Characterisation of an Interaction Involved in Viral Replication

Jolyon Claridge

Submitted in partial fulfilment of the requirements

of the degree of Doctor of Philosophy

Institute of Fundamental Sciences

Massey University

Abstract

Human rhinoviruses (HRVs) are a major cause of illness worldwide and, as members of *Picornaviridae*, are closely related to several other human and animal pathogens that exact a large medical and economic cost on society. Viral infections in general are particularly difficult to treat, as viruses co-opt many of the host's own biochemical pathways, making disabling the virus without harming the host very difficult. Carefully targeted strategies are required and detailed structural information is useful, both to identify new drug targets, and to fully understand interactions. One particular protein expressed by picornaviruses is 3C protease, which is responsible for post-translational processing of the viral capsid. This protease has a cysteine as its active site nucleophile, a functionality not found in eukaryotic proteases. The unusual active site makes 3C an attractive target for pharmaceuticals. Drugs that block the proteolytic action of 3C are currently in clinical trials. In addition to its proteolytic activity, 3C protease also has another function, that of an RNA binding protein. This activity has been shown to be required during replication of the viral RNA genome. In this study, the structure of 3C protease from HRV14 is investigated using NMR and other biophysical techniques. The structural information gained from these studies is used, along with data on 3C protease RNA-binding activity acquired using solution-state NMR and SAXS data, to elucidate a structure of the 3C–RNA complex. In addition, the dynamics of the free protein and of the protein in the presence of a specific inhibitor are investigated by solution-state NMR, and the potential role of dynamics in the function of the protein is explored. Finally, potential allosteric interaction between the RNA-binding and proteolytic functions of 3C is postulated, and further interactions of 3C and the 3C–RNA complex are discussed. It is hoped that a more complete understanding of 3C and its interactions will lead to more effective treatments for picornaviral infections in the future.

Acknowledgements

I would like to thank Dr. Steve Pascal for providing me with a grounding in NMR spectroscopy and for a great deal of helpful advice (mostly taken eventually...) about all aspects of the project. I also wish to thank Dr. Andrew Sutherland-Smith; Professors Geoff Jameson and Dave Harding and Dr. Gill Norris for many discussions and assistance over the years. I wish to thank Trevor Loo for a great deal of help with protein expression and purification when I was a new (and not so new) PhD. student. Dr. Pat Edwards for a phenomenal amount of help with all matters NMR. And also for the homebrew books! Dr. Stephen Headey for help with several aspects of the project and especially for teaching me practical NMR. Dr. David Libich for help with proof reading and for a great deal of input into the biophysical aspects of the project. The other members of the lab, especially Martin and Hari. The structural biology crowd at Massey especially Matt, Simon, Alice and Greg. My parents Gillian and Jolyon Claridge for their support and who can usually be relied upon to deliberately misunderstand my work in humorous and unpredicatable ways. My sister Emma for everything. And finally, my wonderful and long-suffering wife Sarah who was briefly under the impression that a 3 year PhD lasted 3 years...

Glossary of Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
$3C^{B}$	3C protease in a previous study
$3CI^{B}$	$3C^{B}$ with inhibitor
$3C^{pro}$	3C protease
3CI	$3C^{pro}$ with inhibitor
3CD	covalently linked $3C^{pro}$ and $3D^{pol}$
3CIR	3CI with SLD
$3D^{pol}$	3D polymerase
BB1	β-barrel number 1
BB2	β-barrel number 2
BMRB	Biological Magnetic Resonance Data Bank
BTM	bis-Tris propane / MES
CD	circular dichroism
CN1-3C	1st C- and N-labelled sample of 3C produced
CPMG	Carr-Purcell-Meiboom-Gill
CSA	chemical shift anisotropy
CSI	chemical shift index
CSP	chemical shift perturbation
ddH_2O	distilled deionised water
DLS	dynamic Light Scattering
DTT	dithiothreitol

HRV	human rhinovirus
HRV-14	human rhinovirus isotype 14
HRV-2	human rhinovirus isotype 2
HSQC	heteronuclear single quantum coherence
IMAC	immobilised metal ion affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani broth
MES	(2-(*N*-morpholino)ethanesulphonic acid
NOE	nuclear overhauser effect
N2-3C	2nd N-labelled sample of 3C produced
R_1	longitudinal relaxation rate
$R_{1\rho}$	relaxation rate under a transverse spinlock
R_2	transverse relaxation rate
RBD	RNA-binding domain
RCI	random coil index
RDC	residual dipolar coupling
R_{ex}	contribution to relaxation from slow exchange processes
R_{ex}^{CC}	R_{ex} as determined from cross-correlation analysis
R_{ex}^{LS}	R_{ex} as determined from Lipari-Szabo model-free analysis
S^2	order parameter
S^2_{ave}	average order parameter
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SLD	stemloop-d
T_1	characteristic longitudinal relaxation time
T_2	characteristic transverse relaxation time
TROSY	transverse relaxation optimised spectroscopy

Contents

List of Figures

List of Tables

1

Introduction

1.1 Introduction

This chapter provides a brief introduction to the genetics and molecular biology of picornaviruses. The literature relating to the structure and function of $3C^{pro}$ and interaction partner stemloop-d is reviewed. I also delineate the aims of the project and outline the structure of the thesis.

1.2 Picornaviruses

Picornavirus literally means "small RNA virus", referring to the viral genome. The RNA genome is single and positive-stranded, meaning the genome is of the same sense as mRNA. Unlike retroviruses the life-cycle of picornaviruses contains no DNA intermediate. *Picornaviridae* consists of several genera. Four original genera: *Rhinovirus, Enterovirus, Cardiovirus* and *Apthovirus,* were based mainly on pathogenesis in humans and animals and on biophysical properties. With the availability of genetic sequence information five other genera have been described: *Hepatovirus, Parechovirus, Erbovirus, Kobuvirus* and *Teshovirus*. Of these, some contain previously unknown viruses. Other genera, such as *Hepatovirus*, are a reclassification of previously known species on the basis of obviously divergent genetics. Table 1.1 shows the current classification for several members of *Picornaviridae*.

Genus	Species	Serotypes
Enterovirus	*Poliovirus*	PV serotypes 1–3
	Human enterovirus A (HEV-A)	CVA serotypes 2–8, 10, 12, 14, 16; EV-71
	Human enterovirus B (HEV-B)	CVB serotypes 1, 2, 3, 4, 5, 6; CVA9
	.	Echovirus (E) serotypes 1–7, 9, 11–21,
	.	24–27, 29–33; EV-69; SVDV
	Human enterovirus C (HEV-C)	CVA serotypes 1, 11, 13, 15, 17–22, 24
	Human enterovirus D (HEV-D)	EV-68, EV-70
	Bovine enterovirus (BEV)	BEV-1, BEV-2
	Porcine enterovirus A (PEV-A)	PEV-8
	Porcine enterovirus B (PEV-B)	PEV-9, PEV-10
Rhinovirus	*Human rhinovirus A (HRV-A)*	HRV serotypes 1, 2, 7, 9, 11, 15, 16, 21, 29,
	.	36, 39, 49, 50, 58, 62, 65, 85, 89
	Human rhinovirus B (HRV-B)	HRV serotypes 3, 14, 72
Cardiovirus	*Encephalomyocarditis virus (EMCV)*	EMCV
	Theilovirus (ThV)	TMEV, VHEV
Apthovirus	*Foot-and-mouth disease virus (FMDV)*	FMDV serotypes O, A, C, Asia1, SAT-1, SAT-2, SAT-3
	Equine rhinitis A virus (ERAV)	ERAV
Hepatovirus	*Hepatitis A virus (HAV)*	HAV
	Avian encephalomyelitis-like virus (AEV)	AEV
Parechovirus	*Human parechovirus (HPeV)*	HPeV-1, HPeV-2
Erbovirus	*Equine rhinitis B virus (ERBV)*	ERBV
Kobuvirus	*Aichi virus (AiV)*	AiV
Teschovirus	*Porcine teschovirus (PTV)*	PTV serotypes 1–11

Table 1.1: Picornavirus classification. Table adapted from Stanway *et al.* (2002)

Picornaviruses are pathogens of several vertebrate, primarily mammalian, species. Poliomyelitis, paralysis caused by poliovirus, has been known in human society since at least 1300 B.C. The group also contains several other human pathogens such as hepatitis A virus and cardioviruses, as well as economically damaging viruses such as foot and mouth disease virus (FMDV).

1.3 Clinical and Societal Significance

Picornaviral infections are of major medical and economic interest. Although widespread vaccination has reduced the prevalence of polio in most of the world, there have been recent outbreaks in Nigeria and India (Anonymous, 2009). Human rhinoviruses are estimated to be the causative agent in over 50% of upper respiratory tract infections (Greenberg, 2003). They impose a significant economic cost on society and also can have severe medical complications in immune-compromised or elderly patients. Coxsackie virus causes a variety of, mostly minor, conditions but can cause myocarditis, which, in severe cases, can be fatal (Kandolf *et al.*, 1993). Picornaviruses are also responsible for a variety of animal diseases such as foot and mouth, caused by foot and mouth disease virus (FMDV). FMDV is endemic in much of the world and, as such, it is a large burden on farmers in affected countries. FMDV and other picornaviruses infecting livestock are of particular concern for largely agricultural economies such as that of New Zealand. Although vaccination against FMDV is possible, it is very costly and often incompatible with national policy.

1.4 Picornaviral Phylogeny

The earliest attempts to classify picornaviruses were made using clinical measures. Many of the current *genera* were first described as groups that caused similar illnesses. More recently, sequencing of viral genomes has lead to some reorganisation of the picornavirus clade. Most of the original *genera* have been retained but some species (for example hepatitis A virus (HAV)) have been reclassified. In the case of HAV a new genus has been described to accommodate its obviously divergent genetic sequence. Picornaviruses can be classified by the wide variety of surface receptors to which they bind in the process of gaining entry into host cells. This is related to the symptoms of viral infection, as the preferred receptor type can influence tissue specificity of the virus. Poliovirus recognises receptors from the immunoglobulin super-family, many human rhinoviruses recognise intercellular adhesion molecule 1 (ICAM-1) and foot and mouth disease

virus (FMDV) recognises integrins (Rieder and Wimmer, 2002; Evans and Almond, 1998).

Picornaviridae members have been classified by aligning two parts of the genome: the P1 structural proteins and the non-structural 2C/3CD protein sequences. Phylogenetic analysis of these sequences consistently place *rhinovirus* and *enterovirus* into a grouping separate from the remainder of *Picornaviridae* Figure 1.1. Perhaps related to this fact, the $3C^{pro}$ sequences of *Picornaviridae* can be classified by whether their catalytic triad contains a glutamate residue (lineage 'A') or an aspartate (lineage 'B')(Ryan and Flint, 1997). Figure 1.1 (b) shows the *Rhinovirus* and *Enterovirus* phylogeny – lineage 'A'. Cis-acting replication elements (*CREs*), present in many picornaviral species, vary in their position in the non-coding region of the picornaviral genome. The position relative to other strains has been suggested as a way of differentiating between closely related picornavirus species (Cordey *et al.*, 2008).

1.4.1 Rhinoviruses

Nucleotide sequencing of several rhinovirus capsid sequences confirmed that the group contains two species, HRV-A and HRV-B. However, under some analyses, the HRV-B species are more closely related to *Enteroviruses* than to HRV-A (Tapparel *et al.*, 2007). In any case, the genomes of *Rhinovirus* and *Enterovirus* are so similar that if it were not for their different clinical properties they would be classified as one genus. HRV strain 14 belongs to HRV-B. Although recombination between strains is a source of genetic change in human enteroviruses, it seems that this is much less prevalent in HRV strains (Kistler *et al.*, 2007). Figure 1.1 (a) shows a cladogram of $3D^{pol}$ alignments from representative picornavirus species. *Entero-* and *Rhinovirus* form a clearly separate group from other picornaviruses. Figure 1.1 (b) shows a cladogram of only the *Entero-* and *Rhinovirus* (lineage 'A') species. The close relationships in the two figures are slightly different as the cladogram in (a) is calculated using $3D^{pol}$ and (b) is calculated using $3C^{pro}$ sequences.

1.5 Picornaviral Capsid

Picornaviral capsids have a well-conserved overall structure with highly variant inter-domain loops. These loops are under constant selective pressure from the host's immune system leading to this variability (Rossmann and Palmenberg, 1988). The virus has a icosohedral capsid containing 60 repeats of four capsid proteins. The structural proteins VP1-VP3 make up the shell of the virus. A fourth protein VP4 is associated with the internal face of the capsid in *enteroviruses* and

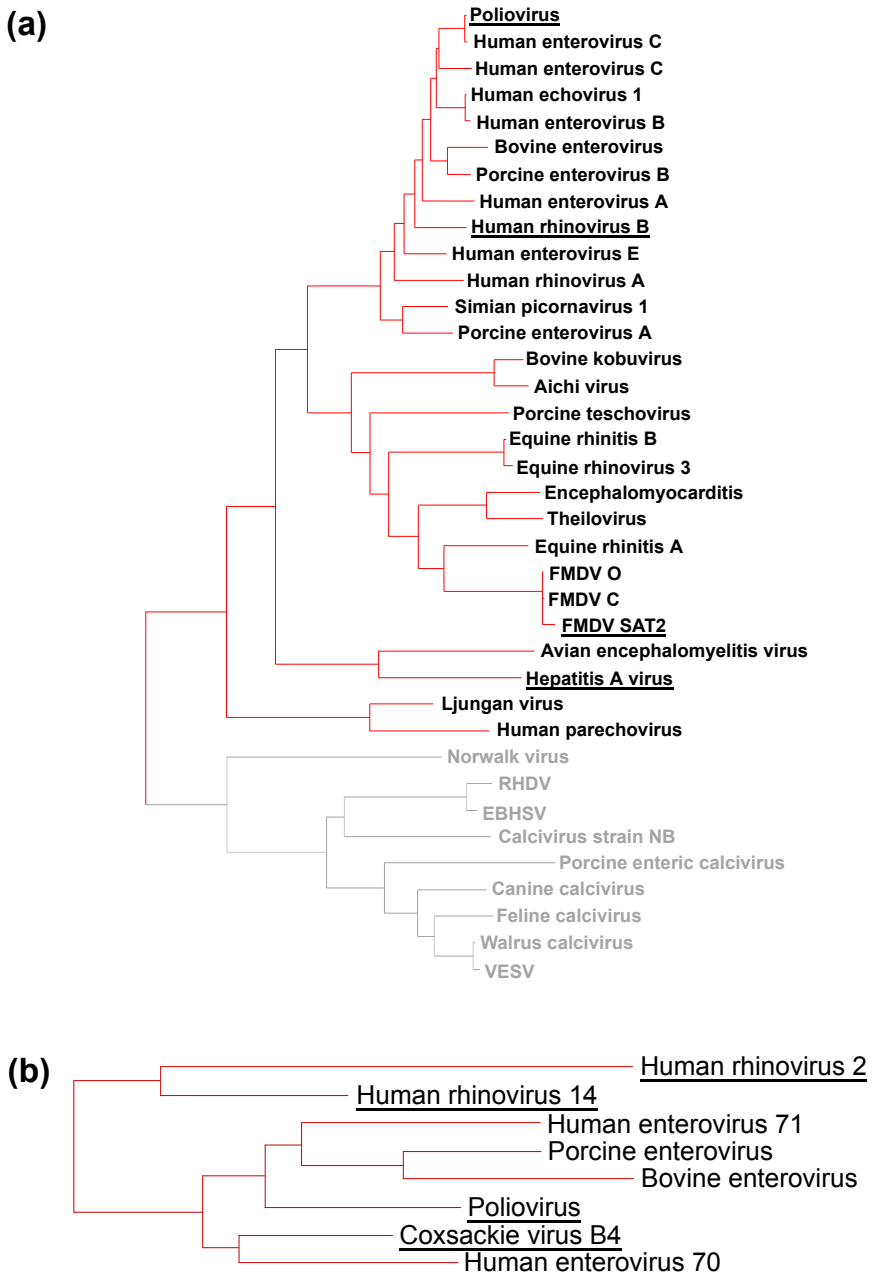


Figure 1.1: Cladograms of Picornaviral Phylogeny. (a) Phylogeny of representative species of picornavirus. The cladogram is rooted with *Calciviridae*. Calciviruses are shown greyed out. Figure adapted from (Hughes, 2004).(b) subset of *Picornaviridae*, the *Rhinovirus* and *Enterovirus*, or lineage 'A' species (see text) (Ryan and Flint, 1997). In both figures, the names of viruses whose $3C^{pro}$ structures have been solved are underlined.

rhinoviruses and is thought to play a role in stabilisation of the RNA genome. In *aphthoviruses* the VP2 and VP4 proteins remain covalently linked leading to only three proteins making up the capsid. The overall fold of each these proteins is similar and highly conserved within the picornaviral family, comprising an eight-stranded β-barrel (Arnold and Rossmann, 1990). The loops joining the strands in the β-barrel are, however, extremely variant and this gives the different viruses their disparate receptor binding properties and antigen specificities. One major feature of the viral capsid is the "canyon", a large invagination in the surface of the capsid surrounding the five-fold symmetry axis. This contains the sites for receptor recognition. An important property of this feature is that it is too small to admit antibodies, meaning that receptor specificity can develop unhindered by immuno-selection. Figure 1.2 (a) shows the viral capsid.

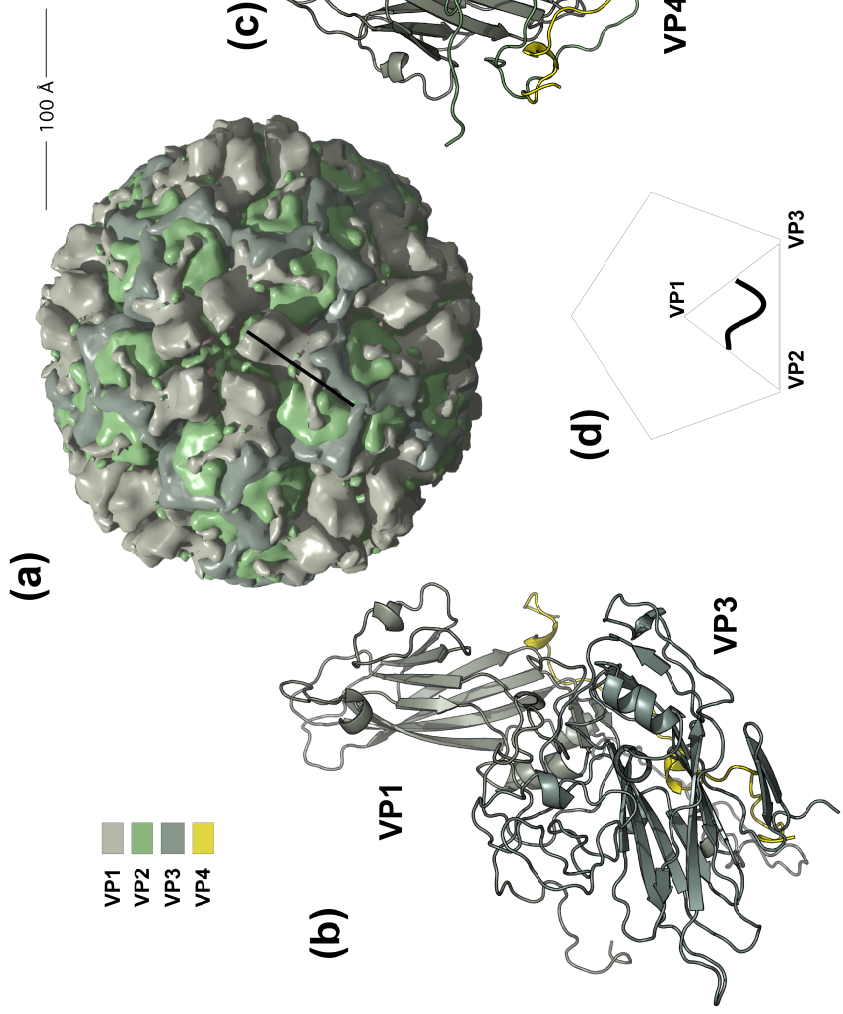


Figure 1.2: View of structure of viral capsid and associated proteins from HRV14. The crystal structure of proteins VP1, VP2, VP3 and VP4 were solved by X-ray crystallography to 3.0 Å and then built into the envelope of the capsid determined by cryo EM. The capsid is shown in (a) The five-fold symmetry is apparent. The black line shown the approximate position of the two side views (b) showing VP1 and VP3 and (c) showing proteins VP1 and VP2. VP4 is shown in yellow in both figures. (d) shows the approximate position of the "canyon" in the asymmetric unit, shown as a triangle (Arnold and Rossmann, 1990).

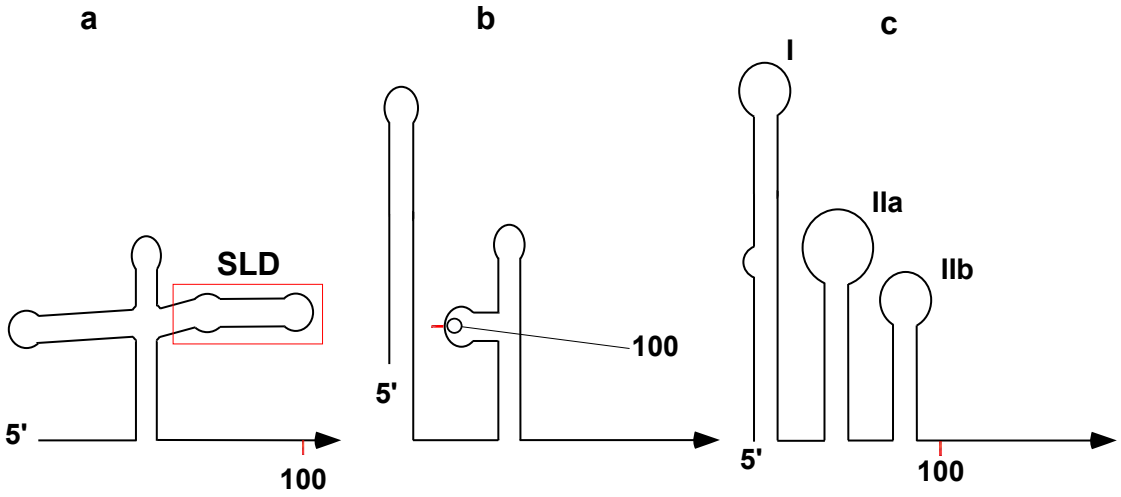


Figure 1.3: Picornaviral 5′ untranslated regions. (a) Poliovirus (Type I) (b) EMCV (Type II) and (c) HAV (Type III). Stemloop-d in poliovirus is annotated 'SLD' and marked with a red box. This motif is conserved within type 1 picornaviruses and the HRV-14 SLD is the RNA target sequence of $3C^{pro}$ in this study.

1.6 Viral Genome

The viral genome contains a 5′ un-translated region (UTR) containing several domains. The first domain provides a binding site for cellular and viral proteins and is involved in initiation of RNA replication. Picornaviral 5′ UTRs can be placed into three groups: one, including enteroviruses and rhinoviruses (see section 1.4) forms a cloverleaf (Figure 1.3(a)). Viruses in the second group, which contains cardioviruses and aphthoviruses (Figure 1.3(b)), and third group, which contains hepatoviruses (Figure 1.3(c)) have less well defined secondary structure. After the initial loops in the 5′ UTR lies the internal ribosome entry site (IRES). This forms a variety of complicated secondary structures, relatively unconserved in *Picornaviridae*. Immediately after the 5′ UTR is the ORF. This consists of approximately 6500 bases (Stanway *et al.*, 1984). Finally a 3′ UTR which contains a poly A tail and a uridylation signal, important in negative strand RNA replication, completes the genome (Stanway *et al.*, 1983, 1984).

1.7 Picornaviral Life Cycle

A general overview of the picornaviral life-cycle is shown in Figure 1.4. These events are largely conserved across *Picornaviridae*; peculiarities of different virus types are mentioned in the text.

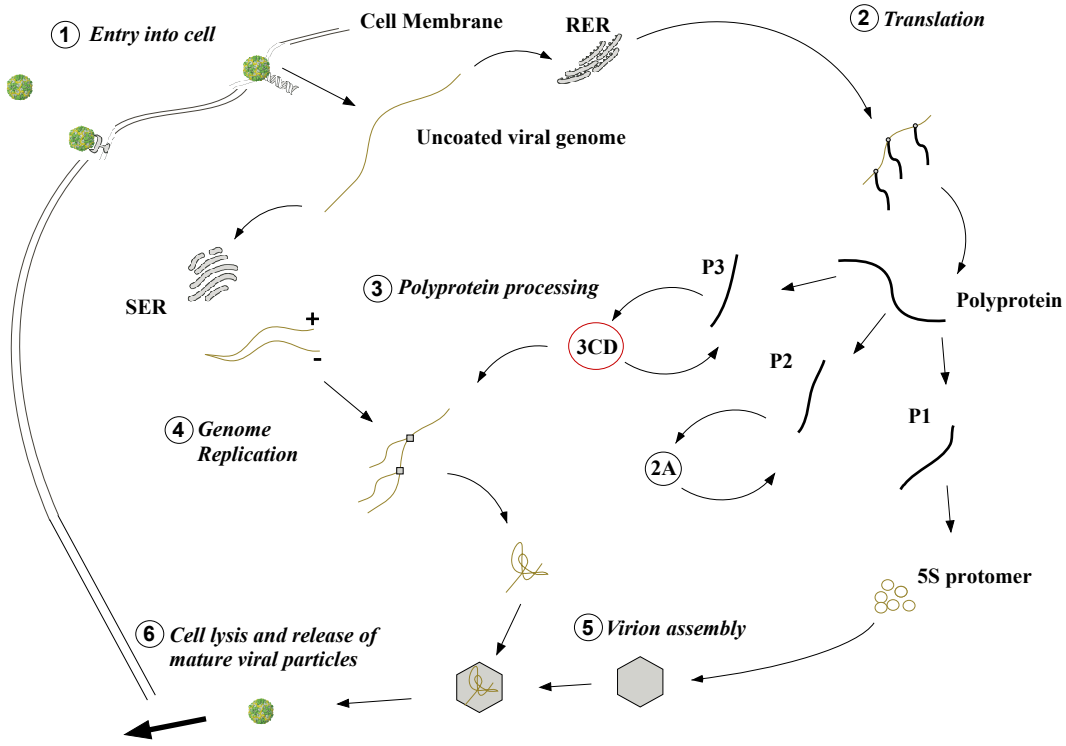


Figure 1.4: Overview of picornaviral life-cycle.

1.7.1 Translation and protein processing

Upon infection of the host, the virus attaches to the cell surface via the interaction with one of a variety of cell-surface receptors (Evans and Almond, 1998) (Figure 1.4). The virus is uncoated and positive-stranded viral RNA enters the cell (Figure 1.4 ①). Once the viral RNA has crossed the cell membrane, it bypasses the requirement for a cap-binding protein by utilising an internal ribosome entry site (IRES) in order to initiate translation (Figure 1.4 ②). This is necessary as the 5′ UTR has secondary structure which would inhibit ribosome scanning as well as several AUG codons which could initiate translation prematurely (Bedard and Semler, 2004). There are several types of IRES. Enteroviruses and rhinoviruses contain a type I IRES and most other picornaviruses have a type II IRES (Jang *et al.*, 1990). Hepatitis A virus has a type III IRES. Although the sequence of the different IRESs is highly variant, the secondary structure seems to be highly conserved. The IRES acts *in cis*, recruiting cellular proteins and rRNAs to form a translation complex. The ORF is translated into a single polyprotein of approximately 243kDa. This then undergoes a series of protease cleavage steps (Figure 1.5). The first is a cleavage by the viral protease $2A^{pro}$ which cleaves its own amino terminus and splits the polyprotein into two

parts, one containing the structural capsid proteins and one containing the non-structural viral proteins (Cordingley *et al.*, 1990). $2A^{pro}$ further cleaves the structural precursor VP0 into VP2 and VP4 (Figure 1.2) as well as cellular proteins such as the cap binding protein eIF-4F (Seipelt *et al.*, 1999). $3C^{pro}$ is also essential for the correct processing of the viral polyprotein. The remaining cleavage steps are all carried out by $3C^{pro}$ or its active precursor 3CD. The non-capsid protein polypeptide is cleaved to form P2 and P3. P2 is further cleaved into $2A^{pro}$ and 2BC. P2 is cleaved into 3A, 3B (also known as VPg, viral protein genome-linked) and 3CD (Figure 1.4 ③). It was originally thought that cleavage of cellular proteins by viral proteases occurred in the cytoplasm but recent research suggests that $3C^{pro}$ and 3CD translocate to the nuclear membrane and cause the arrest in transcription of cellular proteins by hydrolysing transcription factors TFIIIC, OCT-1, CREB and the La autoantigen (Amineva *et al.*, 2004; Kuyumcu-Martinez *et al.*, 2004). This effectively stops cellular RNA transcription. Mutational evidence suggests that the nuclear localisation signal is located on the carboxyl terminus of 3CD and that cleavage of 3CD into $3C^{pro}$ and 3D occurs within the nucleus (Amineva *et al.*, 2004). Binding of the 3CD protein to the 5′ end of the viral genome inhibits the translation process. This is possibly due to the 3D polymerase bringing the 3′ end of the genome into close proximity to the 5′ end and blocking the formation of the translation complex. Alternatively the 3CD protein may sterically inhibit the formation of the ribosome complex (Herold and Andino, 2001).

P3 proteins

Cleavage at the amino terminus of 3A leads to the production of the 3ABCD polypeptide. In poliovirus and HRV14 this peptide is quickly cleaved (probably *in trans*) by $3C^{pro}$, into 3AB and 3CD (Figure 1.5). Both $3C^{pro}$ (in HAV) and $3D^{pol}$ (in poliovirus and FMDV) have been shown to dimerise and even oligomerise in solution. In both cases, a functional purpose for this oligomersiation has been proposed (Pathak *et al.*, 2002; Peters *et al.*, 2004). 3AB has multiple functions both assigned to the separate domains 3A and 3B and also synergistically involving both domains. As a non-specific RNA binding protein it interacts with stemloop-b from the 5′ cloverleaf and increases the affinity of 3CD for the 5′ cloverleaf (Harris *et al.*, 1994). 3AB stimulates the cleavage of 3CD to $3C^{pro}$ and $3D^{pol}$. This could be a regulatory mechanism by which 3CD is turned from a protease into a protease and a polymerase (Molla *et al.*, 1994). The C-terminal of 3A is hydrophobic and interacts with membranes during replication. 3A has been shown to be glycosylated *in vitro* (this modification is not required for replication in PV but does reduce poliovirus RNA

synthesis) 3AB also interacts with the cellular membrane to form part of the replication complex in infected cells. (Datta and Dasgupta, 1994). Inhibition of picornaviral replication by the antiviral drug enviroxime is mediated by 3A or 3AB. It appears that enviroxime blocks the uridylation event required for positive-strand RNA synthesis (Heinz and Vance, 1995).

3B, also known as VPg (viral protein genome linked) is covalently bound to both the 5′ and 3′ end of the viral genome. This linkage is through a highly conserved tyrosine residue at the third position in the 3B peptide. The linkage is shown in Figure 1.6 (inset).

In entero- and rhinoviruses, 3CD is responsible for the cleavage of all sites in the viral polyprotein except VP4/VP2 and P1/P2 (cleaved by $2A^{pro}$. The protease cleaves at Q/G (Cordingley *et al.*, 1990). 3D is thought to be responsible for modulation of $3C^{pro}$ binding to stemloop D of the 5′ cloverleaf, especially in poliovirus, in which $3C^{pro}$ alone has little affinity for SLD.

3D is also responsible for uridylation of 3B and elongation of the replication. 3D requires an RNA template to elongate strands of RNA when using either a polyA sequence or, more likely *in vivo*, the uridylated form of VPg.

1.7.2 RNA replication

The interaction between the 5′ cloverleaf, 3CD and the cellular protein PCBP2 has been demonstrated to be required for both positive and negative strand synthesis in poliovirus, the archetypal enterovirus (Parsley *et al.*, 1997). Figure 1.4 ④. This interaction is probably also important in rhinovirus.

In HAV the polypeptide 3ABC is more prevalent than either 3AB or 3CD and the interaction between the 5′ and 3′ UTRs and 3ABC is stronger than that of 3AB or $3C^{pro}$. $3C^{pro}$ from HRV14 and HAV appears to possess more specific RNA binding activity than that of $3C^{pro}$ in poliovirus.

Several cellular proteins are known to interact with virally encoded proteins and with the viral RNA during genome replication. Nucleolin, polyrC binding protein (PCBP2) and polyA binding protein (PABP) have all been shown to be involved in viral replication. PABP interacts with both the 5′ ribonucleoprotein complex and the 3′ UTR, suggesting that the genome may become circularised during the replication process (Meredith *et al.*, 1999; Herold and Andino, 2001).

cis-acting RNA elements

The RNA genome of picornaviruses was originally shown to contain two structured regions known to be important for replication. These were termed ori(origin of replication)-R (3′) and -L (5′) and were thought to be the origin of replication for positive strand synthesis and negative strand synthesis respectively. This picture has become somewhat clouded with the finding of a third *cis*-acting element, termed oriI (origin of replication internal) which is found within the coding portion of the genome, in the coding region for 2C in poliovirus (Goodfellow *et al.*, 2000) and in the coding region for VP1 (one of the capsid proteins) in HRV14 (McKnight and Lemon, 1998).

The 3′ UTR contains a single stemloop in rhinovirus and two loops in poliovirus. Other viruses contain up to three stem loops in oriR.

VPg uridylation

A necessary step in viral replication is the uridylation of VPg by 3D polymerase. This creates a primer for the polymerase to initiate RNA synthesis. VPg is also known as 3B and is thought to be present in the host cell as 3AB. The uridylation event requires two uridine molecules to be attached to the 5′ end of VPg via a conserved tyrosine residue. The reaction works via a 'slide-back' mechanism (Paul *et al.*, 2003) whereby a conserved A moiety in the cis-acting replication element (CRE) of the virus is used to template the first U and then the polymerase moves the nascent poly-U strand back to template the second U and so on.

Negative strand RNA synthesis

The 3CD protein binds to the 5′ region via an interaction between $3C^{pro}$ and the 5′ clover-leaf stem-loop D. The 3CD protein then is cleaved (this almost certainly occurs *in trans* with the cleavage being mediated by free $3C^{pro}$) and the 3D begins transcription (Xiang *et al.*, 1995). Transcription of the negative RNA strand is begun at the 3′ end of the virus. Polymerase 3D is unable to bind to the RNA without the $3C^{pro}$ protein to which it is attached, however, the polymerase is inactive until the $3C^{pro}$ is removed. The polymerase has no proof-reading function and errors in transcription are as frequent as 1 per genome. This is partly responsible for the high rate of mutation in picornaviruses. Another factor in this is the fact that related viruses that cross-infect a cell can heterologously recombine and form hybrid genomes, the final product being resolved via a copy choice mechanism (Worobey and Holmes, 1999). The production of the negative strand

creates a duplex replicative intermediate (RI). Many copies of the positive viral genome can be transcribed from one negative strand, which accounts for the low copy number of the negative strand in comparison to the positive strand of 1:30 to 1:50 in infected cells (Regoes *et al.*, 2005).

Positive strand RNA synthesis

Transcription of the positive strand is initiated by the uridylation of the terminal tyrosine of VPg which acts as a primer for 3D polymerase. Many cellular proteins interact with the polymerase in order to form the transcription complex. The polymerase appears to also have some helicase activity and is thought to unwind the RNA duplex prior to transcriptional initiation (Cho *et al.*, 1993). Although 2C has a putative helicase domain, and was thought to possess this function, no such activity has been observed for this protein. Transcription proceeds from the 3′ end of the negative strand. Although the 3′ NCR seems to be important in viral replication, it is not necessary as mutated viruses lacking the 3′ region are still viable (Yang *et al.*, 2004). The transcription level of these mutants is, however, greatly reduced. The nascent positive strand can be used for the translation of further polyprotein or can be packaged into a virion.

1.7.3 Virion packaging and cellular apoptosis

This process takes place near the RER (rough endoplasmic reticulum) and many virus particles are assembled within blebs invaginating from the RER membrane (Figure 1.4 ⑤). The final stage in the viral life cycle is the release of viral particles during cell lysis. This lytic phase is thought to be triggered by viral protease-initiated cellular apoptosis shown in Figure 1.4 ⑥.

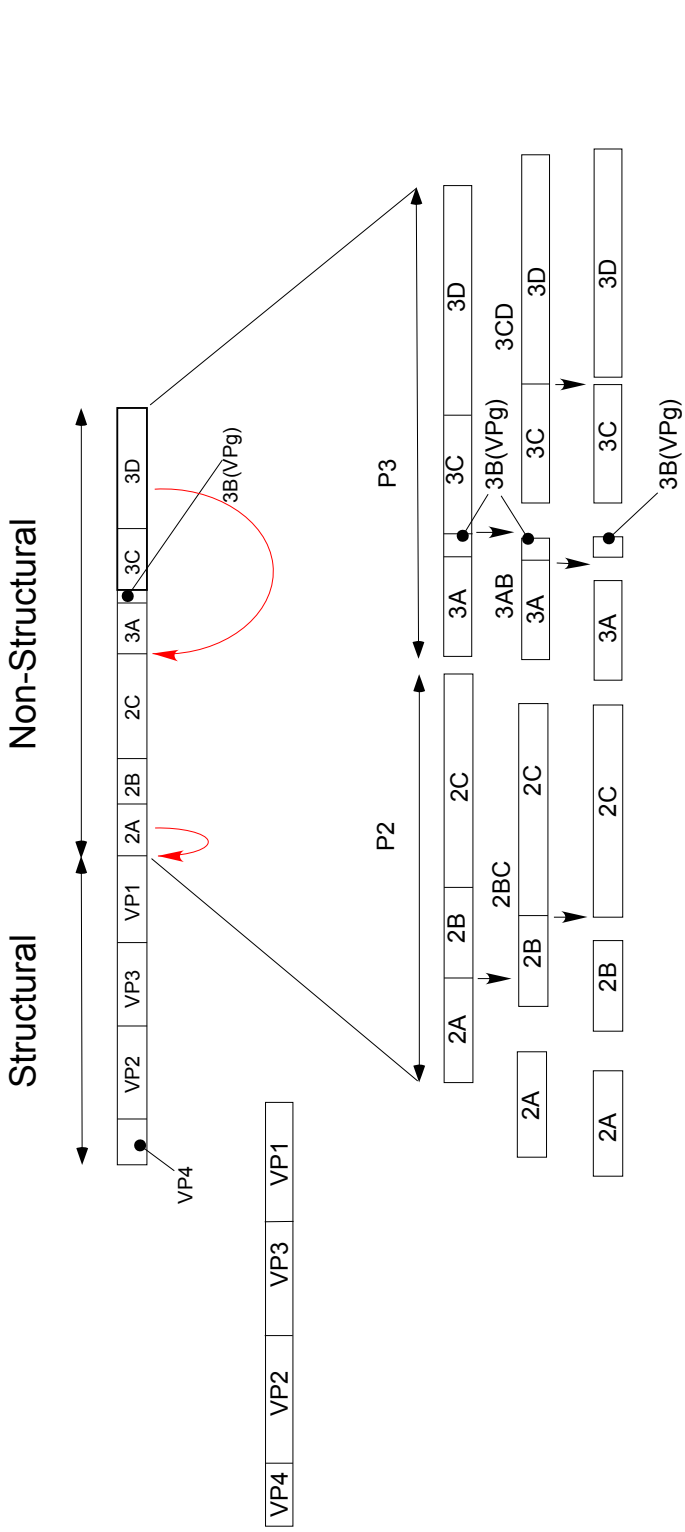


Figure 1.5: Processing of the viral polyprotein by viral proteases. The red arrows show sites of primary (during or immediately after translation) cleavage. Further cleavage events are indicated by black arrows.

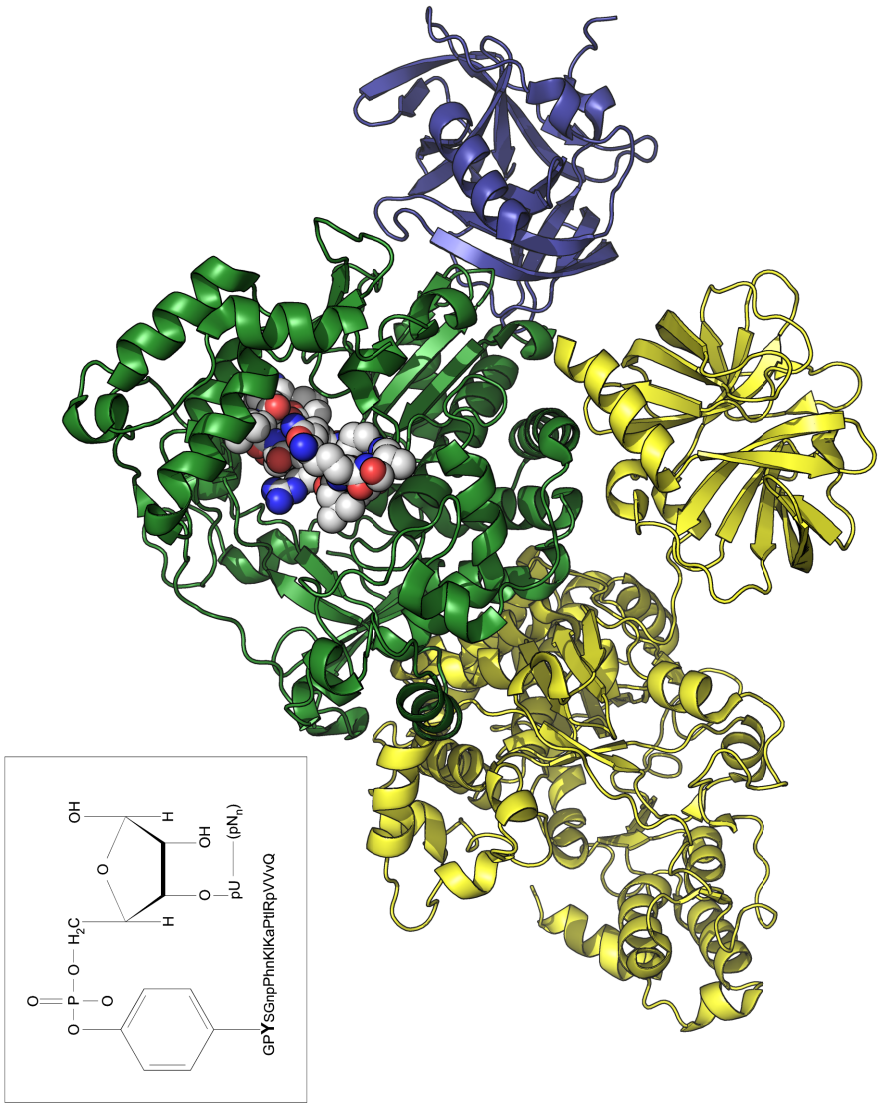


Figure 1.6: Possible formation of 3CD–3D complex during VPg uridylation. 3D is shown in green with VPg as spheres. A 3CD molecule is shown in yellow. The inset shows a tyrosine side chain attached to a uridine moiety. The conserved tyrosine is in bold in the VPg sequence for HRV 14 below the inset.

1.8 3C Protease (3C^{pro})

3C proteases are ubiquitous in picornaviruses. The majority of polyprotein secondary cleavage events in picornaviruses are catalysed by 3C^{pro}. 3C^{pro} is a single peptide protein, 182 amino acids in length. It consists of two anti-parallel β-barrels each containing six β-strands. In the cleft between these two topologically similar domains is the active site. The two β-barrels are connected by a flexible linker (Lawrence and Thach, 1975; Pelham, 1978; Matthews *et al.*, 1994) Although there is variation in sequence between 3C^{pro} sequences from different picornaviruses, the overall fold is very similar. There are also 3C^{pro}-like (3CL) proteases found in many more distantly related viruses such as SARS and Rice tungro spherical virus (RSTV) (Sekiguchi *et al.*, 2005). Although 3C^{pro} is a cysteine protease, its overall fold resembles that of the serine protease chymotrypsin rather than that of other cysteine proteases such as papain.

Figure 1.8 shows a cartoon view of HRV14 3C^{pro} looking down on the two β-barrel domains. The α-helices visible in the top figure are involved in RNA binding. The other protease involved in polyprotein processing, 2A^{pro}, is a cysteine protease and has a similar fold toα-lytic protease, another chymotrypsin-like protease. Interestingly, L protease, coded for by FMDV, and responsible for the cleavage of host cell initiation factors in infection by that virus, has a papain-like fold.

The active site of 3C^{pro} contains a catalytic triad of cysteine, histidine and glutamate. Figure 1.9 shows a view of the HRV14 catalytic triad. This triad is conserved throughout the picornaviral family. The catalytic triad contrasts with that in chymotrypsin, both in the use of an active site cysteine instead of a serine and also in that the third residue in the active site is a glutamate instead of the aspartate. The Cys-146 and His-40 residues face each other across the cleft between the two β-barrel domains. The cysteine provides a nucleophilic centre crucial in catalysis. The protein is much more specific than chymotrypsin (Orr *et al.*, 1989; Long *et al.*, 1989). This seems to be related to binding pockets on either side of the active site which are much smaller than those in chymotrypsin.

In poliovirus, regulation of protease function is mediated by 2C which acts as a protease inhibitor. The protein contains several 'serpin' (serine protease inhibitor) domains which possibly function in a similar fashion in cysteine proteases of this type (Banerjee *et al.*, 2004). The sequence specificity of 3C^{pro} has been used advantageously in the production of recombinant proteins as 3C^{pro} can be used to remove fusion proteins with very little non-specific cleavage. Titrations of

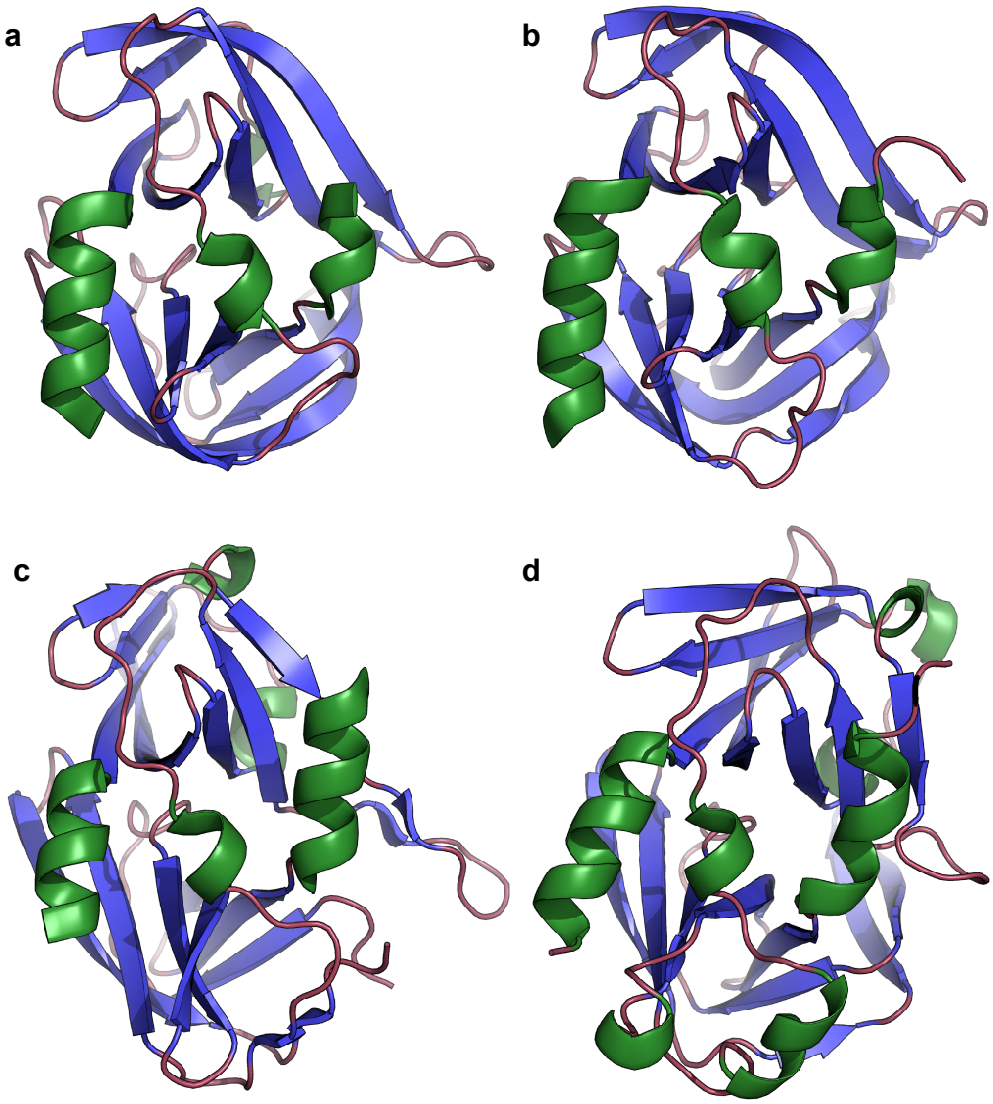


Figure 1.7: 3C proteases from various members of Picornaviridae. (a) poliovirus (Mosimann *et al.*, 1997) (b) human rhinovirus 14(Bjorndahl *et al.*, 2007). (c) FMDV (Birtley *et al.*, 2005) (d) HAV (Bergmann *et al.*, 1997). (a) and (b) are very closely related; the structures show a high degree of similarity. (c) and (d) represent the lineage 'B' picornaviruses and exhibit more divergent structure to the entero- and rhinovirus forms of the protease.

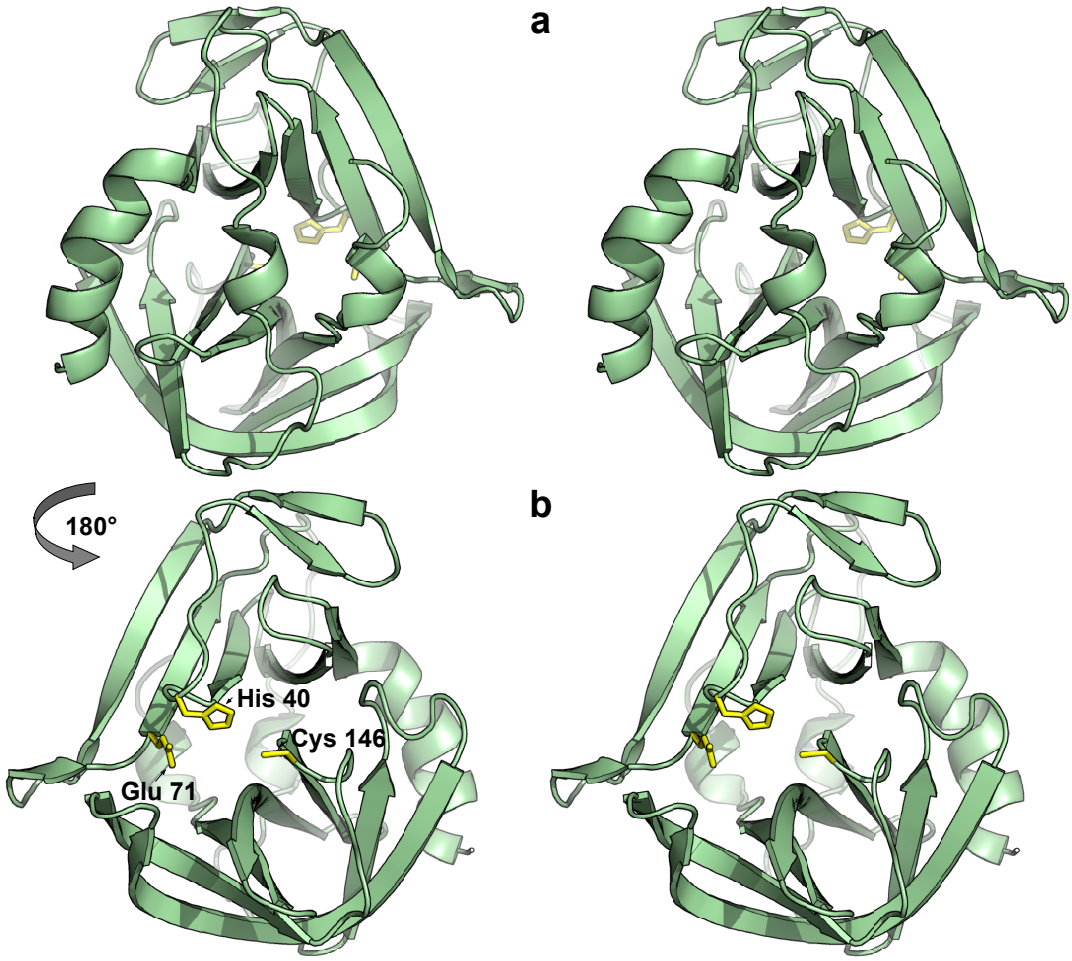


Figure 1.8: Stereo view of 3C^{pro} from HRV14. (a) shows the RNA binding face. The three helical regions of the protein are clearly visible. (b) shows the same representation of 3C^{pro} rotated 180°. The catalytic triad residues (H40, E71 and C146) are labelled.

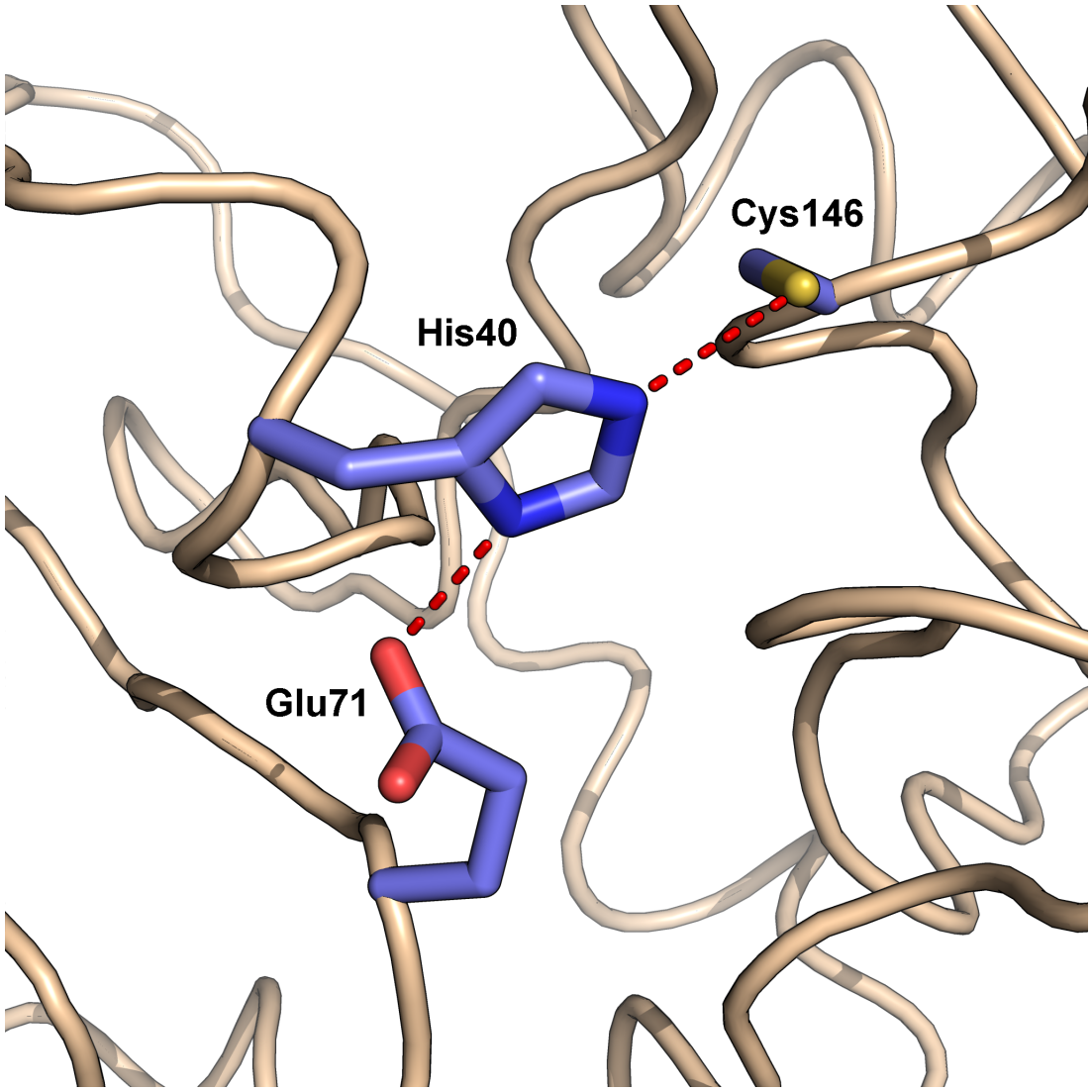


Figure 1.9: Stick view of the catalytic triad of HRV 14 $3C^{pro}$. The bottom residue is glutamate 71, which accepts a hydrogen bond from the middle residue, histidine 40. The topmost residue in the figure, cysteine 146, is responsible for nucleophilic attack on the substrate carbonyl carbon.

3C^{pro} against an irreversible inhibitor coupled with electrophoretic data suggest that the protease can exist in a non-functional form with a lower pI than the active form (8.3 cf. 9.0) this is probably due to deamidation of Asn-164. Deamidated forms of the protein can be detected in infected cells. The deamidated residue is involved in substrate binding and this may indicate a regulatory pathway for protease activity (Cox *et al.*, 1999). Studies of the cleavage sites on the viral polyprotein indicated the sequence preferences for the viral proteases. The sequence specificity of the 3C^{pro} has also been confirmed by competition assays on short peptide sequences.

1.8.1 RNA binding

As has previously been noted, the protease also has a function in RNA binding . 3C^{pro} has been shown to interact specifically with at least two distinct RNA sequences within the picornaviral genome, the 5′ cloverleaf SLD (Leong *et al.*, 1993) and the internal *CRE* reg (Paul *et al.*, 2000). The RNA binding domain is situated on the opposite face to the active site. Mutations that affect RNA binding usually do not adversely affect the catalytic properties of the enzyme. For instance substitutions at the three residues in the catalytic triad (e.g. H40D, E71A or C146S) resulted in non-functioning protease which still bound to RNA. The RNA binding site consists of several basic residues which are sequentially separate. Mutation studies indicate that the highly conserved Asp 85 is vital for RNA binding. 3C^{pro} expressed with this residue replaced by Asn are unable to bind specifically to RNA and produce a non-viable phenotype when introduced to the virus in vivo (Walker *et al.*, 1995). The large positively-charged RNA binding patch is clearly visible in Figure 1.10 panels (c) and (d).

Mutation studies indicate that the residues Gly-154(G154Q) and Lys-155(K155G) also seem to play a vital role in RNA binding along with Arg-176 and Arg-84. Other mutations which affect RNA binding are in His-31(H31A) and Asp-32(D32A). Arg-12, Lys-13, Lys-82, Lys-174 and Lys-175, as well as the previously mentioned Lys-155, are basic amino acids that are thought to interact with the RNA. It should be noted that mutations in His-31 also affect proteolytic activity. This perhaps indicates that these mutations are responsible for protein misfolding rather than the residue being directly involved in RNA interaction. Similarly mutations in Asp-85 affect both proteolytic and RNA-binding function. As this residue is negatively charged is seems unlikely that the residue is directly involved in RNA binding and is more likely to serve a role in the correct folding of the protein. It has been suggested that Phe 6 and Phe 89 stack between base pairs in the RNA duplex (Matthews *et al.*, 1994).

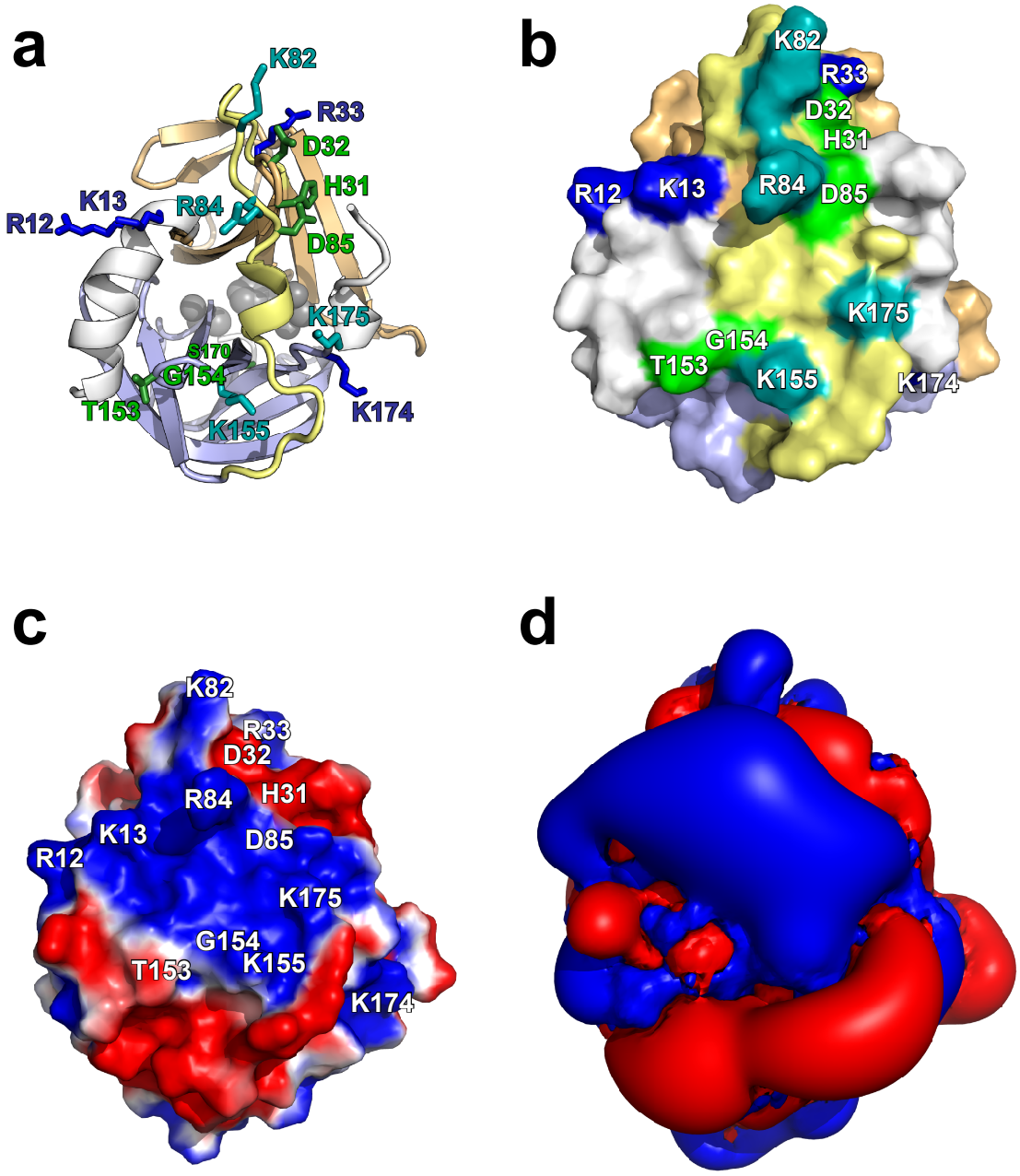


Figure 1.10: Fig. 1. HRV-14 $3C^{pro}$ structure, mutation data and basic residues. **(a)** Cartoon rendering. Two six-stranded β-barrels (peach and lavender ribbons) are joined by a long linker (yellow; residues 76-97). Basic and non-basic residues that, when mutated, inhibit RNA-binding (Andino *et al.*, 1993; Leong *et al.*, 1993) are shown as teal and green sticks, respectively. Sidechains of other nearby basic residues are shown as dark blue sticks. The active site catalytic triad (grey spheres) is in the rear of the molecule in this view and contains the inhibitor JMC-98-3 (inhibitor not shown). **(b)** Surface rendering. Orientation and colouring are as in part A. **(c)** shows the electrostatic potential of the protein surface with the mutation data. **(d)** Shows the iso-electric potential as calculated by APBS (Baker *et al.*, 2001).

The sequence of the RNA bound by 3C^{pro} has been extensively studied. The minimum sequence appears to be a 27-28mer sequence present on stem loop d of the initial cloverleaf of the 5′ UTR. This sequence is not very highly conserved within the family. However, it seems that the overall 3D structure of the RNA is quite well conserved. HRV 14 SLD has been extensively studied and the solution structure has been solved by NMR (Huang *et al.*, 2001; Headey *et al.*, 2006). The consensus sequence for stemloop in renteroviruses (rhino- and enteroviruses) is shown in Figure 1.11

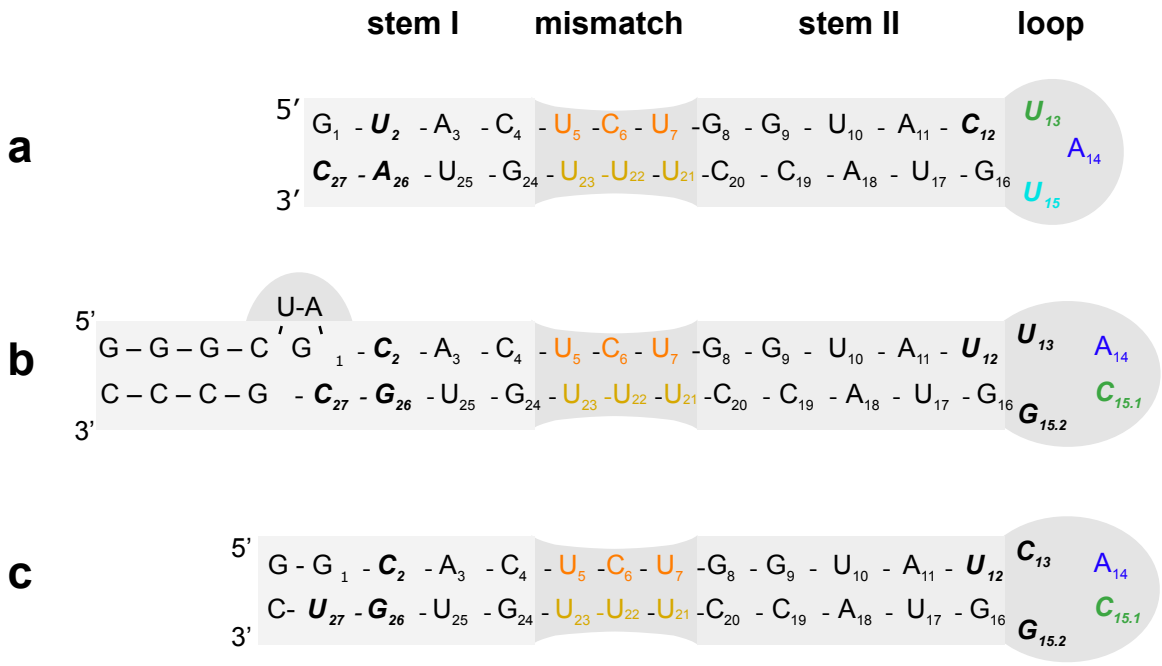


Figure 1.11: SLD sequences for HRV-14 (a) consensus for rhino- and enteroviruses (b) and for coxsackie virus (c). All these sequences have had their structure solved in solution (Figure 1.12).

The stemloop consists of two regions of conventional Watson-Crick base pairing (stem I and stem II) with a pyrimidine mismatch 'cinch' region and an apical tri- or tetraloop. The triloop is only present in rhinovirus serotypes. Other picornaviruses have a tetraloop. Interestingly the rhinoviral 3C^{pro} can bind the stemloop D from coxsackie virus and poliovirus (Walker *et al.*, 1995) but the opposite interaction between poliovirus and coxsackie virus 3C^{pro} and the HRV14 SLD does not occur (Xiang *et al.*, 1995; Blair *et al.*, 1998; Zell *et al.*, 2002; Rieder *et al.*, 2003). The tetraloop motif found in most SLD structures is actually structurally more like a biloop as the first two bases nearly base-pair (Figure 1.11). Thus, the triloop found in HRV-14 SLD is possibly provides more steric hindrance to binding of 3C^{pro} from other species of virus. These data indicate that differences in the structure or sequence of the apical loop region are important in RNA

recognition by $3C^{pro}$.

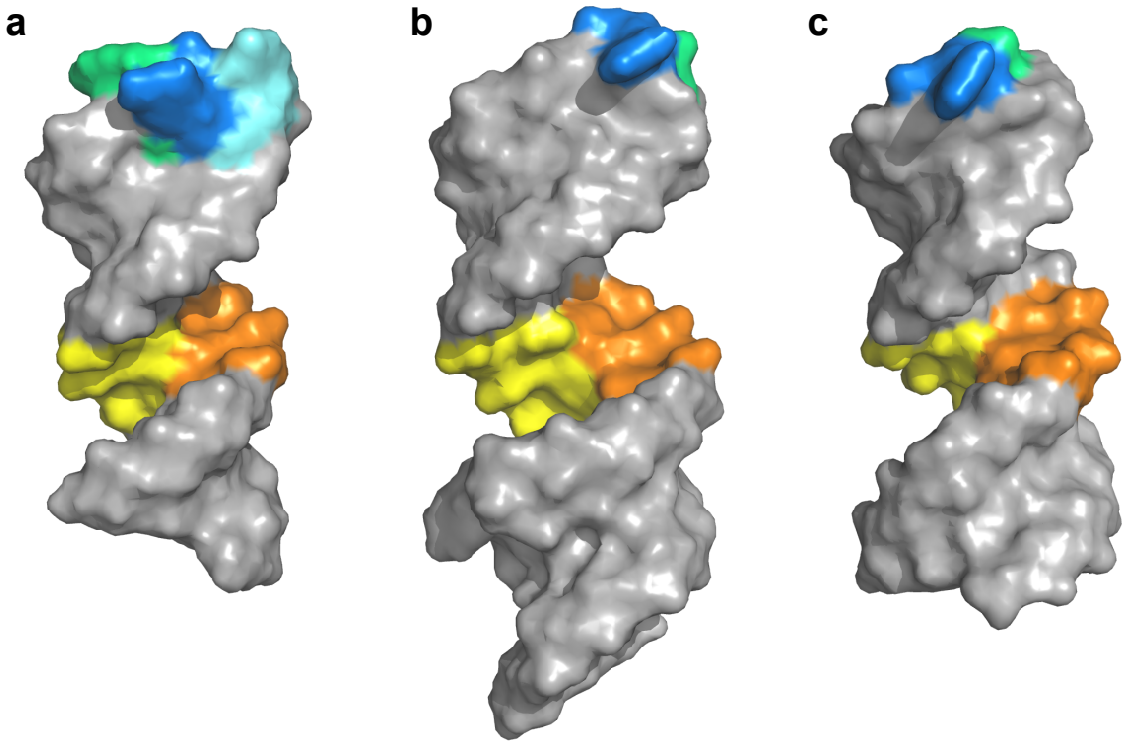


Figure 1.12: Surface representation of (a) human rhinovirus 14 (Headey *et al.*, 2006), (b) rhino- and enterovirus consensus sequence (Du *et al.*, 2004) and (c) coxsackie virus (Ohlenschläger *et al.*, 2004) stemloop-d. The cognate bases are shown in green and blue and the 'extra' residue in the triloop shown in cyan.

1.9 Research Aims

Although structures of several picornaviral 3C^{pro} isotypes have been solved, as well as the solution structure of SLD in human rhinovirus 14, prior to this study no structure for the complex of 3C^{pro} and SLD had been solved. The aim of this research project was to characterise the solution structure of the complex between 3C^{pro} and stemloop-d from human rhinovirus 14. The intended characterisation consisted of four parts:

(i). Elucidation of the NMR-derived structure and dynamics of 3C^{pro} in solution. This aim was modified to focus on the dynamics of the protein after the solution structure of 3C^{pro} was solved (Bjorndahl *et al.*, 2007).
(ii). Mapping of the RNA binding surface of 3C^{pro} by NMR methods.
(iii). Modelling of the structure of the complex using the data from (i) and (ii) as restraints.
(iv). Determination of the solution structure of the complex.

The various stages required several different strategies: (i). Requires the production of milligram quantities of soluble ^{15}N-labelled 3C^{pro} and the use of various NMR pulse sequences to define relaxation parameters of the protein. (ii). Requires NMR-based investigation of interaction of the 27-base SLD from HRV-14 with 3C^{pro}. (iii). Uses software-based methods to model the interaction. (iv). Involves the use of NMR restraints to solve the solution structure of the complex.

Determination of the structural basis of interactions involved in the replication of viruses is necessary to fully understand such interactions and can be used as the basis for rational drug-design.

Longum iter est per praecepta,

breve et efficax per exemplа.

Seneca (4 BC – AD 65)

2

Characterisation of $3C^{pro}$ in Solution

2.1 Introduction

The aim of the research described in this chapter was to produce human rhinovirus 14 (HRV-14) 3C protease (3C^{pro}) in sufficient quantity, and of sufficient quality, for study by NMR. In order to study a protein by solution-state NMR it is necessary to produce milligram quantities of folded, soluble protein. It is also desirable that an NMR sample contain a low concentration of salts as small, charged molecules will affect the matching and sensitivity of the probe. This effect is especially pronounced in spectrometers equipped with a cryoprobe.

The 3C^{pro} sample was intended to be used, along with the backbone assignments of HRV-14 3C^{pro} present in the BMRB (Bjorndahl *et al.*, 2003), to follow the ^{15}N,^{1}H-HSQC shift changes upon addition of stemloop-d RNA (SLD) to the sample. Chemical shift perturbation (CSP) analysis has been shown to be an effective way of determining the interaction surface of proteins in interactions with other molecules (Zuiderweg, 2002). It was not initially deemed necessary to completely assign the protein or do a full NMR structure calculation. However, early spectra revealed that the shifts of protein expressed here differed sufficiently from the published shifts to require backbone assignment. Upon backbone assignment of the apo protein it was discovered that a number of resonances were broadened beyond detection, which necessitated further optimisation of protein purification and preparation in order to get a sample suitable for further study of the 3C^{pro} and study of the 3C^{pro}–RNA complex.

As has been noted in Chapter 1, the structures of several isoforms of 3C^{pro} are available in the Protein DataBank. Initially in this work, a homology model of HRV-14 3C^{pro} was created from a crystal structure of HRV-2 (section 2.2.10). The homology model proved useful for early work on the system but a study published in 2007 (Bjorndahl *et al.*, 2007) provided the solution structure of HRV-14 3C^{pro}. This was used as a basis for further work (section 2.2.10). A structure was solved for HRV-14 3C^{pro} using the Bjorndahl *et al.* solution structure and residual dipolar couplings (RDCs).

In the following, 3C^{pro} from HRV-14 is referred to as 3C in its apo form and 3CI when complexed with a specific inhibitor (JMC-98-3, (Webber *et al.*, 1998)).

2.2 Methods

2.2.1 Transformation of BL21 cod$^+$ *Escherichia coli* cells with HRV-14 3C

H-MBP 3C^{pro} had previously been cloned and the construct was available in the lab. Transformation of cells using the hexamine cobalt method was unsuccessful (Hanahan, 1983). Transformation was achieved by electroporation (Dower *et al.*, 1988). H-MBP-3C^{pro} miniprep cDNA (50 ng μl^{-1}) was diluted 1 in 16. A 5 μl amount of the diluted cDNA of this was added to 50 μl of competent BL21 cod$^+$ *E. coli* and this was put on ice for 1 minute. The cells were electroporated with a 2.5 kV, 25 μF 4.7 mS pulse and then pipetted into 4 ml of warm Luria Broth (LB) containing 50 $\mu g \cdot ml^{-1}$ampicillin. The cells were incubated for 1hr at 37 °C in a 50 ml Falcon tube. 500 μl of this culture was plated out onto an LB (containing 34 $\mu g \cdot ml^{-1}$ chloramphenicol and 50 $\mu g \cdot ml^{-1}$ ampicillin) agar plate.

2.2.2 Unlabelled test expression and production of glycerol stocks of transformed *E. coli*

One colony of BL21^{cod+} *E. coli* transformed with H-MBP 3C^{pro} from a plate of LB agar containing 50 $\mu g \cdot ml^{-1}$ ampicillin and 34 $\mu g \cdot ml^{-1}$ chloramphenicol was inoculated into 10 ml of LB containing 100 $\mu g \cdot ml^{-1}$ ampicillin and grown overnight at 37 °C. Two 850 μl aliquots of this overnight culture were pipetted into separate Eppendorf tubes and 150 μl of sterile glycerol was added to each. The tubes were vortexed thoroughly to mix in the glycerol and the culture was then snap frozen in liquid nitrogen and stored at –80 °C.

A 5 μl aliquot of the overnight culture was inoculated into 50 ml of LB containing 100 $mg \cdot ml^{-1}$ ampicillin and incubated at 37 °C. When the cell density had reached 0.5 the cells were induced by addition of 1 mM IPTG (12 mg). The cells were grown for three hours at 37 °C . Cells were harvested by centrifugation at 3,000 rpm 4 °C for 10 mins. Expression levels were ascertained by SDS-PAGE gel stained with Coomassie blue.

2.2.3 Glucose optimisation procedure

Initial trials yielded good results when the cells were induced and expressed at 37 °C but it was found that the majority (~60 %) of the expressed protein was insoluble. Carrying out the expression at 30 °C increased the solubility such that nearly 100 % of the recombinant protein

was soluble. A trial was set up to optimise the carbon source for labelled growth. 1.8 l of LB containing 34 $\mu g \cdot ml^{-1}$ chloramphenicol and 50 $\mu g \cdot ml^{-1}$ ampicillin was inoculated with a colony of *E.coli* that had been transformed with the H-MBP-3C^{pro} plasmid. This was grown to an OD_{600} of 0.25, 1 g of glucose was added and the growth continued until the OD_{600} was ~2. The cells were then harvested by spinning down at 1541 $\times$ *g* for 15 mins. The pellets were resuspended in 900 ml of M9 minimal media containing 50 $\mu g \cdot ml^{-1}$ ampicillin giving an OD_{600} of 3.9. 300 ml was transferred to each of three Erhlenmeyer flasks, designated 3CU1, 3CU2 and 3CU3. After incubation for half an hour 1, 2 and 3 grams of glucose were added to flasks 3CU1, 3CU2 and 3CU3. This corresponded to 3.3 $g \cdot l^{-1}$, 6.6 $g \cdot l^{-1}$ and 9.9 $g \cdot l^{-1}$, respectively. The cultures were incubated until bacterial growth, monitored by OD_{600}, reached stationary phase. As expected this happened first in 3CU1 then 3CU2 and finally 3CU3. The cells were harvested by spinning down at 1541 $\times$ *g* for 15 mins at 4 °C and frozen at −80 °C. The pellets were each resuspended in 20 ml of lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl) containing one Complete™ EDTA-free protease inhibitor tablet (Roche). The resuspended cells were then lysed by passing twice through a French press (built in-house) at 6,000 psi. The lysates were spun down separately at 17,590 x g for 20 mins. Each cleared lysate was incubated rocking overnight at 4 °C with 2.5 ml of Ni^{2+}–NTA resin (GE Healthcare) pre-equilibrated with lysis buffer. The slurry was spun down at 1784 $\times$ *g* for 5 mins. The pellets were transferred to Poly-Prep© columns (Bio-Rad, CA). The columns were washed sequentially with 4 column volumes of 50 mM Tris pH 7.5, 1 M NaCl buffer. The columns were then washed with 2 column volumes of lysis buffer containing 30 mM imidazole. Finally the columns were eluted with 2 column volumes of 50 mM Tris pH 7.5, 1 M NaCl, 250 mM imidazole buffer.

The 3CU1 growth gave approximately 20 mg of fusion protein. The 3CU2 growth produced approximately 50 mg of fusion protein although a proportion of the expressed fusion protein was insoluble. 3CU3 produced slightly more protein than 3CU2, however almost 40% of this protein was insoluble. Thus, the amount of soluble fusion protein produced per gram of glucose was highest for 3CU2 with 2 g of glucose. This corresponds to 6 g l^{-1} of glucose.

2.2.4 Trials of different affinity media

Initially Talon™ Co^{2+} IMAC (immobilised metal-ion affinity) resin (BD Biosciences, Palo Alto, CA) was used to purify the fusion protein. However, this resulted in insufficient binding, with a significant fraction of the fusion protein being eluted along with impurities at low imidazole

	Glucose Added (g)	Protein Expressed (mg)	Soluble Fraction (%)	Soluble Protein Produced (mg) per gram of Glucose
3CU1	1	20	100	20
3CU2	2	50	90	22.5
3CU3	3	70	60	14

Table 2.1: Table of glucose optimisation parameters. Protein concentrations were calculated using Bradford Assay.

concentrations. A maltoheptaose-agarose column (Sigma Aldrich, St Louis, MO) was tried as an alternative. This was worse, resulting in the loss of a significant fraction of the fusion protein in the flow through. It has been previously shown that Ni^{2+} affinity media has a higher affinity for His-tagged protein than Co^{2+} affinity media (Talon™ manual). The use of Ni^{2+}-NTA affinity media (GE Biosciences) increased the yield of protein and therefore this was used for all subsequent purification procedures.

2.2.5 Cleavage of fusion protein and final polishing

The majority of protein was purified using a Ni^{2+}-NTA affinity column. The fusion protein was eluted from the column with 250 mM imidazole buffer. The MBP was cleaved from the recombinant 3C by incubation with bovine thrombin. The MBP and thrombin as well as other contaminating proteins were removed by anion exchange chromatography. The sample was dialysed into two successive 1 l volumes of 20 mM Tris-base pH 7.8, 1 mM DTT, 0.5 mM EDTA. The sample was then passed through a 5 ml Q-sepharose column (Sigma Aldrich, St Louis, MO). As $3C^{pro}$ has a relatively high pI (7.9 predicted by Swiss-Prot ProtParam tools), nearly all contaminating protein was bound to the column in this step, allowing the pure 3C to flow through the column. The eluate was subsequently concentrated using a centrifugal concentrator (GE Biosciences) with a 10 kDa molecular weight cut-off.

2.2.6 Inhibitor synthesis

Since only limited quantites of the inhibitor were available in the lab, the descision was made to synthesise enough for further studies.

Synthesis was based on the method of Webber (Webber et al, 1998). The following scheme was used:

0.01 moles (mw 265.31, 2.65g) of CBZ-L-Leu were dissolved in 25 ml of dioxane. An equimolar amount of L-Phe (mw 165, 1.65g) was dissolved in 25 ml of de-ionised H_2O containing 0.02 moles of sodium bicarbonate. 0.012 moles of hydroxysuccinamide were dissolved in the aqueous fraction. This was left for an hour to form an active ester (CBZ-L-Leu-O-succinamide). The dioxane and aqueous fractions were then mixed and 0.01 moles of 1-ethyl-3-(3-dimethylaminopropyl) carbo-diimide (EDC) was added. This reaction was left stirring overnight. The reaction was made up to 100 ml with de-ionised H_2O and ethyl acetate extraction was performed to remove any unreacted CBZ-L-Leu. 20 ml of conc. H_2SO_4 was added and a further EtOAc extraction was performed. The solvent was removed under vacuum and the residue was freeze-dried. Yield was 1 g of CBZ-L-Leu-L-Phe-OH. 269 mg of CBZ-L-Leu-L-Phe-OH was dissolved in 20% dimethylformamide and 80% dichloromethane (0.75 ml of DMF and 3 ml DCM) 80 mg of HOSU 133 mg of EDC was added and the reaction was stirred for 3 hrs at room temperature. 100 mg of L-(*N*-Ac-amino-alaninol) was dissolved in 1:1 DMF/DCM (2 ml) and added to the first reaction. This was left stirring overnight. The resulting product was purified by silica flash chromatography. The CBZ-L-Leu-L-Phe-L-(*N*-Ac-amino-alaninol) was then dissolved in DMSO and a 3-fold excess of IBX was added. The mixture was stirred for 3 hours and the IBX was removed under vacuum. The residue was diluted in de-ionised H_2O and freeze-dried several times in order to remove any remaining DMSO. 50 ml of EtOAc was added to the residue and the resulting precipitate was filtered off. The filtrate was diluted to 200 ml in EtOAc and washed with 10% $Na_2S_2O_3$/10% $NaHCO_3$ solution, de-ionised H_2O and finally, brine. The organic phase was dried over Na_2SO_4. The solvent was removed by drying under vacuum at room temperature.

Further synthesis of CBZ-L-Leu-L-Phe-L-(*N*-Ac-amino-alaninal) was performed by D. Lunn and D. Harding. It was found to be more economical to purchase the CBZ-L-Leu-L-Phe-OH rather than synthesise this.

2.2.7 Flame spectroscopy

0.6 mM 3C and 0.6 mM 3CI samples were assayed by Ni^{2+} flame spectroscopy.

2.2.8 NMR sample preparation

Labelled ammonium chloride ($^{15}NH_4Cl$) and glucose (^{13}C-D-glucose) were obtained from Cambridge Isotope Laboratories inc., Andover, MA.

^{15}N labelled 3C

^{15}N-labelled 3C was prepared in a similar manner as described above. Typical preparation procedures are described. Competent BL21^{cod+} *E. coli* were transformed with H-MBP-3C^{pro} using the heat-shock method. One colony of the plasmid-containing bacteria was inoculated into 600 ml of LB containing ampicillin at 100 $\mu g \cdot ml^{-1}$ and chloramphenicol at 34 $\mu g \cdot ml^{-1}$ (LB$^{amp+/cam+}$). The culture was grown overnight at 30 °C (12hrs). The overnight culture was spun down at 1541 $\times$ *g* (Sorvall GS3) for 15 mins and re-suspended in 300 ml of M9 minimal medium containing 100 $\mu g \cdot ml^{-1}$ ampicillin and 34 $\mu g \cdot ml^{-1}$ chloramphenicol (M9$^{amp+/cam}$). The culture was agitated for an hour at 30 °C. 3 g of glucose and 1 g of $^{15}NH_4Cl$ were added via sterile filter. After 30 minutes the cells were induced by the addition of 300 µL 0.5 M IPTG for a final concentration of 0.5 mM. When the stationary growth phase was achieved, the cells were harvested by spinning down at 1541 $\times$ *g* (Sorvall GS3) for 15 mins. The final OD_{600} was between 4 and 7. The resulting cell pellets were re-suspended in 25 ml of lysis buffer containing one Complete™ EDTA-free protease inhibitor tablet (Roche) and placed on ice. The resulting slurry was passed twice through a French press at 6,500 psi. The resulting lysed cells were centrifuged at 17,590 $\times$ *g* (Sorvall SS34) for 30 minutes. The supernatant from this process was decanted from the pellet and sterile filtered with an 0.8 µm filter. The filtered solution was then added to 3.5 ml (bed volume) of Ni^{2+}-NTA affinity medium and nutated overnight at 4 °C. The overnight slurry was poured into a Poly-Prep© column (Bio-Rad, CA) and the flow-through collected. The column was subsequently washed with 8 ml of lysis buffer. The column was then successively washed with 8 ml aliquots of lysis buffer containing 20 mM, 40 mM and 80 mM imidazole respectively. The column was eluted with 8 ml lysis buffer containing 250 mM imidazole. 50 units of bovine thrombin (Sigma) was added to the eluate and this was transferred into a 3 kDa MWCO dialysis bag and left for 24 hrs at 4 °C dialysing into 50 mM Tris-HCl pH 7.4, 300 mM NaCl, 1 mM DTT (Initial work used dialysis into 50 mM Tris-HCl pH 7.4, 300 mM NaCl, 0.25 mM TCEP at this stage. Little cleavage was observed, however. The 50 mM–80 mM fractions were pooled and added to the sample, a further 50 units of thrombin were added and this was then dialysed for 24 hrs into 50 mM Tris-HCl pH 7.4, 300 mM NaCl as previously but with 1 mM DTT as a reducing agent instead of TCEP. This resulted in better cleavage and DTT was used as a reducing agent for later preparations). The sample was passed through a 3.5 ml Ni^{2+}-NTA column and the subjected to two 5 ml washes of lysis buffer. The flow-through and the lysis buffer wash were pooled and dialysed into 20 mM Tris-HCl pH 7.5, 1 mM DTT and then run through a Q-sepharose column.

The flow through from this column was then concentrated to 525 µl. 25 µl of D_2O (5%) was added to produce the NMR sample. Final sample concentration was between 0.3 and 1.3 mM. The nmr tubes were flushed with argon and sealed to prevent oxidation. A representative gel is shown as Figure 2.1. The final 3C sample is pure within the limits of detection by Coomassie staining.

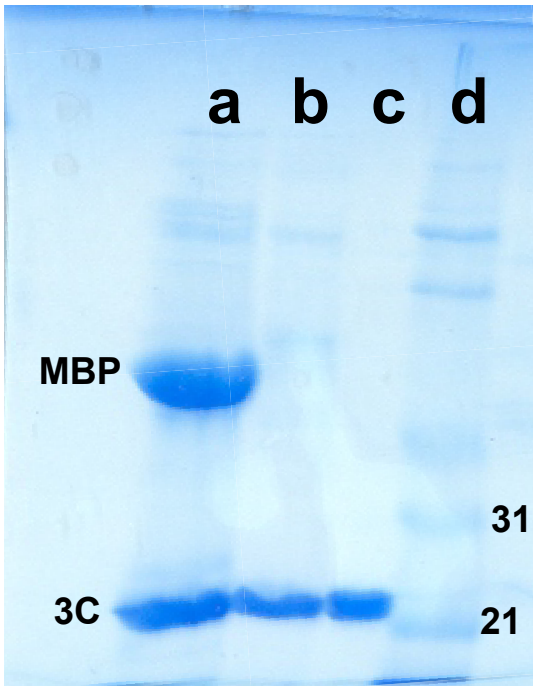


Figure 2.1: SDS-PAGE gel of samples from final 3C purification steps. Lane (a) contains the fusion protein post thrombin cleavage (b) contains 3C post 2nd pass through the Ni^{2+}-NTA column (c) contains 3C after final Q-sepharose polishing and (d) contains molecular weight standards.

^{13}C-^{15}N-labelled 3C

Typically, a flask containing 600ml of LB$^{amp+/cam+}$ was inoculated with 1 colony of H-MBP-3C^{pro}. This was left overnight shaking at 37 °C. The overnight culture was spun down at 1541 $\times$ g and the cells resuspended in 300 ml of M9^{amp+} giving an OD_{600} of 3.0. The culture was left for 30 mins and 3 g of ^{13}C-D-glucose and 1 g of $^{15}NH_4Cl$ were added to the culture via sterile filter. After 1 hour the expression was induced by addition of 0.072 g of IPTG giving a final concentration of 1 mM. The cells were harvested when the growth curve flattened out (final OD_{600} was ~6.5) by spinning down at 1541 $\times$ g for 15 mins at 4 °C. The pellets were resuspended in 30 ml of lysis buffer containing one Complete™ EDTA-free protease inhibitor tablet (Roche). The resuspended cells were passed either through a cell disruptor three times at 13,000 psi or a French press three times at 6,000 psi. The resulting lysed cells were centrifuged at 17,590 $\times$ g (Sorvall SS34) for 30

minutes. The supernatant from this process was decanted from the pellet and sterile filtered with an 0.8 μm filter. The filtered sample was combined with 2 ml (bed volume) of Ni^{2+}-NTA column pre-equilibrated with lysis buffer and left nutating overnight at 4 °C (it was found that efficient binding to Ni^{2+}-NTA could be achieved by passing the filtered sample under gravity through the resin in column format, this was much quicker and thus was used for later preparations). The slurry was then placed into a Poly-Prep© column (Bio-Rad, CA) and washed sequentially with two 4 ml aliquots of lysis buffer. Two further 5 ml washes of lysis buffer containing 50 mM imidazole and 100 mM imidazole were carried out. Finally the protein was eluted with 5 ml of lysis buffer containing 250 mM imidazole. All three imidazole-containing fractions were found to contain fusion protein so they were pooled (further preparations used only one 250 mM imidazole elution after the imidazole-free lysis buffer washes). 50 units of thrombin (Sigma) were added and the sample was dialysed into 2 l of 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM DTT for 72 hrs. The resulting sample was passed through a Ni^{2+}-NTA column pre-equilibrated with lysis buffer. The sample was then dialysed into 2 l of 20mM Tris-HCl pH 7.5, 1 mM DTT overnight and then passed through a 3.5 ml Q-sepharose column pre-equilibrated with 20 mM Tris-HCl pH 7.5, 1 mM DTT. Purity was increased in later work by raising the pH of this buffer to 7.8, this increased the affinity of protein contaminants for the Q-sepharose column. The sample was concentrated to 550 μL and buffer exchanged into 20 mM potassium phosphate pH 6.5, 15 mM DTT, 0.5 mM EDTA. Later preparations used 50 mM BTM (50 mM bis-Tris propane / 50 mM MES (2-(*N*-morpholino)ethanesulphonic acid)) pH 6.0, 15 mM DTT 0.5 mM EDTA in the NMR sample buffer. 5% D_2O was added and the sample was placed into a standard 5 mm NMR tube. Yields of pure 3C were between 13 and 30 mg per litre of cell culture. The typical final protein concentration in NMR samples was between 11–22 mg· ml^{-1} (0.5–1 mM).

2.2.9 NMR experiments

Several NMR techniques were used to characterise the protein in this study. One-dimensional proton NMR was used initially to give an idea of proton peak-dispersion and to assess the concentration of the sample. Subsequently, two-dimensional techniques (principally ^{15}N,^{1}H-HSQC) were used to assess whether the protein was folded. The ^{15}N,^{1}H-HSQC involves the transfer of magnetisation from amide protons to the nuclei of their attached nitrogen atoms via an INEPT (Insensitive Nuclei Enhanced by Polarisation Transfer) sequence. The chemical shift of the nitrogen nucleus is then recorded via a series of increments and the magnetisation is re-

turned to proton via a reverse-INEPT sequence and then detected. The basic HSQC sequence can be added to. The ^{15}N,^{1}H-HSQC thus formed the basis for further experiments such as three-dimensional through-bond (eg. HNCO, HNCA) or through-space (^{15}N,^{1}H-NOESY-HSQC) experiments. The backbone was assigned with HNCA (Kay *et al.*, 1990; Grzesiek and Bax, 1992) and CBCANH (Grzesiek and Bax, 1992) spectra. Assignments were confirmed with HNCO (Kay *et al.*, 1990; Grzesiek and Bax, 1992; Muhandiram and Kay, 1994), HN(CA)CO (Clubb *et al.*, 1992) and CBCA(CO)NH (Grzesiek and Bax, 1993) experiments. TROSY variants of the ^{15}N,^{1}H-HSQC (Pervushin *et al.*, 1997) were used for some samples. Preliminary experiments (data not shown) indicated that the best temperature for acquisition was 295 K as higher temperatures led to sample degradation and lower temperatures to significant loss of signal due to line broadening. Further experiments were carried out at this temperature. Chemical shift indexing (CSI) (Wishart and Sykes, 1994) can be used to determine the secondary structure of proteins from their backbone chemical shift values. CSI is a relatively quick and simple way to assess changes in structure without requiring a full stucture to be solved. Backbone C$^{\alpha}$, C$^{\beta}$ and C′ chemical shift values were used in the CSI analysis. ^{15}N chemical shifts are more dependent on local chemical environment than on the backbone torsion angle and therefore are less informative for use in CSI analysis.

2.2.10 Structural modelling of 3CI

Homology model using HRV-2

Although the crystal structure of HRV-14 3C^{pro} had been solved (Matthews *et al.*, 1994), the co-ordinates for this structure were not available in the PDB. In order to get a HRV-14 3C^{pro} structure, I homology modelled HRV-14 3C^{pro} based on the crystal structure of HRV-2 3C^{pro} (PDB accession 1CCQ). This protein shares 49% identity to the equivalent protein in HRV-14. The SwissPROT homology modelling server was used with default settings.

Homology model using HRV-14 and RDCs

Recently the solution structure of HRV-14 3C^{pro} in apo (3C^{B}) and in complex with a peptide inhibitor (3CI^{B}), was published (Bjorndahl *et al.*, 2007). This work was used as a basis for a homology model of 3C^{pro} in complex with the inhibitor used in this study (CBZ-L-Leu-L-Phe-L-(*N*-Ac-amino-alaninal). The inhibitor was modelled using ChemDraw and the 3D co-ordinates generated using the programme Corina (Molecular Networks) and built into the structure of 3C^{pro} using CNS (Brünger, 2007; Brünger *et al.*, 1998). The structure was then energy-minimised,

using a simulated annealing protocol, and water refined. Finally, ^{1}H-^{15}N RDCs were used to further refine the structure. RDC data were aquired following the method of Ruckert and Otting (Ruckert and Otting, 2000). ^{15}N,^{1}H-HSQC-IPAP spectra (Ottiger *et al.*, 1998; Cordier *et al.*, 1999) of 3C^{pro} yielded a total of 123 well-resolved backbone ^{15}N,^{1}H RDCs. PALES (Zweckstetter and Bax, 2000) was used to generate an alignment tensor for the 3C^{pro}. This alignment tensor was then used in CNS (Brünger, 2007; Brünger *et al.*, 1998), along with the experimental RDC values as restraints, to refine the 3CI homology model.

2.3 Results

2.3.1 Initial sample NMR

One-dimensional (1D) and two-dimensional (2D) experiments were carried out on 3C using a Bruker Avance spectrometer operating at a Larmour frequency of 700 MHz, equipped with a cryoprobe. Examination of the amide region of the 1D spectrum revealed good dispersion, with amide peaks which could be observed from ~6 to ~10 ppm. Random coil chemical shift values for proteins are usually confined to a region between 7.5 and 8.5 ppm, the numerous peaks visible outside this region indicate well defined secondary structure. Figure 2.2 shows a 1D proton spectrum of 3C.

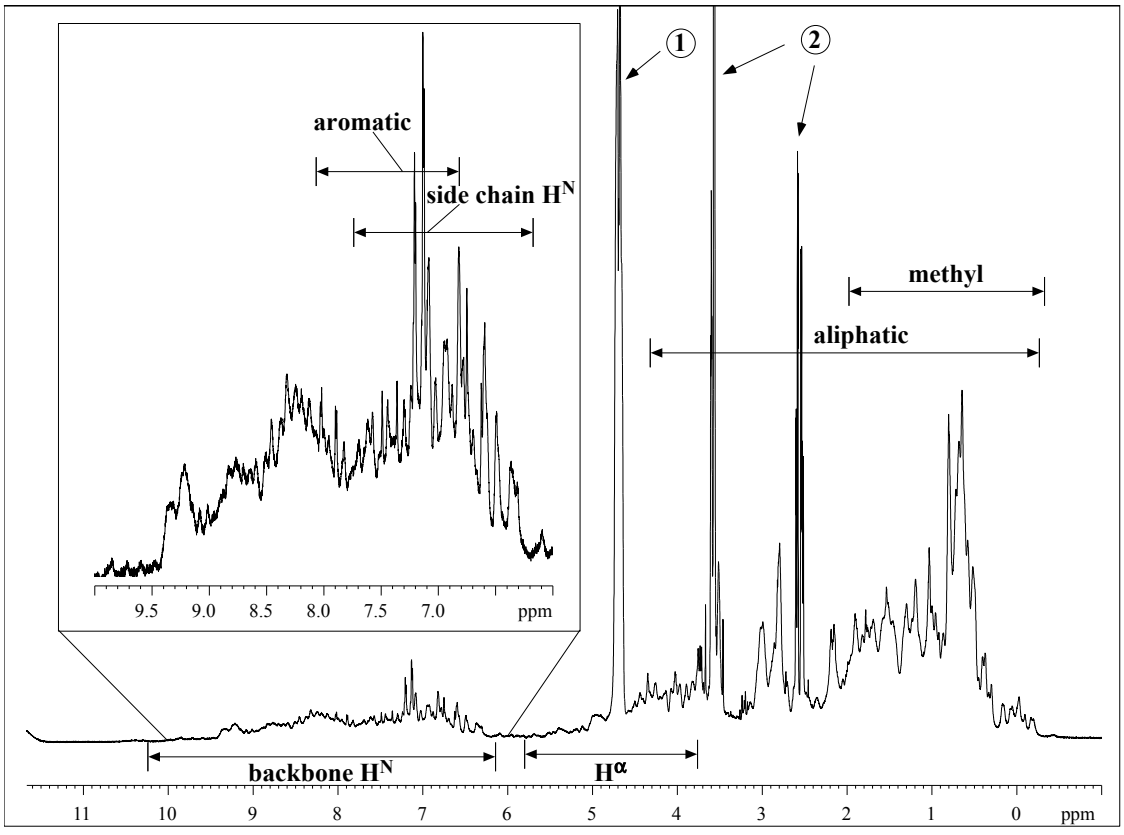


Figure 2.2: Proton 1D spectrum of 3C showing the regions where peaks from various protein functional groups are generally located.① indicates the residual water peak and ② indicates the peaks belonging to the BTM buffer. The inset shows a magnification of the amide proton region.

^{15}N,^{1}H-HSQC spectra of 3C provided further evidence for the folded nature of the protein and confirmed that the peaks in the amide region of the 1D spectra indeed belonged to amide

proton resonances. It can be seen from the spectrum presented in Figure 2.3 that the protein contained approximately the right number of peaks. In this case there should be 182 observable amide proton resonances. Native 3C contains 6 prolines which have no amide protons. However, the construct used in this study also contains an N-terminal cloning artefact consisting of 6 amino acids. Initially it had been hoped to use previously published chemical shifts of $3C^{pro}$ (Bjorndahl *et al.*, 2003) to assign the spectra. Further work could then have been undertaken using 2D experiments on ^{15}N-labelled protein rather than having to utilise 3D (three-dimensional) experiments, which require incorporation of both ^{13}C and ^{15}N labelling. Unfortunately examination of the ^{15}N,^{1}H-HSQC spectra of the 3C sample indicated that the shifts were significantly different to those published (NMR carried out on a sample of 3C in identical buffer conditions to those in the previous work indicated that this was not due to sample conditions alone (Figure 2.4)). In addition to this, peak overlap was considerable in parts of the spectrum, making unambiguous assignment impossible. It was decided to express ^{13}C,^{15}N labelled protein and use 3D spectra to assign the protein backbone. This would then enable unambiguous identification of the amide peaks and permit analysis of titration studies on the 3C–SLD complex. The ^{13}C, ^{15}N sample began to precipitate at higher concentrations required for 3D experiments, however. A further sample was produced using the same protocol but buffer exchanging the final sample into 20 mM potassium phosphate pH 6.5, 15 mM DTT, 0.5 mM EDTA (Bjorndahl *et al.*, 2003). This did not fully alleviate the solubility problems. The solubility issues were solved as a by-product of a change in buffer conditions due to the installation of a cryoprobe on the 700 MHz spectrometer. The cryoprobe is much more sensitive to conductivity of the buffer than a room-temperature probe and it is therefore desirable to use a low-conductivity buffer system (Kelly *et al.*, 2002). By using 50 mM bis-Tris propane / 50 mM MES (BTM) to buffer the sample, the matching of the probe was improved, increasing the sensitivity of the instrument. Screens of protein samples in different buffer systems showed that, as a fortunate side-effect, the protein was more stable and soluble in the BTM buffer.

2.3.2 Assignment of apo 3C

Standard 'backbone walk' approaches using the previously mentioned 3D-spectra were used to assign the apo protein. However, this approach failed to produce assignments for approximately half of the protein. When the ^{15}N,^{1}H-HSQC spectra on the ^{15}N-labelled (N2-3C) sample and the ^{13}C, ^{15}N labelled sample (CN1-3C) were overlayed it became apparent that several of the

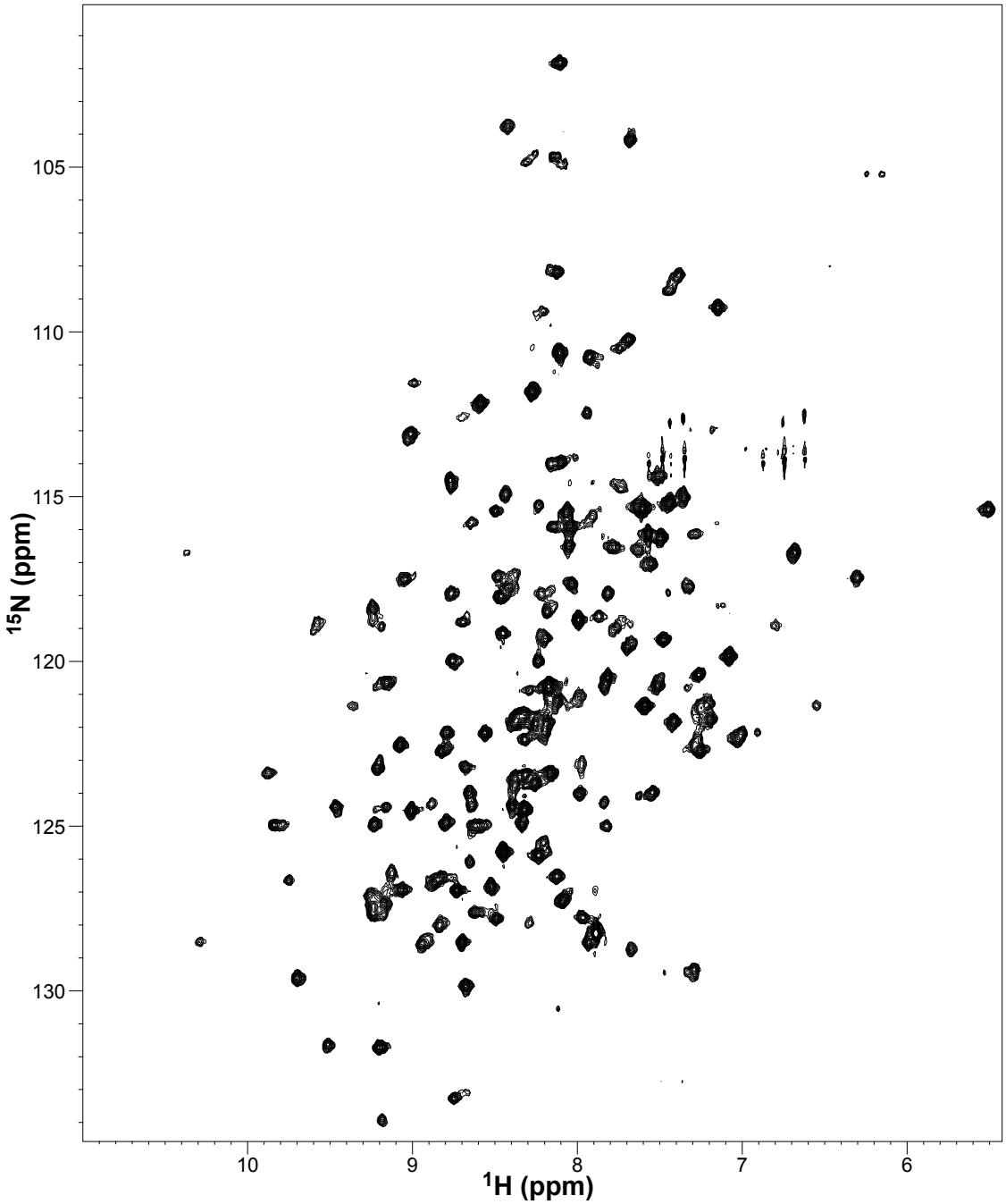


Figure 2.3: ^{15}N,^{1}H-HSQC-TROSY spectrum of 3C^{pro} in 20 mM potassium phosphate pH 6.5, 2 mM DTT, 0.5 mM EDTA, 295 K with full complement of peaks. The TROSY pulse sequence produces extremely narrow side-chain amide peaks.

amide peaks were missing. In Figure 2.6 the N2-3C spectrum is shown in black and the CN1-3C spectrum is shown in red. Using buffer conditions identical to previously published work (Bjorndahl *et al.*, 2003) failed to recover the missing resonances. A spectrum of 3C in these conditions is shown in Figure 2.4.

Once assigment of the backbone had been taken as far as possible, the residues with assigned peaks were plotted onto a representation of a homology model of HRV-14 $3C^{pro}$. The result of this plot is shown in Figure 2.5.

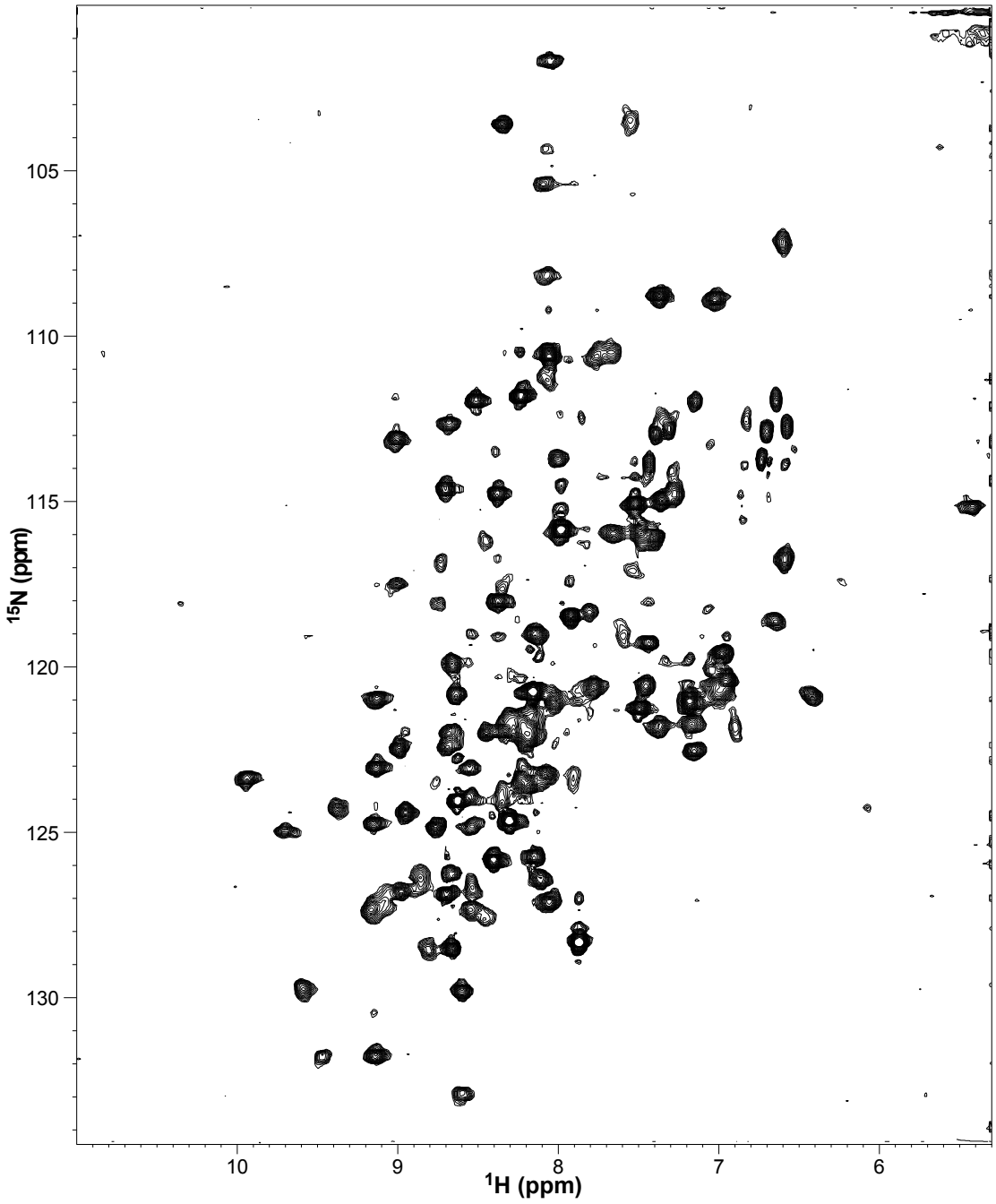


Figure 2.4: ^{15}N,^{1}H-HSQC spectrum of 3C^{pro} in buffer conditions identical to previous studies (20 mM phosphate, pH 6.5,15 mM DTT, 0.5 mM EDTA, 298 K) (Bjorndahl *et al.*, 2003). There are still over 80 peaks that are unobservable.

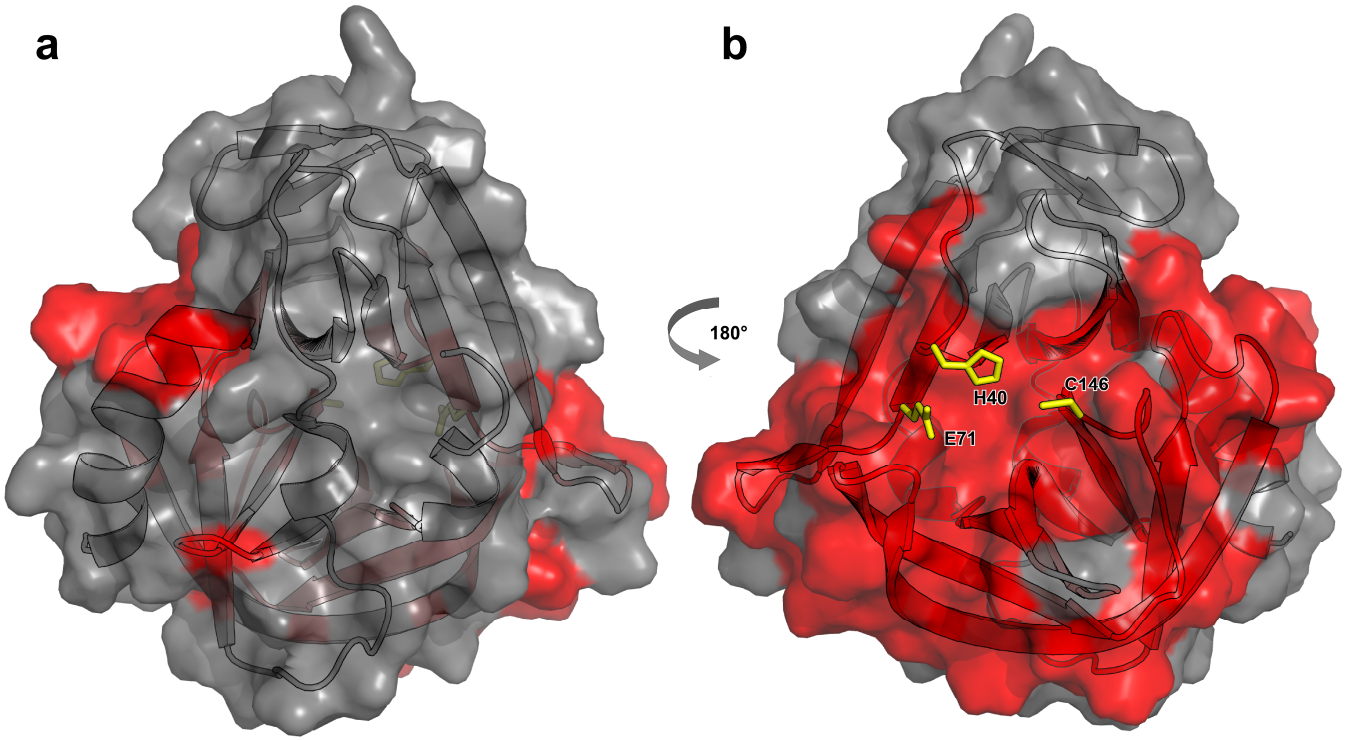


Figure 2.5: Plot of residues with missing amide peaks in ^{15}N,^{1}H-HSQC onto structure of human rhinovirus 14 3C^{pro}. Residues with missing peaks are shown in red and the catalytic triad residues are shown as yellow sticks.

The active site residues shown all have amide peaks missing from the ^{15}N,^{1}H-HSQC spectra as well as the residues coloured red in Figure 2.5. The residues with missing peaks cluster on the active site face of 3C. It is likely the missing peaks are due to exchange dynamics on the NMR timescale. If a resonance is undergoing conformational exchange at a rate much lower than the chemical shift difference between the two states then it is said to be in slow exchange (Cavanagh *et al.*, 2007). Two peaks are observed, one for each conformation. A resonance that is undergoing conformational exchange on a timescale much faster than the difference in shifts is in fast exchange and a single peak is observed at a position weighted by the relative population in each state. If a spin is undergoing exchange on a timescale that is similar to that of the difference in chemical shift, it is in intermediate exchange and a single broad peak will be observed, this will probably be in or below the noise in the ^{15}N,^{1}H-HSQC. In the case of 3C the cause of the intermediate exchange is likely to be conformational exchange of the protein. The two β-barrels of the protein may be also mobile relative to each other in solution (discussed further in Chapter 3).

2.3.3 Flame spectroscopy

Flame spectroscopy of samples of the protein gave a Ni^{2+} content of 0.1 $\mu g \cdot ml^{-1}$ for 3C and 0.3 $\mu g \cdot ml^{-1}$ for 3CI. This gives a Ni^{2+} concentration of 1.92 μM for 3C and 5.77 μM for 3CI. Since a 0.6 mM sample of each protein was used, the 3C sample has 0.0032 Ni^{2+} per 3C molecule and the 3CI sample has 0.0096 Ni^{2+} per 3CI molecule. This indicates that there is not enough Ni^{2+} in solution to cause the exchange-broadening effects seen in the 3C^{pro} samples.

2.3.4 Attempts to recover missing peaks

Increased solubility and maximum sample concentration of protein has been assisted by the use of arginine/glutamate buffer (Golovanov *et al.*, 2004). As transient aggregation or multimerisation of the protein was thought to be a possible reason for the observed peak broadening, 50 mM L-Arg and 50 mM L-Glu was added to the final NMR sample. The resulting spectrum is shown in Figure 2.7. Although a higher degree of protein solubility was achieved, it appeared that unfolded or partially unfolded protein was solubilised as well as folded protein. The unfolded protein in solution can be seen as the sharp peaks within the red box in Figure 2.7. It has also been suggested that shifts in this region are due to multiple C-terminal peaks (Julie Forman-Kay, personal communication). Short peptides will have different chemical shifts due to the effect

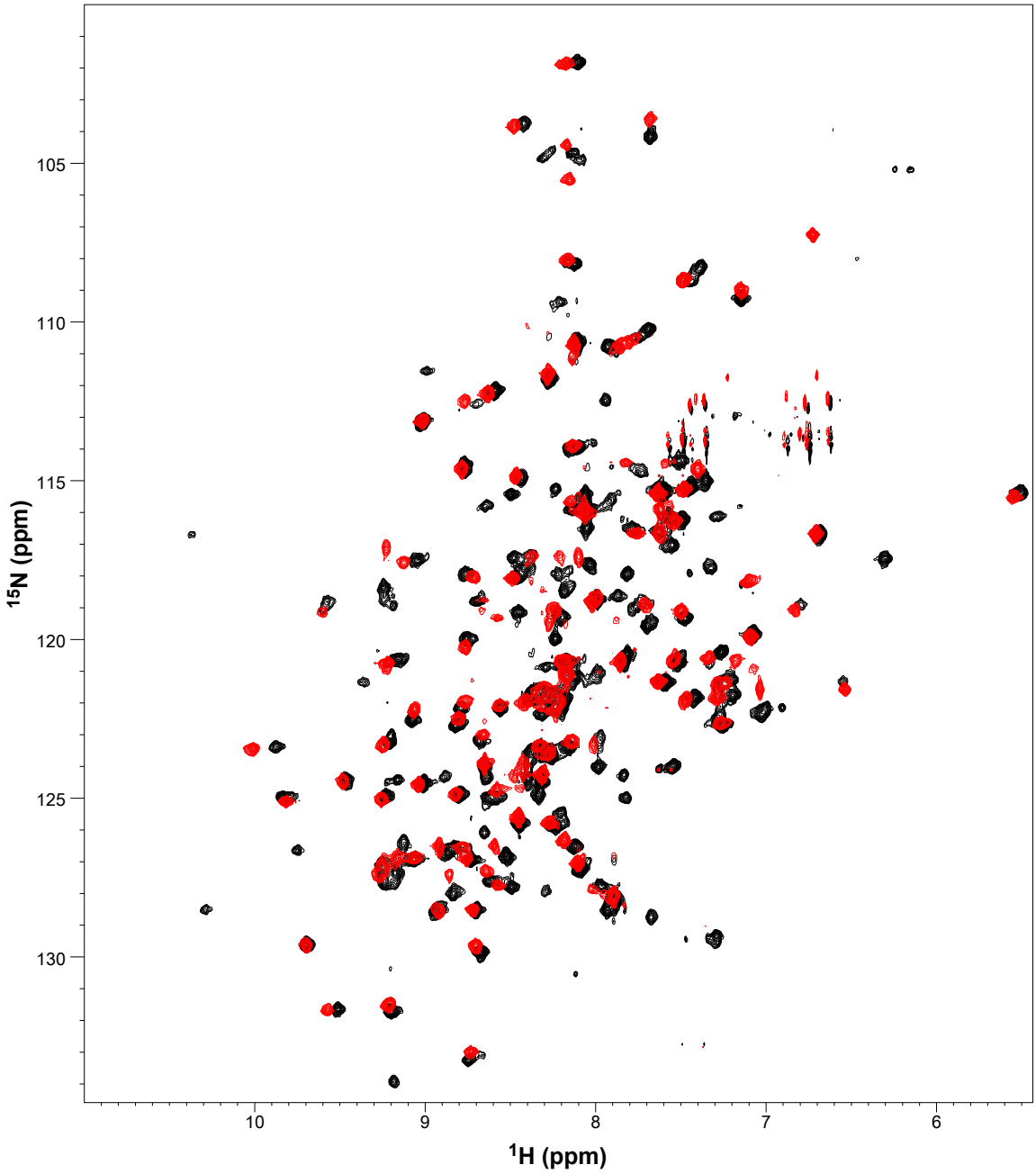


Figure 2.6: Overlay ^{15}N,^{1}H-HSQC-TROSY spectrum of 3C^{pro}(N2-3C) and 3C^{pro} (CN1-3C) sample conditions are 20 mM phosphate, pH 6.5,15 mM DTT, 0.5 mM EDTA, 295 K. N2-3C is shown in black and CN1-3C is shown in red. The side chain amide resonances are extremely narrow in TROSY spectra. It can be seen that there are far fewer peaks in the CN1-3C (red) spectrum.

of the charges at the termini that are distinct from the uncharged NH and CO in the context of a polypeptide chain. In fact, the C-terminal residue of 3C (Q182) is also in this region. Thus there is probably terminal degradation in this case. In addition, none of the unobservable peaks were recovered by addition of 50 mM L-Arg and 50 mM L-Glu. It is possible that the presence of the charged L-Arg and 50 mM L-Glu ions increased the activity of the protease to the extent that self-cleavage occurred. The presence of a high concentration of anions has been shown to increase 3C^{pro} activity (Wang and Johnson, 2001).

2.3.5 CSI

Chemical shift indexing (CSI) (Wishart and Sykes, 1994) was used to assess the secondary structure of the protein. Figure 2.8 shows a comparison of the CSI values for 3C, 3CI and for the values for 3C from the literature (Bjorndahl *et al.*, 2003). Notwithstanding the missing data for 3C (Figure 2.8 (a)), the plots show a similar overall shape and have both highly positive and highly negative values. This indicates that the protein is folded in solution under the conditions of this study and that the solution structure of 3C is similar to that of 3CI, at least in the regions that have detectable peaks in 3C. The absolute values for the 3C are lower than those for the 3CI.

2.3.6 Inhibitor addition

It was hypothesised that the addition of a protease inhibitor would 'lock' the active site into one configuration and lead to the recovery of the missing resonances. The first inhibitor tried was the addition of a small amount of Complete™ protease inhibitor tablet from Roche. This had no effect on the ^{15}N,^{1}H-HSQC spectra of 3C^{pro}. The lab had been gifted a small quantity of specific 3C^{pro} inhibitors (S. Webber, Aguoron Pharmaceuticals now part of Pfizer). A small amount of each inhibitor ($\sim$1mg) was added to separate samples of 0.6 mM 3C. ^{15}N,^{1}H-HSQC experiments were run on each sample. The resulting spectra are shown as Figures 2.9 and 2.10.

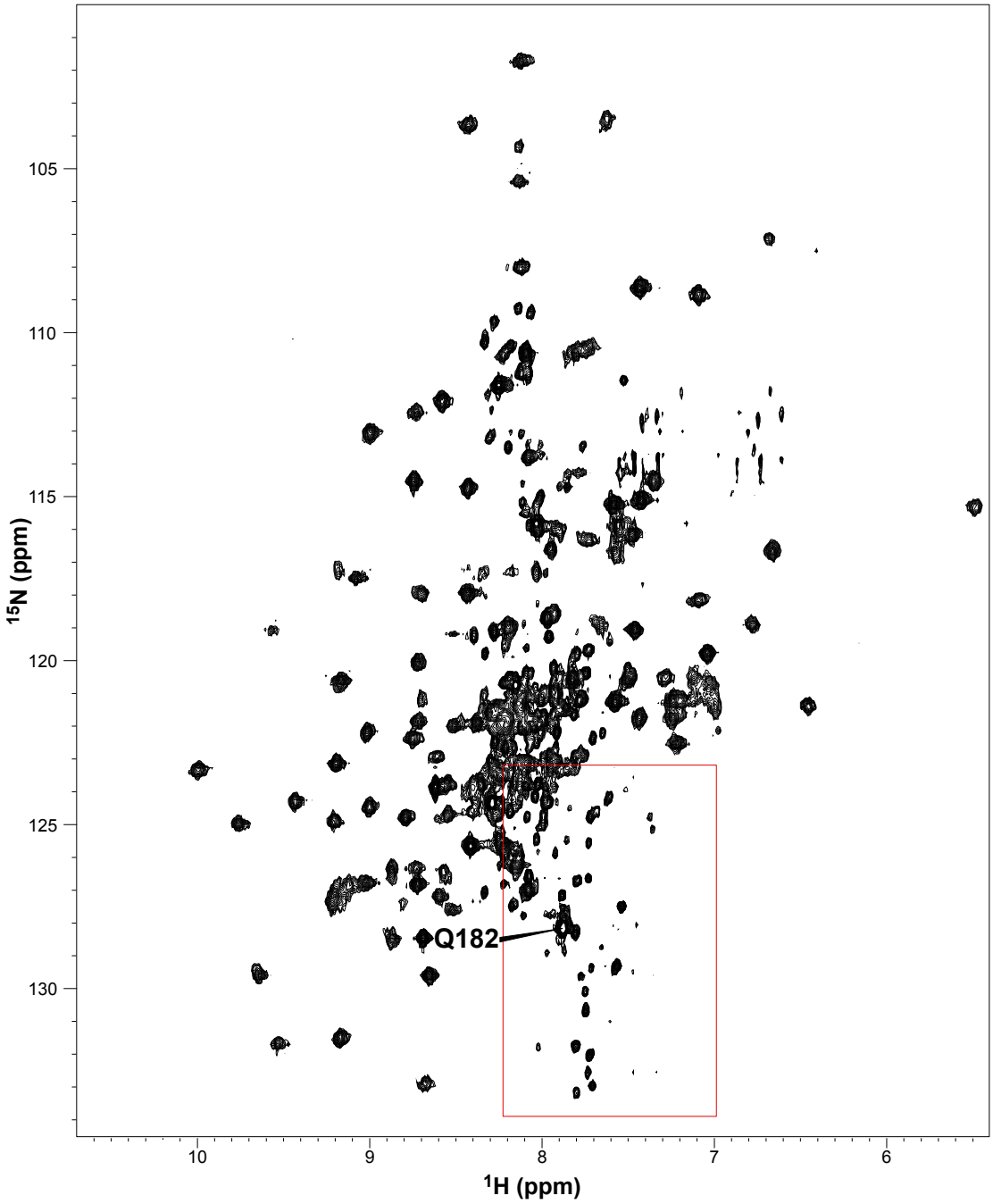


Figure 2.7: ^{15}N,^{1}H-HSQC-TROSY of apo 3C in L-Arg and 50 mM L-Glu buffer pH 6.0, 295 K. The red box marks the region where sharp peaks begin to appear over time when the sample is in this buffer system. The C-terminal residue, Q182, labelled.

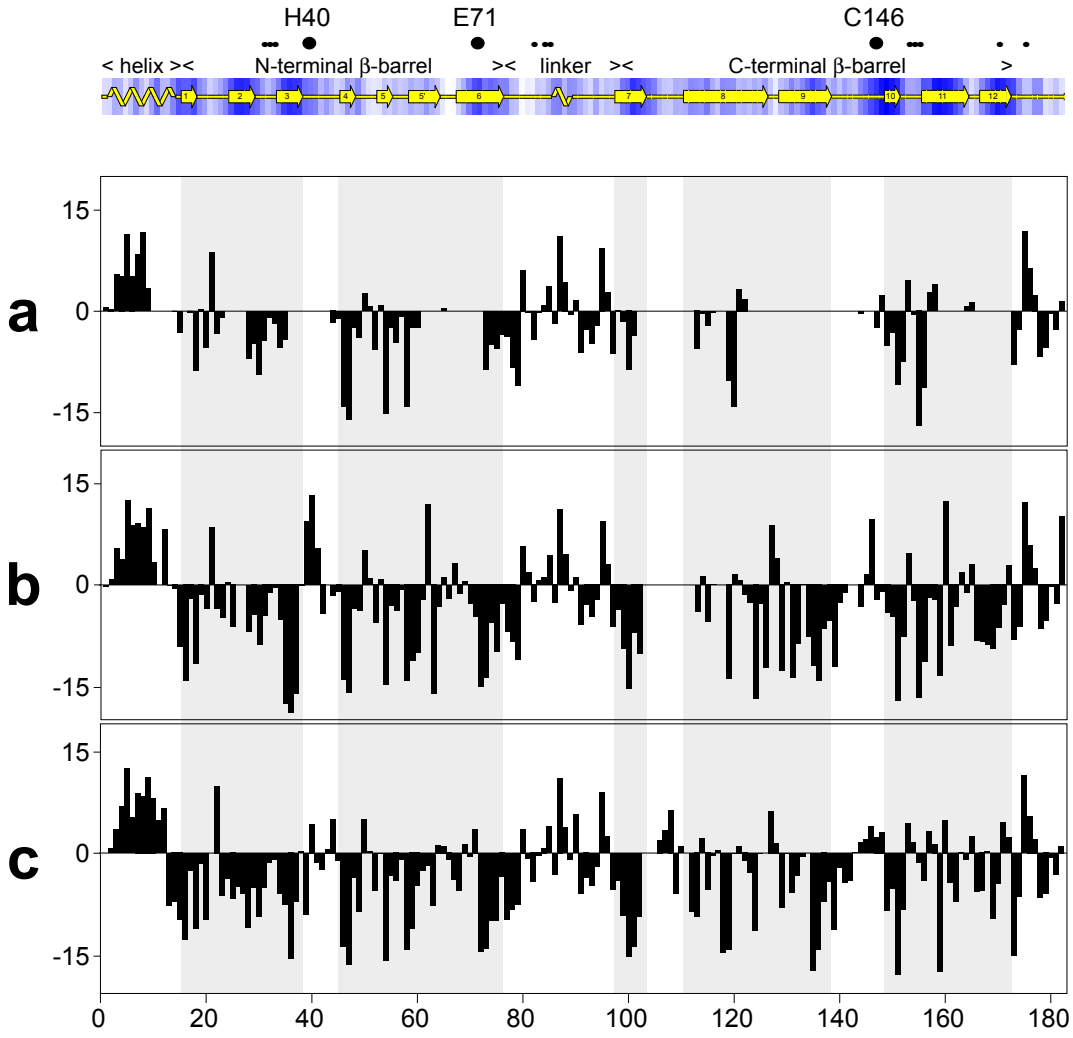


Figure 2.8: CSI plots for (a) 3C (b) 3CI and (c) the shifts deposited in the BMRB as #5659 (Bjorndahl *et al.*, 2003). A Procheck secondary structure prediction of 3CI is shown above (a). Negative values indicate β-sheet propensity and positive values indicate α-helix propensity. Random coil should appear near zero. The secondary structure of 3CI as predicted by PROCHECK is shown as a cartoon above the plots. The catalytic triad (His 40, Glu 71 and Cys 146) is indicated as large dots and smaller dots show residues which contact the inhibitor in 3CI.

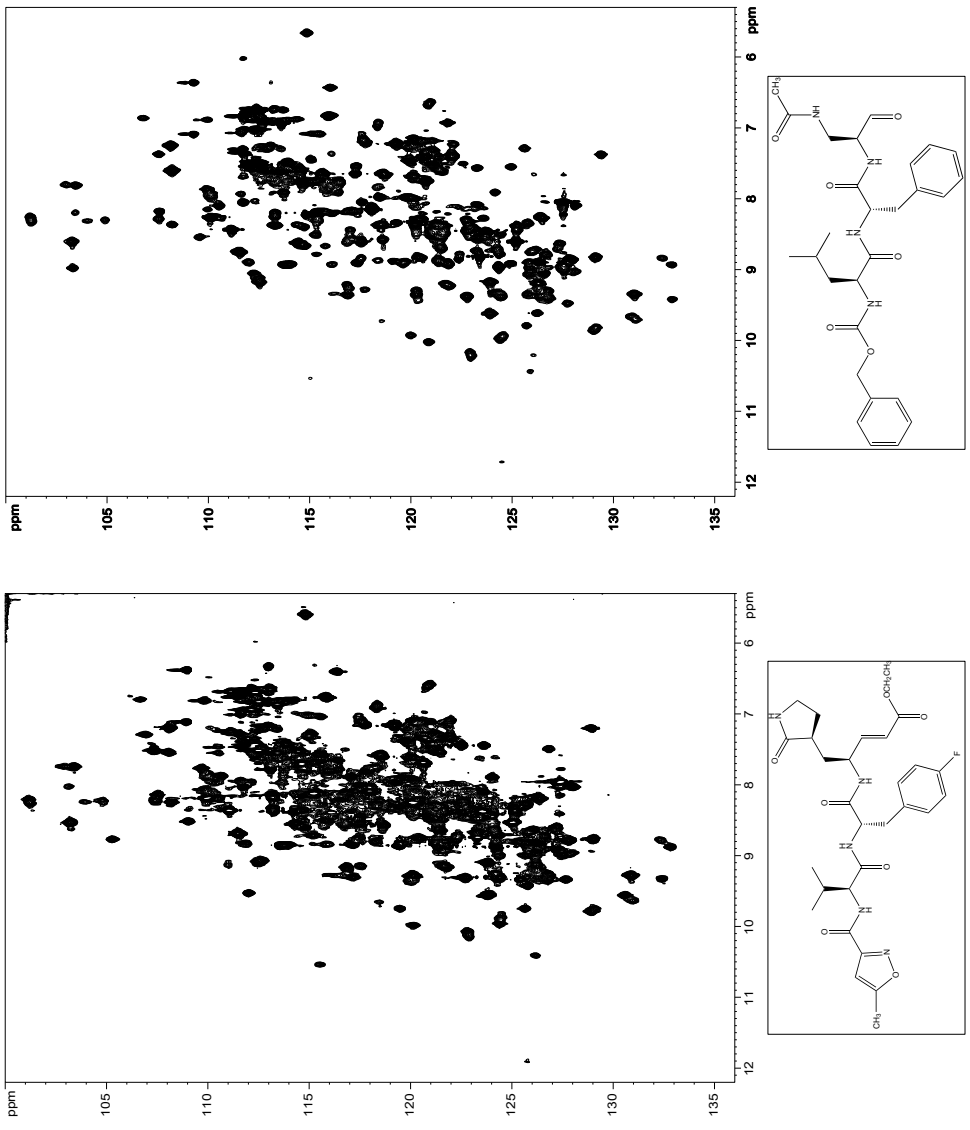


Figure 2.9: JMC 99-9 (Also known as AG7088) and 98-3 specific 3C inhibitors

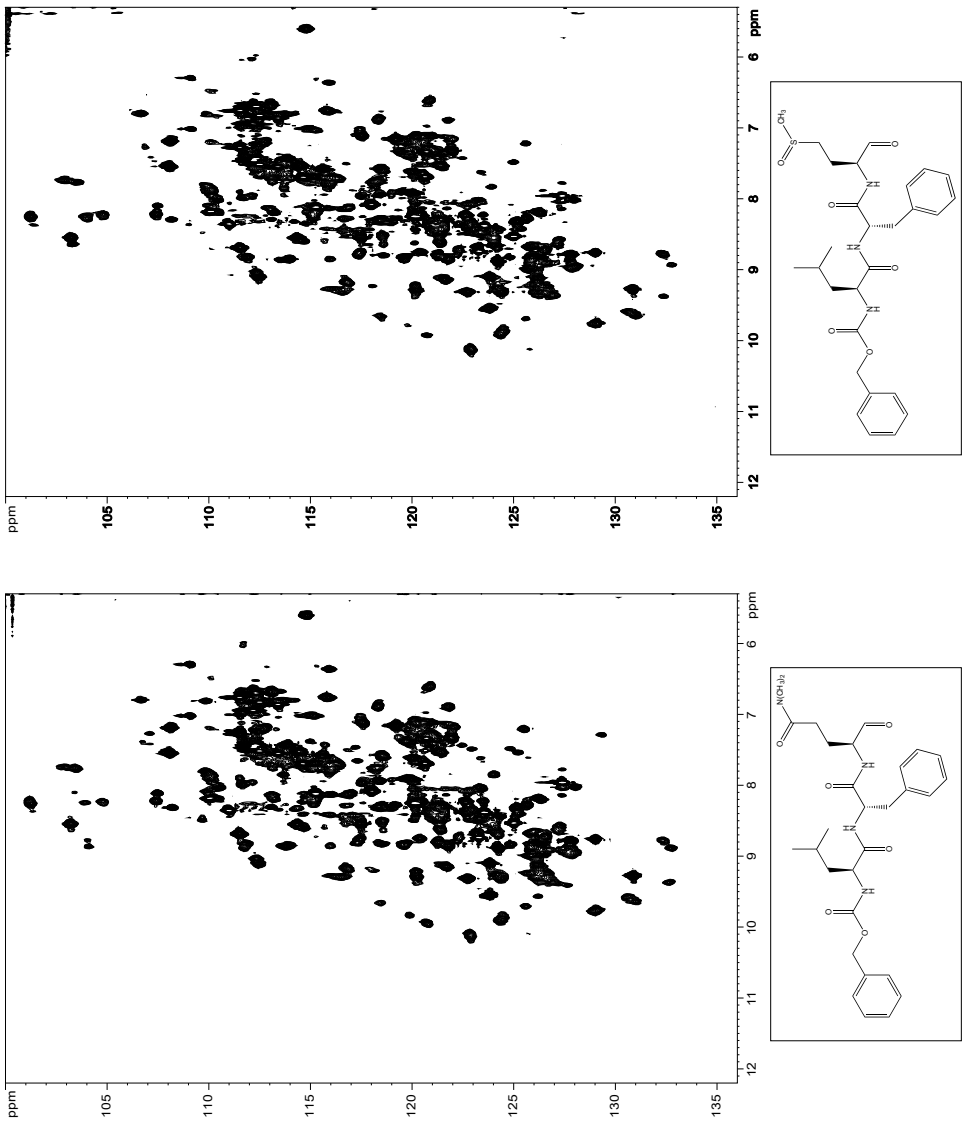


Figure 2.10: JMC 98-15 and 98-16 specific 3C inhibitors

Over time, extra peaks appeared in the spectra indicating a change in chemical shift for some amide moieties. This suggests that the inhibitors were poorly soluble in water but dissolved slowly. After an incubation period of 48 hours, all the spectra showed a difference in number of peaks compared to the spectrum of apo 3C. As the endpoint spectra were quite similar, the inhibitor JMC-98-3 (CBZ-L-Leu-L-Phe-L-(*N*-Ac-amino-alaninal)) was selected for further work as we had the largest quantity available. The K_i of this inhibitor is 6 nM (Webber *et al.*, 1998).

2.3.7 Assignment of inhibited 3C

^{15}N,^{1}H-HSQC spectra of JMC-98-3 inhibited protein (3CI) indicated that the majority of the amide resonances were observable. Consequently the protein was expressed in ^{15}N, ^{13}C labelled form and assigned in the inhibited state using previously mentioned experiments (Section 2.2.9) and using the observable 3C resonances as a guide (see section 2.3.2). After addition of the inhibitor, 90% of backbone shifts were assigned. This figure includes 90% of $^{1}H^{N}$ and 90% of ^{15}N 91% of C^{α} and 89% of CO. In addition to this 83% of $^{1}H^{\alpha}$ and 79% of C^{β} resonances were assigned. In total 53 amide peaks that were absent in the apo form of the protein were recovered in the amide form. 75% of these peaks were within 10 Å of the active site catalytic triad. However, although the majority of the recovered peaks were associated with residues close to the active site, this was not the case for all the recovered peaks. The inhibited protein gave better overall spectra and some isolated residues distal from the active site were also possible to assign only after addition of the inhibitor.

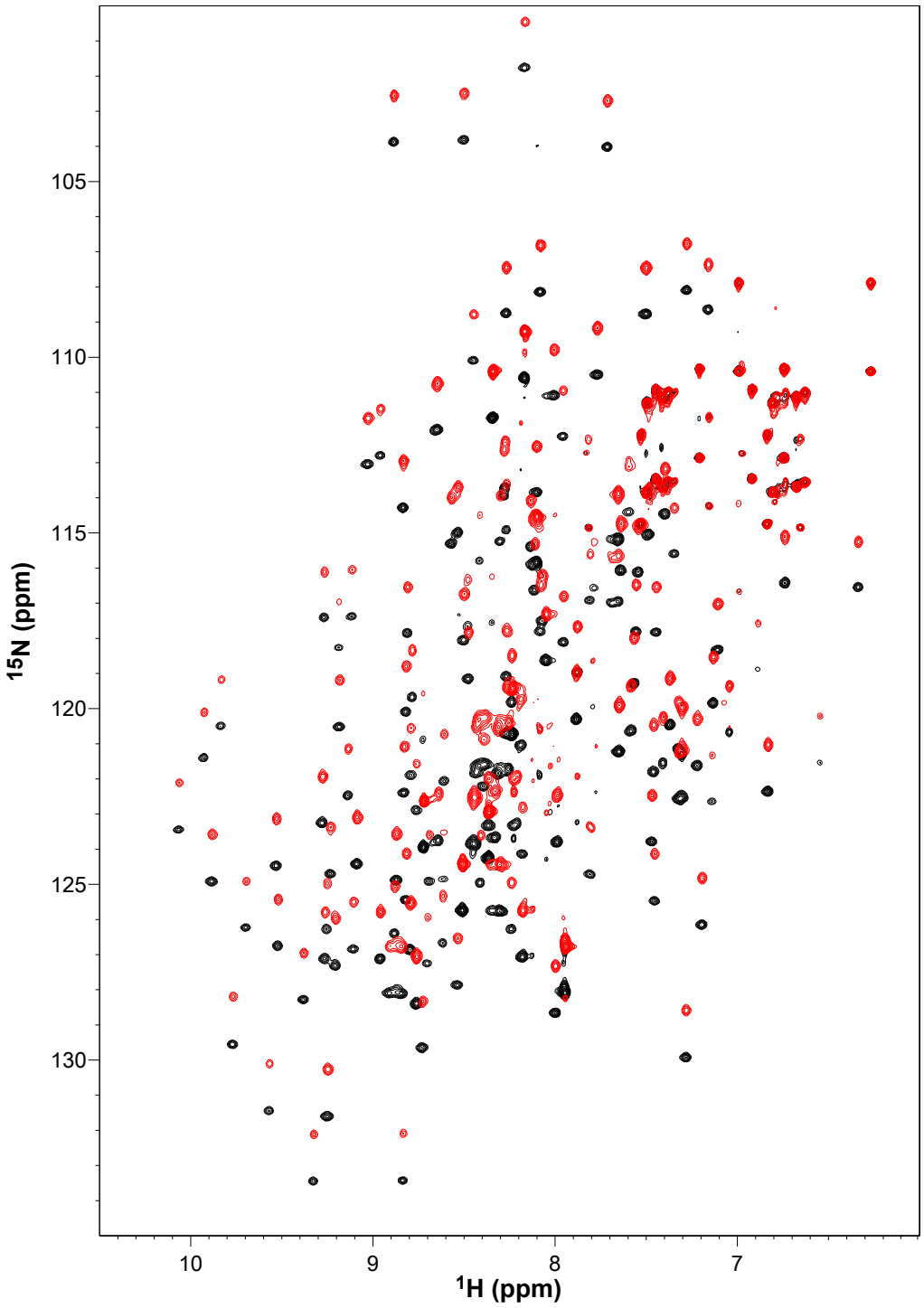


Figure 2.11: Overlay of typical IPAP spectra of 3CI in the absence of alignment media. The addition spectrum is shown in red and the subtraction spectrum in black. The sample is in 50 mM BTM buffer pH 6.0 14mM DTT 0.5 mM EDTA. 295 K

2.4 RDCs

A total of 123 RDCs were collected from the IPAP spectra. As can be seen from Figure 2.11, the IPAP spectra were of high quality. A plot of the distribution of the RDCs along the protein sequence is shown as Figure 2.12. The skew towards negative values is due to the preponderance of β-sheet in the structure and the consequent alignment of large amounts of the secondary structure along one axis.

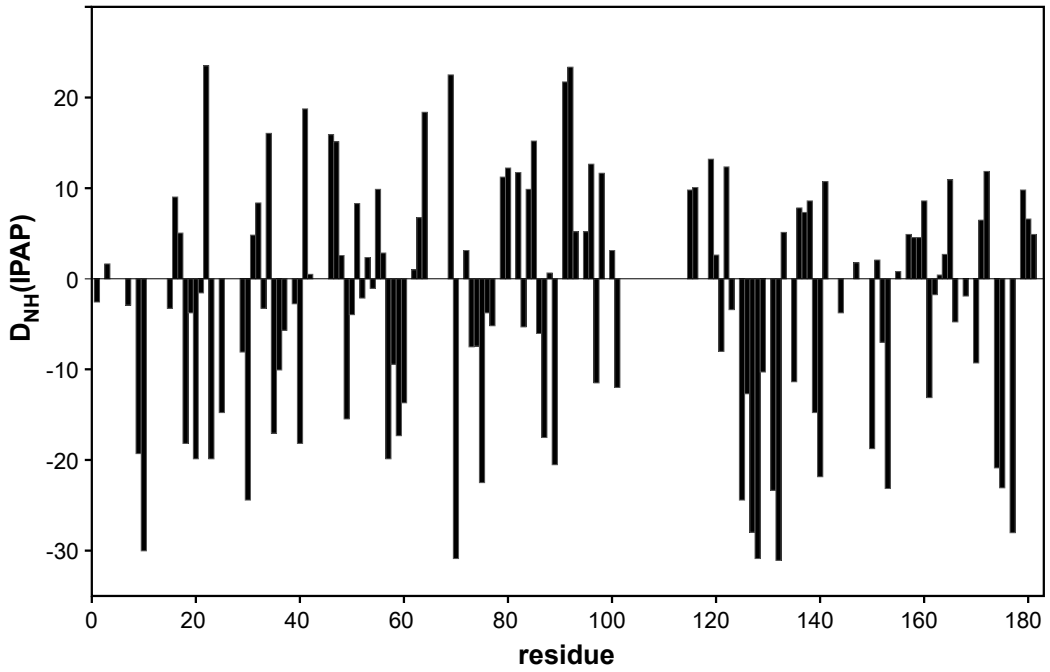


Figure 2.12: Distribution of IPAP-HSQC derived NH RDCs along the protein sequence for 3CI. The gap from 102-114 corresponds to the unassigned mobile loop.

2.5 Discussion

2.5.1 Sample production

Production of quantities of labelled 3C^{pro} of sufficient quality for NMR was not as straightforward as had been anticipated. Two major issues were encountered during the sample preparation process. Firstly, the stability of the sample was an issue. Initial preparations precipitated very rapidly during NMR data collection. This was likely to be partly due to unpaired cysteine residues in the protein (Cys 35, Cys 146 and Cys 151), as sample stability was increased by addition of higher concentrations of DTT. Changing the buffer from Phosphate at pH 6.5 or Tris-HCl at pH 7.5 to BTM at pH 6.0 also increased the stability of the sample. The was possibly due to the higher ionic strength (50 mM versus 20 mM) of the latter. The higher molarity was possible as the BTM is a low conductivity buffer and has less effect on probe matching than Tris. Increasing the ionic strength of a buffer can increase protein solubility. The lower pH may also have had some effect on the protein solubility. 3C has a pI of 7.9 and moving further from this would be expected to increase solubility.

Perhaps of even greater concern, the residues around the active site of the protein were extensively exchange-broadened. Ni^{2+} has been shown to bind non-specifically to amino acids with acidic side chains (Jensen *et al.*, 2004). Figure 2.13 shows the possible sites of interaction of the protein active site face with Ni^{2+}. As Ni^{2+}–NTA affinity chromatography was used during the purification of 3C^{pro} it was possible that Ni^{2+} ions had leached out of the column and were present in the final sample. However, flame spectroscopy indicated that, although some Ni^{2+} ions were present in samples of 3C^{pro}, the level of residual Ni^{2+} ions present was insufficient to cause paramagnetic broadening of the resonances. In order for the Ni^{2+} to cause broadening there would have to be a comparable molar amount to the 3C^{pro}. Overall, it seems unlikely that the sample preparation procedure was responsible for the increased motion around the active site, leading to peak broadening. The procedure followed was similar to that of Bjorndahl and co-workers (Bjorndahl *et al.*, 2007). Only the primary purification, using Ni^{2+}–NTA affinity chromatography differed substantially from the previous study. However, Ni^{2+} does not seem to be present in sufficient quantity in the sample to affect the spectra. It is possible that the 6-residue cloning artefact affects the overall dynamics of the protein but this is on the opposite face of the protein to the catalytic triad. Even assuming some transmission through the protein, it would seem likely that any affects would be strongest nearer to the tag.

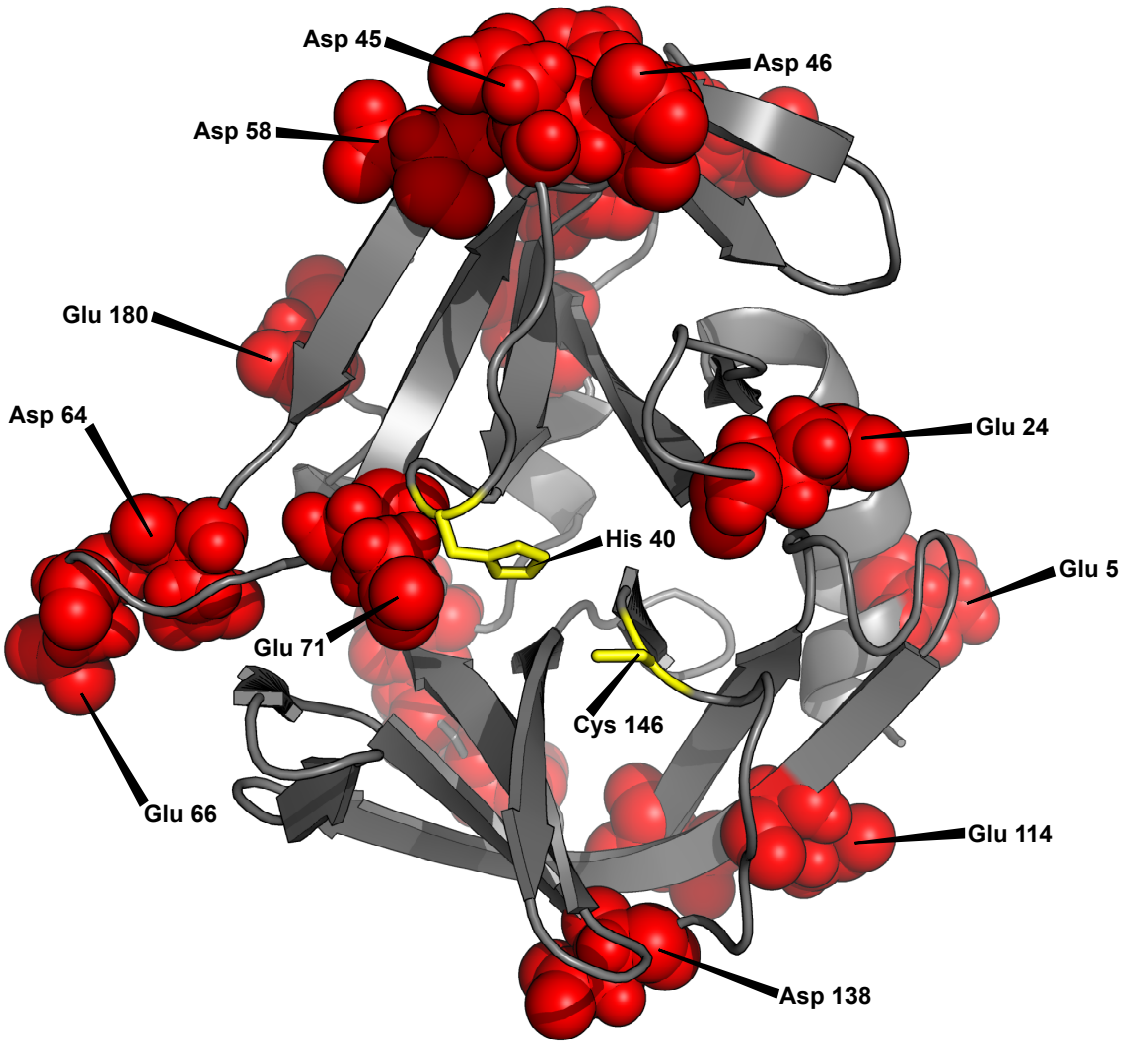


Figure 2.13: Active site of $3C^{pro}$ showing aspartate and glutamate residues (red spheres). The catalytic triad is shown in yellow with the exception of Glu 71 which is shown as red spheres.

The missing peaks were largely recovered by the addition of a specific 3C^{pro} inhibitor. Assuming that reduction in magnitude of active site dynamics is the reason for the appearance of the previously missing peaks in the inhibited form of the protein, there are two dynamic regimes that would account for this observation. Firstly, the protein conformation could exchange between two or more states on the intermediate time-scale. This would cause exchange broadening of the peaks in the spectrum. Alternatively, the protein may be in an ensemble of many different states such that population of no single state gives rise to sufficient signal for these residues. The second scenario would seem less likely as increasing the number of scans in the ^{15}N,^{1}H-HSQC does not reveal any additional peaks. It is possible that a mixture of both intermediate timescale dynamics and a plurality of structures is present. One mechanism for enzyme activity that has been proposed is that the protein samples many different states and a number of these approach the bound state. Upon binding to the substrate, the equilibrium is shifted across to the bound state (Weber, 1972). It seems likely that 3C samples more than one state in its apo form and that one of these states is similar to that of the inhibited form. As the inhibitor becomes covalently linked to the enzyme, it 'locks' the enzyme into one conformation. This changes the dynamic state of the residues with unobservable peaks and the missing resonances become visible. This is known as the "population shift mechanism" of binding. It is unclear why early samples of ^{15}N-labelled 3C gave spectra with the full number of peaks visible. It is possible that an endogenous *E. coli* peptide or other compound was bound to the protein in these samples. The dynamics of 3C^{pro} both in the apo and inhibited form are discussed further in Chapter 3. Evidence of continued intermediate timescale dynamics upon inhibitor binding is provided by the peaks that remain missing in the inhibited form of the enzyme. Residues with amide peaks still unobservable upon addition of the inhibitor are shown in Table 2.2 and Figure 2.14. These residues form a contiguous region near the active site with the exception of V118, G154 and L173. Of these, V118 is sequentially next to a proline residue, which meant that no connectivities could be established going forward through the protein sequence. For both G154 and L173 it was possible to identify carbonyl resonances and for G154 the C^{a} was identified, however, the NH resonances remained unidentified. This is possibly due to ring shift (Johnson and Bovey, 1958; Pauling, 1936) effects in the case of G154 (see Figure 2.15) and it is also possible that the chemical shift of L173 is affected by nearby aromatic residues.

Leu 11	Thr 109
Arg 12	Asn 110
Thr 26	Thr 111
Gly 27	Ile 112
Val 103	Val 118
His 104	Lys 142
Ser 105	Thr 143
Asn 106	Gln 145
Asn 107	Gly 154
Phe 108	Leu 173

Table 2.2: Residues with unobservable NH peaks in ^{15}N,^{1}H-HSQC of 3CI.

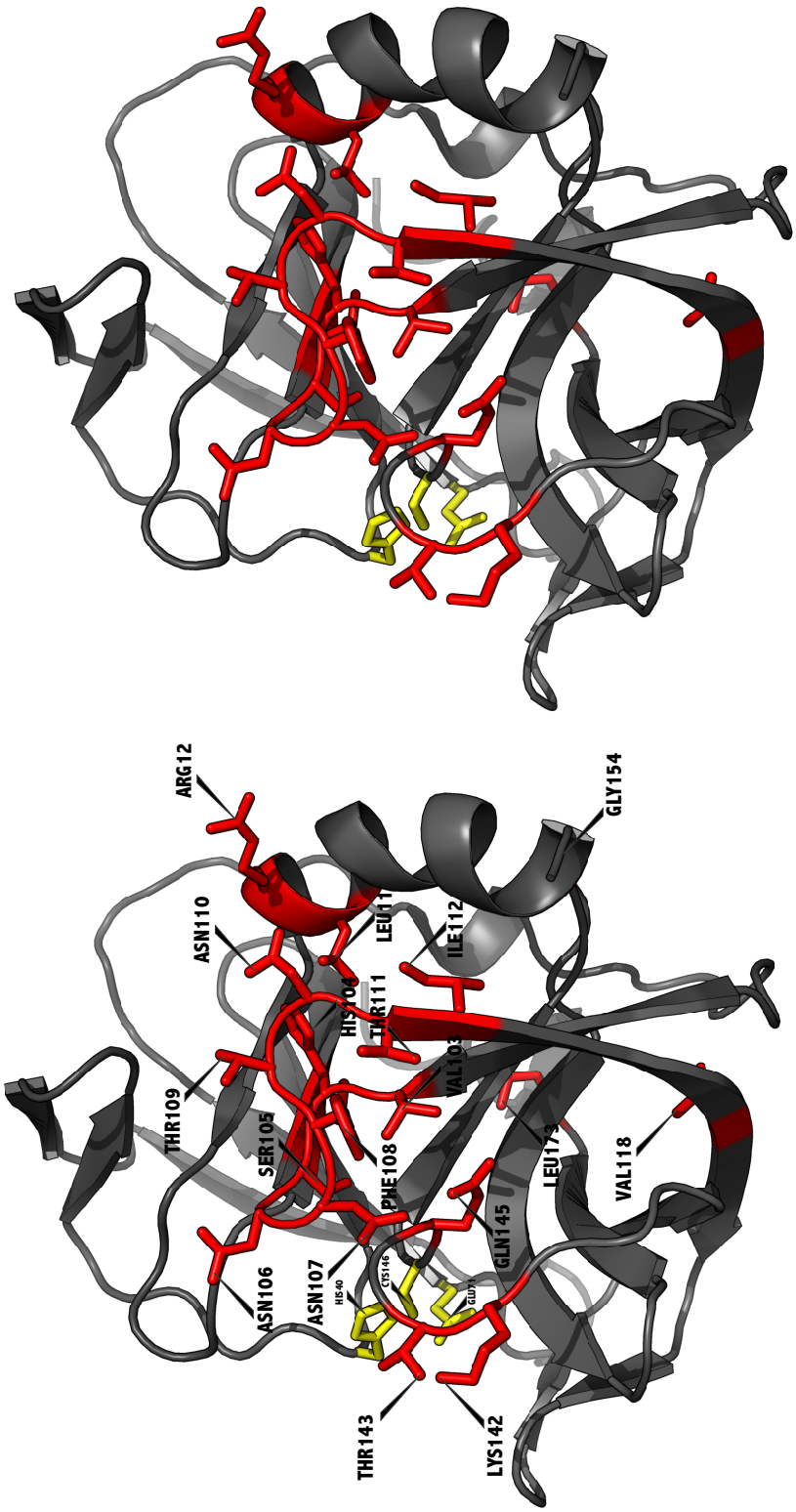


Figure 2.14: Stereo view of residues with missing NH peaks in 3CI. Those residues with unobservable NH peaks in 3CI are shown in stick form in red. The catalytic triad residues are shown in yellow. Residues Thr26 and Gly27 are in the centre of the view behind Ser105. Gly154 is obscured by the N-terminal helix.

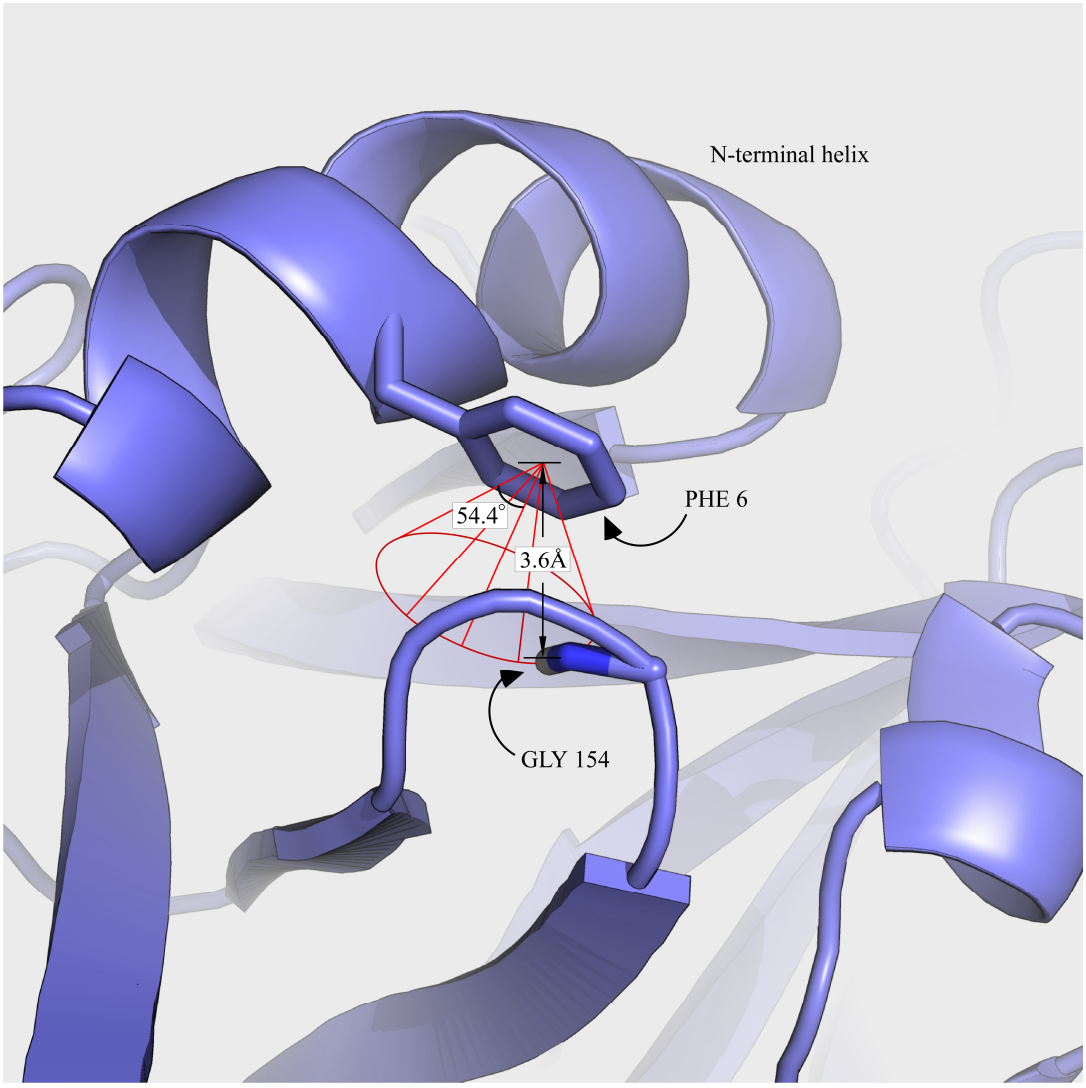


Figure 2.15: View of G154 showing the F6 ring that is hypothesised to cause the disappearance of the amide peak of G154 via ring-shift (Johnson and Bovey, 1958; Pauling, 1936). The cone of the ring-current effect is shown in red. The N-terminal helix is shown at the top of the figure.

2.5.2 Implications of structural model

RDCs are a useful addition to the restraints available for NMR protein structure calculation. The use of a homology model, followed by RDC refinement greatly speeded up the production of a structure suitable for further work. The homology model contained several residues that were in disallowed regions in Ramachandran space, after refinement with RDCs, these were found in allowed regions. The precision of the ensemble was also improved markedly.

The solution structures of HRV-14 $3C^{pro}$, both that of Bjorndahl *et al.*, and the structure resulting from this study, are very similar to that of the crystal structure published previously

	Allowed regions	Additionally allowed regions	Generously allowed regions	Disallowed regions	r.m.s.d. of ensemble
HM	68.8	28.8	1.2	1.9	0.65 ± 0.18
RDC	70.5	29.5	0.0	0.0	0.38 ± 0.04

Table 2.3: Table of structural statistics for homology model (HM) versus RDC refined homology model (RDC).

(Matthews *et al.*, 1994), although direct comparison is difficult as the co-ordinates of this structure were not deposited in the PDB. The CSI plots of 3CI and 3CI^B (Figure 2.8 (b) and (c) respectively) show very similar overall trends. This supports the notion of a similar structure of both forms of 3CI. In the comparison between 3CI and 3CI^B, RMSD between the two structures was highest in the linker, followed by the N-terminal helix, BB2, C-terminus and lowest in BB1. BB2 forms the majority of the interactions with the inhibitor. The two inhibitors are quite distinct and differences between the two structures in this region are to be expected. The N-terminus and the linker exhibit a large amount of exchange broadening.The RMSDs between 3CI and 3CI^B for the secondary structural elements are: N-terminal helix (1.41 Å), BB1 (1.06 Å), linker (2.18 Å), BB2 (1.20 Å) and C-terminus (1.18 Å). The structural differences are visualised in Figure 2.16. In Figure 2.16(a) the deviations in the N-terminal helix and linker are evident. The β-barrels are more similar in the 3CI (gold) and 3CI^B (silver) structures. A surface plot of 3CI^B and 3CI (Figure 2.16 (b) and (c) respectively) shows that the RNA-binding surfaces (facing out from the page) differ markedly. The larger RMSDs in the RNA binding region are interesting and perhaps suggest that the two structures reflect differences in the equilibrium of a dynamic ensemble of states. The similarities and differences in the dynamics of 3CI and 3CI^B are discussed further in Chapter 3.

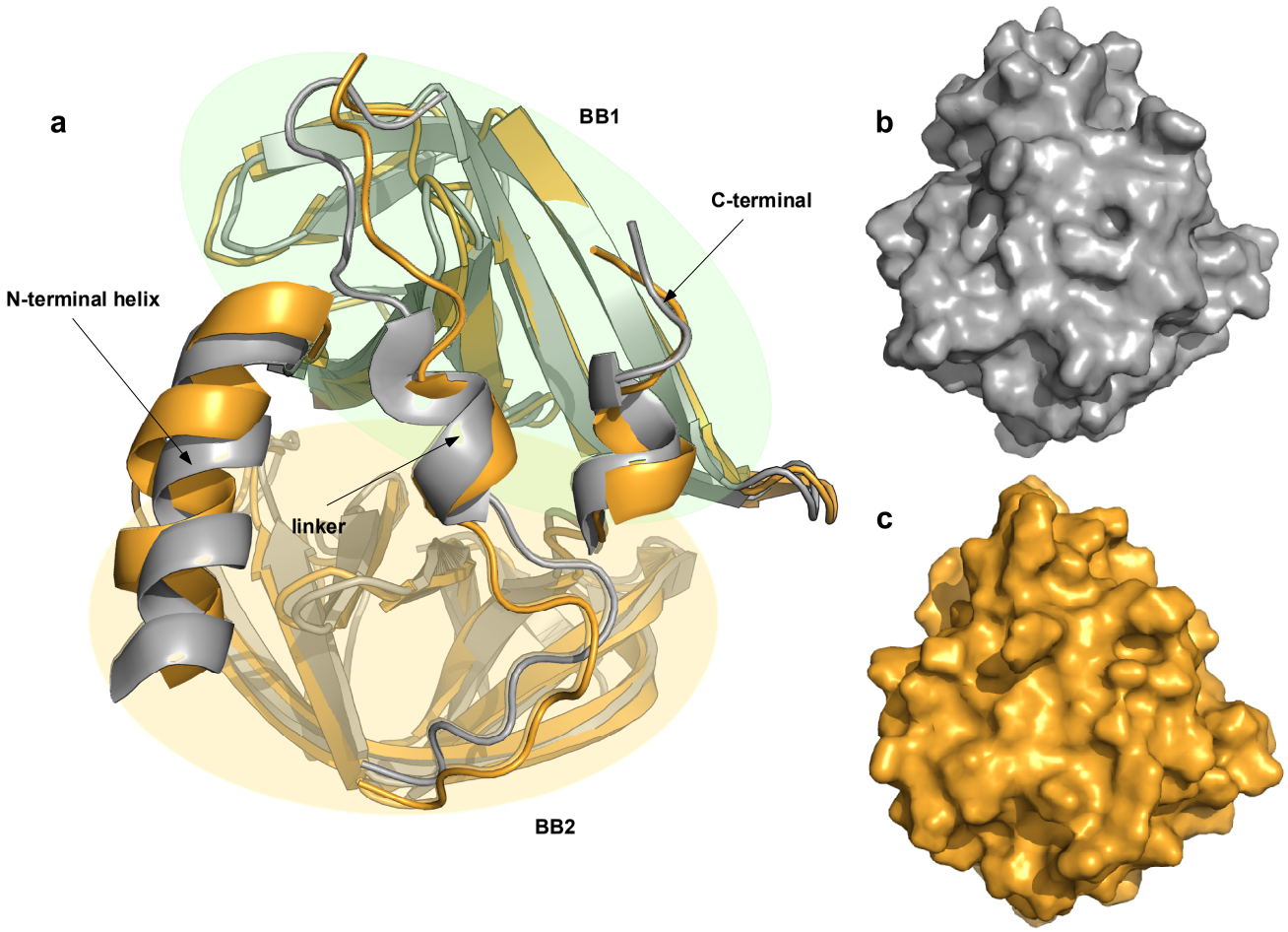


Figure 2.16: 3CI and 3CI^B structural comparison. Alignment of the two structures is shown in (a). 3CI is shown in gold and 3CI^B in silver. (b) and (c) show a surface rendering of 3CI^B and 3CI, respectively.

"Lasciate ogne speranza, voi ch'intrate"

Durante degli Alighieri

(c.1265–1321)

3

Dynamics of Inhibited and Uninhibited 3C

3.1 Introduction

The majority of this chapter has been submitted to the Journal of Molecular Biology as 'Inhibitor-Induced Dynamic Changes in a 3C Protease' *Jolyon K. Claridge, Patrick J. B. Edwards, Stephen J. Headey, David J. Lun, David S. Libich, David R.K. Harding, Andrew J. Dingley and Steven M. Pascal*

The 600 MHz spectra were run by Dr. Andrew Dingley at the University of Auckland. The DLS data were analysed by Dr. David Libich. Dr. Pat Edwards provided scripts for analysing the relaxation parameters for the model-free fits and ran some of the 700 MHz spectra.

The work outlined in this chapter was carried out in order to supplement available structural information on HRV-14 3C^{pro} (Chapter 2, (Bjorndahl *et al.*, 2007)) with dynamics information and to further characterise the biophysical properties of HRV-14 3C^{pro} in solution. As mentioned in Chapter 1 the structure of 3C^{pro} from several viral isotypes has been determined to high resolution (Birtley and Curry, 2005; Mosimann *et al.*, 1997; Matthews *et al.*, 1994; Bjorndahl *et al.*, 2007; Bergmann *et al.*, 1997; Matthews *et al.*, 1999; Sweeney *et al.*, 2007). A crystal structure of HRV 14 3C^{pro} has been determined (Matthews *et al.*, 1994) but the co-ordinates have not been deposited in the PDB. However, it is now accepted that flexibility and internal dynamics of proteins, in addition to tertiary structure, can provide important contributions to activity. Dynamic information on 3C^{pro} is currently available from chemical shift derived S^2 (RCI) data (Bjorndahl *et al.*, 2007) but not from ^{15}N relaxation analysis or other methods to specifically detect the μs–ms timescale motions often implicated in biological activities (McCoy *et al.*, 2001; Epstein *et al.*, 1995; Kristensen *et al.*, 2000; Dutta *et al.*, 2004; Palmer, 2004). Here I discuss the results of an ^{15}N relaxation and RCI analysis along with model-free and cross-correlation analyses of ms–μs dynamics of 3C^{pro} from human rhinovirus 14 (HRV-14) in the apo (uninhibited) state and in the presence of a specific active-site inhibitor. The results indicate that the apo protein is well ordered in regions but displays considerable intermediate time scale dynamics associated with the proteolytic active site and with the RNA-binding surface. Much of the dynamics near the proteolytic site are reduced upon addition of active-site inhibitor. I discuss the implications for the activity of 3C^{pro} and whether the two activities of 3C^{pro} are connected by an allosteric interaction.

3.2 Materials and Methods

Molecular visualisation was carried out with Pymol 1.1 (DeLano, 2008).

3.2.1 Sample preparation

Sample preparation was essentially the same as described in Chapter 2. For the apo 3C sample (3C), glycerol stock of H-MBP-3C^{pro} (Alexandrov *et al.*, 2001) was inoculated into 10 ml of LB$^{amp+/cam+}$ and grown overnight at 37 °C and then used to inoculate 1 l of LB$^{amp+/cam+}$ at 37 °C. Once OD_{600} reached 1.0, the cells were precipitated via centrifugation and re-suspended in 500 ml M9^{amp+} containing 3.5 g of glucose and 0.5 g $^{15}NH_4Cl$. The culture was left at 37 °C for 40 minutes to clear naturally abundant isotopes, then brought to 30 °C and protein expression was induced by addition of 0.4 mM IPTG. Cell growth was allowed to continue until the growth rate slowed (OD_{600} was 7.35). The cells were then harvested by centrifugation at 1541 $\times$ *g* for 15 minutes (Sorvall GS3 rotor). The cell pellets were frozen at -80 °C overnight, then re-suspended in 20 mM Tris-HCl and 150 mM NaCl at pH 7.5 (buffer A) containing 50 µL of protease inhibitor cocktail (Calbiochem) and lysed by passing three times through a French press at 6000 psi. The solution was centrifuged at 17,590 $\times$ *g* for 30 minutes at 4 °C (Sorvall SS34 rotor) and the supernatant was passed through an 80 µm filter and then through a 3 ml Ni^{2+}–NTA column pre-equilibrated with buffer A. The column was washed with 10 ml of buffer A and then the fusion protein was eluted with 11 ml of 250 mM imidazole buffer containing 1 M NaCl. Protein concentration in the eluted fraction was 12 mg ml^{-1} as determined by a Bradford assay, indicating approximately 130 mg in total. A total of 30 units of thrombin (Roche) were added to the eluate and the sample was dialysed for 24 hours into 1 l of buffer A containing 0.5 mM DTT. Subsequently the sample was dialysed for three hours into buffer A containing no DTT. After three hours the sample was passed back through the Ni^{2+}–NTA column and the flow-through was dialysed overnight into 20 mM Tris-HCl at pH 7.8 containing 0.5 mM EDTA and 1 mM DTT (buffer B). The overnight buffer was exchanged and dialysis continued for four hours at 4 °C. The sample was passed though a 3 ml QAE-sepharose column pre-equilibrated with buffer B and the flow through was spin concentrated and exchanged into 50 mM bis-Tris-propane/MES (BTP) buffer at pH 6.0 containing 14 mM DTT and 0.5 mM EDTA. Protein was concentrated to 1 mM and split into two samples of 525 µL each. An aliquot of 25 µL D_2O was added to each sample to give a final concentration of 5% v/v. Inhibited 3C (3CI) sample preparation was similar except

that, after the final concentration step, 525 μL of 1 mM protein was incubated overnight with 0.5 millimoles of JMC-98-3 inhibitor (Matthews *et al.*, 1999) to ensure a higher than 1:1 inhibitor to protein ratio.

3.2.2 NMR-based dynamic analysis

All NMR experiments were performed at 22 °C. A set of ^{15}N T_1, $T_2$and ^{1}H,^{15}N-NOE measurements Kroenke *et al.* (1998) were recorded at 700 MHz on a Bruker Avance spectrometer equipped with pulsed field gradients and a triple resonance cryoprobe. Additionally, T_2 data were acquired at 600 MHz on a similarly equipped Bruker Avance spectrometer. Spectra were processed with Topspin 2.1 (Bruker) and analysed using Analysis 1.15 (Vranken *et al.*, 2005). ^{15}N T_1spectra were acquired with 32 scans and with 1024 complex points in the ^{1}H dimension and 90 complex points in the ^{15}N dimension with relaxation intervals of 51.1, 81.1, 151.1, 251.1, 401.1, 801.1, 1401.1 and 2001.1 ms with duplicate data at 51.1 and 151.1 ms and spectral widths of 11,161 Hz (^{1}H) and 1,828 Hz (^{15}N). T_2 spectra were acquired with 32 scans and with 1024 complex points in the ^{1}H dimension and 90 complex points in the ^{15}N dimension with relaxation intervals of: 16.3, 32.6, 49.0, 81.6, 97.9, 114.2, 130.6 and 163.2 ms with duplicate data at 32.6 and 97.9 ms and a spectral width of 11,161 Hz (^{1}H) and 1,828 Hz (^{15}N). Recycle delays were 1.5 s. Spectral widths for the T_2spectra acquired at 600 MHz were reduced to 9,566 Hz (^{1}H) and 1,587 Hz (^{15}N). $T_1$and T_2 data were collected in an interleaved manner as pseudo 3D experiments. NOE spectra were acquired with 128 scans, 1024 complex points in the ^{1}H dimension and 90 complex points in the ^{15}N dimension, with a six second presaturation period (Farrow *et al.*, 1994). The error in the NOE value (σNOE) was determined via the equation:

$$\sigma_{NOE}=\frac{I_{sat}}{I_{unsat}}\left(\left(\frac{\sigma_{sat}}{I_{sat}}\right)^2+\left(\frac{\sigma_{unsat}}{I_{unsat}}\right)^2\right)^{\frac{1}{2}} \quad (3.1)$$

where, I_{sat} and I_{unsat} represent the measured intensities in the presence and absence of proton presaturation, and σ_{sat} and σ_{unsat} represent the root-mean-square variation in the noise in empty spectral regions of the spectra with and without proton presaturation. Model-free analysis was performed using the program model-free 4.1 (Mandel *et al.*, 1995) and FAST-model-free (Cole and Loria, 2007). In this formalism, the amplitude of fast internal motion (ns-ps time scale) is described by an order parameter (S^2) and the time scale of the motion is described by an internal correlation time (τ_e). The resulting spectral density function is:

$$J_{(\omega)} = \frac{S^2\tau_c}{(1+\omega^2\tau_c^2)} + \frac{(1-S^2)\tau}{(1-\omega^2\tau^2)} \quad (3.2)$$

where τ_c is the global rotational correlation time and the influence of τ_e is included in $\tau = \tau_c \tau_e / (\tau_c + \tau_e)$. Simultaneous internal motion on two time scales differing by at least an order of magnitude can be represented by the following spectral density function (Clore *et al.*, 1990) :

$$J_{(\omega)} = \frac{S^2\tau_c}{(1+\omega^2\tau_c^2)} + \frac{(1-S_f^2)\tau'_f}{(1+\omega^2\tau'_f)} + \frac{(S_f^2-S^2)\tau'_s}{(1+\omega^2\tau'_s)} \quad (3.3)$$

where $S_f^2$and S_s^2represent order parameters for the faster and slower motions, τ_f and τ_s represent the time scales of these two motions, and $S^2 = S_f^2\ S_s^2$. The following formulae relate the relaxation parameters obtained from the NMR experiments to the spectral density function:

$$R_1 = (\frac{d^2}{4})[J(\omega_H-\omega_N) + 3J(\omega_N) + 6J(\omega_H+\omega_N)] + c^2J(\omega_N) \quad (3.4)$$

$$R_2 = (\frac{d^2}{8})[4J(0) + J(\omega_H-\omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H+\omega_N)] + \frac{c^2}{6}[4J(0)+3J(\omega_N)] \quad (3.5)$$

$$NOE = 1 + (\frac{d^2}{4R_1})(\frac{\gamma_N}{\gamma_H})\,[6J(\omega_H+\omega_N) - J(\omega_H-\omega_N)] \quad (3.6)$$

$$d = \frac{\mu_\circ h\gamma_N\gamma_H\left\langle r_{NH}^{-3}\right\rangle}{(8\pi^2)} \quad (3.7)$$

$$c = \frac{\omega_N\Delta\sigma}{\sqrt{3}} \quad (3.8)$$

where μ° is the permeability of free space, h is Plank's constant, γ_H and $\gamma_N$are the gyromagnetic ratios for ^{1} H and ^{15}N, r_{NH} is the N-H bond length (1.02Å), ω_H and ω_N are the Larmor frequencies of ^{1}H and ^{15}N and d is the dipolar constant and $\Delta\sigma$ is the chemical shift anisotropy of ^{15}N (-170 ppm) assuming an axially symmetric chemical shift tensor (Tjandra *et al.*, 1995; Fushman and Cowburn, 1999). After extracting a global correlation time, model-free analysis then attempts to derive a spectral density function for a particular residue that is consistent with the relaxation parameters obtained for that residue. The order parameter and time scales can then be extracted from the spectral density function. Five models of varying complexity are used by

model-free. These models include not only the above possibilities of one or two relatively fast internal motions, but also the possibility of a slower motion (ms-µs) time scale with a contribution to relaxation denoted as R_{ex} . The most easily interpretable results of this analysis are the order parameter (S^2), the R_{ex} value and the complexity of the model required for fitting. Low S^2 indicates a high degree of fast dynamics, high R_{ex} suggests significant slow dynamics and the requirement of a complex model for fitting suggests the presence of more than one type of dynamics at that residue. The FAST-Modelfree model selection process starts with estimates of the rotational diffusion parameters and then fits all the residues using model 1 (S^2) . The sum squared error for each spin Γ_i, is then calculated and a user assigned value (in this case 0.95) used to determine which residues are assigned to model 1. Further model selection is made by comparison of Γ_i and *F*-test statistics. As, in this case, data were available at more than one field strength FAST-modelfree can determine whether models 4 or 5, containing four parameters, are suitable. inability to find a model can also suggest the presence of complex motion. Direct detection of chemical exchange was also performed at 700 MHz using the methodology of Kroenke et al. (Kroenke *et al.*, 1998). The chemical exchange rate (R_{ex}) can be derived from the following equation:

$$R_{ex} = R_2 - R_2^0 \tag{3.9}$$

where R_2 is the experimental rate determined from the ^{15}N T_2 experiment described above and R_2^0 is the transverse relaxation rate in the absence of R_{ex}. R_2^0 was determined for each residue by measuring the cross-correlation rate (η_{xy}) between ^{15}N CSA and ^{15}N–^{1}H dipolar relaxation (Fushman *et al.*, 1998) and then using:

$$R_2^0 = \kappa\eta_{xy} \tag{3.10}$$

(the ratio κ is is independent of chemical exchange and local motions). Cross-correlation was measured using symmetrical reconversion (Pelupessy *et al.*, 2003). The ^{15}N CSA value was assumed to be constant (Wang *et al.*, 2001) and κ was determined to be 1.43 at 700 MHz by comparison of the R_2^0 and η_{xy} values of those residues not involved in exchange processes.

$R_{1\rho}$ experiments were carried out as described in (Dutta *et al.*, 2004). $R_{1\rho}$ values were acquired using relaxation delays of 10, 20, 30, 50, 70, 100, 150 and 200 ms. The spin-lock field of 1519 Hz

was used. The $R_{1\rho}$ values were

$$R_{1\rho} = R_1 cos^2\theta + (R_2 + R_{ex}) sin^2\theta \quad (3.11)$$

corrected for off-resonance effects using:

$$R_{1\rho} = \frac{R_1\rho^{app} - R_1 cos^2\beta}{sin^2\beta} \quad (3.12)$$

$$\beta = \frac{\arctan(\omega_{spinlock})}{\Delta\omega} \quad (3.13)$$

$\delta\omega$ = offset of each resonance from carrier frequency.

RCI analysis was performed as described (Berjanskii and Wishart, 2005; Berjanskii *et al.*, 2007) using the RCI webserver.

3.2.3 Circular Dichrosim (CD)

CD spectroscopy was used to assess the proportion of stable secondary structure and the thermal stability of 3C and 3CI. Spectra were recorded on a Chirascan CD spectropolarimeter (Applied Photophysics, Leatherhead, United Kingdom) equipped with a recirculating water bath. Spectra were recorded in 1 nm steps from 260 to 190 nm with an integration time of 1 s at each wavelength. Three successive scans were recorded, the sample blank subtracted, the scans averaged and smoothed using a sliding window function. Thermal stability was determined by acquiring CD spectra as a function of temperature at 5 °C intervals from 5 °C to 85 °C with two minutes equilibration time at each temperature point. Approximate melting curves were obtained by monitoring the ellipticity at 222 nm (ϑ222) as a function of temperature. Melting temperatures (T_m) were estimated by evaluating the maximum of dϑ222/dT. The Selcon 3 algorithm as part of the DichroWeb server (Whitmore and Wallace, 2004, 2008) was used for spectral deconvolution. The 3C and 3CI concentrations were 0.28 and 0.31 mg ml^{-1}, respectively. Buffer conditions were 20 mM phosphate pH 6.5 0.5 mM EDTA and 1 mM DTT. Lower DTT concentration than in NMR samples was necessary to reduce background absorption. ^{15}N,^{1}H-HSQC spectra recorded under these conditions Figure 3.1 are similar to those recorded under the sample conditions detailed above for the NMR dynamics experiments.

3.2.4 Dynamic Light Scattering (DLS)

The apparent Stokes radii (R_S) of 3C and 3CI were analysed using a Zetasizer Nano ZS (Malvern Instruments, Malvern, United Kingdom). DLS data were obtained at 25 °C using a low-volume disposable 1 cm path length plastic cuvette (Sarstedt, Nürnbrecht, Germany). Five successive scans were collected and averaged for each protein sample. Samples were prepared a day in advance and allowed to sit overnight at 4 °C to permit bubbles to dissipate and then equilibrated to 25 °C before measurements. The diffusion coefficients were extracted from the correlation curve and the hydrodynamic radius was calculated using the Stokes-Einstein equation. Experiments were carried out on 0.3 mg ml^{-1} samples of both 3C and 3CI. Buffer conditions were as in the CD spectroscopy detailed above except 14 mM DTT was present as in the relaxation experiments.

3.3 Results

3.3.1 HSQC Spectra of 3C and 3CI

^{15}N,^{1}H-HSQC spectra of 3C display a degree of dispersion in each dimension consistent with folded protein. Upon assignment of the backbone residues, however, several gaps were present between assigned stretches of residues. Figure 3.1 shows an ^{15}N,^{1}H-HSQC overlay of 3C (black) and 3CI (red). Several residues are only present in the 3CI spectrum. The residues with unobservable amide resonances in 3C were plotted onto a PDB of the protein (modelled as described previously (Chapter 4) (Claridge *et al.*, 2009)). The affected residues cluster near the active site of the protein (Figure 3.2(a), gold surface). In addition, resonances from residues sequentially neighbouring those that are unobservable are reduced in intensity relative to those associated with other assigned residues. The unobservable resonances in 3CI cluster mainly to a region to the right of the active site in Figure 3.2(b), consisting of the loop from V103 to T111 and the spatially neighbouring residues K142 and T143.

3.3.2 CD spectra of 3C and 3CI

Deconvolution of CD spectra of 3C and 3CI (Figure 3.3(a)) shows similar α-helical (7 % and 5% respectively) and β-strand (55 % and 57 % respectively) content. A temperature series of the two forms of the protein reveals that the 3CI is slightly more thermo-stable than 3C with the transition

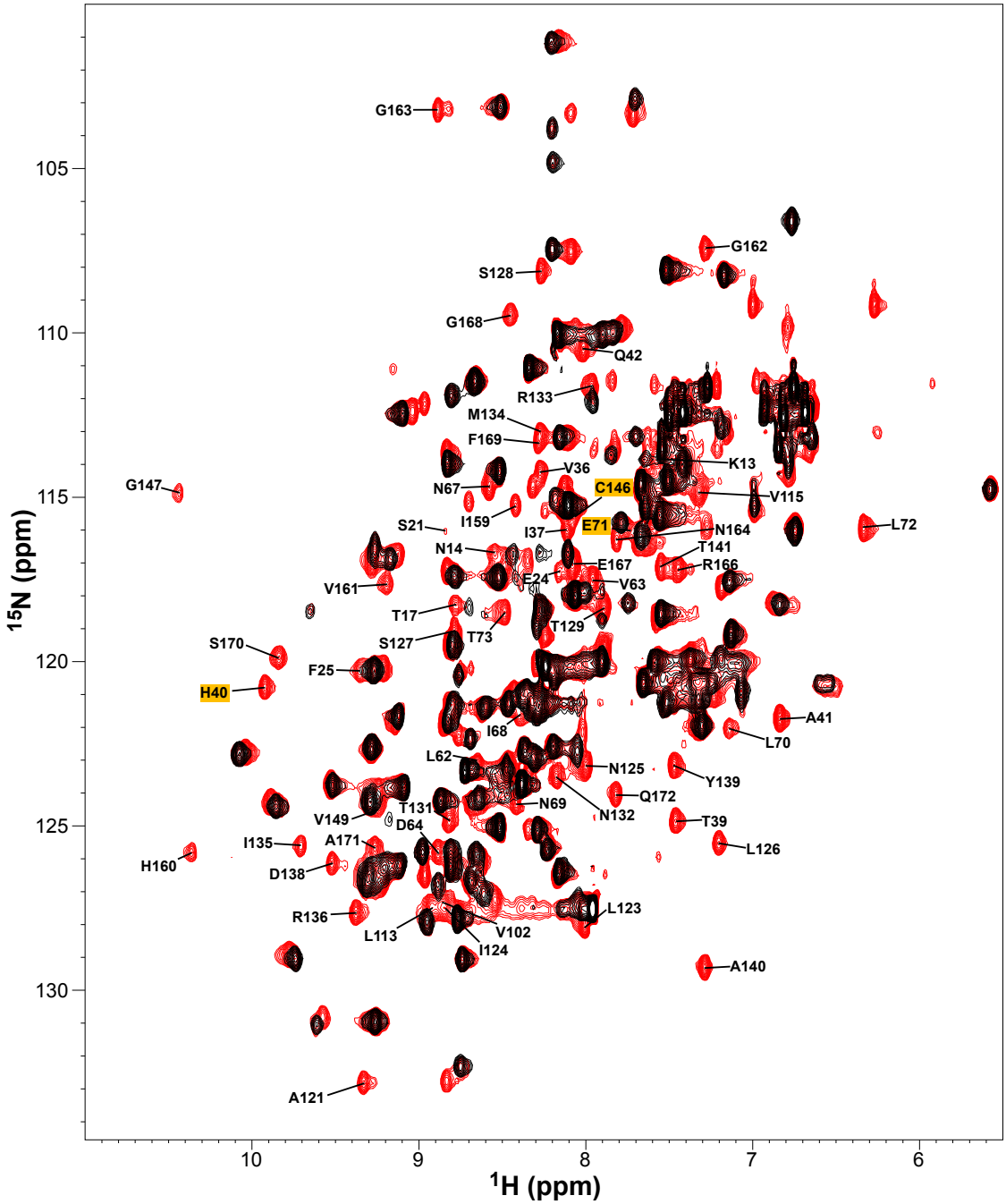


Figure 3.1: Overlay of ^{15}N,^{1}H-HSQC spectra of 3C (black) and 3CI (red) Spectra were recorded at 295 K and in 20 mM phosphate buffer pH 6.5 containing 14 mM DTT and 0.5 mM EDTA.

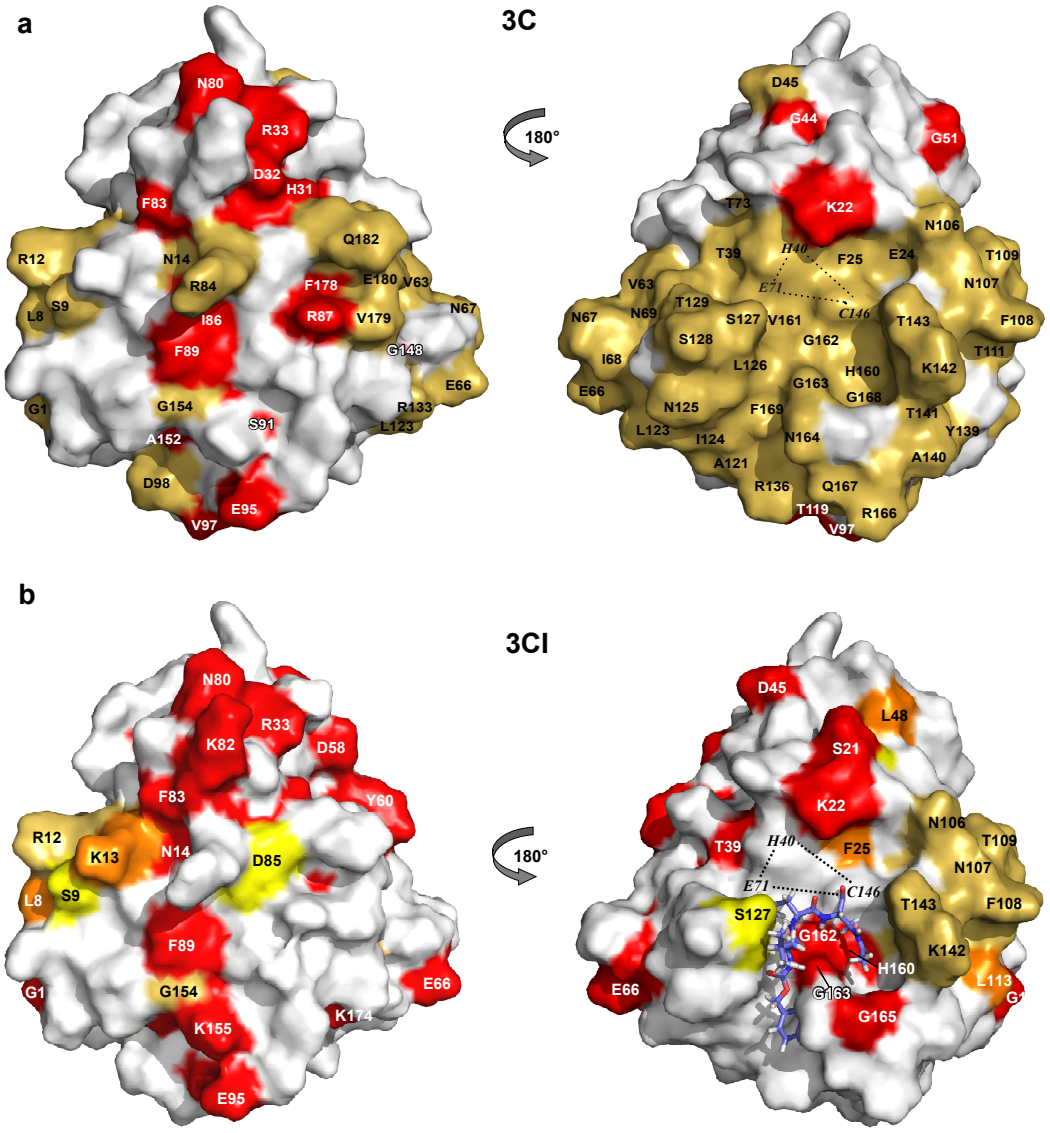


Figure 3.2: Slow to intermediate exchange for (a) 3C and (b) 3CI plotted onto surface representations of the 3CI protease modelled as described in Chapter 2. In each panel, the RNA-binding face is shown in the left hand view and the catalytic face is shown in the right hand view with the catalytic triad residue positions connected via dotted lines. Residues with R_{ex}^{CC} values greater than 1.5 s^{-1} are painted red and residues exchange broadened in 2D ^{15}N,^{1}H-HSQC spectra are painted gold. In panel (b), residues that could not be fit to any model in Lipari-Szabo analysis are also painted gold, residues with R_{ex}^{LS} values greater than 1.5 s^{-1} are yellow and residues with both R_{ex}^{CC} and R_{ex}^{LS} values greater than 1.5 s^{-1} are orange.

from folded to disordered occurring approximately 51 °C in 3C and 56 °C in 3CI (Figure 3.3(b)).

3.3.3 DLS and SDS-PAGE analysis of 3C and 3CI

Dynamic light scattering (DLS) data show that the two forms of the protein have hydrodynamic radii ((R_S)) consistent with a monomeric state (Figure 3.3(c)) with no evidence of significant dimerisation or further self-association. SDS-PAGE analysis reveals indistinguishable apparent molecular weights for 3C and 3CI, under both reducing and non-reducing conditions (data not shown), indicating that there is unlikely to be formation of intermolecular disulphide bonds under the sample conditions used. Further evidence for a monomeric state is provided by fitted rotational correlation times (see next section)

3.4 Dynamics Analysis

Apo 3C^{pro} (3C)

Relaxation data were analyzed (Figure 3.5(a)) for all residues of uninhibited 3C^{pro} with visible and resolved peaks in the 2D ^{15}N,^{1}H-HSQC. The mean and standard deviation of the R_1 values is 0.90 ± 0.10 s^{-1} (10% trimmed mean 0.88 s^{-1}). The mean and standard deviations of the R_2 rates at 600 and 700 MHz are 17.7 ± 4.3 s^{-1} (10% trimmed mean 17.7 s^{-1}) and 17.3 ± 4.9 s^{-1} (10% trimmed mean 17.2 s^{-1}) respectively. The mean NOE is 0.65 ± 0.35 (10% trimmed mean 0.68). The mean R_2/R_1 ratio is 20.1 ± 5.1 (10% trimmed mean 20.3). Model-free analysis of these data gave a global correlation time of 12.27 ns. This is consistent with the protein being monomeric in solution. The order parameter, S^2, provides an indication of the fast (ns-ps) time scale dynamics. An order parameter approaching 1.0 indicates rigidity while progressively lower S^2 indicates progressively increasing internal motion. A significant R_{ex} value suggests the presence of dynamics on a slow (ms-μs) time scale. The molecular shape of 3C and 3CI (modelled as described previously, Chapter 2) is appropriate for the assumption of isotropic tumbling. Model-free analysis of 3C relaxation produces unrealistically high S^2 values. Some S^2 values approach 1.0, indicating total rigidity, which is suggestive of fitting difficulties. This may be related to

As the S^2 values from the model-free fit were unreliable, the alternative RCI method (Berjanskii and Wishart, 2005; Berjanskii *et al.*, 2007) was used to calculate the order parameters for 3C. RCI analysis uses the chemical shifts of backbone resonances to detect backbone flexibility in the form of a chemical shift-derived S^2 value. Advantages of the RCI method are that overlapping peaks, excluded from model-free analysis, can be used, and that proteins displaying complex

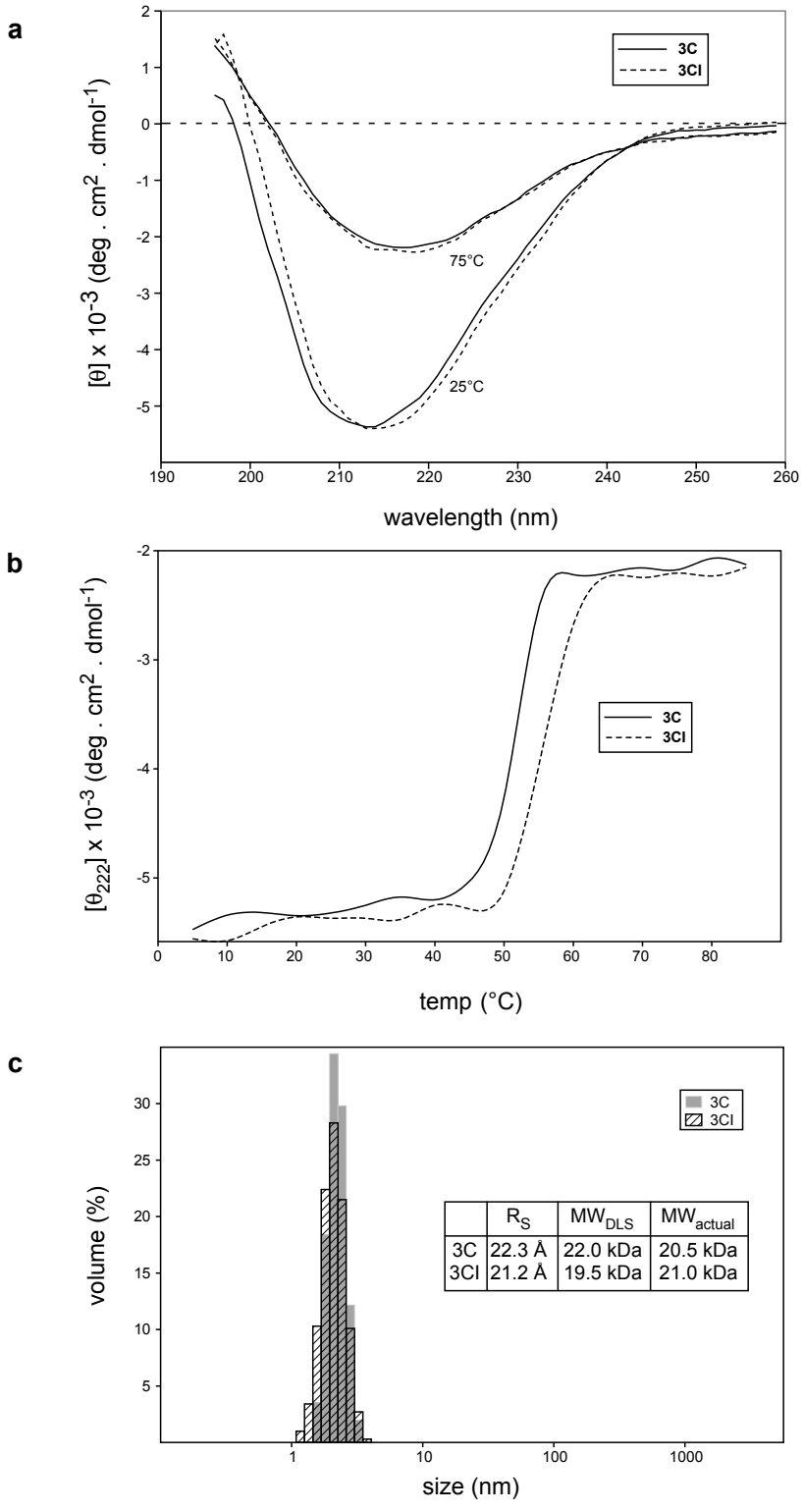


Figure 3.3: CD and DLS for 3C and 3CI. (a) CD spectra of 3C (solid lines) and 3CI (dashed lines) at 25 °C and 75 °C. (b) θ222 vs. temperature for 3C (solid line) and 3CI (dashed line). (c) DLS volume distribution for 3C (filled bars) and 3CI (hashed bars). The inset table shows the R_S and the DLS-derived and actual molecular weight values for 3C and 3CI. Sample conditions were: 0.28 mM 3C and 0.31 mM 3CI in 20 mM phosphate buffer pH 6.5 containing 1 mM DTT and 0.5 mM EDTA (CD) and 0.3 mM 3C and 0.3 mM 3CI in 20 mM phosphate buffer pH 6.5 containing 14 mM DTT and 0.5 mM EDTA (DLS).

dynamic behaviour, such as apo 3C in this study, can nonetheless be analysed (Figure 3.7). Not surprisingly, S^2 values calculated from RCI for 3C diverge significantly from the suspect model-free S^2 values described above (R=0.28).

The RCI derived S^2 data for 3C were plotted onto a cartoon representation of the protein. This plot is shown as Figure 3.4(a). Notable is the number of residues with large order parameters (blue) on the RNA binding face of the protein (left hand panel) compared with the large number of residues with broadened peaks (gold) on the active site face (right hand panel). However, surrounding the helical motifs on the RNA binding face, are several residues with intermediate order parameters. The residues A121, G122, G147, G165, F169 and K181 had S^2 values below 0.5 (yellow).

Directly determined exchange data for the uninhibited protein show a large number of residues with R_{ex}^{CC} terms (Figure 3.6 (a)). Of the 79 residues for which R_{ex}^{CC} terms were calculated, 27 (34%) are above 1.5 s^{-1}. The largest of these values is 13.2 s^{-1} for residue R33. This is in a cluster of several residues (H31, D32, R33 and C35) exhibiting R_{ex}^{CC} terms over 1.5 s^{-1}. The majority of the residues in the linker region (residues 78-96) exhibit significant R_{ex}^{CC} values along with Y177 and F178 in the C-terminal region. Several isolated residues also exhibit R_{ex}^{CC} (N3, K22, G44, G51, V56, V115, T119, G148 and A152).

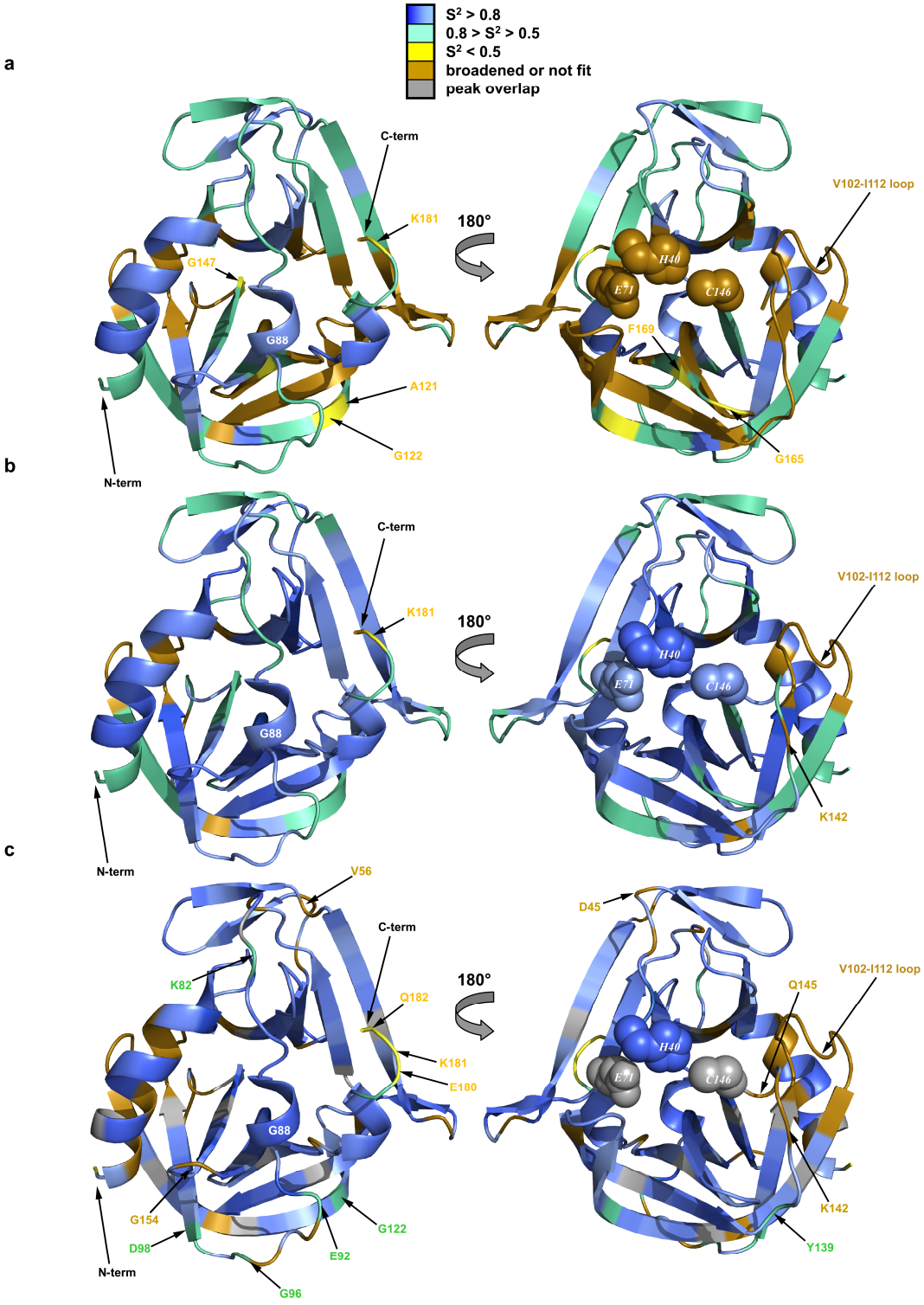


Figure 3.4: (a) RCI-derived S^2 data for 3C, (b) RCI-derived S^2 values for 3CI and (c) Lipari-Szabo-derived S^2 values for 3CI plotted onto cartoon representations of the 3CI protease. In each panel, the RNA-binding face is shown in the left hand view and the catalytic face is shown in the right hand view. The catalytic triad residues (H40, E71 and C146) are depicted as space-filled spheres in the right hand view. S^2 colour-coding is denoted in the bar chart above panel (a).

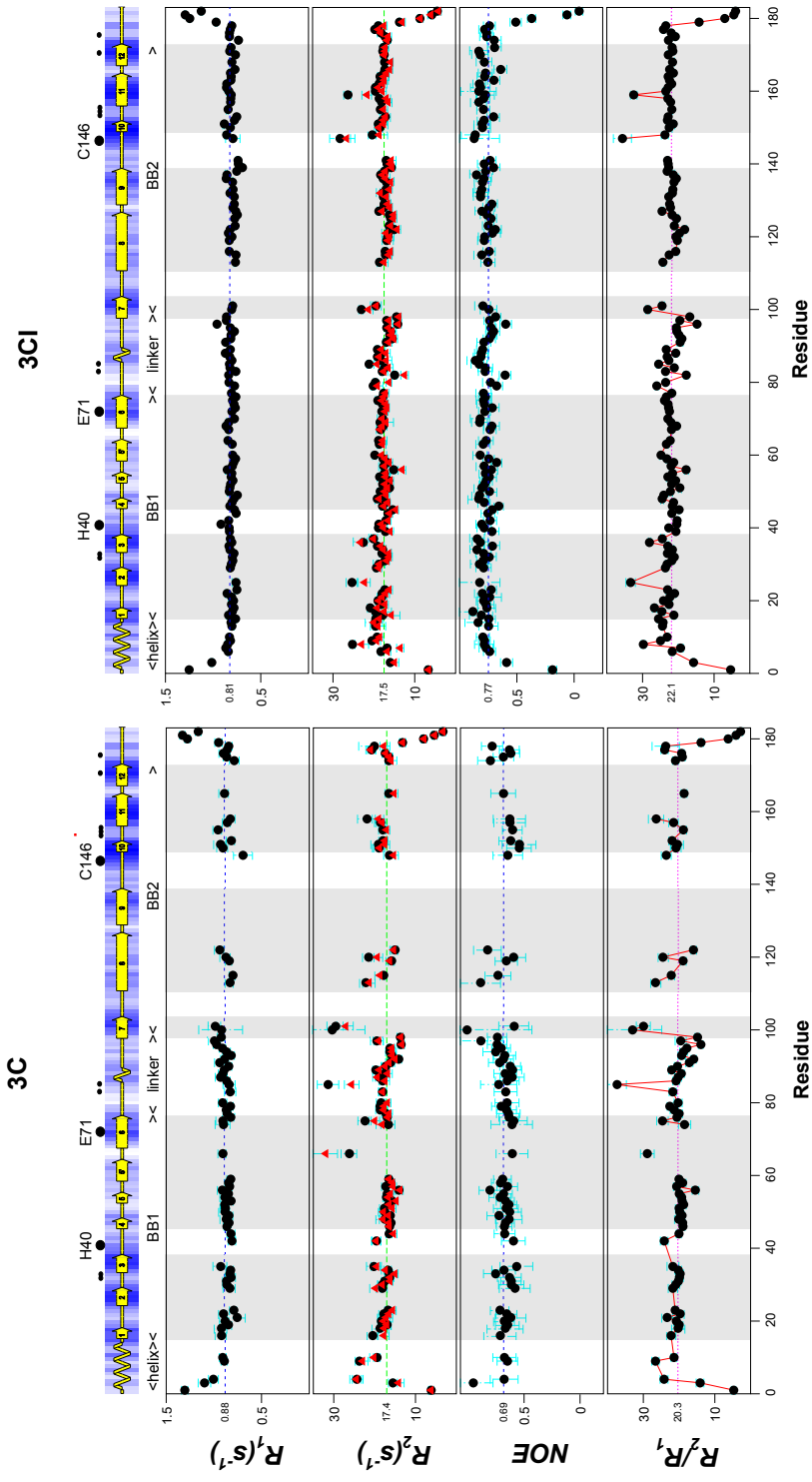


Figure 3.5: (a) R_1, (b) R_2, (c) ^{1}H-^{15}N heteronuclear NOE and (d) the ratio R_2/R_1R1 for 3C (left hand panels) and 3CI (right hand panels), plotted vs. residue number. All plots represent data collected at 700 MHz except the triangles in part (b) that represent R_2 values at 600 MHz. The horizontal dashed lines indicate the trimmed mean value for each parameter. The secondary structure as calculated by PROCHECK (Laskowski *et al.*, 1996, 1993) is shown above the plot. Above the PROCHECK plot, the positions of the catalytic triad residues (H40, E71 and C146) are indicated as large dots and residues implicated in RNA binding by previous mutation studies (Matthews *et al.*, 1994; Hämmerle *et al.*, 1992; Andino *et al.*, 1993) are shown as small dots. The grey boxes in panels a–d highlight the primarily β-strand regions.

Inhibited 3C^{pro} (3CI)

Model-free analysis of ^{15}N relaxation parameters and the ^{1}H-^{15}N NOE was more informative for the 3CI than for the 3C, presumably due to increased rigidity in the presence of inhibitor. The mean R_1 value is 0.83 ± 0.08 s^{-1} (10% trimmed mean 0.81 s^{-1}). The mean R_2 rates at 600 MHz and 700MHz are 17.8 ± 2.9 s^{-1} (10% trimmed mean 17.9 s^{-1}) and 17.0 ± 2.7 s^{-1} (10% trimmed mean 17.1 s^{-1}), respectively. The mean NOE is 0.75 ± 0.12 (10% trimmed mean 0.77). The mean R_2/R_1 ratio is 21.9 ± 4.1 (10% trimmed mean 22.1). Model-free analysis produced a global correlation time (τc) of 12.73 ns. The mean generalised order parameter (S^2: Figure 3.7(b)(open circles)) for fitted residues is 0.85 ± 0.12 (10% trimmed mean 0.87). RCI analysis was also used to calculate order parameters for 3CI. The results of this analysis are presented in Figure 3.7(b) (closed circles). The mean for the RCI derived S^2 values was 0.84 ± 0.08. Figure 3.4(b) shows RCI derived order parameter data plotted onto a cartoon of 3CI. Several residues exhibit S^2 values between 0.5 and 0.8 (shown in green). Only K181 has an RCI derived S^2 value below 0.5 (shown in yellow). Figure 3.4(c) shows model-free derived order parameter data plotted onto 3CI. The model-free and RCI derived S^2 data are largely consistent, although fewer residues show model-free derived S^2 values between 0.5 and 0.8 (shown in green), and attention is drawn to a number of residues that were unable to be used due to peak overlap (shown in grey). The three C-terminal residues (E180, K181 and Q182) all exhibit model-free derived S^2 values below 0.5 (shown in yellow).

Figure 3.6(b) shows the results from analysis of cross-correlation experiments with 3CI. The degree of R_{ex} is greater in the cross-correlation analysis (R_{ex}^{CC}) than in the model-free analysis (R_{ex}^{LS}), although there is some overlap. In all, 45 residues, or 33% of the residues for which data was available, produce an R_{ex}^{CC} term greater than 1.5 s^{-1}. A total of 19 residues, or 15 % of fittable residues, produced R_{ex}^{LS} values greater than 1.5 s^{-1}. In addition, seven residues could not be fit by model-free analysis (denoted by arrows beneath Figure 3.6(c)), typically an indication of complex dynamic behaviour. These residues in many cases correspond to residues with large R_{ex}^{CC} values. More residues with high R_{ex}^{CC} values are found in BB1 (16) than in BB2 (12). Moreover, residues with significant R_{ex}^{CC} terms are relatively evenly distributed throughout BB1, whereas 8 of the 12 residues exhibiting R_{ex}^{CC} over 1.5 s^{-1} in BB2 lie in an 14-residue range spanning part of strands 10 and 11 (residues 152–165). As in the 3C data set, a number of residues in the linker region (D78, R79, N80, K82, F83, F89, E95 and G96) along withY177 and F178 from the C-terminal tail, display significant R_{ex}^{CC} values.

$R_2/R_{1\rho}$ experiments were carried out on the inhibited protein only. The ratio: $R_2/R_{1\rho}$ pro-

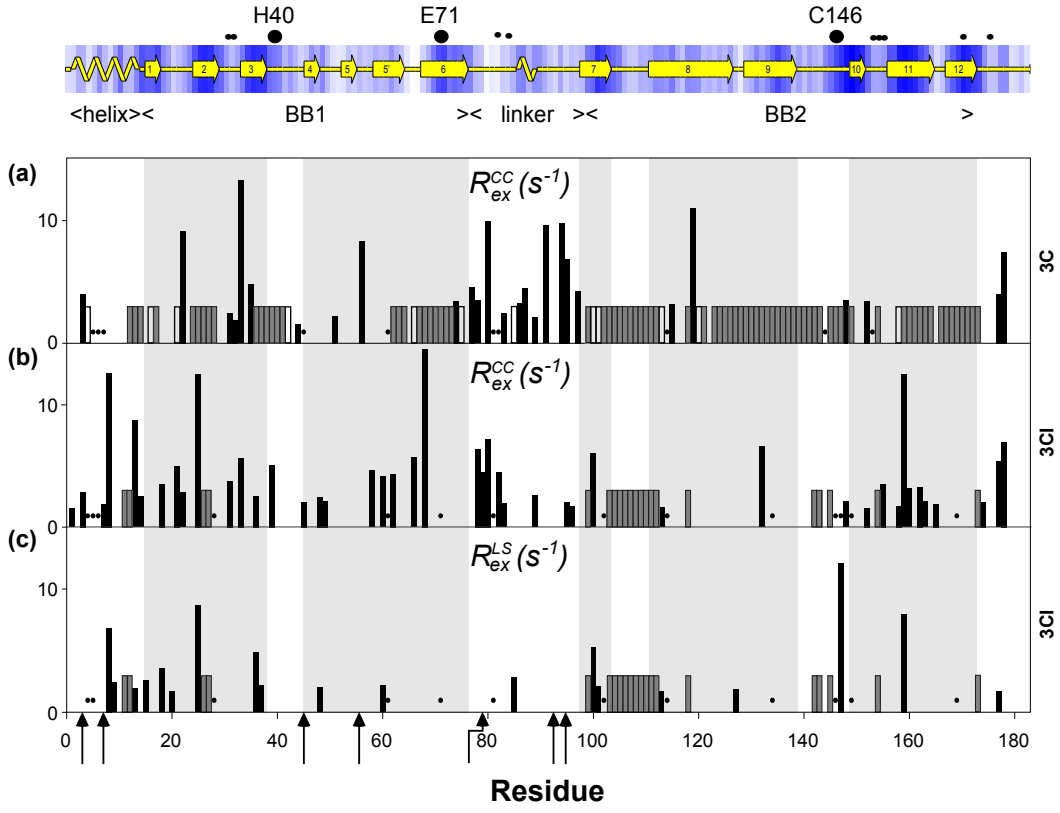


Figure 3.6: R_{ex} values (a) for 3C from cross-correlation analysis, (b) for 3CI from cross-correlation analysis and (c) for 3CI from Lipari-Szabo analysis. Plots represent analysis of data acquired at 700 MHz apart from that shown in (c) which represents analysis of data at both 600 and 700 MHz. Black bars represent residues with R_ex terms greater than 1.5 s^{-1}. Grey bars represent residues exchange broadened beyond detection in 2D ^{15}N,^{1}H-HSQC spectra, unfilled bars indicate peaks that were removed from the analysis due to insufficient signal-to-noise and black dots represent peaks that were overlapped in the ^{15}N,^{1}H-HSQC spectra and therefore removed from the analysis. Vertical arrows below panel (c) represent residues that could not be fitted to any model in Lipari-Szabo analysis. The secondary structure as calculated by PROCHECK along with positions of catalytic triad residues, RNA-binding-implicated residues and the grey shaded boxes are as described in the Legend to Figure 3.5.

vides information on whether intermediate time-scale motions ($\sim$1 ms) are occuring. A value for $R_2/R_{1\rho}$ over 1.1 is significant. Several residues gave significant values in the $R_2/R_{1\rho}$ these are presented in Table 3.1.

Residue	$R_2/R_{1\rho}$
Thr4	1.14 ± 0.06
Ser9	1.10 ± 0.06
Gly23	1.12 ± 0.05
Phe25	1.38 ± 0.45
Ile37	1.16 ± 0.08
Val63	1.10 ± 0.06
Asp64	1.22 ± 0.11
Ser128	1.10 ± 0.06
Gly148	1.26 ± 0.17

Table 3.1: Residues with significant $R_{1\rho}$ values

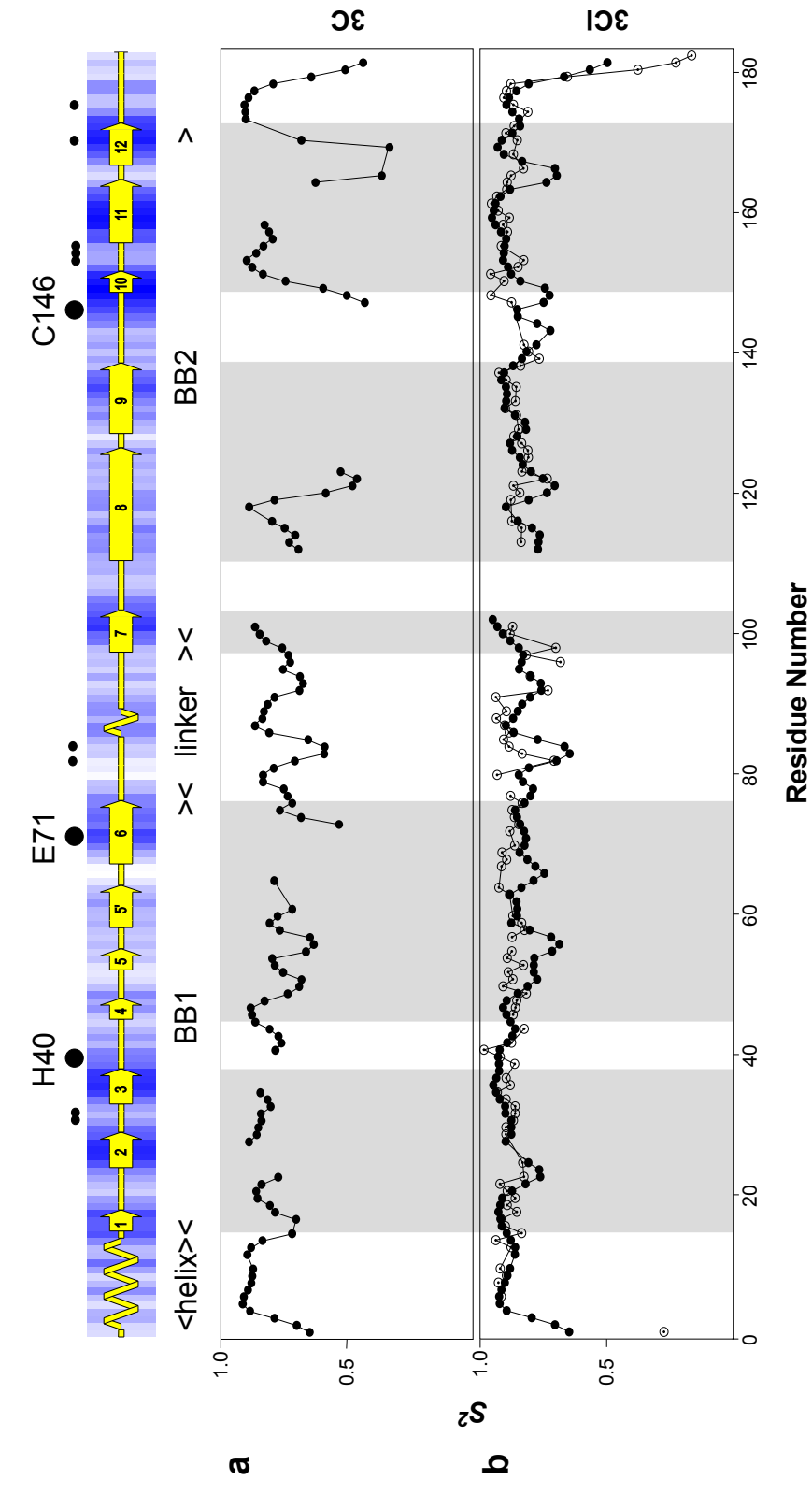


Figure 3.7: Generalised order parameter S^2(a) for 3C, calculated from RCI analysis and (b) for 3CI, calculated from RCI (open circles) and from Lipari-Szabo (filled circles) analyses. The secondary structure as calculated by PROCHECK along with positions of catalytic triad residues, RNA- binding-implicated residues and the grey shaded boxes are provided as described in the Legend to Figure 3.5.

3.5 Discussion

3.5.1 3CI relaxation

The inhibitor-bound 3CI is a well-structured protein. The high S^2 values observed throughout both β-barrels (Figure 3.7(b), open circles) are indicative of low level ns-ps time scale dynamics (Palmer, 1993). This observation reflects the fact that picornaviral proteases have unusually short loops joining secondary structural motifs (Skern *et al.*, 2002). This feature may provide an evolutionary advantage through avoidance of exposure to proteolysis. The lowest S^2 values were observed for residues at the N and C termini and in parts of the linker joining the two barrels. The linker is thus a highly vulnerable part of the protease, and flexibility present in this region must play a functional role. This flexibility is likely to be related to the involvement of the linker in RNA-binding (Andino *et al.*, 1990b; Leong *et al.*, 1993; Walker *et al.*, 1995).

Of the 3CI relaxation parameters, R_1 rates are particularly uniform (Figure 3.5(a)). The R_1 values are higher only at the termini, consistent with an increase in mobility at the chain ends. Concurrent with the increase in R_1 at the termini is a decrease in NOE values (Figure 3.5(c)). The NOE is more variable than R_1 across the rest of the protein. There are several minor decreases in NOE values, the most significant of which are N82 (0.60), G96 (0.60) and $R_1$66 (0.64), consistent with some mobility at these sites. Although NOEs at the termini are lower than the other regions, the values are high relative to termini for many proteins and do not reflect total disorder. The R_2 rates are also highly uniform, though not quite as uniform as the R_1 rates and the NOEs. The low R_2 rates at the termini reflect a greater level of disorder at the termini but are again higher than in totally disordered proteins. The range of R_2 rates in the linker suggests the presence of slower time scale motions in this region.

Six residues have significantly higher than average R_2 rates: L8(25.3), F25(25.4), V36(22.6), L100(23.1), G147(28.3) and I159(26.3). These residues produce the six largest R_{ex}^{LS} terms (Figure 3.6(c)): L8(6.5), F25(8.7), V36(4.8), T100(5.2), G147(12.1) and I159(8.0). Thirteen additional residues (19 in total) fitted to Lipari-Szabo formalism required R_{ex}: S9, K13, I15, I18, T20, I37, L48, Y60, D85, L101, L113, S127 and Y177 (Figure 3.6(c)). These cluster around the active site or the RNA binding surface. Lipari-Szabo analysis produces the following mean S^2 values for various segments of the molecule (a trimmed mean is provided for BB1 and BB2): The N-terminal helix: 0.80 ± 0.26; the N-terminal β-barrel (BB1): 0.88 ± 0.03 (10% trimmed mean 0.88); the linker region: 0.85 ± 0.08; the C-terminal β-barrel (BB2): 0.87 ± 0.05 (10% trimmed mean 0.87); the C-terminus:

0.64 ± 0.30. S^2 values calculated from RCI for 3CI are in moderately good agreement with the S^2 calculated from model-free analysis. (R=0.61). The cross-correlation analysis of 3CI (Figure 3.6(b)) detects a greater degree of exchange dynamics (45 R_{ex}^{CC} rates greater than 1.5 s^{-1}) than model-free analysis. This is not surprising as the cross-correlation approach can detect events on time scales that do not require an R_{ex} parameter in Lipari-Szabo formalism. However, five of the above mentioned six residues with highest R_{ex}^{LS} rates alsoproduce significant R_{ex}^{CC} rates, including three of the four largest R_{ex}^{CC} values: L8(12.5), F25(12.5), and I159(12.4). The sixth of these residues, G147, was excluded from the R_{ex}^{CC} analysis due to low signal-to-noise ratio. The low intensity of this peak clearly suggests a degree of exchange broadening. Of the 19 total residues with R_{ex}^{LS} terms, 11 display R_{ex}^{CC} values, while five of the remaining eight residues sequentially border a residue with significant R_{ex}^{CC} value. This leaves a total of 34 residues exhibiting R_{ex}^{CC} terms but not R_{ex}^{LS} terms. Of the 13 highest R_{ex}^{CC} terms, seven correspond to residues on the RNA-binding face: L8, K13, R33, D78, N80, Y177 and F178. The three residues E66, I68 and N132 are spatially adjacent to each other on the BB1–BB2 interface. F25, T100 and I159 are mentioned below. D78 and N80 from this list are part of a sequentially contiguous stretch (D78–F83) of linker residues with either high R_{ex}^{CC} or an overlapped HSQC peak. I159 is part of the other large nearly contiguous stretch (G158-G163) of high R_{ex}^{CC} terms that form most of β-strand 11 that traverses the centre of the protein. Another residue with a notable R_{ex}^{CC} term, H31, packs against the dynamic D78–F83 region of the linker along with the previously mentioned R33. The number of 3CI ^{15}N,^{1}H-HSQC peaks broadened beyond the detection limit is relatively small (see grey bars, Figure 3.6(b)). The N-terminal helix contains the broadened L11 and R12 residues, and also seven residues for which one or both of the employed methods detect R_{ex}(Figure 3.6(b) and 3.6(c)). The T26-G27 peaks are absent in 3CI spectra. These two residues, along with the highly dynamic (on a ms-μs time scale) F25 residue, spatially border the active site (Figure 3.2(b)). The V103-I112 loop is the largest string of residues with missing (exchange broadened) resonances in the inhibited protein. At one end of the loop, T100 exhibits large R_{ex}^{CC} and R_{ex}^{LS} terms and L101 shows an R_{ex}^{LS} term. At the other end of the loop, both methods employed detected exchange behaviour in L113. This loop, together with the loop containing K142-T143 and Q145, each of which is also exchange broadened (Figure 3.2(b)), also spatially borders the active site. The lack of an HN cross-peak for G154 may be related to its proximity to the aromatic rings of F6 and F89. The amide peak for G154 was also not observed in a previous study of $3C^{pro}$ in solution (Bjorndahl *et al.*, 2007). Seven 3CI residues could not be fit in the Lipari-Szabo formalism: N3, A7, D45,

V56, R79, D93 and E95 (Figure 3.6(c)). Of these seven, five exhibit slow dynamics as detected by cross-correlation experiments (Figure 3.6(b)).

3.5.2 3C relaxation and comparison with 3CI

Significant and pervasive line broadening was observed in ^{15}N,^{1}H-HSQC spectra of 3C, and seems to be largely associated with the active site face of the protease (Figure 3.2(a)). This result has been observed reproducibly in our laboratories over a number of years. Interestingly, the protease is active in this exchange-broadened form (Chapter 2). In addition, the uptake of the JMC-98-3 inhibitor by 3C suggests that there is no significant mis-folding or obstruction of the active site region. A previous study of 3C and a distinct inhibitor (acetyl-LEALFQ-ethylpropionate) also reported pervasive line broadening in apo 3C, but to a somewhat lesser extent, with chemical shift assignment less compromised (Bjorndahl *et al.*, 2007). A number of additional experiments were performed as independent probes of the difference between 3C and 3CI structure. DLS data provide no evidence of 3C or 3CI self-association under the conditions used. Gel data show an absence of intermolecular disulphide bond formation, ruling this out as a possible dimerisation mechanism. Small-angle X-ray scattering (SAXS) studies of 3C^{pro} provide additional evidence that there is no significant self-association of either the apo or inhibited protease at the concentrations used Chapter 4 (personal communication, J. Trewhella). CD spectroscopy indicates that 3C and 3CI share similar secondary structure and that 3CI is slightly more thermo-stable than 3C, consistent with the conclusion that 3CI is slightly more ordered than 3C. Deconvolution indicates the following approximate α-helical and β-strand content: 3C, 7 % helix and 55 % strand; 3CI 5 % helix and 57 % strand. This is consistent with the secondary structure content calculated from the previously published pdbs of 3C^{pro} in solution (Bjorndahl *et al.*, 2007): 3C^{B}(previously published apo form) 9 % helix and 57 % strand; 3CI^{B}(previously published inhibited form) 11 % helix and 59 % strand. The residues for which peaks are observable in the HSQC spectrum and for which relaxation data could be obtained for 3C showed that, for these regions, the protein is also well structured in its uninhibited form. The average R_1 value for 3C is slightly higher than 3CI (0.88 versus 0.81); the average R_2 value for 3C is nearly identical to 3CI (17.4 versus 17.5); the average hetero-NOE value for 3C is lower than for 3CI (0.69 versus 0.77). Taken together, these data indicate slightly more order in 3CI than in the assignable regions of 3C. Cross-correlation analysis of 3C shows several residues with significant R_{ex}^{CC} terms (Figure 3.6(a)). These are most highly concentrated in the linker (residues 78-96) and other regions of the RNA-binding surface (Figure

2(a)). On the RNA-binding face, the residues exhibiting R_{ex}^{CC} in 3C and 3CI overlap (Figure 3.2(a) and (b)). However, there are a larger number of residues in the linker region of 3C exhibiting R_{ex}^{CC} terms than in 3CI (Figure 3.6(a) and (b)), consistent with a global reduction of ms–us dynamics in 3CI. In addition, several residues in 3CI show R_{ex}^{CC} that is absent in 3C, indicating that the ms-us flexibility present in 3CI is not a mere subset of the intermediate time scale dynamics found in 3C.

3.5.3 RCI analysis of 3C and 3CI

As fitting of the 3C relaxation data using model-free was unreliable, RCI analysis was used. RCI data from 3C (Figure 3.7(a)) confirm that, while parts of the protein are ordered, these parts are somewhat less ordered than 3CI (Figure 3.7(b)). The average RCI-derived S^2 value for the observable 3C peaks is 0.76 ± 0.14 compared with 0.84 ± 0.08 for 3CI. Interestingly, the RCI values for 3C show trends towards lower values (indicating a higher degree of motion) in regions sequentially approaching residues with unobservable peaks. This corroborates the supposition of significant slow motion in the unobservable residues and thus reinforces the overall conclusion of mobility in the apo protease.

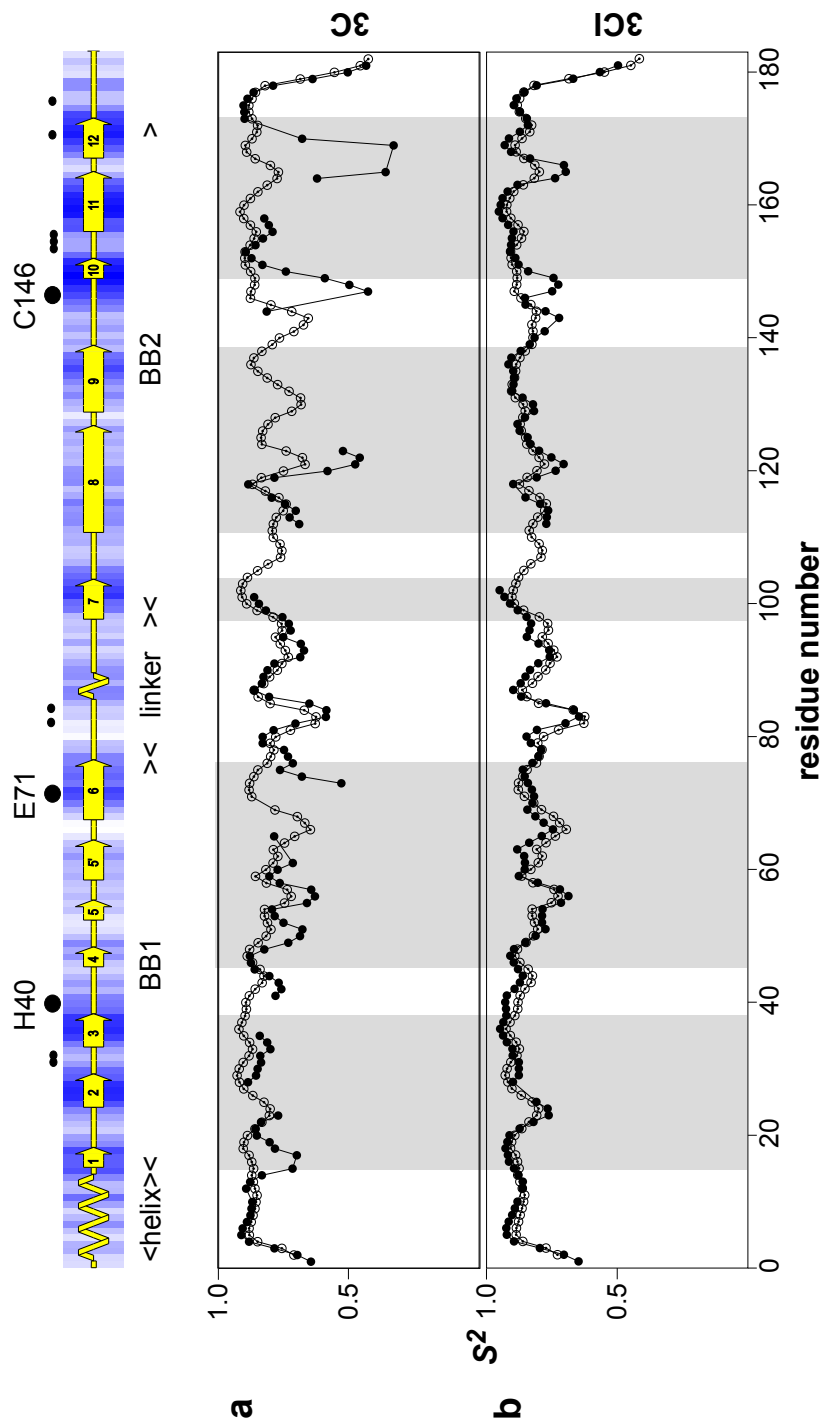


Figure 3.8: RCI-derived S^2 values plotted against residue number for (a) 3C (filled circles) and $3C^B$ (open circles) and (b) 3CI (filled circles) and $3CI^B$(open circles). The $3C^B$ and $3CI^B$ values are as previously reported (Bjorndahl *et al.*, 2007). The secondary structure as calculated by PROCHECK along with positions of catalytic triad residues, RNA-binding-implicated residues and the grey shaded boxes are as described in the Legend to Figure 3.5.

3.5.4 Distinct dynamics states of 3C protease

Comparison between the Bjorndahl results (Bjorndahl *et al.*, 2007) and our data reveal three distinct, but related, states of 3C protease. In the largely similar inhibited forms of the protein (bound to JMC-98-3 in the current study and to acetyl-LEALFQ-ethylpropionate in the previous study) RCI-derived S^2 values (Figure 3.8(b)) are very similar (R=0.83). This indicates very similar dynamic regimes for the protein when inhibited. Three stretches in BB2 of 3CI (141-144, 147-149 and 164-166) show RCI-derived S^2 values significantly lower than those in the inhibited protein in the Bjorndahl study. This probably reflects differences between the two inhibitors used in the studies as these residues are spatially adjacent to the inhibitor in 3CI (Figure 3.2(b)). The RCI values for 3C from this study and for the apo enzyme from the Bjorndahl work, however, differ markedly (R=0.52, Figure 3.8(a)). In the Bjorndahl analysis, 3C^B(the apo form) is less ordered than 3CI^B, (the inhibited form) especially in BB2. It appears from exchange broadening and RCI analysis of 3C (present study), that this form of the protein has a higher level of disorder than 3C^B.

A comparison of RCI-derived S^2 values for 3C, 3CI, 3C^B and 3CI^B, broken down by secondary structure element is shown in Figure 3.9. Several points are evident from this comparison. First, note that the average S^2 (S^2_{ave}) pattern is very similar for each of the inhibited proteins (3CI and 3CI^B), with the linker and C-terminal tail somewhat less ordered than the N-terminus and the β-barrels. However, although both of the apo forms of the protein are less ordered than the inhibited forms, they are somewhat dissimilar to each other. In the 3C protein, each segmental S^2_{ave} value is lower than the corresponding value from 3C^B. The largest difference is in BB2, for which the 3C and 3C^B S^2_{ave} values are 0.70 and 0.83, respectively. Thus, the average dynamic behaviour of the inhibited forms of the protein are similar, but the apo forms differ from each other and from the inhibited forms, most notably in enhanced flexibility in the 3C BB2 segment. It is possible that the 3C^B structure is, in fact, in the inhibited conformation. This could, perhaps, be due to endogenous inhibitors in the sample preparation procedure. It is noted that one preparation in our lab lead to a full peak spectrum of an ostensibly 'apo' protein. The source of the differences in degree of peak broadening in these two studies is not clear at present. However, in both studies the enzyme is capable of binding inhibitor with concomitant loss of broadening. The fact that all currently available crystal structures of rhinoviral 3C proteases (Matthews *et al.*, 1994, 1999) were determined in the presence of active-site inhibitors may be related to the propensity for disorder near the vacant active site that has now been reported from two independent labo-

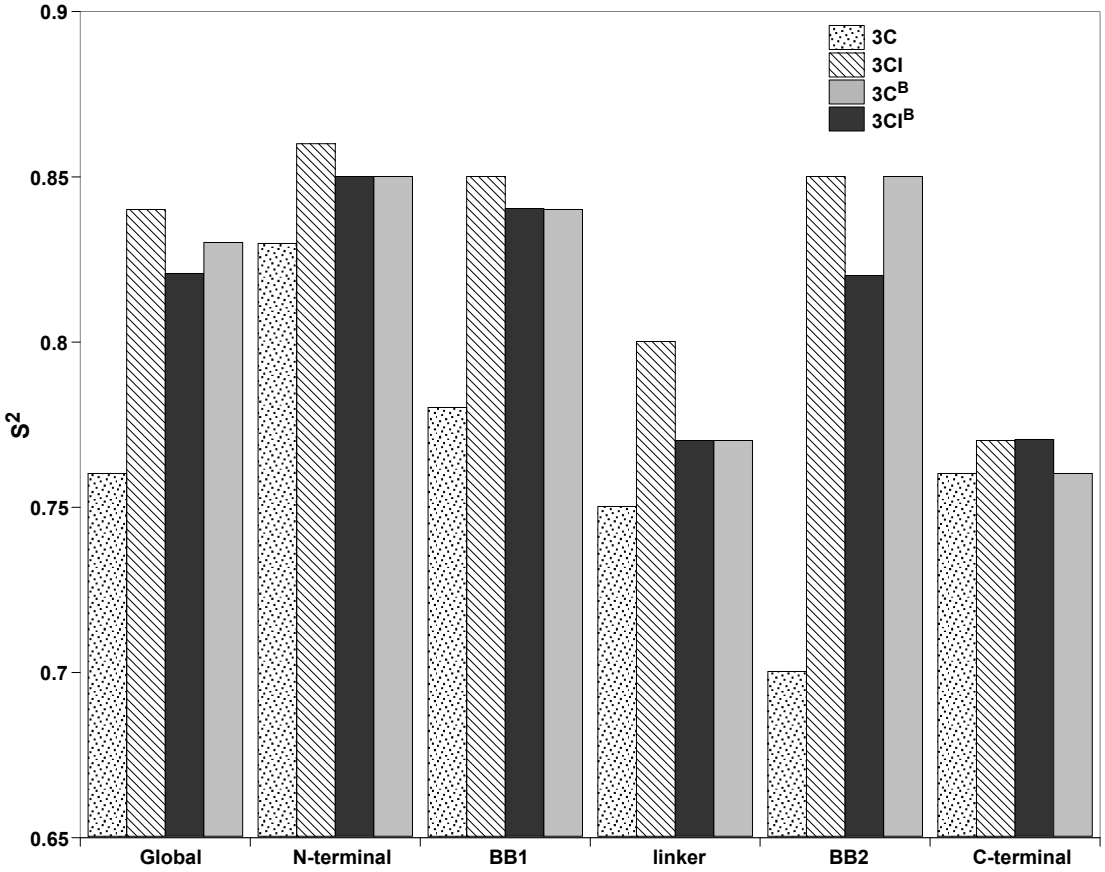


Figure 3.9: RCI-derived S^2 values for 3C, 3CI, 3C^B and 3CI^B grouped by secondary structural motifs. Only residues observable in all states are included. N-terminal: residues 1–15; BB1: residues 16–76; linker: residues 77–96, BB2: residues 97–172; C-term: 173–182. The 3C^B and 3CI^B values are as previously reported (Bjorndahl *et al.*, 2007).

ratories. Active-site inhibitors have previously been shown to stabilize protease folds and active site conformation (Barbato *et al.*, 2000; Zahnley, 1980)and thus occupancy of the active site may make $3C^{pro}$ more amenable to crystallisation

3.5.5 Substrate-recognition loop mobility

It has been suggested that the 140-147 loop connecting β-strands 9 and 10 in 3C proteases is more flexible relative to the homologous loop in otherwise similar serine proteases (Matthews *et al.*, 1994). From this loop, the active site nucleophile, C146, resides near β-strand 10; T141, which is involved in substrate recognition, resides near β-strand 9; and the more central G144 is involved in stabilisation of the transition state. This loop is highly dynamic on a ms-μs time scale as shown by broadening in both 3C and 3CI and R_{ex} terms in 3CI. In 3CI, the K142, T143 and Q145 peaks remain broadened; G147 was assigned the largest value for R_{ex}^{LS} in the analysis (12.11) and G147 along with the nucleophilic C146 produce peaks of low intensity. R_{ex} is detected in the neighbouring G148 residue via cross-correlation analysis. In uninhibited 3C, residues A140-G147 are all unobservable with the exception of G144. The dynamic properties of the loop found in this study are indicative of some such functional role in substrate recognition, indeed, it has previously been suggested that increase in catalytic efficiency of the enzyme in the presence of high levels of anions is due to a conformational change in this loop (Wang and Johnson, 2001).

3.5.6 Other active-site-associated dynamics

There are many additional instances of slow time scale dynamics in and around the active site in 3CI including S21, K22, T39, S127, H160, G162, G163 and G165. These residues, along with the substrate-recognition loop mentioned above, nearly encircle the catalytic triad (Figure 3.2(b)). Together with the widespread exchange broadening observed on the active site face of apo 3C, indicates that the active site of 3C protease possesses the ability to sample two or more conformational substates. This suggests a possible population shift mechanism of binding, wherein the target peptide selectively binds to and stabilises a preferred conformational substate of the enzyme. The loop from residues 138 to 150 that connects β-strands 8 and 9 in $3C^{pro}$ from foot-and-mouth disease virus folds over to contact the substrate in the binding site, and is highly flexible (Sweeney *et al.*, 2007). However, crystallographic studies in HRV2 (Matthews *et al.*, 1999) and HRV14 (Matthews *et al.*, 1994) and data from the current work indicate that the analogous but much shorter loop (residues 126-129) from $3C^{pro}$ is more rigid. In the present study, the 3CI S^2

values for this loop (S^2_{ave}= 0.84) are similar to those from the surrounding β-strands. In the presence of the inhibitor, the relatively modest R^{LS}_{ex} value of 1.71 ± 0.54 for S127 (Figure 3.2(b)), which is not accompanied by a significant R^{CC}_{ex} value, is the only datum not consistent with a relatively high degree of order in this loop. $R_{1\rho}$ analysis indicates that several residues are mobile on the time-scale ($\sim$ 1 ms) probed by this technique. The residues affected are plotted in Figure 3.10. The figure shows that several residues with ms motion cluster around the interface between the two β-barrels. This may suggest that some interdomain motion is present in 3CI.

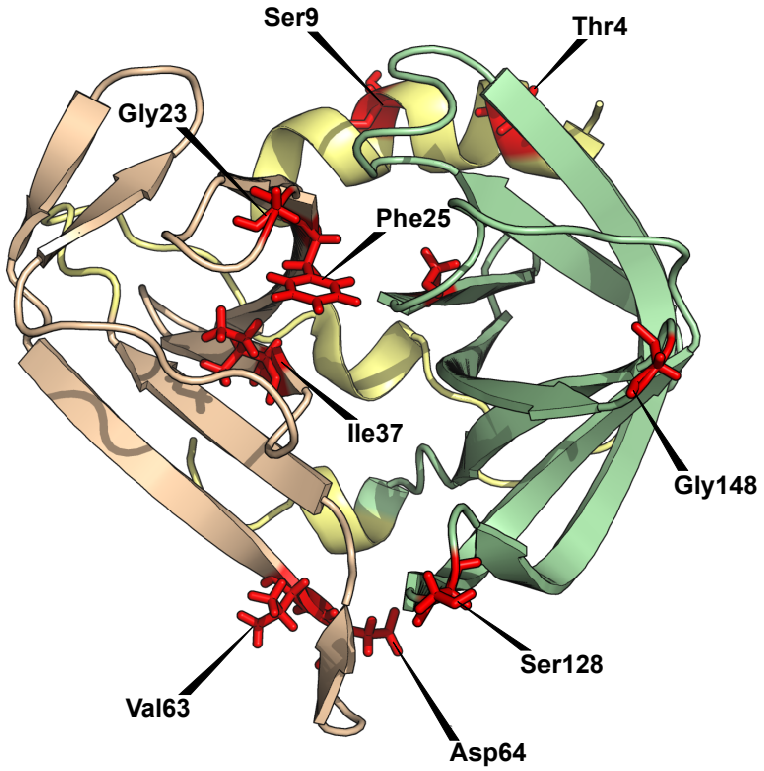


Figure 3.10: $R_{1\rho}$ detected-exchange in 3CI. BB1 is shown in cream and BB2 in pale green. The helices and interdomain linker are shown in yellow. Residues with $R_2/R_{1\rho}$ greater than 1.1 are shown as red sticks.

3.5.7 Mobility of RNA-binding surface

The opposite face of the protein has been implicated in RNA binding (Leong *et al.*, 1993; Andino *et al.*, 1990b,a; Walker *et al.*, 1995; Paul *et al.*, 2000; Amero *et al.*, 2008; Claridge *et al.*, 2009). Two of the major features of this face are the N-terminal α-helix and the long linker joining the two β-barrels. As discussed above, the linker (residues 77-97) is one of the few segments of

3CI displaying fast dynamics in some residues, but some sections of the linker are less mobile. L77-N80 at the N-terminal end of the linker and F83-S91 in the linker centre have higher than average S^2 in the 0.9 range. Conversely, K82 is relatively disordered, as are the residues C-terminal to S91. E92, G96 and D98 have S^2 values less than 0.8 and D93 and E95 could not be fit to Lipari-Szabo formalism. In 3C the linker also exhibits a degree of fast timescale dynamics, especially in the residues adjacent to the helical region in its centre. In 3C, the residues N80, F83, I86, F89, S91, E95 and V97 all display significant values and R84 is broadened beyond detection. The single turn of helix in the centre of the linker (S^2_{ave} = 0.91 in 3CI) and the N-terminal helix (S^2_{ave} = 0.90 in 3CI) could be critical, relatively rigid recognition elements for target RNA. The RCI derived values for both 3C and 3CI also suggest that these elements are relatively rigid in both forms of the protein (N-terminal helix S^2_{ave} = 0.90 in 3CI, 0.87 in 3C; linker helical region S^2_{ave} = 0.84 in 3CI, 0.80 in 3C). The lack of chemical exchange on the internal face of the linker helix suggests that the position of this helix is stable in 3CI and 3C. Conversely, the outward facing side of the N-terminal helix is highly ordered in both 3C and 3CI, but extensive broadening (L11 and R12) and R_{ex} (G1, N3, A7, L8 and S9) is seen on the face of the N-terminal helix packing against the protein in 3CI. In apo 3C, R12 and K13 are exchange broadened. Together, these characteristics suggest that motions near the N-terminus may not disrupt the formation of the N-terminal helix, but the position of this helix relative to the rest of the protein may be unstable in both forms of the protease. The N-terminal helix may even possess the potential to pivot away from the remainder of the protein and interact with the target RNA. $R_{1\rho}$ terms were assigned to T4 and S9 in 3CI, indicating motion in the $\sim$1 ms range. Mobility of the N-terminal helix was also reported in crystallographic studies of HRV14 3C (Matthews *et al.*, 1994). B-factors for the N-terminal helix averaged 50 Å^2 vs. 26 Å^2 for the remainder of the protein. In structures of inhibited 3C (Claridge *et al.*, 2009; Bjorndahl *et al.*, 2007; Matthews *et al.*, 1994), the back of this helix packs against the highly mobile loop connecting β-strands 9 and 10. As discussed above, this loop approaches the active site and thus changes induced in this loop may be responsible for transmitting inhibitor-induced changes to the N-terminal helix. The B-factors for HRV14 linker helix were not reported. However, the HRV2 linker helix B-factors were on average 17 Å^2. In 3CI, a cluster of high R_{ex} residues near the N-terminus of the linker, D78-N80 and K82-F83 (The E81 peak is overlapped in the 2D ^{15}N,^{1}H-HSQC) pack against the loop connecting β-strands 2 and 3. This loop contains H31 and R33, which also exhibit large R_{ex} terms. K82 and F83 are part of the KFRDIR motif (residues 82-87) that is highly conserved in picornavirus and has previously been implicated in

RNA binding (Leong *et al.*, 1993; Shih *et al.*, 2004; Gorbalenya *et al.*, 1989). The D78-R87 stretch includes five basic residues with potential to interact with the RNA phosphodiester backbone and F83, which could, along with F89 (also exhibits R_{ex}), participate in base stacking with RNA (Walker *et al.*, 1995; Kenan *et al.*, 1991). Although similar motions are present on the RNA-binding face of the protein in both 3C and 3CI states of the protein, there is a change in the The difference in distribution and magnitude of R_{ex} terms between 3C and 3CI raises the possibility of allostery between the catalytic site of 3C and the RNA-binding surface. The possibility of a "communication pathway" between the active site and the RNA-binding region in discussed further in Chapter 5. Further evidence for the passage of information between the active site and the RNA-binding surface of $3C^{pro}$ comes from studies on the closely related hepatitis A virus (HAV) and enterovirus $3C^{pro}$. Surface plasmon resonance studies in HAV showed that binding of specific peptides increased the RNA affinity of $3C^{pro}$ (Peters *et al.*, 2004). In addition, mutations in the RNA-binding regions of enterovirus 71 $3C^{pro}$ affected the catalytic activity of the enzyme as well as RNA-binding efficiency (Shih *et al.*, 2004).

3.6 Conclusion

It is now well accepted that a delicate balance must be maintained between order and flexibility in proteins to achieve specificity and affinity without suffering a debilitating loss of entropy upon binding. The structural dynamics of $3C^{pro}$ from human rhinovirus 14, in the presence and absence of the specific inhibitor JMC-98-3 have been examined using both ^{15}N relaxation and chemical shift based approaches. The results clearly show that the inhibited protease is highly ordered, with mean order parameters of 0.85 and 0.84 via relaxation and chemical shift analyses, respectively. The highest degree of ns-ps time scale motion is evident at the chain termini. A lesser degree of fast motion is present in the linker region joining the two β-barrels. The order parameters derived via relaxation and chemical shift-based approaches are very similar, and they largely agree with chemical shift-based order parameters previously reported for HRV14 $3C^{pro}$ bound to a distinct peptidyl inhibitor (Bjorndahl *et al.*, 2007). However, in the present study, model-free analysis and cross-correlation studies permit the additional analysis of ms-μs time scale dynamics. Biologically significant events such as enzyme catalysis, complex formation and functionally relevant conformational rearrangement have been observed to occur on this time scale (Epstein *et al.*, 1995; Kristensen *et al.*, 2000; McCoy *et al.*, 2001; Dutta *et al.*, 2004; Mukherjee *et al.*, 2006; Kempf *et al.*, 2007; Palmer, 2004; Ishima and Torchia, 2000; Mittermaier and Kay, 2006). Significant slow motions in the inhibited protease are most prevalent in the linker, in the N-terminal helix, and in regions spatially bordering the proteolytic active site. The linker region and the N-terminal helix are both involved in RNA-binding. The apo 3C protease was found to be considerably more dynamic than the inhibited form, with an average order parameter of 0.76 determined via chemical shift analysis. In addition, severe resonance broadening was observed across nearly the entire catalytic face of the protease, particularly those regions of this face that form the surface of the C-terminal β-barrel. This result is similar qualitatively, but more pronounced quantitatively, to broadening previously reported for the apo protease (Bjorndahl *et al.*, 2007). Apo 3C order parameters derived from chemical shift data are in general lower in the present study than in the previous analysis, particularly within the C-terminal β-barrel. In addition, ms-μs dynamics was specifically detected at the RNA-binding surface as was found for the inhibited protease. However, distribution of slow dynamics on this surface is distinct in the inhibited vs. uninhibited forms. As the picornaviral 3C protease is a multi-functional protein, with opposite faces participating in proteolytic cleavage and RNA-binding activities, this change

in dynamic behaviour upon inhibitor binding suggests that information may be transmitted from the active site to the RNA-binding surface. This potential allosteric pathway may be linked to the balance and progression between protein synthesis and RNA replication in the viral cycle. Further work is required to ascertain whether the allostery in HRV14 $3C^{pro}$ is uni- or bi-directional and whether this interaction can be harnessed to increase the efficacy of pharmaceuticals for the treatment of picornaviral infections.

The chemical and physical properties of a protein cannot be fully interpreted until all three levels of structure are understood, for these properties depend on the spatial relationships between the amino-acids, and these in turn depend on the tertiary and secondary structures as much as on the primary. Only X-ray diffraction methods seem capable, even in principle, of unravelling the tertiary and secondary structures.

John Cowdrey Kendrew
(1917–97)

4

Interaction of 3C with Stemloop-D RNA

4.1 Introduction

This chapter relates work done to characterise the interaction between HRV 14 $3C^{pro}$ and a 27mer RNA fragment from the 5′ cloverleaf, from the UTR of the HRV 14 genome. The structure of the protein had already been determined, both in previous studies (Matthews *et al.*, 1994; Bjorndahl *et al.*, 2007) and by the author and co-workers (Chapter 2). The structure of the 27mer RNA fragment had also previously been solved in solution (Headey *et al.*, 2006). In addition to this structural work, a number of mutational studies had defined residues and bases likely to be involved the interaction between the RNA and protein constituents of the complex (Andino *et al.*, 1993; Leong *et al.*, 1993; Matthews *et al.*, 1994; Ohlenschläger *et al.*, 2004).

The work described in Chapter 2 determined conditions that enabled the study of the complex, first by NMR and, subsequently, by a combination of SAXS and NMR.

Attempts to crystallise a complex of $3C^{pro}$ and the 27mer failed to yield any diffracting crystals.

In the absence of diffracting crystals, NMR is the method of choice for structural determination of macromolecules. In this case, the usual sequence of events — backbone assignment, side-chain assignment, NOEs and then structural calculation — was not feasible and other methods were brought into play.

The SAXS data acquisition and analyses described in this chapter were carried out by Jill Trewhella, John Chow and Cy Jeffries at the University of Sydney using samples and NMR data provided by the author. Much of the contents of this chapter has been published separately (Claridge *et al.*, 2009).

4.2 Materials and Methods

4.2.1 NMR sample preparation

NMR experiments were recorded with either uniformly $^{15}N/^{13}C$-labelled or $^{2}D/^{15}N/^{13}C$-labelled $3C^{pro}$ in 50 mM bis-tris propane/MES containing 0.5 mM EDTA and 1 mM DTT at pH 6.0.

Isotope-labelled $3C^{pro}$ from HRV-14 was expressed in *E. coli* and purified using standard methods. Typically, a 50 ml overnight growth of *E. coli* BL21 CodonPlus cells containing the HMBP-3C plasmid was used to inoculate 1l of LB containing chloramphenicol and ampicillin.

This construct produces a histidine-tagged MBP fusion with 3Cpro that can be cleaved with thrombin (Alexandrov *et al.*, 2001). When absorbance at 600 nm reached approximately 1.0, the cells were harvested by centrifugation at 1541 $\times$ *g* (GS3 rotor) and resuspended in 500 ml of M9 minimal medium supplemented with 3.5 g of ^{13}C-glucose and 0.5 g of 15N-ammonium chloride. After 30 min at 30 °C, induction was initiated by the addition of 0.5 mM isopropyl β-D galactopyranoside (IPTG). For production of deuterated protein, the M9 minimal medium utilised D_2O with either ^{2}H,^{13}C-glucose or ^{2}H,^{12}C-glycerol as a carbon source. The culture was grown at 30 °C until the growth curve flattened (approximately 6 h). Cells were harvested by centrifugation at 1784 $\times$ *g* at 4 °C for 20 min and resuspended in 30 ml of aqueous buffer containing 20 mM Tris and 150 mM NaCl at pH 7.5 (buffer A) with one pre-dissolved Complete™ EDTA free protease inhibitor tablet (Roche). Cells were lysed by passing twice through a French press at a pressure of 6000 psi. The lysate was spun down at 17,590 $\times$ *g* for 30 minutes at 4 °C (Sorvall SS34 rotor) and the supernatant was passed through an 80 μm filter and loaded onto a Ni^{2+}-NTA (GE Healthcare) column equilibrated with buffer A. The column was washed with three column volumes of 20 mM imidazole in buffer A. H6-MBP- tagged $3C^{pro}$ was then eluted with two column volumes of buffer A containing 250 mM imidazole. A total of 30 units of bovine thrombin (Roche) was added to the eluted fraction and this was dialysed against 1l of aqueous buffer containing 20 mM Tris, 1 mM DTT and0.5 mM EDTA at pH 7.8 (buffer B) for 8 h with a change of buffer at 4 h. A QAE-Sephadex (Sigma-Aldrich) column was equilibrated with buffer B and the dialysed protein fraction was passed through the column. The column was washed with a further three column volumes of buffer B. The flow- through and washes were combined, concentrated to approximately 250 μl and buffer exchanged to 50 mM bis-tris-propane MES (BTM), 14mM DTT, 0.5mM EDTA at pH 6.0 (buffer C) via a 10 kDa molecular weight cut-off spin concentrator. The final protein concentration was determined by absorbance at 280 nm. JMC-98-3 (Webber *et al.*, 1998) $3C^{pro}$-specific inhibitor was a gift from S. Webber, (Agouron Pharmaceuticals, CA, USA, now part of Pfizer, CA, USA) and was added in equimolar or greater ratio to purified $3C^{pro}$ to create the 3CI sample. Partial alignment for RDC data collection was achieved via addition of a penta-ethylene glycol mono-decyl ether:hexanol mixture (Ruckert and Otting, 2000). SLD RNA was purchased from IDT, dissolved in the same buffer as the protein, adjusted to pH 6.0 and annealed. Aliquots of 2 mM SLD were added to a sample containing 0.15 mM $3C^{pro}$ until the RNA to protein ratio was 3:1 to create the 3CIR sample (3CI plus SLD RNA).

4.2.2 X-ray scattering sample preparation

Samples used for X-ray scattering were dialyzed overnight against 50 mM MES, 0.5 mM EDTA, 1 mM DTT. The MES buffer solution contained pure water (MilliQ) treated with diethyl-pyrocarbonate (DEPC) to remove RNase activity and adjusted to pH 6.0 with 50 mM bis-tris methane. The dialysate was retained for solvent blank measurements. The post-dialysis concentration for each sample was determined by UV absorbance using E_{260nm} = 22.83 and E_{280nm} = 0.298 for 1 mg.ml^{-1} solutions with a 10 mm path length for the SLD and 3C protease, respectively. The sample concentrations were: $3C^{pro}$, 0.765 mg/ml; SLD, 0.456 mg/ml; $3C^{pro}$- SLD complex, 1.07 mg/ml. A lysozyme standard was prepared by dissolving and dialyzing the lyophilized protein (USB Corporation) in 150 mM NaCl, 40 mM sodium acetate (pH 3.8).

4.2.3 NMR spectroscopy

NMR experiments were performed on a 700 MHz Bruker Avance spectrometer equipped with a cryoprobe, four r.f. channels and gradient pulse capabilities. In general, spectra were acquired at 22 °C using a WATERGATE solvent suppression scheme and processed in Topspin 2.0 (Bruker). The backbone resonances of 3CI were assigned via ^{15}N,^{1}H-HSQC, HNCA (Kay *et al.*, 1990; Grzesiek and Bax, 1992), HNCO (Kay *et al.*, 1990; Grzesiek and Bax, 1992; Muhandiram and Kay, 1994), HN(CA)CO (Clubb *et al.*, 1992), CBCA(CO)NH (Grzesiek and Bax, 1993), and CBCANH (Grzesiek and Bax, 1992) spectra using XEASY (Bartels *et al.*, 1995) analysis software. In order to obtain the HN chemical shift perturbation values, ^{15}N,^{1}H-HSQC-TROSY (Pervushin *et al.*, 1998) spectra were acquired with the 3CI and 3CIR samples. TROSY versions of the HSQC, HNCA (Eletsky *et al.*, 2001; Salzmann *et al.*, 1998) and HNCO (Permi *et al.*, 2000) experiments were used to assign the 3CIR backbone resonances via the software package Analysis (Vranken *et al.*, 2005).

4.2.4 Calculation of chemical shift indexing and perturbation

By combining ^{15}N,^{1}H-HSQC, HNCO and HNCA data, a single chemical shift perturbation (CSP) value on addition of RNA was calculated for each residue. Chemical Shift Indexing (CSI) (Wishart and Sykes, 1994) was performed using the C_a and C-carbonyl shifts from 3CI and 3CIR.

4.2.5 Docking with HADDOCK

HADDOCK (Dominguez *et al.*, 2003) is a molecular docking program which makes use of NMR restraints as well as charge and shape complementarity. Data from chemical shift perturbation studies can be used to drive the docking process. Ambiguous interaction restraints (AIRs) are defined such that any surface residue that is perturbed over a threshold value (in this case the average CSP for the whole molecule) is "active" and any neighbouring surface residue is "passive". Restraints are satisfied in the complex structure if each of these residues is within a specified distance of an active or passive residue (or in this case base) of the docking partner. Although, in principle, any structure calculation program could be used for the HADDOCK protocol, the default program, CNS (Brünger *et al.*, 1998; Brünger, 2007) was used in this study. HADDOCK was used to dock the $3C^{pro}$ to SLD. However, the HADDOCK results conflicted with the SAXS data and further interaction studies were based on this instead. It is possible that the restraints available for the HADDOCK docking were insufficient to restrain the interaction to a single conformation.

4.2.6 Small-angle X-ray scattering data acquisition, reduction, and analysis

Small-angle X-ray scattering data were acquired, reduced and analysed using an Anton Paar SAXSess with line-collimation and CCD detector as described previously (Jeffries *et al.*, 2008). Sample exposure times were 60 minutes at 20 °C using a 10 mm slit and 10 mm integration length. Data reduction to $I(q)$ versus q (where $q = (4\pi\sin\theta)/\lambda$, 2θ is the scattering angle and λ = 1.54 Å, the wavelength of the radiation) and solvent subtractions were performed using SAXSquant1D (Anton-Paar, Austria) and the desmearing calculations were performed using GNOM (Semenyuk and Svergun, 1991). Radius of gyration (R_g) and zero-angle scattering (I (0)) values were initially estimated from Guinier analysis (Guinier and Fournet, 1955) using PRIMUS (Konarev *et al.*, 2003). Indirect Fourier transformations of I (q) were performed using GNOM and GIFT (Bergmann *et al.*, 2000) to calculate P (r) versus r and associated values for I (0), R_g, and maximum dimension ($Dmax$). The molecular weights of the scattering particles in solution of the $3C^{pro}$, SLD, and $3C^{pro}$– SLD complex were determined using the method of Glatter and the equation (Orthaber *et al.*, 2000):

$$MW = [I(0).N_A]/[c.\Delta\rho M^2] \quad (4.1)$$

where N_A is Avogadro's number and c is the concentration of the scattering particle in g/cm^3, $\Delta\rho M^2$ is the square of the scattering contrast per mass and is calculated as $\Delta\rho M^2 = (\Delta\rho\nu)^2$ where $\Delta\rho$ is the difference in scattering length density between the particle and solvent (termed the contrast and given in cm^{-2}), and ν is the partial specific volume of the scattering particle (in cm^3/g). Values for $\Delta\rho$ and ν were calculated using the program suite MULCh (Whitten *et al.*, 2008) and the NucProt Parameter set (Voss and Gerstein, 2005), respectively.

4.2.7 Shape restoration calculations from scattering data

Ab initio shape restoration calculations from experimental scattering data for the $3C^{pro}$– SLD complex were performed using MONSA (Svergun, 2000) and the smoothed $I(q)$ profiles output for the complex and the $3C^{pro}$ in the GNOM $P(r)$ calculations and an SLD R_g constraint, calculated for the NMR structure, while the experimentally determined R_g values were used as constraints for the $3C^{pro}$ and $3C^{pro}$–SLD. MONSA calculations were performed for the complex and evaluated for consistency using DAMAVER (Volkov and Svergun, 2003). The mean normalized spatial discrepancy between the shapes was 0.45 ± 0.01 and 0.51 ± 0.01 for the $3C^{pro}$ and SLD, respectively, indicating that all restored shape solutions were very similar.

4.2.8 Rigid-body refinement of models against the scattering data

Rigid body refinement of the positions and orientations of atomic models for the $3C^{pro}$ and SLD in the $3C^{pro}$–SLD complex was done using SASREF6 (Petoukhov and Svergun, 2005) using the RDC-refined NMR structure of HRV-14 $3C^{pro}$ and each of the 9 different NMR structures within the SLD structure ensemble. These SLD structures differ primarily in the details of the tri-loop conformation. Ten independent calculations were performed for each combination of $3C^{pro}$ and one of the SLD structures. The SLD component was constrained to have at least one nucleotide within 10 Å of the Cα atom of the $3C^{pro}$ residues strongly implicated in RNA binding by at least two of the following criteria: mutagenesis, chemical shift perturbation, signal loss, signal doubling. This distance constraint was set to 20 Å for more moderately implicated residues to allow for the possibility that residues not directly contacting the binding interface region may be affected by binding, but typically to a lesser extent. The best model was selected based on the fit to the scattering data determined by the minimum χ^2 value, which for an ideal fit is 1.0. The best fit model was energy minimized (CNS (Brünger, 2007; Brünger *et al.*, 1998)) to relieve steric clashes.

4.3 Results

4.3.1 NMR data

The chemical shifts and peak intensities in NMR spectra of isotopically-labelled $3C^{pro}$ in the absence and presence of SLD RNA were compared. Spectral quality was improved by inclusion of the active site inhibitor JMC-98-3 (Webber *et al.*, 1998) as described in Chapter 2. A homology model of the HRV 14 $3C^{pro}$ in the presence of the inhibitor was created as described in Chapter 2. As mentioned in Chapter 2 the HRV14 3C protease consists of 182 residues and a total of 156 backbone amide peaks were assigned in the 3CI spectrum. In addition to the six prolines, twenty residues did not produce identifiable backbone amide peaks. However, the unassigned residues are sequentially isolated with the exception of the 103–112 loop that is adjacent to the proteolytic active site and is not implicated in RNA-binding (Matthews *et al.*, 1994). Thus, the assigned peaks provide a suitable sampling of the potential interactions with RNA.

Initial work was carried out on ^{15}N,^{13}C-labelled 3CI. The ^{15}N,^{1}H-HSQC spectrum resulting from an addition of 1.2:1 RNA to protein molar ratio is shown in Figure 4.1. Line broadening is evident in the red spectrum as a result of increased molecular weight of the complex in comparison to the protein alone. Although some peaks have shifted in the complex spectrum in comparison to the uncomplexed, a number have also disappeared and many of the shifted peaks are not possible to assign due to peak overlap or due to movement sufficiently far from their origin as to make assignment ambiguous. The residues with the greatest chemical-shift perturbation, and those that have amide peaks that disappear upon RNA binding, are plotted onto the 3CI structure (Figure 4.3). A large number of residues are painted brown, indicating that there was too much spectral overlap to reliably identify the peak for that residue. Several residues had peaks that disappeared completely after addition of SLD. α-helical or loop regions in proteins often show less dispersed peaks than β-strands, as the interaction residues are primarily in this type of secondary structure, peak overlap was a major problem. It was decided that the incorporation of ^{2}H into the 3CI sample would reduce the line broadening and increase resolution of peaks in the crowded middle of the spectrum. In addition, it was necessary to produce ^{13}C,^{15}N-labelled protein to assign the backbone resonances in complex. A sample of deuterated, ^{15}N-labelled 3C was produced to test the tolerance of the expression system to deuteration. As this worked well, triple-labelled protein was produced and used for assignment of the complex backbone shifts.

Unlabelled SLD RNA was titrated into a sample of ^{2}H,^{15}N,^{13}C-labelled 3CI with amide pro-

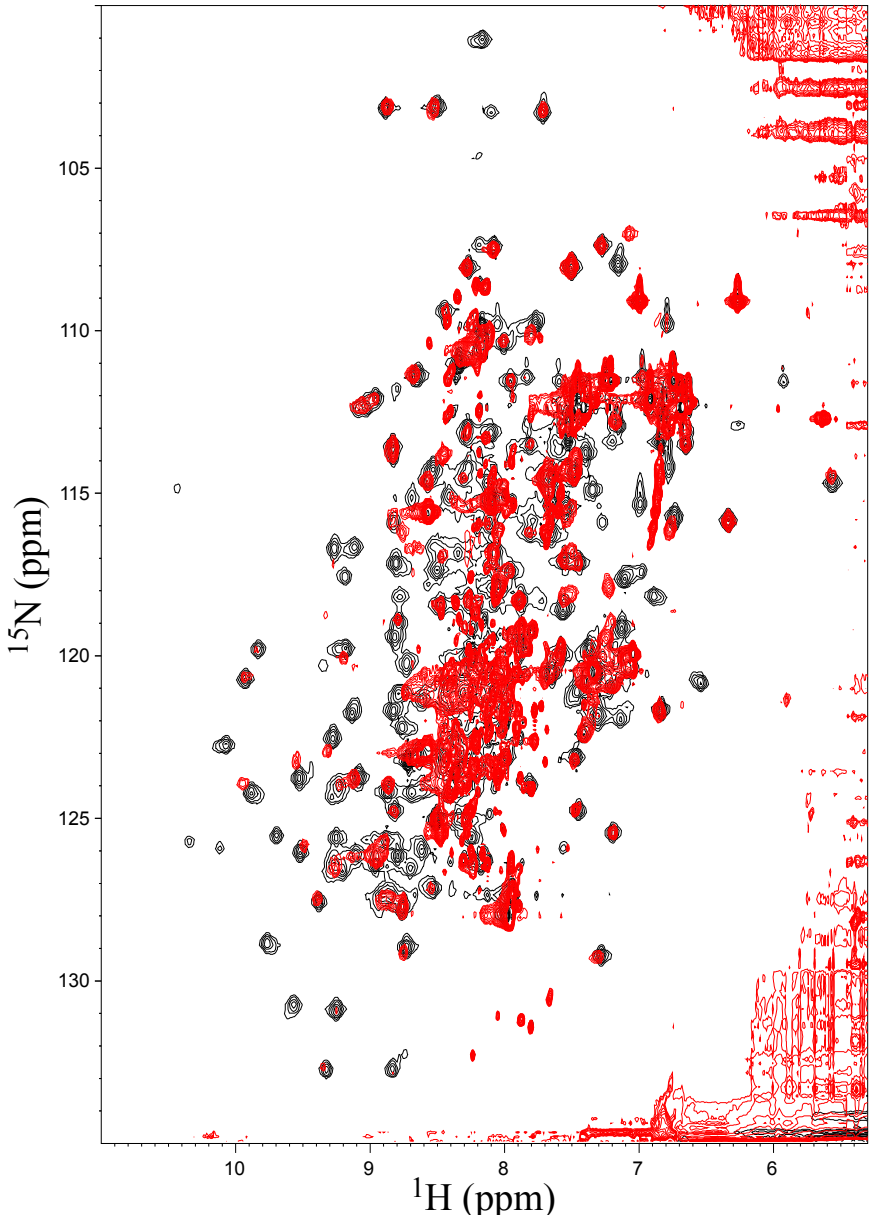


Figure 4.1: ^{15}N,^{1}H-HSQC of 0.5 mM 3CI in complex with 1.2 fold molar amount of 27mer (red) overlayed with 3CI alone (black). Both spectra were carried out at 22 °C and the samples were in 50 mM BTP buffer pH 6.0 15 mM DTT 0.5 mM EDTA.

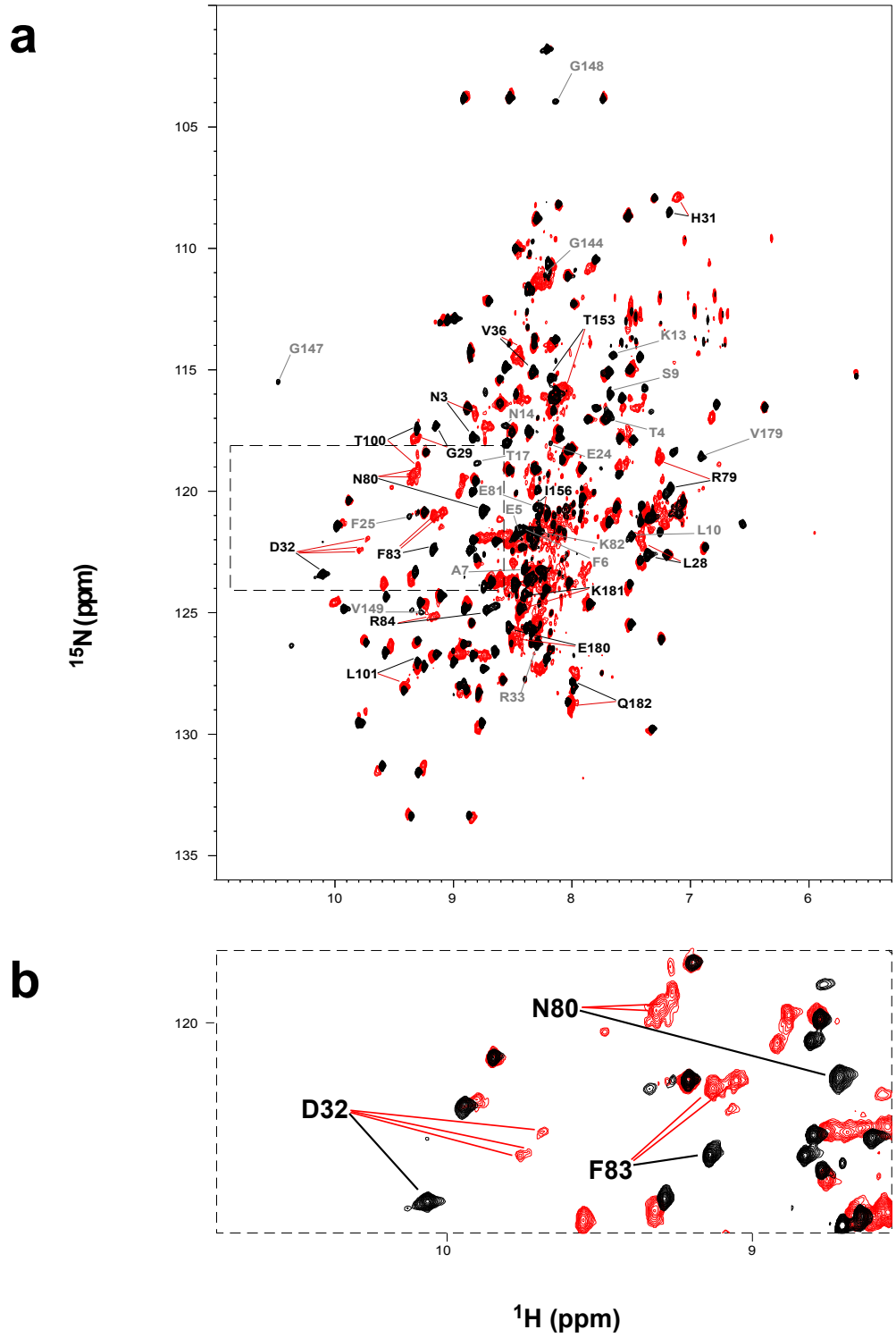


Figure 4.2: Overlay of 3CI/3CIR ^{15}N,^{1}H-HSQC-TROSY. **(a)** shows the overall spectrum and **(b)** shows a close up of a region of the spectrum that contains the three residues that exhibit multiple conformations in 3CIR

tons restored via use of H_2O-containing buffer and the RNA:protein ratio was increased to 3:1 to ensure full complexation of the protein component. NMR spectra obtained at 1:1 and 3:1 molar ratios are similar, clearly identifying the binding stoichiometry as 1:1. Subsequent NMR analysis of 3CIR (used to designate 3CI in the presence of unlabelled SLD-RNA) was performed with the 3:1 sample. Changes induced in the 3CI ^{15}N,^{1}H-HSQC-TROSY spectrum upon addition of unlabelled SLD RNA are shown in Figure 4.2. The 3CIR spectrum produced 139 assignable peaks, with peaks from 19 additional residues unassigned either due to exchange broadening or shifting of peaks into crowded regions of the spectra. Conformational heterogeneity at D32, N80 and F83 (Figure 4.2) is consistent with the presence of additional local conformational exchange processes at the protein/RNA interface.

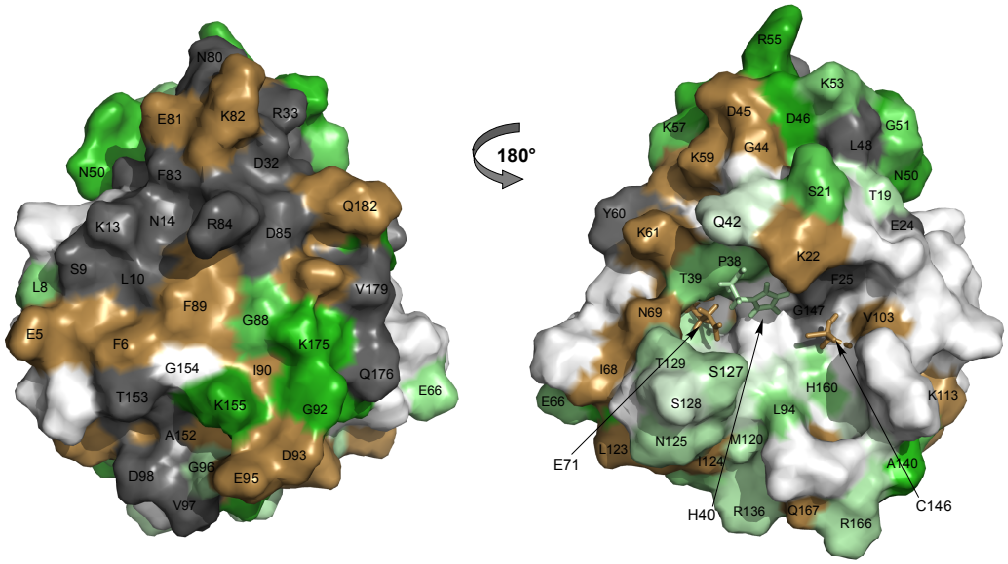


Figure 4.3: CSP from complex of protonated $3C^{pro}$–SLD. Residues are coloured with increasing intensity of green for increasing CSP. Residues coloured brown were unable to be used in the analysis due to peak overlap. Residues coloured grey exhibit disappearance of amide peaks upon RNA binding.

The secondary structure of the 3CI RDC-refined homology model was extracted via Procheck (Laskowski *et al.*, 1993) (shown above Figure 4.4). The results are consistent with previous analyses (Bjorndahl *et al.*, 2007). 3CI consists of an N-terminal α-helix followed by two six- stranded β-barrels. Strand 5 and 5′ surround a one amino acid irregularity that may or may not actually

be present. If combined, these two strands would form the canonical strand 5. The presence of a partial helical turn from K175 to Y177 was not detected by Procheck but was displayed by Pymol using default parameters and thus appears in our figures. Various loops separate the strands. The most extensive interstrand interval occurs between strands 6 and 7 and includes residues T76 through to V97. This lengthy loop, situated along the face opposite to the proteolytic active site, contains a short α-helix and serves as a linker between the N-terminal and C-terminal β-barrel domains (Chapter 1).

An alternative secondary structure analysis of 3CI on the basis of chemical shift indexing (CSI) (Wishart and Sykes, 1994) was performed (Figure 4.4(a)). Contiguous significant positive CSI values implicate α-helix, while contiguous negative stretches indicate β-strand. The Procheck and CSI modes of secondary structure analysis, performed independently, produce largely consistent results, apart from the gap in the CSI graph due to the unassigned 103-112 stretch. The chemical shift indexing analysis of $3C^{pro}$ in the 3CI SLD-RNA complex (3CIR, Figure 4.4 (b)) differs little from (3CI, Figure 4.4(a)) apart from additional concentrations of unassigned residues within the N-terminal helix and in the loop between β-strands 9 and 10. Figure 4.4 (c) is a difference plot (Figure 4.4 (b) minus Figure 4.4 (a)). The largest contiguously changing stretch occurs in the linker region connecting the two β-barrels and contains three residues (K82, R84, D85) implicated in RNA-binding via mutational analysis (Leong *et al.*, 1993; Andino *et al.*, 1993). The CSI changes alternate in sign in this region, suggesting that little regular secondary structure is formed. Nonetheless the magnitude and continuity of change suggest a significant alteration in backbone conformation. Residues 116–120 within β-strand 8 and residues 150–163 spanning β- strands 10 and 11 also show significant CSI change, in each case largely consistent with a reduction in β-strand character. Changes in strand 8 (shown along the bottom of Figure 1.8) could indicate allosteric effects as this strand does not appear to approach the RNA-binding site (see below). The loop between strands 10 and 11 contains three residues (T153, G154, K155) implicated in RNA-binding (Matthews *et al.*, 1994; Andino *et al.*, 1993). In confirmation of this analysis, an alternative implementation of CSI was run on the Pecan webserver (Eghbalnia *et al.*, 2005). Overall, the plots for 3CI and 3CIR are very similar (with the exception of the unobservable shifts for the N-terminal α-helix) but localised changes are evident in the residues around 30, from 80–95 and around 155. These regions are all implicated in RNA-binding.

A plot of chemical shift perturbations introduced by addition of SLD is presented in Figure 4.6. Each black bar is a weighted average of ^{1}H, ^{15}N and ^{13}C shifts for that residue while grey

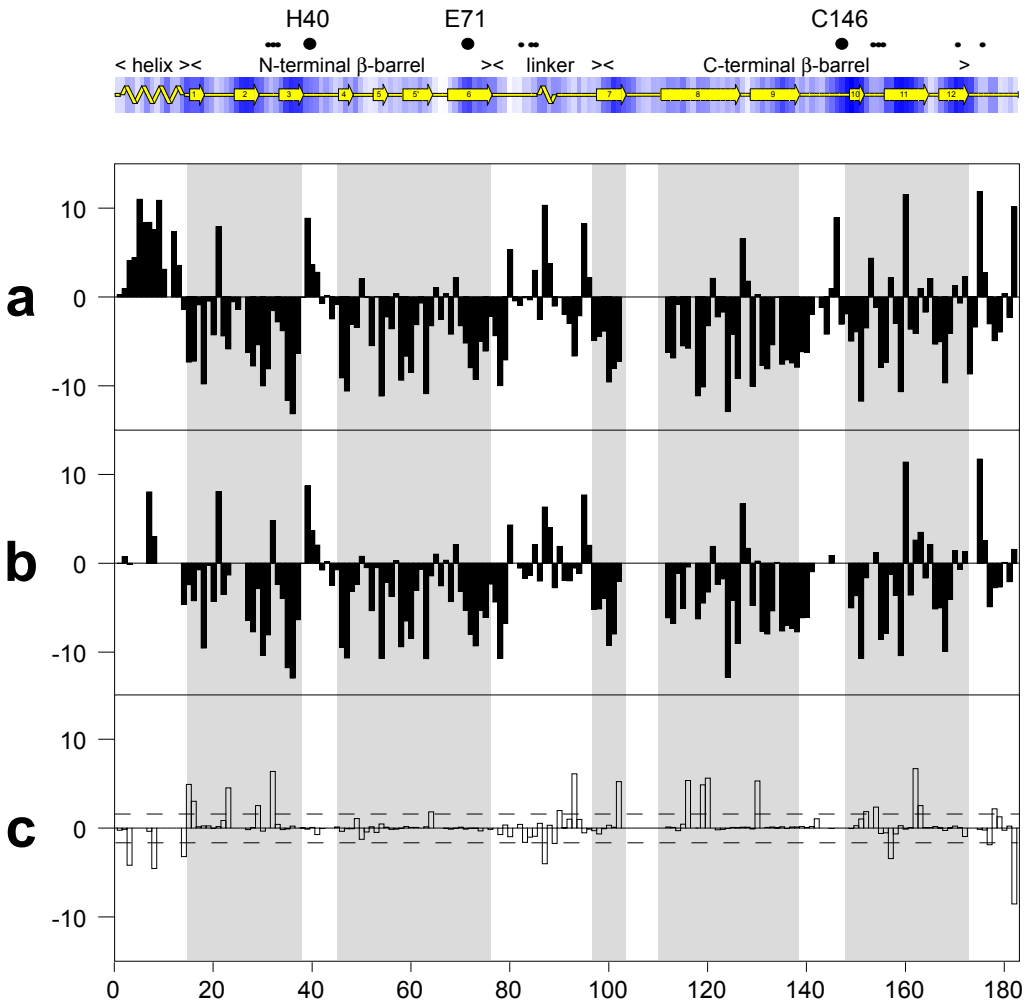


Figure 4.4: CSI plots for **(a)** 3CI, **(b)** 3C and **(c)** a difference plot between the two. More negative values indicate β-strand tendency and more positive values indicate α-helical tendency. Random coil will give near zero values.

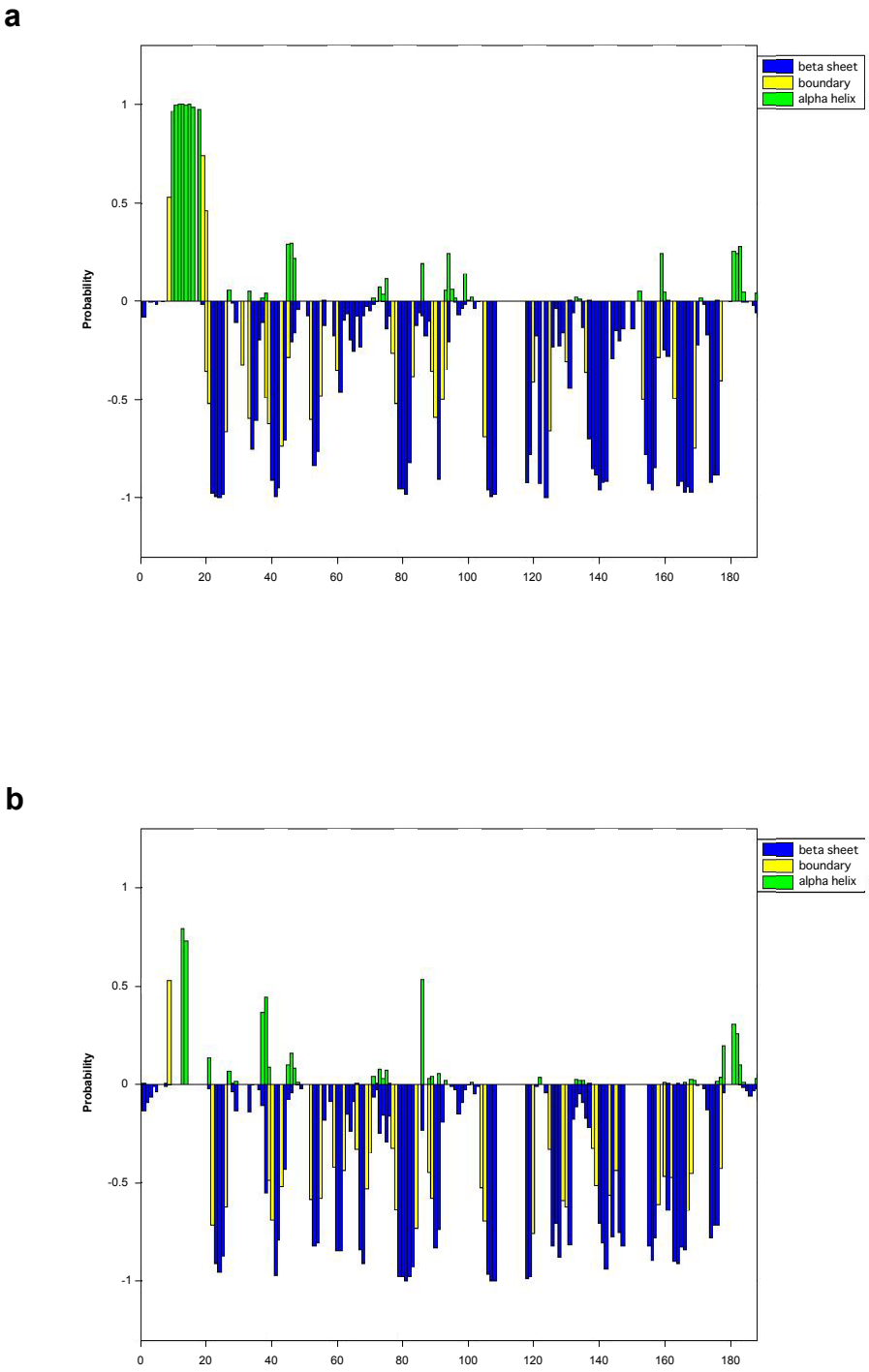


Figure 4.5: (a) 3CI (b) 3CIR results from the Pecan webserver. Pecan is an implementation of CSI. A value near -1 indicates β-strand and a value near 1 indicates α-helix.

bars indicate residues for which 3CI shifts were assigned but 3CIR shifts were not. The largest concentration of grey bars occurs within the N-terminal α-helix and in the loop between β-strands 9 and 10. The largest measurable perturbations occur in the long central linker region connecting the two β-barrel domains of the protein, with significant shifts also noted in β-strands 1–3 and 10–11 and near the C-terminus. Note that the localised CSI changes (Figure 4.4(c)) correlate well with regions of large chemical shift change (Figure 4.6).

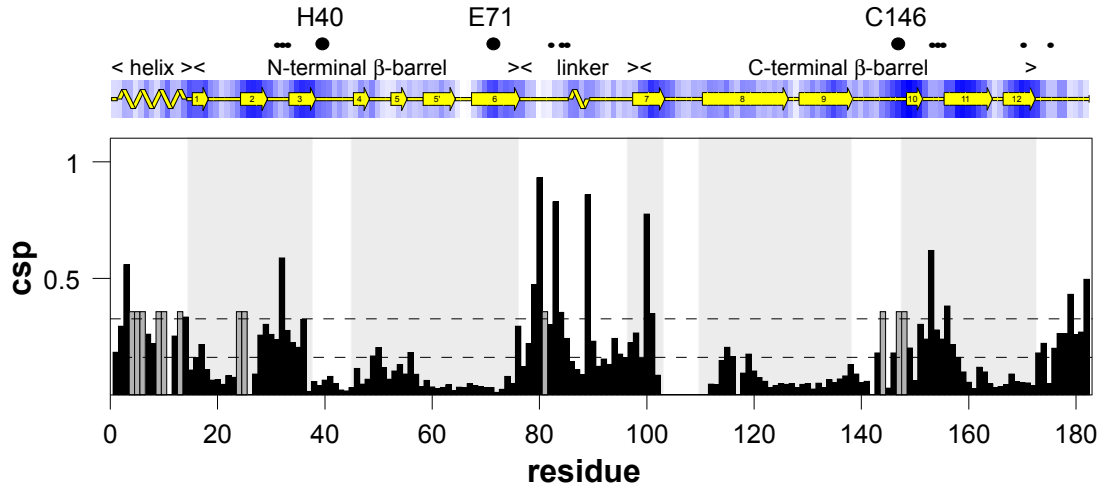


Figure 4.6: Chemical shift perturbation of 3CI on addition of a 1:1.2 molar ratio of SLD. The lower dashed line represents one s.d. above average perturbation and the upper dashed line represents two s.d. above average perturbation. Light grey bars represent residues for which no chemical shift assignments could be made in the bound state.

The largest SLD-induced chemical shift perturbations on the $3C^{pro}$ surface cluster primarily to a patch (dark red in Figure 4.7) that includes D32 from the loop connecting β-strands 2-3, N80, F83 and F89 from the inter-domain linker and V179 from the C-terminal region. Much smaller perturbations are observed on the opposite face of the protein where the catalytic site is located (Figure 4.7, right hand view). Residues for which peaks become unobservable upon addition of SLD are shown in grey and include several amino acids from the N-terminal helix and E81 from the linker, as well as F25 and G144 from the catalytic face.

Changes in intensity of HSQC-TROSY peaks upon addition of SLD were monitored as an additional indication of changes in chemical environment. Dark green patches in Figure 4.8 indicate a large reduction in intensity for resonances assigned to residues in the inter-domain linker, specifically R84, D85, G88 and F89, as well as for N50 from the loop connecting β- strands 4-5. The slight intensity increase for resonances arising from residues I90 and G96 (yellow) could indicate an increase in mobility in the C-terminal half of the linker upon SLD- binding. The three

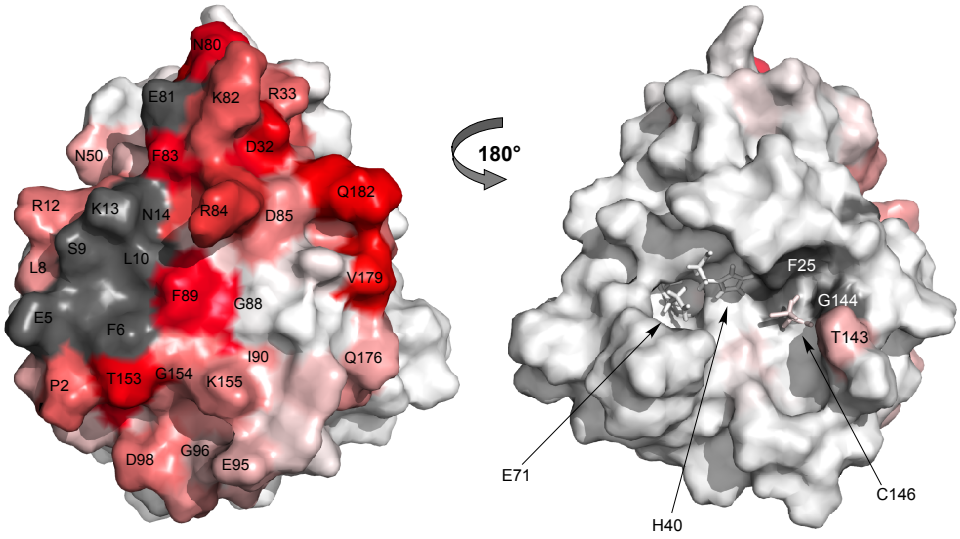


Figure 4.7: Surface rendering of the three-dimensional structure of 3C^{pro} showing the chemical shift perturbation upon binding SLD RNA. Increasing intensity of red corresponds to greater chemical shift change. Residues that could not be assigned in the RNA-bound protein are shown in grey

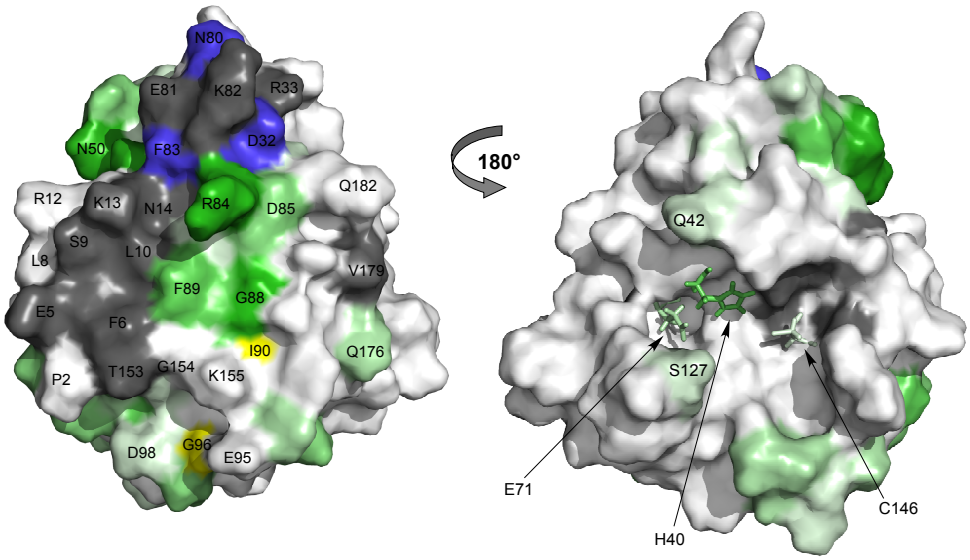


Figure 4.8: Surface rendering of the three-dimensional structure of 3C^{pro} showing the change in HSQC-TROSY peak intensity upon binding SLD RNA (**Figure 4.2**). Increasing intensity of green corresponds to greater reduction in intensity. Increases in intensity are shown in yellow. Residues with no HSQC-TROSY peak in the RNA-bound protein are shown in grey. Navy blue indicates the presence of multiple HSQC-TROSY peaks (**Figure4.2(b)**).

residues giving rise to multiple peaks (Figure 4.2) also produce significantly decreased intensity (dark blue in Figure 4.8). Grey patches in Figure 4.8 represent residues for which the 3CIR HSQC-TROSY peak could not be assigned and include all grey patches from Figure 4.7 in addition to N14, R33, K82 and T153. The resonances arising from residues at the catalytic face (Figure 4.8, right hand view) produce far fewer significant changes in intensity.

4.3.2 Analysis of NMR results for use with SAXS refinement

The analyses summarized in Figures 4.7 and 4.8 suggest that the top central region of the 3CI face opposite to the proteolytic active site is a key part of the SLD binding interface. This result is consistent with pre-existing mutational data (Andino *et al.*, 1993; Leong *et al.*, 1993)) and sequence conservation analysis. The KFRDIR motif (residues 82–87) from the inter-domain linker (residues 76-97) is conserved across picornaviral 3C sequences and has been implicated in RNA binding (Leong *et al.*, 1993; Shih *et al.*, 2004; Gorbalenya *et al.*, 1989) From this linker, N80, K82, F83, R84, D85 and F89 each exhibit large chemical shift changes upon complex formation while N80 and F83 produce multiple peaks, E81 and K82 peaks become unobservable and R84, D85, G88 and F89 peaks reduce significantly in intensity. Residues I86 and R87, which complete the KFRDIR motif but are partially buried along the back side of the small yellow helix in Figure 1.10 (b), show little change on binding.

The NMR data also suggest the N-terminal α-helix (residues 2–14) is involved in RNA binding, as many of the corresponding TROSY peaks are shifted or broadened in the 3CIR spectrum. Strands 2–3 and the intervening loop also appear to be involved. This loop includes D32, which, as noted above, produces three 3CIR HSQC-TROSY peaks (Figure 4.2(b)) and R33. Residues from strand 10 to the midpoint of strand 11, a region that connects C146 of the catalytic triad through the protein core to the RNA-binding face, also display marked changes between the free and bound form. This region spans residues 153–155 which have been implicated in RNA binding by mutagenesis studies in poliovirus (Andino *et al.*, 1993). The C-terminus exhibits above average shift changes in residues 176–182. Somewhat inconsistently with the same work, we do not observe significant perturbation at K175 upon RNA binding.

Taken together, these results show that a surface region representing several non-sequential segments of 3CI interacts with SLD RNA. The linker region from residues 76–97 that contains the conserved KFRDIR motif is most strongly implicated, with the N-terminal helix and β-strand and loop regions from 25–38 and 145–155 also playing a role. The C-terminal residues are also

perturbed. However, NMR peak intensities indicate high mobility of the C-terminus. Therefore we have limited the NMR/mutation-implicated surface used for SAXS analysis to $3C^{pro}$ residues 5–6, 9–10, 13, 32–33, 80–85, 89 and 153–155, bearing in mind that the C-terminus may fold toward the interface upon complex formation. Of these residues, D32, N80, F83, R84 and F89 were treated as strongly-implicated in RNA-binding (see section 4.3.1).

4.3.3 SAXS data

Due to its phosphate backbone, RNA has a higher electron density and hence X-ray scattering density compared to that of a protein. The small-angle scattering from the $3C^{pro}$–SLD is therefore sensitive to the positions and orientations of RNA and protein components. Scattering data were acquired for the $3C^{pro}$, SLD and the $3C^{pro}$-SLD–RNA complex (Figure 4.9). These data yield basic parameters such as molecular weight, radius of gyration, R_g , and the contrast-weighted distribution of interatomic distances, P (r), within the scattering particle (Supplementary Table 1). The experimentally determined molecular weights for $3C^{pro}$ and $3C^{pro}$–SLD, as well as a lysozyme standard (determined using the measured I (0) values and equation (1)) agree well with the expected values based on amino acid and nucleotide sequences and the P(r)-derived R_g value for $3C^{pro}$ agrees with that expected from the NMR structure (16.7 Å, calculated using the program CRYSOL (Svergun and Nierhaus, 2000)). In contrast, the SLD molecular weight determined from I (0) was approximately twice the expected value indicating that under our experimental conditions the SLD-RNA is not a monomer in solution. Nonetheless, consistent with the NMR results, by mixing equimolar amounts of $3C^{pro}$ to SLD a 1:1 $3C^{pro}$–SLD complex formed with the expected experimentally derived scattering particle molecular weight.

Figure 4.14(a) shows the molecular envelope for the $3C^{pro}$–SLD complex calculated by *ab initio* shape restoration from the scattering data superimposed with an atomic model obtained by rigid body refinement of the positions and orientations of NMR-derived $3C^{pro}$ and SLD component structures against the scattering data. The agreement between these two models calculated using different methods is excellent. The shape restoration model (transparent surface in Figure 4.14(a).) fits the scattering data with a statistically ideal χ^2 value of 1.00 (Figure 4.10). The molecular envelope of the complex has dimensions of approximately 80 Åx 50 Åx 50 Åwith the SLD component extending away from the surface of the $3C^{pro}$ at an angle of ~ 30°. In turn, the best-fit structure from the rigid body refinement of the component structures (ribbon representation, Figure 4.14(a)) fits the scattering data with a χ^2 value of 1.26. This fit, although very

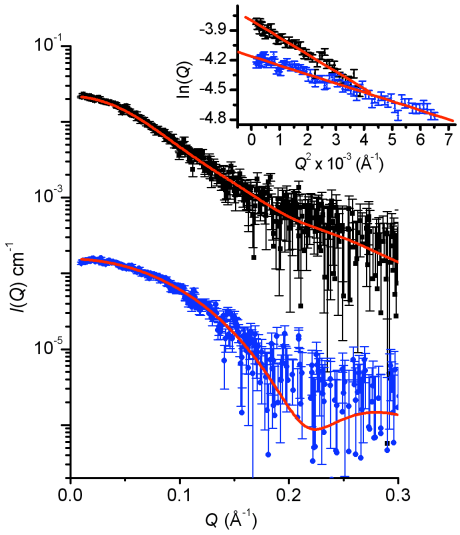


Figure 4.9: Scattering data for 3Cpro and its complex with SLD. Desmeared SAXS profiles for the $3C^{pro}$ (blue) and $3C^{pro}$–SLD complex (black) with the former profile shifted on the vertical axis for clarity. The predicted scattering profiles for the MONSA model for the $3C^{pro}$–SLD complex and the NMR structure of $3C^{pro}$ are shown as red lines superposed with the corresponding experimental scattering data. The inset shows the expected linear Guinier plots for samples that are monodisperse, identical particles.

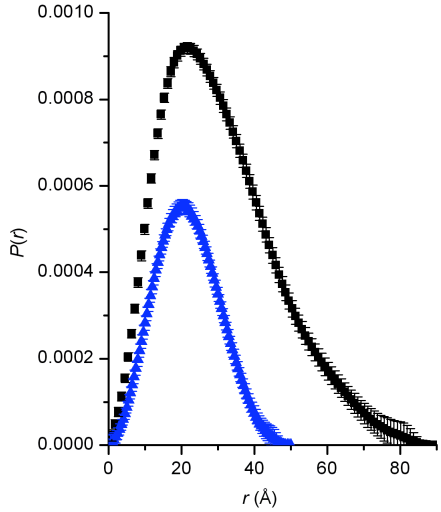


Figure 4.10: *P(r)* profiles calculated from the scattering data for $3C^{pro}$ (blue) with a D_{max} of 50 Åand $3C^{pro}$–SLD (black) with a D_{max} of 80 Å.

good, is somewhat higher than ideal indicating small differences between the conformation of SLD and/or $3C^{pro}$ in the complex compared with the individual component structures used in the calculation. Furthermore, by taking the protein component of the shape restoration model and comparing its predicted scattering profile to the experimental scattering data for the uncomplexed $3C^{pro}$, we find that the fit is also less than ideal (χ^2 = 1.44). This comparison provides further evidence that the $3C^{pro}$ undergoes a small conformational change upon forming a complex with SLD.

The SAXS and NMR data both indicate the formation of a 1:1 complex between $3C^{pro}$ and SLD. The SAXS data provide information on the global shape of the complex and the positions and orientations of the protein and RNA components. When combined with constraints from the NMR data concerning the contacts between the protein and RNA, we are able to optimize the positions and orientations of these two components in the 1:1 complex.

In the structural model (Figures 4.14 and 4.11), the triloop end of the SLD nestles between the end of the $3C^{pro}$ N-terminal α-helix and the inter-domain linker that contains the conserved KFRDIR motif. Contact residues of $3C^{pro}$ include L10 and R12–M16 from the N-terminal helix and the following strand and G27-R33 and C35 from β-strands 2 and 3. Residues V49 and Q52 from the loop connecting strands 4 and 5 and L77, R79–R84 and I86 from the linker that connects the two β- barrels of the protein also contact the SLD RNA. No residues from the C-terminal β-barrel directly contact the RNA in this model. SLD nucleotides from uracil-13 to uracil-17 contact the protein. This stretch includes the triloop and two bases, guanine-16 and uracil-17, from the 3′- side of stem II. Interestingly, the ^{31}P chemical shift values of guanine-16 and uracil-17 in the uncomplexed SLD are atypical, indicating a significant departure from standard A-form geometry (Headey *et al.*, 2006). The intermolecular interface covers an area of 653 Å^2 and lies within the range of typical RNA-protein interaction interface areas (Johnson and Donaldson, 2006; Allain *et al.*, 2000).

The SLD also closely approaches the loop that connects β-strands 2 and 3 (H31–R33). The H31–R33 loop is spatially positioned between residues R79–R84 of the inter-domain linker and the C-terminus of the protein. The N-terminal α-helix may become sufficiently mobile to approach the nearby SLD stem region, possibly explaining the loss of the corresponding peaks in the 3CIR NMR spectra if the association is transient or heterogeneous. A combination of electrostatic and hydrophobic interactions play a role in complex formation. In addition to R84, other basic residues near the binding surface that may contact the SLD phosphodiester backbone in-

clude R12, K13, R33, R79 and K82. The F6, F83 and F89 residues are exposed in the uncomplexed protease, and may stabilize the interaction via hydrophobic interactions such as base stacking. In the present model, the highly conserved F83 is particularly well situated for interaction with the triloop bases. The V49, L77 and I86 side chains could make additional hydrophobic contacts, as may the aliphatic regions of the many lysine and arginine residues in the area.

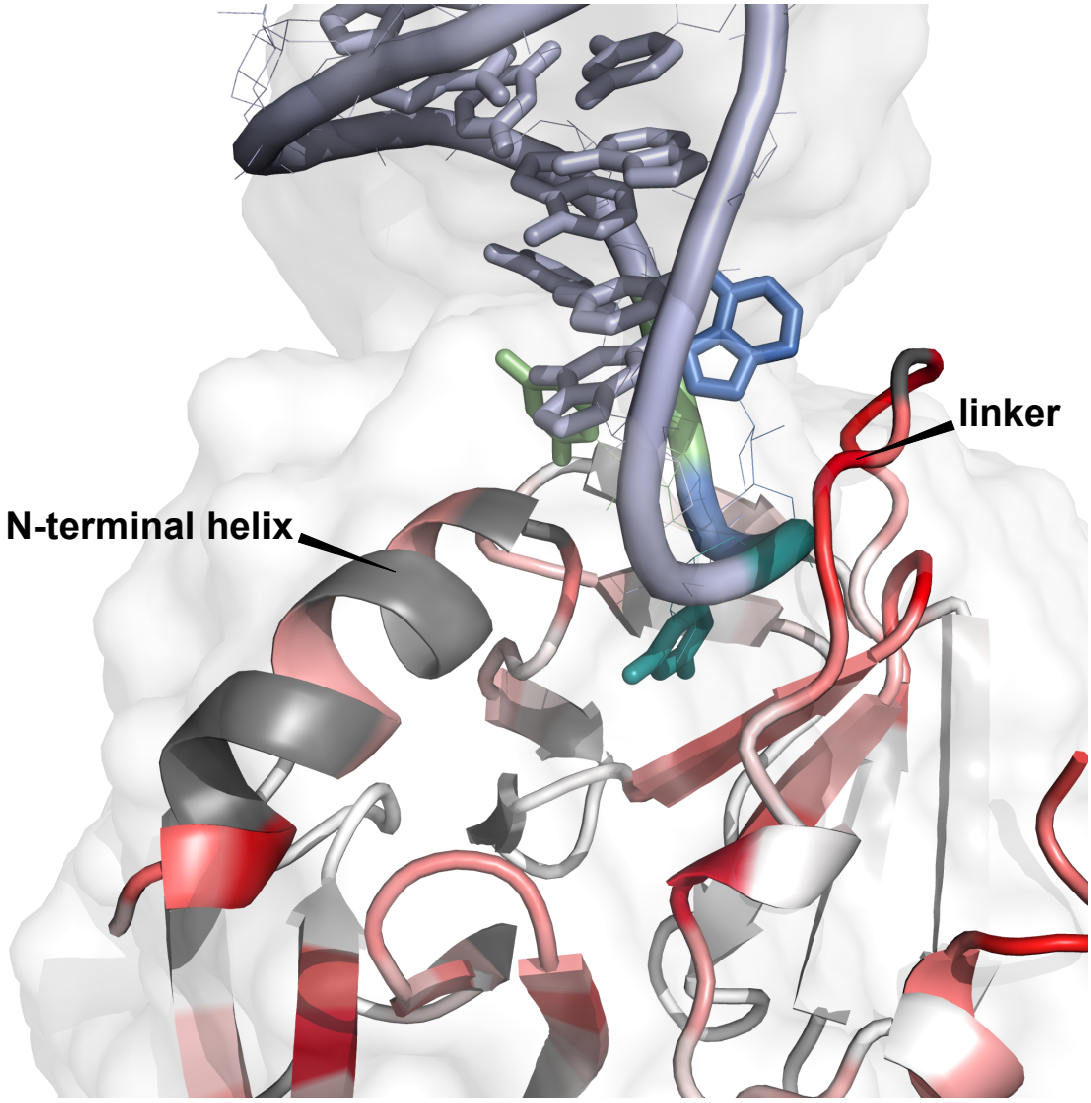


Figure 4.11: Interaction between $3C^{pro}$ and stemloop-d. The bases are coloured as in Figure 1.11. Residues in $3C^{pro}$ are coloured according to the scheme in Figure 4.7. The close proximity of the linker loop and the triloop is clearly apparent. It is possible that the N-terminal helix (on the left of the figure) can interact with the RNA.

A conserved domain database alignment of ten diverse $3C^{pro}$ and $3C^{pro}$-like proteases (alignment pfam00548; (Marchler-Bauer *et al.*, 2007)) was used to examine conserved residues in the contact regions. Of the above RNA-contacting residues, conserved basic positions include: R12

(50% conserved within alignment), R79 (70%), K82 (70%) and R84 (100%). Conserved hydrophobic residues include L10 (60%), I15(80%), I30(100%), V49 (90%), L77(100%), F83 (100%) and I86(90%). The N14 and G29 positions are conserved and the D32 position is acidic in 70% of the ten sequences, while the H31 position is aromatic or H in seven instances.

4.4 Discussion

RNA-protein interactions are by nature diverse due to the variety of secondary and tertiary structures that each of these biopolymers is capable of adopting. The $3C^{pro}$ tertiary structure does not appear to fit neatly into one of the previously defined classes of RNA-binding domains. However, some commonalities are apparent. For instance, we have shown that the linker connecting the two β-barrel domains of $3C^{pro}$, which lacks regular secondary structure and is likely to possess some degree of mobility, is a key to the interaction with SLD. Proteins containing the most prolific RNA-binding domain, dubbed simply the RNA-binding domain (RBD) or RNA recognition motif (RRM), often use the flexible linker between tandem RBDs to contact a flexible region of RNA(Allain *et al.*, 2000; Wang *et al.*, 2001). After complex formation, each of these flexible regions changes conformation, typically becoming more highly ordered (Bouvet *et al.*, 2001). This raises the possibility that a conformational change and an increase in rigidity may occur in the $3C^{pro}$ linker and SLD triloop. This potential rearrangement could, among other effects, help translocate the F89 aromatic ring of the connecting linker into closer contact with SLD. In addition, RBD domains possess two to three solvent-exposed aromatic bases which can contact the RNA bases (Birney *et al.*, 1993). This is consistent with the potential involvement of the exposed F83, F89 and F6 residues from $3C^{pro}$, although the effects at F6 may simply be due to an overall shift in the position of the N-terminal α-helix. Precedence here comes from the U1A-UTR complex (Oubridge *et al.*, 1995; Varani *et al.*, 2000), in which the U1A C-terminal α-helix shifts position to allow the RNA UTR fragment more complete access to the RNA-binding sheet beneath.

The non-Watson-Crick base pairs (uracil-5 • uracil-23, cytosine-6 • uracil-22 and uracil-7 • uracil-21) do not appear to be directly involved in the interaction but may play an indirect role. Ideal A-form RNA contains a deep and narrow major groove which precludes the possibility of specific recognition of the base pairs within. The minor groove side provides greater access, but the bases present an essentially uniform, non-sequence-specific face in the minor groove. Specific recognition of bases must occur via the major groove. RNA stem loops containing non-canonical base pairs can present a widened major groove, as has been shown for HRV-14 SLD (Headey *et al.*, 2006) and in other instances (Du *et al.*, 2004; Ohlenschläger *et al.*, 2004). Thus, contacts between the stem region closest to the triloop and the protease may benefit from the presence of the non-canonical base pairs further up the stem. The contacts between the triloop bases and $3C^{pro}$ have some analogy to the binding of the sex-lethal protein to a non-paired uracil-

rich tract from the drosophila *transformer* mRNA precursor (Handa *et al.*, 1999). The *transformer* uracil-rich tract is extended but is ordered, including stable sugar puckers in the atypical C2'-endo configuration. Thus it is interesting to note that each sugar of the uracil-rich stretch of implicated SLD nucleotides (uracil-13 to uracil-17) adopts primarily a C2'-endo configuration (Headey *et al.*, 2006). In summary, while the global fold of $3C^{pro}$ may be distinct from other RNA-binding proteins investigated to date, the mechanisms of interaction may be largely conserved relative to previous studies.

SLD stem I is highly conserved among rhinoviruses. Mutational work in which stem I adenine • uracil pairs were replaced with guanine • cytosine pairs suggests that this conservation is necessary to retain the ability to bind $3C^{pro}$ (Walker *et al.*, 1995). This would seem at odds with the model (Figure 4.14), in which the triloop and stem II regions contact $3C^{pro}$. On the other hand, NMR shift perturbation data for a coxsackie virus $3C^{pro}$–SLD interaction (Ohlenschläger *et al.*, 2004). suggest that the triloop and the nearby stem II elements contact the protein (Figure 4.14(b)4.11). Also, although $3C^{pro}$ from HRV-14 can bind SLD from other picornaviral species, the opposite is not true with, for example, coxsackie virus $3C^{pro}$ unable to bind HRV-14 SLD. Evidence suggests that this disparity is due to the presence of a triloop in HRV-14 SLD vs. a tetraloop in most other picornaviral SLD (Zell *et al.*, 2002). These results implicate the triloop region in binding. The two loops are shown in Figure 4.13. The first tetraloop pair almost forms a base pair making the structure more like a biloop. Thus the triloop has an 'extra' base. This could sterically hinder the interaction between coxsackie virus $3C^{pro}$ and the HRV SLD. The somewhat contradictory data regarding stem I could suggest that alterations close to the chain termini influence the global fold or stability of SLD, affecting the binding at the triloop end. Sequence conservation in stem I could also to some degree reflect other interactions/activities that are not well characterized at present.

A recent NMR-based study implicated a similar region of poliovirus $3C^{pro}$ in interactions with single-stranded RNA oligomers representing the poliovirus *oriI* element (Amero *et al.*, 2008). This study and others raise the possibility of dimerization of $3C^{pro}$ from Hepatitis A Virus (Peters *et al.*, 2004) and in poliovirus as suggested by crystal contacts Mosimann *et al.* (1997) or in the context of the Vpg-uridylation complex (Pathak *et al.*, 2007). It should be noted that unlike in HRV, the SLD-like element in Hepatitis A Virus does not occur in a cloverleaf structure (Kusov and Gauss-Müller, 1997), possibly indicating differences in the detail of the replication strategy. The present study of the HRV-14 $3C^{pro}$ interaction with SLD (also known as *oriL*) provides no

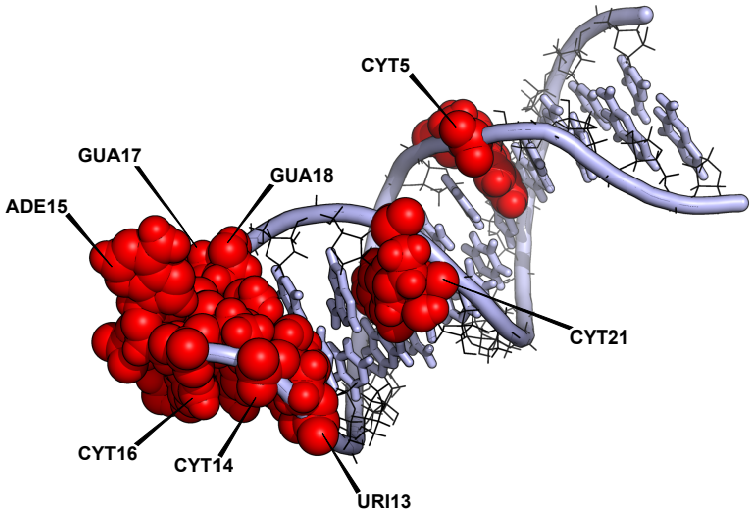


Figure 4.12: Interaction residues from Ohlenschläger and co-workers (Ohlenschläger *et al.,* 2004) plotted onto SLD from HRV14. The residues implicated in 3C-binding in that study are shown as red spheres.

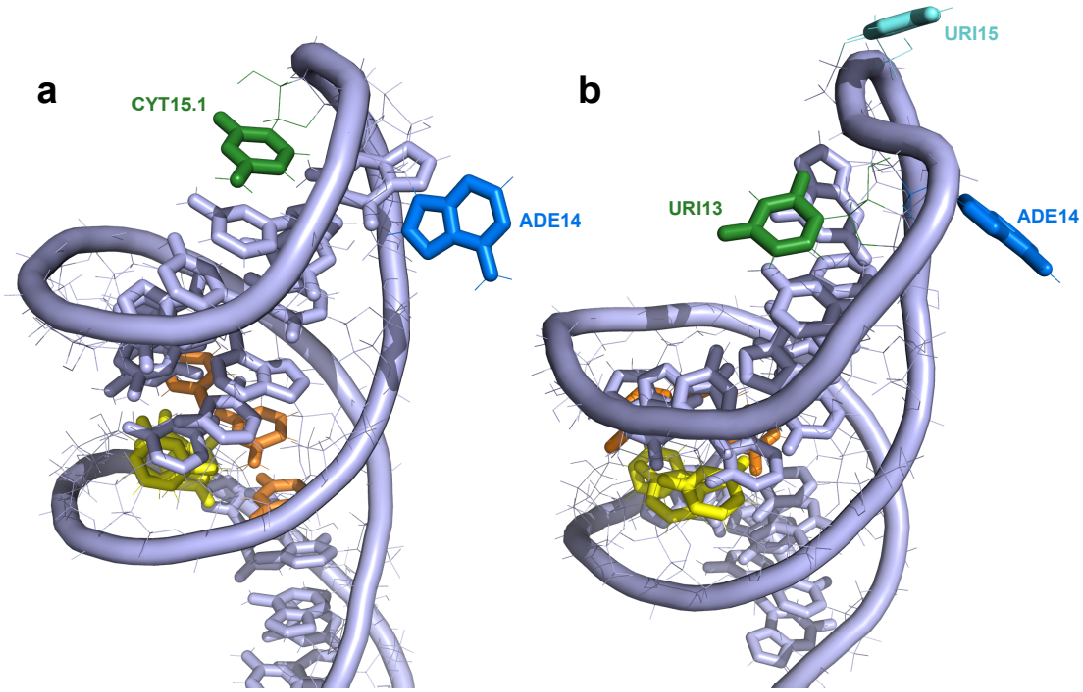


Figure 4.13: Cartoon representation of SLD from (a) coxsackie virus and (b) colouring of bases is as in Figures 1.11.

evidence of $3C^{pro}$ dimerization in the complex. Also, although the bases of the SLD triloop region are unpaired, we detect no significant disruption of base-pairing in HRV-14 SLD. Therefore, while overlapping regions of the $3C^{pro}$ surface from poliovirus and HRV may be involved in interaction with single-stranded *oriI* elements and with double-stranded SLD, respectively, the mode of interaction may differ in other aspects. Interestingly, the $3C^{pro}$ N- terminal α-helix amide chemical shifts were assigned in the poliovirus study, suggesting that the role of the N-terminal helix may differ in binding single vs. double-stranded RNA molecules. This is consistent with the possibility that the N-terminal helix interacts with the SLD stem region.

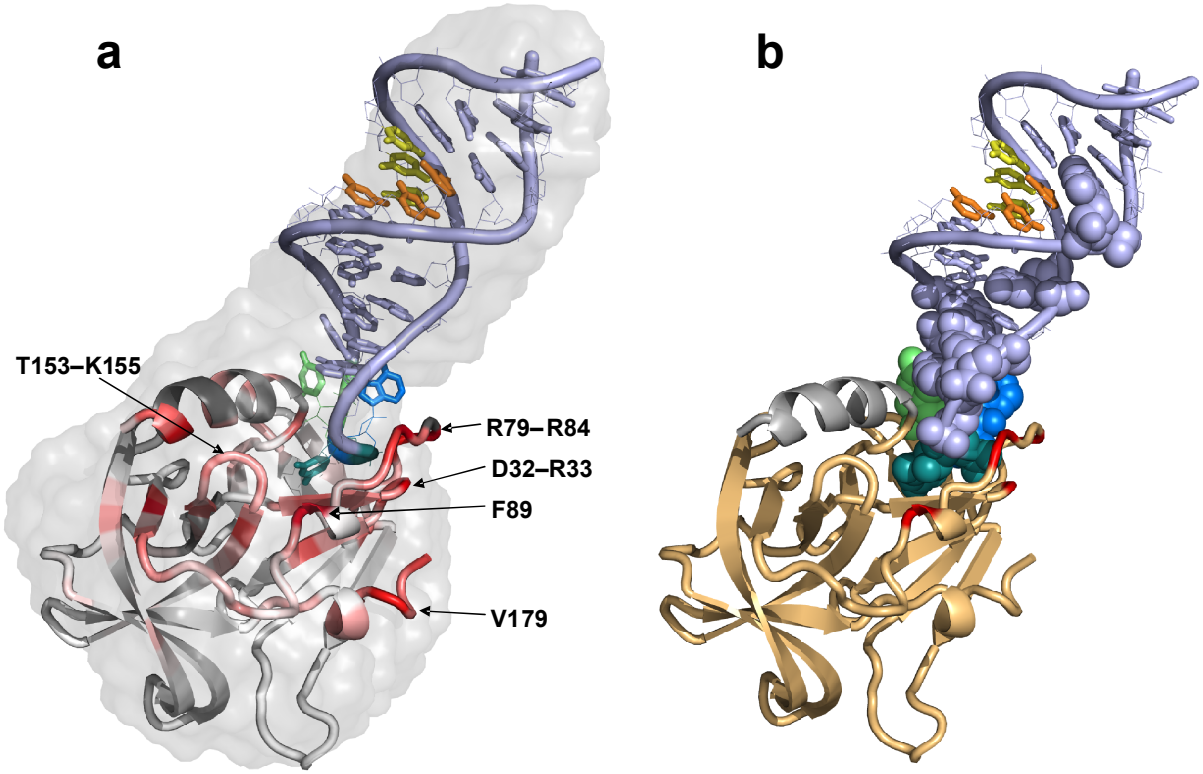


Figure 4.14: (a) Complex showing SAXS envelope in grey. The RNA and protein are shown as cartoon representation. The bases are coloured as in Figure 1.11. Residues implicated in RNA interaction are labelled. (b) Complex in same orientation but with bases equivalent in HRV to those perturbed upon $3C^{pro}$ binding in coxsackie virus (Ohlenschläger *et al.*, 2004) (Figure 4.12 shown as spheres.

While the proteolytic active site face of $3C^{pro}$ experiences relatively little perturbation upon RNA binding, the largest changes on this face cluster to the vicinity of the catalytic triad (H40, E71 and C146). This includes chemical shift changes in T143, loss of peaks for F25 and G144 and intensity reduction in H40, Q42, E71 and S127. It thus appears that the signal of SLD binding to

the linker side of $3C^{pro}$ is transduced into a change at the active site. This could represent a mode of regulation, fine-tuning the specificity or activity of $3C^{pro}$ as previously suggested (Peters et al., 2004; Shih et al., 2004). The mechanism of this transmission may involve the hydrophobic core of the protein. Four β-strands approach both the active site and the RNA-binding site (Figure 4.15): strands 2 (F25–I30) and 3 (V34–I37), connected by the 31–33 loop and strands 10 (G147–L150) and 11 (F157–G163), connected by the 151–156 loop that contains T153, G154 and K155. Both of these loops show changes upon addition of SLD. The H31 and D32 positions of the 31–33 loop that is sandwiched between the C-terminus and residues 79–84 of the linker are relatively well conserved (H or aromatic 70%; D or E 70% in conserved domain database analysis). Perturbation of this loop appears to occur via direct interactions with the RNA triloop through the 79–84 loop, or in a supporting role for this loop. The 151–156 loop is partially conserved at the G154 (50%) and K155 (70%) positions and contains 100% hydrophobic residues at the 156 and 157 positions. Changes induced in this loop may be more indirect in nature, as SLD does not closely approach these residues in the present model. In 3CI, T153 and G154 interact with F6 (90% hydrophobic) of the N-terminal helix and are within ~4 Åof F89 of the opposing FISEDLE motif from the C-terminal half of the linker. K155 closely approaches the acidic residues in the FISEDLE loop. Thus the 151–156 loop is positioned between two important structural elements, the N-terminal α-helix and the inter- domain linker. Both of these elements directly contact the RNA. Changes induced in these elements could result in the observed (indirect) perturbation of the 151–156 loop. Note that I90 and G96 peaks increase in intensity upon SLD-binding (Figure 4.8), consistent with the C-terminal half of the linker increasing in mobility.

In turn, changes in the conformation or position of the 31–33 and 151–156 loops may transmit the observed changes to the proteolytic active site via the four associated β-strands. Strands 2 and 3 extensively hydrogen bond to each other, as do strands 10 and 11 up to the point where the longer strand 11 curves toward the protein periphery. Therefore the induced change is more likely to involve a slight reorientation of the 2–3 pair relative to the 10–11 pair than changes within each pair. The 2–3 to 10–11 contact is in fact the point of junction between the N-terminal and C-terminal β-barrel. Together with the intimate involvement of the 76–97 linker that covalently joins the two barrels, it would appear that $3C^{pro}$–SLD complex formation may induce a subtle change in the relative positions of the two $3C^{pro}$ β-barrels that transmits RNA- binding information to the opposite, catalytic face of the protease. Identification of the $3C^{pro}$– SLD interaction surface along with the detection of RNA-induced changes in the $3C^{pro}$ catalytic site may lead to novel

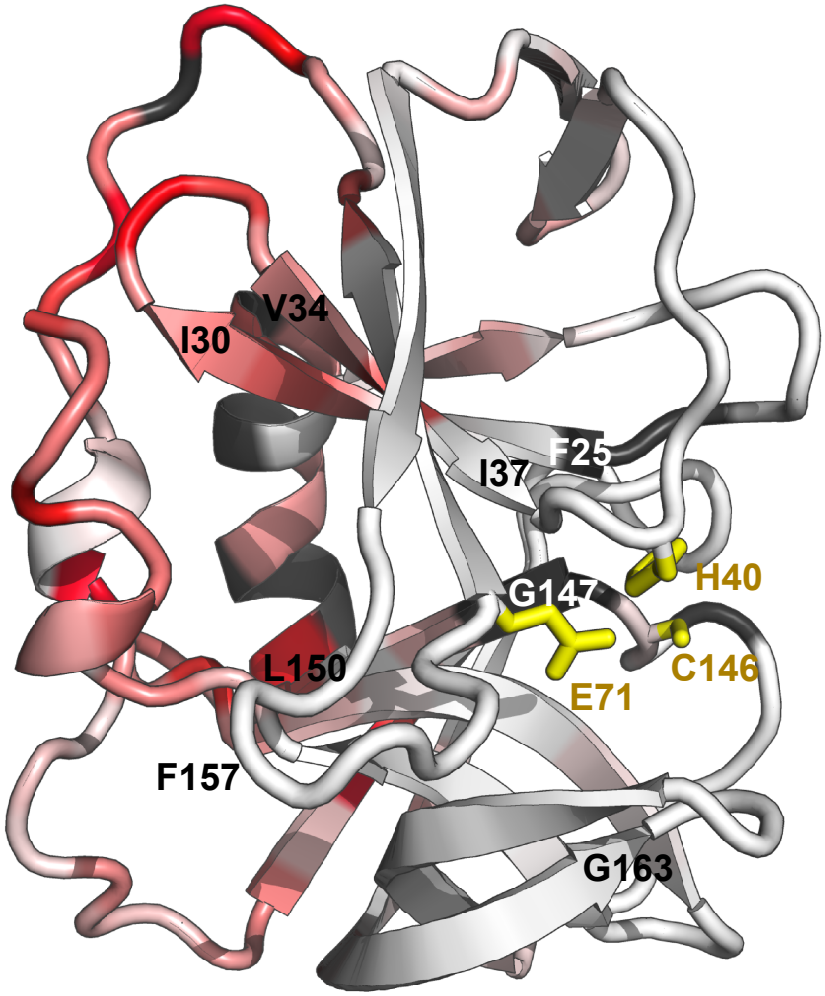


Figure 4.15: Possible allosteric "communication pathway" between the RNA binding surface and the catalytic triad. Residues interacting with the RNA are shown in red and the catalytic triad is shown as yellow sticks.

avenues for drug design.

Crystallographic studies of the $3C^{pro}$–3Dpol fusion protein from poliovirus (Marcotte et al., 2007) position the covalently linked 3Dpol in close contact with the C–terminal region of $3C^{pro}$. This orientation of $3D^{pol}$, if conserved in HRV-14, would not interfere with the SLD–$3C^{pro}$ contacts as determined in the present study, nor would the SLD position sterically interfere with proteolysis of the $3C^{pro}$–3Dpol linkage that is required for $3D^{pol}$ activation. Furthermore, HRV-14 - 23 - A picornaviral loop-to-loop replication complex SLD is positioned well to interact with a symmetry-related $3D^{pol}$ molecule in the poliovirus crystal structure via the surface of $3D^{pol}$ that has been shown to contact VPg during the process of uridylation (Ferrer-Orta *et al.*, 2006). Structural studies of $3C^{pro}$–$3D^{pol}$ in complex with SLD and with larger fragments of the 5′-UTR along with quantitative analyses of RNA-induced effects on $3C^{pro}$ catalytic activity will help to further delineate the molecular mechanism of genomic replication in picornaviruses and coronaviruses.

5

Conclusions and Further Work

5.1 Completed Work

Despite the difficulties inherent in the project, a large amount of useful and novel data was acquired.

5.1.1 Characterisation of 3C in solution

Although the structure of human rhinovirus $3C^{pro}$ has been solved previously, both in solution (Bjorndahl *et al.*, 2007), and by X-ray crystallography (Matthews *et al.*, 1994), the dynamics of the protein had not previously been extensively studied. The process of using a homology model and refining this with residual dipolar coupling data is logically similar to molecular replacement techniques in crystallography and should, in future be able to provide a much quicker and more automated route of NMR structure determination than more conventional techniques.

5.1.2 Dynamics of $3C^{pro}$ in apo and holo forms

The $3C^{pro}$ protein is largely well structured and has relatively uniform S^2 values. This was evident from earlier studies, both crystallographic (Walker *et al.*, 1995) and in solution (Bjorndahl *et al.*, 2007) (see Chapter 3). However, this work has shown dynamics on longer time-scales previously unobserved in $3C^{pro}$. These, longer time-scale, motions have been shown to be related to RNA binding and to proteolytic function.

5.1.3 SAXS structure of complex

Structure calculations by NMR usually proceed via a well-trodden route. Initially, resonances are assigned and NOEs attributed to specific residues. In the present work, this was impractical. The results of the NMR and SAXS structural determination detailed in Chapter 4 show that the complex is not spherical and that the RNA binds to the protein in an end-on fashion. This explains why the peaks in the early ^{15}N,^{1}H-HSQCs, carried out on protonated $3C^{pro}$, were broadened. The large anisotropy of the complex leads to an increase in tumbling time. The estimated correlation time for the complex is 19 ns in comparison to 14 ns for a globular protein of similar molecular weight. The use of deuterated protein samples and TROSY-based pulse-sequences efficiently reduced this broadening and enabled the collection of data that were much more accurate and illuminating than would otherwise have been possible. However, it was still not possible

to acquire inter-molecular NOEs that would have been extremely useful in characterising the complex to a higher resolution.

5.2 Contribution to Current Knowledge

5.2.1 Dynamics

It is perhaps somewhat surprising that a globular protein is so dynamic on a millisecond to microsecond time-scale. The work described in chapters 3 and 4 suggests that communication between the active site of $3C^{pro}$ and the RNA–binding face takes place. In addition to the work described here, a number of sources have postulated or provided evidence for allostery between the two functional regions of the protein and related proteins in other picornaviruses (Peters *et al.*, 2004; Wang and Johnson, 2001; Shih *et al.*, 2004; Bjorndahl *et al.*, 2007). Bjorndahl and co-workers specifically mention that the structure of the RNA–binding face is unchanged in HRV-14 $3C^{pro}$ upon inhibitor binding but leave the possibility open for a change in the RNA–binding face to affect catalytic activity.

5.2.2 Allostery

Allostery (Gr. αλλος: other and στερεος: shape) refers to an alteration in the activity of one protein domain upon binding of an effector to a distal region of the protein. Allosteric interactions in proteins have been studied for a long time. The second crystal structure of a protein to be solved was that of haemoglobin (Perutz *et al.*, 1960), the archetypal allosteric protein. An early model to be proposed for allostery in haemoglobin was that of Monod, Wyman and Changeux (MWC) (Monod *et al.*, 1965). In this model oxygen molecules bind to subunits in haemoglobin and alter the equilibrium between two quaternary states, designated tense (T) and relaxed (R), where T and R forms of haemoglobin have low and high affinity for oxygen respectively. The increase in affinity for binding the second, third and fourth molecules of O_2 to a single haemoglobin molecule is due to the shift of the equilibrium towards the R state. Another model, that of sequential binding was proposed for haemoglobin allostery. In this, the KNF model (Koshland *et al.*, 1966), binding of each O_2 changes the structure of a neighbouring unbound subunit to the R form, with higher affinity for O_2. It is possible for hybrid TR intermediates to occur.

Despite the thermodynamic implications of the MWC model, the majority of proteins have, until recently, been largely considered to have little or no allosteric function. However, several

strands of evidence are beginning to suggest that most, if not all, non-fibrous proteins have the potential for allosteric control. Another way of putting this, in keeping with the classical view of allostery, is that distribution of the conformational ensemble is also not confined to allosteric proteins, it is a characteristic of proteins in general (Gunasekaran *et al.*, 2004; Smock and Gierasch, 2009). A seminal paper by Cooper and Dryden (Cooper and Dryden, 1984) examined the idea that changes in protein dynamics could be of functional consequence. The internal structure of globular proteins is dominated by hydrophobic interactions. Protein folding is often described in terms of minimisation of exposure of hydrophobic residues to the bulk solvent. NMR studies of the dynamics of side chains within the protein interior show that the side chains sample many different conformations and that these states exchange on several time-scales. Upon cooling to around 200K, the exchange between states is 'frozen' and only ps–ns timescale motions (such as bond librations) occur (Lee and Wand, 2001). This so-called 'glass transition' indicates that proteins in solution contain a large amount of residual entropy even when well folded (Wand, 2001b). Although crystal structures, as well as averages of NMR ensembles, suggest a static, well-defined fold for proteins, the reality is that proteins constantly sample the conformational space around the bottom of the 'folding funnel' (Dobson *et al.*, 1998; Lazaridis and Karplus, 1997). Functional consequences of this dynamic situation are beginning to be appreciated. In retrospect it would seem rather surprising if proteins did not take advantage of this large amount of entropy available to them for aspects of their function. S^2 data on folded proteins indicates that backbone motion is low compared to the side-chains and that the backbone functions as a relatively rigid 'skeleton'. This is not to say that large motions involving the backbone do not occur. Many functional motions involve backbone rearrangement. Side chain motions on a ps–ns time-scale are extensive in all proteins (Wand, 2001a; Hu *et al.*, 2003). Interestingly, the degree of motion in the backbone appears to be poorly correlated to the degree of motion in the attached side-chains (Wand, 2001b).

5.2.3 Allostery in single domain proteins

It is now becoming clear that proteins exhibit allosteric behaviour even within domains. Recent NMR studies have shown that mutations in proteins can affect the distribution of states in an ensemble and that this can lead to effects at distal sites. An elegant, NMR-based study of allosteric response in a single domain protein is provided by the work of Volkman and co-workers (Volkman *et al.*, 2001). NtrC is a member of the 'two-component system' signalling family (Stock

et al., 2000). Phosphorylation of the response-regulator region of NtrC promotes oligomer formation in the full length protein, which then stimulates transcription by an RNA polymerase. Mutant NtrC with partial activity when unphosphorylated shows altered chemical shifts in the region that changes structure upon phosphorylation. Both the apo and mutant forms of the protein show significant R_{ex} in this region; upon phosphorylation the R_{ex} is largely abrogated. This leads to the conclusion that the active state is sampled by the apo form but only transiently. The partially active mutant samples the active state more frequently. The conclusion drawn from these data is that the protein samples the active state even in the apo form; phosphorylation of the protein merely changes the equilibrium of a pre-sampled conformational ensemble. It appears that "active" conformers are present in "inactive" ensembles at lower concentrations. The authors of this study also suggest that the μs–ms time-scale motions seen in this study are likely to be common as 'many biological processes occur in micro- to millisecond time scales'. Conversely another study of dynamic pathways in a globular protein found that networks of responses took place mainly on a ps–ns time-scale. Clarkson and co-workers (Clarkson *et al.*, 2006) showed that, by mutating valine residues in eglin C to alanine, they were able to alter the side chain S^2 and τ_e parameters of distal residues. There were found to be two types of response: a 'contiguous network response' involving changes in S^2, transmitted residue to residue down a pathway and a 'disperse network response' where rotomer populations (measured by τ_e) were altered in residues up to 10 Å from the mutated residue with no obvious chain of contact. The signals were not necessarily bidirectional. It was also found that the entropic couplings predict enthalpic couplings, indicating, that even if the changes in dynamics (and therefore entropy in the protein core) are not the root cause of allostery, they may be a good indicator of the presence of allosteric interactions. The difference in time-scales found to be important in these two studies is indicative of two different facets of the interactions. It seems likely that the residual entropy in a protein, as identified by side-chain motions on a ps–ns time-scale has been utilised by proteins to transmit information between distal sites. However, as mentioned above, many biological activities occur on a longer time-scale. It is possible that the time-scale of motions relevant to the allosteric interaction is much faster than the time-scale of the process affected by those motions (Wand, 2001b).

5.2.4 Possibilities of networks for 3C^{pro}

What form does the allosteric interaction take in 3C? There is a small amount of perturbation of the active site upon RNA–binding. I have suggested previously that this could be mediated by a chain of residues, largely following β-strands through the centre of the protein. The network of chemical-shift perturbations is not entirely contiguous. However, it has been suggested that transduction of motion or energy may occur through immobile residues (Ota and Agard, 2005). Figure 5.1 shows the pathway through the centre of the protein of residues that exhibit chemical shift perturbation upon RNA binding. The study of eglin C dynamics mentioned previously (Clarkson *et al.*, 2006), has shown that, although it is possible for contiguous transmission of dynamic changes to cause allosteric effects in proteins, this is by no means necessary. Chemical shifts are probably coupled to selected motions on a ps–ns time-scale(Berjanskii and Wishart, 2005). Examination of R_{ex} found by the cross-correlation method reveals that many of the residues in the two β-strands exhibit significant motion on the milli- to microsecond timescale. It is possible that a change in the rate of exchange in these residues upon binding of the RNA causes changes in the equilibrium population of states, leading to the observed changes in chemical shift. This is certainly suggested by the change in distribution of R_{ex}^{CC} between 3C and 3CI. The RNA-binding region is very dynamic on a ms–μs time-scale, more so than the opposite face. It would be interesting to see whether, in common with previously mentioned work (Volkman *et al.*, 2001), there is a decrease in this motion upon RNA binding. Thus, in 3C^{pro}, it would seem likely that substrate or RNA binding can induce changes in ps–ns dynamics in the protein core, and this can alter the equilibrium of μs–ms motion at either the catalytic or RNA–binding face of the protein.

5.3 Aims of the Project

The original aims of the project were to: (i) Elucidate the dynamics and structure of 3C^{pro} in solution. (ii) Map the RNA–binding surface of 3C^{pro} using chemical shift perturbation (iii) Model the interaction between the RNA and the protein. (iv) Solve the solution structure of the human rhinovirus 3C protease–stemloop-d (3C^{pro}–SLD) complex to high resolution. While the project was successful and a proportion of the aims were achieved, the course of events did not proceed as smoothly as might have been anticipated and, as is often the case, unexpected results were found. The following issues arose during the project:

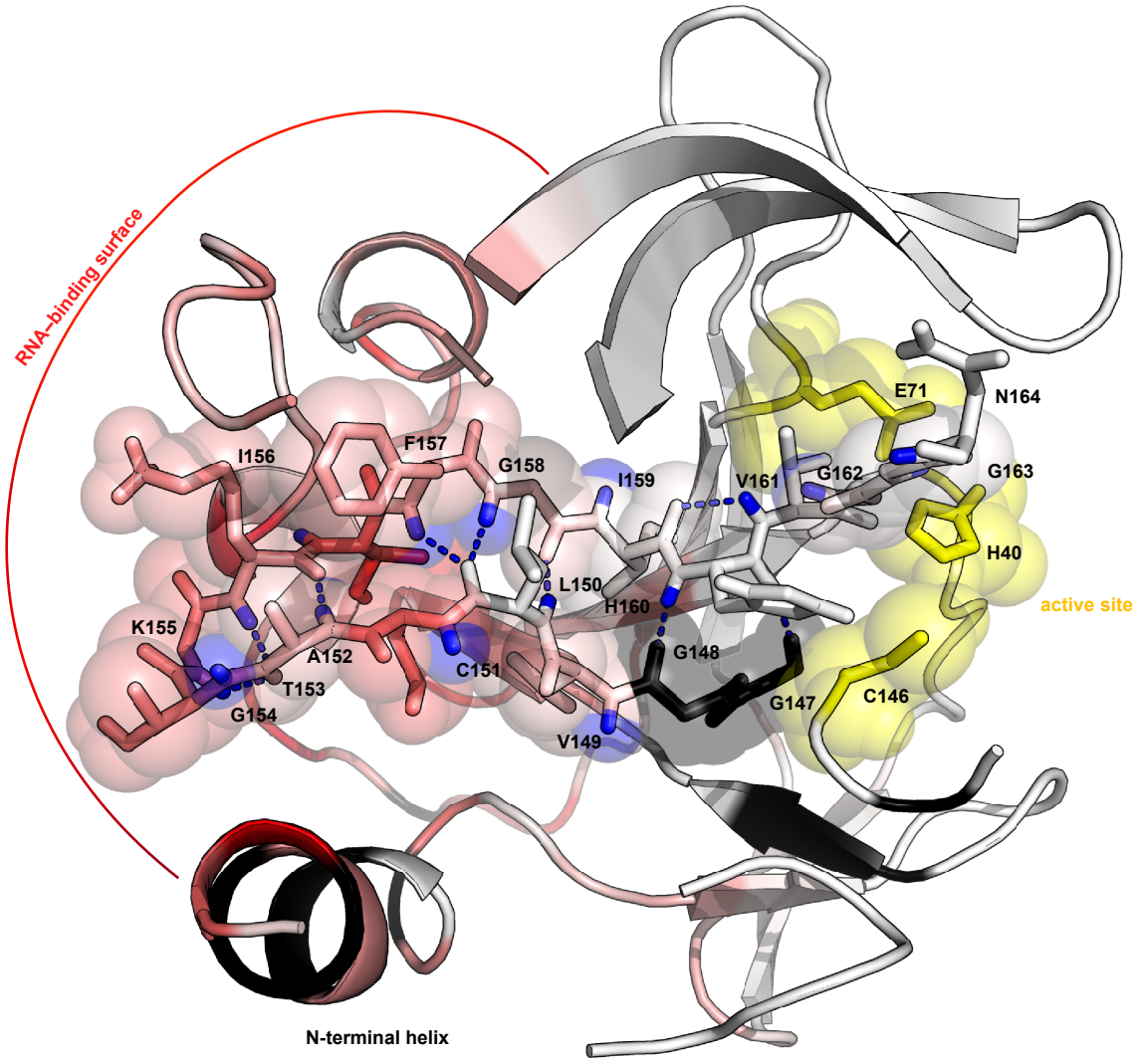


Figure 5.1: Another view of the possible allosteric interaction between RNA binding and proteolytic activity. The hydrogen-bonding network between the two β-strands which connect the T153-K155 loop to the catalytic triad is shown. Residues in the active site and the two β-strands are named. Spheres represent residues that exhibit chemical-shift perturbation upon RNA binding (also shown by increasing intensity of red colouring). Catalytic triad residues are shown in yellow. Residues which are exchange-broadened to the extent that they are undetectable in the complex are shaded black.

5.3.1 Solution structure and Dynamics

At the start of the project, solving the structure of the 3C^{pro} in solution was not intended from scratch as a structure of the protein (Matthews *et al.*, 1994) had been solved by crystallography and the structures of several close homologues were also available(Matthews *et al.*, 1999; Mosimann *et al.*, 1997; Sweeney *et al.*, 2007). This enabled the use of homology modelling along with RDC restraints to produce a structural model of the protein.

5.3.2 RNA-binding surface mapping

Chemical-shift perturbation (CSP) mapping was carried out on the 3C^{pro}–SLD complex giving NH, Cα and C′ perturbations.

5.3.3 Complex Structure

The lack of intermolecular NOEs between the complex partners could be atttibutable to several factors. Although the build-up of NOEs in larger macromolecules is faster than in smaller molecules, the faster T_2 relaxation of larger molecules means that transferred magnetisation is often dephased before it can be detected. This would mean that the NOE would not have time to build up before the signal had become too broad to detect. Another problem is that at least some of the interacting residues in the complex are exchange broadened. This makes it very difficult to get any information about shifts let alone acquire sufficient signal to obtain NOE interactions.

Inability to crystallise complex

Crystallisation of RNA–protein complexes is notoriously difficult. This has been attributed to several factors. Firstly, RNA is often prone to attack by nucleases in solution. Although this does not seem to be the case here, it is very difficult to remove all the RNAses from a sample and they are ubiquitously present in equipment and reagents. Secondly, all the bases in RNA have the same, highly-charged, phosphate moiety. The presence of several identical phosphates can lead to errors in crystal packing. Finally, RNA molecules (including the SLD) are flexible and elongated in solution, this often leads to high temperature factors and disordered regions in RNA and RNA–protein crystals (Ke and Doudna, 2004). As well as the difficulty of crystallising RNA–protein complexes in general, there are further specific issues with this system. Some intermediate time-scale motion is present on the RNA interaction surface of the protein as shown

by the R_{ex} data and by the broadening of some of the protein peaks corresponding to interacting residues. In crystallographic studies such motion is usually indicated by disordered regions where little or no electron density is observed. Disorder across the interaction surface makes production of diffracting crystals difficult. In addition, the multiple peaks associated with D32, N80 and F83 upon titration with RNA are evidence of either chemical exchange between states or of several independent states present in the sample. This inhomogeneity will make crystallisation less likely.

5.3.4 Instability of complex

The complex was found to be quite unstable in solution. This meant that some experiments were unable to be performed before the sample degraded. In particular, RDCs of the protein and RNA in complex were unable to be obtained to a sufficient signal-to-noise before spectral quality had deteriorated too much.

5.4 Future Work

5.4.1 Conditions to get intermolecular NOEs

Since the complex has micromolar affinity, in principle it should be possible to obtain intermolecular NOEs in order to further characterise the interaction between the protein and the RNA. Paramount to this would be increasing the solubility of the complex thus increasing the concentration of complex that could be studied by NMR.

5.4.2 Dynamics of 3C^{pro}

Further work on the dynamics of the protein is required to narrow down the time-scale of the dynamics that are evident from the SAXS and relaxation data. The structure of 3C^{pro} will also be improved by the inclusion of the dynamics data described in this study. This could be achieved by the use of time-averaged NOEs to take the molecular motions into account (Bonvin *et al.*, 1994).

5.4.3 Dynamics of complex

Although the complex structure goes some way to characterising the interaction, it would be very interesting to see what, if any, changes occur in the dynamics of the protein upon binding

to SLD. This would also help, along with biochemical studies, to validate the notion of allosteric interactions between the catalytic and RNA-binding surfaces. A model-free analysis of $3C^{pro}$–SLD would, however, have to take into account the anisotropy of the complex. An assay for $3C^{pro}$ activity is available. It would be useful to determine whether there is any change in activity upon RNA binding. Surface-plasmon resonance studies of HAV $3C^{pro}$ have been used to characterise RNA binding in the presence of $3C^{pro}$ -substrate-mimetic peptides (Peters *et al.*, 2004).

5.4.4 Crystallisation of complex

More thorough investigation of crystallisation conditions for the complex is required to obtain a crystal structure of the complex.

5.4.5 Higher order interactions

With the use of a combination of NMR spectroscopy, SAXS and molecular modelling, including results from previous mutational analysis, it would be possible to build the structure of the replication ribonucleoprotein-complex in human rhinovirus. This putative structure consists of the 3CD and 3AB proteins, the cloverleaf from the 5′ UTR as well as several cellular proteins. Interactions between these structures and the 3′ UTR are thought to be required for the viral proteins to begin RNA synthesis. Elucidation of these interactions would give a much clearer understanding of this crucial part of the viral life cycle.

Bibliography

Alexandrov A., Dutta K., and Pascal S.M. (2001). MBP fusion protein with a viral protease cleavage site: one-step cleavage/purification of insoluble proteins. *BioTechniques*, **30**(6):1194–8.

Allain F.H., Bouvet P., Dieckmann T., and Feigon J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.*, **19**(24):6870–81.

Amero C.D., Arnold J.J., Moustafa I.M., Cameron C.E., and Foster M.P. (2008). Identification of the oriI-binding site of poliovirus 3C protein by nuclear magnetic resonance spectroscopy. *J. Virol.*, **82**(9):4363–70.

Amineva S.P., Aminev A.G., Palmenberg A.C., and Gern J.E. (2004). Rhinovirus 3C protease precursors 3CD and 3CD' localize to the nuclei of infected cells. *J. Gen. Virol.*, **85**(Pt 10):2969–79.

Andino R., Rieckhof G.E., Achacoso P.L., and Baltimore D. (1993). Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. *EMBO J.*, **12**(9):3587–98.

Andino R., Rieckhof G.E., and Baltimore D. (1990a). A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell*, **63**(2):369–80.

Andino R., Rieckhof G.E., Trono D., and Baltimore D. (1990b). Substitutions in the protease (3Cpro) gene of poliovirus can suppress a mutation in the 5' noncoding region. *J. Virol.*, **64**(2):607–12.

Anonymous (2009). Progress towards interrupting wild poliovirus transmission worldwide, 2008. *Wkly Epidemiol Rec*, **84**(14):110–6.

Arnold E. and Rossmann M.G. (1990). Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 A. *J. Mol. Biol.*, **211**(4):763–801.

Baker N.A., Sept D., Joseph S., Holst M.J., and McCammon J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**(18):10037–41.

Banerjee R., Weidman M.K., Echeverri A., Kundu P., and Dasgupta A. (2004). Regulation of poliovirus 3C protease by the 2C polypeptide. *J. Virol.*, **78**(17):9243–56.

Barbato G., Cicero D.O., Cordier F., Narjes F., Gerlach B., Sambucini S., Grzesiek S., Matassa V.G., Francesco R.D., and Bazzo R. (2000). Inhibitor binding induces active site stabilization of the HCV NS3 protein serine protease domain. *EMBO J.*, **19**(6):1195–206.

Bartels C., TH X., Billeter M., Güntert P., and Wuthrich K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR*, **6**(1):1–10.

Bedard K.M. and Semler B.L. (2004). Regulation of picornavirus gene expression. *Microbes Infect*, **6**(7):702–13.

Bergmann A., Fritz G., and Glatter O. (2000). Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated annealing (BSSA). *J Appl Crystallogr*, **33**:1212–1216.

Bergmann E.M., Mosimann S.C., Chernaia M.M., Malcolm B.A., and James M.N. (1997). The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *J. Virol.*, **71**(3):2436–48.

Berjanskii M.V., Berjanskii M.V., Wishart D.S., and Wishart D.S. (2007). The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Research*, **35**(Web Server issue):W531.

Berjanskii M.V. and Wishart D.S. (2005). A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.*, **127**(43):14970–1.

Birney E., Kumar S., and Krainer A. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Research*, **21**:5803–5816.

Birtley J.R. and Curry S. (2005). Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy. *Acta Crystallogr. D Biol. Crystallogr.*, **61**(Pt 5):646–50.

Birtley J.R., Knox S.R., Jaulent A.M., Brick P., Leatherbarrow R.J., and Curry S. (2005). Crystal structure of foot-and-mouth disease virus 3C protease. New insights into catalytic mechanism and cleavage specificity. *J. Biol. Chem.*, **280**(12):11520–7.

Bjorndahl T.C., Andrew L.C., Semenchenko V., and Wishart D.S. (2007). NMR solution structures of the apo and peptide-inhibited human rhinovirus 3C protease (Serotype 14): structural and dynamic comparison. *Biochemistry*, **46**(45):12945–58.

Bjorndahl T.C., Monzavi H., and Wishart D.S. (2003). Backbone 1H, 15N and 13C assignments for the human rhinovirus 3C protease (serotype 14). *J. Biomol. NMR*, **26**(1):85–6.

Blair W.S., Parsley T.B., Bogerd H.P., Towner J.S., Semler B.L., and Cullen B.R. (1998). Utilization of a mammalian cell-based RNA binding assay to characterize the RNA binding properties of picornavirus 3C proteinases. *RNA*, **4**(2):215–25.

Bonvin A.M.J.J., Boelens R., and Kaptein R. (1994). Time- and ensemble-averaged direct NOE restraints. *J. Biomol. NMR*, **4**(1):143–149.

Bouvet P., Allain F., Finger L., Dieckmann T., and Feigon J. (2001). Recognition of pre-formed and flexible elements of an RNA stem-loop by nucleolin. *J. Mol. Biol.*, **309**:763–775.

Brünger A. (2007). Version 1.2 of the Crystallography and NMR system. *Nat Protoc*, **2**(11):2728–33.

Brünger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.S., Kuszewski J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T., and Warren G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**(Pt 5):905–21.

Cavanagh J., Fairbrother W.J., Palmer A.G., III, Skelton N.J., and Rance M. (2007). *Protein NMR spectroscopy: principles and practiceâĂŐ*. Elsevier Academic Press, Amsterdam, Boston, 2nd ed.

Cho M.W., Richards O.C., Dmitrieva T.M., Agol V., and Ehrenfeld E. (1993). RNA duplex unwinding activity of poliovirus RNA-dependent RNA polymerase 3Dpol. *J. Virol.*, **67**(6):3010–8.

Claridge J., Headey S., Chow J., Schwalbe M., Edwards P., Jeffries C., Venugopal H., Trewhella J., and Pascal S. (2009). A picornaviral loop-to-loop replication complex. *J Struct Biol*.

Clarkson M.W., Gilmore S.A., Edgell M.H., and Lee A.L. (2006). Dynamic coupling and allosteric behavior in a nonallosteric protein. *Biochemistry*, **45**(25):7693–9.

Clore G.M., Szabo A., Bax A., Kay L.E., Driscoll P.C., and Gronenborn A.M. (1990). Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.*, **112**(12):4989–4991.

Clubb R.T., Thanabal V., and Wagner G. (1992). A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue $^{1}H^{N}$, ^{15}N, and $^{13}C'$ Chemical-Shifts in ^{15}N–^{13}C-Labeled Proteins. *Journal of Magnetic Resonance*, **97**(1):213–217.

Cole R. and Loria J.P. (2007). *FASTModelfree*.

Cooper A. and Dryden D.T. (1984). Allostery without conformational change. A plausible model. *Eur Biophys J*, **11**(2):103–9.

Cordey S., Gerlach D., Junier T., Zdobnov E.M., Kaiser L., and Tapparel C. (2008). The cis-acting replication elements define human enterovirus and rhinovirus species. *RNA*, **14**(8):1568–78.

Cordier F., Dingley A.J., and Grzesiek S. (1999). A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings. *J. Biomol. NMR*, **13**(2):175–80.

Cordingley M.G., Callahan P.L., Sardana V.V., Garsky V.M., and Colonno R.J. (1990). Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J. Biol. Chem.*, **265**(16):9062–5.

Cox G., Johnson R., Cook J., Wakulchik M., Johnson M., Villarreal E., and Wang Q. (1999). Identification and Characterization of Human Rhinovirus-14 3C Protease Deamidation Isoform. *Journal of Biological Chemistry*, **274**(19):13211.

Datta U. and Dasgupta A. (1994). Expression and subcellular localization of poliovirus VPg-precursor protein 3AB in eukaryotic cells: evidence for glycosylation in vitro. *J. Virol.*, **68**(7):4468–77.

DeLano W. (2008). *The PyMOL Molecular Graphics System*.

Dobson C., Sali A., and Karplus M. (1998). Protein folding: A perspective from theory and experiment. *Angew Chem Int Edit*, **37**(7):868–893.

Dominguez C., Boelens R., and Bonvin A.M.J.J. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**(7):1731–7.

Dower W.J., Miller J.F., and Ragsdale C.W. (1988). High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Research*, **16**(13):6127–45.

Du Z., Yu J., Ulyanov N.B., Andino R., and James T.L. (2004). Solution structure of a consensus stem-loop D RNA domain that plays important roles in regulating translation and replication in enteroviruses and rhinoviruses. *Biochemistry*, **43**(38):11959–72.

Dutta K., Shi H., Cruz-Chu E.R., Kami K., and Ghose R. (2004). Dynamic influences on a high-affinity, high-specificity interaction involving the C-terminal SH3 domain of p67phox. *Biochemistry*, **43**(25):8094–106.

Eghbalnia H.R., Wang L., Bahrami A., Assadi A., and Markley J.L. (2005). Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. *Journal of Biomolecular NMR*, **32**(1):71–81.

Eletsky A., Kienhöfer A., and Pervushin K. (2001). TROSY NMR with partially deuterated proteins. *J. Biomol. NMR*, **20**(2):177–80.

Epstein D.M., Benkovic S.J., and Wright P.E. (1995). Dynamics of the dihydrofolate reductase-folate complex: catalytic sites and regions known to undergo conformational change exhibit diverse dynamical features. *Biochemistry*, **34**(35):11037–48.

Evans D.J. and Almond J.W. (1998). Cell receptors for picornaviruses as determinants of cell tropism and pathogenesis. *Trends Microbiol*, **6**(5):198–202.

Farrow N.A., Muhandiram R., Singer A.U., Pascal S.M., Kay C.M., Gish G., Shoelson S.E., Pawson T., Forman-Kay J.D., and Kay L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, **33**(19):5984–6003.

Ferrer-Orta C., Arias A., Agudo R., Pérez-Luque R., Escarmís C., Domingo E., and Verdaguer N. (2006). The structure of a protein primer-polymerase complex in the initiation of genome replication. *EMBO J.*, **25**(4):880–8.

Fushman D. and Cowburn D. (1999). The effect of noncollinearity of 15N-1H dipolar and 15N CSA tensors and rotational anisotropy on 15N relaxation, CSA/dipolar cross correlation, and TROSY. *Journal of Biomolecular NMR*, **13**(2):139–47.

Fushman D., Tjandra N., and Cowburn D. (1998). Direct Measurement of 15N Chemical Shift Anisotropy in Solution. *J. Am. Chem. Soc*, **120**(42):10947–10952.

Golovanov A.P., Hautbergue G.M., Wilson S.A., and Lian L.Y. (2004). A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.*, **126**(29):8933–9.

Goodfellow I., Chaudhry Y., Richardson A., Meredith J., Almond J.W., Barclay W., and Evans D.J. (2000). Identification of a cis-acting replication element within the poliovirus coding region. *J. Virol.*, **74**(10):4590–600.

Gorbalenya A., Donchenko A., Blinov V., and Koonin E. (1989). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Lett.*, **243**(2):103–114.

Greenberg S. (2003). Respiratory Consequences of Rhinovirus Infection. *Archives of Internal Medicine*, **163**(3):278.

Grzesiek S. and Bax A. (1992). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance*, **99**(1):201–207.

Grzesiek S. and Bax A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J. Biomol. NMR*, **3**(2):185–204.

Guinier A. and Fournet G. (1955). *Small-angle scattering of X-rays*.

Gunasekaran K., Ma B., and Nussinov R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins*, **57**(3):433–43.

Hämmerle T., Molla A., and Wimmer E. (1992). Mutational analysis of the proposed FG loop of poliovirus proteinase 3C identifies amino acids that are necessary for 3CD cleavage and might be determinants of a function distinct from proteolytic activity. *J. Virol.*, **66**(10):6028–34.

Hanahan D. (1983). Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.*, **166**(4):557–80.

Handa N., Nureki O., Kurimoto K., Kim I., Sakamoto H., Shimura Y., Muto Y., and Yokoyama S. (1999). Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature*, **398**:579–585.

Harris K.S., Xiang W., Alexander L., Lane W.S., Paul A.V., and Wimmer E. (1994). Interaction of poliovirus polypeptide 3CDpro with the 5′ and 3′ termini of the poliovirus genome. Identification of viral and cellular cofactors needed for efficient binding. *J. Biol. Chem.*, **269**(43):27004–14.

Headey S.J., Huang H., Claridge J.K., Soares G.A., Dutta K., Schwalbe M., Yang D., and Pascal S.M. (2006). NMR structure of stem-loop D from human rhinovirus-14. *RNA*, **13**(3):351–60.

Heinz B.A. and Vance L.M. (1995). The antiviral compound enviroxime targets the 3A coding region of rhinovirus and poliovirus. *J. Virol.*, **69**(7):4189–97.

Herold J. and Andino R. (2001). Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol Cell*, **7**(3):581–91.

Hu H., Clarkson M.W., Hermans J., and Lee A.L. (2003). Increased rigidity of eglin c at acidic pH: evidence from NMR spin relaxation and MD simulations. *Biochemistry*, **42**(47):13856–68.

Huang H., Alexandrov A., Chen X., Barnes T.W., Zhang H., Dutta K., and Pascal S.M. (2001). Structure of an RNA hairpin from HRV-14. *Biochemistry*, **40**(27):8055–64.

Hughes A. (2004). Phylogeny of the Picornaviridae and differential evolutionary divergence of picornavirus proteins. *Infection, Genetics and Evolution*, **4**(2):143–152.

Ishima R. and Torchia D.A. (2000). Protein dynamics from NMR. *Nat. Struct. Biol.*, **7**(9):740–3.

Jang S.K., Pestova T.V., Hellen C.U., Witherell G.W., and Wimmer E. (1990). Cap-independent translation of picornavirus RNAs: structure and function of the internal ribosomal entry site. *Enzyme*, **44**(1-4):292–309.

Jeffries C., Whitten A., Harris S., and Trewhella J. (2008). Small-angle X-ray scattering reveals the N-terminal domain organization of cardiac myosin binding protein C. *J. Mol. Biol.*, **377**:1186–1199.

Jensen M.R., Lauritzen C., Dahl S.W., Pedersen J., and Led J.J. (2004). Binding ability of a HHP-tagged protein towards Ni2+ studied by paramagnetic NMR relaxation: the possibility of obtaining long-range structure information. *Journal of Biomolecular NMR*, **29**(2):175–85.

Johnson C. and Bovey F. (1958). Calculation of Nuclear Magnetic Resonance Spectra of Aromatic Hydrocarbons. *J Chem Phys*, **29**(5):1012–1014.

Johnson P.E. and Donaldson L.W. (2006). RNA recognition by the Vts1p SAM domain. *Nat. Struct. Mol. Biol.*, **13**(2):177–8.

Kandolf R., Klingel K., Zell R., Canu A., Fortmüller U., Hohenadl C., Albrecht M., Reimann B.Y., Franz W.M., Heim A., Raab U., and McPhee F. (1993). Molecular Mechanisms in the Pathogenesis of Enteroviral Heart Disease: Acute and Persistent Infections. *Clinical Immunology and Immunopathology*, **68**(2):153–158.

Kay L., Ikura M., Tschudin R., and Bax A. (1990). Three-dimensional triple-resonance nmr spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance*.

Ke A. and Doudna J.A. (2004). Crystallization of RNA and RNA-protein complexes. *Methods*, **34**(3):408–14.

Kelly A.E., Ou H.D., Withers R., and Dötsch V. (2002). Low-conductivity buffers for high-sensitivity NMR measurements. *J. Am. Chem. Soc.*, **124**(40):12013–9.

Kempf J.G., Jung J.Y., Ragain C., Sampson N.S., and Loria J.P. (2007). Dynamic requirements for a functional protein hinge. *J. Mol. Biol.*, **368**(1):131–49.

Kenan D.J., Query C.C., and Keene J.D. (1991). RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**(6):214–20.

Kistler A.L., Webster D.R., Rouskin S., Magrini V., Credle J.J., Schnurr D.P., Boushey H.A., Mardis E.R., Li H., and DeRisi J.L. (2007). Genome-wide diversity and selective pressure in the human rhinovirus. *Virol J*, **4**:40.

Konarev P., Volkov V., Sokolova A., Koch M., and Svergun D. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr*, **36**:1277–1282.

Koshland D.E., Némethy G., and Filmer D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, **5**(1):365–85.

Kristensen S.M., Siegal G., Sankar A., and Driscoll P.C. (2000). Backbone dynamics of the C-terminal SH2 domain of the p85alpha subunit of phosphoinositide 3-kinase: effect of phosphotyrosine-peptide binding and characterization of slow conformational exchange processes. *J. Mol. Biol.*, **299**(3):771–88.

Kroenke C., Loria J., Lee L., Rance M., and Palmer A. (1998). Longitudinal and transverse H-1-N-15 dipolar N-15 chemical shift anisotropy relaxation interference: Unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *J. Am. Chem. Soc.*, **120**(31):7905–7915.

Kusov Y.Y. and Gauss-Müller V. (1997). In vitro RNA binding of the hepatitis A virus proteinase 3C (HAV 3Cpro) to secondary structure elements within the 5′ terminus of the HAV genome. *RNA*, **3**(3):291–302.

Kuyumcu-Martinez N.M., Eden M.E.V., Younan P., and Lloyd R.E. (2004). Cleavage of poly(A)-binding protein by poliovirus 3C protease inhibits host cell translation: a novel mechanism for host translation shutoff. *Mol Cell Biol*, **24**(4):1779–90.

Laskowski R., Rullmann J., Macarthur M., Kaptein R., and Thornton J. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**(4).

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

Lawrence C. and Thach R.E. (1975). Identification of a viral protein involved in post-translational maturation of the encephalomyocarditis virus capsid precursor. *J. Virol.*, **15**(4):918–28.

Lazaridis T. and Karplus M. (1997). "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science*, **278**(5345):1928–31.

Lee A.L. and Wand A.J. (2001). Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature*, **411**(6836):501–4.

Leong L.E., Walker P.A., and Porter A.G. (1993). Human rhinovirus-14 protease 3C (3Cpro) binds specifically to the 5′-noncoding region of the viral RNA. Evidence that 3Cpro has different domains for the RNA binding and proteolytic activities. *J. Biol. Chem.*, **268**(34):25735–9.

Long A.C., Orr D.C., Cameron J.M., Dunn B.M., and Kay J. (1989). A consensus sequence for substrate hydrolysis by rhinovirus 3C proteinase. *FEBS Lett.*, **258**(1):75–8.

Mandel A.M., Akke M., and Palmer A.G. (1995). Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.*, **246**(1):144–63.

Marchler-Bauer A., Anderson J.B., Derbyshire M.K., DeWeese-Scott C., Gonzales N.R., Gwadz M., Hao L., He S., Hurwitz D.I., Jackson J.D., Ke Z., Krylov D., Lanczycki C.J., Liebert C.A., Liu C., Lu F., Lu S., Marchler G.H., Mullokandov M., Song J.S., Thanki N., Yamashita R.A., Yin J.J., Zhang D., and Bryant S.H. (2007). CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Research*, **35**(Database issue):D237–40.

Matthews D., Dragovich P., Webber S., Fuhrman S., Patick A., Zalman L., Hendrickson T., Love R., Prins T., Marakovits J., Zhou R., Tikhe J., Ford C., Meador J., Ferre R., Brown E., Binford S., Brothers M., DeLisle D., and Worland S. (1999). Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl. Acad. Sci. U.S.A.*, **96**(20):11000–11007.

Matthews D.A., Smith W.W., Ferre R.A., Condon B., Budahazi G., Sisson W., Villafranca J.E., Janson C.A., McElroy H.E., and Gribskov C.L. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell*, **77**(5):761–71.

McCoy M.A., Senior M.M., Gesell J.J., Ramanathan L., and Wyss D.F. (2001). Solution structure and dynamics of the single-chain hepatitis C virus NS3 protease NS4A cofactor complex. *J. Mol. Biol.*, **305**(5):1099–110.

McKnight K.L. and Lemon S.M. (1998). The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. *RNA*, **4**(12):1569–84.

Meredith J.M., Rohll J.B., Almond J.W., and Evans D.J. (1999). Similar interactions of the poliovirus and rhinovirus 3D polymerases with the 3′ untranslated region of rhinovirus 14. *J. Virol.*, **73**(12):9952–8.

Mittermaier A. and Kay L.E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science*, **312**(5771):224–8.

Molla A., Harris K.S., Paul A.V., Shin S.H., Mugavero J., and Wimmer E. (1994). Stimulation of poliovirus proteinase 3Cpro-related proteolysis by the genome-linked protein VPg and its precursor 3AB. *J. Biol. Chem.*, **269**(43):27015–20.

Monod J., Wyman J., and Changeux J.P. (1965). On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.*, **12**:88–118.

Mosimann S.C., Cherney M.M., Sia S., Plotch S., and James M.N. (1997). Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.*, **273**(5):1032–47.

Muhandiram D. and Kay L. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *Journal of magnetic resonance. Series B(Print)*.

Mukherjee M., Dutta K., White M.A., Cowburn D., and Fox R.O. (2006). NMR solution structure and backbone dynamics of domain III of the E protein of tick-borne Langat flavivirus suggests a potential site for molecular recognition. *Protein Sci.*, **15**(6):1342–55.

Ohlenschläger O., Wöhnert J., Bucci E., Seitz S., Häfner S., Ramachandran R., Zell R., and Görlach M. (2004). The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure*, **12**(2):237–48.

Orr D.C., Long A.C., Kay J., Dunn B.M., and Cameron J.M. (1989). Hydrolysis of a series of synthetic peptide substrates by the human rhinovirus 14 3C proteinase, cloned and expressed in Escherichia coli. *J. Gen. Virol.*, **70 (Pt 11)**:2931–42.

Orthaber D., Bergmann A., and Glatter O. (2000). SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J Appl Crystallogr*, **33**:218–225.

Ota N. and Agard D.A. (2005). Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.*, **351**(2):345–54.

Ottiger M., Delaglio F., and Bax A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson*, **131**(2):373–8.

Oubridge C., Ito N., Teo C.H., Fearnley I., and Nagai K. (1995). Crystallisation of RNA-protein complexes. II. The application of protein engineering for crystallisation of the U1A protein-RNA complex. *J. Mol. Biol.*, **249**(2):409–23.

Palmer A.G. (1993). Dynamic properties of proteins from NMR spectroscopy. *Curr Opin Biotechnol*, **4**(4):385–91.

Palmer A.G. (2004). NMR characterization of the dynamics of biomacromolecules. *Chem Rev*, **104**(8):3623–40.

Parsley T.B., Towner J.S., Blyn L.B., Ehrenfeld E., and Semler B.L. (1997). Poly (rC) binding protein 2 forms a ternary complex with the 5'-terminal sequences of poliovirus RNA and the viral 3CD proteinase. *RNA*, **3**(10):1124–34.

Pathak H.B., Arnold J.J., Wiegand P.N., Hargittai M.R.S., and Cameron C.E. (2007). Picornavirus genome replication: assembly and organization of the VPg uridylylation ribonucleoprotein (initiation) complex. *J. Biol. Chem.*, **282**(22):16202–13.

Pathak H.B., Ghosh S.K.B., Roberts A.W., Sharma S.D., Yoder J.D., Arnold J.J., Gohara D.W., Barton D.J., Paul A.V., and Cameron C.E. (2002). Structure-function relationships of the RNA-dependent RNA polymerase from poliovirus (3Dpol). A surface of the primary oligomerization domain functions in capsid precursor processing and VPg uridylylation. *J. Biol. Chem.*, **277**(35):31551–62.

Paul A.V., Rieder E., Kim D.W., van Boom J.H., and Wimmer E. (2000). Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *J. Virol.*, **74**(22):10359–70.

Paul A.V., Yin J., Mugavero J., Rieder E., Liu Y., and Wimmer E. (2003). A "slide-back" mechanism for the initiation of protein-primed RNA synthesis by the RNA polymerase of poliovirus. *J. Biol. Chem.*, **278**(45):43951–60.

Pauling L. (1936). The Diamagnetic Anisotropy of Aromatic Molecules. *The Journal of chemical physics*, **4**:673–677.

Pelham H.R. (1978). Translation of encephalomyocarditis virus RNA in vitro yields an active proteolytic processing enzyme. *Eur J Biochem*, **85**(2):457–62.

Pelupessy P., Espallargas G.M., and Bodenhausen G. (2003). Symmetrical reconversion: measuring cross-correlation rates with enhanced accuracy. *J Magn Reson*, **161**(2):258–64.

Permi P., Rosevear P.R., and Annila A. (2000). A set of HNCO-based experiments for measurement of residual dipolar couplings in 15N, 13C, (2H)-labeled proteins. *J. Biomol. NMR*, **17**(1):43–54.

Perutz M.F., Perutz M.F., Rossmann M.G., Rossmann M.G., Cullis A.F., Cullis A.F., Muirhead H., Muirhead H., Will G., Will G., North A.C.T., and North A.C.T. (1960). Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature*, **185**(4711):416.

Pervushin K., Riek R., Wider G., and Wüthrich K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue

to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.*, **94**(23):12366–71.

Pervushin K., Wider G., and Wuthrich K. (1998). Single transition-to-single transition polarization transfer (ST2-PT) in [N-15,H-1]-TROSY. *J. Biomol. NMR*, **12**(2):345–348.

Peters H., Kusov Y.Y., Meyer S., Benie A.J., Bäuml E., Wolff M., Rademacher C., Peters T., and Gauss-Müller V. (2004). Hepatitis A virus proteinase 3C binding to viral RNA: correlation with substrate binding and enzyme dimerization. *Biochem. J.*, **385**(Pt 2):363–70.

Petoukhov M. and Svergun D. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.*, **89**:1237–1250.

Regoes R.R., Crotty S., Antia R., and Tanaka M.M. (2005). Optimal replication of poliovirus within cells. *Am Nat*, **165**(3):364–73.

Rieder E. and Wimmer E. (2002). *Molecular Biology of Picornaviruses*, chap. 6. Cellular Receptors of Picornaviruses: an Overview, pp. 61–70. ASM Press.

Rieder E., Xiang W., Paul A., and Wimmer E. (2003). Analysis of the cloverleaf element in a human rhinovirus type 14/poliovirus chimera: correlation of subdomain D structure, ternary protein complex formation and virus replication. *J. Gen. Virol.*, **84**(Pt 8):2203–16.

Rossmann M.G. and Palmenberg A.C. (1988). Conservation of the putative receptor attachment site in picornaviruses. *Virology*, **164**(2):373–82.

Ruckert M. and Otting G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.*, **122**(32):7793–7797.

Ryan M.D. and Flint M. (1997). Virus-encoded proteinases of the picornavirus super-group. *J. Gen. Virol.*, **78 (Pt 4)**:699–723.

Salzmann M., Pervushin K., Wider G., Senn H., and Wüthrich K. (1998). TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **95**(23):13585–90.

Seipelt J., Guarné A., Bergmann E., James M., Sommergruber W., Fita I., and Skern T. (1999). The structures of picornaviral proteinases. *Virus Res*, **62**(2):159–68.

Sekiguchi H., Isogai M., Masuta C., and Uyeda I. (2005). 3C-like protease encoded by Rice tungro spherical virus is autocatalytically processed. *Arch Virol*, **150**(3):595–601.

Semenyuk A. and Svergun D. (1991). GNOM - a program package for small-angle scattering data-processing. *J Appl Crystallogr*, **24**:537–540.

Shih S., Chiang C., Chen T., Wu C., Hsu J., Lee J., Hwang M., Li M., Chen G., and Hoc M. (2004). Mutations at KFRDI and VGK domains of enterovirus 71 3C protease affect its RNA binding and proteolytic activities. *J Biomed Sci*, **11**(2):239–248.

Skern T., Hampölz B., Guarné A., Fita I., Bergmann E., Petersen J., and James M.N. (2002). *Molecular Biology of Picornaviruses*, chap. 17: Structure and Function of Picornaviral Proteinases, pp. 199–212. ASM Press, Washington DC.

Smock R.G. and Gierasch L.M. (2009). Sending signals dynamically. *Science*, **324**(5924):198–203.

Stanway G., Cann A.J., Hauptmann R., Hughes P., Clarke L.D., Mountford R.C., Minor P.D., Schild G.C., and Almond J.W. (1983). The nucleotide sequence of poliovirus type 3 leon 12 a1b: comparison with poliovirus type 1. *Nucleic Acids Research*, **11**(16):5629–43.

Stanway G., Hovi T., Knowles N.J., and Hyypiä T. (2002). *Molecular Biology of Picornaviruses*, chap. 2: Molecular and Biological Basis of Picornaviral Taxonomy, pp. 17–24. ASM Press, Washington DC.

Stanway G., Hughes P.J., Mountford R.C., Minor P.D., and Almond J.W. (1984). The complete nucleotide sequence of a common cold virus: human rhinovirus 14. *Nucleic Acids Research*, **12**(20):7859–75.

Stock A.M., Robinson V.L., and Goudreau P.N. (2000). Two-component signal transduction. *Annu. Rev. Biochem.*, **69**:183–215.

Svergun D. and Nierhaus K. (2000). A map of protein-rRNA distribution in the 70 S Escherichia coli ribosome. *Journal of Biological Chemistry*, **275**:14432–14439.

Sweeney T.R., Roqué-Rosell N., Birtley J.R., Leatherbarrow R.J., and Curry S. (2007). Structural and mutagenic analysis of foot-and-mouth disease virus 3C protease reveals the role of the beta-ribbon in proteolysis. *J. Virol.*, **81**(1):115–24.

Tapparel C., Junier T., Gerlach D., Cordey S., Belle S.V., Perrin L., Zdobnov E.M., and Kaiser L. (2007). New complete genome sequences of human rhinoviruses shed light on their phylogeny and genomic features. *BMC Genomics*, **8**:224.

Tjandra N., Kuboniwa H., Ren H., and Bax A. (1995). Rotational dynamics of calcium-free calmodulin studied by 15N-NMR relaxation measurements. *Eur J Biochem*, **230**(3):1014–24.

Varani L., Gunderson S., Mattaj I., Kay L., Neuhaus D., and Varani G. (2000). The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat. Struct. Biol.*, **7**:329–335.

Volkman B.F., Lipson D., Wemmer D.E., and Kern D. (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science*, **291**(5512):2429–33.

Volkov V. and Svergun D. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr*, **36**:860–864.

Voss N. and Gerstein M. (2005). Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J. Mol. Biol.*, **346**(477-92).

Vranken W.F., Boucher W., Stevens T.J., Fogh R.H., Pajon A., Llinas M., Ulrich E.L., Markley J.L., Ionides J., and Laue E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**(4):687–96.

Walker P.A., Leong L.E., and Porter A.G. (1995). Sequence and structural determinants of the interaction between the 5'-noncoding region of picornavirus RNA and rhinovirus protease 3C. *J. Biol. Chem.*, **270**(24):14510–6.

Wand A.J. (2001a). Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat. Struct. Biol.*, **8**(11):926–31.

Wand A.J. (2001b). On the dynamic origins of allosteric activation. *Science*, **293**(5534):1395.

Wang C., Grey M.J., and Palmer A.G. (2001). CPMG sequences with enhanced sensitivity to chemical exchange. *J. Biomol. NMR*, **21**(4):361–6.

Wang Q.M. and Johnson R.B. (2001). Activation of human rhinovirus-14 3C protease. *Virology*, **280**(1):80–6.

Webber S.E., Okano K., Little T.L., Reich S.H., Xin Y., Fuhrman S.A., Matthews D.A., Love R.A., Hendrickson T.F., Patick A.K., Meador J.W., Ferre R.A., Brown E.L., Ford C.E., Binford S.L., and Worland S.T. (1998). Tripeptide aldehyde inhibitors of human rhinovirus 3C protease: design, synthesis, biological evaluation, and cocrystal structure solution of P1 glutamine isosteric replacements. *J. Med. Chem.*, **41**(15):2786–805.

Weber G. (1972). Ligand binding and internal equilibria in proteins. *Biochemistry*, **11**(5):864–78.

Whitmore L. and Wallace B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, **32**(Web Server issue):W668–73.

Whitmore L. and Wallace B.A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, **89**(5):392–400.

Whitten A., Cai S., and Trewhella J. (2008). MULCh: ModULes for the Analysis of Small-angle Neutron Contrast Variation Data from Biomolecular Complexes. *J Appl Crystallogr*, **41**:222–226.

Wishart D.S. and Sykes B.D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR*, **4**(2):171–80.

Worobey M. and Holmes E.C. (1999). Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.*, **80 (Pt 10)**:2535–43.

Xiang W., Harris K.S., Alexander L., and Wimmer E. (1995). Interaction between the 5′-terminal cloverleaf and 3AB/3CDpro of poliovirus is essential for RNA replication. *J. Virol.*, **69**(6):3658–67.

Yang Y., Rijnbrand R., Watowich S., and Lemon S.M. (2004). Genetic evidence for an interaction between a picornaviral cis-acting RNA replication element and 3CD protein. *J. Biol. Chem.*, **279**(13):12659–67.

Zahnley J.C. (1980). Independent heat stabilization of proteases associated with multiheaded inhibitors. Complexes of chymotrypsin, subtilisin and trypsin with chicken ovoinhibitor and with lima bean protease inhibitor. *Biochim Biophys Acta*, **613**(1):178–90.

Zell R., Sidigi K., Bucci E., Stelzner A., and Görlach M. (2002). Determinants of the recognition of enteroviral cloverleaf RNA by coxsackievirus B3 proteinase 3C. *RNA*, **8**(2):188–201.

Zuiderweg E.R.P. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry*, **41**(1):1–7.

Zweckstetter M. and Bax A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J. Am. Chem. Soc.*, **122**(15):3791–3792.

A

Chemical Shift Tables

A.1 Chemical Shifts for 3C

Table A.1: Table of 3C chemical shifts

Residue	H^N	N	C′	C^α	C^β
Gly 1	8.42	111.16	-	45.38	-
Pro 2	-	-	178.17	64.73	32.47
Asn 3	8.85	117.55	177.05	54.27	-
Thr 4	7.74	116.23	175.13	66.81	-
Glu 5	8.46	121.18	179.33	60.27	-
Phe 6	8.43	121.07	177.61	61.56	39.51
Ala 7	8.27	122.80	179.41	55.89	-
Leu 8	8.51	116.90	179.82	57.97	41.80
Ser 9	7.73	115.42	-	61.54	63.00
Leu 10	7.59	121.37	-	58.10	-
Leu 11	-	-	-	-	-
Arg 12	-	-	178.73	-	-
Lys 13	7.72	113.86	178.13	-	-
Asn 14	8.61	116.83	173.25	56.29	-
Ile 15	7.41	121.29	-	61.32	-
Met 16	8.71	124.25	-	-	-
Thr 17	-	-	175.14	63.23	-
Ile 18	9.69	131.21	175.35	58.78	39.00
Thr 19	9.37	124.45	176.28	61.98	69.49
Thr 20	9.34	120.34	176.06	59.45	-
Ser 21	8.93	116.10	175.61	61.90	62.92
Lys 22	8.36	118.52	176.31	54.73	32.30

Continued on next page

Table A.1 – continued from previous page

Residue	H^N	N	C′	C^α	C^β
Gly 23	7.60	108.21	-	44.43	-
Glu 24	-	-	-	-	-
Phe 25	-	-	-	-	-
Thr 26	-	-	-	-	-
Gly 27	-	-	-	-	-
Leu 28	-	-	175.68	52.91	-
Gly 29	9.25	117.06	-	43.43	-
Ile 30	8.76	126.77	174.09	60.02	37.21
His 31	7.25	108.41	172.18	54.89	-
Asp 32	10.17	122.94	174.39	57.64	41.15
Arg 33	8.32	125.83	175.35	56.68	29.20
Val 34	8.91	121.95	175.64	63.06	-
Cys 35	9.60	123.90	-	55.13	-
Val 36	-	-	-	-	-
Ile 37	-	-	-	-	-
Pro 38	-	-	-	-	-
Thr 39	-	-	-	-	-
His 40	-	-	-	-	-
Ala 41	-	-	177.70	54.42	18.13
Gln 42	7.97	110.22	-	56.04	-
Pro 43	-	-	-	-	-
Gly 44	9.16	112.61	173.45	44.15	-
Asp 45	8.40	121.32	175.66	56.33	41.50
Asp 46	7.75	114.81	174.03	52.89	44.80

Continued on next page

Table A.1 – continued from previous page

Residue	H^N	N	C′	C^α	C^β
Val 47	8.74	111.67	174.92	59.33	36.29
Leu 48	7.87	115.95	177.54	53.55	42.71
Val 49	9.36	122.78	176.82	61.58	32.49
Asn 50	9.81	129.18	176.01	54.42	37.31
Gly 51	8.29	101.32	173.39	45.91	-
Gln 52	7.74	120.83	175.48	54.26	30.74
Lys 53	8.85	128.03	175.77	58.42	32.24
Ile 54	9.34	131.08	174.68	58.57	40.59
Arg 55	8.89	126.35	175.26	57.17	31.45
Val 56	8.38	125.26	176.20	62.45	32.47
Lys 57	8.82	129.17	176.14	58.17	33.60
Asp 58	7.58	114.68	173.17	53.83	44.81
Lys 59	8.13	118.26	174.75	55.17	-
Tyr 60	8.38	119.00	173.28	56.81	-
Lys 61	8.55	123.46	-	-	-
Leu 62	-	-	-	-	-
Val 63	-	-	-	-	-
Asp 64	-	-	-	-	-
Pro 65	-	-	177.61	-	-
Glu 66	8.20	116.91	-	-	-
Asn 67	-	-	-	-	-
Ile 68	-	-	-	-	-
Asn 69	-	-	-	-	-
Leu 70	-	-	-	-	-

Continued on next page

Table A.1 – continued from previous page

Residue	H^N	N	C′	C^α	C^β
Glu 71	-	-	-	-	-
Leu 72	-	-	-	-	-
Thr 73	-	-	171.63	62.18	72.00
Val 74	8.69	127.20	175.83	61.62	-
Leu 75	9.28	126.40	176.09	53.46	45.11
Thr 76	8.60	117.57	175.24	62.18	69.12
Leu 77	9.19	126.35	176.46	54.16	44.38
Asp 78	8.54	121.50	173.74	52.63	40.18
Arg 79	7.21	119.34	174.81	54.39	33.56
Asn 80	8.85	120.69	176.61	55.85	39.28
Glu 81	8.36	120.22	-	57.87	30.41
Lys 82	8.35	121.44	-	55.05	35.20
Phe 83	9.18	121.76	176.39	57.75	40.00
Arg 84	8.71	124.25	176.44	56.96	-
Asp 85	8.74	126.04	179.05	54.27	40.54
Ile 86	8.88	121.47	177.97	59.95	-
Arg 87	8.67	121.55	178.74	61.07	30.25
Gly 88	8.60	103.26	175.18	45.99	-
Phe 89	7.64	115.75	176.61	56.73	39.71
Ile 90	7.38	122.07	176.28	64.07	-
Ser 91	8.44	122.79	174.19	56.92	66.08
Glu 92	8.95	124.36	175.88	55.66	30.33
Asp 93	8.40	123.10	174.53	53.85	40.00
Leu 94	7.67	120.16	176.94	55.09	42.53

Continued on next page

Table A.1 – continued from previous page

Residue	H^N	N	C′	C^α	C^β
Glu 95	7.98	120.13	179.00	59.00	29.35
Gly 96	8.91	114.11	174.25	46.27	-
Val 97	7.45	120.11	175.53	62.44	33.18
Asp 98	8.24	126.55	175.29	54.48	41.11
Ala 99	8.11	127.28	176.42	51.46	23.41
Thr 100	9.34	116.83	172.49	60.92	72.63
Leu 101	9.37	126.67	-	53.23	-
Val 102	-	-	-	-	-
Val 103	-	-	-	-	-
His 104	-	-	-	-	-
Ser 105	-	-	-	-	-
Asn 106	-	-	-	-	-
Asn 107	-	-	-	-	-
Phe 108	-	-	-	-	-
Thr 109	-	-	-	-	-
Asn 110	-	-	-	-	-
Thr 111	-	-	-	-	-
Ile 112	-	-	175.36	60.94	-
Leu 113	9.03	128.10	-	53.09	-
Glu 114	8.56	123.52	-	55.99	-
Val 115	7.32	115.25	176.13	61.45	-
Gly 116	7.89	110.04	-	44.88	-
Pro 117	-	-	-	-	-
Val 118	-	-	-	-	-

Continued on next page

Table A.1 – continued from previous page

Residue	H^N	N	C′	C^α	C^β
Thr 119	8.24	113.38	172.70	60.14	71.37
Met 120	9.04	125.97	-	55.54	-
Ala 121	-	-	178.55	52.87	20.10
Gly 122	8.28	107.60	171.73	46.34	-
Leu 123	8.09	128.02	-	-	-
Ile 124	-	-	-	-	-
Asn 125	-	-	-	-	-
Leu 126	-	-	-	-	-
Ser 127	-	-	-	-	-
Ser 128	-	-	-	-	-
Thr 129	-	-	-	-	-
Pro 130	-	-	-	-	-
Thr 131	-	-	-	-	-
Asn 132	-	-	-	-	-
Arg 133	-	-	-	-	-
Met 134	-	-	-	-	-
Ile 135	-	-	-	-	-
Arg 136	-	-	173.94	-	-
Tyr 137	7.40	120.75	-	-	-
Asp 138	-	-	-	-	-
Tyr 139	-	-	-	-	-
Ala 140	-	-	-	-	-
Thr 141	-	-	-	-	-
Lys 142	-	-	-	-	-

Continued on next page

Table A.1 – continued from previous page

Residue	H^N	N	C′	C^α	C^β
Thr 143	-	-	174.90	-	-
Gly 144	8.28	110.72	-	44.86	-
Gln 145	-	-	-	-	-
Cys 146	-	-	-	-	-
Gly 147	-	-	170.13	-	-
Gly 148	8.29	103.89	-	46.83	-
Val 149	-	-	174.63	63.10	-
Leu 150	8.84	132.46	176.72	54.27	-
Cys 151	9.92	124.58	170.80	56.61	-
Ala 152	8.88	119.69	176.45	51.71	-
Thr 153	8.21	114.80	-	66.40	-
Gly 154	-	-	172.46	-	-
Lys 155	7.62	118.66	173.97	55.33	39.19
Ile 156	8.27	120.57	174.71	61.10	40.76
Phe 157	8.78	122.54	177.96	56.91	42.43
Gly 158	7.79	103.09	-	47.84	-
Ile 159	-	-	-	-	-
His 160	-	-	-	-	-
Val 161	-	-	-	-	-
Gly 162	-	-	-	-	-
Gly 163	-	-	-	-	-
Asn 164	-	-	176.68	51.94	-
Gly 165	8.89	112.04	-	45.89	-
Arg 166	-	-	-	-	-

Continued on next page

Table A.1 – continued from previous page

Residue	H^N	N	C′	C^α	C^β
Gln 167	-	-	-	-	-
Gly 168	-	-	-	-	-
Phe 169	-	-	175.81	-	-
Ser 170	9.93	119.85	174.48	-	-
Ala 171	-	-	-	-	-
Gln 172	-	-	-	-	-
Leu 173	-	-	173.14	-	-
Lys 174	5.65	114.97	177.11	54.13	34.89
Lys 175	9.16	124.02	179.72	61.06	32.66
Gln 176	8.60	114.36	176.69	58.10	28.39
Tyr 177	6.82	116.19	176.75	55.62	35.47
Phe 178	7.51	114.14	174.83	56.32	40.97
Val 179	7.24	117.70	176.17	62.13	33.28
Glu 180	8.58	125.21	176.56	56.96	30.70
Lys 181	8.46	123.85	175.84	56.68	33.31
Gln 182	8.04	127.67	-	-	-

A.2 Chemical Shifts for 3CI

Table A.2: Table of 3CI chemical shifts

Residue	H^N	N	C′	C^α	C^β	H^α	$H^{\alpha a}$	$H^{\alpha b}$	H^β	$H^{\beta a}$	$H^{\beta b}$
Gly 1	8.44	111.06	-	45.17	-	-	-	-	-	-	-
Pro 2	-	-	177.95	64.51	32.42	4.33	-	-	-	1.87	2.26
Asn 3	8.89	117.19	176.83	54.03	38.68	4.76	-	-	-	-	-
Thr 4	7.77	116.26	174.85	66.69	69.09	3.74	-	-	4.11	-	-
Glu 5	8.51	121.02	179.16	60.11	29.13	3.84	-	-	-	1.95	1.95
Phe 6	8.49	120.96	177.47	61.42	39.53	4.16	-	-	-	-	-
Ala 7	8.34	122.67	179.14	55.53	19.13	3.6	-	-	-	-	-
Leu 8	8.59	116.96	179.48	57.68	41.78	3.84	-	-	-	-	-
Ser 9	7.75	115.43	176.96	61.37	62.9	4.07	-	-	-	3.83	3.83
Leu 10	7.57	121.22	-	57.88	-	-	-	-	-	-	-
Leu 11	-	-	-	-	-	-	-	-	-	-	-
Arg 12	-	-	178.58	58.55	-	4.03	-	-	-	1.82	1.82
Lys 13	7.73	113.83	177.95	57.17	35.49	4.59	-	-	-	-	-
Asn 14	8.62	116.86	172.68	56.01	-	5	-	-	-	-	-
Ile 15	7.33	121.04	174.47	60.86	39.3	5.01	-	-	1.76	-	-
Met 16	8.69	124.11	174.29	53.66	38.26	-	-	-	-	-	-
Thr 17	8.87	118.26	174.86	62.97	69.54	4.96	-	-	-	-	-
Ile 18	9.67	130.69	174.88	58.58	39.27	5.4	-	-	1.7	-	-
Thr 19	9.35	124.01	175.96	61.71	69.41	5.39	-	-	3.91	-	-
Thr 20	9.31	120.19	175.88	59.16	71.23	5.25	-	-	4.63	-	-
Ser 21	8.96	116.09	175.31	61.58	62.77	4.36	-	-	-	4.02	4.02
Lys 22	8.38	118.48	176.04	54.28	32.17	4.49	-	-	-	1.68	2.24
Gly 23	7.59	108.08	171.32	44.1	-	-	3.81	4.41	-	-	-
Glu 24	8.26	117.44	176.35	55.98	-	-	-	-	-	-	-
Phe 25	9.47	120.22	-	56.92	42.85	-	-	-	-	-	-

Continued on next page

Table A.2 – continued from previous page

Residue	H^N	N	C′	C^α	C^β	H^α	$H^{\alpha a}$	$H^{\alpha b}$	H^β	$H^{\beta a}$	$H^{\beta b}$
Thr 26	-	-	-	-	-	-	-	-	-	-	-
Gly 27	-	-	171.19	43.99	-	-	3.11	4.93	-	-	-
Leu 28	7.43	121.98	175.34	52.74	-	4.89	-	-	-	-	-
Gly 29	9.22	116.69	172.19	43.21	-	-	1.08	3.04	-	-	-
Ile 30	8.82	126.63	173.93	59.75	37.14	3.83	-	-	1.47	-	-
His 31	7.24	107.8	171.88	54.65	30.1	4.39	-	-	-	2.99	3.7
Asp 32	10.2	122.85	174.07	57.4	41.08	3.98	-	-	-	2.61	2.61
Arg 33	8.37	125.63	175.01	56.43	29.26	4.63	-	-	-	1.55	3
Val 34	8.92	121.77	175.41	62.71	33.33	4.96	-	-	1.82	-	-
Cys 35	9.64	123.85	171.52	54.84	33.39	5.46	-	-	-	2.46	2.46
Val 36	8.4	114.47	173.33	59.06	36.23	5.54	-	-	1.99	-	-
Ile 37	8.22	116.08	171.66	58.15	-	-	-	-	-	-	-
Pro 38	-	-	176.73	61.98	32.55	4.83	-	-	-	0.85	2.32
Thr 39	7.58	124.62	177.76	65.71	68.25	3.65	-	-	3.97	-	-
His 40	10.05	120.79	174.79	58.76	25.78	4.26	-	-	-	2.1	2.92
Ala 41	6.95	121.7	177.47	53.92	17.77	3.25	-	-	-0.13	Hb*	-
Gln 42	8.1	110.52	172.65	55.7	28	-	-	-	-	-	-
Pro 43	-	-	176.85	62.77	-	4.22	-	-	-	1.56	1.73
Gly 44	9.11	112.34	173.18	43.86	-	-	3.7	4.33	-	-	-
Asp 45	8.42	121.2	175.37	56.06	41.46	4.46	-	-	-	2.42	2.81
Asp 46	7.76	114.54	173.74	52.57	44.82	5.11	-	-	-	2.27	2.5
Val 47	8.77	111.55	174.61	59.07	36.04	4.6	-	-	1.98	-	-
Leu 48	7.87	116	177.25	53.34	42.75	5.18	-	-	-	0.81	1.86
Val 49	9.4	122.74	176.62	61.29	32.49	4.49	-	-	1.97	-	-
Asn 50	9.87	128.92	175.73	54.14	37.4	4.43	-	-	-	3.03	3.03
Gly 51	8.28	101.21	173.11	45.66	-	-	3.46	4.24	-	-	-

Continued on next page

Table A.2 – continued from previous page

Residue	H^N	N	C′	C^α	C^β	H^α	$H^{\alpha a}$	$H^{\alpha b}$	H^β	$H^{\beta a}$	$H^{\beta b}$
Gln 52	7.76	120.64	175.19	53.93	30.81	4.52	-	-	-	1.94	2.08
Lys 53	8.86	127.68	175.48	58.13	32.3	4.38	-	-	-	1.49	1.7
Ile 54	9.37	130.92	174.4	58.29	40.62	4.42	-	-	1.73	-	-
Arg 55	8.91	126.14	174.97	56.88	31.52	4.43	-	-	-	1.89	1.89
Val 56	8.4	125.11	175.9	62.17	32.42	3.96	-	-	1.96	-	-
Lys 57	8.83	128.93	175.85	57.87	33.65	4.1	-	-	-	1.38	1.62
Asp 58	7.58	114.41	172.91	53.55	44.82	4.72	-	-	-	2.49	2.6
Lys 59	8.15	118.11	174.4	54.97	36.04	5.29	-	-	-	1.29	1.66
Tyr 60	8.35	119.32	172.89	56.59	40.38	4.73	-	-	-	2.85	3.04
Lys 61	8.54	123.19	176.08	55.1	33.58	4.61	-	-	-	1.67	1.67
Leu 62	8.74	123.09	176.81	55.62	33.58	4.46	-	-	-	0.9	1.55
Val 63	8.07	117.64	174.29	59.32	36.13	4.6	-	-	1.74	-	-
Asp 64	8.99	125.7	176.28	53.13	41.36	-	-	-	-	-	-
Pro 65	-	-	177.35	65.61	32.2	4.31	-	-	-	1.53	1.71
Glu 66	8.17	116.9	175.51	55.75	29.93	4.32	-	-	-	1.96	2.21
Asn 67	8.68	114.8	174.65	54.48	37.63	4.12	-	-	-	2.77	3.12
Ile 68	8.47	121.71	176.04	60.87	36.06	3.94	-	-	1.87	-	-
Asn 69	8.5	124.12	174.43	56.04	40.64	4.3	-	-	-	2.89	3.42
Leu 70	7.26	122.02	176.53	54.24	42.12	-	-	-	-	-	-
Glu 71	7.79	116.44	174.65	55.1	30.04	4.19	-	-	-	1.84	2.64
Leu 72	6.44	115.94	175.09	52.94	47.47	4.7	-	-	-	0.9	1.34
Thr 73	8.6	118.53	171.35	61.99	71.8	4.83	-	-	3.5	-	-
Val 74	8.65	127.19	175.72	61.45	32.62	4.38	-	-	1.35	-	-
Leu 75	9.31	126.58	175.78	53.29	45.21	5	-	-	-	1.19	1.91
Thr 76	8.61	117.42	174.94	61.93	69.22	4.95	-	-	3.86	-	-
Leu 77	9.21	126.12	176.23	53.87	44.24	4.5	-	-	-	1.15	1.7

Continued on next page

Table A.2 – continued from previous page

Residue	H^N	N	C′	C^α	C^β	H^α	$H^{\alpha a}$	$H^{\alpha b}$	H^β	$H^{\beta a}$	$H^{\beta b}$
Asp 78	8.54	121.19	173.43	52.38	40.24	4.7	-	-	-	2	3.11
Arg 79	7.24	119.31	174.52	54.1	33.65	4.62	-	-	-	1.64	1.77
Asn 80	8.82	120.3	176.37	55.54	39.33	4.39	-	-	-	2.64	2.64
Glu 81	8.37	120.06	175.26	57.57	30.2	3.9	-	-	-	1.56	1.56
Lys 82	8.36	121.08	177.42	54.75	35.23	4.13	-	-	-	1.61	1.61
Phe 83	9.24	121.8	176.08	57.44	39.53	4.58	-	-	-	1.99	2.81
Arg 84	8.79	124.28	176.06	56.69	29.97	4.05	-	-	-	1.72	1.83
Asp 85	8.72	125.97	178.78	53.98	40.37	5.06	-	-	-	2.97	2.97
Ile 86	8.89	121.32	177.69	59.72	38.1	4.57	-	-	2.54	-	-
Arg 87	8.71	121.45	178.48	60.76	30.29	3.57	-	-	-	1.68	1.95
Gly 88	8.6	103.21	174.91	45.75	-	-	3.36	3.72	-	-	-
Phe 89	7.64	115.57	176.34	56.42	39.72	4.71	-	-	-	2.87	3.22
Ile 90	7.38	121.98	175.97	63.84	37.14	4.46	-	-	-	-	-
Ser 91	8.47	122.62	173.93	56.61	66.06	4.8	-	-	-	3.26	3.81
Glu 92	8.97	124.22	175.55	55.4	30.29	4.5	-	-	-	1.8	2.31
Asp 93	8.44	123.08	174.27	53.56	39.98	4.76	-	-	-	2.42	2.69
Leu 94	7.69	120.06	176.61	54.91	42.62	3.72	-	-	-	0.76	0.9
Glu 95	7.99	119.71	178.79	58.7	29.39	3.88	-	-	-	1.93	1.93
Gly 96	8.93	113.65	173.96	46.01	-	-	3.69	4	-	-	-
Val 97	7.48	119.97	175.26	62.14	33.2	3.87	-	-	1.82	-	-
Asp 98	8.29	126.3	174.89	54.22	40.97	4.36	-	-	-	2.44	2.63
Ala 99	8.06	127.29	176.09	51.21	23.52	4.8	-	-	1.23	Hb*	-
Thr 100	9.38	116.71	172.22	60.57	72.71	4.98	-	-	3.64	-	-
Leu 101	9.37	126.42	175.03	52.95	46.57	5.05	-	-	-	0.65	0.65
Val 102	8.98	127.33	174.42	61.68	-	-	-	-	-	-	-
Val 103	-	-	-	-	-	-	-	-	-	-	-

Continued on next page

Table A.2 – continued from previous page

Residue	H^N	N	C′	C^α	C^β	H^α	$H^{\alpha a}$	$H^{\alpha b}$	H^β	$H^{\beta a}$	$H^{\beta b}$
His 104	-	-	-	-	-	-	-	-	-	-	-
Ser 105	-	-	-	-	-	-	-	-	-	-	-
Asn 106	-	-	-	-	-	-	-	-	-	-	-
Asn 107	-	-	-	-	-	-	-	-	-	-	-
Phe 108	-	-	-	-	-	-	-	-	-	-	-
Thr 109	-	-	-	-	-	-	-	-	-	-	-
Asn 110	-	-	-	-	-	-	-	-	-	-	-
Thr 111	-	-	-	-	-	-	-	-	-	-	-
Ile 112	-	-	175.12	60.7	60.93	4.44	-	-	1.66	-	-
Leu 113	9.01	127.42	175.81	52.7	-	4.53	-	-	-	0.61	1.3
Glu 114	8.56	123.16	176.45	55.63	28.54	4.49	-	-	-	1.81	2
Val 115	7.47	115.22	175.77	61.02	32.36	4.22	-	-	2.02	-	-
Gly 116	7.87	109.85	170.85	44.82	-	-	-	-	-	-	-
Pro 117	-	-	178.41	62.46	-	-	-	-		-	-
Val 118	8.23	115.41	174.66	58.62	-	5.26	-	-	1.88	-	-
Thr 119	8.21	113.21	172.48	59.82	71.37	4.53	-	-	3.82	-	-
Met 120	9.07	126.41	174.95	55.51	30.1	4.63	-	-	-	1.9	2.15
Ala 121	9.43	132.75	178.19	52.44	20.55	4.45	-	-	1.31	Hb*	-
Gly 122	8.18	107.54	171.48	46.4	-	-	3.5	4.16	-	-	-
Leu 123	8.1	128.06	176.73	55.02	43.23	4.83	-	-	-	1.3	1.66
Ile 124	8.96	127.53	173.88	57.8	40.88	4.12	-	-	1.3	-	-
Asn 125	8.09	123.15	173.99	52.21	38.53	4.52	-	-		2.18	2.32
Leu 126	7.31	125.47	175.16	52.01	44.7	4.18	-	-		0.39	1.09
Ser 127	8.89	119.06	175.87	59.86	61.69	3.74	-	-	-	2.82	2.82
Ser 128	8.37	108.22	173.35	60.05	61.7	3.99	-	-	-	3.85	3.85
Thr 129	7.99	118.46	172.57	59.75	70.6	-	-	-	-	-	-

Continued on next page

Table A.2 – continued from previous page

Residue	H^N	N	C′	C^α	C^β	H^α	$H^{\alpha a}$	$H^{\alpha b}$	H^β	$H^{\beta a}$	$H^{\beta b}$
Pro 130	-	-	175.93	64.02	31.91	4.29	-	-	-	1.8	2.21
Thr 131	8.92	124.81	171.77	62.51	72.92	5.09	-	-	4.03	-	-
Asn 132	8.29	123.43	173.99	49.5	41.91	5.34	-	-	-	2.36	2.67
Arg 133	8.06	111.69	173.19	57.16	27.22	4.93	-	-	-	3.52	3.52
Met 134	8.39	113.28	176.56	55.49	38.66	5.01	-	-	-	1.52	2.41
Ile 135	9.81	125.62	174.55	60.59	41.18	4.57	-	-	1.62	-	-
Arg 136	9.49	127.61	174.54	54.08	35.9	5.21	-	-	-	1.34	1.62
Tyr 137	7.43	120.54	176.2	52.7	38.58	4.84	-	-	-	2.55	3.13
Asp 138	9.64	126.01	173.53	53.7	39.63	4.9	-	-	-	2.7	2.7
Tyr 139	7.58	123.26	173.13	57.88	43.46	4.27	-	-	-	2.55	3.04
Ala 140	7.39	129.22	175.35	50.56	17	3.91	-	-	0.96	Hb*	-
Thr 141	7.66	117.22	174.66	62.47	-	-	-	-		-	-
Lys 142	-	-	-	-	-	-	-	-	-	-	-
Thr 143	-	-	174.62	-	-	4.46	-	-	3.83	-	-
Gly 144	8.3	110.5	171.9	44.45	-	-	-	-	-	-	-
Gln 145	-	-	176.77	-	-	4.32	-	-	-	-	-
Cys 146	8.2	115.53	175.97	63.03	-	-	-	-	-	-	-
Gly 147	10.55	114.89	172.18	44.84	-	-	3.53	5.72	-	-	-
Gly 148	8.21	103.38	171.7	46.33	-	-	3.28	4.26	-	-	-
Val 149	9.34	124.38	174.5	63.15	-	4.24	-	-	1.95	-	-
Leu 150	8.94	132.74	176.43	53.89	43.14	5.04	-	-	-	1.35	1.73
Cys 151	10	124.29	170.48	56.26	33.13	5.01	-	-	-	2.87	3.15
Ala 152	8.92	119.49	176.18	51.36	22.61	4.43	-	-	1.36	Hb*	-
Thr 153	8.25	114.81	-	66.15	-	-	-	-	-	-	-
Gly 154	-	-	172.16	46.2	-	-	2.6	3.65	-	-	-
Lys 155	7.67	118.72	173.65	55.13	39.15	4.96	-	-	-	0.92	0.92

Continued on next page

Table A.2 – continued from previous page

Residue	H^N	N	C′	C^α	C^β	H^α	$H^{\alpha a}$	$H^{\alpha b}$	H^β	$H^{\beta a}$	$H^{\beta b}$
Ile 156	8.34	120.72	174.48	60.84	40.86	3.9	-	-	1.26	-	-
Phe 157	8.88	122.17	177.72	56.75	42.56	5.47	-	-	-	2.45	3.24
Gly 158	7.81	103.28	170.19	47.71	-	-	4.11	4.75	-	-	-
Ile 159	8.53	115.3	175.36	57.29	41.78	5.5	-	-	1.83	-	-
His 160	10.44	125.76	177.78	60.12	31.91	4.72	-	-	-	3.19	3.19
Val 161	9.3	117.77	176.49	61.32	36.1	5.38	-	-	2.73	-	-
Gly 162	7.37	107.34	170.15	46.96	-	-	4.07	4.64	-	-	-
Gly 163	8.99	103.23	172.28	47.51	-	-	4.41	4.41	-	-	-
Asn 164	7.94	116.45	175.83	51.39	39.25	4.93	-	-		1.32	2.87
Gly 165	9.06	112.26	174.35	45.4	-	-	3.5	4.5	-	-	-
Arg 166	7.55	117.31	173.84	56.31	32.92	4.3	-	-	-	1.53	1.7
Gln 167	8.18	117.18	175.22	54.18	33.08	4.84	-	-	-	1.4	1.57
Gly 168	8.55	109.4	169.32	44.22	-	-	3.52	4.92	-	-	-
Phe 169	8.39	113.11	175.53	55.41	43.73	5.69	-	-	-	2.07	3.04
Ser 170	9.96	119.85	174.22	58.44	68.58	5.68	-	-	-	3.45	3.88
Ala 171	9.37	125.61	177.41	51.59	21.26	4.6	-	-	1.41	Hb*	-
Gln 172	7.93	124.03	176.75	57.16	30.37	-	-	-	-	-	-
Leu 173	-	-	172.76	-	-	-	-	-	-	-	-
Lys 174	5.67	114.72	176.88	53.81	34.81	4.42	-	-	-	-0.37	1.44
Lys 175	9.18	123.72	179.53	60.78	32.68	3.75	-	-	-	1.68	1.68
Gln 176	8.63	114.33	176.46	57.91	28.49	3.95	-	-	-	1.88	1.88
Tyr 177	6.85	115.78	176.56	55.29	35.59	4.38	-	-	-	1.84	2.23
Phe 178	7.5	113.81	174.61	56.11	40.88	5.03	-	-	-	2.55	3.41
Val 179	7.19	117.77	175.91	61.9	33.2	4.05	-	-	1.99	-	-
Glu 180	8.6	125	176.27	56.71	30.75	4.2	-	-	-	1.95	1.95
Lys 181	8.46	123.61	175.55	56.43	33.24	4.22	-	-	-	1.7	1.7

Continued on next page

Table A.2 – continued from previous page

Residue	H^N	N	C′	C^α	C^β	H^α	$H^{\alpha a}$	$H^{\alpha b}$	H^β	$H^{\beta a}$	$H^{\beta b}$
Gln 182	8.05	127.44	180.53	57.37	30.66	-	-	-	-	-	-

A.3 Chemical Shifts for 3CIR

Table A.3: Table of 3CIR chemical shifts

Residue	H^N	N	C′	C^α
Gly1	8.4	110.9	-	45.01
Pro2	-	-	177.47	64.41
Asn3	8.89	116.12	-	53.52
Thr4	-	-	-	-
Glu5	-	-	-	-
Phe6	-	-	-	-
Ala7	-	-	179.02	55.43
Leu8	8.81	116.72	-	57.82
Ser9	-	-	-	-
Leu10	-	-	-	-
Leu11	-	-	-	-
Arg12	-	-	-	-
Lys13	-	-	-	-
Asn14	-	-	172.79	-
Ile15	7.4	120.9	-	60.91
Met16	8.71	123.83	-	53.65
Thr17	-	-	174.91	63
Ile18	9.71	130.79	174.99	58.59
Thr19	9.31	123.89	176.04	61.78
Thr20	9.32	120.2	175.82	59.24
Ser21	8.96	115.98	175.34	61.66
Lys22	8.39	118.5	-	54.24
Gly23	7.6	108.01	-	44.09
Glu24	-	-	-	-

Continued on next page

Table A.3 – continued from previous page

Residue	H^N	N	C′	C^α
Phe25	-	-	-	-
Thr26	-	-	-	-
Gly27	-	-	171.22	43.84
Leu28	7.48	121.33	175.36	52.73
Gly29	9.39	117.1	-	43.01
Ile30	8.79	126.1	173.87	59.59
His31	7.18	107.32	171.92	54.6
Asp32	9.8	121.43	-	57.48
Arg33	-	-	174.9	56.87
Val34	8.8	121.26	175.43	62.57
Cys35	9.66	123.16	171.49	54.81
Val36	8.52	113.69	173.45	59.05
Ile37	8.25	116.02	-	58.15
Pro38	-	-	176.83	61.99
Thr39	7.58	124.8	177.76	65.63
His40	10.01	120.57	174.83	58.75
Ala41	6.96	121.66	-	53.94
Gln42	8.1	110.42	-	55.67
Pro43	-	-	176.85	62.73
Gly44	9.16	112.3	173.19	43.86
Asp45	8.44	121.24	175.41	56.1
Asp46	7.79	114.55	173.55	52.59
Val47	8.78	111.47	174.54	59.13
Leu48	7.8	115.87	177.19	53.36
Val49	9.43	122.91	-	61.35
Asn50	9.81	128.41	-	54.14

Continued on next page

Table A.3 – continued from previous page

Residue	H^N	N	C′	C^α
Gly51	8.25	101.07	172.91	45.63
Gln52	7.76	120.45	175.23	53.97
Lys53	8.86	127.63	175.3	58.04
Ile54	9.32	130.73	174.65	58.27
Arg55	9.03	126.02	174.96	56.97
Val56	8.37	124.75	175.94	62.01
Lys57	8.86	129.14	175.82	57.84
Asp58	7.58	114.39	172.88	53.56
Lys59	8.13	117.9	174.45	55.01
Tyr60	8.4	119.31	172.9	56.63
Lys61	8.54	123.12	176.13	55.09
Leu62	8.77	123.06	176.84	55.58
Val63	8.08	117.55	174.32	59.38
Asp64	8.99	125.65	-	53.12
Pro65	-	-	177.39	65.58
Glu66	8.19	116.68	175.54	55.71
Asn67	8.69	114.82	174.61	54.52
Ile68	8.5	121.7	176.04	60.79
Asn69	8.51	124.19	174.42	56.04
Leu70	7.26	121.9	176.56	54.26
Glu71	7.79	116.45	174.65	55.03
Leu72	6.44	115.91	175.08	52.92
Thr73	8.61	118.44	171.35	61.98
Val74	8.65	127.05	175.72	61.23
Leu75	9.37	126.56	175.74	53.33
Thr76	8.57	116.77	175.04	61.68

Continued on next page

Table A.3 – continued from previous page

Residue	H^N	N	C′	C^α
Leu77	9.25	126.13	176.11	54.03
Asp78	8.6	121.44	173.14	52.29
Arg79	7.33	117.96	174.6	54.23
Asn80	9.43	118.73	-	56.62
Glu81	-	-	-	-
Lys82	-	-	177.68	54.67
Phe83	9.17	120.23	-	56.7
Arg84	9.24	124.54	175.62	56.59
Asp85	8.85	126.08	178.48	53.75
Ile86	8.88	121.29	177.82	59.92
Arg87	8.61	121.49	-	60.73
Gly88	8.58	103.11	175.09	45.73
Phe89	7.64	115.94	176.75	54.64
Ile90	7.48	122.27	-	63.92
Ser91	8.4	122.66	173.84	56.76
Glu92	8.96	123.97	-	55.32
Asp93	8.43	123.04	-	53.72
Leu94	7.71	120.17	177.03	55
Glu95	7.96	119.41	178.69	58.46
Gly96	8.92	113.49	173.94	45.94
Val97	7.33	119.54	175.27	61.93
Asp98	8.33	126.67	174.66	54.07
Ala99	8.05	126.8	175.99	51.31
Thr100	9.37	118.25	172	61.1
Leu101	9.48	127.36	175.03	53.02
Val102	9	127.37	-	61.58

Continued on next page

Table A.3 – continued from previous page

Residue	H^N	N	C′	C^α
Val103	-	-	-	-
His104	-	-	-	-
Ser105	-	-	-	-
Asn106	-	-	-	-
Asn107	-	-	-	-
Phe108	-	-	-	-
Thr109	-	-	-	-
Asn110	-	-	-	-
Thr111	-	-	-	-
Ile112	-	-	175.12	60.79
Leu113	9.03	127.41	175.8	52.77
Glu114	8.59	122.99	176.5	55.35
Val115	7.45	115.39	175.91	61.12
Gly116	7.89	109.98	-	44.72
Pro117	-	-	-	-
Val118	-	-	-	58.62
Thr119	8.24	113.34	-	59.96
Met120	9.05	126.03	174.94	55.54
Ala121	9.45	132.57	178.15	52.35
Gly122	8.18	107.59	171.47	46.32
Leu123	8.09	127.99	176.71	55.01
Ile124	8.96	127.57	173.91	57.8
Asn125	8.1	123.16	173.96	52.29
Leu126	7.31	125.39	175.19	52.04
Ser127	8.9	118.93	175.9	59.91
Ser128	8.38	108.16	173.34	59.98

Continued on next page

Table A.3 – continued from previous page

Residue	H^N	N	C′	C^α
Thr129	8	118.46	-	59.78
Pro130	-	-	175.96	63.95
Thr131	8.92	124.77	171.78	62.54
Asn132	8.29	123.34	173.98	49.59
Arg133	8.05	111.63	173.23	57.13
Met134	8.38	113.21	176.62	55.52
Ile135	9.82	125.41	174.54	60.57
Arg136	9.49	127.4	174.61	54.07
Tyr137	7.43	120.41	176.17	52.83
Asp138	9.61	125.74	173.56	53.78
Tyr139	7.59	123.28	173.1	57.95
Ala140	7.41	129.18	175.39	50.63
Thr141	7.66	117.1	-	62.44
Lys142	-	-	-	-
Thr143	-	-	-	-
Gly144	-	-	-	-
Gln145	-	-	176.74	-
Cys146	8.23	115.56	-	-
Gly147	-	-	-	-
Gly148	-	-	-	-
Val149	-	-	174.5	63.13
Leu150	8.9	132.73	176.56	53.89
Cys151	10.09	123.98	170.7	56.66
Ala152	8.99	118.88	176.02	51.35
Thr153	8.13	115.22	-	66.88
Gly154	-	-	171.19	45.83

Continued on next page

Table A.3 – continued from previous page

Residue	H^N	N	C′	C^α
Lys155	7.68	118.39	173.59	54.8
Ile156	8.37	120.89	174.54	60.39
Phe157	8.86	122.24	-	57.03
Gly158	7.79	103.1	170.04	47.44
Ile159	8.51	115.27	175.45	57.36
His160	10.48	125.8	177.76	60.06
Val161	9.31	117.74	176.53	61.29
Gly162	7.37	107.28	170.18	46.82
Gly163	8.96	103.08	172.26	47.44
Asn164	7.93	116.31	175.79	51.47
Gly165	9.06	112.31	174.39	45.37
Arg166	7.56	117.28	173.93	56.31
Gln167	8.2	117.16	175.2	54.23
Gly168	8.53	109.38	169.28	44.09
Phe169	8.39	113.26	175.5	55.48
Ser170	9.96	119.79	174.26	58.53
Ala171	9.39	125.64	177.41	51.62
Gln172	7.94	123.96	-	57.12
Leu173	-	-	-	-
Lys174	5.67	114.47	-	-
Lys175	9.19	123.64	179.54	60.67
Gln176	8.54	113.78	176.46	57.75
Tyr177	6.9	116.33	-	55.17
Phe178	7.56	114.29	-	55.97
Val179	-	-	176.31	62.23
Glu180	8.58	125.52	176.28	56.52

Continued on next page

Table A.3 – continued from previous page

Residue	H^N	N	C′	C^α
Lys181	8.51	124.28	175.55	56.58
Gln182	8.09	128.33	-	57.3

B

Modelfree 3CI

Table B.1: 3CI modelfree parameters

Res	Model	S^2	S^2*err*	S_f^2	S_f^2*err*	S_s^2	S_s^2*err*	τ_e	τ_e*err*	R_{ex}	R_{ex}*err*	SSE
1	5	0.272	0.008	0.784	0.008	0.347	0.008	893.66	16.106			0.65
3												0
6	2	0.897	0.011			0.897	0.011	38.964	12.429			0.077
7												0
8	3	0.932	0.042			0.932	0.042			6.294	1.091	0.545
9	3	0.899	0.011			0.899	0.011			2.109	0.273	3.428
10	1	0.923	0.01			0.923	0.01					5.667
13	1	0.925	0.021			0.925	0.021					4.846
14	1	0.945	0.024			0.945	0.024					2.529
15	3	0.84	0.012			0.84	0.012			2.45	0.367	4.736
16	1	0.898	0.02			0.898	0.02					1.676
17	1	0.923	0.029			0.923	0.029					2.452
18	3	0.858	0.014			0.858	0.014			3.404	0.779	0.978
19	1	0.894	0.01			0.894	0.01					1.482
20												0
21	1	0.895	0.026			0.895	0.026					1.498
22	2	0.899	0.011			0.899	0.011	48.947	15.087			3.83
23	2	0.81	0.01			0.81	0.01	16.588	5.477			4.044
25	3	0.835	0.061			0.835	0.061			8.581	2.111	0.051
29	3	0.902	0.016			0.902	0.016			1.139	0.489	2.912
30	1	0.9	0.015			0.9	0.015					1.911
31	1	0.872	0.008			0.872	0.008					2.394
32	2	0.839	0.01			0.839	0.01	30.482	9.826			1.981
33	1	0.861	0.007			0.861	0.007					1.88
34	1	0.9	0.01			0.9	0.01					4.294
35	1	0.937	0.004			0.937	0.004					6.15
36	3	0.884	0.016			0.884	0.016			4.6	0.438	2.446
37	3	0.901	0.015			0.901	0.015			2.061	0.597	2.806

Continued on next page

Table B.1 – continued from previous page

Res	Model	S^2	S^2*err*	S_f^2	S_f^2*err*	S_s^2	S_s^2*err*	τ_e	τ_e*err*	R_{ex}	R_{ex}*err*	SSE
39	2	0.863	0.012			0.863	0.012	48.626	11.106			3.037
40	1	0.922	0.01			0.922	0.01					1.482
41												0
42	2	0.879	0.012			0.879	0.012	44.231	12.041			3.436
44	2	0.813	0.006			0.813	0.006	15.986	6.349			3.221
45												0
46	2	0.868	0.006			0.868	0.006	46.397	5.034			4.449
47	1	0.85	0.007			0.85	0.007					9.956
48	3	0.858	0.011			0.858	0.011			1.861	0.488	1.719
49	3	0.82	0.013			0.82	0.013			1.337	0.409	0.873
50	1	0.911	0.015			0.911	0.015					2.558
51	1	0.868	0.011			0.868	0.011					18.759
52	2	0.877	0.004			0.877	0.004	22.181	6.156			1.947
53												0
54	1	0.895	0.01			0.895	0.01					3.567
55	1	0.873	0.008			0.873	0.008					6.501
56												0
57	1	0.874	0.013			0.874	0.013					1.401
58	2	0.821	0.007			0.821	0.007	32.168	5.779			1.196
59	1	0.85	0.005			0.85	0.005					18.619
60	3	0.873	0.01			0.873	0.01			1.981	0.466	0.242
63	1	0.885	0.008			0.885	0.008					4.747
64	1	0.925	0.01			0.925	0.01					1.618
67	1	0.918	0.003			0.918	0.003					11.709
68	2	0.892	0.014			0.892	0.014	72.837	17.751			4.266
69	1	0.912	0.013			0.912	0.013					0.328
70	1	0.866	0.011			0.866	0.011					3.296
72	1	0.885	0.005			0.885	0.005					3.504
73	1	0.851	0.007			0.851	0.007					8.948

Continued on next page

Table B.1 – continued from previous page

Res	Model	S^2	S^2err	S_f^2	S_f^2err	S_s^2	S_s^2err	τ_e	τ_eerr	R_{ex}	$R_{ex}err$	SSE
74	1	0.868	0.007			0.868	0.007					5.149
75	3	0.876	0.005			0.876	0.005			1.492	0.448	2.133
76	3	0.837	0.005			0.837	0.005			0.857	0.268	5.37
77	1	0.879	0.007			0.879	0.007					3.546
79												0
80	1	0.924	0.014			0.924	0.014					11.389
82	5	0.697	0.027	0.808	0.021	0.862	0.017	775.71	141.2			1.471
83	3	0.837	0.005			0.837	0.005			1.018	0.349	3.267
84	1	0.882	0.01			0.882	0.01					7.278
85	3	0.91	0.008			0.91	0.008			2.696	0.499	0.176
86	1	0.888	0.008			0.888	0.008					0.536
87	3	0.907	0.003			0.907	0.003			0.67	0.21	2.153
88	1	0.938	0.012			0.938	0.012					7.775
89	1	0.901	0.003			0.901	0.003					6.47
91	1	0.941	0.011			0.941	0.011					9.575
92	5	0.725	0.023	0.81	0.009	0.895	0.032	8410.8	4077.3			7.936
93												0
94												0
95												0
96												0
97												0
98	5	0.697	0.006	0.801	0.009	0.869	0.009	1063.7	198.62			1.848
100	3	0.886	0.011			0.886	0.011			5.117	0.671	2.134
101	1	0.896	0.019			0.896	0.019					6.608
113	3	0.841	0.009			0.841	0.009			1.57	0.504	2.725
115	1	0.837	0.007			0.837	0.007					1.27
116	2	0.877	0.01			0.877	0.01	41.879	9.541			3.764
119												0
120	1	0.851	0.011			0.851	0.011					9.814

Continued on next page

Table B.1 – continued from previous page

Res	Model	S^2	S^2err	S_f^2	S_f^2err	S_s^2	S_s^2err	τ_e	$\tau_e err$	R_{ex}	$R_{ex}err$	SSE
121	2	0.874	0.019			0.874	0.019	49.167	17.581	3.856		
122												0
123	2	0.819	0.009			0.819	0.009	20.009	5.999			2.177
125	2	0.813	0.009			0.813	0.009	25.301	5.597			4.356
126	1	0.808	0.006			0.808	0.006					7.122
127	3	0.839	0.021			0.839	0.021			1.706	0.536	1.313
128	1	0.866	0.012			0.866	0.012					4.997
129	2	0.853	0.005			0.853	0.005	22.153	7.738			3.624
131	1	0.859	0.009			0.859	0.009					2.586
132	1	0.903	0.014			0.903	0.014					4.369
133	1	0.862	0.013			0.862	0.013					1.312
135	2	0.837	0.01			0.837	0.01	26.364	11.342			2.35
136	1	0.906	0.01			0.906	0.01					16.597
137	1	0.924	0.008			0.924	0.008					6.85
138	1	0.847	0.008			0.847	0.008					11.938
139	1	0.768	0.009			0.768	0.009					10.971
140	1	0.81	0.007			0.81	0.007					3.942
141	1	0.831	0.008			0.831	0.008					7.096
147	3	0.882	0.119			0.882	0.119			12.226	4.046	0.387
148	1	0.959	0.032			0.959	0.032					0.572
150	1	0.909	0.015			0.909	0.015					3.553
151	1	0.96	0.01			0.96	0.01					3.842
152	3	0.854	0.006			0.854	0.006			0.525	0.148	3.476
153	1	0.829	0.005			0.829	0.005					11.448
155	1	0.909	0.006			0.909	0.006					4.632
157	1	0.895	0.007			0.895	0.007					0.281
158	1	0.914	0.007			0.914	0.007					2.378
159	3	0.889	0.029			0.889	0.029			7.678	1.236	1.031
160	1	0.932	0.023			0.932	0.023					0.885

Continued on next page

Table B.1 – continued from previous page

Res	Model	S^2	S^2*err*	S_f^2	S_f^2*err*	S_s^2	S_s^2*err*	τ_e	τ_e*err*	R_{ex}	R_{ex}*err*	SSE
161	1	0.96	0.014			0.96	0.014					1.74
162	1	0.935	0.006			0.935	0.006					2.429
163	2	0.885	0.008			0.885	0.008	27.964	10.806			4.655
164	1	0.895	0.009			0.895	0.009					2.249
165	1	0.866	0.006			0.866	0.006					7.448
166	2	0.834	0.009			0.834	0.009	29.638	8.822			4.625
168	1	0.863	0.015			0.863	0.015					1.902
170	1	0.857	0.007			0.857	0.007					5.769
171	1	0.899	0.01			0.899	0.01					2.757
172	2	0.864	0.01			0.864	0.01	33.766	10.269			1.253
174	1	0.813	0.006			0.813	0.006					11.571
175	2	0.864	0.012			0.864	0.012	35.223	12.452			3.565
176	2	0.895	0.008			0.895	0.008	24.379	10.791			3.714
177	3	0.9	0.006			0.9	0.006			1.519	0.309	3.803
178	1	0.891	0.006			0.891	0.006					14.775
179	5	0.651	0.003	0.847	0.01	0.769	0.009	858.35	68.774			0.645
180	5	0.374	0.003	0.803	0.006	0.465	0.005	1031.9	22.468			1.634
181	5	0.226	0.013	0.806	0.01	0.281	0.013	821.01	12.012			1.204
182	5	0.163	0.019	0.702	0.015	0.233	0.023	770.04	11.632			0

C

Relaxation Parameters for 3C and 3CI

	3C		3CI (present in both)		3CI (all)	
	Mean	Trimmed Mean	Mean	Trimmed Mean	Mean	Trimmed Mean
			Global			
R_1	0.90 ± 0.10	0.88	0.84 ± 0.10	0.82	0.83 ± 0.08	0.81
R_2^{700}	17.7 ± 4.3	17.7	17.3 ± 3.0	17.6	17.8 ± 2.9	17.9
R_2^{600}	17.3 ± 4.9	17.18	16.5 ± 2.8	16.9	17.0 ± 2.7	17.1
hetero-NOE	0.79 ± 0.98	0.69	0.73 ± 0.15	0.75	0.75 ± 0.12	0.77
R_2/R_1	20.1 ± 0.5	20.3	21.1 ± 4.4	21.6	21.9 ± 0.4	22.1
S^2	0.91 ± 0.07	0.91	0.83 ± 0.15	0.85	0.85 ± 0.12	0.87
τ_c	12.3		–		12.7	
			N-terminal Helix			
R_1	1.05 ± 0.19		0.98 ± 0.20		0.91 ± 0.15	
R_2^{700}	16.2 ± 7.5		15.6 ± 6.2		17.8 ± 5.3	
R_2^{600}	15.8 ± 7.4		14.8 ± 5.8		16.4 ± 4.9	
hetero-NOE	1.23 ± 0.93		0.59 ± 0.28		0.68 ± 0.21	
R_2/R_1	16.7 ± 9.5		17.3 ± 8.9		20.6 ± 7.4	
S^2	0.94 ± 0.10		–		0.80 ± 0.26	
			β-barrel 1			
R_1	0.86 ± 0.1	0.86	0.80 ± 0.03	0.80	0.80 ± 0.03	0.81
R_2^{700}	17.4 ± 1.6	17.3	17.9 ± 1.1	17.9	18.3 ± 1.6	18.2
R_2^{600}	16.9 ± 1.5	16.9	17.1 ± 1.2	17.2	17.5 ± 1.6	17.4
hetero-NOE	0.67 ± 0.05	0.67	0.77 ± 0.04	0.77	0.78 ± 0.05	0.78
R_2/R_1	20.3 ± 1.7	20.3	22.3 ± 1.6	22.3	22.7 ± 2.3	22.56
S^2	0.90 ± 0.05	0.90	0.87 ± 0.03	0.87	0.88 ± 0.03	0.88
			Linker			
R_1	0.89 ± 0.05		0.83 ± 0.05		0.83 ± 0.04	
R_2^{700}	18.1 ± 4.0		17.7 ± 1.9		17.6 ± 1.9	
R_2^{600}	17.4 ± 2.7		16.9 ± 1.5		16.6 ± 1.7	
hetero-NOE	0.70 ± 0.07		0.76 ± 0.07		0.75 ± 0.07	
R_2/R_1	20.4 ± 5.1		21.5 ± 2.8		21.3 ± 2.8	
S^2	0.90 ± 0.09		0.86 ± 0.08		0.85 ± 0.08	
			β-barrel 2			
R_1	0.88 ± 0.07		0.82 ± 0.03		0.80 ± 0.04	0.80
R_2^{700}	19.6 ± 4.7		18.1 ± 2.0		18.2 ± 2.4	17.9
R_2^{600}	19.6 ± 7.1		17.2 ± 1.8		17.3 ± 2.0	17.1
hetero-NOE	0.69 ± 0.14		0.78 ± 0.05		0.78 ± 0.05	0.78
R_2/R_1	22.4 ± 4.9		22.2 ± 2.7		22.6 ± 3.0	22.4
S^2	0.93 ± 0.06		0.88 ± 0.06		0.87 ± 0.05	0.87
			C-terminus			
R_1	1.00 ± 0.20		0.96 ± 0.21		0.96 ± 0.21	
R_2^{700}	13.5 ± 6.4		13.8 ± 5.9		13.8 ± 5.9	
R_2^{600}	13.1 ± 6.2		13.3 ± 5.6		13.3 ± 5.6	
hetero-NOE	1.39 ± 2.94		0.52 ± 0.32		0.52 ± 0.32	
R_2/R_1	14.9 ± 8.5		15.9 ± 8.6		15.9 ± 8.6	
S^2	0.89 ± 0.04		0.87 ± 0.04		0.64 ± 0.30	

D

R_{ex} Protocol

The pulse sequence used was nnhc13cryo.kd.pe Four spectra are generated (I, II, II, IV), the phasing requirements are detailed in the pulse sequence. The peak heights give the η_{xy} via the following relationship:

$$\eta_{xy} = \frac{arctanh\sqrt{\frac{II}{I}\frac{III}{IV}}}{t} \tag{D.1}$$

$$\kappa = R_2^0/\eta_{xy} \tag{D.2}$$

for all residues without exchange. Calculate this by taking average for all model 1 residues and then trimming outliers ($\pm$ 1 s.d.) Once η_{xy} has been calculated this is subtracted from the CPMG-derived R_2 to give the exchange contribution to transverse relaxation.

E

HADDOCK

E.1 Introduction to HADDOCK

HADDOCK is a molecular docking program that uses ARIA scripts and CNS. HADDOCK stands for High Ambiguity Driven protein–protein Docking. HADDOCK can make use of biochemical and biophysical interaction data to drive the docking process. The set-up of a HADDOCK run using the structures and data described in Chapter 4 is presented.

E.2 Set-up

HADDOCK requires several input files. PDB files of both interaction partners are required. These can be NMR or crystallography-derived structures and can contain protons although this is not necessary. In this case a PDB file from the the RDC refined ensemble of 3CI

```
wi_jmc_accept_1.pdb
```

was used along with a PDB file from the the ensemble of SLD from HRV-14

```
HRV-SLD_time_ave_rmh2o.pdb
```

This structure is RDC and time averaged NOE refined. In addition to PDBs of the docking partners, HADDOCK requires a set of restraint files. In this case, chemical shift perturbation (CSP) and mutation data were combined to create the restraint lists. The AIRs for the protein partner 3CI were defined as residues with greater than 50 % surface accessibility (calculated using AreaIMol from ccp4. The HADDOCK manual recommends NACCESS but this does not run on recent Linux flavours) that exhibited CSP greater than 2 s.d. above average or were shown to be important for RNA-binding by previous mutation studies, or both. This gave a list of "active" residues. "Passive" residues were then defined as residues that were surface neighbours of active residues and also had greater than 50 % surface accessibility. This defines the interaction surface of the protein. The RNA interaction surface was defined in a similar fashion with active residues defined as those exhibiting a significant loss of intensity (us) or significant CSP (ohl) in RNA titration experiments. All RNA bases were assumed to be solvent accessible.

3CI active: 5, 6,9, 10, 13, 32, 33, 80, 81, 82, 83, 84, 85, 89, 153, 154, 155

3CI passive: 4, 7, 8, 11, 12, 14, 31, 34, 79, 86, 88, 90, 152, 156

RNA active: 5,13,14, 15,16,17,18,21

RNA passive: 4,6,12,19,20,22

E.3 Run

The run is controlled by a file called *run.cns*. An example of a functional *run.cns* is at

```
/home/jkclarid/haddock/haddock/SAXS_dock/run3/run.cns
```

the paths for all the files should be checked and a directory called logfiles should be created in the run*x* directory. The following shows the relevant part of the *run.cns* file:

```
{* Logfile directory *}
{* specify a directory for the large CNS log files *}
{===>} temptrash_dir="/home/jkclarid/haddock/SAXS_dock/run3/logfiles";
```

In this run no segments were assigned as either fully or partly flexible. If dynamics data are available (as they are now for 3C but were not used for this run) or if a region is known to be flexible then this can be allowed to move throughout the docking.

for the use of RNA the DNA-RNA restraints file needs to be enabled:

```
{======================DNA-RNA restraints ============================}
{* Use DNA/RNA restraints (dna-rna_restraints.def in data/sequence)? *}
{+ choice: true false +}
{===>} dnarest_on=true;
```

the section:

```
{============================ parallel jobs ================================}
{* How many nodes do you want to use in parallel? *}
{* leave unused fields blank, make sure that the queues are actually running *}
{+ table: rows=10 "1" "2" "3" "4" "5" "6" "7" "8" "9" "10"
 cols=3 "queue command" "cns executable" "number of jobs" +}

{===>} queue_1="csh";
{===>} cns_exe_1="/usr/bin/cns";
{===>} cpunumber_1=4;

{===>} queue_2="";
```

```
{===>} cns_exe_2="";
{===>} cpunumber_2=0;
```

defines the CNS executable and processors used. If both computers are utilised then

```
queue_2
```

should read:

```
{===>} queue_2="ssh minos ~/bin/haddock_wrapper";
{===>} cns_exe_2="/usr/bin/cns";
{===>} cpunumber_2=4;
```

this uses a script to pass the command to the other computer.

Most of this set up is bypassed by use of the HADDOCK webserver but it is slightly harder to fine-tune the process. A particular advantage of the webserver is that it generates a DNA-RNA restraint file automatically if one of the structures present is an oligonucleotide.

E.4 Analysis

The HADDOCK analysis scripts assume that the minimum energy structure is the 'best'. Thus it is important to visually check the results as well as use the analysis tools provided. Various analyses can be performed by HADDOCK on the final docked structures. The HADDOCK score is the weighted sum of various energy terms. the command:

```
$HADDOCKTOOLS/ana_structure.csh
```

should be run in order to generate individual structure analyses. The structures are clustered by pairwise RMSD with a 7.5 Å cut-off. The clusters are then scored. A plot of the RMSD from the lowest energy structure versus HADDOCK score is shown in Figure E.1. The lowest energy cluster (cluster 4) from the run is shown in Figure E.2.

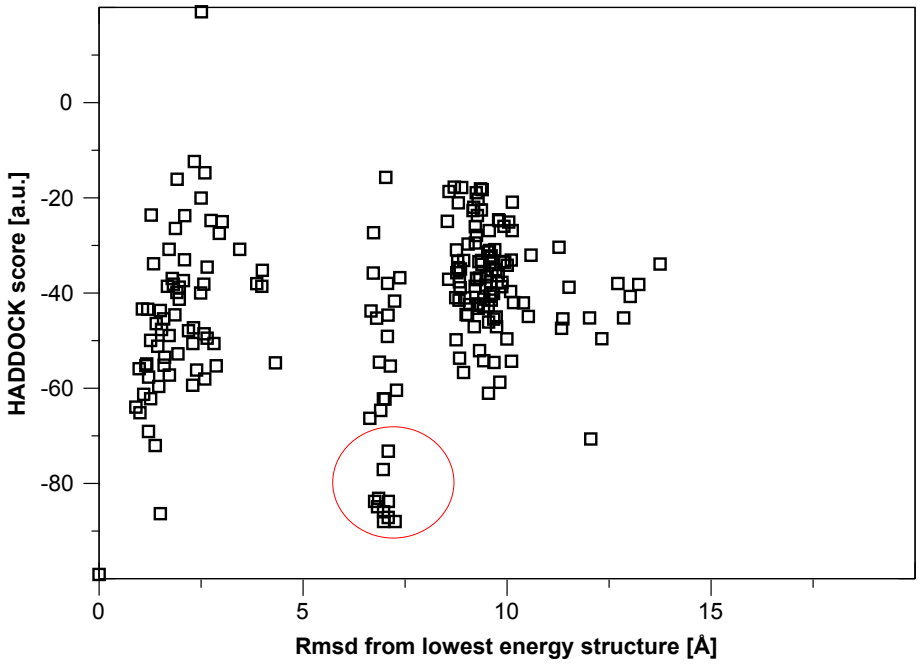


Figure E.1: RMSD from lowest energy structure versus HADDOCK score. The red circle delineates the lowest energy cluster.

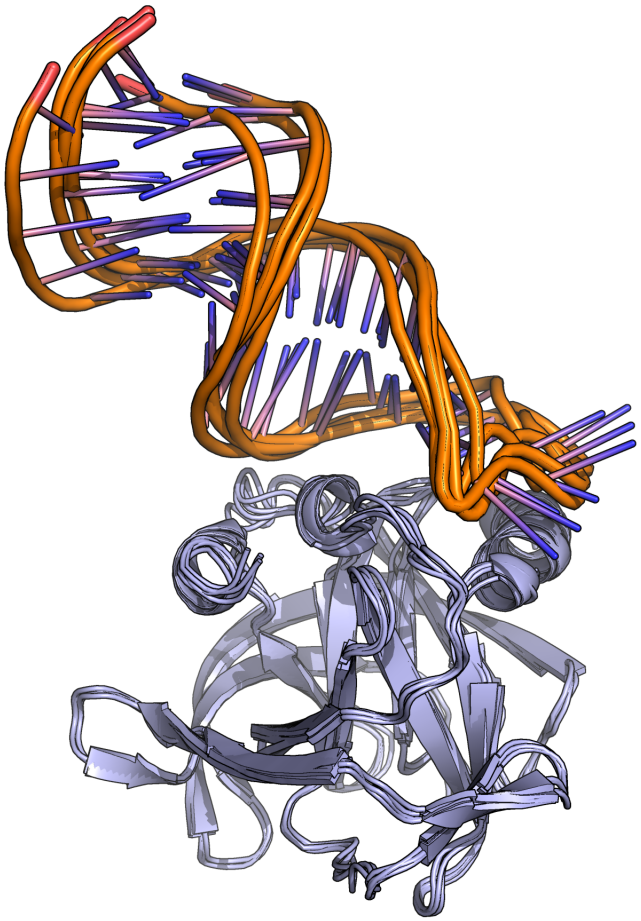


Figure E.2: Overlay of four lowest energy structures from cluster 4

F

3C Purification Protocol

F.1 Two (long!) Day 3C purification protocol

Day 1

Resuspend pellets in 30 ml lysis buffer (20 mM Tris-HCl pH 7.5, 25 mM imidazole) containing one aliquot of protease inhibitor cocktail (Calbiochem). Keep on ice.

Lyse cells with French press (3 $\times$ 5-6,000 psi)

Spin down at 17, 590 $\times$ g for 30 minutes at 4°C (Sorvall, SS34).

Decant supernatant and filter (0.8 μm).

Pass through 3 ml Ni_{2+}-NTA column (pre-equilibrated with lysis buffer)

Wash column with 10 ml of lysis buffer

Wash column with 10 ml of lysis buffer (plus 50mM imidazole)

Elute column with 2 $\times$ 5 ml 250 mM imidazole pH 7.5, 1M NaCl

Transfer to dialysis tubing and add 30 units of thrombin (Roche).

Dialyse into 1 L of lysis buffer containing 0.5 mM DTT (4 hrs)

Change buffer to lysis buffer with no DTT (4 hrs)

Pass through 3 ml Ni_{2+}-NTA column (pre-equilibrated with lysis buffer)

Rinse column with 2 $\times$ 3 ml lysis buffer

Dialyse into 1 L 20 mM Tris-HCl pH 7.8, 0.5 mM EDTA, 1 mM DTT (overnight)

Day 2

Change buffer, dialyse (4 hrs)

Pass through 3 ml Q-sepharose column (Sigma) (pre-equilibrated with 20 mM Tris-HCl pH 7.8, 0.5 mM EDTA, 1 mM DTT

Spin concentrate and buffer exchange into 50 mM BTP/MES pH 6.0, 0.5mM EDTA

Add DTT to 14 mM

Concentrate to desired molarity and add inhibitor (JMC-98-3)

Samples Produced

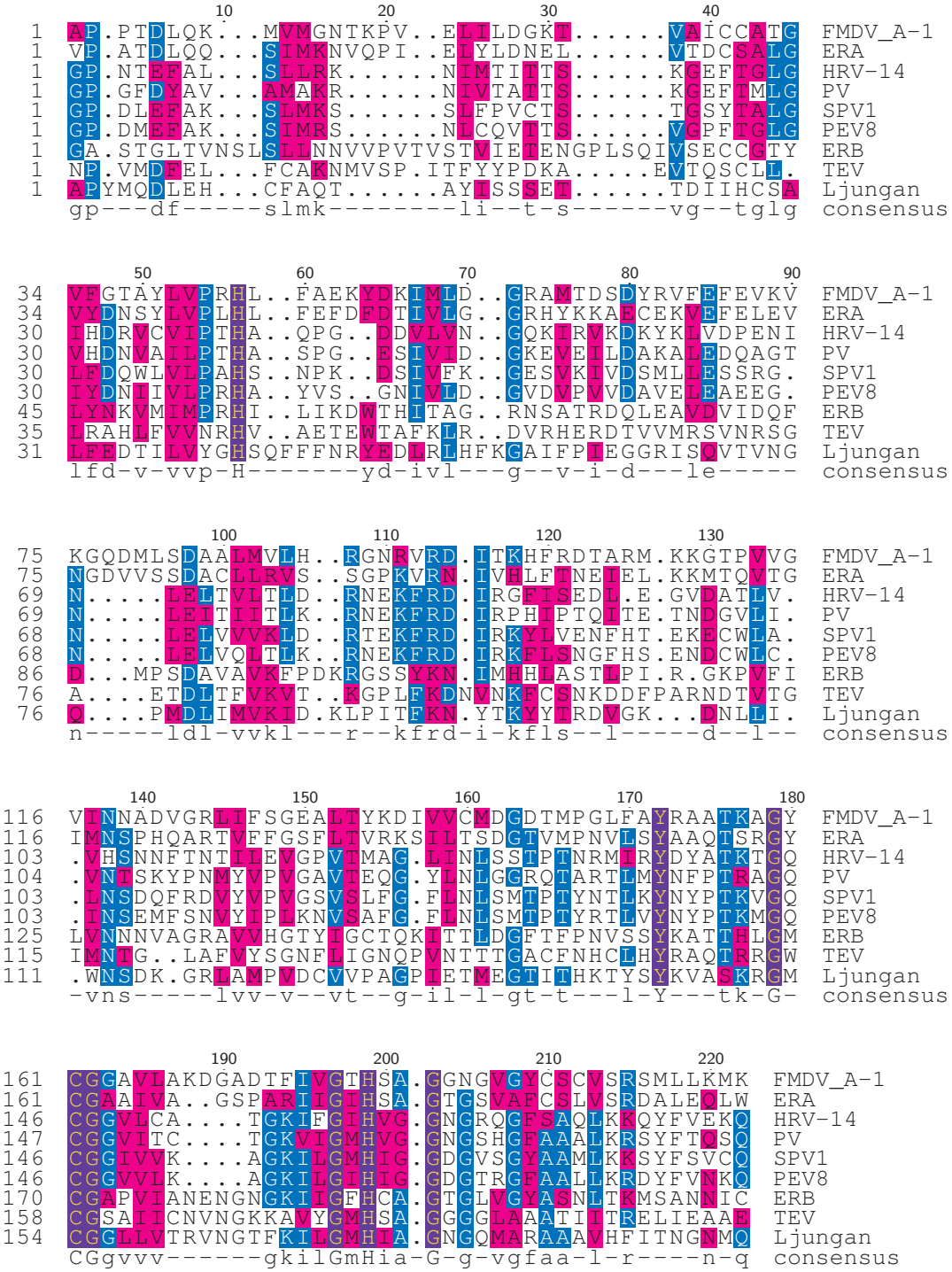
Table G.1: Sample preparations with experiments carried out. Samples marked with an asterisk contain specific $3C^{pro}$ inhibitor.

Sample	Prep. date	Buffer	Experiments run	Remarks
N2-3C (1.5 mM)	041104	20 mM phosphate pH 6.5	hsqcetf3gpsi, noesyhsqcf3gpsi3d	initial ^{15}N labelled growth full peak count
N3-3C (0.85 mM)	041117	20 mM phosphate pH 6.5	hsqcetf3gpsi, trosyetf3gpsi	
CN1-3C (1.3 mM)	050307	20 mM phosphate pH 6.5	hsqcetgpsp, hncogp3d, hncagp3d, hncacbgp3d	^{13}C^{15}N-labelled assignment of 3C
N4-3C (0.1 mM)	050610	20 mM acetate pH 5.5	trosyetf3gpsi	
N5-3C (0.8 mM)*	050623	20 mM Tris-base pH 7.5	trosyetf3gpsi	addition of JMC-98-3C resulted in recovery of peaks
N7-3C (0.3 mM)*	050811	50 mM bis-Tris-MES pH 6.0	preliminary dynamics	
CN5-3C (1.1 mM)*	051016	50 mM bis-Tris-MES pH 6.0	hsqcetfpf3gpsi, hncogpwg3d, hncacogpwg3d, cbcanhgpwg3d, cbcaconhgpwg3d, hncagp3d, noesyhsqcgp3d, SLD RNA added (1:1.2) hsqcetf3gpsi	^{13}C^{15}N-labelled assignment of 3CI
ND1-3C (0.2 mM)*	060412	50 mM bis-Tris-MES pH 6.0	stebpgp1s191d, trosyetf3gpphsi19	unlabelled RNA added to this sample
CND1-3C (0.3 mM)*	060525	50 mM bis-Tris-MES pH 6.0		perdeuterated ^{13}C labelled
N10-3C (0.6 mM)*	060827	50 mM bis-Tris-MES pH 6.0	hsqcetf3gpsi2	different inhibitors tried
REDPRO1-3C*	070322	50 mM bis-Tris-MES pH 6.0	SLD RNA added (1:2) trosyetf3gpsi2 trhncogp2h3d,trhncagp2h3d	Ca deuterated
N16-3C (1 mM)*	071025	50 mM bis-Tris-MES pH 6.0	dynamics experiments	3CI dynamics data
N17-3C (0.3 mM)	071219	50 mM bis-Tris-MES pH 6.0	dynamics experiments	3C dynamics data
N19-3C (0.3 mM)*	090325	20 mM phosphate pH 6.5	CD ,DLS, hsqcetf3gpsi	biophysical data collection two samples: 3C and 3CI

H

3C Alignment

APPENDIX H. 3C ALIGNMENT



X non conserved
X similar
X ≥ 50% conserved
X all match