that comprises the calreticulin binding site for defence colla

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Alexandre R. Gingras BSc, MSc
Department of Microbiology and Immunology

May 2004

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Title : NMR structure of the S-domain of calreticulin, the sub-fragment that comprises the calreticulin binding site for defence collagens.

Author : Alexandre R. Gingras

Calreticulin (CRT) is a highly conserved protein present in every cell of higher organisms, except erythrocytes. It is associated with lectin-like chaperoning, calcium storage and signalling, modulation of gene expression, induction of phagocytosis of apoptotic cells, mediation of autoimmunity, anti-angiogenesis, and inhibition of C1q-dependent complement activation *in vitro*. The human CRT binds to the collagenous region of C1q and members of the collectin family, such as MBL, lung surfactant protein A, and collectin 43.

Using overlapping sub-fragments, i.e. P-, S-, and PS-domains of CRT in different binding assays, we determined that the S-domain has the strongest affinity to bind to C1q and MBL. This makes the structure of the S-domain most relevant for understanding the binding interactions to C1q and MBL and the steric inhibition of complement activation by competing with the respective C1q or MBL associated serine proteases for binding. The CRT S-domain structure determined in this work has an extended hairpin fold stabilised by a short 3 residue β-sheet holding the two opposite strands together. Two hydrophobic cores, each containing two conserved tryptophan residues, are located on either side of this β-sheet and further stabilise this fold. Comparison of 2D [^{1}H,^{15}N]-HSQC spectra indicated higher internal mobility of the S-domain than of the other active fragments, suggesting that the dynamic properties are important for the protein activity. Detailed multi-field relaxation analysis showed high dynamics in the hairpin with more restrictions where secondary structure elements are present.

The interaction with MBL and C1q was determined to be optimal under low ionic strength conditions, indicative of electrostatic interactions. The interaction probably involves the tip of the hairpin in the CRT S-domain structure that has a high concentration of negatively charged residues, and a complementary positively charged region of the MBL and C1q collagen-like domains. The flexibility of the hairpin is likely to be a key element for CRT, allowing it to adopt its conformation for binding to multiple ligands and is expected to be the key factor for the enhanced activity of the S-domain.

Acknowledgements

First I would like to thank my supervisor, Professor Wilhelm Schwaeble, for allowing me to do a Ph.D. in his laboratory which allowed me to learn protein crystallography and NMR. He also provided funding for the research and supervised this project. In the x-ray crystallography lab, Dr. Peter Moody, provided lab space in the crystallography laboratory for two years and also provided funding for the research. Dr. Geoff Briggs showed me how to use the FPLC and other techniques. Dr. Jonas Emsley read and made positive comments on the crystallography chapter. I thank Dr. Igor Barsukov, in the NMR centre, for his patience in teaching me protein NMR. He also gave me access to a computer, collected data for NMR structure determination, and showed me how to use multiple programs to process the data. He also helped in the determination of the protein structure and the processing of the relaxation data. I thank Dr. Mark Pfuhl, who designed the relaxation experiments performed in Leicester and gave me access to his MATHEMATICA scripts for the model-free analysis, and Dr. Geoff Kelly at the MRC Biomedical Centre, Mill Hill, London, who designed NMR experiments and collected data for me on the 800MHz.

In the Microbiology and Immunology lab, I am enormously grateful to Sotiria Tzima for helping me to do my first successful cloning and for her enormous help in designing relevant binding assays for studying CRT-MBL and CRT-C1q complex formation, Dr. Cordula Stover who read and made comments on chapters of this thesis, and Dr. Nick Lynch who provided the first clones of CRT S-domain and who read chapters of this thesis.

Special thanks to Catherine Vial, my girlfriend, who supported me through so many years of writing up. Finally, I would like to thank my parents for their love and support.

CONTENTS

FIGURES

TABLES

Abbreviations

°C: Degrees Celsius

A_{600nm}: Absorbance at 600nm

Amp: Ampicillin

ARIA: Ambiguous Restraints for Iterative Assignment

CD: Circular Dichroism

CNX: Calnexin

CRT: Calreticulin

CUB: a 110-amino acid protein module occurring in C1r, C1s, spermadhesins, and other proteins.

EDTA: EthyleneDiamineTetraacetic Acid

ELISA: Enzyme-Linked ImmunoSorbent Assay

ER: Endoplasmic Reticulum

ERp57: Endoplasmic Reticulum protein of a 57 kDa molecular weight

HSQC: Heteronuclear Single Quantum Coherence (a 2D NMR spectrum often involving ^{15}N and ^{1}H chemical shifts)

HMQC: Heteronuclear Multiple Quantum Coherence (a 2D NMR spectrum often involving ^{15}N and ^{1}H chemical shifts)

ICP-OES: Inductively Coupled Plasma-Optical Emission Spectrometry

Hz: Hertz

IPTG: IsoPropyl-beta-D-ThioGalactopyranoside

K: Kelvin

kDa: KiloDaltons (1000 Daltons)

MASP: MBL-Associated Serine Proteases

MBL: Mannan-Binding Lectin

NMR: Nuclear Magnetic Resonance

NOE: Nuclear Overhauser Effect

NOESY: Nuclear Overhauser Effect SpectroscopY

PDB: Protein Data Bank

RA: Rheumatoid Arthritis

SDS-PAGE: SDS PolyAcrylamide Gel Electrophoresis

SP: Serine Protease

TALOS: Torsion Angle Likelihood Obtained from Shift and sequence similarity

TOCSY: TOtal Correlated SpectroscopY

CHAPTER 1 :
INTRODUCTION

1.1 Calreticulin

Calreticulin (CRT) is an ancient and highly conserved protein. It was first identified by Ostwald and MacLennan [1] in 1974 as a Ca^{2+}-binding protein of the muscle sarcoplasmic reticulum. The molecular cloning of the protein followed in 1989 by Smith and Koch [2] and Fliegel *et al.* [3]. Since then, it has been intensively studied by many laboratories and has been assigned multiple functions (see Table I.I). It became clear that this protein is highly abundant also in nonmuscle tissues and that it is one of the major Ca^{2+}-binding proteins of the endoplasmic reticulum (ER). Subsequently, calreticulin has been characterised as a molecular chaperone, a lectin, an intracellular mediator of integrin function, an inhibitor of steroid hormone-regulated gene expression, a C1q- and collectin-binding protein with several other cell-surface receptor-like functions. The fact that one protein can perform so many functions is at once intriguing and controversial and further investigations are required in order fully to understand the functions of calreticulin. The aim of the present introduction is to consider the latest developments in the calreticulin field, with focus on the physiological functions of the protein.

Table I.I : Putative functions of calreticulin domains. The figure 1.1 shows a schematic representation of the human calreticulin the N-, P- and C-domains. Table adapted from reference [4].

(a) Structural features and function

N-domain	P-domain	C-domain
1. Proceeded by the N-terminal signal sequence targeting the protein to the ER lumen	1. Proline-rich domain	1. Rich in acidic amino acids
2. Highly conserved amino acid sequence	2. Amino acid sequence similarity to calnexin, calmegin and CANLUC	2. ER retrieval signal
3. Potential phosphorylation site	3. Putative glycosylation site (*Leishmania* protein)	3. Putative glycosylation site
4. Potential glycosylation site (bovine proteins)		4. Antithrombotic activity
5. Putative autokinase activity		5. Prevents restenosis
6. Inhibits PDI activity		6. Ca2+ 'sensor' of calreticulin-protein interactions
7. Suppresses tumours		
8. Inhibits angiogenesis		

(b) Ion binding

N-domain	P-domain	C-domain
1. Binds Zn2+	1. High-affinity Ca2+ -binding site	1. High-capacity Ca2+-binding site

(c) Molecules binding

N-domain	P-domain	C-domain
1. Binds to the DNA-binding domain of steroid receptor	1. Binds to a set of ER proteins	1. Binds a set of ER proteins
2. Binds to α-subunit of integrin	2. Strong interactions with PDI	2. Binds Factor IX and Factor X
3. Binds rubella RNA	3. Strong interactions with perforin	3. Binds to cell surface
4. Interacts with PDI	4. Lectin-like chaperone site	
5. Interacts with ERp57	5. Interacts with ERp57	
6. Weak interactions with perforin		

1.1.1 The gene

The human calreticulin gene has been localised to chromosome 19 and consists of nine exons that span approximately 3.6 kb of genomic DNA (Figure 1.1) [5]. The nucleotide sequence of the mouse and the human gene show more than 70% identity, indicating a strong evolutionary conservation of the gene. To date, cDNA and genomic sequences of genes encoding calreticulin have been characterised from several mammals, insects, nematodes, *protozoa* and also plants. The degree of conservation indicates that it has an important role in cellular function. There is no calreticulin gene in yeast and prokaryotes, whose genomes have been fully sequenced, suggesting that the protein must have emerged later during evolution.

1.1.2 The protein

The human calreticulin amino acid sequence indicates a 17 amino acid hydrophobic signal sequence at its N-terminus [3] and that the mature protein contains 400 amino acids with a calculated molecular weight of 46.6 kDa. Anomalous migration on SDS-PAGE, likely due to its highly negative charge (pI = 4.7) and/or its high proline content, results in calreticulin's appearance as an approximately 60 kDa protein. Analysis of the primary sequence of the protein has led to it being divided into three domains (Figure 1.1) [2, 3]. The N-terminal 180 residues are termed the N-domain and are predicted to form a globular structure with eight anti-parallel β-sheets and a helix-turn-helix motif. Residues 198-308 are rich in proline and contain two sets of repeated amino acid sequences unique to CRT and its molecular homologues calnexin (CNX) and calmegin. The C-terminal 110 amino acids of calreticulin are predominantly negatively charged with 37 of its final 57 residues are aspartic or glutamic acid. The C-terminus of calreticulin contains a KDEL endoplasmic reticulum retrieval sequence, responsible for targeting and retention of calreticulin in the ER lumen.

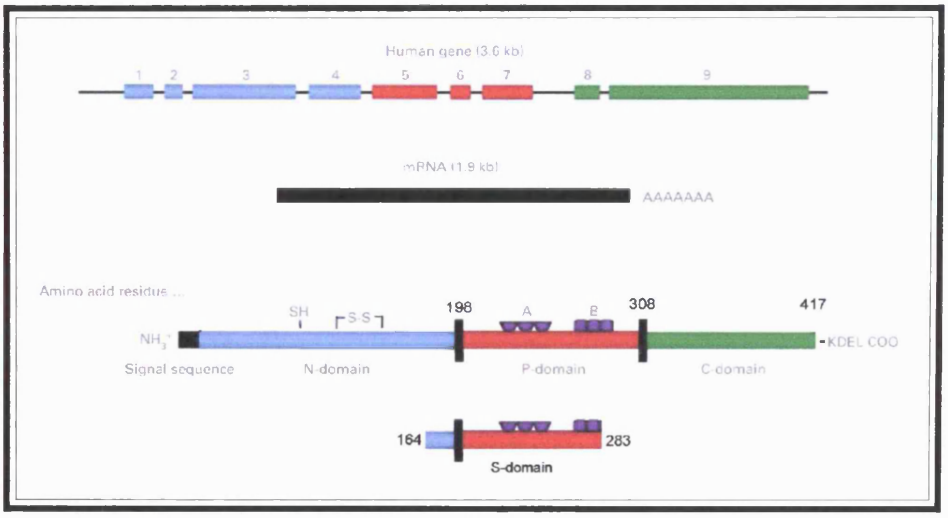

Figure 1.1 : The calreticulin gene, mRNA and protein. The figure shows a schematic representation of the genomic configuration of the human calreticulin gene, mRNA and the N-, S-, P- and C-domains of the calreticulin protein. The protein contains an N-terminal signal sequence, shown in black and a C-terminal KDEL endoplasmic reticulum retrieval signal. The location of the three cysteine residues and the disulfide bridge are indicated. Repeats A (amino acid sequence PXXIXDPDAXKPEDWDE) and B (amino acid sequence GXWXPPXIXNPXYX) are indicated by the purple triangles and purple squares respectively. Figure adapted from reference [4].

1.1.2.1 N-domain

The amino acid sequence of the N-domain is extremely conserved in all calreticulins (Figure 1.2). Within this domain are a binding site for rubella virus RNA [6], a putative phosphorylation site and a segment which binds to steroid receptors [7] and the cytoplasmic domains of integrin α subunits [8] (Table I.I). It also contains low-affinity, high-capacity Zinc binding sites (K_d = 310 μM and 14 mol of zinc/mol of protein) [9]. It has been shown that the N-domain mediates interactions between CRT and the ER folding catalysts protein disulphide isomerase (PDI) and ERp57 [10, 11].

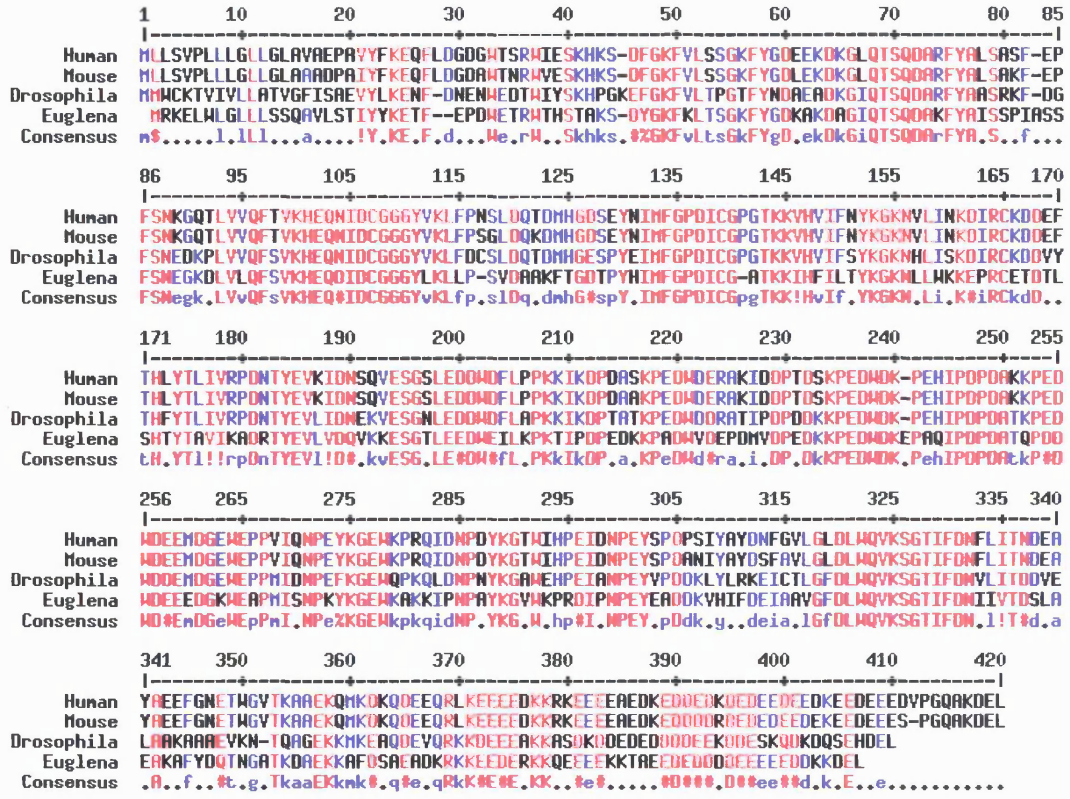

Figure 1.2 : Comparison of the amino acid sequences of selected calreticulin proteins. Amino acid sequence of selected calreticulins were compared using MultAlin [12]. The amino acid sequences of different calreticulins are extremely similar. The red represent 100% consensus, the blue represents 50-90% consensus and the black <50% consensus. cDNAs encoding calreticulin are as follows : human (GenBank accession number M84739), mouse (X14926), Euglena (a unicellular alga) (Y09816) and *Drosophila melanogaster* (fruitfly) (X64461).

1.1.2.2 P-domain

The P-domain is made of a proline-, aspartic acid- and glutamic acid-rich segment with three repeats A of the amino acid sequence PXXIXDPDAXKPEDWDE followed by three repeats B of the sequence GXWXPPXIXNPXYX (Figure 1.1). This domain contains a high-affinity, low capacity Ca^{2+}-binding site (1 mol Ca^{2+}/mol of protein, K_d = ~1μM) [13], but does not have an EF-hand consensus sequence typical of several other well known high-affinity Ca^{2+}-binding proteins. The repeats A and B may be essential for the high-affinity Ca^{2+}-binding of CRT. More importantly, those repeats are critical for the lectin-like chaperone activity of the protein [14]. The P-domain is one of the most interesting regions of the protein because of its lectin-like activity and amino acid sequence similarity with calnexin, an integral ER membrane Ca^{2+}-binding chaperone (Figure 1.3) [15]. Calnexin shares several regions of similarity with the CRT, but the most striking sequences that are highly conserved among different calreticulins are the repeats A and B found in the P-domain of calreticulin, both of which are present in calnexin. The P-domain of calreticulin interacts with PDI [10], ERp57 and perforin [16], a component of the cytotoxic T-cell granules (Table I.I).

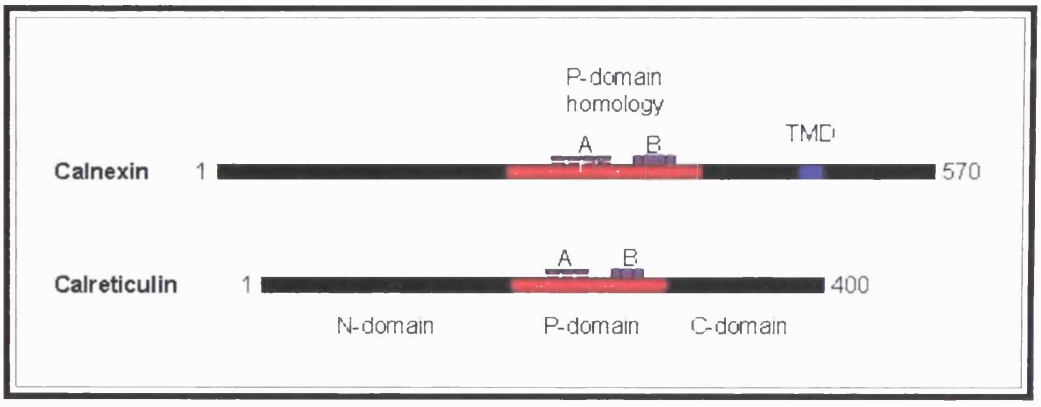

Figure 1.3 : Schematic representation of CRT and CNX. Calreticulin is a soluble 400 amino acid protein, whereas calnexin is a membrane-bound protein of 570 residues. The transmembrane domain (TMD) of calnexin is shown in blue. There is a high degree of amino acid sequence similarity between the P-domain of both proteins. The P-domain is a site of chaperone activity and oligosaccharide binding in both proteins with two sequence repeat types, designated A and B, each repeated three times in CRT and four times in CNX.

1.1.2.3 C-domain

The C-terminal region of calreticulin (the C-domain, figure 1.1) contains the low-affinity, high capacity Ca^{2+}-binding site (25 mol Ca^{2+}/mol protein, K_d = 2 mM) [13]. Calcium binding to this domain has a regulatory role in the control of CRT interactions with PDI, ERp57 and perhaps other chaperones [11]. The C-domain of CRT also binds to vitamin K-dependent blood-clotting factors [17] and inhibits injury-induced restenosis [18] (Table I.I). This domain also contains a potential site for N-linked glycosylation (Asp-327).

1.1.2.4 S-domain

It was demonstrated that the C1q binding site within CRT lies across the intersection of the N- and P-domains (figure 1.1) [19]. CRT and $C1r_2C1s_2$ cross-compete for binding to C1q, indicating that they possibly have overlapping binding site on C1q [20]. Given that CRT, C1r and C1s all interact with C1q, a sequence comparison was performed to investigate the structural basis for this interaction [21]. A region that may correspond to a CUB module was identified in CRT and was analysed by multiple sequence alignment. This region, termed the S region (C1s-like CUB domain), spans the intersection of the N- and P-domains (residues 160-283). Furthermore, competitive inhibition studies of the S-domain-C1q interaction revealed that the S-domain binds to C1q collagen tails and to the collectin proteins, SP-A, MBL, CL43 and conglutinin [19, 21].

1.1.3 Cellular localisation

Calreticulin contains a C-terminal KDEL ER retrieval sequence and was originally identified as a muscle sarcoplasmic reticulum protein [1]. Numerous studies confirmed the ER localisation of the protein and not surprisingly it is the most frequently reported localisation in many species, including plants. However, the protein has also been localised on the cell surface, in the cytosol or nucleus or both and within the secretory pathway. There is still considerable controversy in the literature concerning the cellular localisation of the protein. It is reported that calreticulin interacts *in vitro* with the cytoplasmic tail of the α subunit of integrins [8] and with the DNA-binding domain of steroid receptors [7]. However, for CRT to bind to these molecules, the protein would have to be present in the nucleus and/or cytosol.

Calreticulin-like immunoreactivity has been detected in the nucleus of some cells [22]. However, recent studies indicate that CRT is not a nuclear resident protein, and earlier identification of the protein in the nucleus was likely an artefact of immunostaining [23]. Furthermore, there was no immunologically detectable CRT in the cytoplasm, nor was it associated with focal adhesion sites. The major cellular site of localisation for CRT is the ER, but one cannot exclude the possibility that there may be small quantities of CRT present at extra-ER sites.

1.1.3.1Cell-surface calreticulin

The presence of the KDEL sequence at the C-terminus of some proteins is thought to be responsible for the retention of KDEL proteins inside the ER mediated by the KDEL receptor [24]. Among the soluble ER KDEL proteins are calreticulin, protein disulfide isomerase (PDI), the 94-kDa glucose-regulated protein (GRP94) and the 78-kDa glucose-regulated protein (GRP78). Although KDEL proteins are primarily localised to the ER, there is growing evidence that some of these ER proteins occur on the cell surface of certain cells [25-28]. Several reports have suggested possible functions for these ER proteins on the cell surface.

It is becoming clear that beyond its role as a molecular chaperone of the ER, CRT also regulates multiple cellular functions as a cell surface protein [25, 26, 29, 30]. It was demonstrated that CRT may act as a cell surface receptor for penetrating anti-DNA antibodies in erythrocytes (K562), T lymphocytes (CEM and Jurkat), and B lymphocytes cell lines [29]. In human fibroblasts, surface CRT functions as a receptor for fibrinogen and is essential for its mitogenic effects [26]. In melanoma cells which adhere and spread on a laminin substratum, surface CRT is responsible for initiation of cell spreading [25]. CRT was detected at the cell surface of intestinal cells (IEC-6) [31] and reported to be associated with CD59 at the cell surface of polymorphonuclear neutrophils (PMN) [32].

Thrombospondin (TSP) signals focal adhesion disassembly through interactions with cell surface calreticulin. Because CRT is not a transmembrane protein, it is likely that CRT signals as a part of a coreceptor complex. It has been shown by Orr et al., 2003, [33] that CD91 mediates focal adhesion disassembly initiated by TSP binding to CRT. Furthermore, they have provided the first direct evidence that CD91 and CRT interact and that TSP, CRT, and CD91 are complexed in endothelial cells [33]. These data establish a mechanism of cell surface CRT signaling through its coreceptor, CD91, and suggest a novel function for CRT in regulating cell adhesion.

Using NG108-15 cells (a mouse neuroblastoma-rat glioma hybrid cell line) which can produce and form functional synapses, Xiao et al., 1999 [34] have demonstrated that although most of the total CRT is within the ER, a small but significant quantity is present on the cell surface. In fact, about half of the newly sythesised CRT reaches the cell surface rapidly. Their explanation for this phenomenon is that in NG108-15 cells, the synthesis of KDEL proteins is very efficient, while the amount of KDEL receptor is limited. Under these conditions, only part of the newly synthesised CRT molecules can be retained within the ER permanently. The rest of the newly synthesised CRT molecules escape the ER and reach the cell surface. The second possible mechanism by which CRT can escape from the ER is that the KDEL retention sequence has been sterically blocked by other proteins. They have also shown that KDEL proteins GRP78 and PDI are also found on the surface of NG108-15 cells [35]. However, GRP94 and CNX are not present on the cell surface. In addition, they have demonstrated that surface CRT, but not GRP78 or PDI, is essential for neurite formation in NG108-15 cells [35].

Calreticulin has been reported to be a C1q and collectin receptor on the surface of neutrophils and other cell types [21, 36]. However, CRT lacks a transmembrane domain and therefore must associate with a cell surface receptor. CRT plays a role in C1q signaling of apoptotic cell ingestion through association with CD91, and the two proteins colocalise on the surface of macrophages [37, 38]. Binding of CRT involves both the collagenous tails and the globular domains of C1q [19, 39] and the collagenous tails of MBL and occurs through a region in the N- and P-domain of CRT, termed the S-domain [21]. It is suggested that C1q changes conformation upon binding to immunoglobulins and thereby enables CRT to bind, leading to the formation of a stable complex, which may be internalised by macrophages [40]. This supports a physiologically important role of CRT as a cell-surface receptor with relevance for the immune system.

1.1.4 Calreticulin, a lectin-like chaperone

1.1.4.1 The endoplasmic reticulum

The endoplasmic reticulum (ER) plays an essential role in the folding and maturation of newly synthesised proteins. It provides an environment optimised for protein folding due to the presence of a large variety of folding enzymes, molecular chaperones and folding sensors. Many of these associate with the growing nascent chains and continue to assist the protein folding until it has acquired its native

structure. Folded proteins exit the ER to the *cis*-Golgi and are transported along the secretory pathway while persistently misfolded proteins either aggregate or become degraded [41].

1.1.4.2 Quality control

The endoplasmic reticulum quality control includes two groups of proteins. The first is the molecular chaperones and the second is the glycan modification enzymes. Molecular chaperones prevent the aggregation of partially folded proteins, increase the yield of correctly folded proteins and assembly, and also increase the rate of correctly folded intermediates by recruiting other folding enzymes [4]. The presence of N-linked oligosaccharides is required for the correct folding of many glycoproteins, so the glycosylation machinery also regulates protein folding [42].

The molecular chaperone family includes both the classical and non-classical chaperones. Classical chaperones such as immunoglobulin-heavy-chain-binding protein (BiP, GRP78), the ER homologue of Hsp70, are involved in binding unfolded proteins and probably in their degradation. The non-classical chaperones CNX and CRT are constituents of a molecular machine known as the CRT/CNX cycle [14, 42]. In the last few years there has been a large increase in the number of publications documenting that calreticulin functions as a lectin-like molecular chaperone.

1.1.4.3 The calreticulin/calnexin cycle

CRT and CNX participate in a well characterised cycle of binding and release of newly folded proteins and glycoproteins via a $Glc_1Man_9GlcNAc_2$ carbohydrate ligand or through direct interaction with misfolded polypeptides [14, 42] (Figure 1.4). They bind to $Glc_1Man_9GlcNAc_2$ oligosaccharides and recognise the terminal glucose and four mannose moieties [43, 44]. The carbohydrate attached to newly synthesized proteins in the ER lumen is of the form $Glc_3Man_9GlcNAc_2$. The glucose moieties are removed by glucosidase I and II and, if the glycoproteins are folded correctly they may escape from the folding cycle. However, if the glycoproteins are incorrectly folded, the terminal glucose is once again attached by the UDP-glucose:glycoprotein glucosyltransferase (UGGT), which provides a discrimination between folded and unfolded substrates [45]. Therefore unfolded glycoproteins undergo cycles of interaction with calnexin and calreticulin. In this

model, CNX and CRT do not function as classical chaperones, but rather the lectin-oligosaccharide interaction itself is thought to enhance folding or subunit assembly, stabilise intermediates, and exert quality control. CNX and CRT may also promote folding by recruiting folding catalysts such as thiol oxidoreductase, ERp57, to the vicinity of the folding glycoproteins.

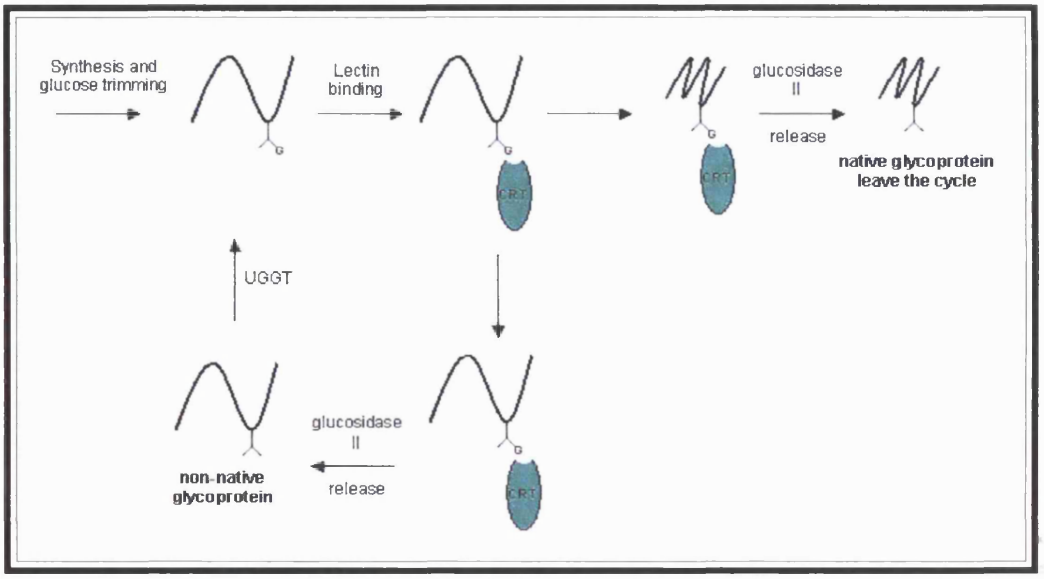

Figure 1.4 : Calreticulin/calnexin chaperone cycle. A proposed model of glycoprotein folding. Calreticulin (and calnexin, its integral membrane homologue; not shown) acts as lectin by binding to monoglucosylated oligosaccharide chains present on newly synthesised glycoproteins. The glycoprotein is released from calreticulin upon glucose trimming by glucosidase II. If the folding has been sucessful, the native protein is free to continue along the secretory pathway. If the glycoprotein is incompletely folded, the folding sensor UGGT adds back a single glucose residue, and the glycoprotein undergoes another round of lectin binding. Figure adapted from reference [50].

The lectin site of calreticulin and calnexin is localised to the Ca^{2+} binding P-domain of the proteins and the bound Ca^{2+} is essential for the lectin-like function of these proteins [14, 46]. Using deletion mutants of calreticulin the lectin site was localised to the proline-rich P-domain that contains two sequence motifs, each repeated three times in tandem. Truncation of even a portion of one of the repeats abolished oligosaccharide binding to calreticulin. Although many glycoproteins bind to both CRT and CNX, there are differences in the spectrum of glycoproteins they bind. At the molecular level, the two main differences are that the P-domain in CNX is longer than in CRT (figure 1.3 and 4.35) and that CRT is a soluble lumenal protein and CNX a type I membrane protein of the ER membrane (figure 1.3). Indeed, their

functions are largely interchangeable. The substrate specificity of both chaperones was essentially inverted when producing a soluble CNX mutant and a membrane-bound version of CRT [47]. So the different topological environments of CRT and CNX are important in determining their distinct substrate specificity.

1.1.4.4 Calreticulin also a classical chaperone

There is considerable evidence that in conjunction with the regulated lectin binding and release cycle, CRT and CNX bind also to unfolded polypeptide segments and promote folding in a manner analogous to classical chaperones. Purified CRT and CNX were able to form stable complexes with nonglycosylated proteins *in vitro* and prevent their thermally induced aggregation in a manner similar to other chaperone families [48]. Furthermore, calreticulin did not associate with folded proteins, but engagement of its lectin site with purified oligosaccharide attenuates its function as a classical chaperone.

1.1.4.5 Chaperone selection in the ER

Proteins interact in different ways with the molecular chaperones present in the endoplasmic reticulum (ER). Some interact first with the binding protein (BiP) and then with calnexin and/or calreticulin, some bind to CRT and CNX but do not associate with BiP, some are sequentially assisted by BiP and glucose regulated protein 94 (GRP94) [49]. Also the formation of disulfide bonds can be catalysed by protein disulfide isomerase (PDI), ERp57, or by both oxidoreductases acting together [10, 50]. The rules that govern which chaperone a protein will associate with and in which order are poorly understood at the moment. Recent work shows that for glycoproteins the choice of chaperone depends, in part, on the position of the glycans in the sequence. Direct interaction with CNX and CRT without prior interaction with BiP occurred if glycans were present within about 50 residues of the protein N-terminus [49].

1.1.4.6 ERp57-catalysed disulfide bond formation

The oxidizing environment and the presence of several different thiol oxidoreductases allow formation of disulfide bonds in the ER. If CRT and CNX act principally as lectins [14, 44], then additional ER components may be required to directly influence glycoproteins folding during the calnexin/calreticulin cycle. The principal candidate for this role is Erp57 [51]. ERp57 is a member of the protein

disulfide isomerase (PDI) family which, in addition to the conventional PDI, also includes ERp72, P5 and PDIp. Interestingly, the substrate binding properties of ERp57 mimicked those of CRT and CNX, i.e. both glycosylation and glucose trimming were required. In addition, ERp57 was coimmunoprecipitated with CNX and CRT, suggesting that it exerts its functions as a co-chaperone with those two lectins [50]. Olivier *et al.* (1999) [50] demonstrated that both CRT and CNX can be isolated in complex with ERp57 from semi-permeabilised cells and microsomes and that the interaction is independent of the presence of substrate glycoproteins. Since ERp57 is not glycosylated, these complexes are clearly not the result of "typical" interactions between CNX/CRT and a glycoproteins substrate. These data are also in agreement with the study of Corbett *et al.*, 1999 [11], which showed that the purified recombinant CRT and ERp57 interact specifically in solution.

The interaction between calreticulin and PDI was also reported [10]. The PDI-calreticulin complex was described to be zinc dependent and involving the P-domain and the NH_2-terminal region of the N-domain of calreticulin. Interestingly, the interaction between these two proteins reduced the calcium binding to the high-affinity site in the P-domain of CRT and inhibited the PDI activity determined by its ability to refold scrambled RNase A. Also, expression of calreticulin domains and PDI as fusion proteins with GAL4 in the yeast two-hybrid system revealed that CRT interacted with PDI also under normal cellular conditions.

It has been shown that calreticulin associates with PDI only at very low Ca^{2+} concentrations, i.e. below 100 μM [11]. The very low calcium concentration needed to observe an interaction would require an almost complete emptying of the ER calcium store considering the steady state $[Ca^{2+}]$ in the ER lumen is approximately 500-800 μM [52]. Finally, the results of Olivier et al. (1999) [50] using intact ER-derived microsomes and semi-intact mammalian cells, underlined the fact that, under the conditions observed *in vivo*, ERp57 forms a specific complex with the ER lectins, CNX and CRT (Figure 1.5), while there was no evidence to suggest that PDI is associated with these components.

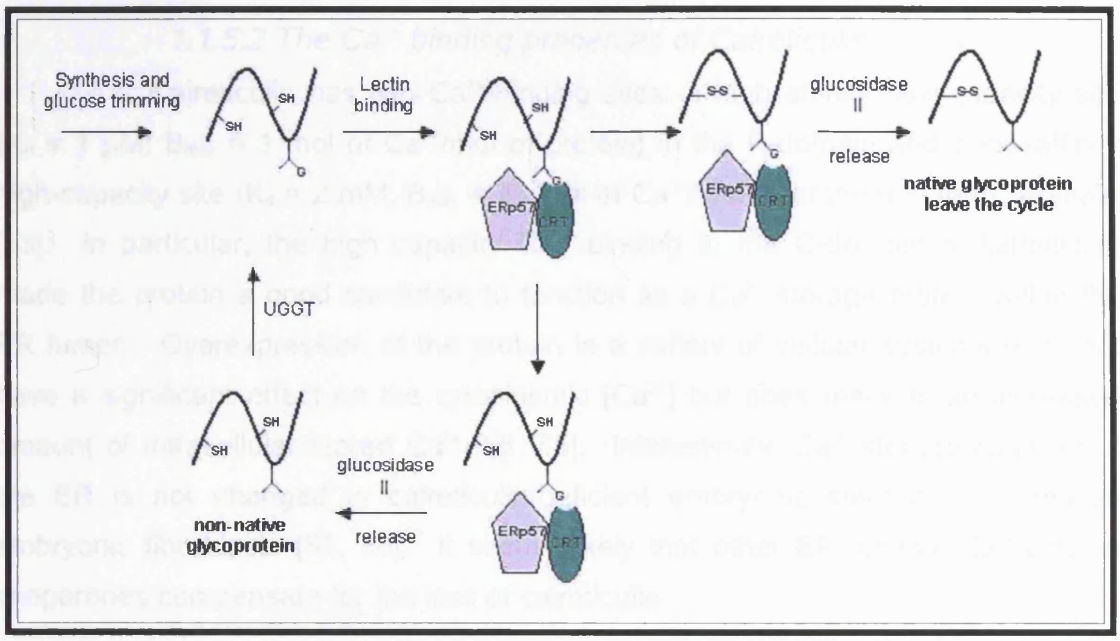

Figure 1.5 : Calreticulin recruits ERp57 to assist disulfide-bond formation. A proposed model of glycoprotein folding. Calreticulin (and calnexin, its integral membrane homologue; not shown) acts as lectin by binding to monoglucosylated oligosaccharide chains present on newly synthesised glycoproteins. By virtue of its association with calreticulin, ERp57 is brought into contact with the glycoproteins. ERp57 modulates the folding of the glycoproteins, illustrated here as disulfide bond formation, but possibly by other mechanisms. The glycoprotein is released from calreticulin upon glucose trimming by glucosidase II. If the folding has been sucessful, the native protein is free to continue along the secretory pathway. If the glycoprotein is incompletely folded, the folding sensor UGGT adds back a single glucose residue, and the glycoprotein undergoes another round of lectin binding. Figure adapted from reference [50].

1.1.5 Calreticulin and Ca^{2+} signalling

1.1.5.1 ER luminal Ca^{2+}

The ER luminal [Ca^{2+}] at steady state, reported to be 500-800 μM, is two to three orders of magnitude higher than in the cytosol [53]. The Ca^{2+} storage capacity of the ER lumen is enhanced by Ca^{2+}-binding chaperones. These include calreticulin, Grp94, BiP, PDI, Erp72 and ER/calcistorin. Besides important effects on Ca^{2+} homeostasis, reduction of the ER calcium leads to accumulation of misfolded proteins, activation of expression of ER chaperones and ER-nucleus and ER-plasma membrane signalling [54]. Clearly, changes of the ER luminal [Ca^{2+}] have profound effects at multiple cellular sites, including the structure and function of the ER luminal Ca^{2+}-binding chaperones.

1.1.5.2 The Ca^{2+} binding properties of Calreticulin

Calreticulin has two Ca^{2+}-binding sites: a high affinity, low capacity site (K_d = 1 μM; B_{max} = 1 mol of Ca^{2+}/mol of protein) in the P-domain and a low-affinity high-capacity site (K_d = 2 mM; B_{max} = 25 mol of Ca^{2+}/mol of protein) in the C-domain [13]. In particular, the high capacity Ca^{2+} binding to the C-domain of calreticulin made the protein a good candidate to function as a Ca^{2+} storage protein within the ER lumen. Overexpression of the protein in a variety of cellular systems does not have a significant effect on the cytoplasmic [Ca^{2+}] but does result in an increased amount of intracellular stored Ca^{2+} [55, 56]. Interestingly, Ca^{2+}-storage capacity of the ER is not changed in calreticulin-deficient embryonic stem cells or mouse embryonic fibroblasts [57, 58]. It seems likely that other ER luminal Ca^{2+}-binding chaperones compensate for the loss of calreticulin.

1.1.5.3 Conformational flexibility and thermal stability

It has been shown from sedimentation analysis that CRT adopts an elongated shape in solution, a characteristic that can be attributed to the extended hairpin structure of the P-domain [59]. This unusual elongated shape suggest that the protein has a high intrinsic structural flexibility (see figure 4.36a) which might be required for its chaperone function. Using different biochemical and biophysical techniques, the effects of calcium and zinc ions on the structural properties of CRT have been determined [60]. Circular dichroism (CD) and analytical ultracentrifugation analysis have shown that occupancy of the high-affinity Ca^{2+} binding site in the P-domain of CRT increases the thermal stability and conformational rigidity. However, upon the binding of zinc ions in the N-domain, a more loosely packed and thermally destabilised structure was adopted by the protein. Results from sedimentation analysis indicated that the overall three-dimensional structure of CRT is essentially unchanged upon binding calcium ions in contrast with the reversible dimerisation upon binding zinc ions. Furthermore, the functional importance of the high-affinity Ca^{2+} binding site is underscored by considering that the concentration of calcium ions in the lumen of the ER is sufficiently high under all conditions, that it is constantly occupied by a calcium ion. Finally, the lectin binding site is dependent on occupancy of this calcium ion, consequently, the lectin recognition site is intrinsically present in CRT.

Interestingly, the occupancy of the low-affinity calcium site in the C-domain further increases the thermal stability and conformational rigidity of the protein without changing its overall 3D structure [60]. This suggest that calcium ions

may serve to spatially organise and stabilise the highly negatively charged C-domain, which was shown to be a more conformationally flexible and destabilised region of the protein in the absence of added calcium ions [60]. The functional implications of these structural rearrangements in the C-domain has been suggested to play a role in regulating calcium-dependent interactions between CRT and PDI or ERp57. The fluctuating calcium concentrations in the ER would provide a mechanism to control this regulatory process as also previously suggested [11]. Finally, the structural properties of CRT are highly dependent upon interactions with calcium and zinc ions pointing out the importance of these metal ions in modulating the function on CRT in the cellular environment of the ER.

1.1.6 The calreticulin-deficient mouse

With the abundance of functions with which calreticulin has been associated, it was anticipated that the protein will be essential for organism survival. Indeed, the most important feature of the knockout is that it is an embryonic lethal mutation [57, 61]. The embryos were viable until 14.5 days into gestation. The most likely cause of embryonic lethality is the cardiac failure of the embryos at 12.5, 14.5 and 18 days [57]. This indicates that CRT must play a role during cardiac development [61], but it is still not clear which of its many proposed functions is responsible for mediating this role. The most likely mechanism is through the ability of CRT to modulate Ca^{2+} homeostasis [61].

1.1.7 Extracellular soluble calreticulin

Calreticulin has been implicated in a number of pathological processes including systemic lupus erythematosus (SLE). Autoantibodies against CRT have been found in approximately 40% of SLE patients [62] and detected in rheumatoid arthritis (RA) [63], celiac disease [64], complete congenital heart block [65] and halothane hepatitis [66] (figure 1.6). The major contributing factor to autoimmune diseases such as SLE is the failure to clear immune complexes, a process largely mediated by the first component of the classical pathway of complement, C1q. Patients who lack C1q almost certainly develop SLE [67], accentuating the importance of C1q in the clearance of immune complexes. Furthermore, it has been demonstrated that CRT can bind to C1q and can compete with antibodies for binding to C1q [68] (figure 1.6). Extracellular calreticulin may contribute to autoimmune disease by inhibiting immune complex clearance. Several parasites, including

Onchocerca volvulus [69], *Necator americanus* [70], *Trypanosoma cruzi* [71] and the ixodid tick *Amblyomma americanum* [72] have been shown to secrete homologues of CRT to evade the immune system, by preventing the complement system to act against the parasite [73] (figure 1.6).

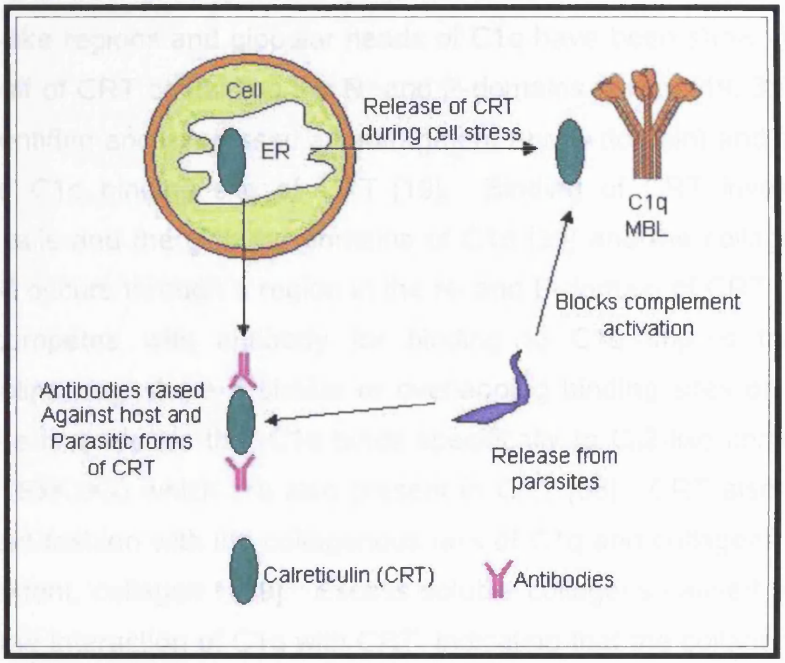

Figure 1.6 : Roles of extracellular CRT in disease.

1.1.8 CRT binds to C1q and mannan-binding lectin

It has been shown that C1q, a subunit of the first component of complement (C1), binds to calreticulin [74]. Potentially, the association of calreticulin with C1q may interfere with complement activation. The interaction between CRT and C1q is ionic strength-dependent [75], and under various ionic conditions, both the collagen-like regions and globular heads of C1q have been shown to bind to the N-terminal half of CRT containing the N- and P-domains *in-vitro* [19, 39]. Within this region we identified and expressed a subfragment (the S-domain) and showed that it harbours the C1q binding site of CRT [19]. Binding of CRT involves both the collagenous tails and the globular domains of C1q [39] and the collagenous tails of MBL [76] and occurs through a region in the N- and P-domain of CRT [21]. The fact that CRT competes with antibody for binding to C1q implies that CRT and immunoglobulins may share a similar or overlapping binding sites on C1q globular heads. There is evidence that C1q binds specifically to C_H2-like immunoglobulin γ (IgG) motifs (ExKxKx) which are also present in CRT [68]. CRT also interacts in a salt dependent fashion with the collagenous tails of C1q and collagens III, V, VI and, to a lesser extent, collagen I [19]. Excess soluble collagens caused near-complete inhibition of the interaction of C1q with CRT, indicating that the collagens bind via an identical, or overlapping site, to CRT. Previous work has shown, that the recombinant CRT S-domain may act as a potent inhibitor of complement activation, as in serum assays, C1q and MBL-mediated complement activation can be blocked in a dose-dependent manner by addition of small quantities of recombinant protein. It has been well demonstrated that the collectins C1q, MBL, SP-A and CL43, with the exception of SP-D, all bind to the recombinant S-domain of CRT [21]. These results were consistent with previous observations with CRT [77]. Comparison of the N-terminal regions of the collagenous regions of C1q, MBL, CL43, SP-A and conglutinin led to the identification of a possible CRT binding site, composed of five collagen repeats (Gly-X-Y triplets) with many charged residues in the X and Y positions (figure 1.7) [76].

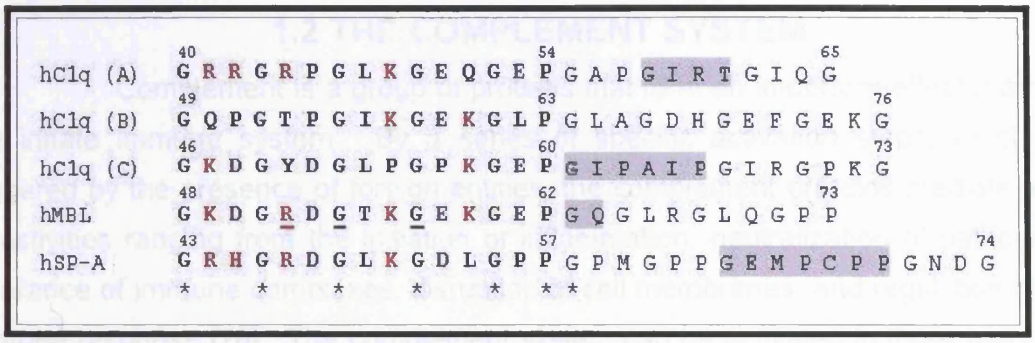

Figure 1.7 : Amino acid sequence of the human C1q (A chain, B chain, C chain), MBL , and SP-A from a region close to the bend in the middle of the collagen region. Positively and negatively charged amino acids are shown in red and blue respectively, and the proposed binding site in bold. The regions in C1q, MBP, and SP-A in which there is an interruption to the Gly-Xaa-Yaa triplet are highlighted in grey. Figure adapted from reference [76].

The S-domain-C1q interaction was found to be ionic-strength-dependent indicating that it is predominantly a protein-protein interaction based upon a region of charged residues on the collagen stalks of the C1q ligand [76]. The collectin-binding site was localised in the S-domain of CRT, which has a putative C1r/C1s (also termed CUB domain) module based upon amino acid sequence alignments [21]. Calreticulin inhibits $C1r_2s_2$ binding to C1q [20], and CRT does not interact with C1q when $C1r_2s_2$ are already bound. This suggests that CRT recognises a binding site on C1q overlapping, or similar to, the site for $C1r_2s_2$ binding. Subsequent structural information on the calreticulin P-domain [59], the sequence of which overlaps with the S-domain, shows it does not fold as a CUB domain. Structural studies will be required to confirm that the S-domain sequence folds as an independent module similar to a CUB domain or most likely comparable to the P-domain.

1.2 THE COMPLEMENT SYSTEM

Complement is a group of proteins that form an important effector arm of the innate immune system. By a series of specific activation steps, which are triggered by the presence of foreign entities, the complement proteins mediate a set of activities ranging from the initiation of inflammation, neutralization of pathogens, clearance of immune complexes, disruption of cell membranes, and regulation of the immune response [78]. The complement system can be activated in three ways: the classical pathway which is initiated by binding of the multimolecular C1 complex to foreign materials, by the alternative pathway initiated by certain structures on microbial surfaces, and by a third pathway that is initiated by the binding carbohydrate recognition complexes composed of either mannan-binding lectin (MBL) or L-ficolin or H-ficolin and the lectin pathway specific serine protease MASP-2 (for MBL-associated serine protease 2) (figure 1.8). The first component of the classical pathway is C1, a complex of proteins that were subsequently named C1q, C1r and C1s (figure 1.9). C1q is a hexamer with globular heads and a rod-like portion very similar to collagen including containing hydroxyproline (figure 1.10). These globular heads actually bind to the targets to initiate complement fixation. Dimers of serine proteases, C1r and C1s, fit in and around the rod-like portions of C1q to form intact C1.

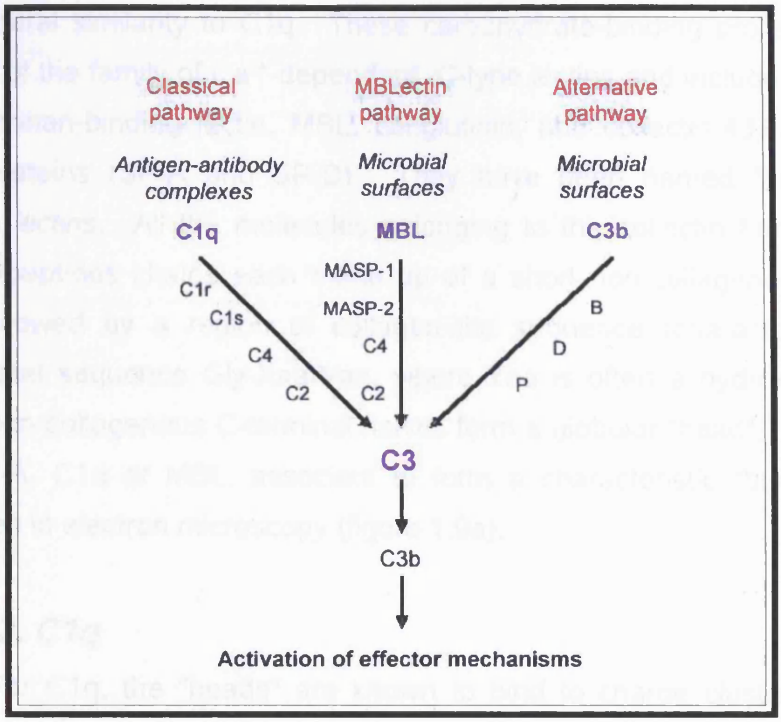

Figure 1.8 : The three pathways of complement activation. All pathways converge on the activation of the central component C3 into C3b, which mediates the effector functions of complement. Figure adapted from reference [80].

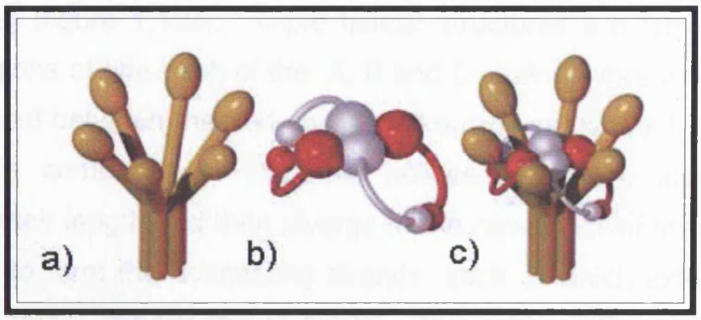

Figure 1.9 : The classical pathway molecule C1. A = C1q, B = $C1r_2$-$C1s_2$ tetramer, and C = intact C1 or $C1qr_2s_2$.

1.2.1. Collectins

Collectins are a group of soluble proteins each of which has collagenous and non-collagenous domains, and may be complement associated with lectin activity. C1q has an unusual modular structure consisting of collagen-like triple-helical regions connected to globular "head" regions (figure 1.9a). The members of a family of mammalian lectins, containing collagen-like structures, all show a strong

overall structural similarity to C1q. These carbohydrate-binding proteins belong to the group III of the family of Ca^{2+}-dependent, C-type lectins and include three plasma proteins (mannan-binding lectin, MBL; conglutinin; and collectin-43) and two lung surfactant proteins (SP-A and SP-D). They have been named "collectins", i.e. *col*lagen-like *lectins*. All the molecules belonging to the collectin family consist of multiple polypeptides chains each made up of a short non-collagenous N-terminal segment, followed by a region of collagen-like sequence (characterised by the repeating triplet sequence Gly-Xaa-Yaa, where Yaa is often a hydroxylated amino acid). The non-collagenous C-terminal halves form a globular "head". Six such sub-units, in SP-A, C1q or MBL, associate to form a characteristic "bunch of tulips" structure seen in electron microscopy (figure 1.9a).

1.2.2. C1q

For C1q, the "heads" are known to bind to charge clusters on a wide range of targets including sequences in the Fc regions of IgG and IgM. The C1q molecule is composed of 18 polypeptide chains (six A, six B, and six C) with each of the three types of chains containing a region of 81 amino acids of a collagen-like $(\text{-Gly-Xaa-Yaa-})_n$ repeating sequence starting close to the N-terminal end and which is followed by the C-terminal portion of about 136 amino acids which are non-collagen-like [79] (figure 1.10a). Triple helical structures are formed between the collagen-like regions of one each of the A, B and C chains, while the globular "head" regions are formed between the non-collagen like regions (figure 1.10b). Thus, one C1q molecule is composed of six triple helices which are aligned in parallel throughout half their length and then diverge to the remainder of their length of triple helical structure to form the connecting strands, each of which extends into one of the six globular "head" regions (figure 1.10c). The collagen-like connecting strands interact with the $C1r_2$-$C1s_2$ tetramer to give the C1 complex (figure 1.9c), which after activation, cleaves and activates C4 and C2 (figure 1.8).

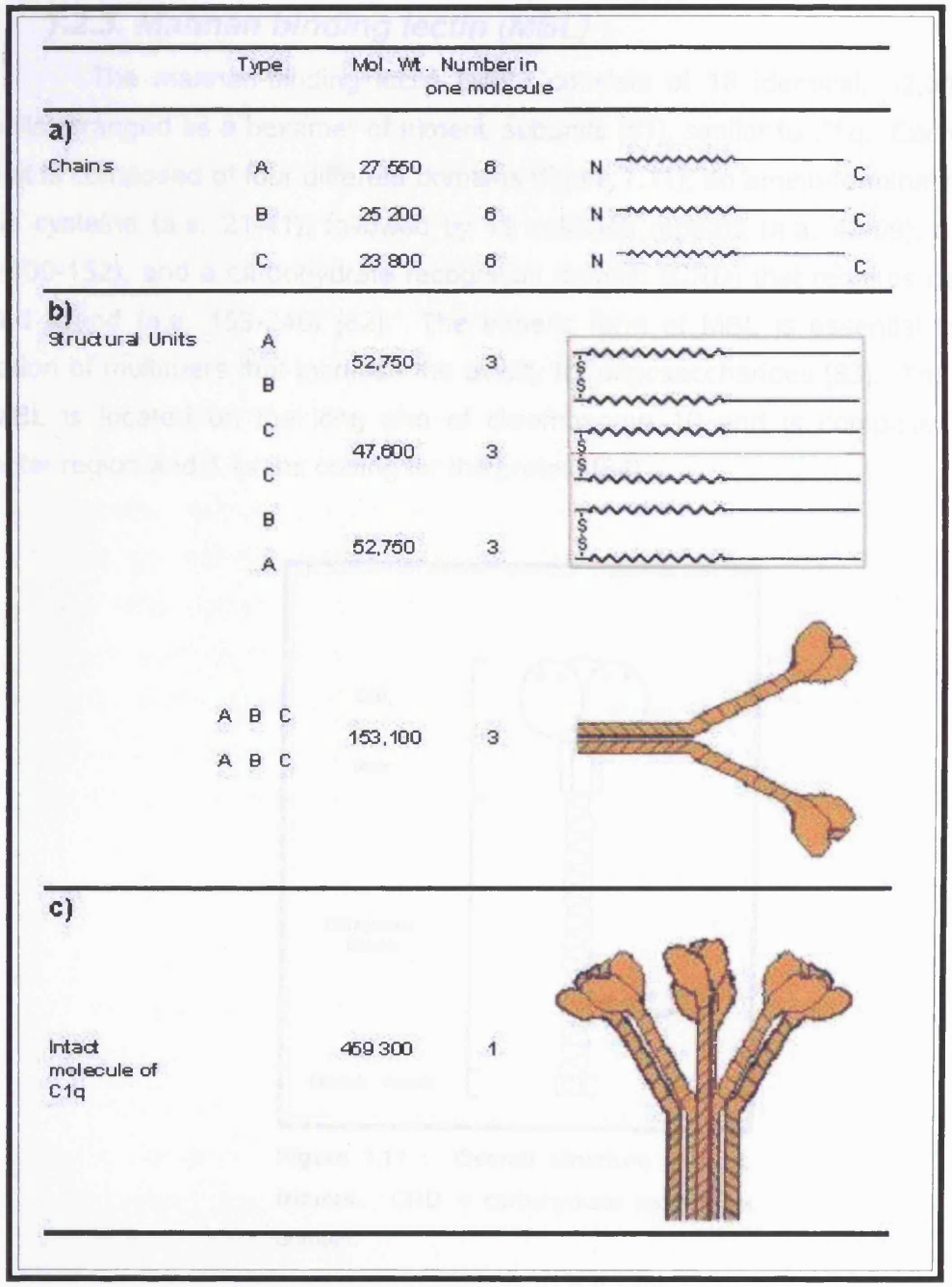

Figure 1.10 : A molecular model of C1q. This ( ) represents the collagen like triple helices and ( ) the collagen-like regions. A) the three C1q chains, B) The dimer of trimers, and C) The hexamer of trimers. Figure adapted from reference [80].

1.2.3. Mannan binding lectin (MBL)

The mannan-binding lectin (MBL) consists of 18 identical, 32,000 Da subunits arranged as a hexamer of trimeric subunits [81], similar to C1q. Each MBL subunit is composed of four different domains (figure 1.11); an amino-terminal region rich in cysteine (a.a. 21-41), followed by 19 collagen repeats (a.a. 42-99), a neck (a.a. 100-152), and a carbohydrate recognition domain (CRD) that requires calcium to bind ligand (a.a. 153-246) [82]. The trimeric form of MBL is essential for the formation of multimers that increase the avidity for oligosaccharides [83]. The gene for MBL is located on the long arm of chromosome 10 and is composed of a promoter region and 4 exons coding for the protein [84].

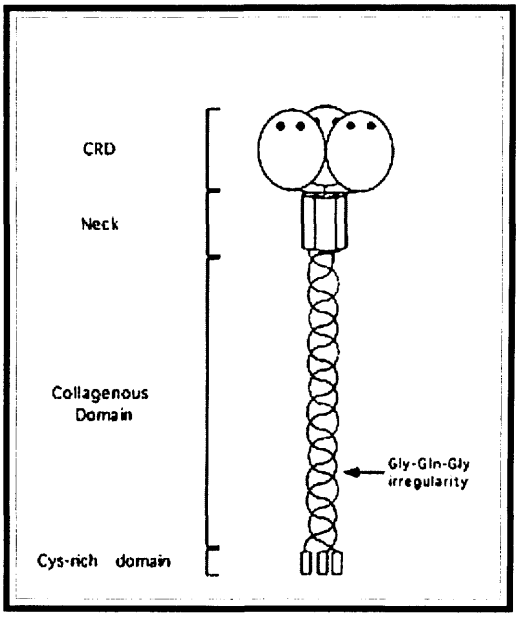

Figure 1.11 : Overall structure of MBL trimers. CRD = carbohydrate recognition domain.

The human MBL is a recognition molecule that plays a role in the innate host immune defence. It binds to the surface of many pathogens via surface carbohydrates and induces cell lysis via complement activation (figure 1.8). The binding of MBL to the surface of a microorganism results in the activation of the complement cascade via MBL-associated serine proteases (MASP-1 and MASP-2) that are coupled to MBL [85, 86]. These enzymes can activate C4 and C2 so that the C3 convertase of the classical route (C4b2b) is produced before there are any specific antibodies (figure 1.8). It has been proposed as the lectin route of complement activation in which MBL plays an essential role. MBL is related to the

complement C1 subcomponent, C1q, and the MASPs are homologous to the C1 serine proteases, C1r and C1s [87].

The probability of mutations in the MBL gene is about 10% [88]. MBL deficiency predisposes both children and adults to all sorts of infectious diseases [88]. Remarkably, MBL deficiency may actually be advantageous in some infections, because certain microorganisms use MBL or complement to invade the cell. A number of distinct MBL allelic forms exist in different population groups, Arg-52→Cys, Gly-54→Asp [89] and Gly-57→Glu [90] (underlined in figure 1.7). These substitutions were predicted to disrupt the collagen helix leading to dysfunctional MBL. It was predicted that both homozygous and heterozygous individuals would have profoundly reduced serum levels of the protein. It was confirmed by immunoassay as was the reduced capacity of such sera to activate complement through the MBL pathway [91]. It appears that the resultant MBL from these mutations are able to form high-order multimers that binds bacteria but do not support complement activation and that MASP-1 and its proenzyme proMASP-1 are unable to bind to recombinant MBL Gly-54→Asp [88]. This suggests that the binding site of the MBL serine proteases maps from the fourth to the sixth collagen repeat of MBL.

1.2.3.1. Ficolins

Ficolins are a group of proteins containing both a collagen-like domain and a fibrinogen-like domain and are found in a number of different tissues. Ficolins present in sera have a lectin activity toward N-acetylglucosamine through their fibrinogen-like domains. The domain organisations between ficolins and mannose-binding lectin (MBL) are very similar in that both consist of a collagen-like domain and a carbohydrate-binding domain, although their carbohydrate-binding moieties are different [92]. MBL acts as an opsonin and activates complement in association with MBL-associated serine proteases (MASPs) via the lectin pathway [93]. Like MBL, two types of human serum ficolins, L-ficolin/P35 and H-ficolin, are associated with MASPs and activate the lectin pathway. In addition, L-ficolin/P35 acts as an opsonin [94]. These findings indicate that serum ficolins play an important role in innate immunity in a similar manner to MBL [92].

1.3 AIMS OF THIS THESIS

At the time of commencement of this research, no 3D structure of calreticulin had been solved. The aim of the research work summarised in this thesis was to determine the structure of the C1q and collectin binding domain of calreticulin. It had been demonstrated that the C1q binding site within CRT lies across the intersection of the N- and P-domains, termed the S-domain [19]. Furthermore, competitive inhibition studies of the S-domain-C1q interaction revealed that the S-domain binds to C1q collagen tails and to the collectin proteins, SP-A, MBL, CL43 and conglutinin [19, 21].

Initial and optimised purification of multiple calreticulin domains and valuable information on the CRT-MBL, -C1q interactions are described in chapter 2. Efforts were mainly devoted to the structural analysis of the calreticulin S-domain. To achieve this goal, X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy were employed. The crystallogenesis of the CRT S-domain is presented in chapter 3. Details of the structure determination by NMR, the structure itself and relaxation experiments are discussed in chapter 4. A general conclusion is made in chapter 5.

CHAPTER 2 :
CLONING AND CHARACTERISATION OF CALRETICULIN CLONES

2.1 INTRODUCTION

Inhibition of the complement system is a promising therapeutic approach in diseases where uncontrolled complement activation plays a significant role in the pathogenesis of the disease such as SLE [62, 67]. Calreticulin (CRT) is a multifunctional protein which has been reported to bind to C1q, prevents the formation of C1 and so inhibits activation of the classical pathway. To assess the usefulness of CRT as a complement inhibitor, recombinant CRT domains were cloned, expressed, and purified for biochemical and structural studies.

Biochemical and structural studies on proteins require that large quantities of homogenous protein be produced as efficiently, in term of time and resources as possible. For the purpose of crystallographic and NMR studies, maximising protein yields is of high importance. The quality, in terms of protein activity and stability is also critical. In order to get high amounts of recombinant human calreticulin domains for functional and structural studies, the protein of interest has been cloned into the pET vector (Novagen). The pET system is developed for the cloning and expression of recombinant proteins in *E.coli*, which is under control of the highly efficient T7 promoter. Once established, plasmids are transfered into expression hosts containing a chromosomal copy of the T7 RNA polymerase gene under *lacUV5* control, and expression is induced by the addition of IPTG. Levels of protein expression in procaryotic systems may be enhanced by manipulation of the host strain and plasmid genetic structure, cell growth conditions and protein purification strategy. In order to localise C1q and MBL binding activity within CRT, recombinant CRT S- (residues 164-283), P- (198-308), and PS-domain (164-308)] were produced.

The pET-15b has a very useful hexahistidine-tagged (His-Tag) fusion partner for easy purification of proteins by metal chelation chromatography under native or denaturing conditions and has a thrombin cleavage site for removal of the His-Tag. Further purification steps using the classical chromatography methods provided us with some relatively pure proteins for biochemical characterisation and structural studies. The integrity of the samples were analysed by protein sequencing and mass spectroscopy. Other more refined methods, as Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES) and circular dichroism (CD), were used to asses metal binding capacity and the presence of secondary structure in the proteins of interest.

Furthermore, obtaining large amount of pure protein is irrelevant if the isolated molecule is not biologically active . The different clones were tested for their biological activity using haemolytic and ELISA based assays. These verified that the different calreticulin clones purified could bind to C1q and MBL and also inhibit their complement activity. All CRT S-, P-, and PS-domains bound to C1q and MBL, demonstrating that the binding site spans the intersection of these domains. However, only the S-domain of calreticulin seems to inhibit the classical and lectin pathway.

2.2 MATERIAL AND METHODS

2.2.1 DNA methods

2.2.1.1 PCR

The polymerase chain reaction (PCR) can be used to synthesise and modify target genes for cloning into pET vectors. With this approach, it is possible to design primers that will isolate the translated portion of the cDNA, add convenient restriction sites and place the coding region into the right reading frame. The full length human calreticulin cDNA, clone phCAR-1 [19], was used to generate by PCR amplification domains of different sizes that could be cloned into the pET vectors. An *Nde*I restriction site has been added in the forward primers for in phase cloning into the pET vectors. For the reverse primers, a stop codon was added at the end of the desired coding sequence followed by a *Bgl*II restriction site. The *Bgl*II restriction site produces after digestion a compatible sticky end with the *Bam*HI restriction site present in the pET vectors.

The general PCR conditions used were as follows (Table II.I) with the Promega; *Taq* DNA polymerase, 10x reaction buffer without $MgCl_2$ and 25 mM $MgCl_2$ solution. The primers used are described in table II.II and were purchased from vhBio Limited, Newcastle upon Tyne, UK.

Table II.I : General PCR conditions using the *taq* DNA polymerase.

Final concentration	
To final volume 50 µl	Sterile water
1x	10x reaction buffer without $MgCl_2$
2 mM	25 mM $MgCl_2$
1 µM	Forward primer
1 µM	Reverse primer
0.2 mM each	dNTP mix (10 mM of each dNTP)
1 pg/µl	Template
0,025 u/µl	*Taq* DNA polymerase (5 u/µl)

Table II.II : Human careticulin primers used. Underlined are the modifications to introduce; restriction sites and stop condons

	Primers name	5'→3' sequence
hCRT S-domain (+2)	s2CRTFWD	ATAGAATTCCATATGCGTTGCAAGGATGATGAGTT
hCRT S-domain	sCRTREV	AGAATTCAGATCTTTATGGGTTGTCGATCTGCCGGG
hCRT P-domain	pCRTFWD	ATAGAATTCCATATGGACGATTGGGACTTCCTGC
hCRT P-domain	pCRTREV	AGAATTCAGATCTTTAATAGGCATAGATACTGGGAT
hCRT S-domain	sCRTFWD	ATAGAATTCCATATGAAGGATGATGAGTTTACACA

The general cycles used for PCR were as described in table II.III. For best results, the PCR products were purified from an agarose gel using the QIAGEN QIAquick gel extraction kit (appendix X.1).

Table II.III : PCR cycles with their related temperature and duration. The annealing temperature, shown by an asterisk, could have varied slightly depending on the primers used.

Cycles	Temperature in °C	Time in seconds
1	96	60
10	96	30
	55*	30
	72	30
20	96	30
	70	30
	72	30
1	72	240
1	4	Until the end

2.2.1.2 Ligation

The Promega pGEM-T Easy vector system (appendix X.2) provides an efficient and convenient system for the cloning of PCR products. The ligation takes advantage of the template-independent addition of a single adenosine to the 3'-end of PCR products by certain thermostable polymerases such as the *Taq* DNA polymerase used. Ligations of inserts prepared by double restriction digestion followed by gel purification into the pET vectors was straightforward. One consistently successful protocol for ligation is presented here in table II.IV, using the Promega T4 DNA ligase and Ligase 10x buffer.

Table II.IV : General ligation conditions used. The reactions were incubated overnight at 14°C.

	Standard reaction
Ligase 10x buffer	1x
vector	50 ng
PCR product	1 µl
T4 DNA ligase (3u/µl)	3 u
Sterile water	Final volume to 10 µl

2.2.1.3 Competent cells, transformation and blue/white selection

The efficiency of this transformation method was high enough to allow all routine cloning in plasmids to be performed with ease. The *E.coli* competent cells used for transformation were prepared using the calcium chloride method [95]. The *E.coli* TOP10F' strain was used for general cloning purposes and BL21(DE3) for protein expression. All plasmids were transformed into chemically competent cells using standard heat shock protocol [95] and plated onto Luria-Bertani (LB)-Agar plates (appendix X.3) with the appropriate antibiotics for selection of recombinant clones, as described by the manufacturers. All plasmids used had an ampicillin (Amp) resistant marker and vector-transfected colonies were grown in the presence of 100 µg of Amp per ml of culture medium.

The pGEM-T Easy vector contains several features to facilitate identification of recombinant clones (appendix X.2). The T7 and the SP6 RNA promoters flank the multiple cloning regions within the α-peptide coding region of the enzyme β-galactosidase. Insertional inactivation of the α-peptide allows recombinant clones to be directly identified by blue/white colour screening. To use blue/white colour screening for recombinants and to maintain antibiotic selection, the transformed TOP10F' cells (appendix X.4) were plated on LB-agar plates containing 100 µg/ml Amp, 0.1 mM IPTG and 40 µg/ml X-Gal. The plates were incubated overnight at 37°C and for a few hours at 4°C to let the colour develop, in order to facilitate the blue/white screening. A few single white colonies, likely to contain the insert as it disrupts the coding region of the enzyme β-galactosidase, were analysed by small-scale plasmid DNA minipreps and identified the desired clones by restriction digestion and gel analysis.

Once the recombinant pET vectors (appendix X.2) were identified and purified from the non-expressing strain TOP10F' they could be transformed into the expression host strain, BL21(DE3).

2.2.1.4 Plasmids DNA preparations

All plasmid were transformed into competent *E.coli* cells (appendix X.4). A single colony was grown overnight into LB medium containing the appropriate antibiotics. Stocks were purified using a commercially available QIAGEN plasmid mini kit (appendix X.5) and stored in the -20°C freezer until used for transformation.

2.2.1.5 DNA digestion, bandprep and gel electrophoresis

Plasmid DNA preparations were analysed by restriction digest using the appropriate restriction enzymes and the restriction fragments separated by electrophoresis. This procedure is the method of choice for analysing recombinant clones and extraction of inserts for ligation. Digestions were performed for 1 hour at 37°C using the appropriate restriction enzymes and buffers. The enzymes and buffers were purchased from Promega (Southampton, UK). Agarose gel electrophoresis was performed for DNA analysis in Tris-acetate EDTA (TAE) buffer, as described by Sambrock, Fritsch and Maniatis (1989) [95]. Linearised plasmids and inserts were purified from agarose gel using the QIAGEN QIAquick gel extraction kit (appendix X.1) and stored in the freezer until use.

2.2.1.6 DNA sequencing

Plasmid minipreps of new recombinant pET vectors were sent for automatic DNA sequencing (Protein and Nucleic Acid Chemistry Laboratory, PNACL, University of Leicester, Leicester, UK) using the T7 promoter and T7 terminator primers, as described by the manufacturer (Novagen).

2.2.2 Protein methods

2.2.2.1 Protein expression

A bacterial culture of 50 ml LB-media containing 100 µg/ml of Amp was inoculated with a glycerol stock (appendix X.4) of *E.coli* BL21(DE3) harbouring the appropriate pET-15b construct. The bacterial culture was incubated overnight at 37°C under shaking (250 x rpm). The appropriate number of 2 L Erlenmeyer flasks containing 500 ml of LB-medium without antibiotics were inoculated with 0.05 A_{600nm} of the overnight culture. Expression constructs were induced in late-log phase, A_{600nm} 0.6, by addition of 1 mM IPTG for 3 hours at 37°C. The cells were pelleted by centrifugation at 5000 x g for 10 minutes at 4°C and frozen at -20°C until needed for the purification step.

2.2.2.2 Protein purification

The pellet was resuspended in ice cold 20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 5 mM Imidazole, 0.1% Triton X-100 and sonicated on ice. The sonicate was then centrifuged at 10 000 x g for 10 minutes at 4°C and decanted into a new tube. The supernatant was filtered through a 0.2 µM acrodisc (Gelman Sciences) and applied to a Ni^{2+} charged iminodiacetic acid-Sepharose column (5 ml Hi-Trap Chelating; Amersham Pharmacia Biotech) and eluted with an Imidazole gradient. The binding buffer was 50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM Imidazole and the elution buffer was containing 0.5 M Imidazole. The fractions were analysed by SDS-PAGE (polyacrylamide gel electrophoresis) [96] and the fractions containing the protein of interest were pooled. The protein was then dialysed in two steps at 4°C (20 mM Tris-HCl pH 8.0, 150 mM NaCl) using a Spectra/Por 3 tubing: 3.5 kDa MWCO (Spectrum Laboratories, US). After dialysis, the protein was digested with 3.5 units of Thrombin per litre of culture for one hour at room temperature with gentle agitation (Thrombin from human plasma; Roche Diagnostics Limited). After incubation, the thrombin was inhibited by the addition of 1 mM PMSF and incubated for 15 minutes at room temperature. The NaCl concentration was increased to 0.5 M by addition of a 5 M NaCl solution and the sample passed through a second Nickel column to remove the His-tag and some impurities. The flow through was collected and concentrated to a volume <4 ml using a Centriprep YM-3 (Millipore Limited, U.K.). The sample was then applied to a Superdex 75 (26/60) gel filtration column (Amersham Pharmacia Biotech) pre-equilibrated in 20 mM MES, pH 6.0, 150 mM NaCl. The fractions containing the protein were identified by SDS-PAGE and

pooled. The recombinant CRT domain was then purified using a Uno Q-6 column (Bio-Rad) using a gradient of NaCl. The binding buffer was 20 mM MES pH 6.0, 150 mM NaCl and the elution buffer was containing 1 M NaCl. The final pool was dialysed in two steps at 4°C (20 mM Tris-HCl pH 8.0, 150 mM NaCl) and the protein concentration determined by UV_{280} before flash freezing in liquid nitrogen (appendix X.7).

2.2.2.3 Protein sequencing

Protein sequencing provides information about the amino acids that make up a protein. During the sequencing process, amino acids are sequentially removed from the N-terminal end of the protein strand, and identified in the order they occur in the protein. The use of extremely pure starting protein is critical for successful sequencing.

Sample has been run on SDS-PAGE and blotted on a PVDF membrane for protein sequencing (PNACL, Leicester University). The PNACL uses an ABI 476 protein sequencer.

2.2.2.4 Mass spectroscopy

The molecular weight is a highly specific characteristic of a molecule and is often the first physiochemical property that is measured of a protein. Traditionally, SDS-PAGE, analytical ultracentrifugation, or gel permeation chromatography were used for molecular weight measurements. Recently, mass spectrometric techniques have been developed that permit mass determination of intact proteins with an accuracy far superior to the above mentioned methods. Only a few picomoles of a protein are necessary to obtain its molecular weight with an accuracy of 0.01-0.1%. Mass spectroscopy is a versatile and powerful analytical technique [97].

The purified proteins at a concentration of 1 to 5 mg/ml have been dialysed in water and sent for mass spectroscopy analysis (PNACL, Leicester University). The PNACL has a Kratos Kompact MALDI-ToF III (Matrix Assisted Laser Desorption Time of Flight) mass spectrometer.

2.2.2.5 Circular Dichroism (CD) Spectroscopy

When linearly-polarised light passes through an optically-active substance, its two circularly-polarised components (right and left circularly-polarised beams of light) travel at different speeds, and are absorbed in differing degrees by the substance. Thus, the light passing through the substance is elliptically polarised, and the substance is said to have "Circular Dichroism (CD)". The magnitude of circular dichroism is usually expressed in terms of molecular ellipticity [θ]. The phenomenon of circular dichroism is very sensitive to the secondary structure of polypeptides and proteins. It has been shown that CD spectra between 260 and 180 nm can be analysed for the different secondary structural types: alpha helix, parallel and antiparallel beta sheet, turn, and other. This has been realised quite early after CD was introduced and the standard curves shown (figure 2.1) were published in 1969 by Greenfield and Fasman [98]. CD spectra for distinct types of secondary structure present in proteins are different. The analysis of CD spectra can therefore yield valuable information about secondary structure of biological macromolecules.

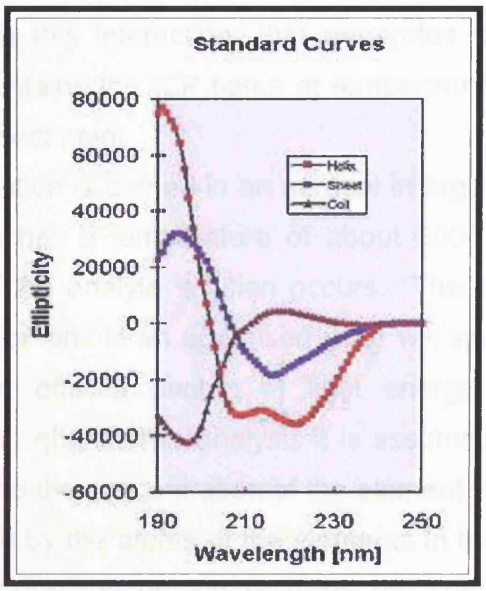

Figure 2.1 : CD standard curves of Greenfield and Fasman. Figure from reference [98].

Sample solutions at a concentration of 0.2 mg/ml were used for the CD experiments. The protein sample was dialysed against 20 mM phosphate buffer, pH 6.5, 50 mM NaCl. CD experiments were performed with a Jasco J-715 spectropolarimeter (JASCO Corporation) using a 1mm pathlength cuvette with the parameters described in table II.V. The data were processed using Jasco spectra

manager.

Table II.V : The CD spectropolarimeter setup.

Parameter	Setup
Range	195-260 nm
Sensitivity	50 mdeg
Resolution	0.5 nm
Accumulation	5
Band width	1.0 nm
Response time	1 sec
Scan speed	50 nm/min

2.2.2.6 ICP-OES

Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES) is a conventional spectroscopic technique; its unique properties are a due to the spectroscopic source used, the inductively coupled plasma. The ICP source is produced by a two turn induction coil which carries a high frequency current. It is the heating, which results from this interaction, that generates the ICP flame. The inductive heating effect maintains the ICP flame at temperatures of 6000 Kelvin (K) and up to 10000 K at its hottest point

The sample solution is carried in an aerosol in argon to the centre of the plasma flame where it reaches a temperature of about 8000 K. At this extreme temperature atomisation of the analyte solution occurs. The basis for all emission spectrometry is that atoms or ions in an energised state will spontaneously revert to a lower energy state and emit a photon of light energy, at a characteristic wavelength, as they do. For quantitative analysis it is assumed that the intensity of light emitted is proportional to the concentration of the element in solution.

The light emitted by the atoms of the elements in the ICP is focused into a spectrometer where a diffraction grating resolves the light into its component wavelengths. The intensity of light emitted at each given wavelength is then converted to an electrical signal by photomultiplier tubes located at specific wavelengths for each element line. Using calibration lines which relate elemental concentration with intensity of light emitted, the electrical signal is converted into a concentration measurement.

The protein samples were at a 0.02 mM concentration in a 10% HCl solution and analysed in the department of geology with the help of Emma Mansley, Leicester University.

2.2.3 Activity assays

Solid phase binding studies and haemolytic assays were done in collaboration with Sotiria Tzima, University of Leicester, UK and Anja Roos, Leiden University, NL.

2.2.3.1 C1q ELISA

Recombinant human calreticulin domains were tested for binding to purified native C1q (in collaboration with Anja Roos, Leiden University, NL). To assess the binding of C1q to the CRT S-, P- and PS-domains, 5 µg/ml of recombinant CRT domains was coated to an ELISA plate. Different concentrations of C1q were added to the wells and the plate incubated overnight at 4°C. The C1q concentrations used were ; 10, 5, 2.5, 1.25, 0.75, 0.375, 0.188, 0.094, 0.047, 0.023, and 0.012 µg/ml. After extensive washing with Tris-Buffered saline (TBS) (see apendix x.3), C1q was detected using biotinylated anti-C1q antibody. A_{415} values were measured.

2.2.3.2 C1q Haemolytic complement assay

The ability of recombinant calreticulin S-domain to bind to C1q and interfere with binding of $C1r_2C1s_2$ was determined by haemolysis assays (in collaboration with Sotiria Tzima, University of Leicester, UK). Pre-coating of sensitised sheep erythrocytes (EA) with complement components from early in the activation pathways allows activation and/or binding of the component under test. Lysis is achieved by adding the remaining components, which will only act if the assayed component is present. As all the components with the exception of the test component are present in excess, the degree of lysis is proportional to the concentration of the test component.

Sheep erythrocytes (E) and rabbit anti-sheep erythrocyte antiserum (A) were provided by TCS, Boltoph Claydon, Oxon, UK. For the preparation of EAC1q cells, 2 ml of EA cells at 10^9/ml in $DGVB^{++}$ (appendix X.3) were mixed with 1ml of fresh human serum diluted in 20 ml PBS-5mM EDTA. The mixture was incubated on ice for 30 min, centrifuged, and washed twice in $DGVB^{++}$ and finally resuspended in 18 ml of $DGVB^{++}$. C1q-deficient serum was obtained from a patient with homozygous C1q deficiency. EA and EAC1q cells were incubated, at 37°C for 1 hour, with serial dilutions of C1q-deficient serum, in order to establish the minimum serum concentration required to cause complete lysis of EAC1q. EAC1q cells (100

µl) were then preincubated, at 37°C for 1 hour, with recombinant CRT domains (100 µl: 100µg/ml). The cells were then incubated with the appropriate dilution of C1q deficient serum at 37°C for 1 hour. The extent of the lysis was established by the addition of cold $DGVB^{++}$ (1ml), centrifugation, and measurement of the A_{412nm} of the supernatant. Controls for 100% lysis comprised 100 µl cells and 1.1 ml water. Initially, fixed concentrations of recombinant CRT domains were tested. Based upon this evidence, the concentration dependence of inhibitory effects was then tested by the use of varying concentrations of recombinant CRT domains.

2.2.3.3 MBL ELISA

Recombinant human calreticulin domains were tested for binding to purified native MBL (in collaboration with Sotiria Tzima, University of Leicester, UK). A 96 well plate was coated with the recombinant CRT domains at 10 µg/ml in 15mM Na_2CO_3, 35 mM $NaHCO_3$, pH 9.6. Purified native MBL (in CRT buffer-10mM Tris, 10mM $CaCl_2$, 0.05% Triton X100, 0.1% HSA, pH 7.4, containing increasing concentrations of NaCl) was added to the wells and the plate incubated at 4°C, overnight. MBL binding to CRT domains was detected using mouse monoclonal anti-MBL antibody (Hyb 131-01, Statens Serum Institut).

2.3 RESULTS

2.3.1 Cloning of calreticulin domains into the pET vectors

Clone phCaR-1 [19] which contained the cDNA of human calreticulin was used as a template for PCR with primers CRTFWD forward and CRTREV reverse to produce different CRT recombinant domains (table II.VI and figure 4.3). The CRTFWD forward primers (table II.II) included a *Nde*I restriction site designed to allow in-frame subcloning into the expression vectors pET-15b and pET-11a. The CRTREV reverse primers (table II.II) included a stop codon and a *Bgl*II restriction site for cloning into the expression vectors. PCR products of the right size were amplified and isolated for cloning into pGEM-T Easy vector. After ligation and transformation, the recombinant colonies were screened using blue/white colour and the plasmids isolated from a single white colony, containing an insert. The plasmids were then digested using *Nde*I and *Bgl*II restriction enzymes and the inserts isolated from the agarose gel. The inserts were then ligated into the linearised pET-15b and pET-11a vectors. The recombinant plasmids were isolated from a single colony and sequenced with the T7 promoter and T7 terminator primers, to confirm that the insert was cloned in frame and did not contain any mutations.

The CRT S-domain (+2), amino acids 162-283, contained a cysteine in position 163. To prevent protein dimerisation, the only cysteine in the CRT S-domain was removed during the cloning process by removal of the two first amino acids at the N-terminus of the protein, CRT 164-283.

Table II.VI : Recombinant human calreticulin clones produced. The table shows the name of the different clones produced, their boundaries and length in amino acids, and the forward and reverse primers used for the PCR. For the primers sequence refer to table II.II in the material and methods.

Clone	*Residues (a.a)*	*Length (a.a)*	*Forward primer*	*Reverse primer*
S-domain (+2)	162-283	122	sCRTFWD	sCRTREV
S-domain	164-283	120	s2CRTFWD	sCRTREV
P-domain	198-308	111	pCRTFWD	pVRTREV
PS-domain	164-308	145	s2CRTFWD	pVRTREV

2.3.2 Purification of the calreticulin S-domain from pET-15b

The protein has been purified successfully using the protocol described in the material and methods and about 60 mg of pure protein can be obtained from one litre of culture. The CRT S-domain (+2) preparations following this purification technique were judged to be >95% homogenous by SDS-PAGE (figure 2.2, lane 7). No problems were encountered with thrombin proteolysis of the His-Tag (lane 5) and a ~29 kDa protein was isolated (lane 7). The identity of the CRT S-domain (+2) was confirmed by electrospray mass spectroscopy and N-terminal sequencing (table II.VII). An unusual migration in SDS-PAGE of the protein, results in a calreticulin S-domain appearance as approximately 30 kDa. This is likely to be due to its highly negative charge (pI = 4.7) and/or its high proline content. Similar observations have been made for all other CRT constructs purified.

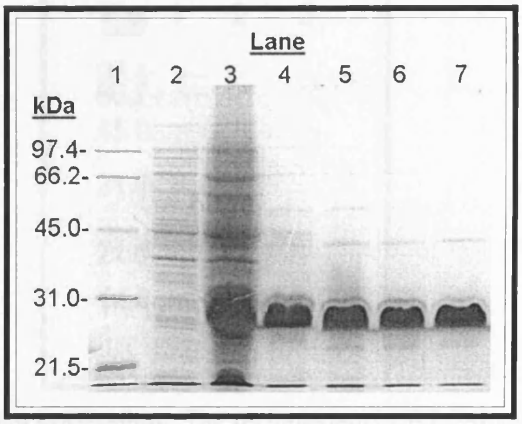

Figure 2.2 : CRT S-domain purification. The recombinant hCRT S-domain was subjected to 12% SDS-PAGE under reducing conditions containing : Low molecular weight marker as indicated at the left, lane 1; before induction, lane 2; induced cells, lane 3; Nickel column fraction, lane 4; after thrombin digestion, lane 5; second nickel column flow-through, lane 6; and Uno Q-6 fraction, lane 7.

Table II.VII : Data of electrospray mass spectroscopy and N-terminal sequencing for the hCRT domains purified from pET-15b. Mass determined by electrospray mass spectroscopy and their expected mass in brackets. Data for the N-terminal sequencing of the clones and their expected N-terminal sequence. The first four residues are cloning artefacts and underlined is the proteins sequence.

Clone	Mass in Daltons		N-terminal sequence	
	Measured	Expected	Measured	Expected
S-domain (+2)	14827.30 ± 7.06	14860.1	GSHMRC	GSHMRC
P-domain	13618,66 ± 1.20	13615.6	GSHMDD	GSHMDD
	8768.10 ± 0.41	unknown	GSHMDD	GSHMDD
PS-domain	Not determined	17813.2	GSHMRC	GSHMRC

2.3.3 Purification of the CRT P-domain from pET-15b

About 15 mg of pure CRT P-domain could be purified from one litre of culture. After thrombin digestion, two bands could be detected by SDS-PAGE with an apparent molecular weight of ~19kDa and ~14.4 kDa (figure 2.3, lane 2). The mass was determined by electrospray mass spectroscopy and the sequence of the first 6 N-terminal residues confirmed by N-terminal sequencing for the two fragments (table II.VII). The larger fragment could be purified using normal chromatography techniques (figure 2.3, lane 3) and used for characterisation. The preparations were judged to be >95% homogenous by SDS-PAGE (figure 2.3, lane 3).

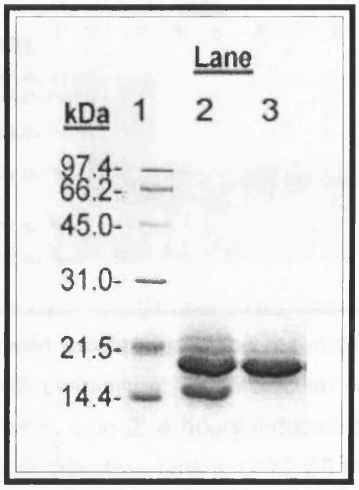

Figure 2.3 : CRT P-domain purfication. The recombinant hCRT P-domain was subjected to 15% SDS-PAGE containing : low molecular weight marker is indicated at the left; lane 1; after thrombin digestion, lane 2; and Uno Q-6 purified P-domain, lane 3.

2.3.4 Purification of the CRT PS-domain from pET-15b

The CRT PS-domain was isolated as described in the materials and methods and about 20 mg of pure protein could be obtained from one litre of culture. The CRT PS-domain preparations following this purification technique were judged to be less than 90% homogenous by SDS-PAGE (figure 2.4, lane 8). After thrombin digestion, one band could be detected by SDS-PAGE with an apparent molecular weight of ~31kDa (figure 2.4, lane 5). The sequence of the first 6 N-terminal residues of recombinant CRT PS-domain were confirmed by N-terminal sequencing.

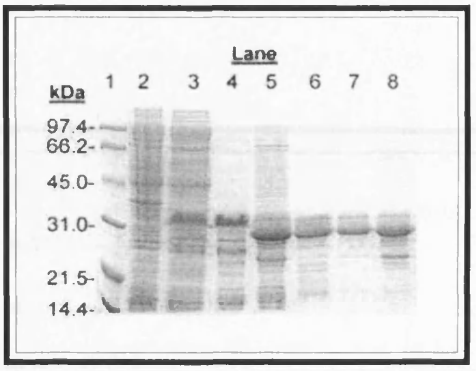

Figure 2.4 : CRT PS-domain purfication. The recombinant hCRT PS-domain was subjected to 15% SDS-PAGE containing : Low molecular weight marker is indicated at the left; lane 1;uninduced cells, lane 2; 4 hours induced cells, lane 3; Nickel column fraction, lane 4; after thrombin digestion, lane 5; UNO Q6 fraction, lane 6; Superdex-75 fraction, lane 7; and final concentrated sample, lane 8.

2.3.5 Circular dichroism

As mentioned in the material and methods, each of the three basic secondary structures of a polypeptide chain (helix, sheet, coil) show a characteristic CD spectrum. The spectrum of the CRT P- and S-domain show a trough at about 228 nm and a second one at 202 nm (figure 2.5). The trough at 202 nm predicts a mainly random coil, while a mainly helical or β-sheet structure should have a peak in this area of the spectrum (figure 2.1). The second trough at 228 nm suggest that some substantial local secondary structure exists and is more characteristic of the presence of β-sheet. By combining those two observations, we should predict that the CRT P- and S-domain have a limited secondary structure, with the presence of some β-sheet.

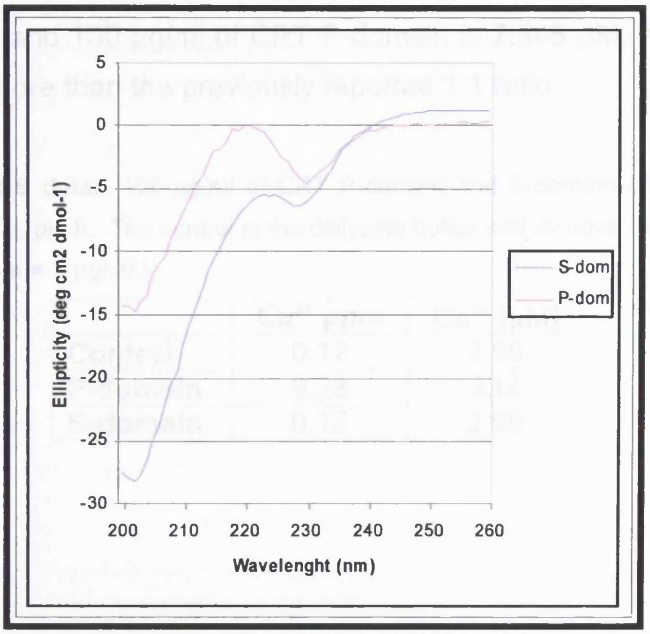

Figure 2.5 : Circular dichroism spectra after baseline subtraction of the recombinant CRT S- and P-domains. Circular dichroism spectra were collected in; 20 mM phosphate buffer, pH 6.5, 50 mM NaCl at 22°C with a protein concentration of 0.08 mg/ml for the P-domain and 50 mM phosphate buffer, pH 7.8, 50 mM NaCl at 22°C with a protein concentration of 0.4 mg/ml for the S-domain.

2.3.6 Calcium analysis

After extensive dialysis against metal-free buffer to remove loosely bound metal ions, preparations of CRT S-, and P-domain were analysed for calcium content by inductively coupled plasma-optical emission spectrometry (ICP-OES). As described in the introduction, a low-capacity and high-affinity calcium-binding site is present in the CRT P-domain, shown to contain 1 mol of Ca^{2+}/mol of protein [13, 99]. To assess whether calcium binding can be detected with the recombinant CRT domains, ICP-OES has been used as previously reported for the cytidine deaminase [100]. Table II.VIII shows that the CRT S-domain does not bind any calcium, compared with the control. However, the CRT P-domain bound 9.28 µg of calcium per 100 µg of protein, ratio of approximately 1:10 (W/W). ICP analysis revealed the presence of 32 atom of Ca^{2+} per molecule of CRT P-domain. (If 9.28 µg/ml of calcium is 232 µM and 100 µg/ml of CRT P-domain is 7.345 µM, we get a ratio of ~ 32:1). This is far more than the previously reported 1:1 ratio.

Table II.VIII : ICP-OES data. 100 µg/ml of CRT P-domain and S-domain after extensive dialysis against 10 mM Tris-HCl, pH 8. The control is the dialysate buffer and all other ions tested were below detection levels. (1 ppm = 1 µg/ml.)

	Ca^{2+} ppm	Ca^{2+} [µM]
Control	0.12	2.99
P-domain	9.28	232
S-domain	0.12	2.99

2.3.7 Activity assays

In order to confirm that the recombinant proteins purified are biologically active, different assays have been used. The following were used to test the multiple CRT clones for their potential complement inhibitory capacity.

2.3.7.1 C1q bind to CRT domains

To assess whether C1q binding to recombinant CRT S-, P- and PS-domains could be detected in an ELISA, CRT domains were coated on ELISA plates and incubated with C1q. Figure 2.6 shows a concentration dependent binding of C1q to coated CRT domains as to the control IgG (koIgG). These data shows that C1q can bind to the CRT domains tested and that C1q can be detected when bound to these CRT domains.

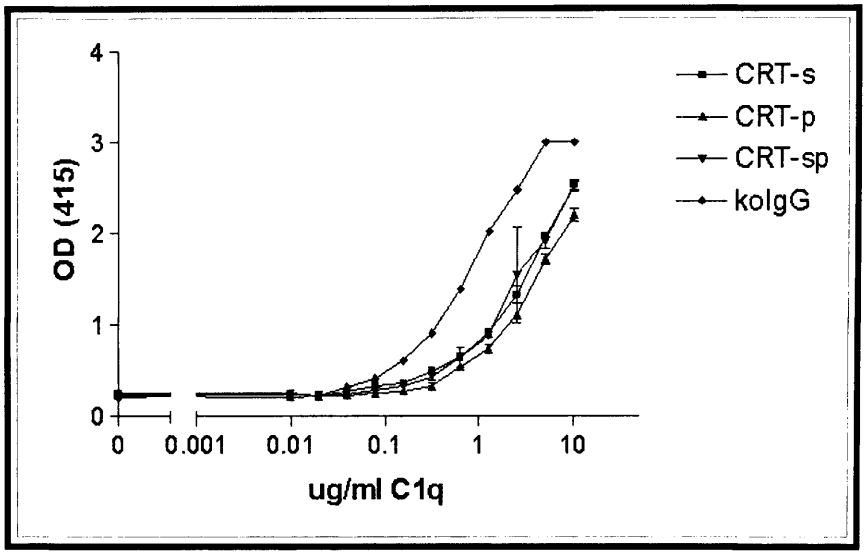

Figure 2.6 : Binding of C1q to immobilised CRT domains. To assess the binding of C1q to the CRT S-, P- and PS-domains, 5 μg/ml of recombinant CRT domains was coated to ELISA plates and different concentrations of C1q were added. After extensive washing, C1q was then detected with biotinylated anti-C1q antibody. A_{415} values were measured.

2.3.7.2 Inhibition of C1q-dependent haemolysis

As previously described in the literature, we used a C1q-dependent haemolytic assay to determine if CRT domains can specifically inhibit the classical pathway of complement. The assay requires purified C1q to be added back to C1q-deficient serum in order to reconstitute the C1 complex. In this present study, the addition of C1q back to C1q-deficient serum was given a relative value of 1 in the absence of CRT domains, because the C1q presented on the erythrocytes activates

the classical pathway in the C1q supplemented serum. This relative value was then used as standard in a series of studies to determine whether the CRT domains could specifically inhibit C1q-dependent haemolysis. Various concentration of CRT domains ranging from 0.1 to 100 µg/ml were incubated with C1q. C1q-deficient serum was then added, together with sheep erythrocytes sensitised with Rabbit antibodies, and incubated at 37°C for one hour.

It was observed that the S-domain of CRT inhibits haemolysis in a concentration dependent manner (figure 2.7). At low concentration of S-domain (<100 ng/ml), high percentage of erythrocyte lysis is observed due to classical pathway activation from the C1q contained on the erythrocytes. At higher concentration of S-domain, low percentage of lysis is observed, because the S-domain binds C1q and prevents complement activation. The CRT S-domain inhibits the C1q activity with a 50% inhibition (IC50) of ~ 20 µg/ml. Interestingly, the CRT P- and PS-domains did not show any inhibition of C1q activity in the protein concentration range tested.

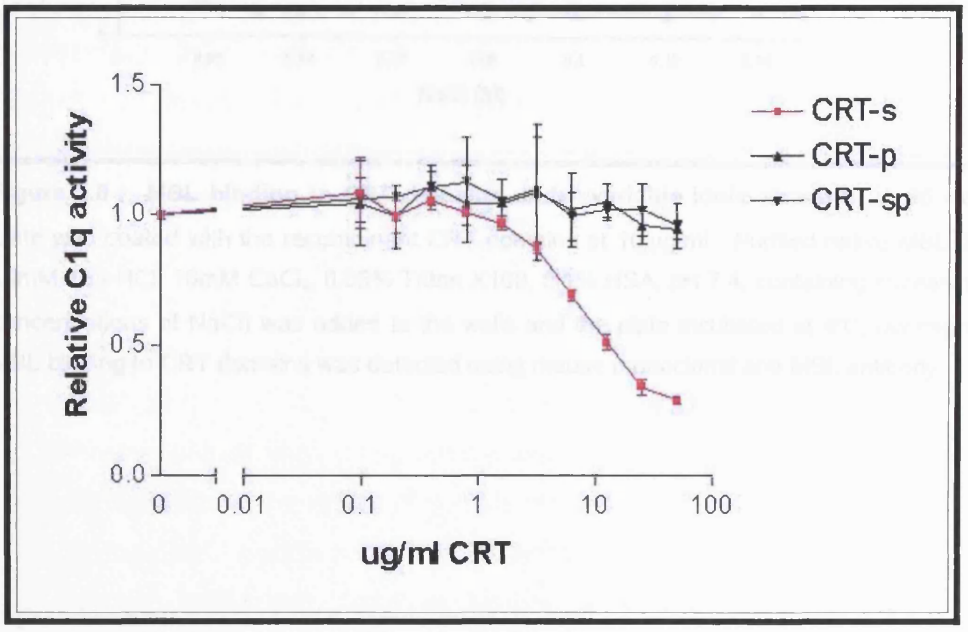

Figure 2.7 : Inhibition of C1q haemolytic assay by CRT domains. Different concentrations of CRT S-, P- and PS-domains were pre-incubated with C1q. C1q-deficient serum was then added, together with sheep erythrocytes sensitised with Rb antibodies, and incubated at 37°C. After 1 hour, release of haemoglobin was measured by A_{414}. The percentage of inhibition of haemolysis was determined relative to lysis in the presence of C1q without CRT domains. Data obtained in collaboration with A. Roos *et al* and S. Tzima.

2.3.7.3 Binding of recombinant calreticulin domains to MBL

Recombinant human calreticulin S- and P-domain were tested for binding to purified native MBL. As shown in figure 2.8, binding of recombinant S-domain and P-domain to MBL can be detected. It can be noticed that as the concentration of NaCl increases, the binding to MBL decreases. The maximum binding observed for the P-domain is at 10 mM NaCl and without any NaCl for the S-domain (figure 2.8).

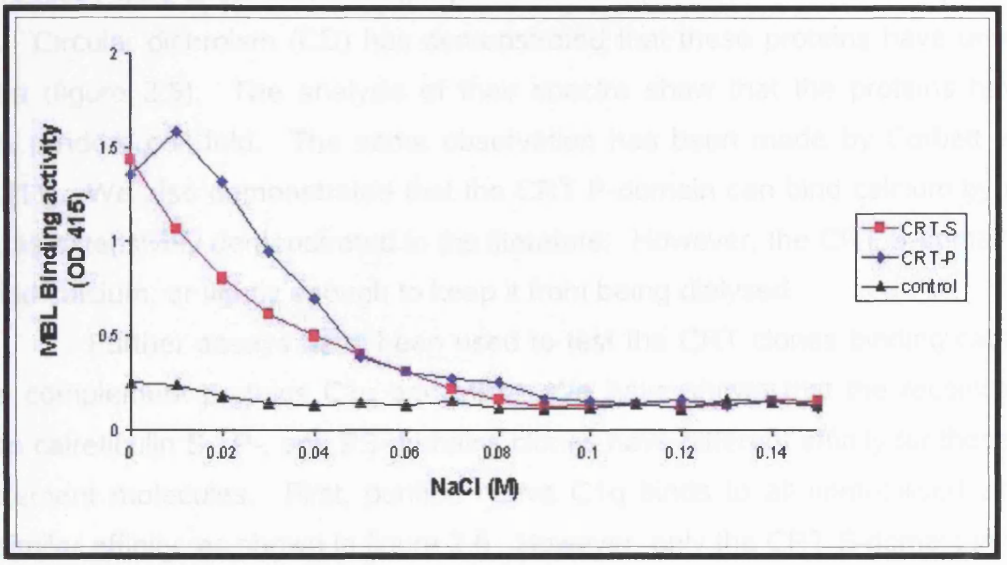

Figure 2.8 : MBL binding to CRT domains under variable ionic strength. A 96 well plate was coated with the recombinant CRT domains at 10 μg/ml. Purified native MBL (in 10mM Tris-HCl, 10mM $CaCl_2$, 0.05% Triton X100, 0.1% HSA, pH 7.4, containing increasing concentrations of NaCl) was added to the wells and the plate incubated at 4°C, overnight. MBL binding to CRT domains was detected using mouse monoclonal anti-MBL antibody.

2.4 DISCUSSION

Multiple human calreticulin clones have been successfully cloned into the pET vectors. Theses clones included the hCRT; S-domain (+2) (residues 164-283), P-domain (198-308), and PS-domain (164-308). Also, theses three clones have been expressed in *E.coli* and purified to a high standard, more than 95% pure. They were all highly soluble and stable. N-terminal protein sequencing and mass spectroscopy also show that the right proteins were purified and have the predicted mass. Circular dichroism (CD) has demonstrated that these proteins have unusual spectra (figure 2.5). The analysis of their spectra show that the proteins have a mainly random coil fold. The same observation has been made by Corbett *et al*, 1999 [11]. We also demonstrated that the CRT P-domain can bind calcium by ICP-OES, as extensively demonstrated in the literature. However, the CRT S-domain did not bind calcium, or tightly enough to keep it from being dialysed.

Further assays have been used to test the CRT clones binding capacity to the complement proteins C1q and MBL. We have shown that the recombinant human calreticulin S-, P-, and PS-domains clones have different affinity for these two complement molecules. First, purified native C1q binds to all immobilised clones with similar affinity, as shown in figure 2.6. However, only the CRT S-domain inhibits the classical complement pathway in the C1q-dependent haemolytic assay performed (figure 2.7). A higher concentration of the P- and PS-domain could be needed to observe the same inhibition (experiment not performed) suggesting that the S-domain affinity for C1q is higher. These data indicate that the calreticulin domains either bind to C1q ; in the collagen stalk and interfere with the $C1r_2s_2$ binding, thereby preventing complement activation or to the globular heads, preventing recognition of the immune complexes.

In addition, purified native MBL binds to the CRT S- and P-domains at low ionic strength buffer with optimum binding at 10 mM NaCl (figure 2.8). The relevance of this binding might be minimal, considering the physiological salt concentration in the blood is about 150 mM. All CRT clones, including the native calreticulin, are potential inhibitors of the MBL mediated lectin pathway. However, only the S-domain in the experimental conditions performed seems to be a C1q mediated classical complement pathway inhibitor. More experimental data are needed to understand this mechanism in details. Considering the 85 amino acids overlapping protein sequence in the three CRT clones (a.a. 198-283), we have partly localised the binding site. The structure of one of theses CRT clones could help to understand more on these interactions.

CHAPTER 3 :
CRYSTALLOGRAPHY

3.1 INTRODUCTION

The prerequisite for an X-ray crystallographic study is the crystallisation of the macromolecule of interest as an ordered crystal that diffracts to a resolution high enough to answer the relevant biological questions [101]. Although X-ray crystallography has become a far more powerful technique during the last decades the 'art' of crystallising a macromolecule is still but poorly understood and is certainly not trivial.

The fact that the crystallisation of a macromolecule has become the primary obstacle for a successful crystallographic study can be explained by the extreme complexity of the molecules themselves. In contrast to small molecules that often crystallise easily by desiccation, macromolecules are structurally dynamic, micro-heterogeneous, and aggregating systems that only maintain their conformation in the fully hydrated state. Their complex surface properties that are a mixture of the specific combinations of amino-acid side-chains protruding into the solvent, depends on many environmental parameters, such as salt concentration and pH.

A crystallisation experiment typically begins with the protein sample in a stabilising solution of salted buffer and possibly other additives such as reducing agent, EDTA, etc. Prior to mixing the sample with crystallisation reagent, this sample solution is highly concentrated (2-20 mg/ml) with respect to the macromolecule in question. In an undersaturated sample solution, no crystals can nucleate, nor can crystals grow from seeds. Upon addition of a precipitant reagent the relative supersaturation of the sample is increased. Assuming the crystallisation reagent decreases the solubility of the sample to increase the relative supersaturation, three events can take place [102]. In the first stage of supersaturation, the metastable zone, spontaneous homogenous nucleation cannot occur, but crystal growth from seeds can occur (figure 3.1). Moving further into supersaturation, the labile zone, spontaneous homogenous nucleation and crystal growth can occur. Further into supersaturation, the precipitation zone, precipitation of the sample from solution occurs, see figure 3.1.

Crystals can grow from a supersaturated solution with respect to the protein (or DNA) until the system reaches equilibrium. The spontaneous nucleation of crystals, however, is only possible in the labile region whereas nuclei or crystals can grow but not emerge in the meta-stable region (figure 3.1). Although crystal-growth is thermodynamically more favourable, the all too frequent emergence of amorphous precipitate is kinetically favoured. It is therefore essential to slowly bring the system to a supersaturated state that, upon passing the borderline between

meta-stable and labile regions, returns to equilibrium by the nucleation and successive growth of crystal(s). If the penetration of the labile region is too far or too fast, showers of crystals arise that all grow to a size unsuitable for X-ray studies and often are plagued by the occurrence of flaws and dislocations associated with rapid growth. In growing crystals, one therefore attempts, by either dehydratation or alteration of physical conditions, to transport the solution into a labile, supersaturated state as close as possible to the meta-stable state. It must be kept in mind that the properties of macromolecules are changing as the conditions are shifting.

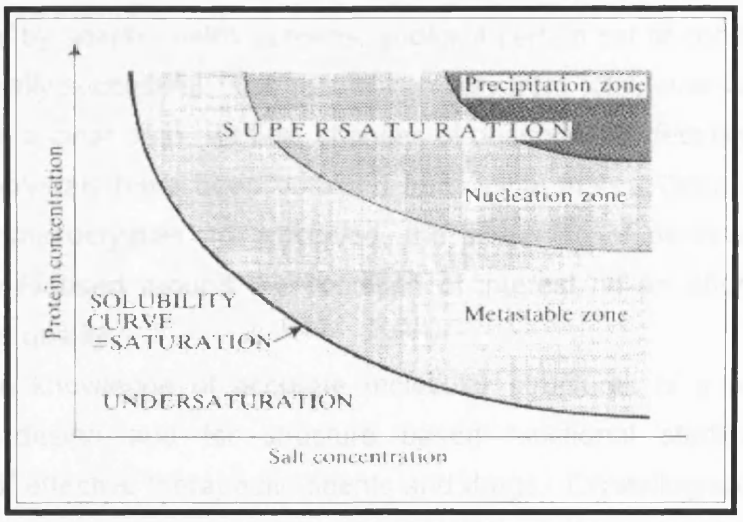

Figure 3.1 : Theoretical phase diagram of a protein in solution. Figure adapted from reference [102].

The most often used techniques to slowly bring a protein solution to supersaturation, is the addition of a gentle precipitant that is slowly increased in concentration, mimicking dehydration without complete physical removal of water. Since macromolecules are 'sensitive', the only conditions that can support crystal growth are those that cause little or no perturbation of the molecular properties. Whereas salts at high concentrations compete with the protein for water molecules to satisfy their hydration-needs (which effectively forces protein out of solution), organic polymers like polyethylene glycol dehydrate proteins and alter the dielectric properties of the medium.

The supersaturated state can be created by a range of methods, including dialysis, batch crystallisation, free interface diffusion and vapour diffusion. The latter method is the most prevalent because it requires a minimum amount of protein. A drop of the protein solution, mixed with a low amount of precipitant, is sealed in a reservoir together with a larger volume of precipitant solution. As the

concentration of precipitant in the reservoir solution is greater than that in the protein drop, water will evaporate from the drop to the reservoir solution thereby decreasing the volume of the drop and increasing both protein and precipitant concentration. In the case of volatile precipitants, both distillation of water out, and diffusion of precipitant in the droplet occur.

In practice, one starts a crystallisation experiment by setting up a series of trials that each vary in one or more parameters. As the protein concentration is often 10mg/mL and 2µL is routinely used for each trial, substantial amounts of pure protein are needed. Initial tests are often conducted to probe the behaviour of the macromolecule by sparse matrix screens, and/or a certain set of conditions that has proven statistically successful. The results can vary from total, over various forms of precipitation, to a clear drop. At first, chances of observing perfect crystals are slim, nevertheless, crystals have been obtained from initial trials. Once an 'interesting precipitate' or microcrystals are observed, the screening of parameters becomes more narrowly focused around the condition of interest, in an effort to maximise crystal size and quality.

The knowledge of accurate molecular structures is a prerequisite for rational drug design and for structure based functional studies to aid the development of effective therapeutic agents and drugs. Crystallography can reliably provide the answer to many structure related questions, from global folds to atomic details of bonding. In contrast to NMR, which is an indirect spectroscopic method, no size limitation exists for the molecule or complex to be studied. The price for the high accuracy of crystallographic structures is that good crystals must be found, and that only limited information about the molecule's dynamic behaviour is available from one single diffraction experiment.

3.2 MATERIAL AND METHODS

3.2.1 Protein purification

The recombinant protein was purified as described previously in chapter 2. The fractions containing the protein were pooled and dialysed in two steps at 4°C (20 mM Tris-HCl, pH 8.0, 50 mM NaCl) using a Spectra/Por 3 tubing: 3.5kDa MWCO (Spectrum Laboratories, US). After dialysis, the protein was concentrated to 10 mg/ml using a Centriprep YM-3 (Millipore Limited, U.K.). Small aliquots were flash frozen in liquid nitrogen and stored at -80°C.

3.2.2 Purification with DTT

The recombinant protein was purified as described previously with the addition of DTT. The flow through of the second nickel column was collected and the disulfide bonds reduced with 20 mM DTT for 1 hour at room temperature. The reduced sample was then applied to a Superdex 75 gel filtration column (Amersham Pharmacia Biotech) pre-equilibrated with 20 mM MES, pH 6.0, 150 mM NaCl, 2 mM DTT. The fractions were analysed by SDS-PAGE and the fractions containing the protein were pooled. The CRT S-domain was then purified on a Uno Q-6 column (Bio-Rad) and eluted with a salt gradient. The binding buffer was 20 mM MES pH 6.0, 150 mM NaCl, 2 mM DTT and the elution buffer was containing 1 M NaCl. The fractions of interest were pooled, dialysed and concentrated as described before with a modified buffer containing 2 mM DTT.

3.2.3 Crystallisation

Initial screens were used to identify the nucleation and/or crystallisation conditions. Six different screening packages were used; Wizard 1 and 2, Cryo 1 and 2 (Emerald BioStructures, Bainbridge Island, USA) and the structure screen 1 and 2 (Molecular Dimensions Limited, Soham, UK). These different screens were set up at 20 and 4 °C with ratios of 1 µl : 1 µl and 2 µl : 1 µl (volume of protein solution to reservoir solution).

Certain conditions have been optimised, based on the initial screen, to obtain large single crystals (>0.2 mm) suitable for X-ray. Different salt, PEGs, buffers, pH, precipitant have been tried at various concentrations. To further improve the crystals, certain additives and detergent have been added to the drops. The additive screens 1, 2 and 3 and the detergent screens 1, 2 and 3 from Hampton Research (Aliso Viejo, USA) were used.

3.2.4 Proteolysis

3.2.4.1 Introduction

In considering the action of proteinases on unfolded proteins, the main factor to be taken into account is the bond specificity exhibited by the proteinase. Any given proteinase will follow a characteristic digestion pattern to yield a defined set of peptide fragments, which can be separated and subsequently characterised. However in the case of folded or native proteins an extra factor must be taken into account, namely the accessibility of susceptible bonds to the proteinase. This can be a complex subject, since some proteinases have extended active sites and thus the accessibility of the bond to be cleaved cannot always be considered in isolation [103].

In general, native domains are rather resistant to the action of proteinases, a consequence of the tight, compact or "domain" structures adopted by most globular proteins. The parts of the polypeptide chain that are likely to be accessible to proteinases are exposed loops within domains or the linking regions of polypeptide chain between domains. These accessible regions could (1) be present in the native structure of the protein; (2) arise from conformational changes in the protein; or (3) be generated during the folding or unfolding of the protein. So the proteolysis of the calreticulin S-domain by different kinds of proteases could give us some information about its folding.

3.2.4.2 Protocol

Purified recombinant calreticulin S-domain was cleaved with trypsin and chymotrypsin (Sigma). The concentration of the proteinases used were of 1:70 (w/w) of the target protein, and time course reactions have been done over 100 minutes at 4°C. The buffer conditions used were 20 mM Tris-HCl, pH 7.5, 150 mM NaCl. The proteinases could be inactivated by addition of SDS to a final concentration of 0.1%, rapidly raising the temperature of the samples to close to 100°C and incubating at this temperature for 2 min. Samples have been taken at different times; 0, 1, 5, 10, 20, 30, 100 minutes and analysed by 15% SDS-PAGE. The fragments have then been sequenced.

3.3RESULTS

Initial screening for crystallisation conditions were carried out with and without DTT added to the protein buffer. Due to the presence of one cysteine in the recombinant protein, non-physiological dimers were formed (figure 3.2) and this may have interfered with the crystal formation. For this reason, the CRT S-domain has been purified with and without DTT. From the difference in precipitation behaviour during the first tests it was concluded that the presence of DTT had a slight effect on the solubility of the protein. A first crystallisation condition was found to be 10% (w/v) PEG-3000, Cacodylate, pH 6.5, 200 mM MgCl2. Despite extensive screening of variables such as PEG, pH, temperature, salt, concentration of precipitant used and of protein and experimental set up, only short needles were obtained that were too small to diffract (Figure 3.3). Other small crystals have been obtained in the following conditions; 20% (w/v) PEG-3000, HEPES pH 7.5, 200 mM NaCl. Again, extensive screening did not improve the initial conditions to diffraction quality crystals.

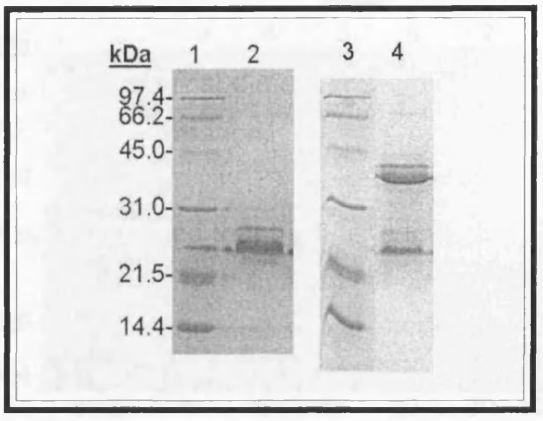

Figure 3.2 : SDS-PAGE of the CRT S-domain under non-reducing conditions. The recombinant hCRT S-domain was subjected to 12% SDS-PAGE containing β-mercaptoethanol, lane 2, and without, lane 4. Low molecular weight marker is indicated at the left; lane 1 and 3.

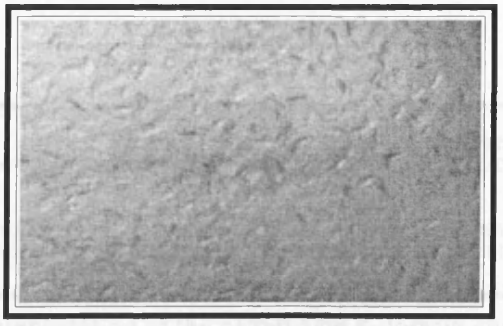

Figure 3.3 : Picture of the calreticulin S-domain needles.

The seeding technique has been tried with the above crystals, but the crystals were extremely fragile and sticky. This technique did not help in our case.

The action of proteases, chymotrypsin and trypsin, on the protein has been assessed to determine the stability of the protein. The incubation of the CRT S-domain with trypsin produced an almost complete digestion of the protein (data not shown). However, incubation with chymotrypsin leads to the formation of two smaller fragments of the protein (figure 3.4). Due to the unusual migration of the protein on SDS-PAGE, the actual size of the fragments cannot be approximated. The N-terminal sequencing of the smaller fragment gave the sequence Thr-Leu-Ile-Val-Arg-Pro, indicating cleavage from the N-terminal of 15 amino acids (figure 3.5). The chymotryptic fragment of the CRT S-domain could not be separated from the full recombinant protein with standard chromatography techniques, preventing mass spectroscopy being performed.

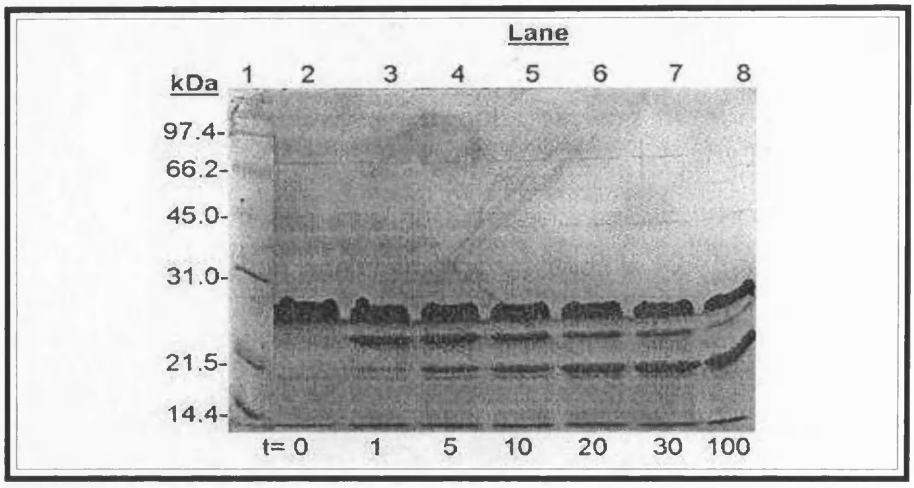

Figure 3.4 : The recombinant hCRT S-domain was subjected to chymotrypsin digestion for different time periods. 15% SDS-PAGE containing : Low molecular weight marker is indicated at the left; lane 1; undigested protein, lane 2; 1 min digestion, lane 3; 5 min digestion, lane 4; 10 min digestion, lane 5; 20 min digestion, lane 6; 30 min digestion, lane 7 and 100 min digestion, lane 8.

```
S-chymo         ---------------TLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKDPDASKPEDWDERAKIDD    55
S-Whole         gshmrckddefthlyTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKDPDASKPEDWDERAKIDD    70

S-chymo         PTDSKPEDWDKPEHIPDPDAKKPEDWDEEMDGEWEPPVIQNPEYKGEWKPRQIDNP                   111
S-Whole         PTDSKPEDWDKPEHIPDPDAKKPEDWDEEMDGEWEPPVIQNPEYKGEWKPRQIDNP                   126
```

Figure 3.5 : Schematic representation of the calreticulin S-domain chymotryptic fragment.

3.4DISCUSSION

Only small crystals of the human calreticulin S-domain could be obtained after extensive screening. Purification and crystallisation of the protein under reducing conditions did not improve the quality of the crystals. Proteolysis showed that the protein is relatively stable to chymotrypsin and very sensitive to trypsin. Furthermore, as we have shown in the previous chapter, the CRT S-domain is active. These data combined suggest that the protein must be properly folded, but the crystallogenesis of this protein seems problematic.

CHAPTER 4 :
THE SOLUTION NMR STRUCTURE AND DYNAMICS OF THE CALRETICULIN S-DOMAIN

4.1 INTRODUCTION

Calreticulin (CRT) is an abundant, soluble molecular chaperone of the endoplasmic reticulum. Similar to its membrane-bound homolog calnexin (CNX), it is a lectin that promotes the folding of proteins carrying N-linked glycans. Previously, in chapter 2, it is shown that a small recombinant domain of CRT, the S-domain, binds to C1q and MBL *in vitro*. Following binding to these molecules, the CRT S-domain acts as a potent inhibitor of the classical and lectin pathway of complement. To understand more in details the mechanism by which calreticulin interacts with the complement molecules C1q and MBL, the three-dimensional structure of the calreticulin S-domain has been determined by NMR. This structure will provide the basis for designing a low-molecular weight compound with a therapeutic potential in the treatment of diseases in which complement activation contributes considerably to the pathology, such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE).

For determining solution protein structures from NMR data, many steps are required. First, samples were purified to homogeneity following multiple labelling, i.e. ^{15}N and $^{13}C/^{15}N$, and spectra optimised by choosing an appropriate buffer and temperature. The protein stability was also increased by addition of protease inhibitors, changing expression vectors and optimisation of purification conditions. The NMR experiments were then acquired and the individual peaks in the spectra assigned to sequence-specific locations in the chemical structure of the protein. For this purpose, the 3D triple-resonance experiments [104] were used to assign the backbone atoms followed by the assignments of the side-chains, using ^{13}C-resolved H(C)CH-COSY and (H)CCH-COSY spectra. The NOESY spectra were then interpreted for the determination of the three-dimensional solution structure of the protein. For this purpose ARIA (**A**mbiguous **R**estraints for **I**terative **A**ssignment) was used [105], a software protocol that integrates automated NOE assignments into structure calculations. The refined structure shows an unusual extended hairpin topology, with one short anti-parallel β-sheet and a long disordered N-terminal. This short structured hairpin, residues 239 to 269, may be used as a low-molecular mass mimic of CRT for further investigation of the interactions with C1q and MBL.

Finally, to further characterise the CRT S-domain structure, relaxation experiments were performed to study the dynamics of the protein in solution. The heteronuclear relaxation studies have focused on the amide $^{15}N/^{1}H$ spin system [106, 107], thereby allowing the local dynamics along the protein backbone to be explored residue by residue.

4.2 NMR METHODOLOGY

The following section describes the main steps in the determination of a protein structure from NMR data. First, a description of the triple-resonance experiments used for the backbone assignment of the protein. Then, the basis of the NOE effect used in the ^{15}N- and ^{13}C-resolved NOESY experiments and based on those spectra, the strategy used for the structure determination. Finally, an introduction to NMR relaxation and the model-free analysis which was used to study the intra-molecular motions in the protein structure determined.

4.2.1 Triple-resonance experiments

NMR spectra contain sufficient information to determine protein structures in solution. However, none of the information can be used without sequence-specific resonance assignment that associates NMR signals with specific atoms in the protein. Three-dimensional heteronuclear triple-resonance experiments are the method of choice for the sequence-specific resonance assignment of large proteins (>100 a.a.). These experiments are called triple-resonance because they correlate ^{1}H, ^{15}N, and ^{13}C spins using one-bond and two-bond scalar coupling interactions [104]. A variety of different heteronuclear triple-resonance experiments were performed on the doubly labelled [$^{13}C/^{15}N$] calreticulin S-domain. Each experiment provided different chemical shift information using different through-bond couplings for magnetisation transfer, as shown in table IV.I and figure 4.1. For instance in the HNCO experiment magnetisation was transferred from the amide proton (H^N), after which the signal was detected. Only protons coupled in this way to carbonyl carbons (CO) of the preceding residue (*i*-1) showed up as cross-peaks. The position of a single cross-peak in the 3D spectrum gave the chemical shifts information of the intra-residue (*i*) amide proton and amide nitrogen and of the carbonyl carbon of residue *i*-1.

Table IV.I : Chemical shift information obtained from each triple-resonance experiment. H^N: amide proton, N: amide nitrogen, CO: carbonyl carbon, C^α: backbone alpha carbon, C^β: side-chain beta carbon. Table interpreted from reference [104].

Triple-resonance experiment	*Chemical shift information obtained for residue i*	*Chemical shift information obtained for residue i-1*
HNCO	H^N, N	CO
HN(CO)CA	H^N, N	C^α
HNCA	H^N, N, C^α	C^α
CBCA(CO)NH	H^N, N	C^α, C^β
CBCANH	H^N, N, C^α, C^β	C^α, C^β

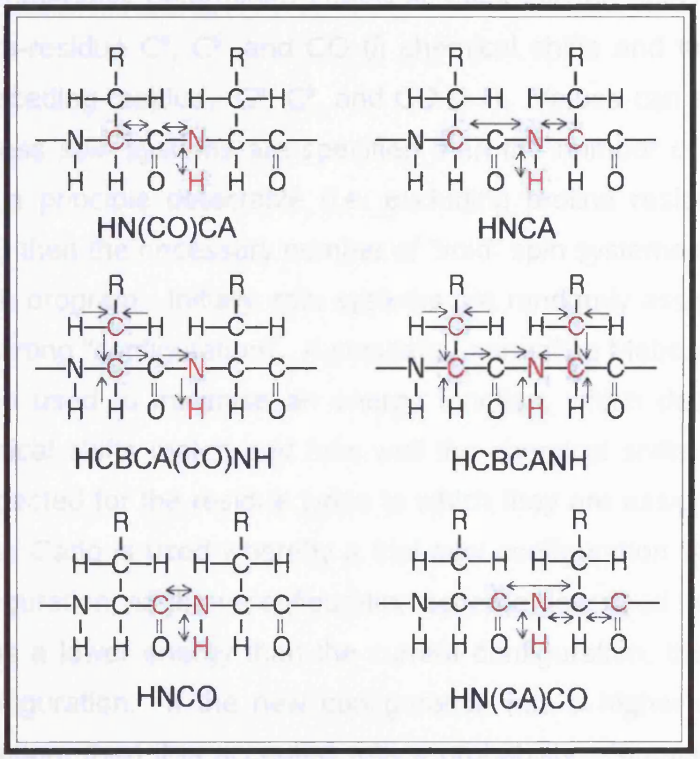

Figure 4.1 : The magnetisation transfers associated with each triple-resonance experiment used on the [$^{13}C/^{15}N$] CRT S-domain. Pairs of experiments are shown side by side. Figure adapted from reference [104].

Spin system identification was the first part of the assignment procedure following data collection. A particular chemical shift peak was classified as belonging to a particular atom type and the peaks then grouped into dipeptide spin systems. In each of these systems there was chemical shift information about H^N, N, CO, C^α and C^β from the current residue (*i*) and the previous residue (*i*-1) in the sequence.

4.2.1.1 Simulated annealing assignment

The next step was the sequential assignment, ordering these dipeptide spin systems in a manner consistent with the primary sequence of the protein, and chemical shift information for the residue type. For this, a global optimisation procedure, which employs a simulated annealing algorithm was used [108]. This determined the sequential arrangement of spins which had the lowest energy in terms of both chemical shift linkage between adjacent spin systems, and the agreement of these systems with the known protein sequence.

The primary input for the simulated annealing program is the amino acid sequence and a table of arbitrarily numbered spin systems. For each spin system, a number of experimentally determined chemical shifts can be specified. These are typically the intra-residue C^{α}, C^{β}, and CO (*i*) chemical shifts and the corresponding shifts for the preceding residue, C^{α}, C^{β}, and CO (*i-1*). Values can also be specified as missing. If less spin systems are specified than the number of residues for NH correlation are in principle detectable (i.e. excluding proline residues, and the N-terminal residue) then the necessary number of "void" spin systems are automatically generated by the program. Initially spin systems are randomly assigned to residues to generate a starting "configuration". A simulated annealing Metropolis Monte Carlo algorithm is then used to minimise an energy function, which describes how well sequential chemical shifts match and how well the chemical shifts of spin systems match those expected for the residue types to which they are assigned. A standard Metropolis Monte Carlo is used whereby a trial new configuration is generated from the current configuration via the reconfiguration scheme described below. If this new configuration has a lower energy than the current configuration, then it is accepted as the new configuration. If the new configuration has a higher energy than the current configuration, then it is accepted with a probability. Simulated annealing is achieved by iterating the Monte Carlo procedure at an elevated 'temperature' and then reducing it gradually. The output of the simulated annealing was analysed manually and the good sequential spin systems could be 'locked' into the simulated annealing to increase the accuracy of the following outputs.

4.2.2 Nuclear Overhauser enhancement

Experiments based on the Nuclear Overhauser Effect (NOE) occupy a very special place in the repertoire of modern NMR methods. The interaction involved is the direct magnetic coupling between nuclei (the dipolar coupling). The NOE provides an indirect way to extract information about this dipolar coupling, which in turn can be related to internuclear distances and molecular motions. Measurements of these parameters for molecules in solution are hard to obtain by other means, making the NOE an extremely important phenomenon. The NOE is generally summarised by the statement : 'the NOE is proportional to the inverse of the distance between nuclei raised to the sixth power' [106].

As a brief description consider the determination of protein structures from NMR data. The structural information mainly comes from NMR cross peaks. An NOE peak between two hydrogen atoms (or groups of hydrogen atoms) is observed if these hydrogens are located at a shorter distance than approximately 5.0 Å from each other. Without sequence-specific resonance assignment it is impossible to determine to which pair of hydrogen atoms a specific distance constraint refers. On the other hand, combined with resonance assignment these distance constraints can be attributed to specific sites along the protein chain and therefore the three dimensional structure can be formed. Analysis of the NOESY spectrum, the 'NOE assignment', provides many distance constraints between the hydrogen atoms of a protein. The inter-proton distance can be calculated from the intensity of the NOE cross peaks. Generally speaking, an NOE peak with strong intensity may indicate that two protons are within 2.5 Å of each other while a weak NOE peak corresponds to an upper limit of 5.0 Å.

4.2.2.1 3D ^{15}N-resolved NOESY experiments

2D NOESY spectra of large proteins, like the CRT S-domain, get crowded with signal. However, the separation of the NOE cross-peaks involving ^{15}N-bound protons can be increased using the 3D [^{1}H,^{15}N]-NOESY-HSQC experiment. This third dimension is achieved by combining two type of experiments, like NOESY and HSQC, in a single 3D experiment by extending the NOESY experiment by an HSQC step. The resulting experiment is called 3D NOESY-HSQC, and in a similar way, a TOCSY-HSQC can be obtained. Cross-peaks in such spectra have three coordinates, two of which are the intra-residue ^{1}H and ^{15}N chemical shifts of the backbone HN group and the third is the ^{1}H chemical shift of the spatially close

protons, as summarised in table IV.II. The resolution increase in going from 2D to 3D spectrum is due to the high dispersion of ^{15}N chemical shifts in the HSQC step. Most of the HN groups produce well-resolved and easily identified strips of signals.

The TOCSY experiment is useful for determining which signals arise from protons within a spin system, especially when the multiplets overlap. TOCSY yields long range as well as short range correlation and the cross peak intensity is not an indicator of distance.

Table IV.II : Chemical shift information obtained from each experiment. H^N: amide proton, N: amide nitrogen, H^α: backbone alpha hydrogen, H^β: side-chain beta hydrogen(s). An entry of 'NA' indicates that no chemical shift are observed in those experiments.

Experiment	*Chemical shift information obtained for residue i*	*Chemical shift information obtained for residue i-1*
[^1H,^{15}N]-TOCSY-HSQC	H^N, N, H^α, H^β, H^γ, etc.	NA
[^1H,^{15}N]-NOESY-HSQC	H^N, N, H^α, H^β, H^γ, etc.	Atoms close in space
HBHA(CO)NH	NA	H^α, H^β

4.2.2.2 3D ^{13}C-resolved NOESY experiments

Resolution of NOE cross-peaks involving carbon-bound protons is improved in 3D [^1H,^{13}C]-NOESY-HSQC spectra. In theory, there is no necessity to perform ^{13}C-resolved NOESY experiment using samples in H_2O, the NOESY experiment in D_2O is often more appropriate. However, in practice, it is more desirable to do both experiments to notice any possible spectral changes due to sample variations.

Performing the HMQC experiment on a sample in D_2O gives the advantages of small base-line distortions due to the decrease of solvent signal and also the chance of observing resonnances close to the solvent signal. The large dispersion of ^{13}C chemical shifts is often limiting the digital resolution that can be achieved in reasonable experimental time. Sufficient resolution is usually obtained by folding the spectrum in the indirect ^{13}C dimension.

4.2.3 Structure determination

The first step is to determine an initial protein structure which is consistent with the thousands of NOE constraints and with some other conformational constraints. For this purpose ARIA (**A**mbiguous **R**estraints for **I**terative **A**ssignment) was used, a software protocol that integrates automated NOE

assignments into structure calculations. The user provides a list of assigned chemical shifts and uninterpreted or partly assigned multidimensional homonuclear or heteronuclear resolved NOE cross-peaks lists. Additionally, torsion angles, *J* couplings, residual dipolar coupling, H-bonds, disulfide bridges, and planarity restrains can be specified. ARIA then merges the distance restraint lists and sets up all restraints for automated structure calculation. Explicit assignments are obtained iteratively from chemical shift assignments and successive generations of calculated structures. The initial structures calculated by ARIA almost always violate many of the experimental constraints. Subsequently structure refinement is required to obtain a high resolution protein structure.

4.2.4 Relaxation and dynamic processes

NMR spectroscopy is sensitive to motional processes in molecules through the phenomenon of nuclear spin relaxation. These studies rely on the observation of certain spectral properties in distinct NMR peaks that can be correlated with intra-molecular motions. Once the NMR peaks responsible for the study region (such as an amide proton) have been assigned, it is possible to investigate the desired spectral properties in the corresponding spectra.

As relaxation is one of the most fundamental aspects of magnetic resonance, an extensive description would be irrelevant for the present work. Once a radiofrequency (RF) pulse has been applied, the populations of α and β states are disturbed from their equilibrium values. Those perturbed populations will relax back to their equilibrium position through two types of relaxation processes, known as T_1 and T_2 relaxation (figure 4.2) [106]. The time constant T_1 is known as the longitudinal relaxation time, it is the time taken for the population of the energy states to return to the original Boltzman distribution, and hence the time taken for the net *z* magnetisation to return to its original value along +*z*. T_2, also known as transversal relaxation, is a measure of the time taken for the coherence in the *xy* plane (generated after a 90° pulse on +*z* magnetisation) to dephase to zero.

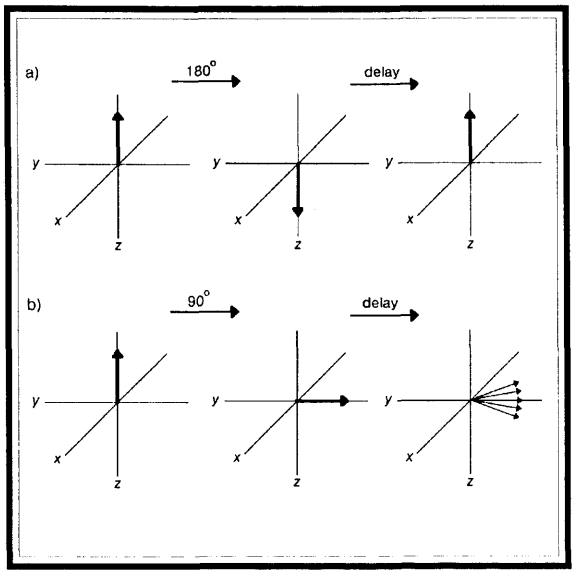

Figure 4.2 : Relaxation processes. a) T_1 relaxation where the number of spins in the α and β states return to their Boltzman distribution equilibrium values. This leads to a return of magnetisation to the $+z$ axis. b) T_2 relaxation taking place in xy plane magnetisation. The spins dephase and net magnetisation in the xy plane decreases. Figure adapted from reference [106].

Typically, dynamical information is derived from the longitudinal and transverse relaxation rates, R_1 (or $1/T_1$) and R_2 (or $1/T_2$) respectively, and also from the cross-relaxation ($^1H^N/^{15}N$) rate (denoted by NOE in the following). Such studies are aimed at characterising ; (1) the protein overall tumbling, (2) local fast motions, (3) restrictions of these local motions defined by a generalised 'order parameter' S, (4) possibly more complicated slow motions relevant to 'chemical exchange'. Moreover, the so-called chemical shift anisotropy (csa) at the ^{15}N nucleus is in principle accessible through measurements of relaxation parameters at several magnetic fields strengths, and should possibly provide further information. It is well known that multiple-fields data has been extensively used in the past for studying complex systems, such as proteins, and made possible the accurate descriptions of different types of motions involved. Relaxation parameters have been determined at three different magnetic field strengths for the CRT S-domain to consider critically the reliability of each parameters.

4.2.4.1 The model-free analysis

For analysis of the relaxation data we used the well recognised 'model-free' analysis of Lipari and Szabo [107]. The model-free analysis does not rely on motional models, therefore producing unbiassed results. In general, the fast internal motions can be described by two model independent quantities: a general order parameter S^2, which is the measure of the degree of spatial restriction of the motion, and an effective correlation time τ_c, which is a measure of the rate of the motion. Table IV.III describes the motional models used in this study. Model 1 is the simplified Lipari-Szabo model assuming only very fast internal motions (thermal vibrations). Model 2 is compensated for contributions from exchange processes (R_{ex}) that are too slow to be accounted for by the Lipari-Szabo model by an exchange term for T_2 correction. Model 3 and Model 4 are the complete Lipari-Szabo Model. Model 5 is the extended Lipari-Szabo model to include internal motion on an intermediate time scale.

Table IV.III : Spectral density function models used in the numerical fit to the experimental T_1, T_2 and NOE values. In the analysis the simplified version was used assuming that for residues with internal motions on an intermediate time scale the fast internal motions would be faster than a few picoseconds. S^2 is the generalised order parameter; S^2_s is the order parameter for slow internal motions; R_{ex} is the fluctuation of the chemical shift evaluated as an exchange term; τ_c is the rotational correlation time of the molecule; τ_e is the correlation time of internal motion. S^2_f is given by $S^2 = S^2_f S^2_s$. R_{ex} is given as an additional term in the transverse relaxation time. For consistency in the spectral density functions and the parameters, we used those described by Ishima, R. *et al.* [109].

Model	*Spectral density function*	*Parameters*
1	$J(\omega)=\frac{2}{5}\cdot\left[\frac{S^2\tau_c}{1+(\omega\tau_c)^2}\right]$	S^2
2	$J(\omega)=\frac{2}{5}\cdot\left[\frac{S^2\tau_c}{1+(\omega\tau_c)^2}\right]$ $R_{2,\exp}=R_2+R_{ex}$	S^2 , R_{ex}
3	$J(\omega)=\frac{2}{5}\cdot\left[\frac{S^2\tau_c}{1+(\omega\tau_c)^2}+\frac{(1-S^2)\tau_i}{1+(\omega\tau_i)^2}\right]$	S^2 , τ_e $\tau_i=\tau_c+\tau_e$
4	$J(\omega)=\frac{2}{5}\cdot\left[\frac{S^2\tau_c}{1+(\omega\tau_c)^2}+\frac{(1-S^2)\tau_i}{1+(\omega\tau_i)^2}\right]$ $R_{2,\exp}=R_2+R_{ex}$	S^2 , τ_e , R_{ex} $\tau_i=\tau_c+\tau_e$
5	$J(\omega)=\frac{2}{5}S_f^2\cdot\left[\frac{S_s^2\tau_c}{1+(\omega\tau_c)^2}+\frac{(1-S_s^2)\tau_i}{1+(\omega\tau_i)^2}\right]$	S^2 , S_s^2 , τ_e $\tau_i=\tau_c+\tau_e$

4.2.4.2 Rotational diffusion anisotropy

The majority of the relaxation studies concern approximately spherical globular proteins and isotropic overall rotational diffusion has been assumed. Rotational diffusion anisotropy has a profound effect on the physics of spin relaxation and on the interpretation of experimental studies of intramolecular dynamics. This is because the spectral density function depends on the relative orientations of the principal axis systems and the diffusion tensor. Consequently, an experimental knowledge of the rotational diffusion tensor is essential for a detailed analysis of intramolecular motions in nonspherical proteins.

4.3 MATERIAL AND METHODS

4.3.1 Protein expression and purification

4.3.1.1 Unlabelled protein

The pET-11a vector containing the human CRT S-domain (a.a. 164-283) and the pET-15b vector containing the CRT P-domain (a.a. 198-308), or PS-domain (a.a. 164-308) were expressed in *Escherichia coli* strain BL21(DE3) as described previously, see section 2.2.2.1. The recombinant pET-11a protein was purified and concentrated as described below. However, the pET-15b constructs were purified as described in section 2.2.2.2.

4.3.1.2 [$^{13}C/^{15}N$]-labelled CRT domains

To produce [$^{13}C/^{15}N$]-labelled CRT domains, the culture was grown in minimum growth medium (MME) supplemented with 2 g/l of ^{13}C-glucose and 1 g/l ^{15}N-NH_4Cl for isotope labelling (Cambridge isotope laboratories, inc., Andover, USA). A culture was inoculated with a glycerol stock of *E.coli* BL21(DE3) harbouring the appropriate pET construct into 50 ml MME media containing 100 µg/ml of Amp. The culture was incubated overnight at 37°C, shaking (250 x rpm). The desired amount of 2 L Erlenmeyer flasks containing 500 ml of MME-medium containing antibiotics were inoculated with 0.05 A_{600nm} of the overnight culture. Expression constructs were induced in late-log phase, A_{600nm} 0.5-1.0 (ideally 0.6), by addition of 1 mM IPTG for 4 hours at 37°C. The cells were harvested by centrifugation at 5000 x g (5500 x rpm, JA-10 rotor) for 5 minutes at 4°C. The supernatant was discarded and the cells were frozen at -20°C until needed for the purification step.

4.3.1.3 Purification of CRT S-domain from pET-11a

The pellet was resuspended in ice cold 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM EDTA, 0.1% Triton X-100, 10 u/ml of DNAse I, 1 protease inhibitor cocktail tablet (complete, mini, EDTA free, Roche) and sonicated on ice. The sonicate was then centrifuged at 10 000 x g for 15 minutes at 4°C and decanted in a new tube. The supernatant was filtered through a 0.2 µM acrodisc (Gelman Sciences) and applied on a 10 ml DE52 column (Whatman) and eluted with a NaCl gradient. The binding buffer was 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 2 mM EDTA and the elution buffer was containing 1 M NaCl. The fractions were analysed by SDS-PAGE and the fractions containing the protein of interest were pooled. The protein was then dialysed in 20 mM MES pH 5.5, 50 mM NaCl. The recombinant CRT was then purified on a Uno Q-6 column (Bio-Rad) using a gradient of NaCl.

The binding buffer was 20 mM MES pH 5.5, 50 mM NaCl, 2 mM EDTA and the elution buffer was containing 1 M NaCl. The fractions of interest from the previous purification step were then concentrated to approximately 3 ml (centriprep) and applied to a Superdex-75 (26/60) gel filtration column (Amersham Pharmacia Biotech) pre-equilibrated in 20 mM phosphate buffer, pH 6.5, 50 mM NaCl. After SDS-PAGE electrophoresis, the fractions containing pure protein were pooled and generally concentrated to 0.6 to 1 mM into 90% 20 mM phosphate buffer, pH 6.5, 50 mM NaCl, 0.01% (w/v) sodium azide in water and 10% D_2O. Protease inhibitor were also added to the sample to a final concentration of 5% (v/v) from a stock solution. The protease inhibitor stock solution was composed of one 'complete mini EDTA-free protease inhibitor cocktail tablet' (Roche Diagnostics GmbH, Mannheim, Germany) dissolved in 1ml of protein buffer (90% 20 mM phosphate buffer, pH 6.5, 50 mM NaCl in water and 10% D_2O).

4.3.2 NMR spectroscopy

Multiple protein preparations of calreticulin S-domain were made for NMR studies. The NMR spectra were acquired with uniformly [^{15}N] or [$^{13}C/^{15}N$]-labelled hCRT S-domain at a protein concentration of 0.6 to 1 mM in 90% H_2O/10% D_2O and at 0.66 mM protein concentration in 100% D_2O. D_2O was present in a proportion of 10% for field-frequency stabilisation purposes. The NMR measurements were performed on Bruker DRX500, DRX600 and Varian INOVA800 spectrometers. Water flip-back pulses in combination with WATERGATE method for water suppression were employed to minimise saturation of the water signal. The ^1H chemical shifts were referenced to the external DSS standard at 0.000 ppm. The ^{15}N reference was calculated using $^{15}N/^1H$ γ-ratios of 0.101329118.

Most of the experiments are summarised in table IV.IV including the total number of points acquired for each dimension. For the triple-resonance assignments the following experiments were recorded ; 2D [$^{15}N/^1H$]-HSQC, 3D HNCO, 3D HN(CO)CA, 3D HNCA, 3D HCBCA(CO)NH, and 3D HCBCANH. The ^{15}N-resolved experiments were also used to facilitate the backbone and side-chain assignment ; 2D [$^{15}N/^1H$]-HSQC, 3D [$^1H,^{15}N$]-NOESY-HSQC, 3D [$^1H,^{15}N$]-TOCSY-HSQC, and 3D-HB(CO)NH. For the structure determination and side-chain assignment multiple ^{13}C-resolved experiments were used; 2D [$^1H,^1H$]-HSQC, 3D [$^1H,^1H$]-NOESY-HSQC, 3D [$^1H,^1H$]-HMQC-NOESY in D_2O, 3D-(H)CCH-COSY, and 3D-H(C)CH-COSY. Data were processed using the NMRPipe [110] package and analysed using NMRView. The data were typically zero-filled twice in each

dimension, generally to 1024 x 256 x 256 points, and linear prediction of up to 50% was used.

Table IV.IV : NMR experiments performed on the CRT S-domain. The total number of points acquired for each dimension (F3, F2, and F1).

Experiment	*F3*	*F2*	*F1*
2D [^{1}H,^{15}N]-HSQC	N/A	2048	512
3D HN(CO)CA	1024	108	100
3D HNCA	1024	108	100
3D HCBCA(CO)NH	2048	48	130
3D HCBCANH	1024	108	118
3D HNCO	1024	108	70
3D HB(CO)NH	1024	108	158
3D [^{1}H,^{15}N]-NOESY-HSQC	3072	64	300
3D [^{1}H,^{15}N]-TOCSY-HSQC	3072	72	300
2D [^{1}H,^{13}C]-HSQC	N/A	3300	256
2D [^{1}H,^{13}C]-HSQC (full sweep)	N/A	3300	400
3D [^{1}H,^{13}C]-HMQC-NOESY (D_2O)	1032	128	320
3D [^{1}H,^{13}C]-NOESY-HSQC	1032	128	320
3D [^{1}H,^{13}C]-NOESY-HSQC (aromatics)	1024	100	400
3D (H)CCH-COSY	2048	164	164
3D H(C)CH-COSY	2048	132	200

4.3.2.1 Calcium titration

For the titration of calcium, samples of CRT S- and P-domain were used at 0.3 mM protein concentration in 10 mM Bis-Tris, pH 6.5, 50 mM NaCl. Amounts of $CaCl_2$ were added by small increments to ; 0.1, 0.2, 0.4, 0.8 and 1.0 mM followed by 1D and 2D NOESY spectra. The calcium was then chelated by addition of EDTA to 0.5, 1 and 2 mM before experiments were recorded as previously. The same type of experiment was performed by adding $CaCl_2$ in small increments to a sample containing 1 mM EDTA.

4.3.2.2 Triple-resonance experiments

The backbone assignment of the human CRT S-domain has been performed using the simulated annealing assignment described in the section 4.2.1.1 (developed by I. Barsukov and K.H. Sze) [108]. The input consisted of the amino acid sequence and a table of arbitrarily numbered spin systems. This spin system table was obtained by picking the peaks in the 5 triple resonance experiments

acquired using NMRView [111], followed by deletion of all duplicated peaks present in other spectra, i.e. the HN(CO)CA peaks from HNCA, using a python script (developed by Dr. I. Barsukov). Those 5 peak tables were correlated into a single spin table, checked and classified as different spin systems; ambiguous, unique, weak,etc. From this refined spin table, the simulated annealing could finally be performed.

4.3.3 Structure calculation

For the human Calreticulin S-domain, combined automated NOE cross-peak assignment and 3D protein structure calculation were performed using the program ARIA [105] version 2.0a (**A**mbiguous **R**estraints for **I**terative **A**ssignment). Briefly, ARIA consists of a series of routines to assign and calibrate NOE data, analyse distance violations and merge the data into a table of distance restraints read by CNS. NOESY cross peaks were assigned by ARIA using peak lists generated from raw NOESY spectra together with chemical shifts lists for the target protein. Standard ARIA protocols were used during the calculations [105]. Violations and unassigned peaks were checked manually after every cycle and cycles were repeated. Four peak lists were generated by interactive peak picking of the NOESY spectra with the program NMRView version 5 [111]. Additionally, The prediction of ϕ (phi) and ψ (psi) backbone torsion angles has been performed using TALOS [112] (**T**orsion **A**ngle **L**ikelihood **O**btained from **S**hift and sequence similarity). The computer program TALOS was developed to search a database for strings of residues with chemical shift and residue type homology. TALOS yields 10 triplets which have the closest similarity in secondary chemical shift and amino acid sequence to those of the query sequence. If the central residues in these 10 triplets exhibit similar ϕ and ψ backbone angles, their averages can reliably be used as angular restraints for the protein whose structure is being studied.

4.3.4 Relaxation

For the relaxation experiments, uniformly [^{15}N]-labelled samples of hCRT S-domain were used at 0.7 mM protein concentration. For multiple field analysis, the NOE, T1 and T2 experiments were performed at 500, 600 and 800 MHz at 25°C (298 K). T_1 data sets were recorded in such a way that the signal intensity decays exponentially to zero as a function of the relaxation delay, enabling a simple two parameter fit. T_1 experiments were performed with 9 relaxation delays (see table

IV.V) during which a train of 180° ^{1}H pulses (separated by delay d15, see table IV.V) is applied to remove dipolar (^{1}H-^{15}N)-csa(^{15}N) cross-correlation effects [113]. T_2 experiments were performed employing a CPMG pulse train with an interval of 0.6 milliseconds (ms) between two consecutive ^{15}N 180° pulses; ^{1}H decoupling 180° pulses are applied between two ^{15}N pulses every 2.4 ms [114]. Nine experiments were run with the duration of the CPMG train described in table IV.V. All T_1 and T_2 measurements were performed with an additional control point to asses reproducibility and long term stability of the sample. The heteronuclear NOE was measured from a pair of spectra recorded with and without proton saturation as reported previously by Barbato *et al.*, 1992 [115]. Proton saturation was achieved by 120° pulses applied for the duration of the NOE relaxation delay described in table IV.VI. In the experiment without presaturation a carefully optimised water flip-back pulse was added before the ^{1}H 90° pulse in order to avoid undesirable effects related to exchange between water and amide protons [116].

Table IV.V: Delays used for the relaxation experiments in milliseconds. Table showing delays used at the different fields for the T_1 and T_2 experiments.

	T_1 500MHz	T_2 500MHz	T_1 600MHz	T_2 600MHz	T_1 800MHz	T_2 800MHz
1	20.1136	4.8	20.05	4.8	16.653	8.528
2	100.5680	14.4	100.25	14.1	96.919	17.056
3	201.1360	28.8	200.50	28.8	193.238	25.584
4	301.7040	43.2	300.75	43.2	297.584	34.112
5	502.8400	62.4	501.25	62.4	498.249	51.168
6	703.9760	96.0	701.75	96.0	698.914	76.752
7	1005.6800	120.0	1002.50	120.0	995.898	93.808
8	1307.3800	144.0	1303.25	144.0	1292.883	119.392
9	1508.5200	216.0	1503.75	216.0	1493.548	144.976
10	301.7040	28.8	300.75	28.8	297.584	25.584

Table IV.VI : Relaxation parameters used for the CRT S-domain experiments. Relaxation experiments were performed at 500, 600 and 800 MHz with the following parameters.

Parameters	*500 MHz*	*600 MHz*	*800 MHz*
NOE relaxation delay	3.1 s	3 s	4.2 s
T1 d15	2.5 ms	2.5 ms	2 ms
T1 180° pulse	14.2 µs	12.5 µs	13.4 µs
T2 d15	0.6 ms	0.6 ms	0.5 ms

NMR spectra were processed with NMRPipe and analysed with NMRView. The peak intensities were determined from peak heights using the NMRView peak analysis. The Jitter mode used determines the height of the most intense point found within a certain range (+/- 25% of the peak bounds) of the peak centre. This allows a certain level of variation of the peak position from one spectrum to the next.

The NOE values were calculated by taking the ratio of the ^{15}N intensities from experiments performed with and without proton presaturation. Relaxation rates were determined by fitting the delay dependent peak intensities to an exponential function in MATHEMATICA (Wolfram Research Europe Ltd., Long Hanborough, UK). The uncertainties due to random errors in the measured heights were deduced from 100 Monte Carlo simulations. The values of noise level were evaluated in peak-free regions and used to estimate the standard deviation of the peak intensities. All MATHEMATICA scripts used were developed by Dr. M. Pfuhl.

An estimation of the overall rotational correlation time (τ_c) was obtained from R_2/R_1 ratios using MATHEMATICA. A filter was applied to ensure that only residues that do not suffer from large amplitude fast motions or chemical exchange are included in the τ_c determination. For this purpose a filter [standard deviation cut-off, NOE cut-off] of [1.5, 0.4] was used for all the data. The data were simulated using the simplest spectral density function (table IV.III, equation 1) and the analysis completed by a Monte-Carlo analysis to produce a Gaussian distribution for the estimation of the uncertainty in the determined τ_c.

The model-free analysis of local relaxation data according to the method of Lipari-Szabo [107] in its extended form [117] was performed using MATHEMATICA. The five different motional models detailed in table IV.III were used in the numerical fit to the experimental R_1, R_2 and NOE values. The equation describing the best the experimental data was used to calculate the appropriate variables necessary to describe the motion; i.e. S^2, R_{ex}, t_i, S^2_f.

For the determination of the protein anisotropy, the parallel ($D_\parallel$) and perpendicular ($D_\perp$) diffusion tensor were calculated using HYDRONMR (version 5a) [118]. HYDRONMR, whose main input was the atomic co-ordinates contained in a protein data bank (PDB) file, builds an appropriate hydrodynamic model of the protein and computes the fully anisotropic rotational diffusion tensor. Pdbinertia [119] was then used to reorient the asymmetric CRT S-domain hairpin along the long axis, or z-axis. The angles (α) between the N-H vector and the long axis of the rotational diffusion tensor were finally calculated using the new PDB file produced.

The equations 6 to 9 described below were used to estimate the expected contribution into the relaxation parameters from the molecule anisotropy. Equations from Baber, J.L. *et al.* 2001 [120].

Equation 6:

$$J(w)=2/5\sum_{k=1}^{3}A_k$$
$$\left\{S_f^2S_s^2[\tau_k/(1+(\omega\tau_k)^2)]+S_f^2(1-S_s^2)[\tau_{s,k}/(1+(\omega\tau_{s,k})^2)]+(1-S_f^2)[\tau_{f,k}/(1+(\omega\tau_{f,k})^2)]\right\}$$

With:

Equation 7: $A_1=0.75\sin^4\alpha$, $A_2=3(\sin^2\alpha)(\cos^2\alpha)$, *and.* $A_3=(1.5\cos^2\alpha-0.5)^2$

Equation 8: $\tau_1=(4D\|+2D\perp)^{-1}$, $\tau_2=(D\|+5D\perp)^{-1}$, *and.* $\tau_3=(6D\perp)^{-1}$

Equation 9: $1/\tau_{i,k}=1/\tau_i+1/\tau_k$ *where* $i=s$ *or.* f *and.* $k=1,2,3$

4.4 RESULTS

4.4.1 Sample optimisation

The sample conditions were optimised using rapidly obtainable 1D and 2D NOESY spectra on a number of CRT S-domain samples. By analysing the spectra for peak dispersion and intensity, the conditions were optimised. Phosphate buffer was used, as it is the most common buffer used for NMR studies and it has the advantage of not giving any signals, as it does not contain any slowly-exchanging protons. The pH did not affect the quality of the spectra, so pH 6.5 was used as it is the pH where the NH group exchange is the slowest. Various protein, DTT and NaCl concentrations were tried (see table IV.VII). As we increased the concentration of the protein up to 1.23 mM there was no degradation of the quality of the spectra as the protein could have started to aggregate. However, the NaCl concentration had a mild effect on the peak dispersion in the spectra and a concentration of 50 mM was judged to be an acceptable choice. It was critical to asses those two, as the protein could have aggregated or precipitated at low salt or high protein concentrations. The presence of DTT in the sample did not improve the quality of the spectra, even if the protein has only one cysteine and non-physiological dimers could have formed. This analysis revealed that 20 mM phosphate buffer, pH 6.5, 50 mM NaCl was the optimal condition for performing future experiments.

Table IV.VII : Buffer conditions tested for the sample optimisation. The buffer used was 20 mM Phosphate and the pH was equilibrated by mixing different ratios of Na_2HPO_4/NaH_2PO_4.

pH	*[NaCl] mM*	*[DTT] mM*	*[Protein] mM*
8	50	2	0.2
6.5	50	2	0.2, 0.4, 1
6.5	100	5	0.2, 1
6.5	25	5	1.5
6.5	8	5	1.23
6.5	50	0	0.2, 0.4, 1, 1.23

Comparison of 2D [$^{15}N/^{1}H$]-HSQC spectra collected over a period of one week showed sample deterioration. After approximately one week, extra peaks would appear with a tendency to move towards the middle of the spectrum, indicating deterioration of the sample quality. The protein sample was then sequenced to notice that the first fifteen residues were cut after the Arg 177. The CRT S-domain was then re-cloned into the pET-11a vector, which does not have any N-terminal His-Tag. This new clone could be purified without thrombin cleavage of

the His-Tag, also permitted the use of protease inhibitors all along the purification procedure and the final yield of pure protein obtained was similar to the His-Tagged protein. Moreover, the presence of non-physiological CRT S-domain dimers in the absence of reducing agent was observed in chapter 3. To prevent protein dimerisation, the only cysteine in the CRT S-domain was removed during the cloning process by removal of the two first amino acids at the N-terminus of the protein (RCKDDE). This new protein sequence did not contain any cysteine residues, preventing dimerisation and also making the use of DTT unecessary. This new clone had almost identical spectra to the previous pET-15b clone, and more importantly, it was less susceptible to degradation and more stable. This new clone allowed more spectra to be run on a single sample, facilitating the acquisition of multiple spectra for the structure determination.

Due to the relatively poor stability of the protein over long period of time, the experiments were performed at 25ºC (298K), in the presence of protease inhibitors. However, it was noticed by running 2D [^{1}H,^{15}N]-HSQC at different temperatures, 15 and 25 ºC, that the protein was extremely temperature sensitive with a tendency for the well resolved peaks to move towards the centre of the spectra at higher temperature. This effect was probably due to an increase in the rigidity of the molecule at low temperature. However, at low temperature, the signals were broader and therefore weaker, making the assignment of the resonances harder. As a compromise for having a more dynamic molecule in solution at higher temperature, we collected data at 25 ºC for an increase in sensitivity.

The calreticulin S-domain 2D [^{1}H,^{15}N]-HSQC spectra show many sharp peaks, indicative of highly flexible regions and more dispersed and broad peaks indicating the presence of some secondary structure (figure 4.4). Moreover, some overlapping peaks can be noticed in centre of the spectra but the spectral quality is good enough for assignment of the signals. The CRT P-domain was cloned to assess whether the sharp peaks present in the S-domain spectra are the product of a wrong choice of boundaries or if this molecule has highly flexible and dynamic regions. The CRT P-domain spectra has the best dispersion of signal, with a smaller fraction of sharp cross-peaks, suggesting a more folded molecule (figure 4.5). This better cross-peak dispersion could be explained by the presence of all the three repeats B shown in figure 4.3, which have been highly conserved through evolution and/or due to the removal of the residues at the N-terminus. To further test this hypothesis, the PS-domain was cloned, which included the boundaries of both the S- and P-domains (figure 4.3). The PS-domain keep most of the spectral

characteristics of the P-domain, with extra sharp cross-peaks probably from the N-terminus of the domain, a.a. 164-198 (figure 4.3 and 4.6). The results indicated that the P-domain is more structured. At the same time, the S-domain showed enhanced complement inhibition (see chapter 2), making the structural information on this domain more biologically relevant.

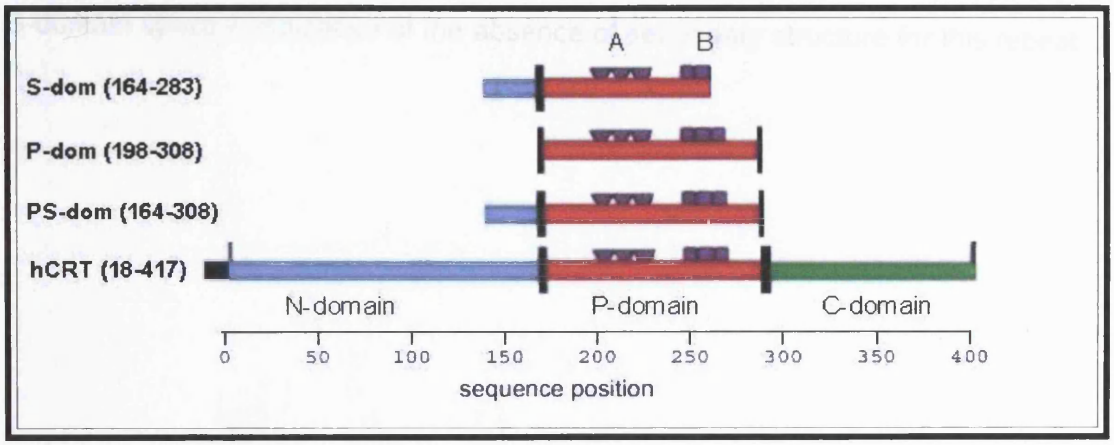

Figure 4.3 : Alignment of the three calreticulin clones subjected to NMR (S-, P-, and PS-domains) on the complete mature chain (amino acids 18-417). The figure shows a schematic representation of the human calreticulin N-, P- and C-domains of the protein. The protein contains an N-terminal signal sequence, shown in black and a C-terminal KDEL endoplasmic reticulum retrieval signal. Form left to right, repeats A1 to 3 (amino acid sequence PXXIXDPDAXKPEDWDE) and B1 to 3 (amino acid sequence GXWXPPXIXNPXYX) are indicated by the purple triangles and squares respectively.

The P-domain was studied by Ellgaard *et al.* who published the resonance assignment [121] and later the NMR structure [59]. The three-dimensional structure showed an unusual fold containing a single extended hairpin formed by the entire polypeptide chain. To understand more in detail the enhanced activity of the S-domain of CRT in inhibiting complement, the structure of the S-domain was important. The structural information of the two domains, S- and P-, would allow us to understand more in detail the interaction between the complement proteins, C1q and MBL with CRT.

It can be seen from spectra comparison in figure 4.5 and 4.6 that the complete lack of the repeat B3, figure 4.3, has a strong impact on the CRT S-domain spectra. First, it can be noticed in the $\Delta\delta$ plot, figure 4.7, that the chemical shift of the residues located into repeats A3 to B1 (red bars) have almost the same chemical shift in the P- and S-domains. However, as indicated by blue lines in figure 4.5, the

cross-peaks for Ile225, Trp275, Lys276, and Ile280 move towards the middle of the S-domain spectra, indicative of the loss of secondary structure. This is also summarised in figure 4.7, where we can notice large chemical shift changes in the residues outside repeats A3 to B1 (blue bars). The Ile225 is located in repeat A2, figure 4.3, and the Trp275, Lys276, and Ile280 in repeat B2. Finally, all the residues located in repeat A1 of the CRT S-domain were concentrated into the middle of the S-domain spectra, indicative of the absence of secondary structure for this repeat.

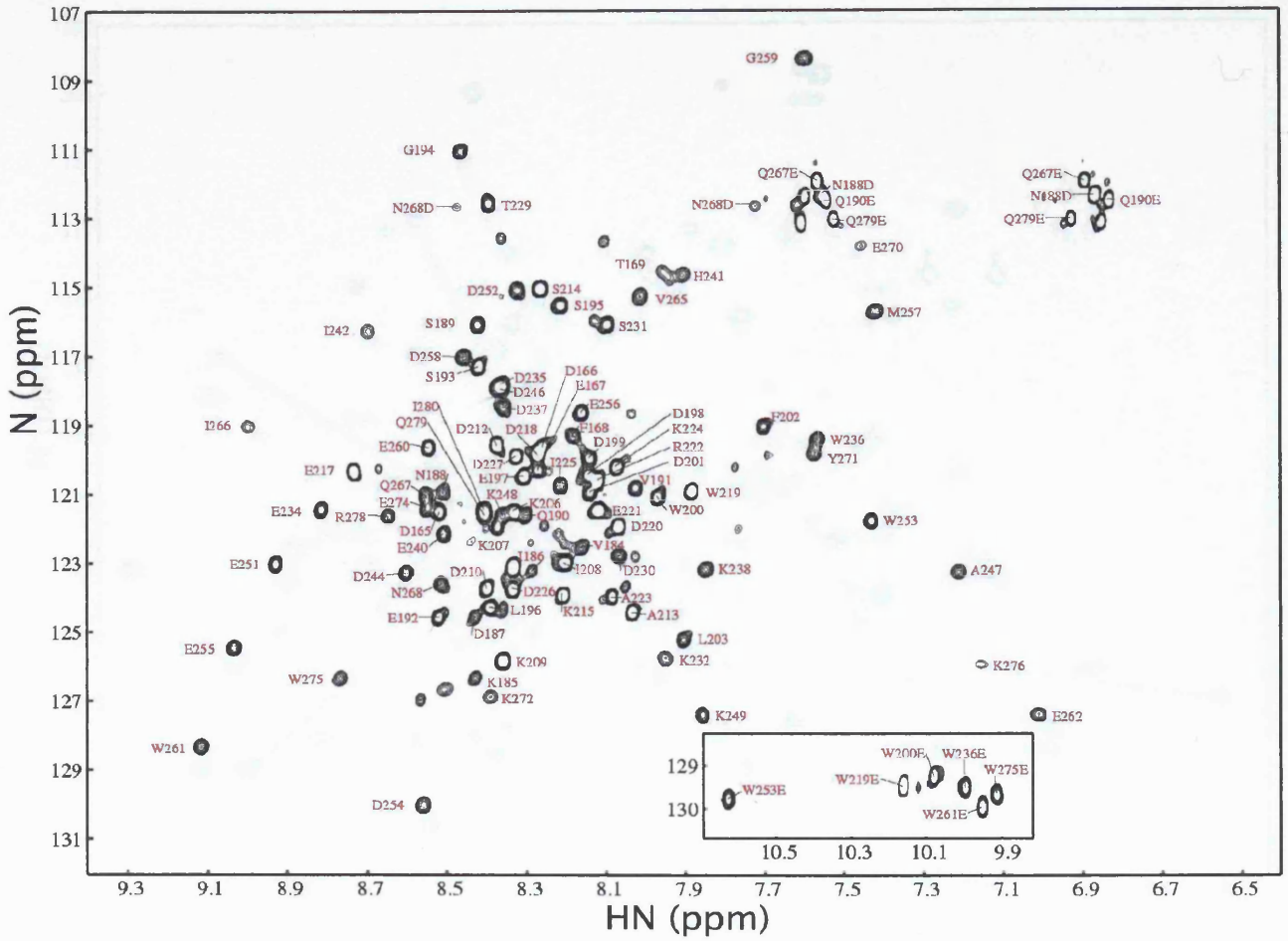

Figure 4.4 : 2D [^{15}N/^{1}H]-HSQC spectrum of hCRT S-domain (164-283) with backbone ^{15}N-^{1}H resonance assignments. The insert in the lower right contains the indole $^{15}N^{\varepsilon}$-^{1}H cross peaks of the six tryptophans. The sequence numbering is that of human calreticulin, SwissProt accession number P27797, with an extra Met163 from the cloning.

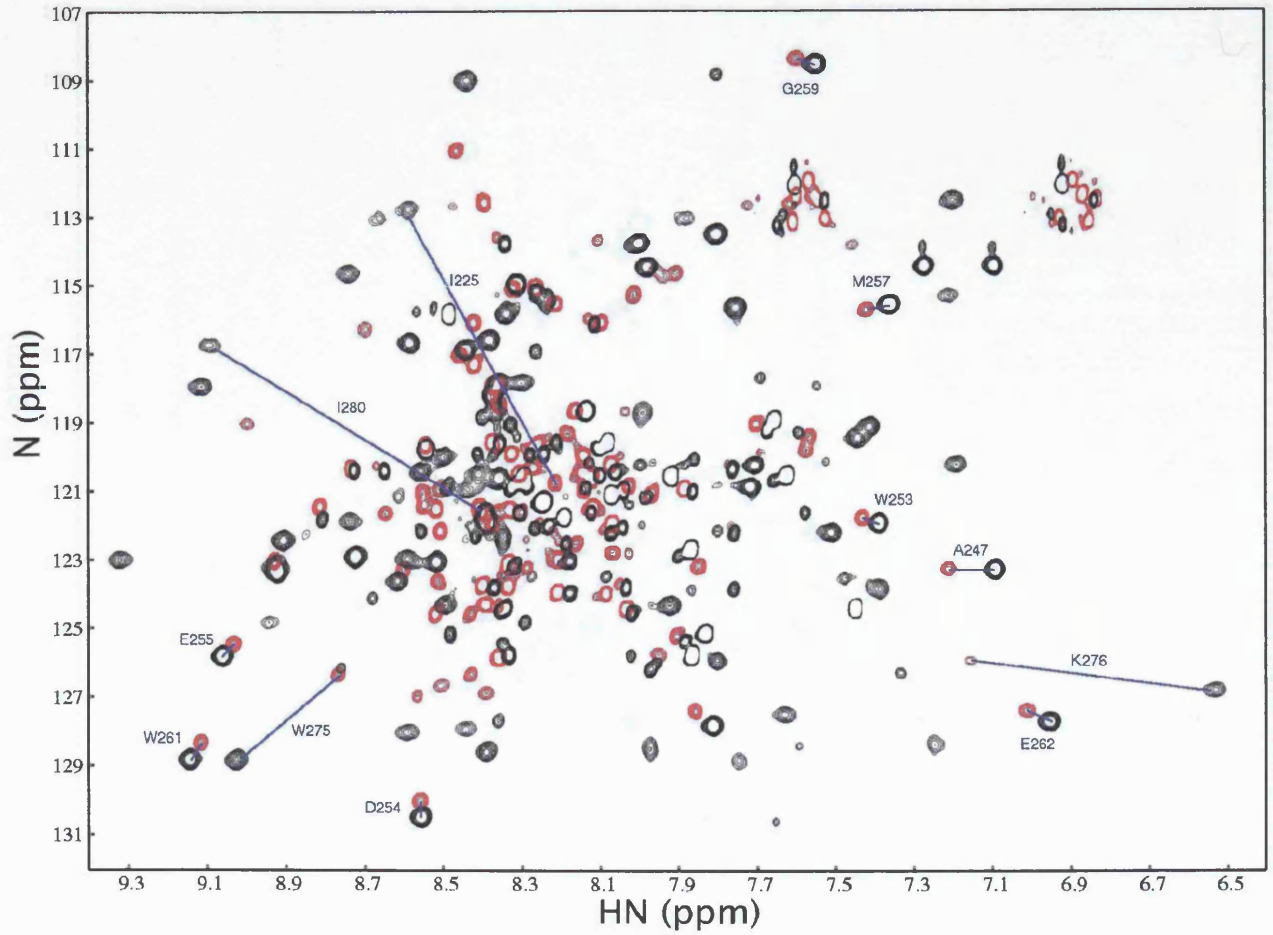

Figure 4.5 : Overlay of 2D [^{15}N/^{1}H]-HSQC spectrum of hCRT P-domain (198-308) in black and S-domain (164-283) in red. Backbone ^{15}N-^{1}H resonance assignments are displayed in blue for a few amino acids to highlight important changes between the two clones. The rat CRT P-domain assignment has been published by Ellgaard, L. *et al.* (2001).

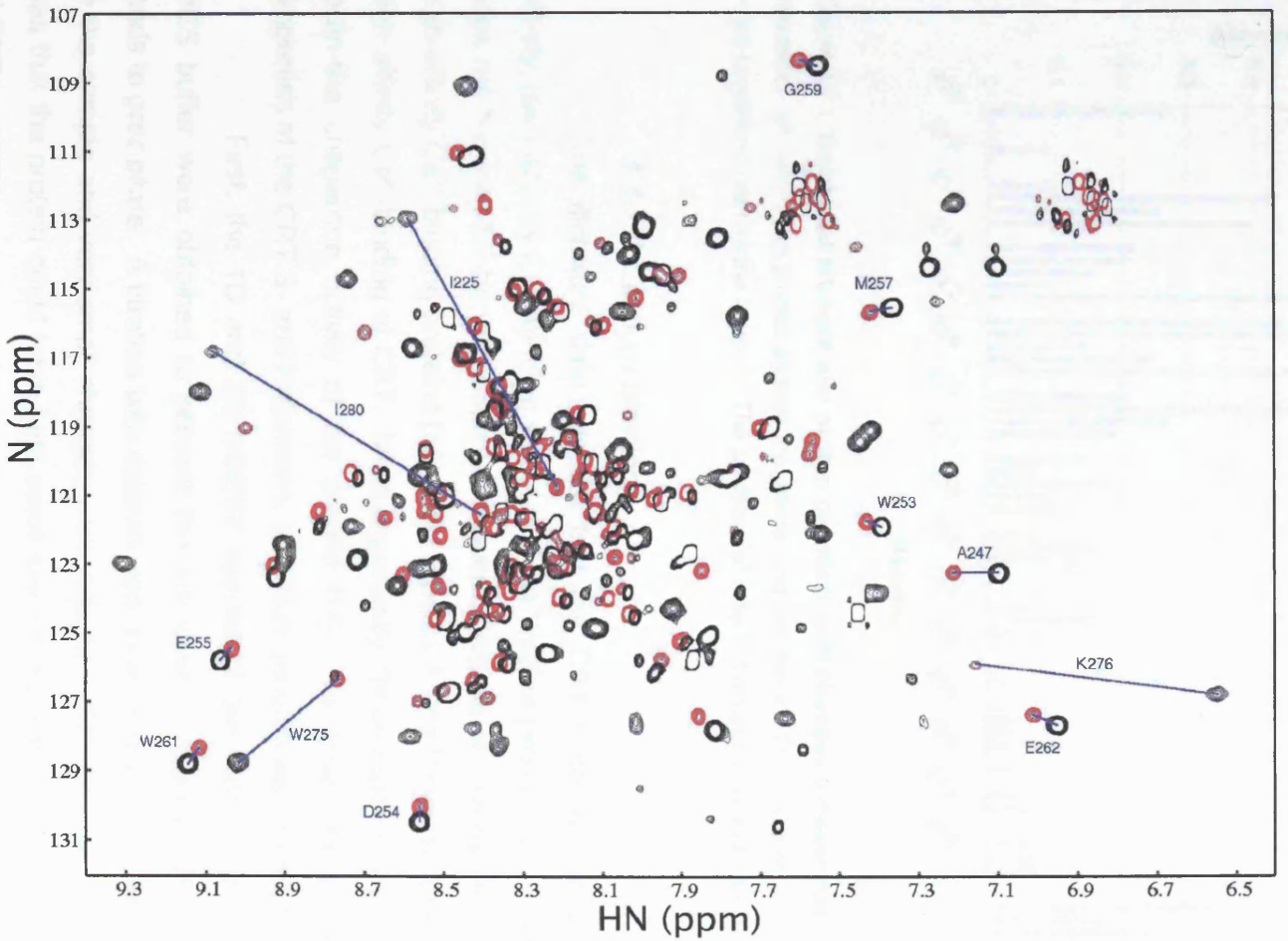

Figure 4.6 : Overlay of 2D [^{15}N/^{1}H]-HSQC spectrum of hCRT PS-domain (164-308) in black and S-domain (164-283) in red. Backbone ^{15}N-^{1}H resonance assignments are displayed in blue for a few amino acids to highlight important changes between the two clones.

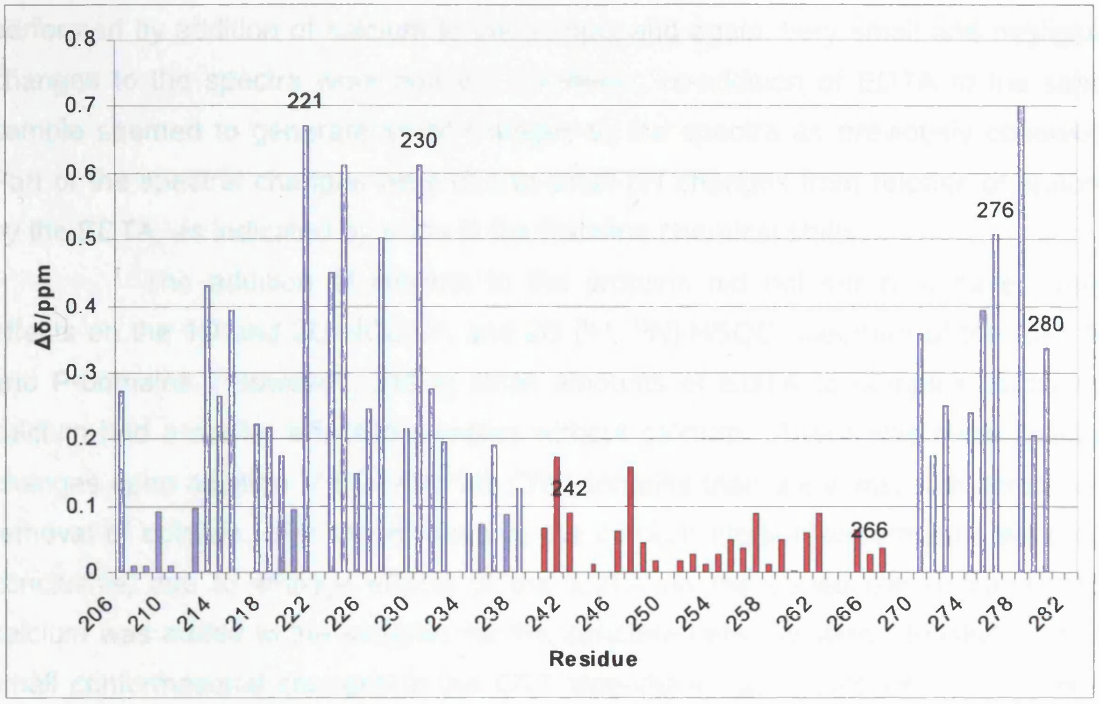

Figure 4.7 : Combined nitrogen and proton chemical shift changes between the CRT S- and P-domains. Δδ values are plotted against the amino acid sequence. The residues 239-269 are shown in red (described later as the hairpin). The positions of selected residues are indicated.

4.4.1.1 Calcium titration

The literature data suggest that the CRT P-domain contains a high-affinity, low capacity Ca^{2+}-binding site (1 mol Ca^{2+}/mol of protein, K_d = ~1μM) [13], but does not have an EF-hand consensus sequence typical of several other well known high-affinity Ca^{2+}-binding proteins [59]. The repeats A and B may be essential for the high-affinity Ca^{2+}-binding of CRT. More importantly, those repeats are critical for the lectin-like chaperone activity of the protein [14]. To asses the calcium binding properties of the CRT S- and P-domains, a calcium titration was performed.

First, the 1D and 2D NOESY spectra of the freshly purified protein in MES buffer were obtained to prevent the formation of calcium phosphate which tends to precipitate. A titration with calcium was done by gradual addition of calcium to the sample and very small changes were noticed in 1D spectra. The hypothesis was that the protein could have kept some calcium from the purification in absence of EDTA. EDTA was then added to the sample to chelate any calcium present into the sample and small changes could be noticed in the 1D spectra. To make sure that no traces of calcium are present in the sample, the protein was purified in the presence of EDTA and a reference spectrum was taken. A new titration was

performed by addition of calcium to the sample and again, very small and negligible changes to the spectra were noticed. However, re-addition of EDTA to the same sample seemed to generate small changes to the spectra as previously observed. Part of the spectral changes were due to small pH changes from release of protons by the EDTA, as indicated by shifts in the histidine chemical shifts.

The addition of calcium to the proteins did not seem to have strong effects on the 1D and 2D NOESY, and 2D [^{1}H, ^{15}N]-HSQC spectrum of the CRT S- and P-domains. However, adding small amounts of EDTA to samples containing calcium had a similar effect to samples without calcium. There was more spectral changes upon addition of EDTA to the CRT domains than there was with addition or removal of calcium. For those reasons, the calcium titration experiments were not conclusive, due to strange effects of the EDTA on the calreticulin domains. No calcium was added to the samples for the structure determination. Finally, to study small conformational changes in the CRT side-chains upon calcium binding, more suitable 3D ^{13}C-resolved NOESY-HSQC experiments could be used.

4.4.2 Backbone resonance assignment of the CRT S-domain

The peaks in the triple-resonance spectra were picked using NMRVIEW and the statistics of peak picking are presented in table IV.VIII. The human CRT S-domain contains 121 amino acids, including the extra Met163 from the cloning into pET-11a, which is the first amino acid of the chain. Excluding the 16 prolines and the first N-terminal residue, 104 backbone cross-peaks should be observed in the 2D [^{1}H,^{15}N]-HSQC of the recombinant protein. However, the CRT S-domain has three Arg side-chain containing an NH group each, and four Asn and three Gln, which produce a doublet due to the presence of an NH_2 group on their side-chain bringing the total of observable cross-peaks to 121. The observed number of cross-peaks in the spectra described in table IV.VIII agrees with the number of predictable cross-peaks.

Table IV.VIII : Number of peaks picked in each spectrum used for triple-resonance experiments. The peaks were picked using NMRVIEW.

Spectra	*No. peaks*
HN(CO)CA	121
HNCA	216
HCBCA(CO)NH	217
HCBCANH	364
HNCO	124

In the 2D [$^{15}N/^{1}H$]-HSQC spectrum (figure 4.4) some additional weak cross-peaks are observed, mostly located next to strong signals from flexible residues near the N-terminus. These may reflect the presence of minor populations of one or several conformation isoforms of CRT S-domain, which may arise from *cis/trans* isomerisation of Xxx-Pro peptide bonds and/or from degradation of the sample in the N-terminus during the experiments.

The next three figures describe how the assignment of residues 163 to 283 of the CRT S-domain backbone has been achieved (figures 4.8, 4.9, and 4.10). Some distortion in the shape of the peaks in the ^{15}N dimension was noticed in some of the spectra (HNCO, HNCA, and HNCACO), which was due to a bug in the XWINNMR software at the time. This unfortunate problem did not affect the assignment.

First, by the use of the HN(CO)CA spectra (figure 4.8), the chemical shift of the H^N and N of residue *i* and the C^α of residue *i-1* could be determined (also see table IV.I and figure 4.1). The HNCA spectra gives the same information as the HN (CO)CA with the C^α chemical shift of the residue *i* in extra. Second, the HCBCANH and HCBCA(CO)NH spectra (figure 4.9) show C^β chemical shifts for the residue *i*, *i-1* and *i-1*, respectively. Finally, the HNCO spectrum (figure 4.10) gives the CO chemical shift of the reside *i-1*.

All this chemical shift information was compiled into a table and the sequential assignment of these di-peptides achieved by a simulated annealing algorithm. A more detailed protocol on how this step was performed is described in the introduction, see chapter 4.2.1.1. The output of the simulated annealing was checked manually for inter- and intra-residual connectivities, by displaying strips as in figures 4.8, 4.9 and 4.10. The proline CO, C^α and C^β could also be assigned using those experiments.

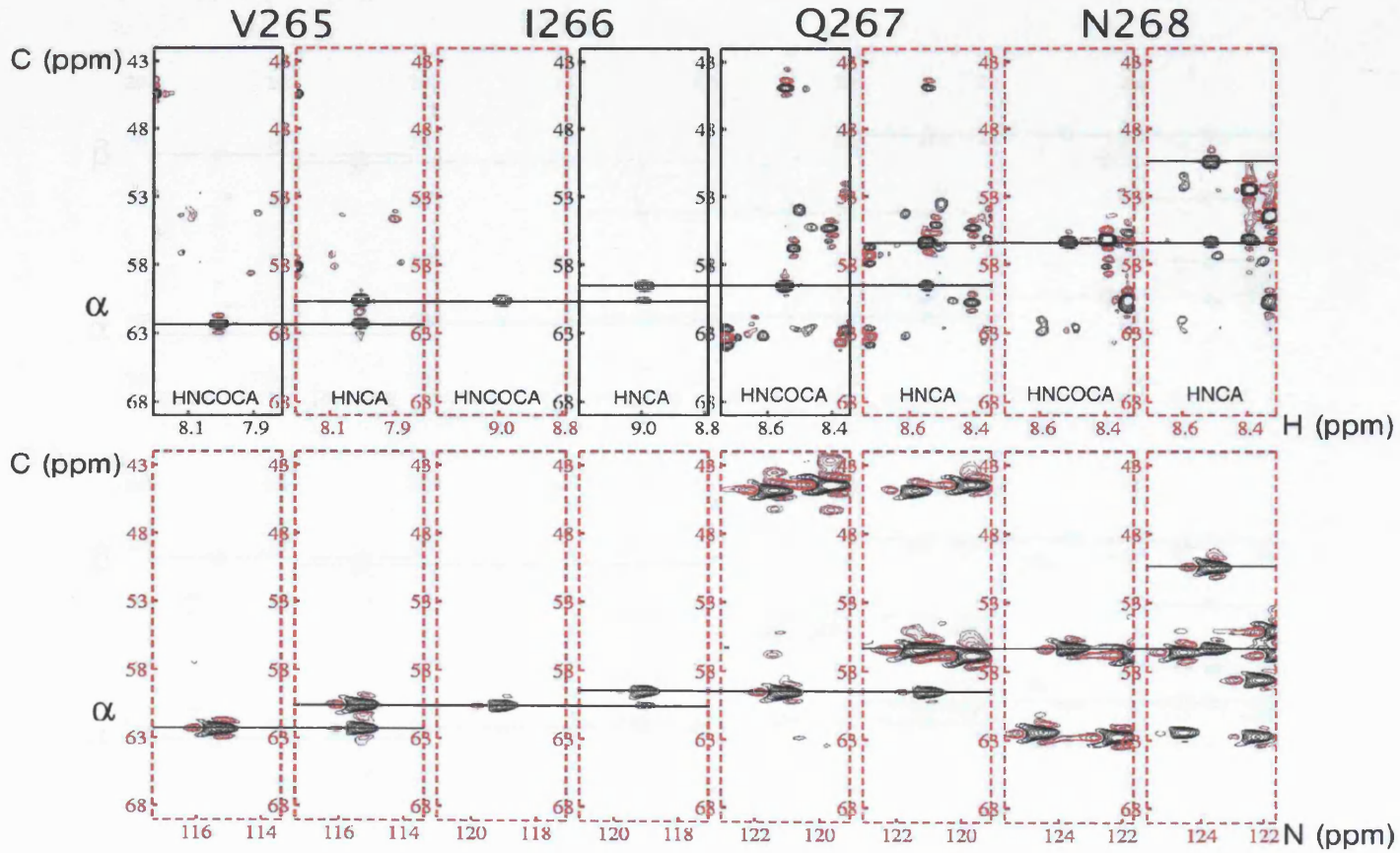

Figure 4.8 : Illustration of (C^{α}, N, and H^{N})-assignment for the backbone resonance of residues V265, I266, Q267 and N268 (indicated at the top of the spectra). For each residue, strips from the [$^{13}C/^{1}H$]-planes (top row) and [$^{13}C/^{15}N$]-planes (bottom row) of 3D-HNCA and -HN(CO)CA spectra are shown. Horizontal lines trace the sequential pathway of inter- and intra-residual connectivities.

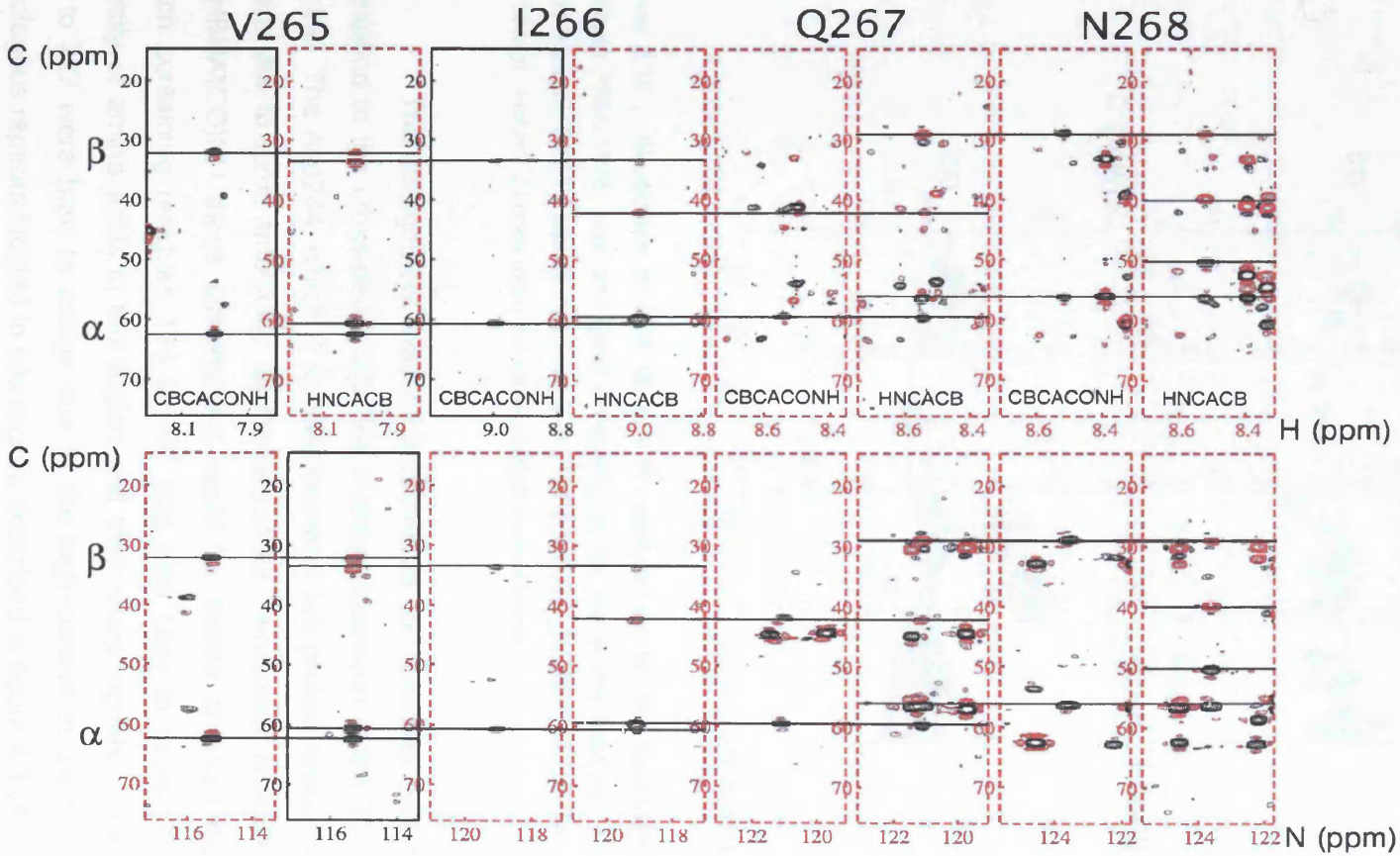

Figure 4.9 : Illustration of (C^{α},C^{β}, N, and H^{N})-assignment for the backbone resonance of residues V265, I266, Q267 and N268 (indicated at the top of the spectra). For each residue, strips from the [$^{13}C/^{1}H$]-planes (top row) and [$^{13}C/^{15}N$]-planes (bottom row) of 3D-HCBCANH and -HCBCA(CO)NH spectra are shown. Horizontal lines trace the sequential pathway of inter- and intra-residual connectivities.

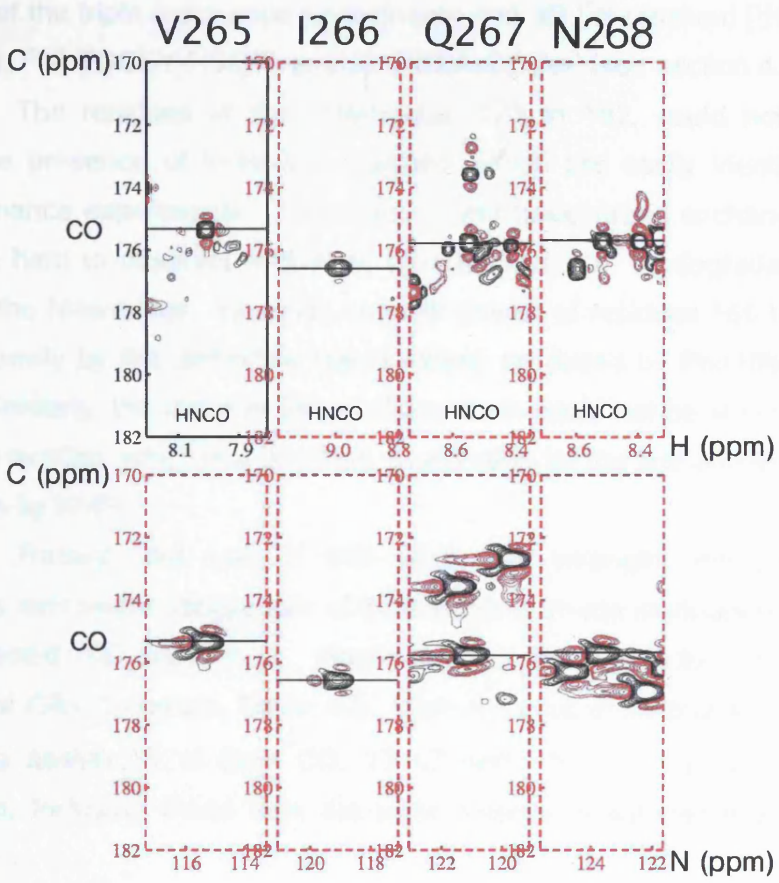

Figure 4.10 : Illustration of (CO, N, and H^N)-assignment for the backbone resonance of residues P264, V265, I266 and Q267 (indicated at the top of the spectra). For each residue, strips from the [$^{13}C/^{1}H$]-planes (top row) and [$^{13}C/^{15}N$]-planes (bottom row) of 3D-HNCO spectrum are shown. Horizontal lines trace the inter-residual connectivities.

The assignment was straightforward for residues 229 to 276, which correspond to the cross-peaks with the highest dispersion in the 2D [$^{1}H,^{15}N$]-HSQC spectra. The Asp244, which is located between two proline residues, was left until the end due to some ambiguity. By the use of 3D ^{15}N-resolved NOESY, TOCSY, and 3D HBHA(CO)NH those assignments could be double checked rapidly. A long stretch containing residues 184 to 197 was also easy to assign due to the high diversity of amino acids in this region and their sharp signals. However, residues 198 to 227 were hard to assign due to the high content in aspartic acid and two homologous repeats located in this region, described in figure 4.3 (A1 and A2). Also, this region has sharp peaks, which tend to indicate that those residues are dynamic and for the same reason, most chemical shifts for those aspartic acids were almost

identical. Multiple assignment possibilities had to be carefully checked by displaying the strips of the triple resonance experiments and 3D ^{15}N-resolved [^{1}H, ^{15}N]-NOESY-HSQC, [^{1}H, ^{15}N]-TOCSY-HSQC, and 3D HBHA(CO)NH (see section 4.3.3.1).

The residues at the N-terminus, 170 to 182, could not be assigned, despite the presence of threonine residues, which are easily identifiable with the triple-resonance experiments. This region might have strong exchange, making the resonance hard to observe, and/or as described before, by degradation of the the protein at the N-terminus. However, a small stretch of residues 165 to 169 could be assigned easily by the distinctive spin systems produced by Phe168 followed by a Thr169. Similarly, the last 4 residues of the chain could not be assigned due to the resonance overlap, which is a common observation for the first and last few residues of a protein by NMR.

Through the use of 600 MHz field strength and [^{13}C/^{15}N]-labelled samples, a successful assignment of 84 backbone amide resonance was made out of an expected 105, figure 4.11. These are also detailed in the 2D [^{1}H,^{15}N]-HSQC spectrum of CRT S-domain, figure 4.4. Moreover, out of 16 proline residues, it was possible to assign 11 of their CO, 13 C^{α} and 13 C^{β}. The full chemical shift assignment, including those from the triple resonance experiments, is detailed in table IV.IX.

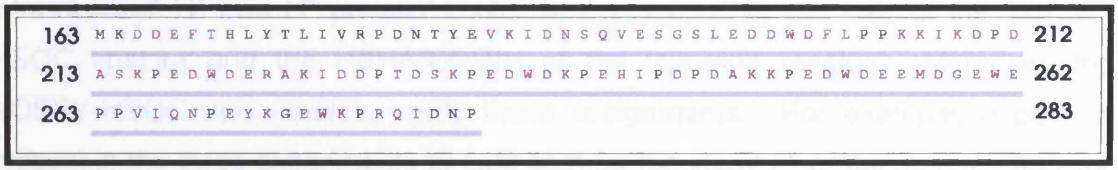

Figure 4.11 : The sequence of hCRT S-domain (164-283). Assigned backbone amides residues are shown in red. Residues with unassigned backbone amides, including prolines, are shown in black. The sequence numbering is that of human CRT, SwissProt accession number P27797, with an extra Met in position 163 from the cloning.

4.4.3 Side-chain resonance assignment of the CRT S-domain

The side-chain assignment was done in two steps through the use of 3D ^{15}N- and ^{13}C-resolved NOESY experiments. The first one mainly allowing the assignment of the α and β protons and the second one the complete side-chain, including the prolines and the aromatics.

4.4.3.1 Assignment of the H^{α} and H^{β} protons

Through the use of 3D [^{1}H,^{15}N]-NOESY-HSQC, [^{1}H,^{15}N]-TOCSY-HSQC and HBHA(CO)NH experiments, most of the H^{α} and H^{β} protons of the previously assigned backbone C^{α} and side-chain C^{β} were identified (see table IV.II). First, by the use of the 3D [^{1}H,^{15}N]-TOCSY-HSQC and HBHA(CO)NH experiments, the α and β protons chemical shifts of a particular residue *i* and *i-1*, respectively, were determined. Finally, the [^{1}H,^{15}N]-NOESY-HSQC showed the NOEs from spatially close protons, which in general includes the α and β protons of residue *i* and more (figure 4.12 and 4.13). The proline H^{α}, $H^{\beta1}$ and $H^{\beta2}$ could also be easily assigned by the use of the 3D HBHA(CO)NH.

An example of the inter- and intra-residual connectivities is shown in figures 4.12 and 4.13. By centring strips onto the the HN spin of each residues, the intra-residual H^{α} and H^{β} protons could be determined by the use of the TOCSY-HSQC spectra and the HBHA(CO)NH of the following residue. Moreover, the NOESY-HSQC was used to verify those assignments. For example, a peak is present in the three experiments at 4.20 ppm for the Q267 H^{α} proton and similarly for the H^{β} protons at 1.99 ppm. At the same time the backbone assignment could be put to the challenge allowing us to correct some possible wrong assignments. Wrong assignments had no coherence in their connectivities between consecutive spins, i.e. H^{α} and H^{β} peaks in the TOCSY were different to the HBHA(CO)NH peaks of the following residue. Overlapped and ambiguous cross-peaks were left aside until they could be resolved at a later stage with the use of the ^{13}C-resolved experiments.

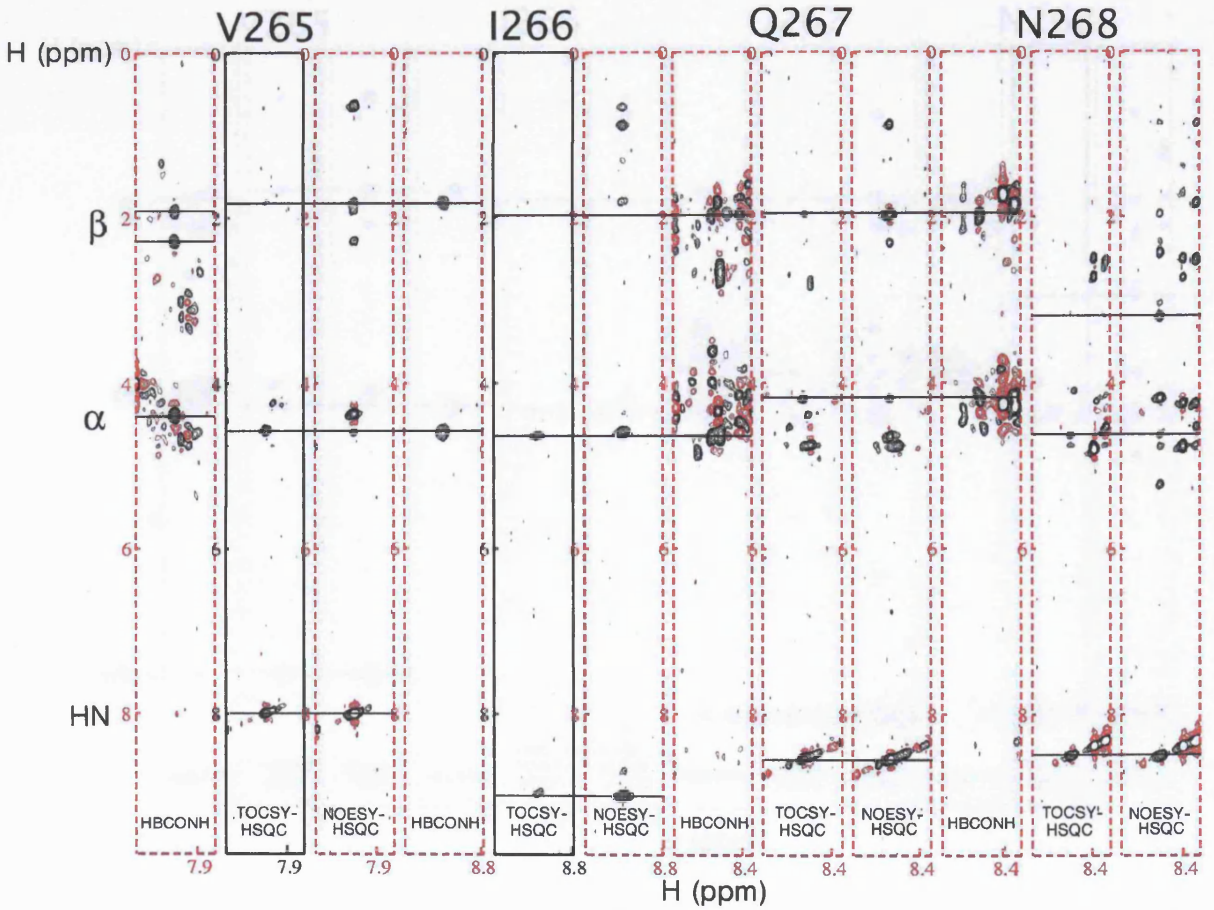

Figure 4.12 : Contour plot from 3D [^{1}H/^{1}H] ^{15}N-resolved spectra. The twelve spectral regions are centred about the H^{N} chemical shifts of the residues V265, I266, Q267, and N268, as indicated at the top. The first and second strips shows the HBHA(CO)NH and TOCSY-HSQC experiment, as indicated at the bottom, providing the α and β protons chemical shifts of the residue *i*-1 and *i* respectively. The third strip shows the 3D ^{15}N-resolved NOESY-HSQC spectrum of the CRT S-domain.

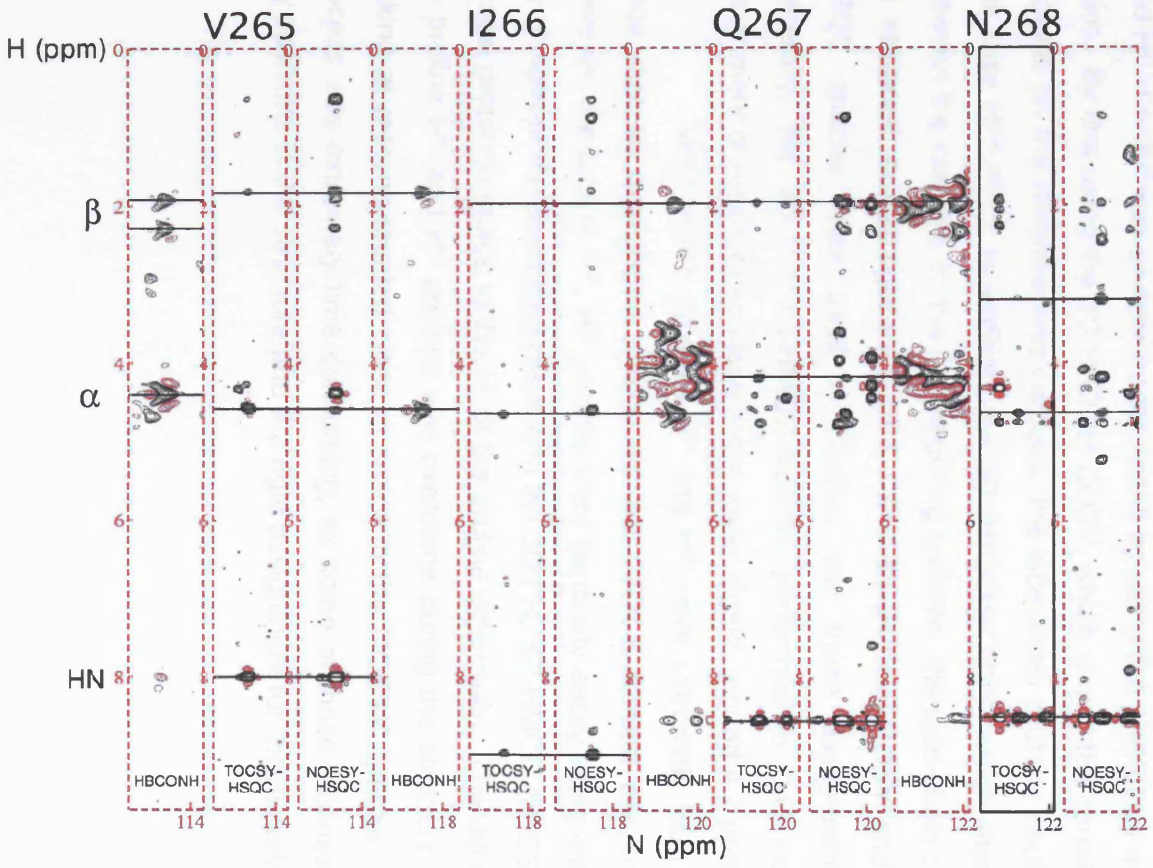

Figure 4.13 : Contour plot from 3D [^{1}H/^{15}N] ^{15}N-resolved spectra. The twelve spectral regions are centred about the N chemical shifts of the residues V265, I266, Q267, and N268, as indicated at the top. The first and second strips shows the HBHA(CO)NH and TOCSY-HSQC experiment, as indicated at the bottom, providing the α and β protons chemical shifts of the residue *i*-1 and *i* respectively. The third strip shows the 3D ^{15}N-resolved NOESY-HSQC spectrum of the CRT S-domain. The peak shape in the ^{15}N dimension of the HBHA(CO)NH experiment is distorted due to a bug in an XWINNMR software at the time.

4.4.3.2 Assignment of the rest of the chains

The side-chain assignment was based on eight different ^{13}C-resolved experiments ; first the 2D [^{1}H,^{13}C]-HSQC and full sweep 2D [^{1}H,^{13}C]-HSQC for the aromatic region and then the 3D [^{1}H,^{13}C]-NOESY-HSQC (on the aliphatic region), 3D [^{1}H,^{13}C]-NOESY-HSQC (on the aromatic region), 3D [^{1}H,^{13}C]-HMQC-NOESY (in D_2O), 3D (H)CCH-COSY and 3D H(C)CH-COSY.

Most of the side-chains could be assigned starting from the previously assigned α and β spins. Figure 4.14 is an example of how the assignment of $C^{\gamma 1,\ \gamma 2}$ and $H^{\gamma 11,\ \gamma 12,\ \text{and}\ \gamma 2}$ was performed for Ile266 by using the previously assigned C^{β} and H^{β} spins. By first using the 3D H(C)CH-COSY, which shows the correlation between the protons on the neighbouring carbons, the side-chain protons could be assigned, in this case $H^{\gamma 12,\ \text{and}\ \gamma 2}$. In addition, the 3D (H)CCH-COSY, which shows the correlation between the carbons on the neighbouring carbons, the side-chain carbons could also be assigned, in this figure $C^{\gamma 1,\ \text{and}\ \gamma 2}$. The 3D (H)CCH-COSY and [^{1}H,^{13}C]-NOESY-HSQC spectra were used to confirm that those assignments were correct. Generally, the 3D [^{1}H,^{13}C]-HMQC-NOESY performed in D_2O was only used for assignment of resonances close to the water signal, around 4.7 ppm.

Most of the prolines $H^{\gamma 1}$ and $H^{\gamma 2}$ were extremely hard to assign at this stage due to overlapped peaks, as the CRT S-domain contained 16 prolines. However, the proline $H^{\delta 1}$, $H^{\delta 2}$ protons were generally easily assigned by centring onto the H^{α} spin of the preceding residue in the 3D [^{1}H,^{13}C]-HMQC-NOESY, which has the closest proton in space to the H^{δ} of the proline side-chain. The difficulty in assigning the proline $H^{\gamma 1}$ and $H^{\gamma 2}$ protons was overcome during the structure determination by looking at protons located close in space in the NOESY spectra. That part of the process was extremely time consuming, as some of those prolines were located at key positions in the structure and the right assignment for those prolines was critical.

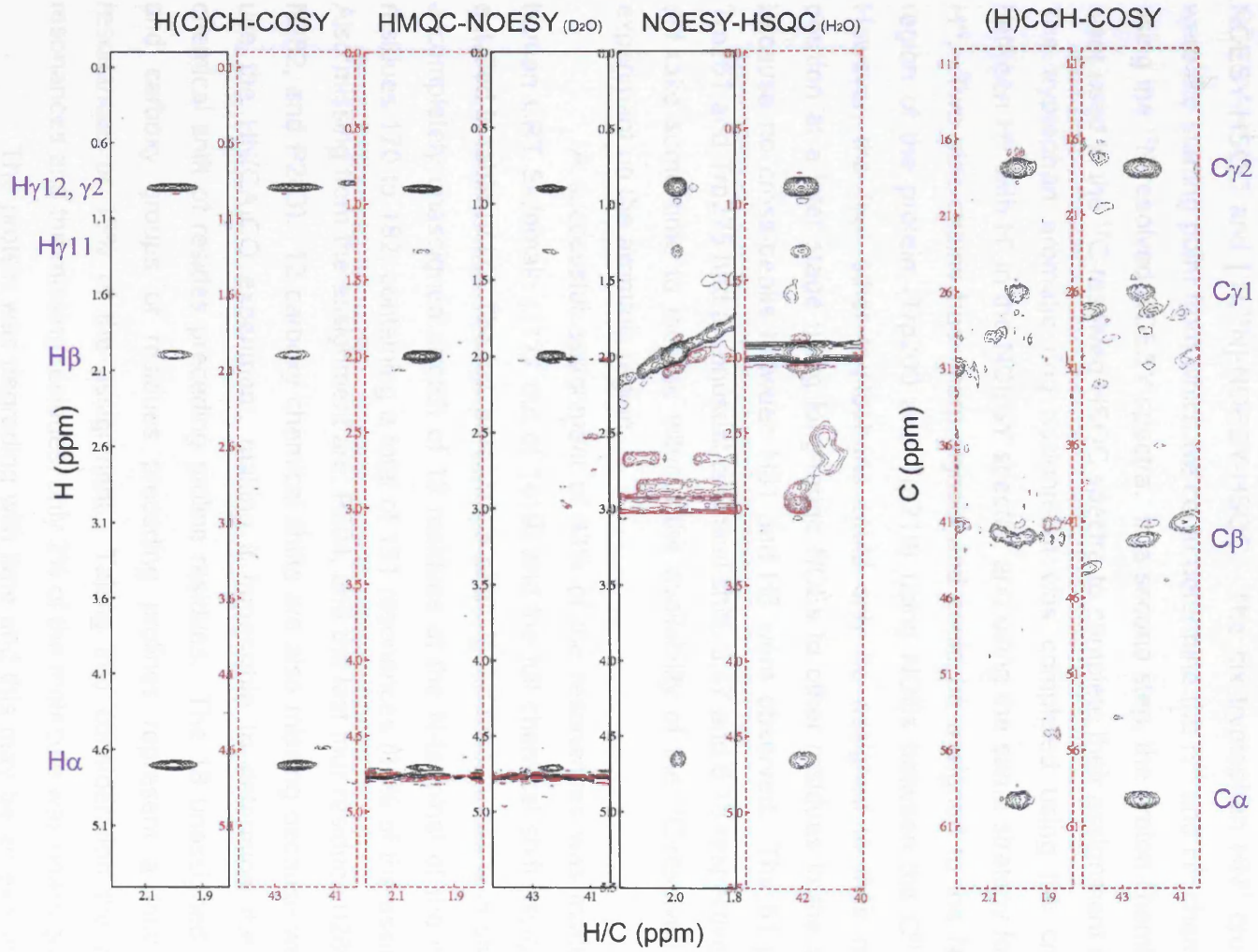

Figure 4.14 : Side-chain assignment using 4 different 3D experiments. Example for Ile266 centred on the C^{β} spin from which assignment of the $C^{\gamma 1, \gamma 2}$ and $H^{\gamma 11, \gamma 12, \text{and } \gamma 2}$ was obtained. From left to right; 3D [^{1}H,^{1}H]-H(C)CH-COSY, 3D [^{1}H,^{1}H]-HMQC-NOESY (D_2O), 3D [^{1}H,^{1}H]-NOESY-HSQC, and 3D (H)CCH-COSY.

The full sweep 2D [^{1}H,^{13}C]-HSQC, and 3D [^{1}H,^{13}C]-NOESY-HSQC on the aromatic region made possible the assignment of the side-chains of tryptophans, histidines, phenyalanines, and tyrosines. This was straightforward due to the low content of those aromatic amino acids in the CRT S-domain (6Trp, 2His, 2Phe, and 3Tyr). The tryptophan aromatic rings were easily assigned using the 3D [^{1}H,^{13}C]-NOESY-HSQC and [^{1}H,^{15}N]-NOESY-HSQC. The six tryptophan HN$^{\varepsilon 1}$ cross-peaks were the starting point from which we could determine the H$^{\delta 1}$ and H$^{\zeta 2}$ chemical shift using the ^{15}N-resolved NOESY spectra. In a second step, the proton chemical shifts were used in the ^{13}C-resolved HSQC spectra to complete their assignment (C$^{\delta 1}$, C$^{\zeta 2}$). The tryptophan aromatic ring assignment was completed using the cross-peak between H$^{\zeta 2}$ with H$^{\eta}$ in the NOESY spectra, and using the same strategy for H$^{\zeta 3}$ and H$^{\varepsilon 3}$. Two side-chains had sharp signals and could be assigned to the N-terminal region of the protein (Trp200 and Trp219) using NOEs between the C$^{\delta 1}$ and C$^{\beta}$. However, the four other tryptophans could only be assigned to their respective position at a later stage using long-range NOEs to other residues in the structure, because no cross-peaks between Hδ1 and Hβ were observed. The δ1 proton in Trp261 and Trp275 had an unusual chemical shift, 5.47 and 6.18 respectively, which did take some time to localise without the availability of the ^{13}C-resolved TOCSY experiment on the aromatic region.

A successful assignment of 83% of the resonances was made for the human CRT S-domain (1177 out of 1419) and the full chemical shift assignment is detailed in table IV.IX. The low percentage of assigned resonances is in part due to a completely unassigned stretch of 13 residues at the N-terminal of the molecule, residues 170 to 182, containing a total of 151 resonances (10% of the assignment). Also missing from the assignment are; P204, and the last four residues (I280, D281, N282, and P283). 12 carboxy chemical shifts are also missing because we did not use the HN(CA)CO experiment, making it impossible to determine the carboxy chemical shift of resides preceding proline residues. The 18 unassigned residues and carboxy groups of residues preceding prolines represent a total of 220 resonances or 15% of the assignment. Taking into consideration the assigned resonances and the missing residues, only 2% of the molecule was unassigned.

The protein was degrading with time and this may be an explanation for the missing residues 170 to 182 from the assignment. It can be observed in the 2D-HSQC spectra of the protein (figure 4.4), that multiple unassigned small and sharp peaks indicative of degradation are present. Also, there are some overlapped peaks in the middle of the spectra that were impossible to assign due to multiple forms,

also indicative of sample degradation.

Table IV.IX : Chemical shifts (ppm) assignment of the CRT S-domain. Amide proton (H^N), amide nitrogen (N), carbons (C^α, C^β, etc.), protons (H^α, H^β, etc.) and carbonyl carbon (CO). An entry of 'NA' indicates that the chemical shift could not be assigned due to insufficient data.

Residue	*H^N (N)*	*CO*	*H^α (C^α)*	*H^β (C^β)*	*H^γ (C^γ)*	*Others*
M163	NA	NA	4.09 (55.1)	2.16, 2.16 (33.2)	2.61, 2.61 (31.1)	H^ε 2.09 (16.9)
K164	NA	176.2	4.34 (56.8)	1.86, 1.78 (33.1)	1.43, 1.43 (24.5)	H^δ 1.68, 1.68 (28.8)
						H^ε 2.98, 2.98 (41.8)
D165	8.51 (121.5)	176.1	4.53 (55.1)	2.72, 2.63 (40.9)		
D166	8.25 (119.5)	176.6	4.55 (54.5)	2.64, 2.64 (40.9)		
E167	8.26 (120.3)	176.6	4.14 (57.3)	1.90, 1.86 (29.9)	2.25, 2.25 (36.1)	
F168	8.17 (119.3)	176.3	4.63 (58.2)	3.19, 3.02 (39.3)		H^δ NA, H^ε NA, H^ζ NA
T169	7.94 (114.5)	NA	4.16 (63.0)	4.13 (69.7)	1.16 (21.5)	
H170	NA	NA	NA	NA		H^δ 7.07 (119.8)
						H^ε 8.21 (137.3)
L171	NA	NA	NA	NA	NA	$H^{\delta 1}$ NA, $H^{\delta 2}$ NA
Y172	NA	NA	NA	NA		H^δ NA, H^ε NA
T173	NA	NA	NA	NA	NA	
L174	NA	NA	NA	NA	NA	$H^{\delta 1}$ NA, $H^{\delta 2}$ NA
I175	NA	NA	NA	NA	$H^{\gamma 1}$ NA, $H^{\gamma 2}$ NA	H^δ NA
V176	NA	NA	NA	NA	$H^{\gamma 1}$ NA, $H^{\gamma 2}$ NA	
R177	NA	NA	NA	NA	NA	H^δ NA, HN^ε NA
P178		NA	NA	NA	NA	H^δ NA
D179	NA	NA	NA	NA		
N180	NA	NA	NA	NA		HN^δ NA
T181	NA	NA	NA	NA	NA	
Y182	NA	NA	NA	NA		H^δ NA, H^ε NA
E183	NA	175.8	4.26 (56.1)	1.87, 1.96 (31.1)	2.12, 2.18 (36.2)	
V184	8.14 (122.5)	175.7	4.04 (62.3)	2.01 (32.7)	$H^{\gamma 1}$ 0.94 (20.8)	
					$H^{\gamma 2}$ NA	
K185	8.42 (126.4)	176.1	4.42 (55.9)	1.77, 1.72 (33.4)	1.37, 1.37 (24.6)	H^δ 1.59, 1.63 (28.8)
						H^ε 2.96, 2.96 (41.8)
I186	8.29 (123.2)	175.7	4.14 (61.0)	1.82 (38.8)	$H^{\gamma 1}$ 1.44, 1.16 (27.1)	H^δ 0.83 (12.7)
					$H^{\gamma 2}$ 0.89 (17.3)	
D187	8.42 (124.5)	175.9	4.61 (54.0)	2.72, 2.62 (41.4)		
N188	8.49 (120.9)	175.6	4.76 (53.4)	2.86, 2.76 (38.6)		HN^δ 7.55, 6.86 (112.4)
S189	8.41 (116.1)	174.7	4.34 (59.5)	3.89, 3.89 (63.6)		
Q190	8.29 (121.6)	175.9	4.39 (55.9)	2.14, 1.99 (29.3)	2.32, 2.35 (33.7)	HN^ε 7.53, 6.82 (112.6)
V191	8.02 (120.8)	176.3	4.11 (62.5)	2.08 (32.8)	$H^{\gamma 1}$ 0.92 (20.6)	
					$H^{\gamma 2}$ 0.92 (21.2)	
E192	8.51 (124.6)	176.6	4.31 (56.5)	2.04, 1.92 (30.1)	2.26, 2.26 (36.0)	
S193	8.41 (117.3)	175.1	4.43 (58.7)	3.90, 3.90 (63.9)		
G194	8.45 (111.0)	174.1	3.98, 4.00 (45.2)			
S195	8.21 (115.5)	174.5	4.48 (58.1)	3.83, 4.00 (64.0)		
L196	8.39 (124.3)	177.4	4.36 (55.4)	1.64, 1.58 (42.2)	1.62 (26.9)	$H^{\delta 1}$ 0.87 (24.9)
						$H^{\delta 2}$ 0.83 (23.3)
E197	8.30 (120.5)	176.2	4.13 (56.6)	2.01, 1.84 (30.2)	2.19, 2.19 (36.1)	

Residue	H^N (N)	CO	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	Others
D198	8.13 (120.5)	175.8	4.48 (54.6)	2.49, 2.49 (41.1)		
D199	8.13 (120.0)	176.3	4.54 (54.4)	2.63, 2.54 (40.8)		
W200	7.96 (121.1)	176.1	4.58 (57.4)	3.18, 3.23 (29.5)		
	$H^{\delta1}$ 7.20 (127.2) $HN^{\varepsilon1}$ 10.08 (129.2) $H^{\varepsilon3}$ NA (NA) $H^{\zeta2}$ 7.45 (114.4) $H^{\zeta3}$ 7.48 (NA) $H^{\eta2}$ 7.22 (124.6)					
D201	8.13 (121.0)	175.5	4.46 (54.3)	2.36, 2.36 (40.6)		
F202	7.69 (119.0)	175.1	4.45 (57.6)	3.02, 3.02 (39.4)		H^δ NA, H^ε NA, H^ζ NA
L203	7.89 (125.1)	NA	4.56 (52.6)	1.44, 1.51 (42.0)	1.51 (26.7)	$H^{\delta1}$ 0.86 (25.0)
						$H^{\delta2}$ 0.86 (23.5)
P204		NA	NA	NA	NA	H^δ NA
P205		176.9	4.37 (62.8)	2.27, 1.92 (32.2)	2.02, 2.02 (27.2)	H^δ 3.76, 3.56 (50.6)
K206	8.32 (121.5)	176.5	4.21 (56.2)	1.72, 1.72 (33.0)	1.36, 1.36 (24.7)	H^δ 1.67, 1.67 (28.9)
						H^ε 2.96, 2.96 (41.9)
K207	8.28 (122.3)	176.2	4.32 (56.1)	1.77, 0.71 (33.0)	1.38, 1.38 (24.7)	H^δ 1.68, 1.68 (28.9)
						H^ε 2.97, 2.97 (41.8)
I208	8.20 (123.0)	175.9	4.14 (60.8)	1.82 (38.8)	$H^{\gamma1}$ 1.45, 1.17 (27.2)	H^δ 0.84 (12.7)
					$H^{\gamma2}$ 0.88 (17.3)	
K209	8.34 (125.8)	175.7	4.26 (56.2)	1.71, 1.71 (33.1)	1.38, 1.38 (24.5)	H^δ 1.65, 1.65 (28.9)
						H^ε 2.97, 2.97 (41.9)
D210	8.39 (123.8)	NA	4.80 (52.0)	2.72, 2.48 (40.9)		
P211		NA	4.36 (63.8)	2.27, 1.93 (32.0)	2.01, 2.01 (27.1)	H^δ 3.84, 3.79 (50.8)
D212	8.36 (119.6)	176.1	4.56 (54.4)	2.72, 2.57 (41.1)		
A213	8.03 (124.4)	177.7	4.32 (52.6)	1.41 (19.1)		
S214	8.26 (115.1)	174	4.40 (58.5)	3.83, 3.83 (64.0)		
K215	8.20 (123.9)	NA	4.58 (54.1)	1.67, 1.74 (32.6)	1.33, 1.33 (24.2)	H^δ 1.53, 1.53 (28.8)
						H^ε 2.85, 2.85 (41.7)
P216		177.4	4.31 (63.3)	2.27, 1.92 (32.0)	2.01, 2.01 (27.6)	H^δ 3.78, 3.57 (50.7)
E217	8.72 (120.3)	176.5	4.15 (57.3)	1.85, 1.91 (29.8)	2.23, 2.23 (36.1)	
D218	8.26 (119.8)	176.2	4.56 (54.2)	2.65, 2.65 (40.8)		
W219	7.88 (120.9)	176.1	4.54 (57.9)	3.29, 3.29 (29.5)		
	$H^{\delta1}$ 7.25 (127.2) $HN^{\varepsilon1}$ 10.16 (129.4) $H^{\varepsilon3}$ NA (NA) $H^{\zeta2}$ 7.47 (114.6) $H^{\zeta3}$ NA (NA) $H^{\eta2}$ 7.22 (124.6)					
D220	8.06 (122.0)	176.2	4.45 (54.3)	2.51, 2.42 (41.1)		
E221	8.12 (121.5)	176.8	4.07 (57.0)	2.03, 1.92 (29.8)	2.20, 2.20 (36.1)	
R222	8.12 (120.6)	176.3	4.21 (56.4)	1.82, 1.74 (30.3)	1.61, 1.61 (26.9)	H^δ 3.12, 3.12 (43.1)
						HN^ε NA
A223	8.08 (124.0)	177.6	4.24 (52.6)	1.35 (18.9)		
K224	8.07 (120.3)	176.3	4.36 (56.0)	1.79, 1.71 (33.4)	1.35, 1.35 (24.5)	H^δ 1.65, 1.65 (28.8)
						H^ε 2.98, 2.98 (41.7)
I225	8.21 (120.8)	175.6	4.28 (60.7)	1.87 (39.6)	$H^{\gamma1}$ 1.41, 1.08 (26.8)	H^δ 0.85 (13.1)
					$H^{\gamma2}$ 0.88 (17.5)	
D226	8.33 (123.8)	175.4	4.55 (54.3)	2.51, 2.59 (41.3)		
D227	8.32 (119.9)	NA	4.76 (51.5)	2.80, 2.56 (39.1)		
P228		177.5	4.52 (63.8)	2.32, 2.04 (32.2)	1.93, 1.93 (27.1)	H^δ 3.66, 3.81 (50.4)
T229	8.38 (112.6)	174.5	4.33 (62.2)	4.28 (69.5)	1.24 (21.5)	
D230	8.06 (122.8)	175.6	4.75 (54.1)	2.77, 2.55 (41.6)		
S231	8.09 (116.1)	173.1	4.37 (57.7)	3.71, 3.66 (64.4)		
K232	7.94 (125.7)	NA	4.34 (54.2)	1.51, 1.59 (34.8)	1.31, 1.31 (24.1)	H^δ 1.60, 1.60 (28.7)
						H^ε 2.95, 2.95 (41.8)
P233		177.5	4.28 (63.1)	2.40, 1.98 (32.4)	2.06, 2.06 (27.5)	H^δ NA (50.9)

Residue	H^N (N)	CO	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	Others
E234	8.81 (121.4)	176.6	4.06 (58.2)	1.94, 1.94 (29.6)	2.26, 2.26 (36.1)	
D235	8.35 (117.9)	175.9	4.65 (53.5)	2.76, 2.60 (40.3)		
W236	7.57 (119.4)	177.1	4.14 (59.8)	3.04, 3.22 (30.2)		
	$H^{\delta1}$ 7.42 (127.5) $HN^{\varepsilon1}$ 10.00 (129.5) $H^{\varepsilon3}$ 7.35 (122.2) $H^{\zeta2}$ 7.37 (115.2) $H^{\zeta3}$ 7.59 (120.5) $H^{\eta2}$ 7.23 (122.1)					
D237	8.35 (118.4)	175.2	4.53 (55.1)	2.52, 2.52 (40.4)		
K238	7.84 (123.2)	NA	4.14 (52.7)	1.49, 1.49 (33.6)	1.16, 1.16 (24.1)	H^δ 1.33, 1.33 (28.8)
						H^ε 2.85, 2.85 (41.8)
P239		176.9	4.40 (62.7)	2.34, 1.99 (31.9)	2.05, 2.05 (27.2)	H^δ 3.74, 3.55 (50.5)
E240	8.50 (122.1)	176.1	3.15 (58.7)	1.37, 1.29 (29.8)	1.18, 1.32 (35.6)	
H241	7.89 (114.8)	174.1	5.29 (54.6)	2.88, 2.77 (32.2)		H^δ 6.77 (119.3)
						H^ε 8.00 (137.1)
I242	8.69 (116.3)	NA	4.80 (58.3)	1.93 (40.2)	$H^{\gamma1}$ 1.44, 1.44 (25.3)	H^δ 0.87 (13.6)
					$H^{\gamma2}$ 0.86 (18.5)	
P243		176.6	4.39 (62.5)	2.06, 1.68 (31.8)	2.05, 1.67 (27.7)	H^δ 3.60, 3.73 (50.6)
D244	8.60 (123.3)	NA	4.47 (51.9)	2.43, 2.79 (42.0)		
P245		177.3	4.42 (64.1)	2.30, 2.01 (32.1)	1.95, 2.04 (26.6)	H^δ 4.08, 3.98 (51.1)
D246	8.36 (117.9)	175.9	4.75 (54.4)	2.77, 2.56 (41.7)		
A247	7.21 (123.3)	176.7	4.37 (52.5)	1.62 (20.0)		
K248	8.35 (121.6)	174.7	4.34 (54.4)	1.52, 1.59 (34.8)	1.31, 1.31 (24.2)	H^δ 1.61, 1.61 (28.9)
						H^ε 2.94, 2.94 (41.9)
K249	7.85 (127.4)	NA	4.32 (53.7)	0.98, 0.42 (32.5)	0.04, 0.93 (22.8)	H^δ 1.43, 1.33 (29.6)
						H^ε 2.75, 2.83 (41.3)
P250		178.1	4.27 (63.1)	2.50, 2.09 (32.7)	2.32, 2.06 (27.9)	H^δ 3.28, 2.58 (50.8)
E251	8.92 (123.0)	176.5	4.00 (58.8)	2.05, 1.98 (29.6)	2.32, 2.32 (36.3)	
D252	8.31 (115.1)	175.8	4.63 (52.6)	2.89, 2.59 (39.7)		
W253	7.43 (121.8)	175.3	4.09 (58.8)	3.19, 2.94 (29.7)		
	$H^{\delta1}$ 7.04 (126.2) $HN^{\varepsilon1}$ 10.62 (129.7) $H^{\varepsilon3}$ 7.14 (122.3) $H^{\zeta2}$ 7.26 (115.7) $H^{\zeta3}$ 7.59 (121.0) $H^{\eta2}$ 7.37 (124.9)					
D254	8.56 (130.0)	175.4	4.80 (52.4)	2.70, 2.31 (41.9)		
E255	9.02 (125.4)	178.8	4.36 (59.0)	2.04, 2.14 (29.6)	2.55, 2.58 (37.1)	
E256	8.16 (118.6)	177.6	4.04 (50.0)	2.05, 2.09 (29.3)	2.31, 2.22 (36.3)	
M257	7.42 (115.8)	177.4	4.43 (56.0)	1.86, 1.74 (33.5)	2.50, 2.39 (31.9)	H^ε 2.02 (16.8)
D258	8.45 (117.1)	176.7	4.58 (55.2)	1.86, 1.37 (41.3)		
G259	7.59 (108.4)	172.7	4.30, 3.94 (44.2)			
E260	8.53 (119.6)	177.1	4.46 (56.8)	2.00, 2.02 (30.1)	2.30, 2.38 (36.1)	
W261	9.10 (128.3)	175.1	3.83 (59.4)	2.59, 1.39 (27.5)		
	$H^{\delta1}$ 5.46 (126.6) $HN^{\varepsilon1}$ 9.95 (129.9) $H^{\varepsilon3}$ 7.01 (120.3) $H^{\zeta2}$ 7.08 (113.6) $H^{\zeta3}$ 6.90 (121.7) $H^{\eta2}$ 6.80 (123.6)					
E262	7.01 (127.4)	NA	4.17 (52.2)	1.64, 1.49 (31.2)	1.95, 1.95 (35.4)	
P263		NA	3.61 (60.5)	2.15, 1.69 (30.8)	1.77, 1.85 (26.9)	H^δ 3.06, 3.32 (50.0)
P264		175.3	4.40 (62.6)	2.28, 1.93 (32.1)	2.06, 2.06 (27.2)	H^δ 3.41, 3.66 (50.7)
V265	8.00 (115.3)	176.5	4.66 (60.5)	1.82 (33.8)	$H^{\gamma1}$ 0.67 (21.3)	
					$H^{\gamma2}$ 0.62 (19.1)	
I266	8.99 (119.0)	175.7	4.71 (59.5)	2.00 42.3	$H^{\gamma1}$ 1.31, 0.90 (26.0)	H^δ 0.90 (14.0)
					$H^{\gamma2}$ 0.90 (18.0)	
Q267	8.54 (121.0)	175.6	4.20 (56.4)	1.99, 1.99 (28.8)	2.32, 2.35 (33.9)	HN^ε 7.56, 6.88 (112.0)
N268	8.51 (123.6)	NA	4.63 (50.4)	3.25, 2.43 (39.7)		HN^δ 8.45, 7.71 (112.7)
P269		177.2	4.45 (64.6)	2.41, 1.99 (32.3)	2.07, 2.07 (27.3)	H^δ 4.05, 4.05 (51.4)
E270	7.45 (113.9)	175.9	4.17 (56.0)	2.12, 1.70 (30.4)	2.28, 2.13 (36.8)	
Y271	7.57 (119.8)	176.1	4.21 (58.9)	3.04, 3.09 (38.6)		H^δ NA, H^ε NA

Residue	H^N (N)	CO	H^α (C^α)	H^β (C^β)	H^γ (C^γ)	Others
K272	8.38 (126.9)	NA	4.09 (55.7)	1.45, 1.45 (33.1)	0.95, 1.01 (24.1)	H^δ 1.03, 0.95 (28.0)
						H^ε 2.62, 2.62 (41.7)
G273	NA	173.5	4.00, 3.58 (44.6)			
E274	8.54 (121.4)	176.4	4.43 (56.6)	1.96, 2.00 (30.4)	2.33, 2.26 (36.0)	
W275	8.76 (126.3)	175.2	4.10 (58.8)	2.95, 3.19 (29.8)		
	$H^{\delta1}$ 6.18 (126.7) $HN^{\varepsilon1}$ 9.92 (129.7) $H^{\varepsilon3}$ 6.81 (118.3) $H^{\zeta2}$ 6.81 (114.4) $H^{\zeta3}$ 6.90 (121.7) $H^{\eta2}$ 6.99 (124.2)					
K276	7.15 (125.9)	NA	4.14 (53.0)	1.31, 1.50 (33.5)	1.17, 1.17 (24.0)	H^δ 1.53, 1.53 (28.8)
						H^ε 2.87, 2.87 (41.7)
P277		176.7	3.96 (62.2)	2.23, 1.78 (32.4)	1.84, 1.77 (27.0)	H^δ 3.09, 3.28 (50.2)
R278	8.64 (121.6)	175.8	4.14 (56.2)	1.77, 1.77 (30.7)	1.61, 1.61 (27.0)	H^δ 3.12, 3.12 (43.2)
						HN^ε NA
Q279	8.39 (121.6)	175.8	4.54 (55.1)	1.93, 1.93 (30.0)	2.27, 2.27 (33.8)	HN^ε 7.52, 6.91 (113.0)
I280	8.40 (121.4)	NA	NA	NA	$H^{\gamma1}$ NA, $H^{\gamma2}$ NA	H^δ NA
D281	NA	NA	NA	NA		
N282	NA	NA	NA	NA		HN^δ NA
P283		NA	NA	NA	NA	H^δ NA

4.4.4 Structure

4.4.4.1 Secondary structure

Long range NOEs indicative of an anti-parallel β-sheet were first identified (see figure 4.15). They include H^N-H^N cross-peaks between Ile242 and Ile266 found in the 3D [^{1}H,^{15}N]-NOESY-HSQC (figure 4.15a, middle). Also, long-range H^α-H^α NOEs indicative of anti-parallel β-sheet were found in the 3D [^{1}H,^{13}C]-HMQC-NOESY spectra between the residue pairs His241, Gln267 and Pro243, Val265 (figure 4.15a, left and right respectively), which define the short β-sheet shown in figure 4.15b. The presence of this β-sheet was supported by further NOE constraints, and they are summarised in figure 4.15b by black arrows.

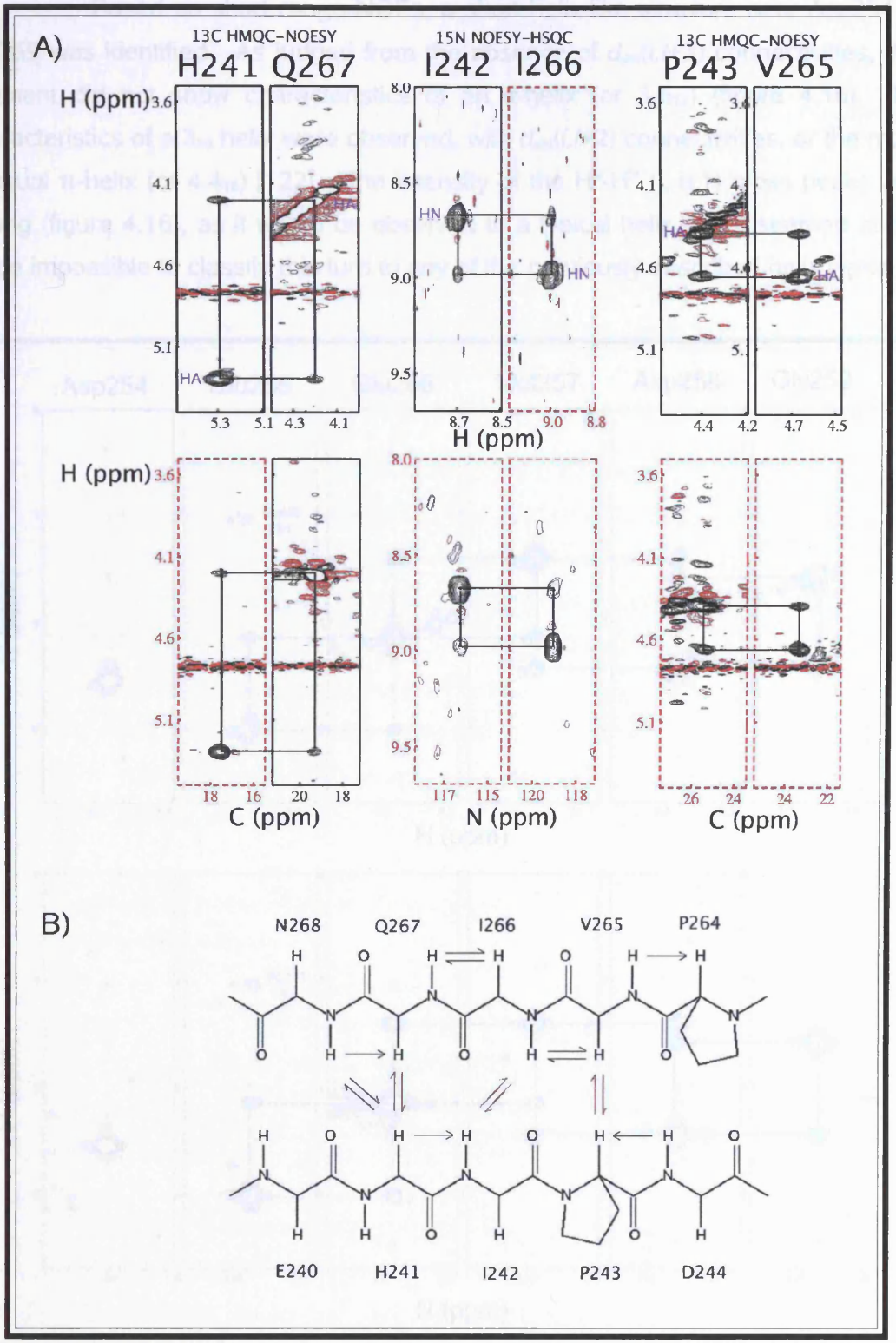

Figure 4.15 : Identification of one anti-parallel β-sheet in the hCRT S-domain. A) Spectral regions from 3D [^{1}H,^{13}C]-HMQC-NOESY recorded in D_2O, left and right, and in the middle 3D [^{1}H,^{15}N]-NOESY-HSQC. Squares connect the cross-peaks. B) Uniquely assigned NOEs used to deduce the ordering and alignment of the β-strand are shown as arrows.

Based on short range NOEs, a short helix-like structure near Asp254 to Gly259 was identified. As judged from the absence of $d_{\alpha N}(i,i+3)$ connectivities, this segment did not show characteristics of an α-helix (or 3.6_{13}) (figure 4.16). No characteristics of a 3_{10} helix were observed, with $d_{\alpha N}(i,i+2)$ connectivities, or the more unusual π-helix (or 4.4_{16}) [122]. The intensity of the H^N-H^N (i, i±1) cross peaks was strong (figure 4.16), as it would be observed in a typical helix, but it seemed at this stage impossible to classify this turn to any of the previously described helix types.

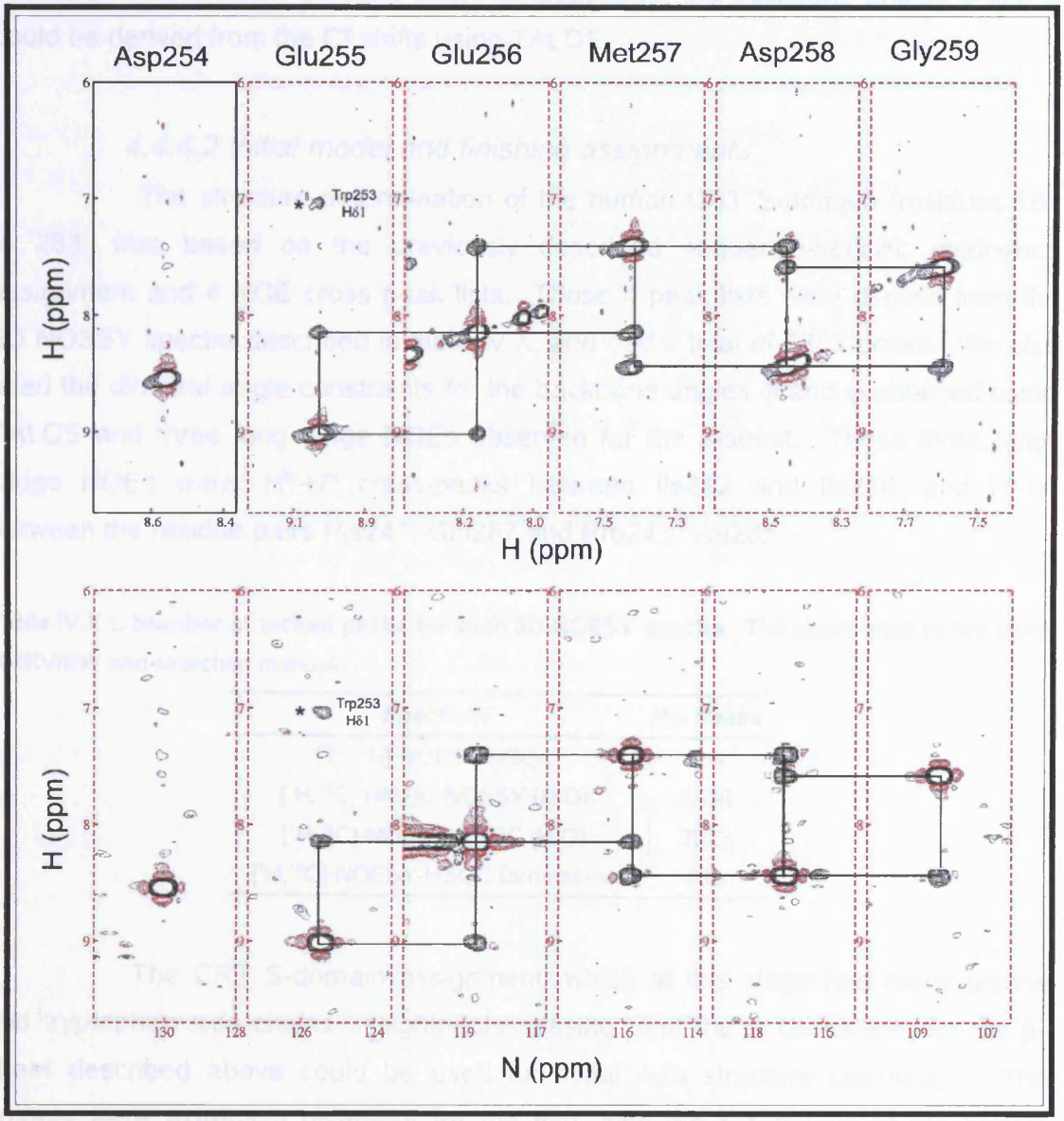

Figure 4.16 : Identification of the short helical turn in the hCRT S-domain. Spectral regions from 3D[^{1}H,^{15}N]-NOESY-HSQC. Squares connect the H^N-H^N cross-peaks.

At the N-terminus of the protein chain, residues 163 to 229, very sharp

resonances could be observed in the 2D [^{1}H,^{15}N]-HSQC spectra indicative of a long unfolded tail (figure 4.4). Moreover, in the 3D [^{1}H,^{15}N]-NOESY-HSQC and [^{1}H,^{13}C]-NOESY-HSQC spectra, only intra-residual and sequential NOEs could be observed also supporting the absence of secondary structure elements in this region of the protein.

Secondary structure analysis on the CRT S-domain based on the resonance chemical shift did not produce any usueful information on the secondary structure. However, 30 dihedral angle constraints for the backbone angles ϕ and ψ could be derived from the C^{α} shifts using TALOS.

4.4.4.2 Initial model and finishing assignments

The structure determination of the human CRT S-domain (residues 164 to 283) was based on the previously described sequence-specific resonance assignment and 4 NOE cross peak lists. Those 4 peak lists were derived from the 3D NOESY spectra described in table IV.X, and had a total of 5363 peaks. We also used the dihedral angle constraints for the backbone angles ϕ and ψ obtained using TALOS and three long-range NOEs observed for the β-sheet. Those three long-range NOEs were; H^{N}-H^{N} cross-peaks between Ile242 and Ile266, and H^{α}-H^{α} between the residue pairs His241, Gln267 and Pro243, Val265.

Table IV.X : Number of picked peaks for each 3D NOESY spectra. The peaks were picked using NMRVIEW and searched manually.

Spectrum	*No. Peaks*
[^{1}H,^{15}N]-NOESY-HSQC	664
[^{1}H,^{13}C]-HMQC-NOESY (D_2O)	2438
[^{1}H,^{13}C]-NOESY-HSQC (H_2O)	2025
[^{1}H,^{13}C]-NOESY-HSQC (aromatics)	236

The CRT S-domain assignment, which at this stage had many proline and tryptophan side-chains assignments missing, and the 3 constraints for the β-sheet described above could be used for initial Aria structure calculation. The outputs were extremely divergent for the fold, with the N-terminus of the protein folding in strange ways. At this stage, only residues 217 to 283 were used for the structure calculations, which contained the most resolved peaks in the protein (see figures 4.4 and 4.11). That process speeded up the structure calculations and the analysis of the structures. The Aria output started to form the expected β-sheet after

a few cycles with a very flexible hairpin. The structures and the Aria assignment were used to facilitate the assignment of the two tryptophans (Trp253 and 261) and 5 proline side-chains (Pro243, 245, 250, 263, and 264) contained in this hairpin.

Several long-range NOEs were observed between the side-chains of residues; Trp253, Trp261, Lys249, Pro263 and Pro250, which helped in the first instance to assign the tryptophan side-chains to their respective positions and also for the proline side-chain assignment. As illustrated in figure 4.17 by the black lines, they include inter-residual NOEs between Trp253 $HN^{\varepsilon 1}$ and Trp261 H^{N}, H^{α}, and H^{β}. Some long-range NOEs could also be seen between the Trp253, and Trp261 aromatic ring and the side-chain protons of Pro250 and Pro263, illustrated by asterisks in figure 4.17. Additionally, long-range NOEs were observed in the ^{13}C-resolved NOESY-HSQC between Trp253 and Trp261, and side-chain protons of Lys249, Pro263 and Pro250. After approximately 10 Aria cycles and the assignment of the two tryptophan side-chains to their respective position and complete assignment of the proline side-chains, the hairpin was taking shape. The assignment of those residues was important for the folding of the structure as Trp253, Trp261, Lys249, Pro250, and Pro263 make multiple contacts (figure 4.17) and form an hydrophobic cluster.

The Aria output was converging for the hairpin region, but the two strands on the other side of the β-sheet caused some difficulties for the folding and had to be characterised further. This region contained very few long-range NOEs and was also containing two unassigned tryptophans (Trp236 and 275) and five proline (Pro228, 233, 239, 269, and 277) side-chains. The assignment of the two tryptophan side-chains to their respective position in the chain was the key, but this time the NOEs were much weaker. They include inter-residual NOEs between Trp275 H^{N}, H^{α}, and H^{β} and Trp236 $HN^{\varepsilon 1}$ as illustrated in figure 4.18 by the black lines. Those two tryptophan side-chains were also making contact to some of the proline side-chains, which helped for their assignment. In figure 4.18, the asterisks put in evidence the long-range NOEs between Trp253 and Trp261, and side-chain protons of Lys272 and Pro233. As also described in the previous case and to a lower extent, long-range NOEs in the ^{13}C-resolved NOESY-HSQC were observed between those residues. After another 10 Aria cycles, the long-range NOEs to the tryptophan side-chains were used after their assignment and the structure calculations converged towards another hydrophobic core with very similar features to the first described one. A second hydrophobic cluster involving residues ; Trp236, Trp275, Lys272 and Pro233 could be seen but was less well defined.

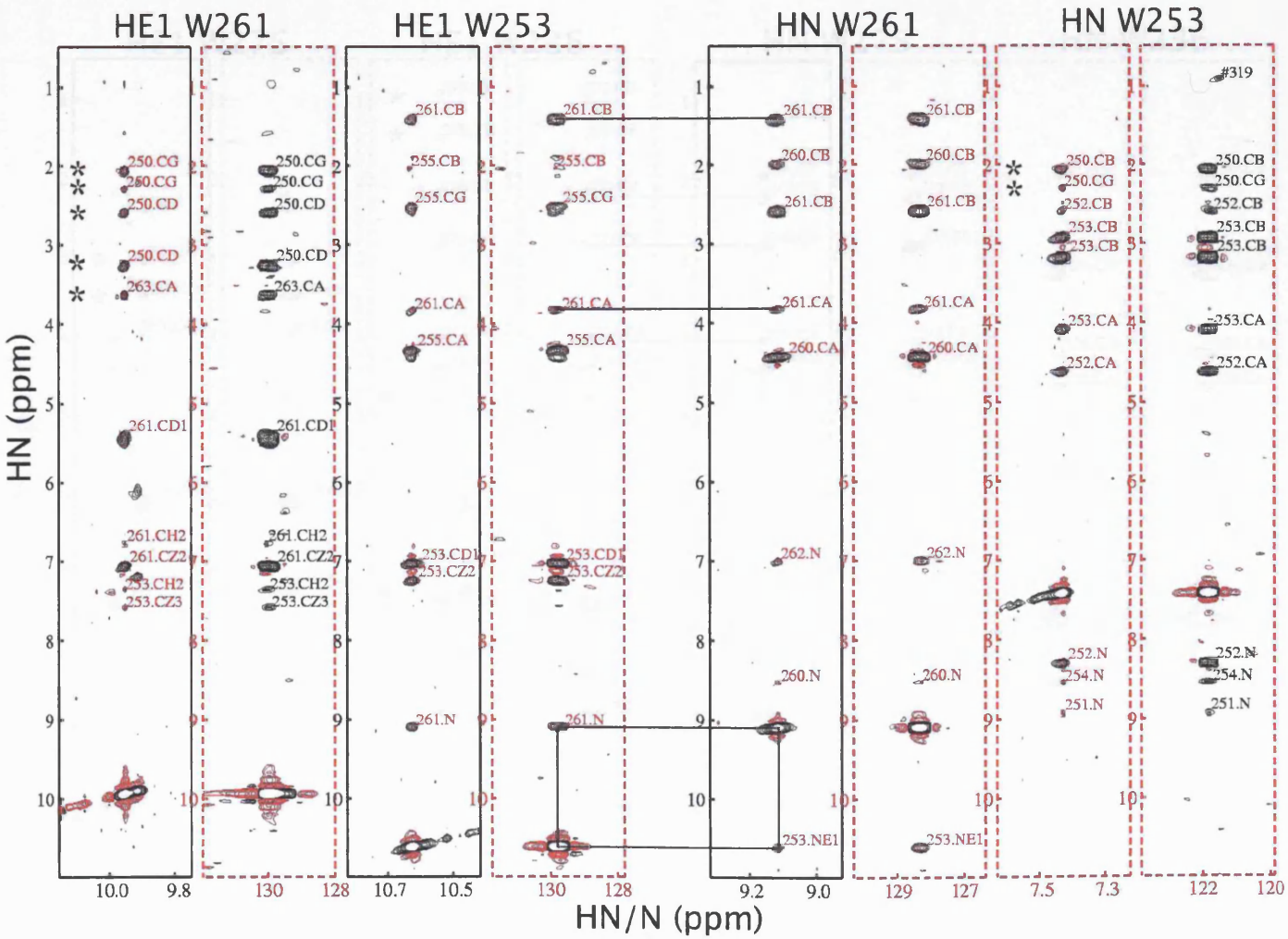

Figure 4.17 : Identification of the first hydrophobic core in the CRT S-domain. Spectral regions from 3D ^{15}N-resolved NOESY-HSQC showing strips for HN and HN$^{\varepsilon 1}$ of residues Trp261 and Trp253. The squares and lines connect the cross-peaks. The asterisks shows the long-range NOEs to Pro250 and Pro263.

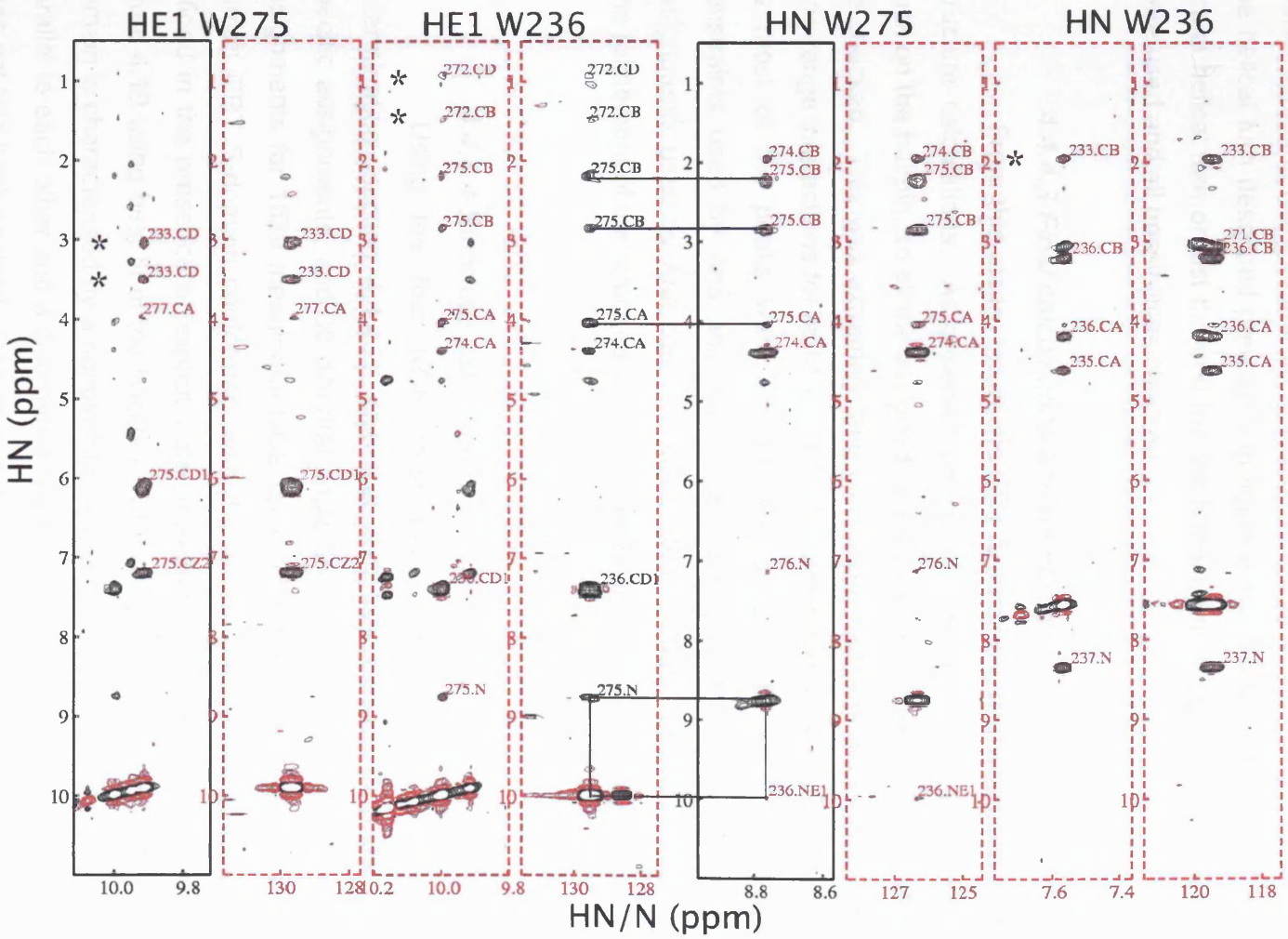

Figure 4.18 : Identification of the second hydrophobic core in the CRT S-domain. Spectral regions from 3D ^{15}N-resolved NOESY-HSQC showing strips for HN and HN$^{\varepsilon 1}$ of residues Trp236 and Trp275. The squares and lines connect the cross-peaks. The asterisks shows the long-range NOEs to P233 and K272.

The structure calculations started to converge after another 5 Aria cycles to a more compact hairpin, which helped to complete the proline side-chain assignment. At this stage the structures converged towards a hairpin, with a central β-sheet and two hydrophobic cores, one located on each side of this β-sheet. Some more interesting features started to appear at the tip of the hairpin, which looked like the helical turn described previously in figure 4.16. To further define if this was an actual helical turn or just the end the the hairpin, the assignments produced by Aria were used and all possibilities checked.

4.4.4.3 Final calculations and refinement

From this stage, the whole CRT S-domain chain was used for the ARIA structure calculations. As observed previously, the N-terminal tail seemed to fold back on the hairpin, so all the assigned peaks by Aria causing this fold to form had to be checked. This was extremely time consuming, but was necessary. The lack of long-range interactions for residues 163-230 suggested an N-terminal disordered tail as most of the peaks were intra-residual or $i\pm1$. By carefully analysing the constraints used by Aria and displaying the strips with NMRView, those wrong assignments used by Aria could be inactivated to finally form the 'reliable' folding. The refinement of the structure was achieved by another 20 Aria cycles.

4.4.4.4 Structure validation

Using the four NOE cross peak lists described in the structure determination (section 4.4.4.2), chemical shift lists derived from the sequence-specific assignments, and 30 dihedral angle constraints, the program ARIA yielded assignments for 1639 meaningful NOE upper distance limits (table IV.XI). The 20 human CRT S-domain conformers with the lowest energy structures were further refined in the presence of explicit water molecules using ARIA and superposed in figure 4.19 using best fit of the backbone atoms. The NMR structure of the CRT S-domain is characterised by a hairpin fold, where the two sides of the hairpin run anti-parallel to each other and a disordered N- and C-terminus (figure 4.19). The hairpin does not fold back on itself. One short anti-parallel β-sheet stabilises the structure.

Table IV.XI : Input of the structure calculation and characterisation of the energy-minimised NMR structures of CRT S-domain. Except for the top 5 entries, the data characterise the group of 20 conformers that is used to represent the NMR structure. The energies and the NOE violations have been extracted from the header of the 'pdb' file produced by ARIA, and the RMS difference of matched atoms in Angstrom (Å) have been obtained by running the XtalView 'lsq fit' command. The mean value and the standard deviation are given

Quantity	*Value*
Total NOE distance restraints	1639
Short-range (intra-residue)	1049
Medium-range (≤ 4)	433
Long-range (≥ 5)	157
Dihedral angle constraints	30
Energies (Kcal/Mol)	-2845 ± 300
NOE violations ≥ 0.5 Å	5.9 ± 1.1
RMS difference for matched atoms (Å)	
Backbone (242-266)	0.74 ± 0.22
Backbone (230-275)	3.69 ± 0.86
Heavy atoms (242-266)	1.54 ± 0.24
Heavy atoms (230-275)	4.46 ± 0.91

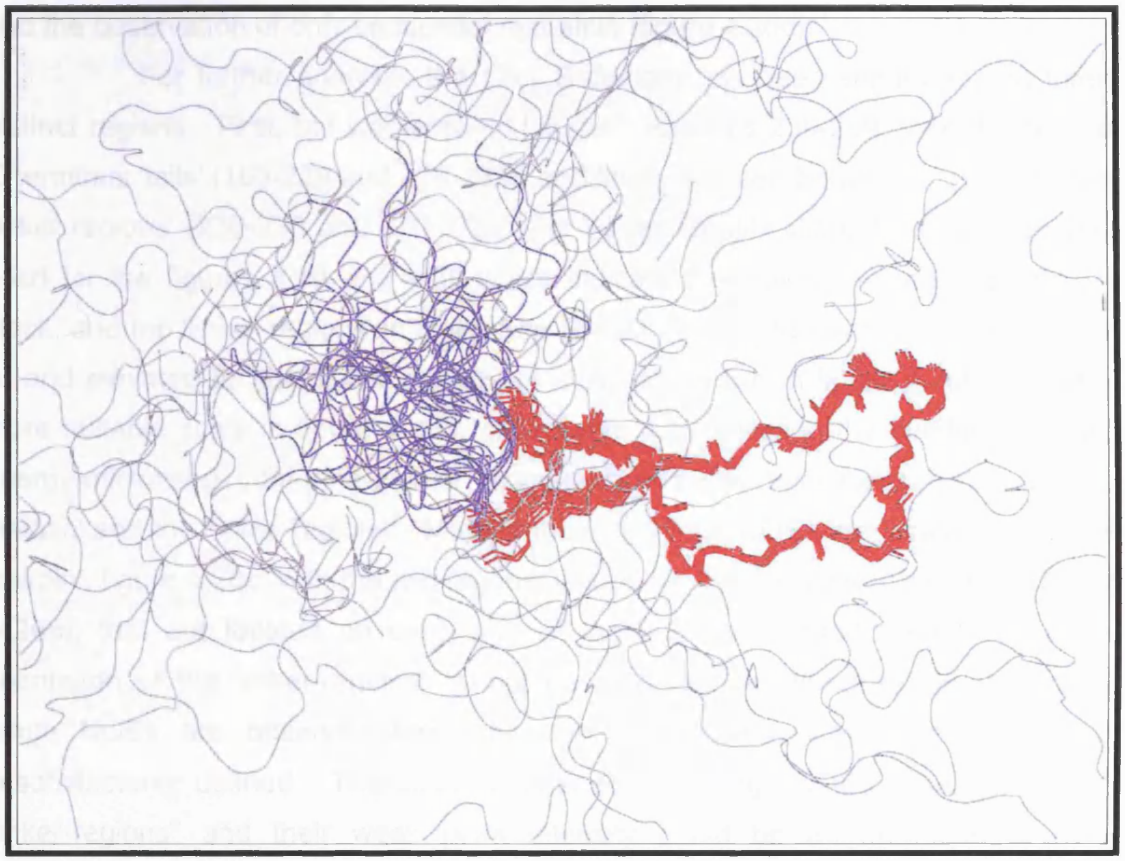

Figure 4.19 : Bundle of the 20 water refined energy-minimised conformers used to represent the NMR structure of CRT S-domain. Superposition of the polypeptide chain using best fit of the backbone atoms for residues 242-266 using CCP4i. The backbone residues 239-269 are shown in red (hairpin), residues 163-229 and 276-283 are represented as black ribbons (tails), and blue ribbons represent residues 230-238 and 270-275 (linker regions).

The final 20 CRT S-domain energy-minimised conformers shown in figure 4.19 have been analysed using Procheck NMR (figure 4.20). The CRT S-domain topology is summarised in figure 4.20a, with two short anti-parallel 3-residue β-strands and a 4-residue helical turn. In principle, the global hairpin arrangement of the polypeptide chain could accommodate extensive formation of anti-parallel β-sheet structure. However, long-range H^{α}-H^{α} NOEs indicative of anti-parallel β-sheet were found in the 3D ^{13}C-resolved [^{1}H,^{1}H]-NOESY-HSQC spectra only between the residue pairs (241, 267) and (243, 265), which define the short β-sheet depicted in figure 4.20a. The figure 4.20b shows on a residue per residue basis the Root-Mean-Square Deviation (RMSD) analysis for the main-chain (black) and side-chain (grey). High local RMSD values relative to the mean coordinates can be noticed for the long N-terminal tail (163-229), with values ranging from 15 to 80Å as anticipated from figure 4.19. This can be explained by the lack of long-range NOEs for that region

and the observation of only sequential restraints (figure 4.20c).

For further analysis, the CRT S-domain has been separated into three distinct regions. First, the well-ordered 'hairpin', residues 239-269, then the N- and C-terminus 'tails' (163-229 and 276-283), and finally situated between those two , the 'linker regions' (230-238 and 270-275). For easier identification, a colour code was used for the figures 4.19 and 4.20 which indicate the 'hairpin' in red, the 'tails' in black, and the 'linker-regions' in blue. The RMSD values are low for the 'hairpin' (<2 Å) and elevated for the 'linker regions' (<15 Å), as shown in figure 4.20b and on a more suitable scale in figure 4.20d. The figure 4.19 showing the bundle of the 20 energy-minimised conformers also represents this low backbone RMSD for the 'hairpin' and the 'linker-regions'. Most of the long-range NOEs are located within the 'hairpin', figure 4.20c, with the two highest values for the two Isoleucines (Ile242 and Ile266), that are located on each side of the β-sheet. Unfortunately, the exact orientation of the 'linker-regions' cannot be precisely described as only few long-range NOEs are observed between some side-chains, leaving the backbone unsatisfactorily defined. The small number of long-range NOEs observed for the "linker-regions" and their weak peak intensity could be an indication of slow conformational exchange. To some extent, this could explain why so few NOEs were observed as the broadening of the signals makes the NOEs hard to detect.

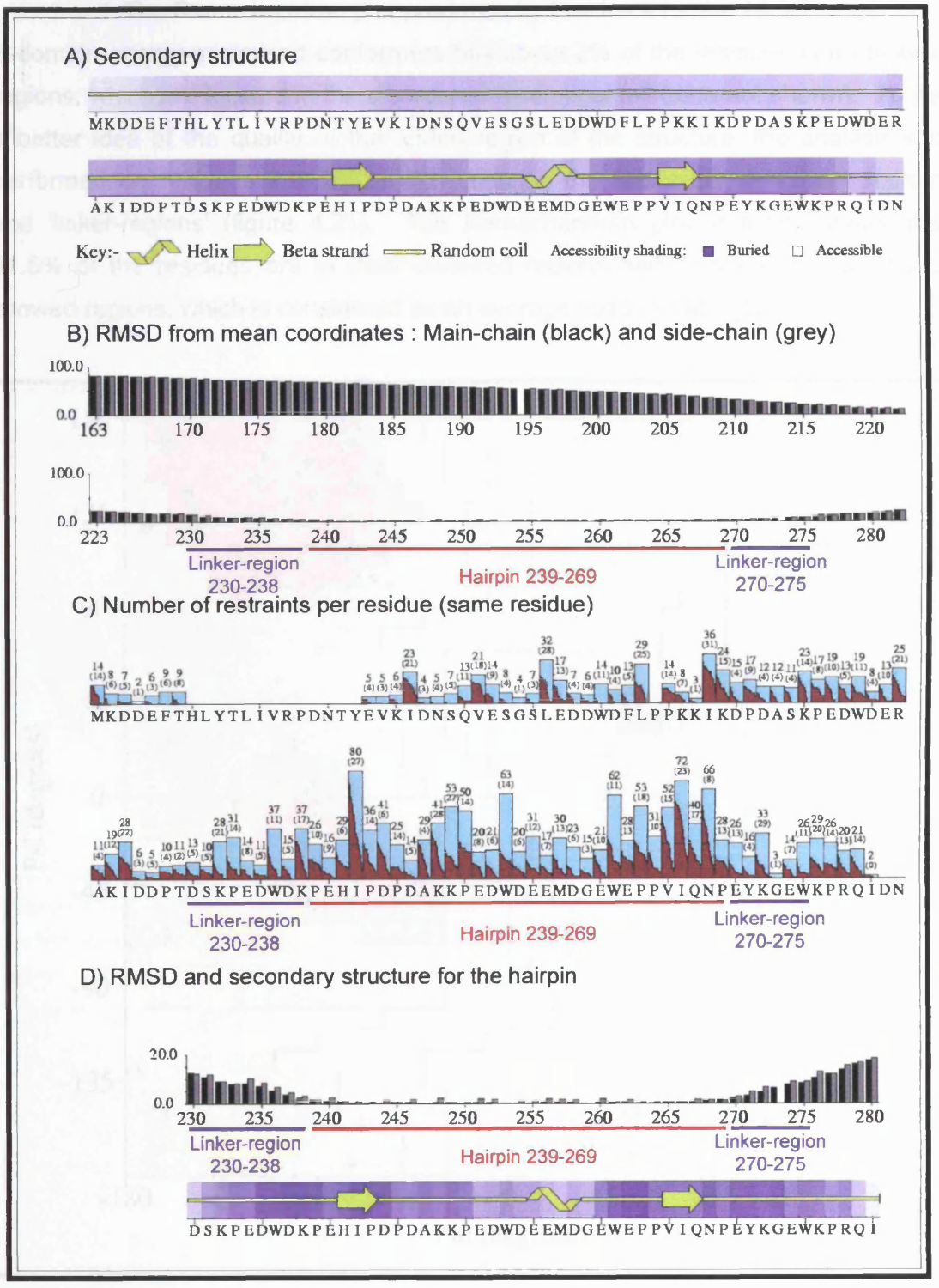

Figure 4.20 : Statistics on a residue per residue basis for the 20 energy-minimised conformers used to represent the NMR structure of the CRT S-domain. a) Secondary structure of the 3D structure. b) Average local RMSD for the backbone in black and the side-chain in grey. c) Number of NOE restraints per residue are displayed as blue bars and the intra-residual in brackets. d) Average local RMSD and secondary structure focused on residues 230-280. These analysis were performed using Procheck NMR.

The Ramachandran plot produced by Procheck for the 20 full length CRT S-domain energy-minimised conformers has about 2% of the residues in disallowed regions, which are located in the disordered N-terminal tail (data not shown). To get a better idea of the quality of the folded region of the structure, the analysis was performed on residues 230-280, which contains the previously described 'hairpin' and 'linker-regions' (figure 4.21). The Ramachandran plot statistics shows that 81.6% of the residues are in most favoured regions and 18.4% in the additional allowed regions, which is considered as an average quality model [122].

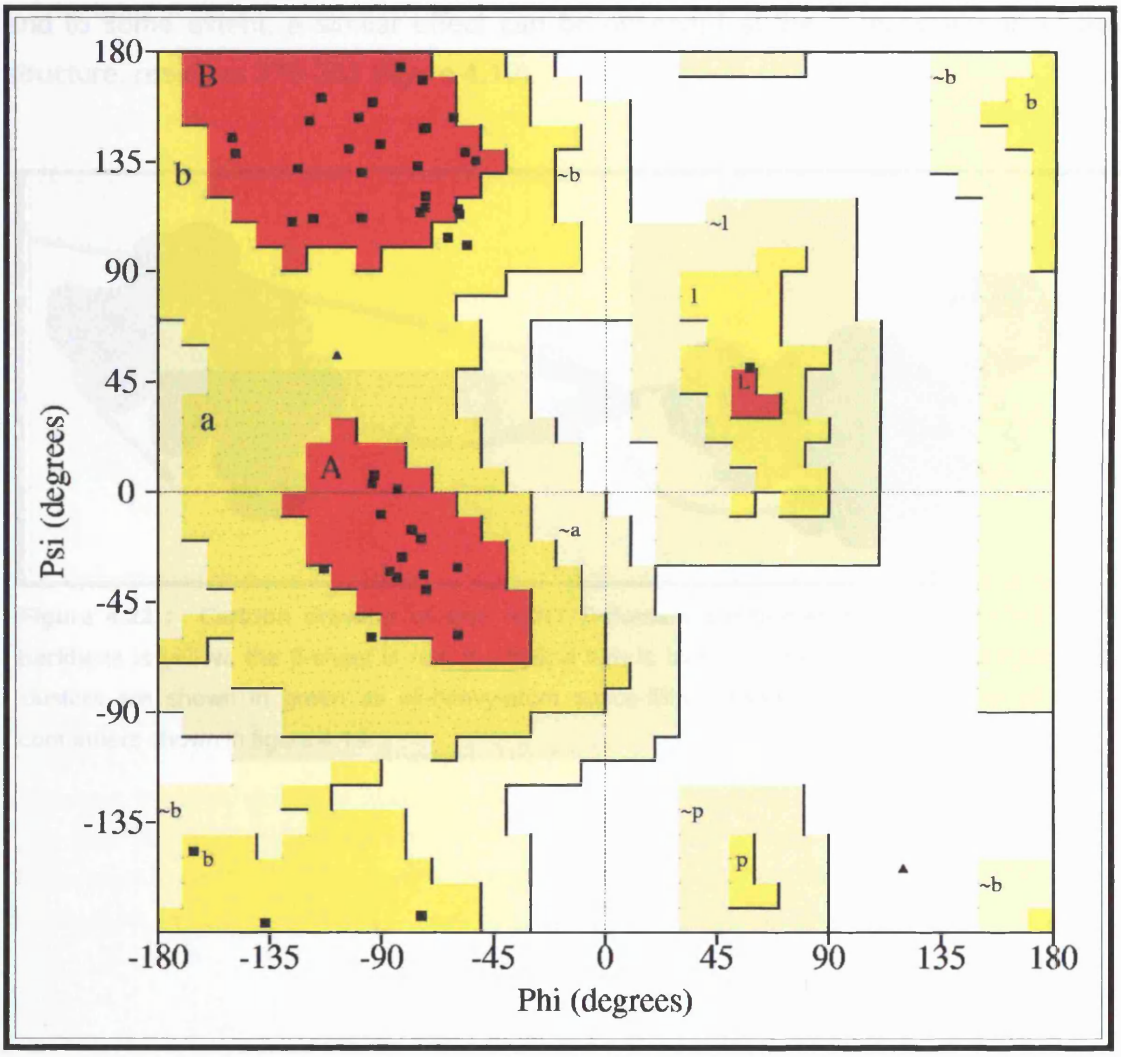

Figure 4.21 : Ramachandran plot for the residues 230-280 of the CRT S-domain. The analysis was performed using Procheck for the 51 residues, excluding 2 glycines (shown as triangles), 9 prolines, and 2 end-residues, for a total of 38 residues. The most favoured regions are shown in red (A,B,L), additional allowed regions in yellow (a,b,l,p), generously allowed regions in beige (~a,~b,~l,~p), and disallowed regions in white.

4.4.4.5 The human calreticulin S-domain structure

The NMR structure of the CRT S-domain is characterised by an extended hairpin fold. The non-classical hairpin is formed by a short anti-parallel β-sheet and a well-defined hydrophobic cluster involving the rings of two tryptophans side-chains stabilise the structure (figure 4.22). At the tip of the hairpin there is a rather loose single helical turn. At the location of the β-sheet the two hairpin strands cross over and then run anti-parallel to a second hydrophobic cluster, with very similar characteristics to the first one. The lack of long-range stabilising interactions for residues 163-230 goes along with the presence of an N-terminal disordered tail and to some extent, a similar effect can be observed at the C-terminal part of the structure, residues 276-283 (figure 4.19).

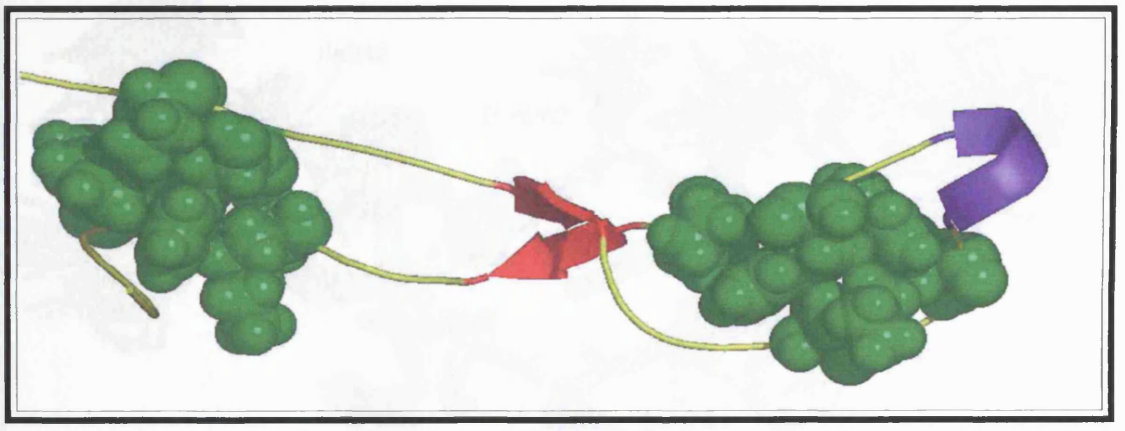

Figure 4.22 : Cartoon drawing of one hCRT S-domain conformer (residues 234-285). The backbone is yellow, the β-sheet is red, the helical turn is blue, and the residues of the hydrophobic clusters are shown in green as all-heavy-atom space-filling models. Drawing of one of the 20 conformers shown in figure 4.19.

In the hCRT S-domain structure the hairpin loop is the best defined part of the structure, as shown after superposition for best fit of the 20 energy-minimised conformers in figure 4.23. The backbone is well defined as previously described by the low RMSD (0.74 ± 0.22 Å) with more solvent exposed side-chains being less ordered (1.54 ± 0.24 Å) (see table IV.XI). The two isoleucine side-chains (Ile242 and 266) are located on each side of the β-sheet, where the highest number of NOE restraints have been observed (figure 4.20c). However, the definition of the structure is less in the 'linker-regions', which are located on the left hand side of those two isoleucines in figure 4.23. It can also be noticed that some of the buried side-chains like Pro263, Trp261, and Lys249 are well defined.

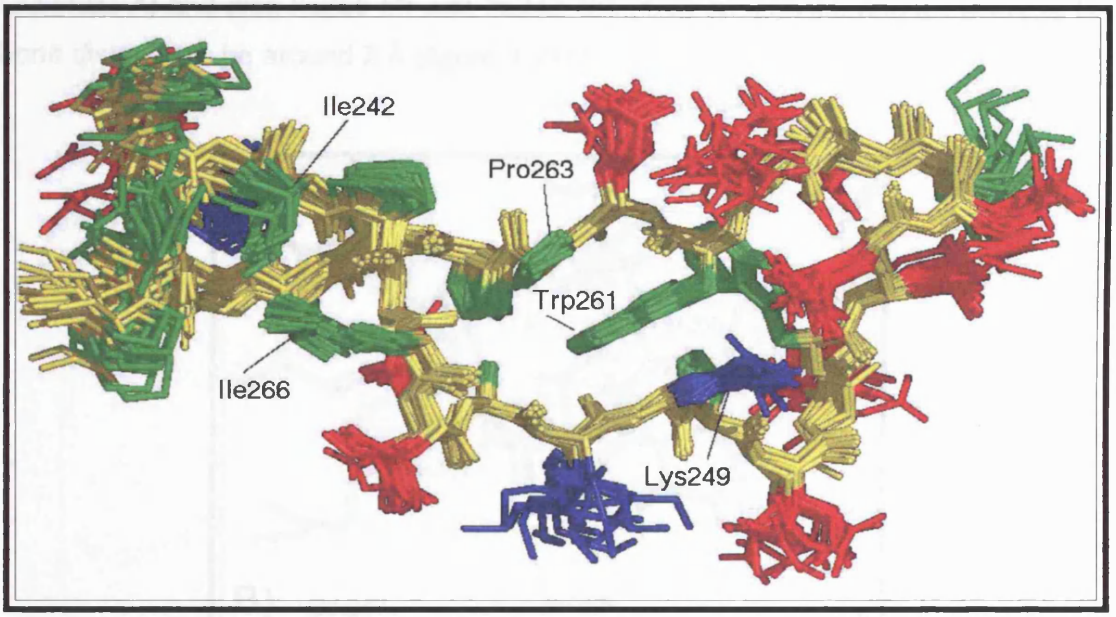

Figure 4.23 : Bundle of the 20 energy-minimised conformers for residues 239-269 used to represent the NMR structure of CRT S-domain hairpin. Superposition of the polypeptide chain for best fit of the backbone has been performed for residues 242-266, using CCP4i. The backbone is represented in yellow, positively charged residues are in blue, negatively charged residues are red, and hydrophobic and polar residues are green. The positions of selected residues are indicated.

The two strands of the hairpin cross over and interact along a short anti-parallel 3 residue β-sheet (figure 4.24a). This short β-sheet was identified by two H^{α}-H^{α} interactions between residue pairs (His241, Gln267) and (Pro243, Val265), and H^{N}-H^{N} interaction between (Ile242 and Ile266) (see figure 4.15). The mean distance for the 20 energy-minimised conformers used to represent the NMR structure of CRT S-domain, are 2.8Å for H^{α}-H^{α} of residue pair (241, 267), 2.6Å for H^{N}-H^{N} of (242, 266), and 2.5Å for H^{α}-H^{α} of residue (243, 265), as shown in figure 4.24a by red dotted lines. Also, the H^{α}-H^{N} (*i+1*) NOEs are intense for this region, while the H^{α}-H^{N} intra-residue interactions are weak or absent. This comes in agreement with the typical arrangement of the backbone residues for an anti-parallel β-sheet. Moreover, possible hydrogen bonding could form between residue pairs Ile242 H^{N} with Ile266 CO (1.92 Å) and also Ile266 H^{N} with Ile242 CO (1.75 Å), considering an average H-bond distance to be around 2 Å (figure 4.24b).

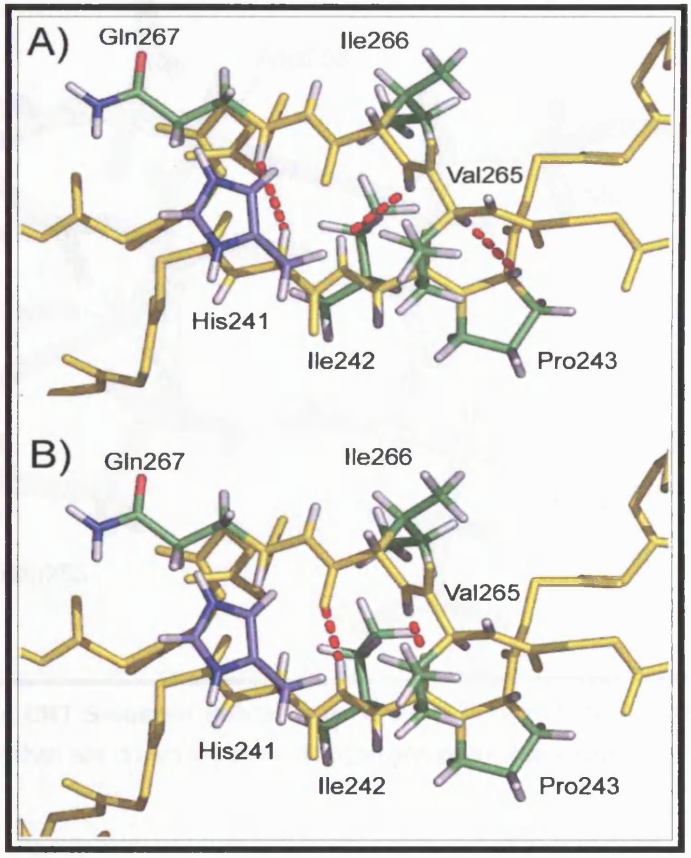

Figure 4.24 : The CRT S-domain anti-parallel β-sheet. A) The red dotted lines show the NOEs between ; H^{α}-H^{α} of residue 241 and 267, HN-HN of 242 and 266, and H^{α}-H^{α} of residue 243 and 265. B) The red dotted lines show the potential hydrogen bonds between Ile242 and Ile266. Drawing of one of the 20 conformers shown in figure 4.19. The residue side-chains are colour-coded as in figure 4.23.

At the tip of the hairpin a single helical turn is present, as shown in figure 4.25. This region looks similar to a helix but without having the appropriate hydrogen bonds. From the Procheck-NMR analysis (figure 4.20d), only the residues 255 to 258 have the right ϕ and ψ angles to be part of an α-helix. The Asp254 and Gly259 side-chain seems to be going out of the turn, but the H^N group of Gly259 is in the right position to create an hydrogen bond with the carboxy group of Asp254, which would indicate the presence of an unusual π-helix. The π-helix has 4.4 residues and 16 atoms per turn and is thus called the 4.4_{16} helix [122]. However, in the CRT S-domain structure the backbone H^N of Gly259 is too distant from the CO of Asp254 to form an hydrogen bond (2.7 Å apart). As no hydrogen bond forms to stabilise the helical turn in a closed conformation, it can be considered as a 'loose' helical turn rather than to a specific type of helix.

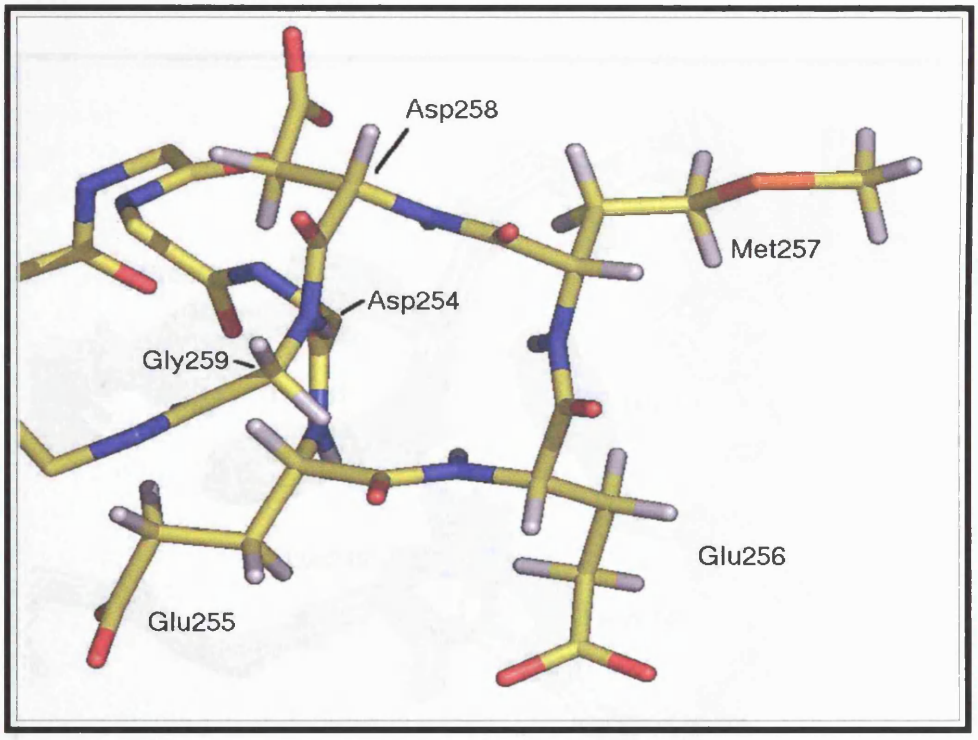

Figure 4.25 : The CRT S-domain helical turn. Drawing of one of the 20 conformers shown in figure 4.19. The carbon are drawn in yellow, nitrogen groups in blue, oxygen in red and hydrogen in grey.

In the structure, the two strands of the hairpin make contacts not only along the short β-sheet, but also through side-chain interactions in two hydrophobic clusters (figure 4.22). Each cluster involves two tryptophan residues, one from each side of the hairpin, and also share a similar arrangement of the side-chains (figures 4.26 and 4.27).

The two indole rings of residues Trp253 and Trp261 are packed against the aliphatic side-chains of Pro250, and Pro263, and the Lys249 side-chains, forming the first hydrophobic core (figure 4.26). Those two conserved tryptophan side-chains are the main driving force for this region of the protein as there is no evidence of H-bond formation into the periphery of this hydrophobic core to either the backbone or side-chains. The backbone is well defined due to the position of the two tryptophan and proline side-chains restricting any large amplitude motion for the backbone.

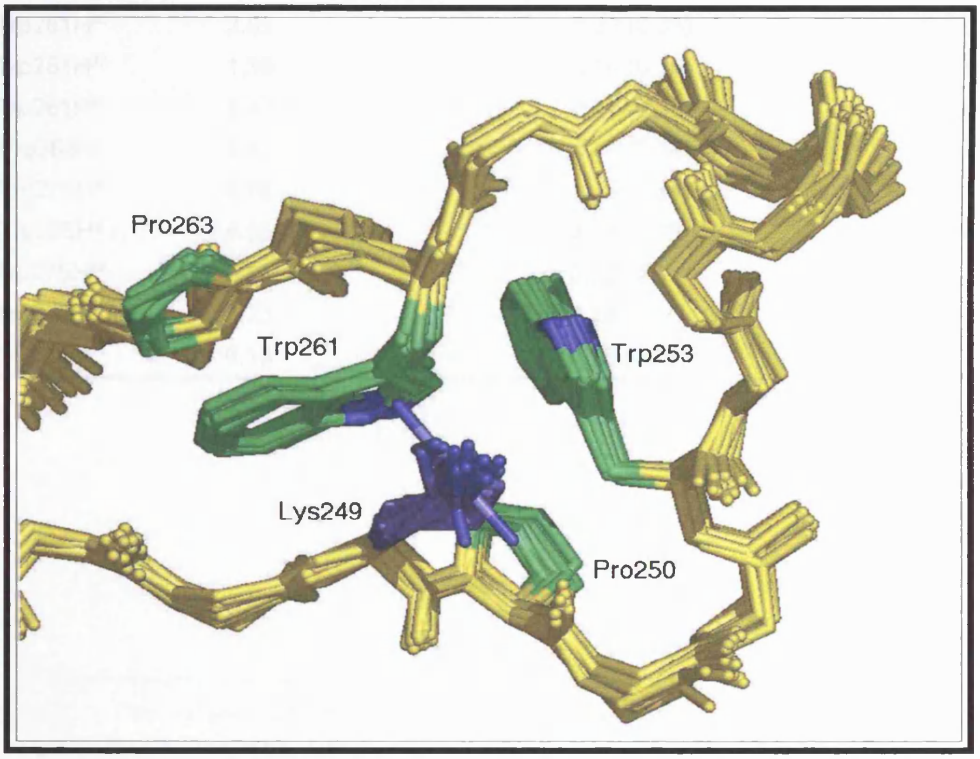

Figure 4.26 : The first hydrophobic core in the CRT S-domain. Close-up view showing the side-chain arrangement of the residues Lys249, Pro250, Trp253, Trp261 and Pro263. The residue side-chains are colour-coded as in figure 4.23. Superposition of the polypeptide chain for best fit of the backbone atoms N, C^{α} and CO has been performed for residues 242-266, using CCP4i.

Moreover, this structural arrangement gives rise to unusual ^1H chemical shifts. For example, the Lys249 H^{β} proton chemical shifts are 0.99 and 0.42 ppm, as compared with a typical value near 1.8 ppm, and the $H^{\delta 1}$ resonance of Trp261 is shifted from a typical value near 7.16 ppm to 5.47 ppm (table IV.XII). All these unusual chemical shifts can be explained by ring current effects from the nearby indole rings in the hydrophobic cluster.

Table IV.XII : Ring current shifts in the CRT S-domain. Observed chemical shift, the average shift and their standard deviation used from the BMRB database (http://www.bmrb.wisc.edu), and the deviation from the average shift value.

^1H atom	*Observed shift (ppm)*	*Average shift (ppm) (standard deviation)*	*Δδ (ppm)*
Lys249$H^{\beta 2}$	0.99	1.79 (0.25)	-0.8
Lys249$H^{\beta 3}$	0.42	1.78 (0.26)	-1.36
Trp261H^{N}	9.11	8.35 (0.84)	0.76
Trp261H^{α}	3.84	4.74 (0.55)	-0.9
Trp261$H^{\beta 2}$	2.59	3.22 (0.36)	-0.63
Trp261$H^{\beta 3}$	1.39	3.18 (0.36)	-1.79
Trp261$H^{\delta 1}$	5.47	7.16 (0.34)	-1.69
Pro263H^{α}	3.61	4.41 (0.36)	-0.8
Trp275H^{N}	8.76	8.35 (0.84)	0.41
Trp275H^{α}	4.05	4.74 (0.55)	-0.69
Trp275$H^{\beta 2}$	2.82	3.22 (0.36)	-0.4
Trp275$H^{\beta 3}$	2.23	3.18 (0.36)	-0.95
Trp275$H^{\delta 1}$	6.18	7.16 (0.34)	-0.98

Similarly, a second hydrophobic cluster is present in the 'linker region' involving the indole rings of two tryptophans (Trp236 and Trp275), and the side-chains of Lys272, and Pro233 (figure 4.27). As in the previous hydrophobic cluster, this structural arrangement gives rise to unusual ^{1}H chemical shifts (table IV.XII) and can be explained by ring current effects from the nearby indole rings in the hydrophobic cluster. However, this 'linker region' is less well defined. Only few long-range NOEs could be observed in this hydrophobic cluster (see figure 4.17); these NOEs mainly involve the four side-chains described in figure 4.27. Moreover, the lack of stabilising long-range NOEs for the backbone in this region and the presence of a long disordered N-terminal tail further compromise the definition of this hydrophobic cluster in the structure. The main stabilising force for the "linker-regions" is this hydrophobic cluster which form between the two opposite strands.

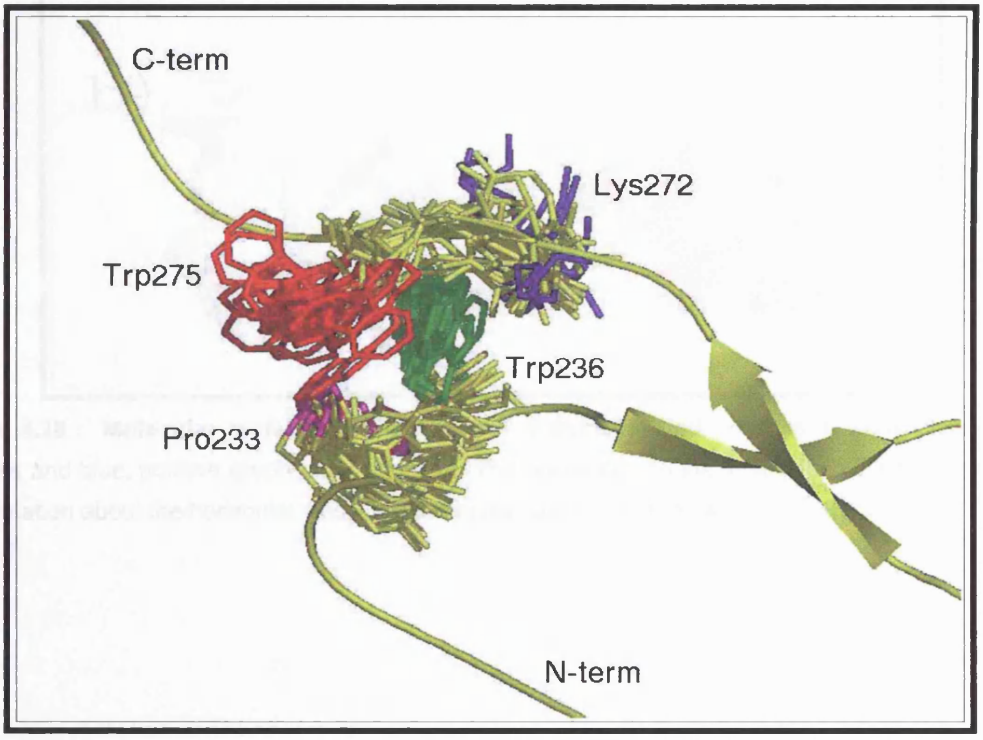

Figure 4.27 : The second hydrophobic core in the CRT S-domain. Close-up view showing the side-chain arrangement of the residues Lys272 (blue), Pro233 (violet), Trp275 (red), and Trp236 (green). Superposition of the polypeptide chain for best fit of the side-chains has been performed for residues 233-236 and 272-275, using CCP4i.

The molecular surface of the CRT S-domain, figure 4.28, shows a pronounced negative charge. The tip of the hairpin loop has an accentuated negative charge, due to the high content of aspartic and glutamic acid (shown in red in figure 4.23).

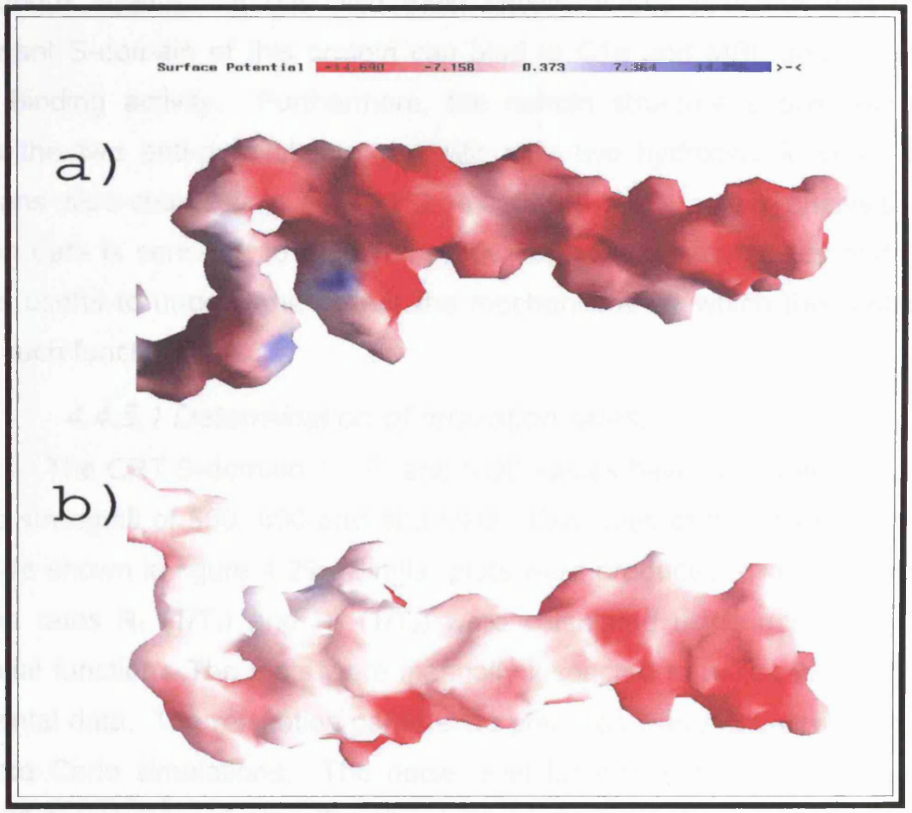

Figure 4.28 : Molecular surface view of the CRT S-domain. Red indicates negative electrostatic charges and blue, positive electrostatic charges. The orientation on the molecule in b was obtained by 180° rotation about the horizontal axis. The molecular surface was drawn using GRASP [123].

4.4.5 Relaxation

The structure of the calreticulin S-domain has special features that are not commonly found in other proteins. In the first instance, CRT is a molecular chaperone that can bind to multiple ligands and indeed must be able to adapt to those various ligands. It has also been shown in this research that the small recombinant S-domain of this protein can bind to C1q and MBL and therefore has kept its binding activity. Furthermore, the hairpin structure shows few contacts between the two anti-parallel strands, with only two hydrophobic cores involving tryptophans side-chains and a short 3 residues β-sheet. The analysis of NMR relaxation data is sensitive to motional processes in macromolecules and therefore becomes useful to understand further the mechanisms by which this molecule can achieve such functions.

4.4.5.1 Determination of relaxation rates.

The CRT S-domain T_1, T_2, and NOE values have been measured at 25°C with field strengths of 500, 600 and 800 MHz. Examples of the observed relaxation decays are shown in figure 4.29. Similar plots were produced for each residue from which the rates R_1 ($1/T_1$) and R_2 ($1/T_2$) were calculated using non-linear fit to an exponential function. The plots were manually inspected to asses the quality of the experimental data. The relaxation parameters precision was evaluated using a set of 100 Monte Carlo simulations. The noise level for different signals is indicated in figure 4.29 by the error bars. The figure shows the non-linear fit at 800 MHz for Trp200, which is located in the N-terminal tail of the structure, and for Ile266, located in the β-sheet. Most of the observed relaxation curves were close to one of these examples. The first example, Trp200, very sharp peaks were observed in the spectra, giving high intensity values that did not relax completely to values close to zero after the longest delays in the T_1 and T_2 experiments. In the example of Ile266, the spectra show broad peaks with small intensity which relax close to zero after the long delays. The non-linear fits were contained within the experimental error as shown in figure 4.29. The noise level reached 10% of the intensity of the the first delay point for the weak peaks and 0.6% for the intense peaks. The duplicated delays also verified the quality of the data and reflected the experimental error in most cases, i.e both points were contained within the error. Some weak peaks show a large error after long delays, for example on the T_2 plot of Ile266, values can reach 0.02 ± 0.04. However, the important points for the determination of the rates are not those close to zero but those located in the higher part of the curve and the quality of

the non-linear fits were satisfying for those regions. It has to be kept in mind that the non-linear fits shown in these plots do not represent the full analysis which contained a 100 Monte Carlo simulations.

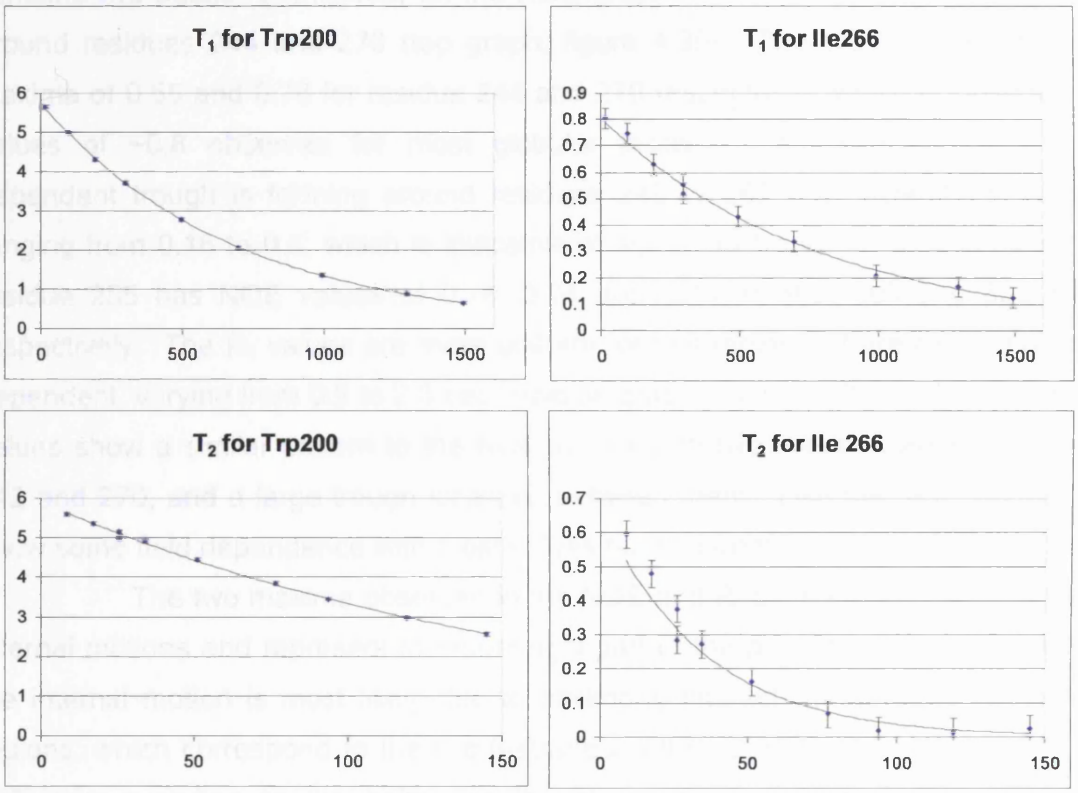

Figure 4.29 : Peak intensity versus delay in milliseconds. Example of the non-linear fit for residues Trp200 and Ile266 from which the rate constants have been obtained at 800 MHz. The noise level in the spectra is ±0.04 and shown by the error bars in the figures. The fourth and third delay were duplicated for the T_1 and T_2 respectively as described in the material and methods.

4.4.5.2 Relaxation rate analysis

The experimental results are summarised in figure 4.30 which shows the NOE and relaxation rates values (R_1 and R_2) on a residue per residue basis for the three fields used. A more detailed analysis, including error bars reflecting the experimental error, is also described in the appendix figures X.4 to X.6.

Three distinct regions can be identified from the relaxation data shown in figure 4.30. First, for residues 163 to 225, regular R_1 values ranging from about 1.3 to 1.8 sec^{-1}, more uniform R_2 values from 2.6 to 12 sec^{-1}, and extremely low NOEs with variation from -1.0 to 0.37 were observed. Relaxation data obtained at three different magnetic fields showed little field dependence of R_1 and R_2 values for this region, in contrast with the large field dependence for the NOE values, typical of an

unfolded protein [124]. This agrees with the absence of long-range NOEs and a high rmsd for residues 163-229, clearly indicating that the N-terminal tail is unfolded.

The second distinct region is localised around residues 240 to 274, and contains two peaks, or maxima, on the NOE graph with the highest values centred around residues 244 and 270 (top graph, figure 4.30). The NOE values reach a maxima of 0.65 and 0.78 for residue 244 and 270 respectively, which is close to the values of ~0.8 observed for most globular proteins. However, a more field dependent trough is forming around residues 248 to 262 with lower NOE values ranging from 0.16 to 0.5, which is indicative of some flexibility. As an example, the residue 255 has NOE values of 0.16, 0.24 and 0.44 at 500, 600 and 800 MHz respectively. The R_1 values are more uniform for this region but are extremely field dependent, varying from 0.9 to 2.0 sec^{-1} (middle graph, figure 4.30). However, the R_2 values show a similar pattern to the NOE values with two maxima around residues 242 and 270, and a large trough localised between them. Like the NOE values, R_2 show some field dependence with most values ranging from 8 to 31 sec^{-1}.

The two maxima observed in the NOE and R_2 plots are indicative of slow internal motions and represent the most rigid part of the protein. The restrictions in the internal motion is most likely due to backbone interactions between those two regions, which correspond to the two β-strands in the structure. The region 248 to 262, corresponding to the hairpin in the structure, shows interesting relaxation properties. The NOEs have intermediate values, indicative of some flexibility, while the presence of an hydrophobic core, a high amount of long-range NOEs and a low local rmsd corresponds to a well-defined structure. Those intermediate NOE values in the helical turn located at the tip of the hairpin, residues 254 to 259, further support the absence of stabilising interactions in the helical turn. Also, long-range NOEs indicative of possible hydrogen bonding between residues Asp254 and Gly259 could not be observed during the structure determination (figure 4.25). These data put together suggest the presence of a 'loose' helical turn at the tip of the hairpin, which in fact is flexible compared to typical helical structures due to the lack of stabilising interactions. Interestingly, the two tryptophans located in the core of the hairpin, Trp253 and 261, have an NOE value of 0.36 ± 0.09 for their side-chain $HN^{\varepsilon 1}$ group (appendix figure X.7). This latter observation suggest more restrictions for those two side-chains, which are located within the hydrophobic core of the 'hairpin'. Those two observations indicate that this unusual 'hairpin' structure has little backbone restrictions with a more restricted core, which also correlates with extensive interactions in the hydrophobic core.

Finally, the third region is composed of two short residue stretches with similar features. The first stretch is localised between the two previously described regions, residues 226 to 239 and the second one at the C-terminus of the protein, residues 275 to 280 (figure 4.30). These regions are 'linkers' between the region containing the highest NOE and R_2 values (residues 240 to 274) to either the long disordered N-terminus or short C-terminus tail. The data for these regions shows a transition from large to small amplitude motions. The NOE values change almost linearly from roughly 0.1 to 0.5 for the first stretch and from 0.5 to -0.1 for the second one. A similar effect is clearly observed for the R_2 values, with more field sensitive and variable R_1 values. The data indicate that those regions are also flexible, but to a lower extent in comparison to the N-terminal tail, indicating that these regions could serve as a 'linker regions' in this recombinant protein.

The above two stretches are on the opposite side of the hairpin in the structure and almost overlap with the residues in the vicinity of the second hydrophobic core (residues 230-238 and 270-275) which was also called the 'linker region' in the structure description (shown in blue in figure 4.19). The intermediate NOE and R_2 values suggest the presence of some backbone flexibility for this region. However, the NOE values for the two tryptophan side-chains are 0.33 ± 0.06 (appendix, figure X.7), indicating that the side-chains are less dynamic than the backbone. This can be explained with the structure (figure 4.27), as the two tryptophan side-chains make contacts within an hydrophobic core with the backbone located on either side of this core. Those side-chain interactions between the two opposite strands are strong enough to keep the two opposite strands in close proximity and leave the backbone more motional freedom, because they are too distant from each other to make stabilising contacts (also well described by figure 4.27). Moreover, few long-range NOEs could be assigned to the side-chains during the structure determination and practically none to the backbone, also supporting the absence of stabilising interactions for the backbone in this region of the protein. The mean backbone local rmsd for this region is about 10Å, also supporting this flexibility.

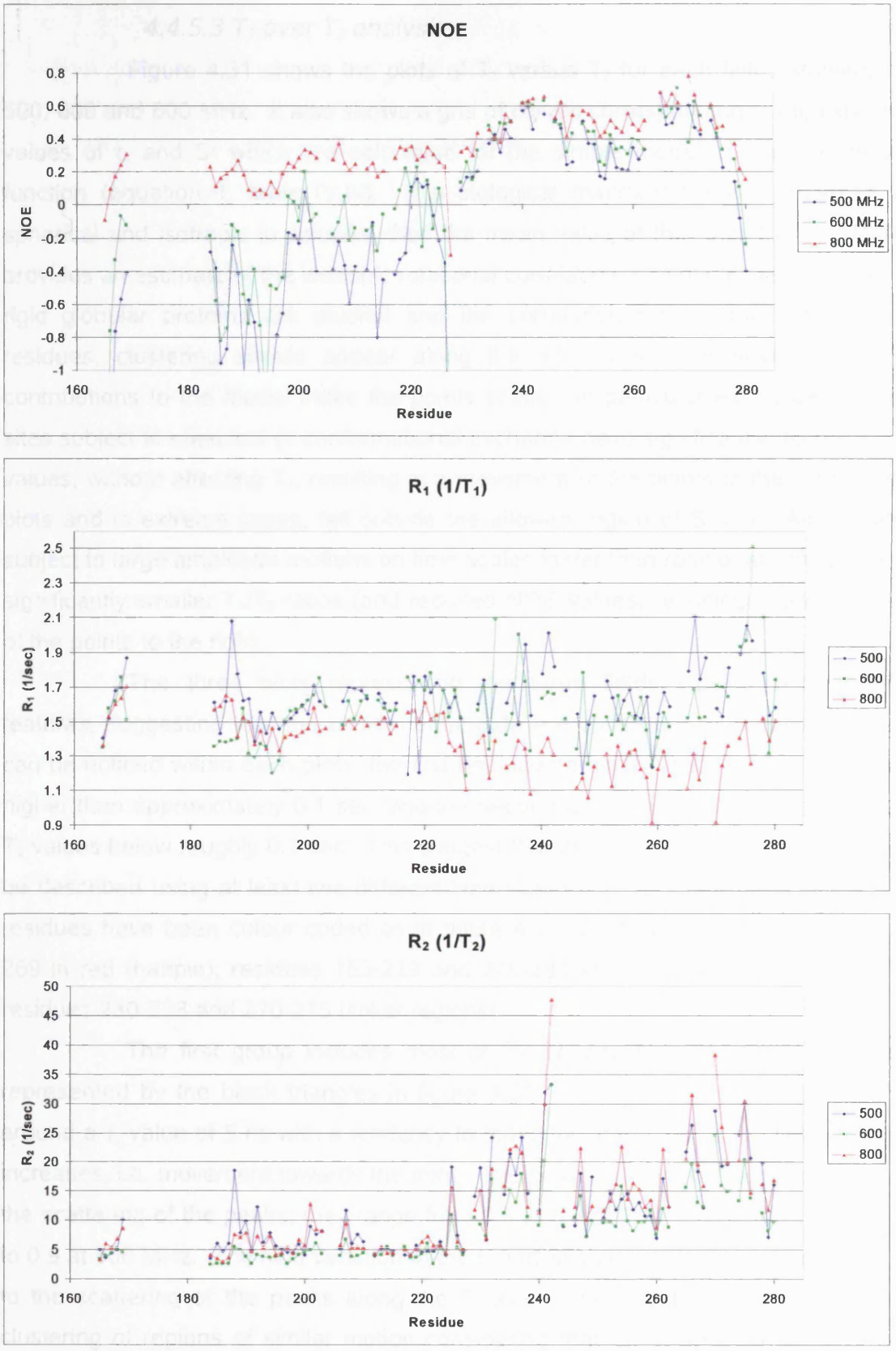

Figure 4.30 : ^{15}N relaxation data for the hCRT S-domain measured at 25 °C and at ^{1}H frequencies of 500 MHz (Blue), 600 MHz (green), and 800 MHz (red). The same data are shown in the appendix per field with error bars.

4.4.5.3 T_1 over T_2 analysis

Figure 4.31 shows the plots of T_1 versus T_2 for each fields studied, i.e. 500, 600 and 800 MHz. It also shows a grid of curves corresponding to the expected values of τ_c and S^2 which are calculated for the simple model of spectral density function (equation 1, table IV.III). If a biological macromolecule is regarded as spherical and isotropic in solution, then the mean value of the ratio T_1/T_2 (or R_2/R_1) provides an estimate of the isotropic rotational correlation time (τ_c) of the molecule. If rigid globular proteins are studied and the correlation time is the same for all residues, clustering should appear along the iso-τ_c lines. However, additional contributions to the model make the points scatter, in particular exchange. Atomic sites subject to chemical or conformational exchange have significantly decreased T_2 values, without affecting T_1, resulting in a movement of the points to the left in those plots and in extreme cases, fall outside the allowed region of $S^2 > 1$. Atomic sites subject to large amplitude motions on time scales faster than rotational diffusion yield significantly smaller T_1/T_2 ratios (and reduced NOE values) resulting in a movement of the points to the right.

The three plots representing the three fields studied show similar features, suggesting that the experimental data are consistent. Two main clusters can be noticed within each plots; the first one situated in the middle, or at T_2 values higher than approximately 0.1 sec, and the second group on the left hand side with T_2 values below roughly 0.1 sec. This suggest that the CRT S-domain motion has to be described using at least two different time scales. To facilitate the analysis, the residues have been colour coded as in figure 4.19 which shows the residues 239-269 in red (hairpin), residues 163-229 and 276-283 in black (tails), and in blue for residues 230-238 and 270-275 (linker regions).

The first group includes most of the N- and C-terminal tails residues, represented by the black triangles in figure 4.31. This group is clustering roughly around a τ_c value of 5 ns with a tendency to move towards 2 ns as the field strength increases, i.e. movement towards the right. The S^2 values vary in the plots reflecting the scattering of the peaks; they range from 0.5 to 0.8 at 500 and 600 MHz, and 0.6 to 0.9 at 800 MHz. The field variation of the τ_c and S^2 values seems to be mainly due to the scattering of the peaks along the T_2 axis. These data are consistent with clustering of regions of similar motion considering that the N-terminal tail is clearly unfolded.

More interestingly, the second group contains most of the folded residues, represented by red and blue peaks in figure 4.31. This group shows scattering of the peaks in the T_1 axis with a strong field dependence, as observed in the R_1 plot in figure 4.30 (residues 230-275). The T_2 values are more uniform and show little field dependence. However, this group is clustering in an area close to a correlation time (τ_c) of approximately 10 ns for the three fields studied (the third green line from the left representing 10 ns in figure 4.31). The scattering of this group to the left of the plot, small T_2 values, indicates the presence of a strong exchange contribution, which reduce T_2 values without affecting T_1 values. Surprisingly, most of these peaks are also located outside the allowed value for the order parameter (S^2) = 1, red line on the left in figure 4.31, which can be explained by the small T_2 values, likely due to exchange contribution. These data suggest that the 'hairpin' region of the CRT S-domain has a limited motional freedom of the backbone amide vectors with a strong exchange contribution.

As follows from the analysis of T_1/T_2, the simple model of spectral density function, equation 1, (see table IV.III) does not describe the data. The motions present in the CRT S-domain are more complex. Many peaks fall outside the grid of curves corresponding to the expected values of S^2 (>1) which are calculated for the simple model of spectral density function and the τ_c values obtained vary as function of the field (figure 4.31). However, some useful information can be extracted from this simple analysis. Two main groups seems to emerge, first, a flexible region in the area of low S^2 and low τ_c (black triangles). The second group, the folded part of the structure, is quite separate but mainly pushed away into the S^2 forbidden region (>1), most likely due to exchange contribution. More complex models are required to describe the spectral density function.

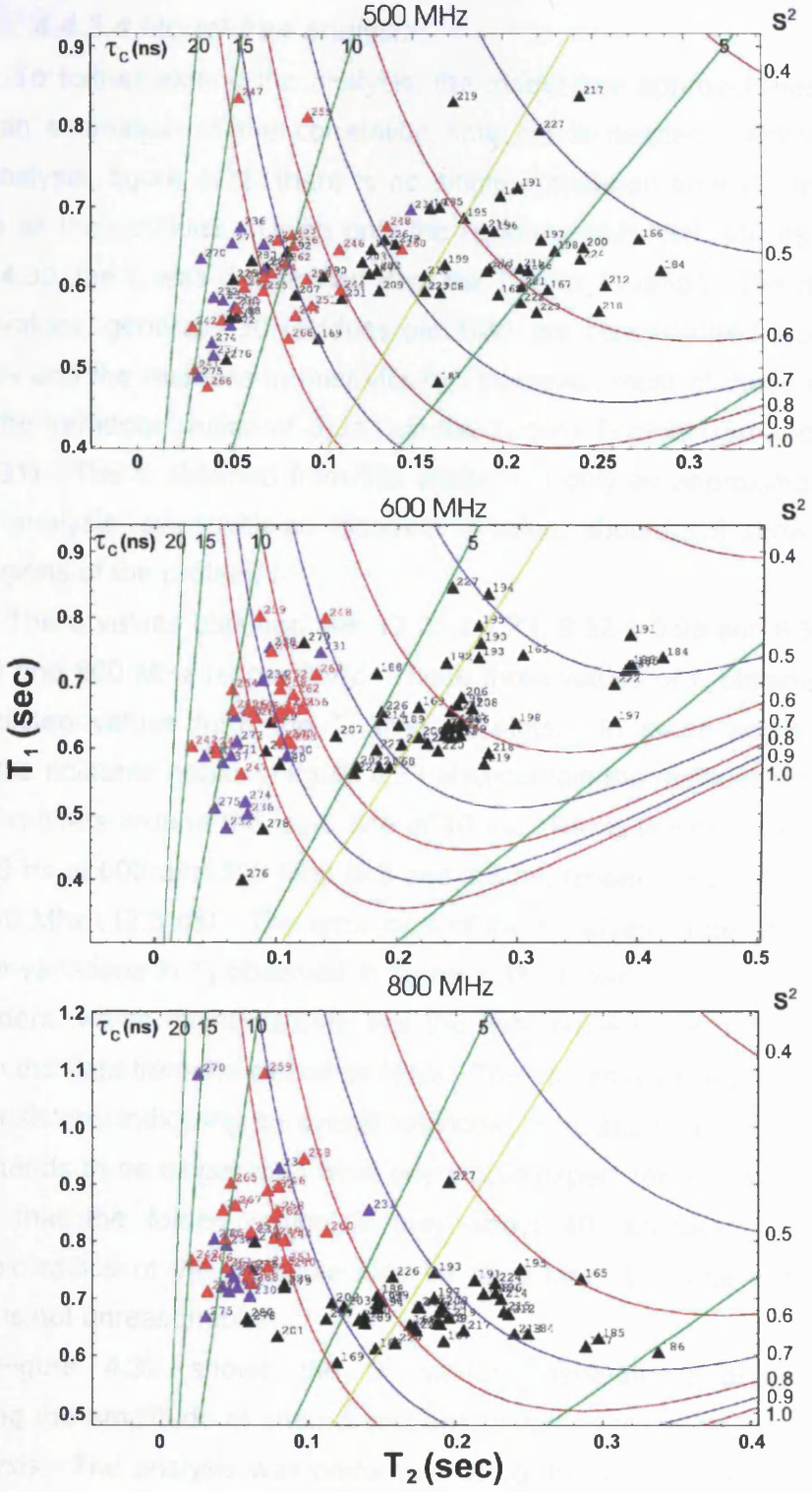

Figure 4.31 : 2D plots of T_1 against T_2 values for the three fields studied. It overlays a grid of curves corresponding to expected values of τ_c and S^2 which were calculated from the simplest model of spectral density. It plots alternating in red and blue curves at constant S^2 in steps of 0.1 from 0.4 (top right) to 1.0 (low left). In addition lines corresponding to constant τ_c at different values of S^2 are plotted in green for values of 2, 5, 10, 15 & 20 ns. The triangles colour reflects the regions described in figure 4.19.

4.4.5.4 Model-free analysis

To further extend the analysis, the model-free approach has been used for which an estimation of the correlation time (τ_c) is needed. According to the previous analysis, figure 4.31, there is no single correlation time (τ_c) that could be ascribed to all the residues. Using only the residues with NOE values higher than 0.4, figure 4.30, the τ_c was determined from the T_1 over T_2 ratios. The residues with high NOE values, generally 30 residues per field, are concentrated around the two beta strands and the residues in their vicinity. However, most of these residues are located in the forbidden region of S^2 (>1) in the T_1 over T_2 plots (red and blue peaks in figure 4.31). The τ_c obtained from this analysis is only an approximation, but the model-free analysis, even with an incorrect τ_c value, should still show differences between regions of the protein.

The τ_c values obtained are; 12.25 ± 1.73, 8.52 ± 0.98 and 8.80 ± 1.24 ns at 500, 600 and 800 MHz respectively. These three values of τ_c obtained are close to the predicted values from the T_1 over T_2 plots. In other words, the group containing the coloured peaks in figure 4.31 also contain the residues with high NOE values and clusters around the iso-τ_c line of 10 ns. The τ_c values were expected to be below 10 ns at 600 and 800 Mhz (8.5 and 8.8 ns, respectively), and slightly over 10 ns at 500 Mhz (12.3 ns). The error bars of the τ_c values, larger than 1 ns, also illustrate the variations in T_2 observed in figure 4.31. However, the τ_c should not be field dependent, which might indicate that the data obtained at 500 MHz are less reliable than the data from the two other fields. The values obtained for 600 and 800 MHz are consistent, indicating an overall rotational correlation time around 8.6 ± 1.0 ns. This τ_c tends to be longer than what one would expect for a protein of this size, considering that the folded region is only about 40 residues. However the hydrodynamic radius of the molecule may be fairly large because of the unfolded region, so it is not unreasonable.

Figure 4.32 shows the S^2 values, generalised order parameter characterising the amplitude of internal motions, obtained from the extended Lipari-Szabo analysis. The analysis was performed using the NOE, T_1, and T_2 values for the three fields studied and a τ_c value of 8.6 ± 1.0 ns. The experimental data were simulated using the spectral density functions models described by five equations (see table IV.III). The results were presented, on a residue per residue basis, by the equation used and their respective parameters; S^2, S^2_s, R_{ex}, and τ_e. The qualitative analysis described in the previous section suggested the presence of exchange contribution. Accurate assessment of chemical exchange contributions to R_2 is

essential for a proper interpretation of the model-free results and for this purpose we used data at three different fields.

The analysis shows differences in dynamics between regions in the protein (figure 4.32), with a similar profile to the NOE and R_2 plots shown in figure 4.30. Two distinct regions can be identified from the order parameters (S^2) plots shown in figure 4.32. First, the lowest S^2 values are found for the poorly defined region in the structure, the N-terminal region of the protein, residues 163 to 227. The average S^2 for this region is 0.29 ± 0.13, indicating that large rmsd values in this region are due to the high mobility. These values are similar to those found in other unfolded proteins [124]. The relaxation data for 90% at 600 MHz and 97% at 800MHz are satisfactorily represented by equation 5 (S^2, S^2_f, τ_i). Dynamics of amide vectors located in this region have a large contribution from a fast internal motion represented in equation 5 by τ_i. However, the data for 43% of the spins at 500 MHz are represented by equation 4 (S^2, τ_i, R_{ex}) which has an additional contribution from exchange in comparison to equation 5 (red bars in the upper graph, figure 4.32). As noted previously, the data obtained at 500 MHz did not seem reliable due to inconsistency in τ_c which makes S^2 less reliable.

The second distinct region is localised around residues 229 to 280, and contains the hairpin and the linker regions (figure 4.19). The average values of S^2 for this region is 0.87 ± 0.25 for the three fields studied and the data for most of the spins are satisfactorily represented by the equations 2 (S^2, R_{ex}) and 4 (S^2, τ_i, R_{ex}) (yellow and red bars respectively in figure 4.32). On the graphs, the highest S^2 values are centred around residues; 236, 242, 253, 261, 266 and 275, which include the secondary structure elements. The Ile242 and Ile266 are located in the middle of the two β-strands and make an hydrogen bond to each other, explaining their restricted amide vector dynamic. The Trp236 side-chain is forming an hydrophobic core with Trp275, as for the Trp253 with Trp261. These interactions between the tryptophans side-chains are the most likely reason for the restricted motions of the backbone NH group. On the other hand, the low S^2 values are concentrated in regions of the hairpin (residues 244-262) and linker regions (residues 228-232 and 276-280) where no backbone interactions are observed. In general, those two regions of the protein have a dynamic backbone, with more restriction around the secondary structure elements. However, the analysis is not so reliable as many spins are in the forbidden region of $S^2 > 1$, as expected from the T_1 over T_2 plots, which could be misinterpreted as having a highly restricted internal dynamics for those spins ($S^2 > 0.8$). Those high S^2 values are most likely due to exchange

contribution to the T_2 values.

Amide vectors that have additional exchange contributions (R_{ex}) are also concentrated in this second area of the plot, more specifically in proximity of the two β-strands (residues 242 and 266) (see figure in appendix X.8). The largest exchange terms (R_{ex}) in the model-free analysis are found for residues Ile242, Ile266 and Glu270. As described before, the two first residues are located in the middle of the β-strands and make an hydrogen bond to each other. The most likely explanation for the additional exchange contribution is formation and breaking of the hydrogen bond between the two strands. However, reside Glu270 is located in a bulge and may also form dynamic hydrogen bonds to other residues in proximity. It is also well known that simplified isotropic models, in which anisotropy is neglected, can wrongly be interpreted as an exchange contribution and will be looked into further in this work.

The field dependence observed in the S^2 values can come from an error in the determination of the τ_c. First, the data obtained at 500 MHz do not seem consistent with the other fields, a variation of approximately 3.5 ns is observed in the τ_c determined (12.5 ns instead of 8.6 ns). Analysing the data with a τ_c of 12.5 ns has strong effect on the S^2 values obtained (data not shown). The general profile of the plot is identical to figure 4.32, but the S^2 values are, as expected, different. The choice of model for the first region of the plot (left) change for equation 5, as observed at 600 and 800 MHz. However, the region on the right hand side of the plot is still generally using equation 4, in contrast with equation 2 for the other fields. Other values for τ_c were also tested for the other fields without improvement in the consistence of the data.

The average value of S^2 for the N-terminal region of the protein (residues 163 to 227) is similar to values found for other unfolded proteins [124]. For the hairpin and linker regions, residues 228-280, the average value of S^2 obtained is close to folded proteins [120, 125]. A similar S^2 profile is observed for the three fields studied, however, a strong field dependence can be detected. Also, some residues reach S^2 values higher than the allowed maxima of 1 which make analysis less reliable. Qualitatively, the results are similar to the simple relaxation rate analysis, but also indicate that the traditional approach does not describe the motion satisfactorily. The 'model-free' analysis has been developed for globular folded proteins and our data show no self-consistent field dependence, which might again indicate that the analysis is not appropriate for this kind of unusual protein fold. It is interesting to notice that the equation 5 was mainly used for the unfolded region,

suggesting the presence of internal motions on the intermediate time scale in contrast to the equations 2 and 4 used for the folded region, suggesting the presence of some exchange (table IV.III). The main outcome of this analysis is that the motions are not restricted for the N-terminal part of the protein, with small S^2 values. The residues in proximity of the β-strands are the most restricted, with less restriction at the tip of the 'hairpin'; and 'linker-regions'.

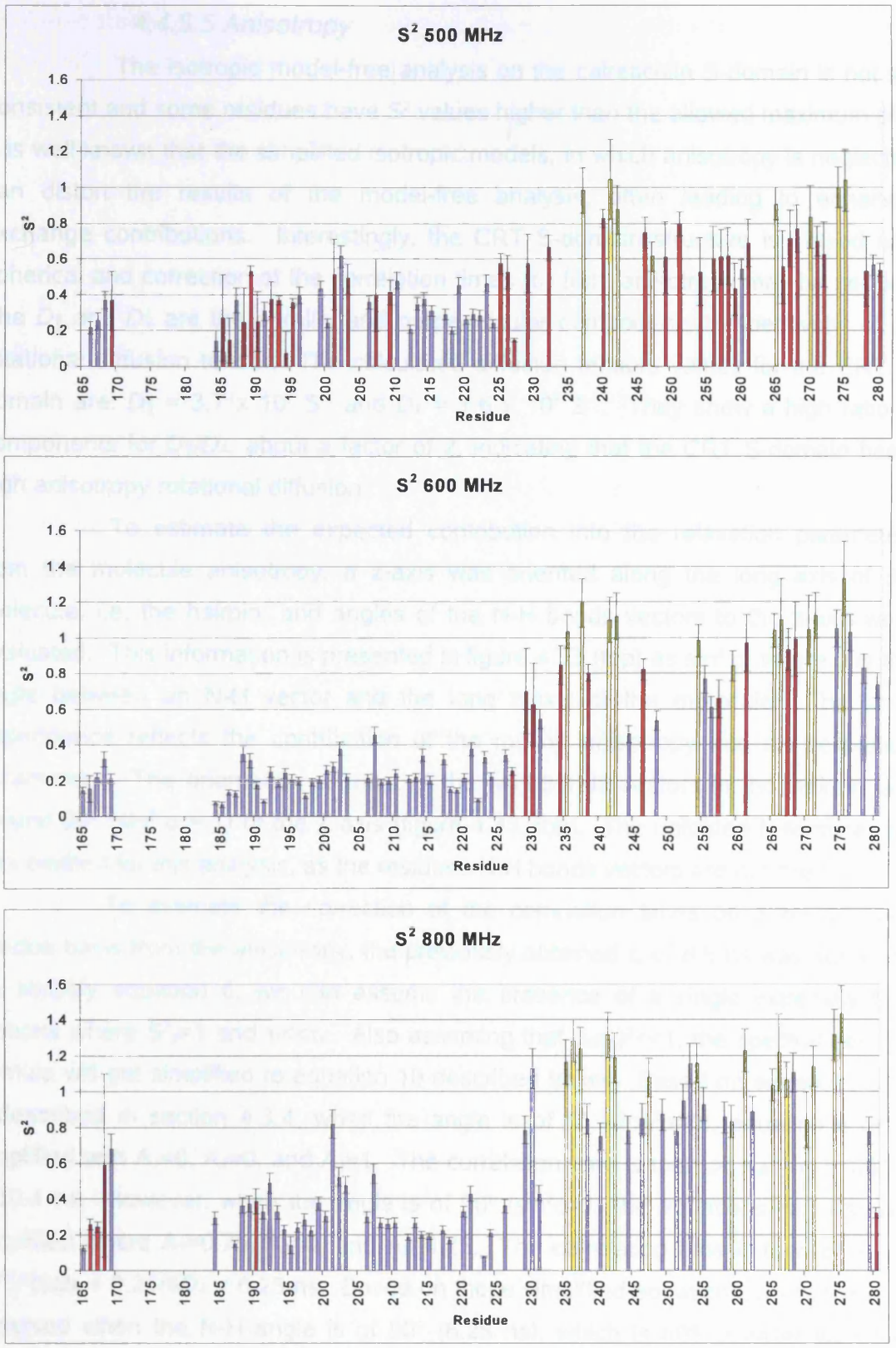

Figure 4.32 : Profile of the order parameter S^2 determined from different models. The S^2 are shown for each residues using a τ_c of 8.52 ± 1 ns for all the fields (500, 600 and 800 MHz). The bars have been coloured according to the spectral density model used in the analysis: model 5 (S^2, S^2_f, τ_i) in blue, model 4 (S^2, τ_i, R_{ex}) in red and model 2 (S^2, R_{ex}) in yellow.

4.4.5.5 Anisotropy

The isotropic model-free analysis on the calreticulin S-domain is not self consistent and some residues have S^2 values higher than the allowed maximum of 1. It is well known that the simplified isotropic models, in which anisotropy is neglected, can distort the results of the model-free analysis, often leading to enhanced exchange contributions. Interestingly, the CRT S-domain structure is indeed non-spherical and correction of the correlation times (τ_c) from anisotropy may be needed. The $D_{\parallel}$ and $D_{\perp}$ are the parallel and perpendicular components, respectively, of the rotational diffusion tensor. The calculated diffusion tensors values for the CRT S-domain are: $D_{\parallel}$ = 3.7 x 10^7 S^{-1} and $D_{\perp}$ = 1.6 x 10^7 S^{-1}. They show a high ratio of components for $D_{\parallel}/D_{\perp}$, about a factor of 2, indicating that the CRT S-domain has a high anisotropy rotational diffusion.

To estimate the expected contribution into the relaxation parameters from the molecule anisotropy, a z-axis was oriented along the long axis of the molecule, i.e. the hairpin, and angles of the N-H bonds vectors to the z-axis were evaluated. This information is presented in figure 4.33 (top) as $\sin^2 \alpha$, where α is the angle between an N-H vector and the long z-axis of the molecule. The sine^2 dependence reflects the contribution of the motion anisotropy into the relaxation parameters. The orientation of most of the N-H bonds vectors in the 'hairpin' are around 90° ($\sin^2 \alpha \approx 1$) of the Z-axis (figure 4.33, top). The unfolded N-terminal tail was omitted for this analysis, as the residues N-H bonds vectors are not fixed.

To estimate the correction of the correlation times on a residue per residue basis from the anisotropy, the previously obtained τ_c of 8.5 ns was adjusted. To simplify equation 6, we can assume the presence of a single extremely fast motions where $S^2_s=1$ and $\tau_f << \tau_k$. Also assuming that $(\omega\tau_k)^2 << 1$, the spectral density formula will get simplified to equation 10 described below. Based on equations 6 to 9 described in section 4.3.4, when the angle is of 0° ($\text{sine}^2\alpha=0$), equation 6 gets simplified with $A_1=0$, $A_2=0$, and $A_3=1$. The correlation time is then of; $\tau_{eff} = \tau_3 = 1/6D_{\perp}$ = 10.4 ns. However, when the angle is of 90° ($\sin^2\alpha=1$) the equations can also be simplified where $A_1=0.75$, $A_2=0$, and $A_3=0.25$. The correlation time is then of; τ_{eff} = $0.75/10D_{\perp} + 0.25/6D_{\perp}$ = 6.25 ns. Based on those simplified equations, a smaller τ_c is observed when the N-H angle is of 90° (6.25 ns), which is 60% smaller than the highest τ_{eff} when the angle is of 0° (10.4 ns). To keep the data to the same scale, a τ_c of 8 ns was used for $\text{sine}^2 \alpha = 1$ and the others adjusted to reach a maximum of 14 ns in the cases where $\text{sine}^2 \alpha = 0$. To estimate the effective τ_c at the intermediate angles, a simple formula was used; $\tau_c=((1-\sin^2 \alpha)$ x14ns) + ($\sin^2 \alpha$ x 8ns). This

simplified formula gives the right values at the extremes, α=0 and 90°, and a good approximation for the intermediate angles.

Equation 10:

$$J(w)=\sum_{k=1}^{3} A_k\left\{S_f^2 \tau_k\right\}$$

The figure 4.33 (bottom) show an estimated correction factor to the correlation times from anisotropy. From this plot we can notice that the same correction for anisotropy is required for most of the residues. Although the CRT S-domain is anisotropic, the isotropic model can be used because most N-H vectors have a similar orientation and no improvement is expected by introduction of anisotropy. Finally, the residues for which the T_1/T_2 ratios are elevated is most likely due to exchange rather than rotational anisotropy.

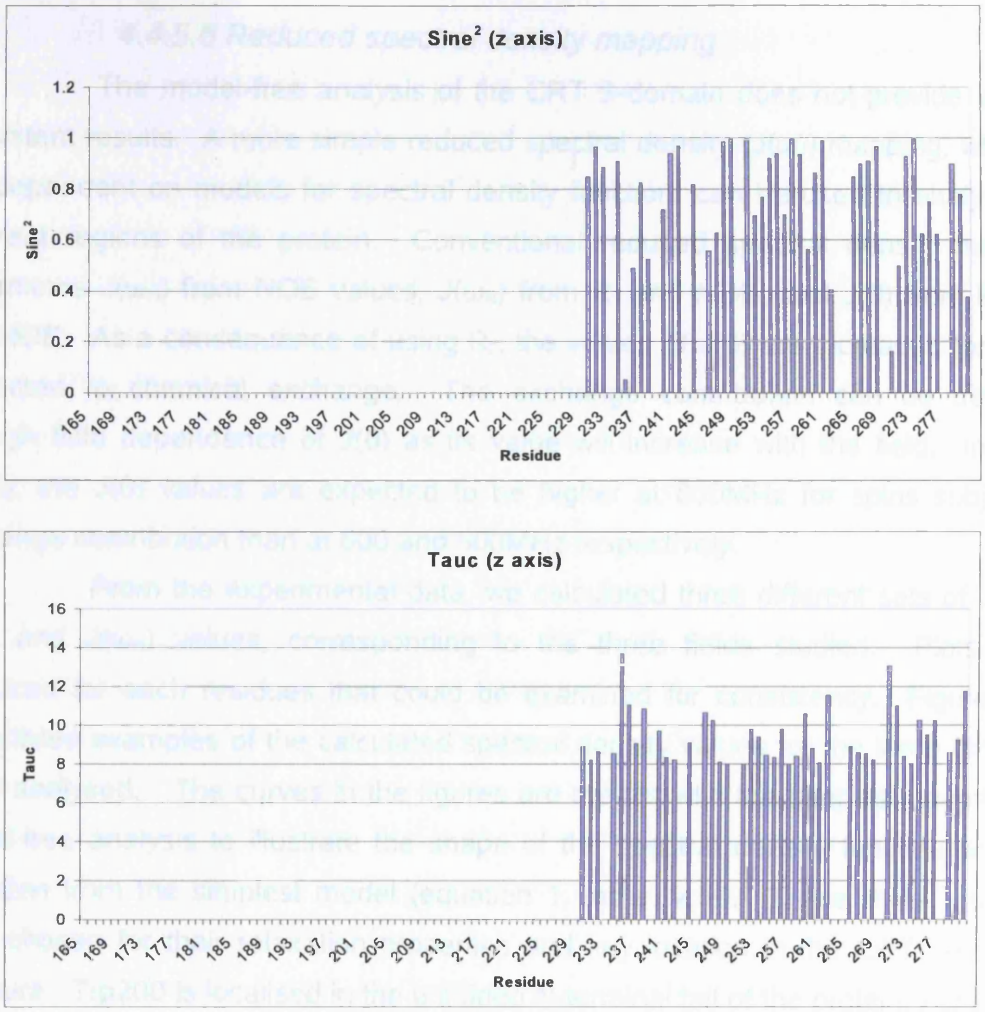

Figure 4.33 : Sine² α and Tauc (τ_c) on a residue basis. Top: Angle of the backbone amide vectors to the z-axis expressed as sine². Data obtained with HYDRONMR and pdbinertia. Bottom: Estimated correction to Tauc (τ_c) on a residue basis from anisotropy.

4.4.5.6 Reduced spectral density mapping

The model-free analysis of the CRT S-domain does not provide us with consistent results. A more simple reduced spectral density ($J(\omega)$) mapping, which is not dependent on models for spectral density function, can be used to analyse the different regions of the protein. Conventional reduced spectral density mapping determines $J(\omega_H)$ from NOE values, $J(\omega_N)$ from R_1 and NOE, and $J(0)$ from R_1, R_2, and NOE. As a consequence of using R_2, the values of $J(0)$ are increased for spins subjected to chemical exchange. The exchange contribution can be detected through field dependence of $J(0)$ as its value will increase with the field. In other words, the $J(0)$ values are expected to be higher at 800MHz for spins subject to exchange contribution than at 600 and 500MHz respectively.

From the experimental data, we calculated three different sets of $J(0)$, $J(\omega_N)$, and $J(\omega_H)$ values, corresponding to the three fields studied. Plots were produced for each residues that could be examined for consistency. Figure 4.34 show three examples of the calculated spectral density values for the three different fields analysed. The curves in the figures are the fits with the simplest equation of model-free analysis to illustrate the shape of the spectral density function and it's deviation from the simplest model (equation 1, table IV.III). These three residues were chosen for their relaxation properties and key location in the CRT S-domain structure. Trp200 is localised in the unfolded N-terminal tail of the protein, Asp254 at the tip of the 'hairpin', more precisely in the loose helical turn, and finally, Ile266 in the most rigid part of the protein, the β-sheet.

It can be noticed from results of the analysis, that the quality of the values for the $J(0)$ calculated are not fully consistent. The values for the three fields are in the wrong order in contrast to the values for $J(\omega_N)$ and $J(\omega_H)$, with larger values at 800MHz and decreasing with the field, possibly due to exchange contribution. In the case of Asp254 the fit with the simplest model of spectral density cover all the $J(\omega_N)$ and $J(\omega_H)$ points (figure 4.34, middle). However, the $J(0)$ points are high and in the wrong order, suggesting the presence of exchange contribution. In agreement with this observation, the model-free analysis used equation 2 for this residue (S^2, R_{ex}). For Ile266 (figure 4.34, bottom), the $J(\omega_N)$ values are below the fitted line, suggesting the presence of motion on a slower time scale. This effect is most probably due to the very high values for $J(0)$ making the fitted line for the simplest model stretch to cover those points. In this case, the data indicate the presence of strong contribution from exchange, also in agreement with the model-free analysis which used equation 2 to describe the motion. This strong exchange contribution for

Ile266 could be explained by the presence of a hydrogen bond forming and breaking with Ile242 in the centre of the β-sheet. In the case of Trp200, the $J(\omega_N)$ values are below the fitted line for the simplest model, suggesting the presence of a motion on the intermediate time scale (figure 4.34, top). The J(0) were also in the wrong order in this case, however they are not as high as in the two previous cases, about 1.1 ns instead of 4.5 and 7 ns. The model-free analysis used equation 5 to describe the motion of this residue (S^2, S^2_f, τ_i), which is in agreement with the presence of a motion on the fast to intermediate time scale without exchange contribution. We can also notice a small variation in the $J(\omega_H)$ values for Trp200, which is in agreement with the presence of a fast motion described by S^2_f in the model-free analysis. Most of the residues located in the unfolded N-terminal region of the protein have similar profiles. In general, the data are not fully consistent, especially for J(0), but the analysis indicates the presence of motions on different time-scales.

The figure 4.35 shows a summary of the reduced spectral density mapping for the values of; J(0), $J(\omega_N)$, and $J(\omega_H)$. It can be concluded that the residues located in the unfolded N-terminus of the protein (163-230), have a high field dependence for $J(\omega_N)$ and $J(\omega_H)$, which indicate the presence of fast motions (<0.1 ns). Considering that this region is unfolded the presence of a fast motion is not unreasonable, because of the absence of any backbone interactions. The residues displaying the exchange contribution can be identified by visual inspection of the J(0) graphs, since they have large and field dependent values. Peaks centred around residues 236, 242, 266, 270 and 275, show high values for J(0) that suggest the presence of strong exchange contribution. Those peaks coincide with the residues involved in the β-sheet and the second hydrophobic core. The $J(\omega_N)$ values provide valuable information that is free of exchange contribution for the folded region of the protein (residues 230 to 280). The data at J(80) show little variation, however below this frequency a pattern for J(60) and J(50) is found similar to that of J(0), with low values for the flexible regions and high values for the rigid ones. The secondary structure elements, β-strands, are among these last residue but, interestingly, high values are also observed around residues Trp236 and Trp275 for which only their side-chains are involved in a hydrophobic core. The large variation in $J(\omega_N)$ values for residues 230 to 280 is indicative of a motion on the intermediate time scale (1 to 0.1 ns) for the folded region of the protein.

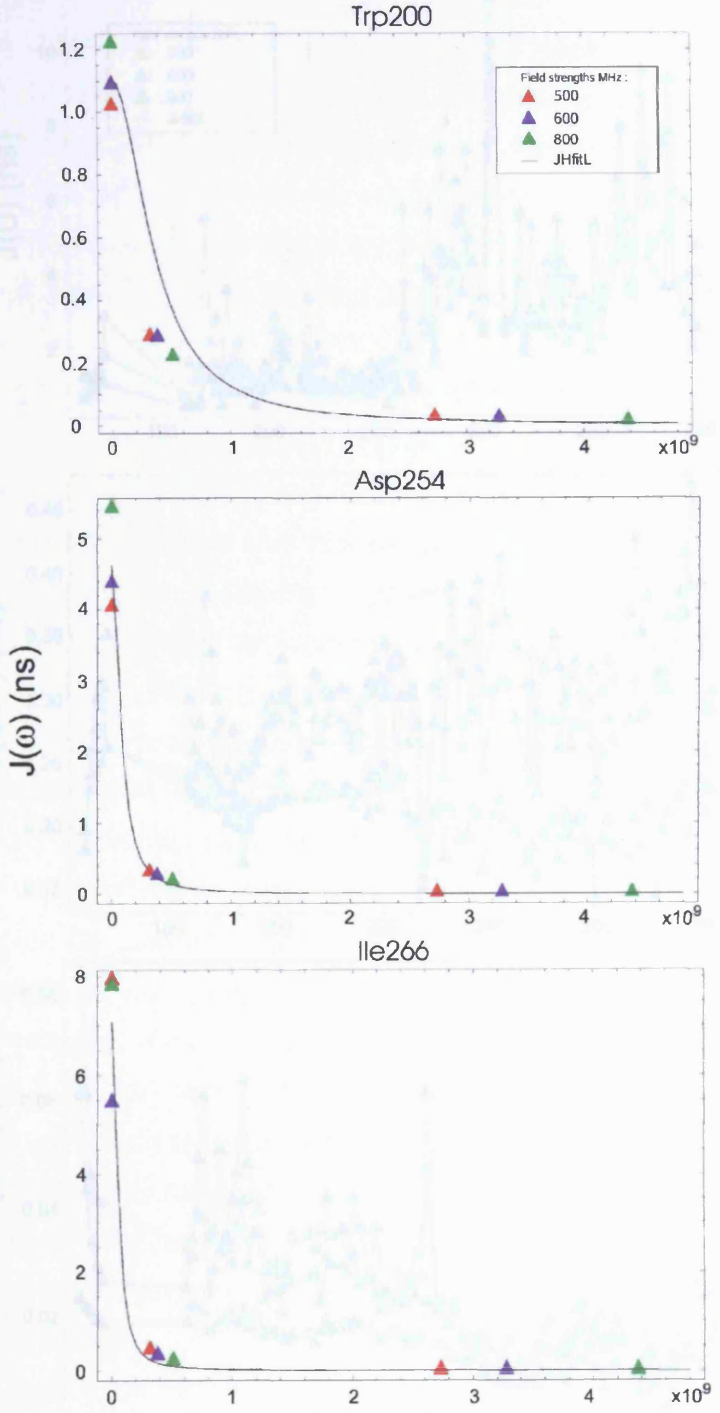

Figure 4.34 : Plots for the calculated reduced spectral density values J(0), J(ω_N) and J (ω_H) at different frequencies. Data obtained at 500 (red), 600 (blue) and 800 MHz (green).

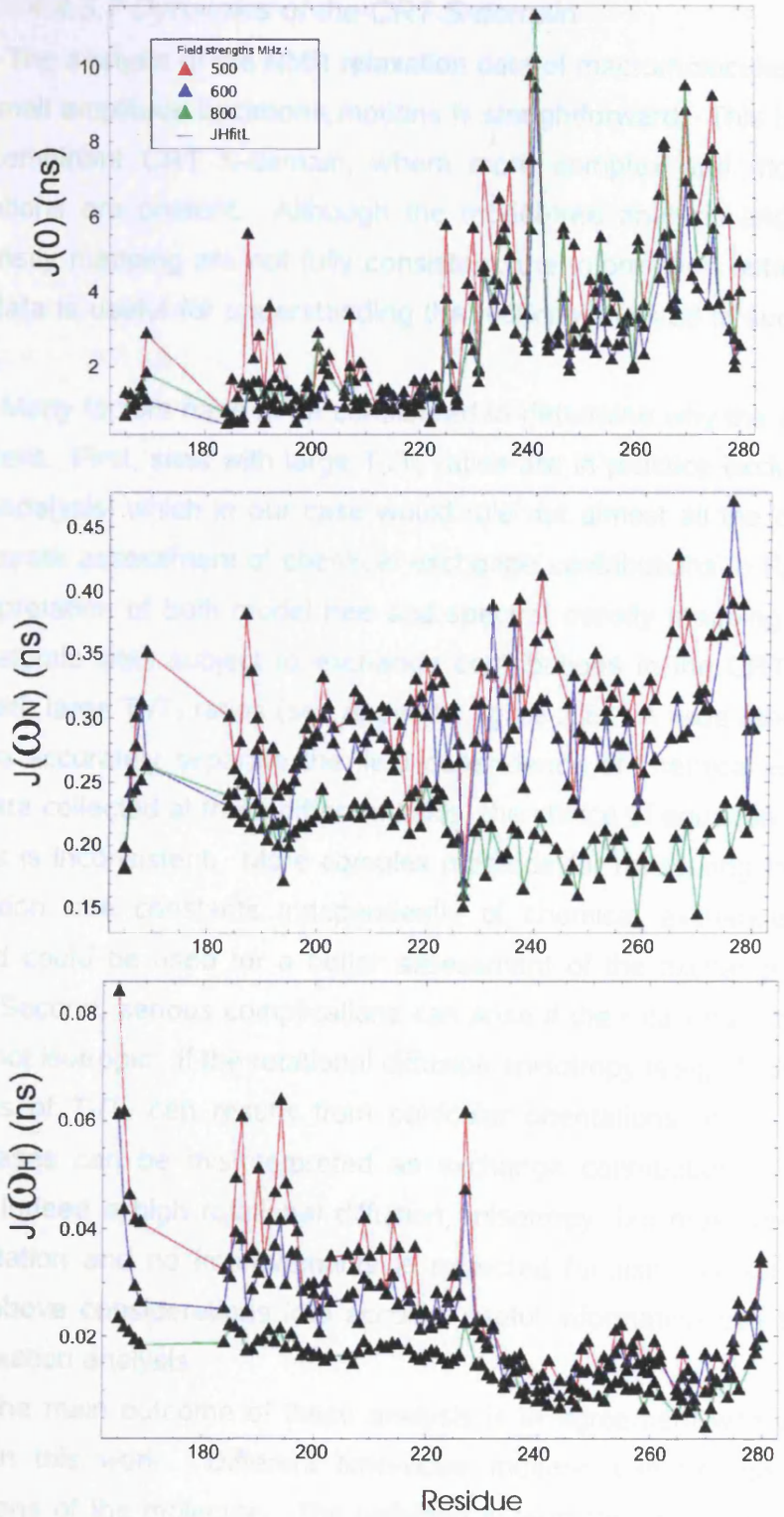

Figure 4.35 : CRT S-domain sequence distribution of the reduced spectral density mapping at different frequencies. Distribution of the spectral density values, J(0), $J(\omega_N)$ and $J(\omega_H)$, mapped at different frequencies. Filled triangles represent the data obtained at 500 (red), 600 (blue) and 800 MHz (green).

4.4.5.7 Dynamics of the CRT S-domain

The analysis of the NMR relaxation data of macromolecules that undergo only fast, small amplitude backbone motions is straightforward. This is not the case for the recombinant CRT S-domain, where more complex and more interesting internal motions are present. Although the model-free analysis and the reduced spectral density mapping are not fully consistent, the information obtained from the relaxation data is useful for understanding the motions involved in such an unusual protein fold.

Many factors have to be considered to determine why the analysis is not fully consistent. First, sites with large T_1/T_2 ratios are in practice excluded from the model-free analysis, which in our case would rule out almost all the data. For this reason, accurate assessment of chemical exchange contributions to R_2 is critical for proper interpretation of both model-free and spectral density mapping results. The number of atomic sites subject to exchange contributions in the CRT S-domain is significant with large T_1/T_2 ratios (see appendix figure X.8). A wide range of fields is necessary to accurately separate the field dependence of chemical exchange and even with data collected at three different fields, the choice of equation in the model-free analysis is inconsistent. More complex methods for measuring the transverse cross-relaxation rate constants independently of chemical exchange have been reported and could be used for a better assessment of the exchange contribution [126, 127]. Second, serious complications can arise if the rotational diffusion of the molecule is not isotropic. If the rotational diffusion anisotropy is significant, increases in the values of T_1/T_2 can results from particular orientations of the N-H vector. These increases can be misinterpreted as exchange contribution. The CRT S-domain has indeed a high rotational diffusion anisotropy, but most vectors have a similar orientation and no improvements is expected for using anisotropic model. Taking the above considerations into account useful information can be extracted from the ralaxation analysis.

The main outcome of these analysis is in agreement with the structure determined in this work. Different time-scale motions can be associated with different regions of the molecule. The unfolded N-terminal tail of the protein show fast motions (residues 163-230), supporting a high motional freedom of the NH-vectors rather than a lack of data. In contrast, the two β-strands show the highest motional restrictions, but have also the strongest exchange contributions probably due to the formation and breaking of the hydrogen-bonds. Most interestingly, the CRT S-domain structure shows a low rmsd for the backbone in the hairpin region,

however, the relaxation data indicate local dynamics of the NH-vectors. Although the structure is rigid for this region of the protein, there is an absence of stabilising hydrogen-bonds, leaving a high degree of motional freedom to the backbone NH-vectors. The average NOE value in the hairpin is approximately 0.5 in contrast to values of 0.8 for most globular proteins, which indicates that the backbone is flexible. The core of the hairpin appears less dynamic than the backbone as the main stabilising force is consisted of two hydrophobic tryptophan side-chains. Taking this in consideration, one would expect that the backbone of the CRT S-domain hairpin described in the structure could be much more dynamic in solution and this could have some biological relevance. An evident negative charge also observed for this part of the protein, could indicate that this region of the protein needs some flexibility to remodel itself to adopt different conformations for binding to multiple ligands.

4.5 DISCUSSION

The present study shows that the unusual hairpin-type fold of the 100-residue CRT P-domain [59] is maintained in the S-domain (figure 4.36). The P-domain of CRT shows an extended hairpin fold that involves the entire polypeptide chain, and has the two chain ends in close spatial proximity (figure 4.36a). This globally extended structure is stabilised by three short anti-parallel β-sheets between repeats pairs A1-B3, A2-B2, and A3-B1. The hairpin loop and the two connecting regions between the β-sheets contain a hydrophobic cluster, where each of the three clusters includes two highly conserved tryptophan residues, one from each strand of the hairpin. There is a high degree of amino acid sequence similarity between the P-domain of CRT and calnexin (CNX) (see section 1.1.2). The crystal structure of the CNX [128] shows the same topology and overall tertiary structure as CRT for the P-domain with an elongated hairpin fold (figure 4.36c). The P-domain is a site of chaperone activity and oligosaccharide binding to both proteins with two sequence repeat types, designated A and B, each repeated three times in CRT and four times in CNX (figure 1.3). In this structural topology, repeats A (PXXIXDPDAXKPEDWDE) run anti-parallel to repeats B (GXWXPPXIXNPXYX) in the P-domain structures, where their conserved Isoleucines face each other to form the centre of the β-sheets and similarly, their conserved tryptophans face each other to form a hydrophobic cluster.

The CRT S-domain structure determined in this work forms a stable hairpin fold with a core of roughly 39 residues, which retains the structural features of the corresponding polypeptide segment in the P-domain (figure 4.36b). The main driving force for this fold is a short 3 residue β-sheet between repeats A3 and B1, holding the two strands together, with an hydrophobic cluster at the tip of the hairpin. The second β-sheet between repeats A2 and B2 present in the P-domain is not formed in the S-domain probably because the repeat B2 is too close to the C-terminus of the chain and thus not stable. However, a second hydrophobic core is formed between repeats A2 and B2, but is less well defined possibly due to the non-formation of the upstream β-sheet. There are no other stabilising forces for the backbone in the hairpin region, giving little rigidity to the molecule which has been confirmed in this study by relaxation experiments. The fact that the CRT S-domain also contains an N-terminal disordered tail of approximately 66 residues localised between the N-terminus to repeat A2, further illustrates that the formation of this fold-type depends critically on inter-strand contacts in the hairpin since the partner strand segment of repeat A1, repeat B3, is missing in this construct (figure 4.36b).

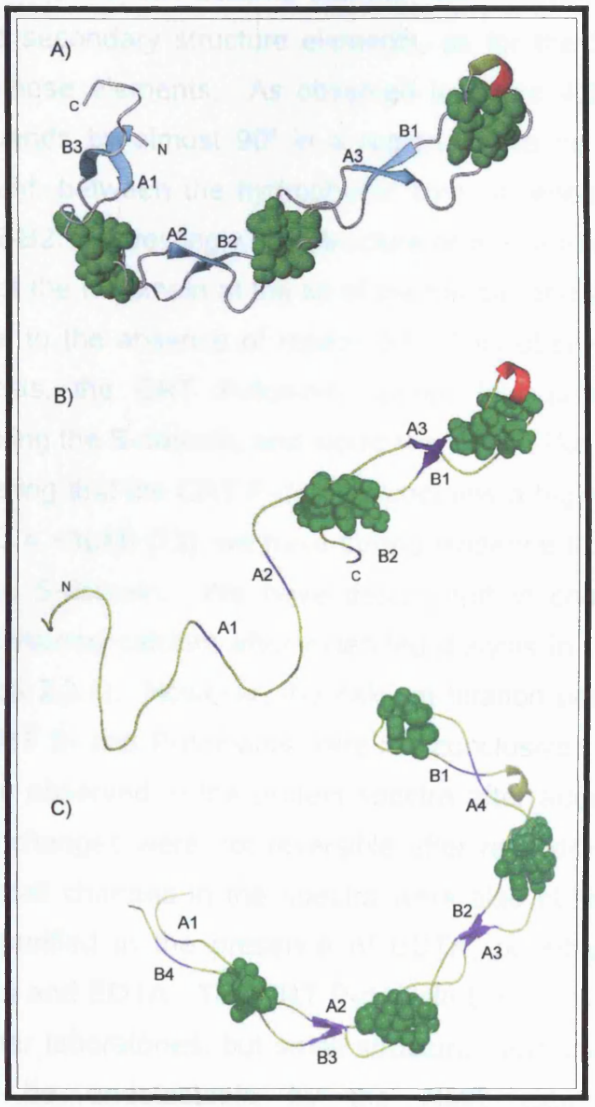

Figure 4.36 : Three dimensional structure of the calreticulin P-domain, calreticulin S-domain and calnexin P-domain. In; A) the rat CRT P-domain (PDB accession 1HHN), in B) the human CRT S-domain, and in C) the *Canis Familiaris* (dog) CNX P-domain (PDB accession 1JHN). The ribbon drawing shows the α-helical turns in red and the β-sheets in blue. The hydrophobic clusters are shown in green all-heavy-atoms space-filling models. The β-strands are numbered according to their relative repeats, as described in figures 1.3 and 4.3.

The relaxation data clearly differentiate three regions in the CRT S-domain structure. First, the N-terminal which has a high backbone motional freedom, then the linker region, composed of repeats A2 and B2, which makes the transition between the unfolded N-terminal and the more restricted short hairpin. Without relaxation data for the CRT P-domain, it is difficult to have an idea on its backbone motional freedom, but its elongated structure indicates that this domain of CRT has the possibility to change its overall shape, like an arm. Using the CRT S-

domain relaxation data, one can assume that the restricted regions in the P-domain co-localise with the secondary structure elements, as for the S-domain, with more flexibility between those elements. As observed in figure 4.36, the elongated P-domain structure bends by almost 90° in a region where no secondary structure elements are present, between the hydrophobic core of repeats A1-B3 and the β-sheet of repeats A2-B2. Interestingly, the structure of the S-domain keeps the same structural features of the P-domain at the tip of the hairpin and a more disordered N-terminal mainly due to the absence of repeat B3. This observation could indicate that during synthesis, the CRT P-domain cannot fold until the repeat B1 is synthesised, mimicking the S-domain, and would then close like a zipper.

Considering that the CRT P-domain contains a high-affinity, low capacity Ca^{2+}-binding site (K_d = ~1μM) [13], we have strong evidence that this binding site is compromised in the S-domain. We have determined in chapter 2 that CRT P-domain samples contained calcium after extended dialysis in contrast to S-domain samples (see section 2.3.6). However, the calcium titration performed in this study by NMR with the CRT S- and P-domains were not conclusive (see section 4.3.1.1). Small changes were observed in the protein spectra after addition of EDTA to the samples, but those changes were not reversible after re-addition of calcium to the samples. Those small changes in the spectra were also observed by addition of EDTA to samples purified in the presence of EDTA, possibly due to interaction between the proteins and EDTA. The CRT P-domain binds calcium, as determined in this work and other laboratories, but small structural rearrangements involved to the protein must be undetectable by the NMR experiments performed. Unfortunately, the NMR structure of the P-domain does not show evidence of calcium binding site, although the structure was determined in the presence of 10 mM $CaCl_2$. Moreover, the calnexin P-domain which is highly homologous to the CRT P-domain, does not contain any calcium ions in the crystal structure [128]. Using the structure of the CRT and CNX P-domains, the exact location of the calcium-binding site is still unclear. The additional information on the CRT S-domain structure could imply that the calcium-binding site is localised in a region between repeats A1-B3 to A2-B2 (figure 4.36) or that the non-formation of secondary structure element in those repeats destabilise repeats A3-B1, therefore reducing binding affinities for calcium.

The small CRT S-domain hairpin is thus of keen interest, because it comprises all residues involved in direct interactions between the CRT P-domain and Erp57 [51] , and on the other hand there is strong evidence that the C1q and MBL binding sites are also localised in this S-domain. The relaxation data indicate that

this hairpin backbone has a slow to intermediate motion, which is faster in comparison with other more globular proteins described in the literature. This backbone flexibility could be the key for the calreticulin chaperone activity, which would allow it to adopt its conformation for binding to multiple ligands. Moreover, small motional restrictions for the backbone would support larger motional freedom to the negatively charged side-chains, therefore facilitating intermolecular interactions through electrostatic forces with multiple protein ligands.

CHAPTER 5 :
CONCLUSIONS

5.1 Conclusions

Calreticulin (CRT) is a lectin-like chaperone in the endoplasmic reticulum (ER) of eukaryotic cells. It binds to the majority of nascent and newly synthesised glycoproteins in cells and the association promotes efficient folding of the proteins until they have acquired their native structure [14, 42]. Calreticulin also binds calcium and functions as a calcium storage protein within the ER lumen [13]. More importantly, CRT is also located on the surface of many cell types, including macrophages, and has been shown to specifically bind to the collagenous tail of C1q and MBL, and has been identified as the collectin receptor, cC1qR [21]. However, CRT lacks a transmembrane domain and therefore must associate with a cell surface receptor, possibly CD91 [29, 38]. CRT binds to the collagenous tails of C1q and MBL through a 12 kDa S-domain (spanning the N- and P-domains) [19]. Furthermore, it has been demonstrated that CRT can compete with antibodies for binding to C1q and this interaction may contribute to autoimmune diseases by inhibiting clearance of immune complexes [38]. Several parasites, including *Trypanosoma cruzi* have been shown to secrete and express on their surface homologues of CRT to evade the immune system by preventing the complement system acting against the parasite [129]. In the present study, we focused onto the interaction between the human calreticulin (hCRT) and the complement proteins C1q and MBL.

To further characterise the CRT interaction with the C1q and MBL proteins, three human calreticulin domains have been cloned, expressed in *E.coli* and purified to a high standard. Theses clones included the hCRT; S-domain (residues 164-283), P-domain (residues 198-308), and PS-domain which spans both the S- and P-domains (residues 164-308). Multiple biochemical methods have been used to analyse the quality of theses clones, such as N-terminal protein sequencing, mass spectroscopy, proteolysis, circular dichroism (CD), and NMR spectroscopy. Further assays have been used to test their binding capacity to the complement proteins C1q and MBL, such as; C1q-dependent haemolytic assay, and ELISA based assays (see chapter 2). We have shown that the recombinant human calreticulin S-, P-, and PS-domains clones have different affinity for these two complement proteins. Moreover, all hCRT clones, including the native calreticulin bind specifically to C1q and MBL and are potential inhibitors of the classical and lectin pathways of complement. Considering the 85 amino acids overlapping protein sequence in the three CRT clones, residues 198-283, the binding site to C1q and MBL has been partly localised to a smaller region of CRT than the previously

characterised S-domain.

To further understand the interaction between the collagen-like region of C1q and MBL with CRT, the structure determination of one of these CRT domains was essential. Of the three CRT clones produced, the S-domain has the highest affinity for C1q and MBL and was chosen for structure determination. The method of choice to solve the structure of the protein was crystallography. Unfortunately, only small crystals of the human calreticulin S-domain could be obtained after extensive screening (see chapter 3). The combination of all biochemical and activity assays suggest that the protein is active and folded, but the crystallogenesis of this protein seems problematic.

To further characterise the three CRT clones, NMR spectroscopy was used. The P-domain shows the best dispersion of NMR signals of the three clones in 2D ^{15}N-resolved HSQC and its three dimensional structure was judged possible to solve by NMR. At the same time, another group was also studying this domain by NMR who published the resonance assignment [121] and later the structure [59]. We then pursued the structure determination of the CRT S-domain by NMR, which was to us, the most interesting clone for its enhanced ability to inhibit complement activation. The refined structure shows an extended hairpin topology, with one short anti-parallel β-sheet, two small hydrophobic clusters, one helical turn at the tip of the hairpin, and a long disordered N-terminal (see chapter 4). The hairpin is conserved between the CRT S- and P-domain structures (figure 4.36), however, the S-domain has a long disordered N-terminal tail. Relaxation data are in agreement with the protein structure obtained, indicating the presence of highly dynamic backbone in the N-terminus region of the protein that could explain the problems in the crystallogenesis of this domain.

Interestingly, the hairpin is contained within the previously observed 85 residues overlap between the three CRT clones that bind to C1q and MBL. These data strongly suggest that the interaction between CRT and the collagen-like region of C1q and MBL, is localised within the CRT hairpin. The CRT S-domain structure in figure 4.36 shows the importance of repeats A3 and B1 in stabilising the hairpin fold. However, the structure clearly indicate that the repeats A2 and B2 are not trivial, as they only contribute to the S-domain fold by a small hydrophobic cluster that does not look well defined in the structure with a clearly dynamic backbone by relaxation studies. To confirm the hypothesis that the hairpin is the binding site for the collagen-like region of C1q and MBL, a short 5.5 kDa CRT construct composed of residues 235 to 273, including repeats A3 and B1, has been cloned and purified in

Professor Schwaeble laboratory. C1q and MBL binding assays have confirmed our theory that the binding site of CRT is located in this short hairpin (original data not shown in this thesis). This new hCRT clone, residues 235 to 273, could be used as a low-molecular mass mimic of CRT to further investigate the CRT interaction with C1q and MBL.

With the recently published structure of the rat MASP-2 CUB1-EGF-CUB2 [130] and the structure of the CCP1-CCP2-SP [131] region from the homologous C1r containing the inactive zymogen form of the serine protease (SP) domain, Feinberg *et al.*, 2003 [130], constructed a full six-domain model for MASP-2 (figure 5.1 a). They also constructed a model for the dimer of homotrimer MBL molecule from coordinates of a collagen peptide and generated the correct number of Gly-X-Y repeats before and after the kink. The MBL dimer model was docked onto the MASP-2 model. Presumably, the catalytic regions are brought to proximity upon binding of MBL to a target cell surface, where upon cleavage of one zymogen by the other produces an active protease, which subsequently activates the partner (figure 5.1b).

The two possible CRT docking sites on MBL and C1q proteins are the globular heads or collagen-like region. The first hypothesis, where CRT would bind to the globular head of C1q and the C-type lectin domain of MBL, is less probable. The C1q and MBL have almost no sequence identity in their C-terminal regions. In the second hypothesis, the calreticulin hairpin would possibly dock on the same or overlapping region as the serine proteases on the collagen stalk of MBL or C1q, therefore inhibiting complement activation (figure 5.1c). In the case of MBL, the MASPs bind on the C-terminal side of the hinge region formed by an interruption in the Gly-X-Y repeat pattern of the collagen-like domain [132].

The third possible docking site for CRT on MBL and C1q is also located on the collagen-like region, but at the N-terminal of the hinge (figure 5.1d). This region of the collagen stalk of MBL and C1q is positively charged in the region located before the hinge with arginine and lysine side-chains (figure 1.7). This positive charge would then make a sensible binding site for the CRT hairpin which has a high negative charge, then allowing electrostatic interactions to form. Moreover, we have shown that the interaction between MBL and CRT is inhibited at high NaCl concentration (figure 2.8), and in a more recent publication, that the *Trypanosoma cruzi* (Tc) CRT S-domain, which has 77% identity to the human S-domain, binds to purified C1q collagen tails optimally under low ionic strength conditions [129], as observed for the human CRT S-domain [21]. The above two

support that the CRT binding to C1q and MBL involves electrostatic interactions.

All those data suggest that CRT binds to the N-terminal side of the hinge of the MBL and C1q proteins. The most likely site is a positively charged region, complementary to the negatively charged hairpin of CRT. Inhibition can either occur by steric hindrance or by destabilisation of the collagen triple helix, probably resulting from distortion of the long positively charged side-chains in the complex. This destabilisation to the collagen triple-helix, possibly also destabilises the C-terminal side of the collagen-like region of MBL and C1q. Two possible mechanisms of inhibition can occur. First, because the serine proteases presumably bind only to the helical conformation, this destabilisation in the triple helix could account for the loss in affinity for the MASPs and $C1r_2$-$C1s_2$ serine proteases for their respective ligands, therefore inhibiting complement activation by the MBL and classical pathways (figure 5.1d, arrow 1). The second mechanism could be that the interaction of CRT with the collagenous tails of both MBL and C1q mediates a steric change at the globular head level, resulting in inhibition of their interactions with their natural ligands (figure 5.1d, arrow 2).. The later one seems the most probable, as we have shown in a recent publication that TcCRT S-domain inhibit the binding of human MBL to mannose [129].

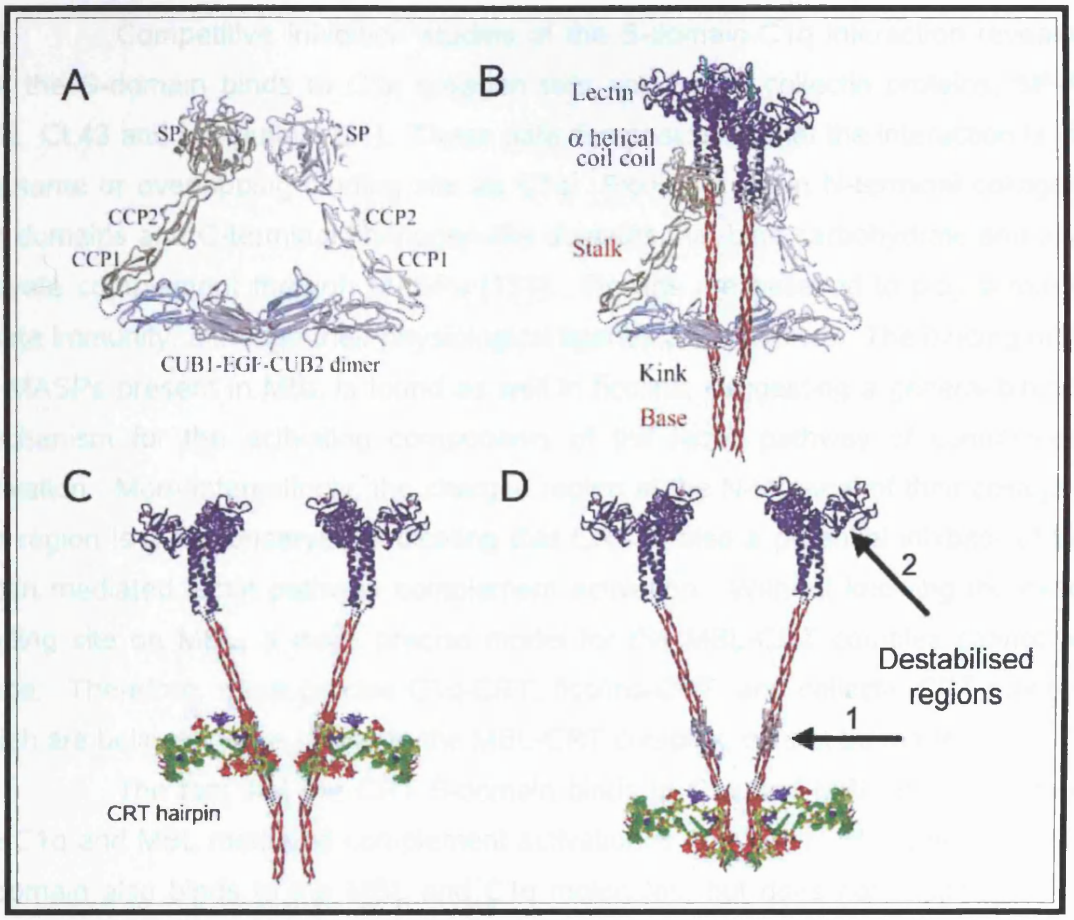

Figure 5.1 : Models of the CRT inhibition of the MASP-2 interaction with the MBL dimer. A) The N-termini of two molecules of human C1r CCP1-CCP2-SP (PDB ID 1GPZ) were positioned near the C-termini of the rat MASP-2 CUB1-EGF-CUB2 dimer (PDB ID 1NTO). B) Model of a dimer of trimeric MBL subunits bound to the MASP2 model shown in (A). Coordinates of the trimeric coiled-coil neck and lectin region of rat MBL (blue) (PDB ID 1RTM) were placed at the C-terminus of a model of the collagen-like region (red) generated from a collagen peptide $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ (PDB ID 1K6F); the dashed lines represent the flexible joint between these two regions of the MBP primary structure. The gap in the collagen connected by three dashed lines represents the interruption in the collagen consensus sequence, where the trimeric stalks splay away from the kink. Other dashed lines in the model represent protein links between the different domains. C) Model of CRT hairpin domain binding to the C-terminal side of the hinge region of MBL collagen-like domain. Inhibition of the MBL/MASPs complex by CRT binding on the same or overlapping region. D) Model of CRT hairpin domain binding to the N-terminal side of the hinge region of MBL collagen-like domain. Inhibition can either occur by steric hindrance or by conformational changes in the MASPs binding site induced by CRT binding to MBL, therefore reducing MASPs affinity for MBL. Figure adapted from Feinberg *et al.*, 2003 [130].

Competitive inhibition studies of the S-domain-C1q interaction revealed that the S-domain binds to C1q collagen tails and to the collectin proteins, SP-A, MBL, CL43 and conglutinin [21]. These data demonstrating that the interaction is via the same or overlapping binding site as C1q. Ficolins contain N-terminal collagen-like domains and C-terminal fibrinogen-like domains that bind carbohydrate and also activate complement through MASPs [131]. Ficolins are believed to play a role in innate immunity, although their physiological ligands are unknown. The binding motif for MASPs present in MBL is found as well in ficolins, suggesting a general binding mechanism for the activating components of the lectin pathway of complement activation. More interestingly, the charged region at the N-terminal of their collagen-like region is also conserved, indicating that CRT is also a potential inhibitor of the ficolin mediated lectin pathway complement activation. Without knowing the exact binding site on MBL, a more precise model for the MBL-CRT complex cannot be made. Therefore, more precise C1q-CRT, ficolins-CRT, and collectin-CRT models, which are believed to be similar to the MBL-CRT complex, cannot be made.

The fact that the CRT S-domain binds to C1q and MBL and also inhibit the C1q and MBL mediated complement activation is now clear. However, the CRT P-domain also binds to the MBL and C1q molecules, but does not inhibit the C1q haemolytic assay in the experimental conditions tested. Moreover, less C4b deposition was observed on CRT P-domain coated plates, indicating that the affinity of this domain for the complement molecules C1q and MBL is significantly less than the S-domain. With the structure of the two domains available, it is still unclear why the S-domain has such an enhanced activity. The most probable hypothesis is that the CRT binding site for those complement molecules, i.e. the tip of the hairpin, has more backbone freedom in the S-domain, allowing it to adopt its conformation more easily and/or more efficiently for binding than the P-domain. From the structural point of view, both clones have their negatively side-chains pointing out of the hairpin, making them very accessible for electrostatic interactions. Relaxation experiments performed on the S-domain suggest that the backbone in the hairpin region has a high motional freedom, also confirmed in the structure by the absence of backbone interactions in this region. Unfortunately, relaxation experiments have not been performed onto the P-domain, but the additional interactions between repeats A1 with B3 and A2 with B2 (figure 4.36) may introduce some extra stability into the hairpin region. Without additional relaxation data on the CRT P-domain it is difficult to be certain of those assumptions.

5.2 Further work

To understand in more detail the interaction between CRT and either C1q or MBL, the determination of their structure in complex is tempting. The size of those complexes is the limiting factor for NMR spectroscopy. A low molecular weight fragment of the large multimeric complement molecules C1q and MBL containing the CRT binding site would allow us to determine the key residues in the CRT hairpin by NMR spectroscopy. Those structural information would help us to understand more on how the hCRT S-domain binds and inhibit the classical and MBL pathway of the complement. The size of theses large molecules is not a limiting factor for crystallography, but obtaining large amounts of the complement proteins.

We have synthesised the 'GFPGKDGRDGTKGEKGEPGQG' peptide based on the charged region of the collagen-like region of MBL, just before the hinge. This peptide did not form the collagen trimer expected based on circular dichroism (CD) analysis performed from 4 to 30°C and NMR studies (data not shown). Moreover, no changes were observed in the 2D ^{15}N-resolved HSQC of the CRT S-domain upon addition of the peptide (data not shown). The above peptide has an unusual high content in long and positively charged side-chains (Arg and Lys), probably destabilising the collagen triple-helix formation. By adding at the two extremities of this peptide some repeats of the amino-acids sequence Gly-X-Y, where Gly represents glycine and X and Y proline or hydroxyproline residues, we could increase the thermal stability of the triple-helical peptide [133]. This extra stability could allow this peptide to form a tight triple helix, which would allow us to study the binding by NMR. By synthesising overlapping peptides of the MBL collagen-like region, Wallis *et al.*, 2004 [132], have mapped the MASPs binding site. Using the same approach it would be possible to determine the exact binding site of CRT in the MBL collagen-like domain. Moreover, using NMR it would be possible to determine exactly which residues in the CRT S-domain interact with those peptides. To overcome the instability of those short collagen-like peptides, we have cloned the human MBL collagen-like region (59 residues) in pET-15b for expression in *E.coli* and easy purification with the His-Tag. Unfortunately, the expression was low and purification difficult (data not shown). One of the reason for the low expression of this construct in *E.coli*, is that proper modification of proline residues to hydroxyproline does not take place.

APPENDIX

X.1 DNA gel extraction

Linearised Plasmids, PCR products and DNA inserts have been purified by Agarose gel extraction using the QIAGEN (West Sussex) QIAquick gel extraction kit. DNA fragments have been excised from the agarose gel with a clean, sharp scalpel and weighted in an eppendorf. The DNA extraction was performed as described by the manufacturer. The Plasmid was eluted in 50 µl of buffer EB (10 mM Tris-HCl, pH 8.5) and quantified using the standard of 50 µg/ml of double stranded DNA creates an A_{260nm} of 1.

X.2 Vectors

The vectors that were used in this work were, pET-15b and pET-11a (Novagen, CN Biosciences Ltd., Nottingham, UK) and the pGEM-T Easy (Promega, Southampton, UK). The following figures represent the vector maps where the multiple cloning sites are shown.

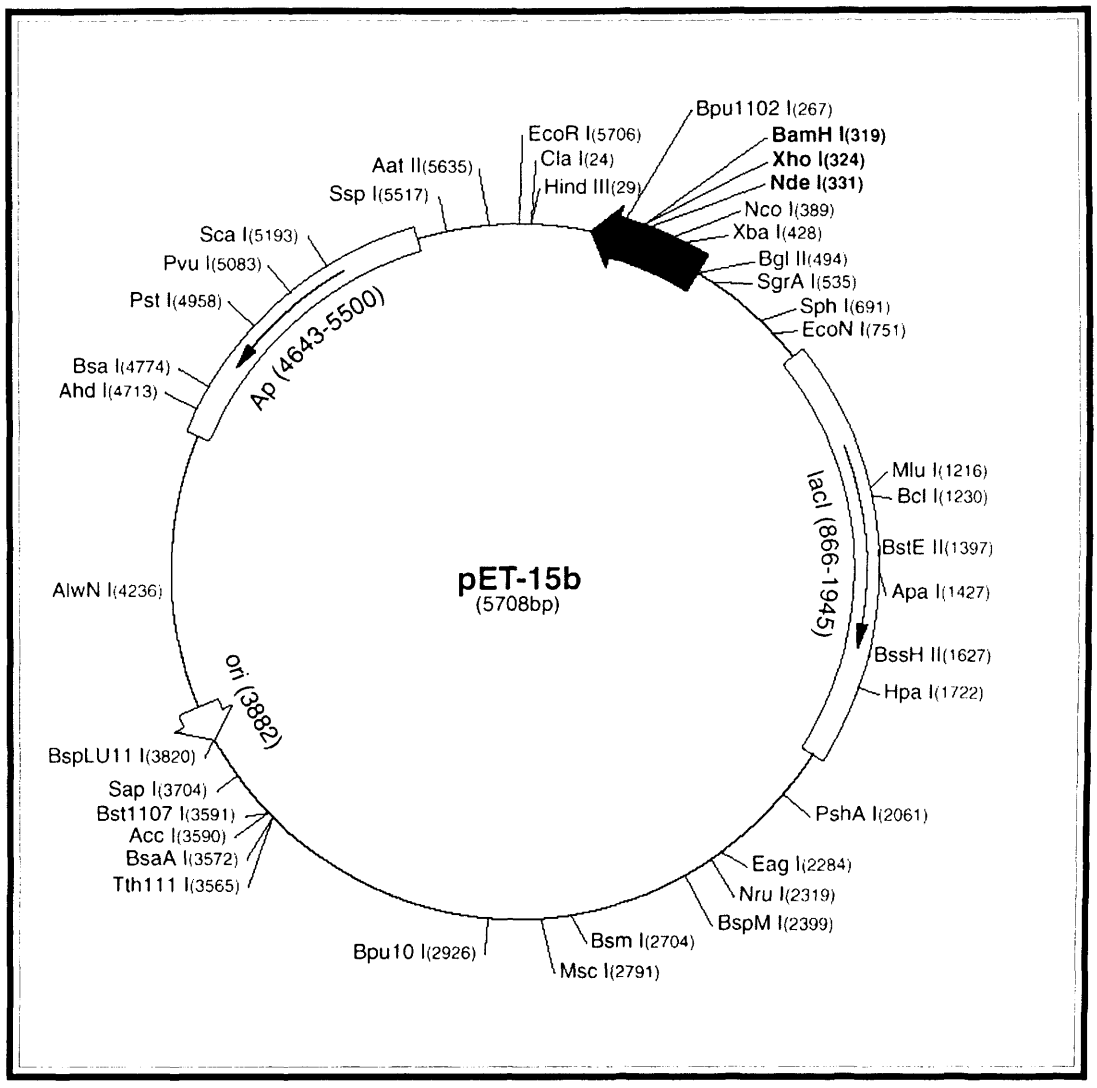

Figure X.1 : pET-15b plasmid map. The pET-15b vector where calreticulin clones have been inserted between the *Bam*HI and *Nde*I restriction sites.

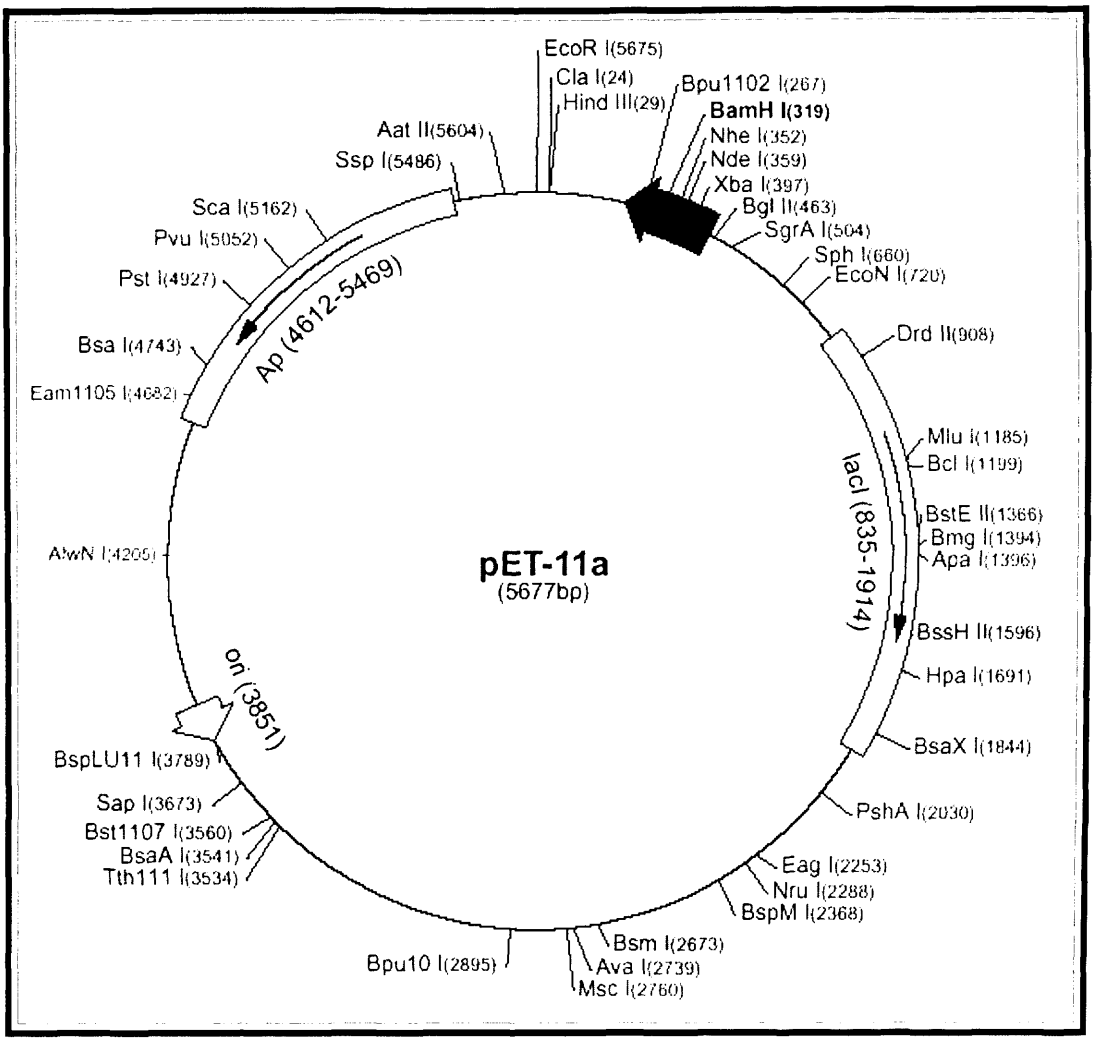

Figure X.2 : The pET-11 plasmid map. The pET-11a plasmid where calreticulin clones have been inserted between the *Bam*HI and *Nde*I restriction sites.

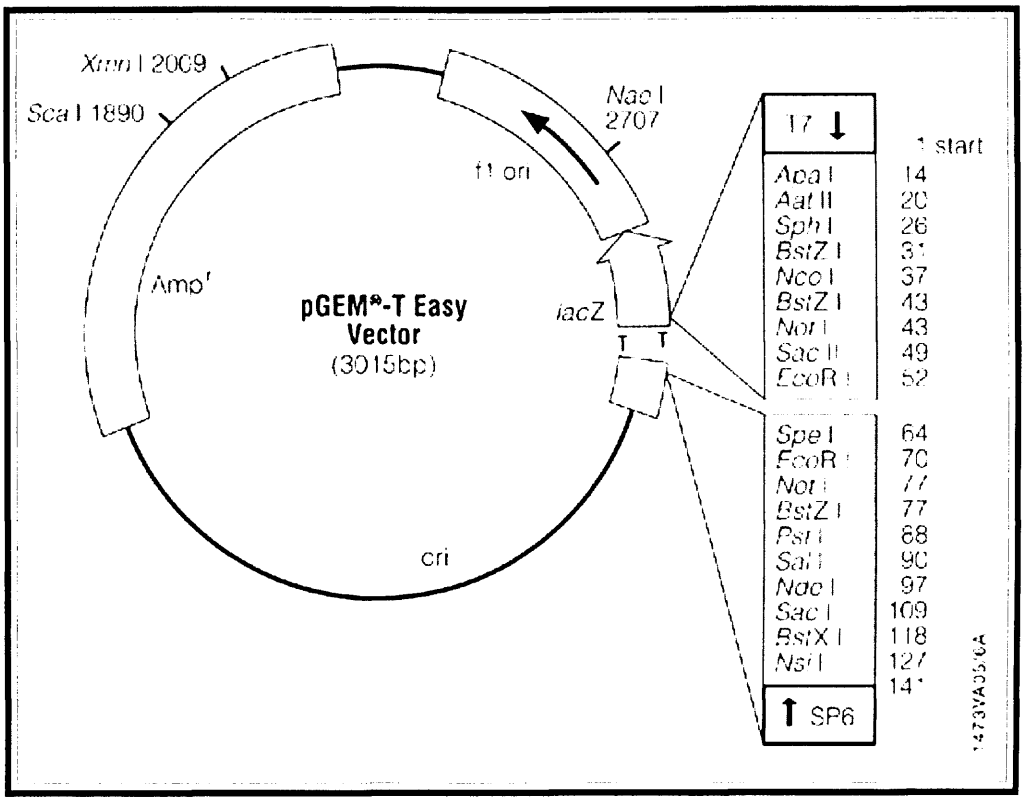

Figure X.3 : The pGEM-T Easy plasmid map. The pGEM-T Easy plasmid where the PCR clones have been inserted between the two single 3'-T overhangs.

X.3 Recipes (alphabetical order)

20% APS :
100 mg Ammonium persulphate
Complete to 0.5 ml dH_2O.

DGVB^{++}:

71 mM	NaCl
2.5 mM	sodium-diethylbarbiturate
0.05% (v/v)	gelatin
3% (w/v)	D-glucose
1 mM	$MgCl_2$
0.15 mM	$CaCl_2$

LB Medium :
10 g Tryptone
5 g Yeast extract
10 g NaCl
Complete to 1 litre with dH_20.
Autoclave and store at RT.

LB Agar :
As LB medium, plus;
15 g Agar
Complete to 1 litre with dH_20.
Autoclave and store at 4°C.

Lower Tris 4x :
90.9 g Tris
2 g SDS
Adjust pH to 8.8 with HCl.
Complete to 0.5 litre with dH_2O.

PBS (phosphate buffer saline) :
150 mM NaCl
10 mM Na_2HPO_4
5 mM KH_2PO_4
Adjust pH to 7.4.

Protein sample buffer 5x :
5.8 ml Upper Tris
2.5 ml Glycerol
0.83 g SDS
1 mg Bromophenol blue
Complete to 10 ml with dH_2O.
Add 5% (v/v) β-mercaptoethanol before use.

Table X.1 : Minimal growth medium (MME). Recipe for 1 L of supplemented minimal growth medium and 1 L of M9 salts (autoclaved).

Minimal growth medium	
M9 Salts	1000.0 ml
D-glucose stock (20 g/100ml)(0.2 μm filter sterilised)[a]	20.0 ml
100 x BME vitamins[b]	10.0 ml
1 M $MgSO_4$ (autoclaved)	2.0 ml
1 M $CaCl_2$ (autoclaved)	0.1 ml
Ampicillin stock (100 mg/ml)(in 20% ethanol)	1.0 ml
1 L M9 Salts (autoclaved)	
KH_2PO_4	7.5 g
$Na_2HPO_4 \cdot 7H_2O$	12.5 g
NH_4Cl[a] (pH adjusted to 7.2 with NaOH)	1 g

[a] For isotopic labelling $^{15}NH_4Cl$ and ^{13}C D-glucose were substituted for unlabeled NH_4Cl and D-glucose respectively.
[b] Basal Vitamins Eagle Media.
The above recipe is from reference [134].

SDS-PAGE destaining solution :
50 ml acetic acid
25 ml methanol
425 ml dH_2O

SDS-PAGE running buffer 10x :
30.2 g Tris
144 g Glycine
8 g SDS
Complete to 1 litre with dH_2O.

SDS-PAGE Staining solution :
225 ml methanol
225 ml dH_2O
50 ml acetic acid
125 mg Coomassie blue

Table X.2 : Separating gel.

Components	Volume required for gels (ml)	
	12%	15%
Protogel	4.0	5.0
Lower Tris 4x	2.5	2.5
dH_2O	3.5	2.5

Immediately before use, add 15 µl of TEMED and 15 µl of 20% APS.

Table X.3 : Stacking gel.

Components	Volume required for gels (ml)	
	4.5%	6%
Protogel	1.5	2.0
Upper Tris 4x	2.5	2.5
dH_2O	6.0	5.5

Immediately before use, add 15 µl of TEMED and 40 µl of 20% APS.

TAE 10x :
48.4 g Tris
20 ml EDTA (0.5 M, pH 8.0)
11.42 ml glacial acetic acid
Adjust the pH to 7.2 with HCl.
Complete to 1 litre.

TBS (Tris buffered saline):
150 mM NaCl
140mM KCl
10mM Na_2HPO_4
5mM KH_2PO_4, pH 7.4.
Autoclave and store at RT

TE :
10 mM Tris-HCl
1 mM EDTA
Adjust ph to 8.0 with HCl.

Upper Tris 4x :
30.2 g Tris
2 g SDS
Adjust the pH to 6.8 with HCl.
Complete to 0.5 litre with dH_2O.

X.4 The cellular transformation and the glycerol storage protocol

Competent BL21(DE3) and TOP10F' cells were prepared with the calcium chloride method [95]. A 50 µl aliquot of competent cells was thaw on ice. Approximately 10 ng of DNA was added to the competent cells and mixed by gently swirling the pipette tip followed by 30 minutes incubation on ice. The cells were then heat shocked for 5 minutes at 37°C and placed on ice to cool for 2 minutes. 0.75 ml of LB media was then added to the cells and incubated for 1 hour at 37°C with shaking (200 rpm). Following incubation, 50 and 200 µl of the transformation mixture was plated on LB-Amp plates and incubated O/N at 37°C.

Glycerol stocks of all transformed cells were made by growing the cells up to A_{600} of 0.6 (37°C, vigorous shaking). 0.86 ml of the culture was transferred to a cryovial and mixed gently with 0.14 ml of sterile 50% glycerol for a final 7% glycerol stock. The cells were then transferred to a -80°C freezer.

X.5 Plasmid preparation

Plasmids were transformed into non-expressing strain, TOP10F', in accordance with the transformation protocol described in appendix 1.2. 5 ml volume of LB media was inoculated from a single colony and the culture allowed to attain stationary phase overnight by incubation at 37°C and 200 x rpm shaking. The plasmid was purified using the QIAGEN plasmid mini kit, for a detailed protocol refer to the kit. The Plasmid was eluted in 50 µl of buffer EB (10 mM Tris-HCl, pH 8.5) and quantified using the standard of 50 µg/ml of double stranded DNA creates an A_{260nm} of 1.

X.6 Reagents

Unless stated otherwise, all reagents were purchased from Fisher, Loughborough, England, and were of analytical reagent quality or better.

MES Hydrate, Basal Vitamins Eagle Media (catno. B-6891), and BIS-TRIS. Sigma-aldrich, UK.

Ammonium-N^{15} Chloride and C^{13}-D-Glucose. Cambridge isotope laboratories, inc., Andover, USA.

Deuterium oxide, and precision NMR sample tubes. GOSS Scientific Instruments Ltd.

Protogel (30% (w/v) Acrylamide: 0.8% (w/v) Bis-Acrylamide). National diagnostics, Hull, England.

β-Mercaptoethanol, and SDS-PAGE standards (low range). Bio-Rad Laboratories Ltd., Hemel Hempstead, England.

Complete mini EDTA-free protease inhibitor cocktail tablets, Thrombin from human plasma, and DNase I (RNase-free). Roche Diagnostics GmbH, Mannheim, germany.

Molecular biology enzymes, *Bam*HI, *Nde*I, *Bgl*II, *taq* DNA polymerase, Promega.

The following reagent kits were obtained from the sources indicated.

BugBuster plus Benzonase	CN Biosciences Ltd, Nottingham, UK.
Non-Interfering protein assay	CN Biosciences Ltd, Nottingham, UK.
QIAGEN plasmid mini kit	QIAGEN Ltd., West Sussex, England.
QIAquick gel extraction kit	QIAGEN Ltd., West Sussex, England.

X.7 Recombinant proteins theoretical molar absorbance coefficient

The protein concentration of pure CRT clones was determined using the extinction coefficient A_{280nm} as described in table X.4. The extinction coefficient were calculated using the Gill and Von Hippel relation [135].

Table X.4 : The molar extinction coefficients of the calreticulin domains.

	$A_{280\ nm}$ ($M^{-1}cm^{-1}$)
S-domain	38 040
S-domain 2	37 980
P-domain	46 230
PS-domain	48 790

X.8 Relaxation data

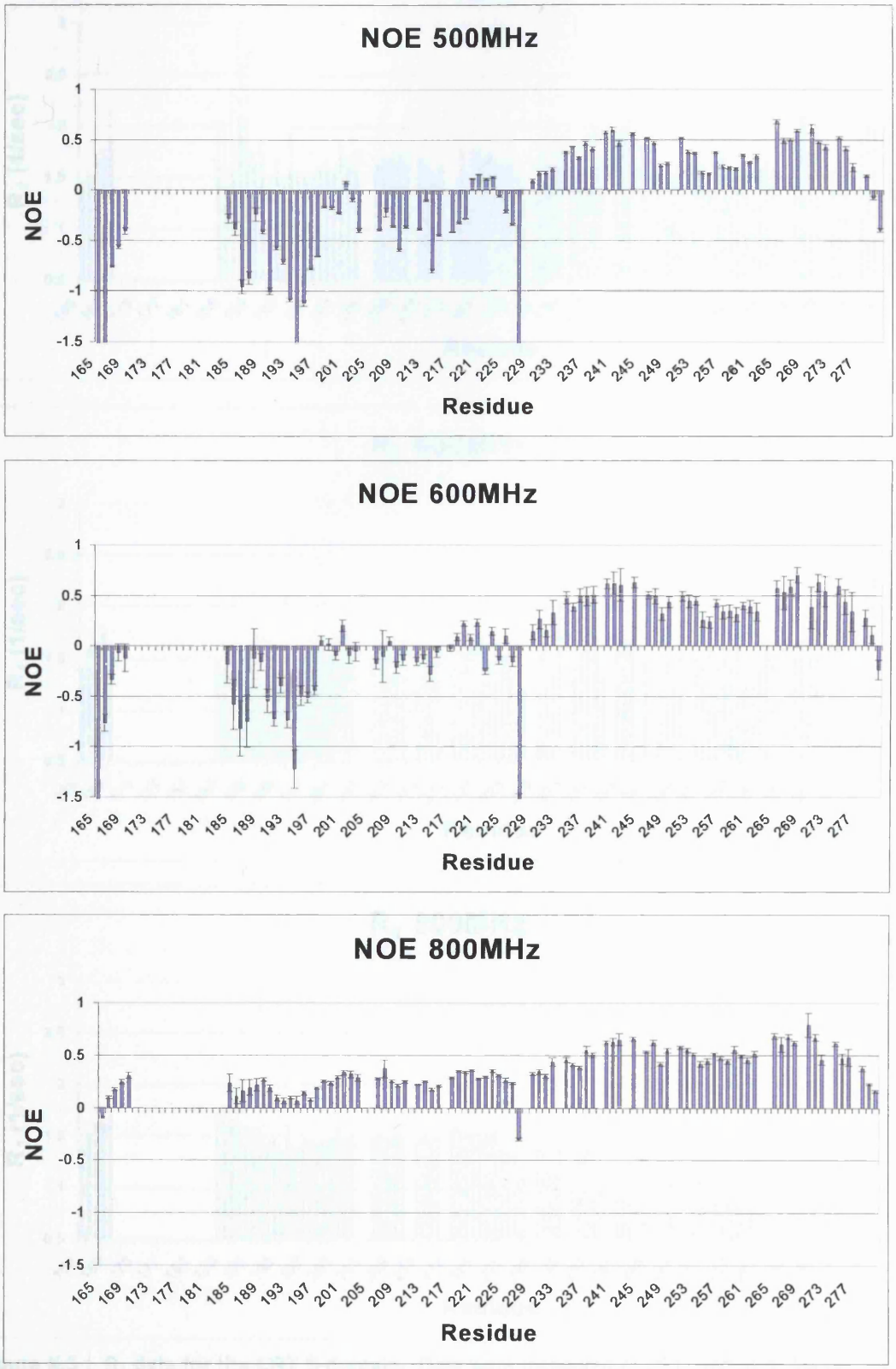

Figure X.4 : ^{15}N NOE data for the CRT S-domain. Data were measured at 25°C and at ^{1}H frequencies of 500, 600 and 800 MHz.

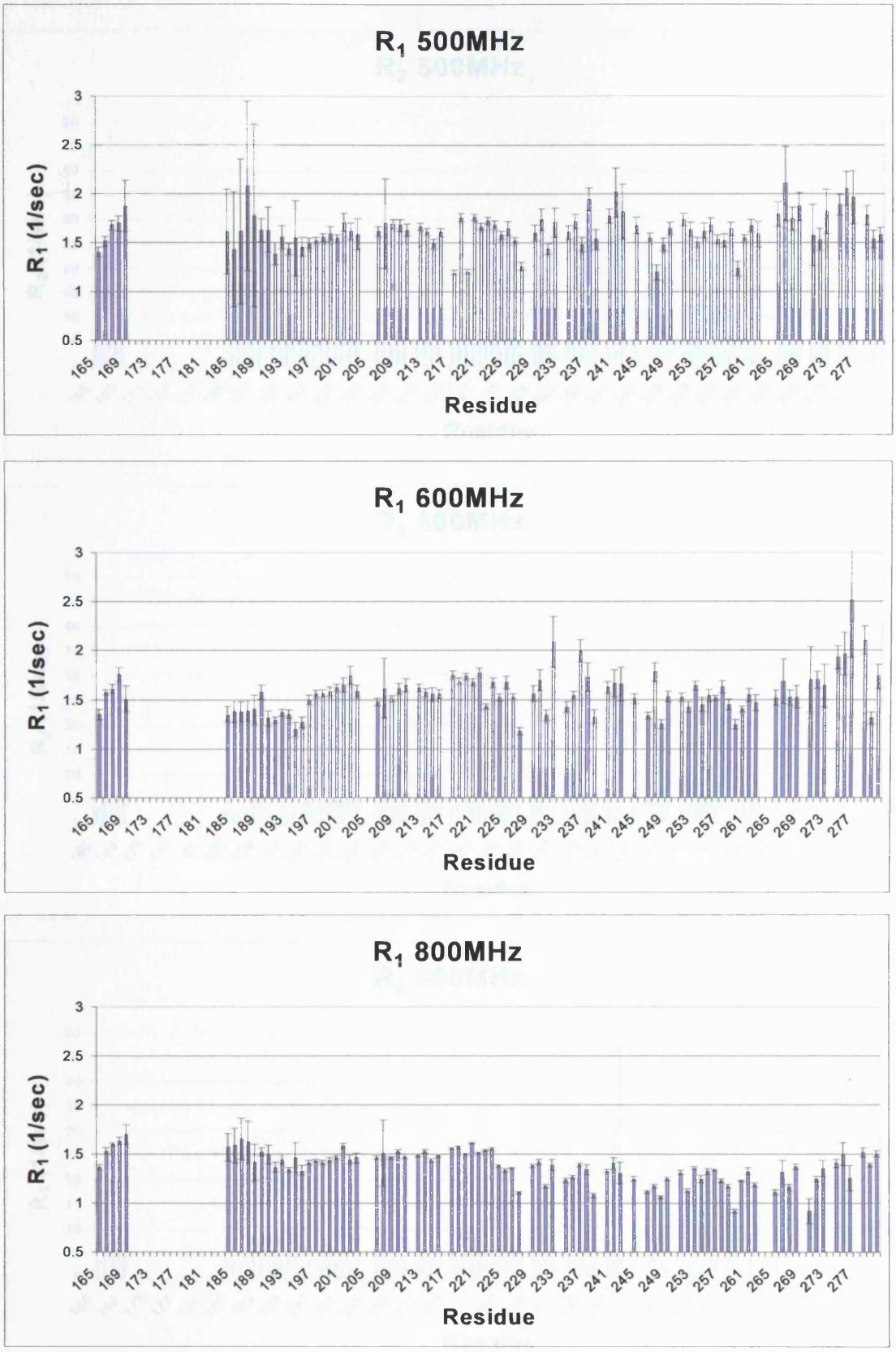

Figure X.5 : R_1 data for the CRT S-domain. Data were measured at 25°C and at ^1H frequencies of 500, 600 and 800 MHz.

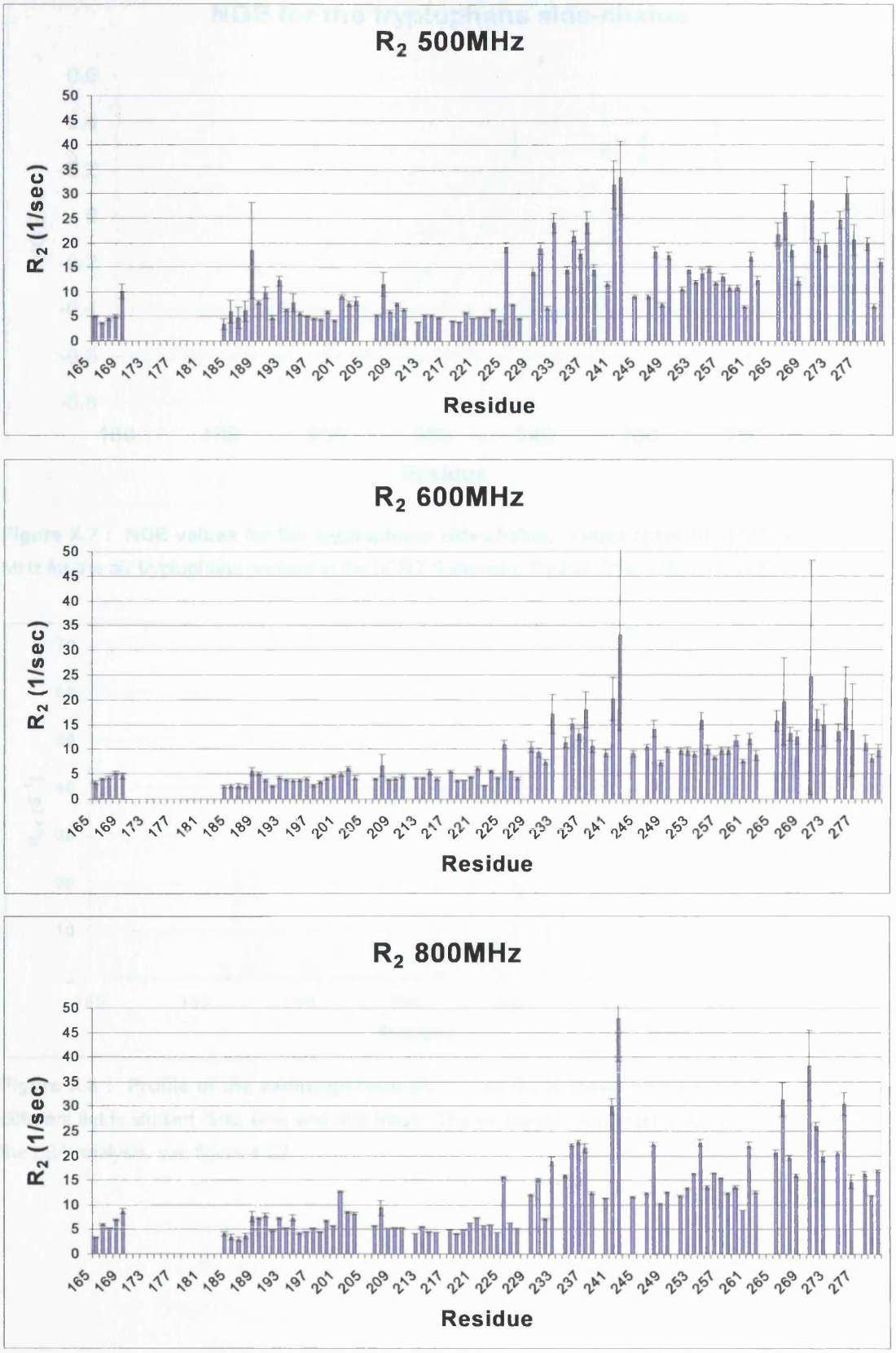

Figure X.6 : R_2 data for the CRT S-domain. Data were measured at 25°C and at ^1H frequencies of 500, 600 and 800 MHz.

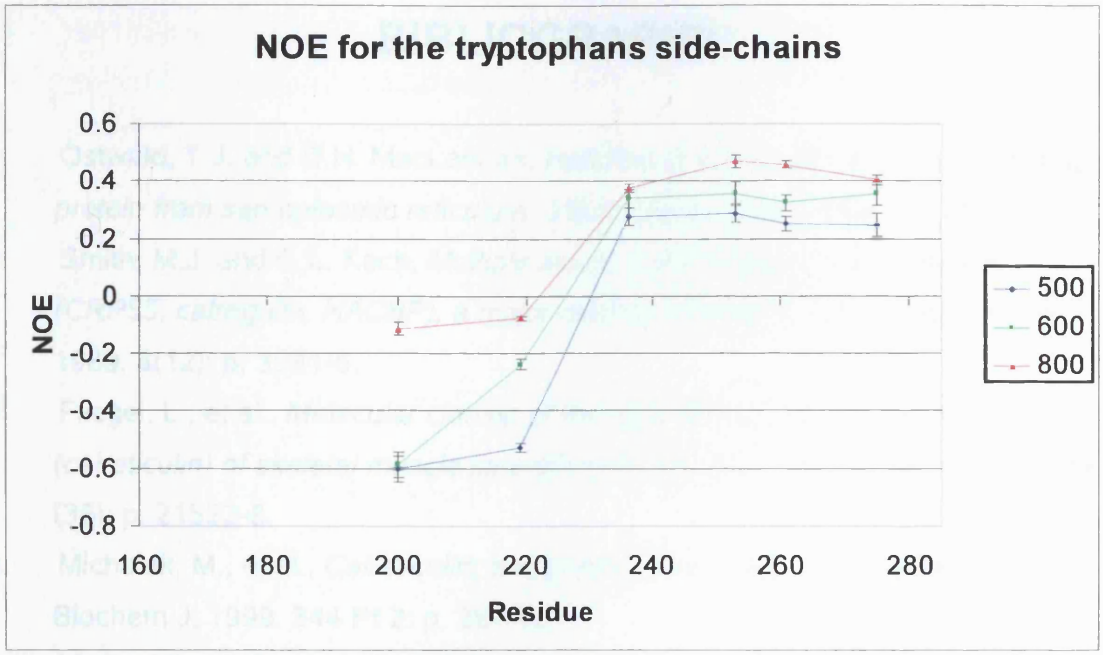

Figure X.7 : NOE values for the tryptophans side-chains. Values obtained at 500, 600, and 800 MHz for the six tryptophans present in the hCRT S-domain; Trp200, 219, 236, 253, 261, and 275.

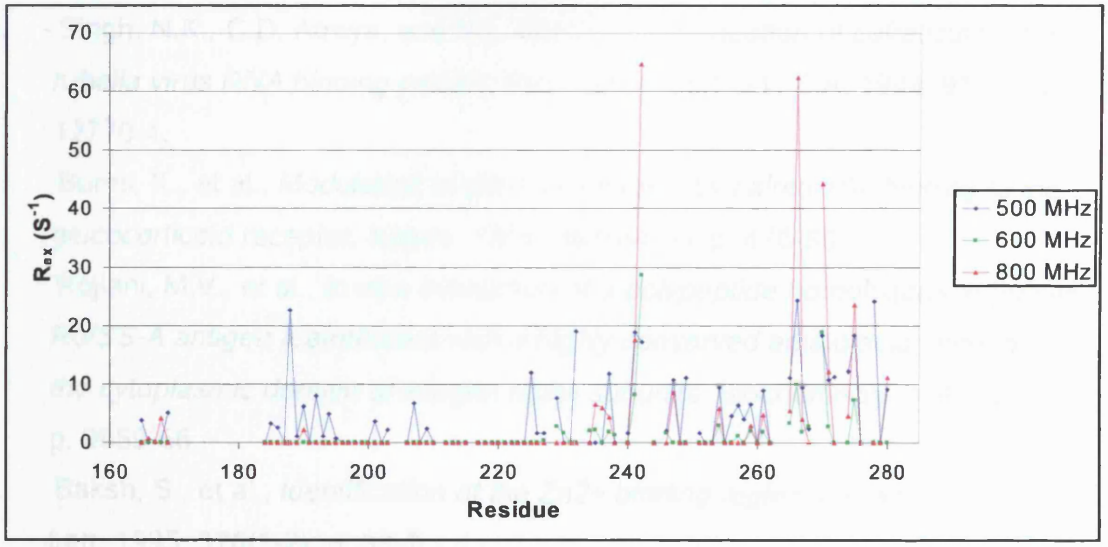

Figure X.8 : Profile of the exchange term (R_{ex}). The R_{ex} is shown for each residues at the three different fields studied (500, 600, and 800 Mhz). The exchange term applies to the models 2 and 4 of the LSA analysis, see figure 4.32.

BIBLIOGRAPHY

1. Ostwald, T.J. and D.H. MacLennan, *Isolation of a high affinity calcium-binding protein from sarcoplasmic reticulum.* J Biol Chem, 1974. **249**(3): p. 974-9.
2. Smith, M.J. and G.L. Koch, *Multiple zones in the sequence of calreticulin (CRP55, calregulin, HACBP), a major calcium binding ER/SR protein.* Embo J, 1989. **8**(12): p. 3581-6.
3. Fliegel, L., et al., *Molecular cloning of the high affinity calcium-binding protein (calreticulin) of skeletal muscle sarcoplasmic reticulum.* J Biol Chem, 1989. **264** (36): p. 21522-8.
4. Michalak, M., et al., *Calreticulin: one protein, one gene, many functions.* Biochem J, 1999. **344 Pt 2**: p. 281-92.
5. McCauliffe, D.P., et al., *The 5'-flanking region of the human calreticulin gene shares homology with the human GRP78, GRP94, and protein disulfide isomerase promoters.* J Biol Chem, 1992. **267**(4): p. 2557-62.
6. Singh, N.K., C.D. Atreya, and H.L. Nakhasi, *Identification of calreticulin as a rubella virus RNA binding protein.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12770-4.
7. Burns, K., et al., *Modulation of gene expression by calreticulin binding to the glucocorticoid receptor.* Nature, 1994. **367**(6462): p. 476-80.
8. Rojiani, M.V., et al., *In vitro interaction of a polypeptide homologous to human Ro/SS-A antigen (calreticulin) with a highly conserved amino acid sequence in the cytoplasmic domain of integrin alpha subunits.* Biochemistry, 1991. **30**(41): p. 9859-66.
9. Baksh, S., et al., *Identification of the Zn2+ binding region in calreticulin.* FEBS Lett, 1995. **376**(1-2): p. 53-7.
10. Baksh, S., et al., *Interaction of calreticulin with protein disulfide isomerase.* J Biol Chem, 1995. **270**(52): p. 31338-44.
11. Corbett, E.F., et al., *Ca2+ regulation of interactions between endoplasmic reticulum chaperones.* J Biol Chem, 1999. **274**(10): p. 6203-11.
12. Corpet, F., *Multiple sequence alignment with hierarchical clustering.* Nucleic Acids Res, 1988. **16**(22): p. 10881-90.
13. Baksh, S. and M. Michalak, *Expression of calreticulin in Escherichia coli and identification of its Ca2+ binding domains.* J Biol Chem, 1991. **266**(32): p. 21458-65.

14. Vassilakos, A., et al., *Oligosaccharide binding characteristics of the molecular chaperones calnexin and calreticulin.* Biochemistry, 1998. **37**(10): p. 3480-90.
15. Bergeron, J.J., et al., *Calnexin: a membrane-bound chaperone of the endoplasmic reticulum.* Trends Biochem Sci, 1994. **19**(3): p. 124-8.
16. Andrin, C., et al., *Interaction between a Ca2+-binding protein calreticulin and perforin, a component of the cytotoxic T-cell granules.* Biochemistry, 1998. **37** (29): p. 10386-94.
17. Kuwabara, K., et al., *Calreticulin, an antithrombotic agent which binds to vitamin K-dependent coagulation factors, stimulates endothelial nitric oxide production, and limits thrombosis in canine coronary arteries.* J Biol Chem, 1995. **270**(14): p. 8179-87.
18. Dai, E., et al., *Calreticulin, a potential vascular regulatory protein, reduces intimal hyperplasia after arterial injury.* Arterioscler Thromb Vasc Biol, 1997. **17** (11): p. 2359-68.
19. Stuart, G.R., et al., *Localisation of the C1q binding site within C1q receptor/calreticulin.* FEBS Lett, 1996. **397**(2-3): p. 245-9.
20. van den Berg, R.H., et al., *Regulation of the function of the first component of complement by human C1q receptor.* Eur J Immunol, 1995. **25**(8): p. 2206-10.
21. Stuart, G.R., et al., *The C1q and collectin binding site within C1q receptor (cell surface calreticulin).* Immunopharmacology, 1997. **38**(1-2): p. 73-80.
22. Opas, M., et al., *Regulation of expression and intracellular distribution of calreticulin, a major calcium binding protein of nonmuscle cells.* J Cell Physiol, 1991. **149**(1): p. 160-71.
23. Michalak, M., et al., *Endoplasmic reticulum form of calreticulin modulates glucocorticoid-sensitive gene expression.* J Biol Chem, 1996. **271**(46): p. 29436-45.
24. Gomord, V., et al., *The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER.* Plant J, 1997. **11**(2): p. 313-25.
25. White, T.K., Q. Zhu, and M.L. Tanzer, *Cell surface calreticulin is a putative mannoside lectin which triggers mouse melanoma cell spreading.* J Biol Chem, 1995. **270**(27): p. 15926-9.
26. Gray, A.J., et al., *The mitogenic effects of the B beta chain of fibrinogen are mediated through cell surface calreticulin.* J Biol Chem, 1995. **270**(44): p. 26602-6.

27. Essex, D.W., K. Chen, and M. Swiatkowska, *Localization of protein disulfide isomerase to the external surface of the platelet plasma membrane.* Blood, 1995. **86**(6): p. 2168-73.
28. Wiest, D.L., et al., *Incomplete endoplasmic reticulum (ER) retention in immature thymocytes as revealed by surface expression of "ER-resident" molecular chaperones.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1884-9.
29. Seddiki, N., et al., *Calreticulin, a potential cell surface receptor involved in cell penetration of anti-DNA antibodies.* J Immunol, 2001. **166**(10): p. 6423-9.
30. Arosa, F.A., et al., *Calreticulin is expressed on the cell surface of activated human peripheral blood T lymphocytes in association with major histocompatibility complex class I molecules.* J Biol Chem, 1999. **274**(24): p. 16917-22.
31. Simovich, M.J., et al., *Cellular location of proteins related to iron absorption and transport.* Am J Hematol, 2002. **69**(3): p. 164-70.
32. Ghiran, I., L.B. Klickstein, and A. Nicholson-Weller, *Calreticulin is at the surface of circulating neutrophils and uses CD59 as an adaptor molecule.* J Biol Chem, 2003. **278**(23): p. 21024-31.
33. Orr, A.W., et al., *Low density lipoprotein receptor-related protein is a calreticulin coreceptor that signals focal adhesion disassembly.* J Cell Biol, 2003. **161**(6): p. 1179-89.
34. Xiao, G., et al., *Calreticulin is transported to the surface of NG108-15 cells where it forms surface patches and is partially degraded in an acidic compartment.* J Neurosci Res, 1999. **58**(5): p. 652-62.
35. Xiao, G., et al., *KDEL proteins are found on the surface of NG108-15 cells.* Brain Res Mol Brain Res, 1999. **72**(2): p. 121-8.
36. Sim, R.B., et al., *Interaction of C1q and the collectins with the potential receptors calreticulin (cC1qR/collectin receptor) and megalin.* Immunobiology, 1998. **199**(2): p. 208-24.
37. Ogden, C.A., et al., *C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells.* J Exp Med, 2001. **194**(6): p. 781-95.
38. Vandivier, R.W., et al., *Role of Surfactant Proteins A, D, and C1q in the clearance of Apoptotic Cells In Vivo and In Vitro: Calreticulin and CD91 as a Common Collectin Receptor Complex.* The Journal of Immunology, 2002. **169**: p. 3978–3986.

39. Kishore, U., et al., *Release of calreticulin from neutrophils may alter C1q-mediated immune functions.* Biochem J, 1997. **322 (Pt 2)**: p. 543-50.
40. Steino, A., et al., *Interaction of C1q with the receptor calreticulin requires a conformational change in C1q.* Scand J Immunol, 2004. **59**(5): p. 485-95.
41. Kostova, Z. and D.H. Wolf, *For whom the bell tolls: protein quality control of the endoplasmic reticulum and the ubiquitin-proteasome connection.* Embo J, 2003. **22**(10): p. 2309-17.
42. Spiro, R.G., et al., *Definition of the lectin-like properties of the molecular chaperone, calreticulin, and demonstration of its copurification with endomannosidase from rat liver Golgi.* J Biol Chem, 1996. **271**(19): p. 11588-94.
43. Petrescu, A.J., et al., *The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding.* Embo J, 1997. **16**(14): p. 4302-10.
44. Patil, A.R., C.J. Thomas, and A. Surolia, *Kinetics and the mechanism of interaction of the endoplasmic reticulum chaperone, calreticulin, with monoglucosylated (Glc1Man9GlcNAc2) substrate.* J Biol Chem, 2000. **275**(32): p. 24348-56.
45. Sousa, M. and A.J. Parodi, *The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase.* Embo J, 1995. **14**(17): p. 4196-203.
46. Peterson, J.R. and A. Helenius, *In vitro reconstitution of calreticulin-substrate interactions.* J Cell Sci, 1999. **112 (Pt 16)**: p. 2775-84.
47. Danilczyk, U.G., M.F. Cohen-Doyle, and D.B. Williams, *Functional relationship between calreticulin, calnexin, and the endoplasmic reticulum luminal domain of calnexin.* J Biol Chem, 2000. **275**(17): p. 13089-97.
48. Saito, Y., et al., *Calreticulin functions in vitro as a molecular chaperone for both glycosylated and non-glycosylated proteins.* Embo J, 1999. **18**(23): p. 6718-29.
49. Molinari, M. and A. Helenius, *Chaperone selection during glycoprotein translocation into the endoplasmic reticulum.* Science, 2000. **288**(5464): p. 331-3.
50. Oliver, J.D., et al., *ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin.* Mol Biol Cell, 1999. **10**(8): p. 2573-82.
51. Frickel, E.M., et al., *TROSY-NMR reveals interaction between ERp57 and the*

tip of the calreticulin P-domain. Proc Natl Acad Sci U S A, 2002. **99**(4): p. 1954-9.

52. Alonso, M.T., et al., *Ca2+-induced Ca2+ release in chromaffin cells seen from inside the ER with targeted aequorin.* J Cell Biol, 1999. **144**(2): p. 241-54.
53. Miyawaki, A., et al., *Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin.* Nature, 1997. **388**(6645): p. 882-7.
54. Krause, K.H. and M. Michalak, *Calreticulin.* Cell, 1997. **88**(4): p. 439-43.
55. Opas, M., et al., *Calreticulin modulates cell adhesiveness via regulation of vinculin expression.* J Cell Biol, 1996. **135**(6 Pt 2): p. 1913-23.
56. Mery, L., et al., *Overexpression of calreticulin increases intracellular Ca2+ storage and decreases store-operated Ca2+ influx.* J Biol Chem, 1996. **271**(16): p. 9332-9.
57. Mesaeli, N., et al., *Calreticulin is essential for cardiac development.* J Cell Biol, 1999. **144**(5): p. 857-68.
58. Coppolino, M.G., et al., *Calreticulin is essential for integrin-mediated calcium signalling and cell adhesion.* Nature, 1997. **386**(6627): p. 843-7.
59. Ellgaard, L., et al., *NMR structure of the calreticulin P-domain.* Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3133-8.
60. Li, Z., W.F. Stafford, and M. Bouvier, *The metal ion binding properties of calreticulin modulate its conformational flexibility and thermal stability.* Biochemistry, 2001. **40**(37): p. 11193-201.
61. Rauch, F., et al., *Heart, brain, and body wall defects in mice lacking calreticulin.* Exp Cell Res, 2000. **256**(1): p. 105-11.
62. Eggleton, P., et al., *Fine specificity of autoantibodies to calreticulin: epitope mapping and characterization.* Clin Exp Immunol, 2000. **120**(2): p. 384-91.
63. Verreck, F.A., et al., *DR4Dw4/DR53 molecules contain a peptide from the autoantigen calreticulin.* Tissue Antigens, 1995. **45**(4): p. 270-5.
64. Tuckova, L., et al., *Anti-gliadin antibodies in patients with celiac disease cross-react with enterocytes and human calreticulin.* Clin Immunol Immunopathol, 1997. **85**(3): p. 289-96.
65. Orth, T., et al., *Complete congenital heart block is associated with increased autoantibody titers against calreticulin.* Eur J Clin Invest, 1996. **26**(3): p. 205-15.
66. Gut, J., U. Christen, and J. Huwyler, *Mechanisms of halothane toxicity: novel insights.* Pharmacol Ther, 1993. **58**(2): p. 133-55.
67. Walport, M.J., K.A. Davies, and M. Botto, *C1q and systemic lupus*

erythematosus. Immunobiology, 1998. **199**(2): p. 265-85.

68. Kovacs, H., et al., *Evidence that C1q binds specifically to CH2-like immunoglobulin gamma motifs present in the autoantigen calreticulin and interferes with complement activation*. Biochemistry, 1998. **37**(51): p. 17865-74.
69. McCauliffe, D.P., et al., *A human Ro/SS-A autoantigen is the homologue of calreticulin and is highly homologous with onchocercal RAL-1 antigen and an aplysia "memory molecule"*. J Clin Invest, 1990. **86**(1): p. 332-5.
70. Pritchard, D.I., et al., *A hookworm allergen which strongly resembles calreticulin*. Parasite Immunol, 1999. **21**(9): p. 439-50.
71. Marcelain, K., et al., *Development of an immunoenzymatic assay for the detection of human antibodies against Trypanosoma cruzi calreticulin, an immunodominant antigen*. Acta Trop, 2000. **75**(3): p. 291-300.
72. Sanders, M.L., et al., *Antibody to a cDNA-derived calreticulin protein from Amblyomma americanum as a biomarker of tick exposure in humans*. Am J Trop Med Hyg, 1998. **59**(2): p. 279-85.
73. Kasper, G., et al., *A calreticulin-like molecule from the human hookworm Necator americanus interacts with C1q and the cytoplasmic signalling domains of some integrins*. Parasite Immunol, 2001. **23**(3): p. 141-52.
74. Eggleton, P., et al., *Calreticulin is released from activated neutrophils and binds to C1q and mannan-binding protein*. Clin Immunol Immunopathol, 1994. **72**(3): p. 405-9.
75. Arvieux, J., et al., *Characterization of the C1q receptor on a human macrophage cell line, U937*. Biochem J, 1984. **218**(2): p. 547-55.
76. Malhotra, R., et al., *Localization of the receptor-binding site in the collectin family of proteins*. Biochem J, 1993. **293 (Pt 1)**: p. 15-9.
77. Malhotra, R., et al., *Human leukocyte C1q receptor binds other soluble proteins with collagen domains*. J Exp Med, 1990. **172**(3): p. 955-9.
78. Law, S.K.A. and K.B.M. Reid, *Complement 2nd edition*, ed. O. IRL. 1996. 1-88.
79. Reid, K.B., *Proteins involved in the activation and control of the two pathways of human complement*. Biochem Soc Trans, 1983. **11**(1): p. 1-12.
80. Dodds, A. and R. Sim, *Complement : a practical approach*. The practical approach series. 1997, Oxford: Oxford University Press 1997.
81. Kawasaki, N., T. Kawasaki, and I. Yamashina, *Isolation and characterization of a mannan-binding protein from human serum*. J Biochem (Tokyo), 1983. **94**(3): p. 937-47.

82. Sheriff, S., C.Y. Chang, and R.A. Ezekowitz, *Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil.* Nat Struct Biol, 1994. **1**(11): p. 789-94.
83. Schweinle, J.E., et al., *Truncated forms of mannose-binding protein multimerize and bind to mannose-rich Salmonella montevideo but fail to activate complement in vitro.* J Biol Chem, 1993. **268**(1): p. 364-70.
84. Schuffenecker, I., et al., *The gene for mannose-binding protein maps to chromosome 10 and is a marker for multiple endocrine neoplasia type 2.* Cytogenet Cell Genet, 1991. **56**(2): p. 99-102.
85. Thiel, S., et al., *A second serine protease associated with mannan-binding lectin that activates complement.* Nature, 1997. **386**(6624): p. 506-10.
86. Matsushita, M. and T. Fujita, *Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation.* Immunobiology, 1995. **194**(4-5): p. 443-48.
87. Gadjeva, M., S. Thiel, and J.C. Jensenius, *The mannan-binding-lectin pathway of the innate immune response.* Curr Opin Immunol, 2001. **13**(1): p. 74-8.
88. Super, M., et al., *Association of low levels of mannan-binding protein with a common defect of opsonisation.* Lancet, 1989. **2**(8674): p. 1236-9.
89. Matsushita, M., R.A. Ezekowitz, and T. Fujita, *The Gly-54-->Asp allelic form of human mannose-binding protein (MBP) fails to bind MBP-associated serine protease.* Biochem J, 1995. **311 (Pt 3)**: p. 1021-3.
90. Sumiya, M., et al., *Molecular basis of opsonic defect in immunodeficient children.* Lancet, 1991. **337**(8757): p. 1569-70.
91. Lipscombe, R.J., et al., *High frequencies in African and non-African populations of independent mutations in the mannose binding protein gene.* Hum Mol Genet, 1992. **1**(9): p. 709-15.
92. Fujita, T., M. Matsushita, and Y. Endo, *The lectin-complement pathway--its role in innate immunity and evolution.* Immunol Rev, 2004. **198**: p. 185-202.
93. Schwaeble, W., et al., *The mannan-binding lectin-associated serine proteases (MASPs) and MAp19: four components of the lectin pathway activation complex encoded by two genes.* Immunobiology, 2002. **205**(4-5): p. 455-66.
94. Teh, C., et al., *M-ficolin is expressed on monocytes and is a lectin binding to N-acetyl-D-glucosamine and mediates monocyte adhesion and phagocytosis of Escherichia coli.* Immunology, 2000. **101**(2): p. 225-32.
95. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular cloning : a laboratory*

manual. 2nd ed. 1989, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

96. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*. Nature, 1970. **227**(259): p. 680-5.
97. Creighton, T.E., *Protein structure : a practical approach*. 1989, Oxford ; Washington, DC: IRL Press. xviii, 355.
98. Greenfield, N. and G.D. Fasman, *Computed circular dichroism spectra for the evaluation of protein conformation*. Biochemistry, 1969. **8**(10): p. 4108-16.
99. Khanna, N.C., M. Tokuda, and D.M. Waisman, *Conformational changes induced by binding of divalent cations to calregulin*. J Biol Chem, 1986. **261**(19): p. 8883-7.
100. Vincenzetti, S., et al., *Recombinant human cytidine deaminase: expression, purification, and characterization*. Protein Expr Purif, 1996. **8**(2): p. 247-53.
101. Giacovazzo, C., et al., *Fundamentals of Crystallography*. Second editions ed, ed. C. Giacovazzo. 2002: Oxford University Press.
102. Ducruix, A. and R. Giege, *Crystallization of nucleic acids and proteins : apractical approach*. Practical approach series. 1992, Oxford: IRL Press at Oxford University Press.
103. Beynon, R.J. and J.S. Bond, *Proteolytic enzymes : a practical approach*. 1990, Oxford: IRL Press.
104. Cavanagh, J., *Protein NMR spectroscopy : principles and practice*. 1996, San Diego: Academic Press. xxiii, 587.
105. Nilges, M., et al., *Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin*. J Mol Biol, 1997. **269**(3): p. 408-22.
106. Derome, A.E., *Modern NMR techniques for chemistry research*, ed. J.E. Baldwin. Vol. 6. 1987, Oxford: Pergamon Press. 85-109.
107. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity*. J. Am. Chem. Soc., 1982. **104**: p. 4546-4570.
108. Reed, M.A., et al., *Effects of domain dissection on the folding and stability of the 43 kDa protein PGK probed by NMR*. J Mol Biol, 2003. **330**(5): p. 1189-201.
109. Ishima, R., K. Yamasaki, and K. Nagayama, *Application of the quasi-spectral density function of 15N nuclei to the selection of a motional model for model-free analysis.* Journal of Biomolecular NMR, 1995. **6**: p. 423-6.
110. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system*

based on UNIX pipes. J Biomol NMR, 1995. **6**(3): p. 277-93.

111. Johnson, B.A. and R.A. Blevins, *Software for the Visualization and Analysis of Nuclear Magnetic Resonance Data.* Journal of Biomolecular NMR, 1994. **4**: p. 603-14.
112. Cornilescu, G., F. Delaglio, and A. Bax, *Protein backbone angle restraints from searching a database for chemical shift and sequence homology.* J Biomol NMR, 1999. **13**(3): p. 289-302.
113. Kay, L.E., D.A. Torchia, and A. Bax, *Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease.* Biochemistry, 1989. **28**(23): p. 8972-9.
114. Kay, L.E., et al., *Pulse sequences for removal of the effects of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T1 and T2 values in proteins.* J. Magn. Reson., 1992. **97**: p. 359-375.
115. Barbato, G., et al., *Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible.* Biochemistry, 1992. **31**(23): p. 5269-78.
116. Piotto, M., V. Saudek, and V. Sklenar, *Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions.* J Biomol NMR, 1992. **2**(6): p. 661-5.
117. Clore, G.M., et al., *Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins.* J. Am. Chem. Soc., 1990. **112**: p. 4989-4991.
118. Garcia de la Torre, J., M.L. Huertas, and B. Carrasco, *HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations.* J Magn Reson, 2000. **147**(1): p. 138-46.
119. Lee, L.K., et al., *Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13C alpha nuclear spin relaxation.* J Biomol NMR, 1997. **9** (3): p. 287-98.
120. Baber, J.L., A. Szabo, and N. Tjandra, *Analysis of slow interdomain motion of macromolecules using NMR relaxation data.* J Am Chem Soc, 2001. **123**(17): p. 3953-9.
121. Ellgaard, L., et al., *Three-dimensional structure topology of the calreticulin P-domain based on NMR assignment.* FEBS Lett, 2001. **488**(1-2): p. 69-73.
122. Mathews Christopher, K. and K.E. Van Holde, *Biochemistry.*

Benjamin/Cummings series in the life sciences and chemistry. 1996, Menlo Park, Calif. ; Wokingham: Benjamin/Cummings 1996.

123. Nicholls, A., K.A. Sharp, and B. Honig, *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.* Proteins, 1991. **11**(4): p. 281-96.
124. Buevich, A.V. and J. Baum, *Dynamics of unfolded proteins: incorporation of distributions of correlation times in the model free analysis of NMR data.* J. Am. Chem. Soc., 1999. **121**: p. 8671-2.
125. Canet, D., et al., *A comprehensive analysis of multifield 15N relaxation parameters in proteins: determination of 15N chemical shift anisotropies.* J Am Chem Soc, 2001. **123**(19): p. 4567-76.
126. Tjandra, N., A. Szabo, and A. Bax, *Protein Backbone Dynamics and 15N Chemical Shift Anisotropy from Quantitative Measurement of Relaxation Interference Effects.* J Am Chem Soc, 1996. **118**: p. 6986-6991.
127. Kroenke, C.D., et al., *Longitudinal and transverse 1H-15N dipolar/15N chemical shift anisotropy relaxation interference: unambiguous determination of rotational diffusion tensor and chemical exchange effects in biological macromolecules.* J. Am. Chem. Soc., 1998. **120**: p. 7905-7915.
128. Schrag, J.D., et al., *The Structure of Calnexin, an ER Chaperone Involved in Quality Control of Protein Folding.* Molecular Cell, 2001. **8**: p. 633-644.
129. Ferreira, V., et al., *The classical activation pathway of the human complement system is specifically inhibited by calreticulin from Trypanosoma cruzi.* J Immunol, 2004. **172**(5): p. 3042-50.
130. Feinberg, H., et al., *Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2.* Embo J, 2003. **22**(10): p. 2348-59.
131. Matsushita, M., Y. Endo, and T. Fujita, *Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine protease.* J Immunol, 2000. **164**(5): p. 2281-4.
132. Wallis, R., et al., *Localization of the serine protease-binding sites in the collagen-like domain of mannose-binding protein: indirect effects of naturally occurring mutations on protease binding and activation.* J Biol Chem, 2004. **279** (14): p. 14065-73.
133. Kramer, R.Z., et al., *X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly).* J Mol Biol, 1998. **280**(4): p. 623-38.

134. Marley, J., M. Lu, and C. Bracken, *A method for efficient isotopic labeling of recombinant proteins*. J Biomol NMR, 2001. **20**(1): p. 71-5.
135. Gill, S.C. and P.H. von Hippel, *Calculation of protein extinction coefficients from amino acid sequence data*. Anal Biochem, 1989. **182**(2): p. 319-26.