```

Figure 6.1 ClustalW sequence alignment result of M2AP from *Neospora caninum* and *Toxoplasma gondii.* The solved domain is highlighted in red. Propeptide is colored in blue. "*" means identical residues ":" means that conserved substitutions. "." means that semi-conserved substitutions.

As no other functional or sequence annotation for M2AP protein family has been reported, the NMR structure of TgM2AP reveals the first insight into the protein family.

6.3.1 M2AP probably contains the largest galectin-like fold structure

In order to futher analyze the M2AP structure, the structural coordinates of 47m2apn were submited to the Dali Server (Holm and Park, 2000), a network service for comparing protein structures in 3D against those deposited already in the Protein Data Bank. The analysis yielded 520 structures having a Z-score > 2, indicative of structural similarity and 310 structures having Z-score higher than 5, indicative of highly structural similarity. Similarities with a Z-score lower than 2 are spurious (Holm and Sander, 1995). Most of the structures of the highest scores were found to belong to the carbohydrate-binding family of galectins and every structure has a galectin –like fold (See table 6.2).

Description	Z score	RMSD	PDB ID
Galectin-3 (Human)	8.8	3.2	2nmo
Lectin (Mouse)	8.7	3.2	2d6p
Galectin-9 (Human)	8.6	3.2	1a3k
Galectin-1 (Human)	8.5	3.2	2nmn

Table 6.2 Top four structures that have highest similarity with M2AP. All structures possess a galectin-like fold.

More interestingly, no protein possessing only a galectin-like fold has more than 165 residues based on the information by Dali server. With a total number of 182 residues (including 5-7 flexible C termimal residues), M2AP appears to have the largest galectin-like fold published to date. Compared to the largest galectin-like structure found in protein data bank – the C terminal Gal-bind lectin domain 2 from human galectin-4 (PDB 1X50), M2AP has one additional strand (See figure 6.3). Also, some longer strands are found in M2AP, which appear to be important for the interaction with MIC2 (See section 6.5). The three disulfphide bridges in M2AP provide additional support for a larger structure scaffold. In contrast, no disulfphide bridges are present in galectin-4 (multiple disulfphide bridges are not commonly found in galectin-like proteins).

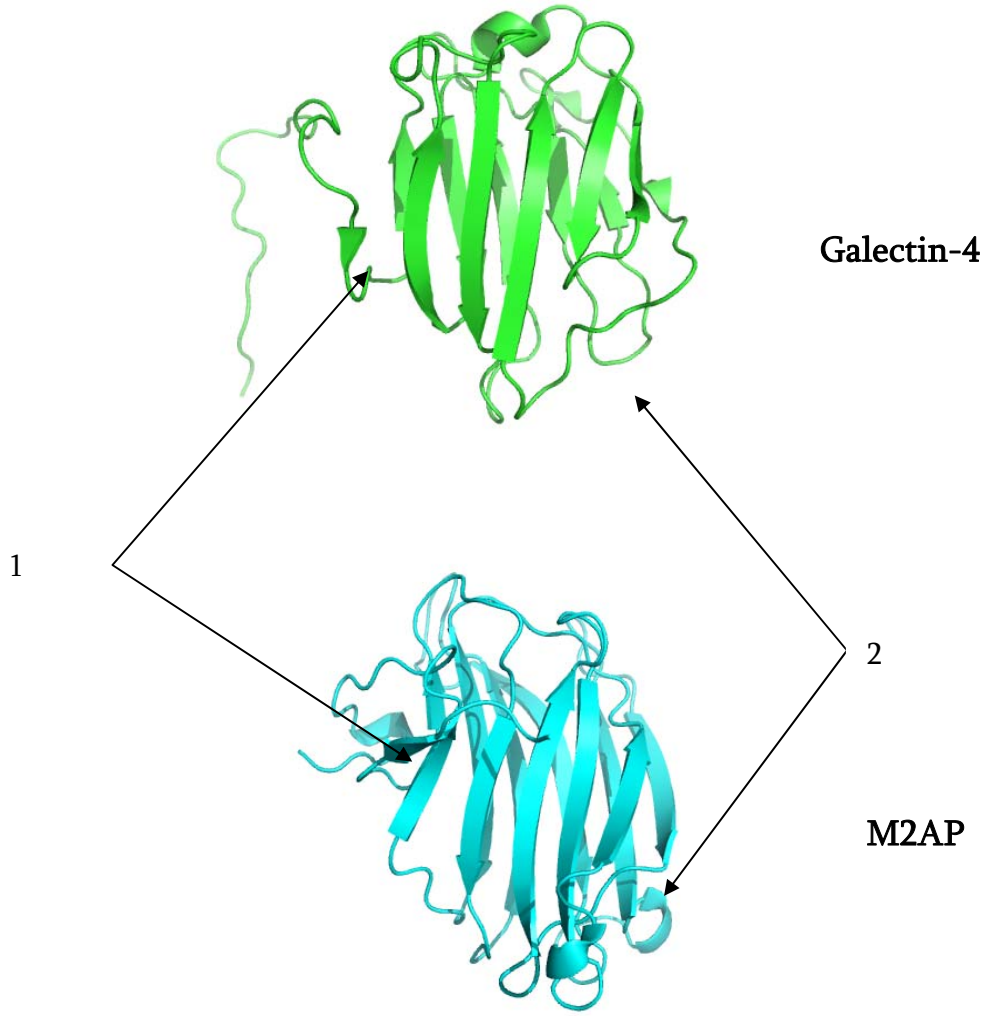


Figure 6.3 M2AP and the C domain of human Galectin-4 are aligned in the same orentation. Blue (bottom) is M2AP and green (top) is human Galectin-4. The first arrows indicate an extra strand and the second arrows indicate longer stands in M2AP. The flexible N terminal of human Galectin is not counted as part of the galectin fold. Figure generated by Pymol.

6.3.2 M2AP does not contain a Carbohydrate Recognition Domain (CRD)

The soluble-type lectins or galectins constitute the galectin family defined by shared consensus amino acid sequence and affinity for beta-galactose-containing

oligosaccharides. This family of proteins is involved in many biological processes including cell-cell and cell-matrix adhesion, growth regulation, signaling, and cytokine secretion (Demetter et al., 2008). In fact, the core of each galectin is dictated by the carbohydrate recognition domain (CRD). The CRD is made up of around 135 aa and provides a scaffold onto which conserved residues are located that mediate carbohydrate interactions.

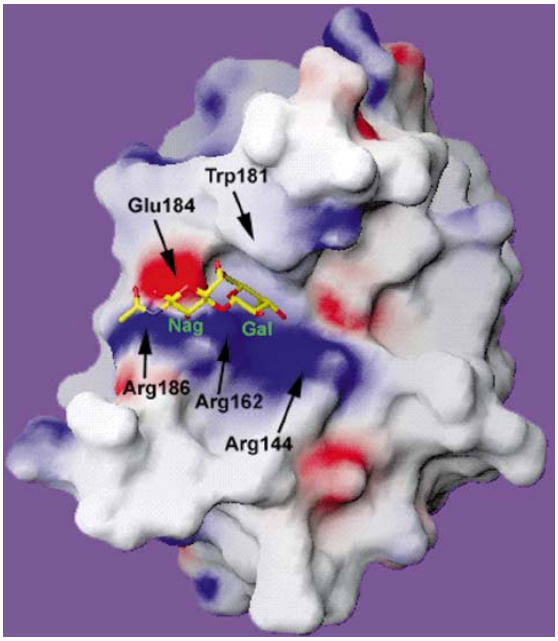
As described in section 6.2, 47m2apn does not show significant affinities for sugars. Thus, it is very likely that the CRD is not present in this protein. To understand the absence of galectin activity of M2AP, Galectin 3 - its closet structural homologue (Z-scoure 8.8) was chosen to make a detailed comparison. Galectin 3 posseses a number of conserved amino acids at its carbohydrate recognition domain (CRD) that mediate interactions with sugars (Collins et al., 2007; Seetharaman et al., 1998). The conserved sugar binding motif (Lac/LacNAc binding site) contains six highly conserved residues including His158, Arg162, Glu165, Asn174, Glu184 and Arg186 (See figure 6.4). A "hydrophilic core" is created in the CRD by these conserved residues. Direct and indirect hydrogen bonds (through three water molecules) mediate the interaction. Van der Waals contacts with the equally conserved Trp181 stabilise the interaction even further.

These positions are not conserved in the equivalent locations in M2AP, except for Glu94 from M2AP which is equivalent to residue 184 of Gal-3 (See figure 6.4). These residues in M2AP do not form any equivalent hydrophilic site like those in other Galectin-like proteins. In contrast, M2AP presents a predominantly hydrophobic environment and negatively charged surface in this region that is more likely to be a

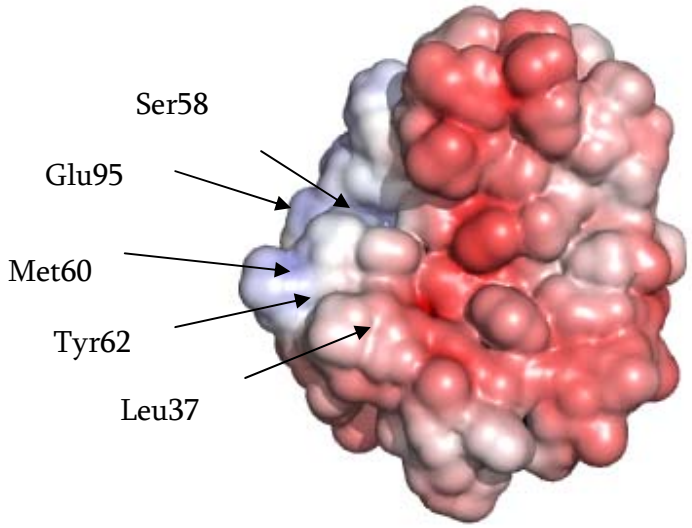
protein-protein interaction interface (see figure 6.5). This suggests that carbohydrate recognition is not retained by M2AP.

```
M2AP          TFLELVEVPCNSVHVQGVMTPNQMVKVTGAGWDNGVLEFYVTRPTKTGGDTSRSHLAS
Galectin-3    ----------YNLPLPG----RMLITILGTVK-ANRIALDFQRG--------NDVAFH

M2AP          IMCYSKDIDGVPSDKAGKCFLKNFSGEDSSEIDEKEVSLPIKSHNDAFMFVCSSNDGS
Galectin-3    FNPRF----NENN---RVIVCNTK---DNNWGREERQVFPFES-GKPFKIQVLVE-PD

M2AP          ALQCDVFALDNTNSSDGWKVNTVDLGVSVSPDLAFGLTADGVKVKKLYASSGLTAIND
Galectin-3    HFKVAVN------------AHLLQYNH--LNEIKLGISGIDLT--SASYTMI------

M2AP          DPSLGCK
Galectin-3    -------
```

Figure 6.4 Structural alignment between M2AP and Galectin-3. The strands in Galectin-3 are separated by "-". The residues of Galectin-3 involved in the carbohydrate binding are colored in red. The residues which occupy the corresponding places in M2AP are shown in blue. Cysteines are coloured in dark blue. Alignment was performed by DaliLite (Holm and Park, 2000).



Galectin-3



M2AP

Figure 6.5 Electrostatic potential of the galectin-3 surface (top) and 47M2APN (bottom) viewed into the carbohydrate binding site. Blue and red indicate positive and negative electrostatic potentials respectively. The bound LacNAc moiety is shown in the stick representation in yellow. The Lac/LacNAc binding site is formed by beta-strands S4–S6a/S6b. The equivalent residues are displayed on the surface of 47M2APN with the similar orentiation as galectin-3 (No similar moiety formed). Upper figure adapted from (Seetharaman et al., 1998). The lower figure was generated by Pymol.

6.4 NMR titration and Chemical shift mapping

The chemical shift change observed upon ligand interaction with a protein can be due to two phenomena: when adding the binding partner, chemical shifts of the residues at the interface are perturbed by the proximity of the partner (change in local chemical environment); or the addition of the partner induces structural changes around the binding site (e.g. side-chain reorientations or even allosteric effects). NMR is very sensitive to changes in the chemical environment. These changes can be conveniently monitored in the backbone amide (NH) chemical shifts of an ^{15}N-labelled protein, usually via the ^{1}H-^{15}N HSQC spectrum, during titrations with unlabelled ligands.

6.4.1 NMR titration

MIC2 and M2AP fulfill an important role in the parasite virulence only when they are in complex with one another. The interaction between the two partners provides molecular basis of the complex. In order to map the interface, two sets of experiments were performed. ^{15}N labelled M2AP and unlabelled TSR12, TSR34, and TSR56 were used for the first titration.

In the first titration, unlabeled TSR12 was gradually added into ^{15}N Labeled M2AP to achieve ratios from 4:1 to 1:4 (See figure 6.6). The interaction only caused small chemical changes and peaks moving could be observed from 2:1ratio (M2AP: TSR12). M2AP was saturated at an equivalent amount of TSR12.

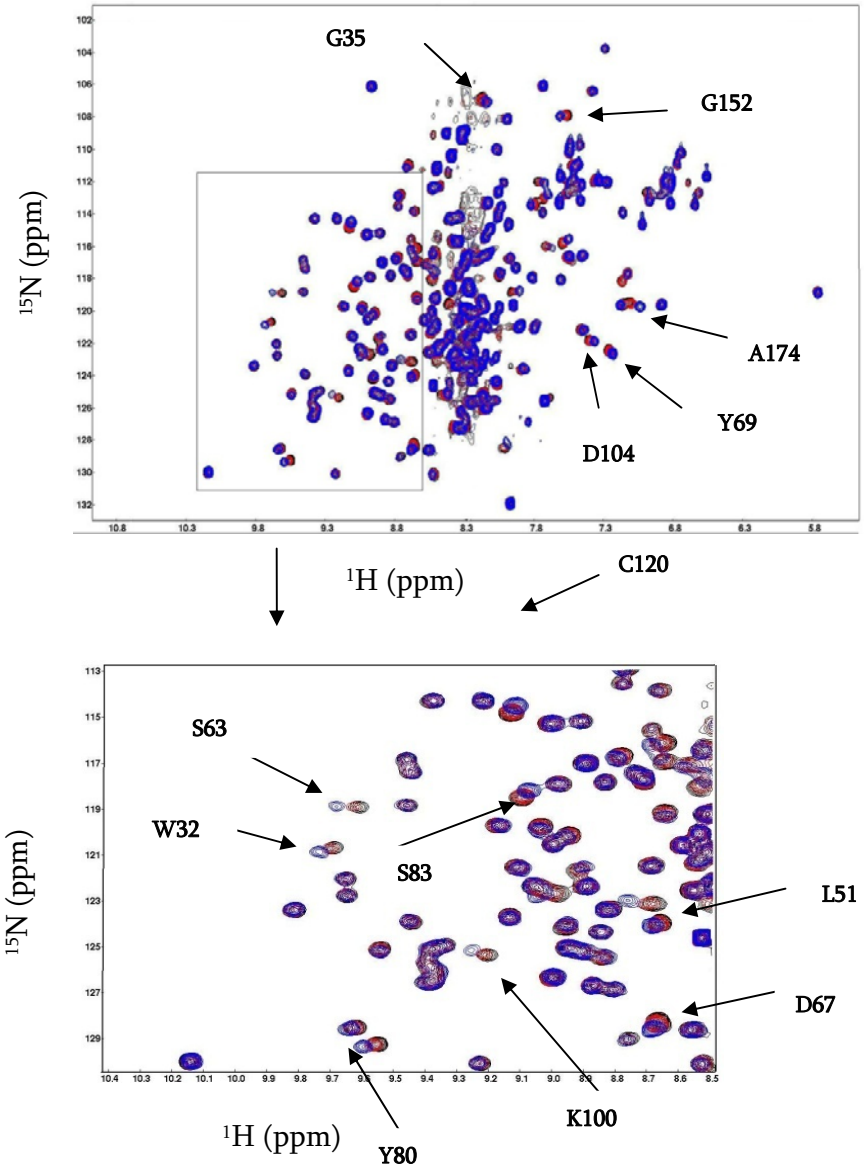


Figure 6.6 (^{1}H, ^{15}N) HSQC superimposed spectrum of ^{15}N-labelled M2AP: Black in the free state; Red in the transit state when bound with TSR12 at M2AP versus TSR12 = 2:1; Indigo in the saturated state when bound with TSR12 at M2AP versus TSR12 = 1: 1. Spectra were collected at 308 K.

The titration between TSR34 and ^{15}N-labelled 4 M2AP showed no significant chemical shift changes. This is probably because TSR34 recombinant protein is not folded or there is no real interaction between these two domains, therefore the results are not shown here.

The titration between ^{15}N-labeled M2AP and unlabeled TSR56 was also performed under the same condition. There was no intermediate state between free form and bound form during the titration (indicating a slow exchange, see figure 6.7). The chemical shift perturbation ended at M2AP: TSR56 = 1:1. The chemical shifts perturbation is very similar to that of the titration (M2AP and TSR12).

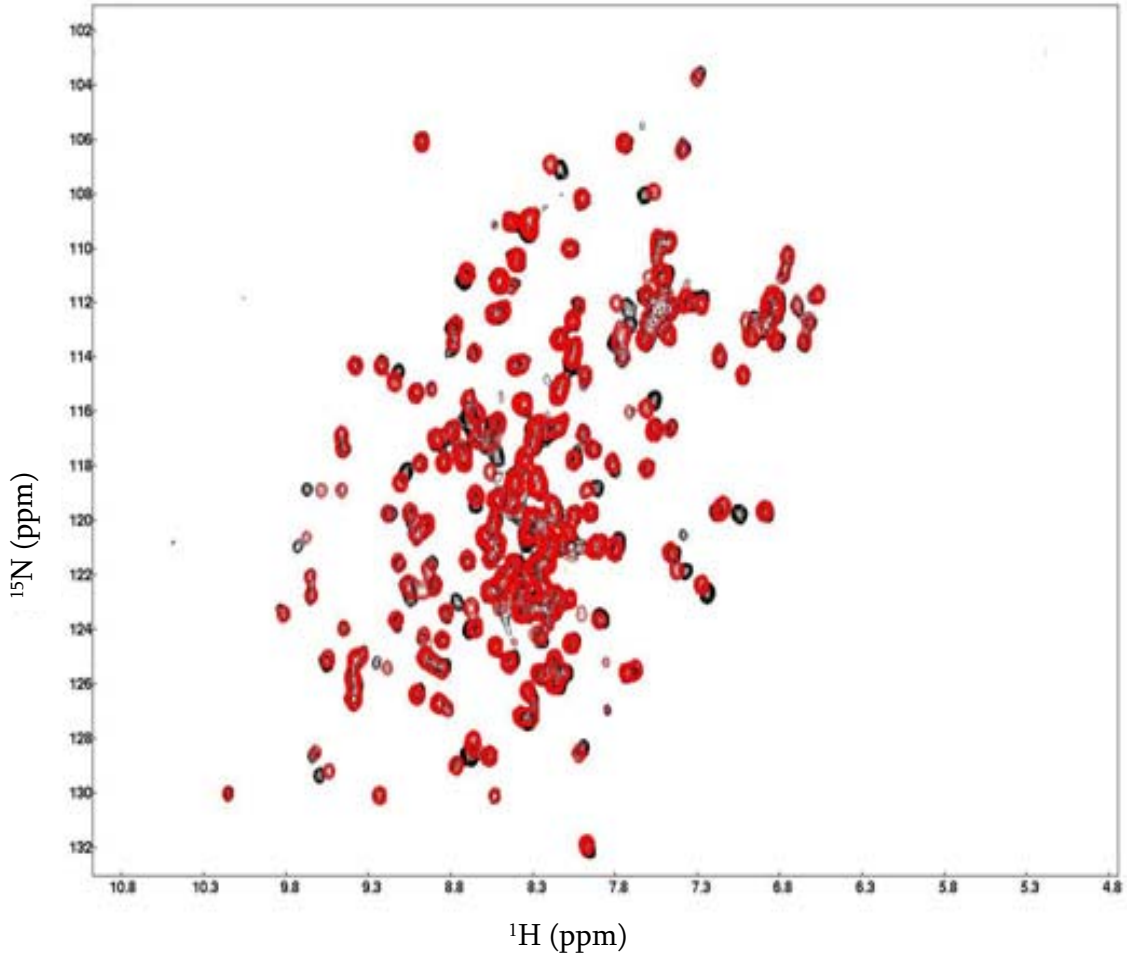


Figure 6.7 Chemical shift perturbation analysis of TSR56 binding to M2AP. An overlay of the ^{1}H–^{15}N HSQC spectra of the M2AP in the absence (black contour levels) or the presence (red contour levels) of TSR56 at molar concentration ratio of 1:1. The chemical shifts perturbation is very similar to that of the titration (M2AP and TSR12) thus specific peaks are not shown in this figure.

6.4.2 Reverse titration

To confirm the interaction between M2AP and TSR12, a reverse titration experiment using ^{15}N labelled TSR12 and unlabelled M2AP was carried out. Similar to ^{15}N labelled

M2AP, TSR12 was saturated when an equivalent amount of M2AP was added (See figure 6.8). The interaction only caused small chemical changes which imply there are no aromatic residues involved.

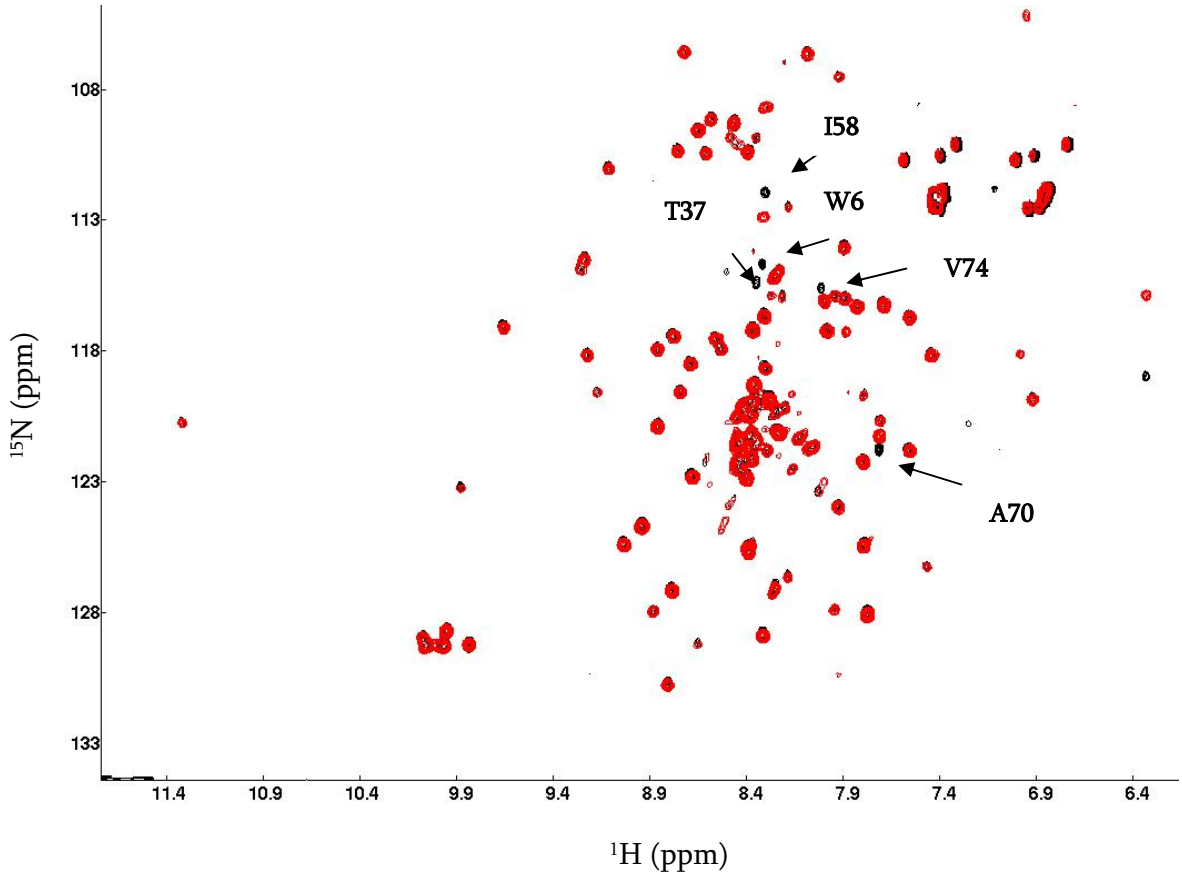


Figure 6.8 ^{1}H, ^{15}N HSQC superimposed spectra of ^{15}N-labeled TSR12: Black in the free state; Red in saturated state at M2AP versus TSR12 = 1:1; Spectra were collected at 308K. Peaks with significant chemical shift differences are indicated.

6.4.3 Chemical shift mapping

^{1}H-^{15}N HSQC NMR titration is a powerful technique to map the interacting site of a protein upon binding to its partner. Chemical shift perturbations can be mapped onto the protein sequence and 3D structure to delineate the residues that constitute the

putative binding site. These data can be used to direct amino acid mutagenesis and other studies can follow to investigate the binding site even further. Thus, a titration would give information about the binding interface and conformation changes.

Apart from some minor differences, chemical shift changes profile are very similar between the M2AP-TSR12 titration and M2AP-TSR56 titration (See figure 6.9). Chemical shift mappings on the 3D structure of M2AP are mainly based on M2AP-TSR12 titration since the weighted chemical shift perturbation ($\Delta\delta^{av}=\Delta\delta N/5+ \Delta\delta H$) differences between the two titrations are minor (See figure 6.10). This weighted chemical shift perturbation method is to average the chemical shift changes between Nitrogen and Proton dimensions (Hajduk et al. 1997). The interface on M2AP is mainly located on one side of the β-sandwich, where residues Thr 42, Ser 58 and Ala 150 form an interface. The loop region, especially Trp 32, Asp 33 and Ile 66 are also main contributors for the interaction.

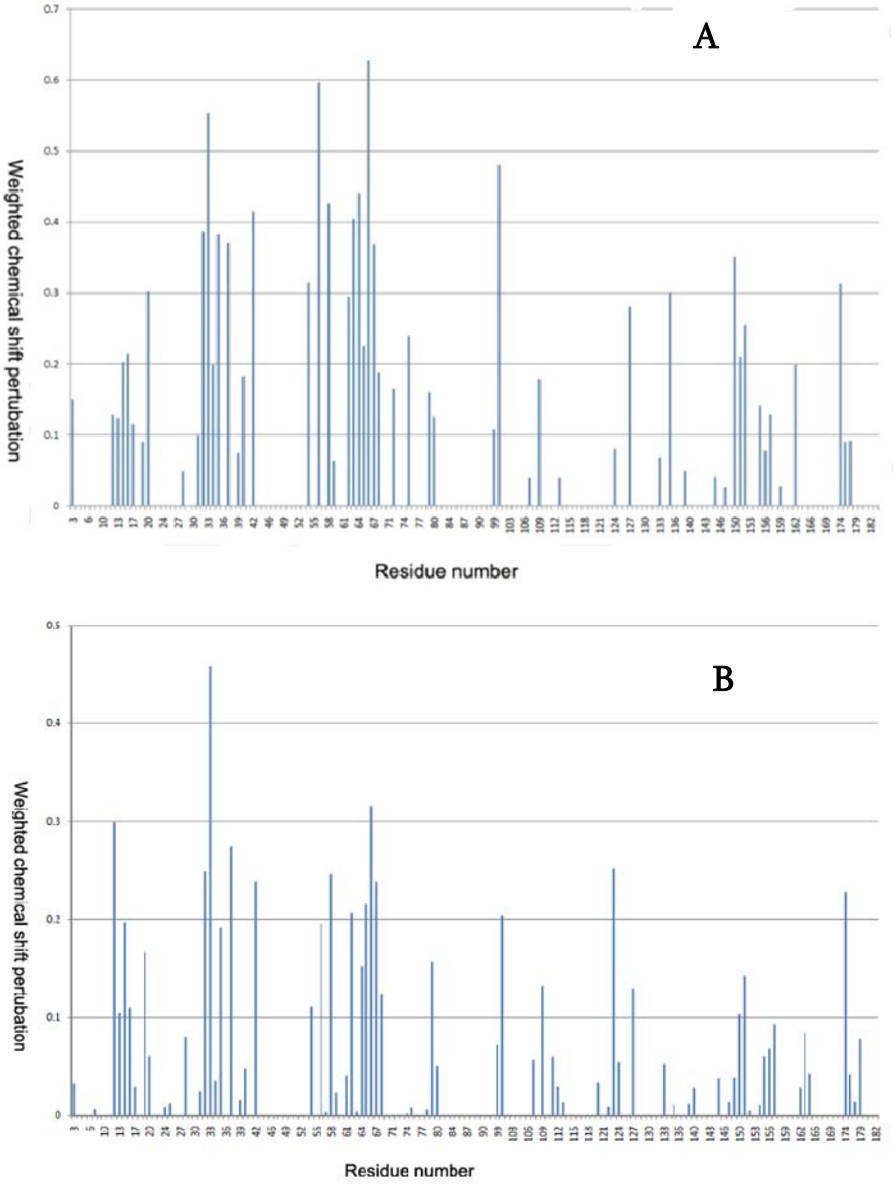


Figure 6.9 Weighted-average chemical shift perturbation ($\Delta\delta^{av}=\Delta\delta N/5+\Delta\delta H$) versus residue number: (A) Weighted chemcial shift pertubation of M2AP due to TSR12 binding; (B) Weighted chemcial shift pertubation of M2AP due to TSR56 binding.

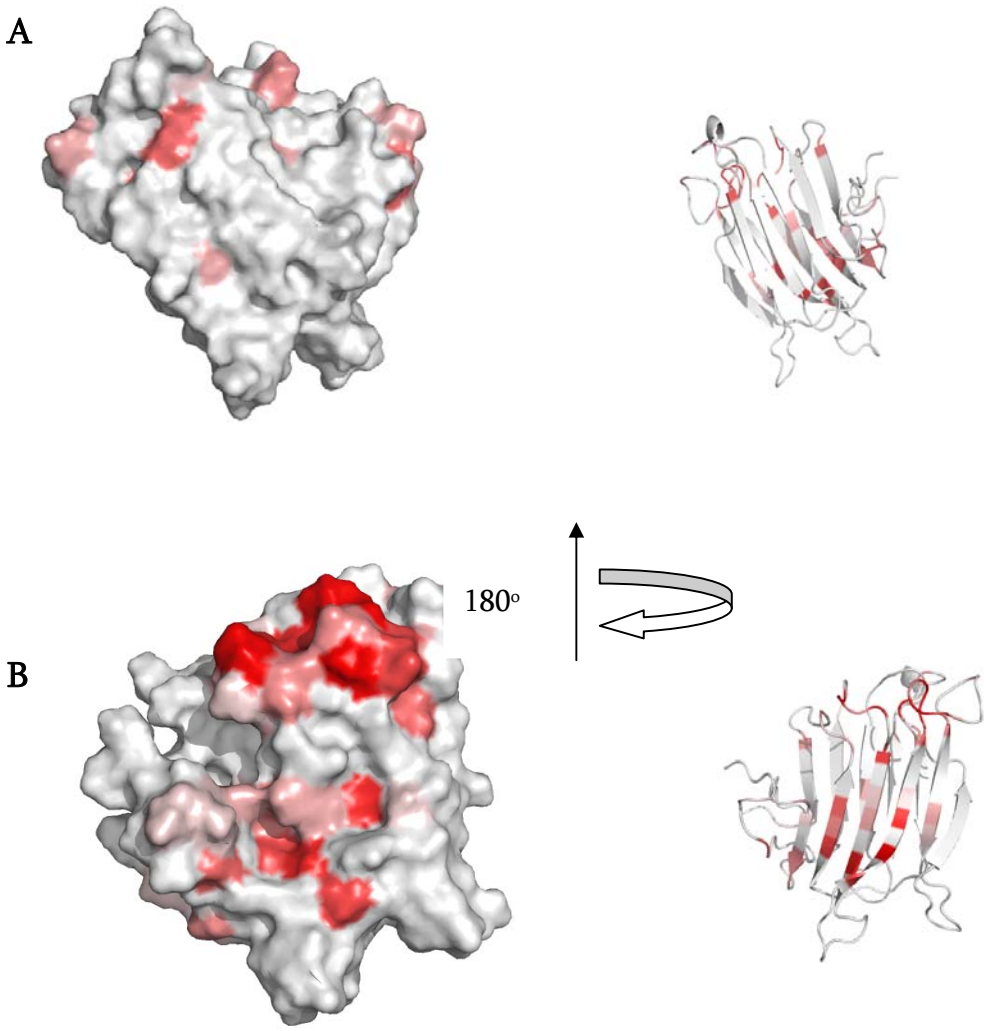


Figure 6.10 Chemical shifts mapped on the surface of M2AP upon TSR12 binds M2AP. Two orientations of the protein are presented (ribbon representation of the protein). Residues with $\Delta\delta > 0.3$ are coloured in red, residues with $0.2 < \Delta\delta < 0.3$ are in light red and residues with $\Delta\delta < 0.2$ are in white. (A) Back view of the binding site on the molecular surface as well as ribbons. (B) Front view of the binding site on the molecular surface as well as ribbons. Figures were generated by Pymol.

6.4.4 M2AP and MIC1-CT

MIC1 carboxyl terminal domain (MIC1-CT) from *Toxoplasma gondii* is another known microneme protein with a galectin-like fold. Furthermore, this 14 kDa domain does not contain a CRD, as in M2AP. As described in Chapter 1.6, MIC1-4-6 form a complex that works synergistically with the MIC2-M2AP complex. In particular, MIC1-CT is involved in interactions with the EGF2 and EGF3 motifs of MIC6 (Sawmynaden, 2008). M2AP and MIC1-CT share a certain degree of similarity both in terms of structure and function (*i.e.* galectin like folds that bind partner proteins in place of a carbohydrate). .

Both M2AP and MIC1-CT contain disulphide bridges, whereas a conserved, reduced cysteine is normally found in other galectin-like fold proteins. Three disulphide bridges in M2AP could provide more constraints to support a larger structure, especially the distortion caused by longer loops. In MIC1-CT, one pair of cysteines is found in a loop between two strands. Considering the fact that microneme proteins are generally cysteine-rich, the conserved disulphide bridge is clearly another feature adopted by the parasite.

Not suprisingly, these proteins use different residues to bind to their partner respectively. Two principal interfacial regions can be delineated in the MIC1-CT and EGF3 interface (See figure 6.11a). The first comprises an exposed hydrophobic surface on one of the large β strands. The second important contact interface is at one edge of the β-sandwich. In the centre of this region are the hydrophobic residues Phe 49, Val 60 and Tyr 99, which form a binding pocket for EGF2. However, the interfacial region in M2AP is mainly located on one side of the β-sandwich (including residues Thr 42, Ser 58 and Ala 150) (See figure 6.11b). In particular, the loops (including Trp

32, Asp 33 and Ile 66) between four main strands participate in the interaction (based on chemical shift mapping in section 6.4).

Together with MIC1-CT, M2AP are novel members of the Galectin-fold family (i.e. lack of carbohydrate-binding property). Instead of traditional carbohydrate-binding site, they contain a surface as the protein-protein interaction interface.

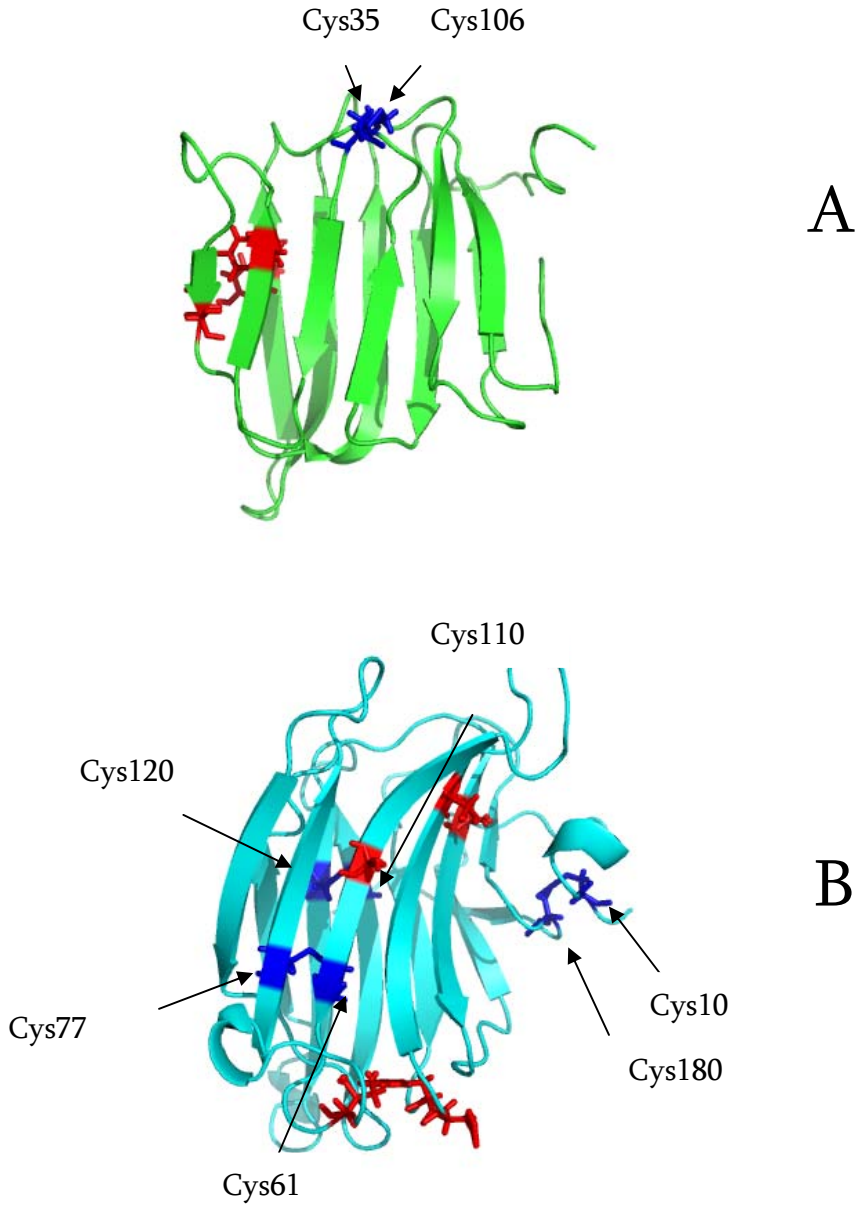


Figure 6.11 Binding interfaces of MIC1-CT and M2AP (place in the same orentation). (A) The binding interface on MIC1-CT (between MIC-CT and EGF-2) (Sawmynaden, 2008) and (B) The binding interaface on M2AP (between M2AP and TSR12): The residues invovled are coloured red and the disulphide bridges are in dark blue. The interface of M2AP was based on chemical shift mapping (See section 6.4.3)

6.5 Conclusion

The studies of M2AP and MIC2 reveal the first structure on the M2AP protein family and structural information on TSR domain. The similarity between M2AP and MIC1-CT structures presents a conserved structure – function relationship in *T. gondii*. Furthermore, the structural studies provide the substrate for study interaction and complex formation. As a result, M2AP and TSR domain were found to interact with each other in a previous unreported way. This result provides the first details of the interaction interface between M2AP and MIC2.

Chapter 7 Conclusions and future perspectives

7.1 Biological aspects

Toxoplasma gondii is a protozoan parasite and model organism for studying other members of Apicomplexan. Its sophisticated nature during its invasion state means all mammals and birds are its targets. The unique 'gliding motility' involves an actomyosin motor powering both host cell invasion and locomotion of these invasive stages. Both properties originate from the family of microneme proteins. Among all microneme proteins involved in the host cell recognition, MIC2 and M2AP form one such motor complexes that is essential for parasite virulence and even survival. This complex has demonstrated significance to the parasite's capacity for invasion and made such proteins important targets for biological study.

The project began with an early study on the MIC3 and MIC8 complex. Data obtained from this work suggested that current *E. coli* expression system employed in the laboratory might be an obstacle to express folded protein. Regardless of its eukaryotic nature, recombinant protein M2AP and MIC2 constructs showed several successes in the described prokaryotic system, such as the M2AP β-sheet domain and TSR12 domain. Together, these works provided a platform that led to better understanding of each individual protein, protein complexes and other biological aspects, such as protein structure, interactions.

7.2 M2AP structure – the first of its kind

M2AP from *Toxoplasma gondii* (TgM2AP) is a member of a protein family that is expressed by many coccidian parasites including *Neospora caninum* and *Eimeria tenella* (Rabenau et al. 2001). MIC1 protein from *E. tenella* (EtMIC1) is thought to be functionally equivalent to TgM2AP (Note that microneme proteins have been named according to the order in which they were discovered, so protein counterparts do not always share the same number designation). M2AP from *N. caninum* (NcM2AP) shows even higher sequence similarity to M2AP (See chapter 6.3). Without any published structure of this family, M2AP could serve as the first model leading to a better understanding of this group of proteins.

7.3 M2AP and MIC1-CT - similarities and differences

M2AP possesses the largest published galectin-like fold (see chapter 6.3.1), which is structurally similar to MIC1-CT. Its extended loops are generally longer than any other galectin-like folds, and these loops appear to be important for the interaction with TSR domain. More interestingly, the carbohydrate recognition domain (CRD) that is commonly found in other galectin-like proteins is not present in either of these microneme proteins (See chapter 6.3.2).

M2AP is known to facilitate the transport of MIC2 through the secretory network. It complexes with MIC2 within 15 minutes of synthesis and remains associated with MIC2 in the micronemes, on the parasite surface and in the culture medium after

release from the parasite plasma membrane. Mutant parasites that do not express M2AP have a defect in parasite invasion, presumably the consequence of insufficient MIC2 transported to the cell surface, but accumulated in the parasite endoplasmic reticulum and Golgi (Jewett and Sibley 2004b). The Galectin-like domain of MIC1-CT presumably interacts with and assists the folding of MIC6, thereby allowing the complex to pass the ER quality control checkpoint, exit and continue its journey to the microneme via the targeting information within the C-terminus of MIC6 (Di Cristina et al. 2000). Therefore, MIC1-CT and M2AP have a similar role in quality control of other microneme protein(s).

Chemical shift mapping of M2AP upon binding to MIC2 TSR domain shows that residues are affected by the interaction are not from the same motif as in MIC1-CT (see chapter 6.3.3). The amino acids involved in the interaction in M2AP are from three adjacent stands and the loops. In MIC1-CT, the interacting residues are mainly from one strand.

In summary, these two microneme proteins share similar structural topology without typical CRD motifs, but have evolved with different ways of interacting with their partners to achieve similar functions.

7.4 TSR12 from MIC2 has a TSP fold

The TSR domain is part of the thrombospondin-related anonymous protein, TRAP, an essential protein in host cell invasion for many apicomplexans. The TSR domain from the TRAP protein in *Plasmodium falciparum* has an elongated structure similar to those of thrombospondin type-1 and F-spondin TSRs (Tossavainen et al. 2006). The first two TSR repeats (TSR12) from MIC2 are very similar structurally to its counterpart. Although further refinement is required, the structure extends the

understanding of TRAP proteins. This is particular important for those none-*Plasmodium* type TRAP proteins, i.e., requiring a partner protein.

By comparing with other TSP domains, TSR-1 repeats of MIC2 were thought to participate in the interaction with M2AP and the formation of the MIC2 and M2AP complex, although the exact mechanism was unclear. The structure became even more important in light of the NMR studies and biological data from our collaborators which confirm the interaction (data not shown).

7.5 M2AP interacts with TRS domains of MIC2

An investigation of the functional properties of the M2AP led to the identification of the M2AP - TSR interaction (*i.e.* by NMR spectroscopy and other biological assays from our collaborator). Such an interaction had previously been unreported and represented a pivotal finding in the project. Interactions were found between M2AP and TSR12 and TSR56, respectively. The lack of folded or even partially folded TSR34 limited our effort to show its potential interaction with M2AP.

A detailed insight into the nature of the M2AP-TSR interaction was provided by chemical shift mapping of both proteins. A structural model will be generated after refinement of TSR12 structure. From the current NMR data, it is clear that the TSR12 docks exclusively onto one side of M2AP (Detailed discussion of the binding interface has been shown in Chapter 6). In conclusion, the nature of M2AP-TSR interaction represents a new and probably general mode of TRAP protein complex formation.

7.6 Collaborative work

In order to confirm our findings *in vivo,* parasite knockout experiments were conducted in collaboration with Vern Carruthers (University of Michigan Medical School). These have shown that parasites conditionally deficient in one MIC2-M2AP complex are invasion incompetent, partially defective in gliding motility, and fail to kill mice even at high doses. The deficient strain also acts as a live-attenuated vaccine since surviving animals are protected from subsequent challenge with a lethal parasite strain (Huynh and Carruthers 2006). Besides the critical importance of the complex, the roles of both components were reported (Huynh et al. 2003). MIC2 is a major virulence factor for the parasite and M2AP facilitates MIC2 during the secretion pathway. It has also been demonstrated that M2AP interacts with TSR56 in the context of the complex *in vivo* (unpublished data).

7.7 Future perspectives

The future direction of the project and some potential follow-up work (including preliminary data and analysis) are detailed below.

7.7.1 Stoichiometry of MIC2 and M2AP

MIC2-M2AP complex was reported to be a hexameric complex, consisting three αβ dimers (Jewett and Sibley 2004a). The hypothesis was suggested by size exclusive gel filtration and native blue gel. However, this stoichiometry has been questioned in the parasite community. Furthermore, MIC2 TSR domains do demonstrate unusual behaviour in many size determining experiments, including SDS gel and size exclusion chromatography. Constructs were often found to behave much larger than

expected. This is probably due to the elongated structure of TSR repeats. The complex which contains these repeats behaves larger in native gel and gel filtration assays. This may explain the controversy between published data (native gel and gel filtration) and data shown here. The titration data evidenced that M2AP binds to at least two different sites in TSR domain of MIC2. This may provide additional evidence that the published stoichiometry (3:3 ratios) need be to be readdressed. A cryogenic electron microscopy approach is currently being pursued in order to address this question.

7.7.2 M2AP and TSR34 – Interaction?

Since no folded TSR34 recombinant protein was produced using the current *E. coli* expression system, the interaction was not able to be confirmed. The new expression system may be able to shed the light on this putative interaction. However, it is possible that TSR34 are intrinsically disordered like EGF2 domain from MIC6 (Sawmynaden et al. 2008).

7.7.3 New expression system

The *E. coli* bacterial expression system is a prokaryotic system that lacks post-translational modifications performed by higher eukaryotic cells such as proteolytic processing, glycosylation and formation of stable complexes. Due to these intrinsic drawbacks, eukaryotic proteins such as MIC2 are often not expressed properly. *Pichia* and mammalian expression system are being tested as an alternative to address these issues.

7.7.4 Model of the complex

To build a MIC2-M2AP complex model is a more ambitious target for this project. The structures of TSR domains and the interaction between TSR34 and M2AP need to be addressed in the future studies. With these structural and interaction information, the model will be able to be built in the near future.

7.8 Publications

1 Camara, B., Liu, M., Shadrin, A., Liu, B., Simpson, P, Weinzierl, R., Severinov, K., Ernesto Cota, Matthews, S., Wigneshweraraj, S. (2010) T7 Phage protein Gp2 inhibits the Escherichia coli RNA polymerase by antagonising stable DNA strand-separation near the transcription start site. Proc Natl Acad Sci U S A 107:2247-52

2 Liu, B., Marchant, J., Simpson, P. and Matthews, S. (2009) Complete resonance assignments for the MIC2 associated protein from Toxoplasma gondii. Biomolecular Nmr Assignments. **3**, 81-83

3 Saouros, S., Blumenschein, T. M., Sawmynaden, K., Marchant, J., Koutroukides, T., Liu, B., Simpson, P., Carpenter, E. P. and Matthews, S. J. (2007) High-level bacterial expression and purification of apicomplexan micronemal proteins for structural studies. Protein Pept Lett. **14**, 411-415

4 Blumenschein, T. M. A., Friedrich, N., Childs, R. A., Saouros, S., Carpenter, E. P., Campanero-Rhodes, M. A., Simpson, P., Chai, W. G., Koutroukides, T., Blackman, M. J., Feizi, T., Soldati-Favre, D. and Matthews, S. (2007) Atomic resolution insight into host cell recognition by Toxoplasma gondii. Embo Journal. **26**, 2808-2820

References

Achbarou A, Mercereau-Puijalon O, Autheman JM, Fortier B, Camus D, Dubremetz JF (1991) Characterization of microneme proteins of Toxoplasma gondii. Mol Biochem Parasitol 47:223-33

Adams, J. C., R. Tucker and J. Lawler (1995). "The thrombospondin gene family." Austin: RG Landes: p.191.

Adams, J. C. and R. P. Tucker (2000). "The thrombospondin type 1 repeat (TSR) superfamily: Diverse proteins with related roles in neuronal development." Developmental Dynamics **218**(2): 280-299.

Ahmad, N., H. J. Gabius, H. Kaltner, S. Andre, I. Kuwabara, F. T. Liu, S. Oscarson, T. Norberg and C. F. Brewer (2002). "Thermodynamic binding studies of cell surface carbohydrate epitopes to galectins-1,-3, and-7: Evidence for differential binding specificities." Canadian Journal of Chemistry-Revue Canadienne De Chimie **80**(8): 1096-1104.

Ahmad, N., H. J. Gabius, S. Sabesan, S. Oscarson and C. F. Brewer (2004). "Thermodynamic binding studies of bivalent oligosaccharides to galectin-1, galectin-3, and the carbohydrate recognition domain of galectin-3." Glycobiology **14**(9): 817-825.

Ajioka, J. W., Fitzpatrick, J. M. Reitter, C. P. (2001). "*Toxoplasma gondii* genomics: shedding light on pathogenesis and chemotherapy." Exp. Rev. Mol. Med: 1-19.

Alexander, D. L., J. Mital, G. E. Ward, P. Bradley and J. C. Boothroyd (2005). "Identification of the moving junction complex of Toxoplasma gondii: A collaboration between distinct secretory organelles." Plos Pathogens **1**(2): 137-149.

Almond DS, Szwandt IS, Edwards G, Lee MG, Winstanley PA (2000) “Disposition of intravenous pyrimethamine in healthy volunteers.” Antimicrob Agents Chemother **44**:1691-3

Asai, T., S. Miura, L. D. Sibley, H. Okabayashi and T. Takeuchi (1995). "Biochemical and Molecular Characterization of Nucleoside Triphosphate Hydrolase Isozymes from the Parasitic Protozoan Toxoplasma-Gondii." Journal of Biological Chemistry **270**(19): 11391-11397.

Asakura, T., K. Taoka, M. Demura and M. P. Williamson (1995). "The Relationship between Amide Proton Chemical-Shifts and Secondary Structure in Proteins." Journal of Biomolecular Nmr **6**(3): 227-236.

Baglia, F. A. and P. N. Walsh (1996). "A binding site for thrombin in the apple 1 domain of factor XI." Journal of Biological Chemistry **271**(7): 3652-3658.

Bannister, L. H., J. M. Hopkins, R. E. Fowler, S. Krishna and G. H. Mitchell (2000). "Ultrastructure of rhoptry development in Plasmodium falciparum erythrocytic schizonts." Parasitology **121**: 273-287.

Baum J, Richard D, Healer J, Rug M, Krnajski Z, Gilberger TW, Green JL, Holder AA, Cowman AF (2006) A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. J Biol Chem 281:5197-208

Barondes, S. H., D. N. W. Cooper, M. A. Gitt and H. Leffler (1994). "Galectins - Structure and Function of a Large Family of Animal Lectins." Journal of Biological Chemistry **269**(33): 20807-20810.

Beghetto, E., H. V. Nielsen, P. Del Porto, W. Buffolano, S. Guglietta, F. Felici, E. Petersen and N. Gargano (2005). "A combination of antigenic regions of Toxoplasma gondii microneme proteins induces protective immunity against oral infection with parasite cysts." Journal of Infectious Diseases **191**(4): 637-645.

Berdoy, M., J. P. Webster and D. W. MacDonald (1995). "The manipulation of rat behaviour by Toxoplasma gondii." Mammalia **59**(4): 605-613.

Berdoy, M., J. P. Webster and D. W. Macdonald (2000). "Fatal attraction in rats infected with Toxoplasma gondii." Proceedings of the Royal Society of London Series B-Biological Sciences **267**(1452): 1591-1594

Bhattacharya A, Tejero R, Montelione GT (2007) "Evaluating protein structures determined by structural genomics consortia." Proteins **66**:778-95

Bhopale, G. M. (2003). "Development of a vaccine for toxoplasmosis: current status." Microbes and Infection **5**(5): 457-462.

Black, M. W. and J. C. Boothroyd (2000). "Lytic cycle of Toxoplasma gondii." Microbiology and Molecular Biology Reviews **64**(3): 607-+.

Blumenschein TM, F. N., Childs RA, Saouros S, Carpenter EP, Campanero-Rhodes MA, Simpson P, Chai W, Koutroukides T, Blackman MJ, Feizi T, Soldati-Favre D, Matthews S (2007). "Atomic resolution insight into host cell recognition by Toxoplasma gondii." The EMBO Journal.

Bonhomme, A., A. Bouchot, N. Pezzella, J. Gomez, H. Le Moal and J. M. Pinon (1999). "Signaling during the invasion of host cells by Toxoplasma gondii." Fems Microbiology Reviews **23**(5): 551-561.

Botero-Kleiven, S., V. Fernandez, J. Lindh, A. Richter-Dahlfors, A. von Euler and M. Wahlgren (2001). "Receptor-mediated endocytosis in an apicomplexan parasite (Toxoplasma gondii)." Experimental Parasitology **98**(3): 134-144.

Bowie, W. R., A. S. King, D. H. Werker, J. L. IsaacRenton, A. Bell, S. B. Eng and S. A. Marion (1997). "Outbreak of toxoplasmosis associated with municipal drinking water." Lancet **350**(9072): 173-177.

Bradley, P. J., C. Ward, S. J. Cheng, D. L. Alexander, S. Coller, G. H. Coombs, J. D. Dunn, D. J. Ferguson, S. J. Sanderson, J. M. Wastling and J. C. Boothroyd (2005). "Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in Toxoplasma gondii." Journal of Biological Chemistry **280**(40): 34245-34258.

Brecht, S., V. B. Carruthers, D. J. P. Ferguson, O. K. Giddings, G. Wang, U. Jakle, J. M. Harper, L. D. Sibley and D. Soldati (2001). "The toxoplasma micronemal

protein MIC4 is an adhesin composed of six conserved apple domains." Journal of Biological Chemistry **276**(6): 4119-4127.

Brossier, F., T. J. Jewett, L. D. Sibley and S. Urban (2005). "A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma." Proceedings of the National Academy of Sciences of the United States of America **102**(11): 4146-4151.

Brossier, F., T. J. Jewettt, J. L. Lovett and L. D. Sibley (2003). "C-terminal processing of the toxoplasma protein MIC2 is essential for invasion into host cells." Journal of Biological Chemistry **278**(8): 6229-6234.

Brossier F, David Sibley L (2005) "Toxoplasma gondii: microneme protein MIC2." Int J Biochem Cell Biol **37**:2266-72

Brown, A. S., C. A. Schaefer, C. P. Quesenberry, L. Y. Liu, V. P. Babulas and E. S. Susser (2005). "Maternal exposure to toxoplasmosis and risk of schizophrenia in adult offspring." American Journal of Psychiatry **162**(4): 767-773.

Buscaglia, C. A., I. Coppens, W. G. J. Hol and V. Nussenzweig (2003). "Sites of interaction between aldolase and thrombospondin-related anonymous protein in Plasmodium." Molecular Biology of the Cell **14**(12): 4947-4957.

Buxton, D. and E. A. Innes (1995). "A Commercial Vaccine for Ovine Toxoplasmosis." Parasitology **110**: S11-S16.

Cantrell, A. S., S. G. Burgett, J. A. Cook, M. C. Smith and H. M. Hsiung (1991). "Effects of 2nd-Codon Mutations on Expression of the Insulin-Like Growth Factor-Ii-Encoding Gene in Escherichia-Coli." Gene **98**(2): 217-223.

Carruthers, V. B. (1999). "Armed and dangerous: *Toxoplasma gondii* uses an arsenal of secretory proteins to infect host cells." Parasitology International **48**: 1-10.

Carruthers, V. B. (2002). "Host cell invasion by the opportunistic pathogen Toxoplasma gondii." Acta Tropica **81**(2): 111-122.

Carruthers, V. B. and J. C. Boothroyd (2007). "Pulling together: an integrated model of Toxoplasma cell invasion." Current Opinion in Microbiology **10**(1): 82-89.

Carruthers, V. B., O. K. Giddings and L. D. Sibley (1999). "Secretion of micronemal proteins is associated with toxoplasma invasion of host cells." Cellular Microbiology **1**(3): 225-235.

Carruthers, V. B., S. Hakansson, O. K. Giddings and L. D. Sibley (2000). "Toxoplasma gondii uses sulfated proteoglycans for substrate and host cell attachment." Infection and Immunity **68**(7): 4005-4011.

Carruthers, V. B. and L. D. Sibley (1997). "Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts." European Journal of Cell Biology **73**(2): 114-123.

Carruthers, V. B. and L. D. Sibley (1999). "Mobilization of intracellular calcium stimulates microneme discharge in Toxoplasma gondii." Molecular Microbiology **31**(2): 421-428.

Catalano-Pons, C., S. Bargy, D. Schlecht, M. D. Tabone, G. Deschenes, A. Bensman and T. Ulinski (2004). "Sulfadiazine-induced nephrolithiasis in children." Pediatric Nephrology **19**(8): 928-931.

Cavanagh J (2007) Protein NMR spectroscopy : principles and practice, 2nd edn. Elsevier Academic Press, Burlington, Mass. ; London

Cerede, O., J. F. Dubremetz, D. Bout and M. Lebrun (2002). "The Toxoplasma gondii protein MIC3 requires pro-peptide cleavage and dimerization to function as adhesin." Embo Journal **21**(11): 2526-2536.

Cerede, O., J. F. Dubremetz, M. Soete, D. Deslee, H. Vial, D. Bout and M. Lebrun (2005). "Synergistic role of micronemal proteins in Toxoplasma gondii virulence." Journal of Experimental Medicine **201**(3): 453-463.

Charron, A. J. and L. D. Sibley (2004). "Molecular partitioning during host cell penetration by Toxoplasma gondii." Traffic **5**(11): 855-867.

Chen, G. F. T. and M. Inouye (1994). "Role of the Aga/Agg Codons, the Rarest Codons in Global Gene-Expression in Escherichia-Coli." Genes & Development **8**(21): 2641-2652.

Chen, H., M. E. Herndon and J. Lawler (2000). "The cell biology of thrombospondin-1." Matrix Biology **19**(7): 597-614.

Chitarra, V., I. Holm, G. A. Bentley, S. Petres and S. Longacre (1999). "The crystal structure of C-terminal merozoite surface protein 1 at 1.8 angstrom resolution, a highly protective malaria vaccine candidate." Molecular Cell **3**(4): 457-464.

Clore GM, Gronenborn AM (1998) "NMR structure determination of proteins and protein complexes larger than 20 kDa." Curr Opin Chem Biol 2:564-70

Comstock GW, Ganley JP (1973) "Association of toxoplasmosis and cats." Am J Epidemiol **97**:424

Coppens, I., M. Andries, J. L. L. Liu and M. F. Cesbron-Delauw (1999). "Intracellular trafficking of dense granule proteins in Toxoplasma gondii and experimental evidences for a regulated exocytosis." European Journal of Cell Biology **78**(7): 463-472.

Coppens I, Joiner KA (2001) "Parasitehost cell interactions in toxoplasmosis: new avenues for intervention?" Expert Rev Mol Med 2001:1-20

Coppens, I., J. D. Dunn, J. D. Romano, M. Pypaert, H. Zhang, J. C. Boothroyd and K. A. Joiner (2006). "Toxoplasma gondii sequesters lysosomes from mammalian hosts in the vacuolar space." Cell **125**(2): 261-274.

Cornilescu, G., F. Delaglio and A. Bax (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." Journal of Biomolecular Nmr **13**(3): 289-302.

Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) "JPred: a consensus secondary structure prediction server." Bioinformatics **14**:892-3

Daragan, V. A. and K. H. Mayo (1999). "Using the model free approach to analyze NMR relaxation data in cases of anisotropic molecular diffusion." Journal of Physical Chemistry B **103**(32): 6829-6834.

Dawson, D. W., S. F. A. Pearce, R. Q. Zhong, R. L. Silverstein, W. A. Frazier and N. P. Bouck (1997). "CD36 mediates the in vitro inhibitory effects of thrombospondin-1 on endothelial cells." Journal of Cell Biology **138**(3): 707-717.

Dawson, D. W., O. V. Volpert, S. F. A. Pearce, A. J. Schneider, R. L. Silverstein, J. Henkin and N. P. Bouck (1999). "Three distinct D-amino acid substitutions confer potent antiangiogenic activity on an inactive peptide derived from a thrombospondin-1 type 1 repeat." Molecular Pharmacology **55**(2): 332-338.

De Peredo, A. G., D. Klein, B. Macek, D. Hess, J. Peter-Katalinic and J. Hofsteenge (2002). "C-mannosylation and O-fucosylation of thrombospondin type 1 repeats." Molecular & Cellular Proteomics **1**(1): 11-18.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax (1995). "Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes." Journal of Biomolecular Nmr **6**(3): 277-293.

DeLano WL (2004) “Use of PYMOL as a communications tool for molecular science.” Abstracts of Papers of the American Chemical Society 228:U313-U314

Desmonts, G. and J. Couvreur (1974). "Toxoplasmosis in Pregnancy and Its Transmission to Fetus." Bulletin of the New York Academy of Medicine **50**(2): 146-159.

Di Cristina M, Spaccapelo R, Soldati D, Bistoni F, Crisanti A (2000) “Two conserved amino acid motifs mediate protein targeting to the micronemes of the apicomplexan parasite Toxoplasma gondii.” Mol Cell Biol 20:7332-41

Djurkovic-Djakovic, O., V. Milenkovic, A. Nikolic, B. Bobic and J. Grujic (2002). "Efficacy of atovaquone combined with clindamycin against murine infection with a cystogenic (Me49) strain of Toxoplasma gondii." Journal of Antimicrobial Chemotherapy **50**(6): 981-987.

Dlugonska H (2008) “Toxoplasma rhoptries: unique secretory organelles and source of promising vaccine proteins for immunoprevention of toxoplasmosis.” J Biomed Biotechnol 2008:632424

Dobrowolski, J. M., V. B. Carruthers and L. D. Sibley (1997). "Participation of myosin in gliding motility and host cell invasion by Toxoplasma gondii." Molecular Microbiology **26**(1): 163-173.

Dobrowolski J, Sibley LD (1997) "The role of the cytoskeleton in host cell invasion by Toxoplasma gondii." Behring Inst Mitt:90-6

Dosset P, Hus JC, Marion D, "Blackledge M (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings." J Biomol NMR 20:223-31

Dowse T, Soldati D (2004) Host cell invasion by the apicomplexans: the significance of microneme protein proteolysis. Curr Opin Microbiol 7:388-96

Driscoll, P. C., A. M. Gronenborn, P. T. Wingfield and G. M. Clore (1990). "Determination of the Secondary Structure and Molecular Topology of Interleukin-1-Beta by Use of 2-Dimensional and 3-Dimensional Heteronuclear N-15-H-1 Nmr-Spectroscopy." Biochemistry **29**(19): 4668-4682.

Dubey, J. P., A. P. Gendronfitzpatrick, A. L. Lenhard and D. Bowman (1988). "Fatal Toxoplasmosis and Enteroepithelial Stages of Toxoplasma-Gondii in a Pallas Cat (Felis-Manul)." Journal of Protozoology **35**(4): 528-530.

Dumetre, A. and M. L. Darde (2005). "Immunomagnetic separation of Toxoplasma gondii oocysts using a monoclonal antibody directed against the oocyst wall." Journal of Microbiological Methods **61**(2): 209-217.

Dutta, C., J. Grimwood and L. H. Kasper (2000). "Attachment of Toxoplasma gondii to a specific membrane fraction of CHO cells." Infection and Immunity **68**(12): 7198-7201.

Fabri, L., E. C. Nice, L. D. Ward, H. Maruta, A. W. Burgess and R. J. Simpson (1992). "Characterization of Bovine Heparin-Binding Neurotrophic Factor (Hbnf) - Assignment of Disulfide Bonds." Biochemistry International **28**(1): 1-9.

Ferguson, D. J. P., S. Brecht and D. Soldati (2000). "The microneme protein MIC4, or an MIC4-like protein, is expressed within the macrogamete and associated

with oocyst wall formation in Toxoplasma gondii." International Journal for Parasitology **30**(11): 1203-1209.

Flegr, J., J. Havlicek, P. Kodym, M. Maly and Z. Smahel (2002). "Increased risk of traffic accidents in subjects with latent toxoplasmosis: a retrospective case-control study." Bmc Infectious Diseases **2**: art. no.-11.

Flegr, J., P. Kodym and V. Tolarova (2000). "Correlation of duration of latent Toxoplasma gondii infection with personality changes in women." Biological Psychology **53**(1): 57-68.

Flegr, J., M. Preiss, J. Klose, J. Havlicek, M. Vitakova and P. Kodym (2003). "Decreased level of psychobiological factor novelty seeking and lower intelligence in men latently infected with the protozoan parasite Toxoplasma gondii Dopamine, a missing link between schizophrenia and toxoplasmosis?" Biological Psychology **63**(3): 253-268.

Fourmaux, M. N., A. Achbarou, O. MercereauPuijalon, C. Biderre, I. Briche, A. Loyens, C. OdbergFerragut, D. Camus and J. F. Dubremetz (1996). "The MIC1 microneme protein of Toxoplasma gondii contains a duplicated receptor-like domain and binds to host cell surface." Molecular and Biochemical Parasitology **83**(2): 201-210.

Fourmaux, M. N., N. GarciaReguet, O. MercereauPuijalon and J. F. Dubremetz (1996). Toxoplasma gondii microneme proteins: Gene cloning and possible function. Toxoplasma Gondii. **219:** 55-58.

Freeman, M. (2004). "Proteolysis within the membrane: Rhomboids revealed." Nature Reviews Molecular Cell Biology **5**(3): 188-197.

Gantt, S. M., P. Clavijo, X. M. Bai, J. D. Esko and P. Sinnis (1997). "Cell adhesion to a motif shared by the malaria circumsporozoite protein and thrombospondin is mediated by its glycosaminoglycan-binding region and not by CSVTCG." Journal of Biological Chemistry **272**(31): 19205-19213.

Garcia-Reguet N, Lebrun M, Fourmaux MN, Mercereau-Puijalon O, Mann T, Beckers CJ, Samyn B, Van Beeumen J, Bout D, Dubremetz JF (2000) “The microneme

protein MIC3 of Toxoplasma gondii is a secretory adhesin that binds to both the surface of the host cells and the surface of the parasite." Cell Microbiol **2**:353-64

Gardner KH, Kay LE (1998) "The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins." Annu Rev Biophys Biomol Struct **27**:357-406

Gaskins, E., S. Gilk, N. DeVore, T. Mann, G. Ward and C. Beckers (2004). "Identification of the membrane receptor of a class XIV myosin in Toxoplasma gondii." Journal of Cell Biology **165**(3): 383-393.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) "ExPASy: The proteomics server for in-depth protein knowledge and analysis." Nucleic Acids Res 31:3784-8

Gheysen, D., D. Iserentant, C. Derom and W. Fiers (1982). "Systematic Alteration of the Nucleotide-Sequence Preceding the Translation Initiation Codon and the Effects on Bacterial Expression of the Cloned Sv40 Small-T Antigen Gene." Gene **17**(1): 55-63.

Grimwood, J. and J. E. Smith (1996). "Toxoplasma gondii: The role of parasite surface and secreted proteins in host cell invasion." International Journal for Parasitology **26**(2): 169-173.

Groenendyk, J. and M. Michalak (2005). "Endoplasmic reticulum quality control and apoptosis." Acta Biochimica Polonica **52**(2): 381-395.

Guerina, N. G., H. W. Hsu, H. C. Meissner, J. H. Maguire, R. Lynfield, B. Stechenberg, I. Abroms, M. S. Pasternack, R. Hoff, R. B. Eaton, G. F. Grady, S. H. Cheeseman, K. McIntosh, D. N. Medearis, R. Robb and B. J. Weiblen (1994). "Neonatal Serologic Screening and Early Treatment for Congenital Toxoplasma-Gondii Infection." New England Journal of Medicine **330**(26): 1858-1863.

Guitard, J., N. Kamar, M. Mouzin, J. S. Borde, T. Tran-Van, D. Durand and L. Rostaing (2005). "Sulfadiazine-related obstructive urinary tract lithiasis: an

unusual cause of acute renal failure after kidney transplantation." Clinical Nephrology **63**(5): 405-407.

Guo, N. H., H. C. Krutzsch, E. Negre, V. S. Zabrenetzky and D. D. Roberts (1992). "Heparin-Binding Peptides from the Type-I Repeats of Thrombospondin - Structural Requirements for Heparin Binding and Promotion of Melanoma Cell-Adhesion and Chemotaxis." Journal of Biological Chemistry **267**(27): 19349-19355.

Hajduk PJ, Dinges J, Miknis GF, Merlock M, Middleton T, Kempf DJ, Egan DA, Walter KA, Robins TS, Shuker SB, Holzman TF, Fesik SW (1997) NMR-based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein. J Med Chem 40:3144-50

Hakansson, S., A. J. Charron and L. D. Sibley (2001). "Toxoplasma evacuoles: a two-step process of secretion and fusion forms the parasitophorous vacuole." Embo Journal **20**(12): 3132-3144.

Harper, J. M., E. F. Hoff and V. B. Carruthers (2004). "Multimerization of the Toxoplasma gondii MIC2 integrin-like A-domain is required for binding to heparin and human cells." Molecular and Biochemical Parasitology **134**(2): 201-212.

Havlicek, J., Z. Gasova, A. P. Smith, K. Zvara and J. Flegr (2001). "Decrease of psychomotor performance in subjects with latent 'asymptomatic' toxoplasmosis." Parasitology **122**: 515-520.

Hay, J., P. P. Aitken, D. M. Hair, W. M. Hutchison and D. I. Graham (1984). "The Effect of Congenital Toxoplasma Infection on Mouse Activity and Relative Preference for Exposed Areas over a Series of Trials." Annals of Tropical Medicine and Parasitology **78**(6): 611-618.

Heath P, Zuelzer WW (1944) “Toxoplasmosis (Report of Eye Findings in Infant Twins).” Trans Am Ophthalmol Soc **42**:119-32

Hehl, A. B., C. Lekutis, M. E. Grigg, P. J. Bradley, J. F. Dubremetz, E. Ortega-Barria and J. C. Boothroyd (2000). "Toxoplasma gondii homologue of plasmodium

apical membrane antigen 1 is involved in invasion of host cells." Infection and Immunity **68**(12): 7078-7086.

Heintzelman, M. B. and J. D. Schwartzman (1997). "A novel class of unconventional myosins from Toxoplasma gondii." Journal of Molecular Biology **271**(1): 139-146.

Herm-Gotz, A., S. Weiss, R. Stratmann, S. Fujita-Becker, C. Ruff, E. Meyhofer, T. Soldati, D. J. Manstein, M. A. Geeves and D. Soldati (2002). "Toxoplasma gondii myosin A and its light chain: a fast, single-headed, plus-end-directed motor." Embo Journal **21**(9): 2149-2158.

Hoff, E. F. and V. B. Carruthers (2002). "Is Toxoplasma egress the first step in invasion?" Trends in Parasitology **18**(6): 251-255.

Hofsteenge, J., M. Blommers, D. Hess, A. Furmanek and O. Miroshnichenko (1999). "The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues." Journal of Biological Chemistry **274**(46): 32786-32794.

Holliman RE (1988) “Toxoplasmosis and the acquired immune deficiency syndrome.” J Infect **16**:121-8

Holm, L. and C. Sander (1993). "Protein-Structure Comparison by Alignment of Distance Matrices." Journal of Molecular Biology **233**(1): 123-138.

Hooft, R. W. W., G. Vriend, C. Sander and E. E. Abola (1996). "Errors in protein structures." Nature **381**(6580): 272-272.

Howe DK, Honore S, Derouin F, Sibley LD (1997) “Determination of genotypes of Toxoplasma gondii strains isolated from patients with toxoplasmosis.” J Clin Microbiol **35**:1411-4

Hughes, R. C. (2001). "Galectins as modulators of cell adhesion." Biochimie **83**(7): 667-676.

Huwiler, K. G., M. M. Vestling, D. S. Annis and D. F. Mosher (2002). "Biophysical characterization, including disulfide bond assignments, of the anti-angiogenic

type 1 domains of human thrombospondin-1." Biochemistry **41**(48): 14329-14339.

Huynh, M. H., C. Opitz, L. Y. Kwok, F. M. Tomley, V. B. Carruthers and D. Soldati (2004). "Trans-genera reconstitution and complementation of an adhesion complex in Toxoplasma gondii." Cellular Microbiology **6**(8): 771-782.

Huynh, M. H., K. E. Rabenau, J. M. Harper, W. L. Beatty, L. D. Sibley and V. B. Carruthers (2003). "Rapid invasion of host cells by Toxoplasma requires secretion of the MIC2-M2AP adhesive protein complex." Embo Journal **22**(9): 2082-2090.

Huynh MH, Carruthers VB (2006) "Toxoplasma MIC2 is a major determinant of invasion and virulence." PLoS Pathog **2**:e84

Inoue, H., H. Nojima and H. Okayama (1990). "High-Efficiency Transformation of Escherichia-Coli with Plasmids." Gene **96**(1): 23-28.

Iwasaki, W., K. Nagata, H. Hatanaka, T. Inui, T. Kimura, T. Muramatsu, K. Yoshida, M. Tasumi and F. Inagaki (1997). "Solution structure of midkine, a new heparin-binding growth factor." Embo Journal **16**(23): 6936-6946.

Isaacson RL, Simpson PJ, Liu M, Cota E, Zhang X, Freemont P, Matthews S (2007) "A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues." J Am Chem Soc **129**:15428-9

Jewett, T. J. and L. D. Sibley (2003). "Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites." Molecular Cell **11**(4): 885-894.

Jewett, T. J. and L. D. Sibley (2004). "The toxoplasma proteins MIC2 and M2AP form a hexameric complex necessary for intracellular survival." Journal of Biological Chemistry **279**(10): 9362-9369.

Johnson, B. A. and R. A. Blevins (1994). "Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data." Journal of Biomolecular Nmr **4**(5): 603-614.

Johnson BA (2004) “Using NMRView to visualize and analyze the NMR spectra of macromolecules.” Methods Mol Biol **278**:313-52

Joiner, K. A. and D. S. Roos (2002). "Secretory traffic in the eukaryotic parasite Toxoplasma gondii: less is more." Journal of Cell Biology **157**(4): 557-563.

Julander, I., C. Martin, M. Lappalainen, E. Guy, B. Isberg and B. Evengard (2001). "Polymerase chain reaction for diagnosis of cerebral toxoplasmosis in cerebrospinal fluid in HIV-positive patients." Scandinavian Journal of Infectious Diseases **33**(7): 538-541.

Jung, Y. S. and M. Zweckstetter (2004). "Mars - robust automatic backbone assignment of proteins." Journal of Biomolecular Nmr **30**(1): 11-23.

Kane, J. F. (1995). "Effects of Rare Codon Clusters on High-Level Expression of Heterologous Proteins in Escherichia-Coli." Current Opinion in Biotechnology **6**(5): 494-500.

Kappe, S., T. Bruderer, S. Gantt, H. Fujioka, V. Nussenzweig and R. Menard (1999). "Conservation of a gliding motility and cell invasion machinery in Apicomplexan parasites." Journal of Cell Biology **147**(5): 937-943.

Kats, L. M., C. G. Black, N. I. Proellocks and R. L. Coppel (2006). "Plasmodium rhoptries: how things went pear-shaped." Trends in Parasitology **22**(6): 269-276.

Kay, L. E., D. A. Torchia and A. Bax (1989). "Backbone Dynamics of Proteins as Studied by N-15 Inverse Detected Heteronuclear Nmr-Spectroscopy - Application to Staphylococcal Nuclease." Biochemistry **28**(23): 8972-8979.

Keeler J (2005) “Understanding NMR spectroscopy.” John Wiley & Sons, Chichester

Keeley, A. and D. Soldati (2004). "The glideosome: a molecular machine powering motility and host-cell invasion by apicomplexa." Trends in Cell Biology **14**(10): 528-532.

Keller, N., A. Naguleswaran, A. Cannas, N. Vonlaufen, M. Bienz, C. Bjorkman, W. Bohne and A. Hemphill (2002). "Identification of a Neospora caninum

microneme protein (NcMIC1) which interacts with sulfated host cell surface glycosaminoglycans." Infection and Immunity **70**(6): 3187-3198.

Kilpelainen, I., M. Kaksonen, T. Kinnunen, H. Avikainen, M. Fath, R. J. Linhardt, E. Raulo and H. Rauvala (2000). "Heparin-binding growth-associated molecule contains two heparin-binding beta-sheet domains that are homologous to the thrombospondin type I repeat." Journal of Biological Chemistry **275**(18): 13564-13570.

Kravetz, J. D. and D. G. Federman (2005). "Toxoplasmosis in pregnancy." American Journal of Medicine **118**(3): 212-216.

Kvansakul, M., J. C. Adams and E. Hohenester (2004). "Structure of a thrombospondin C-terminal fragment reveals a novel calcium core in the type 3 repeats." Embo Journal **23**(6): 1223-1233.

Lafferty, K. D. (2006). "Can the common brain parasite, Toxoplasma gondii, influence human culture?" Proc Biol Sci **273**(1602): 2749-55.

Labbe, M., P. de Venevelles, F. Girard-Misguich, C. Bourdieu, A. Guillaume and P. Pery (2005). "Eimeria tenella microneme protein EtMIC3: identification, localisation and role in host cell infection." Molecular and Biochemical Parasitology **140**(1): 43-53.

Laskowski, R. A., J. A. C. Rullmann, M. W. MacArthur, R. Kaptein and J. M. Thornton (1996). "AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR." Journal of Biomolecular Nmr **8**(4): 477-486.

Lawler, J. (1986). "The Structural and Functional-Properties of Thrombospondin." Blood **67**(5): 1197-1209.

Lee, J. O., P. Rieu, M. A. Arnaout and R. Liddington (1995). "Crystal-Structure of the a-Domain from the a-Subunit of Integrin Cr3 (Cd11b/Cd18)." Cell **80**(4): 631-638.

Leonidas, D. D., B. L. Elbert, Z. Zhou, H. Leffler, S. J. Ackerman and K. R. Acharya (1995). "Crystal structure of human Charcot-Leyden crystal protein, an

eosinophil lysophospholipase, identifies it as a new member of the carbohydrate-binding family of galectins." Structure **3**(12): 1379-1393.

Leonidas, D. D., E. H. Vatzaki, H. Vorum, J. E. Celis, P. Madsen and K. R. Acharya (1998). "Structural basis for the recognition of carbohydrates by human galectin-7." Biochemistry **37**(40): 13930-13940.

Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P (2006) "SMART 5: domains in the context of genomes and networks." Nucleic Acids Res **34**:D257-60

Liao, D. and F. A. Ferrone (1994). "Measurements of Solubilities of Sickle Hemoglobin in Small Sample Volumes." Biophysical Journal **66**(2): A395-A395.

Linge, J. P., S. I. O'Donoghue and M. Nilges (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B. **339:** 71-90.

Linge JP, Habeck M, Rieping W, Nilges M (2003a) "ARIA: automated NOE assignment and NMR structure calculation." Bioinformatics **19**:315-6

Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M (2003b) "Refinement of protein structures in explicit solvent." Proteins: Structure, Function, and Genetics 50:496-506

Lipsitz RS, Tjandra N (2004) "Residual dipolar couplings in NMR structure analysis." Annual Review of Biophysics and Biomolecular Structure **33**:387-413

Liu B, Sawmynaden K, Marchant J, Simpson P, Matthews S (2009) Complete resonance assignments for the MIC2 associated protein from Toxoplasma gondii. Biomolecular Nmr Assignments 3:81-83

Lobsanov, Y. D., M. A. Gitt, H. Leffler, S. H. Barondes and J. M. Rini (1993). "X-Ray Crystal-Structure of the Human Dimeric S-Lac Lectin, L-14-Ii, in Complex with Lactose at 2.9-Angstrom Resolution." Journal of Biological Chemistry **268**(36): 27034-27038.

Lopez, R., J. Valbuena, H. Curtidor, A. Puentes, A. E. Rodriguez, J. Garcia, J. Suarez, R. Vera, M. Ocampo, M. Trujillo, L. E. Ramirez and M. E. Patarroyo (2004).

"Plasmodium falciparum: red blood cell binding studies using peptides derived from rhoptry-associated protein 2 (RAP2)." Biochimie **86**(1): 1-6.

Lourenco, E. V., A. Panunto-Castelo, J. B. Molfetta, N. C. Avanci, M. H. S. Goldman and M. C. Roque-Barreira (2006). "Immunization with MIC1 and MIC4 induces protective immunity against toxoplasma gondii." Glycobiology **16**(11): 1112-1112.

Lourenco, E. V., S. R. Pereira, V. M. Faca, A. A. M. Coelho-Castelo, J. R. Mineo, M. C. Roque-Barreira, L. J. Greene and A. Panunto-Castelo (2001). "Toxoplasma gondii micronemal protein MIC1 is a lactose-binding lectin." Glycobiology **11**(7): 541-547.

Luft, B. J. and J. S. Remington (1992). "Toxoplasmic Encephalitis in Aids." Clinical Infectious Diseases **15**(2): 211-222.

Lynfield R, G. N. (1997). "Toxoplasmosis." Pediatr Rev **18**: 75-83.

Makrides, S. C. (1996). "Strategies for achieving high-level expression of genes in Escherichia coli." Microbiological Reviews **60**(3): 512-&.

Marchant J, Sawmynaden K, Saouros S, Simpson P, Matthews S (2008) “Complete resonance assignment of the first and second apple domains of MIC4 from Toxoplasma gondii, using a new NMRView-based assignment aid.” Biomolecular Nmr Assignments **2**:119-121

Marechal, E. and M. F. Cesbron-Delauw (2001). "The apicoplast: a new member of the plastid family." Trends in Plant Science **6**(5): 200-205.

Matuschewski, K., A. C. Nunes, V. Nussenzweig and R. Menard (2002). "Plasmodium sporozoite invasion into insect and mammalian cells is directed by the same dual binding system." Embo Journal **21**(7): 1597-1606.

McCormick, C. J., D. S. Tuckwell, A. Crisanti, M. J. Humphries and M. R. Hollingdale (1999). "Identification of heparin as a ligand for the A-domain of Plasmodium falciparum thrombospondin-related adhesion protein." Molecular and Biochemical Parasitology **100**(1): 111-124.

Meissner, M., M. Reiss, N. Viebig, V. B. Carruthers, C. Toursel, S. Tomavo, J. W. Ajioka and D. Soldati (2002). "A family of transmembrane microneme proteins of Toxoplasma gondii contain EGF-like domains and function as escorters." Journal of Cell Science **115**(3): 563-574.

Meissner, M., D. Schluter and D. Soldati (2002). "Role of Toxoplasma gondii myosin A in powering parasite gliding and host cell invasion." Science **298**(5594): 837-840.

Mercier, C., K. D. Z. Adjogble, W. Daubener and M. F. C. Delauw (2005). "Dense granules: Are they key organelles to help understand the parasitophorous vacuole of all apicomplexa parasites?" International Journal for Parasitology **35**(8): 829-849.

Mercier, C., L. Lecordier, F. Darcy, D. Deslee, A. Murray, B. Tourvieille, P. Maes, A. Capron and M. F. Cesbrondelauw (1993). "Molecular Characterization of a Dense Granule Antigen (Gra-2) Associated with the Network of the Parasitophorous Vacuole in Toxoplasma-Gondii." Molecular and Biochemical Parasitology **58**(1): 71-82.

Mevelec, M. N., D. Bout, B. Desolme, H. Marchand, R. Magne, O. Bruneel and D. Buzoni-Gatel (2005). "Evaluation of protective effect of DNA vaccination with genes encoding antigens GRA4 and SAG1 associated with GM-CSF plasmid, against acute, chronical and congenital toxoplasmosis in mice." Vaccine **23**(36): 4489-4499.

Mikkelsen, R. B., M. Kamber, K. S. Wadwa, P. S. Lin and R. Schmidtullrich (1988). "The Role of Lipids in Plasmodium-Falciparum Invasion of Erythrocytes - a Coordinated Biochemical and Microscopic Analysis." Proceedings of the National Academy of Sciences of the United States of America **85**(16): 5956-5960.

Mineo, J. R. and L. H. Kasper (1994). "Attachment of Toxoplasma-Gondii to Host-Cells Involves Major Surface Protein, Sag-1 (P-30)." Experimental Parasitology **79**(1): 11-20.

Misenheimer, T. M., K. G. Huwiler, D. S. Annis and D. F. Mosher (2000). "Physical

characterization of the procollagen module of human thrombospondin 1 expressed in insect cells." Journal of Biological Chemistry **275**(52): 40938-40945.

Mital, J., M. Meissner, D. Soldati and G. E. Ward (2005). "Conditional expression of Toxoplasma gondii apical membrane antigen-1 (TgAMA1) demonstrates that TgAMA1 plays a critical role in host cell invasion." Molecular Biology of the Cell **16**(9): 4341-4349.

Mondragon, R. and E. Frixione (1996). "Ca2+-dependence of conoid extrusion in Toxoplasma gondii tachyzoites." Journal of Eukaryotic Microbiology **43**(2): 120-127.

Montelione GT, Zheng D, Huang YJ, Gunsalus KC, Szyperski T (2000) Protein NMR spectroscopy in structural genomics. Nat Struct Biol 7 Suppl:982-5

Montoya JG, Liesenfeld O (2004) "Toxoplasmosis." Lancet 363:1965-76

Mordue, D. G., N. Desai, M. Dustin and L. D. Sibley (1999). "Invasion by Toxoplasma gondii establishes a moving junction that selectively excludes host cell plasma membrane proteins on the basis of their membrane anchoring." Journal of Experimental Medicine **190**(12): 1783-1792.

Mordue, D. G., S. Hakansson, I. Niesman and L. D. Sibley (1999). "Toxoplasma gondii resides in a vacuole that avoids fusion with host cell endocytic and exocytic vesicular trafficking pathways." Experimental Parasitology **92**(2): 87-99.

Mordue, D. G. and L. D. Sibley (1997). "Intracellular fate of vacuoles containing Toxoplasma gondii is determined at the time of formation and depends on the mechanism of entry." Journal of Immunology **159**(9): 4452-4459.

Morgan, W. D., B. Birdsall, T. A. Frenkiel, M. G. Gradwell, P. A. Burghaus, S. E. H. Syed, C. Uthaipibull, A. A. Holder and J. Feeney (1999). "Solution structure of an EGF module pair from the Plasmodium falciparum merozoite surface protein 1." Journal of Molecular Biology **289**(1): 113-122.

Morrissette, N. S. and L. D. Sibley (2002). "Cytoskeleton of apicomplexan parasites."

Microbiology and Molecular Biology Reviews **66**(1): 21-+.

Muhandiram, D. R., N. A. Farrow, G. Y. Xu, S. H. Smallcombe and L. E. Kay (1993). "A Gradient C-13 Noesy-Hsqc Experiment for Recording Noesy Spectra of C-13-Labeled Proteins Dissolved in H2o." Journal of Magnetic Resonance Series B **102**(3): 317-321.

Naitza, S., F. Spano, K. J. H. Robson and A. Crisanti (1998). "The thrombospondin-related protein family of apicomplexan parasites: The gears of the cell invasion machinery." Parasitology Today **14**(12): 479-484.

Neidhard.Fc, P. L. Bloch and D. F. Smith (1974). "Culture Medium for Enterobacteria." Journal of Bacteriology **119**(3): 736-747.

Neudeck, A., S. Stachelhaus, N. Nischik, B. Striepen, G. Reichmann and H. G. Fischer (2002). "Expression variance, biochemical and immunological properties of Toxoplasma gondii dense granule protein GRA7." Microbes and Infection **4**(6): 581-590.

Neuhaus D, Williamson MP (1989) “The nuclear Overhauser effect in structural and conformational analysis.” VCH, Weinheim ; Cambridge

Novotna M, Hanusova J, Klose J, Preiss M, Havlicek J, Roubalova K, Flegr J (2005) Probable neuroimmunological link between Toxoplasma and cytomegalovirus infections and personality changes in the human host. BMC Infect Dis 5:54

Opitz, C., M. Di Cristina, M. Reiss, T. Ruppert, A. Crisanti and D. Soldati (2002). "Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite Toxoplasma gondii." Embo Journal **21**(7): 1577-1585.

Ortega-Barria, E. and J. C. Boothroyd (1999). "A Toxoplasma lectin-like activity specific for sulfated polysaccharides is involved in host cell infection." Journal of Biological Chemistry **274**(3): 1267-1276.

Pancake, S. J., G. D. Holt, S. Mellouk and S. L. Hoffman (1992). "Malaria Sporozoites and Circumsporozoite Proteins Bind Specifically to Sulfated Glycoconjugates." Journal of Cell Biology **117**(6): 1351-1357.

Parodi, A. J. (2000). "Recognition of misfolded conformers by the UDP-Glc: Glycoprotein glucosyltransferase." Glycobiology **10**(10): 1077-1077.

Philip J. Brown, D. M., Jennifer R. Potts2 Fiona M. Tomley and Iain D. Campbell (2003). "Solution structure of a PAN module from the apicomplexan parasite Eimeria tenella " Journal of Structural and Functional Genomics **4**: 227-234.

Pollock, S., G. Kozlov, M. F. Pelletier, J. F. Trempe, G. Jansen, D. Sitnikov, J. J. M. Bergeron, K. Gehring, I. Ekiel and D. Y. Thomas (2004). "Specific interaction of ERp57 and calnexin determined by NMR spectroscopy and an ER two-hybrid system." Embo Journal **23**(5): 1020-1029.

Ponticelli, C. and M. R. Campise (2005). "Neurological complications in kidney transplant recipients." Journal of Nephrology **18**(5): 521-528.

Porchethennere, E. and G. Torpier (1983). "Relations between Toxoplasma and Its Host-Cell." Protistologica **19**(3): 357-370.

Qu, A. D. and D. J. Leahy (1995). "Crystal-Structure of the I-Domain from the Cd11a/Cd18 (Lfa-1, Alpha(L)Beta-2) Integrin." Proceedings of the National Academy of Sciences of the United States of America **92**(22): 10277-10281.

Rabenau, K. E., A. Sohrabi, A. Tripathy, C. Reitter, J. W. Ajioka, F. M. Tomley and V. B. Carruthers (2001). "TgM2AP participates in Toxoplasma gondii host cells and is tightly associated with the invasion of adhesive protein TgMIC2." Molecular Microbiology **41**(3): 537-547.

Rajagopal, P., E. B. Waygood, J. Reizer, M. H. Saier and R. E. Klevit (1997). "Demonstration of protein-protein interaction specificity by NMR chemical shift mapping." Protein Science **6**(12): 2624-2627.

Reiss, M., N. Viebig, S. Brecht, M. N. Fourmaux, M. Soete, M. Di Cristina, J. F. Dubremetz and D. Soldati (2001). "Identification and characterization of an escorter for two secretory adhesins in Toxoplasma gondii." Journal of Cell Biology **152**(3): 563-578.

Remington, J. S. and S. Y. Wong (1995). "Prevention of Toxoplasma Infection in Pregnant-Women and Their Fetuses - Reply." Clinical Infectious Diseases

20(3): 728-729.

Renold, C., A. Sugar, J. P. Chave, L. Perrin, J. Delavelle, G. Pizzolato, P. Burkhard, V. Gabriel and B. Hirschel (1992). "Toxoplasma Encephalitis in Patients with the Acquired-Immunodeficiency-Syndrome." Medicine **71**(4): 224-239.

Ribeiro, D. A., P. C. M. Pereira, J. M. Machado, S. B. Silva, A. W. P. Pessoa and D. M. F. Salvadori (2004). "Does toxoplasmosis cause DNA damage? An evaluation in isogenic mice under normal diet or dietary restriction." Mutation Research-Genetic Toxicology and Environmental Mutagenesis **559**(1-2): 169-176.

Rich, K. A., F. W. George, J. L. Law and W. J. Martin (1990). "Cell-Adhesive Motif in Region-Ii of Malarial Circumsporozoite Protein." Science **249**(4976): 1574-1577.

Roberts, T. and J. K. Frenkel (1990). "Estimating Income Losses and Other Preventable Costs Caused by Congenital Toxoplasmosis in People in the United-States." Journal of the American Veterinary Medical Association **196**(2): 249-256.

Robson, K. J. H., J. R. S. Hall, M. W. Jennings, T. J. R. Harris, K. Marsh, C. I. Newbold, V. E. Tate and D. J. Weatherall (1988). "A Highly Conserved Amino-Acid-Sequence in Thrombospondin, Properdin and in Proteins from Sporozoites and Blood Stages of a Human Malaria Parasite." Nature **335**(6185): 79-82.

S. Saouros, T. M. A. B., K. Sawmynaden, J. Marchant, T. Koutroukides, B. Liu, P. Simpson, E.P. Carpenter and S.J. Matthews (2007). "High-Level Bacterial Expression and Purification of Apicomplexan Micronemal Proteins for Structural Studies." Protein & Peptide Letters **14**(5): 411-415.

Saffer, L. D., O. Mercereaupuijalon, J. F. Dubremetz and J. D. Schwartzman (1992). "Localization of a Toxoplasma-Gondii Rhoptry Protein by Immunoelectron Microscopy During and after Host-Cell Penetration." Journal of Protozoology **39**(4): 526-530.

Sager, H., M. Gloor, A. Tenter, S. Maley, M. Hassig and B. Gottstein (2003). "Immunodiagnosis of primary Toxoplasma gondii infection in sheep by the use

of a P30 IgG avidity ELISA." Parasitology Research **91**(2): 171-174.

Saouros S, Blumenschein TM, Sawmynaden K, Marchant J, Koutroukides T, Liu B, Simpson P, Carpenter EP, Matthews SJ (2007) High-level bacterial expression and purification of apicomplexan micronemal proteins for structural studies. Protein Pept Lett 14:411-5

Saouros, S., H. A. Chen, P. Simpson, E. Cota, B. Edwards-Jones, D. Soldati-Favre and S. Matthews (2005). "Letter to the Editor: Complete resonance assignments of the C-terminal domain from MIC1: A micronemal protein from Toxoplasma gondii." Journal of Biomolecular Nmr **31**(2): 177-178.

Saouros, S., B. Edwards-Jones, M. Reiss, K. Sawmynaden, E. Cota, P. Simpson, T. J. Dowse, U. Jakle, S. Ramboarina, T. Shivarattan, S. Matthews and D. Soldati-Favre (2005). "A novel galectin-like domain from Toxoplasma gondii micronemal protein 1 assists the folding, assembly, and transport of a cell adhesion complex." Journal of Biological Chemistry **280**(46): 38583-38591.

Sawmynaden K, Saouros S, Friedrich N, Marchant J, Simpson P, Bleijlevens B, Blackman MJ, Soldati-Favre D, Matthews S (2008) Structural insights into microneme protein assembly reveal a new mode of EGF domain recognition. EMBO Rep 9:1149-55

Schrag, J. D., J. J. M. Bergeron, Y. G. Li, S. Borisova, M. Hahn, D. Y. Thomas and M. Cygler (2001). "The structure of calnexin, an ER chaperone involved in quality control of protein folding." Molecular Cell **8**(3): 633-644.

Seetharaman, J., A. Kanigsberg, R. Slaaby, H. Leffler, S. H. Barondes and J. M. Rini (1998). "X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-angstrom resolution." Journal of Biological Chemistry **273**(21): 13047-13052.

Sibley LD (2004) Intracellular parasite invasion strategies. Science 304:248-53

Sibley, L. D., S. Hakansson and V. B. Carruthers (1998). "Gliding motility: An efficient mechanism for cell penetration." Current Biology **8**(1): R12-R14.

Sibley, L. D., I. R. Niesman, T. Asai and T. Takeuchi (1994). "Toxoplasma-Gondii - Secretion of a Potent Nucleoside Triphosphate Hydrolase into the Parasitophorous Vacuole." Experimental Parasitology **79**(3): 301-311.

Sinai, A. P. and K. A. Joiner (2001). "The Toxoplasma gondii protein ROP2 mediates host organelle association with the parasitophorous vacuole membrane." Journal of Cell Biology **154**(1): 95-108.

Sinai, A. P., P. Webster and K. A. Joiner (1997). "Association of host cell endoplasmic reticulum and mitochondria with the Toxoplasma gondii parasitophorous vacuole membrane: a high affinity interaction." Journal of Cell Science **110**: 2117-2128.

Soldati, D., J. F. Dubremetz and M. Lebrun (2001). "Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite Toxoplasma gondii." International Journal for Parasitology **31**(12): 1293-1302.

Soldati, D., K. Kim, J. Kampmeier, J. F. Dubremetz and J. C. Boothroyd (1995). "Complementation of a Toxoplasma gondii ROP1 knock-out mutant using phleomycin selection." Molecular and Biochemical Parasitology **74**(1): 87-97.

Sparrow, C. P., H. Leffler and S. H. Barondes (1987). "Multiple Soluble Beta-Galactoside-Binding Lectins from Human-Lung." Journal of Biological Chemistry **262**(15): 7383-7390.

Stebbins, C. E. and J. E. Galan (2003). "Priming virulence factors for delivery into the host." Nature Reviews Molecular Cell Biology **4**(9): 738-743.

Stibbs, H. H. (1985). "Changes in Brain Concentrations of Catecholamines and Indoleamines in Toxoplasma-Gondii Infected Mice." Annals of Tropical Medicine and Parasitology **79**(2): 153-157.

Stormo, G. D. (1986). "Maximizing protein expression." (Reznikoff, W &Gold, L., Eds.) Butterworth, Stoneham, Massachusetts.

Striepen, B., D. Soldati, N. Garcia-Reguet, J. F. Dubremetz and D. S. Roos (2001). "Targeting of soluble proteins to the rhoptries and micronemes in Toxoplasma

gondii." Molecular and Biochemical Parasitology **113**(1): 45-53.

Studier, F. W. and B. A. Moffatt (1986). "Use of Bacteriophage-T7 Rna-Polymerase to Direct Selective High-Level Expression of Cloned Genes." Journal of Molecular Biology **189**(1): 113-130.

SussToby, E., J. Zimmerberg and G. E. Ward (1996). "Toxoplasma invasion: The parasitophorous vacuole is formed from host cell plasma membrane and pinches off via a fission pore." Proceedings of the National Academy of Sciences of the United States of America **93**(16): 8413-8418.

Swaminathan, G. J., D. D. Leonidas, M. P. Savage, S. J. Ackerman and K. R. Acharya (1999). "Selective recognition of mannose by the human eosinophil Charcot-Leyden crystal protein (galectin-10): A crystallographic study at 1.8 angstrom resolution." Biochemistry **38**(42): 13837-13843.

Takagi J, DeBottis DP, Erickson HP, Springer TA (2002) The role of the specificity-determining loop of the integrin beta subunit I-like domain in autonomous expression, association with the alpha subunit, and ligand binding. Biochemistry 41:4339-47

Tan, K. M., M. Duquette, J. H. Liu, Y. C. Dong, R. G. Zhang, A. Joachimiak, J. Lawler and J. H. Wang (2002). "Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication." Journal of Cell Biology **159**(2): 373-382.

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

Tobias, J. W., T. E. Shrader, G. Rocap and A. Varshavsky (1991). "The N-End Rule in Bacteria." Science **254**(5036): 1374-1377.

Tolman, J. R., J. M. Flanagan, M. A. Kennedy and J. H. Prestegard (1995). "Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution." Proceedings of the National Academy of Sciences of the United States of America **92**(20): 9279-9283.

Tomley, F. M., K. J. Billington, J. M. Bumstead, J. D. Clark and P. Monaghan (2001). "EtMIC4: a microneme protein from Eimeria tenella that contains tandem arrays of epidermal growth factor-like repeats and thrombospondin type-1 repeats." International Journal for Parasitology **31**(12): 1303-1310.

Tomley, F. M. and D. S. Soldati (2001). "Mix and match modules: structure and function of microneme proteins in apicomplexan parasites." Trends in Parasitology **17**(2): 81-88.

Torda A (2001) Toxoplasmosis. Are cats really the source? Aust Fam Physician 30:743-7

Torrey, E. F. and R. H. Yolken (2003). "Toxoplasma gondii and schizophrenia." Emerging Infectious Diseases **9**(11): 1375-1380.

Tucker, R. P. (2004). "The thrombospondin type 1 repeat superfamily." International Journal of Biochemistry & Cell Biology **36**(6): 969-974.

Umemiya, T., M. Takeichi and A. Nose (1997). "M-spondin, a novel ECM protein highly homologous to vertebrate F-spondin, is localized at the muscle attachment sites in the Drosophila embryo." Developmental Biology **186**(2): 165-176.

Umemoto, K., H. Leffler, A. Venot, H. Valafar and J. H. Prestegard (2003). "Conformational differences in liganded and unliganded states of Galectin-3." Biochemistry **42**(13): 3688-3695.

Urban, S. and M. Freeman (2003). "Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain." Molecular Cell **11**(6): 1425-1434.

Uyetake, L., E. Ortega-Barria and J. C. Boothroyd (2001). "Isolation and characterization of a cold-sensitive attachment/invasion mutant of Toxoplasma gondii." Experimental Parasitology **97**(1): 55-59.

Wan, K. L., V. B. Carruthers, L. D. Sibley and J. W. Ajioka (1997). "Molecular characterisation of an expressed sequence tag locus of Toxoplasma gondii encoding the micronemal protein MIC2." Molecular and Biochemical

Parasitology **84**(2): 203-214.

Webster, J. P., C. F. A. Brunton and D. W. Macdonald (1994). "Effect of Toxoplasma-Gondii Upon Neophobic Behavior in Wild Brown-Rats, Rattus-Norvegicus." Parasitology **109**: 37-43.

Wengelnik, K., R. Spaccapelo, S. Naitza, K. J. H. Robson, C. J. Janse, F. Bistoni, A. P. Waters and A. Crisanti (1999). "The A-domain and the thrombospondin-related motif of Plasmodium falciparum TRAP are implicated in the invasion process of mosquito salivary glands." Embo Journal **18**(19): 5195-5204.

Wetzel, D. M., S. Hakansson, K. Hu, D. Roos and L. D. Sibley (2003). "Actin filament polymerization regulates gliding motility by apicomplexan parasites." Molecular Biology of the Cell **14**(2): 396-406.

Whittaker, C. A. and R. O. Hynes (2002). "Distribution and evolution of von Willebrand/integrin a domains: Widely dispersed adhesion and elsewhere." Molecular Biology of the Cell **13**(10): 3369-3387.

Wishart, D. S., C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes (1995). "H-1, C-13 and N-15 Random Coil Nmr Chemical-Shifts of the Common Amino-Acids .1. Investigations of Nearest-Neighbor Effects." Journal of Biomolecular Nmr **5**(1): 67-81.

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

Wishart, D. S. and B. D. Sykes (1994). "The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data." Journal of Biomolecular Nmr **4**(2): 171-180.

Wishart, D. S., B. D. Sykes and F. M. Richards (1992). "The Chemical-Shift Index - a Fast and Simple Method for the Assignment of Protein Secondary Structure through Nmr-Spectroscopy." Biochemistry **31**(6): 1647-1651.

Wright, H. T., G. Sandrasegaram and C. S. Wright (1991). "Evolution of a Family of

N-Acetylglucosamine Binding-Proteins Containing the Disulfide-Rich Domain of Wheat-Germ-Agglutinin." Journal of Molecular Evolution **33**(3): 283-294.

Zhou, X. W., M. J. Blackman, S. A. Howell and V. B. Carruthers (2004). "Proteomic analysis of cleavage events reveals a dynamic two-step mechanism for proteolysis of a key parasite adhesive complex." Molecular & Cellular Proteomics **3**(6): 565-576.

Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 269:592-610

Appendix

A1. Luria-Berani (LB) media

10 g tryptone, 5 g yeast extract and 5 g NaCl per litre.

A2. Minimal (M9) media

6 g Na2HP4, 3 g KH2PO4 and 0.5 g NaCl. Adjust pH to 7 – 7.4 and autoclave.

Filter-sterile (0.2 µm) and add the following: 2ml MgSO4, 10 µl CaCl2, 1ml Fe SO4 (0.01M), vitamin and micronutrient mixture, 0.7g 15NH_4Cl and 2g 13$C_6H_{12}O_6$.

A3. Protein purification buffers

All purification buffers are been filter-sterilised before use.

Component	Binding buffer	Wash buffer	Elution buffer
Tris-HCl (pH 7.9)	20mM		
NaCl	0.5mM		
Imidazole	5mM	20 mM	250 mM

A4. Primers

A4.1 M2AP PRIMERS FOR pET32 system

M2AP_F023: *GGT ATT GAG GGT CGC* AGG AAA GTT GGA AAT CCG GCG

M2AP-F38: *GGT ATT GAG GGT CGC* GAA CCG GTG GCC CTA GCT CAG C

M2AP-F47: *GGT ATT GAG GGT CGC* ACA TTC CTC GAG CTC GTC GAG G

M2AP-F53: *GGT ATT GAG GGT CGC* GAG GTG CCA TGT AAC TCT GTT C

M2AP_R330: *AGA GGA GAG TTA GAG CCT TA*C GCC TCA TCG TCA CTC GGC

M2AP_R228: *AGA GGA GAG TTA GAG CCT TA* A GCC TTG CAC CCC AAG GAA GG

A4.2 MIC2 PRIMERS FOR pET32 system

MIC2_F001: *GGT ATT GAG GGT CGC* ATG AGA CTC CAA CGC GAG GC

MIC2_F027: *GGT ATT GAG GGT CGC* GGA GGG TGG AGT ATT GTG G

MIC2_F271: *GGT ATT GAG GGT CGC* GCC ATT TGC TCG GAT TGG TC

MIC2_F073: *GGT ATT GAG GGT CGC* AAC CAG CTG GAC ATT TGC TTT C

MIC2_F067: *GGT ATT GAG GGT CGC* GCC GCA GAG GGG TGC ACC AAC

MIC2_F482: *GGT ATT GAG GGT CGC* TGC TCG GTC TCA TGT GGT GGT GG

MIC2_R270:*AGA GGA GAG TTA GAG CCT TA*A TCC TGG GGG AGT GTC TTA C

MIC2_R702:*AGA GGA GAG TTA GAG CCT TA*A GCC CCA GCT ATG CCA CTG C

MIC2_R265: *AGA GGA GAG TTA GAG CCT TA*C TTA CAA ACC TCT TTA AGC

MIC2_R228: *AGA GGA GAG TTA GAG CCT TA*C ATG CTT CGG CAC TCA CTA TG

MIC2_R232: *AGA GGA GAG TTA GAG CCT TA*A CTG CAC CCA CAC ATG CTT CGG

MIC2_R245: *AGA GGA GAG TTA GAG CCT TA*A AGA TAC AGG GGA CAC GGA CTG

A4.3 MIC2 and M2AP Co-expression design 1

MIC2_F027Co: *GAC GAC GAC AAG ATT* GGA GGG TGG AGT ATT GTG G

MIC2_R702Co: *CGC GGG CGG CCG T*AG CCC CAG CTA TGC CAC TGC

M2AP_F023Co: *GCG GGC CCG GCC T* AGG AAA GTT GGA AAT CCG GCG

M2AP_R330Co : *GAG GAG AAG CCC GGT* C GCC TCA TCG TCA CTC GGC

MIC2 and M2AP Co-expression design 2

M2AP_F023Co1: *GAC GAC GAC AAG ATT* AGG AAA GTT GGA AAT CCG GCG

M2AP_R330Co2: *CGC GGG CGG CCG T* C GCC TCA TCG TCA CTC GGC

MIC2_F027Co: *GCG GGC CCG GCC T* GGA GGG TGG AGT ATT GTG G

MIC2_R702Co: *GAG GAG AAG CCC GGT*AG CCC CAG CTA TGC CAC TGC

A4.4 MIC3 primers

MIC3F067: *GGT ATT GAG GGT CGC* TCC CCC AGC AAG CAG GAG ACG CAA C

MIC3F077: *GGT ATT GAG GGT CGC* GCT ATC TCG AGT GAA GGC AAG CCA TGT C

MIC3R144: *AGA GGA GAG TTA GAG CC TTA GTC* TCC TCC ATA GCT TTT GTC AG

MIC3R146: *AGA GGA GAG TTA GAG CC TTA GCT* GCA GTC TCC TCC ATA GCT TTT G

MIC3R189: *AGA GGA GAG TTA GAG CC TTA* CAA GGA TCC TCG GAG CAA GTC AAC CCA G

A5. Resonance assignment for 47M2APN

BMRB accession number 15961

Atom_shift_assign_ID

Residue_author_seq_code

Residue_seq_code

Residue_label

Atom_name

Atom_type

Chem_shift_value

Chem_shift_value_error

Chem_shift_ambiguity_code

1 1 1 THR HB H 4.005 . 1
2 2 2 PHE HA H 4.64 . 1
3 2 2 PHE HB2 H 3.063 . 2
4 2 2 PHE HB3 H 2.997 . 2
5 2 2 PHE HD1 H 7.24 . 3
6 2 2 PHE HD2 H 7.24 . 3
7 2 2 PHE HE1 H 7.21 . 3
8 2 2 PHE HE2 H 7.21 . 3
9 2 2 PHE C C 174.851 . 1
10 2 2 PHE CA C 57.831 . 1
11 2 2 PHE CB C 39.665 . 1
12 2 2 PHE CD1 C 132.8 . 3
13 2 2 PHE CD2 C 132.8 . 3
14 3 3 LEU H H 8.019 . 1
15 3 3 LEU HA H 4.257 . 1
16 3 3 LEU HB2 H 1.479 . 2
17 3 3 LEU HB3 H 1.479 . 2
18 3 3 LEU HD1 H 0.838 . 2
19 3 3 LEU HD2 H 0.785 . 2
20 3 3 LEU HG H 1.424 . 1
21 3 3 LEU C C 176.075 . 1
22 3 3 LEU CA C 54.846 . 1
23 3 3 LEU CB C 43.036 . 1
24 3 3 LEU CD1 C 24.864 . 1
25 3 3 LEU CD2 C 23.830 . 1
26 3 3 LEU CG C 27.203 . 1
27 3 3 LEU N N 124.428 . 1
28 4 4 GLU H H 8.188 . 1
29 4 4 GLU HA H 4.188 . 1
30 4 4 GLU HB2 H 2.0 . 2
31 4 4 GLU HB3 H 1.893 . 2
32 4 4 GLU HG2 H 2.157 . 2
33 4 4 GLU HG3 H 2.239 . 2
34 4 4 GLU C C 175.404 . 1
35 4 4 GLU CA C 56.274 . 1
36 4 4 GLU CB C 30.495 . 1
37 4 4 GLU CG C 36.176 . 1
38 4 4 GLU N N 122.557 . 1
39 5 5 LEU H H 8.230 . 1
40 5 5 LEU HA H 4.50 . 1
41 5 5 LEU HB2 H 1.554 . 2
42 5 5 LEU HB3 H 1.554 . 2
43 5 5 LEU HD1 H 0.491 . 2
44 5 5 LEU HD2 H 0.566 . 2
45 5 5 LEU HG H 1.475 . 1
46 5 5 LEU C C 176.120 . 1
47 5 5 LEU CA C 54.318 . 1
48 5 5 LEU CB C 43.8 . 1
49 5 5 LEU CD1 C 25.9 . 1
50 5 5 LEU CD2 C 23.65 . 1
51 5 5 LEU CG C 27.203 . 1

52 5 5 LEU N N 123.052 . 1
53 6 6 VAL H H 8.675 . 1
54 6 6 VAL HA H 4.451 . 1
55 6 6 VAL HB H 2.188 . 1
56 6 6 VAL HG1 H 1.020 . 2
57 6 6 VAL HG2 H 0.956 . 2
58 6 6 VAL C C 175.419 . 1
59 6 6 VAL CA C 60.816 . 1
60 6 6 VAL CB C 34.252 . 1
61 6 6 VAL CG1 C 21.403 . 1
62 6 6 VAL CG2 C 20.128 . 1
63 6 6 VAL N N 121.389 . 1
64 7 7 GLU H H 8.813 . 1
65 7 7 GLU HA H 4.282 . 1
66 7 7 GLU HB2 H 1.864 . 2
67 7 7 GLU HB3 H 1.864 . 2
68 7 7 GLU HG2 H 2.198 . 2
69 7 7 GLU HG3 H 2.198 . 2
70 7 7 GLU C C 176.255 . 1
71 7 7 GLU CA C 56.735 . 1
72 7 7 GLU CB C 29.999 . 1
73 7 7 GLU CG C 36.176 . 1
74 7 7 GLU N N 126.572 . 1
75 8 8 VAL H H 8.719 . 1
76 8 8 VAL HA H 4.5 . 1
77 8 8 VAL HB H 1.857 . 1
78 8 8 VAL HG1 H 0.895 . 2
79 8 8 VAL HG2 H 0.895 . 2
80 8 8 VAL C C 175.747 . 1
81 8 8 VAL CA C 59.174 . 1
82 8 8 VAL CB C 33.058 . 1
83 8 8 VAL CG1 C 21.062 . 1
84 8 8 VAL CG2 C 20.947 . 1
85 8 8 VAL N N 128.892 . 1
86 9 9 PRO HA H 4.90 . 1
87 9 9 PRO HB2 H 2.61 . 2
88 9 9 PRO HB3 H 1.823 . 2
89 9 9 PRO HD2 H 3.7 . 2
90 9 9 PRO HD3 H 3.852 . 2
91 9 9 PRO HG2 H 2.176 . 2
92 9 9 PRO HG3 H 2.013 . 2
93 9 9 PRO C C 175.449 . 1
94 9 9 PRO CA C 62.905 . 1
95 9 9 PRO CB C 28.5 . 1
96 9 9 PRO CD C 51.1 . 1
97 9 9 PRO CG C 27 . 1
98 10 10 CYS H H 7.909 . 1
99 10 10 CYS HA H 5.003 . 1
100 10 10 CYS HB2 H 3.333 . 2
101 10 10 CYS HB3 H 3.333 . 2
102 10 10 CYS C C 172.777 . 1
103 10 10 CYS CA C 54.542 . 1
104 10 10 CYS CB C 48.729 . 1
105 10 10 CYS N N 117.279 . 1
106 11 11 ASN H H 8.644 . 1
107 11 11 ASN HA H 4.685 . 1
108 11 11 ASN HB2 H 2.891 . 2
109 11 11 ASN HB3 H 2.891 . 2
110 11 11 ASN HD21 H 6.625 . 2
111 11 11 ASN HD22 H 6.625 . 2
112 11 11 ASN C C 174.613 . 1
113 11 11 ASN CA C 54.697 . 1
114 11 11 ASN CB C 39.401 . 1
115 11 11 ASN N N 115.501 . 1
116 11 11 ASN ND2 N 112.650 . 1
117 12 12 SER H H 7.497 . 1
118 12 12 SER HA H 4.62 . 1
119 12 12 SER HB2 H 3.818 . 2

120 12 12 SER HB3 H 3.646 . 2
121 12 12 SER C C 172.359 . 1
122 12 12 SER CA C 57.458 . 1
123 12 12 SER CB C 65.367 . 1
124 12 12 SER N N 110.920 . 1
125 13 13 VAL H H 8.996 . 1
126 13 13 VAL HA H 4.419 . 1
127 13 13 VAL HB H 2.206 . 1
128 13 13 VAL HG1 H 0.93 . 2
129 13 13 VAL HG2 H 0.93 . 2
130 13 13 VAL C C 174.344 . 1
131 13 13 VAL CA C 61.406 . 1
132 13 13 VAL CB C 33.506 . 1
133 13 13 VAL CG1 C 21.7 . 1
134 13 13 VAL CG2 C 21.7 . 1
135 13 13 VAL N N 119.681 . 1
136 14 14 HIS HA H 5.009 . 1
137 14 14 HIS HB2 H 3.178 . 2
138 14 14 HIS HB3 H 2.953 . 2
139 14 14 HIS HD2 H 7.379 . 3
140 14 14 HIS C C 173.956 . 1
141 14 14 HIS CA C 54.500 . 1
142 14 14 HIS CB C 29.749 . 1
143 15 15 VAL H H 8.829 . 1
144 15 15 VAL HA H 4.087 . 1
145 15 15 VAL HB H 2.375 . 1
146 15 15 VAL HG1 H 1.07 . 2
147 15 15 VAL HG2 H 0.875 . 2
148 15 15 VAL C C 174.986 . 1
149 15 15 VAL CA C 62.532 . 1
150 15 15 VAL CB C 32.634 . 1
151 15 15 VAL CG1 C 21.8 . 1
152 15 15 VAL CG2 C 22.3 . 1
153 15 15 VAL N N 125.456 . 1
154 16 16 GLN H H 8.400 . 1
155 16 16 GLN HA H 4.606 . 1
156 16 16 GLN HB2 H 1.981 . 2
157 16 16 GLN HB3 H 1.842 . 2
158 16 16 GLN HG2 H 2.247 . 2
159 16 16 GLN HG3 H 2.191 . 2
160 16 16 GLN C C 175.837 . 1
161 16 16 GLN CA C 55.518 . 1
162 16 16 GLN CB C 30.148 . 1
163 16 16 GLN CG C 35.907 . 1
164 16 16 GLN N N 125.016 . 1
165 16 16 GLN NE2 N 111.297 . 1
166 17 17 GLY H H 8.717 . 1
167 17 17 GLY HA2 H 4.248 . 2
168 17 17 GLY HA3 H 3.816 . 2
169 17 17 GLY C C 173.941 . 1
170 17 17 GLY CA C 44.324 . 1
171 17 17 GLY N N 112.678 . 1
172 18 18 VAL H H 8.462 . 1
173 18 18 VAL HA H 4.047 . 1
174 18 18 VAL HB H 1.893 . 1
175 18 18 VAL HG1 H 0.827 . 2
176 18 18 VAL HG2 H 0.827 . 2
177 18 18 VAL C C 175.479 . 1
178 18 18 VAL CA C 62.18 . 1
179 18 18 VAL CB C 33.6 . 1
180 18 18 VAL CG1 C 20.860 . 1
181 18 18 VAL CG2 C 20.860 . 1
182 18 18 VAL N N 120.705 . 1
183 19 19 MET H H 7.630 . 1
184 19 19 MET HA H 4.315 . 1
185 19 19 MET HB2 H 2.281 . 2
186 19 19 MET HB3 H 1.888 . 2
187 19 19 MET HE H 0.881 . 1

188 19 19 MET HG2 H 2.661 . 2
189 19 19 MET HG3 H 2.526 . 2
190 19 19 MET C C 174.419 . 1
191 19 19 MET CA C 56.786 . 1
192 19 19 MET CB C 34.103 . 1
193 19 19 MET CG C 32.625 . 1
194 19 19 MET N N 125.351 . 1
195 20 20 THR H H 7.658 . 1
196 20 20 THR HA H 4.593 . 1
197 20 20 THR HB H 4.43 . 1
198 20 20 THR HG2 H 1.218 . 1
199 20 20 THR C C 173.195 . 1
200 20 20 THR CA C 59.805 . 1
201 20 20 THR CB C 68.105 . 1
202 20 20 THR CG2 C 22.083 . 1
203 20 20 THR N N 115.892 . 1
204 21 21 PRO HA H 4.584 . 1
205 21 21 PRO HB2 H 2.29 . 2
206 21 21 PRO HB3 H 1.787 . 2
207 21 21 PRO HD2 H 3.77 . 2
208 21 21 PRO HD3 H 3.638 . 2
209 21 21 PRO HG2 H 1.75 . 2
210 21 21 PRO HG3 H 2.047 . 2
211 21 21 PRO C C 176.613 . 1
212 21 21 PRO CA C 64.900 . 1
213 21 21 PRO CB C 32.123 . 1
214 21 21 PRO CD C 51.342 . 1
215 21 21 PRO CG C 28 . 1
216 22 22 ASN H H 8.610 . 1
217 22 22 ASN HA H 4.349 . 1
218 22 22 ASN HB2 H 3.0 . 2
219 22 22 ASN HB3 H 2.8 . 2
220 22 22 ASN HD21 H 6.923 . 2
221 22 22 ASN HD22 H 6.923 . 2
222 22 22 ASN C C 174.270 . 1
223 22 22 ASN CA C 55.070 . 1
224 22 22 ASN CB C 37.833 . 1
225 22 22 ASN N N 113.766 . 1
226 22 22 ASN ND2 N 112.581 . 1
227 23 23 GLN H H 7.538 . 1
228 23 23 GLN HA H 4.516 . 1
229 23 23 GLN HB2 H 1.88 . 2
230 23 23 GLN HB3 H 2.096 . 2
231 23 23 GLN HG2 H 2.166 . 2
232 23 23 GLN HG3 H 2.250 . 2
233 23 23 GLN C C 175.374 . 1
234 23 23 GLN CA C 55.164 . 1
235 23 23 GLN CB C 30.122 . 1
236 23 23 GLN CG C 33.6 . 1
237 23 23 GLN N N 116.580 . 1
238 23 23 GLN NE2 N 111.297 . 1
239 24 24 MET H H 8.939 . 1
240 24 24 MET HA H 5.187 . 1
241 24 24 MET HB2 H 1.832 . 2
242 24 24 MET HB3 H 1.699 . 2
243 24 24 MET HE H 0.558 . 1
244 24 24 MET HG2 H 2.158 . 2
245 24 24 MET HG3 H 2.158 . 2
246 24 24 MET C C 173.912 . 1
247 24 24 MET CA C 55.019 . 1
248 24 24 MET CB C 37.251 . 1
249 24 24 MET CE C 23.56 . 1
250 24 24 MET CG C 31.271 . 1
251 24 24 MET N N 117.807 . 1
252 25 25 VAL H H 8.090 . 1
253 25 25 VAL HA H 4.543 . 1
254 25 25 VAL HB H 1.130 . 1
255 25 25 VAL HG1 H 0.536 . 2

256 25 25 VAL HG2 H 0.290 . 2
257 25 25 VAL C C 173.494 . 1
258 25 25 VAL CA C 60.937 . 1
259 25 25 VAL CB C 34.949 . 1
260 25 25 VAL CG1 C 20.733 . 1
261 25 25 VAL CG2 C 20.233 . 1
262 25 25 VAL N N 116.295 . 1
263 26 26 LYS H H 8.780 . 1
264 26 26 LYS HA H 5.009 . 1
265 26 26 LYS HB2 H 1.585 . 2
266 26 26 LYS HB3 H 1.427 . 2
267 26 26 LYS C C 173.628 . 1
268 26 26 LYS CA C 53.951 . 1
269 26 26 LYS CB C 37.610 . 1
270 26 26 LYS N N 126.778 . 1
271 27 27 VAL H H 9.340 . 1
272 27 27 VAL HA H 5.256 . 1
273 27 27 VAL HB H 2.277 . 1
274 27 27 VAL HG1 H 1.094 . 2
275 27 27 VAL HG2 H 1.094 . 2
276 27 27 VAL C C 175.359 . 1
277 27 27 VAL CA C 60.765 . 1
278 27 27 VAL CB C 35.471 . 1
279 27 27 VAL CG1 C 20.7 . 1
280 27 27 VAL CG2 C 20.7 . 1
281 27 27 VAL N N 126.465 . 1
282 28 28 THR H H 9.413 . 1
283 28 28 THR HA H 5.364 . 1
284 28 28 THR HB H 4.474 . 1
285 28 28 THR HG2 H 1.07 . 1
286 28 28 THR C C 171.300 . 1
287 28 28 THR CA C 59.920 . 1
288 28 28 THR CB C 71.584 . 1
289 28 28 THR CG2 C 22.204 . 1
290 28 28 THR N N 118.768 . 1
291 29 29 GLY H H 7.274 . 1
292 29 29 GLY HA2 H 4.416 . 2
293 29 29 GLY HA3 H 3.331 . 2
294 29 29 GLY C C 172.345 . 1
295 29 29 GLY CA C 45.468 . 1
296 29 29 GLY N N 103.721 . 1
297 30 30 ALA H H 9.017 . 1
298 30 30 ALA HA H 4.658 . 1
299 30 30 ALA HB H 1.297 . 1
300 30 30 ALA C C 174.926 . 1
301 30 30 ALA CA C 51.488 . 1
302 30 30 ALA CB C 19.128 . 1
303 30 30 ALA N N 122.308 . 1
304 31 31 GLY H H 7.353 . 1
305 31 31 GLY HA2 H 4.601 . 2
306 31 31 GLY HA3 H 4.371 . 2
307 31 31 GLY C C 172.553 . 1
308 31 31 GLY CA C 44.698 . 1
309 31 31 GLY N N 106.348 . 1
310 32 32 TRP H H 9.635 . 1
311 32 32 TRP HA H 5.278 . 1
312 32 32 TRP HB2 H 3.379 . 2
313 32 32 TRP HB3 H 3.246 . 2
314 32 32 TRP HD1 H 7.311 . 1
315 32 32 TRP HE1 H 9.530 . 3
316 32 32 TRP HH2 H 7.34 . 1
317 32 32 TRP HZ2 H 6.71 . 3
318 32 32 TRP HZ3 H 7.24 . 3
319 32 32 TRP C C 173.330 . 1
320 32 32 TRP CA C 58.377 . 1
321 32 32 TRP CB C 30.670 . 1
322 32 32 TRP CD1 C 129.5 . 3
323 32 32 TRP CH2 C 129.5 . 1

324 32 32 TRP N N 120.519 . 1

325 32 32 TRP NE1 N 129.1 . 1

326 33 33 ASP H H 8.515 . 1

327 33 33 ASP HA H 4.922 . 1

328 33 33 ASP HB2 H 2.646 . 2

329 33 33 ASP HB3 H 2.646 . 2

330 33 33 ASP C C 175.210 . 1

331 33 33 ASP CA C 56.1 . 1

332 33 33 ASP CB C 43.754 . 1

333 33 33 ASP N N 118.150 . 1

334 34 34 ASN H H 7.722 . 1

335 34 34 ASN HA H 4.639 . 1

336 34 34 ASN HB2 H 2.873 . 2

337 34 34 ASN HB3 H 2.873 . 2

338 34 34 ASN C C 174.508 . 1

339 34 34 ASN CA C 52.682 . 1

340 34 34 ASN CB C 41.191 . 1

341 34 34 ASN N N 113.186 . 1

342 34 34 ASN ND2 N 113.866 . 1

343 35 35 GLY H H 8.153 . 1

344 35 35 GLY HA2 H 4.202 . 2

345 35 35 GLY HA3 H 3.958 . 2

346 35 35 GLY C C 174.180 . 1

347 35 35 GLY CA C 46.489 . 1

348 35 35 GLY N N 106.757 . 1

349 36 36 VAL H H 8.238 . 1

350 36 36 VAL HA H 4.299 . 1

351 36 36 VAL HB H 1.676 . 1

352 36 36 VAL HG1 H 0.454 . 2

353 36 36 VAL HG2 H 0.415 . 2

354 36 36 VAL C C 174.911 . 1

355 36 36 VAL CA C 61.287 . 1

356 36 36 VAL CB C 34.109 . 1

357 36 36 VAL CG1 C 21.398 . 1

358 36 36 VAL CG2 C 20.771 . 1

359 36 36 VAL N N 121.650 . 1

360 37 37 LEU H H 7.974 . 1

361 37 37 LEU HA H 3.868 . 1

362 37 37 LEU HB2 H 0.509 . 2

363 37 37 LEU HB3 H 0.509 . 2

364 37 37 LEU HD1 H -0.15 . 2

365 37 37 LEU HD2 H -0.05 . 2

366 37 37 LEU HG H -1.0 . 1

367 37 37 LEU C C 174.046 . 1

368 37 37 LEU CA C 54.00 . 1

369 37 37 LEU CB C 42.3 . 1

370 37 37 LEU CD1 C 24.3 . 1

371 37 37 LEU CD2 C 24.8 . 1

372 37 37 LEU CG C 42.2 . 1

373 37 37 LEU N N 128.459 . 1

374 38 38 GLU H H 8.299 . 1

375 38 38 GLU HA H 5.44 . 1

376 38 38 GLU HB2 H 1.552 . 2

377 38 38 GLU HB3 H 1.552 . 2

378 38 38 GLU HG2 H 1.95 . 2

379 38 38 GLU HG3 H 1.956 . 2

380 38 38 GLU C C 174.225 . 1

381 38 38 GLU CA C 54.268 . 1

382 38 38 GLU CB C 32.567 . 1

383 38 38 GLU CG C 36.993 . 1

384 38 38 GLU N N 127.069 . 1

385 39 39 PHE H H 9.188 . 1

386 39 39 PHE HA H 4.953 . 1

387 39 39 PHE HB2 H 2.30 . 2

388 39 39 PHE HB3 H 1.777 . 2

389 39 39 PHE HD1 H 6.561 . 3

390 39 39 PHE HD2 H 6.561 . 3

391 39 39 PHE HE1 H 6.972 . 3

392 39 39 PHE HE2 H 6.972 . 3
393 39 39 PHE HZ H 7.30 . 1
394 39 39 PHE C C 173.762 . 1
395 39 39 PHE CA C 56.115 . 1
396 39 39 PHE CB C 41.564 . 1
397 39 39 PHE CD1 C 131.72 . 3
398 39 39 PHE CD2 C 131.72 . 3
399 39 39 PHE CE1 C 130.5 . 3
400 39 39 PHE CE2 C 130.5 . 3
401 39 39 PHE CZ C 130.5 . 1
402 39 39 PHE N N 129.999 . 1
403 40 40 TYR H H 7.849 . 1
404 40 40 TYR HA H 5.303 . 1
405 40 40 TYR HB2 H 2.611 . 2
406 40 40 TYR HB3 H 2.491 . 2
407 40 40 TYR HD1 H 6.35 . 3
408 40 40 TYR HD2 H 6.35 . 3
409 40 40 TYR HE1 H 6.31 . 3
410 40 40 TYR HE2 H 6.31 . 3
411 40 40 TYR C C 171.583 . 1
412 40 40 TYR CA C 55.000 . 1
413 40 40 TYR CB C 41.990 . 1
414 40 40 TYR CD1 C 133.92 . 3
415 40 40 TYR CD2 C 133.92 . 3
416 40 40 TYR CE1 C 118.5 . 3
417 40 40 TYR CE2 C 118.5 . 3
418 40 40 TYR N N 123.556 . 1
419 41 41 VAL H H 8.179 . 1
420 41 41 VAL HA H 4.927 . 1
421 41 41 VAL HB H 1.677 . 1
422 41 41 VAL HG1 H 0.61 . 2
423 41 41 VAL HG2 H 0.481 . 2
424 41 41 VAL C C 176.911 . 1
425 41 41 VAL CA C 59.448 . 1
426 41 41 VAL CB C 34.743 . 1
427 41 41 VAL CG1 C 21.666 . 1
428 41 41 VAL CG2 C 21.4 . 1
429 41 41 VAL N N 116.582 . 1
430 42 42 THR H H 9.064 . 1
431 42 42 THR HA H 4.922 . 1
432 42 42 THR HB H 3.903 . 1
433 42 42 THR HG2 H 0.708 . 1
434 42 42 THR C C 171.897 . 1
435 42 42 THR CA C 59.398 . 1
436 42 42 THR CB C 72.381 . 1
437 42 42 THR CG2 C 20.860 . 1
438 42 42 THR N N 118.511 . 1
439 43 43 ARG H H 8.594 . 1
440 43 43 ARG HA H 4.934 . 1
441 43 43 ARG HB2 H 1.867 . 2
442 43 43 ARG HB3 H 1.867 . 2
443 43 43 ARG HD2 H 3.095 . 2
444 43 43 ARG HD3 H 3.052 . 2
445 43 43 ARG HG2 H 1.483 . 2
446 43 43 ARG HG3 H 1.425 . 2
447 43 43 ARG C C 172.539 . 1
448 43 43 ARG CA C 53.279 . 1
449 43 43 ARG CB C 31.491 . 1
450 43 43 ARG CD C 43.538 . 1
451 43 43 ARG CG C 25.9 . 1
452 43 43 ARG N N 116.961 . 1
453 44 44 PRO HA H 4.684 . 1
454 44 44 PRO HB2 H 2.303 . 2
455 44 44 PRO HB3 H 1.900 . 2
456 44 44 PRO HD2 H 3.707 . 2
457 44 44 PRO HD3 H 3.707 . 2
458 44 44 PRO HG2 H 1.93 . 2
459 44 44 PRO HG3 H 2.055 . 2

460 44 44 PRO C C 176.508 . 1

461 44 44 PRO CA C 63.396 . 1

462 44 44 PRO CB C 32.139 . 1

463 44 44 PRO CD C 50.307 . 1

464 44 44 PRO CG C 28.2 . 1

465 45 45 THR H H 8.236 . 1

466 45 45 THR HA H 4.267 . 1

467 45 45 THR HB H 4.064 . 1

468 45 45 THR HG2 H 1.07 . 1

469 45 45 THR C C 174.494 . 1

470 45 45 THR CA C 62.018 . 1

471 45 45 THR CB C 69.919 . 1

472 45 45 THR CG2 C 21.777 . 1

473 45 45 THR N N 116.277 . 1

474 46 46 LYS H H 8.500 . 1

475 46 46 LYS HA H 4.438 . 1

476 46 46 LYS HB2 H 1.889 . 2

477 46 46 LYS HB3 H 1.777 . 2

478 46 46 LYS HD2 H 1.647 . 2

479 46 46 LYS HD3 H 1.655 . 2

480 46 46 LYS HE2 H 2.961 . 2

481 46 46 LYS HE3 H 2.982 . 2

482 46 46 LYS HG2 H 1.386 . 2

483 46 46 LYS HG3 H 1.386 . 2

484 46 46 LYS C C 176.732 . 1

485 46 46 LYS CA C 56.661 . 1

486 46 46 LYS CB C 33.482 . 1

487 46 46 LYS CD C 29.000 . 1

488 46 46 LYS CE C 41.908 . 1

489 46 46 LYS CG C 24.822 . 1

490 46 46 LYS N N 122.400 . 1

491 47 47 THR H H 8.010 . 1

492 47 47 THR HA H 4.315 . 1

493 47 47 THR HB H 4.200 . 1

494 47 47 THR HG2 H 1.163 . 1

495 47 47 THR C C 175.120 . 1

496 47 47 THR CA C 62.117 . 1

497 47 47 THR CB C 69.919 . 1

498 47 47 THR CG2 C 21.666 . 1

499 47 47 THR N N 113.888 . 1

500 48 48 GLY H H 8.435 . 1

501 48 48 GLY HA2 H 4.058 . 2

502 48 48 GLY HA3 H 3.857 . 2

503 48 48 GLY C C 174.688 . 1

504 48 48 GLY CA C 45.967 . 1

505 48 48 GLY N N 112.146 . 1

506 49 49 GLY H H 8.272 . 1

507 49 49 GLY HA2 H 3.946 . 2

508 49 49 GLY HA3 H 3.946 . 2

509 49 49 GLY C C 173.912 . 1

510 49 49 GLY CA C 45.519 . 1

511 49 49 GLY N N 109.240 . 1

512 50 50 ASP H H 8.149 . 1

513 50 50 ASP HA H 4.643 . 1

514 50 50 ASP HB2 H 2.672 . 2

515 50 50 ASP HB3 H 2.672 . 2

516 50 50 ASP C C 176.613 . 1

517 50 50 ASP CA C 54.548 . 1

518 50 50 ASP CB C 41.714 . 1

519 50 50 ASP N N 120.713 . 1

520 51 51 THR H H 8.096 . 1

521 51 51 THR HA H 4.387 . 1

522 51 51 THR HB H 4.335 . 1

523 51 51 THR HG2 H 1.163 . 1

524 51 51 THR C C 174.896 . 1

525 51 51 THR CA C 62.009 . 1

526 51 51 THR CB C 69.620 . 1

527 51 51 THR CG2 C 21.666 . 1

528 51 51 THR N N 113.264 . 1

529 52 52 SER H H 8.224 . 1

530 52 52 SER HA H 4.362 . 1

531 52 52 SER HB2 H 3.823 . 2

532 52 52 SER HB3 H 3.823 . 2

533 52 52 SER C C 174.225 . 1

534 52 52 SER CA C 58.888 . 1

535 52 52 SER CB C 63.875 . 1

536 52 52 SER N N 118.667 . 1

537 53 53 ARG H H 8.171 . 1

538 53 53 ARG HA H 4.684 . 1

539 53 53 ARG HB2 H 1.565 . 2

540 53 53 ARG HB3 H 1.565 . 2

541 53 53 ARG HD2 H 3.007 . 2

542 53 53 ARG HD3 H 3.007 . 2

543 53 53 ARG C C 175.717 . 1

544 53 53 ARG CA C 55.070 . 1

545 53 53 ARG CB C 32.162 . 1

546 53 53 ARG CD C 42.535 . 1

547 53 53 ARG CG C 27.130 . 1

548 53 53 ARG N N 121.395 . 1

549 54 54 SER H H 8.581 . 1

550 54 54 SER HA H 4.530 . 1

551 54 54 SER HB2 H 3.678 . 2

552 54 54 SER HB3 H 3.510 . 2

553 54 54 SER C C 173.718 . 1

554 54 54 SER CA C 57.234 . 1

555 54 54 SER CB C 65.069 . 1

556 54 54 SER N N 116.002 . 1

557 55 55 HIS HA H 4.00 . 1

558 55 55 HIS HB2 H 2.169 . 2

559 55 55 HIS HB3 H 2.169 . 2

560 55 55 HIS C C 173.628 . 1

561 55 55 HIS CA C 57.345 . 1

562 55 55 HIS CB C 28.730 . 1

563 56 56 LEU H H 8.640 . 1

564 56 56 LEU HA H 4.662 . 1

565 56 56 LEU HB2 H 1.543 . 2

566 56 56 LEU HB3 H 1.25 . 2

567 56 56 LEU HD1 H 0.824 . 2

568 56 56 LEU HD2 H 0.894 . 2

569 56 56 LEU C C 176.001 . 1

570 56 56 LEU CA C 56.338 . 1

571 56 56 LEU CB C 44.855 . 1

572 56 56 LEU CD1 C 24.801 . 1

573 56 56 LEU CD2 C 24.801 . 1

574 56 56 LEU CG C 27.309 . 1

575 56 56 LEU N N 123.071 . 1

576 57 57 ALA H H 7.781 . 1

577 57 57 ALA HA H 5.512 . 1

578 57 57 ALA HB H 1.423 . 1

579 57 57 ALA C C 175.822 . 1

580 57 57 ALA CA C 51.861 . 1

581 57 57 ALA CB C 22.961 . 1

582 57 57 ALA N N 117.806 . 1

583 58 58 SER H H 9.093 . 1

584 58 58 SER HA H 5.757 . 1

585 58 58 SER HB2 H 4.047 . 2

586 58 58 SER HB3 H 3.868 . 2

587 58 58 SER C C 172.927 . 1

588 58 58 SER CA C 57.855 . 1

589 58 58 SER CB C 67.555 . 1

590 58 58 SER N N 114.788 . 1

591 59 59 ILE H H 9.781 . 1

592 59 59 ILE HA H 5.258 . 1

593 59 59 ILE HB H 1.889 . 1

594 59 59 ILE HD1 H 0.950 . 1

595 59 59 ILE HG12 H 1.084 . 1

596 59 59 ILE HG13 H 1.084 . 1

597 59 59 ILE HG2 H 0.945 . 2

598 59 59 ILE C C 174.598 . 1

599 59 59 ILE CA C 60.073 . 1

600 59 59 ILE CB C 40.967 . 1

601 59 59 ILE CD1 C 13.860 . 1

602 59 59 ILE CG1 C 28.799 . 2

603 59 59 ILE CG2 C 18.415 . 1

604 59 59 ILE N N 123.353 . 1

605 60 60 MET H H 9.356 . 1

606 60 60 MET HA H 5.668 . 1

607 60 60 MET HB2 H 2.046 . 2

608 60 60 MET HB3 H 2.046 . 2

609 60 60 MET HG2 H 2.269 . 2

610 60 60 MET HG3 H 2.269 . 2

611 60 60 MET C C 176.016 . 1

612 60 60 MET CA C 53.460 . 1

613 60 60 MET CB C 36.176 . 1

614 60 60 MET CG C 32.704 . 1

615 60 60 MET N N 125.889 . 1

616 61 61 CYS H H 9.089 . 1

617 61 61 CYS HA H 5.717 . 1

618 61 61 CYS HB2 H 3.499 . 2

619 61 61 CYS HB3 H 2.393 . 2

620 61 61 CYS C C 172.225 . 1

621 61 61 CYS CA C 55.891 . 1

622 61 61 CYS CB C 45.876 . 1

623 61 61 CYS N N 121.442 . 1

624 62 62 TYR H H 8.904 . 1

625 62 62 TYR HA H 4.964 . 1

626 62 62 TYR HB2 H 2.635 . 2

627 62 62 TYR HB3 H 2.635 . 2

628 62 62 TYR HD1 H 7.32 . 3

629 62 62 TYR HD2 H 7.32 . 3

630 62 62 TYR HE1 H 6.68 . 3

631 62 62 TYR HE2 H 6.68 . 3

632 62 62 TYR C C 177.523 . 1

633 62 62 TYR CA C 56.313 . 1

634 62 62 TYR CB C 40.445 . 1

635 62 62 TYR CD1 C 134.92 . 3

636 62 62 TYR CD2 C 134.92 . 3

637 62 62 TYR CE1 C 117.5 . 3

638 62 62 TYR CE2 C 117.5 . 3

639 62 62 TYR N N 121.680 . 1

640 63 63 SER H H 9.503 . 1

641 63 63 SER HA H 4.411 . 1

642 63 63 SER HB2 H 3.938 . 2

643 63 63 SER HB3 H 3.946 . 2

644 63 63 SER C C 176.120 . 1

645 63 63 SER CA C 58.055 . 1

646 63 63 SER CB C 63.726 . 1

647 63 63 SER N N 118.820 . 1

648 64 64 LYS H H 8.914 . 1

649 64 64 LYS HA H 4.390 . 1

650 64 64 LYS HB2 H 2.046 . 2

651 64 64 LYS HB3 H 1.678 . 2

652 64 64 LYS C C 176.001 . 1

653 64 64 LYS CA C 57.532 . 1

654 64 64 LYS CB C 32.386 . 1

655 64 64 LYS N N 122.428 . 1

656 65 65 ASP H H 8.602 . 1

657 65 65 ASP HA H 4.961 . 1

658 65 65 ASP HB2 H 3.060 . 2

659 65 65 ASP HB3 H 3.060 . 2

660 65 65 ASP C C 174.762 . 1

661 65 65 ASP CA C 54.800 . 1

662 65 65 ASP CB C 42.660 . 1

663 65 65 ASP N N 118.860 . 1

664 66 66 ILE H H 7.094 . 1
665 66 66 ILE HA H 4.293 . 1
666 66 66 ILE HB H 1.643 . 1
667 66 66 ILE HD1 H 0.745 . 1
668 66 66 ILE HG12 H 0.849 . 1
669 66 66 ILE HG13 H 0.849 . 1
670 66 66 ILE HG2 H 0.601 . 2
671 66 66 ILE C C 175.389 . 1
672 66 66 ILE CA C 60.238 . 1
673 66 66 ILE CB C 39.848 . 1
674 66 66 ILE CD1 C 13.605 . 1
675 66 66 ILE CG1 C 27.488 . 2
676 66 66 ILE CG2 C 17.786 . 1
677 66 66 ILE N N 119.288 . 1
678 67 67 ASP H H 8.600 . 1
679 67 67 ASP HA H 4.461 . 1
680 67 67 ASP HB2 H 2.608 . 2
681 67 67 ASP HB3 H 2.608 . 2
682 67 67 ASP C C 177.135 . 1
683 67 67 ASP CA C 55.442 . 1
684 67 67 ASP CB C 41.340 . 1
685 67 67 ASP N N 127.996 . 1
686 68 68 GLY H H 8.645 . 1
687 68 68 GLY HA2 H 4.170 . 2
688 68 68 GLY HA3 H 3.678 . 2
689 68 68 GLY C C 174.673 . 1
690 68 68 GLY CA C 45.668 . 1
691 68 68 GLY N N 110.803 . 1
692 69 69 VAL H H 7.245 . 1
693 69 69 VAL HA H 4.438 . 1
694 69 69 VAL HB H 2.191 . 1
695 69 69 VAL HG1 H 0.915 . 2
696 69 69 VAL HG2 H 1.04 . 2
697 69 69 VAL C C 174.106 . 1
698 69 69 VAL CA C 60.568 . 1
699 69 69 VAL CB C 33.506 . 1
700 69 69 VAL CG1 C 21.577 . 1
701 69 69 VAL CG2 C 21.4 . 1
702 69 69 VAL N N 122.149 . 1
703 70 70 PRO HA H 4.438 . 1
704 70 70 PRO HB2 H 2.354 . 2
705 70 70 PRO HB3 H 1.833 . 2
706 70 70 PRO HD2 H 3.992 . 2
707 70 70 PRO HD3 H 4.08 . 2
708 70 70 PRO HG2 H 2.04 . 2
709 70 70 PRO HG3 H 2.063 . 2
710 70 70 PRO C C 178.941 . 1
711 70 70 PRO CA C 63.203 . 1
712 70 70 PRO CB C 32.231 . 1
713 70 70 PRO CD C 51.3 . 1
714 70 70 PRO CG C 27.20 . 1
715 71 71 SER H H 8.896 . 1
716 71 71 SER HA H 4.148 . 1
717 71 71 SER HB2 H 3.980 . 2
718 71 71 SER HB3 H 3.980 . 2
719 71 71 SER C C 175.568 . 1
720 71 71 SER CA C 61.200 . 1
721 71 71 SER CB C 62.756 . 1
722 71 71 SER N N 120.026 . 1
723 72 72 ASP H H 8.686 . 1
724 72 72 ASP HA H 4.444 . 1
725 72 72 ASP HB2 H 2.661 . 2
726 72 72 ASP HB3 H 2.661 . 2
727 72 72 ASP C C 175.852 . 1
728 72 72 ASP CA C 55.219 . 1
729 72 72 ASP CB C 39.550 . 1
730 72 72 ASP N N 117.611 . 1
731 73 73 LYS H H 7.420 . 1

732 73 73 LYS HA H 4.313 . 1

733 73 73 LYS HB2 H 0.907 . 2

734 73 73 LYS HB3 H 0.907 . 2

735 73 73 LYS HD2 H 0.66 . 2

736 73 73 LYS HD3 H 0.66 . 2

737 73 73 LYS HE2 H 2.37 . 2

738 73 73 LYS HE3 H 2.37 . 2

739 73 73 LYS HG2 H 0.79 . 2

740 73 73 LYS HG3 H 0.79 . 2

741 73 73 LYS C C 174.926 . 1

742 73 73 LYS CA C 56.264 . 1

743 73 73 LYS CB C 34.476 . 1

744 73 73 LYS CD C 28.473 . 1

745 73 73 LYS CE C 41.92 . 1

746 73 73 LYS CG C 24.8 . 1

747 73 73 LYS N N 116.540 . 1

748 74 74 ALA H H 7.121 . 1

749 74 74 ALA HA H 3.646 . 1

750 74 74 ALA HB H 1.420 . 1

751 74 74 ALA C C 178.418 . 1

752 74 74 ALA CA C 54.804 . 1

753 74 74 ALA CB C 19.427 . 1

754 74 74 ALA N N 119.554 . 1

755 75 75 GLY H H 7.772 . 1

756 75 75 GLY HA2 H 4.371 . 2

757 75 75 GLY HA3 H 4.181 . 2

758 75 75 GLY C C 173.300 . 1

759 75 75 GLY CA C 46.662 . 1

760 75 75 GLY N N 111.924 . 1

761 76 76 LYS H H 8.236 . 1

762 76 76 LYS HA H 4.751 . 1

763 76 76 LYS HB2 H 1.956 . 2

764 76 76 LYS HB3 H 1.420 . 2

765 76 76 LYS HD2 H 1.49 . 2

766 76 76 LYS HD3 H 1.61 . 2

767 76 76 LYS HE2 H 2.966 . 2

768 76 76 LYS HE3 H 3.03 . 2

769 76 76 LYS HG2 H 1.70 . 2

770 76 76 LYS HG3 H 1.66 . 2

771 76 76 LYS HZ H 3.03 . 1

772 76 76 LYS C C 174.986 . 1

773 76 76 LYS CA C 55.392 . 1

774 76 76 LYS CB C 37.386 . 1

775 76 76 LYS CD C 25.3 . 1

776 76 76 LYS CE C 41.7 . 1

777 76 76 LYS CG C 28.9 . 1

778 76 76 LYS N N 120.134 . 1

779 77 77 CYS H H 8.333 . 1

780 77 77 CYS HA H 6.238 . 1

781 77 77 CYS HB2 H 2.962 . 2

782 77 77 CYS HB3 H 2.582 . 2

783 77 77 CYS C C 172.580 . 1

784 77 77 CYS CA C 55.542 . 1

785 77 77 CYS CB C 50.817 . 1

786 77 77 CYS N N 115.633 . 1

787 78 78 PHE H H 9.423 . 1

788 78 78 PHE HA H 5.14 . 1

789 78 78 PHE HB2 H 2.83 . 2

790 78 78 PHE HB3 H 2.83 . 2

791 78 78 PHE HD1 H 7.150 . 3

792 78 78 PHE HD2 H 7.15 . 3

793 78 78 PHE HE1 H 7.11 . 3

794 78 78 PHE HE2 H 7.11 . 3

795 78 78 PHE C C 173.941 . 1

796 78 78 PHE CA C 56.910 . 1

797 78 78 PHE CB C 43.801 . 1

798 78 78 PHE CD1 C 131.72 . 3

799 78 78 PHE CD2 C 131.72 . 3

800 78 78 PHE N N 117.314 . 1
801 79 79 LEU H H 9.506 . 1
802 79 79 LEU HA H 5.346 . 1
803 79 79 LEU HB2 H 1.86 . 2
804 79 79 LEU HB3 H 1.25 . 2
805 79 79 LEU HD1 H 1.08 . 2
806 79 79 LEU HD2 H 1.08 . 2
807 79 79 LEU HG H 1.46 . 1
808 79 79 LEU C C 175.389 . 1
809 79 79 LEU CA C 53.876 . 1
810 79 79 LEU CB C 46.140 . 1
811 79 79 LEU CD1 C 22.293 . 1
812 79 79 LEU CD2 C 28.76 . 1
813 79 79 LEU CG C 27.98 . 1
814 79 79 LEU N N 125.001 . 1
815 80 80 LYS H H 9.579 . 1
816 80 80 LYS HA H 5.120 . 1
817 80 80 LYS HB2 H 1.618 . 2
818 80 80 LYS HB3 H 1.618 . 2
819 80 80 LYS HD2 H 1.460 . 2
820 80 80 LYS HD3 H 1.460 . 2
821 80 80 LYS HE3 H 2.709 . 2
822 80 80 LYS HG2 H 1.273 . 2
823 80 80 LYS HG3 H 1.273 . 2
824 80 80 LYS C C 174.598 . 1
825 80 80 LYS CA C 54.972 . 1
826 80 80 LYS CB C 35.774 . 1
827 80 80 LYS CD C 29.96 . 1
828 80 80 LYS CE C 41.7 . 1
829 80 80 LYS CG C 25.502 . 1
830 80 80 LYS N N 128.403 . 1
831 81 81 ASN H H 8.508 . 1
832 81 81 ASN HA H 5.127 . 1
833 81 81 ASN HB2 H 2.975 . 2
834 81 81 ASN HB3 H 2.975 . 2
835 81 81 ASN HD21 H 7.500 . 2
836 81 81 ASN HD22 H 7.500 . 2
837 81 81 ASN C C 173.643 . 1
838 81 81 ASN CA C 53.279 . 1
839 81 81 ASN CB C 42.00 . 1
840 81 81 ASN N N 121.296 . 1
841 81 81 ASN ND2 N 109.242 . 1
842 82 82 PHE H H 8.156 . 1
843 82 82 PHE HA H 4.927 . 1
844 82 82 PHE HB2 H 2.995 . 2
845 82 82 PHE HB3 H 2.874 . 2
846 82 82 PHE HD1 H 7.02 . 3
847 82 82 PHE HD2 H 7.02 . 3
848 82 82 PHE HE1 H 7.02 . 3
849 82 82 PHE HE2 H 7.02 . 3
850 82 82 PHE C C 175.732 . 1
851 82 82 PHE CA C 57.383 . 1
852 82 82 PHE CB C 40.669 . 1
853 82 82 PHE CD1 C 131.72 . 3
854 82 82 PHE CD2 C 131.72 . 3
855 82 82 PHE CE1 C 131.72 . 3
856 82 82 PHE CE2 C 131.72 . 3
857 82 82 PHE N N 120.713 . 1
858 83 83 SER H H 8.831 . 1
859 83 83 SER HA H 4.627 . 1
860 83 83 SER HB2 H 3.92 . 2
861 83 83 SER HB3 H 4.08 . 2
862 83 83 SER C C 174.941 . 1
863 83 83 SER CA C 58.428 . 1
864 83 83 SER CB C 63.94 . 1
865 83 83 SER N N 116.911 . 1
866 84 84 GLY H H 8.351 . 1
867 84 84 GLY HA2 H 4.161 . 2

868 84 84 GLY HA3 H 3.968 . 2
869 84 84 GLY C C 174.732 . 1
870 84 84 GLY CA C 45.594 . 1
871 84 84 GLY N N 110.308 . 1
872 85 85 GLU H H 8.533 . 1
873 85 85 GLU HA H 4.200 . 1
874 85 85 GLU HB2 H 2.060 . 2
875 85 85 GLU HB3 H 1.957 . 2
876 85 85 GLU HG2 H 2.237 . 2
877 85 85 GLU HG3 H 2.242 . 2
878 85 85 GLU C C 176.822 . 1
879 85 85 GLU CA C 57.780 . 1
880 85 85 GLU CB C 29.999 . 1
881 85 85 GLU CG C 36.176 . 1
882 85 85 GLU N N 120.362 . 1
883 86 86 ASP H H 8.232 . 1
884 86 86 ASP HA H 4.697 . 1
885 86 86 ASP HB2 H 2.758 . 2
886 86 86 ASP HB3 H 2.758 . 2
887 86 86 ASP C C 176.210 . 1
888 86 86 ASP CA C 53.876 . 1
889 86 86 ASP CB C 41.042 . 1
890 86 86 ASP N N 118.213 . 1
891 87 87 SER H H 8.013 . 1
892 87 87 SER HA H 4.19 . 1
893 87 87 SER HB2 H 3.97 . 2
894 87 87 SER HB3 H 3.970 . 2
895 87 87 SER C C 174.180 . 1
896 87 87 SER CA C 59.270 . 1
897 87 87 SER CB C 62.572 . 1
898 87 87 SER N N 113.535 . 1
899 88 88 SER H H 8.092 . 1
900 88 88 SER HA H 4.42 . 1
901 88 88 SER HB2 H 3.969 . 2
902 88 88 SER HB3 H 3.841 . 2
903 88 88 SER C C 174.494 . 1
904 88 88 SER CA C 59.123 . 1
905 88 88 SER CB C 64.173 . 1
906 88 88 SER N N 114.923 . 1
907 89 89 GLU H H 8.324 . 1
908 89 89 GLU HA H 4.294 . 1
909 89 89 GLU HB2 H 1.880 . 2
910 89 89 GLU HB3 H 1.889 . 2
911 89 89 GLU HG2 H 2.132 . 2
912 89 89 GLU HG3 H 2.132 . 2
913 89 89 GLU C C 175.643 . 1
914 89 89 GLU CA C 56.592 . 1
915 89 89 GLU CB C 29.729 . 1
916 89 89 GLU CG C 35.997 . 1
917 89 89 GLU N N 123.265 . 1
918 90 90 ILE H H 8.179 . 1
919 90 90 ILE HA H 4.304 . 1
920 90 90 ILE HB H 1.616 . 1
921 90 90 ILE HD1 H 0.742 . 1
922 90 90 ILE HG12 H 1.323 . 1
923 90 90 ILE HG13 H 1.323 . 1
924 90 90 ILE HG2 H 0.585 . 2
925 90 90 ILE C C 175.419 . 1
926 90 90 ILE CA C 60.840 . 1
927 90 90 ILE CB C 39.973 . 1
928 90 90 ILE CD1 C 13.695 . 1
929 90 90 ILE CG1 C 27.399 . 2
930 90 90 ILE CG2 C 17.746 . 1
931 90 90 ILE N N 123.140 . 1
932 91 91 ASP H H 8.118 . 1
933 91 91 ASP HA H 4.773 . 1
934 91 91 ASP HB2 H 2.657 . 2
935 91 91 ASP HB3 H 2.537 . 2

936 91 91 ASP C C 174.329 . 1
937 91 91 ASP CA C 53.603 . 1
938 91 91 ASP CB C 43.182 . 1
939 91 91 ASP N N 124.945 . 1
940 92 92 GLU H H 8.310 . 1
941 92 92 GLU HA H 5.203 . 1
942 92 92 GLU HB2 H 1.883 . 2
943 92 92 GLU HB3 H 1.770 . 2
944 92 92 GLU HG2 H 1.988 . 2
945 92 92 GLU HG3 H 1.988 . 2
946 92 92 GLU C C 174.852 . 1
947 92 92 GLU CA C 55.219 . 1
948 92 92 GLU CB C 32.983 . 1
949 92 92 GLU CG C 36.982 . 1
950 92 92 GLU N N 122.401 . 1
951 93 93 LYS H H 8.804 . 1
952 93 93 LYS HA H 4.883 . 1
953 93 93 LYS HB2 H 1.912 . 2
954 93 93 LYS HB3 H 1.800 . 2
955 93 93 LYS HD2 H 1.7 . 2
956 93 93 LYS HD3 H 1.7 . 2
957 93 93 LYS HE2 H 2.995 . 2
958 93 93 LYS HE3 H 2.995 . 2
959 93 93 LYS HG2 H 1.501 . 2
960 93 93 LYS HG3 H 1.41 . 2
961 93 93 LYS C C 174.762 . 1
962 93 93 LYS CA C 54.795 . 1
963 93 93 LYS CB C 35.669 . 1
964 93 93 LYS CD C 29.16 . 1
965 93 93 LYS CE C 41.99 . 1
966 93 93 LYS CG C 23.758 . 1
967 93 93 LYS N N 123.294 . 1
968 94 94 GLU H H 8.960 . 1
969 94 94 GLU HA H 4.074 . 1
970 94 94 GLU HB2 H 1.781 . 2
971 94 94 GLU HB3 H 1.805 . 2
972 94 94 GLU HG2 H 1.923 . 2
973 94 94 GLU HG3 H 1.923 . 2
974 94 94 GLU C C 176.150 . 1
975 94 94 GLU CA C 57.258 . 1
976 94 94 GLU CB C 29.727 . 1
977 94 94 GLU CG C 36.28 . 1
978 94 94 GLU N N 126.272 . 1
979 95 95 VAL H H 8.140 . 1
980 95 95 VAL HA H 4.627 . 1
981 95 95 VAL HB H 1.933 . 1
982 95 95 VAL HG1 H 0.65 . 2
983 95 95 VAL HG2 H 0.65 . 2
984 95 95 VAL C C 174.673 . 1
985 95 95 VAL CA C 59.249 . 1
986 95 95 VAL CB C 35.147 . 1
987 95 95 VAL CG1 C 21.5 . 1
988 95 95 VAL CG2 C 21.5 . 1
989 95 95 VAL N N 119.312 . 1
990 96 96 SER H H 8.469 . 1
991 96 96 SER HA H 4.461 . 1
992 96 96 SER HB2 H 3.756 . 2
993 96 96 SER HB3 H 3.673 . 2
994 96 96 SER C C 172.897 . 1
995 96 96 SER CA C 57.581 . 1
996 96 96 SER CB C 62.397 . 1
997 96 96 SER N N 119.167 . 1
998 97 97 LEU H H 8.210 . 1
999 97 97 LEU HA H 4.7 . 1
1000 97 97 LEU HB2 H 1.81 . 2
1001 97 97 LEU HB3 H 1.14 . 2
1002 97 97 LEU HD1 H 0.852 . 2
1003 97 97 LEU HD2 H 0.852 . 2

1004 97 97 LEU HG H 1.48 . 1
1005 97 97 LEU C C 175.896 . 1
1006 97 97 LEU CA C 51.787 . 1
1007 97 97 LEU CB C 44.8 . 1
1008 97 97 LEU CD1 C 25.1 . 1
1009 97 97 LEU CD2 C 25.1 . 1
1010 97 97 LEU CG C 27.3 . 1
1011 97 97 LEU N N 125.505 . 1
1012 98 98 PRO HA H 4.435 . 1
1013 98 98 PRO HB2 H 1.86 . 2
1014 98 98 PRO HB3 H 2.37 . 2
1015 98 98 PRO HD2 H 3.76 . 2
1016 98 98 PRO HD3 H 3.76 . 2
1017 98 98 PRO HG2 H 2.078 . 2
1018 98 98 PRO HG3 H 2.034 . 2
1019 98 98 PRO C C 175.255 . 1
1020 98 98 PRO CA C 63.383 . 1
1021 98 98 PRO CB C 32.139 . 1
1022 98 98 PRO CD C 50.3 . 1
1023 98 98 PRO CG C 27.219 . 1
1024 99 99 ILE H H 5.743 . 1
1025 99 99 ILE HA H 3.85 . 1
1026 99 99 ILE HB H 1.658 . 1
1027 99 99 ILE HD1 H 0.562 . 1
1028 99 99 ILE HG12 H 1.048 . 1
1029 99 99 ILE HG13 H 1.048 . 1
1030 99 99 ILE HG2 H 1.0 . 2
1031 99 99 ILE C C 174.270 . 1
1032 99 99 ILE CA C 62.419 . 1
1033 99 99 ILE CB C 39.4 . 1
1034 99 99 ILE CD1 C 15.26 . 1
1035 99 99 ILE CG1 C 28.781 . 2
1036 99 99 ILE CG2 C 17.122 . 1
1037 99 99 ILE N N 118.729 . 1
1038 100 100 LYS H H 9.142 . 1
1039 100 100 LYS HA H 4.776 . 1
1040 100 100 LYS HB2 H 1.879 . 2
1041 100 100 LYS HB3 H 1.82 . 2
1042 100 100 LYS HD2 H 1.60 . 2
1043 100 100 LYS HD3 H 1.50 . 2
1044 100 100 LYS HE2 H 2.97 . 2
1045 100 100 LYS HE3 H 3.01 . 2
1046 100 100 LYS HG2 H 1.66 . 2
1047 100 100 LYS HG3 H 1.71 . 2
1048 100 100 LYS C C 175.270 . 1
1049 100 100 LYS CA C 55.791 . 1
1050 100 100 LYS CB C 35.853 . 1
1051 100 100 LYS CG C 29 . 1
1052 100 100 LYS N N 125.316 . 1
1053 101 101 SER H H 8.002 . 1
1054 101 101 SER HA H 4.932 . 1
1055 101 101 SER HB2 H 3.418 . 2
1056 101 101 SER HB3 H 3.63 . 2
1057 101 101 SER C C 173.479 . 1
1058 101 101 SER CA C 56.040 . 1
1059 101 101 SER CB C 64.845 . 1
1060 101 101 SER N N 112.093 . 1
1061 102 102 HIS C C 173.971 . 1
1062 102 102 HIS CA C 56.413 . 1
1063 102 102 HIS CB C 26.790 . 1
1064 103 103 ASN H H 7.568 . 1
1065 103 103 ASN HA H 4.569 . 1
1066 103 103 ASN HB2 H 2.257 . 2
1067 103 103 ASN HB3 H 2.71 . 2
1068 103 103 ASN C C 174.598 . 1
1069 103 103 ASN CA C 52.983 . 1
1070 103 103 ASN CB C 39.251 . 1
1071 103 103 ASN N N 115.814 . 1

1072 104 104 ASP H H 7.410 . 1
1073 104 104 ASP HA H 4.81 . 1
1074 104 104 ASP HB2 H 2.78 . 2
1075 104 104 ASP HB3 H 3.05 . 2
1076 104 104 ASP C C 175.180 . 1
1077 104 104 ASP CA C 54.846 . 1
1078 104 104 ASP CB C 41.564 . 1
1079 104 104 ASP N N 121.759 . 1
1080 105 105 ALA H H 7.769 . 1
1081 105 105 ALA HA H 4.55 . 1
1082 105 105 ALA HB H 1.479 . 1
1083 105 105 ALA C C 176.090 . 1
1084 105 105 ALA CA C 51.450 . 1
1085 105 105 ALA CB C 20.200 . 1
1086 105 105 ALA N N 120.909 . 1
1087 106 106 PHE H H 8.315 . 1
1088 106 106 PHE HA H 6.2 . 1
1089 106 106 PHE HB2 H 3.114 . 2
1090 106 106 PHE HB3 H 3.04 . 2
1091 106 106 PHE HD1 H 7.36 . 3
1092 106 106 PHE HD2 H 7.36 . 3
1093 106 106 PHE HE1 H 7.308 . 3
1094 106 106 PHE HE2 H 7.308 . 3
1095 106 106 PHE HZ H 6.614 . 1
1096 106 106 PHE C C 175.016 . 1
1097 106 106 PHE CA C 56.089 . 1
1098 106 106 PHE CB C 43.280 . 1
1099 106 106 PHE CD1 C 132.8 . 3
1100 106 106 PHE CD2 C 132.8 . 3
1101 106 106 PHE CE1 C 133.8 . 3
1102 106 106 PHE CZ C 129.5 . 1
1103 106 106 PHE N N 114.099 . 1
1104 107 107 MET H H 8.874 . 1
1105 107 107 MET HA H 5.380 . 1
1106 107 107 MET HB2 H 1.815 . 2
1107 107 107 MET HB3 H 1.441 . 2
1108 107 107 MET HE H 0.148 . 1
1109 107 107 MET HG2 H 2.22 . 2
1110 107 107 MET HG3 H 1.96 . 2
1111 107 107 MET C C 173.732 . 1
1112 107 107 MET CA C 55.294 . 1
1113 107 107 MET CB C 38.8 . 1
1114 107 107 MET CE C 14.5 . 1
1115 107 107 MET CG C 31.4 . 1
1116 107 107 MET N N 115.099 . 1
1117 108 108 PHE H H 9.609 . 1
1118 108 108 PHE HA H 5.002 . 1
1119 108 108 PHE HB2 H 2.8 . 2
1120 108 108 PHE HB3 H 2.26 . 2
1121 108 108 PHE HD1 H 6.68 . 3
1122 108 108 PHE HD2 H 6.68 . 3
1123 108 108 PHE HE1 H 7.28 . 3
1124 108 108 PHE HE2 H 7.28 . 3
1125 108 108 PHE HZ H 7.22 . 1
1126 108 108 PHE C C 175.046 . 1
1127 108 108 PHE CA C 57.398 . 1
1128 108 108 PHE CB C 41.963 . 1
1129 108 108 PHE CD1 C 132.8 . 3
1130 108 108 PHE CD2 C 132.8 . 3
1131 108 108 PHE CE1 C 130.5 . 3
1132 108 108 PHE CE2 C 130.5 . 3
1133 108 108 PHE CZ C 132.8 . 1
1134 108 108 PHE N N 122.006 . 1
1135 109 109 VAL H H 9.406 . 1
1136 109 109 VAL HA H 5.012 . 1
1137 109 109 VAL HB H 2.336 . 1
1138 109 109 VAL HG1 H 0.79 . 2
1139 109 109 VAL HG2 H 0.919 . 2

1140 109 109 VAL C C 176.807 . 1
1141 109 109 VAL CA C 61.8 . 1
1142 109 109 VAL CB C 33.580 . 1
1143 109 109 VAL CG1 C 20.9 . 1
1144 109 109 VAL CG2 C 22.08 . 1
1145 109 109 VAL N N 123.891 . 1
1146 110 110 CYS H H 9.345 . 1
1147 110 110 CYS HA H 5.914 . 1
1148 110 110 CYS HB2 H 2.916 . 2
1149 110 110 CYS HB3 H 2.6 . 2
1150 110 110 CYS C C 172.076 . 1
1151 110 110 CYS CA C 55.791 . 1
1152 110 110 CYS CB C 48.280 . 1
1153 110 110 CYS N N 126.465 . 1
1154 111 111 SER H H 9.180 . 1
1155 111 111 SER HA H 4.638 . 1
1156 111 111 SER HB2 H 3.82 . 2
1157 111 111 SER HB3 H 3.82 . 2
1158 111 111 SER C C 171.986 . 1
1159 111 111 SER CA C 57.607 . 1
1160 111 111 SER CB C 66.263 . 1
1161 111 111 SER N N 114.252 . 1
1162 112 112 SER H H 9.305 . 1
1163 112 112 SER HA H 5.248 . 1
1164 112 112 SER HB2 H 3.512 . 2
1165 112 112 SER HB3 H 3.449 . 2
1166 112 112 SER C C 175.613 . 1
1167 112 112 SER CA C 56.115 . 1
1168 112 112 SER CB C 65.316 . 1
1169 112 112 SER N N 114.194 . 1
1170 113 113 ASN H H 8.913 . 1
1171 113 113 ASN HA H 4.756 . 1
1172 113 113 ASN HB2 H 3.004 . 2
1173 113 113 ASN HB3 H 2.806 . 2
1174 113 113 ASN HD21 H 6.830 . 2
1175 113 113 ASN HD22 H 6.830 . 2
1176 113 113 ASN C C 176.523 . 1
1177 113 113 ASN CA C 54.005 . 1
1178 113 113 ASN CB C 38.580 . 1
1179 113 113 ASN N N 124.943 . 1
1180 113 113 ASN ND2 N 111.297 . 1
1181 114 114 ASP H H 8.249 . 1
1182 114 114 ASP HA H 4.61 . 1
1183 114 114 ASP HB2 H 2.91 . 2
1184 114 114 ASP HB3 H 2.595 . 2
1185 114 114 ASP C C 177.195 . 1
1186 114 114 ASP CA C 54.298 . 1
1187 114 114 ASP CB C 41.340 . 1
1188 114 114 ASP N N 116.477 . 1
1189 115 115 GLY H H 8.034 . 1
1190 115 115 GLY HA2 H 4.107 . 2
1191 115 115 GLY HA3 H 3.802 . 2
1192 115 115 GLY C C 173.568 . 1
1193 115 115 GLY CA C 45.892 . 1
1194 115 115 GLY N N 109.844 . 1
1195 116 116 SER H H 8.378 . 1
1196 116 116 SER HA H 4.481 . 1
1197 116 116 SER HB2 H 3.889 . 2
1198 116 116 SER HB3 H 3.889 . 2
1199 116 116 SER C C 173.106 . 1
1200 116 116 SER CA C 59.547 . 1
1201 116 116 SER CB C 63.950 . 1
1202 116 116 SER N N 114.224 . 1
1203 117 117 ALA H H 8.867 . 1
1204 117 117 ALA HA H 4.747 . 1
1205 117 117 ALA HB H 1.490 . 1
1206 117 117 ALA C C 177.225 . 1
1207 117 117 ALA CA C 51.960 . 1

1208 117 117 ALA CB C 21.094 . 1

1209 117 117 ALA N N 125.138 . 1

1210 118 118 LEU H H 8.433 . 1

1211 118 118 LEU HA H 4.84 . 1

1212 118 118 LEU HB2 H 1.608 . 2

1213 118 118 LEU HB3 H 1.46 . 2

1214 118 118 LEU HD1 H 0.893 . 2

1215 118 118 LEU HD2 H 0.893 . 2

1216 118 118 LEU HG H 1.57 . 1

1217 118 118 LEU C C 175.255 . 1

1218 118 118 LEU CA C 54.821 . 1

1219 118 118 LEU CB C 45.270 . 1

1220 118 118 LEU CD1 C 25.2 . 1

1221 118 118 LEU CD2 C 25.2 . 1

1222 118 118 LEU CG C 27.509 . 1

1223 118 118 LEU N N 121.989 . 1

1224 119 119 GLN H H 8.861 . 1

1225 119 119 GLN HA H 5.258 . 1

1226 119 119 GLN HB2 H 2.018 . 2

1227 119 119 GLN HB3 H 2.018 . 2

1228 119 119 GLN HG2 H 2.326 . 2

1229 119 119 GLN HG3 H 2.255 . 2

1230 119 119 GLN C C 174.688 . 1

1231 119 119 GLN CA C 54.473 . 1

1232 119 119 GLN CB C 32.661 . 1

1233 119 119 GLN CG C 33.400 . 1

1234 119 119 GLN N N 122.289 . 1

1235 120 120 CYS H H 8.971 . 1

1236 120 120 CYS HA H 5.681 . 1

1237 120 120 CYS HB2 H 2.759 . 2

1238 120 120 CYS HB3 H 2.562 . 2

1239 120 120 CYS C C 172.718 . 1

1240 120 120 CYS CA C 55.692 . 1

1241 120 120 CYS CB C 49.548 . 1

1242 120 120 CYS N N 115.343 . 1

1243 121 121 ASP H H 9.375 . 1

1244 121 121 ASP HA H 4.963 . 1

1245 121 121 ASP HB2 H 2.565 . 2

1246 121 121 ASP HB3 H 1.996 . 2

1247 121 121 ASP C C 174.523 . 1

1248 121 121 ASP CA C 53.876 . 1

1249 121 121 ASP CB C 42.161 . 1

1250 121 121 ASP N N 125.683 . 1

1251 122 122 VAL H H 8.816 . 1

1252 122 122 VAL HA H 4.64 . 1

1253 122 122 VAL HB H 1.795 . 1

1254 122 122 VAL HG1 H 0.371 . 2

1255 122 122 VAL HG2 H 0.655 . 2

1256 122 122 VAL C C 173.016 . 1

1257 122 122 VAL CA C 61.6 . 1

1258 122 122 VAL CB C 32.8 . 1

1259 122 122 VAL CG1 C 20.97 . 1

1260 122 122 VAL CG2 C 21.58 . 1

1261 122 122 VAL N N 124.277 . 1

1262 123 123 PHE H H 9.304 . 1

1263 123 123 PHE HA H 5.943 . 1

1264 123 123 PHE HB2 H 3.162 . 2

1265 123 123 PHE HB3 H 2.617 . 2

1266 123 123 PHE HD1 H 6.866 . 3

1267 123 123 PHE HD2 H 6.866 . 3

1268 123 123 PHE HE1 H 7.013 . 3

1269 123 123 PHE HE2 H 7.013 . 3

1270 123 123 PHE C C 176.583 . 1

1271 123 123 PHE CA C 55.019 . 1

1272 123 123 PHE CB C 43.341 . 1

1273 123 123 PHE CD1 C 132.7 . 3

1274 123 123 PHE CD2 C 132.7 . 3

1275 123 123 PHE CE1 C 132.7 . 3

1276 123 123 PHE CE2 C 132.7 . 3
1277 123 123 PHE N N 124.876 . 1
1278 124 124 ALA H H 9.154 . 1
1279 124 124 ALA HA H 5.789 . 1
1280 124 124 ALA HB H 1.332 . 1
1281 124 124 ALA C C 175.553 . 1
1282 124 124 ALA CA C 51.9 . 1
1283 124 124 ALA CB C 23.971 . 1
1284 124 124 ALA N N 119.636 . 1
1285 125 125 LEU H H 8.102 . 1
1286 125 125 LEU HA H 4.058 . 1
1287 125 125 LEU HB2 H 0.7 . 2
1288 125 125 LEU HB3 H 0.7 . 2
1289 125 125 LEU HD1 H 0.2 . 2
1290 125 125 LEU HD2 H -0.03 . 2
1291 125 125 LEU HG H -0.762 . 1
1292 125 125 LEU C C 174.553 . 1
1293 125 125 LEU CA C 53.627 . 1
1294 125 125 LEU CB C 43.728 . 1
1295 125 125 LEU CD1 C 24.980 . 1
1296 125 125 LEU CD2 C 22.752 . 1
1297 125 125 LEU CG C 43.7 . 1
1298 125 125 LEU N N 125.444 . 1
1299 126 126 ASP H H 8.130 . 1
1300 126 126 ASP HA H 4.638 . 1
1301 126 126 ASP HB2 H 2.759 . 2
1302 126 126 ASP HB3 H 2.641 . 2
1303 126 126 ASP C C 176.807 . 1
1304 126 126 ASP CA C 53.354 . 1
1305 126 126 ASP CB C 41.340 . 1
1306 126 126 ASP N N 125.911 . 1
1307 127 127 ASN H H 8.594 . 1
1308 127 127 ASN HA H 4.500 . 1
1309 127 127 ASN HB2 H 2.769 . 2
1310 127 127 ASN HB3 H 2.769 . 2
1311 127 127 ASN HD21 H 6.612 . 2
1312 127 127 ASN HD22 H 6.612 . 2
1313 127 127 ASN C C 176.434 . 1
1314 127 127 ASN CA C 55.145 . 1
1315 127 127 ASN CB C 39.251 . 1
1316 127 127 ASN N N 123.715 . 1
1317 127 127 ASN ND2 N 112.605 . 1
1318 128 128 THR H H 8.496 . 1
1319 128 128 THR HA H 4.232 . 1
1320 128 128 THR HB H 4.235 . 1
1321 128 128 THR HG2 H 1.155 . 1
1322 128 128 THR C C 174.673 . 1
1323 128 128 THR CA C 63.303 . 1
1324 128 128 THR CB C 69.397 . 1
1325 128 128 THR CG2 C 21.846 . 1
1326 128 128 THR N N 112.323 . 1
1327 129 129 ASN H H 7.885 . 1
1328 129 129 ASN HA H 4.815 . 1
1329 129 129 ASN HB2 H 2.794 . 2
1330 129 129 ASN HB3 H 2.582 . 2
1331 129 129 ASN HD21 H 7.532 . 2
1332 129 129 ASN HD22 H 7.532 . 2
1333 129 129 ASN C C 175.449 . 1
1334 129 129 ASN CA C 52.906 . 1
1335 129 129 ASN CB C 38.953 . 1
1336 129 129 ASN N N 120.909 . 1
1337 129 129 ASN ND2 N 112.067 . 1
1338 130 130 SER H H 8.243 . 1
1339 130 130 SER HA H 4.195 . 1
1340 130 130 SER HB2 H 3.870 . 2
1341 130 130 SER HB3 H 3.811 . 2
1342 130 130 SER C C 176.120 . 1
1343 130 130 SER CA C 59.864 . 1

1344 130 130 SER CB C 63.353 . 1

1345 130 130 SER N N 117.077 . 1

1346 131 131 SER H H 8.303 . 1

1347 131 131 SER HA H 4.304 . 1

1348 131 131 SER HB2 H 3.930 . 2

1349 131 131 SER HB3 H 3.88 . 2

1350 131 131 SER C C 174.822 . 1

1351 131 131 SER CA C 60.95 . 1

1352 131 131 SER CB C 63.353 . 1

1353 131 131 SER N N 117.570 . 1

1354 132 132 ASP H H 7.871 . 1

1355 132 132 ASP HA H 4.599 . 1

1356 132 132 ASP HB2 H 2.680 . 2

1357 132 132 ASP HB3 H 2.680 . 2

1358 132 132 ASP C C 176.762 . 1

1359 132 132 ASP CA C 55.070 . 1

1360 132 132 ASP CB C 41.739 . 1

1361 132 132 ASP N N 120.909 . 1

1362 133 133 GLY H H 7.712 . 1

1363 133 133 GLY HA2 H 4.097 . 2

1364 133 133 GLY HA3 H 3.753 . 2

1365 133 133 GLY C C 173.225 . 1

1366 133 133 GLY CA C 44.922 . 1

1367 133 133 GLY N N 106.126 . 1

1368 134 134 TRP H H 8.191 . 1

1369 134 134 TRP HA H 4.481 . 1

1370 134 134 TRP HB2 H 3.098 . 2

1371 134 134 TRP HB3 H 3.162 . 2

1372 134 134 TRP HD1 H 7.416 . 1

1373 134 134 TRP HE1 H 10.1 . 3

1374 134 134 TRP HE3 H 6.590 . 3

1375 134 134 TRP HH2 H 6.970 . 1

1376 134 134 TRP HZ2 H 7.04 . 3

1377 134 134 TRP HZ3 H 6.86 . 3

1378 134 134 TRP C C 177.254 . 1

1379 134 134 TRP CA C 58.353 . 1

1380 134 134 TRP CB C 30.2 . 1

1381 134 134 TRP CD1 C 120.7 . 3

1382 134 134 TRP CH2 C 124.9 . 1

1383 134 134 TRP CZ2 C 115.2 . 3

1384 134 134 TRP CZ3 C 124.9 . 3

1385 134 134 TRP N N 119.836 . 1

1386 134 134 TRP NE1 N 129.1 . 1

1387 135 135 LYS H H 8.932 . 1

1388 135 135 LYS HA H 4.786 . 1

1389 135 135 LYS HB2 H 1.92 . 2

1390 135 135 LYS HB3 H 1.77 . 2

1391 135 135 LYS HD2 H 1.46 . 2

1392 135 135 LYS HD3 H 1.46 . 2

1393 135 135 LYS HE2 H 3 . 2

1394 135 135 LYS HG2 H 1.77 . 2

1395 135 135 LYS HG3 H 1.65 . 2

1396 135 135 LYS C C 175.016 . 1

1397 135 135 LYS CA C 55.019 . 1

1398 135 135 LYS CB C 35.371 . 1

1399 135 135 LYS CD C 24 . 1

1400 135 135 LYS CE C 41.990 . 1

1401 135 135 LYS CG C 28.9 . 1

1402 135 135 LYS N N 124.161 . 1

1403 136 136 VAL H H 8.230 . 1

1404 136 136 VAL HA H 5.356 . 1

1405 136 136 VAL HB H 1.391 . 1

1406 136 136 VAL HG1 H -0.222 . 2

1407 136 136 VAL HG2 H 0.477 . 2

1408 136 136 VAL C C 174.240 . 1

1409 136 136 VAL CA C 59.222 . 1

1410 136 136 VAL CB C 35.680 . 1

1411 136 136 VAL CG1 C 20.713 . 1

1412 136 136 VAL CG2 C 18.084 . 1
1413 136 136 VAL N N 118.667 . 1
1414 137 137 ASN H H 8.204 . 1
1415 137 137 ASN HA H 4.855 . 1
1416 137 137 ASN HB2 H 2.694 . 2
1417 137 137 ASN HB3 H 2.444 . 2
1418 137 137 ASN HD21 H 7.513 . 2
1419 137 137 ASN HD22 H 7.513 . 2
1420 137 137 ASN C C 172.986 . 1
1421 137 137 ASN CA C 53.205 . 1
1422 137 137 ASN CB C 44.325 . 1
1423 137 137 ASN N N 121.395 . 1
1424 138 138 THR H H 8.361 . 1
1425 138 138 THR HA H 5.179 . 1
1426 138 138 THR HB H 3.654 . 1
1427 138 138 THR HG2 H 1.106 . 1
1428 138 138 THR C C 173.792 . 1
1429 138 138 THR CA C 62.308 . 1
1430 138 138 THR CB C 70.814 . 1
1431 138 138 THR CG2 C 21.577 . 1
1432 138 138 THR N N 118.581 . 1
1433 139 139 VAL H H 9.333 . 1
1434 139 139 VAL HA H 4.579 . 1
1435 139 139 VAL HB H 1.952 . 1
1436 139 139 VAL HG1 H 0.909 . 2
1437 139 139 VAL HG2 H 0.909 . 2
1438 139 139 VAL C C 172.494 . 1
1439 139 139 VAL CA C 59.229 . 1
1440 139 139 VAL CB C 35.819 . 1
1441 139 139 VAL CG1 C 21.756 . 1
1442 139 139 VAL CG2 C 20.412 . 1
1443 139 139 VAL N N 124.995 . 1
1444 140 140 ASP H H 8.507 . 1
1445 140 140 ASP HA H 4.79 . 1
1446 140 140 ASP HB2 H 2.81 . 2
1447 140 140 ASP HB3 H 2.523 . 2
1448 140 140 ASP C C 176.628 . 1
1449 140 140 ASP CA C 53.802 . 1
1450 140 140 ASP CB C 41.863 . 1
1451 140 140 ASP N N 128.523 . 1
1452 141 141 LEU H H 8.155 . 1
1453 141 141 LEU HA H 4.235 . 1
1454 141 141 LEU HB2 H 1.72 . 2
1455 141 141 LEU HB3 H 1.628 . 2
1456 141 141 LEU HD1 H 0.93 . 2
1457 141 141 LEU HD2 H 1.006 . 2
1458 141 141 LEU HG H 1.008 . 1
1459 141 141 LEU C C 176.553 . 1
1460 141 141 LEU CA C 57 . 1
1461 141 141 LEU CB C 42.609 . 1
1462 141 141 LEU CD1 C 27.2 . 1
1463 141 141 LEU CD2 C 23.8 . 1
1464 141 141 LEU CG C 26.4 . 1
1465 141 141 LEU N N 123.606 . 1
1466 142 142 GLY H H 8.939 . 1
1467 142 142 GLY HA2 H 3.999 . 2
1468 142 142 GLY HA3 H 3.699 . 2
1469 142 142 GLY C C 174.180 . 1
1470 142 142 GLY CA C 46.414 . 1
1471 142 142 GLY N N 106.037 . 1
1472 143 143 VAL H H 7.226 . 1
1473 143 143 VAL HA H 4.579 . 1
1474 143 143 VAL HB H 2.041 . 1
1475 143 143 VAL HG1 H 0.919 . 2
1476 143 143 VAL HG2 H 0.732 . 2
1477 143 143 VAL C C 174.255 . 1
1478 143 143 VAL CA C 59.153 . 1
1479 143 143 VAL CB C 35.744 . 1

1480 143 143 VAL CG1 C 21.858 . 1
1481 143 143 VAL CG2 C 20.401 . 1
1482 143 143 VAL N N 112.058 . 1
1483 144 144 SER H H 8.361 . 1
1484 144 144 SER HA H 5.111 . 1
1485 144 144 SER HB2 H 3.901 . 2
1486 144 144 SER HB3 H 3.753 . 2
1487 144 144 SER C C 174.300 . 1
1488 144 144 SER CA C 57.158 . 1
1489 144 144 SER CB C 64.545 . 1
1490 144 144 SER N N 118.581 . 1
1491 145 145 VAL H H 8.487 . 1
1492 145 145 VAL HA H 3.644 . 1
1493 145 145 VAL HB H 2.027 . 1
1494 145 145 VAL HG1 H 0.866 . 2
1495 145 145 VAL HG2 H 0.909 . 2
1496 145 145 VAL C C 174.762 . 1
1497 145 145 VAL CA C 64.796 . 1
1498 145 145 VAL CB C 32.386 . 1
1499 145 145 VAL CG1 C 22.1 . 1
1500 145 145 VAL CG2 C 22.1 . 1
1501 145 145 VAL N N 120.026 . 1
1502 146 146 SER H H 6.862 . 1
1503 146 146 SER HA H 4.864 . 1
1504 146 146 SER HB2 H 3.841 . 2
1505 146 146 SER HB3 H 3.761 . 2
1506 146 146 SER C C 170.912 . 1
1507 146 146 SER CA C 56.861 . 1
1508 146 146 SER CB C 63.743 . 1
1509 146 146 SER N N 119.542 . 1
1510 147 147 PRO HA H 4.205 . 1
1511 147 147 PRO HB2 H 2.129 . 2
1512 147 147 PRO HB3 H 1.805 . 2
1513 147 147 PRO HD2 H 3.76 . 2
1514 147 147 PRO HD3 H 3.66 . 2
1515 147 147 PRO HG2 H 1.803 . 2
1516 147 147 PRO HG3 H 1.803 . 2
1517 147 147 PRO C C 175.628 . 1
1518 147 147 PRO CA C 64.126 . 1
1519 147 147 PRO CB C 32.064 . 1
1520 147 147 PRO CD C 51 . 1
1521 147 147 PRO CG C 28 . 1
1522 148 148 ASP H H 8.460 . 1
1523 148 148 ASP HA H 4.973 . 1
1524 148 148 ASP HB2 H 2.85 . 2
1525 148 148 ASP HB3 H 2.773 . 2
1526 148 148 ASP C C 175.747 . 1
1527 148 148 ASP CA C 54.756 . 1
1528 148 148 ASP CB C 38.878 . 1
1529 148 148 ASP N N 116.313 . 1
1530 149 149 LEU H H 7.561 . 1
1531 149 149 LEU HA H 4.481 . 1
1532 149 149 LEU HB2 H 1.698 . 2
1533 149 149 LEU HB3 H 0.9 . 2
1534 149 149 LEU HD1 H 0.952 . 2
1535 149 149 LEU HD2 H 0.633 . 2
1536 149 149 LEU HG H 0.84 . 1
1537 149 149 LEU C C 175.120 . 1
1538 149 149 LEU CA C 55.290 . 1
1539 149 149 LEU CB C 43.952 . 1
1540 149 149 LEU CD1 C 27.219 . 1
1541 149 149 LEU CD2 C 23.316 . 1
1542 149 149 LEU N N 117.994 . 1
1543 150 150 ALA H H 7.941 . 1
1544 150 150 ALA HA H 5.615 . 1
1545 150 150 ALA HB H 0.624 . 1
1546 150 150 ALA C C 175.180 . 1
1547 150 150 ALA CA C 49.700 . 1

1548 150 150 ALA CB C 21.398 . 1
1549 150 150 ALA N N 118.843 . 1
1550 151 151 PHE H H 8.538 . 1
1551 151 151 PHE HA H 5.032 . 1
1552 151 151 PHE HB2 H 2.690 . 2
1553 151 151 PHE HB3 H 2.63 . 2
1554 151 151 PHE HD1 H 7.24 . 3
1555 151 151 PHE HD2 H 7.24 . 3
1556 151 151 PHE HE1 H 7.115 . 3
1557 151 151 PHE HE2 H 7.115 . 3
1558 151 151 PHE HZ H 7.244 . 1
1559 151 151 PHE C C 175.061 . 1
1560 151 151 PHE CA C 55.592 . 1
1561 151 151 PHE CB C 42.758 . 1
1562 151 151 PHE CD1 C 132.8 . 3
1563 151 151 PHE CD2 C 132.8 . 3
1564 151 151 PHE CE1 C 132.8 . 3
1565 151 151 PHE CE2 C 132.8 . 3
1566 151 151 PHE N N 117.045 . 1
1567 152 152 GLY H H 7.546 . 1
1568 152 152 GLY HA2 H 3.893 . 2
1569 152 152 GLY HA3 H 3.147 . 2
1570 152 152 GLY C C 182.075 . 1
1571 152 152 GLY CA C 44.176 . 1
1572 152 152 GLY N N 107.829 . 1
1573 153 153 LEU H H 7.933 . 1
1574 153 153 LEU HA H 5.366 . 1
1575 153 153 LEU HB2 H 1.569 . 2
1576 153 153 LEU HB3 H 1.440 . 2
1577 153 153 LEU HD1 H 0.906 . 2
1578 153 153 LEU HD2 H 0.906 . 2
1579 153 153 LEU HG H 1.51 . 1
1580 153 153 LEU C C 175.926 . 1
1581 153 153 LEU CA C 53.300 . 1
1582 153 153 LEU CB C 47.161 . 1
1583 153 153 LEU CD1 C 25.90 . 1
1584 153 153 LEU CD2 C 25.90 . 1
1585 153 153 LEU CG C 26.960 . 1
1586 153 153 LEU N N 116.664 . 1
1587 154 154 THR H H 9.610 . 1
1588 154 154 THR HA H 4.65 . 1
1589 154 154 THR HB H 4.107 . 1
1590 154 154 THR HG2 H 1.205 . 1
1591 154 154 THR C C 172.628 . 1
1592 154 154 THR CA C 61.862 . 1
1593 154 154 THR CB C 70.995 . 1
1594 154 154 THR CG2 C 21.676 . 1
1595 154 154 THR N N 122.745 . 1
1596 155 155 ALA H H 8.104 . 1
1597 155 155 ALA HA H 4.402 . 1
1598 155 155 ALA HB H 1.146 . 1
1599 155 155 ALA C C 175.687 . 1
1600 155 155 ALA CA C 52.184 . 1
1601 155 155 ALA CB C 21.044 . 1
1602 155 155 ALA N N 125.911 . 1
1603 156 156 ASP H H 9.069 . 1
1604 156 156 ASP HA H 4.667 . 1
1605 156 156 ASP HB2 H 2.902 . 2
1606 156 156 ASP HB3 H 2.683 . 2
1607 156 156 ASP C C 177.269 . 1
1608 156 156 ASP CA C 55.443 . 1
1609 156 156 ASP CB C 40.759 . 1
1610 156 156 ASP N N 123.464 . 1
1611 157 157 GLY H H 8.723 . 1
1612 157 157 GLY HA2 H 4.094 . 2
1613 157 157 GLY HA3 H 3.929 . 2
1614 157 157 GLY C C 174.896 . 1
1615 157 157 GLY CA C 47.235 . 1

1616 157 157 GLY N N 113.184 . 1

1617 158 158 VAL H H 8.117 . 1

1618 158 158 VAL HA H 4.618 . 1

1619 158 158 VAL HB H 2.327 . 1

1620 158 158 VAL HG1 H 1.21 . 2

1621 158 158 VAL HG2 H 1.255 . 2

1622 158 158 VAL C C 175.747 . 1

1623 158 158 VAL CA C 63.032 . 1

1624 158 158 VAL CB C 31.819 . 1

1625 158 158 VAL CG1 C 22.831 . 1

1626 158 158 VAL CG2 C 21.5 . 1

1627 158 158 VAL N N 119.742 . 1

1628 159 159 LYS H H 8.494 . 1

1629 159 159 LYS HA H 4.46 . 1

1630 159 159 LYS HB2 H 1.84 . 2

1631 159 159 LYS HB3 H 1.84 . 2

1632 159 159 LYS C C 174.852 . 1

1633 159 159 LYS CA C 55.754 . 1

1634 159 159 LYS CB C 34.326 . 1

1635 159 159 LYS CD C 29.055 . 1

1636 159 159 LYS CE C 42.081 . 1

1637 159 159 LYS CG C 24.773 . 1

1638 159 159 LYS N N 129.878 . 1

1639 160 160 VAL H H 8.988 . 1

1640 160 160 VAL HA H 3.4 . 1

1641 160 160 VAL HB H 2.26 . 1

1642 160 160 VAL HG1 H 0.919 . 2

1643 160 160 VAL HG2 H 0.849 . 2

1644 160 160 VAL C C 175.837 . 1

1645 160 160 VAL CA C 61.935 . 1

1646 160 160 VAL CB C 29.849 . 1

1647 160 160 VAL CG1 C 23.607 . 1

1648 160 160 VAL CG2 C 21.312 . 1

1649 160 160 VAL N N 122.308 . 1

1650 162 162 LYS H H 8.353 . 1

1651 162 162 LYS C C 173.941 . 1

1652 162 162 LYS CA C 56.264 . 1

1653 162 162 LYS CB C 33.207 . 1

1654 162 162 LYS N N 122.710 . 1

1655 163 163 LEU H H 8.474 . 1

1656 163 163 LEU HA H 5.317 . 1

1657 163 163 LEU HB2 H 1.933 . 2

1658 163 163 LEU HB3 H 1.37 . 2

1659 163 163 LEU HD1 H 0.894 . 2

1660 163 163 LEU HD2 H 0.894 . 2

1661 163 163 LEU HG H 1.77 . 1

1662 163 163 LEU C C 175.344 . 1

1663 163 163 LEU CA C 53.951 . 1

1664 163 163 LEU CB C 44.027 . 1

1665 163 163 LEU CD1 C 26.4 . 1

1666 163 163 LEU CD2 C 26.4 . 1

1667 163 163 LEU CG C 28.5 . 1

1668 163 163 LEU N N 122.390 . 1

1669 164 164 TYR H H 9.426 . 1

1670 164 164 TYR HA H 5.504 . 1

1671 164 164 TYR HB2 H 2.781 . 2

1672 164 164 TYR HB3 H 2.582 . 2

1673 164 164 TYR HD1 H 7.34 . 3

1674 164 164 TYR HD2 H 7.34 . 3

1675 164 164 TYR HE1 H 6.70 . 3

1676 164 164 TYR HE2 H 6.70 . 3

1677 164 164 TYR C C 174.523 . 1

1678 164 164 TYR CA C 56.057 . 1

1679 164 164 TYR CB C 44.778 . 1

1680 164 164 TYR CD1 C 131.7 . 3

1681 164 164 TYR CD2 C 131.7 . 3

1682 164 164 TYR CE1 C 116.4 . 3

1683 164 164 TYR CE2 C 116.4 . 3

1684 164 164 TYR N N 116.766 . 1
1685 165 165 ALA H H 8.974 . 1
1686 165 165 ALA HA H 5.209 . 1
1687 165 165 ALA HB H 1.365 . 1
1688 165 165 ALA C C 176.941 . 1
1689 165 165 ALA CA C 50.597 . 1
1690 165 165 ALA CB C 22.980 . 1
1691 165 165 ALA N N 120.465 . 1
1692 166 166 SER H H 8.690 . 1
1693 166 166 SER HA H 4.737 . 1
1694 166 166 SER HB2 H 4.100 . 2
1695 166 166 SER HB3 H 3.944 . 2
1696 166 166 SER C C 175.807 . 1
1697 166 166 SER CA C 60.200 . 1
1698 166 166 SER CB C 64.148 . 1
1699 166 166 SER N N 117.319 . 1
1700 167 167 SER H H 8.780 . 1
1701 167 167 SER HA H 4.147 . 1
1702 167 167 SER HB2 H 3.991 . 2
1703 167 167 SER HB3 H 3.878 . 2
1704 167 167 SER C C 175.568 . 1
1705 167 167 SER CA C 60.990 . 1
1706 167 167 SER CB C 63.202 . 1
1707 167 167 SER N N 117.695 . 1
1708 168 168 GLY H H 8.375 . 1
1709 168 168 GLY HA2 H 3.956 . 2
1710 168 168 GLY HA3 H 3.956 . 2
1711 168 168 GLY C C 174.076 . 1
1712 168 168 GLY CA C 45.592 . 1
1713 168 168 GLY N N 111.116 . 1
1714 169 169 LEU H H 7.766 . 1
1715 169 169 LEU HA H 4.469 . 1
1716 169 169 LEU HB2 H 1.667 . 2
1717 169 169 LEU HB3 H 1.667 . 2
1718 169 169 LEU HD1 H 0.838 . 2
1719 169 169 LEU HD2 H 0.838 . 2
1720 169 169 LEU HG H 1.613 . 1
1721 169 169 LEU C C 177.463 . 1
1722 169 169 LEU CA C 54.45 . 1
1723 169 169 LEU CB C 43.2 . 1
1724 169 169 LEU CD1 C 24.773 . 1
1725 169 169 LEU CD2 C 24.773 . 1
1726 169 169 LEU CG C 27. . 1
1727 169 169 LEU N N 120.884 . 1
1728 170 170 THR H H 8.100 . 1
1729 170 170 THR HA H 4.30 . 1
1730 170 170 THR HB H 4.173 . 1
1731 170 170 THR HG2 H 1.155 . 1
1732 170 170 THR C C 172.986 . 1
1733 170 170 THR CA C 61.562 . 1
1734 170 170 THR CB C 69.695 . 1
1735 170 170 THR CG2 C 21.403 . 1
1736 170 170 THR N N 114.923 . 1
1737 171 171 ALA H H 8.313 . 1
1738 171 171 ALA HA H 4.95 . 1
1739 171 171 ALA HB H 1.178 . 1
1740 171 171 ALA C C 177.433 . 1
1741 171 171 ALA CA C 50.200 . 1
1742 171 171 ALA CB C 20.229 . 1
1743 171 171 ALA N N 127.031 . 1
1744 172 172 ILE H H 8.203 . 1
1745 172 172 ILE HA H 4.000 . 1
1746 172 172 ILE HB H 1.846 . 1
1747 172 172 ILE HD1 H 0.704 . 1
1748 172 172 ILE HG12 H 1.451 . 1
1749 172 172 ILE HG13 H 1.451 . 1
1750 172 172 ILE HG2 H 0.998 . 2
1751 172 172 ILE C C 176.105 . 1

1752 172 172 ILE CA C 61.611 . 1

1753 172 172 ILE CB C 37.833 . 1

1754 172 172 ILE CD1 C 12.3 . 1

1755 172 172 ILE CG1 C 27.615 . 2

1756 172 172 ILE CG2 C 19.00 . 1

1757 172 172 ILE N N 122.113 . 1

1758 173 173 ASN H H 8.228 . 1

1759 173 173 ASN HA H 4.547 . 1

1760 173 173 ASN HB2 H 2.680 . 2

1761 173 173 ASN HB3 H 2.680 . 2

1762 173 173 ASN C C 175.792 . 1

1763 173 173 ASN CA C 52.981 . 1

1764 173 173 ASN CB C 39.15 . 1

1765 173 173 ASN N N 121.65 . 1

1766 174 174 ASP H H 8.023 . 1

1767 174 174 ASP HA H 4.566 . 1

1768 174 174 ASP HB2 H 2.697 . 2

1769 174 174 ASP HB3 H 2.544 . 2

1770 174 174 ASP C C 175.434 . 1

1771 174 174 ASP CA C 53.503 . 1

1772 174 174 ASP CB C 40.967 . 1

1773 174 174 ASP N N 117.736 . 1

1774 175 175 ASP H H 8.115 . 1

1775 175 175 ASP HA H 4.472 . 1

1776 175 175 ASP HB2 H 1.678 . 2

1777 175 175 ASP HB3 H 1.987 . 2

1778 175 175 ASP C C 178.105 . 1

1779 175 175 ASP CA C 55.255 . 1

1780 175 175 ASP CB C 42.25 . 1

1781 175 175 ASP N N 123.119 . 1

1782 176 176 PRO HA H 4.444 . 1

1783 176 176 PRO HB2 H 2.383 . 2

1784 176 176 PRO HB3 H 2.052 . 2

1785 176 176 PRO HD2 H 4.095 . 2

1786 176 176 PRO HD3 H 4.005 . 2

1787 176 176 PRO HG2 H 2.062 . 2

1788 176 176 PRO HG3 H 2.03 . 2

1789 176 176 PRO C C 179.463 . 1

1790 176 176 PRO CA C 64.623 . 1

1791 176 176 PRO CB C 32.089 . 1

1792 176 176 PRO CD C 51.602 . 1

1793 176 176 PRO CG C 27.233 . 1

1794 177 177 SER H H 8.760 . 1

1795 177 177 SER HA H 4.322 . 1

1796 177 177 SER HB2 H 3.988 . 2

1797 177 177 SER HB3 H 3.878 . 2

1798 177 177 SER C C 175.508 . 1

1799 177 177 SER CA C 60.965 . 1

1800 177 177 SER CB C 62.930 . 1

1801 177 177 SER N N 116.575 . 1

1802 178 178 LEU H H 7.437 . 1

1803 178 178 LEU HA H 4.477 . 1

1804 178 178 LEU HB2 H 1.984 . 2

1805 178 178 LEU HB3 H 1.668 . 2

1806 178 178 LEU HD1 H 0.65 . 2

1807 178 178 LEU HD2 H 0.502 . 2

1808 178 178 LEU HG H 1.406 . 1

1809 178 178 LEU C C 177.359 . 1

1810 178 178 LEU CA C 54.427 . 1

1811 178 178 LEU CB C 42.317 . 1

1812 178 178 LEU CD1 C 23.5 . 1

1813 178 178 LEU CD2 C 22.5 . 1

1814 178 178 LEU CG C 26.58 . 1

1815 178 178 LEU N N 121.089 . 1

1816 179 179 GLY H H 7.970 . 1

1817 179 179 GLY HA2 H 3.932 . 2

1818 179 179 GLY HA3 H 3.932 . 2

1819 179 179 GLY C C 174.389 . 1

1820 179 179 GLY CA C 46.340 . 1

1821 179 179 GLY N N 108.064 . 1

1822 180 180 CYS H H 7.524 . 1

1823 180 180 CYS HA H 5.180 . 1

1824 180 180 CYS HB2 H 2.893 . 2

1825 180 180 CYS HB3 H 2.429 . 2

1826 180 180 CYS C C 175.016 . 1

1827 180 180 CYS CA C 55.589 . 1

1828 180 180 CYS CB C 41.510 . 1

1829 180 180 CYS N N 116.942 . 1

1830 181 181 LYS H H 8.206 . 1

1831 181 181 LYS HA H 4.342 . 1

1832 181 181 LYS HB2 H 1.867 . 2

1833 181 181 LYS HB3 H 1.753 . 2

1834 181 181 LYS C C 174.538 . 1

1835 181 181 LYS CA C 56.003 . 1

1836 181 181 LYS CB C 33.341 . 1

1837 181 181 LYS CD C 29.100 . 1

1838 181 181 LYS CE C 42.267 . 1

1839 181 181 LYS CG C 24.533 . 1

1840 181 181 LYS N N 124.083 . 1

1841 182 182 ALA H H 7.915 . 1

1842 182 182 ALA HA H 4.14 . 1

1843 182 182 ALA HB H 1.3 . 1

1844 182 182 ALA C C 182.343 . 1

1845 182 182 ALA CA C 53.788 . 1

1846 182 182 ALA CB C 20.149 . 1

1847 182 182 ALA N N 131.820 . 1

A6. Resonance assignment for TSR12

Atom_shift_assign_ID

Residue_author_seq_code

Residue_seq_code

Residue_label

Atom_name

Atom_type

Chem_shift_value

Chem_shift_value_error

Chem_shift_ambiguity_code

1 1 ALA CA C 4.000 . 1

2 1 ALA HA H 4.002 . 1

3 1 ALA CB C 4.000 . 1

4 2 ILE N N 125.960 . 1

5 2 ILE HN H 7.900 . 1

6 2 ILE CA C 61.482 . 1

7 2 ILE HA H 4.174 . 1

8 2 ILE CB C 38.605 . 1

9 2 ILE HB H 1.850 . 1

10 2 ILE CG1 C 27.488 . 2

11 2 ILE HG12 H 1.174 . 1

12 2 ILE HG11 H 1.478 . 1

13 2 ILE CD1 C 12.790 . 1

14 2 ILE HD11 H 0.848 . 1

15 2 ILE CG2 C 17.598 . 1

16 2 ILE HG21 H 0.880 . 2

17 2 ILE C C 176.101 . 1

18	3	CYS	N	N	130.326	.	1
19	3	CYS	HN	H	7.905	.	1
20	3	CYS	CA	C	57.311	.	1
21	3	CYS	HA	H	4.557	.	1
22	3	CYS	CB	C	32.128	.	1
23	3	CYS	C	C	180.705	.	1
24	4	SER	N	N	118.129	.	1
25	4	SER	HN	H	8.439	.	1
26	4	SER	CA	C	58.706	.	1
27	4	SER	CB	C	64.025	.	1
28	4	SER	C	C	173.853	.	1
29	5	ASP	N	N	122.106	.	1
30	5	ASP	HN	H	8.227	.	1
31	5	ASP	CA	C	54.520	.	1
32	5	ASP	HA	H	4.625	.	1
33	5	ASP	CB	C	41.342	.	1
34	5	ASP	HB2	H	2.595	.	2
35	5	ASP	HB1	H	2.702	.	2
36	5	ASP	C	C	176.008	.	1
37	6	TRP	N	N	121.593	.	1
38	6	TRP	HN	H	8.054	.	1
39	6	TRP	CA	C	57.698	.	1
40	6	TRP	HA	H	4.548	.	1
41	6	TRP	CB	C	29.559	.	1
42	6	TRP	HB2	H	3.263	.	2
43	6	TRP	HB1	H	3.263	.	2
44	6	TRP	HD1	H	7.154	.	1
45	6	TRP	HE3	H	7.495	.	3
46	6	TRP	C	C	176.489	.	1
47	7	SER	N	N	117.226	.	1
48	7	SER	HN	H	8.000	.	1
49	7	SER	CA	C	59.198	.	1
50	7	SER	HA	H	4.182	.	1
51	7	SER	CB	C	63.822	.	1
52	7	SER	HB2	H	3.597	.	2
53	7	SER	HB1	H	3.711	.	2
54	7	SER	C	C	174.216	.	1
55	8	ALA	N	N	124.675	.	1
56	8	ALA	HN	H	7.937	.	1
57	8	ALA	CA	C	52.737	.	1
58	8	ALA	HA	H	4.181	.	1
59	8	ALA	CB	C	19.177	.	1
60	8	ALA	HB1	H	1.229	.	1
61	8	ALA	C	C	177.093	.	1
62	9	TRP	N	N	119.584	.	1
63	9	TRP	HN	H	7.857	.	1
64	9	TRP	CA	C	57.388	.	1
65	9	TRP	HA	H	4.587	.	1
66	9	TRP	CB	C	29.980	.	1
67	9	TRP	HB2	H	3.287	.	2
68	9	TRP	HB1	H	3.284	.	2
69	9	TRP	HD1	H	7.170	.	1
70	9	TRP	HE3	H	7.512	.	3
71	9	TRP	C	C	175.636	.	1
72	10	SER	N	N	118.763	.	1
73	10	SER	HN	H	7.785	.	1
74	10	SER	CA	C	59.481	.	1
75	10	SER	HA	H	4.565	.	1
76	10	SER	CB	C	63.745	.	1
77	10	SER	HB2	H	3.907	.	2
78	10	SER	HB1	H	4.192	.	2
79	11	PRO	CA	C	63.427	.	1
80	11	PRO	HA	H	4.427	.	1
81	11	PRO	CB	C	32.166	.	1
82	11	PRO	HB2	H	1.898	.	2
83	11	PRO	HB1	H	2.261	.	2
84	11	PRO	CG	C	27.622	.	1
85	11	PRO	HG2	H	2.046	.	2

86	11	PRO	HG1	H	2.046	.	2
87	11	PRO	CD	C	50.523	.	1
88	11	PRO	HD2	H	3.672	.	2
89	11	PRO	HD1	H	3.839	.	2
90	12	CYS	N	N	119.538	.	1
91	12	CYS	HN	H	8.149	.	1
92	12	CYS	CA	C	53.402	.	1
93	12	CYS	HA	H	4.812	.	1
94	12	CYS	CB	C	30.310	.	1
95	12	CYS	HB2	H	3.084	.	2
96	12	CYS	HB1	H	3.203	.	2
97	12	CYS	C	C	174.489	.	1
98	16	CYS	CA	C	55.490	.	1
99	16	CYS	HA	H	4.483	.	1
100	16	CYS	CB	C	33.158	.	1
101	16	CYS	HB2	H	2.043	.	2
102	16	CYS	HB1	H	2.043	.	2
103	17	GLY	N	N	109.860	.	1
104	17	GLY	HN	H	8.334	.	1
105	17	GLY	CA	C	45.528	.	1
106	17	GLY	HA2	H	3.938	.	2
107	17	GLY	HA1	H	3.971	.	2
108	17	GLY	C	C	173.667	.	1
109	18	ASP	N	N	120.500	.	1
110	18	ASP	HN	H	8.230	.	1
111	18	ASP	CA	C	54.442	.	1
112	18	ASP	HA	H	4.589	.	1
113	18	ASP	CB	C	41.109	.	1
114	18	ASP	HB2	H	2.710	.	2
115	18	ASP	HB1	H	2.710	.	2
116	18	ASP	C	C	177.124	.	1
117	19	GLY	N	N	109.843	.	1
118	19	GLY	HN	H	8.467	.	1
119	19	GLY	CA	C	45.724	.	1
120	19	GLY	HA2	H	3.944	.	2
121	19	GLY	HA1	H	3.958	.	2
122	19	GLY	C	C	174.722	.	1
123	20	SER	N	N	115.876	.	1
124	20	SER	HN	H	8.261	.	1
125	20	SER	CA	C	59.093	.	1
126	20	SER	HA	H	4.371	.	1
127	20	SER	CB	C	63.745	.	1
128	20	SER	HB2	H	3.888	.	2
129	20	SER	HB1	H	3.888	.	2
130	20	SER	C	C	174.691	.	1
131	21	GLN	N	N	121.529	.	1
132	21	GLN	HN	H	8.323	.	1
133	21	GLN	CA	C	55.838	.	1
134	21	GLN	HA	H	4.344	.	1
135	21	GLN	CB	C	29.636	.	1
136	21	GLN	HB2	H	1.945	.	2
137	21	GLN	HB1	H	2.096	.	2
138	21	GLN	HG2	H	2.305	.	2
139	21	GLN	HG1	H	2.399	.	2
140	21	GLN	C	C	175.714	.	1
141	22	ILE	N	N	121.701	.	1
142	22	ILE	HN	H	8.044	.	1
143	22	ILE	CA	C	61.186	.	1
144	22	ILE	HA	H	4.145	.	1
145	22	ILE	CB	C	38.706	.	1
146	22	ILE	HB	H	1.868	.	1
147	22	ILE	CG1	C	27.695	.	2
148	22	ILE	HG12	H	1.194	.	1
149	22	ILE	HG11	H	1.458	.	1
150	22	ILE	CD1	C	13.003	.	1
151	22	ILE	HD11	H	0.845	.	1
152	22	ILE	CG2	C	18.113	.	1
153	22	ILE	HG21	H	0.861	.	2

154 22 ILE C C 176.101 . 1
155 23 ARG N N 125.600 . 1
156 23 ARG HN H 8.370 . 1
157 23 ARG CA C 56.070 . 1
158 23 ARG HA H 4.410 . 1
159 23 ARG CB C 31.161 . 1
160 23 ARG HB2 H 1.758 . 2
161 23 ARG HB1 H 1.856 . 2
162 23 ARG CG C 27.125 . 1
163 23 ARG CD C 43.275 . 1
164 23 ARG C C 176.210 . 1
165 24 THR N N 115.938 . 1
166 24 THR HN H 8.197 . 1
167 24 THR CA C 61.962 . 1
168 24 THR HA H 4.329 . 1
169 24 THR CB C 69.791 . 1
170 24 THR HB H 4.199 . 1
171 24 THR CG2 C 21.681 . 1
172 24 THR HG21 H 1.174 . 1
173 24 THR C C 174.303 . 1
174 25 ARG N N 121.366 . 1
175 25 ARG HN H 8.402 . 1
176 25 ARG CA C 53.674 . 1
177 25 ARG CB C 34.021 . 1
178 25 ARG CG C 36.652 . 1
179 27 GLU CA C 56.398 . 1
180 27 GLU HA H 4.450 . 1
181 27 GLU CB C 30.525 . 1
182 27 GLU HB2 H 1.937 . 2
183 27 GLU HB1 H 2.065 . 2
184 27 GLU CG C 36.289 . 1
185 27 GLU HG2 H 2.261 . 2
186 27 GLU HG1 H 2.290 . 2
187 27 GLU C C 177.388 . 1
188 28 VAL N N 121.850 . 1
189 28 VAL HN H 8.359 . 1
190 28 VAL CA C 62.246 . 1
191 28 VAL HA H 4.150 . 1
192 28 VAL CB C 32.893 . 1
193 28 VAL HB H 2.084 . 1
194 28 VAL CG2 C 21.170 . 1
195 28 VAL HG21 H 0.929 . 2
196 28 VAL CG1 C 21.170 . 1
197 28 VAL HG11 H 0.929 . 2
198 28 VAL C C 175.723 . 1
199 29 SER N N 119.945 . 1
200 29 SER HN H 8.340 . 1
201 29 SER CA C 58.123 . 1
202 29 SER HA H 4.443 . 1
203 29 SER CB C 64.058 . 1
204 29 SER HB2 H 3.839 . 2
205 29 SER HB1 H 3.825 . 2
206 29 SER C C 173.619 . 1
207 30 ALA N N 127.111 . 1
208 30 ALA HN H 8.233 . 1
209 30 ALA CA C 50.309 . 1
210 30 ALA HA H 4.582 . 1
211 30 ALA CB C 18.826 . 1
212 30 ALA HB1 H 1.368 . 1
213 30 ALA C C 175.094 . 1
214 31 PRO CA C 63.010 . 1
215 31 PRO HA H 4.429 . 1
216 31 PRO CB C 32.159 . 1
217 31 PRO HB2 H 1.884 . 2
218 31 PRO HB1 H 2.259 . 2
219 31 PRO CG C 27.441 . 1
220 31 PRO HG2 H 2.026 . 2
221 31 PRO HG1 H 2.026 . 2

222	31	PRO	CD	C	50.432	.	1
223	31	PRO	HD2	H	3.662	.	2
224	31	PRO	HD1	H	3.898	.	2
225	31	PRO	C	C	176.695	.	1
226	32	GLN	N	N	121.593	.	1
227	32	GLN	HN	H	8.437	.	1
228	32	GLN	CA	C	53.708	.	1
229	32	GLN	HA	H	4.606	.	1
230	32	GLN	CB	C	29.309	.	1
231	32	GLN	HB2	H	1.951	.	2
232	32	GLN	HB1	H	2.111	.	2
233	32	GLN	CG	C	33.749	.	1
234	32	GLN	HG2	H	2.273	.	2
235	32	GLN	HG1	H	2.273	.	2
236	32	GLN	C	C	174.216	.	1
237	33	PRO	CA	C	63.518	.	1
238	33	PRO	HA	H	4.271	.	1
239	33	PRO	CB	C	32.159	.	1
240	33	PRO	HB2	H	1.957	.	2
241	33	PRO	HB1	H	2.290	.	2
242	33	PRO	CG	C	27.804	.	1
243	33	PRO	HG2	H	2.035	.	2
244	33	PRO	HG1	H	2.055	.	2
245	33	PRO	CD	C	51.068	.	1
246	33	PRO	HD2	H	3.672	.	2
247	33	PRO	HD1	H	3.888	.	2
248	33	PRO	C	C	177.350	.	1
249	34	GLY	N	N	109.270	.	1
250	34	GLY	HN	H	8.459	.	1
251	34	GLY	CA	C	45.129	.	1
252	34	GLY	HA2	H	3.939	.	2
253	34	GLY	HA1	H	3.939	.	2
254	34	GLY	C	C	173.202	.	1
255	35	THR	N	N	116.077	.	1
256	35	THR	HN	H	7.986	.	1
257	35	THR	CA	C	59.947	.	1
258	35	THR	HA	H	4.647	.	1
259	35	THR	CB	C	69.765	.	1
260	35	THR	HB	H	4.134	.	1
261	35	THR	CG2	C	21.310	.	1
262	35	THR	HG21	H	1.247	.	1
263	35	THR	C	C	172.691	.	1
264	36	PRO	CA	C	63.609	.	1
265	36	PRO	HA	H	4.488	.	1
266	36	PRO	CB	C	32.439	.	1
267	36	PRO	HB2	H	1.888	.	2
268	36	PRO	HB1	H	2.339	.	2
269	36	PRO	CG	C	27.713	.	1
270	36	PRO	HG2	H	2.035	.	2
271	36	PRO	HG1	H	2.035	.	2
272	36	PRO	CD	C	51.341	.	1
273	36	PRO	HD2	H	3.752	.	2
274	36	PRO	HD1	H	3.899	.	2
275	36	PRO	C	C	177.155	.	1
276	37	THR	N	N	113.479	.	1
277	37	THR	HN	H	8.192	.	1
278	37	THR	CA	C	61.311	.	1
279	37	THR	HA	H	3.839	.	1
280	37	THR	CB	C	69.311	.	1
281	37	THR	HB	H	4.155	.	1
282	37	THR	CG2	C	21.401	.	1
283	37	THR	HG21	H	1.290	.	1
284	37	THR	C	C	174.023	.	1
285	38	CYS	N	N	121.075	.	1
286	38	CYS	HN	H	8.219	.	1
287	38	CYS	CA	C	53.900	.	1
288	38	CYS	HA	H	5.045	.	1
289	38	CYS	CB	C	40.811	.	1

290 38 CYS HB2 H 2.671 . 2
291 38 CYS HB1 H 2.702 . 2
292 38 CYS C C 172.985 . 1
293 39 PRO CA C 65.217 . 1
294 39 PRO HA H 4.410 . 1
295 39 PRO CB C 31.962 . 1
296 39 PRO HB2 H 1.986 . 2
297 39 PRO HB1 H 2.310 . 2
298 39 PRO HG2 H 2.035 . 2
299 39 PRO HG1 H 2.035 . 2
300 39 PRO HD2 H 3.682 . 2
301 39 PRO HD1 H 3.800 . 2
302 39 PRO C C 176.876 . 1
303 40 ASP N N 115.200 . 1
304 40 ASP HN H 8.237 . 1
305 40 ASP CA C 53.340 . 1
306 40 ASP HA H 4.719 . 1
307 40 ASP CB C 40.254 . 1
308 40 ASP HB2 H 2.718 . 2
309 40 ASP HB1 H 2.702 . 2
310 40 ASP C C 175.342 . 1
311 41 CYS N N 121.336 . 1
312 41 CYS HN H 7.693 . 1
313 41 CYS CA C 58.402 . 1
314 41 CYS HA H 4.810 . 1
315 41 CYS CB C 36.947 . 1
316 41 CYS HB2 H 3.353 . 2
317 41 CYS HB1 H 3.353 . 2
318 41 CYS C C 173.319 . 1
319 42 PRO CA C 63.116 . 1
320 42 PRO HA H 4.418 . 1
321 42 PRO CB C 31.978 . 1
322 42 PRO HB2 H 1.880 . 2
323 42 PRO HB1 H 2.270 . 2
324 42 PRO CG C 27.438 . 1
325 42 PRO HG2 H 2.036 . 2
326 42 PRO HG1 H 2.036 . 2
327 42 PRO CD C 50.497 . 1
328 42 PRO HD2 H 3.682 . 2
329 42 PRO HD1 H 3.800 . 2
330 42 PRO C C 176.693 . 1
331 43 ALA N N 125.659 . 1
332 43 ALA HN H 8.373 . 1
333 43 ALA CA C 50.256 . 1
334 43 ALA HA H 4.361 . 1
335 43 ALA CB C 18.474 . 1
336 43 ALA HB1 H 1.335 . 1
337 43 ALA C C 175.636 . 1
338 44 PRO CA C 63.245 . 1
339 44 PRO HA H 4.410 . 1
340 44 PRO CB C 32.348 . 1
341 44 PRO HB2 H 1.908 . 2
342 44 PRO HB1 H 2.290 . 2
343 44 PRO CG C 27.804 . 1
344 44 PRO HG2 H 2.045 . 2
345 44 PRO HG1 H 2.045 . 2
346 44 PRO CD C 50.705 . 1
347 44 PRO HD2 H 3.653 . 2
348 44 PRO HD1 H 3.858 . 2
349 44 PRO C C 175.962 . 1
350 45 MET N N 120.308 . 1
351 45 MET HN H 8.405 . 1
352 45 MET CA C 55.583 . 1
353 45 MET HA H 4.478 . 1
354 45 MET CB C 33.158 . 1
355 45 MET HB2 H 2.074 . 2
356 45 MET HB1 H 2.101 . 2
357 45 MET CG C 31.984 . 1

358	45	MET	HG2	H	2.604	.	2
359	45	MET	HG1	H	2.663	.	2
360	45	MET	C	C	177.698	.	1
361	46	GLY	N	N	110.574	.	1
362	46	GLY	HN	H	8.385	.	1
363	46	GLY	CA	C	47.311	.	1
364	46	GLY	HA2	H	3.922	.	2
365	46	GLY	HA1	H	3.922	.	2
366	46	GLY	C	C	174.908	.	1
367	47	ARG	N	N	125.447	.	1
368	47	ARG	HN	H	7.777	.	1
369	47	ARG	CA	C	57.129	.	1
370	47	ARG	HA	H	4.202	.	1
371	47	ARG	CB	C	31.583	.	1
372	47	ARG	HB2	H	1.732	.	2
373	47	ARG	HB1	H	1.847	.	2
374	47	ARG	CG	C	27.034	.	1
375	47	ARG	HG2	H	1.585	.	2
376	47	ARG	HG1	H	1.585	.	2
377	47	ARG	CD	C	43.366	.	1
378	47	ARG	HD2	H	3.192	.	2
379	47	ARG	HD1	H	3.192	.	2
380	48	THR	N	N	115.942	.	1
381	48	THR	HN	H	8.200	.	1
382	49	CYS	CA	C	55.944	.	1
383	49	CYS	HA	H	4.752	.	1
384	49	CYS	CB	C	41.964	.	1
385	49	CYS	HB2	H	2.970	.	2
386	49	CYS	HB1	H	3.165	.	2
387	50	VAL	N	N	122.359	.	1
388	50	VAL	HN	H	8.423	.	1
389	50	VAL	CA	C	62.512	.	1
390	50	VAL	HA	H	4.160	.	1
391	50	VAL	CB	C	33.040	.	1
392	50	VAL	HB	H	2.066	.	1
393	50	VAL	CG2	C	20.865	.	1
394	50	VAL	HG21	H	0.945	.	2
395	50	VAL	CG1	C	20.865	.	1
396	50	VAL	HG11	H	0.945	.	2
397	50	VAL	C	C	175.887	.	1
398	51	GLU	N	N	124.823	.	1
399	51	GLU	HN	H	8.512	.	1
400	51	GLU	CA	C	56.900	.	1
401	51	GLU	HA	H	4.353	.	1
402	51	GLU	CB	C	30.428	.	1
403	51	GLU	HB2	H	1.945	.	2
404	51	GLU	HB1	H	2.059	.	2
405	51	GLU	CG	C	36.289	.	1
406	51	GLU	HG2	H	2.271	.	2
407	51	GLU	HG1	H	2.281	.	2
408	51	GLU	C	C	176.365	.	1
409	52	GLN	N	N	121.079	.	1
410	52	GLN	HN	H	8.160	.	1
411	52	GLN	CA	C	55.827	.	1
412	52	GLN	HA	H	4.296	.	1
413	52	GLN	CB	C	29.702	.	1
414	52	GLN	HB2	H	1.961	.	2
415	52	GLN	HB1	H	2.108	.	2
416	52	GLN	CG	C	33.839	.	1
417	52	GLN	HG2	H	2.330	.	2
418	52	GLN	HG1	H	2.330	.	2
419	52	GLN	C	C	176.352	.	1
420	53	GLY	N	N	110.144	.	1
421	53	GLY	HN	H	8.428	.	1
422	53	GLY	CA	C	45.500	.	1
423	53	GLY	HA2	H	3.971	.	2
424	53	GLY	HA1	H	3.938	.	2
425	53	GLY	C	C	174.551	.	1

426 54 GLY N N 108.682 . 1
427 54 GLY HN H 8.281 . 1
428 54 GLY CA C 45.528 . 1
429 54 GLY HA2 H 3.955 . 2
430 54 GLY HA1 H 3.955 . 2
431 54 GLY C C 174.163 . 1
432 55 LEU N N 121.341 . 1
433 55 LEU HN H 8.112 . 1
434 55 LEU CA C 55.605 . 1
435 55 LEU HA H 4.329 . 1
436 55 LEU CB C 42.731 . 1
437 55 LEU HB2 H 1.595 . 2
438 55 LEU HB1 H 1.595 . 2
439 55 LEU CG C 27.034 . 1
440 55 LEU HG H 1.625 . 1
441 55 LEU CD1 C 24.948 . 1
442 55 LEU HD11 H 0.916 . 2
443 55 LEU CD2 C 23.496 . 1
444 55 LEU HD21 H 0.880 . 2
445 55 LEU C C 177.465 . 1
446 56 GLU N N 121.069 . 1
447 56 GLU HN H 8.432 . 1
448 56 GLU CA C 57.000 . 1
449 56 GLU HA H 4.231 . 1
450 56 GLU CB C 30.290 . 1
451 56 GLU HB2 H 1.921 . 2
452 56 GLU HB1 H 2.018 . 2
453 56 GLU CG C 36.289 . 1
454 56 GLU HG2 H 2.281 . 2
455 56 GLU HG1 H 2.271 . 2
456 56 GLU C C 176.349 . 1
457 57 GLU N N 121.814 . 1
458 57 GLU HN H 8.278 . 1
459 57 GLU CA C 56.845 . 1
460 57 GLU HA H 4.317 . 1
461 57 GLU CB C 30.525 . 1
462 57 GLU HB2 H 1.947 . 2
463 57 GLU HB1 H 2.045 . 2
464 57 GLU CG C 36.455 . 1
465 57 GLU HG2 H 2.214 . 2
466 57 GLU HG1 H 2.271 . 2
467 57 GLU C C 176.163 . 1
468 58 ILE N N 122.503 . 1
469 58 ILE HN H 8.144 . 1
470 58 ILE CA C 61.236 . 1
471 58 ILE HA H 4.157 . 1
472 58 ILE CB C 38.907 . 1
473 58 ILE HB H 1.862 . 1
474 58 ILE CG1 C 27.397 . 2
475 58 ILE HG12 H 1.161 . 1
476 58 ILE HG11 H 1.465 . 1
477 58 ILE CD1 C 13.003 . 1
478 58 ILE HD11 H 0.856 . 1
479 58 ILE CG2 C 17.508 . 1
480 58 ILE HG21 H 0.902 . 2
481 58 ILE C C 176.155 . 1
482 59 ARG N N 125.446 . 1
483 59 ARG HN H 8.372 . 1
484 59 ARG CA C 55.295 . 1
485 59 ARG HA H 4.716 . 1
486 59 ARG CB C 41.546 . 1
487 59 ARG HB2 H 3.011 . 2
488 59 ARG HB1 H 3.219 . 2
489 59 ARG C C 174.505 . 1
490 60 GLU N N 119.538 . 1
491 60 GLU HN H 8.143 . 1
492 60 GLU CA C 53.100 . 1
493 60 GLU HA H 4.678 . 1

494 60 GLU CB C 30.412 . 1
495 60 GLU HB2 H 2.026 . 2
496 60 GLU HB1 H 2.677 . 2
497 60 GLU C C 174.489 . 1
498 61 CYS N N 121.850 . 1
499 61 CYS HN H 8.082 . 1
500 61 CYS CA C 57.698 . 1
501 61 CYS HA H 4.547 . 1
502 61 CYS CB C 29.559 . 1
503 61 CYS HB2 H 1.856 . 2
504 61 CYS HB1 H 1.856 . 2
505 61 CYS C C 176.483 . 1
506 62 SER CA C 59.129 . 1
507 62 SER HA H 4.380 . 1
508 62 SER CB C 63.674 . 1
509 62 SER HB2 H 3.880 . 2
510 62 SER HB1 H 3.880 . 2
511 62 SER C C 176.548 . 1
512 63 ALA N N 125.189 . 1
513 63 ALA HN H 8.424 . 1
514 63 ALA CA C 53.047 . 1
515 63 ALA HA H 4.271 . 1
516 63 ALA CB C 19.016 . 1
517 63 ALA HB1 H 1.416 . 1
518 63 ALA C C 177.233 . 1
519 64 GLY N N 106.979 . 1
520 64 GLY HN H 8.187 . 1
521 64 GLY CA C 45.838 . 1
522 64 GLY HA2 H 3.737 . 2
523 64 GLY HA1 H 4.123 . 2
524 64 GLY C C 173.900 . 1
525 65 VAL N N 119.701 . 1
526 65 VAL HN H 7.776 . 1
527 65 VAL CA C 62.272 . 1
528 65 VAL HA H 4.160 . 1
529 65 VAL CB C 32.885 . 1
530 65 VAL HB H 2.097 . 1
531 65 VAL CG2 C 21.155 . 1
532 65 VAL HG21 H 0.900 . 2
533 65 VAL CG1 C 20.956 . 1
534 65 VAL HG11 H 0.900 . 2
535 65 VAL C C 175.838 . 1
536 66 CYS N N 123.861 . 1
537 66 CYS HN H 8.479 . 1
538 66 CYS CA C 55.722 . 1
539 66 CYS HA H 4.719 . 1
540 66 CYS CB C 42.143 . 1
541 66 CYS HB2 H 2.907 . 2
542 66 CYS HB1 H 3.096 . 2
543 66 CYS C C 173.466 . 1
544 67 ALA N N 125.446 . 1
545 67 ALA HN H 8.359 . 1
546 67 ALA CA C 52.504 . 1
547 67 ALA HA H 4.492 . 1
548 67 ALA CB C 19.631 . 1
549 67 ALA HB1 H 1.221 . 1
550 67 ALA C C 176.411 . 1
551 68 VAL N N 122.974 . 1
552 68 VAL HN H 7.980 . 1
553 68 VAL CA C 61.341 . 1
554 68 VAL HA H 4.166 . 1
555 68 VAL CB C 34.985 . 1
556 68 VAL HB H 1.742 . 1
557 68 VAL CG2 C 20.593 . 1
558 68 VAL HG21 H 0.766 . 2
559 68 VAL CG1 C 20.865 . 1
560 68 VAL HG11 H 0.961 . 2
561 68 VAL C C 175.667 . 1

562 69 ASP N N 127.955 . 1
563 69 ASP HN H 8.865 . 1
564 69 ASP CA C 55.126 . 1
565 69 ASP HA H 4.570 . 1
566 69 ASP CB C 40.576 . 1
567 69 ASP HB2 H 2.505 . 2
568 69 ASP HB1 H 2.897 . 2
569 69 ASP C C 176.086 . 1
570 70 ALA N N 120.676 . 1
571 70 ALA HN H 7.690 . 1
572 70 ALA CA C 51.962 . 1
573 70 ALA HA H 3.947 . 1
574 70 ALA CB C 22.737 . 1
575 70 ALA HB1 H 1.327 . 1
576 70 ALA C C 177.419 . 1
577 71 GLY N N 106.590 . 1
578 71 GLY HN H 8.704 . 1
579 71 GLY CA C 45.426 . 1
580 71 GLY HA2 H 3.532 . 2
581 71 GLY HA1 H 4.101 . 2
582 71 GLY C C 174.784 . 1
583 72 CYS N N 118.678 . 1
584 72 CYS HN H 8.280 . 1
585 72 CYS CA C 53.435 . 1
586 72 CYS HA H 4.654 . 1
587 72 CYS CB C 41.962 . 1
588 72 CYS HB2 H 2.637 . 2
589 72 CYS HB1 H 3.092 . 2
590 72 CYS C C 175.326 . 1
591 73 GLY N N 109.155 . 1
592 73 GLY HN H 8.565 . 1
593 73 GLY CA C 46.246 . 1
594 73 GLY HA2 H 3.776 . 2
595 73 GLY HA1 H 3.987 . 2
596 73 GLY C C 172.954 . 1
597 74 VAL N N 114.082 . 1
598 74 VAL HN H 7.875 . 1
599 74 VAL CA C 59.016 . 1
600 74 VAL HA H 4.411 . 1
601 74 VAL CB C 35.528 . 1
602 74 VAL HB H 2.242 . 1
603 74 VAL CG2 C 18.415 . 1
604 74 VAL HG21 H 0.714 . 2
605 74 VAL CG1 C 21.863 . 1
606 74 VAL HG11 H 0.880 . 2
607 74 VAL C C 176.117 . 1
608 75 TRP N N 120.050 . 1
609 75 TRP HN H 8.387 . 1
610 75 TRP HA H 4.540 . 1
611 75 TRP HB2 H 2.908 . 2
612 75 TRP HB1 H 3.085 . 2
613 76 GLY N N 110.458 . 1
614 76 GLY HN H 8.591 . 1
615 76 GLY CA C 43.635 . 1
616 76 GLY HA2 H 3.998 . 2
617 76 GLY HA1 H 4.204 . 2
618 76 GLY C C 172.582 . 1
619 77 GLU N N 116.712 . 1
620 77 GLU HN H 8.290 . 1
621 77 GLU CA C 55.993 . 1
622 77 GLU HA H 4.074 . 1
623 77 GLU CB C 30.492 . 1
624 77 GLU HB2 H 1.868 . 2
625 77 GLU HB1 H 2.055 . 2
626 77 GLU CG C 36.289 . 1
627 77 GLU HG2 H 2.359 . 2
628 77 GLU HG1 H 2.438 . 2
629 77 GLU C C 178.178 . 1

630 78 TRP N N 122.359 . 1
631 78 TRP HN H 8.426 . 1
632 78 TRP CA C 58.402 . 1
633 78 TRP HA H 4.418 . 1
634 78 TRP CB C 31.070 . 1
635 78 TRP HB2 H 2.801 . 2
636 78 TRP HB1 H 3.085 . 2
637 78 TRP HD1 H 6.862 . 1
638 78 TRP HE3 H 7.187 . 3
639 78 TRP C C 175.693 . 1
640 79 SER N N 118.253 . 1
641 79 SER HN H 9.211 . 1
642 79 SER CA C 58.241 . 1
643 79 SER HA H 4.427 . 1
644 79 SER CB C 63.977 . 1
645 79 SER HB2 H 3.841 . 2
646 79 SER HB1 H 3.841 . 2
647 79 SER C C 173.466 . 1
648 80 ALA N N 120.454 . 1
649 80 ALA HN H 8.352 . 1
650 80 ALA CA C 52.504 . 1
651 80 ALA HA H 4.215 . 1
652 80 ALA CB C 19.326 . 1
653 80 ALA HB1 H 1.416 . 1
654 80 ALA C C 179.438 . 1
655 81 TRP N N 122.063 . 1
656 81 TRP HN H 8.356 . 1
657 81 TRP CA C 58.551 . 1
658 81 TRP HA H 4.684 . 1
659 81 TRP CB C 30.877 . 1
660 81 TRP HB2 H 3.060 . 2
661 81 TRP HB1 H 3.271 . 2
662 81 TRP HD1 H 7.170 . 1
663 81 TRP CE3 C 49.184 . 3
664 81 TRP HE3 H 7.300 . 3
665 81 TRP C C 179.806 . 1
666 82 SER N N 117.100 . 1
667 82 SER HN H 9.639 . 1
668 82 SER CA C 60.945 . 1
669 82 SER HA H 4.312 . 1
670 82 SER CB C 63.245 . 1
671 82 SER HB2 H 4.150 . 2
672 82 SER HB1 H 4.150 . 2
673 82 SER C C 173.249 . 1
674 83 ALA N N 121.809 . 1
675 83 ALA HN H 7.536 . 1
676 83 ALA CA C 51.419 . 1
677 83 ALA HA H 4.700 . 1
678 83 ALA CB C 23.667 . 1
679 83 ALA HB1 H 1.660 . 1
680 83 ALA C C 175.574 . 1
681 84 SER N N 110.384 . 1
682 84 SER HN H 8.738 . 1
683 84 SER CA C 57.886 . 1
684 84 SER HA H 4.426 . 1
685 84 SER CB C 63.930 . 1
686 84 SER HB2 H 3.808 . 2
687 84 SER HB1 H 3.922 . 2
688 84 SER C C 173.016 . 1
689 85 CYS N N 116.280 . 1
690 85 CYS HN H 7.669 . 1
691 85 CYS CA C 53.590 . 1
692 85 CYS HA H 4.727 . 1
693 85 CYS CB C 45.683 . 1
694 85 CYS HB2 H 3.124 . 2
695 85 CYS HB1 H 3.219 . 2
696 85 CYS C C 172.071 . 1
697 86 GLY N N 109.592 . 1

698	86	GLY	HN	H	8.632	.	1
699	86	GLY	CA	C	45.575	.	1
700	86	GLY	HA2	H	3.532	.	2
701	86	GLY	HA1	H	3.841	.	2
702	86	GLY	C	C	172.551	.	1
703	87	ASN	N	N	121.129	.	1
704	87	ASN	HN	H	8.216	.	1
705	87	ASN	CA	C	54.210	.	1
706	87	ASN	HA	H	5.025	.	1
707	87	ASN	CB	C	38.784	.	1
708	87	ASN	HB2	H	2.637	.	2
709	87	ASN	HB1	H	2.765	.	2
710	87	ASN	C	C	174.226	.	1
711	88	ALA	N	N	127.154	.	1
712	88	ALA	HN	H	8.768	.	1
713	88	ALA	CA	C	51.246	.	1
714	88	ALA	HA	H	4.790	.	1
715	88	ALA	CB	C	25.205	.	1
716	88	ALA	HB1	H	1.399	.	1
717	88	ALA	C	C	175.807	.	1
718	89	THR	N	N	111.051	.	1
719	89	THR	HN	H	9.095	.	1
720	89	THR	CA	C	59.826	.	1
721	89	THR	HA	H	5.948	.	1
722	89	THR	CB	C	73.108	.	1
723	89	THR	HB	H	4.182	.	1
724	89	THR	CG2	C	21.715	.	1
725	89	THR	HG21	H	1.233	.	1
726	89	THR	C	C	173.931	.	1
727	90	ARG	N	N	118.486	.	1
728	90	ARG	HN	H	8.673	.	1
729	90	ARG	CA	C	55.349	.	1
730	90	ARG	HA	H	4.508	.	1
731	90	ARG	CB	C	33.040	.	1
732	90	ARG	HB2	H	0.635	.	2
733	90	ARG	HB1	H	1.384	.	2
734	90	ARG	CD	C	43.344	.	1
735	90	ARG	C	C	173.481	.	1
736	91	LYS	N	N	117.924	.	1
737	91	LYS	HN	H	8.513	.	1
738	91	LYS	CA	C	55.275	.	1
739	91	LYS	HA	H	5.876	.	1
740	91	LYS	CB	C	37.650	.	1
741	91	LYS	HB2	H	1.749	.	2
742	91	LYS	HB1	H	1.749	.	2
743	91	LYS	CG	C	24.222	.	1
744	91	LYS	HG2	H	1.318	.	2
745	91	LYS	HG1	H	1.396	.	2
746	91	LYS	CD	C	29.303	.	1
747	91	LYS	HD2	H	1.557	.	2
748	91	LYS	HD1	H	1.557	.	2
749	91	LYS	CE	C	41.914	.	1
750	91	LYS	HE2	H	2.822	.	2
751	91	LYS	HE1	H	2.839	.	2
752	91	LYS	C	C	175.233	.	1
753	92	ARG	N	N	119.563	.	1
754	92	ARG	HN	H	8.724	.	1
755	92	ARG	CA	C	54.365	.	1
756	92	ARG	HA	H	4.719	.	1
757	92	ARG	CB	C	33.786	.	1
758	92	ARG	HB2	H	0.603	.	2
759	92	ARG	HB1	H	1.010	.	2
760	92	ARG	HG1	H	1.108	.	2
761	92	ARG	C	C	173.311	.	1
762	93	GLU	N	N	117.955	.	1
763	93	GLU	HN	H	8.842	.	1
764	93	GLU	CA	C	54.305	.	1
765	93	GLU	HA	H	5.663	.	1

766 93 GLU CB C 34.457 . 1
767 93 GLU HB2 H 1.927 . 2
768 93 GLU HB1 H 2.124 . 2
769 93 GLU CG C 36.380 . 1
770 93 GLU HG2 H 2.259 . 2
771 93 GLU HG1 H 2.259 . 2
772 93 GLU C C 177.248 . 1
773 94 ARG N N 122.100 . 1
774 94 ARG HN H 8.403 . 1
775 94 ARG CA C 54.156 . 1
776 94 ARG HA H 4.564 . 1
777 94 ARG CB C 29.890 . 1
778 94 ARG HB2 H 1.479 . 2
779 94 ARG HB1 H 1.896 . 2
780 94 ARG CG C 27.307 . 1
781 94 ARG HG1 H 1.445 . 2
782 94 ARG CD C 43.275 . 1
783 94 ARG C C 178.457 . 1
784 95 THR N N 114.504 . 1
785 95 THR HN H 9.222 . 1
786 95 THR CA C 64.024 . 1
787 95 THR HA H 4.262 . 1
788 95 THR CB C 69.153 . 1
789 95 THR HB H 4.003 . 1
790 95 THR CG2 C 22.806 . 1
791 95 THR HG21 H 1.243 . 1
792 95 THR C C 174.799 . 1
793 96 ARG N N 118.158 . 1
794 96 ARG HN H 7.420 . 1
795 96 ARG CA C 55.311 . 1
796 96 ARG HA H 4.313 . 1
797 96 ARG CB C 33.310 . 1
798 96 ARG HB2 H 1.573 . 2
799 96 ARG HB1 H 1.938 . 2
800 96 ARG CG C 26.944 . 1
801 96 ARG HG2 H 1.606 . 2
802 96 ARG HG1 H 1.606 . 2
803 96 ARG CD C 43.457 . 1
804 96 ARG HD2 H 3.149 . 2
805 96 ARG HD1 H 3.149 . 2
806 96 ARG C C 173.776 . 1
807 97 TYR N N 119.295 . 1
808 97 TYR HN H 8.349 . 1
809 97 TYR CA C 60.489 . 1
810 97 TYR HA H 4.284 . 1
811 97 TYR CB C 41.962 . 1
812 97 TYR HB2 H 2.555 . 2
813 97 TYR HB1 H 2.816 . 2
814 97 TYR CD1 C 62.853 . 3
815 97 TYR HD1 H 6.910 . 3
816 97 TYR CD2 C 62.853 . 3
817 97 TYR HD2 H 6.910 . 3
818 97 TYR C C 174.055 . 1
819 98 ASN N N 126.250 . 1
820 98 ASN HN H 7.447 . 1
821 98 ASN CA C 55.683 . 1
822 98 ASN HA H 4.246 . 1
823 98 ASN CB C 36.458 . 1
824 98 ASN HB2 H 2.617 . 2
825 98 ASN HB1 H 2.767 . 2
826 98 ASN C C 173.001 . 1
827 99 ASP N N 116.730 . 1
828 99 ASP HN H 7.537 . 1
829 99 ASP CA C 51.574 . 1
830 99 ASP HA H 4.492 . 1
831 99 ASP CB C 44.132 . 1
832 99 ASP HB2 H 2.344 . 2
833 99 ASP HB1 H 2.376 . 2

834 99 ASP C C 173.202 . 1
835 102 PRO CA C 62.662 . 1
836 102 PRO HA H 4.329 . 1
837 102 PRO CB C 31.796 . 1
838 102 PRO HB2 H 1.672 . 2
839 102 PRO HB1 H 1.770 . 2
840 102 PRO CG C 27.488 . 1
841 102 PRO HG2 H 1.191 . 2
842 102 PRO HG1 H 1.634 . 2
843 102 PRO CD C 49.614 . 1
844 103 GLN N N 117.483 . 1
845 103 GLN HN H 8.767 . 1
846 103 GLN CA C 55.600 . 1
847 103 GLN HA H 4.675 . 1
848 103 GLN CB C 31.947 . 1
849 103 GLN HB2 H 1.702 . 2
850 103 GLN HB1 H 2.114 . 2
851 103 GLN CG C 37.015 . 1
852 103 GLN HG2 H 2.595 . 2
853 103 GLN HG1 H 2.653 . 2
854 103 GLN C C 176.675 . 1
855 104 GLY N N 114.914 . 1
856 104 GLY HN H 9.232 . 1
857 104 GLY CA C 45.760 . 1
858 104 GLY HA2 H 3.759 . 2
859 104 GLY HA1 H 3.800 . 2
860 104 GLY C C 173.125 . 1
861 105 ALA N N 129.300 . 1
862 105 ALA HN H 8.642 . 1
863 105 ALA CA C 51.962 . 1
864 105 ALA HA H 4.524 . 1
865 105 ALA CB C 17.621 . 1
866 105 ALA HB1 H 1.368 . 1
867 105 ALA C C 177.403 . 1
868 106 GLY N N 106.660 . 1
869 106 GLY HN H 8.067 . 1
870 106 GLY CA C 44.505 . 1
871 106 GLY HA2 H 3.808 . 2
872 106 GLY HA1 H 3.808 . 2
873 106 GLY C C 172.179 . 1
874 107 ARG N N 119.582 . 1
875 107 ARG HN H 9.155 . 1
876 107 ARG CA C 55.581 . 1
877 107 ARG HA H 4.605 . 1
878 107 ARG CB C 32.128 . 1
879 107 ARG HB2 H 1.419 . 2
880 107 ARG HB1 H 1.583 . 2
881 107 ARG CG C 26.441 . 1
882 107 ARG CD C 43.616 . 1
883 107 ARG HD2 H 3.813 . 2
884 107 ARG HD1 H 3.869 . 2
885 107 ARG C C 179.124 . 1
886 108 ARG N N 124.693 . 1
887 108 ARG HN H 8.924 . 1
888 108 ARG CA C 57.140 . 1
889 108 ARG HA H 3.986 . 1
890 108 ARG CB C 31.400 . 1
891 108 ARG HB2 H 1.445 . 2
892 108 ARG HB1 H 2.367 . 2
893 108 ARG CG C 29.031 . 1
894 108 ARG HG2 H 1.711 . 2
895 108 ARG HG1 H 2.036 . 2
896 108 ARG CD C 43.366 . 1
897 108 ARG HD2 H 3.353 . 2
898 108 ARG HD1 H 3.353 . 2
899 108 ARG C C 177.341 . 1
900 109 CYS N N 123.248 . 1
901 109 CYS HN H 9.863 . 1

902 109 CYS CA C 61.990 . 1
903 109 CYS HA H 3.776 . 1
904 109 CYS CB C 43.859 . 1
905 109 CYS HB2 H 1.986 . 2
906 109 CYS HB1 H 2.734 . 2
907 109 CYS C C 176.752 . 1
908 110 GLU N N 120.735 . 1
909 110 GLU HN H 11.303 . 1
910 110 GLU CA C 58.707 . 1
911 110 GLU HA H 4.032 . 1
912 110 GLU CB C 27.765 . 1
913 110 GLU HB2 H 1.955 . 2
914 110 GLU HB1 H 2.330 . 2
915 110 GLU CG C 35.654 . 1
916 110 GLU HG2 H 2.166 . 2
917 110 GLU HG1 H 2.556 . 2
918 110 GLU C C 175.295 . 1
919 111 ASN N N 116.009 . 1
920 111 ASN HN H 7.871 . 1
921 111 ASN CA C 52.340 . 1
922 111 ASN HA H 5.166 . 1
923 111 ASN CB C 40.799 . 1
924 111 ASN HB2 H 2.388 . 2
925 111 ASN HB1 H 3.230 . 2
926 111 ASN C C 174.660 . 1
927 112 GLN N N 119.848 . 1
928 112 GLN HN H 6.899 . 1
929 112 GLN CA C 56.761 . 1
930 112 GLN HA H 3.870 . 1
931 112 GLN CB C 28.492 . 1
932 112 GLN HB2 H 2.124 . 2
933 112 GLN HB1 H 2.124 . 2
934 112 GLN CG C 35.347 . 1
935 112 GLN HG2 H 2.686 . 2
936 112 GLN HG1 H 2.683 . 2
937 112 GLN C C 172.303 . 1
938 113 ASP N N 116.313 . 1
939 113 ASP HN H 7.807 . 1
940 113 ASP CA C 50.101 . 1
941 113 ASP HA H 5.012 . 1
942 113 ASP CB C 43.047 . 1
943 113 ASP HB2 H 2.416 . 2
944 113 ASP HB1 H 2.767 . 2
945 113 ASP C C 173.729 . 1
946 115 PRO CA C 63.667 . 1
947 115 PRO HA H 4.036 . 1
948 115 PRO CB C 32.431 . 1
949 115 PRO HB2 H 1.682 . 2
950 115 PRO HB1 H 2.310 . 2
951 115 PRO CG C 27.216 . 1
952 115 PRO HG2 H 2.049 . 2
953 115 PRO HG1 H 2.049 . 2
954 115 PRO CD C 50.250 . 1
955 115 PRO C C 177.016 . 1
956 116 VAL N N 123.964 . 1
957 116 VAL HN H 7.906 . 1
958 116 VAL CA C 61.962 . 1
959 116 VAL HA H 4.101 . 1
960 116 VAL CB C 33.047 . 1
961 116 VAL HB H 1.896 . 1
962 116 VAL CG2 C 20.774 . 1
963 116 VAL HG21 H 0.766 . 2
964 116 VAL CG1 C 20.774 . 1
965 116 VAL HG11 H 0.879 . 2
966 116 VAL C C 175.683 . 1
967 117 LEU N N 130.738 . 1
968 117 LEU HN H 8.790 . 1
969 117 LEU CA C 54.379 . 1

970 117 LEU HA H 4.506 . 1
971 117 LEU CB C 42.217 . 1
972 117 LEU HB2 H 1.829 . 2
973 117 LEU HB1 H 1.919 . 2
974 117 LEU CG C 25.038 . 1
975 117 LEU HG H 1.102 . 1
976 117 LEU CD1 C 22.679 . 1
977 117 LEU HD11 H 0.974 . 2
978 117 LEU CD2 C 24.948 . 1
979 117 LEU HD21 H 1.096 . 2
980 117 LEU C C 175.233 . 1
981 118 GLN N N 117.249 . 1
982 118 GLN HN H 7.962 . 1
983 118 GLN CA C 58.765 . 1
984 118 GLN HA H 3.143 . 1
985 118 GLN CB C 28.255 . 1
986 118 GLN HB2 H 0.642 . 2
987 118 GLN HB1 H 1.439 . 2
988 118 GLN CG C 33.711 . 1
989 118 GLN HG2 H 2.073 . 2
990 118 GLN HG1 H 2.073 . 2
991 118 GLN C C 174.846 . 1
992 119 GLU N N 112.328 . 1
993 119 GLU HN H 7.469 . 1
994 119 GLU CA C 53.976 . 1
995 119 GLU HA H 5.492 . 1
996 119 GLU CB C 34.066 . 1
997 119 GLU HB2 H 1.929 . 2
998 119 GLU HB1 H 1.929 . 2
999 119 GLU CG C 36.198 . 1
1000 119 GLU HG2 H 2.100 . 2
1001 119 GLU HG1 H 2.100 . 2
1002 119 GLU C C 174.412 . 1
1003 120 GLN N N 122.804 . 1
1004 120 GLN HN H 8.661 . 1
1005 120 GLN CA C 55.275 . 1
1006 120 GLN HA H 4.607 . 1
1007 120 GLN CB C 32.765 . 1
1008 120 GLN HB2 H 1.156 . 2
1009 120 GLN HB1 H 1.416 . 2
1010 120 GLN CG C 33.749 . 1
1011 120 GLN HG2 H 2.290 . 2
1012 120 GLN HG1 H 2.281 . 2
1013 120 GLN C C 173.962 . 1
1014 121 THR N N 117.548 . 1
1015 121 THR HN H 8.540 . 1
1016 121 THR CA C 61.020 . 1
1017 121 THR HA H 5.459 . 1
1018 121 THR CB C 71.392 . 1
1019 121 THR HB H 4.069 . 1
1020 121 THR CG2 C 21.854 . 1
1021 121 THR HG21 H 1.223 . 1
1022 121 THR C C 174.024 . 1
1023 122 GLU N N 124.697 . 1
1024 122 GLU HN H 8.938 . 1
1025 122 GLU CA C 55.311 . 1
1026 122 GLU HA H 4.694 . 1
1027 122 GLU CB C 34.129 . 1
1028 122 GLU HB2 H 1.448 . 2
1029 122 GLU HB1 H 1.539 . 2
1030 122 GLU CG C 37.015 . 1
1031 122 GLU HG2 H 2.017 . 2
1032 122 GLU HG1 H 2.115 . 2
1033 122 GLU C C 173.932 . 1
1034 123 GLU N N 121.484 . 1
1035 123 GLU HN H 8.411 . 1
1036 123 GLU CA C 54.603 . 1
1037 123 GLU HA H 5.467 . 1

1038 123 GLU CB C 32.856 . 1
1039 123 GLU HB2 H 1.905 . 2
1040 123 GLU HB1 H 2.063 . 2
1041 123 GLU CG C 36.561 . 1
1042 123 GLU HG2 H 2.201 . 2
1043 123 GLU HG1 H 2.201 . 2
1044 123 GLU C C 175.295 . 1
1045 124 ALA N N 125.396 . 1
1046 124 ALA HN H 9.017 . 1
1047 124 ALA CA C 51.619 . 1
1048 124 ALA HA H 4.524 . 1
1049 124 ALA CB C 22.967 . 1
1050 124 ALA HB1 H 0.710 . 1
1051 124 ALA C C 175.652 . 1
1052 125 THR N N 117.216 . 1
1053 125 THR HN H 8.350 . 1
1054 125 THR CA C 61.962 . 1
1055 125 THR HA H 4.313 . 1
1056 125 THR CB C 69.171 . 1
1057 125 THR HB H 3.890 . 1
1058 125 THR CG2 C 22.037 . 1
1059 125 THR HG21 H 1.213 . 1
1060 125 THR C C 174.179 . 1
1061 126 LEU N N 128.873 . 1
1062 126 LEU HN H 8.296 . 1
1063 126 LEU CA C 52.857 . 1
1064 126 LEU HA H 4.296 . 1
1065 126 LEU CB C 41.807 . 1
1066 126 LEU HB2 H 1.357 . 2
1067 126 LEU HB1 H 1.622 . 2
1068 126 LEU CG C 26.873 . 1
1069 126 LEU HG H 1.107 . 1
1070 126 LEU CD1 C 25.231 . 1
1071 126 LEU HD11 H 0.163 . 2
1072 126 LEU CD2 C 21.409 . 1
1073 126 LEU HD21 H 0.144 . 2
1074 126 LEU C C 175.419 . 1
1075 127 ALA N N 122.241 . 1
1076 127 ALA HN H 7.778 . 1
1077 127 ALA CA C 51.419 . 1
1078 127 ALA HA H 4.339 . 1
1079 127 ALA CB C 17.388 . 1
1080 127 ALA HB1 H 1.360 . 1
1081 127 ALA C C 174.923 . 1
1082 128 PRO CA C 63.245 . 1
1083 128 PRO HA H 4.288 . 1
1084 128 PRO CB C 32.257 . 1
1085 128 PRO HB2 H 1.789 . 2
1086 128 PRO HB1 H 2.279 . 2
1087 128 PRO CG C 27.986 . 1
1088 128 PRO HG2 H 1.975 . 2
1089 128 PRO HG1 H 2.024 . 2
1090 128 PRO CD C 50.250 . 1
1091 128 PRO C C 177.450 . 1
1092 129 CYS N N 120.884 . 1
1093 129 CYS HN H 8.842 . 1
1094 129 CYS CA C 54.365 . 1
1095 129 CYS HA H 4.518 . 1
1096 129 CYS CB C 41.342 . 1
1097 129 CYS HB2 H 2.691 . 2
1098 129 CYS HB1 H 3.172 . 2
1099 129 CYS C C 174.985 . 1
1100 130 ILE N N 122.845 . 1
1101 130 ILE HN H 8.383 . 1
1102 130 ILE CA C 61.064 . 1
1103 130 ILE HA H 4.271 . 1
1104 130 ILE CB C 38.800 . 1
1105 130 ILE HB H 1.880 . 1

1106 130 ILE CG1 C 27.350 . 2

1107 130 ILE HG12 H 1.164 . 1

1108 130 ILE HG11 H 1.449 . 1

1109 130 ILE CD1 C 13.082 . 1

1110 130 ILE HD11 H 0.848 . 1

1111 130 ILE CG2 C 17.808 . 1

1112 130 ILE HG21 H 0.896 . 2

1113 130 ILE C C 176.008 . 1

1114 131 THR N N 119.903 . 1

1115 131 THR HN H 8.261 . 1

1116 131 THR CA C 61.542 . 1

1117 131 THR HA H 4.377 . 1

1118 131 THR CB C 69.791 . 1

1119 131 THR HB H 4.170 . 1

1120 131 THR CG2 C 21.534 . 1

1121 131 THR HG21 H 1.182 . 1

1122 131 THR C C 173.373 . 1

1123 132 ILE N N 128.071 . 1

1124 132 ILE HN H 7.756 . 1

1125 132 ILE CA C 62.814 . 1

1126 132 ILE HA H 4.084 . 1

1127 132 ILE CB C 39.481 . 1

1128 132 ILE HB H 1.838 . 1

1129 132 ILE CG1 C 27.216 . 2

1130 132 ILE HG12 H 1.154 . 1

1131 132 ILE HG11 H 1.439 . 1

1132 132 ILE CD1 C 13.425 . 1

1133 132 ILE HD11 H 0.856 . 1

1134 132 ILE CG2 C 17.689 . 1

1135 132 ILE HG21 H 0.896 . 2

1136 132 ILE C C 180.876 . 1