UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE HEALTH AND LIFE SCIECES

School of Biological Sciences

Investigation into the mechanism of immature HIV-1 capsid assembly

By

Michael John Knight

Thesis for degree of doctor of philosophy

February 2010

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
SCHOOL OF BIOLOGICAL SCIENCES

Doctor of Philosophy

INVESTIGATION INTO THE MECHANSIM OF IMMATURE HIV-1 CAPSID ASSEMBLY

By Michael John Knight

The major structural protein of the retrovirus HIV-1 is called Gag and is expressed as a 55 kDa poly-protein with six contiguous domains. These are labelled from the N-terminus as MA, CA, SP1, NC, SP2 and P6. There are two distinct assembly steps in the lifecycle of HIV-1, termed immature and mature assembly, both of which are essential to the production of infectious viral progeny and are as such potential targets of therapeutic intervention. The immature assembly step involves self-association of, typically, 1000-2500 copies of Gag in a nucleic acid-dependent manner, resulting in formation of a spherical capsid immediately below the host cell membrane. The resulting immature, non-infectious virions are released from the cell and the viral protease, PR, hydrolyses Gag into its component domains. MA remains at the membrane and NC remains in complex with the genome, whilst CA reassembles as a mature capsid with a conical shape and 5,7-Fullerene geometry. In the immature and mature state, CA forms a lattice in which N-CA is arranged as hexamers linked to one another by C-CA dimerisation, but the exact interfaces and CA conformations are different between the two states. In this thesis, experiments are described which seek to establish how a single protein, CA, can form two distinct lattices, and what the role of NC-nucleic acid interactions are in immature assembly. Several Gag mutants are studied using a combination of NMR spectroscopy, fluorescence spectroscopy, electron microscopy and *in vitro* virus capsid assembly assays. It is shown that the NC domain does not intrinsically effect any modulation of the C-CA domain at the level of the first intermediates in the assembly pathways, and that nucleic acid is required to link two Gag molecules together in order to promote immature assembly.

Contents

List of figures

List of tables

DECLARATION OF AUTHORSHIP

I, Michael John Knight

declare that the thesis entitled 'Investigation into the mechanism of immature HIV-1 capsid assembly' and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed:

Date:

Acknowledgements

Firstly, I would like to acknowledge the help of my supervisor, Joern Werner.

I would also like to acknowledge Stuart Findlow and Phil Williamson for helping me with countless experiments and for many useful discussions.

For assistance in electron microscopy, I would like to thank Anton Page.

For assistance with fluorescence measurements, and for critically reading several pieces of draft work, I would like to thank Tony Lee.

For assisting with several experiments, I would like to thank the NMR staff at Mill Hill, particularly Geoff Kelly, Alain Oregioni and Tom Frenkiel.

For help with solid-state NMR, I would like to thank Malcolm Levitt.

For providing the plasmids used in this work, I would like to thank Hans-Georg krausslich and Vanda Bartanova at the university of Heidelberg.

Many people have been helpful out of the lab, and I'd particularly like to acknowledge those people unlucky enough to share an office with me, namely Helen Watson, Nik Rogers, Juan Bolivar-Gonzalez, Garrick Taylor and Jon Butler.

Of course, I'd like to acknowledge my parents, Carol and John Knight, my brothers Justin and Daniel Knight, and my various friends who helped keep me sane whilst preparing to submit this thesis.

List of abbreviations

A260	Absorbance and 260 nm
A280	Absorbance and 280 nm
AIDS	Acquired immunodeficiency syndrome
CAI	Capsid assembly inhibitor
CA	Capsid protein of Gag
CP	Cross-polarisation
cEM	Cryo-electron microscopy
cET	Cryo-electron tomography
C-CA	C-terminal domain of CA
DNA	Deoxyribonucleic acid
DTT	Di-thio threitol
dSL3	DNA version of SL3
FID	Free inductive decay
HIV-1	Human immunodeficiency virus type 1
MA	Matrix protein of Gag
N-CA	N-terminal domain of CA
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NC	Nucleocapsid protein of Gag
PDB	Protein database
PDSD	Proton-driven spin diffusion
RDC	Residual dipolar coupling
RNA	Ribonucleic acis
RNP	Ribo-nucleoprotein
SL2	RNA stem-loop 2 of HIV-1 psi-site
SL3	RNA stem-loop 3 of HIV-1 psi-site
RSV	Rous sarcoma virus
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
ssNMR	Solid-state NMR
SP1	Spacer peptide 1 of Gag
SP2	Spacer peptide 2 of Gag
tRNA	Transfer RNA
TCEP	tris(2-carboxyethyl)phosphine
UV	Ultraviolet
VLP	Virus-like particle

Chapter 1 Introduction

1.1. Assembly and maturation in retroviruses

Retroviruses are Group VI (ssRNA –RT) viruses of the family *retroviridae*. There are two subfamilies of retroviruses, namely *orthoretrovirinae* and *spumaretrovirinae*. Human immunodeficiency virus type 1 (HIV-1) is of the *orthoretrovirinae* subfamily of *retroviridae*, and of the *lentivirus* genus. It is to HIV-1 that all experiments in this thesis pertain, though features common to retroviruses will be introduced prior to specific consideration of HIV-1. At the most basic level, retroviruses comprise two compartments, the outermost housing the capsid which in turn envelopes the single-stranded RNA genome. The outer compartment is comprised of a host cell-derived plasma membrane wherein are situated a number of proteins involved in cell entry and exit, and is closely associated with the matrix, comprised of the viral MA protein. The inner compartment, the capsid or core, exists as a protective layer for the genetic material. It dissociates upon cell entry, releasing the single-stranded RNA genome into the host cell cytoplasm where it is reverse transcribed, the resulting double-stranded DNA being incorporated into the host genome, serving as a template for synthesis of the all the necessary proteins for the assembly of viral progeny as shown in Figure 1.1.

The major structural protein of retroviruses is the Gag polyprotein, which assembles as a spherical capsid before targeting to the host cell membrane prior to particle budding and exit from the host cell in the form of immature virions (26, 157). The precise cellular location of immature assembly is not well defined. It has been observed for the related retrovirus Rous sarcoma virus (RSV) that immature assembly begins in the host cell cytoplasm, with large complexes being formed before targeting to the plasma membrane (87), the targeting process itself taking advantage of the host cell cytoskeletal remodelling systems (55).

Icosahedral symmetry has been suggested for the immature capsid which, whilst well established for certain other virus classifications, remains a contested issue in the case of retroviruses. It is for this reason that the classical Kaspar-Klug system of classification does not apply to retroviruses.

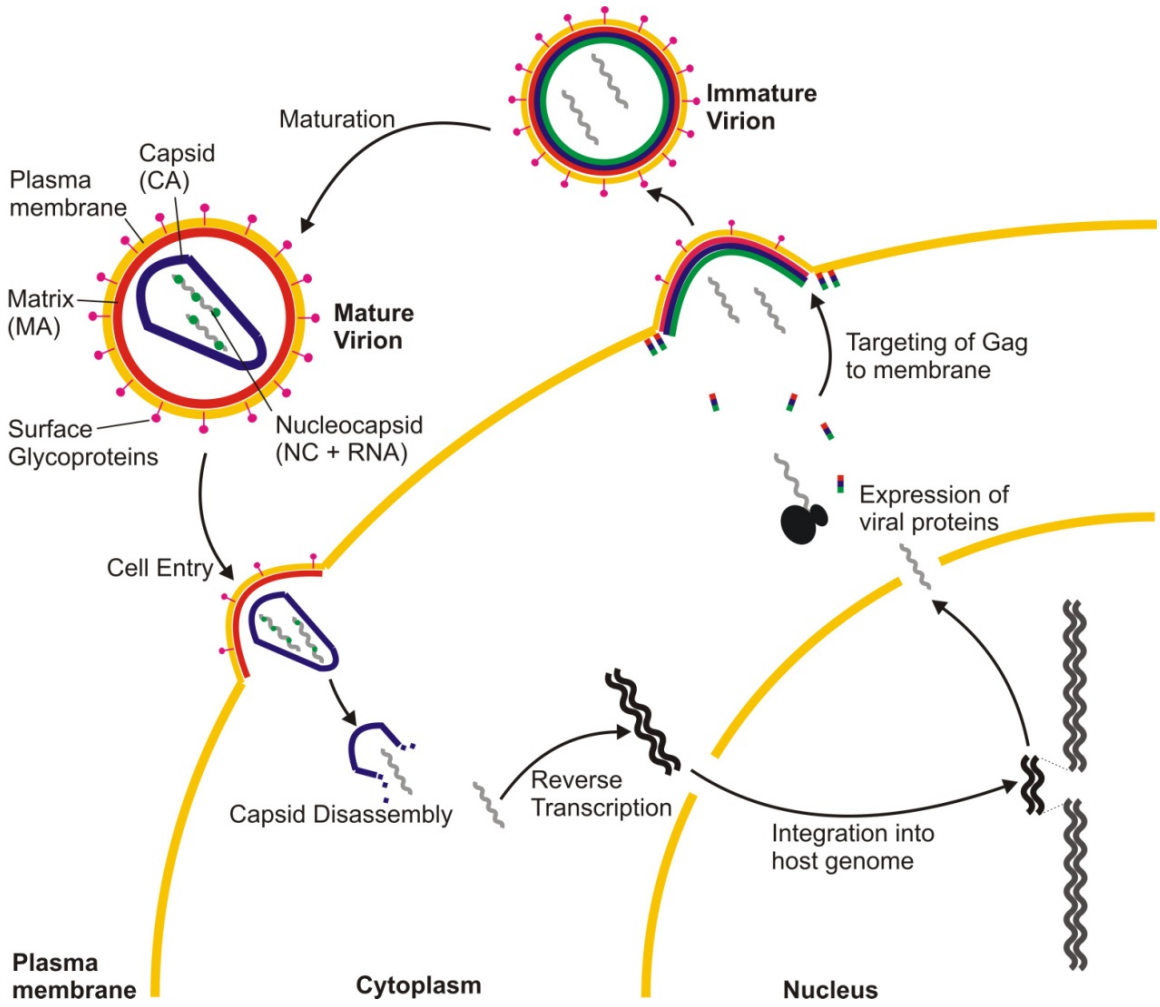


Figure 1.1: The lifecycle of HIV-1 as a prototypical retrovirus. Only the steps relevant to assembly are shown. RNA is shown in grey, DNA in black.

1.2. Maturation

A defining characteristic of retroviruses is the maturation process whereby an ordered series of cleavage reactions catalysed by the viral protease (PR) separates Gag into its component domains (156) after release of immature progeny. The various domains of Gag are indicated in Figure 1.2. The first PR-catalysed hydrolysis removes NC, P1 and P6 from the C-terminal of the SP1 spacer. At a 10-fold lower rate CA is cleaved from MA, and a final step, at a rate around 400-fold lower than the first step, cleaves SP1 from CA, leaving the latter to reassemble as the mature capsid. It is thought that only around one third of the CA molecules present contribute to the mature capsid (26). The fate of the remainder is unclear.

Maturation is necessary for infectivity (152), and interference therewith presents a potential point of therapeutic intervention in the treatment of retroviral diseases such as

HIV. Following maturation, the MA domain of Gag remains associated with the membrane, whilst CA molecules reassemble as a morphologically distinct capsid with geometry defined by the relevant class of retrovirus. The third domain of Gag, NC, remains within the capsid in complex with the retroviral RNA genome (57, 149). A variety of core morphologies are seen across retroviruses; for example that of the gammaretrovirus Moloney-murine leukaemia virus (M-MuLV) is spherical (160), the betaretrovirus Mason-Pfizer monkey virus (MPMV) is spherocylindrical (27), whilst capsids of lentiviruses such as Rous Sarcoma virus (RSV) and human immunodeficiency virus type 1 (HIV-1) are conical.

1.3. Difficulties in studying the structural biology of retroviruses

In all classes of retrovirus there is considerable pleomorphicity amongst capsids, which applies to both mature and immature assemblies. This confounds the study of retroviral capsids by established techniques such as X-ray crystallography and cryo-electron microscopy (cEM) image reconstruction. The former technique relies on the recording of a pattern of discrete diffraction spots amplified by symmetry elements present within an ordered structure. The major translational, rotational and mirror-plane elements are generally absent from the randomly aligned pleomorphic structures associated with retroviral capsids; hence, no meaningful diffraction pattern is observable. The latter technique is reliant on averaging of images at different orientations of copies of some fundamental unit. With that unit being capsids, each ‘copy’ has a different structure, such that an average image is no longer meaningful (154). In the case of immature virions, spherical symmetry is not present due to variation in the number of molecules comprising each virion, such that a distribution of immature structures is present in any sample.

Progress has nevertheless been made in understanding capsid structure using transmission electron microscopy (TEM), and in three dimensions using cryo-electron microscopy (cEM), cryo-electron tomography (cET) (15, 23), electron crystallography and computer modelling. The molecular basis for capsid formation remains poorly understood; the mechanism whereby some 1500 copies of a single protein associate to form a closed structure of variable dimensions and shape requires an understanding of the manner in which Gag molecules interact with one another. Thus, in the following sections, aspects of biochemical and structural studies on retroviral assembly are reviewed, with particular emphasis on HIV-1.

1.4. The Gag polyprotein

Gag comprises three major contiguous segments; MA at the N-terminus, which forms the virion matrix beneath the plasma membrane; CA, which forms the capsid and NC, which condenses with RNA to form the ribonucleoprotein core (RNP) (57, 149). Gag is thought to be highly flexible, which precludes the determination of a unique structure for the entire molecule. Nonetheless, the individual domain structures, along with a number of dual-domain constructs, have been determined by X-ray crystallography, NMR or in some cases by both techniques and are shown in Figure 1.2. Gag flexibility has been demonstrated using small angle neutron scattering and computational modelling (38).

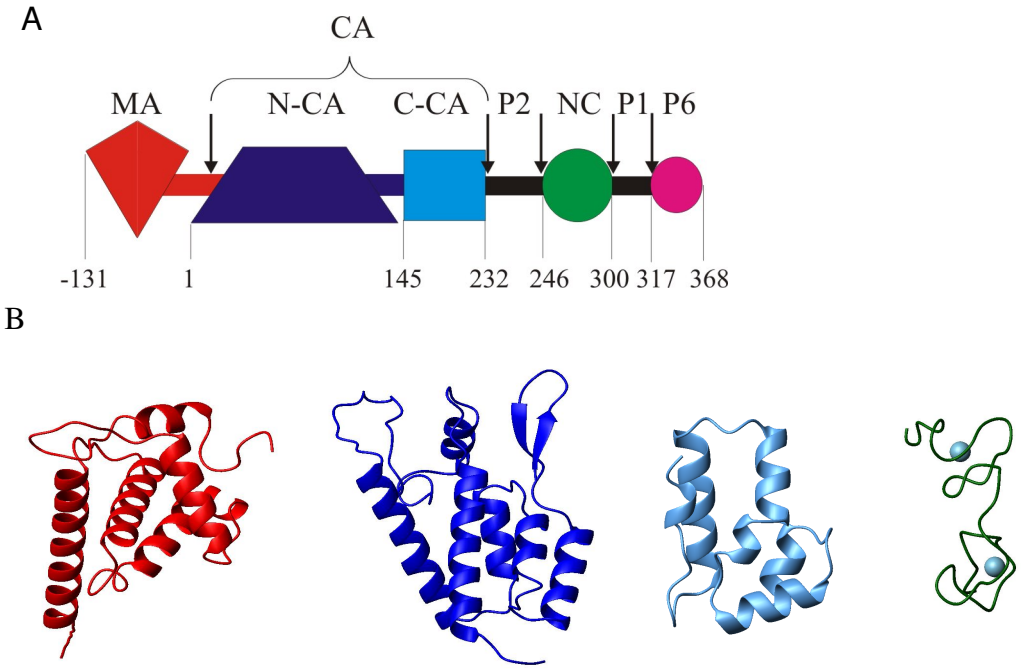


Figure 1.2: A; The domain structures of HIV-1 Gag.

The N and C-terminal residues of each domain are indicated and numbered with respect to the N-terminus of CA . B; Ribbon diagrams of the structures of isolated Gag domains. MA (red; pdb code: 2GOL), N-CA (dark blue; pdb code: 2GON), C-CA (light blue; pdb code: 1A43) and NC (green; pdb code:1ESK) Zinc ions are shown as spheres. MA, N-CA and C-CA as shown were determined by X-ray crystallography, and NC by NMR. Figure was created using MolMol (82).

1.4.1. The MA domain

Whilst all retroviruses undergo analogous assembly and maturation processes, discussion is henceforth restricted to the proteins of HIV-1. The role of the MA domain is largely to target Gag to the plasma membrane and maintain the gross structure of the resultant virions. It has a myristoylation site which facilitates its binding to the plasma membrane (101, 136). This is required for targeting newly synthesised Gag to the host cell plasma membrane prior to assembly and for maintaining close contact between the matrix and viral membrane after the new viruses exit their host cell (65, 126). The structure of MA has been solved by both NMR and X-ray crystallography (79, 102). Its structure, as seen in Figure 1.2, is a globular six-helix bundle with a flexible disordered C-terminal tail which serves as a linker between itself and N-CA (79).

1.4.2. The CA domain

CA comprises two domains, with a short linker separating them. In CA numbering (which will be used throughout), the N-terminal domain (N-CA) spans residues 1-145 and the C-terminal domain (C-CA) spans residues 146-231. NMR data are supportive of the two domains being flexible with respect to one another in solution (Ludovic Pecqueur, unpublished data).There is considerable disorder at the C-terminus of C-CA; the last 12 residues are not seen in electron density maps (19, 46), and ^{15}N-NMR relaxation data are also supportive of a flexible tail (Jana Sticht, personal communication). The structures of N-CA and C-CA have been solved in isolation (46, 53, 158), and the entire CA structure solved in the presence of a Fab fragment (105), however in this structure most of the C-CA domain was not visible. N-CA is a seven-helix bundle adopting an arrowhead-like shape (53, 79, 105); helices 1, 2, 3 and 7 form a globular bundle with helix 4 joining at an angle to impart the tapered shape on the molecule. Helices 5 and 6 are shorter, and pack roughly perpendicular to the rest of the molecule, at the wide end of the arrowhead. In the ‘free’ form of N-CA (without MA attached at its N-terminus) a short β-hairpin is formed at the N-terminus. This is believed to be stabilised by a salt bridge between the protonated N-terminal amine group and D51 (53), and may be a morphological switch between the immature and mature arrangements of assembled CA molecules since a D51A Gag mutant does not form conical (mature) capsids (152).

The C-terminal CA domain, C-CA, is a smaller, four-helix bundle whose structure has been determined by X-ray crystallography (46, 158). Helices 1 and 2 (H1 and H2) pack antiparallel to one another at an angle of ~30° with a short loop connecting them. The shorter 3rd and 4th helices (H3 and H4) are packed next to one another at the opposite end of the molecules to the H1-H2 loop (Figures 1.2 and 7). At the N-terminus, residues 154-174 of C-CA form the major homology domain (MHR), a stretch of 20 residues highly conserved amongst retroviruses, which adopts a strand-turn-helix motif and packs against the C-terminal of helix 2. Its role is not understood, but mutational analysis reveals it to be important for retroviral assembly and infectivity (144, 152).

1.4.3. The NC domain

NC is a short protein domain of around 55 residues, whose structure has been determined by NMR spectroscopy (109, 147). It comprises two 'zinc knuckle' domains, which are similar in structure and function to zinc fingers; *in vivo* each binds RNA using loops constrained by a tetrahedrally coordinated zinc ion in a Cys_2-His-Cys motif (20), shown in Figure 1.3, though a wide variety of both RNA and DNA sequences can be recognised. The affinity of NC for Zn^{2+} has been reported as 3.5 x 10^{-15} M (104), such that Zn^{2+} is extremely tightly bound. The CCHC motifs are short, and the termini of NC are highly flexible, as is the section of polypeptide, ^{274}RAPRKKG, between the more ordered zinc knuckle motifs. NMR data suggest that, when not bound to RNA/DNA, the two zinc knuckles act as small, independent domains and NC may become a more rigid molecule upon RNA/DNA binding (90).

NC has numerous roles besides contributing to immature assembly. For example, it is required to package the viral RNA genome, fulfilling its role in assembly by doing so, (5, 35, 122, 138) and, after maturation, appears to direct the dimerisation and stabilisation of a stable structure in the 5' region of the retroviral genome (44). NC is also involved in reverse transcription (9, 75, 80, 81, 125, 132), mediating strand transfer, unwinding tRNA and fulfilling chaperone-like roles for RNA and DNA secondary structure, as reviewed in (130). Its role in reverse transcription requires NC-DNA interactions in addition to NC-RNA interactions.

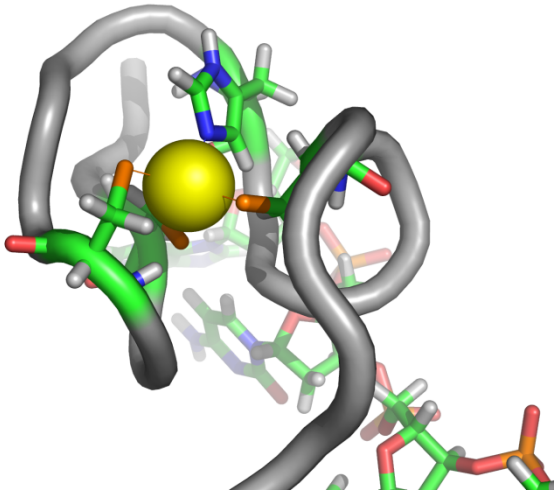


Figure 1.3:Coordination of Zn2+ and DNA binding by the N-terminal domain of NC from HIV-1.

1.5. Gag arrangement in immature virions

Immature virions, like mature virions, appear to be comprised of a hexameric CA lattice; however the MA, SP1, NC, P1 and P6 domains are also present, and the CA arrangement is different between immature and mature virions. The current knowledge of Gag stoichiometry and arrangement in immature virions is discussed in this section.

1.5.1. The number of Gag molecules in an immature virion

The number of Gag molecules contributing to an HIV-1 virion is highly variable, and several different estimates of the number of Gag molecules have been made. The largest estimate is that around 5000 copies of Gag form the 'average' HIV-1 immature virion (26), this figure being derived from scanning transmission EM (STEM) of *in vitro*-assembled Gag particles. A lower estimate of 2500 Gag molecules was made based on the ratio of the CA concentration to genomic RNA concentration in virus preparations (120). Three separate studies provided estimates of around 1200-1500 (89, 151, 166), whilst a recent study employed fluorescence fluctuation spectroscopy to provide an estimate of 750-2500 for VLPs produced in COS-1 cells (31).

In the immature virion, Gag molecules arrange at the plasma membrane by host-mediated myristoylation of MA (65, 101, 136). The C-terminus of Gag then points into the centre of the virion. Cryo-electron microscopy and cryo-electron tomography have been

instrumental in understanding the Gag arrangement. Early electron microscopic work on HIV-1 was conducted under the assumption of icosahedral symmetry, as expected from Caspar-Klug theory, and yielded a model mature capsid with triangulation number *T*=7 (49-51, 100, 112, 113). It was considered by some to be logical to extend the assumption of icosahedral symmetry to the immature capsid (113). However, it became apparent that HIV-1, like most retroviruses, is heterogeneous in size, and lacks global symmetry. The major papers contributing to the emergence of such views are reviewed in the following section.

1.5.2. Gag lattice structure from electron microscopy

Various models of immature and mature virions have been studied using electron microscopy, which include authentic virions and PR^- virions (which do not undergo maturation) isolated from infected T-cells, *in vitro* assembled capsids from purified Gag mutants and virus-like particles (VLPs). By VLP is meant any capsid obtained from and *in vivo* system which is not a complete virion. Often, VLPs are produced by expressing Gag in insect cells by means of a baculovirus vector, resulting in the release of spherical particles from the transfected cells enveloped by host cell-derived plasma membrane (52, 115).

A hexameric (p6 symmetric) arrangement of CA with local order, but not global symmetry, was first observed in VLPs prepared from baculovirus-infected insect cells (45). In a later study by the same authors, again using VLPs (157), it was shown that MA, CA and NC form distinct layers, implied by radial density profiles. This study also demonstrated that almost the same radial density profile is seen for VLPs in which the p6 domain is absent, demonstrating that p6 does not contribute to the ordered part of the immature Gag lattice, and that VLPs from insect cells expressing full-length Gag have a very similar radial density profile to PR^- HIV-1 virions from infected T-cells. In all radial density profiles, the MA, N-CA, C-CA and NC domains were clearly separated. The authors report a mean virion size of 133 ± 17 nm for PR^- virions, whilst VLPs (which lack the envelope proteins) appear slightly smaller at 112 ± 7.3 nm.

A separate study, again using cEM, also demonstrated the presence of p6 symmetry in the Gag lattice using immature HIV-1 prepared from T-cells treated with a protease inhibitor, and in *in vitro* assembled particles (26). All of these cEM studies thus demonstrate the

presence of a CA hexameric lattice with a centre-to-centre spacing of 8 nm, and that the MA, N-CA, C-CA and NC domains are indeed distinct in the immature virion. They also demonstrate that *in vitro* models, using either VLPs or Gag mutants containing at least the CA and NC domains, faithfully represent the arrangement of such domains in the immature capsid.

A Gag mutant used in this thesis for HIV-1 assembly assays, ΔMACANCSP2, has also been subject to cEM, as has a similar mutant lacking the SP2 spacer (60). Both such mutants have residues 15-99 of MA deleted, but contain the entire CA, SP1, and NC domains. (ΔMACANCSP2 is used for various experiments in this thesis, and is summarised in Figure 3.1.) The *in vitro* assembled particles had an outer diameter of 114 ± 6 nm as determined from cEM, and 90 nm from negative-stain TEM, the smaller measurement being attributed to shrinkage as a result of the staining process. Radial density profiles from cEM revealed almost the same density for the CA region as for PR^- virions from T-cells, and the NC density, although occurring at the same radial displacement from the CA layer, to be slightly higher for the *in vitro* particles. Not surprisingly, there was a much-reduced density for ΔMACANCSP2 at the expected level of MA, owing to the large deletion, though this observation has the useful corollary of justifying the use of this particular mutant as a model for immature capsids, at least for the study of the CA and NC interactions. This observation also provides a further convincing demonstration of the orientation of Gag in the immature state, i.e. that it is oriented with its C-terminus pointing towards the centre of the virion (or VLP etc).

Hexameric CA packing in prebudding assemblies at the plasma membrane of baculovirus-infected insect cells expressing Gag, that is, a VLP system studied before particles are fully assembled, also shows a hexameric lattice, using negative-stain TEM (113). The centre-to-centre spacing of such hexamers was 7 nm, the reduced value as compared to the cEM consensus value of 8 nm possibly being an artefact of staining, which is avoided in cEM. The authors of this study were the last to publish a claim of icosahedral symmetry in immature HIV-1 Gag arrangement.

HIV-1 CA with the 11 N-terminal residues of NC, when using a short his-tag at the CA N-terminus (which the authors call his-$CANC_{11}$), assembled on lipid monolayers in 2D (103). Hexameric order was seen with 64 Å spacing, interpreted as N-CA hexamerisation since the resolution was too poor to resolved N-CA and C-CA separately. A similar approach

was applied to a different Gag mutant MACANC, which has the entire domains listed in its name and the SP1 domain (70). From 50-200 nm patches of 2D crystals on lipid monolayers, image reconstructions by tilt series were used to generate a model. Unlike work with VLPs, NC did not contribute to ordered regions significantly and, unlike his-$CANC_{11}$ monolayer assembly, hexamer spacing was 8 ± 0.2 nm, consistent with the consensus measurements of immature CA hexamer spacing. The MA domain was also poorly resolved. Despite such limitations a model was proposed similar to that for mature HIV; N-CA forms hexamer contacts linked by symmetric C-CA dimers. The arrangement of MA was suggested to be trimeric, with such trimers linking three N-CA hexamers. NC was suggested to be a dimer beneath the C-CA dimer.

Different again is recent cET data of inactivated HIV-1 immature virions (159). The electron density profile generated in this study accommodates a model in which N-CA forms a hexamer with ~ 8 nm spacing, and C-CA arranges beneath the N-CA layer, also with hexameric order. Hexamer contacts *via* C-CA dimerisation are also accommodated. Hexameric order in the SP1 domain was also detected, implying that SP1 forms hexamers beneath CA, but not at the level of NC. The MA layer was disordered and not included in the model for immature Gag arrangement. This data also demonstrated that the immature lattice may not be continuous, that is, there may be gaps on the CA surface of the lattice rather than being a fully closed structure as typically envisaged. It was estimated that only 40% of the viral surface is composed of an ordered lattice.

Recently, cET has been applied to immature HIV-1virions and to *in vitro*-assembled particles (25), improving the best resolution structure of the immature Gag arrangement to 17 Å. The likeness of the Gag lattice in authentic virions and *in vitro*-assembled particles was confirmed once again. The data presented in (25) largely support the model presented in (159), except with improved resolution, again demonstrating that N-CA forms hexamers with centre-to-centre spacing of ~ 8 nm, and that C-CA packs below N-CA. Six-fold (p6) and three-fold (p3) symmetry were locally present, and two-fold (p2) symmetry at the C-CA radius in reconstructed tomograms was taken as indicative of C-CA dimerisation, linking the hexamers. Atomic models based upon published CA domain structures were suggested, but remain ambiguous. Notably so is their proposed N-CA orientation, which differs from that of Wright et al (159), even though both place helices 1-3 of N-CA at the centre of the helix.

The proposed SP1 six-helix bundle could be accommodated, but the authors note that other arrangements are also possible. The immature lattice was again found to be discontinuous, less so for *in vitro*-assembled particles than for authentic virions, and the important observation was made that such discontinuities do not accommodate CA pentamers, implying that icosahedral symmetry is absent.

A highly discontinuous Gag lattice in the immature capsid is also supported by fluorescence fluctuation spectroscopy of VLPs (31). The consistent observation of a discontinuous lattice has the consequence of making it very difficult to model immature capsids mathematically, for example using the approaches of Zlotnick and co-workers, or using equilibrium polymerisation models (42, 43, 76, 167).

Both the recent cET-derived immature hexamers are compared in Figure 1.4, which also includes models for the mature hexamer. The two models for the immature hexamer are different, as has already been noted, but it is not clear whether such differences in domain orientations are real, or a consequence of the inherently low resolution of cryo-electron tomography.

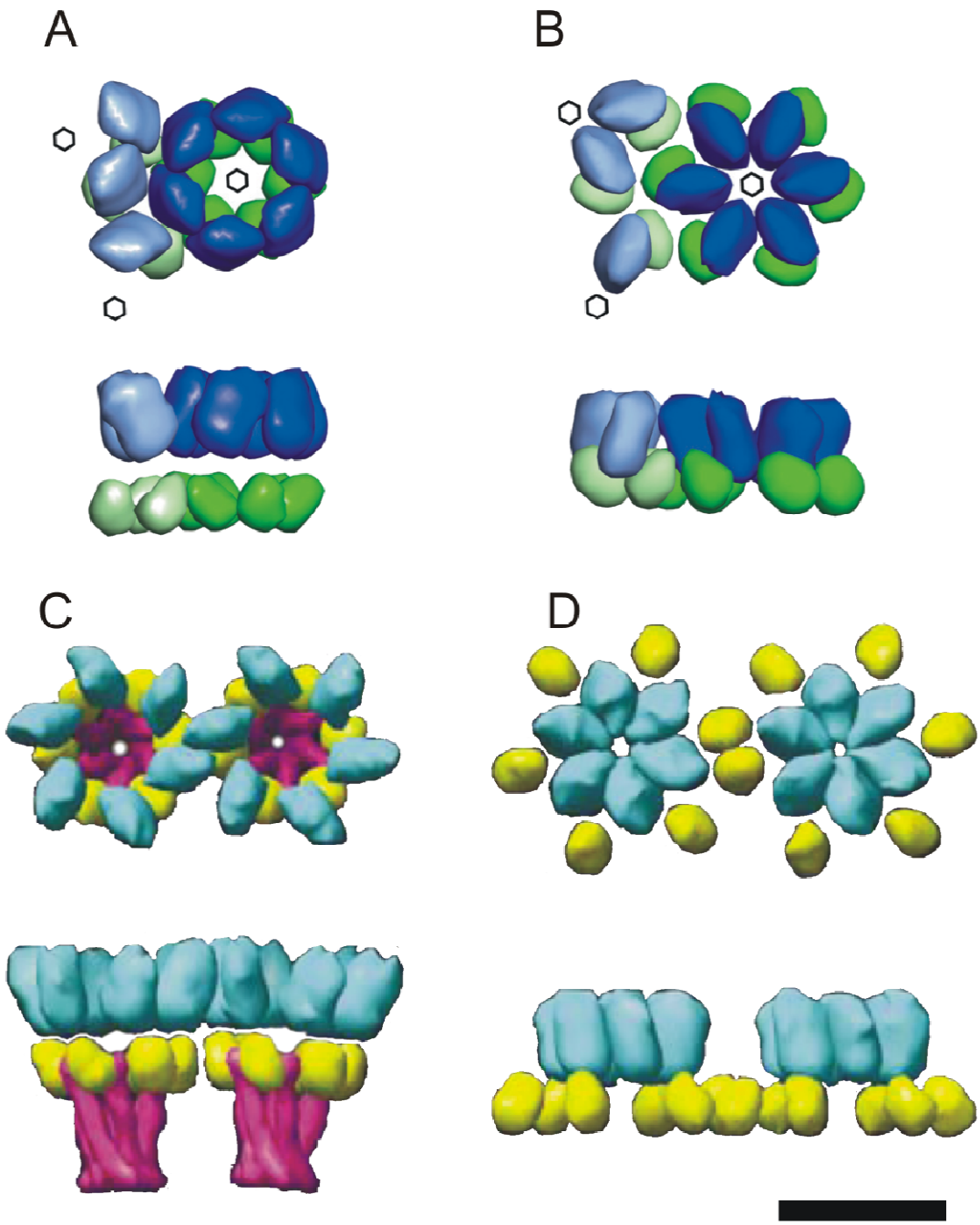


Figure 1.4: Cryo-electron tomographic models of the immature and mature CA arrangement.

A and C) models of the immature hexamer, as proposed by (25) and (159) respectively (from which the figure has been adapted). B and D) models of the mature hexamer by the same authors. In A and B, N-CA is in blue and C-CA in green. In C and D, N-CA is turquoise, C-CA is yellow and the linker, up to the C-terminus of SP1, is magenta. Scale bar: 8 nm

1.5.3. The consensus view of immature capsid structure

The consensus view of the immature capsid structure is that CA forms hexamers with a centre-to centre distance of 8 nm, with the C-CA domain directly below the N-CA domain, and hexamers are linked by C-CA dimerisation. This basic arrangement is illustrated in Figure 1.5. The precise orientations and structures of both the N-CA and C-CA domains in the immature state are not known. The C-CA domains are closer to one another than occurs in mature virions, being underneath the N-CA domains rather than pushed outwards. The SP1 arrangement may also be hexameric, directly below the CA hexamers, and NC does not arrange in an ordered fashion, or at least its arrangement cannot be reconciled with a CA hexamer lattice. The arrangement of MA is also unclear, if it is arranged at all. According to a study in which myristoylated MA or myristoylated MACA was assembled as 2D crystals on membranes, MA could form hexamers above those of CA (8), though the proteins used are not capable of assembling to give particles resembling immature virions. It has also been suggested to form trimers above the CA lattice (70). It is of course possible that MA does not arrange in an ordered fashion at all, or that it arranges as trimers which do not have strict symmetry relative to the CA lattice, in which the threefold or sixfold averaging used in many cEM studies would ‘smear’ its electron density, giving the impression of a disordered arrangement. A similar argument can be applied to the NC domain. It has been proposed that NC is dimeric in immature virions (70), with such dimers arranged directly below C-CA dimers on the basis that replacement of NC with a dimerising domain removes the nucleic acid dependence of assembly (73). The existence of dimeric NC in virions has not been observed, neither has dimeric NC been observed in solution, such that this arrangement is unlikely. Alternatives, and a review of the deficiencies of current models, are reserved for the discussion chapter.

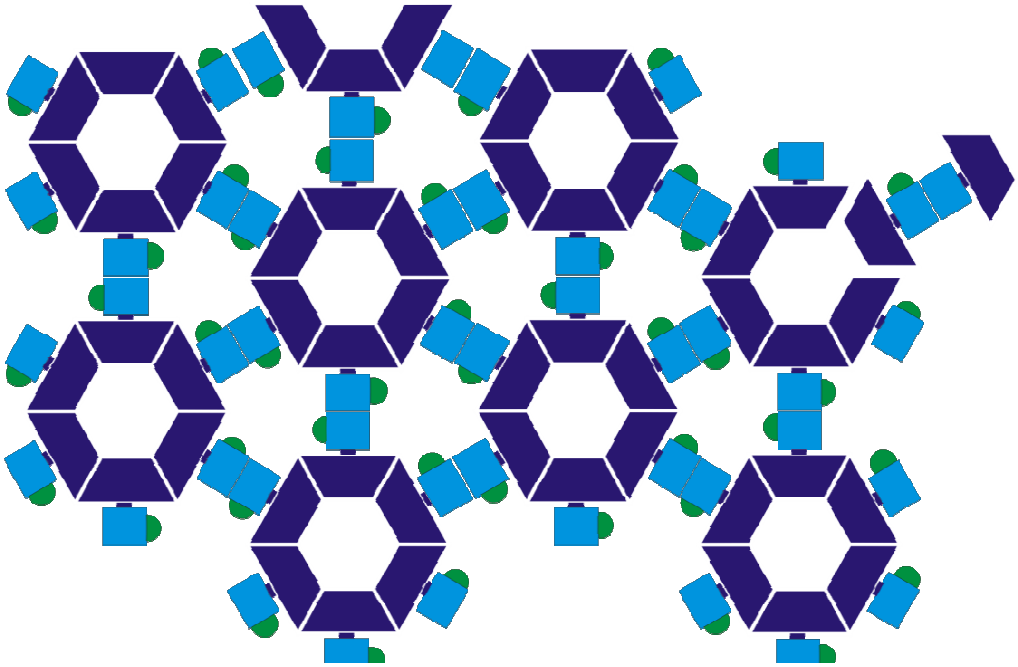


Figure 1.5: Schematic arrangement of CA and NC in the immature lattice.

N-CA is shown as a dark blue rhomboid, C-CA as a pale blue square and NC as a green triangle. The lattice has been flattened, to show the symmetry more clearly. NC has been shown in an arbitrary position, since it seems not to have a symmetric arrangement.

1.6. Atomic models for the assembled states

1.6.1. C-CA dimerisation new version

The C-terminal domain of HIV-1 CA, C-CA, dimerises in solution with a K_d in the micromolar range, and is thus unique amongst retroviral CA proteins which generally do not form dimers in solution. Several studies have sought to determine the dimerisation K_d for C-CA both alone and as a component of larger proteins, all such studies making use of analytical ultracentrifugation. For C-CA alone (residues 146-231) the dimerisation K_d has been reported as 10 ± 3 μM (46), whilst independent studies have published affinities for full-length CA of 18 ± 1 μM (161), 13 μM (134) and 2.94 ± 0.8 μM (114). For C-CA-NC, the K_d has been reported as 1.78 ± 0.5 μM (114).

Several different structures for the C-CA dimer have been proposed from X-ray crystallography and, although the C-CA domain structure is generally similar in each, there is considerable discrepancy as to the precise orientations of C-CA monomers relative to one another in the dimer. To date, the structures available are for C-CA alone (146-231)

(158), C-CA lacking the first five residues (151-231) (46), C-CA in complex with a peptide inhibitor of assembly called CAI (148) (see chapter 1.9), various structures of single site mutants of C-CA in complex with CAI (13) and a single site deletion mutant Δ177 as a domain-swapped dimer (72). It is consistently observed that dimerisation is mediated by approximately parallel packing of helix 2, involving hydrophobic contacts between V181, W184 and M185, whose role in dimerisation is corroborated by mutational analysis (152). The major exception is the domain-swapped dimer (72) in which the N-terminal region of C-CA, including helix 1, is exchanged with the symmetry-matched partner, though the helix 2 interactions are still present.

The available dimer structures have, until recently, left open the questions of whether the dimer interfaces in the immature and mature assembled states are distinct from one another, and whether the solution-state dimer interface is different again. It may be the case that the dimer interface is inherently ill-defined and the monomers contributing to the dimer are able to move relative to one another. It is thus possible that the solution-state dimer represents an ensemble of dimer structures, all able to exchange amongst dimeric conformations, of which particular conformations may be selected by the assembly processes. This is consistent with published NMR data for C-CA-NC (114), with the NMR data for the same protein presented in this thesis and with unpublished NMR data for C-CA (Ludovic Pecqueur). That is, the broad and often invisible lineshapes for the helix1-helix 2 loop and helix 2 of C-CA may be due to exchange between dimeric conformations, rather than between monomeric and dimeric states.

1.6.2. The mature state

Recently, significant improvements have been made in the level of detail with which the arrangement of CA in the mature state is known, and are largely attributable to two papers (47, 123). In the first of these, a CA mutant with an R18L substitution was expressed (CA_{R18L}), the mutation allowing for assembly of CA_{R18L} as spheres, cones and cylinders, rather than the long tubes expected for wild-type CA. Conditions were chosen under which CA_{R18L} assembled as large spheres, which were flattened on an EM grid and examined as 2D crystals by electron crystallography, from which a model for the mature CA arrangement was obtained at 9 Å resolution, which is shown in Figure 1.6 A. At this resolution, individual helices were visible, enabling unambiguous placement of high-resolution C-CA and N-CA structures within the observed electron density map. As a

result, many features of the mature lattice were characterised in more detail than has been previously observed. N-CA and C-CA were individually docked into the density. The N-CA hexamer interface comprised helices 1, 2 and 3, which associated at the centre of the N-CA hexamer as an 18 helix bundle. Of the available C-CA dimer structures, pdb entry 1a43 (158) was the best fit, though the monomers still required slight changes in their orientations to fit the density map. No other pdb entry of a dimeric C-CA structure could be accommodated, including C-CA bound to CAI (13, 148), C-CA lacking its 5 N-terminal residues (46) and domain-swapped C-CA (72). Thus the mature state appears to select a particular dimer conformation of C-CA, closely resembling that of C-CA crystallised at pH 8 (158) (pdb entry 1a43). The other available dimer conformations, notably including that of CAI-bound C-CA, are therefore presumably incompatible with the dimer conformation selected by mature assembly.

It was also found (47) that the C-CA and N-CA domain make very few intramolecular contacts with one another, but intermolecular contacts between the N-CA and C-CA domains are extensive. Previous analyses of mature capsids had suggested that intermolecular N-CA/C-CA interactions occur (85, 86), and that K70 (of N-CA) and K182 (at the N-terminal end of helix 2 of C-CA) are in close proximity in the mature state by means of an intermolecular interaction. The electron crystallography study confirms this to be so: helix $4_{N\text{-}CA}$ (i.e. helix 4 of N-CA) was found to insert into a groove in an adjacent C-CA, making putative contacts with helices $1_{C\text{-}CA}$ and $2_{C\text{-}CA}$ on 1 side of the groove and helices $3_{C\text{-}CA}$ and $4_{C\text{-}CA}$ on the other.

The second major contributor to current knowledge of the atomic structure of the mature HIV-1 CA_{AA} hexamer are the X-ray crystal structures of a CA_{AA} hexamer stabilised by covalently linked cysteine residues in N-CA (A14C, E45C substitutions were made in N-CA) and the CA_{AA} hexamer structure of a chimera of CA_{AA} with the stable hexamer-forming protein CcmK4 at the C-terminus (123). By CA_{AA} is meant that C-CA was the mutant form with W184A/M185A substitutions, to enable discrete hexamers to be formed. Both the cysteine mutant and CcmK4 chimera formed identical hexamers, and gave structures to 1.9 Å (crosslinked) and 7 Å (chimeric). A hexamer of cross-linked CA_{AA} is shown in various forms in Figure 1.6 B-D. All details previously observed were present, though the improved resolution showed several previously unobserved features of the mature CA state. The N-CA hexamer interface involved predominantly hydrophilic

interactions, and is a highly hydrated interface. Despite using the CA_{AA} mutant, the C-CA_{AA} domains were in a dimeric state with orientations consistent with the electron crystallography, vindicating the use of C-CA_{AA} as a model for C-CA. In some crystals, no density was visible at the N-terminus of helix $2_{C\text{-}CA}$ between residues 176 and 187, which includes the location of the alanine substitutions, whereas in others two disctinct conformations could be seen. One conformation has the helix $1_{C\text{-}CA}$/helix $2_{C\text{-}CA}$ loop in contact with N-CA (a putative interaction not previously seen) whilst the other has no contact between the $1_{C\text{-}CA}$/helix $2_{C\text{-}CA}$ loop and N-CA. Most of the contacts at the intermolecular N-CA/C-CA interface were between extended side chains, and corroborate mutagenesis results that CA with Y169A or L211A substitutions are incompatible with mature assembly (13). The variability of the C-CA_{AA} dimer conformation in the mature state is presumably important for enabling the variable curvature required by the conical morphology of the mature capsid, and for allowing pentamers to be incorporated into the otherwise hexameric lattice.

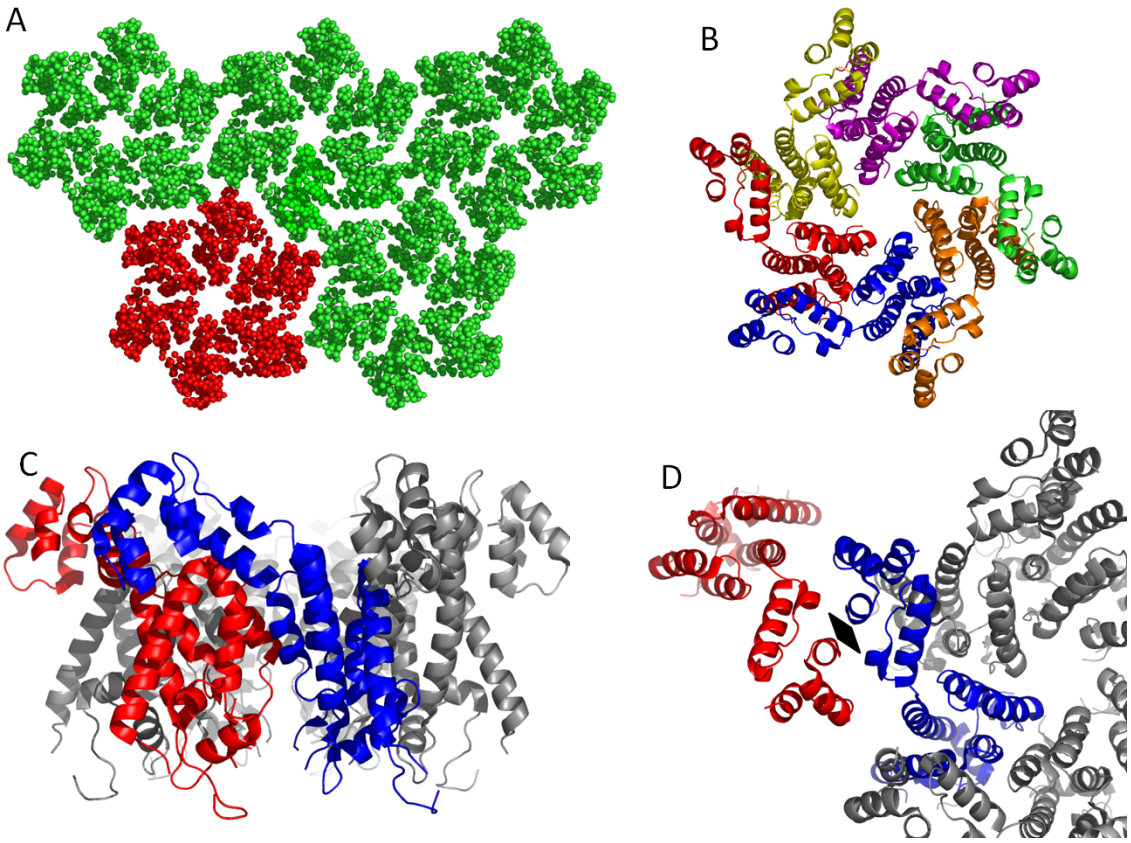


Figure 1.6: The mature CA hexamer.

A), pseudoatomic model of CA lattice flattened on a carbon grid, PDB code 3dik (47), shown from within the virion, 'below' the lattice. Five hexamers are shown; one entire hexamer is highlighted in red. B) X-ray crystal structure of a CA_{AA} mutant with A14C/E45C mutations to allow N-CA covalent cross-linking, viewed from 'below' with C-CA_{AA} above the plane of the page, projecting outwards from the body of the N-CA hexamer, PDB code 3h47 (123). C) The same CA hexamer from the side, showing intermolecular N-CA/C-CA interactions between the molecules highlighted in red and blue. The C-CA_{AA} domain is at the top of the picture, representing the inside of the capsid. D) The C-CA_{AA} dimer interaction linking a hexamer to an adjacent CA molecule, with the C_2 symmetry axis represented by the black diamond. Note the large distance between N-CA domains of the blue/grey hexamer and the red neighbour, characteristic of the 'loose' CA packing in the mature lattice.

1.6.3. The immature state

There is not yet an atomic model for the immature state with the same resolution as that for the mature state, the closest so far being the tomographic reconstruction of ΔMACANCSP2 at 17 Å resolution (25). The authors attempted to determine which of the published C-CA dimer crystal structures best fit their electron density map by constraining the two-fold symmetry axis of the crystal structures to be coincident with that of their cET electron density map, and by determining the number of intermolecular contacts and clashes expected between adjacent C-CA dimers as a result. They found that the dimer structures most consistent with their own observations, subject to the additional constraint that the C-CA N-terminus is in reasonable proximity to the N-CA density, were the CAI-bound structure (13) and the domain-swapped structure (72). Notably, pdb entry 1a43, the crystal structure of dimeric C-CA (146-231), which was the closest match for the mature state, fitted in such a fashion as to place its N-terminus at a position considerably displaced from the N-CA domain, making such an arrangement unrealistic for the immature state. Thus it is possible that the immature state selects a C-CA dimer conformation which is distinct from the mature state, and more closely resembles that selected by CAI. That the CAI-bound C-CA dimer conformation is similar to the C-CA dimer conformation in the immature state may explain why CAI is less efficient at inhibiting immature assembly.

1.7 The HIV-1 genome and the packaging role of NC

1.7.1. Genome structure

The genome of HIV-1 is composed of a single RNA molecule of length 9.7 kilobases, of which, on average, two copies are found per virion (129). A region of ~120 bases important for genome packaging is located near the 5' end of the genome, between the untranslated 5'-LTR (long terminal repeat) and *gag* initiation codon, which is termed the ψ-site (6). This region has been extensively studied and found to contain four stem-loop structures, labelled SL1 to SL4 (34, 62-64, 163), and is implicated as directing the selective packaging of HIV-1 genomic RNA during immature assembly (6, 33, 91, 137) (reviewed in (94)). Studies which detected stem-loops were predominantly carried out *in vitro*, though more recently the suggested stem-loops were also observed *in vivo* (118, 162). Different overall secondary structures for the entire ψ-site were found, but the consensus remains that there are four stable stem-loops (34). Of these, SL1 appears to be involved in dimerisation of the genome in the absence of any other signals (35, 67, 99, 141), although genomic RNA dimerisation appears to be a complex process involving major structural rearrangements within the ψ-site (71) and recent evidence suggests that dimerisation is not necessary for genome packaging (142), although this remains contested (117). It is likely that secondary structure in HIV-1 genomic RNA is necessary for reverse transcription (17).

It is thought that SL3 is the most important determinant of selective recognition of the viral genome in preference to the multitude of other nucleic acids present in the host cell, since it alone can direct the packaging of heterologous RNAs into VLPs (63). Various studies indicate that mutations SL3 result in reduced levels of genome packaging (94).

Recently, a thorough analysis using the SHAPE (Selective 2'-Hydroxyl Acylation analysed by Primer Extension) method has provided a model for the secondary structure of the entire HIV-1 genome (155). This shows many stem-loops punctuated by unstructured regions, the prevalence of stem-loops as secondary structure elements perhaps explaining selective genome recognition, since it is known that the packaging signal has such secondary structure.

1.7.2. Genomic RNA is not necessary for immature assembly *in vitro* or *in vivo*

Selective genome recognition is not prerequisite for immature assembly; if the entire ψ-site is deleted from the HIV-1 genome, assembly proceeds nonetheless, though the resulting particles generally lack genomic RNA (6), indicating that the mechanism of immature assembly is not nucleic acid sequence-dependent. This is supported by the observation that HIV-1 often packages high levels of cellular RNA and DNA in addition to its own genome (68, 111), and that if mutations are made in the NC domain which reduce specificity for the ψ-site, normal assembly proceeds but yields non-infectious virions (56). Similar results were obtained in a study in which systematic mutations were made in the NC domain and, by expressing the mutant viral DNA in HeLa cells, only slight reductions in assembly were observed even though the resulting virions were not infectious (40), that is, packaged very little genomic RNA. The same observation was made using Gag truncation mutants expressed in insect cells, wherein reduction in viral RNA packaging did not correlate with a reduction in assembly of VLPs (74).

The literature described in the preceding paragraphs hence indicates that neither genomic dimerisation nor sequence are determinants of immature assembly, and that assembly can be studied with simple model nucleic acids systems rather than the entire HIV-1 ψ-site.

1.7.3. The NC domain is essential for immature assembly

The NC domain is essential for immature assembly, as has been demonstrated by many studies. For example, using an insect cell VLP system, deletion of the NC abolishes assembly (135), which is corroborated by other similar studies making systematic truncation mutations in Gag and expressing in insect cells (66, 74), COS-1 cells (16), or using a cell-free expression system (121). *In vitro* assembly studies making use of purified Gag mutants also indicate that in the absence of the NC domain the resulting particles are tubular rather than spherical (28, 29, 59). The only exception occurs for *in vitro* assembly of a Gag mutant containing the entire CA domain with an N-terminal extension of the basic regions of MA (the same MA residues as used in ΔMACANCSP2), which assembles as heterogeneous approximately spherical particles (58), but the insect cell VLP studies show that truncation mutants lacking NC are not released from the host cell.

1.7.4. The role of NC-nucleic acid interactions in immature assembly

The molecular basis for the requirement of nucleic acid in immature assembly has remained elusive. Although, as discussed, NC-nucleic acid interactions are required, there remains some debate as to why this is so. In certain simple viruses, with only a single assembly step, interactions between a small number of contributing capsid proteins with nucleic acid seems to aid in the formation of a nucleus, to which capsid proteins may add without further interaction with nucleic acid. Later in this thesis, data will be presented which demonstrates that this is not so for HIV-1.

It has been observed that a purified protein containing the entire CA and SP1 domains, but with the NC domain replaced only by a glycine and a cysteine can, by means of placing the protein in oxidising conditions, assemble as approximately spherical particles, albeit in low yield (7). It is worth noting that the protein also had a His tag, which may mimic the role of the basic regions of MA in ΔMACANCSP2, and that the addition of a cysteine-reactive cross-linking reagent promoted the formation of small oligomers (analysed by non-reducing SDS-PAGE), but a very low yield of pelletable material. The authors concluded that the role of NC-nucleic acid interactions is to promote the formation of Gag dimers, which then undergo a conformational change and form assembly-competent oligomers. Such oligomers essentially act as a nucleus from which assembly proceeds. A similar conclusion was reached using an RSV (Rous Sarcoma Virus) VLP system in which an RSV Gag mutant was expressed with the NC domain replaced by a leucine zipper domain which is known to dimerise (73). Assembly of spherical particles in high yield and with good homogeneity was observed, and the authors concluded that NC-nucleic acid interactions serve only to dimerise Gag, such dimers being the unit of assembly.

In contrast to the models outlined above, several studies have concluded that NC-nucleic acid interactions act as a scaffold upon which Gag-Gag interactions yielding immature capsids can proceed. By this is meant that many Gag molecules may be aggregated upon a single oligonucleotide, which by some mechanism leads to the formation of immature capsids, presumably since a high local Gag concentration favours otherwise weak protein-protein interactions (29, 59). Another alternative suggestion is that NC-nucleic acid interactions promote curvature in a Gag lattice anchored to a plasma membrane (133). This model, however, makes the implicit assumption that Gag assembles as flat sheets on membranes in the absence of nucleic acid.

A systematic study of the effects of nucleic acid upon immature assembly has been conducted by means of *in vitro* assembly of a purified Gag mutant lacking P6 (GagΔP6) (28). In this study it was observed that RNA of variable length and sequence could promote assembly of small (25-30nm) spheres. It was also observed that single-stranded DNA could serve the same purpose, and the dependence on oligonucleotide length was examined using poly(dA) and poly(dTG), with gel densitometry used to quantify the fraction of assembled material. Particle assembly (more accurately, the fraction of material in the insoluble fraction) was dependent on both the length and concentration of DNA. Using poly(dTG), it was seen that 10 nucleotides were sufficient for assembly, albeit inefficient, and that a mass ratio of at least 4% DNA was needed. Longer DNA molecules up to 30 bases in length for poly(dTG) generally favoured assembly but only at mass ratios <5%, above which assembly was more efficient with only 15-20 bases. This may suggest that immature assembly proceeds from a nucleation event which must be sufficient to link at least 2 Gag molecules (scaffold nucleation), but that such nuclei must only be allowed to form in limiting amounts. Prevalent nuclei may favour rapid elongation but, by virtue of the rapid diminution of Gag concentration, lead only to incomplete particles. It is estimated, based on the dimensions of Gag, that 10 bases is the minimum requirement to span 2 Gag molecules. The interpretation given here of the author's results that nucleic acids must link at least two Gag molecules to one another is tested rigorously in this thesis.

1.7.5. Structures of NC free and in complex with nucleic acids

NC-nucleic acid interactions have been examined structurally, and NMR has revealed the structures of NC bound to a 20-nucleotide construct including the SL3 RNA sequence (39), a similar 20-nucleotide sequence including the SL2 RNA element (11) and to a minimal 5-nucleotide DNA, dACGCC (107). In all (bound) cases, Trp282 (Trp37 in NC numbering) stacks against a guanine residue, and the mutation of Trp282 is deleterious to this interaction (153). Common interactions to both structures are hydrophobic contacts to nucleic acid made by V258, F261, T269, A270, W282, Q290 and M291, and hydrogen bonds to nucleic acid made by K259, R271, R277, K278, K279, G280 and K292. A notable difference in binding mode is that F261 participates in base stacking in the dACGCC complex but not the SL3 complex.

The conformations of the zinc knuckles are individually similar between all structures, but their relative orientations differ due to a different linker conformation. This region of

protein between the two zinc knuckles, ^{274}RAPRKKG, has a number of different conformations compatible with nucleic acid binding. This linker may therefore act as a hinge to bring the two zinc knuckles into an orientation compatible with nucleic acid binding, such that many different DNA or RNA sequences may be bound. All nucleic acid-bound structures imply that binding promotes the formation of intra-protein contacts within NC, particularly between the two zinc knuckles, leading to a more motionally restricted protein-nucleic acid complex. This is on the basis of increased intensity of NOEs between residues in the zinc knuckles when bound, relative to free NC (106, 109). NMR relaxation data using the NC-dACGCC complex also support a more rigid structure in the bound state (128). This study, however, examined only 14 positions in NC, which had been ^{15}N-isotope-labelled, and does not therefore pertain to the entire protein.

A potentially important difference between the two RNA-bound structures and the DNA-bound structure is that full length 245-300 NC was used in the SL3/SL2 structures, but that used for the dACGCC structure lacked the N-terminal 12 residues and 2 C-terminal residues. The N-terminal section in the SL3/SL2 structures forms a 3_{10} helix spanning residues K248-R255, which packs against the N-terminal zinc knuckle and fits into the major groove of the SL3/SL2 RNA. The SL3/SL2 RNA is thus held in a position which could conceivably be very near the C-CA domain of Gag or distort the solution structure of Gag relative to the unbound form. The implications of this are discussed in more detail in a later section. The 3_{10} helix differs in orientation between the SL2-bound and SL3-bound structures by approximately 90°, but the relative zinc knuckle orientations are comparable.

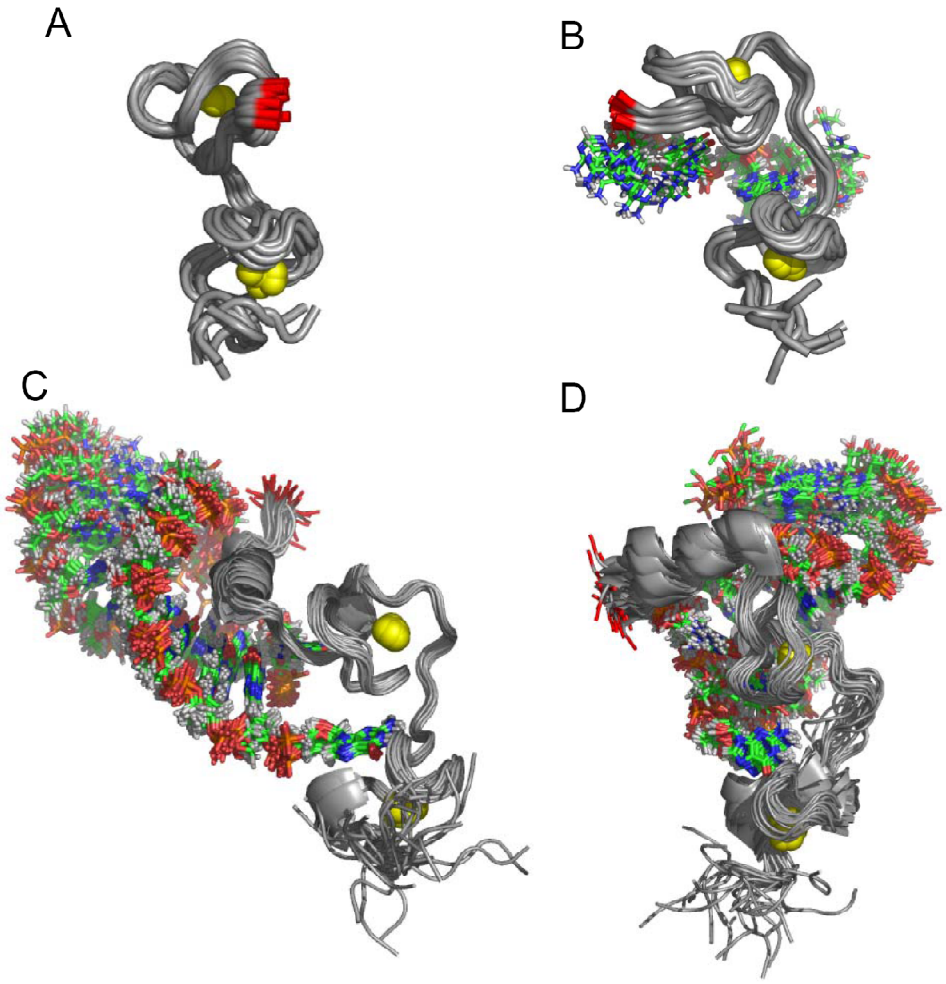


Figure 1.7: Families of NMR structures of HIV-1 NC with and without various nucleic acid ligands.

(A), HIV-1 NC bound to dACGCC (B), bound to SL3 RNA (C) and to SL2 RNA (D). In each case the N-terminal residue is shown in red, and the orientations are chosen such that the C-terminal zinc knuckles approximately superimpose. PDB codes: NC, 1ESK; NC-SL2, 1F6U; NC-SL3, 1A1T (39); NC-dACGCC, 1BJ6 (107)

1.7.6. NC-nucleic acid interactions revisited

There is no clear consensus as to the stoichiometry of NC-RNA or NC-DNA interactions, nor whether DNA is indeed a good model for RNA in this context, at least if it is the precise nature of the interactions one wishes to study. Using SL3 (RNA) with 3-base complementary extensions at both the 5' and 3' termini to extend the stem length, the structure of a 1:1 NC-SL3 complex has been solved by NMR spectroscopy (39). The

stoichiometry was determined by gel filtration chromatography, whilst a K_d of ~100 nM was determined by native PAGE. The strong binding constant was supported by the observation of slow exchange in NMR spectra. A different study also used SL3 RNA, though it was not stated whether this had the extended stem, and by ITC NC was found to bind with a K_d of 170 ± 65 nM to NC, the data being consistent with a 1:1 stoichiometry, but no other models were fit to the data. Analysis of NMR chemical exchange data in the same study yielded an off-rate of 5 ± 1 Hz (10). The conditions used for ITC were pH 6.5, 25mM NaCl.

The structure of NC in a 1:1 complex with SL2 RNA is also available (10). In this study, NMR spectra of a mixture of NC and SL2 at a 1:1 ratio were found to contain only a single set of signals, indicating that a second site of comparable affinity to the first is not present. However, when excess NC is added, line-broadening of the NC signals is observed, indicating low-affinity binding in addition to a high-affinity site. A 1:1 complex was also supported by gel filtration chromatography, whilst chemical exchange in 2D NOESY spectra were analysed to give an off-rate of 6 ± 1 Hz. This analysis assumes a 1:1 stoichiometry, rather than allowing optimisation of the number of binding sites. In the chemical exchange analyses, as for NC-SL3, the RNA concentration was in 50% excess over NC, which should favour 1:1 binding. The same study applied ITC, again assuming a 1:1 binding model, to give a NC-SL2 K_d of 110 ± 50 nM, and again using conditions of pH 6.5, 25 mM NaCl.

A different study used RNA molecules closely based upon the SL2, SL3 and SL4 stem-loops (97). The SL2 oligo had the same loop sequence as wild-type, but a slightly modified stem, whilst the SL3 and SL4 RNAs each had extended stems by 6 bases (3 base-pairs), in total having 20 bases each. Fluorescence spectroscopy was used to examine binding, at pH 7 but no mention was made of salt concentration, which is known in influence protein-nucleic acid interactions. The authors determined the binding site size by extrapolation of the initial slope of a single binding isotherm (at 1 μM NC) and the limiting fluorescence. Direct linear plots of 1/[free nucleotide] against 1/(quenching) were used to generate affinities, from which the stoichiometry was estimated with the binding site size as a restraint. The authors report NC:RNA stoichiometries of 2:1 for SL2, SL3 and SL4, and likewise for their DNA counterparts. The data are summarised in Table 1.1, adapted from (97), showing that the results for RNA and the DNA counterparts are similar. The binding

constants are much stronger than those reported from ITC, which may be due to either the absence of NaCl, or the use of an inappropriate model. However, the means of analysis is questionable for several reasons. Firstly, the initial slope of a binding isotherm is generally not well-determined, and if non-specific recognition is present then the 'limiting fluorescence' does not represent the true maximum extent of fluorescence quenching, since lattice saturation is not reached in practical concentration regimes. Where the assumption of a single binding site cannot be justified, titrations at multiple protein concentrations can be used to generate a more accurate estimate of the minimum number of binding sites.

Table 1.1: Equilibrium binding constants of NC for SL(2-4) DNA and RNA. Such data are reproduced from (97).

Oligo	NC:Oligo stoichiometry	K_d (nM)
SL2	2:1	1.49
dSL2	2:1	5.56
SL3	2:1	0.769
dSL3	2:1	10.0
SL4	2:1	11.5
dSL4	2:1	25.0

It is possible that binding appears stronger than it is since the assumption of non-overlapping binding sites is used. If a model assuming non-specific binding is applied, one expects that the binding constants would be weaker. This is also the case if there are in reality more non-overlapping binding sites than are assumed in the model, but in the absence of any salt, non-specific binding of proteins to nucleic acids is common.

The binding of NC to RNA stem-loops was investigated in a similar paper, again using fluorescence quenching, with the inclusion of 200 mM NaCl in the buffer (140). This study is under conditions very close to those used for fluorescence titrations in this thesis. The authors note that titrations were not reproducible if the NaCl concentration was below 150 mM. Titrations were performed using a fixed NC concentration of 0.3 μM. The SL2 variant used was exactly the 19 base sequence which occurs *in vivo*, whilst the SL3 variant again had an extended stem of 3 base pairs, the entire molecule being 20 bases in total. SL4 was also used, with 16 bases in total. Data were only described using a model assuming a single specific binding site, and the results of fitting such a model are

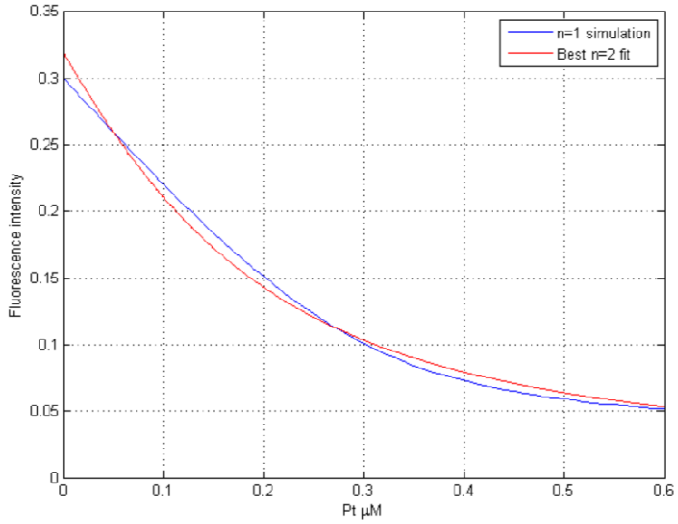
summarised in table 1.2. Two SL3 mutants were also used, one in which the GGAG tetraloop was replaced with the sequence UUCG, and one with the sequence GAUA, causing reduced affinities.

Table 1.2: Affinities of the NC domain for stem-loop RNAs, assuming a 1:1 stoichiometry, as determined by fluorescence quenching. The data are reproduced from (140).

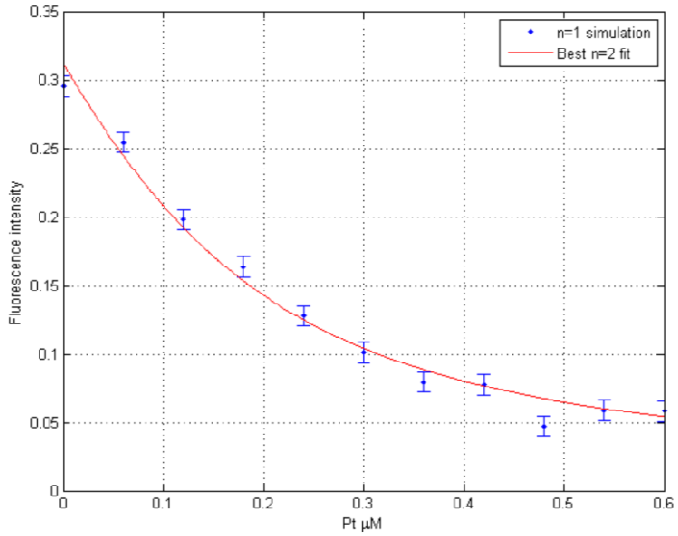
Oligo	NC:Oligo stoichiometry	K_d (nM)
SL2	1:1	23 ± 2
SL3	1:1	28 ± 3
SL3(UUCG)	1:1	~7500
SL3(GAUA)	1:1	~10000
SL4	1:1	320 ± 30

In Figure 1.8 A, the titration performed in (140) to determine the affinity and stoichiometry of NC for the SL3(20-mer) variant is simulated using a 1:1 model with the reported K_d of 28 nM and 90% fluorescence quenching at saturation. A 2:1 model is then fitted to this simulation, and the reader will appreciate that, even if modest experimental error were present, discriminating between 2:1 and 1:1 stoichiometries would be very difficult from this single titration only. It is to be noted that, in the fitting, the protein concentration was fixed at the simulation value of 0.3 μM, as used experimentally. The simulation extends to a SL3 (20-mer) concentration of 0.6 μM, again as reported by the authors, and results in a 2:1 K_d of 200 nM, which is around 7 times weaker than the reported 1:1 affinity of 28 nM. In Figure 1.8 B, the estimated 2:1 K_d is used to simulate a 2:1 fluorescence titration with 5% noise, then a 1:1 model is fit to this. The parameters reported in (140) are reproduced very well, again demonstrating that deriving a binding stoichiometry from a single titration is not practical. Non-specific binding is also possible, and such a model can be fitted to the one-site specific model reported by the authors to give a binding isotherm in close agreement with the one-site simulation. The non-specific K_d corresponding to a one-site model of 28 nM affinity binding, as occurs for 20-mer SL3 RNA, is ~1.43 μM. This is also shown in Figure 1.8 C, indicating the difficulty in discriminating between binding models.

A


n=1 simulation
Best n=2 fit
Fluorescence intensity
Pt µM
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.1
0.2
0.3
0.4
0.5
0.6

B


n=1 simulation
Best n=2 fit
Fluorescence intensity
Pt µM
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.1
0.2
0.3
0.4
0.5
0.6

C

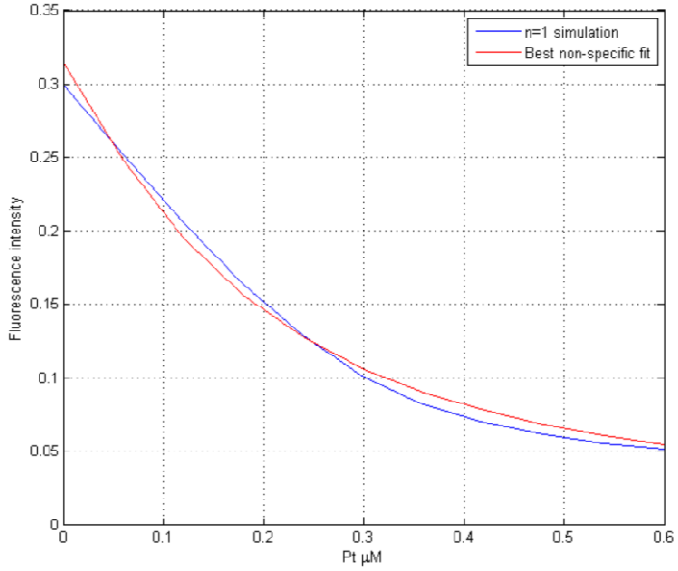

n=1 simulation
Best non-specific fit
Fluorescence intensity
Pt µM
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.1
0.2
0.3
0.4
0.5
0.6

Figure 1.8: Simulated fluorescence quenching curves.

A) Fluorescence titration simulated with a single binding site with 28 nM affinity (blue) as reported for SL3 RNA (140), and the best fit to this simulation of a two site model (red), giving a K_d of 200 nM for each site. B) Two site simulation with K_d = 200 nM including 5% Gaussian noise (blue), and the best fit to this simulation of a single site model, giving a K_d of 25 nM, demonstrating that binding stoichiometry cannot be determined from a single titration. C) Single-site 28 nM binding is simulated and a non-specific model with a site size of 6 bases is fitted to the simulation with a resulting K_d of 1.43 μM, showing that weak non-specific binding cannot be discriminated from strong specific binding from a single titration only.

The binding of NC to various DNA molecules including SL3 mutants has also been studied (153), using fluorescence quenching and time-resolved fluorescence. The SL3 variant used was the DNA analogue of SL3 extended by three base-pairs, and fluorescence titrations were performed in 50 mM Hepes pH 7.5 and 100 mM NaCl at 20 °C, using a constant NC concentration of 0.8 μM. Two NC variants were used, one containing residues 1-72 (met-247-317 extending seventeen residues beyond the actual C-terminus of NC using CA numbering) and the other containing residues 12-53 (257-298 in CA numbering). In both cases the authors report a 1:1 stoichiometry, and give K_d values of 0.071 μM for NC(1-72) and 1 μM for NC(12-53). Again, simulations show that the stoichiometry is poorly determined by this means, as shown in Figure 1.9, and a 2:1 model would give a K_d of 0.35 μM for NC(1-72) and 2.44 μM for NC(12-53). In the case of NC(12-53), the 1:1 and 2:1 models give virtually identical binding isotherms in the range of ligand concentrations used by the authors, such that discriminating between binding models would be impossible. It is similarly difficult to distinguish between specific and non-specific binding. If it is assumed that NC occludes six bases when binding to an oligonucleotide, then, in the case of NC(12-53), with an apparent 1:1 K_d of 1 μM for the 20-base SL3 DNA analogue, the non-specific binding model would give a K_d of 17 μM. In this case, specific and non-specific binding isotherms would be virtually identical, as shown in Figure 1.9. In the case of NC(1-72), with an apparent 1:1 K_d of 0.071 μM, the non-specific equivalent is ~3.1 μM, which is in very close agreement with the non-specific K_d derived in the current study to describe

binding to dSL3 of 3.2 µM for C-CA-NC. Hence, it is possible that NC indeed recognises DNA non-specifically, and the (apparent) tight binding to long oligonucleotides is just an artefact from the assumption of too few binding sites.

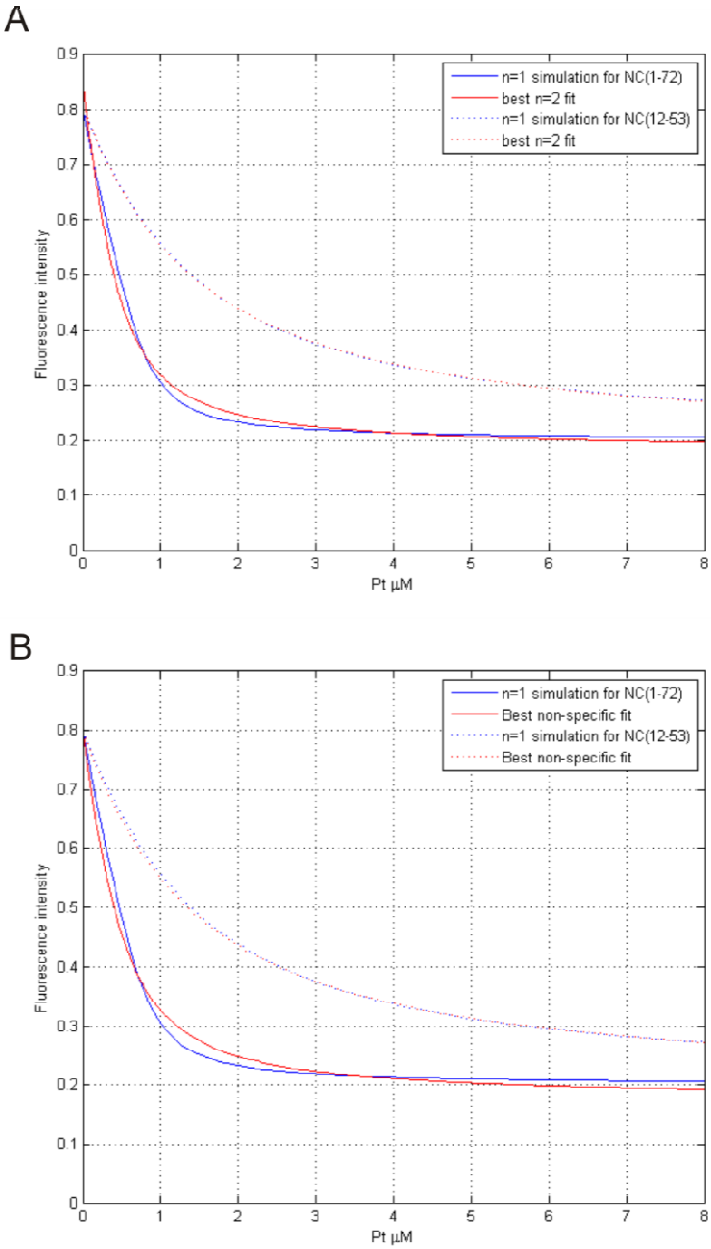


Figure 1.9: Simulated fluorescence titrations for SL3 DNA and two NC variants, as reported in (153).

A) Two-site binding is fit to one-site binding simulated with the reported single-site affinities. B) Non-specific binding is fit to the same one-site simulations. In all cases the number of binding sites, n, is not well determined, and reported stoichiometries should be treated with caution. It is equally apparent that non-specific binding models can also be difficult to distinguish from specific binding models.

1.8. The role of the SP1 domain

Residues 232-245 of Gag define the SP1 linker peptide that separates CA from NC. These 14 residues, which are part of a flexible linker spanning residues 220-255 in C-CA-NC and C-CA_{AA}-NC, are excised by proteolysis during maturation, and several studies indicate an important role for the SP1 domain both *in vitro* and *in vivo*. Similar sequences also occur in other retroviruses (37, 61). Assembly of VLPs by transfection of a proviral DNA encoding CA-NC with SP1 deleted is reduced, resulting in an aberrant heterogeneous phenotype and limited particle release from cells (48, 83). Using *in vitro* assembly of a Gag mutant ΔMACANC(ΔSP1), in which residues 16-99 of MA are deleted along with the 14 residues of SP1, limited assembly of heterogeneous cone-like structures was observed, rather than the 100 nm spheres yielded by the mutants ΔMACANC and ΔMACANCSP2 (60). The same study subjected the *E.Coli* in which such proteins were expressed to thin section EM, revealing that ΔMACANC forms 100 nm spheres within such bacteria, whereas ΔMACANC(ΔSP1) forms only tubes, indicating a possible role of SP1 in assembly phenotype selection.

Computational modelling has suggested a preference for an α-helical conformation in SP1 (2), consistent with NMR data that provide evidence for a transient helix in equilibrium with random coil conformations in SP1 (114). Our own NMR data are consistent with the latter study. This helix appears to be necessary for assembly insomuch as helix-breaking mutations of a glutamate residue to glycine or proline in the SP1 domain have an inhibitory effect on particle formation (93). A helical model for SP1, and the flanking residues, has been proposed, but TFE was included in the media, essentially forcing a helical conformation (108). Recent cET data provide a possible explanation for its role (159). As has been alluded to already (in section 1.3.3), the CA component of Gag arranges with 'tight' hexameric periodicity. Hexameric density was also observed where the SP1 domain should be located, with the 'asymmetric unit' of such hexamers being of such dimensions as to accommodate a helix. Such groups of 6 helices are located 'below' the C-CA layer of density, and a low resolution model was proposed, shown in Figure 1.4. It was suggested that the hexameric helical packing of SP1 stabilises the hexamer-forming C-CA interactions which would otherwise be too weak to enable immature assembly. This arrangement, however, is far from unambiguous and awaits a rigorous demonstration.

1.9. Molecules which inhibit HIV-1 capsid assembly

Inhibition of capsid assembly at either the immature or mature stage is an attractive means of treating HIV-1 infection. Inhibition at the immature stage would prevent release of particles, whilst inhibition at the mature stage (but not the preceding immature stage) would result in non-infectious particles. Molecules have been discovered which bind different regions of CA.

The first of these (146, 148) was discovered by phage display screening and is a 12-residue peptide with a binding site between residues 169-191 in C-CA, as identified by NMR spectroscopy using C-CA_{AA} (i.e. with W184A/M185A mutations) and corroborated by X-ray structural analysis using wild-type C-CA. It has been called Capsid Assembly Inhibitor (CAI), and the structure of the C-CA/CAI complex is shown in Figure 1.10**.** Using ΔMACANCSP2 and CANC as models for immature and mature assembly respectively, it was found that a greater concentration of CAI was required to inhibit immature assembly than mature assembly. It is of note that the ΔMACANCSP2 and CANC were 25 μM and 15 μM respectively, and that assembly assays were performed using a dialysis protocol. Analysis of NMR relaxation data using C-CA and CAI showed that dimerisation is not abrogated by CAI binding. An NMR titration revealed the K_d of C-CA_{AA} for CAI to be 15 ± 7 μM. The mechanism whereby CAI inhibits immature and mature assembly is likely to involve blocking the N-CA/C-CA intermolecular interactions, a hypothesis for which good evidence exists in the case of the mature state (123). However, its binding may also result in a C-CA dimer conformation which is incompatible with either the immature or mature lattice. Recent electron microscopy analysis of immature HIV-1 showed that the CAI-bound C-CA dimer was a good fit to the low resolution (17 Å) cET electron density map for ΔMACANCSP2 spheres, such that the CAI-bound dimer may closely resemble the immature C-CA dimer, thus potentially explaining the higher concentration of CAI required to inhibit immature assembly.

CAI cannot penetrate the cell membrane and as such is ineffective as an inhibitor physiologically. Recently, however, the C-CA/CAI structure was used as the basis for a rational drug design experiment leading to the development of a ‘hydrocarbon stapled’ peptide with a more stable α-helical conformation which can penetrate cell membranes (164). This has been called NYAD-1, and uses a hydrocarbon linker between residues of CAI. It has poor solubility, but is nevertheless effective at inhibiting immature and mature

assembly in cell culture assays, in which it has been shown to co-localise with Gag. A similar molecule with improved solubility has been developed called NYAD-13, and solution NMR has been employed to determine the structure of the C-CA_{AA}/NYAD-13 complex (21), demonstrating it to be very similar to the C-CA/CAI structure determined by X-ray crystallography. The affinity of C-CA_{AA} for NYAD-13 was estimated to be 1.2 ± 0.6 µM from an NMR titration.

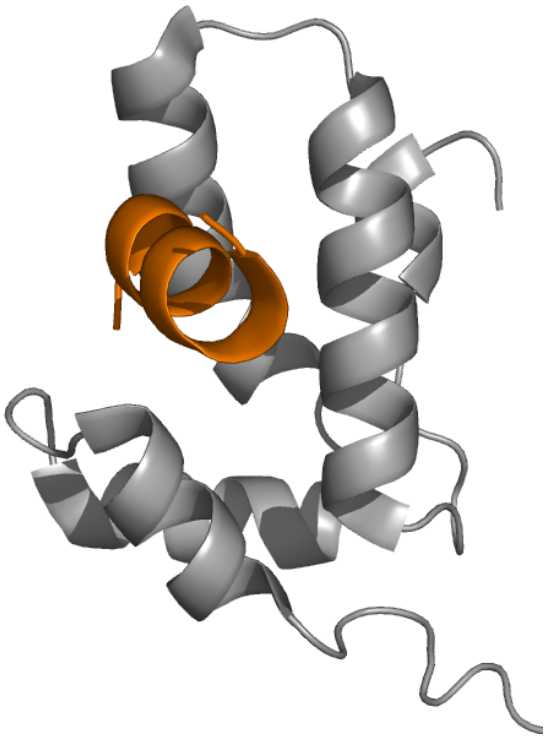


Figure 1.10: The structure of capsid assembly inhibitor (CAI) in complex with C-CA.

A single monomer is shown, but the complex crystallises as a C-CA_2/CAI_2 dimer. C-CA is shown in silver, and CAI in orange.

Assembly can also be inhibited by interference with the N-CA domain, and a molecule termed CAP-1 has been discovered which does so, despite binding only weakly with a K_d of ~800 µM. Its binding site has been identified by NMR and is located at the 'tip' of the arrowhead-like structure of N-CA, where several helices form an apex, as shown in Figure 1.11. Binding was found to be fast on the NMR timescale, and was the same in both affinity and region whether N-CA or full-length CA was used. CAP-1 could inhibit mature assembly *in vitro* using CA as a model, but immature assembly was not examined. It could also penetrate cell membranes and as such was used for *in vivo* cell culture assays, which revealed that although particles were released from infected cells, none had the mature

conical capsid and particles were heterogeneous in size and shape. CAP-1 therefore interferes with, but does not abolish, immature assembly, which is the determinant of particle size upon cellular release. It was also observed by Western blotting that intracellular Gag levels were reduced, presumably meaning that proteolysis of Gag was more rapid in the presence of CAP-1. The structure of the N-CA/CAP-1 complex has been solved using a combination of X-ray and solution NMR methods; there was poor electron density in the known CAP-1 binding region, but NOEs could be observed and so structure factors and NOEs were used as complementary restraints in structure determination. The overall structure of N-CA is affected very little, but the side-chains of H62, Q63 and Y145 are oriented so as to disrupt the N-CA/C-CA intermolecular interface determined for the mature state.

Assembly can also be inhibited by a small molecule called DSB which binds at the CA/SP1 boundary and inhibits the final reaction of the viral protease PR, namely the removal of SP1 from CA. As discussed, this is inhibitory to mature capsid assembly, and thus to infectivity. DSB is one of several betulinic acid derivatives (BA; a naturally occurring molecule in some plants) with similar activity, but has been the focus of research in preference to related molecules since it has only modest cytotoxicity in cell culture and binds with high affinity (4, 165).

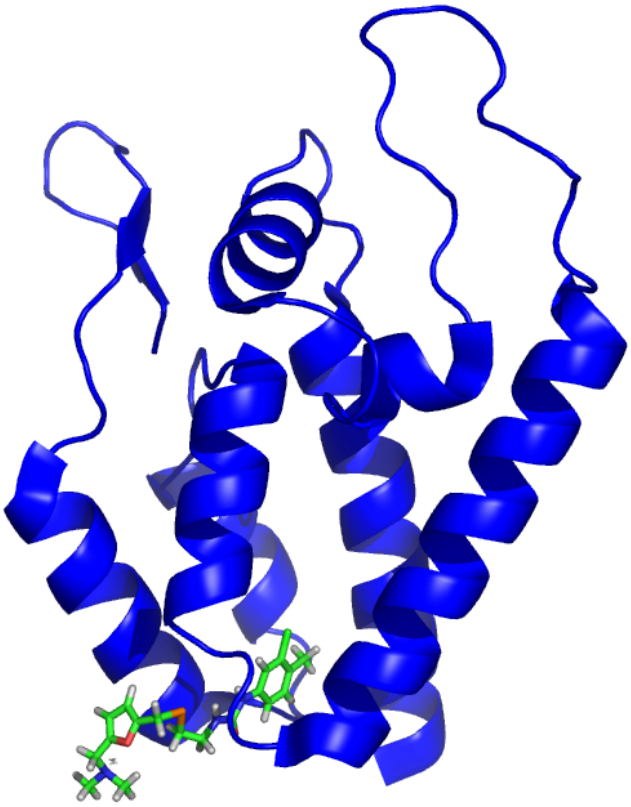


Figure 1.11: The structure of CAP-1 (as a stick diagram) and N-CA.

CAP-1 binds at the apex of the arrow-head structure of N-CA. This structure was solved using a combination of X-ray structure factors and NOEs measured by solution NMR.

1.10 Aims and objectives

The mechanism whereby CA is able to form distinct lattices in the immature and mature states is unknown, though CA differs in the two states in the sense that the SP1 and NC domains are attached in the immature state. The first objective of this thesis was therefore to determine whether the SP1 and NC domains had any interactions with the C-CA domain, in order that one can assess whether such interactions are responsible for directing CA lattice formation in the immature state. It was concluded that there are no interactions between SP1 or NC with C-CA.

The capsid assembly inhibitor peptide, CAI, appears to inhibit immature and mature assembly to different extents. Therefore, another aim of this work was to determine whether the SP1 or NC domains in some way affect its interaction with the C-CA domain, with the conclusion that they do not. This work is described in chapter 4.

Another objective was the determination of what role NC-nucleic acid interactions play in immature assembly, since immature assembly does not occur unless nucleic acid is present. A model of immature assembly is proposed based on the results. This work is described in chapter 5.

It is hoped that the experiments described in this thesis will contribute to a better understanding of immature HIV-1 capsid assembly, in order that inhibitors of assembly can be developed, which may lead to novel therapeutic compounds.

Chapter 2 NMR spectroscopy

2.1. Fundamentals

Nuclei with odd mass numbers, such as ^{1}H, ^{13}C, ^{15}N and many others possess an intrinsic angular momentum term spin and, as a result, have associated with them a magnetic moment when placed in an external magnetic field. Spin obeys the same laws as the more intuitive orbital angular momentum, for example the movement of an electron about a nucleus. As such it is spatially quantised, adopting a discrete set of values of orientations, by which is meant the projection of the spin angular momentum vector onto an arbitrary axis, taken as z by convention. Spin quantisation is described by two quantum numbers, which are described in greater detail in a later section. The spin angular momentum quantum number, *I*, is related to the magnitude of the energy of the spin. It may be zero, integral or half-integral and is a fixed quantity for a particular nucleus. The second quantum number is the magnetic quantum number, *m*, which is related to the energy of each spin state. It can take values

$$m = -I, -I+1, ..., I \qquad 2.1$$

Thus there are $2I+1$ spin states for a nucleus with a particular value of *I*. The current discussion will be limited to nuclei with $I=1/2$, of which ^{1}H, ^{13}C and ^{15}N are examples. Such nuclei therefore have two spin states; $m=+1/2$ and $m=-1/2$

Nuclear magnetic resonance (NMR) spectroscopy pertains to the measurement of the interactions of nuclear spins with one another, and with an external magnetic field. Because such interactions are highly sensitive to the chemical environment about a nucleus, NMR is of great utility in the determination of molecular structure, dynamics and interactions. Relaxation pertains to the processes whereby the system under study returns to equilibrium with its surroundings, and is also highly informative in that it depends upon the dynamics of the system. In following sections the spin interactions which determine the NMR signal, and hence the nature of the information accessible, are described and explained. More thorough discussions are to be found in the literature (1, 30, 92). Such interactions are described accurately only by the use of quantum mechanics, an exposition of which is provided in the latter references. Reviews and further information on protein dynamics are to be found in the following references (77, 78, 95, 96).

2.2. Theoretical description

A theoretical description of an NMR experiment is afforded by first finding the Hamiltonian for the system, which is the quantum mechanical operator for the energy. This will determine the resonance frequencies and lineshapes, and hence the spectrum. To extract such information a second operator, the density operator, must be determined, which contains the statistical information to convert the microscopic Hamiltonian into the macroscopic quantities to be observed. By finding the time-evolution of the density operator we can obtain all the information that quantum mechanics permits one to know about the system (subject to the constraints of the uncertainty principle). Such information constitutes a description of the experiment and may be obtained by solving the Liouville-von Neumann equation, given by

$$\frac{d\hat{\sigma}(t)}{dt} = -i\left[\hat{H}, \hat{\sigma}(t)\right] \qquad 2.2$$

where $\sigma(t)$ is the density operator and H the Hamiltonian for the system. Units of $\hbar$ are assumed. If the Hamiltonian is time-independent or periodic, as may be achieved by an appropriate transformation on the density operator, the solution to the Liouville-von Neumann equation is

$$\hat{\sigma}(t) = \exp\left(-i\hat{H}t\right)\hat{\sigma}(0)\exp\left(i\hat{H}t\right) \qquad 2.3$$

as can easily be verified by differentiation. To make the latter equation useful we must of course be able to express the density operator and Hamiltonian in an appropriate basis. The Cartesian nuclear spin angular momentum operators constitute such a basis, which afford an intuitive picture of the experiment. In following sections the theoretical tools of NMR are reviewed, along with the different spin interactions and the Hamiltonians used to represent them.

2.2.1. Angular momentum

2.2.1.2. Commutators

The commutator relations for angular momentum operators are extremely useful, and the entire angular momentum theory presented here can be derived from them. By defining angular momentum as the generator of rotations we obtain, as a fundamental relation,

$$\left[\hat{I}_i, \hat{I}_j\right] = ie_{ijk}\hat{I}_k \qquad 2.4$$

with e_{ijk} the permutation tensor and $(i,j)=(x,y,z)$. Explicitly this implies

$$\begin{aligned}\left[\hat{I}_x, \hat{I}_y\right] &= i\hat{I}_z \\ \left[\hat{I}_z, \hat{I}_x\right] &= i\hat{I}_y \\ \left[\hat{I}_y, \hat{I}_z\right] &= i\hat{I}_x\end{aligned} \qquad 2.5$$

Thus no two angular momentum operators commute with one another in the Cartesian basis.

2.2.1.1. Cartesian spin operators and the Zeeman Eigenbasis

By convention, the z-axis of the laboratory frame of reference is parallel to the applied magnetic field, which we call B_0. The x and y axes are transverse to this, and stationary in the laboratory frame. The Cartesian spin operators are then given the symbols $\left\{\hat{I}_x, \hat{I}_y, \hat{I}_z\right\}$ for the projections of the spin vector onto the different axes of the laboratory frame. Hence they are the components of a vector operator for spin angular momentum. Insofar as we are restricted by the uncertainty principle to simultaneously determine only two Cartesian components of the angular momentum of a particular entity, we cannot choose basis kets that are eigenfunctions of all three angular momentum operators. Instead, we represent the wavefunction for a single-spin system using a basis of kets that are simultaneous eigenfunctions of the angular momentum about the z-axis, and of the square magnitude of the total angular momentum. Such a basis is called the Zeeman eigenbasis. This is possible since the latter quantity commutes with I_z. We label such kets with two quantum numbers; m for the eigenvalue of I_z and I as the upper limit of m. We then have, by definition,

$$\hat{I}_z\left|I,m\right\rangle = m\hbar\left|I,m\right\rangle \qquad 2.6$$

The second quantum number, I, is put to use through the relation

$$\begin{aligned}\hat{I}^2\left|I,m\right\rangle &= \left(\hat{I}_x^2 + \hat{I}_y^2 + \hat{I}_z^2\right)\left|I,m\right\rangle \\ &= I\left(I+1\right)\hbar^2\left|I,m\right\rangle\end{aligned} \qquad 2.7$$

Generally, we use units of $\hbar$ and re-write these two formulae as

$$\hat{I}_z \left| I,m \right\rangle = m \left| I,m \right\rangle$$
$$\hat{I}^2 \left| I,m \right\rangle = I\left(I+1\right) \left| I,m \right\rangle \qquad 2.8$$

with $m=I,-I+1,\ldots,+I$, which is possible only if I is either integral or half-integral. Hence we see that angular momentum (by implication, spin) is spatially quantised insomuch as the z-component takes a discrete spectrum of values m, and has a discrete spectrum of magnitudes, $\sqrt{I\left(I+1\right)}$, with I also setting an upper and lower limit as to the angular momentum a particular entity can possess.

It is useful to define an alternative set of operators for the transverse magnetisation, namely the shift operators, defined as

$$\hat{I}_+ = \hat{I}_x + i\hat{I}_y$$
$$\hat{I}_- = \hat{I}_x - i\hat{I}_y \qquad 2.9$$

where the + subscript is for the raising operator and the – subscript is for the lowering operator. These names derive from their effects on Zeeman eigenkets; the raising operator increases the magnetic quantum number m by one unit, and vice versa for the lowering operator. These properties can make the properties of the total operators for certain spin interactions more transparent. The transverse operators be expressed in terms of the shift basis as

$$\hat{I}_x = \frac{1}{2}\left(\hat{I}_+ + \hat{I}_-\right)$$
$$\hat{I}_y = \frac{1}{2i}\left(\hat{I}_+ - \hat{I}_-\right) \qquad 2.10$$

2.2.1.3. Matrix representations

It is convenient to represent the wavefunction for a spin system by a vector, and the operators by matrices. This is simple for a single spin-1/2 nucleus and, using the single spin representations, the wavefunctions and operators for coupled systems can be determined. For a single uncoupled spin-1/2 nucleus we make the following identifications:

$$\left|\tfrac{1}{2},\tfrac{1}{2}\right\rangle = \left|\alpha\right\rangle$$
$$\left|\tfrac{1}{2},-\tfrac{1}{2}\right\rangle = \left|\beta\right\rangle \qquad 2.11$$

The total state ket is then

$$|\psi\rangle = c_\alpha |\alpha\rangle + c_\beta |\beta\rangle \qquad 2.12$$

We represent this as a vector whose elements are the coefficients multiplying the state kets. This is written

$$\begin{aligned} |\psi\rangle &= c_\alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_\beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix} \end{aligned} \qquad 2.13$$

In such a system all operators can then be represented by 2x2 square matrices with elements in general given by

$$Q_{ij} = \langle i | \hat{Q} | j \rangle \qquad 2.14$$

The array (matrix) of real or complex scalar elements Q_{ij} is then the representation of the operator $\hat{Q}$. The matrix representations for the uncoupled spin-1/2 operators are then

$$\hat{I}_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\hat{I}_+ = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \hat{I}_- = \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \qquad 2.15$$

$$\hat{I}_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \hat{I}_y = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

It is left to the next section to show how these operators appear in the Hamiltonians relevant to NMR experiments.

2.2.2. The density operator

2.2.2.1 Definitions and the equilibrium density operator

In quantum mechanics a pure ensemble is a system in which all members are described by the same state vector. Where this is not met, one has a mixed ensemble, which is the case in NMR. That is, each member of the ensemble of spins in an NMR sample is described by its own state vector. The measured value of an observable is therefore not the expectation value of its corresponding operator, but the fractional population-weighted sum of all

expectation values over the ensemble. This quantity is the ensemble average. The density operator formalism enables calculations to be performed involving such mixed states. The density operator is defined as

$$\hat{\sigma} = \overline{|\psi\rangle\langle\psi|} \tag{2.16}$$

where the overbar represents the ensemble average. Thus the density operator is the ensemble average of the outer products of the wavefunctions for all the spins in the sample. It is used to extract the measurable magnetisation by first expanding the wavefunction to give

$$\begin{aligned} \hat{\sigma} &= \sum_{ij} \overline{c_i c_j^* |i\rangle\langle j|} \\ &= \sum_{ij} \overline{c_i c_j^*} |i\rangle\langle j| \end{aligned} \tag{2.17}$$

where the second line is justified as the basis kets are the same for each spin, with only the expansion coefficients differing for different members of the ensemble. The matrix elements are then given by

$$\begin{aligned} \langle m|\hat{\sigma}|n\rangle &= \sum_{ij} \overline{c_i c_j^*} \langle m|i\rangle\langle j|n\rangle \\ &= \sum_{ij} \overline{c_i c_j^*} \delta_{mi}\delta_{jn} \\ &= \overline{c_m c_n^*} \end{aligned} \tag{2.18}$$

To extract the ensemble average of the expectation value of an operator Q (implied by an overbar) we then evaluate

$$\begin{aligned} \overline{\langle\hat{Q}\rangle} &= \overline{\langle\psi|\hat{Q}|\psi\rangle} \\ &= \sum_{mn} \overline{c_n^* c_m} \langle n|\hat{Q}|m\rangle \\ &= \sum_{mn} \langle m|\hat{\sigma}|n\rangle\langle n|\hat{Q}|m\rangle \\ &= \sum_{m} \langle m|\hat{\sigma}\hat{Q}|m\rangle \\ &= Tr\{\hat{\sigma}\hat{Q}\} \end{aligned} \tag{2.19}$$

Thus the density operator has removed the need for explicit statistical consideration of the operator Q.

By consideration of the first derivative of the density operator, the Liouville-von Neumann equation 2.2 is determined. The equilibrium density operator is required to solve equation 2.2, which for a single spin is given by

$$\hat{\sigma}_{eq} \propto \hat{I}_z \qquad 2.20$$

Since the absolute value of the magnetisation from a sample is not measured, the equation is generally scaled to give a proportionality constant of 1, such that the LHS and RHS are equal. For a multispin system the equilibrium density operator is the sum of *z*-angular momentum operators in a product space.

2.2.2.2. Expansion of the density operator in an operator space

Any square matrix can be represented as a linear combination of a complete set of mutually orthonormal basis matrices. It transpires that the Cartesian angular momentum operators along with the unit matrix can serve this purpose for a single spin (although other bases can be used) where the density matrix has four elements. In the case of a coupled spin system (*I* and *S* spins for example) a complete basis is obtained by taking all direct products between basis operators in the *I* subspace and those of the *S* subspace. A two-spin system thus has sixteen basis operators and sixteen elements in its density matrix. Such operator expansions are exploited in the product operator formalism, where the simple transformation properties of the basis operators make the results of many NMR experiments transparent.

2.2.3. The rotating frame

It is often the case that Hamiltonians are expressed in a frame of reference rotating about the *z*-axis of the stationary laboratory frame. The reason for this is that the time dependence of the Hamiltonian can often be removed by such a transformation, which leads to a considerable simplification. The rotating frame transformation is achieved by rotating the coordinate system about the *z*-axis with frequency ω_{rf}. Since the rotating frame is more useful for describing NMR experiments, the form of the Liouville-von Neumann equation in the rotating frame is required. To do so, we find the unitary transformation which removes the time-dependence from the Hamiltonian and apply the same transformation to the density operator. The relevant unitary operator is

$$U = \hat{R}_z\left(\omega_{rf} t\right) = \exp\left(-i\omega_{rf} t \hat{I}_z\right) \qquad 2.21$$

The transformed density operator, $\hat{\sigma}'$, then evolves according to a transformed Liouville-von Neumann equation, which now reads

$$\frac{d\hat{\sigma}'}{dt} = -i\left[U\hat{H}U^{-1} - \omega_{rf}\hat{I}_z, \hat{\sigma}'\right] \qquad 2.22$$

2.2.4. The nuclear spin Hamiltonian

The nuclear spin Hamiltonian is the operator representing all of the nuclear spin interactions present, and is partitioned into external and internal interaction terms. By the former is meant the interaction of the spin system with the apparatus, and comprises a term for the static applied field $\mathbf{B_0}$ and for the oscillating RF field $\mathbf{B_1}$. The internal terms pertain to the interactions between spins (both nuclear and electronic). Such interactions are the chemical shift (or chemical shielding), the dipolar coupling, the scalar coupling and, for spins >1/2, the quadrupolar interaction. The last of these is not considered here.

In general, the external terms are of the form

$$\hat{H}_{ext} = -\boldsymbol{\mu}.\mathbf{B} \qquad 2.23$$

in which $\boldsymbol{\mu}$ is a vector operator representing the nuclear spin magnetic moment i.e.

$$\boldsymbol{\mu} = \gamma \begin{pmatrix} \hat{I}_x \\ \hat{I}_y \\ \hat{I}_z \end{pmatrix} \qquad 2.24$$

and $\mathbf{B}$ is the magnetic field vector. The internal terms are of the general form

$$\hat{H}_{int} = -\boldsymbol{\mu}.(\mathbf{C}\mathbf{v}) \qquad 2.25$$

where $\boldsymbol{\mu}$ has the same meaning as before, $\mathbf{C}$ is a second-rank Cartesian tensor (represented as a 3x3 matrix) and $\mathbf{v}$ is either a magnetic field vector or vector operator for a second magnetic moment, depending on the interaction. In constructing matrix representations, it is understood that operators for individual spins are in fact part of a product space.

2.2.4.1. Interaction with the applied magnetic field

The applied magnetic field is taken to be along z with magnitude B_0, such that the Hamiltonian for the interaction between a nuclear spin and the applied field is

$$\begin{aligned}\hat{H}_0 &= -\gamma B_0 \hat{I}_z \\ &= \omega_0 \hat{I}_z\end{aligned} \qquad 2.26$$

This is termed the Zeeman Hamiltonian. This shows that the energy of an uncoupled spin is minimised when it is aligned in the opposite direction to the applied field, and maximised when aligned in the same direction. Note that equation 2.26 is expressed in a stationary laboratory frame of reference. Whilst its form is obviously invariant under a rotation about the *z*-axis, the Zeeman term is removed from the active Hamiltonian, i.e. the form of the Hamiltonian which appears in the rotating-frame Liouville-von Neumann equation. This is clear by inspection of equation 2.2.

2.2.4.2. Interaction with a oscillating radiofrequency field

A linearly polarised RF field of magnitude B_1 oscillating with frequency ω_{rf} and with phase ϕ is described by the vector

$$\mathbf{B_1} = B_1 \begin{pmatrix} \cos\left(\omega_{rf} t + \phi\right) \\ 0 \\ 0 \end{pmatrix} \qquad 2.27$$

This can be decomposed into two counter-rotating fields, one of which is on-resonance with the spin (rotates in the same direction), and one off-resonance (in the opposite sense). Equation 2.23 can then be written

$$\begin{aligned}\mathbf{B_1} &= \frac{1}{2}\mathbf{B_1^{on}} + \frac{1}{2}\mathbf{B_1^{off}} \\ &= \frac{1}{2} B_1 \begin{pmatrix} \cos\left(\omega_{rf} t + \phi\right) \\ \sin\left(\omega_{rf} t + \phi\right) \\ 0 \end{pmatrix} + \frac{1}{2} B_1 \begin{pmatrix} \cos\left(\omega_{rf} t + \phi\right) \\ -\sin\left(\omega_{rf} t + \phi\right) \\ 0 \end{pmatrix}\end{aligned} \qquad 2.28$$

Only the resonant component, rotating with the spins, has a significant effect. The other can be ignored for most purposes. Utilising this fact we then have

$$\hat{H}_{rf}^{lab}(t) = \omega_1 \left\{ \cos\left(\omega_{rf} t + \phi\right) \hat{I}_x + \sin\left(\omega_{rf} t + \phi\right) \hat{I}_y \right\} \qquad 2.29$$

in which

$$\omega_1 = -\frac{1}{2}\gamma B_1 \qquad 2.30$$

whose modulus is called the nutation frequency. Like 2.26, 2.29 is expressed in the laboratory frame. It is more common to work in a frame rotating at frequency ω_{rf}, wherein the RF Hamiltonian is given by, with phase *x* for example,

$$\hat{H}_{rf} = \omega_1 \hat{I}_x \qquad 2.31$$

Hence in the rotating frame, during an RF pulse magnetisation precesses about the pulse axis. The angle through which it is rotated is then $\omega_1 t$, called the flip angle. RF pulses can thus be used to rotate the equilibrium *z*-magnetisation in to the transverse plane where it becomes observable, whose utility is self-evident.

2.2.4.3. Chemical shift

Chemical shift describes the perturbation upon the magnetic field experienced by a nuclear spin as a result of the electrons surrounding such a nucleus. It is important in that it results in a unique Larmor frequency for each spin, even those belonging to the same nuclear species. The magnetic field vector $\mathbf{B}_{local}$ at a nucleus experiencing a field induce by its electronic environment $\mathbf{B}_{induced}$ is given by

$$\mathbf{B}_{local} = \mathbf{B}_0 + \mathbf{B}_{induced} \qquad 2.32$$

The induced field is given by

$$\mathbf{B}_{induced} = \boldsymbol{\delta}\mathbf{B}_0 \qquad 2.33$$

where $\boldsymbol{\delta}$ is the chemical shift tensor. The secular chemical shift Hamiltonian is then

$$\hat{H}_{CS} = \omega_0 \delta_{zz} \hat{I}_z$$

All elements of the chemical shift tensor are functions of orientation, i.e. chemical shift is an anisotropic interaction. This is important in solid-state NMR, but under solution conditions only the isotropic chemical shift is measured due to motional averaging. The chemical shift Hamiltonian in solution is then

$$\hat{H}_{CS} = \omega_0 \delta_{iso} \hat{I}_z \qquad 2.34$$

in which

$$\delta_{iso} = \frac{1}{3} Tr\{\boldsymbol{\delta}\} \qquad 2.35$$

It is common to report a chemically shifted Zeeman Hamiltonian, rather than consider the Zeeman and chemical shift terms separately. The Larmor frequency is then redefined as

$$\omega_0 = -\gamma B_0 \left(1 + \delta_{iso}\right) \qquad 2.36$$

In the rotating frame, the chemically shifted Zeeman term is then

$$\hat{H}_Z = \Omega \hat{I}_z \qquad 2.37$$

in which $\Omega = \omega_0 - \omega_{rf}$ and is called the offset frequency. Clearly the chemical shift is only significant in the rotating frame.

2.2.4.4. Scalar coupling

Scalar coupling (also called *J*-coupling) is an indirect interaction between spins connected by one or more covalent bonds, transmitted *via* the electrons in the bond(s) connecting them. Following the general Hamiltonian given by equation 2.25 and ascribing the symbol **J** to a Cartesian tensor for scalar coupling,

$$\hat{H}_J^{12} = \hat{\mathbf{I}}_\mathbf{1}^\mathrm{T} \mathbf{J} \hat{\mathbf{I}}_\mathbf{2} \qquad 2.38$$

where $\mathbf{I_1}$ is a vector operator for spin 1 (whose transpose is used) and $\mathbf{I_2}$ a vector operator from spin 2, which can be either the same or different types of nuclei. The total *J*-coupling Hamiltonian extends over all spin pairs, so is given by

$$\hat{H}_J^{ij} = \sum_{i>j} \hat{\mathbf{I}}_i^\mathrm{T} \mathbf{J} \hat{\mathbf{I}}_j \qquad 2.39$$

The restricted sum is to ensure that each spin pair is counted only once. In this form, we have allowed for the possibility of anisotropy. In fact, this anisotropy is very small and generally neglected. We then use a scalar coupling constant ${}^k J_{ij}$ with *k* the number of bonds through which the coupling is transmitted and *ij* the spin pair. Such constants are generally given in Hz, so multiplication by 2π is necessary to be rad s^{-1}, consistent with convention. Therefore, in the laboratory frame,

$$\begin{aligned} \hat{H}_J &= \sum_{i>j} 2\pi J_{ij} \hat{\mathbf{I}}_i . \hat{\mathbf{I}}_j \\ &= \sum_{i>j} 2\pi J_{ij} \left(\hat{I}_{ix} \hat{I}_{jx} + \hat{I}_{iy} \hat{I}_{jy} + \hat{I}_{iz} \hat{I}_{jz} \right) \\ &= \sum_{i>j} 2\pi J_{ij} \left(\frac{1}{2} \left(\hat{I}_{i+} \hat{I}_{j-} + \hat{I}_{i-} \hat{I}_{j+} \right) + \hat{I}_{iz} \hat{I}_{jz} \right) \end{aligned} \qquad 2.40$$

This can be simplified by the secular approximation, whereby terms are retained only if they commute with the Zeeman Hamiltonian and their time-dependence is slow. Thus the

z-terms are automatically retained. If the separation between Zeeman interaction frequencies for two spins is large compared to the scalar coupling constant (as occurs in the heteronuclear case), we can, under the secular approximation, neglect the transverse terms. This is the weak coupling condition, which in general is a good approximation in the heteronulcear case, for which we use *I* and *S* notation and write

$$H_J^{IS} = 2\pi J_{IS} \hat{I}_z \hat{S}_z \qquad 2.41$$

Clearly, it is invariant under the rotating frame transformation. For certain homonuclear experiments (TOCSY for example), the full strong coupling Hamiltonian of equation (3.40) must be used and appropriately transformed. In writing the operator products it is understood that the *I* and *S* operators each occupy their own subspaces of a total space of dimensionality equal to the sum of dimensionalities of the subspaces.

The effect of scalar coupling, either in solution or solids, is to split peaks in two, since for each spin in a spin pair, its bonded neighbour can be in either of its spin states. Hence there are four transition frequencies for a spin pair. This effect can be removed by using decoupling RF pulse sequences. Scalar coupling also mediates coherence transfer between bonded spins. Hence coherence generated on, for example, a ^{1}H nucleus can be transferred to a neighbouring ^{15}N nucleus, which is of paramount importance in multidimensional correlation experiments.

2.2.4.5. Dipolar coupling

The dipolar coupling between two spins is a direct through-space interaction, arising since the magnetic dipole at site *I* gives rise to a magnetic field at site *S* and *vice versa*. That is, the magnetic dipoles mutually influence one another. The dipolar Hamiltonian can be written

$$\hat{H}_D^{12} = D^{12}\left(\hat{\mathbf{I}}_1.\hat{\mathbf{I}}_2 - 3\frac{\left(\hat{\mathbf{I}}_1.\vec{\mathbf{r}}_{12}\right)\left(\hat{\mathbf{I}}_2.\vec{\mathbf{r}}_{12}\right)}{r_{12}^2}\right) \qquad 2.42$$

in which $\mathbf{I}_1$ and $\mathbf{I}_2$ are vector operators for the two respective spins, $\vec{\mathbf{r}}_{12}$ is the radial vector connecting the spins, r_{12} is the distance between the two spins, and D_{12} is the dipolar coupling constant, generally reported in Hz, given by

$$D^{12} = -\frac{\mu_0 \hbar \gamma_1 \gamma_2}{4\pi r_{12}^3} \qquad 2.43$$

Expressing this in spherical coordinates and applying the secular approximation simplifies the dipolar Hamiltonian to

$$H_D^{12} = D^{12}\frac{1}{2}\left(1-3\cos^2\theta\right)\left(2\hat{I}_{1z}\hat{I}_{2z}-\frac{1}{2}\left(\hat{I}_{1+}\hat{I}_{2-}+\hat{I}_{1-}\hat{I}_{2+}\right)\right) \qquad 2.44$$

in the homonuclear case, which is expressed above in the laboratory frame. In the heteronuclear it is further simplified to

$$H_D^{IS} = D^{IS}\frac{1}{2}\left(1-3\cos^2\theta\right)2\hat{I}_z\hat{S}_z \qquad 2.45$$

Here, θ is the angle connecting the internuclear bond vector to the magnetic field vector. In isotropic solution the average dipolar coupling is zero, so it is not directly observed. In solids it can be very large and dipolar coupling effects can dominate the spectra of static solids. An intermediate case exists where the molecules in the sample have a slight preferential orientation, which can be achieved in a weak alignment media such as an elongated polyacrylamide gel. In such a case, we measure residual dipolar couplings (RDCs), so-called as they arise through incomplete orientational averaging. In the heteronuclear case the dipolar Hamiltonian is

$$\hat{H}_D^{IS}(t) = D_{\max}^{IS}\frac{1}{2}\left\langle 3\cos^2\theta-1\right\rangle 2\hat{I}_z\hat{S}_z \qquad 2.46$$

in which

$$D_{\max}^{IS} = -\frac{\mu_0 h\gamma_I\gamma_S}{16\pi^3\left\langle r_{IS}\right\rangle^3} \qquad 2.47$$

where angular brackets denote the ensemble average. The residual dipolar coupling is then

$$D^{IS} = D_{\max}^{IS}\left\langle 3\cos^2\theta-1\right\rangle \qquad 2.48$$

The molecular alignment tensor is then defined as

$$A_{ij} = 3\left\langle\cos\beta_i\cos\beta_j\right\rangle-\delta_{ij} \qquad 2.49$$

Where $(i,j)=(x,y,z)$ and $\beta_{i,j}$ are the angles connecting a fixed molecular reference frame to the B_0 field. Defining $\alpha_{i,j}$ as the angles relating a given bond vector to the molecular frame and assuming a rigid molecule, the dipolar coupling constant can be written

$$D^{IS} = D_{\max}^{IS}\sum_{ij}A_{ij}\cos\alpha_i\cos\alpha_j \qquad 2.50$$

The alignment tensor is symmetric and traceless, so requires the specification of 5 independent parameters. Generally we work in the principle axis frame of the alignment tensor (the alignment frame), wherein we require two parameters. The other 3 are then the Euler angles to rotate the molecular frame to the alignment frame. It is easier to describe the orientation of a bond vector in the alignment frame by its polar angles *(θ,Φ)* such that

$$D^{IS}(\theta,\phi) = D^{IS}_{\max}\left(\left(\cos^2\theta - 1\right)A_a + \frac{3}{2}A_r \sin^2\theta\cos 2\phi\right) \qquad 2.51$$

In which $A_a = \frac{3}{2}A_{zz}$ and $A_r = \left(A_{xx} - A_{yy}\right)$ where the components A_{ii} are defined such that $\left|A_{zz}\right| > \left|A_{yy}\right| > \left|A_{zz}\right|$

Like scalar coupling, dipolar coupling mediates coherence transfer. However its mechanism is through space, such that the chemical shifts of nearby unbonded, as well as bonded, spins can be correlated. An important consequence of the dipolar interaction, particularly in solution, is the nuclear Overhauser effect (NOE), which correlates nearby spins in space and provides distance restraints.

2.3. Building blocks for NMR experiments

Multidimensional NMR experiments generally consist of a preparation period during which an appropriate density operator is generated, an evolution period during which coherence between different spins is allowed to evolve, a mixing period to generate observable coherence on the nucleus to be observed and an acquisition period during which a signal is measured.

2.3.1. Polarisation transfer and coherence transfer

Often in the preparation period, polarisation is transferred from ^{1}H to either ^{15}N or ^{13}C, since the higher gyromagnetic ratio of ^{1}H enables stronger signals on heteronuclei to be achieved by its exploitation. Also, polarisation transfer is a prerequisite for coherence transfer, where chemical shifts for different nuclei are correlated. This is of paramount importance in assignment of large molecules and obtaining distance restraints for structural characterisation.

2.3.1.1. The INEPT transfer

INEPT stands for insensitive nucleus enhancement by polarisation transfer. It converts the equilibrium magnetisation on a particular *I* spin (generally ^{1}H) to antiphase coherence on the *S* spin *via* the *J*-coupling interaction. If 90° pulses are applied to the *I* and *S* spins about the *y* and *x* axes respectively, coherence transfer is achieved, resulting in antiphase coherence on the *S* spin. It is in this sense a coherence transfer sequence, and is used in both preparation periods and mixing periods for heteronuclear NMR. The pulse sequence is given in Figure 2.1, which does not include the final 90° pulses.

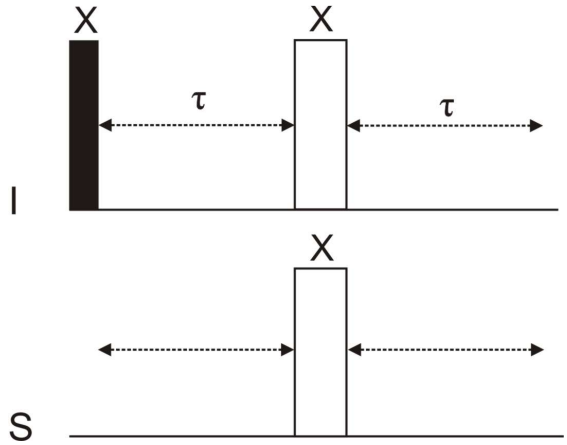


Figure 2.1: The INEPT transfer pulse sequence.

The filled rectangle represents a 90° pulse, the empty rectangles 180° pulses.

Polarisation transfer is optimised when

$$2\tau = \frac{1}{2J_{IS}} \qquad 2.52$$

For a ^{1}H to ^{15}N transfer at amide positions, typically 2τ=5.4ms since J_{NH}=93Hz. For ^{1}H to ^{13}C, 2τ=3.6ms with $^{1}J_{CH}$=140Hz.

2.3.1.2. Mixing periods

In solution, the *J*-coupling interaction can be used to correlate spins using INEPT sequences (and variants upon) for mixing periods. COSY-type mixing periods, for example, comprise an INEPT sequence element followed by 90º pulses on one or more channels. Correlations are observed between scalar-coupled spins. TOCSY-type mixing periods also use the *J*-coupling interaction but correlate all spins in a spin system even when all are not mutually coupled. NOE mixing periods can also be used, which use dipolar cross-relaxation to yield cross-peaks between spins nearby in space.

2.4. Relaxation and protein dynamics

2.4.1. Theory of relaxation

Relaxation is the process whereby the experimental system, having been perturbed during the experiment, returns to equilibrium with the surroundings. All operators have associated with them a characteristic time-modulation and the present discussion will be restricted to in-phase single-quantum coherence operators (e.g. $\hat{I}^{\pm}$) and zero-quantum *z*-operators (e.g. $\hat{I}_z$) for a system in solution. This time-dependence is often oscillatory if only the secular components of the nuclear spin interactions contributing to spin Hamiltonian are retained. The non-secular components, however, manifest themselves in facilitating exchange of energy with the lattice, establishing thermal equilibrium and abolishing coherences. The CSA interaction, for example, is predicted to give only an oscillation of transverse magnetisation operators at the chemically shifted Larmor frequency in isotropic solution if time-dependent perturbations are not accounted for. A more thorough treatment, however, reveals that the CSA interaction influences line-width even in isotropic solution, as well as contributing to the return of longitudinal magnetisation to its equilibrium value.

Relaxation is described theoretically by consideration of the Liouville-von Neumann equation (equation 2.2) in the interaction frame, whence the superscript *I* is used. In such a representation the Hamiltonian is partitioned into two components in the Schrödinger picture as

$$\hat{H}(t) = \hat{H}_0 + \hat{H}_1(t) \qquad 2.53$$

where H_0 does not explicitly depend upon time and H_1 is the time-dependent part (a stationary random operator in the treatment of relaxation). The interaction-frame time-dependent Hamiltonian is then defined as

$$\hat{H}_1^I(t) = \exp\left(i\hat{H}_0 t\right)\hat{H}_1(t)\exp\left(-i\hat{H}_0 t\right) \qquad 2.54$$

The Liouville-von Neumann equation is then

$$\frac{d\hat{\sigma}^I(t)}{dt} = -i\left[\hat{H}_1^I(t), \hat{\sigma}^I(t)\right] \qquad 2.55$$

This equation will be solved subject to a set of simplifying assumptions. Firstly, we shall assume that the density matrix changes slowly as compared to the stochastic part of the

Hamiltonian. Secondly, we will assume that the correlation time, whose meaning will be made definite later, is shorter than the timescale upon which the density operator changes.

Integrating equation 2.56 to second order and assuming symmetry between past and future gives

$$\frac{d\hat{\sigma}^I(t)}{dt} = -\int_{-\infty}^{\infty} \overline{\left[\hat{H}_1^I(t), \left[\hat{H}_1^I(t-\tau), \hat{\sigma}^I(t)\right]\right]} d\tau \qquad 2.56$$

where τ is the difference between *t* and some other point in time *t'* and the overbar denotes the ensemble average. The density operator, as detailed in the first assumption, does not change significantly between *t* and *t'*. The Hamiltonian in the above equation is then expanded as a series of time-independent operators $A^{(q)}$ weighted by random functions of time $F^{(q)}(t)$, written

$$\hat{H}_1^I(t) = \sum_q F^{(q)}(t)\,\hat{A}^{(q)} \qquad 2.57$$

The operators $A^{(q)}$ evolve according to

$$\exp\left(i\hat{H}_0 t\right)\hat{A}^{(q)}\exp\left(i\hat{H}_0 t\right) = \sum_p \exp\left(i\omega_p^{(q)} t\right)\hat{A}_p^{(q)} \qquad 2.58$$

Cross-correlation is then assumed negligible and the autocorrelation function is defined as

$$C_q(\tau) = \overline{F^{(q)}(t)\,F^{(-q)}(t-\tau)} \qquad 2.59$$

which clearly depends on the dynamics of the system and describes the deviation of a system from its initial state. We define also the spectral density function as

$$J_q(\omega) = \int_{-\infty}^{\infty} C_q(\tau)\exp(-i\omega\tau)d\tau \qquad 2.59$$

which is the Fourier transform of the correlation function and represents the density of motion of the functions $F^{(q)}$ as a spectrum of frequencies. With these definitions equation 2.57 is written

$$\frac{d\hat{\sigma}^I(t)}{dt} = -\frac{1}{2}\sum_{p,q} J_q\left(\omega_p^{(q)}\right)\left[\hat{A}_p^{(-q)}, \left[\hat{A}_p^{(-q)}, \hat{\sigma}^I(t)\right]\right] \qquad 2.60$$

This equation represents the link between the system dynamics and the time-evolution of the density operator.

2.4.2. Relation to protein dynamics

The discussion is henceforth cast in a notation appropriate for the description of a ^{15}N nucleus coupled to a ^{1}H nucleus, so that the backbone dynamics of a protein *via* the peptide N-H bond vectors can be described. In such a case the ^{15}N nuclear relaxation is dominated by the dipolar interaction with the directly bonded proton and CSA interactions. The (longitudinal) autorelaxation of the ^{15}N nucleus is described by exponential modulation with time constant T_1 and (transverse) autorelaxation by exponential modulation with time constant T_2, calculable from equation 2.62. Cross-relaxation (given by the NOE) is also measurable. These three parameters are given by

$$\frac{1}{T_1} = \frac{d^2}{4}\left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\right] + \frac{\omega_N^2 \Delta\sigma^2}{3} J(\omega_N)$$

$$\frac{1}{T_2} = \frac{d^2}{8}\left[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\right] + \frac{\omega_N^2 \Delta\sigma^2}{18}\left[4J(0) + 3J(\omega_N)\right] + R_{ex}$$

$$NOE = 1 + \frac{d^2}{4}\frac{\gamma_N}{\gamma_H} T_1 \left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right] \qquad 2.61$$

In which

$$d = \frac{\mu_0 \hbar \gamma_N \gamma_H}{4\pi \left\langle r_{NH}^3 \right\rangle}, \qquad 2.62$$

Longitudinal relaxation is thus dominated by motions at the Larmor frequency of the ^{15}N nucleus, and the sum and differences of the Larmor frequencies of itself and the directly bonded proton. Transverse relaxation is dominated by $J(0)$, meaning motions which do not vary with time, representing the distribution of local magnetic fields along the z axis. The term R_{ex} describes the conformational exchange contribution to transverse relaxation on a μs-ms timescale, and can be related to a particular model of exchange if needed.

A rapidly moving system loses correlation with its initial state rapidly, so has a narrow correlation function. This implies a broad spectral density function, such that $J(0)$ is small for a rapidly moving system since the integral of the spectral density is constant. Hence a rapidly moving system has generally a long T_2 and short T_1. For a protein, however, the overall motion is slow, such that if motion is dominated by rotational diffusion, the system will lose correlation with its initial state slowly, resulting in a narrow spectral density

function. Hence $J(0)$ is large, but the other terms contributing to relaxation are small, and lines will be broad (i.e. T_2 is short), with a relatively long T_1. If internal motions of bond vectors are present and on a faster timescale than overall motion, then we expect more motions at the Larmor frequencies of the H and ^{15}N nuclei, to which T_1 is sensitive but which has little influence on T_2. Thus T_1 reports on fast internal motions, whilst T_2 is a good reporter of overall motion. The following section details a common approach to quantifying the magnitude and timescale of internal motions of bond vectors in an otherwise rigid protein.

2.4.3. The model-free approach

The Lipar-Szabo model-free approach (95, 96) facilitates the interpretation of protein relaxation data by factorising the correlation function into an overall term for tumbling in solution and an internal term for fast internal motion of a bond vector. Defining $C_O(t)$ as the overall correlation function and $C_I(t)$ we have

$$C(t) = C_O(t)C_I(t) \qquad 2.63$$

Here it is assumed that overall motion is isotropic (the generalisation to anisotropic motion, though not rigorous, is possible to a good approximation) and that internal motion is on a faster timescale than overall motion. The overall correlation function is, for isotropic rotational diffusion, given by

$$C_O(t) = \frac{1}{5} exp\left(-\frac{t}{\tau_m}\right) \qquad 2.64$$

Which may be derived from the rotational diffusion equation. Here, τ_m is the rotational correlation time, which is on the order of nanoseconds for the proteins considered in this thesis. The internal term is a heuristic approximation required to be exact at short and long times t, given by

$$C_I(t) = S^2 + (1 - S^2) exp\left(-\frac{t}{\tau_e}\right) \qquad 2.65$$

Here, τ_e is the effective timescale of internal motion, in the ps-ns range, describing internal fluctuation of a bond vector. S^2 is the order parameter, describing the degree of spatial restriction of the motion. Its values are between zero and one, with higher values having the meaning that motion is more restricted. Note that neither τ_e nor S^2 requires the invocation of a model of internal motion. It is sometimes the case that the single exponential approximation above does not hold, for example when the internal motion is

on distinct fast and slow timescales (by slow is meant slow internal motion, but still fast as compared to overall motion) (36). Here, the order parameter can be factorised as

$$S^2 = S_s^2 S_f^2 \qquad 2.66$$

The resulting form of the spectral density function (the Fourier transform of the correlation function) is

$$J(\omega) = \frac{2}{5} S_f^2 \left[\frac{S_s^2 \tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S_s^2)\tau}{1+(\omega\tau)^2} \right] \qquad 2.67$$

in which $\tau = \tau_e \tau_m / (\tau_e + \tau_m)$.

Inasmuch as requiring the specification of, at most, three parameters, the Lipari-Szabo approach is well suited to probing protein backbone dynamics, since often we measure only three experimental parameters, namely T_1, T_2 and the heteronuclear NOE.

If there is no fast internal motion, then $S_f^2 = 1$ and $\tau = 0$, in which case $S^2 = S_s^2$. The spectral density is then a function of the slow order parameter and diffusion tensor only (or correlation time for isotropic motion). In this case, T_2 is shortest when $S^2 = 1$, and lines become narrower as S^2 decreases, meaning that the maximum possible line-width permitted in the absence of an exchange process is calculable given the diffusion tensor. The diffusion tensor is calculable given the structure of the molecule.

The above discussion holds only for isotropic rotational diffusion. If the overall motion is anisotropic, then a diffusion tensor must be determined in order to adequately describe overall motion. As an approximation, an elliptic diffusion tensor will be used for the analyses in this thesis. In this case, the spectral density function becomes

$$J(\omega) = \frac{2}{5} S_f^2 \sum_{j=1}^{3} A_j \left[\frac{S_s^2 \tau_j}{1+(\omega\tau_j)^2} + \frac{(1-S_s^2)\tau_j'}{1+(\omega\tau_j')^2} \right] \qquad 2.69$$

In which

$$A_1 = \frac{(3\cos^2\alpha - 1)^2}{4}$$

$$A_2 = 3\cos^2\alpha \sin^2\alpha$$

$$A_3 = 3\sin^4\alpha / 4$$

$$\tau_1 = 1/6D_\perp$$

$$\tau_2 = \frac{1}{5D_\perp + D_\parallel}$$

$$\tau_3 = \frac{1}{2D_\perp + 4D_\parallel}$$

$$\tau_j' = \frac{\tau_e \tau_j}{\tau_e + \tau_j} \qquad 2.70$$

Here, α is the angle between the NH bond vector and the unique axis of the diffusion tensor in its principle frame. The polar angles (θ,φ), which relate the principal frame of the diffusion tensor to the molecular frame are also determined, such that an axially symmetric diffusion tensor has 4 independent components. The meanings of the angles in the above formulae are shown in Figure 2.2.

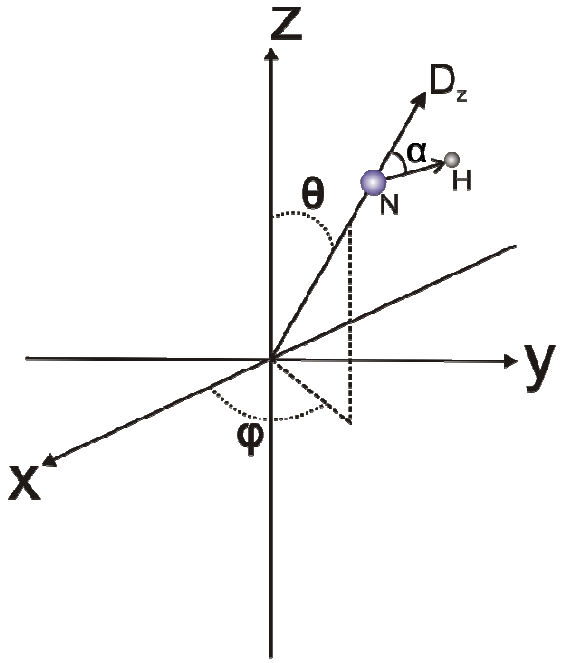


Figure 2.2: The meaning of the angles of an axially symmetric diffusion tensor.

The (X,Y.Z) axes are fixed to the molecule, and D_z is the unique axis of the diffusion tensor. An NH bond vector is shown.

To obtain a diffusion tensor from relaxation data, residues with limited internal mobility are selected, based on having high NOEs. In the limit of structural rigidity, relaxation times are functions of the diffusion tensor only, which is then calculable given the reduced set of 'rigid' relaxation times and a known structure. The isotropic correlation time is obtained by solving the equation

$$\frac{T_1}{T_2} = \frac{4J(0) + J(\omega_-) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_+) + (c^2/d^2)\{4J(0) + 3J(\omega_N)\}}{2J(\omega_-) + 6J(\omega_N) + 12J(\omega_+) + 2(c^2/d^2)J(\omega_N)} \quad 2.71$$

In which

$$\omega_+ = \omega_N + \omega_H \quad 2.72$$

$$\omega_- = \omega_N - \omega_H \quad 2.73$$

Because rigid residues only are included in diffusion tensor analysis, the spectral density function has the simplified form

$$J(\omega) = \frac{2}{5}S^2 \sum_{j=1}^{3} \frac{A_j \tau_j}{1 + \left(\omega\tau_j\right)^2} \quad 2.74$$

2.5. Chemical exchange in NMR

Chemical exchange refers to the phenomena observed in NMR when a spin is able to exchange between different chemical environments. In the simplest example, a molecule may have two distinct conformations, such that a spin may have two distinct chemical environments, between which magnetisation may be transferred by virtue of that spin exchanging between such states. Chemical exchange hence acts as a relaxation mechanism, since each chemical environment implies a distinct Larmor frequency, and as such said spin may experience either Larmor frequency, contributing to the range of spin state transition frequencies accessible. Recall that transverse relaxation is a result of a spin experiencing different Larmor frequencies. The precise lineshape observed for a spin which may undergo chemical exchange is calculable given the chemical reaction rate law, such that accurate lineshape measurements report upon the nature of exchange. Since exchange depends upon the chemical kinetics of the particular exchange reaction, such measurements make rate constants accessible. This makes events such as, for example, conformational flexibility, ligand binding, or self-interactions both qualitatively and quantitatively tractable by NMR.

2.5.1. The McConnell equations for transverse magnetisation

For a single uncoupled spin undergoing exchange, the NMR spectrum is calculable using the McConnell equations. Such formulae are a modification of the classical Bloch equations, and are given by

$$\frac{d}{dt}\boldsymbol{M}^{+}(t) = (i\boldsymbol{\Omega} - \boldsymbol{R} + \boldsymbol{K})\boldsymbol{M}^{+}(t) \qquad 2.76$$

Here, $\mathbf{M}^{+}(t)$ is the complex-valued vector of FIDs for each chemical environment of the spin, $\boldsymbol{\Omega}$ is a diagonal matrix whose diagonal elements are the chemical shifts at each site accessible to the exchanging spin, $\mathbf{R}$ is a diagonal matrix whose diagonal elements are the transverse relaxation rates at each site accessible to the exchanging spin, $\mathbf{K}$ is the exchange matrix which describes the transformation amongst chemical environments, and $\mathbf{M}^{+}(0)$ is a vector of initial transverse magnetisations. Typically, we will assume that the initial state is *x*-magnetisation, such that the elements of $\mathbf{M}^{+}(0)$ are real numbers proportional to the concentrations of the different chemical sites at the start of the NMR experiment, and further assume that the experiment is conducted under conditions of chemical equilibrium. The energies involved in NMR have no effect upon the position of thermodynamic equilibrium of the reaction under study. A discussion of the exchange matrix follows this section, but it is useful first to find the general form of the solution to the McConnell equations. First, make the definition

$$\boldsymbol{V} = i\boldsymbol{\Omega} - \boldsymbol{R} + \boldsymbol{K} \qquad 2.77$$

Therefore

$$\frac{d}{dt}\boldsymbol{M}^{+}(t) = \boldsymbol{V}\boldsymbol{M}^{+}(t) \qquad 2.78$$

The solution is given by

$$\boldsymbol{M}^{+}(t) = exp(\boldsymbol{V}t)\boldsymbol{M}^{+}(0) \qquad 2.79$$

Now let $\mathbf{U}$ be a unitary matrix which diagonalises $\mathbf{V}$ such that

$$\begin{aligned} \boldsymbol{M}^{+}(t) &= \boldsymbol{U}\boldsymbol{U}^{-1}\, exp(\boldsymbol{V}t)\boldsymbol{U}\boldsymbol{U}^{-1}\boldsymbol{M}^{+}(0) \\ &= \boldsymbol{U}\, exp(\boldsymbol{U}^{-1}\boldsymbol{V}\boldsymbol{U}t)\boldsymbol{U}^{-1}\boldsymbol{M}^{+}(0) \\ &= \boldsymbol{U}\, exp(\boldsymbol{D}t)\boldsymbol{U}^{-1}\boldsymbol{M}^{+}(0) \end{aligned} \qquad 2.80$$

Where $\mathbf{D}$ is the diagonal form of $\mathbf{V}$. Thus $\mathbf{U}$ contains the eigenvectors of $\mathbf{V}$ and $\mathbf{D}$ its eigenvalues. The NMR spectrum is the Fourier transform of the sum elements of $\mathbf{M}^{+}(t)$. For simple reactions, the McConnell equations are solvable analytically (30, 92), but such approaches are unwieldy and impractical for most reactions, in which case numerical approaches are used.

It will be important in the interpretation of exchange-induced line-broadening to determine the relative contributions to line-width from the actual T_2 and from the exchange process. For that reason, and to enable rapid calculation of lineshapes in an exchanging system, we will derive a convenient expression for the frequency-domain signal. The propagator acting upon $\mathbf{M}^+(0)$ is

$$\boldsymbol{P} = \boldsymbol{U}\, exp(\boldsymbol{D}t)\boldsymbol{U^{-1}} \tag{2.81}$$

The elements of **P** can be written in the general form

$$P_{il} = \sum_{j,k} U_{ij} U_{kl}^{-1} exp(D_{ik}t) \tag{2.82}$$

We make the definitions

$$a_{il} = \sum_{j} U_{ij} U_{il}^{-1} \tag{2.83}$$

$$b_{il} = \sum_{j,k\neq i} U_{ij} U_{kl}^{-1} \tag{2.84}$$

Such that the elements of **P** can be re-written as

$$P_{il} = a_{il}\, exp(D_{ii}t) + b_{il} \tag{2.85}$$

Since the b_{il} terms are independent of time, we shall ignore them and examine only the time-dependent part of **P**. It would be convenient to resolve the time-dependent part into components representing chemical shift and line-width separately, accomplished by noting that the imaginary and real parts of **D** contain these parameters respectively. This motivates the definitions

$$exp\big(iIm\big(D_{jj}\big)t\big) = exp\big(id_{jj}t\big) \tag{2.86}$$

$$exp\left(Re\big(D_{jj}\big)\right) = exp\big(-c_{jj}t\big) \tag{2.87}$$

The time-dependent part of the propagator is then, in terms of individual elements,

$$P_{il}(t) = a_{il}\, exp(id_{ii}t)\, exp(-c_{ii}t) \tag{2.88}$$

And the full form of the signal is

$$s(t) = \sum_{i,l} a_{il}\, exp(id_{ii}t)\, exp(-c_{ii}t) M_{il}^{+}(0) + \sum_{i,\prime} b_{il} M_{il}^{+}\,(0) \tag{2.89}$$

In this equation, the a_{il} terms are constant multipliers for signal magnitudes, describing the rates of mixing of chemical states between which a spin can exchange, whose values may be complex. They form a symmetric matrix (we shall call it **a**) approximating the identity matrix for slow exchange, meaning that chemical states are not mixed, and that off-diagonal matrix elements of **a** are negligible in this regime. If exchange is fast, **a** is symmetric and real but off-diagonal elements are non-zero. If exchange is on a similar timescale to the chemical shift difference between states, the elements of **a** are complex, such that some terms in the spectrum will have phase distortions. The frequencies at which peaks will appear is given by the d_{ii} terms, with corresponding line-widths by the c_{ii} terms. Hence the c_{ii} terms provide a convenient means of stating the observed line-widths, from which can be subtracted the transverse relaxation rate constants to obtain the contribution from exchange to the line-width. This is essentially the R_{ex} term obtainable by Lipari-Szabo analysis, and means that meaningful model-dependent exchange parameters are accessible from the model-free R_{ex} term.

If the constant terms are neglected, the real part of the frequency domain signal is

$$S(\omega) = \sum_{i,l} \left\{ \frac{Re(a_{il})c_{ii}}{c_{ii}^2 + (\omega - d_{ii})^2} + \frac{Im(a_{il})(\omega - d_{ii})}{c_{ii}^2 + (\omega - d_{ii})^2} \right\} M_l(0) \qquad 2.90$$

Here, the summation on l only provides signal amplitudes rather than determining chemical shifts or linewidths. The terms ‘Re’ and ‘Im’ denote real and imaginary parts respectively, such that the real part of the NMR frequency-domain signal as absorptive and dispersive components. The dispersive parts only make a contribution if exchange is on a similar timescale to the chemical shift difference between states, in which case the phase distortions cancel one another out to give a signal with even symmetry. The signal can be normalised in accord with the condition that its integral is unity if multiplied by a normalisation factor

$$N = \pi \sum_{i,l} Re(a_{il})\, M_l(0) \qquad 2.91$$

Chapter 3 Methods

3.1. Proteins used in this thesis

A total of five different proteins have been used in the research presented in this thesis. The regions of Gag comprising these proteins are summarised in Figure 3.1. The shortest two proteins are C-CA and C-CA_{AA}, each of which span residues 146-231 of Gag, as labelled from the N-terminus of CA, the former with the wild-type sequence and the latter with W1814A/M185A mutations which prevent dimerisation whilst leaving the overall structure very similar. The proteins C-CA-NC and C-CA_{AA}-NC each span residues146-300 of Gag, each containing the SP1 and NC domains in addition to the C-CA domain. The 'AA' subscript has the same meaning as before. The largest protein used is ΔMACANCSP2, which contains a mutant of the MA domain in which residues 15-99 have been deleted, but which retains the basic N-terminus. It also contains the entire CA, SP1, NC and SP2 domains. As described in the introduction, it can assemble as immature capsids when incubated with nucleic acid at pH 8, and such capsids have been shown by cET to be faithfully representative of real immature capsids. All proteins were expressed using the pET-11c vector (Novagen), with the necessary molecular biology having been previously described (145). This expression vector is shown in Appendix 7.6, and contains the T7 operator and promoter recognised by T7 RNA polymerase, so the *E.Coli* strain BL21 (DE3) was used for protein preparation. The vector is IPTG-inducible, with low basal expression levels, and encodes ampicillin resistance.

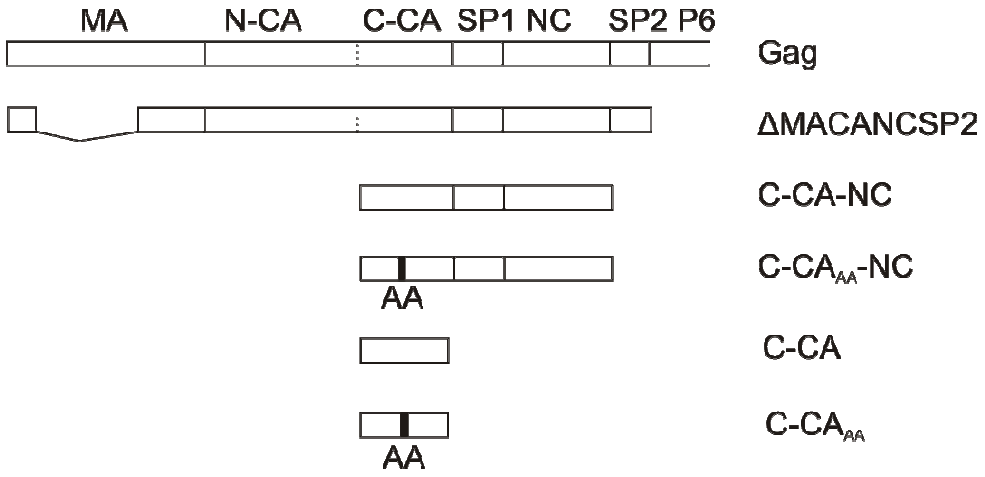


Figure 3.1: The regions of Gag used in this thesis.

3.2. Protein expression and purification

3.2.1. Media

LB media, per L (pH7.5)

10g Bactotryptone

5g Yeast extract

10g NaCl

0.27g $MgCl_2.7H_2O$

2ml Glycerol

Agar plates

LB media as above with 15g Agarose per L

M9 media, per L (pH7.4)

6 g Na_2HPO_4

3 g KH_2PO_4

0.5 g NaCl

0.25 g $MgSO_4.7H_2O$ (1 ml 1 M solution)

44 mg $CaCl_2.2H_2O$ (0.3 ml 1 M solution)

Autoclave, then add:

1 ml Biotin, 1 mg/ml solution

1 ml Thiamin, 0.4 mg/ml solution

1 g [^{15}N]-NH_4Cl

2 g [^{13}C]-Glucose

To all media, ampicillin is added at 1/1000 dilution from a 50 mg/ml solution after autoclaving.

3.2.3. Expression and purification of C-CA_{AA}-NC and C-CA-NC

These proteins were required for NMR and as such were prepared with uniform ^{13}C and ^{15}N labelling. The protocol has been described previously (145). Briefly, BL21(DE3) cells were transformed and 10 ml LB overnight cultures grown from single colonies. These were used to inoculate 500 ml flasks of M9 minimal media supplemented with 1g/L [^{15}N]-NH_4Cl as the sole nitrogen source and 2 g/L [^{13}C]-glucose as the sole carbon source. Such

cultures were grown to mid-log (OD_{600nm}~0.6-0.7) and induced with 0.6 mM IPTG overnight (16hours) at 30°C. Cells were harvested by centrifugation, lysed by sonication (5 s on, 5 s off, 10 min total process time) and insoluble material discarded. C-$CA_{(AA)}$-NC was precipitated with 40% ammonium sulphate and resuspended in AEX buffer. Anion exchange chromatography (Pharmacia HiLoad 26/10 Q sepharose HP) was used to remove nucleic acids, with C-$CA_{(AA)}$-NC remaining in the flow-through. Such fractions were pooled and C-$CA_{(AA)}$-NC purified by heparin affinity chromatography (GE healthcare HiTrap heparain HP), using a linear gradient of 0-1 M NaCl in HE buffer for elution. C-CA_{AA}-NC fractions were concentrated, followed by gel filtration (Pharmacia HiLoad 16/60 sephadex 75) in GF buffer. All buffers contained DTT to ensure reducing conditions, then pure protein exchanged into a buffer containing TCEP instead of DTT for long-term storage.

Using the ProtParam program, the (unlabelled) molar mass of C-CA_{AA}-NC was calculated to be 17253.8 g/mol with 156 amino acid residues, with the molar extinction coefficient at 280 nm calculated as 8480 M^{-1} cm^{-1} assuming there are no disulphide bonds (a condition met by the presence of DTT or TCEP). This predicted molar extinction coefficient was used to determine concentrations. The molecular formula of C-CA_{AA}-NC is $C_{735}H_{1206}N_{228}O_{223}S_{14}$, which was used to calculate the percentage isotopic enrichment. For C-CA-NC, PropParam calculated the molar mass to be 17429.1 g/mol with 156 residues, the molar extinction coefficient without disulphide bonds to be 13980 M^{-1} cm^{-1} and the formula to be $C_{745}H_{1215}N_{229}O_{223}S_{15}$.

Buffers:

Lysis buffer:
50 mM Tris pH 7.5
1 M NaCl
20 μM $ZnCl_2$
5 mM DTT
10% glycerol

AEX buffer:
30 mM Tris pH 8.2
500 mM NaCl

20 μM $ZnCl_2$
5 mM DTT

HE buffer:
30 mM HEPES pH 7.4
40 μM $ZnCl_2$
5 mM DTT

GF buffer:
50 mM Tris pH 8.0
5 mM DTT

3.2.4. Expression and purification of ΔMACANCSP2

A protocol for expression and purification of ΔMACANCSP2 was adapted from that already described (60). Since the protein was not required for NMR, LB media was used for the growth, and the protein expressed in BL21 (DE3) pLysS cells (Invitrogen). Induction was allowed for only 2 hours to maintain assembly-competence and avoid excessive intracellular assembly. Cells were harvested by centrifugation for 20 minutes at 6000 G, lysed by sonication (5 s on, 5 s off, 10 minute total process time) in a buffer containing 50 mM Tris pH 7.5, 1 M NaCl, 20 μM $ZnCl_2$, 10% v/v glycerol and insoluble material discarded. ΔMACANCSP2 was precipitated by addition of ammonium sulphate to 40% saturation at 4°C and collected by centrifugation at 27000 g for 30 minutes. Pelleted material was redissolved in 50 ml AEX buffer (30 mM Tris pH 8.2, 500 mM NaCl, 20 μM $ZnCl_2$, 5 mM DTT) and nucleic acids removed by anion exchange chromatography (Pharmacia HiLoad 26/10 Q sepharose HP). Proteins were collected from the flow-through and precipitated by addition of ammonium sulphate to 40% saturation. Pelleted material was resuspended in 50 ml CEX buffer (50 mM Mes pH 6.0, 200 mM NaCl, 5 mM DTT) and ΔMACANCSP2 purified by cation exchange chromatography (Pharmacia HiLoad 26/10 SP sepharose HP) with a 0.5-2 M NaCl linear elution gradient. Fractions were assayed for purity by SDS-PAGE and those containing pure ΔMACANCSP2 were subject to ammonium sulphate precipitation at 40% saturation. Pelleted, pure ΔMACANCSP2 was re-dissolved to 100 μM in 30 mM Tris pH 8.0, 2 mM TCEP, 0.05%NaN_3. TCEP was used as the reducing agent in the final stage on account of its longer half-life than DTT. Concentrations were determined using the molar extinction coefficient as predicted by the

ProtParam tool, which was 47940 M^{-1} cm^{-1} with a molar mass of 40.06 kDa. Purity was assessed by SDS-PAGE and the A_{280}/A_{260} ratio. SDS-PAGE was used as an assay of purity.

3.3. Sample purity results

3.3.1. Purity of C-CA_{AA}-NC preparation

The progress of purifications was followed by SDS-PAGE as an assay of purity and to detect which fractions contained protein following a column chromatography step. A typical result is shown in Figure 3.2, demonstrating the level of purity typically obtained for C-CA_{AA}-NC after gel filtration. It is highly important that C-CA_{AA}-NC is separated from nucleic acids, which could be assayed using the ratio of absorbance of 280 nm light to 260 nm. Protein was not regarded as pure unless the A_{280}/A_{260} ratio was greater than 1.67, since this indicates that nucleic acid, which absorbs more strongly at 260 nm than 280 nm, is making a negligible contribution to the measured absorbance. C-CA_{AA}-NC was stable at room temperature for several weeks after gel filtration, and could be frozen in liquid nitrogen and stored at -80 °C indefinitely. The yield of purified ^{13}C, ^{15}N C-CA_{AA}-NC from 1 L culture was 10-15 mg.

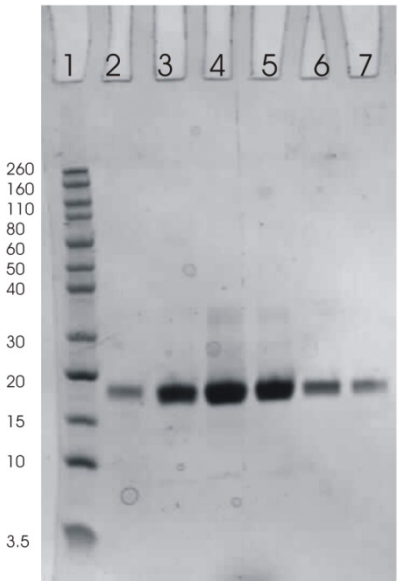


Figure 3.2: SDS-PAGE analysis of purified C-CA_{AA}-NC.

Lane indices: 1: Molecular weight marker, 2-7: Peak fractions after gel filtration.

3.3.2. Purity of C-CA-NC preparation

The same preparation techniques and criteria for purity were applied to C-CA-NC. The SDS-PAGE result in Figure 3.3 gives an indication of sample purity, and again separation from nucleic acids was judged from the A_{280}/A_{260} ratio.

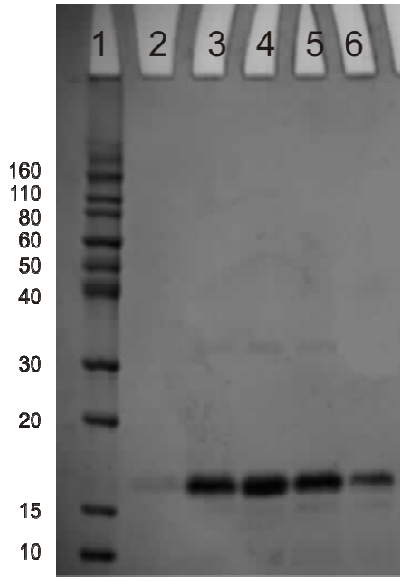


Figure 3.3: SDS-PAGE analysis of purified C-CA-NC

Lane indices: 1: Molecular weight marker, 2-6: Peak fractions after gel filtration.

3.3.3. Purity of ΔMACANCSP2 preparation

After the final cation exchange chromatography step, SDS-PAGE was used to determine the purity of peak fractions, and an example gel is given in Figure 3.4 to demonstrate that the protein was indeed highly pure.

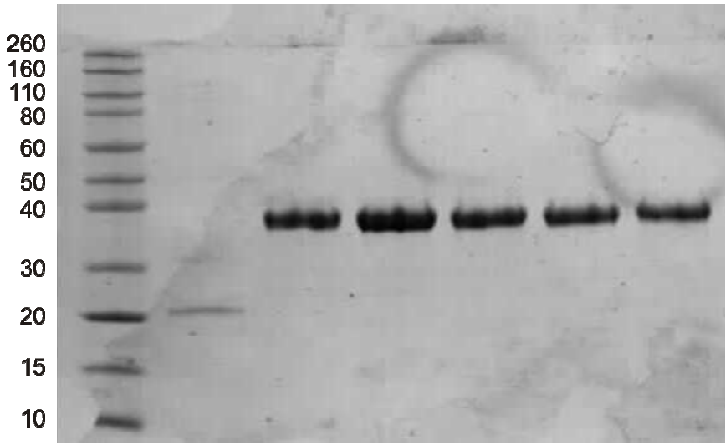


Figure 3.4: SDS-PAGE analysis of purified ΔMACANCSP2.

Gel demonstrating that the protein is highly pure after cation exchange chromatography. The left-most lane is a molecular weight marker. All others are peak fraction from cation exchange chromatography.

Chapter 4 NMR studies of HIV-1 proteins

4.1. Introduction

In this chapter, the NMR experiments used to characterise both the wild-type dimerising protein C-CA-NC and its non-dimerising W184A/M185A mutant, C-CA_{AA}-NC, are described. The experiments performed make use of chemical shifts, relaxation analysis and RDCs, the theoretical background of which was described in chapter 2. Similar studies are performed for the single domain constructs C-CA and C-CA_{AA}, which establish how the wild-type and mutant C-CA domains behave in the absence of the NC and SP1 domains. The extent to which intramolecular and intermolecular interactions amongst HIV-1 Gag domains are influenced by one another has not been studied in detail before. By working with non-dimerising C-CA_{AA} and C-CA_{AA}-NC, it is possible to assess whether dimerisation has any effects upon regions of the protein distal to the dimerisation interface under the same conditions in which the dimer is studied. Inasmuch as immature assembly involves distinct interactions from mature assembly, we can use comparisons of C-CA-NC and C-CA_{AA}-NC to determine whether the NC or SP1 domains have any modulating effect upon C-CA domain dimerisation, and *vice versa*. Various studies have suggested a helical structure in the SP1 region, and we have used secondary Cα shifts and relaxation data to determine secondary structure and dynamics respectively. A previous study examined C-CA-NC secondary Cα shifts and dynamics, and the analysis presented herein for C-CA_{AA}-NC complements this, demonstrating the influence of C-CA dimerisation upon structure and dynamics of the SP1 region.

Experiments in which ligands are used to perturb a single domain of the dual-domain proteins are also described. The binding of CAI to C-CA_{AA}-NC is studied and compared to the equivalent studies with C-CA_{AA}, and the binding of a penta-deoxynucleotide dACGCC to the NC domain of C-CA-NC and C-CA_{AA}-NC is studied also.

4.2. NMR methods

4.2.1. Assignments

The relaxation experiments, RDC measurements and titrations used to study C-CA-NC and C-CA_{AA}-NC are all reliant upon knowing which amide resonance in a 2D ^{1}H-^{15}N HSQC

spectrum corresponds to which amino acid residue in the protein, that is, assigning resonances to particular sites in the protein. To do so, triple resonance experiments are used which also exploit the ^{13}C nuclei present, which is present at almost 100% abundance due to uniform labelling strategies. In general, the triple resonance experiments used to make NH assignments correlate the amide proton shift of the i^{th} residue in the acquisition dimension with the directly bonded amide nitrogen shift and some combination of carbon shifts from the i^{th} and/or $(i\text{-}1)^{th}$ residues. A suite of experiments is generally necessary, since spectral crowding and poor signal-to-noise (S/N) are often complicating factors for information-rich experiments, whereas there may be insufficient basis for unambiguous assignment if too few correlations are made, regardless of S/N. The three experiments used for making assignments are summarised in figure 4.1. Triple resonance experiments are named according the correlations that are made; HNCACB, summarised in Figure 4.1 B, correlates the amide proton and nitrogen shifts with the Cα and Cβ shifts of the i^{th} and $(i\text{-}1)^{th}$ residues, allowing sequential assignments to be made. CBCA(CO)NH correlates the amide proton and nitrogen shifts of the i^{th} residue with the Cα and Cβ shifts of the $(i\text{-}1)^{th}$ residue only. Sequential assignments are not possible, but the experiment lifts the degeneracy in the HNCACB spectrum.

A list of all the experiments used for assignment purposes in this thesis are given in Table 4.1. Note that the assignments for C-CA_{AA} and C-CA as single-domain proteins have been published elsewhere (145) and so details of their assignments are not repeated here. Likewise for C-CA_{AA} + CAI (146). All triple-resonance experiments were gradient-enhanced coherence-selected (30), as implemented in Biopack.

Table 4.1: List of experiments used for assignments.

The Acq time column refers to the acquisition times in the indirect dimensions. The relevant nuclei appear as subscripts. All experiments were conducted at 25 °C, and at pH 7.0 in 20 mM Hepes buffer with a 10-fold molar excess of TCEP and 0.05% NaN_3.

Molecule	Conc. (mM)	Field (MHz)	Experiment	Acq time	Value (ms)	Spectral width	Value (Hz)
C-CA_{AA}-NC	0.5	600	HNCACB	τ_N	24	ω_N	1530
				τ_C	6.7	ω_C	9600
			CBCACONH	τ_N	24	ω_N	1530
				τ_C	6.7	ω_C	9600
			HNCA	τ_N	24	ω_N	1824
				τ_C	6.7	ω_C	4225
C-CA_{AA}-NC + CAI	0.5 0.55	600	HNCACB	τ_N	22	ω_N	1530
				τ_C	6.0	ω_C	9600
			CBCACONH	τ_N	24	ω_N	1530
				τ_C	6.0	ω_C	9600
C-CA_{AA}-NC + dACGCC	1.0 1.1	600	HNCACB	τ_N	24	ω_N	1500
				τ_C	4.6	ω_C	12063
			CBCACONH	τ_N	27	ω_N	1500
				τ_C	5.3	ω_C	12063
C-CA_{AA}-NC + dSL3	0.5 0.55	600	HNCA	τ_N	20	ω_N	2000
				τ_C	9.7	ω_C	4523
C-CA-NC	0.5	800	HNCA	τ_N	18	ω_N	1983
				τ_C	8.0	ω_C	8803

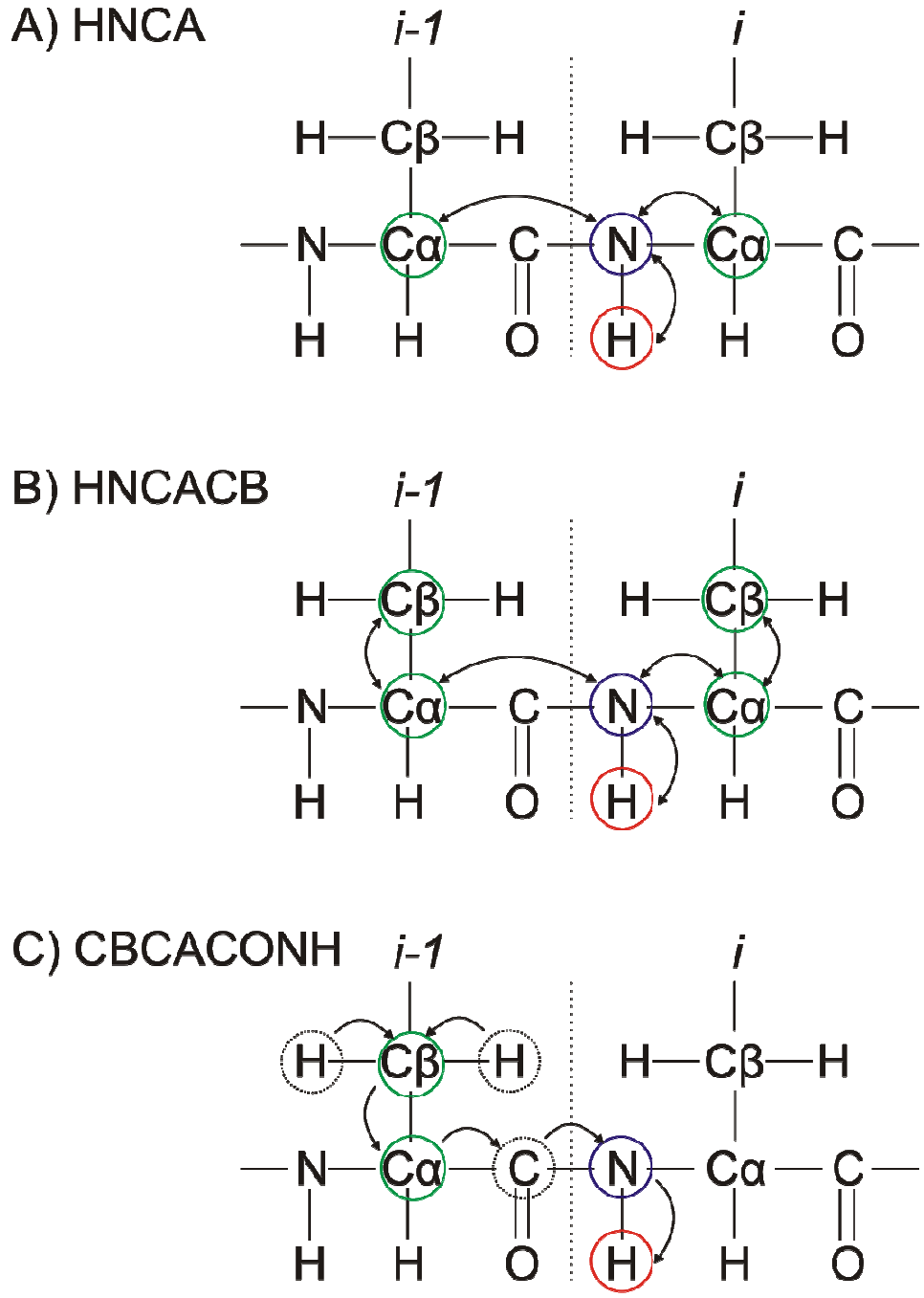


Figure 4.1: Triple resonance experiments used for assignments.

Observed spins are shown in coloured circles according to their dimension, unobserved spins to which coherences are transferred in the experiment are circled with dotted lines. Arrowheads denote the direction of coherence transfer. A and B are out-and-back experiments, so coherences are transferred though the spin system from the amide proton, then back again. C is a straight-through experiment, in which aliphatic protons are excited but the signal acquired on the amide proton. All coherence transfers are by INEPT sequence elements.

4.2.2. Relaxation measurements

The relaxation parameters required were described in chapter 2.4. They are T_1 for amide ^{15}N longitudinal autorelaxation, T_2 for transverse autorelaxation and the [^{1}H]-^{15}N NOE. Full details of the experiments can be found elsewhere (30). Briefly, T_1 is measured using an HSQC-like sequence in which a density operator proportional to S_z (the operator for ^{15}N longitudinal magnetisation) is prepared, then allowed to decay for a time τ. This is the relaxation delay time. The sequence then generates (-1)QC on the ^{15}N spin, allows the chemical shift to evolve then transfers coherence to ^{1}H for acquisition. The pulse sequence for measuring T_2 is similar but with the obvious difference that a density operator proportional to S_x, rather than S_z, is generated during the preparation period. In either case, several spectra are recorded, for different values of the relaxation delay τ (typically around 7 values are used), and intensities fitted to the equation

$$S(\tau) = Aexp\left(-\tau/T_j\right) \qquad 4.1$$

Here, T_j can be either T_1 or T_2. The NOE is measured by recording two HSQC-like spectra, one with saturation of protons during the recycle delay and one without. By saturation is meant that the proton spins are irradiated at their Larmor frequency with weak RF radiation, equalising spin state populations, for which reason this experiment is said to measure the steady-state NOE. The NOE as defined in chapter 2.4 is then the ratio of intensities in the saturated spectrum to the unsaturated. The error for signal intensity is taken to be the standard deviation of the noise. Errors on fitted parameters are estimated using a Monte-Carlo procedure.

The conditions used in relaxation experiments are listed in Table 4.2, which also acts as a list of all molecules and complexes analysed. Note that the relaxation data for the single-domain proteins C-CA and C-CA$_{AA}$ were not measured personally, but all analyses presented in this thesis were conducted personally.

Table 4.2: List of experimental conditions for relaxation measurements. All samples were in 20 mM Hepes, pH 7.0 with a 10-fold molar excess of TCEP and 0.05% NaN_3.

Molecule	Conc (mM)	Field (MHz)	Parameter	Parameter values
C-CA_{AA}-NC	0.5	600	T_1 delay	0.02, 0.08, 0.16, 0.32, 0.60, 1.0, 1.4
			T_2 delay	0.01, 0.03, 0.05 0.07, 0.11, 0.13, 0.15, 0.21,
			NOE sat	3.5
C-CA_{AA}-NC + CAI	0.5 0.55	600	T_1 delay	0.02, 0.08, 0.16, 0.22, 0.6, 1.0, 1.4
			T_2 delay	0.01, 0.03, 0.05, 0.07, 0.11, 0.13, 0.15, 0.21,
			NOE sat	3.5
C-CA_{AA}-NC + dACGCC	0.5 0.55	600	T_1 delay	0.02, 0.08, 0.16, 0.26, 0.6, 1.0, 1.4
			T_2 delay	0.01, 0.03, 0.07, 0.11, 0.15, 0.21, 0.31
			NOE sat	3.5
C-CA_{AA}-NC + dSL3	0.5 0.5	600	T_1 delay	0.02, 0.08, 0.16, 0.26, 0.6, 1.0, 1.4
			T_2 delay	0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.15, 0.21,
			NOE sat	3.5
C-CA-NC	0.1	800	T_1 delay	0.02, 0.1, 0.25, 0.5, 0.75, 1.0, 1.5
			T_2 delay	0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.15
C-CA-NC	0.5	700	T_1 delay	0.01, 0.05, 0.1, 0.1, 0.5, 0.8, 1.4
			T_2 delay	0.008, 0.024, 0.048, 0.072, 0.096, 0.128, 0.2
			NOE sat	5.0
C-CA_{AA}	1.0	600	T_1 delay	0.02, 0.08, 0.16, 0.32, 0.6, 1.0, 1.4
			T_2 delay	0.01, 0.03, 0.05, 0.07, 0.11, 0.13, 0.15
			NOE sat	3.5
C-CA	0.1	600	T_1 delay	0.02, 0.08, 0.16, 0.32, 0.6, 1.0, 1.4
			T_2 delay	0.01, 0.03, 0.05, 0.07, 0.11, 0.13, 0.15
C-CA	0.1	600	T_1 delay	0.02, 0.08, 0.16, 0.32, 0.6, 1.0, 1.4
			T_2 delay	0.01, 0.03, 0.05, 0.07, 0.11, 0.13, 0.15

4.2.3. Interpreting relaxation measurements

Relaxation data, where possible, were interpreted using the Lipari-Szabo model-free formalism, using the procedure of Mandel *et al* and implemented in the program Modelfree4 (98). This requires firstly the determination of a diffusion tensor. To

accomplish this, the methods of chapter 2.4.3 were used. For the C-CA_{AA} domain of C-CA_{AA}-NC, C-CA_{AA} alone, the C-CA domain of C-CA-NC and C-CA alone, the X-ray crystal structure of wild-type C-CA was used as a structural model in diffusion tensor determination (158). If CAI was present, the X-ray crystal structure of C-CA in complex with CAI was used instead (148). For the NC domain, the diffusion tensor was determined using NMR average structure of a peptide fragment containing residues 257-298 of NC (pdb code 1esk). This is the most recently determined unliganded NC structure in the protein databank. When dACGCC was present, the NMR average structure of the NC-dACGCC complex was used (107).

Once the diffusion tensor has been determined, 5 different approximations to the spectral density functions were fitted to the measured relaxation parameters by a χ^2 minimisation, incorporating different motional parameters. These are summarised below:

Model 1: S_s^2

Model 2: S_s^2 and τ_e

Model 3: S_s^2 and R_{ex}

Model 4: S_s^2, τ_e and R_{ex}

Model 5: S_s^2, S_f^2 and τ_e

In all analyses, the quoted order parameter is the product $S^2 = S_s^2 S_f^2$ and, unless model 5 is used, $S_f^2 = 1$

Errors on the fitted parameters were estimated using Monte-Carlo simulations. The simplest model which adequately describes the measured data is accepted, and once models have been selected for all residues, a full optimisation allowing both the diffusion tensor and model-free parameters to vary is conducted. The analysis is considered complete when a convergent solution is found. By this is meant that upon full optimisation the diffusion tensor should not change from its initial set of values, and all residues should remain faithful to their selected models. To determine whether a model adequately describes the relaxation parameters for a particular residue, the χ^2 distribution is simulated, and the measured χ^2 statistic should be lower than that at the 95th percentile. In order to simulate the χ^2 distribution, relaxation parameters are back-calculated from the fitted

spectral density parameters many times, including for each calculation the addition of error, taken as the measured experimental error, drawn from a Gaussian distribution.

4.2.4. RDC measurements

RDC measurements were performed by measuring the sum of dipolar and scalar couplings for C-CA_{AA}-NC firstly in isotropic media, then in an aligning media. Under isotropic conditions, the contribution of the dipolar coupling is zero, such that subtraction of the scalar coupling measured under isotropic conditions from the sum of dipolar and scalar couplings measured under aligning conditions yields the RDC. Alignment was achieved using a laterally compressed polyacrylamide gel. This technique requires the preparation of a cylindrical polyacrylamide gel of 6 mm diameter, which is subsequently dried. The protein solution is then soaked into this, and the rehydrated gel is squeezed into a 5 mm diameter NMR tube, resulting in longitudinally elongated pores in the gel and thus conditions in which a preferential alignment is adopted by protein (14, 32).

The [^{1}H]-^{15}N couplings are measured using the IPAP (In-Phase Anti-Phase) technique. For each media, two HSQC-like experiments are performed. In the first, a density operator is prepared proportional to S_x, the chemical shift and coupling (scalar and dipolar) of which is allowed to evolve during the indirect acquisition period prior to a mixing period for subsequent acquisition on the amide proton. In the second experiment, a density operator proportional to an anti-phase term on the amide ^{15}N is prepared and allowed to evolve. The sum and difference of the two spectra then yield one peak of the doublet each.

All alignment tensors were calculated using Module (41). This software uses rigid body modelling to extract values for the axial (A_a) and rhombic (A_r) components of the alignment tensor, along with the 3 Euler angles defining the orientation of alignment relative to a fixed molecular frame as defined by the PDB for the protein under study. The axial and rhombic components are then obtained by solving equation 2.51 using SVD.

Alignment tensors were calculated for the unliganded C-CA_{AA} domain using the crystal structure of wild-type C-CA (158) and for the CAI-bound C-CA_{AA} domain using the CAI-bound crystal structure (148). In the fitting, the measured RDCs were used directly, but those for residues known to have high internal mobility (as judged by relaxation data) were excluded. This was because RDCs should be weighted by order parameters to make

alignment tensor determination rigorous, and for poorly structured regions the assumption of a rigid body obviously does not hold.

The experimental conditions for all RDC measurements are listed in Table 4.3. All measurements were made at 600 MHz field strength, 25 °C, at pH 7.0, with a soaking time of 3 days to ensure all protein was absorbed into the polyacrylamide gel. As in the case of relaxation measurements, experiments for single-domain C-CA_{AA} were conducted by Dr Joern Werner, but alignment tensors were calculated personally.

Table 4.3: List of RDC experiments.

The gel % cross-linking column refers to the percentage cross-linking of the polyacrylamide gels into which samples were soaked. All samples were buffered with 20 mM Hepes pH 7.0, 5 mM TCEP, 0.05% NaN_3.

Molecule	Concentration (mM)	Gel % cross-linking
C-CA_{AA}-NC	0.5	5
C-CA_{AA}-NC + CAI	0.5 0.55	5
C-CA_{AA}-NC + dACGCC	0.5 0.55	5
C-CA_{AA}	0.5	7
C-CA_{AA} + CAI	0.5 2.0	7

4.2.5. Chemical shift differences

For comparison of amide chemical shifts from HSQC spectra, the ^{1}H and ^{15}N chemical shift changes were measured, and a weighted chemical shift defined by

$$\Delta = \sqrt{\Delta\delta_H^2 + \frac{1}{6}\Delta\delta_N^2} \qquad 4.2$$

Such comparisons are possible only when mutual assignments in the 2 proteins to be compared are available.

4.3. NMR studies of C-CA_{AA}-NC

4.3.1. Assignments for C-CA_{AA}-NC

Assignments were made for 140 backbone resonances in C-CA_{AA}-NC, out of 148 non-proline residues (proline does not have an amide NH due to its cyclic structure). An assigned HSQC is given in Figure 4.2, along with some example strip plots from an HNCACB. Although assigning peaks in the highly crowded centre of the spectrum was difficult, for all but the most highly overlapped resonances an assignment could be made.

4.3.2. Relaxation data for C-CA_{AA}-NC

A total of 134 residues were analysed for T_1, 131 for T_2 and 132 for NOE. Those not analysed are mostly due to spectral crowding, but for T_1 and T_2 particularly short or long values are outside experimental limits and for the NOE values close to zero are very poorly determined. The relaxation data for C-CA_{AA}-NC are shown in Figure 4.3. Even without further analysis, it is apparent that the C-CA_{AA} and NC domain have different correlation times, since the residues in the NC domain have generally shorter T_1 times, longer T_2 times and lower NOEs than those in the C-CA_{AA} domain. It is also clear that the flexible linker (which includes the SP1 domain) is aptly named, since its high T_1 times and high T_2 times, combined with low and often negative NOEs imply rapid overall and/or internal motion. The C-terminal and N-terminal residues of C-CA_{AA}-NC also have high amplitude fast motion, which is often the case in proteins.

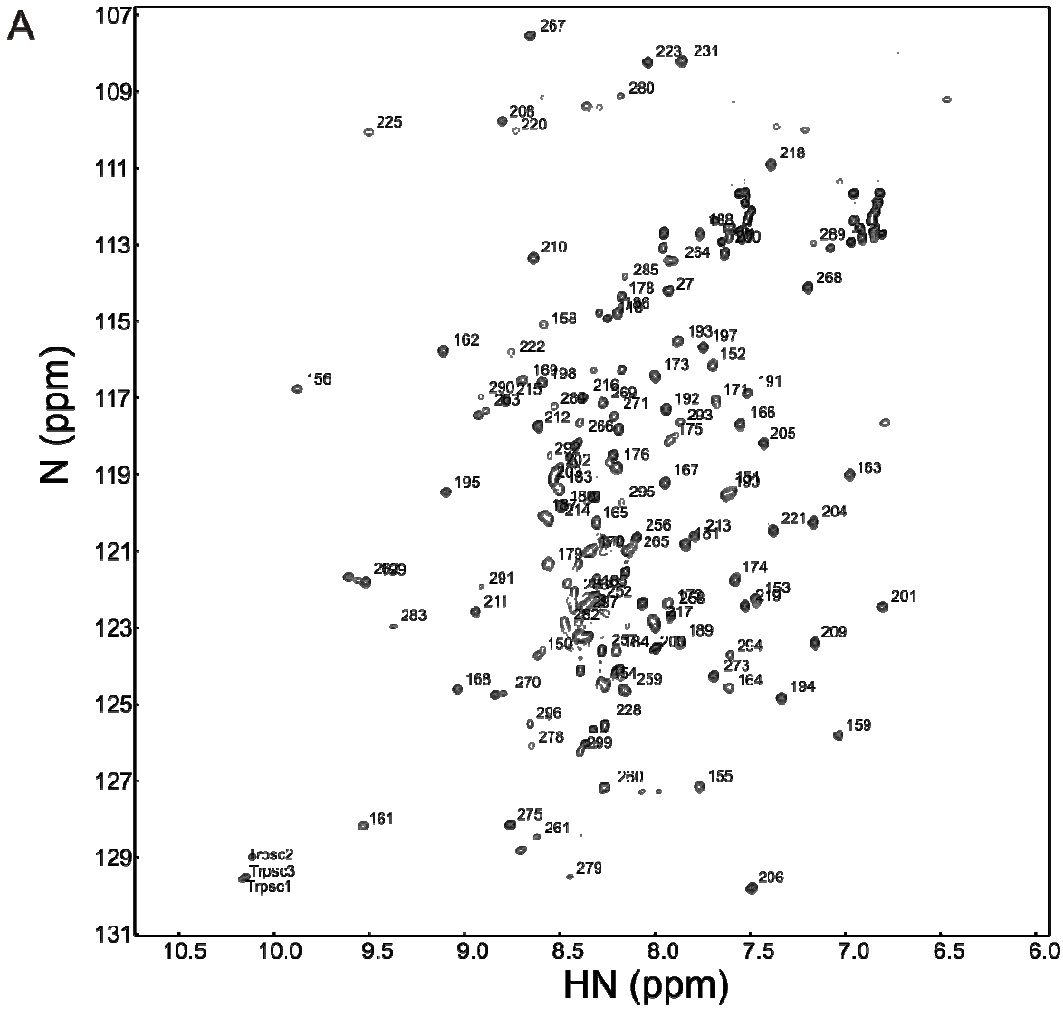

A
N (ppm)
HN (ppm)
B
H226
K227
A228
R229
V230
L231
A232
E233
A234
M235

Figure 4.2: (Previous page) Assignments for C-CA$_{AA}$-NC.

A) ^{1}H-^{15}N HSQC spectrum and B) strips through an HNCACB spectrum showing sequential assignments. Those shown are for a region within the flexible linker (G220-K256) and were amongst the more challenging assignments to make. Sample conditions are concentration: 500 µM, temperature: 25 °C, field: 600 MHz, pH 7.0.

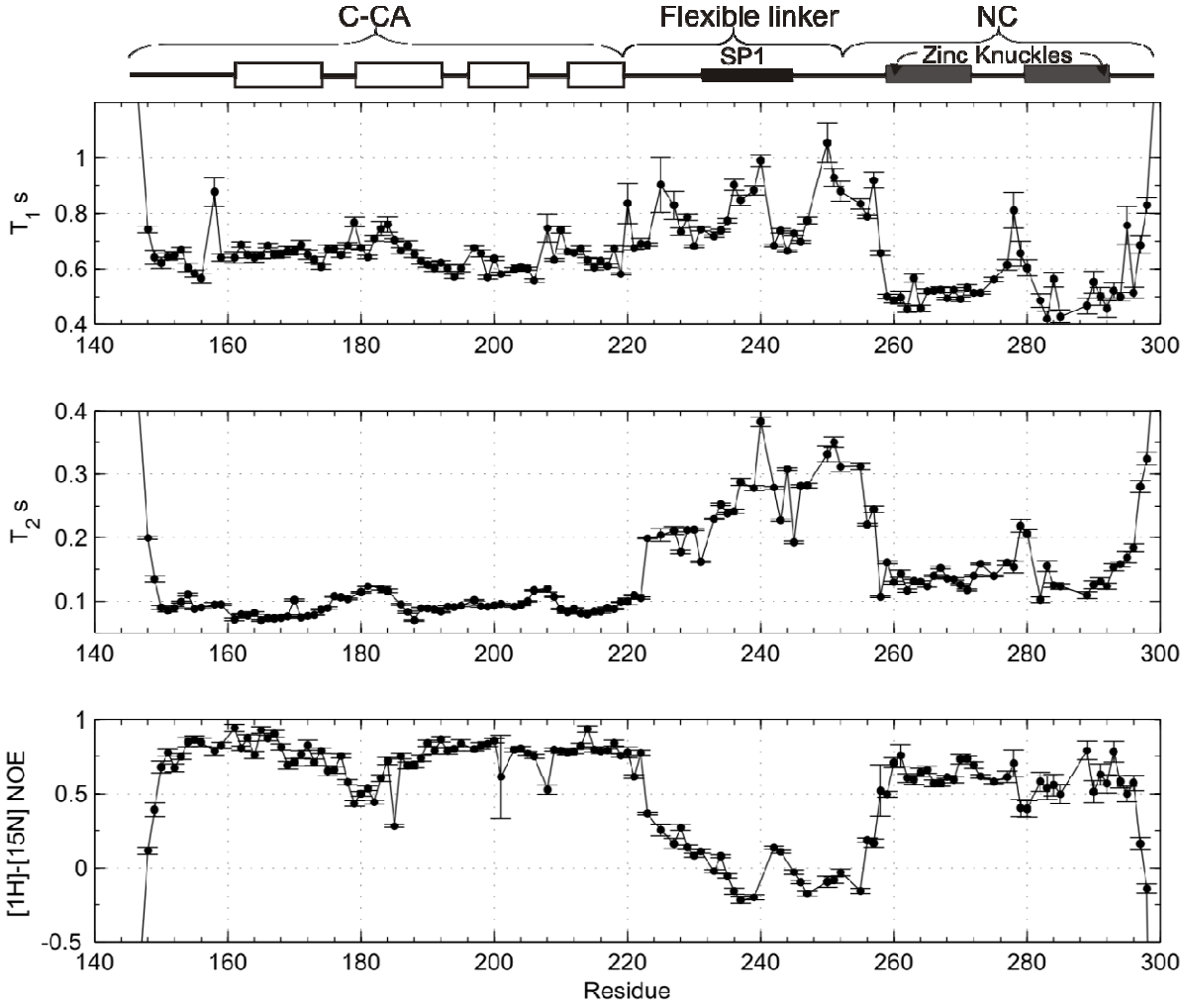


Figure 4.3: Relaxation data for C-CA$_{AA}$-NC.

Sample conditions are concentration: 500 µM, temperature: 25 °C, field: 600 MHz, pH 7.0. The helices of the C-CA$_{AA}$ domain are indicated above with empty boxes, the 14 residues SP1 domain is indicated with a solid black box and the zinc knuckles of NC are indicated with solid grey boxes. The term "flexible linker" refers to the entire flexible region between the well folded parts of C-CA$_{AA}$ and NC, and comprises the 12 C-terminal residues of C-CA$_{AA}$, all of SP1 and the 12 N-terminal residues of NC.

The dynamics of the folded domains of C-CA_{AA}-NC were analysed with the Lipari-Szabo model-free formalism described in chapter 2.4.3, using the program Modelfree4. Since the dynamics are clearly different for the C-CA_{AA} domain and NC domain, a global diffusion tensor was not sought, and instead separate diffusion tensors were determined for each folded domain. Such diffusion tensors are summarised in Table 4.4.

For the C-CA_{AA} domain, the best model for the diffusion tensor is a prolate elliptic tensor, whose principal frame is close to the inertia frame. A total of 36 residues were selected as representative of overall motion, to predict the 4 parameters of the diffusion tensor, using the criteria that only residues with NOE > 0.7 were sufficiently rigid.

For the NC domain, an oblate elliptic diffusion tensor was found to give the best fit to the data, using 14 residues with NOE > 0.6 as representative of overall motion. The lower NOE criteria, as compared to that applied to the C-CA_{AA} domain, was used since the NC domain is smaller than the C-CA_{AA} domain.

Table 4.4: Diffusion tensors for the C-CA_{AA} and NC domains of C-CA_{AA}-NC.

Domain	τ_m (ns)	$D_{\parallel}/D_{\perp}$	θ (°)	φ (°)
C-CA_{AA}	7.9 ± 0.02	1.49 ± 0.015	23 ± 2	-31 ± 6
Linker	7.9 ± 0.02	n/a	n/a	n/a
NC	5.3 ± 0.04	0.64 ± 0.03	35 ± 2	-22 ± 8

For the folded part of the C-CA_{AA} domain (residues 146-219), relaxation data for 57 residues could be quantified using the model-free approach. Those not analysed are due either to spectral crowding or poorly determined relaxation parameters. The fitted parameters are shown in Figure 4.4 and tabulated in the appendix. The flexibility apparent in the relaxation data is clearly demonstrated in a quantitative sense by means of this analysis; of the 57 residues analysed, only 16 could be described using model 1, in which motion is assumed to be too rapid to be detectable in the NMR relaxation parameters. Of the remainder, 11 were accounted for by model 2, generally with τ_e terms on the order of 1 ns. Model 3 could account for 9 residues, mostly with R_{ex} terms of around 1 Hz, meaning that there is also slow motion within C-CA_{AA}. Model 4 was fitted to 2 residues whilst model 5 was fitted to 19 residues. The overall picture of C-CA_{AA} internal motion is thus of a generally well folded protein with bond vector fluctuations in the low ns range. The

region of C-CA_{AA} between helices 1 and 2 and the N-terminal few residues of helix 2, encompassing the mutated dimer interface, showed particular flexibility, as evidenced by low order parameters and many τ_e terms. Often model 5 was needed, in which motions on two distinct fast timescales contribute to the spectral density function. An interesting exception was residue 188, with and R_{ex} contribution of ~3 Hz, which implies conformational exchange on a slow timescale.

For the folded part of the NC domain (residues 257-298), 36 residues were successfully analysed with the model-free approach. It was found that 2 residues were accounted for by model 1, 14 by model 2, 4 by model 3, 10 by model 4 and 6 by model 5. The fitted parameters are presented in Figure 4.4 and tabulated in the appendix. The NC domain is therefore in general flexible on at least two timescales; there is fast internal motion on a timescale of ~100 ps, whilst several other residues show fast motion on a timescale of ~1000ps, with very few τ_e terms in between these values. There are 14 residues for which an R_{ex} term was necessary, distributed throughout the NC domain, typically of around 1-3 Hz, which implies that the domain is also flexible on a slow timescale, in the microsecond range. Since NC is a zinc-binding protein, the R_{ex} terms could be attributed to bound/unbound forms. However, given the very tight binding constants for zinc, the bound population would be too large for the observed exchange terms to be due to a binding event. Thus it is more likely that exchange broadening in the NC domain is due to conformational sampling, consistent with the established fact that NC can adopt different conformations in order to bind different nucleic acids.

For the flexible portion of C-CA_{AA}-NC, from residues 220-256, only 26 of the 36 residues could be accurately analysed for relaxation data due to the spectral crowding. It is clear, even without a model-free analysis, that such residues are generally flexible on a fast timescale. To determine this timescale of motion, the overall motion of the linker residues was approximated by an isotropic diffusion tensor with the same correlation time as for the C-CA_{AA} domain, that is, the C-CA_{AA} domain dominates the overall motion of the linker. Using this approximation, all 26 resolved resonances for residues in the flexible linker could be described using model 5 (but not a simpler model), invoking motion on two distinct timescales, each faster than overall motion. It was found that most residues had τ_e terms in the range 800-1100 ps, thus giving an estimate for the typical timescale of motion in the linker. Inasmuch as requiring model 5 to account for such data, backbone dynamics

in the linker also occur on a timescale faster than 50 ps, though the exact timescales are beyond the limits of the experiments to accurately quantify. The large conformational space accessible to residues in the linker is indicated by the order parameters, which are generally in the range 0.1-0.2, and slightly higher towards the C-CA_{AA} domain.

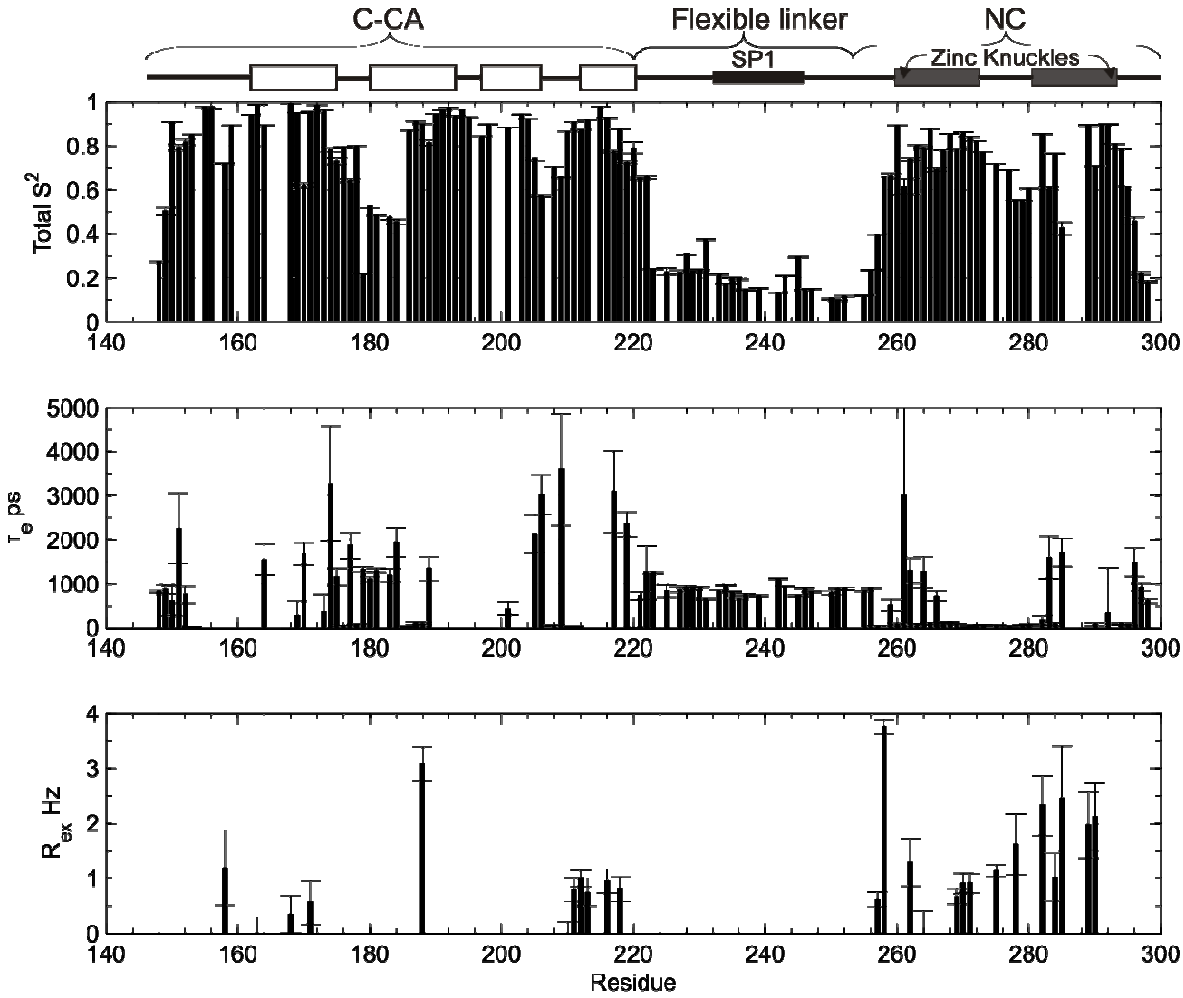


Figure 4.4: The results of a Lipari-Szabo model-free analysis applied to C-CA_{AA}-NC.

Separate diffusion tensors were used to describe the C-CA_{AA} domain, flexible linker and NC domains, as described in the text. There is fast internal motion distributed throughout the entire molecule, and in the NC domain slow exchange also contributes to linewidths for many peaks. Of 156 residues, 119 were analysed and it is for these that the above results apply. Of the remaining residues, 8 are prolines, and the 21 non-proline residues not analysed are either unassigned, or spectral crowding prevented the accurate determination of relaxation times or NOEs.

4.3.3. RDCs for C-CA_{AA}-NC

The alignment tensor parameters for the C-CA_{AA} domain are given in Table 4.5. For the NC domain, an alignment tensor could not be determined. This is partly due to many peaks corresponding to NC domain residues being overlapped with those in the flexible linker, thus preventing RDC measurement, and also likely to be attributable to the inherent flexibility of the NC domain, removing the validity of the rigid body approach. This flexibility is verified by relaxation data. Likewise for the flexible linker, spectral crowding was too severe to accurately measure RDCs. Such crowding is shown in Figure 4.5. The measured RDCs for the C-CA_{AA} domain of C-CA_{AA}-NC are shown in Figure 4.6. A total of 66 RDCs could be determined in the C-CA_{AA} domain, whilst those for the other regions are sparsely distributed and as such excluded from the figure. In alignment tensor calculation, 52 RDCs were used, and the agreement between measured and predicted RDCs for these is shown in Figure 4.7. These data are listed in the appendix, which also shows which RDCs were used in alignment tensor determination.

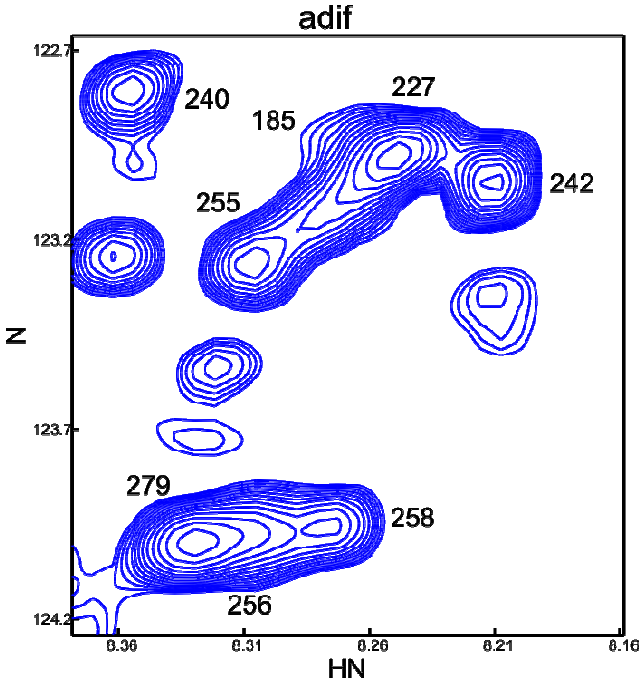


Figure 4.5: Extract from the difference spectrum for aligned C-CA_{AA}-NC.

The degeneracy of peaks makes determination of RDCs very difficult, and nearly all of the peaks shown are overlapped to some extent. Most correspond to NC domain residues or flexible linker residues, exemplifying why RDC analysis was not extended beyond the folded part of the C-CA_{AA} domain of C-CA_{AA}-NC.

Table 4.5. The alignment tensor parameters for the C-CA$_{AA}$ domain of C-CA$_{AA}$-NC. Angles are reported in degrees.

α	β	γ	A_a x 10^4	A_r x 10^4
-158 ± 6	166 ± 1	-74 ± 4	2.23 ± 0.04	0.79 ± 0.09

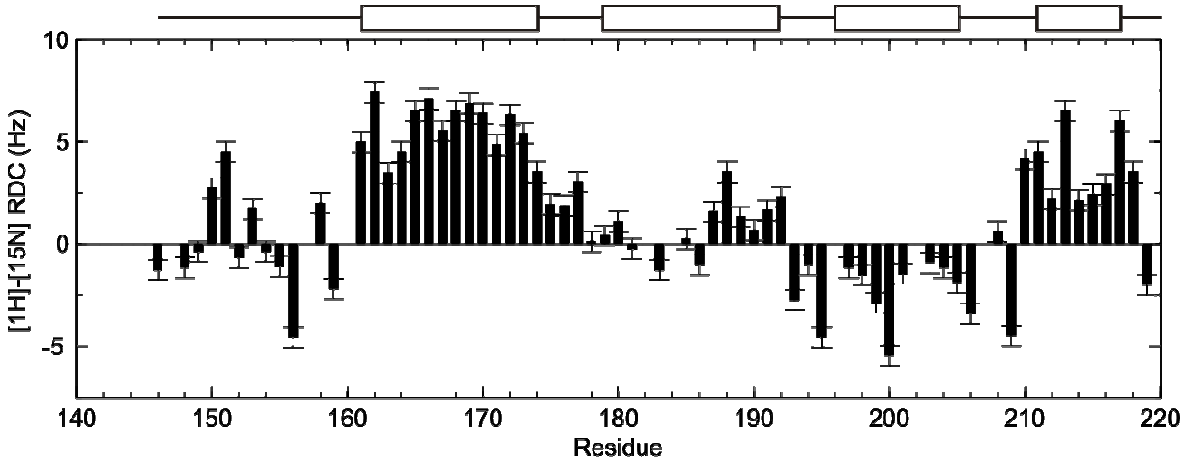


Figure 4.6: The measured RDCs for the C-CA$_{AA}$ domain of C-CA$_{AA}$-NC.

The helices are indicated by boxes above the figure, and a standard error of 0.5 Hz is assumed.

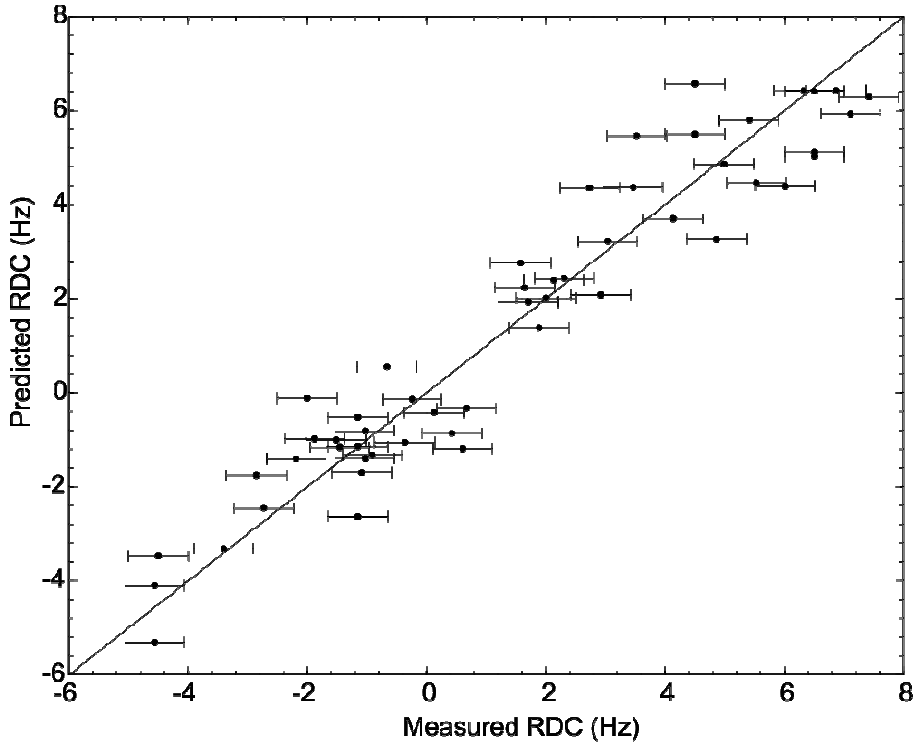


Figure 4.7: Measured and predicted RDCs for the C-CA_{AA} domain of C-CA_{AA}-NC, showing only those which were included in alignment tensor determination.

4.3.4. Cα secondary shifts for C-CA_{AA}-NC

The Cα secondary shift is the difference between the Cα chemical shift for a particular type of residue in a random coil (determined in urea for model peptides) and the observed Cα chemical shift for that type of residue in a protein. Secondary shifts are sensitive to the (φ,ψ) dihedral angles describing the conformation of a particular amino acid residue and therefore report on the secondary structure. It has been found empirically that a positive Cα secondary shift is generally indicative of a helical conformation, whilst a negative Cα secondary shift implies an extended conformation, for example as occurs in a β-strand (143).

A previous analysis of C-CA-NC used secondary shifts to determine whether the flexible linker has a propensity to adopt a particular conformation, and we conducted a similar analysis for C-CA_{AA}-NC, with Cα shifts determined from the same HNCACB experiment used for assignment purposes. The results are shown in Figure 4.8. In the C-CA_{AA} domain, regions of positive secondary shifts correspond to the helices in the C-CA dimer crystal structure (which was used to generate the secondary structure labels above the figure). Secondary shifts in the N-terminal half of helix 2 are smaller than for other helical regions, consistent with the C-CA_{AA} NMR structure and order parameters which show this region to be highly flexible, and probably therefore a dynamic ensemble of rapidly exchanging structures. Some slightly positive secondary shifts are seen in the flexible linker, consistent with those published for C-CA-NC (114). It is therefore possible that this region can transiently form a helix with the C-CA domain in monomeric and dimeric states.

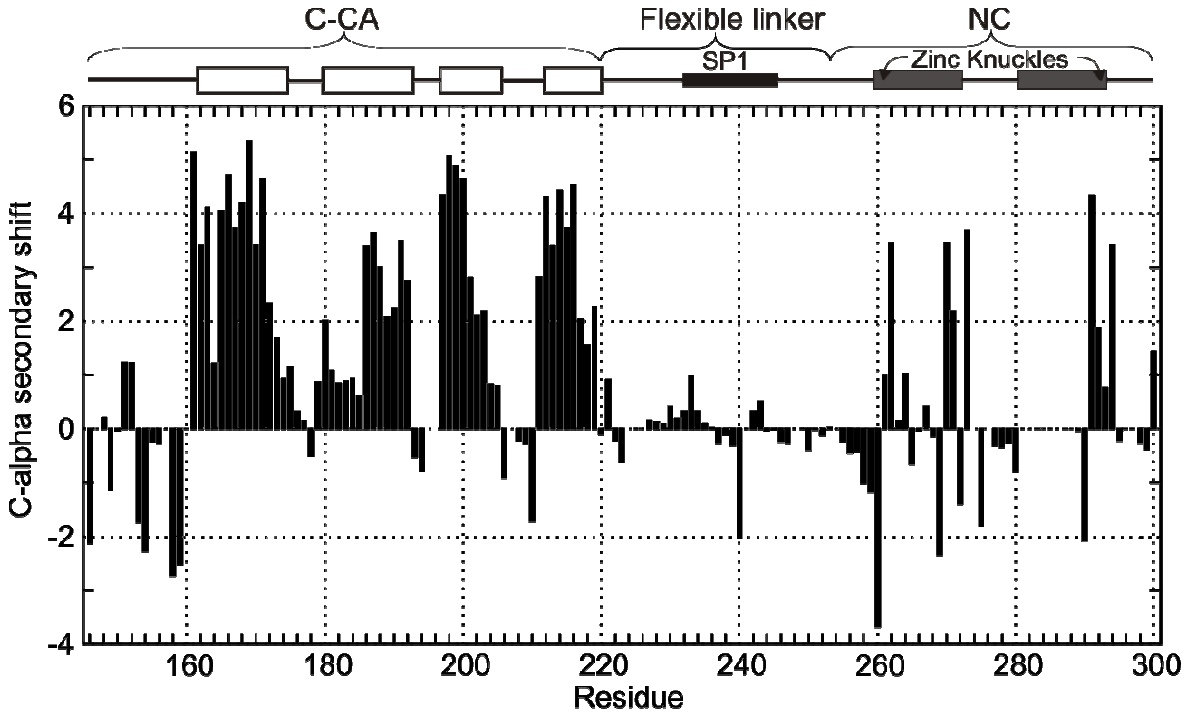


Figure 4.8:Secondary shifts measured for C-CA$_{AA}$-NC.

Secondary structural elements from individual domain structures given above.

4.4. NMR studies of C-CA-NC

C-CA-NC is a more challenging molecule to study than C-CA$_{AA}$-NC, on account of the ability of the wild-type C-CA domain to dimerise. Because the dimeric C-CA domain is considerably larger than the NC domain, peak intensities between residues in the 2 domains are very different, owing to the increased correlation time of the dimeric C-CA domain. The presence of monomer/dimer exchange adds further complication, particularly since, as will be shown, the residues actually involved in dimerisation are not visible in the NMR spectra.

4.4.1. Assignments for C-CA-NC

Out of 148 non-proline residues, assignments were made for 110, but a gap in assignments between residues 171-192 was left, the resonances for which were too weak to be observable. This is the subject of chapter 4.10. An assigned HSQC spectrum for C-CA-NC is given in Figure 4.9.

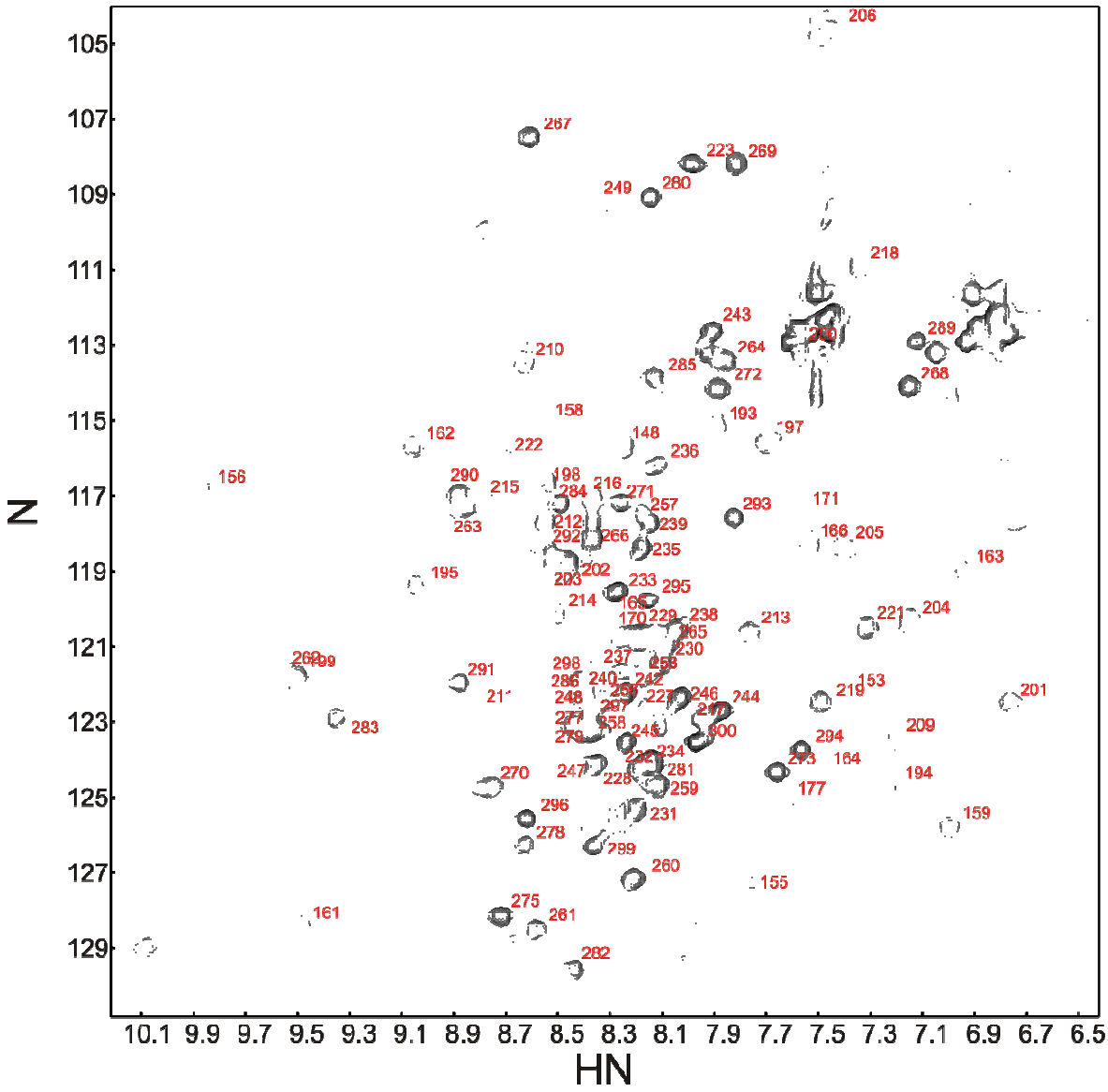


Figure 4.9: Assigned HSQC of C-CA-NC.

The spectrum was measured at 50 µM concentration, so that the C-CA peaks are of comparable intensity to the NC peaks. The spectrum was measured at 600 MHz.

4.4.2. Relaxation data for C-CA-NC

Data were interpreted using the same mono-exponential function as for C-CA$_{AA}$-NC. A total of 109 T_1, T_2 and NOE values were measured, and are shown in Figure 4.10. A qualitative consideration of the data again implies that the NC domain and C-CA domains have distinct correlation times, and that the flexible linker is fast-moving.

To gain insight into the nature of C-CA dimerisation, relaxation measurements of T_1 and T_2 (but not the NOE) were also conducted at 0.1 mM C-CA-NC concentration, using an 800

MHz spectrometer. At this concentration, the dimer population is expected to be ~75% given a K_d of 18 μM, such that the monomer population should make a noticeable contribution to the observed relaxation times. The data were interpreted using a mono-exponential function as before. It is shown in chapter 4.13 that this is not exact if monomer/dimer exchange is on a slower timescale than relaxation of either the monomer or dimer species, but is a good approximation in any case. Bi-exponential relaxation (as occurs in the case of slow exchange with small chemical shift differences) is beyond the limit of experimental resolution to detect. Some examples of relaxation curves, and the results of mono-exponential fits, are given in Figure 4.13 to show the quality of data obtained.

Relaxation data at 0.5 mM

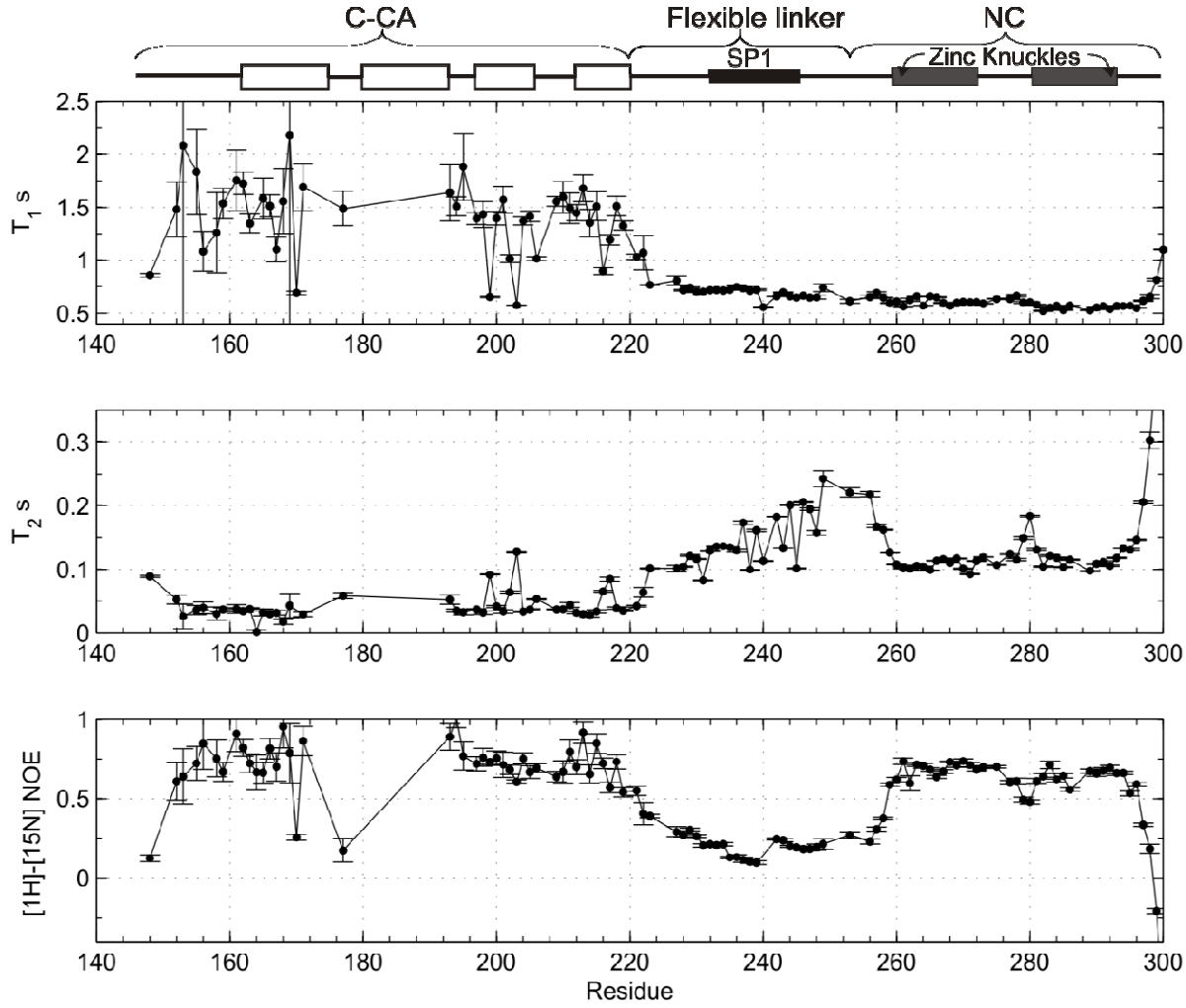


Figure 4.10: Relaxation data for C-CA-NC at 0.5 mM concentration.

Sample conditions are temperature: 25 °C, field: 700 MHz, pH 7.0Elements of secondary structure are indicated above. The T_1 times in the C-CA domain are generally poorly determined since they are often long and the peaks are often weak.

Because many T_1s in the C-CA domain were poorly determined on account of their being long and peak intensities being low, a model-free analysis was not applied to the C-CA domain (it was attempted, but failed). However, the data for the NC domain was of a high quality and amenable to further analysis. As in the case of C-CA_{AA}-NC, a diffusion tensor was determined for the NC domain, tabulated below, using the NMR average structure of a peptide fragment containing residues 257-298 in CA numbering (pdb code 1esk).

Table 4.6: The diffusion tensor for the NC domain in C-CA-NC, determined from relaxation data recorded at 500 μM concentration.

Domain	τ_m (ns)	$D_{\parallel}/D_{\perp}$	θ (°)	φ (°)
NC	5.6 ± 0.05	0.64 ± 0.05	25 ± 4	-22 ± 8

A model-free analysis was conducted on the NC domain, and 38 of the 42 residues in the NMR structure used as a structural model could be accounted for. Of these, 9 were described by model 2, 19 by model 4 and 10 by model 5. The results are shown in Figure 4.11, and are similar to those for C-CA_{AA}-NC; the NC domain is generally well ordered with a flexible section between the two zinc knuckles. Fast internal motions are on very similar timescales to those of the NC domain in C-CA_{AA}-NC, with slow motions again distributed throughout the NC domain.

A model-free analysis was also conducted on the flexible linker region, under the assumption of isotropic rotational diffusion dominated by the C-CA domain. The means of calculation of the correlation time for the C-CA domain at 500 μM are detailed in chapter 4.4.3.

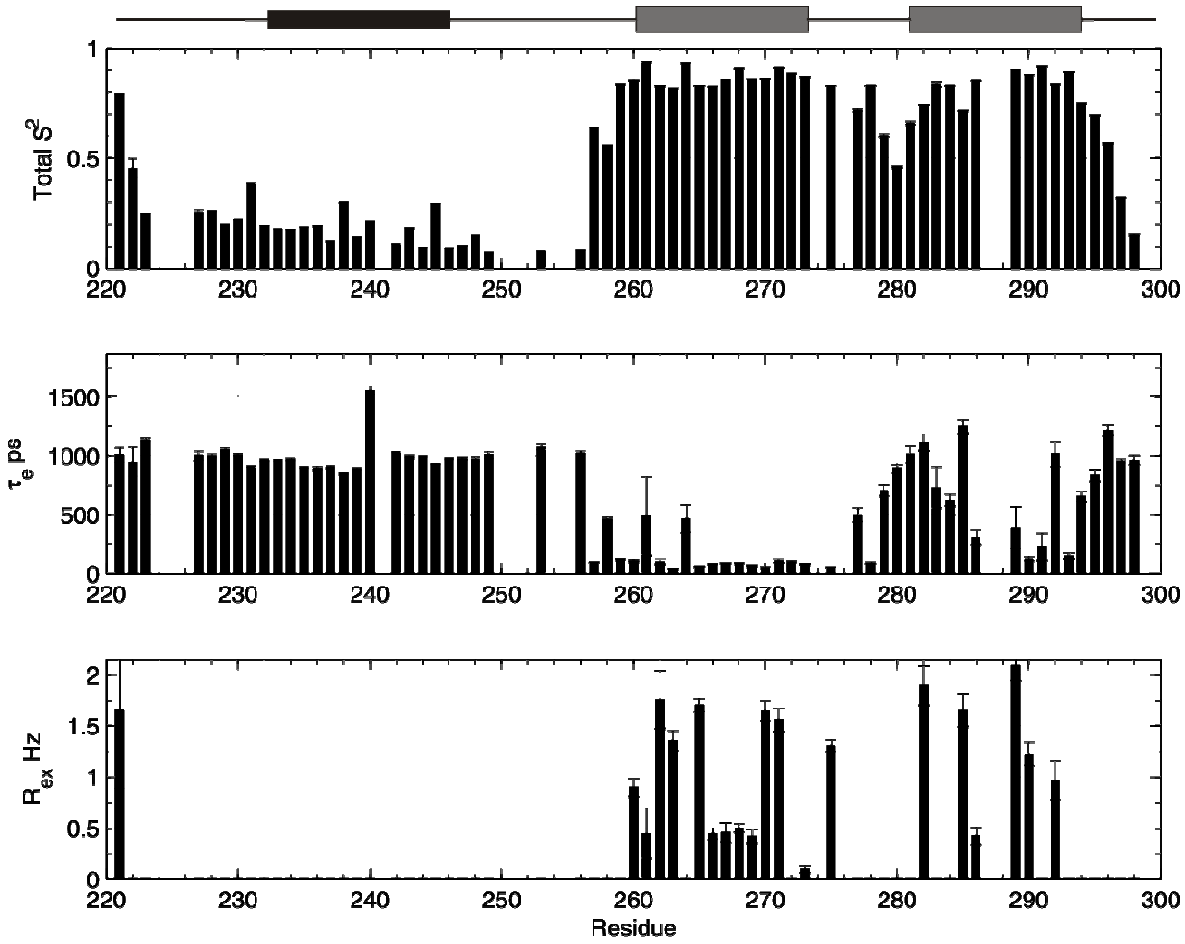


Figure 4.11: Model-free parameters for the linker and NC domain of C-CA-NC at 500 μM.

Only the relaxation data for the linker and NC domain was analysed since that for the C-CA domain was poorly determined.. The SP1 domain is indicated with a black box, and the NC zinc knuckles with grey boxes.

Relaxation data at 0.1 mM

The relaxation times inferred from mono-exponential fitting for 0.1 mM C-CA-NC at 800 MHz are given in Figure 4.12. Relaxation times were determined for 102 residues

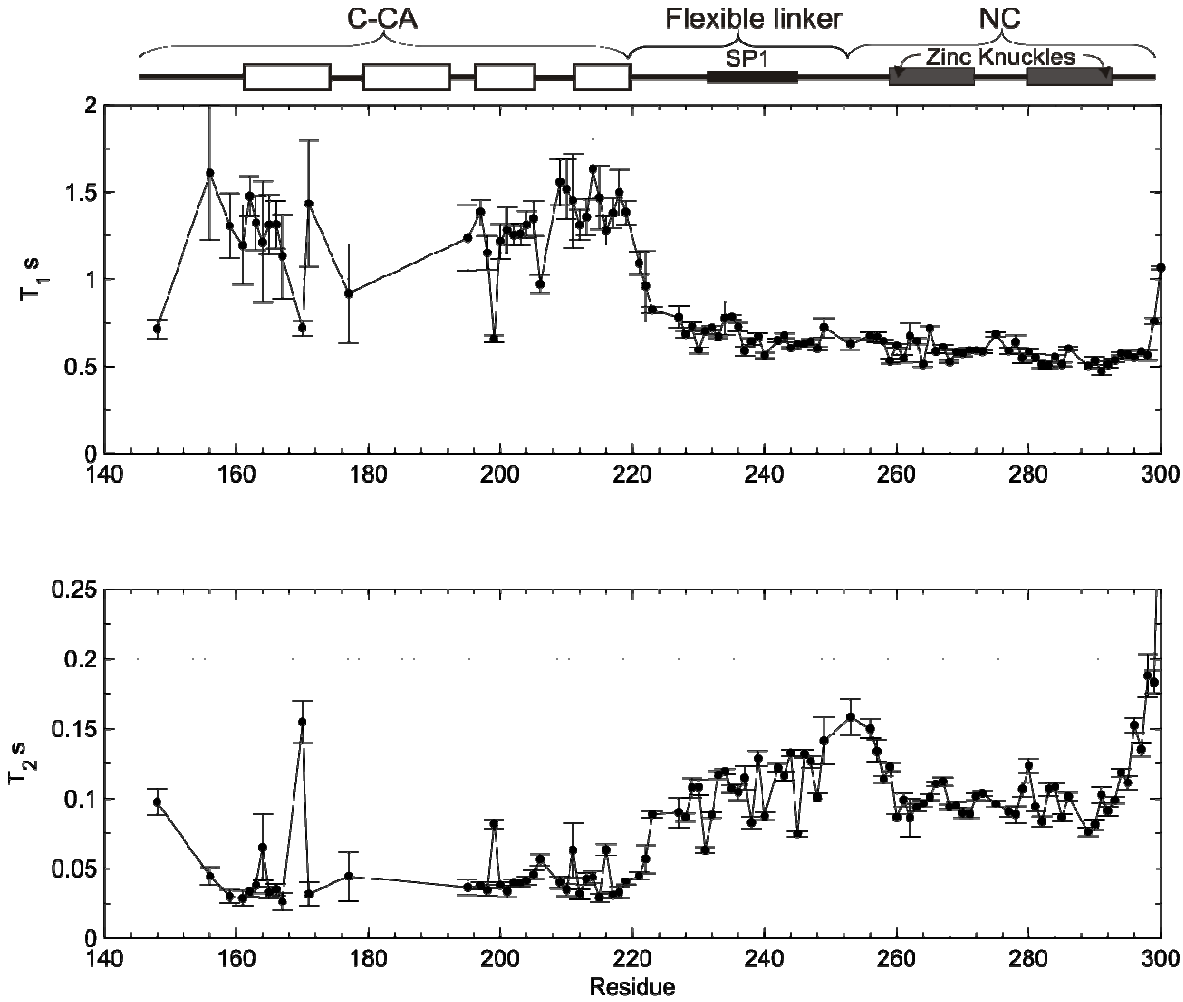


Figure 4.12: Relaxation data for 0.1 mM C-CA-NC.

Sample conditions are, temperature: 25 °C, field: 800 MHz, pH 7.0.

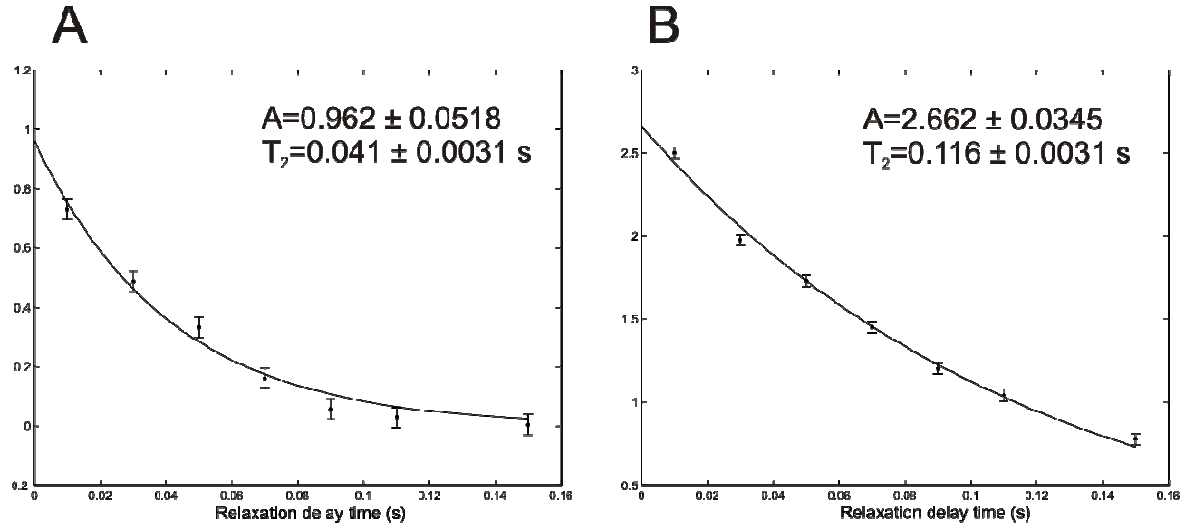


Figure 4.13: Two examples of T2 decay curves with mono-exponential fits for 0.1 mM C-CA-NC, demonstrating a long and short T2 in different regions of the molecule.

A) 204 in the C-CA domain. B) 243 in the flexible linker. The standard error is 0.034, the measured standard deviation of the noise.

4.4.3. C-CA correlation times in C-CA-NC

As a simple first-order approximation to the correlation times for the C-CA domain of C-CA-NC, it was assumed that exchange for peaks which show no measurable chemical shift change in the range of concentrations under which C-CA-NC was examined (25-500 μM) was very fast compared to intrinsic transverse relaxation. In this case a mono-exponential fit to the relaxation data yields relaxation rates which are, to a reasonable approximation, population weighted average relaxation rates. To interpret such data, we assume that the autocorrelation function is

$$C(t) = \frac{1}{5} exp\left(-\frac{t}{\tau_{obs}}\right) \qquad 4.3$$

Here, τ_{obs} is the observed or 'effective' correlation time, which for very fast exchange is assumed to be related to a population-weighted average through the relation

$$\frac{1}{\tau_{obs}} = \frac{P_M}{\tau_M} + \frac{P_D}{\tau_D} \qquad 4.4$$

Here, the subscripts '*M*' and '*D*' denote monomer and dimer respectively, whilst *P* and τ are fractional populations and correlation times respectively. The observed correlation time is therefore

$$\tau_{obs} = \frac{\tau_M \tau_D}{P_M \tau_D + P_D \tau_M} \qquad 4.5$$

Equation 2.72 then holds, from which τ_{obs} is obtained and τ_D determined by least-squares fitting of the above function (equation 4.5) using τ_{obs} for 100 μM C-CA-NC, 500 μM C-CA-NC, and for monomeric C-CA-NC the measured correlation time for the C-CA_{AA} domain of C-CA_{AA}-NC was used. Hence three measured τ_{obs} were used to estimate τ_D, using an assumed K_d for dimerisation of 18 μM, and the resulting τ_D was 20.34 ns. The measured and back-predicted correlation times are listed in Table 4.7. This estimate holds only if the monomer/dimer exchange is fast. Unfortunately, the relaxation data were not good enough to determine the frequency of exchange, for there is insufficient resolution in the relaxation decay curves to distinguish mono-exponential relaxation from bi-exponential, as will be appreciated from Figure 4.13.

Table 4.7. Isotropic correlation times for the C-CA$_{AA}$ and C-CA domains, as measured and predicted under the assumption of fast (k_{ex} > 10 Hz) monomer/dimer exchange and K_d = 18 μM.

The predicted correlation times are in reasonable agreement with the measurements and provide an estimate for the dimeric C-CA-NC correlation time. Errors were estimated using Monte-Carlo simulations.

Molecule	C-CA$_{AA}$-NC	100 μM C-CA-NC	500 μM C-CA-NC	Dimeric C-CA-NC
Observed τ_m (ns)	7.8 ± 0.02	13.6 ± 0.31	17.4 ± 0.59	n/a
Predicted τ_m (ns)	7.8 ± 0.02	14.4 ± 0.35	16.9 ± 0.57	20.3 ± 0.90

It is important that this analysis does not require an exact expression for either T_1 or T_2, nor the approximation of population-averaged relaxation times, for the averaging enters only at the level of the effective correlation time. It is also important that it is the T_1/T_2 ratio which is used to extract the correlation times, and that the T_1 alone can act as an internal control inasmuch as being a function of the correlation time. If the correlation time extracted from the T_1/T_2 ratio is inconsistent with the input T_1 values, then the T_2s used in the analysis are probably affected by an exchange process or some form of internal motion. Therefore, the predicted T_1s for each sample have been calculated from each of the correlation times, using the equations 2.63 and 2.64. The results are shown in Table 4.8, demonstrating that the correlation times inferred from equation 4.5 reproduce the observed mean T_1 values quite well. Thus it is concluded that the inferred correlation times are accurate, and not an overestimate due to exchange processes.

Table 4.8: Observed and calculated T_1 averages for the determination of the correlation time of the dimeric C-CA domain of C-CA-NC. The average T_1 values predicted from the fitted correlation times are in good agreement with the measured T_1 averages.

Molecule	C-CA$_{AA}$-NC	100 μM C-CA-NC	500 μM C-CA-NC
Observed mean T_1 (s)	0.643 ± 0.038	1.327 ± 0.160	1.446 ± 0.297
Predicted mean T_1 (s)	0.581 ± 0.001	1.382 ± 0.030	1.477 ± 0.049

Comparison of relaxation data at 100 μM and 500 μM

Because the relaxation experiments for 100 µm and 500 µM C-CA-NC were conducted at different fields, it is difficult to compare the data directly. However, insight into the dynamics can be obtained by plotting the T_2 against the T_1 for each sample, and plotting the expected T_1 and T_2 for different correlation times and order parameters in the absence of fast internal motion or exchange. This is shown in Figure 4.14. In this figure, the data have been coloured according to domain. In each case, the area around which the data cluster is indicative of the correlation time for that domain. Residues lying to the left of the S^2=1 curve have T_2s which are lower than is predicted by Redfield theory, and so contributions to relaxation due to exchange processes are present. Residues which are to the right of the main cluster of data for a particular domain have fast internal motion, since the T_2 is longer than the correlation time alone would allow. Such conclusions hold at both 700 MHz and 800 MHz, such that the figure allows comparison of the two datasets.
It is clear in both cases plotted that the C-CA domain has a different timescale of overall motion than the NC domain, and that many residues in the C-CA domain are affected by exchange processes. It seems also that the extent to which exchange contributes to C-CA relaxation is the same at both concentrations used.

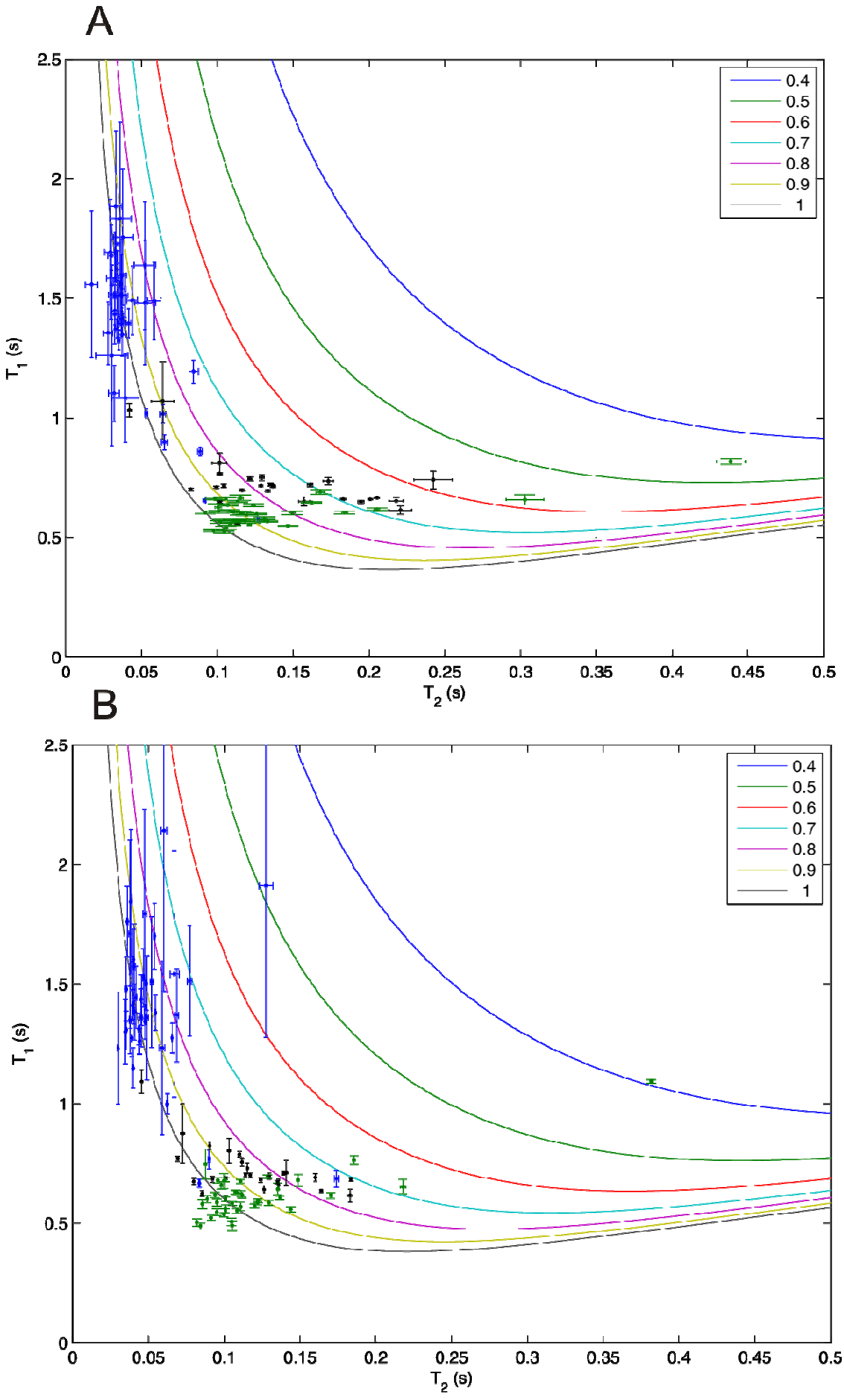

A
B
0.4
0.5
0.6
0.7
0.8
0.9
1
T1 (s)
T2 (s)
0
0.5
1
1.5
2
2.5
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Figure 4.14. Plots of T1 against T2 for A) 500 µM C-CA-NC and B) 100 µM C-CA-NC.

Data for C-CA domain is in blue, for the flexible linker in black and for NC in green. Each smooth curve represents the expected T_1 and T_2 predicted in the absence of fast internal motion or exchange, given different values of the order parameter as listed in the legend.

4.4.4. Effects of concentration on C-CA-NC chemical shifts

Conventionally, if a protein dimerises or binds a ligand, a titration can be used to identify residues which change chemical shift or change in extent of line-broadening. Such shift changes localise the binding site, provided assignments are available. Therefore, to try and determine residues involved in C-CA-NC dimerisation, a series of NH HSQC spectra were obtained, from 50µM to 500 µM. However, chemical shift changes could not be found, except a few small changes. At low concentration, a small number of additional peaks became visible, but these could not be assigned (due to absence of triple resonance data for very low concentration samples, such experiments being impractical at low concentration). Two overlaid spectra for C-CA-NC are shown in Figure 4.15, which demonstrate the invariance of chemical shifts in the C-CA domain to concentration. From this, it is concluded that C-CA dimerisation does not affect chemical shifts of residues distal to the dimer interface itself. Such data also have the important consequence that the absence of peaks at the dimer interface is not caused by monomer/dimer exchange alone, a point which will be returned to and discussed detail in chapters 4.6 and 4.10.

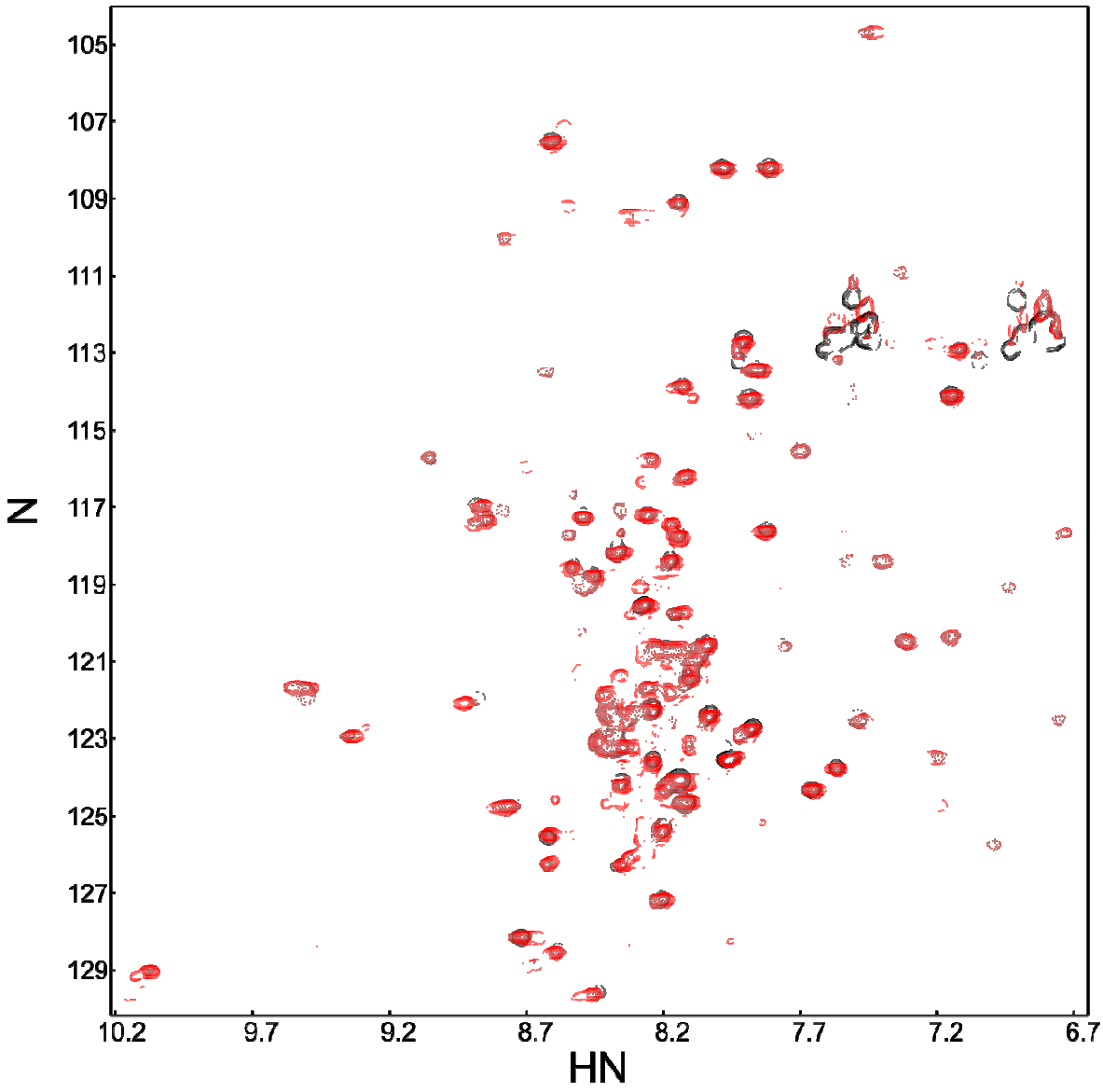


Figure 4.15: Overlaid HSQC spectra of C-CA-NC.

Red: 500 µM, black: 50 µM. Both spectra were recorded at 600 MHz, 25 °C, pH 7.0, and both have 128 complex points in the indirect dimension with 64 additional points derived by linear prediction. The spectra have been normalised so that the C-CA domain peaks have similar intensities. As a result, flexible region and NC peaks in the 500 µM spectrum have higher intensities, since the relative C-CA intensities are reduced by its larger correlation time at high concentration. Assignments are omitted for clarity but can be found in Figure 4.9.

4.5. NMR studies of C-CA$_{AA}$

4.5.1. Relaxation analysis of C-CA$_{AA}$

The data presented in the following section were acquired by Dr Joern Werner and Dr Jana Sticht, and analysed by myself. Details of the protocols for expression, purification and NMR assignments can be found elsewhere (145).

Relaxation data for C-CA$_{AA}$ is presented in Figure 4.16. The NOEs have generally quite large errors, since exchange of water protons with amide protons made accurate measurements difficult.

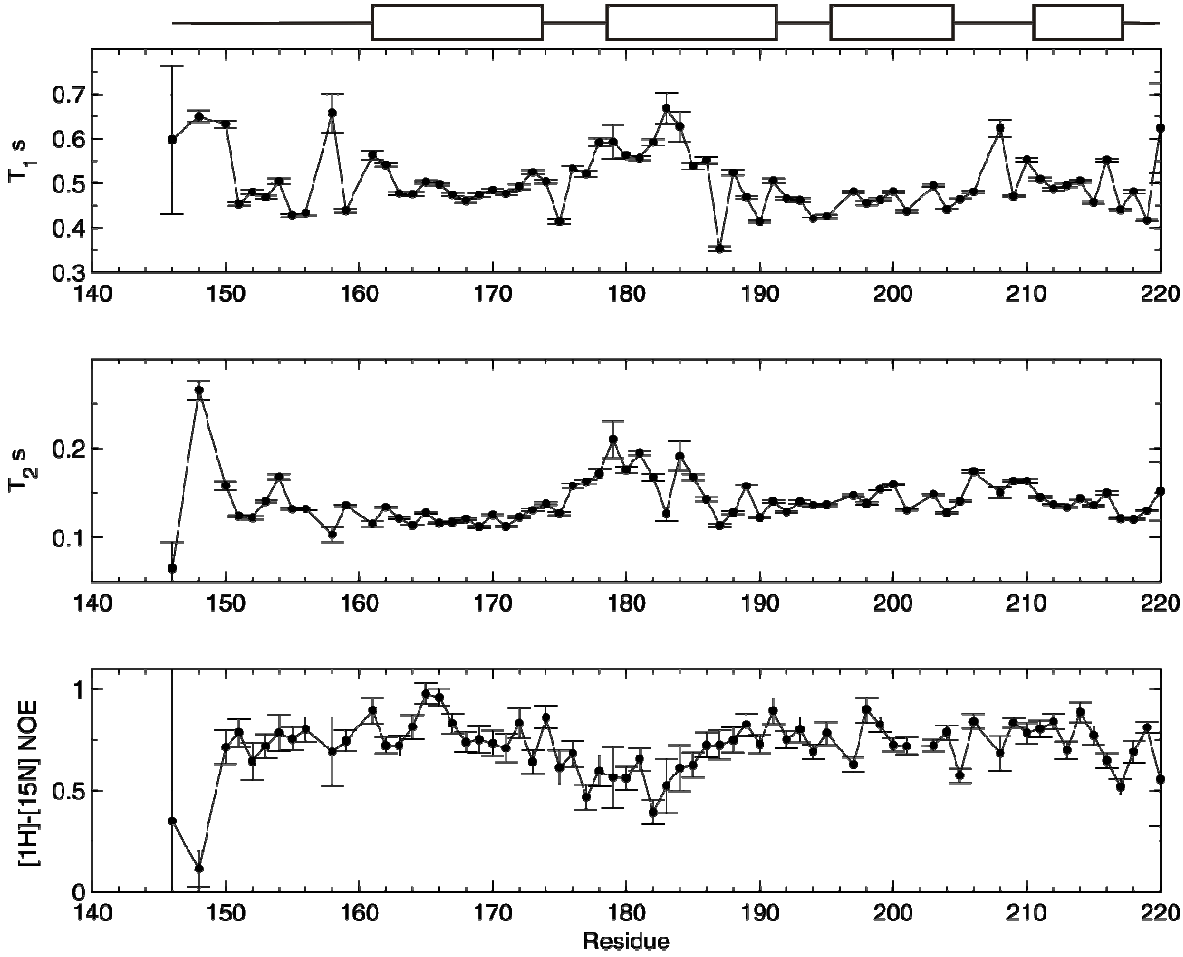


Figure 4.16: Relaxation data for C-CA$_{AA}$.

Sample conditions are concentration: 1 mM, temperature: 25 °C, field: 600 MHz, pH 7.0

A model-free analysis was performed, using the X-ray structure of C-CA as a structural model in diffusion tensor determination, in the same frame as used for C-CA$_{AA}$-NC analysis so as to make diffusion tensor orientations directly comparable. The diffusion

tensor and model-free results are given in Table 4.9 and Figure 4.17 respectively. The analysis was applied to 64 residues, of which 16 were accounted for by model 1, 14 by model 2, 12 by model 3, 6 by model 4 and 6 by model5.

By reference to Table 4.4 and Figure 4.4, it is seen that the diffusion tensors for isolated C-CA_{AA} and the C-CA_{AA} domain of C-CA_{AA}-NC are remarkably consistent, with the exception of τ_m, which differs presumably on account of the increased total size of C-CA_{AA}-NC. The internal dynamics are also very consistent between C-CA_{AA} and the corresponding domain in C-CA_{AA}-NC, as will be demonstrated clearly in chapter 4.12. Therefore, C-CA_{AA} is a highly dynamic protein, flexible on several timescales. There is fast motion on a timescale of nanoseconds distributed throughout the molecule, and also slow motion on a millisecond timescale. The order parameters demonstrate that helices 1 and 4 are well folded, and that much of the N-terminal region leading to helix 1 is also quite rigid. However, the loop residues between helices 1 and 2, along with most of helix 2 itself, are more flexible, as was also the case in C-CA_{AA}-NC. The loop between helices 3 and 4 is also flexible.

Table 4.9: The diffusion tensor for C-CA_{AA}.

τ_m (ns)	$D_{\parallel}/D_{\perp}$	θ (°)	φ (°)
5.0 ± 0.01	1.38 ± 0.02	22 ± 2	-35 ± 5

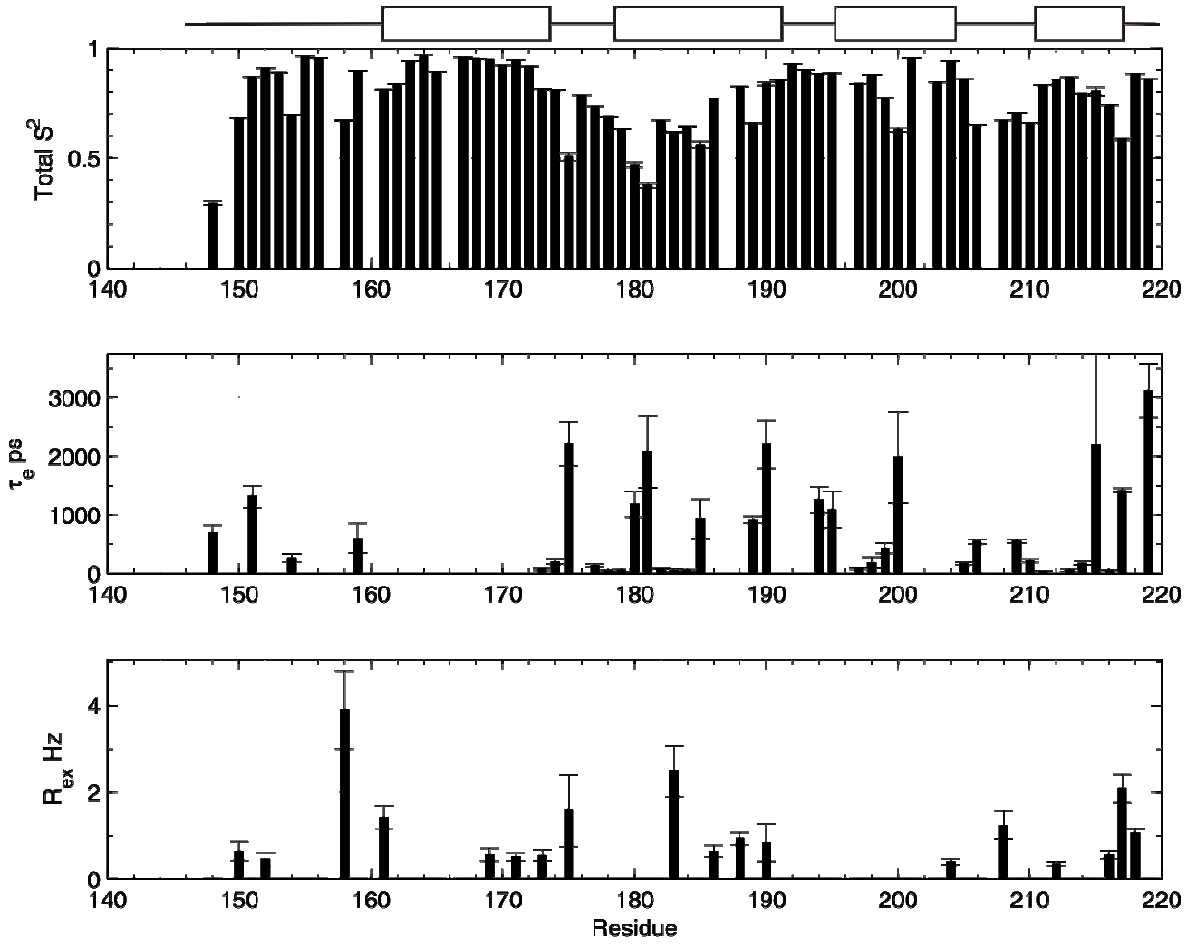


Figure 4.17: The results of a Lipari-Szabo model-free analysis of C-CA$_{AA}$.

A total of 64 residues were included in the analysis.

4.6. NMR studies of C-CA

4.6.1. Experiments and assignments

All experiments on C-CA were carried out by Joern Werner, but the theoretical treatments and analyses presented here were conducted by myself. Relaxation experiments were conducted for 2 C-CA samples, the first at 0.1 mM and the second at 1 mM, since dynamics are likely to depend upon the fraction of protein which is in a dimeric state.

4.6.2. Interpreting relaxation data

For the data recorded at 1 mM concentration, relaxation times were determined using a mono-exponential function. Using the measured K_d of 18 μM, at 1 mM, the dimer population is approximately 91%, and it was assumed that the observed "effective" relaxation times from the mono-exponential fits were representative of the dimer. For the

100 μM data, relaxation times obtained by mono-exponential fits were assumed to be population-weighted average relaxation times. Computer simulations using the theory of exchange show that this is a reasonable approximation to make, provided $k_{ex} > R_2$. To determine a diffusion tensor, the C-CA(146-231) dimer from the X-ray crystal structure was used. For the 100 μM data, an isotropic diffusion tensor was determined, since anything more complex would be difficult to extract from the data.

An implicit assumption made in the determination of relaxation times using a mono-exponential fit is that the unique axis of the elliptic diffusion tensor in its principal frame is coincident with the C_2 axis of symmetry present in the dimer (158). This is because relaxation times for amide ^{15}N nuclei depend upon the angle α between the NH bond vector and the unique axis of the diffusion tensor. If the angle made by a particular NH vector in one monomer of the dimer with the diffusion tensor unique axis differs from that made by the symmetry-related mate, then the symmetry-related spins in each monomer of the dimer will have different relaxation times, in which case relaxation is bi-exponential. This means that if a molecular frame is defined in which the C_2 axis is coincident with the molecular z-axis, then no rotation need be effected to orient the diffusion tensor, such that $\theta = \varphi = 0$ (where the two angles define the orientation of the principle frame of the diffusion tensor relative to the molecular frame) and a two-parameter model is sufficient for the diffusion tensor, under the assumption of axial symmetry, the parameters being D_r and τ_m. The diffusion tensors are given in Table 4.10.

Table 4.10: Diffusion tensors for C-CA as an isolated domain.

An axial diffusion tensor was calculated for 1 mM C-CA, whereas an isotropic model was used for 0.1 mM C-CA.

Concentration	τ_m (ns)	$D_{\parallel}/D_{\perp}$
0.1 mM	8.77 ± 0.18	n/a
1 mM	9.80 ± 0.02	1.16 ± 0.02

4.6.3. Internal dynamics of C-CA at 1 mM

In figure 4.18, the measured relaxation times for C-CA are compared with those in which the model 1-3 forms of the Lipari-Szabo spectral density function have been fitted to the measured data (i.e. an order parameter is calculated, and if necessary either a τ_e or R_{ex}, but

not both). NOEs were not successfully measured for C-CA, and are at any rate very difficult to interpret in the presence of exchange, so it was not possible to use either model 4 or 5, which are underdetermined in the absence of NOEs. In Figure 4.19, the spectral density parameters obtained from the model 1-3 spectral density functions are shown. Out of the 43 residues included in the analysis, 18 could be accounted for by model 1, 9 by model 2 and 16 by model 3. In Figure 4.20, the C-CA dimer structure is coloured according to which model is used to describe the dynamics for each residue.

Such results demonstrate that there is considerable exchange broadening for residues in helix 1, and also in the loop between helices 3 and 4, in the range 2-5 Hz. However, these regions are, for the most part, distal to the dimer interface, and the chemical shifts for residues in this region are the same over a broad range of C-CA concentrations (25 – 1000 μM) as shown in the forthcoming chapter 4.6.4. This means that the observed exchange broadening is certainly not due simply to exchange between monomeric and dimeric states. If this were the case, then the exchange regime would be a function of the protein concentration, as shown by the calculation in chapter 4.14. Hence the lineshape and chemical shift would be functions of protein concentration, which is contrary to observation. Such exchange broadening is presumably therefore due to an inherent flexibility of the C-CA dimer, or of C-CA whether or not it is dimeric. In chapter 4.10, lineshape simulations are presented to make this clear.

This supposed flexibility seems not to be present in C-CA$_{AA}$ mutant, which remains monomeric. This is on the basis that R_{ex} terms are found for fewer residues in this region, and those which are found are lower than those for dimeric C-CA. It is possible that both monomeric C-CA$_{AA}$ and dimeric C-CA undergo similar conformational changes, that is, explore similar conformational spaces, but on different timescales. There is fast motion on a timescale of ~ 1 ns throughout almost the whole of C-CA$_{AA}$, and dimerisation may have simply reduced this timescale of motion to the ms scale. This is speculative, however, and the possibility that the faster timescale of motion in C-CA$_{AA}$ is due to the alanine mutations, rather than its being monomeric, is ever present.

Because of the absence of NOEs, it is difficult to accurately quantify fast internal motions, nor justifiably incorporate them into a description of the dynamics for a particular residue, for the NOE is more sensitive to such motions than the relaxation times. It is perhaps because of this that many fitted relaxation times, particularly in helices 3 and 4, are below

their measured values. Accurately determining fast motion timescales necessary to account for such data will require NOEs, but it is reasonable to assume that there is more fast motion in helices 3 and 4 than has been detected in the analyses presented here.

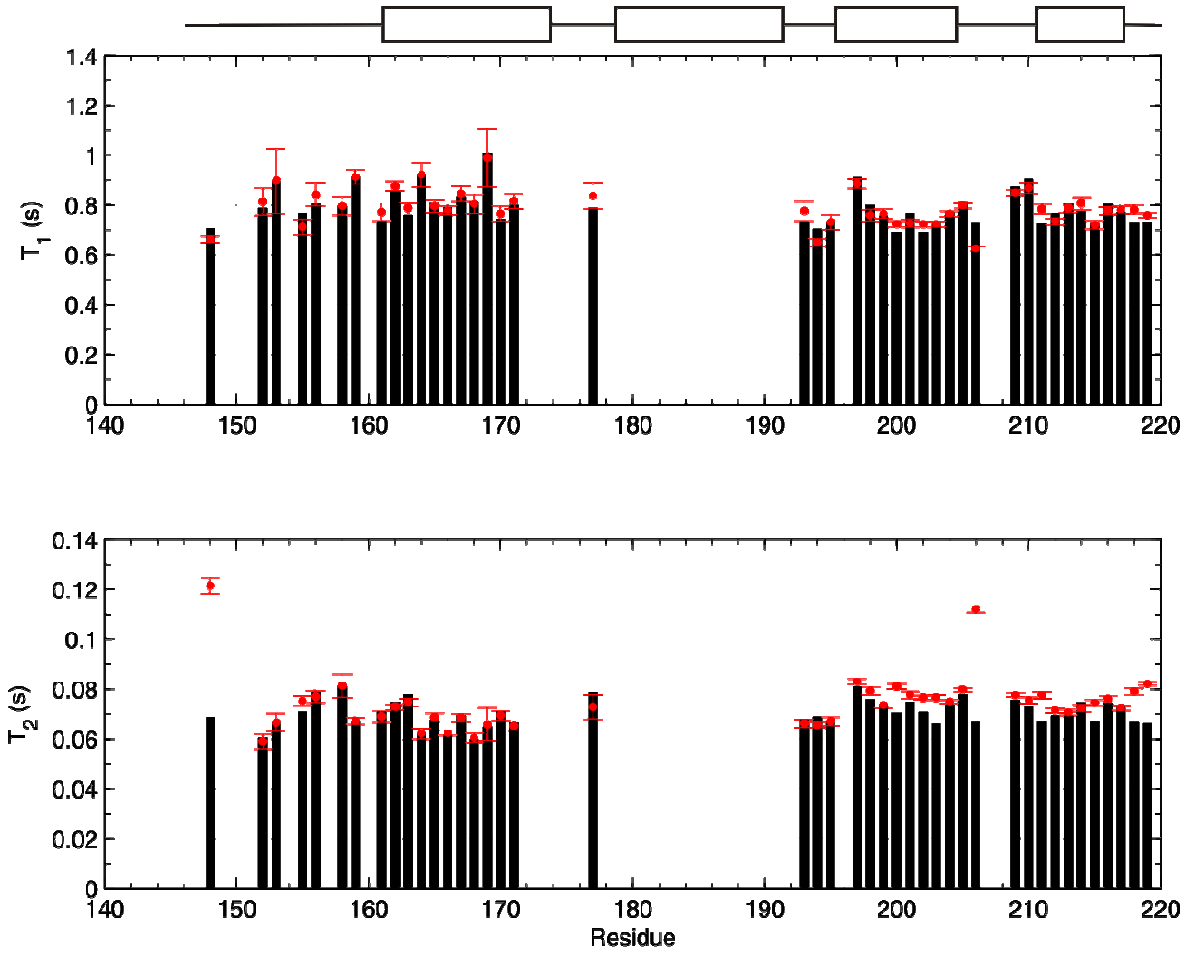


Figure 4.18: Measured relaxation times (red) and those predicted from a model-free analysis for C-CA at 1 mM.

No NOEs were used, so only models 1-3 could be fitted to the data, since models 4 and 5 require more free parameters than there are measured parameters. C-CA helices are indicated with boxes.

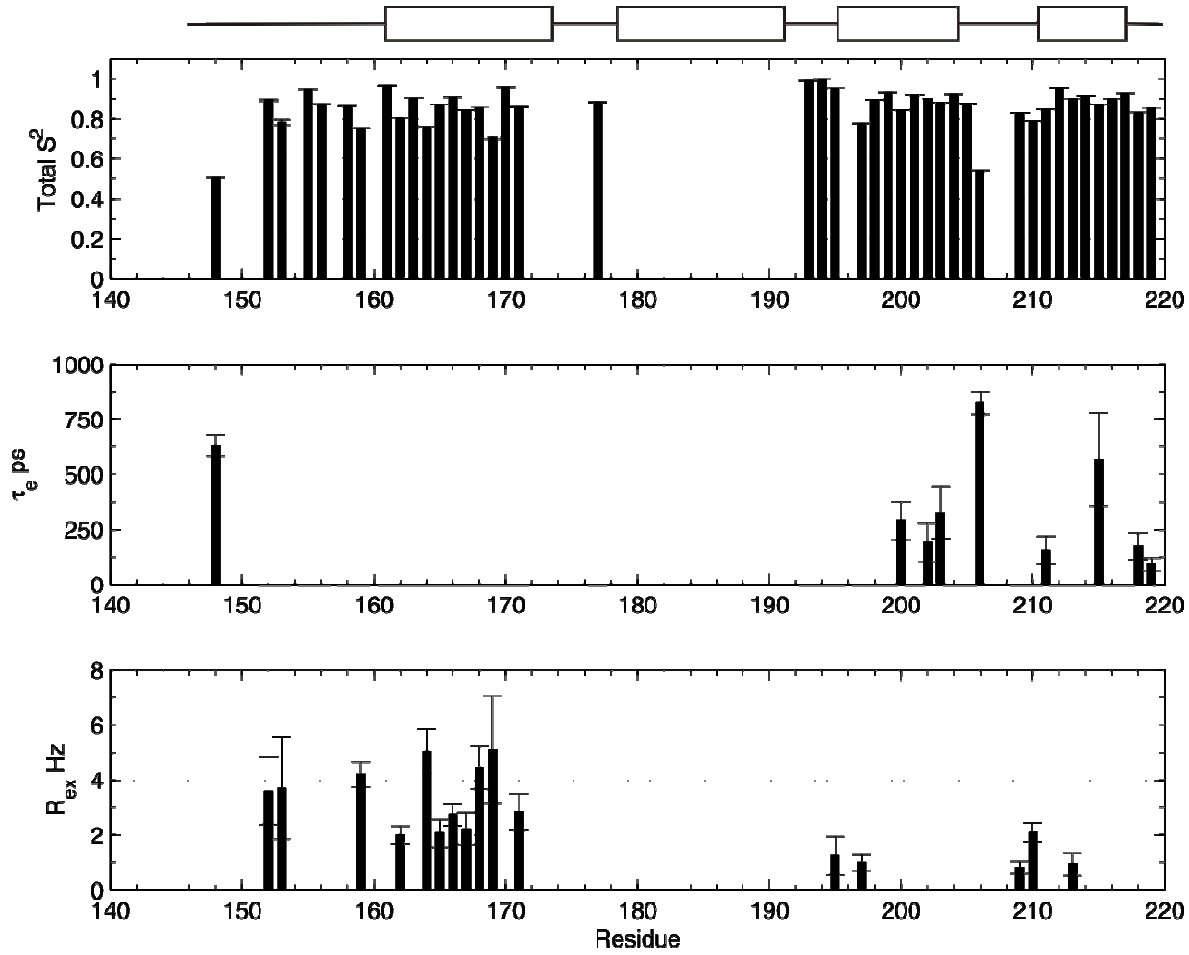


Figure 4.19: Spectral density parameters for C-CA at 1 mM concentration, obtained by fitting the model 1-3 forms of the Lipari-Szabo spectral density function to measured relaxation times.

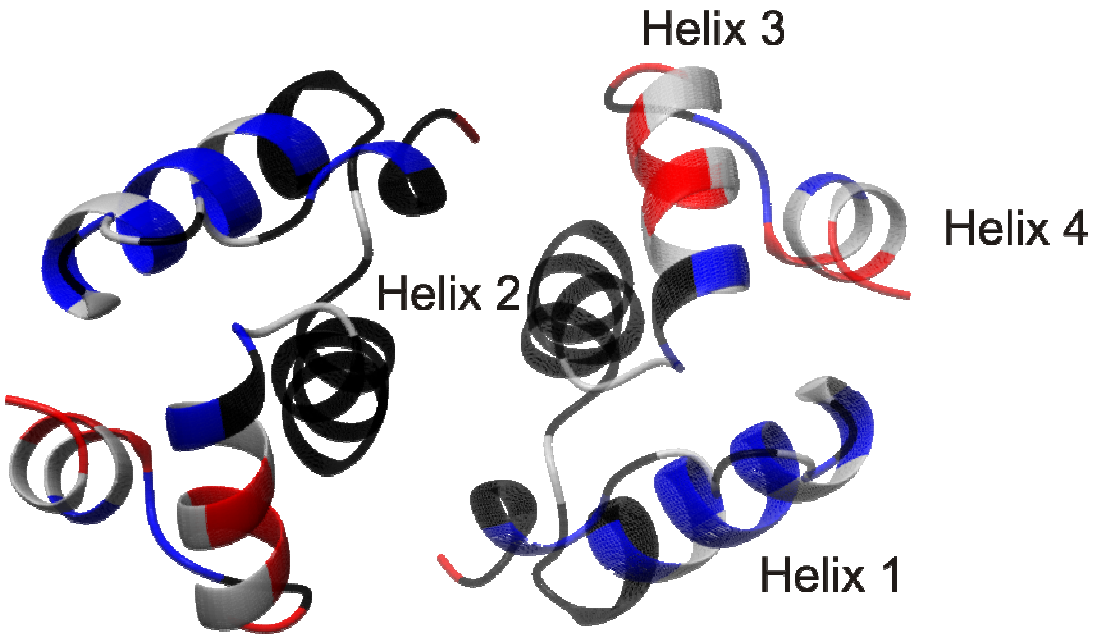


Figure 4.20: Ribbon diagram of the X-ray crystal structure of the C-CA dimer, coloured according to the results of a model-free analysis.

Grey: model 1, red: model 2, blue: model 3, black: not analysed.

4.6.4. Concentration dependence of C-CA chemical shifts

As in the case of C-CA-NC, chemical shift changes were not detectable for C-CA over a broad concentration range. Overlaid HSQC spectra in Figure 4.21 demonstrate this. The assignments are not shown but have been published (145).

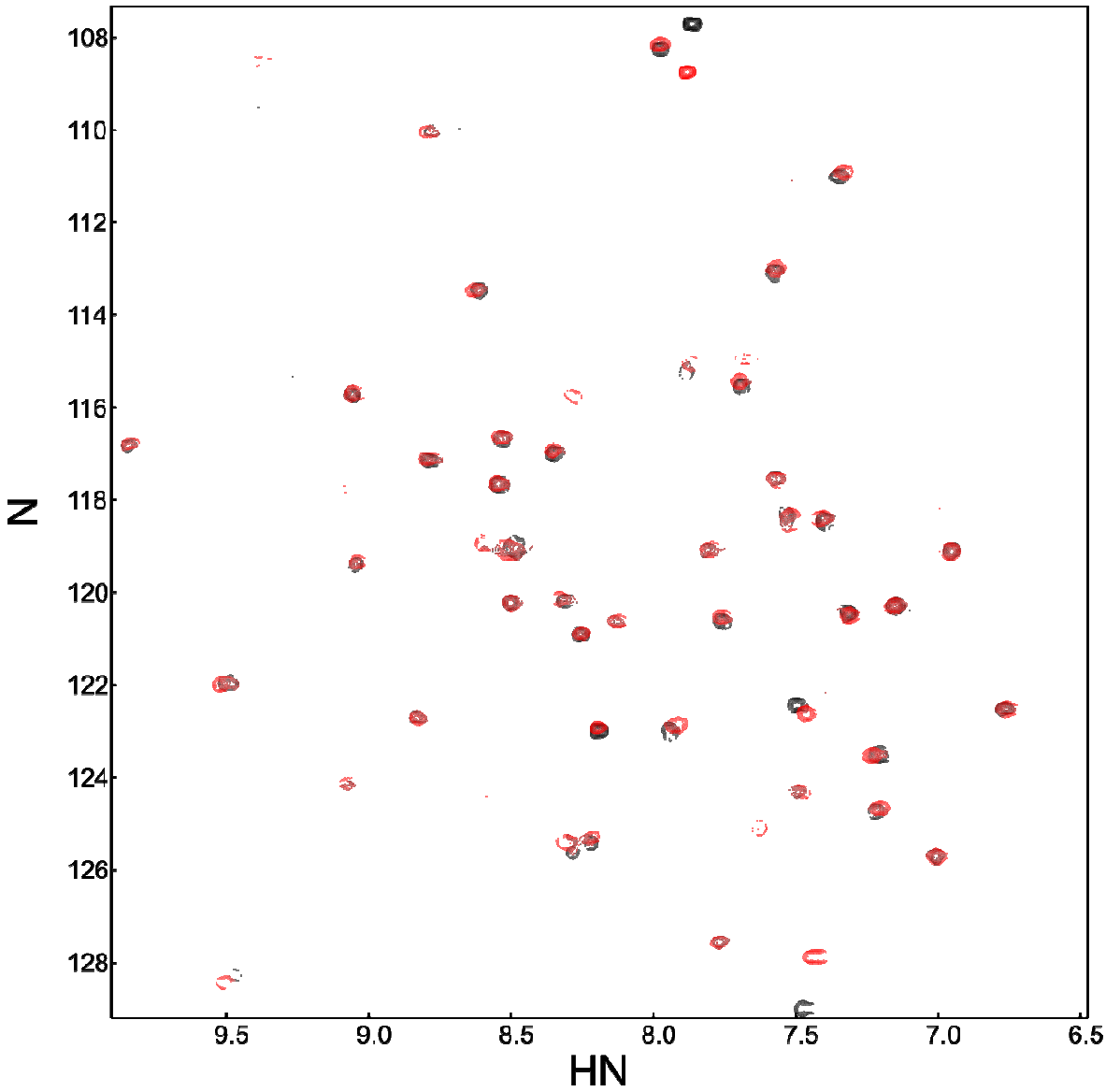


Figure 4.21: Concentration dependence of C-CA chemical shifts.

Black: 1 mM, red: 0.1 mM, both recorded at 600 MHz, 25 °C, pH 7.0.

4.7. NMR studies of C-CA$_{AA}$-NC with CAI

CAI is a 12-residue peptide which binds the C-CA domain and inhibits both immature and mature assembly, as described in detail in chapter 1.9. It is noteworthy that a higher concentration of CAI is required to inhibit immature assembly than mature assembly, and this chapter is dedicated to NMR experiments which seek to establish why. Previous studies have examined interactions between the C-CA$_{AA}$ domain and CAI, without the flexible region and NC domain present. The analysis here is extended to such regions, in order to determine whether they have any modulating effect upon C-CA/CAI interactions. It is found that the binding site for CAI is the same in C-CA$_{AA}$-NC as for C-CA$_{AA}$ alone, and that flexible linker and NC domains are unperturbed by its presence.

4.7.1. Assignments and chemical shift changes

In total, assignments were made for 121 out of 148 non-proline residues. Peak intensities near the known CAI binding site were generally weak, and between residues 183-190 no assignments were made. This is probably due to exchange between bound and free states of C-CA$_{AA}$-NC, since the concentrations of C-CA$_{AA}$-NC and CAI used in the experiments would give a bound population of ~80% given the K_d measured for C-CA$_{AA}$ of 15 μM (146). An HSQC spectrum, showing amide NH assignments, is provided in Figure 4.22. For the assigned residues, chemical shift changes could be determined in both the ^{1}H and ^{15}N dimensions. Such results are given in Figure 4.23, and are consistent with the reported CAI binding site. No second site for CAI was identified, based upon the shift changes.

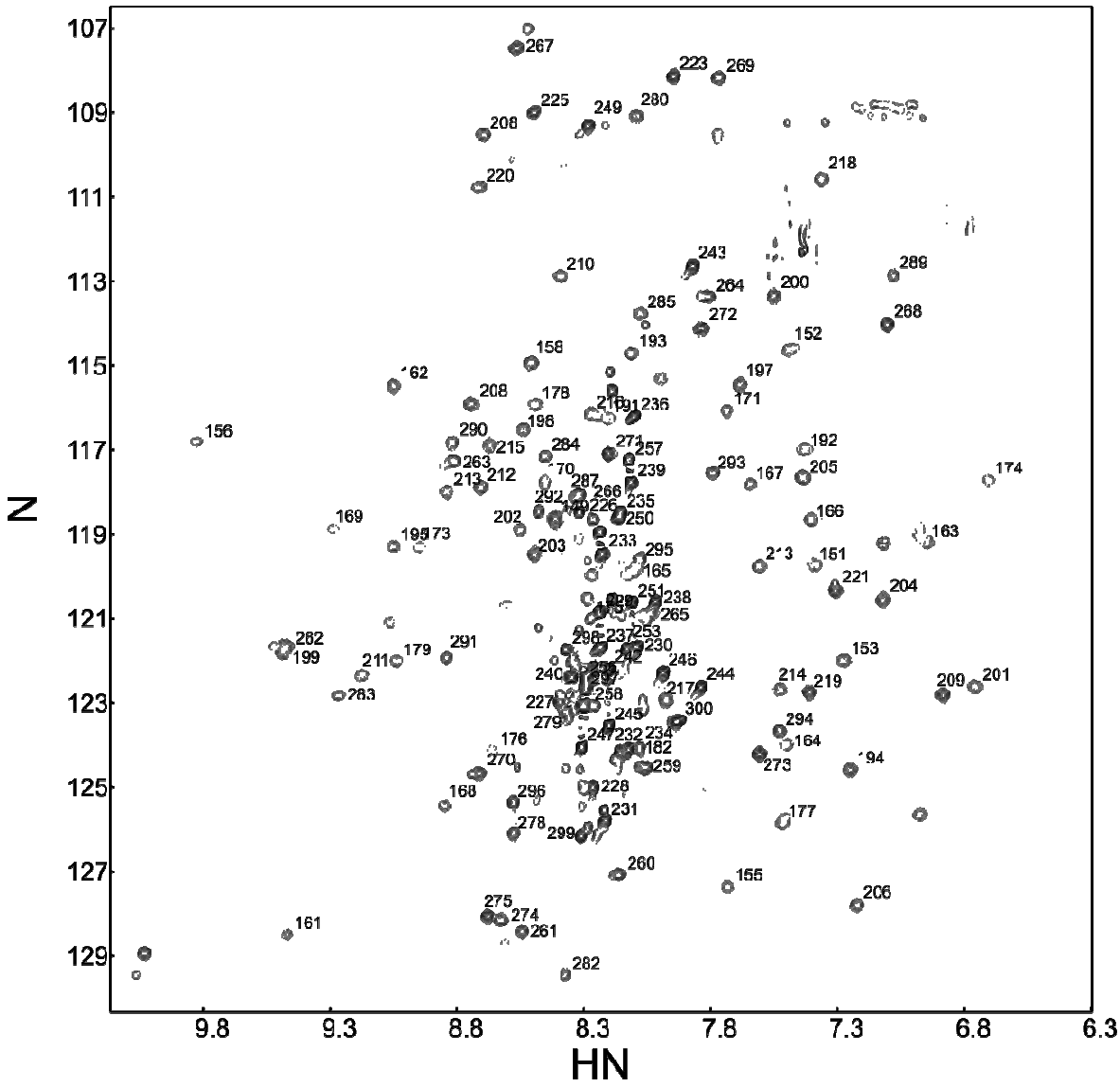


Figure 4.22: Assigned HSQC spectrum of C-CA$_{AA}$-NC in complex with CAI.

Sample conditions are C-CA$_{AA}$-NC concentration: 500 µM CAI concentration 550 µM, temperature: 25 °C, field: 600 MHz, pH 7.0.

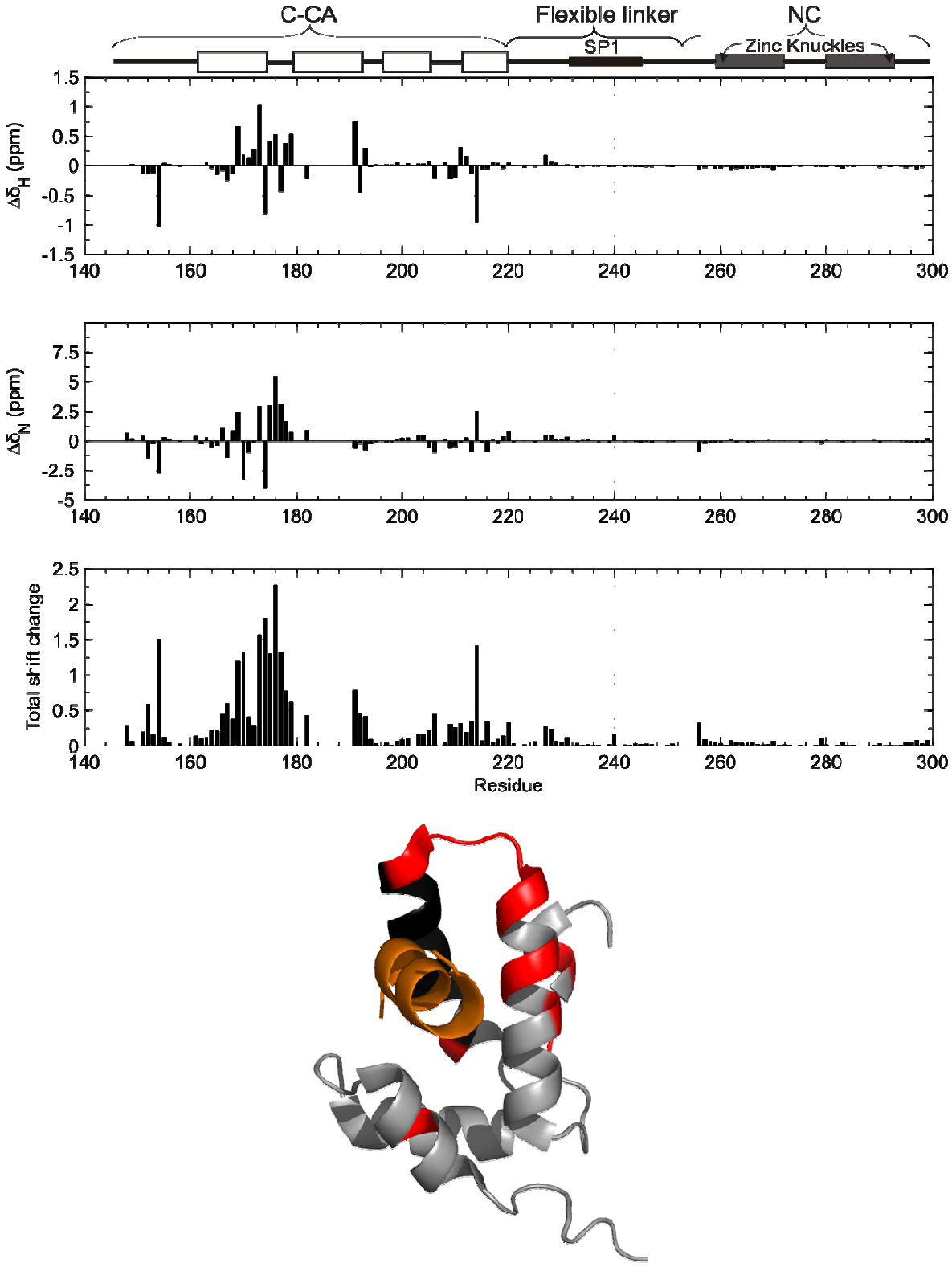


Figure 4.23: Chemical shift changes between free and CAI-bound C-CA$_{AA}$-NC for backbone amide positions.

Only where mutual assignments were made is such an analysis possible, thus excluding all but one residue in the region from 179-190 and several others throughout the molecule.

117 chemical shift comparisons are shown in this figure, since 4 of the 121 residues assigned for the CAI complex were not assigned in free C-CA$_{AA}$-NC. The CAI binding site is restricted to residues in the C-CA$_{AA}$ domain. In the ribbon diagram, the X-ray structure of the C-CA/CAI complex is shown and residues with total shift changes > 0.5 are shown in red. Black indicates the unassigned region.

4.7.2. Relaxation measurements

The ^{15}N T_1, T_2 and NOEs for the C-CA$_{AA}$-NC/CAI complex were measured under the same conditions as for assignment experiments, using the same methods as for unliganded C-CA$_{AA}$-NC. Such data are plotted in Figure 4.24.

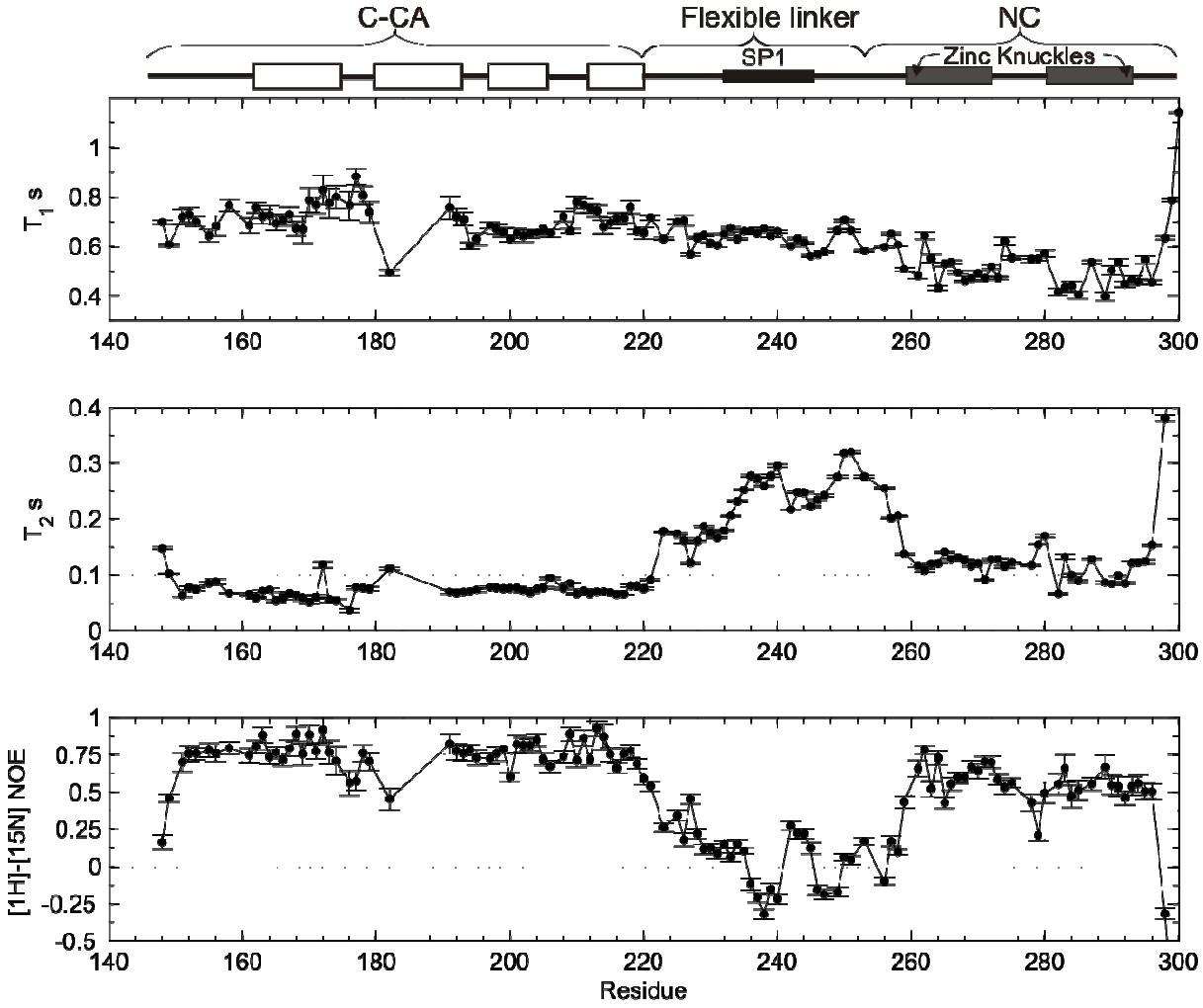


Figure 4.24: ^{15}N T_1 and T_2 data and [^{1}H]-^{15}N NOEs for C-CA$_{AA}$-NC + CAI.

Sample conditions are C-CA$_{AA}$-NC concentration: 500 µM, CAI concentration 550 µM, temperature: 25 °C, field: 600 MHz, pH 7.0. Elements of secondary structure are indicated

by boxes above the figure. Of 148 non-proline residues, 121 were assignable, and relaxation parameters were obtained and plotted for all of these.

Despite the gaps in assignments, there was sufficient relaxation data to proceed with a model-free analysis to determine the timescales of internal motion and degree of flexibility in the C-CA_{AA}-NC/CAI complex. It is clear from Figure 4.24 that the overall motion of the C-CA_{AA} domain is distinct from that of the NC domain, and that the flexible linker again lives up to its name. Therefore, separate diffusion tensors were again used for each domain. For the C-CA_{AA} domain, the X-ray structure of wild-type C-CA in complex with CAI was used as a structural model (pdb 2bu0), whilst for the NC domain the NMR average structure of NC(257-298) was used, the same as when unliganded C-CA_{AA}-NC was analysed. For the flexible linker, as a first approximation the overall motion was treated as being dominated by the C-CA_{AA} domain, and an isotropic diffusion tensor was used, equal to the isotropic correlation time for the C-CA_{AA}/CAI domain. The fitted diffusion tensors are given in Table 4.11. The diffusion tensor for the NC domain is very similar to that for the NC domain in unliganded C-CA_{AA}-NC, with a slightly increased correlation time. For the C-CA_{AA} domain, overall motion is slower than in the absence of CAI, which is to be expected given the larger size of the C-CA_{AA}/CAI complex. The ellipticity and orientation of the C-CA_{AA}/CAI diffusion tensor is in good agreement with that of the C-CA_{AA} domain in unliganded C-CA_{AA}-NC.

Table 4.11: *Diffusion tensors for C-CA_{AA}-NC + CAI*

Domain	τ_m (ns)	$D_{\parallel}/D_{\perp}$	θ (°)	φ (°)
C-CA_{AA}/CAI	9.3 ± 0.04	1.5 ± 0.04	14 ± 3	2 ± 15
Linker(220-256)	9.3 ± 0.04	N/A	N/A	N/A
NC	5.5 ± 0.08	0.60 ± 0.05	55 ± 3	11 ± 4

In the model-free analysis of the C-CA_{AA}/CAI domain, 54 residues were analysed. Of these, 16 were described by model 1, 15 by model 2, 9 by model 3, 10 by model 4 and 4 by model 5. Thus 29 residues have internal motions on a ps-ns timescale of the 54 analysed, and 19 have slower motions on a µs-ms timescale. Most of the fast motions were on a timescale of 1-3 ns, but a subset of residues instead had fast motions on a scale of ~100 ps. The order parameters are generally above 0.9, with exceptions in known coil regions, such that the C-CA_{AA}/CAI complex is a well-folded domain.

For the NC domain, 35 residues were described, of which none were accounted for by model 1, 11 by model 2, 1 by model 3, 18 by model 4 and 5 by model 5. Thus nearly all residues in the NC domain have fast internal motions, and around two thirds also have slow internal motion.

For the linker region, 30 residues were analysed, of which1 residue (220) could be described by model 2, whereas the remainder were described by model 5.

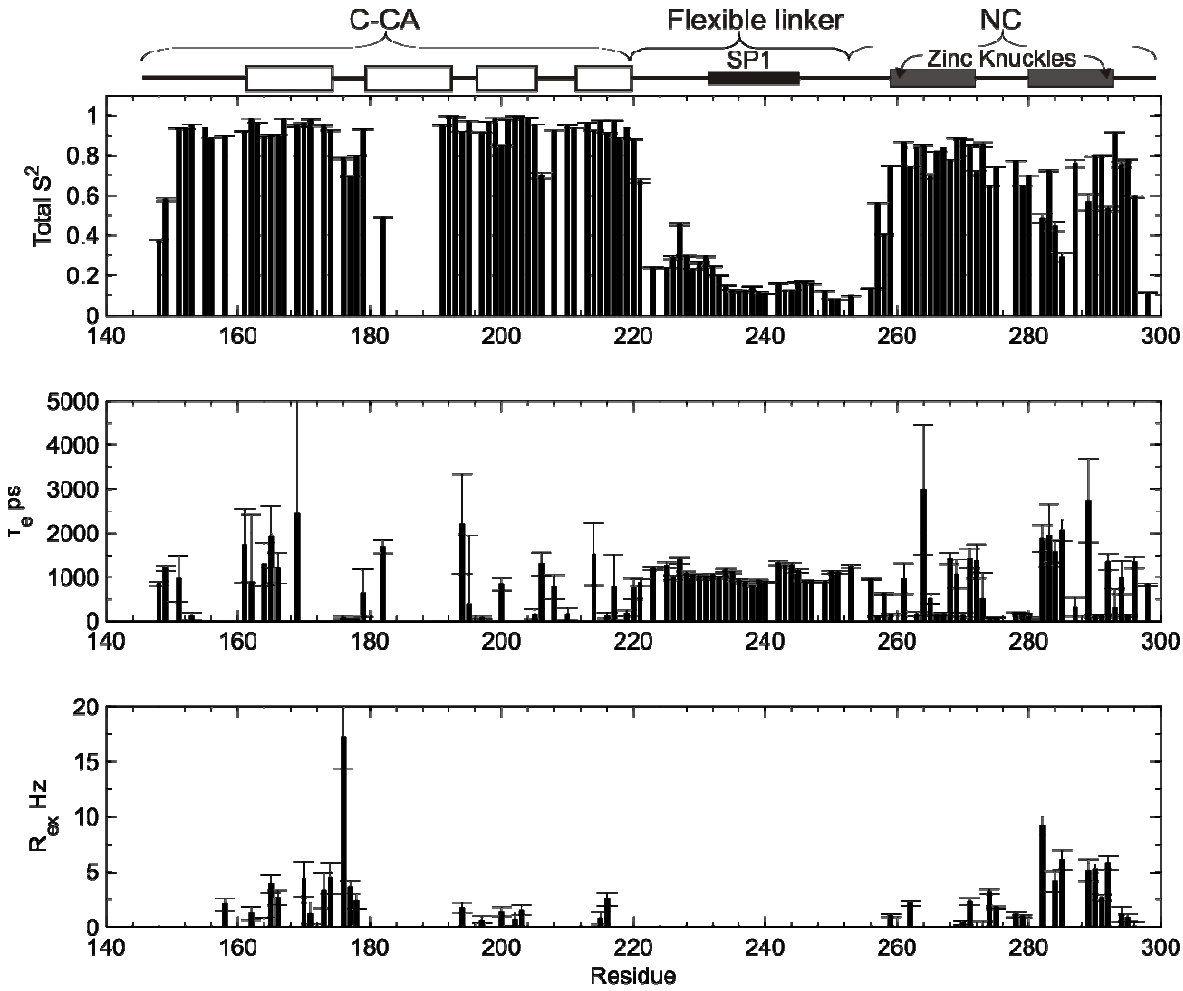


Figure 4.25: The results of a Lipari-Szabo model-free analysis to determine internal dynamics of the C-CA$_{AA}$-NC/CAI complex.

4.7.3. RDC analysis of the C-CA$_{AA}$-NC/CAI complex

RDCs were recorded under the same sample conditions as for the relaxation experiments, using a 5% cross-linked cylindrical polyacrylamide gel as an aligning media. As in the case of unliganded C-CA$_{AA}$-NC, spectral crowding and/or weak signal intensity was prohibitive to meaningful analysis beyond residue G220, so RDC data were analysed for

the C-CA_{AA} domain only. A total of 55 RDCs were measured in the C-CA_{AA} domain, and are shown in Figure 4.26 with back-correlations in Figure 4.27. Of these, 47 were used for alignment tensor determination. The fitted tensor is given in table 4.12. The values of all RDCs are listed in the appendix.

Table 4.12: The measured alignment tensor for the C-CAAA domain of C-CA_{AA}-NC + CAI.

α	β	γ	A_a x 10^4	A_r x 10^4
171 ± 7	9 ± 1	146 ± 4	3.28 ± 0.05	0.70 ± 0.09

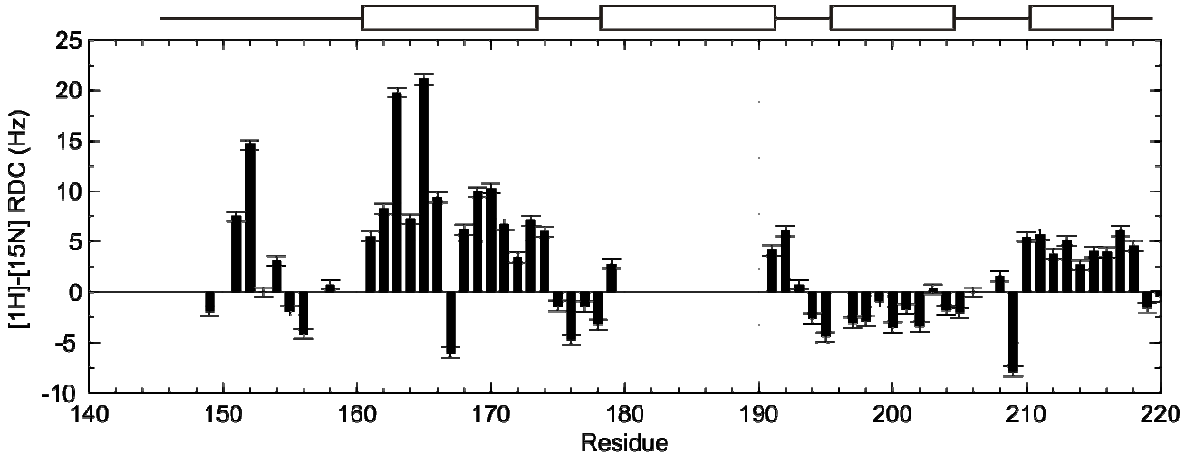


Figure 4.26: Measured RDCs for the C-CA_{AA} domain of the C-CA_{AA}-NC/CAI complex.

The C-CA_{AA}-NC concentration was 500 µM, and CAI 550 µM, with experimental details in the text. The helices of C-CA_{AA} are indicated by boxes.

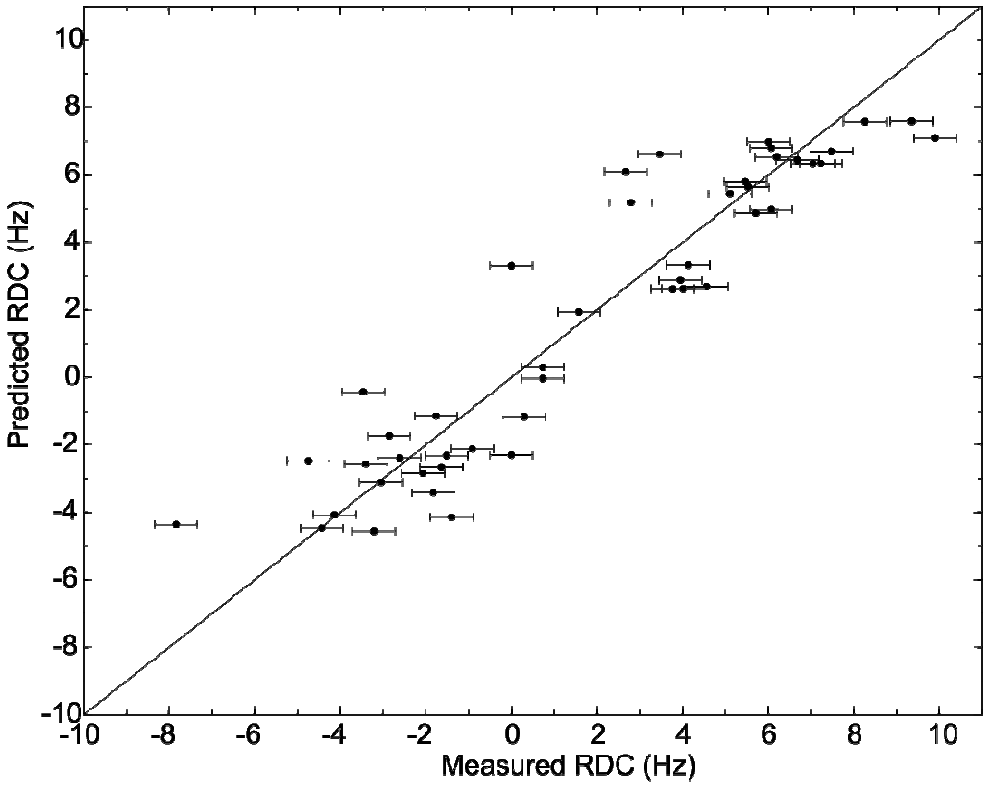


Figure 4.27: The correlation between measured and back-calculated RDCs for the C-CA_{AA} domain of C-CA_{AA}-NC + CAI.

4.7.4. Cα secondary shifts

The HNCACB experiment was used to determine Cα secondary shifts for the C-CA_{AA}-NC/CAI complex, in order to determine changes in secondary structure due to interactions with CAI. The results are given in Figure 4.28. Although most of helix 2 is missing from the analysis on account of gaps in assignments, the other helices of C-CA_{AA} are still seen, and in the NC domain the secondary shifts are very close to those observed for C-CA_{AA}-NC in the absence of CAI. In the flexible linker, generally fewer positive secondary shifts are observed between residues and K227 and S236, which in the absence of CAI had all been slightly positive. However, this may be an artefact since such small secondary shifts, in the presence of spectral crowding, are not easily measurable.

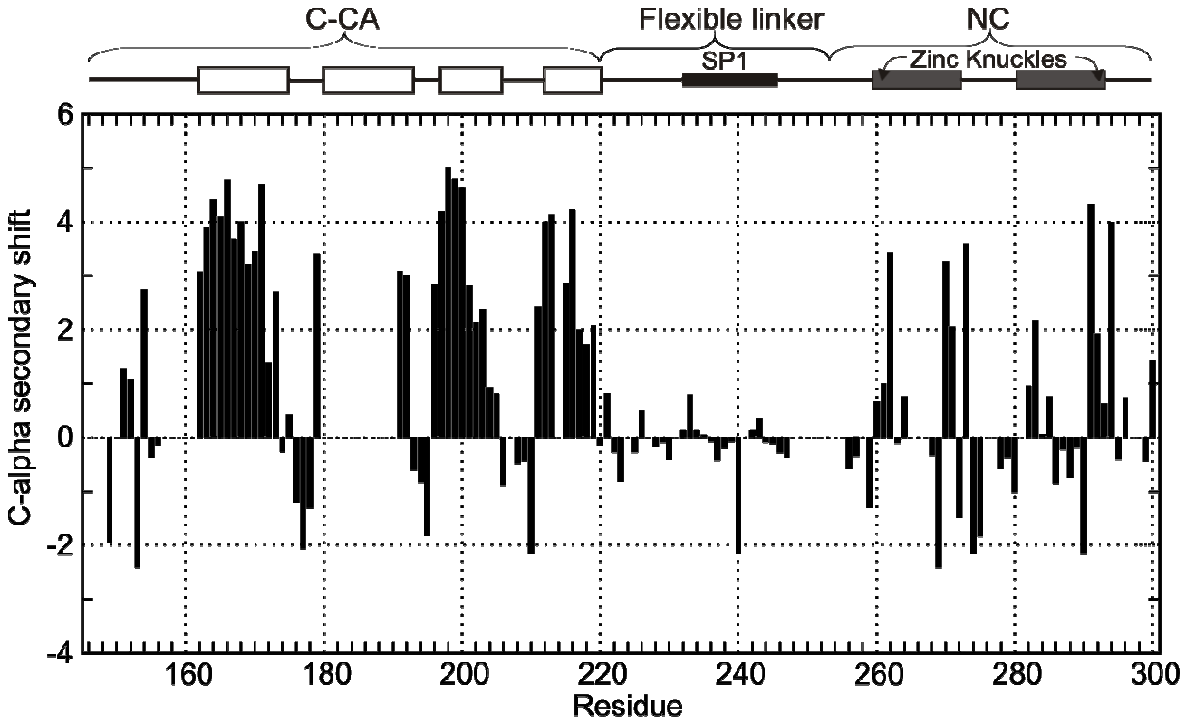


Figure 4.28: Cα secondary shifts for C-CA_{AA}-NC + CAI.

4.7.5. Is CAI binding altered by the SP1 or NC domains?

The chemical shift changes for free C-CA_{AA} upon addition of CAI have been published (146), and are very similar to those observed here for C-CA_{AA}-NC binding to CAI. The (unpublished) Cα secondary shifts for C-CA_{AA} in the presence of CAI are entirely consistent with those for C-CA_{AA}-NC. It is therefore likely that the C-CA_{AA}/CAI interaction is in no way affected by the presence of the SP1 and NC domains, a point which will be returned to in chapter 4.9. The fact that the amount of CAI required to inhibit immature assembly is greater as compared to mature assembly cannot therefore be explained by a perturbed C-CA/CAI interaction in the presence of the NC domain. However, inasmuch as involving energetically favourable NC-nucleic acid interactions in addition to formation of a CA lattice, immature assembly presumably is itself a more energetically favoured process than mature assembly. The increased thermodynamic favourability of immature assembly may therefore be the reason for requiring a greater concentration of CAI to inhibit the process. It was also remarked in the introduction that the CAI-bound C-CA dimer is rather unlike the mature C-CA dimer, but a better match to cET data of immature virions. Therefore the CAI-bound C-CA dimer may have a similar structure to the immature dimer, which may explain its relative inefficacy against immature assembly in comparison to mature.

4.8. NMR studies of interactions between either C-CA_{AA}-NC with dACGCC

The subject of this chapter is the characterisation of the C-CA_{AA}-NC/dACGCC complex using NMR spectroscopy. The structure of the NC/dACGCC complex has been previously solved by NMR, using a peptide fragment spanning residues 257-298 in CA numbering. However, no upstream modulating effects within the Gag polyprotein of dACGCC binding have been previously investigated, and this chapter extends the analysis to the flexible region from G220-K256 (including the SP1 domain) and to the C-CA_{AA} domain. Chemical shift changes are used to locate the binding site, whilst relaxation data are used to characterise the dynamics of the complex, both for overall motion of the globular domains, and for internal protein backbone motion throughout the complex. The binding site comprises most of the NC domain, and is consistent with the NMR structure for the complex. The internal dynamics demonstrate that the NC domain may be rigidified upon dACGCC binding, which is consistent with the previous analysis for the NC(257-298) fragment, whilst the C-CA_{AA} domain dynamics are essentially unchanged. RDC data are used to demonstrate that C-CA_{AA} domain alignment is unchanged by binding of dACGCC to the NC domain, which is consistent with the NC domain being decoupled from the C-CA_{AA} domain in both free and nucleic acid-bound form.

4.8.1. Assignments and chemical shift changes

Expression and purification of proteins was conducted as before. NMR studies were conducted firstly upon the C-CA_{AA}-NC-dACGCC complex, using a protein concentration of 500 µM and a dACGCC concentration of 550 µM, giving a protein-DNA complex population of ~90 % given the affinity determined by fluorescence spectroscopy (chapter 5.4.1). Assignments were made using HNCACB and CBCACONH spectra each recorded using ~72 hours instrument time at 25 °C, in a spectrometer operating at 600 MHz proton Larmor frequency with a room temperature probe, the sample being under the same conditions as in the absence of dACGCC. For residues in the C-CA_{AA} and flexible linker domains, many assignments could be made by comparison of HSQC spectra (then corroborated with the triple resonance data). For the NC domain, however, there were many significant chemical shift changes, which was entirely expected.

Assignments of backbone amide resonances were made for 143 out of 148 non-proline residues, though overlap between peaks often precluded further analysis. A chemical shift

difference plot between bound and unbound C-CA_{AA}-NC is given in Figure 4.30. This analysis showed that dACGCC indeed binds to the NC domain, and that the chemical environments of most residues are altered by its binding to some extent. The residues for which the chemical shifts change the most are either in parts of the zinc knuckles close to dACGCC, or in the linker between the zinc knuckles which is known to change in structure upon dACGCC binding. A titration was also performed, in which C-CA_{AA}-NC was held at a constant concentration of 100 µM, and dACGCC was added from a concentrated stock solution. Peak intensities changed only slightly during the titration, and a single peak was seen for affected residues, indicating that exchange between bound and unbound states was fast compared to typical chemical shift differences.

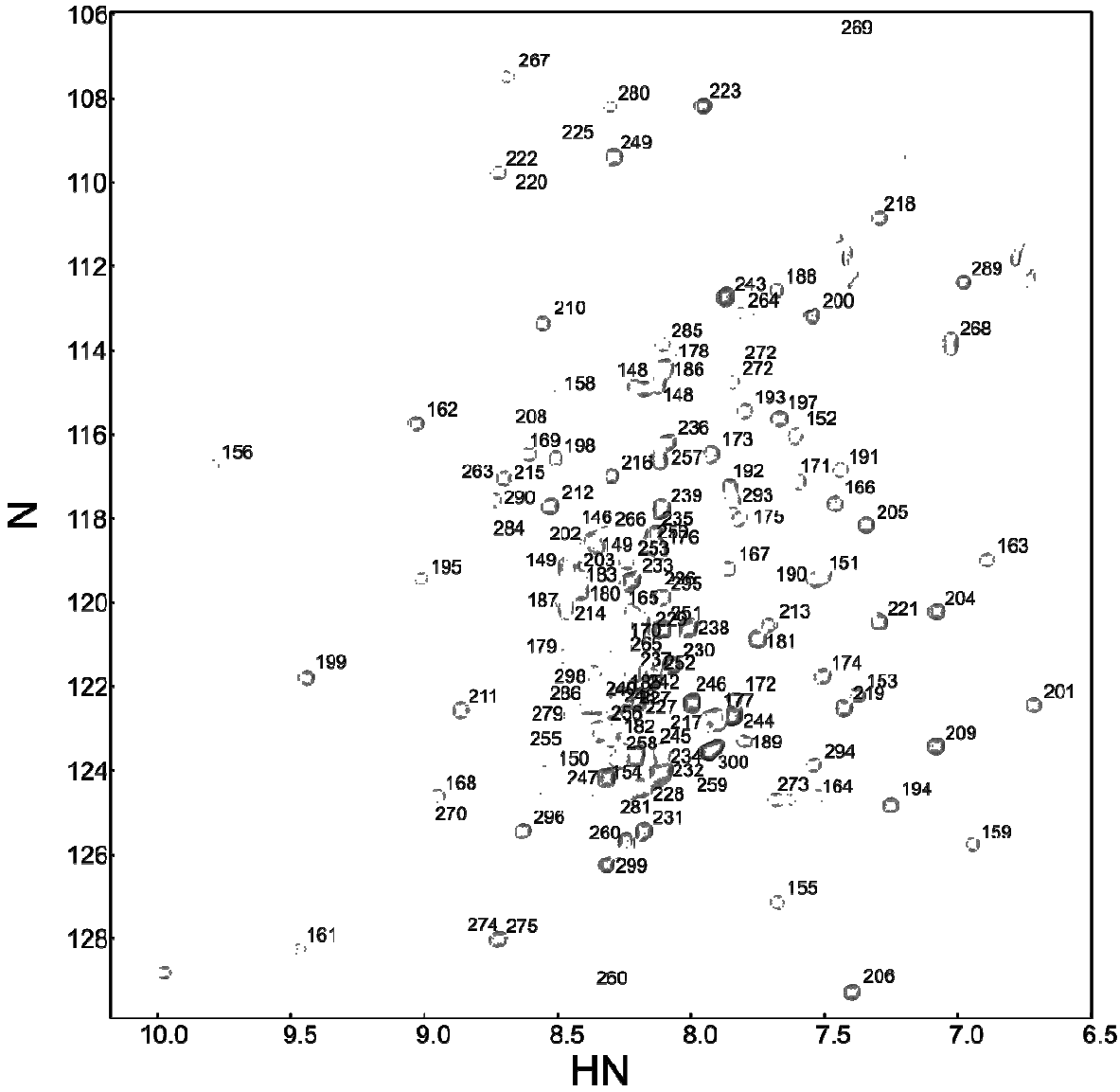


Figure 4.29: Assigned HSQC spectrum of C-CA_{AA}-NC + dACGCC.

Sample conditions are C-CA$_{AA}$-NC concentration: 500 µM, dACGCC concentration 550 µM, temperature: 25 °C, field: 600 MHz, pH 7.0.

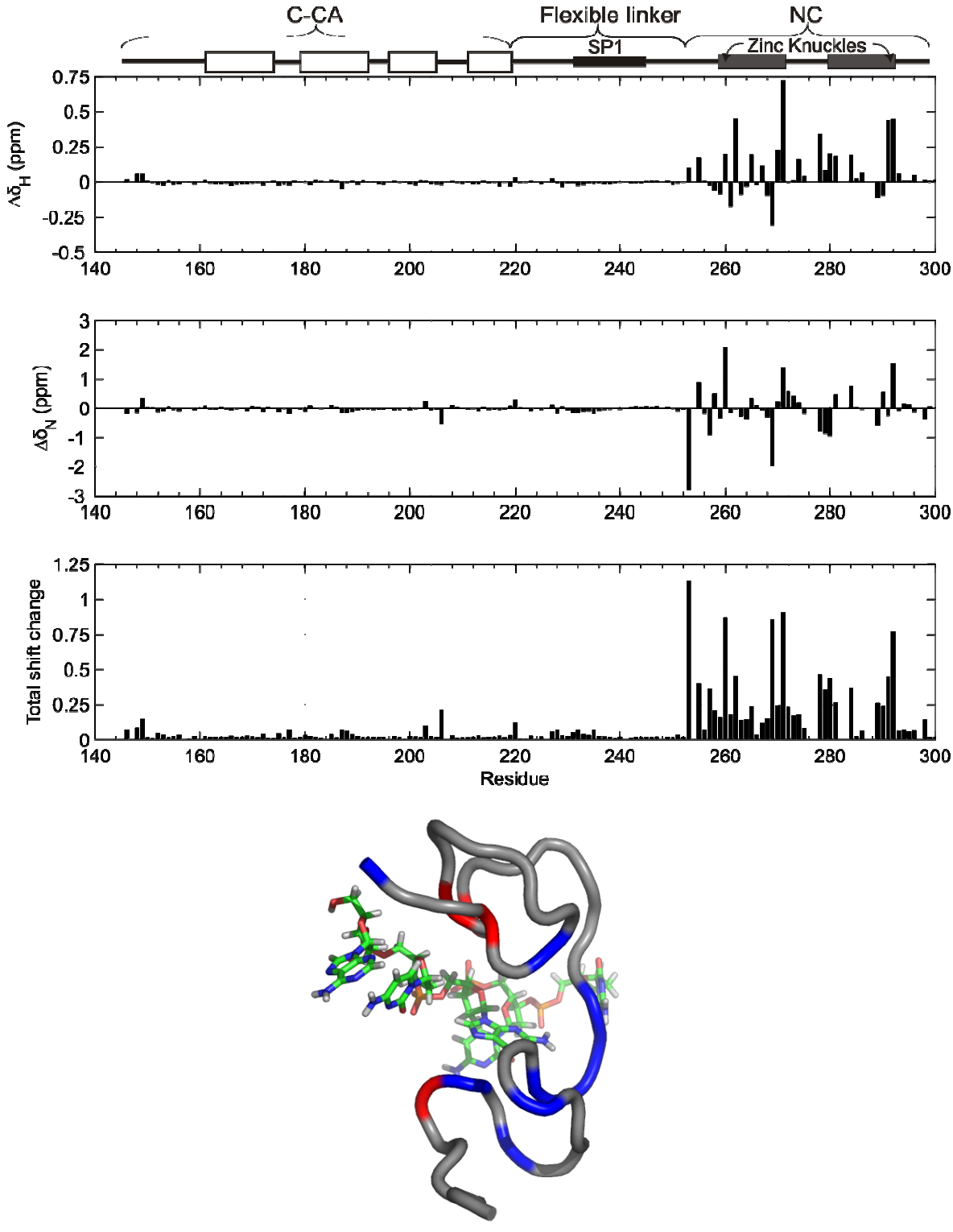


Figure 4.30: Chemical shift changes for C-CAAA-NC and dACGCC.

A) Chemical shift changes for C-CA$_{AA}$-NC in complex with dACGCC in the ^{1}H and ^{15}N dimensions, and weighted total shift changes. Sample conditions are given in the text. B)

Average NMR structure of the NC-dACGCC complex, with 'large' chemical shift changes displayed. Blue: total shift change > 0.25, Red: total shift change > 0.5, with the N-terminus at the top of the diagram.

4.8.2. Relaxation analysis for C-CA$_{AA}$-NC and dACGCC

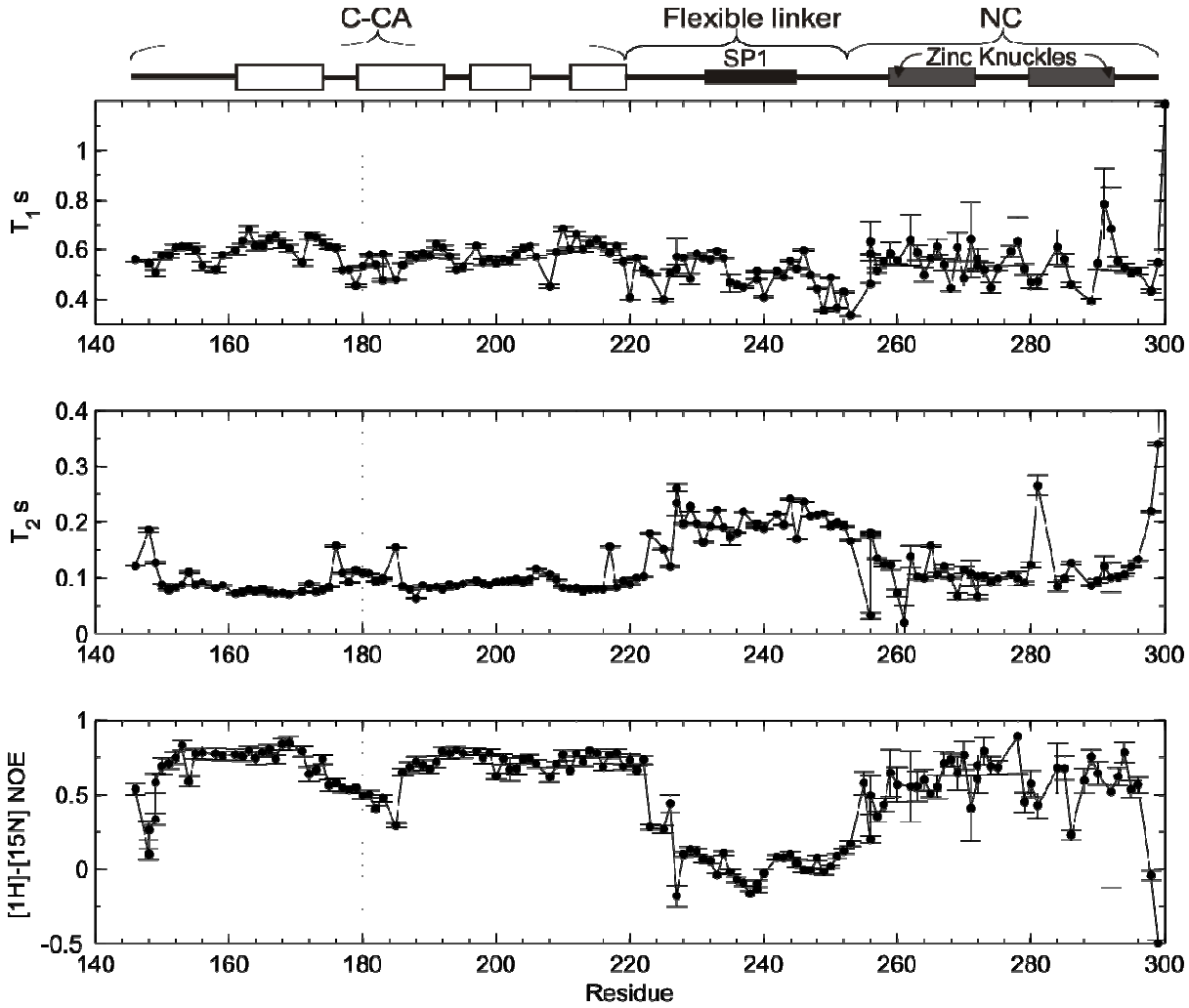


Figure 4.31: Relaxation parameters for C-CAAA-NC + dACGCC.

Sample conditions are CCA$_{AA}$-NC concentration: 500 µM, dACGCC concentration 550 µM, temperature: 25 °C, field: 600 MHz, pH 7.0. Secondary structure is indicated by boxes above the graphs. Fast dynamics in the mutated C-CA$_{AA}$ dimer interface and in the linker are apparent for low NOEs and high T$_2$s.

Diffusion tensors were again calculated for each domain separately, and are tabulated in Table 4.13. For the C-CA$_{AA}$ domain, the dimer X-ray structure was again used as a structural model (pdb code 1a43), whilst for the NC domain the average NMR structure of

the NC-dACGCC complex was used. This structure spans residues 257-298 in CA numbering, encompassing the well-folded part of NC only.

Table 4.13: Diffusion tensors for C-CAAA-NC + dACGCC

Domain	τ_m (ns)	$D_{\parallel}/D_{\perp}$	θ (°)	φ (°)
C-CA_{AA}	7.8 ± 0.04	1.4 ± 0.02	13 ± 4	24 ± 21
NC	6.4 ± 0.09	1.54 ± 0.17	81 ± 6	120 ± 7

Model-free analyses were conducted for each domain separately, in order to quantify timescales of internal motion. For the C-CA_{AA} domain, 66 residues were described using this approach. Of these, 8 were described by model 1, 21 by model 2, 6 by model 3, 24 by model 4 and 7 by model 5. Thus the C-CA_{AA} domain of C-CA_{AA}-NC in complex with dACGCC has very similar internal and overall dynamics to C-CA_{AA}-NC in the absence of dACGCC , which is consistent with there being no change in the C-CA_{AA} domain induced by NC-dACGCC interactions as indicated by chemical shift changes. As in the case of unliganded C-CA_{AA}-NC, the C-CA_{AA} domain shows fast internal motion distributed throughout the entire sequence, mostly with timescales in the range 500-3000 ps. Many resonances are affected by exchange broadening contributing 1-3 Hz to the linewidths, which may be due to slow conformational exchange or slow sidechain motions. The order parameters show that the fast internal motions are of similar amplitudes to those in the unliganded form.

For the NC domain, 31 residues were analysed, since spectral crowding was prohibitive to either assigning or accurately analysing many residues. Of these, 12 were described by model 1, 6 by model 2, 2 by model 3, 9 by model 4 and 3 by model 5. There is generally less fast internal motion than in the unliganded case; 30 out of 36 analysed residues in the unliganded NC domain required fast motion terms to account for their relaxation behaviour, whereas 18 out of 31 residues require such terms in the dACGCC complex. A similar number of residues in the dACGCC complex require R_{ex} terms. Given the concentration regime used, it is to be expected that exchange broadening will occur due to free/bound exchange. The exchange terms measured are consistent with this, and indeed the larger exchange terms correspond to residues which undergo large chemical shift changes upon binding to dACGCC. Inasmuch as a large chemical change makes chemical exchange slower, by definition, than for a residue which undergoes a small shift change,

one expects greater exchange broadening for residues which undergo large shift changes upon binding. The exchange terms observed, therefore, are more likely to be due to free/bound exchange than conformational exchange amongst an ensemble of bound states. The R_{ex} values correlate reasonably well with the ^{15}N chemical shift changes ($R^2 = 0.81$ as plotted in Figure 4.33 using a linear function) , which supports a model for binding in which a unique conformation of NC is selected by dACGCC, with fast internal flexibility, but which does not itself change conformation unless released from the complex. This is an example of conformational selection. A model in which conformational exchange results in the observed R_{ex} terms can account for their existence, but this introduces the assumption of conformational flexibility whilst ignoring the necessary inclusion of bound/free exchange, which must occur since the diffusion tensor and chemical shifts are altered in the presence of dACGCC. Thus we accept that exchange broadening is most likely a result of bound/free exchange.

For the flexible linker, overall motion was again approximated by an isotropic diffusion tensor dominated by C-CA_{AA} overall motion. A model-free analysis was conducted using this approximation, the results of which should be a first approximation to the timescales of internal motion and degree of flexibility. Model 1 was consistent with 4 residues, model 2 was consistent with 19 residues, and model 5 was consistent with 3 residues. Thus 26 residues could be described using the model-free approach. The model-free parameters are plotted in Figure 4.32 and tabulated in the appendix.

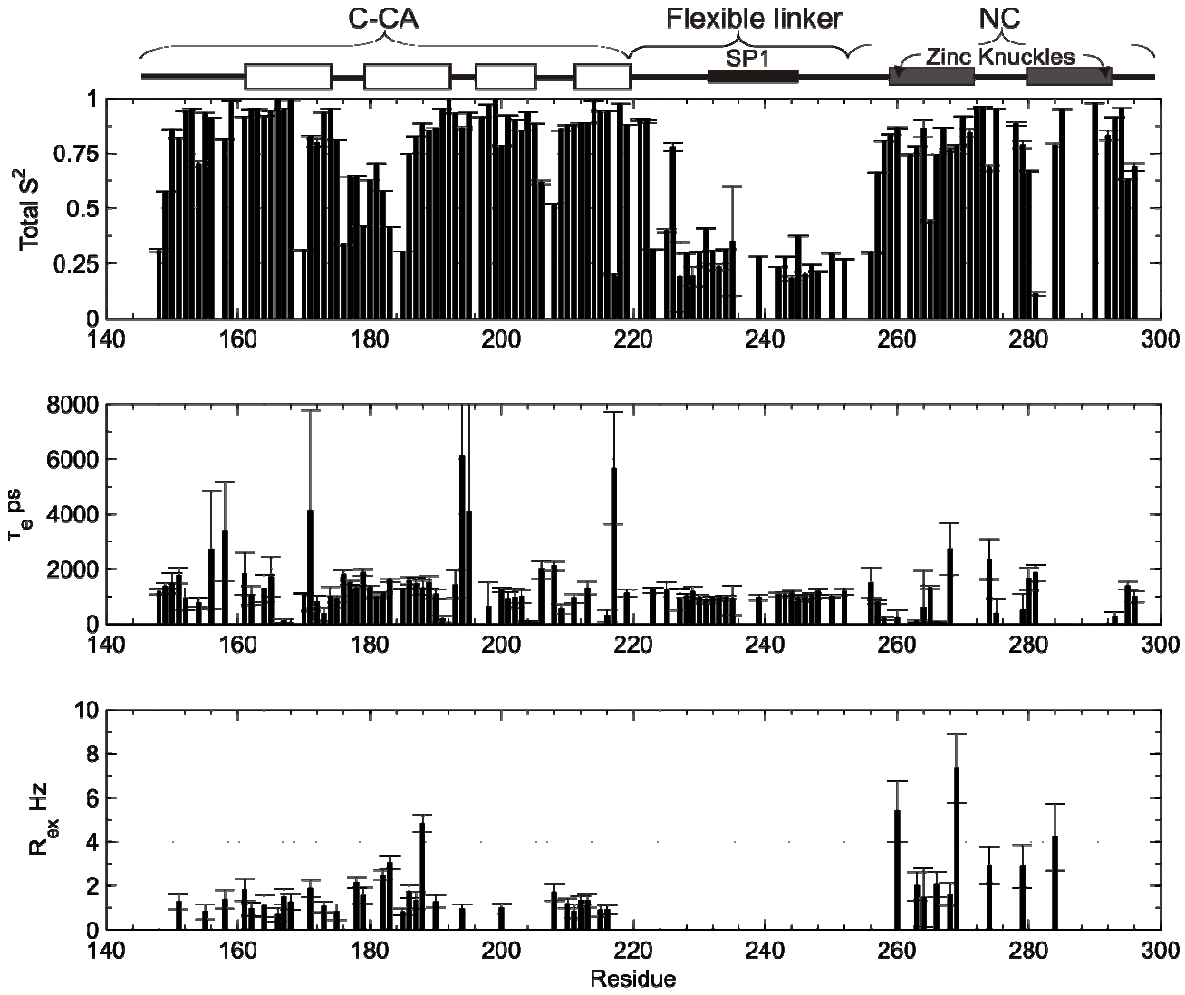


Figure 4.32: Results of a model-free analysis on the C-CA_{AA}-NC/dACGCC complex.

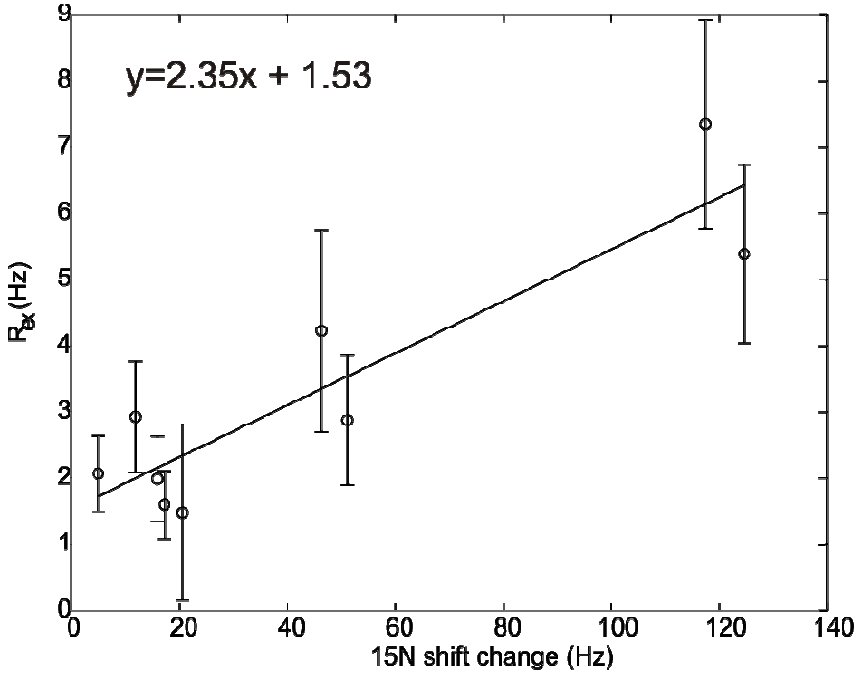


Figure 4.33: R_{ex} terms predicted from a model-free analysis of C-CA_{AA}-NC + dACGCC correlate with 15N chemical shift changes induced by dACGCC binding.

4.8.3. RDC analysis of the C-CA_{AA}-NC/dACGCC complex

As in the cases of C-CA_{AA}-NC and the C-CA_{AA}-NC/CAI complex, broad peaks and spectral crowding were prohibitive to the accurate measurement of RDCs downstream of the folded part of the C-CA_{AA} domain, so the analysis of RDCs here pertains only to residues 148-220, for which 68 RDCs could be measured. Of these, 53 were used in alignment tensor determination. The fitted tensor is given in Table 4.14, the measured couplings in Figure 4.34 and the calculated RDCs in Figure 4.35. In the appendix, all measured and predicted couplings are listed.

Table 4.14: Alignment tensor for the C-CA_{AA} domain of C-CA_{AA}-NC + dACGCC

α	β	γ	A_a x 10^4	A_r x 10^4
-152±6	171±1	-64±5	2.86±0.04	0.94±0.09

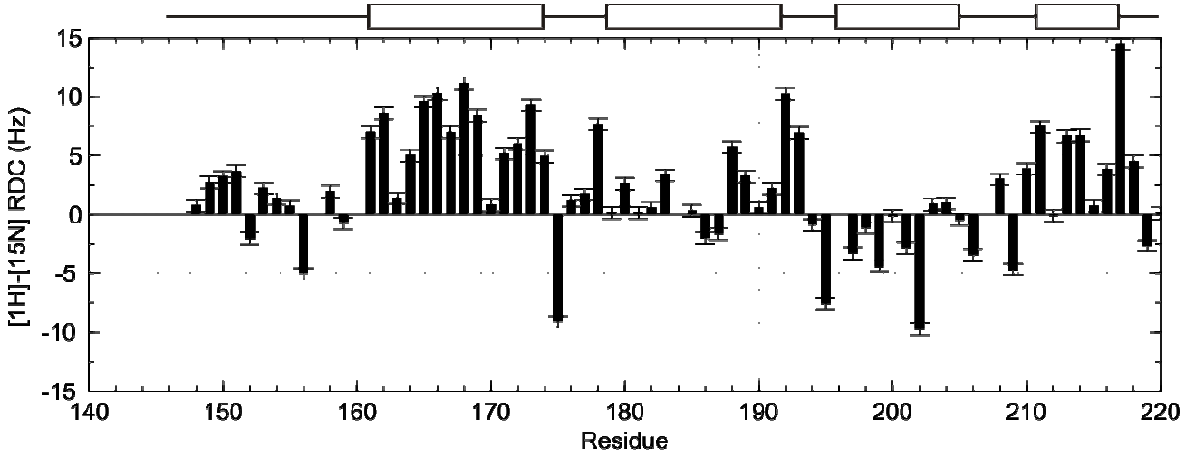


Figure 4.34: Measured RDCs for the C-CAAA domain of the C-CAAA-NC/dACGCC complex. The helices of C-CAAA are indicated with boxes.

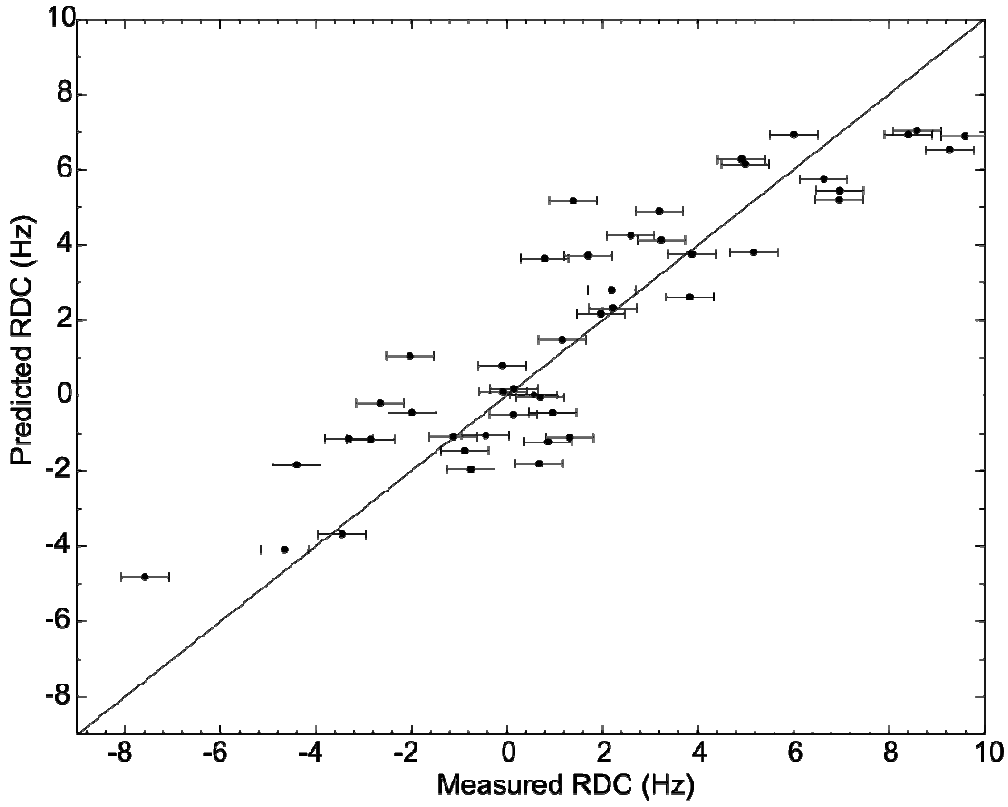


Figure 4.35: Correlation between measured and back-predicted RDCs for C-CA_{AA}+dACGCC.

Only the data used in obtaining the data for the C-CA_{AA} domain are shown.

4.8.4. Cα secondary shifts

Cα secondary shifts were determined from an HNCACB experiment and are plotted in Figure 4.36. Secondary shifts in the C-CA_{AA} and flexible linker domains are very close to the values observed in the absence of dACGCC and, perhaps surprisingly, also generally consistent in the NC domain with the unliganded form.

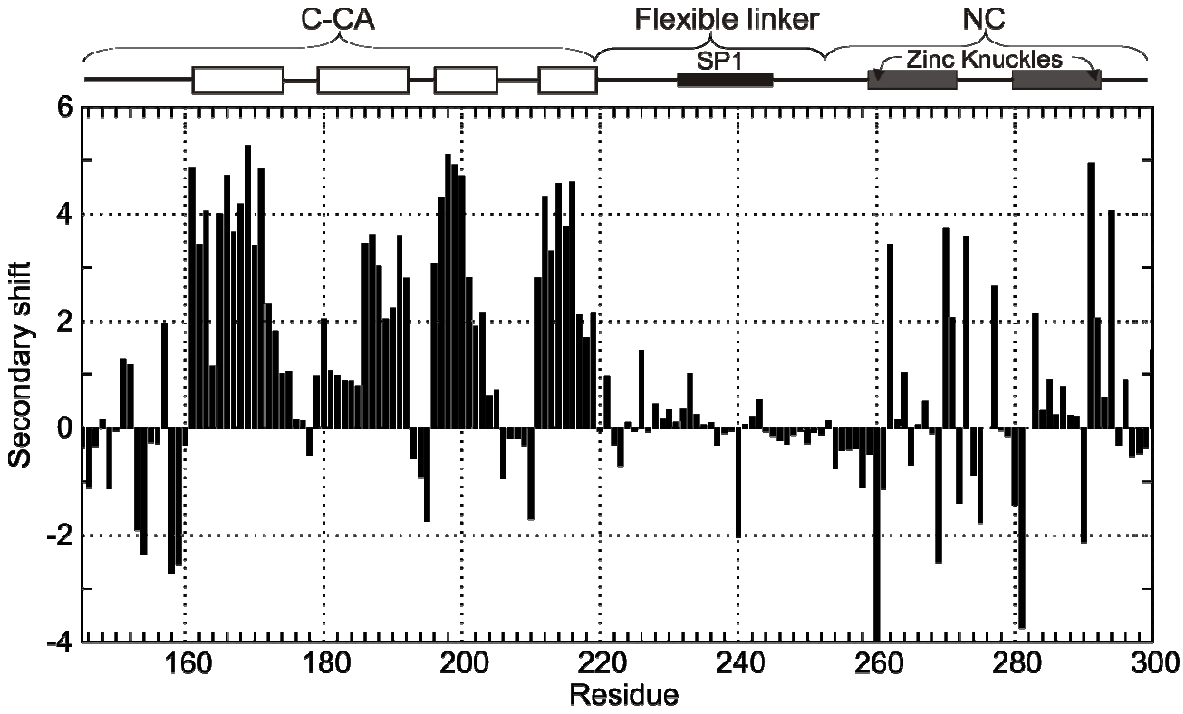


Figure 4.36: Secondary Cα shifts for C-CAAA-NC in complex with dACGCC.

Those for the C-CA$_{AA}$ domain are consistent with its known structure, and the linker region shows similar shifts to those in C-CA$_{AA}$-NC and C-CA-NC.

4.9. NMR studies of interactions between C-CA_{AA}-NC and dSL3

4.9.1. Introduction

In chapter 5, the interactions between various proteins and a DNA deoxyribonucleotide dSL3 are described which seek to determine the affinity and stoichiometry of its interactions with the NC domain. A full description of dSL3 is reserved for chapter 5, and for the present discussion it suffices to say that it is desirable to know the binding site for dSL3 on C-CA_{AA}-NC, and whether any effects upon regions distal to the NC domain can be detected. In order to characterise interactions between C-CA_{AA}-NC and dSL3, NMR was attempted, since it is powerful in detecting binding sites with site specificity, and determining whether other interactions occur as a result of a protein-ligand interaction, for example formation of higher-order complexes or changes in dynamic properties of the protein. However, upon titration of dSL3 into C-CA_{AA}-NC, it was observed that many signals corresponding to residues in the NC domain were very much reduced in intensity, even with low dSL3 concentrations as compared to the C-CA_{AA}-NC. Most chemical shifts for peaks in the C-CA_{AA} domain were not affected, so the behaviour of the C-CA_{AA} domain was tractable. In this chapter, the various attempts to characterise the C-CA_{AA}-NC/dSL3 complex with NMR are described, and models for processes which may be responsible are tested. The outcome is that, although dSL3 is shown to interact only with the NC domain, a detailed characterisation of the complex has not been possible.

4.9.2. Assignments and chemical shift changes

Assignments were made for 101 out of 148 non-proline residues. Initially, to determine the binding site for dSL3, it was titrated into 100 µM C-CA_{AA}-NC using a 600 MHz spectrometer. Many NC peaks disappeared, even with a relatively low dSL3 concentration, and no additional peaks appeared in the spectrum, even with dSL3 in excess. This was not anticipated, for the theory of chemical exchange predicts that if a peak is weakened with low ligand concentration, then the reaction is in the intermediate exchange regime. However, with excess ligand, the reaction will be pushed to the fast exchange regime, in which case the peak will again be visible, but at the chemical shift corresponding to the protein-ligand complex.

In Figure 4.37, the NH chemical shift changes are shown, for those peaks which did not disappear in the course of the experiment. It is clear that the chemical shifts in the C-CA_{AA} domain and flexible linker are essentially unaffected.

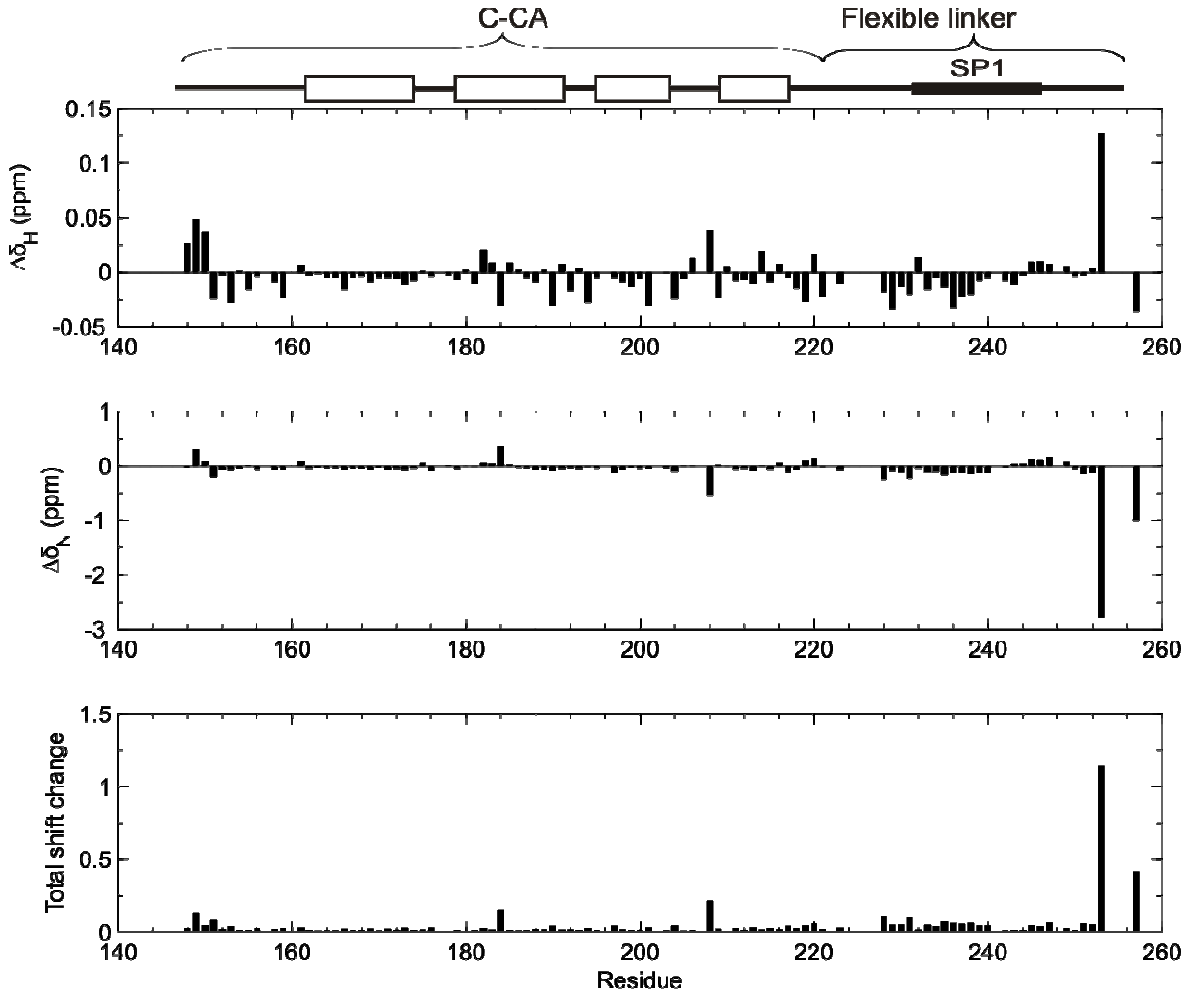


Figure 4.37: Chemical shift changes in C-CAAA-NC upon addition of dSL3.

The sample conditions are: C-CA_{AA}-NC concentration: 500 µM, dSL3 concentration: 500 µM, field: 600 MHz, temperature: 25 °C, pH 7.0. Assignments were not possible beyond the flexible region due to unexplained suppression of NC domain peak intensities.

Since the peak intensities in the presence of dSL3 are very weak or totally absent for so many NC residues, a titration was performed in which the C-CA_{AA}-NC concentration was again kept at 100 µM, but only low concentrations of dSL3 were added, from 1 to 8µM in order to track such changes. To improve resolution and signal-to-noise, a 700 MHz instrument equipped with a cryo-probe was used for this experiment. The only observed effect in the titration was that NC peaks were dramatically reduced in intensity, even with the lowest dSL3 concentration of 1 µM, whilst peaks in the C-CA_{AA} domain were

unaffected. This is illustrated in Figure 4.38 for two examples, C260 in the NC domain and Q155, which are a convenient example on account of their coincident ^{15}N shifts. Overlaid HSQC spectra in the absence and presence of dSL3 are given in Figure 4.39 to illustrate which peaks are affected, and that no additional peaks are visible even with an excess of dSL3.

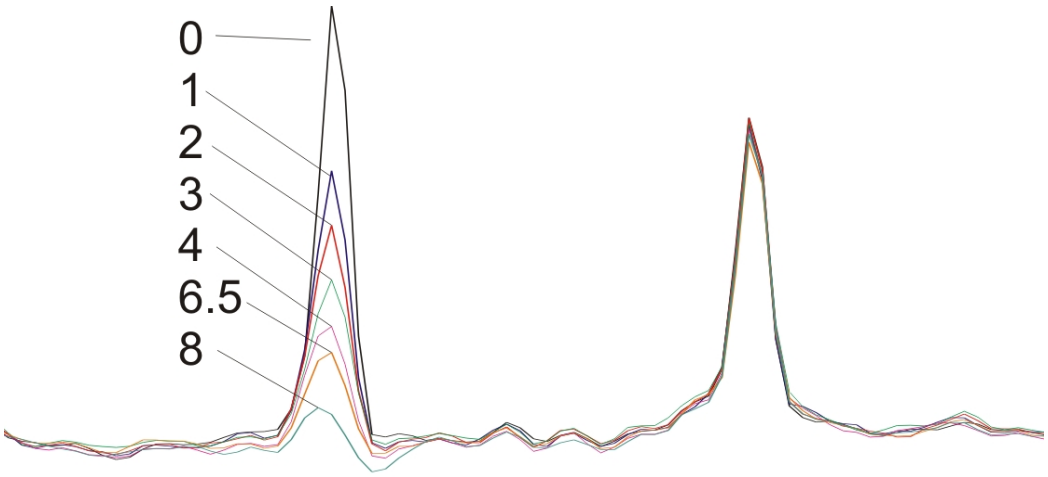


Figure 4.38: Slices through HSQC spectra of C-CA$_{AA}$-NC with dSL3, whose concentration in μM appears on the left.

The peak on the left is for C260 in the NC domain, the peak on the right is for Q155 in the C-CA$_{AA}$ domain. The slices are taken along the ^{1}H dimension of spectra recorded at 700 MHz, with 100 μM C-CA$_{AA}$-NC.

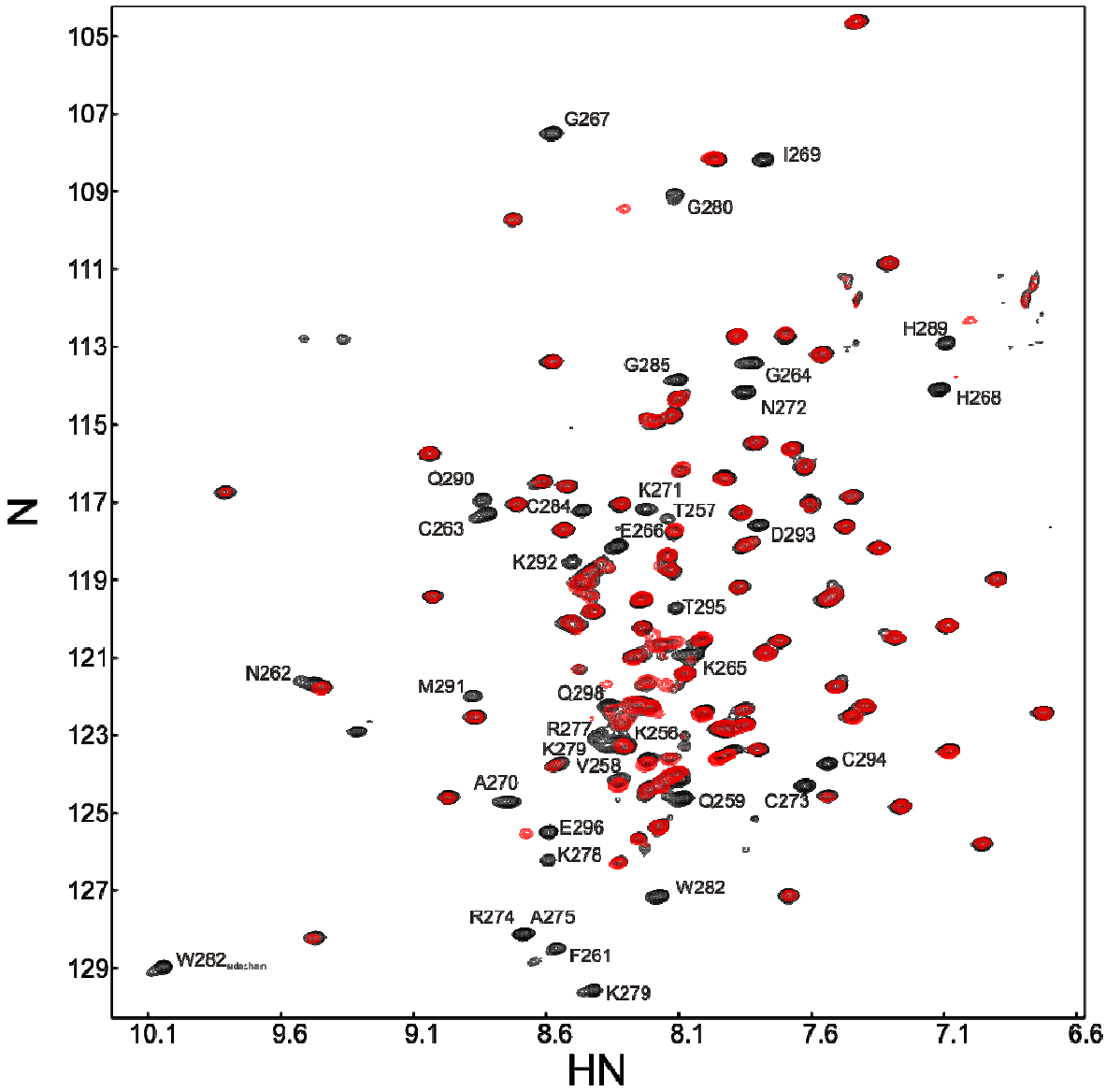


Figure 4.39: Overlaid HSQC spectra of C-CA$_{AA}$-NC with and without dACGCC.

Sample conditions: C-CA$_{AA}$-NC concentration 500 µM, dSL3 concentration 550 µM, 600 MHz, pH 7.0, 25 °C. The missing peaks are labelled and are all in the folded part of the NC domain.

4.9.3. Relaxation data

Relaxation times and NOEs were measured only between residues 148 and 253, since too few assignments existed outside this range to permit analysis. Mono-exponential relaxation was assumed. The data are useful, even though incomplete, since they span the entire C-CA$_{AA}$ domain and most of the flexible linker. The data are plotted in Figure 4.40. It was observed that some T_1s were below the theoretical minimum value, particularly in the N-

terminal region of helix 2 and in the flexible linker. Possible explanations are reserved for the discussion.

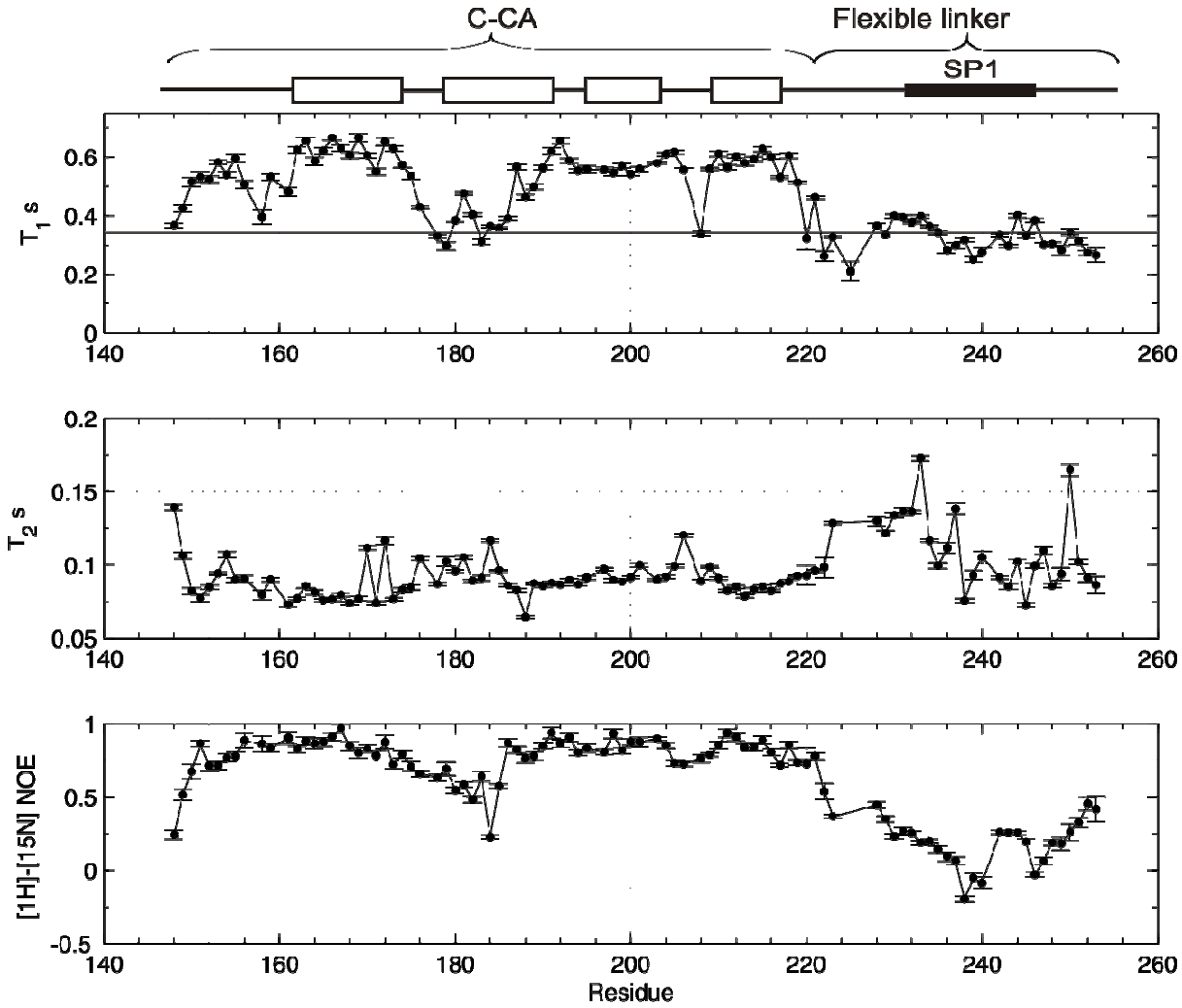


Figure 4.40: *Measured T_1, T_2 and heteronuclear NOEs for the C-CA_{AA}-NC/dSL3 complex.*

Data were recorded at 600 MHz with 500 µM C-CA_{AA}-NC and 500 µM dSL3 at 25 °C, pH 7.0. The data are plotted only for residues 148-253, since no data could be obtained outside this range. In the T_1 plot, the minimum T_1 value permitted by Redfield theory is indicated with a solid line, to highlight those values falling outside.

Since relaxation times and NOEs could be determined for a reasonable number of residues in the C-CA_{AA} domain, a diffusion tensor was calculated, excluding residues with NOE < 0.75 or with a T_1 below the theoretical minimum. The fitted parameters are given in Table 4.15. The diffusion tensor is very similar to that for the C-CA_{AA} domain in C-CA_{AA}-NC without ligands, and in the presence of dACGCC.

Table 4.15: The diffusion tensor for the C-CAAA domain of C-CA_{AA}-NC + dSL3

Domain	τ_m (ns)	$D_{\parallel}/D_{\perp}$	θ (°)	φ (°)
C-CA_{AA}	7.7 ± 0.02	1.34 ± 0.02	35 ± 2	10 ± 4

4.9.4. Secondary shift analysis

As in the case of relaxation analysis, the Cα secondary shift analysis for the C-CA_{AA}-NC/dSL3 complex extended only as far as residue 253, and the results are given in Figure 4.41.

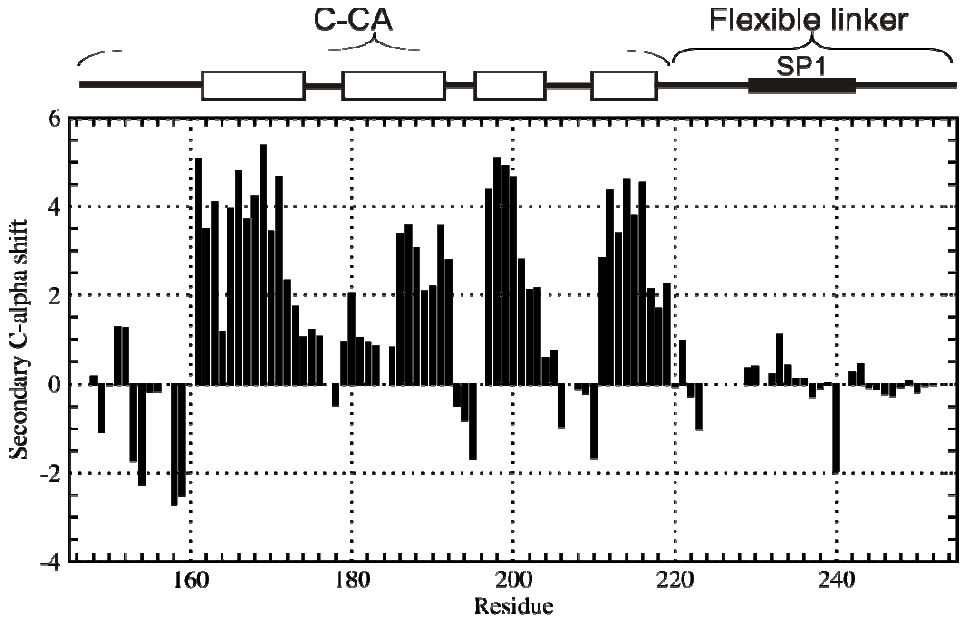


Figure 4.41: Secondary Cα shifts for C-CA_{AA}-NC + dSL3.

As in the case of other analyses with dSL3, data could not be obtained for the NC domain. Those data which could be analysed are consistent with all other secondary shift analyses of the C-CA_{AA} domain.

4.9.5. Why are NC domain signals suppressed?

It is possible to explain the dramatic suppression in signal intensity if it is assumed that dSL3 is in complex with a paramagnetic cation. A paramagnetic cation has one or more unpaired electrons, and a strong dipolar coupling between such electron spins and nuclear spins in the vicinity causes rapid relaxation of such nuclear spins. The T_1 and T_2 for a nuclear spin close to a paramagnetic cation can be very much shorter than usual. By virtue of the negatively charged phosphates on the DNA backbone, dSL3 could bind paramagnetic cations, which may in principle be present as impurities under the conditions of preparation of dSL3.

To try and determine whether a paramagnetic cation was responsible for the observations, a solution of 100 µM C-CA_{AA}-NC with 10 µM dSL3 was treated with EDTA up to concentrations of 100 µM. Since EDTA binds cations, generally with high affinity, it was reasoned that it may remove the suggested paramagnetic cation from dSL3, thus resulting in a normal NMR spectrum for the C-CA_{AA}-NC/dSL3 complex. However, excessive EDTA must be avoided since NC is a zinc-dependent protein, so it would be difficult to interpret changes in the spectrum since NC's DNA-binding capability may be compromised. As it happened, no change at all was seen. The options for explaining the inaction of EDTA are twofold. It may be that there is no paramagnetic impurity, or that it is too tightly bound to dSL3 for EDTA to remove.

Unfortunately, therefore, the basis for signal intensity diminution in the NC domain upon addition of dSL3 remains unexplained. However, it is established by means of chemical shift invariance and diffusion tensor invariance in the C-CA_{AA} domain that the interaction site is restricted to the NC domain.

4.10. Comparison of NMR results

Before making comparisons of results, the following assertions are made, which will be justified *post hoc*. Firstly, the C-CA_{AA} domain and NC domains do not interact with one another, and a change in either domain is not communicated to the other. Secondly, the linker and NC domain do not have any role in C-CA dimerisation, and C-CA dimerisation does not affect either the linker or NC.

4.10.1. Data comparisons demonstrating C-CA_{AA} and NC independence

It has been noted already that the shapes and orientations of the diffusion tensors for C-CA_{AA} and the corresponding domain in C-CA_{AA}-NC are very similar. A comparison of the order parameters shows that the internal dynamics are also very similar, as shown in Figure 4.42. There are some small changes in order parameters, particularly in helices 2 and 4, but the general pattern is the same in both cases.

It was shown in chapter 4.8 that the chemical shifts and diffusion tensor of the C-CA_{AA} domain of C-CA_{AA}-NC are unchanged in the presence of dACGCC. A comparison of order parameters between C-CA_{AA}-NC with and without dACGCC is made in Figure 4.43. It is seen that the internal dynamics of the C-CA_{AA} domain are essentially unaltered by

dACGCC, which is interpreted as meaning that perturbations to the NC domain are not communicated to the C-CA_{AA} domain, and that the two domains are therefore independent of one another and do not interact.

The perturbations to the NC domain resulting from dACGCC binding do not affect the C-CA_{AA} domain. Likewise, the perturbations to the C-CA_{AA} domain resulting from addition of CAI are not communication to the NC domain, as evidence by the small chemical shift changes in NC, its consistent diffusion tensor and the consistent order parameters shown in Figure 4.44. In chapter 4.7, it was noted that there is no difference between the behaviour of C-CA_{AA} and C-CA_{AA}-NC when CAI was added.

A simple visual representation of order parameters is provided in Figures 4.45 and 4.46, in which the C-CA and NC structures are coloured according to order parameters.

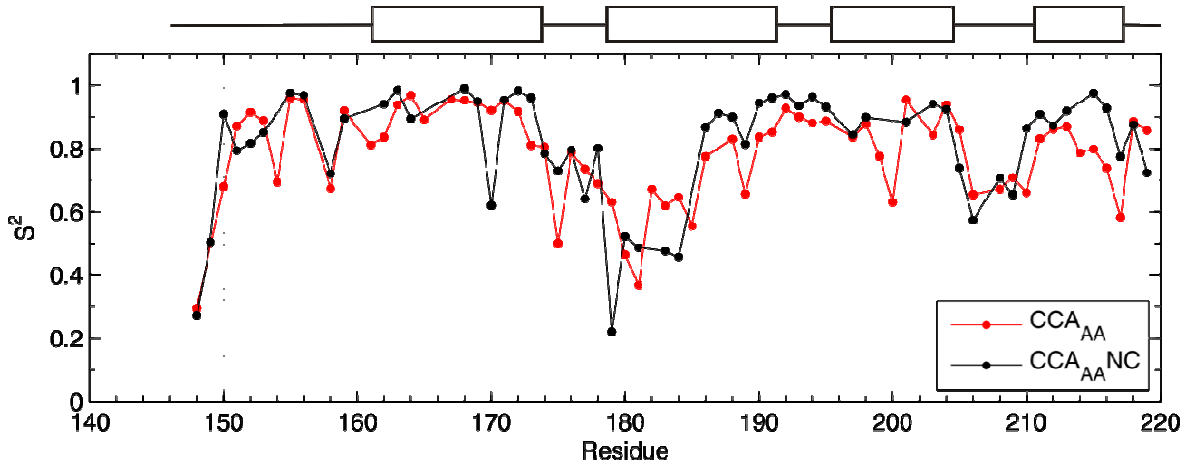


Figure 4.42: Comparison of order parameters for C-CA_{AA} and C-CA_{AA}-NC.

Error bars are omitted for clarity but are shown on the individual datasets elsewhere.

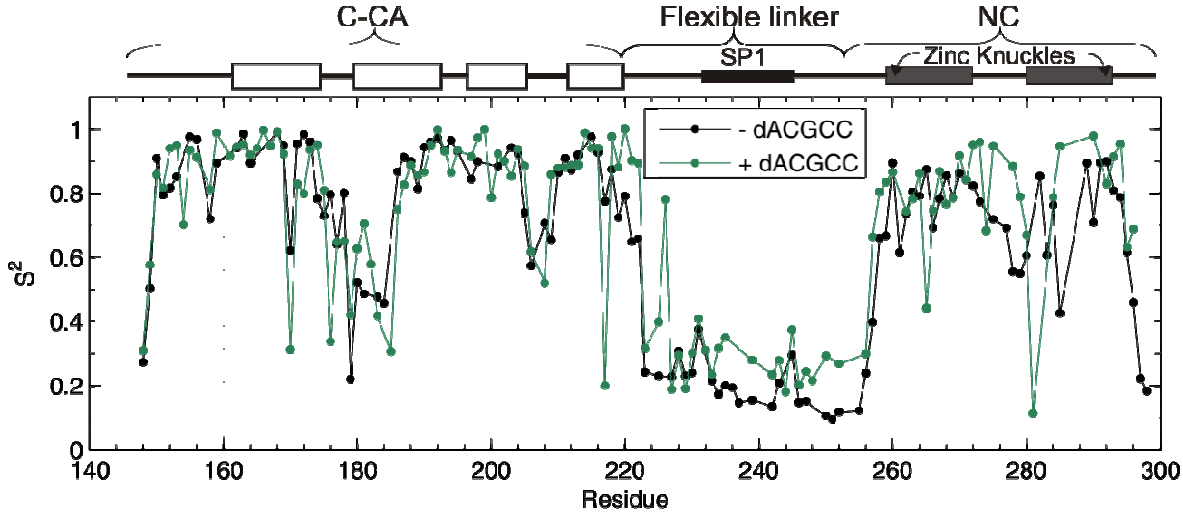


Figure 4.43: Comparison of order parameters for C-CA$_{AA}$-NC with and without dACGCC.

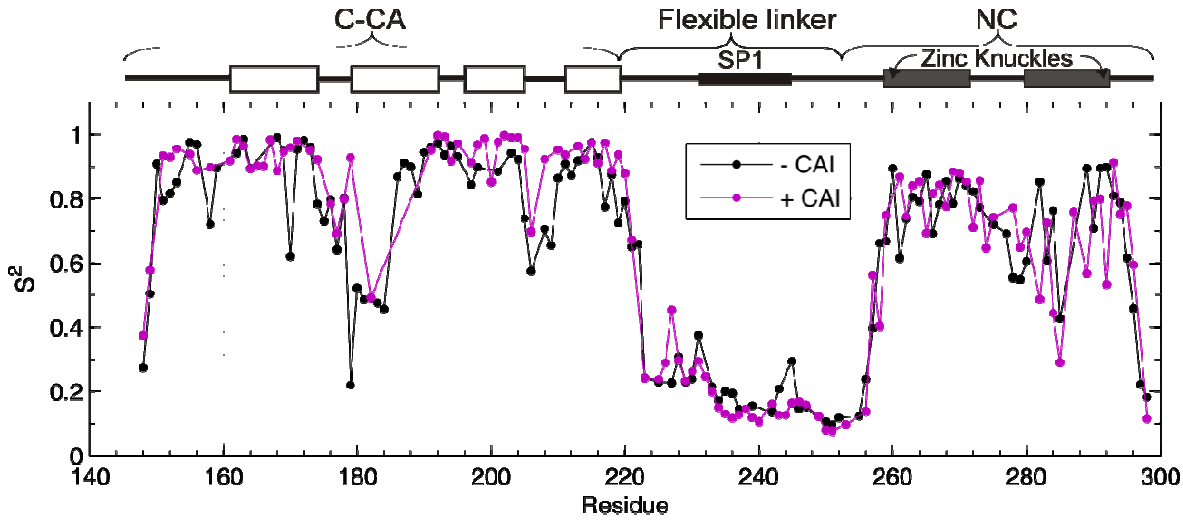


Figure 4.44: Comparison of order parameters for C-CA$_{AA}$-NC with and without CAI.

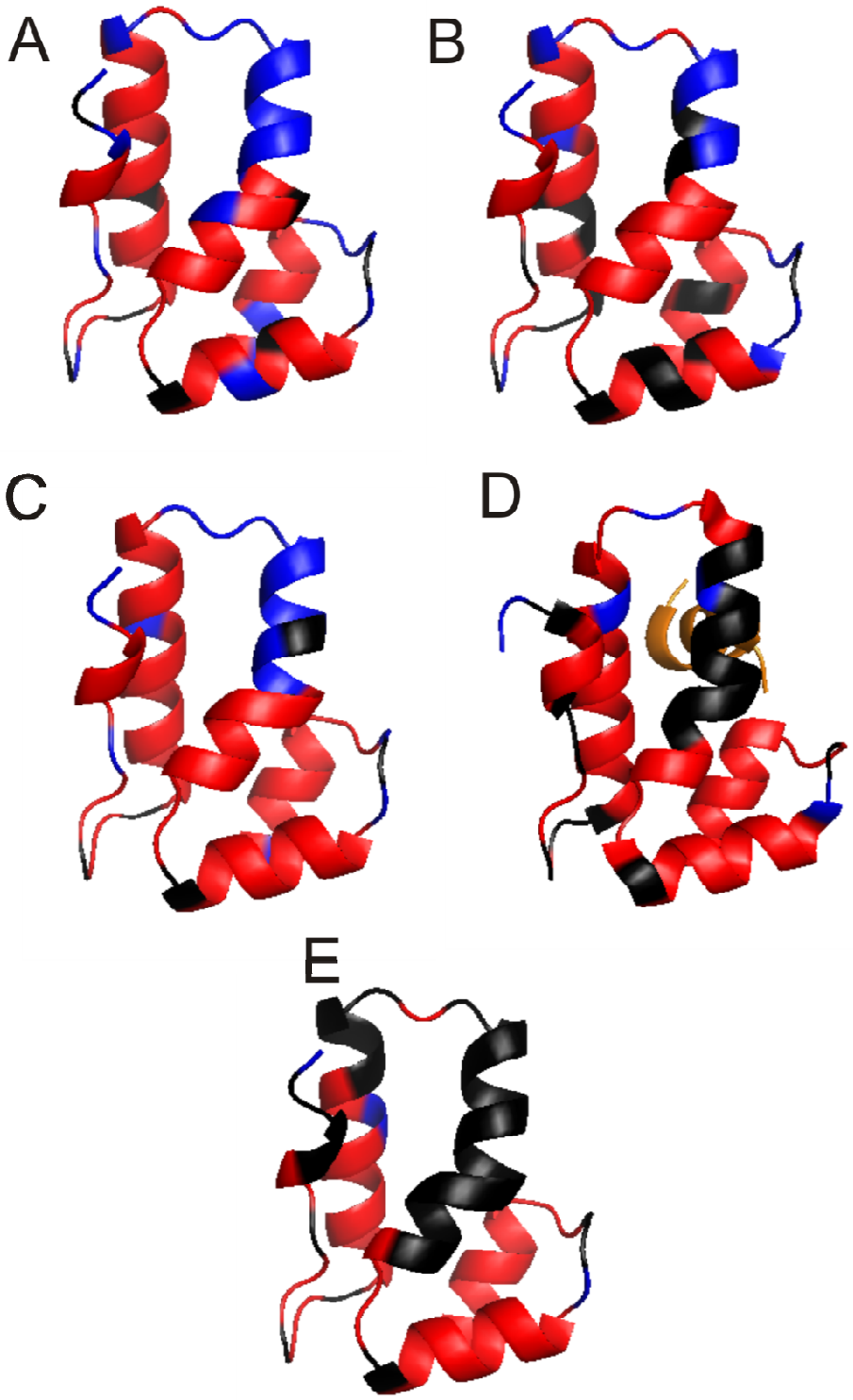


Figure 4.45: Order parameters displayed on C-CA and C-CA_{AA} structures.

Residues with $S^2 \leq 0.75$ are coloured blue and those with $S^2 > 0.75$ are coloured red. Residues which were not analysed due to absence of data (including prolines) are coloured black. A) C-CA_{AA}, B) C-CA_{AA}-NC, C) C-CA_{AA}-NC + dACGCC, D) C-CA_{AA}-NC + CAI, E) C-CA at 1 mM. In D), CAI is shown in orange.

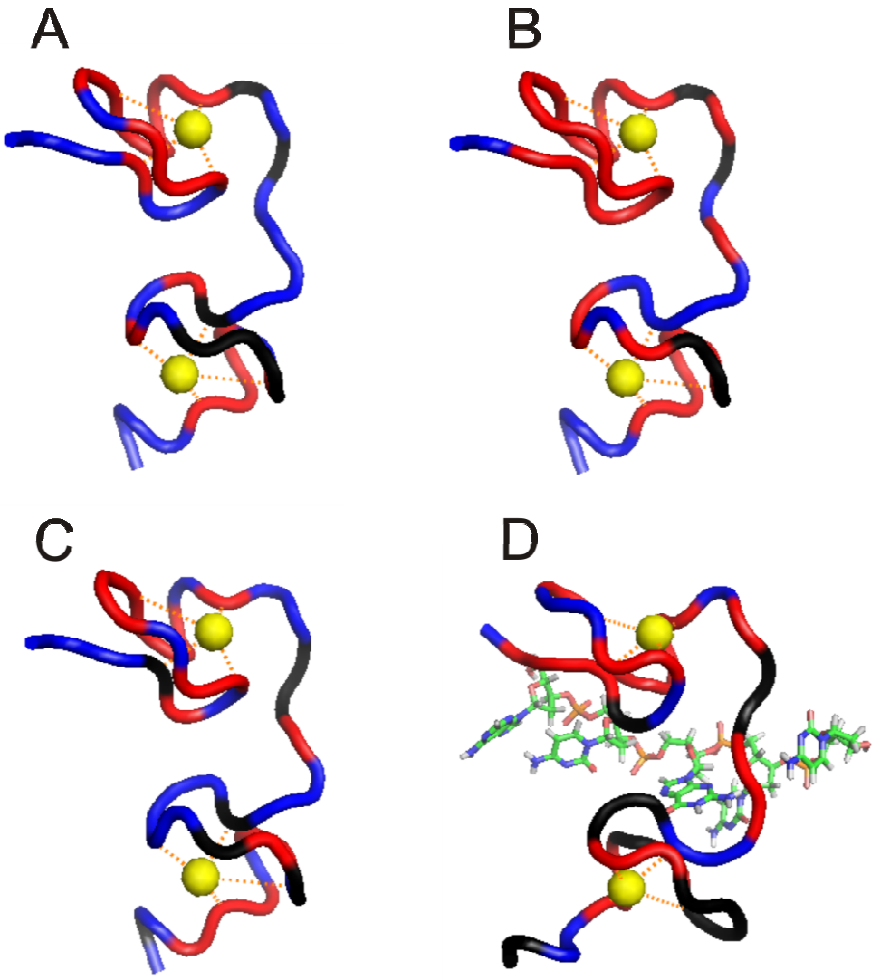


Figure 4.46: Order parameters displayed on NC structures.

Residues with $S^2 \leq 0.75$ are coloured blue and those with $S^2 > 0.75$ are coloured red. Residues which were not analysed due to absence of data (including prolines) are coloured black. A) C-CA_{AA}-NC, B) C-CA-NC, C) C-CA_{AA}-NC + CAI, D) C-CA_{AA}-NC + dACGCC. In D), dACGCC is displayed in stick representation. In all cases, Zn^{2+} ions are shown as yellow spheres, and the residues coordinating the Zn^{2+} are indicated with dotted orange lines.

RDCs show domain independence

The eigenvectors of the alignment tensors for the C-CA_{AA} domain of C-CA_{AA}-NC with and without either CAI or dACGCC are shown projected on ribbon diagrams in Figure 4.47. The elements of the alignment tensors were tabulated in the corresponding chapters. The alignment tensors for isolated C-CA_{AA}, with and without CAI, are tabulated in Table 4.16 and the orientations shown in Figure 4.48.

By comparing the various alignment tensors, the RDC data for -CA_{AA} and C-CA_{AA}-NC, with and without either dACGCC or CAI, are supportive of there being no communication between the different domains, as required by the first assertion of this discussion. This is demonstrated by the alignment tensors of C-CA_{AA}, its corresponding domain in C-CA_{AA}-NC and its corresponding domain in dACGCC-bound C-CA_{AA}-NC having almost the same orientations. That is, the presence of the NC and SP1 domains do not affect C-CA_{AA} alignment, and neither does applying a considerable perturbation to the NC domain.
The alignment tensors with CAI are almost identical for single-domain C-CA_{AA} and the C-CA_{AA} domain in C-CA_{AA}NC. This is consistent with the first assertion made at the outset of the current discussion and, therefore, it is concluded that CAI interacts with isolated C-CA_{AA} and the corresponding domain of C-CA_{AA}-NC in the same way. The results differ only the magnitudes of the alignment tensors, which are most likely due to the relative strengths of aligning media; data for C-CA_{AA} were recorded in a 7% crosslinked gel, whereas a 5% crosslinked gel was used for C-CA_{AA}NC.

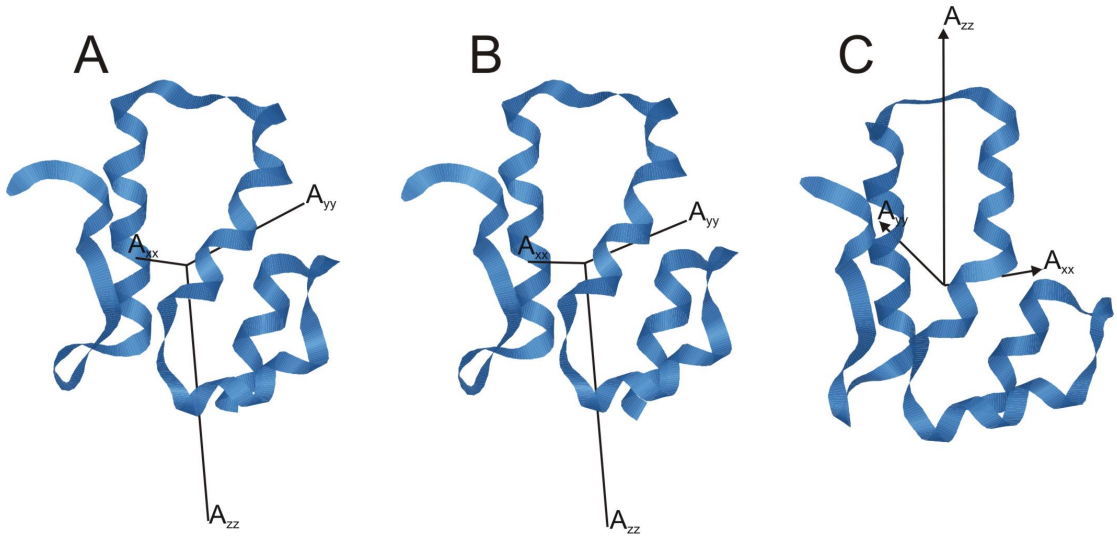


Figure 4.47: Alignment tensors for the C-CA_{AA} domain of C-CA_{AA}NC.

A: free, B: with dACGCC, C: with CAI.

Table 4.16: Alignment tensors for the isolated C-CA_{AA} domain with and without CAI. Angles are reported in degrees.

Molecule	α	β	γ	A_a x 10^4	A_r x 10^4
C-CA_{AA}	-155±2	167±1	-66±2	5.54±0.04	2.03±0.09
C-CA_{AA}+CAI	-164±2	12±1	132±1	9.82±0.04	2.03±0.11

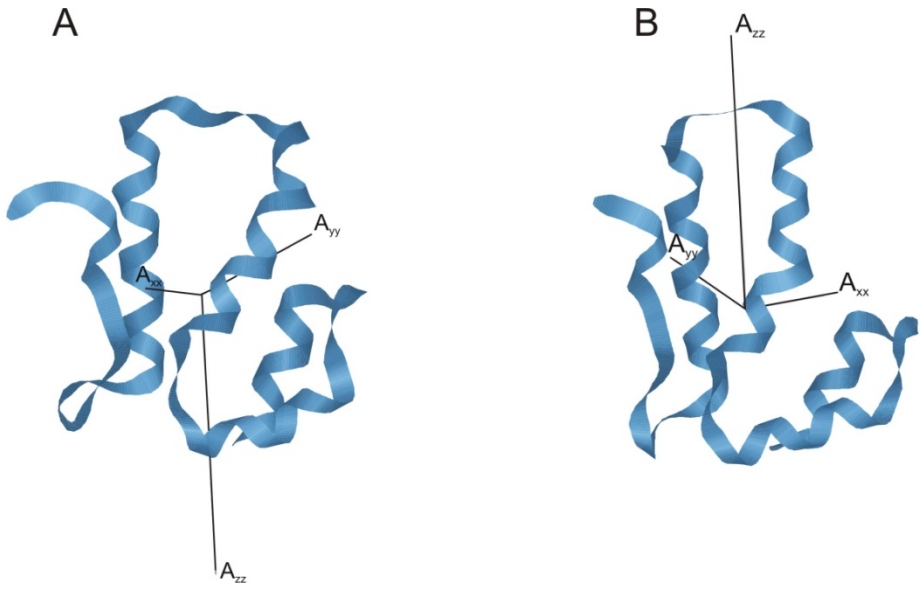


Figure 4.48: Alignment tensors for C-CA$_{AA}$.

A: free, B: wih CAI.

4.10.2. Data comparisons demonstrating the separation of dimerisation and NC domain

The second assertion made at the outset of the present discussion was that the NC domain has no role in C-CA dimerisation. This is evidenced by the similarity in chemical shifts between C-CA and C-CA-NC, which are compared in Figure 4.49. Evidence is also provided by the similarity of diffusion tensors, internal dynamics and chemical shifts between C-CA-NC and C-CA$_{AA}$-NC. Such data all show that the linker and NC domains behave the same whether they are components of monomeric C-CA$_{AA}$-NC or dimeric C-CA-NC, and the comparisons con be seen in Figures 4.50 and 4.51.

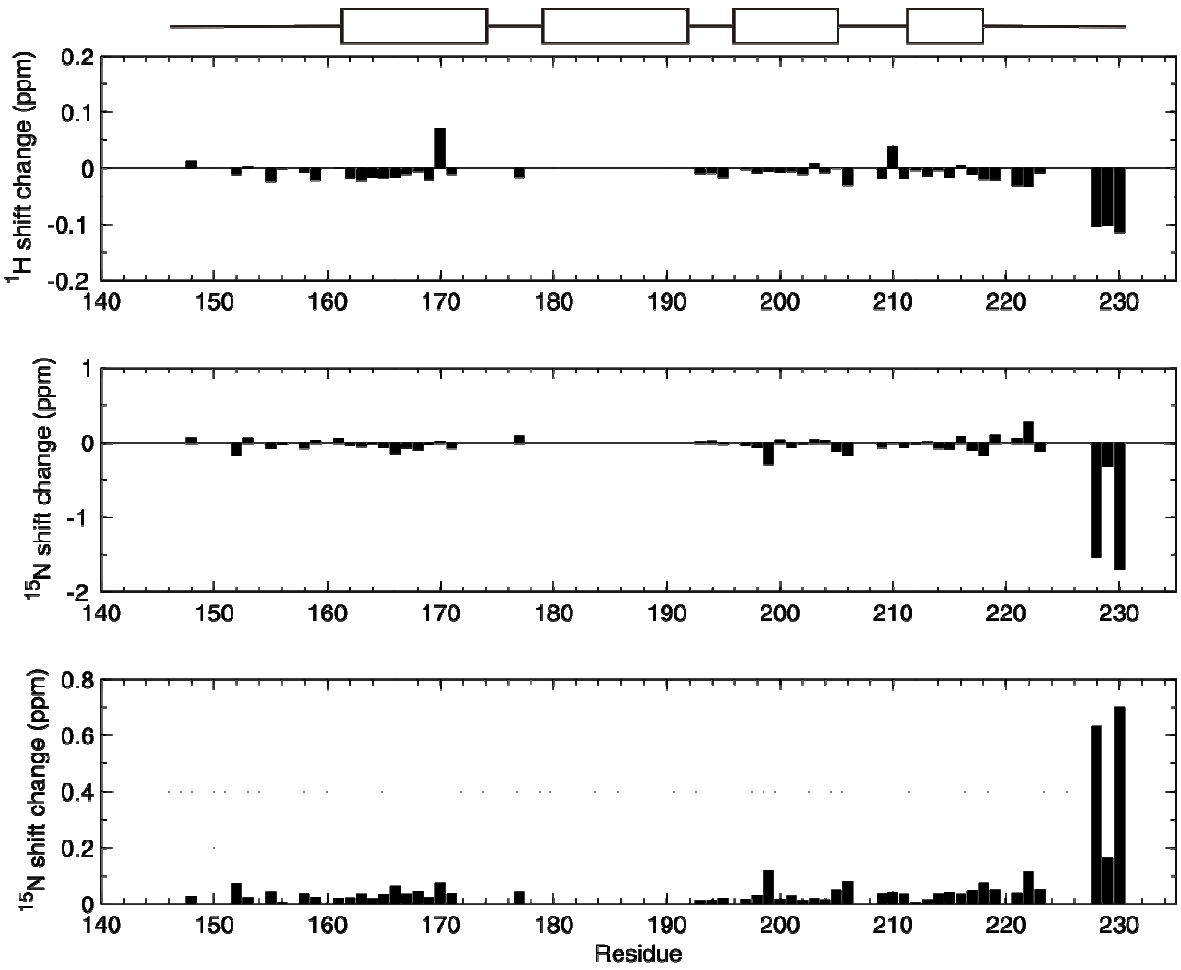


Figure 4.49: Chemical shift differences between C-CA and the C-CA domain of C-CA-NC.

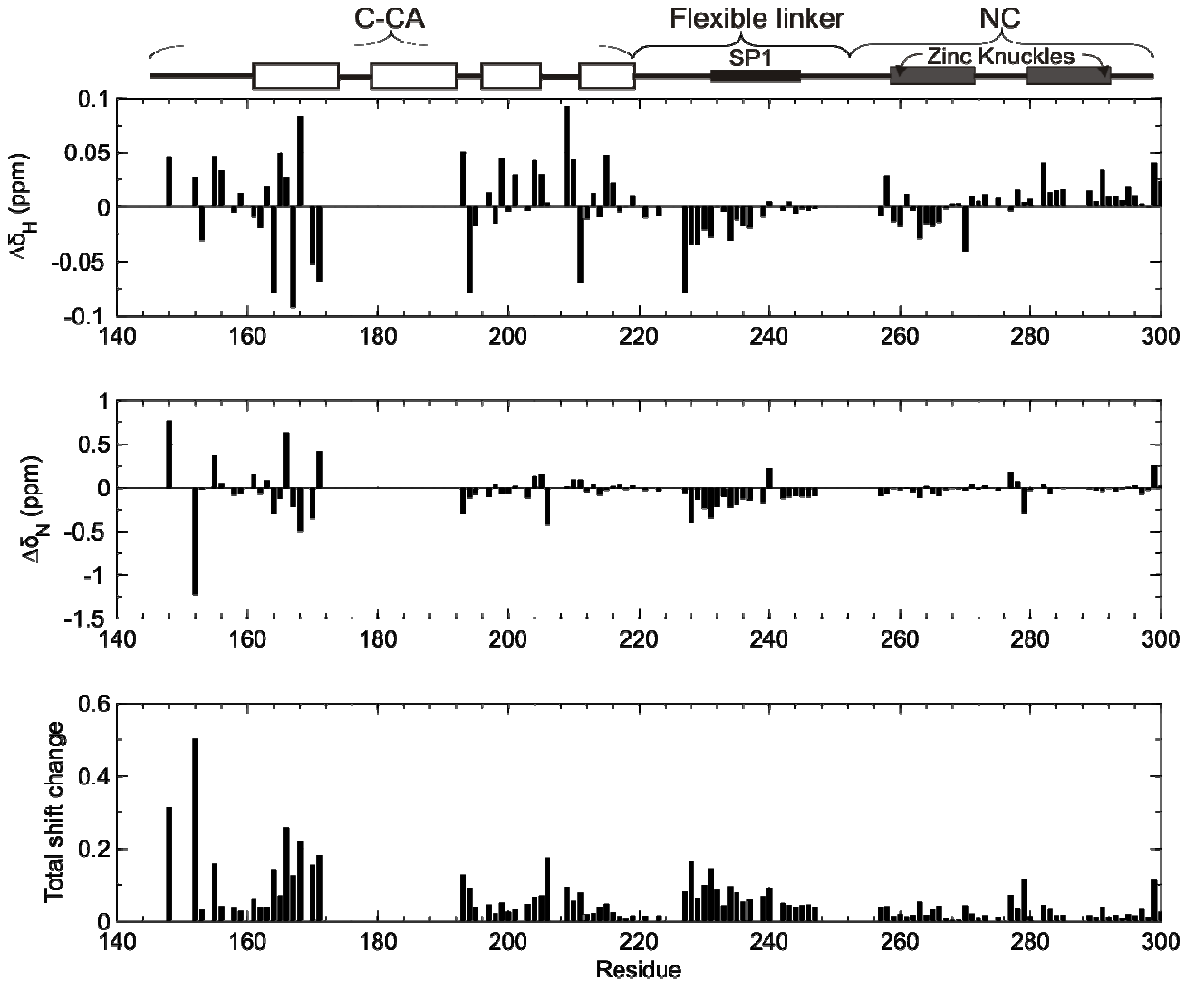


Figure 4.50: Chemical shift differences for amide positions between C-CA-NC and C-CA$_{AA}$-NC.

Where mutual assignments were made for C-CA-NC and C-CA$_{AA}$-NC, ^{15}N and ^{1}H chemical shift changes are shown. Changes are defined as δ(C-CA-NC) - δ(C-CA$_{AA}$-NC). Shift changes were calculated for 99 mutually assigned, resolved resonances.

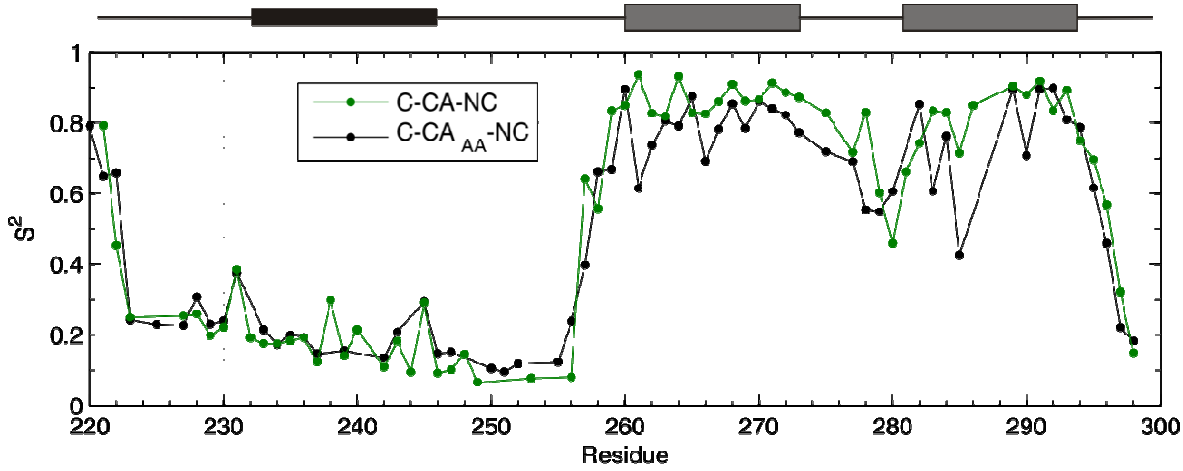


Figure 4.51: Order parameter comparison for the linker and NC domains in C-CA$_{AA}$-NC and C-CA-NC.

Order parameters are generally consistent between the two different constructs. The Zinc knuckles are indicated with grey boxes, and the SP1 domain with a black box.

It was not possible to obtain data of high enough quality to extract spectral density parameters for the C-CA domain of C-CA-NC and, since relaxation data were collected at different fields, direct comparison of relaxation data between C-CA as an isolated domain and as a component of C-CA-NC is difficult. Hence determining whether the internal dynamics of C-CA are different from C-CA-NC is not possible by a direct comparison of measured data. However, a comparison of the internal dynamics of the two molecules is very much desirable, in order to determine whether the linker or NC have any effect upon the C-CA internal dynamics. Therefore, to accomplish such a comparison, the spectral density parameters of C-CA were used in conjunction with the measured correlation time of the corresponding domain in C-CA-NC to predict the relaxation times in C-CA-NC. A comparison of measured and predicted relaxation can thus reveal differences in internal dynamics between the C-CA domain in the two different contexts studied. Such a comparison is shown in Figure 4.52. There are some outliers, but in general the measured and predicted relaxation times are in reasonable agreement, showing that the internal dynamics of C-CA-NC are quite similar to those of C-CA. This provides further evidence that the linker and NC have no effect upon C-CA dimerisation.

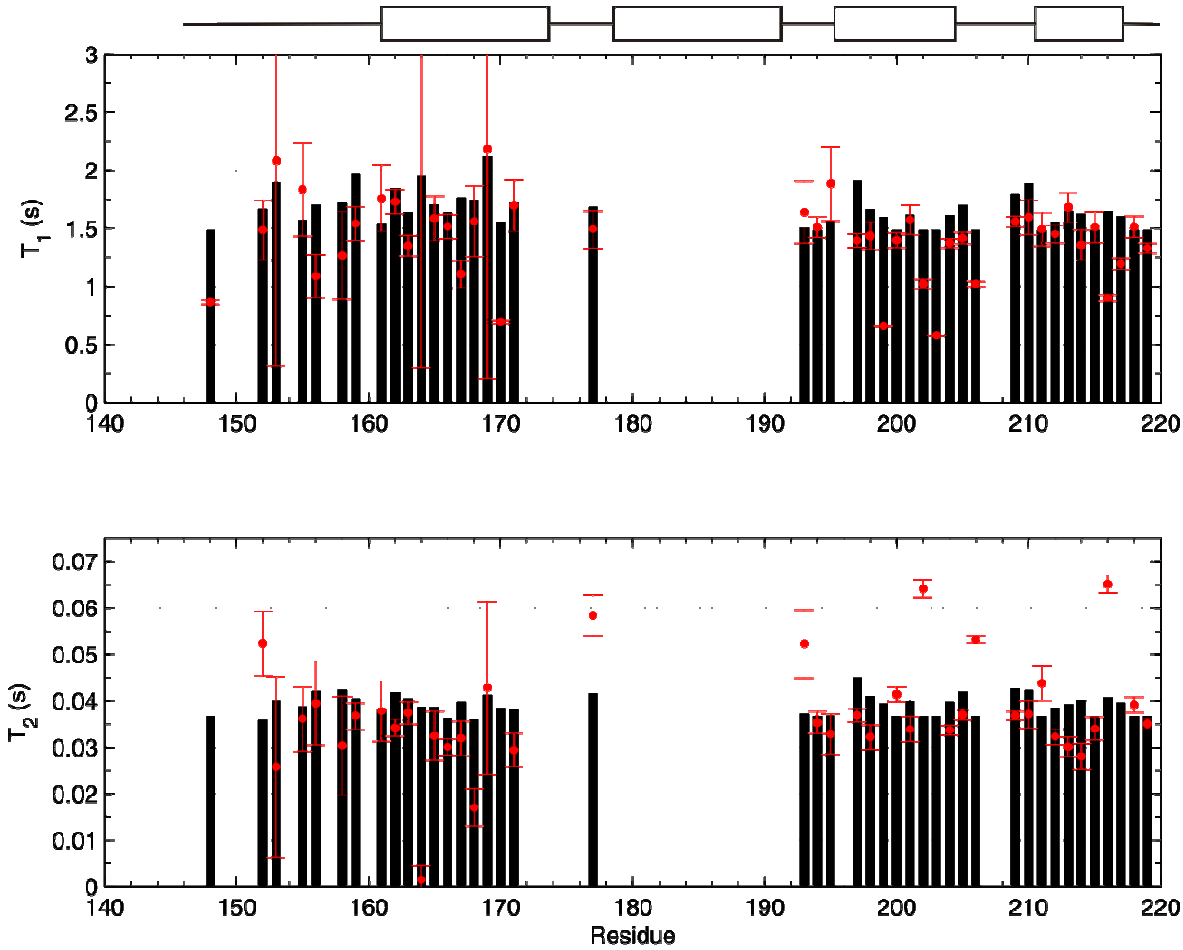


Figure 4.52: Relaxation times for the C-CA domain of C-CA-NC predicted from the spectral density parameters of C-CA.

The measured C-CA-NC relaxation times at 500 μM are shown in red, and the predicted relaxation times from the C-CA model-free analysis are shown in black, assuming an isotropic diffusion tensor with τ_m = 17.4 ns as measured for C-CA-NC.

How different is C-CA to C-CA$_{AA}$?

Various structures of the C-CA dimer have been solved by X-ray crystallography, and that of C-CA$_{AA}$ has been solved by NMR. Collectively, it is apparent that the wild-type C-CA domain and the C-CA$_{AA}$ mutant have similar overall structures, and are both flexible. The model-free analyses of C-CA and C-CA$_{AA}$ show that the order parameters in the regions distal to the dimer interface are very similar between the 2 molecules, even though other features of the internal dynamics are different. The order parameters for C-CA are compared to those for C-CA$_{AA}$ in Figure 4.53.

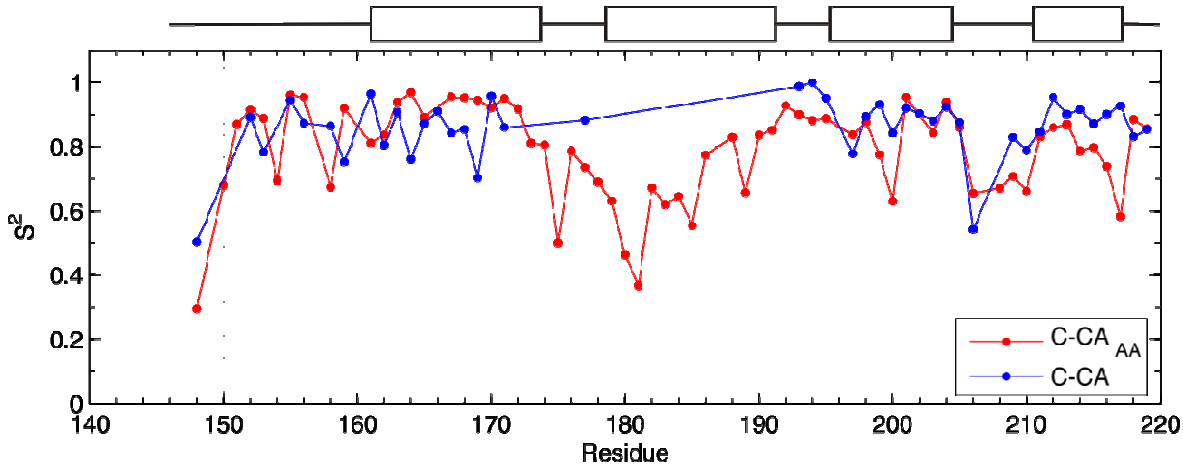


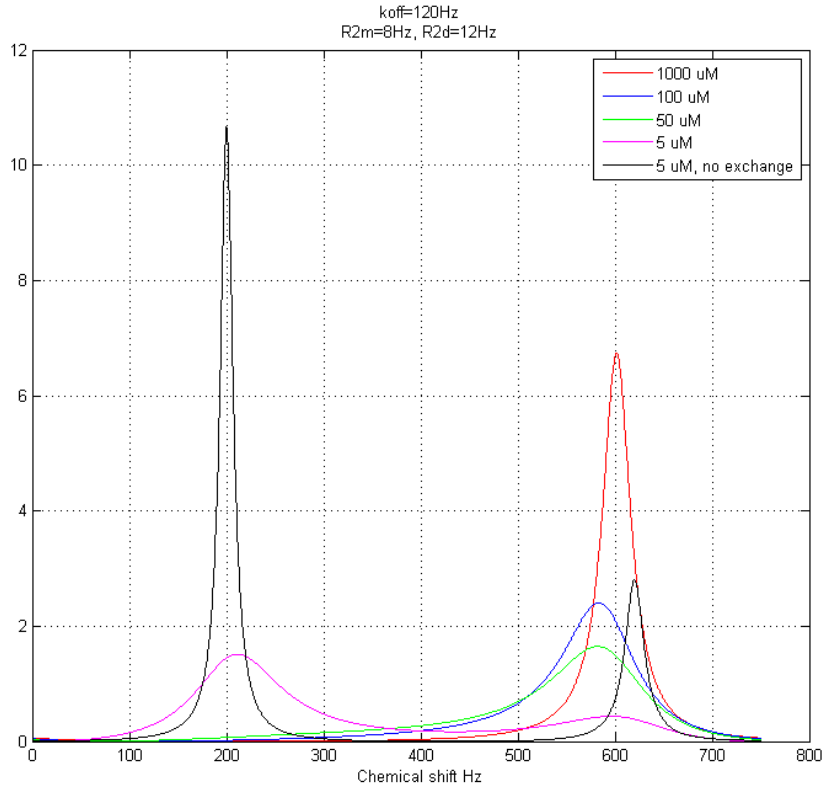
Figure 4.53: Comparison of order parameters for C-CA$_{AA}$ and C-CA.

4.11. The nature of C-CA dimerisation

We wish to determine whether monomer/dimer exchange alone can account for the observed linewidths of C-CA, and will do so using the Bloch-McConnell formalism. Since the linewidths are invariably too broad to be measurable in the dimer interface, the task is essentially to determine the lowest possible signal magnitudes which monomer/dimer exchange permits, in a realistic range of chemical shift differences between monomer and dimer states, using $R_{2,\mathrm{dimer}}$ values predicted from the dimer crystal structures of C-CA and $R_{2,\mathrm{monomer}}$ values as measured for the non-dimerising mutant C-CA$_{AA}$. By reference to chapter 2.4 it is seen that the R_2 predicted by rotational diffusion alone is the upper limit for relaxation rate, or equivalently the lower limit of T_2. A slower R_2 than that predicted by rotational diffusion implies a non-rigid site (S_2<1) whilst a faster R_2 implies an exchange process on a slower timescale than rotational diffusion. The task of finding the minimum signal amplitude is then essentially equivalent to maximising the line-broadening terms given by the **c** matrix described in the introduction, which contains the linewidths. It is generally the case in an NMR spectrum that the intensity maximum of a peak determines whether it will be visible above the noise. Therefore it is the intensity maximum which will be used as the parameter of interest.

In Figure 4.54, simulations are shown which demonstrate that monomer/dimer exchange is insufficient to account for the absence of signals around the dimer interface. The simulations were carried out by modelling dimerisation as a simple reaction, using the equations derived in 4.13. It is also shown that a model in which the monomeric form of C-CA has a single conformation, but the dimer can exist in 2 distinct conformations, can

account for the observed absence of signal intensity. This is referred to as the flexible dimer model, and the equations and method of simulation is detailed in the appendix.



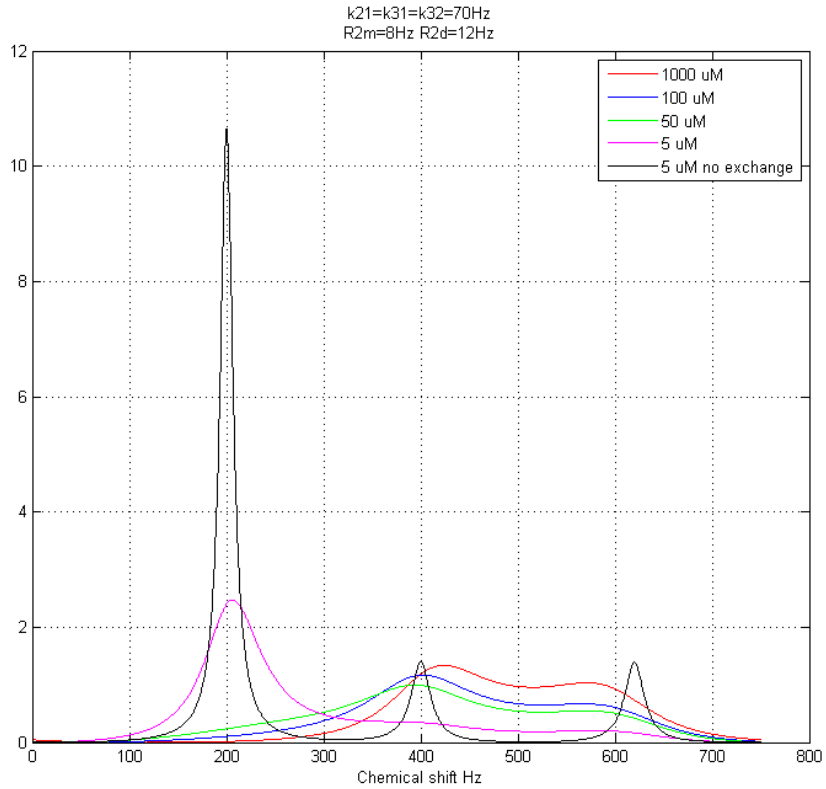


Figure 4.54: Simulations of the expected signals arising from peaks at the C-CA dimer interface.

Simulations were performed using (top) a simple monomer/dimer exchange model and (below) a model in which the dimer has two conformations between which it can exchange. The simple model predicts strongly concentration-dependent signal intensity, whereas the flexible model predicts very low signal over a broad range of concentrations. The intensity scales are in arbitrary units, but are comparable between figures. In each case the black peak is a reference in the absence of exchange. The signal integrals are normalised to one another such that signal intensities are independent of concentration.

4.12. Summary of results

In this chapter, it has been shown that the NC domain of C-CA-NC and of C-CA_{AA}-NC has very similar overall and internal dynamics and structure. That is, dimerisation of the C-CA domain has no modulating effect upon the structure or dynamics of NC. Such dynamics are on both fast and slow timescales.

The dynamics of the C-CA_{AA} domain in the context of C-CA_{AA}-NC have also been determined, demonstrating fast internal motion of backbone NH bonds throughout the C-CA_{AA} domain. Such dynamics are very similar to isolated C-CA_{AA}, which implies that the NC domain does not have any effect upon the C-CA_{AA} domain. Consistent with this observation is the result that the diffusion tensor has a similar orientation and shape in either isolated C-CA_{AA} or in the C-CA_{AA}-NC. Using RDCs, the alignment tensor for the C-CA_{AA} domain of C-CA_{AA}-NC has been determined and found to be in very good agreement with that of isolated C-CA_{AA}, which is again consistent with the C-CA_{AA} domain having no interaction with the NC domain.

It has also been shown that the flexible linker, between residues 220 and 256, is apparently unaffected by C-CA dimerisation. This is on the basis of minor changes in chemical shift for this region between C-CA-NC and C-CA_{AA}-NC, and of only slight changes in the Cα secondary shifts. The dynamics for the linker residues are also very similar between C-CA-NC (at both 100 and 500 μM) and C-CA_{AA}-NC.

4.13. Discussion

The linker is unaffected by dimerisation (probably)

As described in the introduction, the flexible linker region contains the SP1 domain, a 14-residue peptide from 232-245. Various studies have suggested that it can form an α-helix, and that the ability to form an α-helix is necessary for immature assembly. Of the most immediate relevance is a study in which the secondary Cα shifts of C-CA-NC were determined, with slightly positive values in the SP1 region. This was interpreted as a transient helix in equilibrium with random coil conformations. The results presented here for C-CA-NC are in good agreement with this study, and the C-CA_{AA}-NC results are also in good agreement. That is, consecutive slightly positive Cα shifts for the SP1 region of C-CA_{AA}-NC are observed. This implies that if the linker has a tendency to form an α-helix in dimeric C-CA-NC, it has an equal tendency to do so monomeric C-CA_{AA}-NC, and as such dimerisation of the C-CA domain does not affect its structure. The NH chemical shift differences for the linker region between C-CA-NC and C-CA_{AA}-NC are small, which can be interpreted as meaning they adopt similar ensembles of structures. Hence there is no clear evidence in the data presented here that the linker region either adopts a helical structure, in either monomeric or dimeric states of the C-CA domain, or that it is in any way altered by C-CA dimerisation. If self-interactions are to occur at all, then in the C-CA-NC dimer they should be reasonably likely to do so, since they would be intramolecular, so some noticeable difference in NH or Cα shifts should be observed. That this is not the case is reasonable evidence that self-interactions do not occur in low order assembly intermediates at least. However, a cET study of immature HIV-1 presented evidence that the linker region forms a 6-helix bundle in the immature capsid, inferred from 6-fold symmetry in the electron density map for this region (after averaging) (159). If such a helical bundle is to form, therefore, it must do so only when higher-order assembly intermediates than a dimer are reached. It is of course possible that the cET observation is not of a helical bundle, for such an arrangement is one of many which are consistent with the low resolution electron density map. An ensemble of flexible structures could be accommodated, for example, as noted in a similar cET study (25).

It may be possible that, if the linker region is constrained by dimerisation of C-CA at its N-terminus and NC-nucleic acid interactions at its C-terminus, as is believed to occur in the immature virion, then its flexibility would be sufficiently limited to permit self-interactions. This is supported by the observation of a 3_{10} helix in the N-terminal 12 residues of the NC-SL3 and NC-SL2 NMR structures (10, 39). Such a helix could propagate through the rest of the linker in the immature capsid. However, it is difficult to

reconcile this suggestion of the ability to form a helix with the apparent *necessity* for helical structure.

Why is SP1 removal the last PR cleavage?

It is interesting that, in the series of proteolytic reactions that separate Gag into its constituent domains and lead to formation of the mature capsid, removal of SP1 from CA is the slowest reaction catalysed by the viral protease PR (156). There are 2 questions to be addressed; firstly, why is SP1 removed at all? Secondly, why is its removal from CA the slowest reaction catalysed by PR?

It is likely that the proposed folding of the SP1 region into a stable helix in the immature state carries a considerable entropic penalty, that is, results in a loss of entropy. Of course, this makes immature assembly an unfavourable process in the absence of a compensating negative enthalpic change. Even if no such helix forms and the linker remains flexible in the immature assembled state, it probably still loses some entropy on account of at least some restriction of the conformational space accessible to it. This may be the reason for the SP1 region being removed from CA by PR, which appears to be important in mature assembly. In the case of immature assembly, the compensating enthalpic change may come from the NC-nucleic acid interactions, without which immature assembly does not occur. However, for mature assembly, the removal of the SP1 region from CA may result in a sufficient reduction in the entropic penalty concomitant with the assembly process that it becomes thermodynamically viable.

Thus a simple thermodynamic argument may answer the first question proposed at the outset of this discussion; SP1 is removed because the loss in entropy concomitant with mature assembly would, without the compensating NC-nucleic acid interaction's enthalpic contribution, render mature assembly thermodynamically unfavourable.

To answer the second question, it is necessary to consider whether it is actually a requirement of the mature assembly process that CA-SP1 cleavage be slower than the other proteolysis reactions. Since this remains unknown, further investigation is required before a satisfactory answer can be given, but it is possible to speculate nonetheless. The basis for target specificity in HIV-1 PR is not fully understood, and it is thought that it involves a complex interplay between sequence and structure (116, 124). Determining whether slow CA-SP1 cleavage is necessary or coincidental should therefore be solved by mutagenesis

around the CA-SP1 site, in which sequences are sought which allow rapid cleavage but which are not inhibitory to immature assembly. If it transpires to be necessary for CA-SP1 cleavage to be slow, then it is possible that the apparent ability of the region around SP1 to form a transient helix, as well as the primary structure of this region, have roles in regulating the rate of CA-SP1 cleavage. Such studies should keep in mind the fact that the SP1-NC cleavage site may also be helical, as shown by the NC-SL3 and NC-SL2 NMR structures, but that cleavage at this site is much more rapid than at the CA-SP1 site.

Dimerisation is unaffected by linker or NC

The NH and Cα shifts for isolated C-CA are very similar to those of the C-CA domain of C-CA-NC. The absence of signals around the dimer interface was an equal problem in both cases. However, as shown, even an absence of signal can be revealing, for it was shown that a model of dimerisation in which the dimer has at least two conformations between which slow exchange occurs is consistent with the loss of signal intensity over a broad concentration range, whereas a conformational selection event is not. Therefore, it is concluded that C-CA dimerises in the same manner as the C-CA domain of C-CA-NC, with the important implication that neither the linker nor NC effects any conformational selection.

Implications for the mechanism of immature assembly

It is reasonably well established from cEM and cET data that the interactions in which C-CA are involved are different in the immature state from those in the mature state. The data presented in this chapter demonstrate that such differences are in no way mediated by any direct interaction between the C-CA domain and either the linker or NC domains.

4.14. Appendix:

4.14.1. Interpreting relaxation data for C-CA and C-CA-NC

Both C-CA and C-CA-NC exchange between monomeric and dimeric states, and it is likely that their dimeric states have several conformations amongst which exchange also occurs. In the interpretation of relaxation, two major problems are encountered. Firstly, the relaxation decay curves may not be mono-exponential in form, since, as will be proven in the following sections, only if exchange is significantly faster than relaxation are the relaxation decay curves truly mono-exponential. If two-site exchange occurs, relaxation is

bi-exponential, and the meaning of a mono-exponential approximation to bi-exponential relaxation data will be explored.

The second problem is the form of the auto-correlation function, which is required to extract diffusion tensor parameters, including the isotropic correlation time. Only if a mono-exponential approximation to the auto-correlation is made can a diffusion tensor be determined, else the problem is under-determined. Therefore, the meaning of a mono-exponential auto-correlation function is explored, to try and determine whether the diffusion tensor parameters for C-CA and C-CA-NC are meaningful.

The form of the relaxation decay curve in the presence of exchange

Many peaks for residues in C-CA and C-C-NC which are suitably far from the dimer interface do not seem to be affected by exchange. However, in order to interpret relaxation times for these peaks one must still account for exchange in the sense that a single relaxation time (T_1 or T_2) is insufficient to describe linewidth when monomer and dimer species are present. We must therefore determine the form of the NMR signal for peaks which do not change in chemical shift as result of monomer/dimer exchange, but which change in relaxation time on account of the larger size of the dimer as compared to dimer, which will be referred to as "unshifted exchange". We will treat the dimer as a single state for such peaks, since they are apparently unaffected by events at the dimer interface. Hence the monomer/dimer equilibrium model should be sufficient here. No additional simplifications are made, except the obvious constraint that the chemical shift is unaltered by dimerisation for such peaks. We will call this single chemical shift Ω_0. Recall that the propagator for the time-domain signal is given in the classical formalism by

$$\mathbf{P} = \mathbf{U}\exp(\mathbf{D}t)\mathbf{U}^{-1}$$

In which

$$\mathbf{D} = \mathbf{U}^{-1}(i\boldsymbol{\Omega} - \mathbf{R} + \mathbf{K})\mathbf{U}$$

Refer to chapter 2.5 for the definitions of the other symbols. To find an analytical expression for the NMR signal, **U** and **D** are required, for which the characteristic equation for the secular determinant is evaluated. This was accomplished with Matlab, and with some manipulation the real and imaginary parts of the eigenvalues (i.e. the diagonal elements of **D**) are, following the notation of chapter 2.5, given by

$$d_{jj} = \Omega_0$$

$$c_{\pm} = \frac{1}{2}\{R_M + R_D + 2k_{12}A + k_{21} \mp [(R_M - R_D + 2k_{12}A - k_{21})^2 + 8k_{21}k_{12}A]^{1/2}\}$$

In which $c_+=c_{11}$ and $c_-=c_{22}$. These expressions are somewhat lengthy, but can be simplified by use of the following definitions:

$$k_{ex} = k_{21} + 2k_{12}A$$

$$R = R_M + R_D$$

$$\Delta R = R_M - R_D$$

It is also useful to define P_M and P_D as the mole fractions of monomer and dimer respectively, whose sum is unity. By consideration of the reaction at equilibrium, one has

$$k_{21} = P_M k_{ex}$$

$$2k_{12}A = P_D k_{ex}$$

$$\Delta P = P_D - P_M$$

The real parts of the eigenvalues (i.e. relaxation rates) can now be written

$$c_{\pm} = \frac{1}{2}\{R + k_{ex} \mp \psi^{1/2}\}$$

In which

$$\psi = [(\Delta R + \Delta P k_{ex})^2 + 4P_M P_D k_{ex}^2]$$

The forms of the eigenvalues can be heavily simplified if both rate constants are either much greater than, or much less than, either of the relaxation rates, such conditions constituting alternative definitions of fast and slow exchange respectively in the limit of small or zero chemical shift differences.

The eigenvectors are also required, which determine the amplitudes of each contributing signal. Since there is no chemical shift difference, this task is simpler than the general case. The eigenvectors are the columns of **U**, which can be written

$$\mathbf{U} = \begin{pmatrix} 1 & 1 \\ (\Delta R + \Delta P k_{ex} + \psi^{1/2})/2k_{21} & (\Delta R + \Delta P k_{ex} - \psi^{1/2})/2k_{21} \end{pmatrix}$$

In general, if a 2x2 matrix has the form

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then its inverse is given by

$$\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Hence the inverse of **U** is

$$\mathbf{U}^{-1} = \frac{1}{\psi^{1/2}}\begin{pmatrix} (\Delta R + \Delta P k_{ex} - \psi^{1/2})/2 & -k_{21} \\ (-\Delta R - \Delta P k_{ex} - \psi^{1/2})/2 & k_{21} \end{pmatrix}$$

Note that both **U** and $\mathbf{U}^{-1}$ contain only real elements, since the imaginary parts would normally be functions of the chemical shift difference, which is zero in this special case. The signal amplitude coefficients are given by

$$a_{11} = \frac{\left(\Delta R + \Delta P k_{ex} - \psi^{1/2}\right)}{\psi^{1/2}}$$

$$a_{12} = \frac{-2k_{12}}{\psi^{1/2}}$$

$$a_{21} = \frac{\left(-\Delta R - \Delta P k_{ex} - \psi^{1/2}\right)(\Delta R + \Delta P k_{ex})}{2k_{21}\psi^{1/2}}$$

$$a_{22} = \frac{\Delta R + \Delta P k_{ex}}{\psi^{1/2}}$$

Such terms are found using Matlab, then re-written in the above form after some manipulation. In the absence of a chemical shift difference, all signal amplitudes are real numbers. The frequency-domain signal is

$$S(\omega) = \frac{c_+\left(a_{11}M_1(0) + a_{12}M_2(0)\right)}{c_+^2 + (\omega - \Omega_0)^2} + \frac{c_-\left(a_{21}M_1(0) + a_{22}M_2(0)\right)}{c_-^2 + (\omega - \Omega_0)^2}$$

This is a sum of two absorption Lorentzian functions, centred at Ω_0, with linewidths given by the real parts of the eigenvalues of **V**. Note that the monomer and dimer populations contribute to each line.

We wish to describe the signal intensity as a function of a relaxation delay time τ, during which transverse magnetisation is allowed to relax. Let us suppose an initial magnetisation is prepared which is given by

$$M^+(0) = \frac{1}{A_T}\begin{pmatrix} A \\ 2A_2 \end{pmatrix}$$

This state is then allowed to evolve for a time τ under the propagator **P** with elements

$$P_{il}(\tau) = a_{il}\exp(-c_{ii}\tau)$$

Note that chemical shift evolution has been ignored, which can be achieved practically by means of the incorporation of an appropriate phase factor if needed. The magnetisation at the beginning of acquisition is given by

$$\mathbf{M}^+(\tau) = \mathbf{P}(\tau)\mathbf{M}^+(0)$$

The frequency-domain signal at Ω_0 after a delay τ is therefore

$$S(\Omega_0;\tau) = \frac{\left(a_{11}M_1(\tau) + a_{12}M_2(\tau)\right)}{c_+} + \frac{\left(a_{21}M_1(\tau) + a_{22}M_2(\tau)\right)}{c_-}$$

$$M_1(\tau) = \exp(-c_+\tau)\left(a_{11}M_1(0) + a_{12}M_2(0)\right)$$

$$M_2(\tau) = \exp(-c_-\tau)\big(a_{21}M_1(0) + a_{22}M_2(0)\big)$$

Finally, we arrive at a bi-exponential relaxation decay curve given by

$$S(\Omega_0;\tau) = \exp(-c_+\tau)\big(a_{11}M_1(0) + a_{12}M_2(0)\big)\left(\frac{a_{11}}{c_+} + \frac{a_{21}}{c_-}\right) + \exp(-c_-\tau)\big(a_{21}M_1(0) + a_{22}M_2(0)\big)\left(\frac{a_{12}}{c_+} + \frac{a_{22}}{c_-}\right)$$

We will now derive simplified expressions which apply when exchange is either fast or slow, but not when exchange is on a similar timescale to relaxation.

Slow exchange

If exchange is slower than relaxation, **U** becomes the identity matrix, in which case so does **a**. The relaxation rates become

$$c_+ = R_M$$
$$c_- = R_D$$

Hence the signal is

$$S(\omega) = \frac{R_M M_1(0)}{R_M^2 + (\omega - \Omega_0)^2} + \frac{R_D M_2(0)}{R_D^2 + (\omega - \Omega_0)^2}$$

This describes the form of the frequency-domain signal given an initial transverse magnetisation vector $\mathbf{M}^+(0)$. To find the relaxation decay curve, the same simplifications are applied to the bi-exponential decay to give

$$S(\Omega_0;\tau) = \frac{M_1(0)}{R_M}\exp(-R_M\tau) + \frac{M_2(0)}{R_D}\exp(-R_D\tau)$$

This is simple and intuitive result. However, resolving the two components is very difficult, and a mono-exponential approximation must be used practically. The agreement between a mono-exponential function and simulated data is given in Figure 4.56 to demonstrate this.

Fast exchange

If exchange is fast, the two eigenvalues are very different, with c_+ giving a sharp line and c_- a very broad line. The latter line is basically unobservable, so only the c_+ terms need be considered. This eigenvalue reduces to a population-weighted average relaxation rate, which is written

$$c_+ = P_M R_M + P_D R_D$$

Ignoring the amplitude of the decay curve, which is treated as a free parameter in experimental situations, the relaxation decay curve becomes a mono-exponential of the form

$$S(\Omega_0;\tau) \propto \exp(-c_+\tau)$$

Hence, if exchange is faster than relaxation, and there is little or no chemical shift change, relaxation is mono-exponential. The population-weighted average is accurate to within ~ 5% even when k_{21} is around the same as the relaxation times. A simulation to show that the approximation is valid when k_{21} = 100 Hz is given in Figure 4.55.

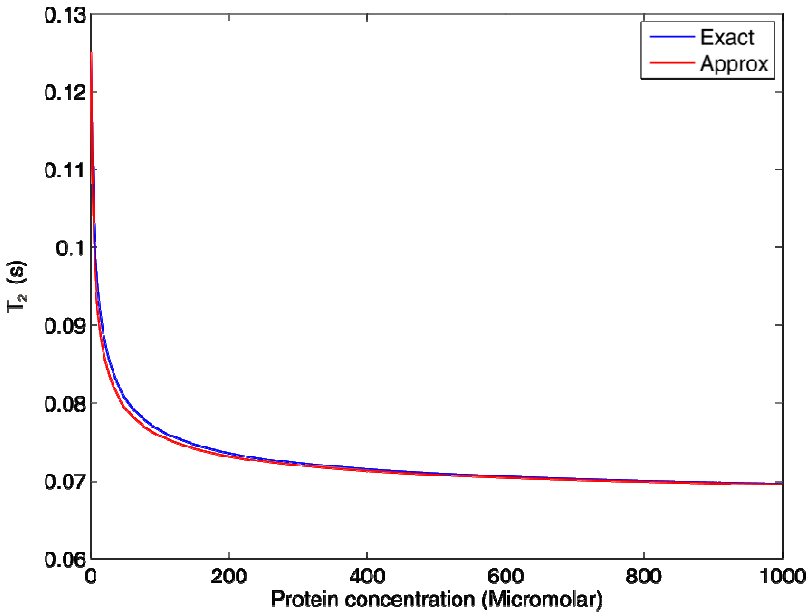


Figure 4.55: Simulated relaxation times in the fast exchange regime.

Simulation of the exact T_2 and approximation by a population-weighted average for monomer/dimer exchange when relaxation is faster than exchange, demonstrating the validity of the population-weighted average approximation when k_{21} = 100 Hz.

Range of validity of the mono-exponential approximation

In analysing relaxation data, the approximation to a mono-exponential function of the relaxation time was made, both for T_1 and T_2 determination. Here, some computer simulations using the theory of unshifted exchange are presented, so as to show the range of validity of this approximation at 0.1 mM and 1 mM C-CA concentration. Simulations of relaxation decay curves in the range of relaxation delays used for actual measurements were conducted for a wide range of k_{ex} values. To each simulated relaxation decay curve, a mono-exponential function was fitted to determine an effective T_2 and see how good a fit to the data the mono-exponential function should be. At 1mM concentration, with the

monomer and dimer T_2 0.125 s and 0.0667 s respectively, it is seen that the mono-exponential approximation gives effective T_2 values close to the dimer T_2 for $k_{ex} > 10$ Hz, consistent with the transition to the fast exchange regime at this value of k_{ex}. If $k_{ex} \leq 10$ Hz, the mono-exponential approximation does not hold, but, as shown in Figure 4.56, still reproduces the simulated data well with an over-estimated T_2. That is, the effective T_2 as determined by a mono-exponential approximation is longer than a population-weighted average T_2. Given the goodness-of-fit of the mono-exponential approximation to bi-exponential relaxation, it is unlikely that distinct relaxation times would be resolvable from real data, making the exact bi-exponential function impractical.

The same results are obtained for simulated T_1 decay curves, so the simulations are omitted for brevity. This has the important implication that, although both T_1 and T_2 are over-estimated if exchange is slow, the T_1/T_2 ratio is still meaningful, and is to a good approximation the population-weighted weighted average T_1/T_2 ratio. If a diffusion tensor is to be calculated from measured relaxation data, slow exchange is not prohibitive to its being meaningful, and the resulting diffusion tensor can be interpreted as a population-weighted diffusion tensor to a reasonable approximation.

An effective T_2 (or T_1) resulting from fitting a mono-exponential function to data which is actually bi-exponential (when $k_{ex} < 10$ Hz) is longer than a population-weighted average. This could be problematic, since in correlation time and internal dynamics analysis, the relaxation times are necessarily interpreted as population-weighted averages. Hence an apparently long T_1 or T_2 can in fact be due to exchange being slower than relaxation, rather than internal dynamics, such that the dynamics analyses for C-CA and C-CA-NC are valid only if exchange is faster than relaxation, in both qualitative and quantitative senses. One must also note that exchange must be faster that longitudinal and transverse relaxation. If, therefore, exchange is on a timescale that falls between T_1 and T_2, then one is in trouble, and the relaxation times will be un-interpretable. This is particularly problematic, at least potentially, for C-CA-NC, where the longitudinal relaxation times are very different to transverse relaxation times. In this case, it is entirely reasonable that k_{ex} may fall between R_1 and R_2.

At 0.1 mM concentration, the same results were found, that is, distinct relaxation times are not resolvable, and fast exchange is a valid assumption with a meaningful effective T_2 for $k_{ex} > 10$ Hz.

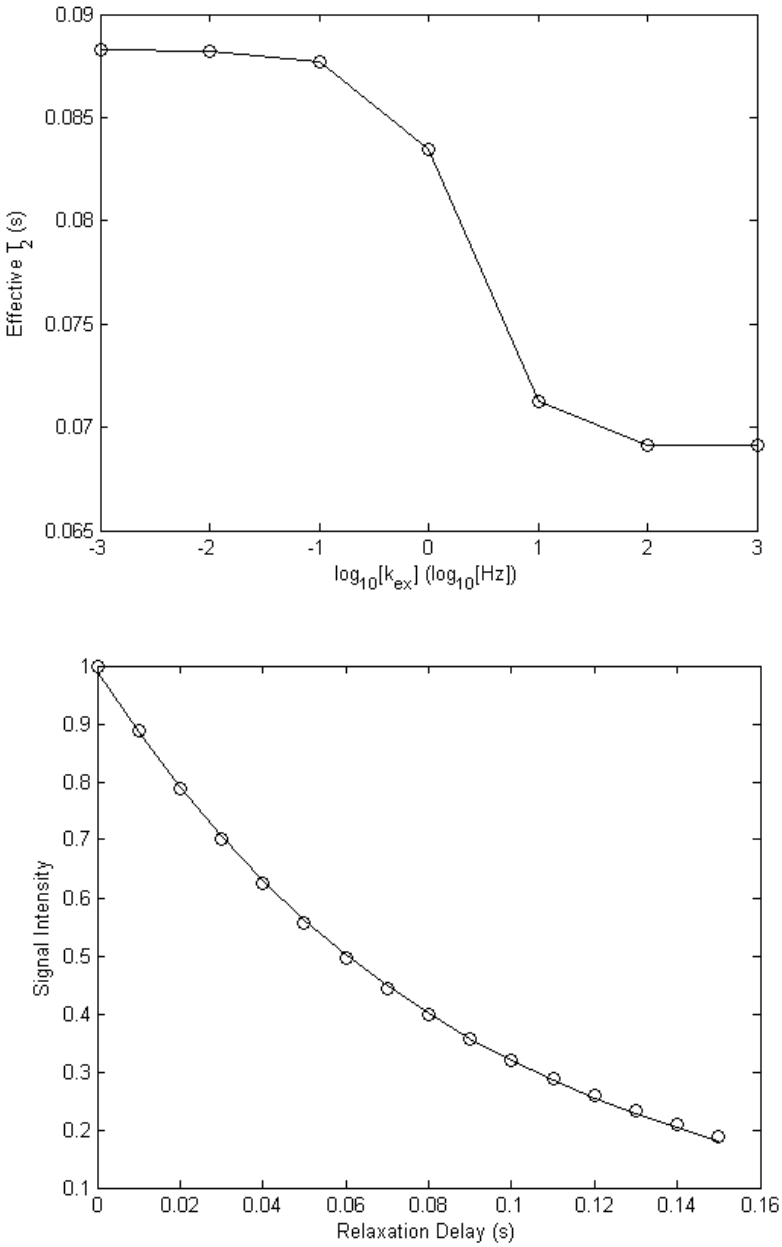


Figure 4.56: Effects of a mono-exponential approximation to bi-exponential relaxation.

A) Effective mono-exponential T_2 as a function of k_{ex} at 1 mM C-CA concentration. If k_{ex} > 10 Hz, the effective T_2 is easily interpretable as a population-weighted average, but not for slower k_{ex}. The population-weighted average T_2 using the simulation parameters is 0.067 s. B) mono-exponential fit (solid line) to slow exchange (k_{ex} = 0.001 Hz) simulated data (circles). The fit is good, but the resulting T_2 cannot be interpreted further. Distinct T_2 times for monomer and dimer would not be resolvable even with modest experimental noise. The relaxation times are in the range used experimentally. Further simulation details are in the text.

The mono-exponential auto-correlation function

Inasmuch as necessarily assuming mono-exponential relaxation for C-CA and C-CA-NC, the assumption of a mono-exponential auto-correlation function must also be made. There are two options for a tractable mono-exponential auto-correlation function. The first has already been described in chapter 4.4, which to re-capitulate assumes exchange to be so fast that the auto-correlation function is truly mono-exponential. This is valid rigorously if exchange is faster than rotational diffusion, which requires $k_{ex} > 1 \times 10^{10}$ Hz, since rotational diffusion is on a nanosecond timescale. The alternative to a rigorous mono-exponential function with a population-weighted average isotropic rotational diffusion rate is the approximation to a bi-exponential function with a mono-exponential function using a single effective correlation time. This is entirely analogous to extracting a single effective relaxation time from bi-exponential data. The effective correlation time can differ significantly from a population-weighted average correlation time, which may compromise the correlation time analysis for C-CA and C-CA-NC.

In Figure 4.57, the observed correlation time as a function of the fractional dimer population is plotted, for the case in which the autocorrelation function is truly mono-exponential, i.e. monomer/dimer exchange is very fast. In this case the isotropic rotational diffusion coefficient is a population-weighted average. This is compared to the effective correlation time, in which a mono-exponential function is fitted to a bi-exponential auto-correlation function. This establishes the relationship between the best mono-exponential approximation to a bi-exponential auto-correlation function, and also shows that the fast-exchange auto-correlation function would yield an observed correlation time which is generally quite close to the effective correlation time of the slow-exchange auto-correlation function, provided that a majority of protein is in either monomeric or dimeric state. In other words, it is still valid to a reasonable approximation (generally within 10 %) to use the fast-exchange mono-exponential auto-correlation function even when it is not valid rigorously, and the resulting correlation time can be interpreted as the reciprocal of a population-weighted average rotational diffusion coefficient.

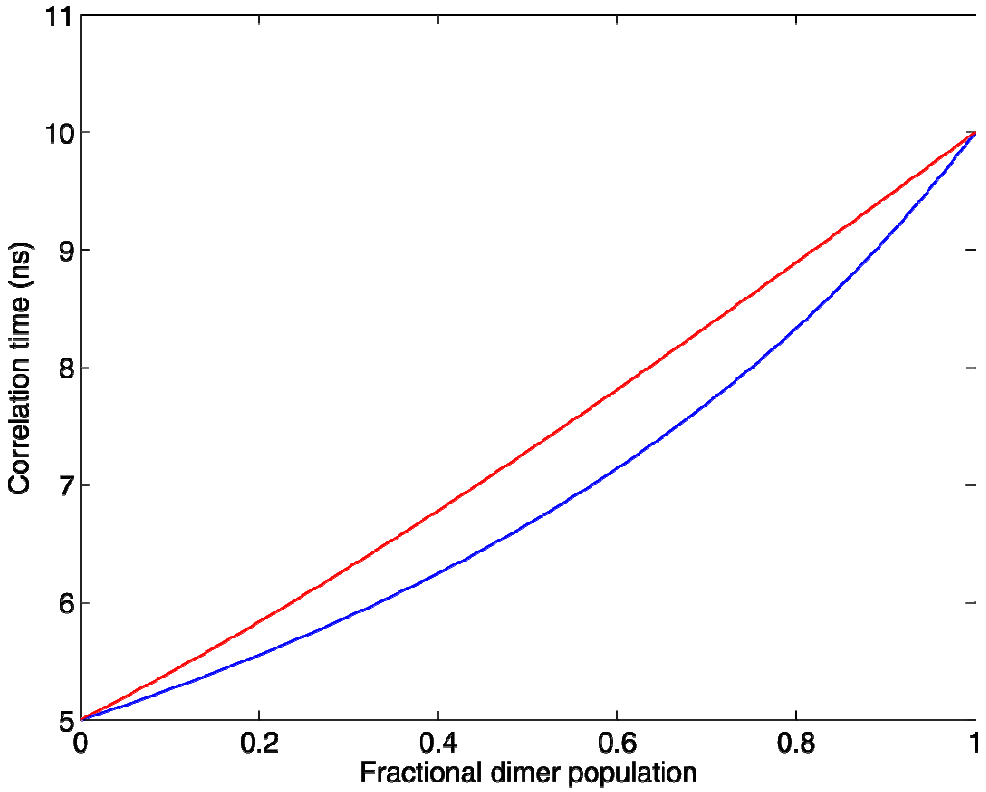


Figure 4.57: Simulated correlation times in the presence of monomer/dimer exchange.

Blue: A mono-exponential autocorrelation is simulated and the correlation time derived from a population-weighted rotational diffusion coefficient plotted. This is how data were interpreted. Red: A more realistic bi-exponential correlation function is simulated, and the effective correlation time from fitting a mono-exponential function is plotted. It is seen that assuming a mono-exponential correlation function leads to reasonably small differences even though it is not rigorously correct.

Appendix:

C-CA dimerisation simulation

We wish to calculate the form of the NMR signal for the C-CA domain, which will be influenced by dimerisation. We wish then to determine whether monomer/dimer exchange alone can account for the observed linewidths and, as will be made clear in coming sections, this is not the case, such that more complex models are sought. Nevertheless, it is important to first justify the statement that monomer/dimer exchange alone cannot account for C-CA linewidths. Therefore we construct a model in which we assume that the only exchange process is the monomer/dimer reaction, and define k_{12} as the forward rate

constant (for dimer formation) and k_{21} as the reverse rate constant (for dimer dissociation). The dissociation constant is then

$$K_d = \frac{A^2}{A_2} = \frac{k_{21}}{k_{12}}$$

In which A is the monomer concentration and A_2 is the dimer concentration. Note that the concentration of C-CA in dimeric state is twice the dimer concentration, since two molecules contribute to each dimer. The initial magnetisation vector is

$$M(0) = \frac{1}{A_T}\begin{pmatrix} A \\ 2A_2 \end{pmatrix}$$

The kinetic matrix is therefore

$$K = \begin{pmatrix} -2k_{12}A & k_{21} \\ 2k_{12}A & -k_{21} \end{pmatrix}$$

The equilibrium concentrations are given in terms of the total protein concentration A_T by

$$A = \frac{K_d}{4}\left\{-1 + \sqrt{1 + 8A_T/K_d}\right\}$$

$$A_2 = \frac{1}{2}\left(A_T + \frac{K_d}{4}\left\{1 - \sqrt{1 + 8A_T/K_d}\right\}\right)$$

The chemical shift and relaxation matrices are given by

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_{monomer} & 0 \\ 0 & \Omega_{dimer} \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} R_{2,monomer} & 0 \\ 0 & R_{2,dimer} \end{pmatrix}$$

With Ω a chemical shift and R_2 a transverse relaxation rate. The simulations presented were performed by obtaining the matrices **a**, **c** and **d** by numerical diagonalisation of the matrix **V** (see chapter 2.5 for the definitions of these terms), from which a frequency-domain signal was calculated.

Flexible dimer model for C-CA dimerisation

The 'chemical states' are monomer (A), and the two dimer conformations (A_2 and A_2') and are indexed, for A, A_2 and A_2' respectively as states 1, 2 and 3. The rate constant for the evolution of a state l to a state m is denoted k_{lm}, and the initial magnetisation vector is defined as

$$M(0) = \begin{pmatrix} A \\ 2A_2 \\ 2A_2' \end{pmatrix}$$

For the calculation of species concentrations at equilibrium, we define association constants by

$$K_{12} = \frac{A_2}{A^2} = \frac{k_{12}}{k_{21}}$$

$$K_{13} = \frac{A_2'}{A^2} = \frac{k_{13}}{k_{31}}$$

$$K_{23} = \frac{A_2'}{A_2} = \frac{k_{23}}{k_{32}}$$

The kinetic matrix is then

$$K = \begin{pmatrix} -2A(k_{12} + k_{13}) & 2k_{21} & 2k_{31} \\ 2k_{12}A & -(k_{23} + k_{21}) & k_{32} \\ 2k_{13}A & k_{23} & -(k_{32} + k_{31}) \end{pmatrix}$$

Equilibrium concentrations are given by

$$A = \frac{-1 + \sqrt{1 + 8A_T(K_{12} + K_{13})}}{4(K_{12} + K_{13})}$$

$$A_2 = \frac{A_T - A}{2(1 + K_{23})}$$

$$A_2' = K_{23}A_2$$

The association constants for the two dimerisation reactions, K_{12} and K_{13} are constrained such that the total dimerisation dissociation K_d constant is the experimentally observed 18 μM, and for simplicity is assumed that $K_{12} = K_{13} = 36$ μM. Since discriminating between dimer conformations will be difficult given the generally low signal intensities, it is further assumed that K_{23}=1, meaning that each dimer conformational state has the same population. Since each dimer state is expected to have a similar molecular architecture, a single relaxation rate is used for both dimer states, $R_{2,\mathrm{dimer}}$ and a distinct rate for the monomer $R_{2,monomer}$. The chemical shift and relaxation matrices are then

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_1 & 0 & 0 \\ 0 & \Omega_2 & 0 \\ 0 & 0 & \Omega_3 \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} R_{2,monomer} & 0 & 0 \\ 0 & R_{2,dimer} & 0 \\ 0 & 0 & R_{2,dimer} \end{pmatrix}$$

As before, the frequency-domain signal was calculated numerically.

Chapter 5 The role of nucleic acid in HIV-1 assembly

Interactions between the NC domain of Gag and nucleic acids are absolute requirements for immature HIV-1 assembly (and also for other retroviruses). As described in detail in the introduction, immature assembly is promoted by both DNA and RNA of variable length and sequence. In this chapter, the role of nucleic acid interactions is studied using fluorescence quenching experiments and quantitative gel densitometry assays, making use of two model DNA oligonucleotides. The first, dACGCC, has only 5 bases and has been previously studied by NMR with a NC peptide fragment spanning residues 257-298 (107), and in chapter 4.8 data were presented characterising its interaction with C-CA_{AA}-NC using NMR. The second model DNA oligonucleotide is termed dSL3, which is the 14-base DNA counterpart of the HIV-1 SL3 packaging element. Various studies have used similar molecules for studies with the NC domain and were detailed in the introduction, but no study has yet sought to determine the effects on assembly. Studies on dSL3 are described in this chapter in which UV melting analysis and 1D NMR are used to establish that it can indeed form a stem-loop. Fluorescence quenching experiments are used to determine the affinity and stoichiometry of interactions between dACGCC and dSL3 with both C-CA-NC and C-CA_{AA}-NC. The effects of the oligonucleotides upon immature assembly are assayed using ΔMACANCSP2 and C-CA-NC. The first of these proteins contains the full CA, SP1, NC and SP2 domains, and a mutant of the MA domain lacking residues 15-99. It has been shown to assemble into reasonably homogeneous spheres of 100 nm radius at pH 8 and tubes at pH 6, with the spheres having been the subject of a cET investigation which revealed them to be faithfully representative of immature virions. The effects of oligonucleotides upon C-CA-NC are also examined in this chapter, and similarities and differences between the effects upon ΔMACANCSP2 are found. It has been previously reported that C-CA-NC can assemble into small virus-like particles, but in the assays reported here only aggregates are found. Based upon the data presented in this chapter, a simple model for the role of nucleic acids in immature assembly is presented.

5.1. Fluorescence studies of NC-nucleic acid interactions

5.1.1. The fluorescence phenomenon

Fluorescence is the phenomenon observed when an electron is promoted to an excited state by absorption of radiation, then returns to its ground state *via* radiative transitions, thus

emitting radiation of longer wavelengths (lower energy) than the excitation radiation as it does so. The emitted radiation will be of interest, since it is sensitive to the molecular environment. In particular, the side-chain of tryptophan residues can absorb strongly in the range 280 nm to 300 nm, and emit at around 350 nm. However, radiative transitions are not the only means by which excited electrons can relax, and fluorescence is said to be quenched if some other non-radiative process dominates the return to the ground state. Quenching of tryptophan fluorescence occurs when the tryptophan side-chain is in close proximity with a nucleic acid base, and for this reason protein-nucleic acid interactions can be characterised by fluorescence quenching (84).

5.1.2. Interpreting the fluorescence quenching signal

The NC domain of C-CA-NC and C-CA_{AA}-NC contains a single tryptophan which, as described in chapter 1.7, stacks parallel to a guanine base of either RNA or DNA in all NC-nucleic acid complexes whose structures have been solved (10, 39, 107). A number of studies have exploited this fact in order to monitor NC-nucleic acid interactions using tryptophan fluorescence (97, 119, 140, 150, 153).

The fluorescence quenching signal observed is a population-weighted sum of time-independent coefficients for the signal amplitudes associated with each chemical state of the system. By 'chemical states' is meant, for example, ligand-bound and ligand-free protein in the case of a protein titrated with a ligand. For the experiments to be described, in which either C-CA-NC or C-CA_{AA}-NC is titrated with a nucleic acid, it is generally found that the oligonucleotides contribute a negligible fluorescence signal. We will also make the assumption that the fluorescence signal contributed by a protein bound to an oligonucleotide is unaffected by the binding of a second protein molecule to the same oligonucleotide. Thus the fluorescence signal observed in titrations of oligonucleotides is

$$I = I_f P + I_b (P_T - P) \tag{5.1}$$

Here, I is the observed intensity, I_f is the signal intensity of unliganded protein, I_b the signal intensity of bound protein, P the unliganded protein concentration and P_T the (experimentally controlled) total protein concentration. Note that P is to be determined by assuming a particular binding model, from which it is determinable given some set of model parameters (often a single association constant) and the total protein and ligand concentrations. When fitting this equation to data, I_f and I_b are treated as free parameters, along the parameter(s) upon which P depends.

It was described in chapter 1.7 that the stoichiometry of a binding interaction is poorly determined if a titration at a single concentration is used (unless evidence can be obtained elsewhere of a particular stoichiometry). However, the titration curve is also sensitive to the total protein concentration, such that using more than one protein concentration provides additional constraints for the stoichiometry. Essentially, the fluorescence signal is determined as a function of L_T and P_T, rather than L_T alone. The stoichiometry and association constant are determined once values are found which account for all titrations, and if a particular combination of K and M can account for a titration at a particular P_T but not another, they are incorrect. It is for that reason that multiple protein concentrations were used for dSL3. In the case of dACGCC, NMR showed convincingly that there was only a single site, so only a single concentration was needed for the fluorescence studies.

Binding models

By binding model is meant some set of assumptions as to the nature of interaction between P and L, and the corresponding expressions establishing the link between the free protein concentration P and the model parameters. There are two types of binding models to be considered. In the first, the assumptions are made that a finite number of protein molecules P can be accommodated on the oligonucleotide L at M specific and non-overlapping sites. The convention is adopted that P and L refer to protein and ligand, and in italicised type their respective concentrations. For simplicity, we assume that each of the binding sites on L has the same intrinsic affinity for P. Note that by 'non-overlapping' sites is meant that a protein binding at an arbitrary site does not occlude any part of any other sites, such that binding sites are separated on L. This will be referred to as simply 'specific binding'. The second type of model is of non-specific binding, in which an oligonucleotide of N bases is considered, of which any stretch of n bases constitute a binding site, that is to say, are occluded by the binding of a single P to L, and any binding site has the same intrinsic affinity for P. This will be referred to as 'non-specific binding'.

General binding equations

We derive here an equation to describe general binding events subject to the assumption that all sites have the same intrinsic association constant for the protein. We can write a binding density equation as

$$\nu = \frac{P_B}{L_T} = \frac{P_T - P}{L_T} \qquad 5.1$$

In which the symbol P_B denotes the total concentration of bound protein. The intrinsic association constant is defined as

$$K = \frac{k_f}{k_r} \qquad 5.2$$

In which k_f is the forward rate constant for the reaction (binding) and k_r is the reverse (un-binding) rate constant. Consideration of the reaction at equilibrium gives

$$f_m k_f P . P_{m-1} L = r_m k_r P_m L \qquad 5.3$$

Where f_m is the number of ways in which P_mL (of which there may be several forms, and P_mL refers to the total concentration of ligand with m protein molecules bound) can be formed from $P_{m-1}L$, and r_m the number of ways P_mL can dissociate to $P_{m-1}L$. This expression readily leads to the equation

$$\frac{P_m L}{P^m . L} = \beta_m K^m \qquad 5.4$$

In which β_m is defined as

$$\beta_m = \begin{cases} \prod_{j=1}^{m} \frac{f_j}{r_j}, 1 \le m \le M \\ 1, \qquad m = 0 \end{cases} \qquad 5.5$$

These terms give the number of distinct permutations of m molecules of P on an oligonucleotide L with an arbitrary number of binding sites for P, irrespective of whether such sites are overlapping. The binding density equation is then

$$\frac{P_T - P}{L_T} = \frac{\sum_{m=1}^{M} m \beta_m (KP)^m}{\sum_{m=0}^{M} \beta_m (KP)^m} \qquad 5.6$$

The symbol M has been used to denote the maximum number of P that can be accommodated on a single L. The β_m coefficients are determined by the particular model invoked for binding and, once found, the free protein concentration P is determinable from this equation, thus relating the observed fluorescence I to the association constant K. This is accomplished by re-writing the binding density as a polynomial of the form

$$\sum_{j=0}^{M+1} a_j P^j = 0 \qquad 5.7$$

in which

$$a_j = \beta_j K^j (P_T - jL_T) - \beta_{j-1} K^{j-1} \quad 5.8$$

Sometimes, this can be solved analytically, in other cases it must be solved numerically, for example by means of a Newton-Raphson procedure . Note that in the results, the dissociation constant K_d is given, defined as the reciprocal of the association constant K. The dissociation constant has the advantage of being more readily interpretable, whereas the association constant makes the binding equations easier to handle.
With K known, we can extract other species concentrations according to the formulae

$$L = \frac{L_T}{\sum_{m=0}^{M} \beta_m (KP)^m} \quad 5.9$$

$$P_m L = L.\beta_m (KP)^m \quad 5.10$$

Note that P_mL refers to the total concentration of complexes in which m P are bound to a single L, and does not refer to the occupancy of a particular site. This has implications for the non-specific binding model.

Specific binding

Following the assumptions of the specific binding model, the β_m coefficients are the number of distinct ways of distributing m molecules P amongst the M separate sites on L. Hence they are the binomial coefficients, given by

$$\beta_m = \frac{M!}{m!\,(M-m)!} \quad 5.11$$

The binding density equation becomes

$$\frac{(P_T - P)}{L_T} = \frac{MKP}{1 + KP} \quad 5.12$$

Such that P is given by

$$P = \frac{1}{2K}\left\{-(MKL_T - KP_T + 1) + \sqrt{((MKL_T - KP_T + 1)^2 + 4KP_T)}\right\} \quad 5.13$$

Non-specific binding

With the definitions given for non-specific binding the β_m coefficients are the number of distinct ways of distributing m sets of n contiguous unoccupied bases amongst (N-mn) unoccupied bases. Hence they are written for non-specific binding as

$$S_m^{Nn} = \begin{cases} \dfrac{(N - mn + m)!}{m!\,(N - mn)!}, 0 \le m \le M \\ 0, \qquad\qquad otherwise \end{cases} \qquad 5.14$$

In this expression, the more informative symbol S_m^{Nn} is substituted for β_m to denote the number of distinct permutations of m molecules P on an oligonucleotide L with N bases and a site size of n contiguous bases. M, as before, is the maximum number of P accommodated by a single L. It is the highest integer satisfying $nM \leq N$ with nM the product of the binding site size n and the total number of bases M. The polynomial for P must be solved numerically for non-specific binding, since the simplifications possible when β_m are binomial coefficients as occurs in specific binding are no longer possible. The non-specific binding model is quite general and the relevant formulae can also be found in the literature (88).

5.1.3. Fluorescence titration procedure

Fluorescence titrations were performed by titrating dSL3 into C-CA_{AA}-NC or C-CA-NC at fixed concentrations of 0.3 μM, 1 μM and 2 μM, or dACGCC into C-CA_{AA}-NC or C-CA-NC at a fixed concentration of 1 μM, in a 3.5 ml reaction volume using a Hitachi F2500 fluorescence spectrometer. Fluorescence was excited at 290 nm and monitored at 350 nm. Protein solutions were buffered with 10 mM Hepes pH 7.0, 200 mM NaCl, 0.05% NaN_3, and TCEP in 10-fold molar excess over protein. Error was accounted for by repeated independent measurements of a limited number of points, from which a well-defined standard error of 5% was found and applied to all titration points.

When fitting models to the data, a global analysis was conducted using Matlab, leaving I_f, I_b and K as continuous free parameters. For the specific binding model, M was treated as an integer-valued free parameter (the binding equations do not hold for non-integer M). For the non-specific binding model, N was fixed (the number of bases were known) and n (the site size) was assumed to be 6, in accord with previous fluorescence and NMR studies which concluded that this is the most likely number of bases occluded by NC binding (107, 119, 140).

5.2. UV melting and 1D NMR analysis of dSL3

Ultraviolet (UV) light is absorbed by DNA on account of the conjugated π-electron systems of the bases. The amount of light absorbed at the absorbance maximum of 260 nm

is dependent upon whether bases are base paired, and base pairing can be broken at high temperatures, referred to as melting. The temperature at which half the DNA molecules present in a sample are in base-paired conformation and half are un-paired is the melting temperature, T_m, of that particular DNA.

It was reasoned that dSL3 could form a stem-loop by intra-molecular base pairing, or a duplex by inter-molecular base pairing. Stem-loop formation is a first order reaction, and at equilibrium the ratio of stem-loop and unfolded DNA is not dependent upon the total concentration of DNA. Hence T_m is invariant to the total DNA concentration. However, formation of a duplex is a second order reaction whilst its dissociation is first order. For this reason, the equilibrium concentrations of monomeric and duplex DNA, if duplex formation is possible, will depend upon the total DNA concentration. Therefore the T_m will be dependent upon the total DNA concentration. To discriminate between duplex formation and stem-loop formation, the T_m can be determined for dSL3 as a function of dSL3 concentration. If the melting temperature is unchanged by dSL3 concentration, it is likely that stem-loop formation is the favoured base-pairing interaction. If there is a dependence upon dSL3 concentration, duplex formation is also possible. If there is no change in absorbance with temperature, then bases were probably not paired in the first place.

To determine melting temperatures, the absorbance at 260 nm, A_{260}, was measured in the range 20 to 80 °C with 138 points per melting curve. The dSL3 concentrations used were 0.5, 1, 2, 5 and 10 μM, and experiments were conducted in 10 mM Hepes pH7 with 200 mM NaCl, which is the same as for fluorescence titrations.

A 1D ^{1}H NMR spectrum was acquired for a 1 mM sample of dSL3 at 12 °C, using 256 scans with an acquisition time (per scan) of 256 ms. The reduced temperature of 12 °C was used (rather than 25 °C as used for proteins) to improved the signal-to-noise ratio. NMR is a useful tool for the detection of DNA base pairing since the imino protons, which have characteristic chemical shifts, are visible only if they are in hydrogen bonds. If they are not in hydrogen bonds, exchange with water protons means that no signals are visible. Hence, if imino proton signals are visible, then there is base pairing, indicative of either duplex or stem-loop formation.

5.3. ΔMACANCSP2 and C-CA-NC assembly assays

ΔMACANCSP2 has been shown previously to assemble into 100 nm spheres which closely resemble immature HIV-1 virions when incubated with a wide range of DNA or RNA molecules in a low-salt buffer at mildly alkaline pH, or as tubes if mildly acidic pH conditions are used (60). For our assays, we exchanged ΔMACANCSP2 or C-CA-NC into an assembly buffer containing 30 mM Tris pH 8.0, 50 mM NaCl, 1 mM TCEP and 0.05% NaN_3 using Vivaspin centrifugal concentrators, and used the same buffer to adjust the protein concentration accordingly. All assembly reactions were carried out in hydrophobic-walled 1.5 ml microfuge tubes (Alpha Laboratories), such that none (or at worst very little) of the reaction volume would adhere to the side of the tube. Appropriate volumes of dSL3 and dACGCC were added from concentrated stock solutions to the ΔMACANCSP2 solutions and reactions incubated for 4 hours at 25°C, after which insoluble material was recovered by centrifugation at 4°C for 45 minutes at 20000 G. The supernatant was carefully removed, and the pellet re-suspended in the same volume of assembly buffer as used in the reaction (typically 12 µl). The re-suspended pellet was either used for transmission electron microscopy, or SDS-PAGE analysis for quantitation of the fraction of pelleted protein using densitometry. Reaction conditions were chosen such that 1-10 µg of protein was loaded in each lane of the SDS-PAGE gel, this being an accurately quantifiable range. For gel densitometry, precast gels were used (invitrogen Novex system), with 14% cross-linking and 1 mm thickness, to ensure that results were comparable. Images of gels were acquired under ultraviolet light with exposure times of 40 ms after staining with coomassie blue and destaining overnight. Densitometry was performed by counting pixel intensities after subtracting the background intensity using the image processing toolbox in Matlab. The fraction of assembled material *f* is given by

$$f = \frac{I_{pellet}}{I_{pellet} + I_{supernatant}} \qquad 5.15$$

In which each *I* is an integrated band intensity.

Transmission electron microscopy was used to examine assembly products. To do so, an assembly assay was carried out as above, and the pelleted material was redissolved in 10 µl assembly buffer. Of this, 5 µl was loaded onto a 3 mm diameter formvar-coated copper grid and left for 5 minutes before blotting with filter paper and left to dry for 2 minutes. For negative staining, 5 µl ammonium molybdate was added to the grid and left for 10 s

before blotting with filter paper, then the sample left for at least 5 minutes to dry fully before examination using a Hitachi H7000 instrument with a 75 KV accelerator voltage (University of Southampton Biomedical imaging unit).

5.4. Results

5.4.1. UV melting and 1D NMR analysis of dSL3

The 1D NMR spectrum for 1 mM dSL3 shows two distinct sets of imino peaks, with one set of signals in the region 12.6-12.9 ppm and the other in the range 13.6-14 ppm. The former is assigned to the guanine imino protons and the latter to the thymine imino protons. Unfortunately, the signals are overlapping, which is prohibitive to counting exactly how many base pairing signals are present. However, their presence alone is indicative of base pairing, such that either an intra-molecular stem-loop or inter-molecular duplex exist (or both). The full 1D spectrum, and a close-up around the imino region, are shown in Figure 5.1.

The UV melt results are plotted in Figure 5.2. The raw data have been normalised to one another. To find melting temperatures, the gradients of the melt curves were computed numerically after smoothing the raw data with an 8-point moving average. The observation of changes in A_{260} with temperature is indicative of an unfolding event, such that there is at least some degree of structure in dSL3, consistent with the NMR data. No change in T_m is detectable in the concentration range used, which spans an order of magnitude, and therefore the folded state is either an intra-molecular stem-loop, or is an inter-molecular duplex. If the latter is the case, the dimerisation dissociation constant must be significantly less than 0.5 μM in order that the duplex is the dominant species at all concentrations used (0.5 – 10 μM). Insomuch as stem-loop formation being an intra-molecular process, this is considered the more likely structure for dSL3.

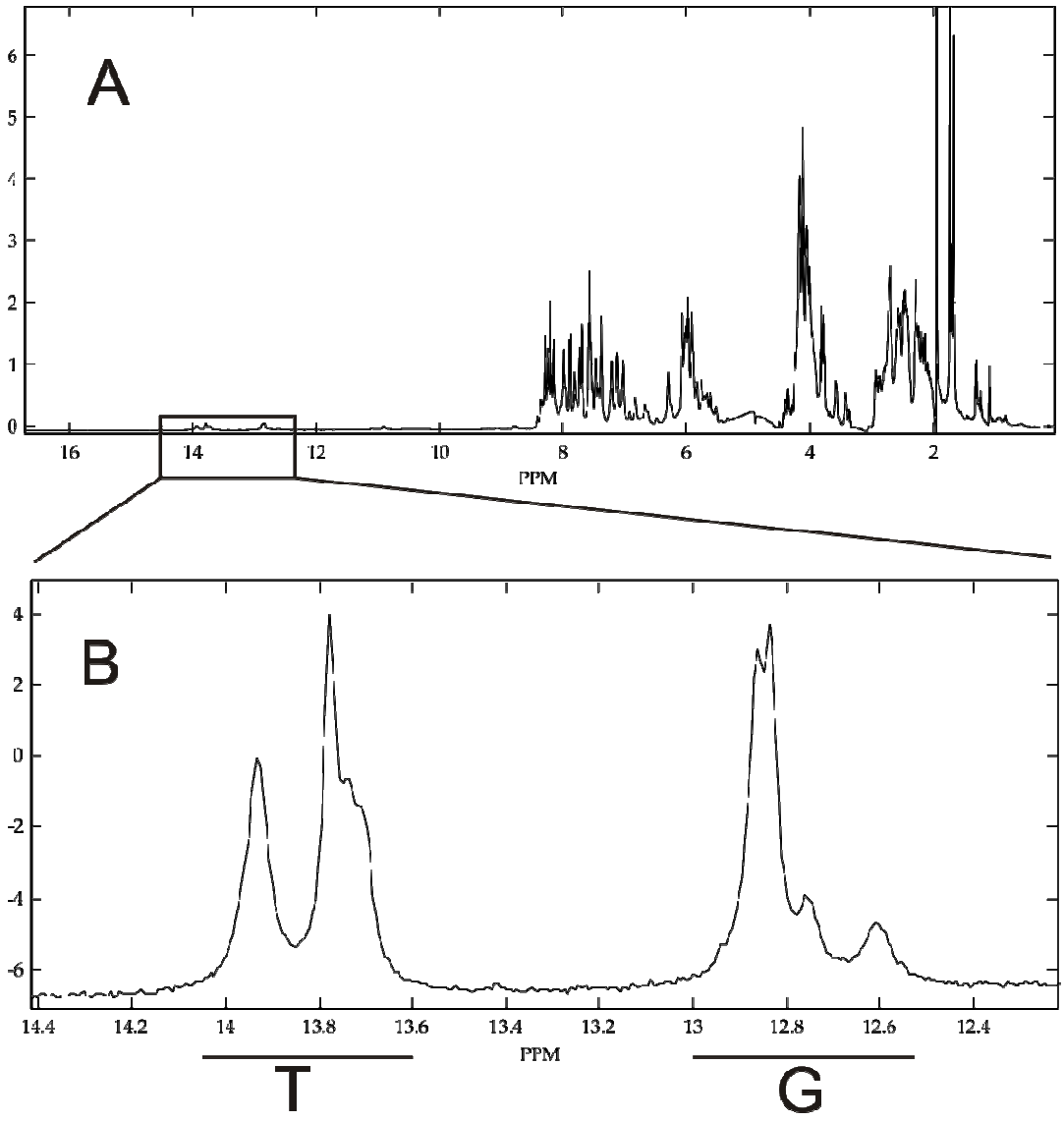


Figure 5.1: ^{1}H NMR spectrum of dSL3.

A) full proton spectrum. B) Close-up of the weak but detectable imino proton region confirms base pairing. The thymine and guanine regions are shown, but specific peak assignments have not been made.

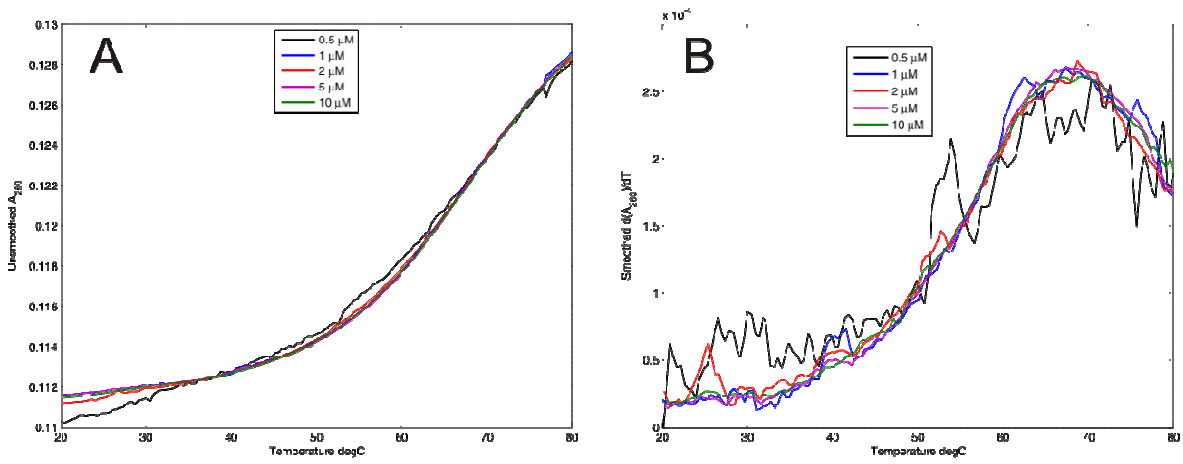


Figure 5.2: A) Melting curves at different dSL3 concentrations after normalising the datasets for comparison.

At all concentration there is a clear shift in absorbance as the temperature is increased. B) Gradients of the normalised melting curves after smoothing the data with an 8-point moving average. The crest of the peak in each case occurs at the melting temperature, which are very similar in each case.

5.4.2. Fluorescence studies

It was thought possible that the emission wavelength of W282, the tryptophan whose fluorescence was monitored, could change upon formation of the NC-DNA complex. This was found not to occur, the signal change upon DNA binding being a quenching effect only (within the limits of experimental accuracy). The background fluorescence signal was negligible, as was the contribution from the DNA oligonucleotides, as established by measuring spectra of buffer alone and buffer with oligonucleotide. The fluorescence intensity at saturation, I_b, was considerably higher than reported for NC alone (119, 140), which is probably attributable to the tyrosine residues in the C-CA domain which do not interact with either oligonucleotide (as shown by NMR in chapters 4.8-4.10). In the case of C-CA-NC, I_b was generally higher again, probably due to W184 at the C-CA dimer interface. A series of emission spectra taken for different concentrations of C-CA in the range 0.3 μM to 5 μM showed very little change in the fluorescence spectrum, meaning that either dimerisation causes negligible change in the W184 signal, or that it is buried beneath the stronger W282 signal. Either way, it was not necessary to account for dimerisation in the range of protein concentrations used, which would require a more complicated expression than that given for quenching by complex formation.

Interaction of the NC domain with dACGCC

The titrations of dACGCC into C-CA-NC and C-CA_{AA}-NC are given in Figure 5.3. The quality of fit is generally very good, and the binding curves sufficiently well sampled as to give well-defined dissociation constants. For C-CA-NC, a K_d of 9.5 ± 2.8 μM was obtained, and for C-CA_{AA}-NC the best K_d was 5.7 ± 0.7 μM, each under the assumption of a single binding site, consistent with the NMR data presented in chapter 4.8.

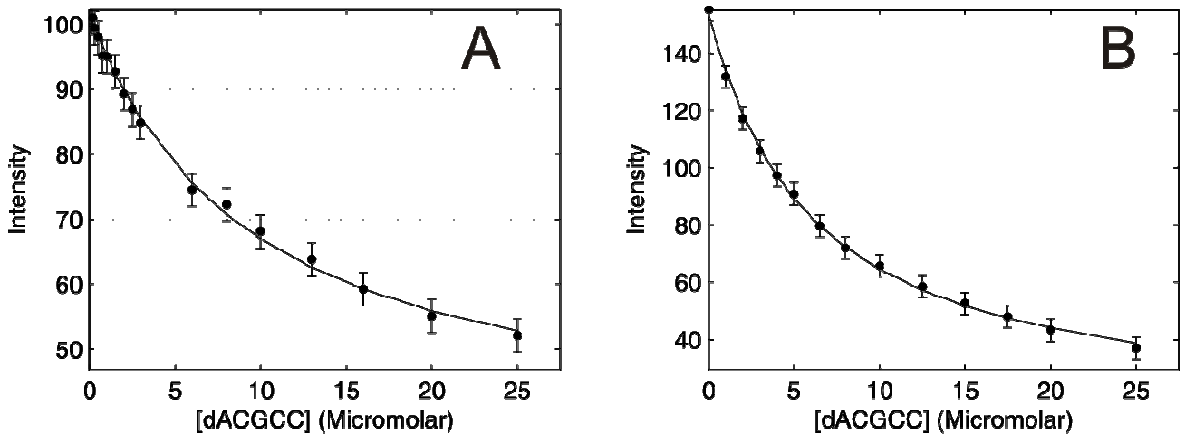


Figure 5.3: Fluorescence data for dACGCC titrated into A) C-CA-NC and B) C-CA$_{AA}$-NC.

The curves are the best fit for a model in which dACGCC binds the NC domain in 1:1 stoichiometry.

Interaction of the NC domain with dSL3

The titration curves for dSL3 are shown for C-CA-NC and C-CA$_{AA}$-NC in Figures 5.4 and 5.5 respectively. We determined the affinity and stoichiometry of the NC-dSL3 interaction for both C-CA-NC and C-CA$_{AA}$-NC, with W282 acting as a reporter in fluorescence titrations. No shift in emission maximum was detected for either protein, but a substantial quenching of fluorescence at the observed emission maximum of 350 nm was seen. It was found by fitting equation 5.13 to the data that a model in which dSL3 is assumed to have a single site for NC domain could not describe the data, since a globally optimised K_d could not be found, indicating that there must be more than one binding site.Thus we fitted a model in which dSL3 is assumed to have two sites of identical affinity for the NC domain, and this adequately described all titrations for either protein, with a K_d of 0.80 ± 0.12 µM for C-CA-NC and a K_d of 0.98 ± 0.17 µM for C-CA$_{AA}$-NC, which are consistent within error. Such datasets, and the fitted curves, are shown in Figures 5.4 and 5.5. Models in which the NC domain has multiple sites for dSL3 were found to give poor fits to the data, and a model in which the NC is assumed to have two sites of different affinity for dSL3 resulted in very similar K_d values for each site. We hence conclude that dSL3 can bind at least two NC domains at once, which has been recently observed for NC in isolation and 12-mer DNA oligonucleotides, which are of comparable length to dSL3 at 14 bases. Our reported K_d values are in good agreement for studies of the NC domain with DNA oligonucleotides conducted under similar conditions (12, 107, 153).

It is possible that the NC domain binds non-specifically at multiple sites along dSL3. We found this to be consistent with fluorescence data by applying a non-specific binding model, in which dSL3 is treated as a lattice of overlapping sites of identical affinity, using equation 5.14 substituted into equation 5.7 and 5.8. We fitted this model to our fluorescence data for C-CA-NC (K_d = 2.7 ± 0.48 μM) and C-CA_{AA}-NC (K_d = 3.1 ± 0.61 μM). However, we were unable to discriminate between this model and the two-site specific model. The two-site specific model gave slightly better χ^2 values for global optimisation, but the improvements are small.

The results of fitting the various models are listed in Tables 5.1 and 5.2.

Table 5.1 Fitted fluorescence parameters for C-CA-NC

Model	Conc (μM)	K_d (μM)	χ^2	Gloabl K_d (μM)	Global χ^2
1:1	0.3	0.36	6.25	0.19	41.6
	1	0.13	13.8		
	2	0.20	12.3		
2:1	0.3	0.89	4.50	0.80	10.6
	1	0.61	3.76		
	2	0.89	0.975		
Non-specific	0.3	3.48	5.49	2.7	18.8
	1	1.48	6.22		
	2	1.52	2.15		

Table 5.2 Fitted fluorescence parameters for C-CA_{AA}-NC

Model	Conc (μM)	K_d (μM)	χ^2	Gloabl K_d (μM)	Global χ^2
1:1	0.3	0.36	0.910	0.30	22.6
	1	0.27	5.00		
	2	0.23	16.1		
2:1	0.3	0.97	0.725	0.97	5.99
	1	1.20	1.50		
	2	0.52	0.646		
Non-specific	0.3	3.6	0.811	3.1	10.2
	1	3.5	2.50		
	2	0.79	1.18		

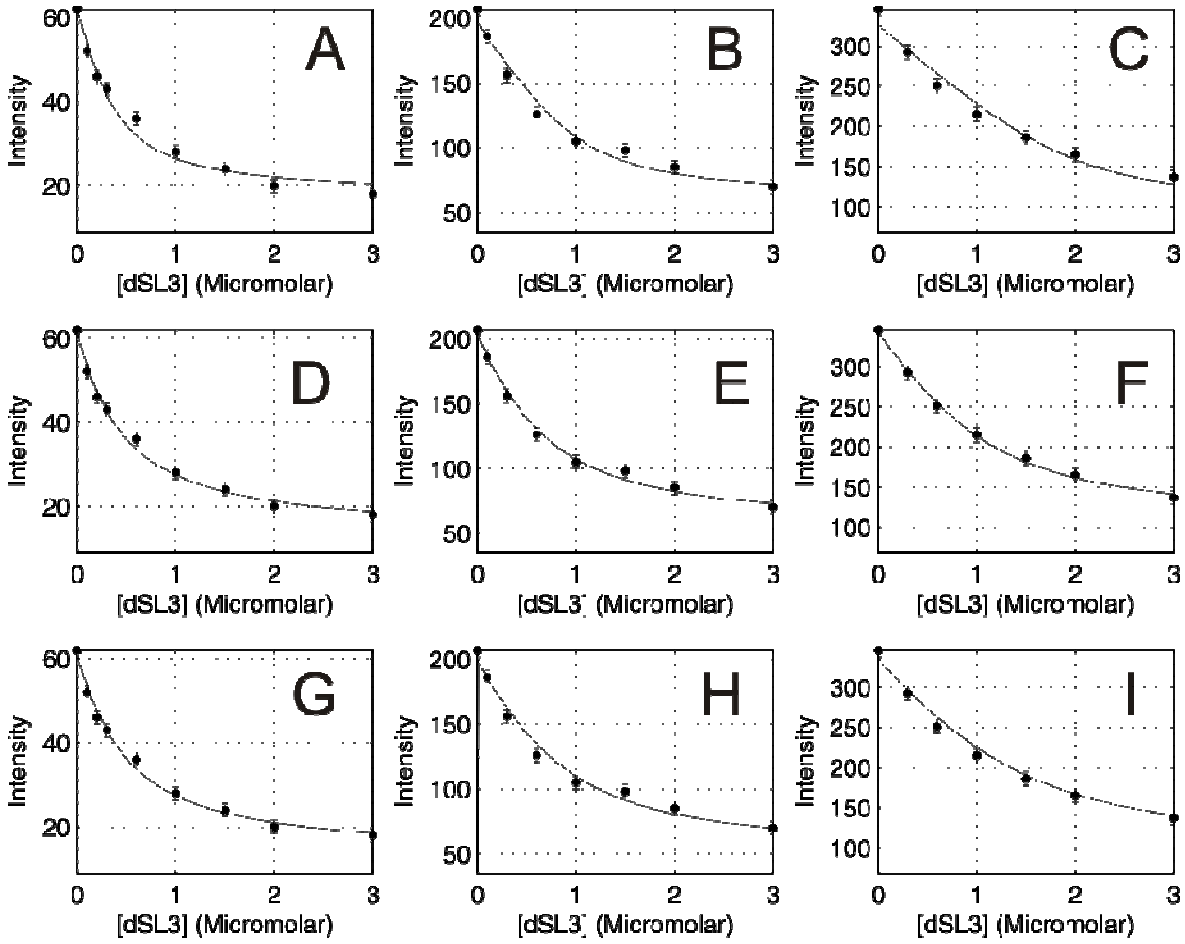


Figure 5.4: Fluorescence quenching titrations for C-CA-NC and dSL3 and, in smooth curves, fitted models.

A-C) Single-site specific binding model. D-F) Two-site specific binding model. G-I) Non-specific binding model with a 6-base binding site. A,D,G) [C-CA-NC] = 0.3 μM, B,E,H) [C-CA-NC] = 1 μM, C,F,I) [C-CA-NC] = 2 μM. For each concentration, the globally optimised affinity is used for displaying the fitted model. Data have not been normalised, but for each concentration the axes are standardised.

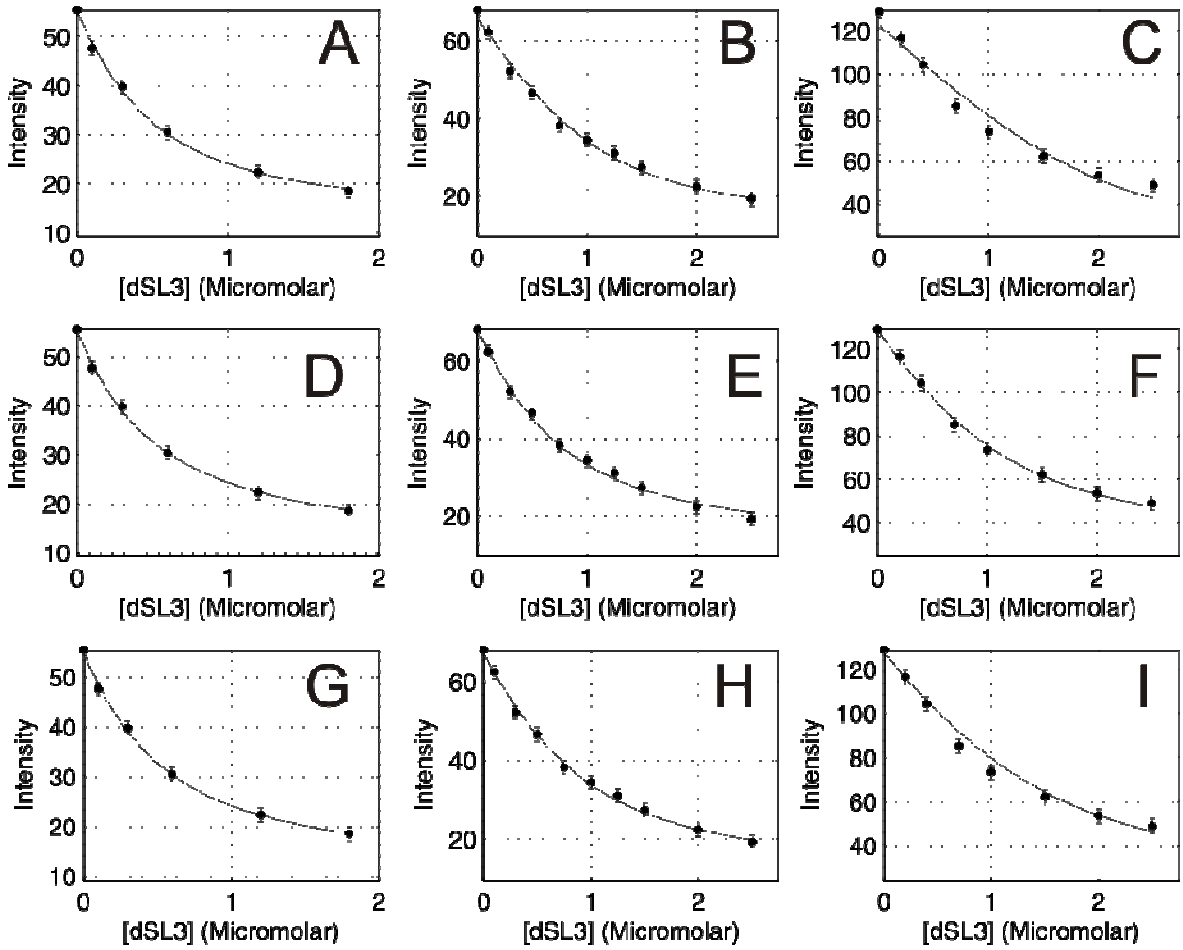


Figure 5.5: Fluorescence quenching titrations for C-CA$_{AA}$-NC and dSL3 and, in smooth curves, fitted models.

A-C) Single-site specific binding model. D-F) Two-site specific binding model. G-I) Non-specific binding model with a 6-base binding site. A,D,G) [C-CA$_{AA}$-NC] = 0.3 μM, B,E,H) [C-CA$_{AA}$-NC] = 1 μM, C,F,I) [C-CA$_{AA}$-NC] = 2 μM. For each concentration, the globally optimised affinity is used for displaying the fitted model. Data have not been normalised, but for each concentration the axes are standardised.

5.4.3. ΔMACANCSP2 assembly

We used rapid dilution assembly assays to determine whether the 14-mer dSL3 and 5-mer dACGCC oligonucleotides could promote assembly of ΔMACANCSP2 into VLPs. To determine whether the rapid-dilution assays used for assembly resulted in spherical particles, negative-stain transmission electron microscopy was used. A representative micrograph is given in figure 5.6, demonstrating that indeed spherical particles are formed. No tubes were observed, neither were obviously incomplete particles or aggregates. In reactions using dSL3 a large number of spherical particles of diameter ~100nm were

observed, but very little disordered aggregates. This is consistent with previous data in which dialysis, rather than rapid dilution, assays were used (60), and we therefore assume that the pellet fraction after rapid dilution assays represents the fraction of correctly assembled virus-like particles.

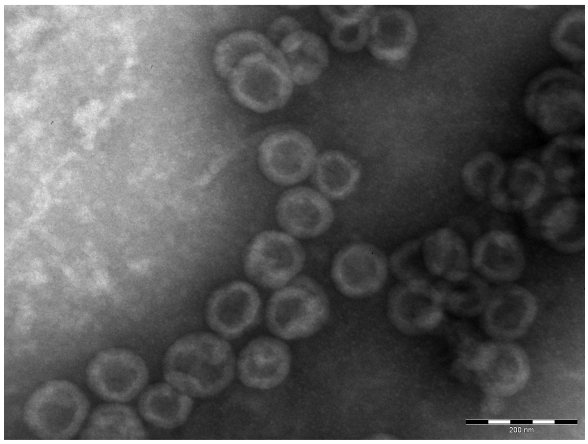


Figure 5.6: Negative-stain transmission electron micrograph of ΔMACANCSP2 spheres assembled by a rapid dilution assay with dSL3.

This image is typical of assay results viewed under the electron microscope. Scale bar: 200 nm.

By varying the ΔMACANCSP2 concentration used in the assays we found that there is an upper limit to the fraction of material that forms VLPs, which was found to be 70 ± 5%, and that this upper limit is independent of the ΔMACANCSP2 concentration. These results are shown in Figure 5.9. By varying the dSL3 concentration we also found that the upper limit of the fraction of assembled material is reached only when the dSL3 concentration is at least half the ΔMACANCSP2 concentration, and that further increasing the dSL3 concentration has no observable effect, as shown in Figure 5.7 and 5.8.

We then tested whether dACGCC, by virtue of its inability to promote assembly, can out-compete the interactions with dSL3, and hence act as an inhibitor of immature assembly. We found that addition of dACGCC at concentrations up to 1 mM to 50 μM ΔMACANCSP2, prior to addition of 25 μM dSL3, did not significantly reduce the VLP fraction. Hence it does not function either as a promoter or an inhibitor of VLP formation. This is also shown in Figure 5.8.

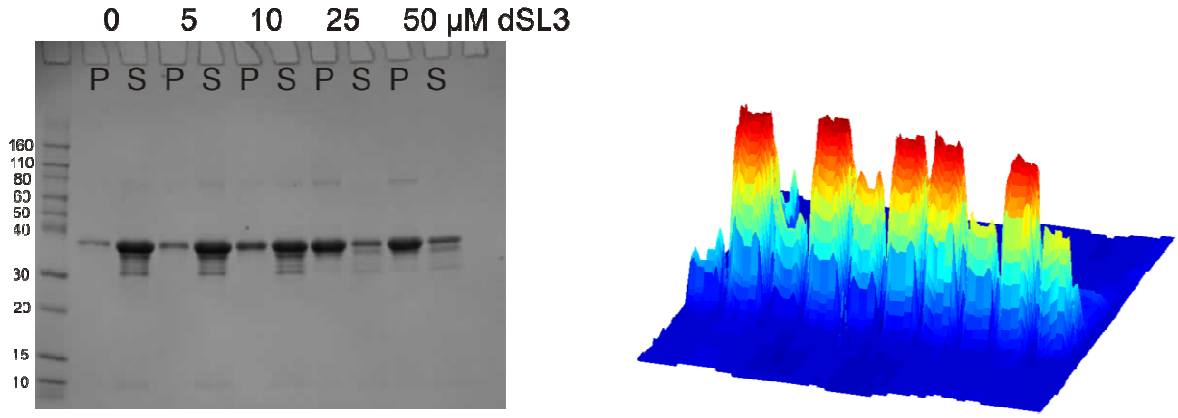


Figure 5.7: Gel densitometry analysis of a rapid-dilution assembly assay to determine the effects of dSL3 concentration upon ΔMACANCSP2 assembly.

The ΔMACANCSP2 was 50 µM. Above the gel, 'S' and 'P' denote supernatant and pellet respectively. The bands have been excised and shown in 3D, displayed in the same order as on the gel. The band integrals were used to determine the fraction of material in the pellet. The gel also demonstrates the quality of the ΔMACANCSP2 preparation.

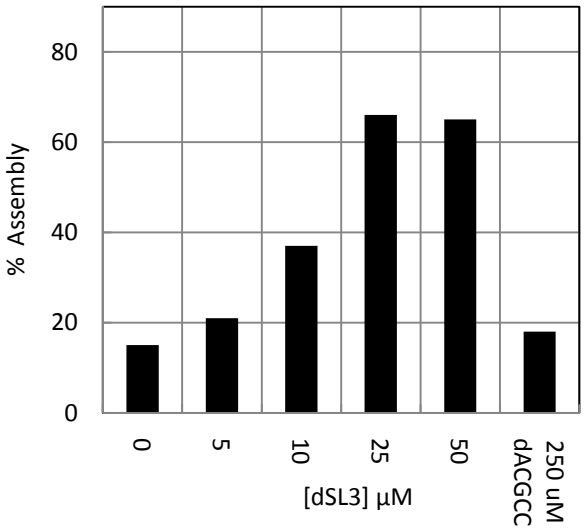


Figure 5.8: ΔMACANCSP2 assembly yields as a function of dSL3 and dACGCC concentration.

The protein concentration was fixed at 50 µM, and the concentrations of nucleic acid as indicated on the x-axis was added.

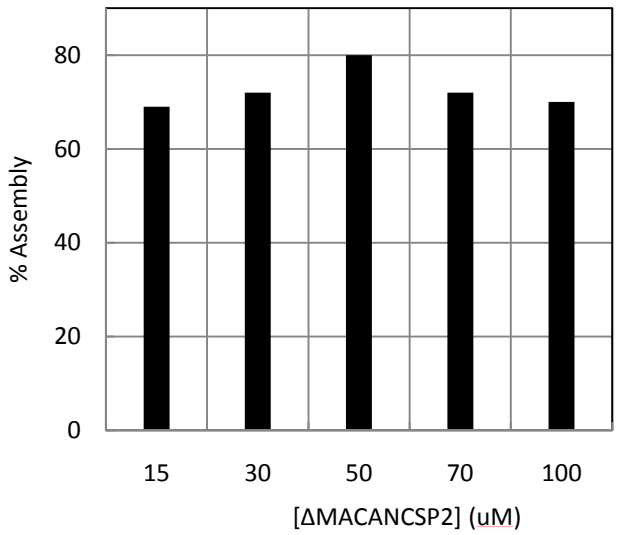


Figure 5.9: ΔMACANCSP2 assembly yields as a function of its own concentration.

In each case dSL3 was present at half the indicated protein concentration. The yield of assembled material is almost constant, regardless of the protein concentration.

5.4.4. C-CA-NC assembly

We used the same rapid-dilution protocol to assess the effects of dSL3 and dACGCC upon C-CA_{AA}-NC and C-CA-NC. Neither oligonucleotide promoted any aggregation or assembly using the non-dimerising mutant C-CA_{AA}-NC, which was consistent with NMR studies which showed that peak intensities in the C-CA_{AA} domain were unperturbed. C-CA-NC, however, showed very similar behaviour to ΔMACANCSP2, that is to say, was not promoted to assemble or aggregate by dACGCC but a significant fraction was found to be in the pellet after dilution assays in which dSL3 was present. Such results are shown in Figures 5.10 and 5.11. Although the pellet fraction was smaller than the corresponding ΔMACANCSP2 assays by a factor of around two, the results under all conditions were proportionately the same; the assembled yield was highest when the dSL3 concentration was half that of C-CA-NC, dACGCC could not prevent association in the presence of dSL3 nor dissociate assembled material, and the amount of pelleted material was invariant to the C-CA-NC concentration. However, in the transmission electron microscope, no ordered particles were visible and the pellet fraction was instead found to comprise sheet-like aggregates. Hence our results indicate that the N-terminal CA domain, N-CA, although not absolutely required for aggregation, is required to coordinate ordered assembly. It is noteworthy that, although it has been suggested that the N-CA domain is

not an absolute requirement of immature assembly, and correct capsid assembly can proceed in its absence (3, 22), our data are contrary to this in the sense that C-CA-NC forms only aberrant particles which bear little resemblance to immature capsids.

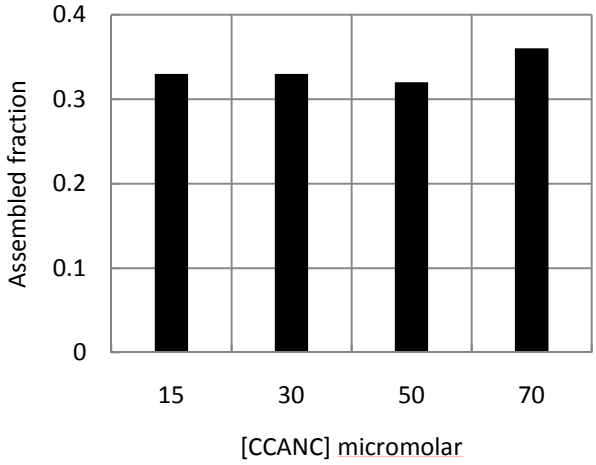


Figure 5.10: C-CA-NC assembly yield as a function of its own concentration.

As in the case of ΔMACANCSP2, the yield is almost constant, and dSL3 was always present at half the indicated protein concentration.

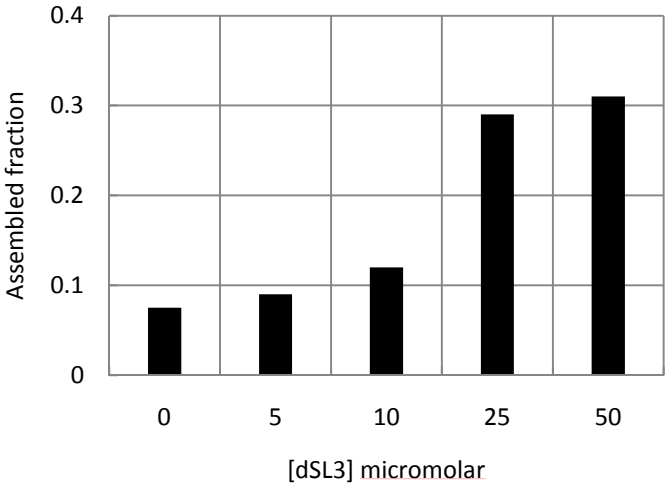


Figure 5.11: C-CA-NC assembly yield as a function of dSL3 concentration.

The C-CA-NC concentration was fixed at 50 μM. The results are consistent with those obtained for ΔMACANCSP2.

5.5. Discussion

5.5.1. Binding mode and stoichiometries

This is the first report demonstrating that, in order for an oligonucleotide to promote immature HIV-1 assembly *in vitro*, such a molecule must be able to accommodate at least two NC domains, that is, facilitate the linking of two molecules to one another. This requirement is perhaps unsurprising, since it has been previously observed that assembly can occur at the plasma membrane of baculovirus-infected insect cells expressing a Rous sarcoma virus (RSV) Gag variant in which the NC domain is replaced with a dimerising leucine zipper domain (73). It has also been observed that chemical cross-linking of cysteine residues at the C-terminal of purified HIV-1 his-tagged CA-SP1 protein can promote assembly (7). The proposed stoichiometry of two NC per dSL3 is consistent with a previous study in which both the RNA and DNA forms of SL3 (with three additional base-pairs to stabilise the stem) were used with the NC domain alone (97), and with the observation that a 12-mer DNA oligonucleotide can accommodate two NC domains (12). The fact that there are NMR structures available for NC bound to SL2 and SL3 RNA in 1:1 complexes (10, 39) was discussed in chapter 1.7.6, the basic argument being that the concentrations of NC and nucleic acid used in such studies would favour 1:1 complexes.

The non-specific dSL3 binding model was tested on account of the range of nucleic acid sequences to which NC can bind, which collectively give the impression that binding is not critically dependent upon sequence. The model is of course an oversimplification inasmuch as assuming that any sequence can be bound with identical affinity, but ascribing all sequences a separate affinity is beyond experimental resolution. The reader will recall from chapter 1.7.6 that NC(1-72) interactions with a DNA oligonucleotide comprising dSL3 with 3 additional base pairs to stabilise the stem have been previously studied using fluorescence quenching, under similar conditions to the experiments performed here (but with an NaCl concentration of 100 mM rather than 200 mM) (153). Their findings were reviewed in the introduction, including a re-analysis of their titration data using the non-specific binding model. The resulting K_d was ~3.1 μM, which is in very good agreement with the values found here of 2.7 ± 0.48 μM for C-CA-NC and 3.1 ± 0.61 μM for C-CA_{AA}-NC.

Unfortunately it is not possible to distinguish between specific and non-specific dSL3 recognition using the data presented here, since the two models predict very similar binding isotherms, as demonstrated in chapter 1.7.6. It is most likely that NC indeed recognises many sequences, but does so with different affinities, such that neither the specific nor non-specific model is a complete picture of NC-nucleic acid interactions. In this case the non-specific model is a closer approximation to reality, since it does not require the assumption that the association constant is zero for most sequences as does the specific model, and is more generally applicable to oligonucleotides of different length than the specific model for which the assumptions of the model are more restrictive in this sense. That is, the specific model requires that only certain sequences are targets for NC recognition, and in the determination of binding stoichiometry for an oligonucleotide of arbitrary length, the number of binding sites is determined without knowledge of the recognised sequences, but under the assumption that they do not overlap with one another. Therefore, titrating oligonucleotides of different lengths and sequences and fitting the non-specific model should aid in demonstrating it to be the more accurate picture of NC-nucleic acid interactions.

5.5.2. Cross-linked dimer model for assembly

It has been proposed that immature assembly proceeds from polymerisation of Gag dimers in which both NC-nucleic acid interactions and C-CA dimerisation contribute to the lowest-order assembling species (7) although precise details of how this could be were not given. We propose a different scheme in which the assembling species (ΔMACANCSP2 or C-CA-NC) can dimerise *via* the C-CA domain, and such dimers are linked to other such complexes by the NC-dSL3 interaction, to which we refer as the cross-linked dimer model. Such a model explains why the dSL3 concentration must be half the protein concentration for optimal assembly, for it is at such a ratio that the fraction of dSL3 associated with two NC domains is highest, provided the protein concentration is considerably in excess of the K_d for the NC-dSL3 interaction (the approximation holds well for [NC]>5K_d). This condition is met in all our experiments, where the typical protein concentration for assembly assays was 50μM and the K_d close to 1μM.

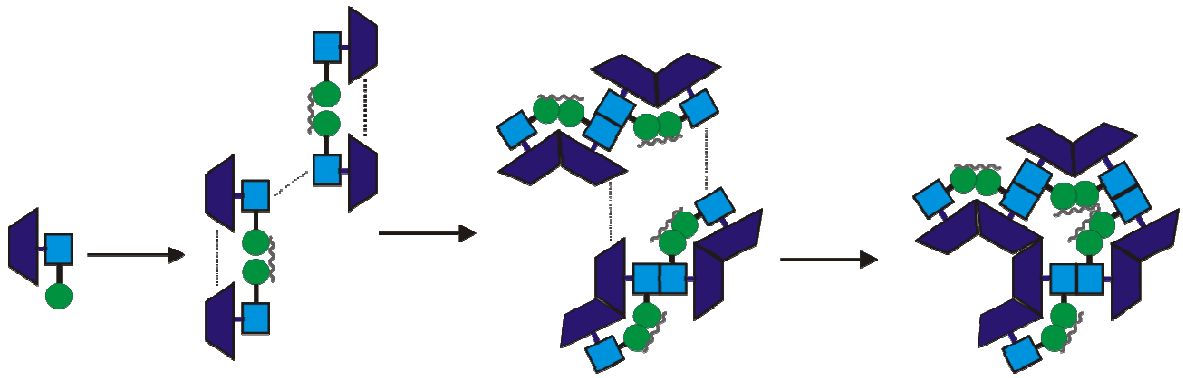


Figure 5.12: The first few steps on the immature assembly pathway according to the cross-linked dimer model.

Only N-CA, C-CA, NC and dSL3 are shown. N-CA is denoted by a dark blue rhomboid, C-CA by a pale blue square, NC by a green circle and dSL3 by a grey waved line.

This cross-linked dimer model may be sufficient to explain C-CA-NC aggregation in the presence of dSL3, but alone is insufficient for explaining ΔMACANCSP2 assembly inasmuch as the consistently observed hexameric arrangement of CA domains in immature virions remains unaccounted for. This discrepancy is corrected for by reasoning that, if small 'linear' polymers of cross-linked dimers form, then elevation of the local N-CA concentration may facilitate weak hexamer-forming interactions. This, we assume, is sufficient to direct assembly away from the aggregates formed by C-CA-NC and towards the reasonably homogeneous 100 nm spherical capsids formed by ΔMACANCSP2. N-CA is therefore assumed to make only a minor contribution to the free energy change of assembly, but is important in keeping the process on the correct pathway. The model is illustrated in Figure 5.12.

This simple model makes no assumption of cooperativity, that is to say, each interaction is not dependent upon any other interaction. It is of note that if the previously favoured model in which a C-CA dimer is the basic unit of assembly is true, then C-CA-NC aggregation upon nucleic acid addition is difficult to explain since there are no other known interaction interfaces remaining upon C-CA-NC to allow such DNA-bound dimers to aggregate with one another. Hence the assumption of an as yet undiscovered interface is necessary to explain the data, but the assumption of such an interface is then to be lifted again in the explanation of ΔMACANCSP2 assembly. Hence we favour the cross-linked dimer model, since it explains and unifies all our results, and introduces no additional assumptions of

interaction surfaces. The cross-linked dimer explanation of C-CA-NC aggregation is shown in Figure 5.13.

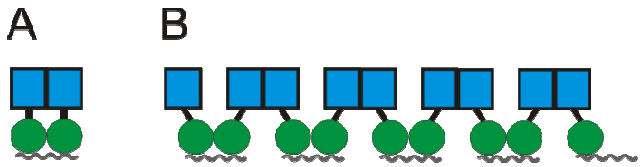


Figure 5.13: Polymeric C-CA-NC.

The C-CA domain is shown as a blue square, NC as a green circle and dSL3 as a grey wiggly line. A) A complex in which a C-CA dimer also involves an NC-nucleic acid dimer. There are no remaining known interfaces to permit elongation. B) A linear polymer according to the cross-linked dimer model, making use of the C-CA dimer interface and NC:dSL3 stoichiometry of 2:1, showing how C-CA-NC may polymerise in the presence of dSL3.

In the cET and cEM studies which have demonstrated CA to exist in the immature state as a hexameric lattice, a 'smeared' constant electron density is observed at the level of NC, that is to say, NC does not arrange with hexameric symmetry in the immature capsid (24, 25, 60, 113, 159). It is noteworthy that, in the cross-linked dimer model, there is no arrangement of cross-linked dimers in which dSL3 can have a 6-fold symmetric arrangement, subject to 2-fold C-CA symmetry across hexamers and 6-fold N-CA symmetry within hexamers. Since 6-fold symmetrisation is used in the generation of reconstructed 3D tomograms, if the dSL3 permutation beneath the CA hexamers lacks 6-fold symmetry, the electron density will be smeared in the final tomogram. This is a testable prediction of the cross-linked dimer model. The observation of 6-fold symmetry at the NC level in a tomogram of ΔMACANCSP2 assembled in the presence of dSL3 would disprove this model, or at least some aspects of it.

5.5.3. Explaining features of assembly using the cross-linked dimer model

The observation that dACGCC cannot out-compete assembly of either ΔMACANCSP2 or C-CA-NC can be explained by consideration of the arrangement of Gag observed in immature virions or VLPs: NC points towards the centre of the particle, and the closed (albeit possibly with gaps) spherical structure of the virion may prevent dSL3 from diffusing out, or dACGCC diffusing in, at least in appreciable amounts.

Similar reasoning may be applied as to why excess concentrations of dSL3 still yield the same amount of assembled material as when it is present at 50% the protein concentration. In the case of C-CA_{AA}-NC, which does not assemble, one expects that an excess of dSL3 would favour the state in which binding occurs with one-to-one stoichiometry between NC domains and dSL3, whilst the two-to-one state would be favoured when the dSL3 concentration is small compared to that of C-CA_{AA}-NC. This is inherent in the binding models used to explain the fluorescence titration data. If it is so that dSL3 is required to cross-link C-CA dimers in the case of C-CA-NC or ΔMACANCSP2, then one would expect NC domains to primarily populate the one-to-one state if they are exposed to bulk solution when the dSL3 concentration is high. Hence, a high dSL3 concentration as compared to protein concentration would be inhibitory to assembly. That this is not the case may be because the NC domains are occluded from bulk solution in the assembled state, making assembly a 'psuedo-irreversible' process inasmuch as the molecules contributing to the assembled state do not exchange with those in soluble oligomeric states. By such means, two-to-one intermediates (two NC domains per dSL3) may form only in low concentrations when the dSL3 concentration is high, but once incorporated into a VLP are unable to revert to one-to-one states.

The inability of dACGCC to out-compete assembly implies that targeting NC with small molecules as antiretroviral agents may not be effective at inhibiting HIV-1 at the immature assembly stage, but may nevertheless be effective at disrupting other stages of the retroviral lifecycle in which NC is involved.

5.5.4. Invariance of VLP yield to protein concentration

The concentrations of ΔMACANCSP2 used were varied from 15 μM to 100 μM, and the dSL3 concentration was always half that of ΔMACANCSP2. Prior to the addition of dSL3, the strongest interaction is expected to be dimerisation of the C-CA domain (K_d = 18 μM for the CA domain alone), in which case the only species present in appreciable concentrations would be monomeric ΔMACANCSP2 and dimeric ΔMACANCSP2, but the ratio of dimer and monomer in this concentration range is a sensitive function of the ΔMACANCSP2 concentration. The invariance of the size of the VLP fraction to the ΔMACANCSP2 concentration therefore implies that the measured C-CA dimerisation strength and NC-dSL3 interaction strength are insufficient to describe immature assembly. Either there must be an additional interaction contributing an appreciable amount of free

energy, or the interaction strengths are altered during the assembly process, being increased as the size of the assembly intermediate increases. The obvious candidate for an additional interaction is N-CA hexamersiation, but this must be extremely weak, for it has not been observed in solution (though crystals of hexameric N-CA can be grown). In explaining the invariance of the yield of assembled protein to the protein concentration, one can construct a similar argument as in the case of the failure of excess nucleic acid concentrations to inhibit assembly. Suppose the protein concentration is low, and for argument's sake dSL3 is present at half the protein concentration or greater. One expects that a complete capsid will form slowly, however once formed will be difficult to disassemble, for liberating a molecule from the capsid will require several interactions to be broken. A complete capsid, as previously argued, is essentially removed from the system, so a low protein concentration should only reduce the rate of capsid formation.

5.5.5. The thermodynamics of immature HIV-1 assembly

The cross-linked dimer model is adequate for qualitatively explaining C-CA-NC assembly in the presence of dSL3. However, the dimerisation K_d and dSL3 affinity are not alone sufficient to describe C-CA-NC assembly quantitatively, such that further parameters must be introduced to fully describe assembly within the cross-linked dimer model. The failure of the known parameters to describe C-CA-NC assembly is explained quantitatively in chapter 5.6. For the present discussion, however, a qualitative consideration is sufficient to see why a two-parameter description of C-CA-NC assembly is inadequate. Let us assume, as is done in the quantitative model in the appendix, that the dimerisation and dSL3 affinities are invariant to the size of a C-CA-NC/dSL3 polymer. As small C-CA-NC/dSL3 complexes form, the number of particles in solution (where particle is intended to mean a C-CA-NC/dSL3 complex of arbitrary size) decreases. Since reaction rates are dependent upon concentrations, the reaction will equilibrate with a distribution of a large number of relatively small oligomers formed. There is no reason, within the assumptions made, for such oligomers to grow any further.

Clearly, however, the distribution of C-CA-NC/dSL3 complexes in reality differs considerably from this prediction. The failure of the two-parameter model motivates the suggestion that the dimerisation and dSL3 affinities, for interactions between two C-CA-NC/dSL3 oligomers of different sizes, depend upon the sizes of the oligomers themselves. In order for large polymers to form, it is to be expected that the affinity of a large oligomer

for another oligomer or monomer is higher than of a small oligomer for another small oligomer or monomer. By this means, the diminution of particle concentration concomitant with formation of a distribution of oligomers is overcome, and the system will be driven towards larger oligomers, as required by observation.

It should be born in mind that the affinity of a particular interaction is related to the free energy change involved in that process. In order that the affinities of dimerisation or dSL3 binding are dependent upon the sizes of complexes formed, the free energy change for adding to a large oligomer must be different from that of the formation of a small oligomer from monomers, and so on. It is reasonable to assume that the enthalpic component of the free energy change for combination of two oligomers of arbitrary size is invariant to the sizes of the oligomers, for the interfaces are not assumed to change in any way. However, the entropic contribution, which is determined by the number of degrees of freedom gained or lost by some process, may be dependent upon the sizes of the combining oligomers. To explain the observations, it is assumed that the entropic loss arising from the combination of two small oligomers is less than the entropic loss arising from combination of two large oligomers. If this is so, then the affinity of two large oligomers for one another will be stronger than the affinity of two small oligomers for one another, as required. In other words, a diminution in entropic loss of each elongation event can drive C-CA-NC/dSL3 assembly and, by the cross-linked dimer model, real immature assembly also, but this is of course unproved at this stage.

5.6. Appendix

5.6.1. The dimerisation and dSL3 affinities are insufficient to fully describe assembly

The case considered here is that of C-CA-NC, in which the C-CA domain can dimerise and the NC domain can bind DNA or RNA. The conditions for application of the model are that the two domains are independent of one another regardless of binding 'state', and that the nucleic acid has two specific and non-overlapping sites for NC.

Let us first define the states which the protein can occupy, following the convention that plain-type refers to a state, and italicised type its concentration. Let the index *j* denote the total number of protein monomers in a particular state and let *k* enumerate the particular permutation of molecules in that *j*-state. Let P denote free protein, L denote free ligand, PL denote a 1:1 protein-ligand complex (with a free site remaining on the ligand), P_2 as

protein dimerised by the C-CA interface, and P_2L as two proteins bound to a single dSL3 molecule without use of the C-CA interface. Such states are given in Figure 5.14. Dimeric protein involving the C-CA interface (P_2) can also be bound to either 1 or 2 dSL3, such states being ascribed the symbols $L(P_2)$ and $L(P_2)L$ respectively. We can then put the j index to use to define higher-order complexes. Those with even and odd values of j are treated slightly differently. For even j the states are $(P_2L)_{j/2}$, $P(P_2L)_{j-2/2}P$, $LP(P_2L)_{j-2/2}P$ and $LP(P_2L)_{j-2/2}PL$. For odd j the states are $P(P_2L)_{j-1/2}$ and $LP(P_2L)_{j-1/2}$. We must of course specify the strengths of the interactions across the C-CA interface, and for ligand binding. Let the intrinsic binding constant for C-CA dimerisation be defined as K_D and for ligand binding K_N. These are invariant to the size of the complex, and each site on the ligand is treated as having the same K_N. These will be put to use presently.

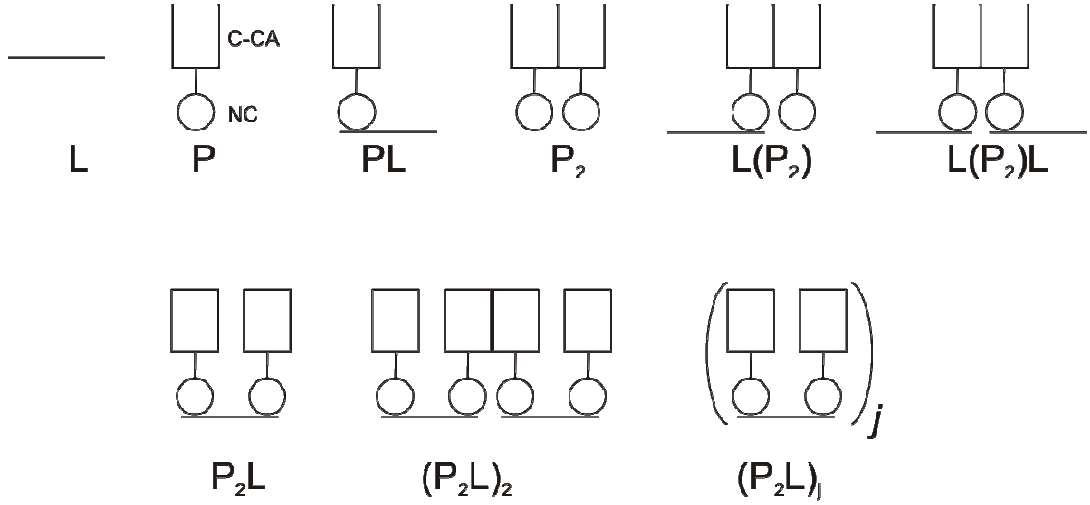


Figure 5.14: The different complexes and linear polymers which can form as a result of C-CA-NC/dSL3 interactions.

We wish to calculate the species distribution function, and begin by writing an expression for the total protein concentration P_T, starting from

$$P_T = (\text{even } j \text{ terms}) + (\text{odd } j \text{ terms})$$

$$(\text{even } j \text{ terms}) = \sum_{j=2,4\ldots}^{N} j\{(P_2L)_{j/2} + P(P_2L)_{(j-2)/2}P + LP(P_2L)_{(j-2)/2}P + LP(P_2L)_{(j-2)/2}PL\}$$

$$(odd\ j\ terms) = \sum_{j=1,3,\ldots}^{N-1} j\{P(P_2L)_{(j-1)/2} + LP(P_2L)_{(j-1)/2}\} \qquad 5.16$$

A similar equation can be written for L_T which, applying the previous definitions, is given by

$$L_T = L + (\text{even } j \text{ terms}) + (\text{odd } j \text{ terms})$$

$$(\text{even } j \text{ terms}) = \sum_{j=2,4,\ldots}^{N} \left\{ \frac{j}{2}(P_2L)_{j/2} + \frac{j-2}{2}P(P_2L)_{(j-2)/2}P + \frac{j}{2}LP(P_2L)_{(j-2)/2}P + \frac{j+2}{2}LP(P_2L)_{(j-2)/2}PL \right\}$$

$$(odd\ j\ terms) = \sum_{j=1}^{N-1} \left\{ \frac{j-1}{2}P(P_2L)_{(j-1)/2} + \frac{j+1}{2}LP(P_2L)_{(j-1)/2} \right\} \qquad 5.17$$

We must eliminate terms to yield an equation in a single unknown only. This is accomplished by use of the equilibrium constants and back-substitution.
The lower-order states (low *j*) can be found in terms of *P* and *L* by considering the association reactions between P and L, as already discussed, and C-CA dimerisation. The first of these considerations yields the equations

$$\frac{2P_2L}{P.PL} = K_N \qquad 5.18$$

$$\frac{PL}{2P.L} = K_N \qquad 5.19$$

These can be combined to give

$$P_2L = K_N^2 P^2.L \qquad 5.20$$

Consideration of C-CA dimerisation yields

$$P_2 = K_D P^2 \qquad 5.21$$

All terms can now be written in terms of *P*, *L* and the equilibrium constants. The term $P(P_2L)_{(j-2)/2}P$, for example, can be found by considering its formation from P and $P(P_2L)_{(j-2)/2}$, which involves addition of P *via* the C-CA interface. The dimerisation constant can hence be written

$$\frac{P(P_2L)_{(j-2)/2}P}{P.P(P_2L)_{(j-2)/2}} = K_D \qquad 5.22$$

Therefore

$$P(P_2L)_{(j-2)/2}P = K_D P.P(P_2L)_{(j-2)/2} \qquad 5.23$$

The term $P(P_2L)_{(j-2)/2}$ can be eliminated by considering its formation by addition of P to $(P_2L)_{(j-2)/2}$, again using the C-CA interface, to give

$$P(P_2L)_{(j-2)/2} = 2K_DP.(P_2L)_{(j-2)/2} \quad 5.24$$

Therefore

$$P(P_2L)_{(j-2)/2}P = 2K_D^2P^2.(P_2L)_{(j-2)/2} \quad 5.25$$

Next, we seek to eliminate $(P_2L)_{(j-2)/2}$. The concentration of such a polymer of repeating (P_2L) subunits can be written

$$\begin{aligned}(P_2L)_k &= (2K_D)^{k-1}(P_2L)^k \\ &= (2K_D)^{k-1}(K_N^2P^2.L)^k\end{aligned} \quad 5.26$$

We then have

$$\begin{aligned}P(P_2L)_{(j-2)/2}P &= 2K_D^2P^2(2K_D)^{(j-4)/2}(P_2L)^{(j-2)/2} \\ &= 2K_D^2P^2(2K_D)^{(j-4)/2}(K_N^2P^2.L)^{(j-2)/2}\end{aligned} \quad 5.27$$

Similar reasoning gives, for the other 'even *j*' terms,

$$LP(P_2L)_{(j-2)/2}P = 4K_NL.P(P_2L)_{(j-2)/2}P \quad 5.28$$

$$LP(P_2L)_{(j-2)/2}PL = K_NL.LP(P_2L)_{(j-2)/2}P \quad 5.29$$

The 'odd *j*' terms are given by

$$P(P_2L)_{(j-1)/2} = K_DP.(2K_D)^{(j-3)/2}(K_N^2P^2.L)^{(j-1)/2} \quad 5.30$$

$$LP(P_2L)_{(j-1)/2} = 2K_NL.P(P_2L)_{(j-1)/2} \quad 5.31$$

In terms of P, L, K_D and K_N only, the equations for P_T and L_T read

$$\begin{aligned}P_T = &\sum_{j=2,4\ldots}^{N} j\{(2K_D)^{(j-2)/2}(K_N^2P^2.L)^{j/2} + 2K_D^2P^2.(2K_D)^{(j-4)/2}(K_N^2P^2.L)^{(j-2)/2} \\ &+ 4K_NL.2K_D^2P^2.(2K_D)^{(j-4)/2}(K_N^2P^2.L)^{(j-2)/2} \\ &+ K_NL.4K_NL.2K_D^2P^2.(2K_D)^{(j-4)/2}(K_N^2P^2.L)^{(j-2)/2}\} \\ &+ \sum_{j=1,3,\ldots}^{N-1} j\{K_DP.(2K_D)^{(j-3)/2}(K_N^2P^2.L)^{(j-1)/2} \\ &+ 2K_NL.K_DP.(2K_D)^{(j-3)/2}(K_N^2P^2.L)^{(j-1)/2}\}\end{aligned} \quad 5.32$$

$$L_T = L + \sum_{j=2,4,\ldots}^{N} \left\{ \frac{j}{2}(2K_D)^{(j-2)/2}(K_N^2P^2.L)^{j/2} + \frac{j-2}{2}2K_D^2P^2.(2K_D)^{(j-4)/2}(K_N^2P^2.L)^{(j-2)/2} \right.$$
$$+ \frac{j}{2}4K_NL.2K_D^2P^2.(2K_D)^{(j-4)/2}(K_N^2P^2.L)^{(j-2)/2}$$
$$\left. + \frac{j+2}{2}K_NL.4K_NL.2K_D^2P^2.(2K_D)^{(j-4)/2}(K_N^2P^2.L)^{(j-2)/2} \right\}$$
$$+ \sum_{j=1}^{N-1} \left\{ \frac{j-1}{2}K_DP.(2K_D)^{(j-3)/2}(K_N^2P^2.L)^{(j-1)/2} \right.$$
$$\left. + \frac{j+1}{2}2K_NL.K_DP.(2K_D)^{(j-3)/2}(K_N^2P^2.L)^{(j-1)/2} \right\} \qquad 5.33$$

Subject to a defined P_T, such equations can be solved numerically, for example by means of a Newton-Raphson procedure, to obtain L and L_T. Back substitution yields concentrations of all intermediates.

The results of a simulation with a total C-CA-NC concentration of 50 µM dSL3 concentration of 25 µM are shown in Figure 5.15. The dSL3 affinity was 1 µM and the C-CA dimerisation constant 18 µM. The upper limit, *N*, of the length of oligomer that can form could not be extended beyond 500, since obtaining solutions to the above equations became impossible above this. However, it is clear from Figure 5.15 that the species distribution function for the two-parameter cross-linked dimer model is broad, and large polymers are rarer than short oligomers. Certainly, large polymers are not favoured, which is not the case in capsid assembly.

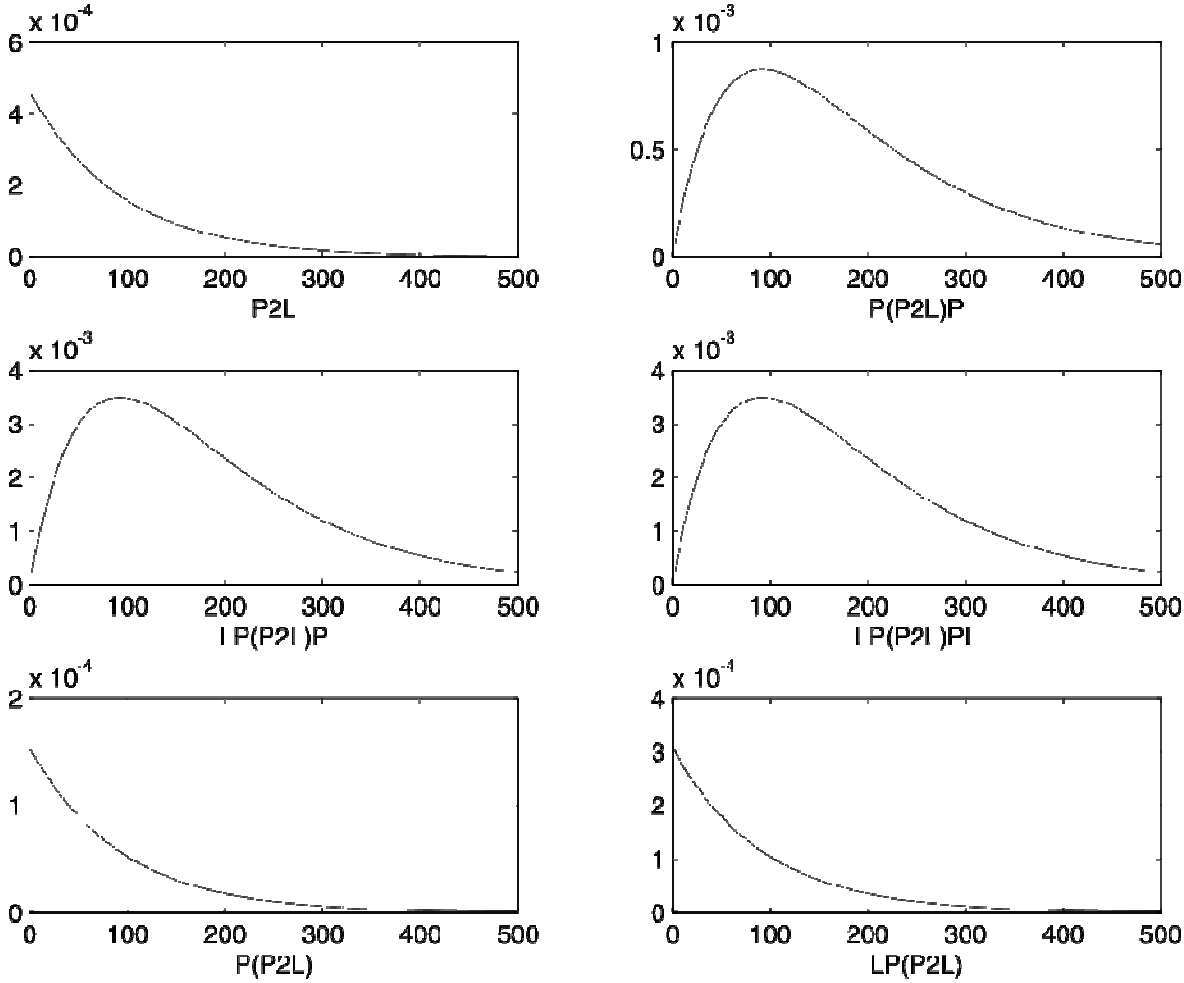


Figure 5.15: Simulated species distribution functions for all possible oligomers formed from C-CA-NC and dSL3, when the concentration of dSL3 is half that of C-CA-NC.

The y-axes give the mole fractions of each oligomer, and their different scale ranges should be noted when examining the figure. The x-axes show the number of C-CA-NC molecules contributing to the particular class of oligomer labelled below the axes.

Chapter 6 General discussion

6.1. The CCA dimer conformation hypothesis

In the introduction, it was explained that C-CA can crystallise as a dimer in several different forms. Although the C-CA domain is very similar in each dimer, the relative orientations of the two contributing molecules are different. The more recent cET and electron crystallography studies have shown that the C-CA dimer conformation in the immature state is different from that of the mature state (25, 47, 123, 159). The NMR data presented in this thesis are consistent with a model in which the C-CA dimer is a dynamic ensemble of dimer conformations. With this in mind, the following hypothesis is made

The conformation of the C-CA dimer is a critical determinant of capsid morphology and a restricted set of dimer conformations are selected by each assembly phenotype.

What is meant by this is that, if the C-CA domain forms a dimer conformation compatible with the immature state, an immature capsid will form, whereas if C-CA forms a dimer conformation compatible with the mature state, a mature capsid will form. Hence the C-CA dimer conformation effects the selection of an appropriate N-CA interface. By "restricted set of dimer conformations" is meant that it is not necessary for a single conformation to exist in either the mature or immature state, but that the two assembly states select regions of conformational space which are distinct from one another.

This leaves the question of how the C-CA dimer conformation is selected. The NC domain does not *per se* result in the selection of the immature dimer in preference to the mature, for C-CA and C-CA-NC behave in basically the same way according to the NMR experiments presented. However, the cross-linked dimer model introduced previously requires that, in order for an immature capsid to form, C-CA dimers be linked to one another by NC-nucleic acid interactions. It is possible to explain C-CA dimer conformational selection if the mature dimer is incompatible with the 2:1 stoichiometry required as the minimum Gag:nucleic acid ratio. For example, in the particular case studied of ΔMACANCSP2 with dSL3, a P_2L complex (in the notation of chapter 5.1) must bind another P_2L complex by C-CA dimerisation. It may simply be that the immature dimer state positions the two P_2L complexes in a more favourable arrangement for further elongation, whereas the mature dimer state positions the two P_2L complexes in such a

fashion as to preclude further elongation. If this suggestion holds, then it is the elongation of a $(P_2L)_n$ to a $(P_2L)_{n+1}$ complex which forces the immature dimer to be selected. Upon selection of the immature dimer, which is supposed to be the only form permissive to further elongation, the immature N-CA interface is selected for hexamerisation, which is also supposed to be the only form permissive to further elongation.

It may be possible to explain C-CA dimer conformational selection using the same simple thermodynamic argument used to account for the necessity of SP1 removal from CA prior to mature assembly. That is, the entropic loss resulting from constraints to the conformational space accessible to the linker could be minimised when C-CA is in the immature dimer conformation.

In the absence of the SP1 and NC domains, the immature assembly phenotype is generally preferred (59, 83). If the C-CA dimer conformation determines the assembly phenotype, how is the mature C-CA dimer selected in preference to the immature? That is, why doesn't CA assemble as a mixture of immature and mature capsids? A possible answer is that, whilst the C-CA dimer is probably an ensemble of structures with near-degenerate energy levels, the mature dimer may have a slightly lower energy level than the immature dimer after proteolysis by PR, giving a slight preference to the mature dimer. The obvious question remains of why the mature dimer should have a lower energy than the immature. A possible answer is provided by the arguments made in chapter 4.13, in which it was suggested that SP1 removal from CA was necessary to minimise the entropic loss concomitant with assembly. The removal of SP1 may lead to a slight preference for the mature dimer, but may not be a noticeable preference when intermediates on the assembly pathway are only at the level of small oligomers. However, as a capsid grows, the effect may become more noticeable. The essential argument here is that distinct C-CA dimer conformations will not be observable until large intermediates on the assembly pathways are reached, when energy levels amongst dimer conformations diverge.

6.2. The immature and mature assembly pathways

It would be useful to summarise the pathways taken to immature and mature assembled states, and identify the similarities and differences between the pathways. In particular, the point at which C-CA dimer conformational selection takes place, and the point at which N-CA hexamer conformational selection take place are of great interest. It is reasonable to

assume that C-CA is dimeric from the earliest point in both immature and mature assembly, for C-CA dimerisation in solution is much stronger than any other CA interactions (no others are observable in solution). It is also reasonable to assume that, in the early stages of both immature and mature assembly, small oligomers must form in which C-CA still exists as an ensemble of dimeric structures, based on the observation that ensembles of C-CA dimers exist in both isolated C-CA and C-CA-NC. The point at which the assembly pathways diverge, at least in terms of the CA lattices formed, thus remains unknown. There must exist a point on either pathway, however, at which C-CA adopts a particular conformation, and subsequent elongation steps lead naturally to the incorporation of subunits in which the C-CA conformation corresponds to that adopted by those already C-CA domains already present in the growing structure. A similar argument applies to N-CA hexamer conformation, which may or may not be concomitant with the C-CA dimer conformational selection. Once a growing oligomer reaches a point at which conformational selection in either C-CA or N-CA occurs, therefore, the immature and mature pathways diverge, and the final assembled phenotype is determined as of that point. One could regard the early stages of these hypothetical pathways as nucleation events, where nucleation now has the meaning of the formation of an oligomer in which the first conformational selection events in C-CA and N-CA occur. Of course, the sizes of such nuclei are unknown.

As a unified (albeit speculative) picture of immature and mature assembly, therefore, one can picture the pathways as having several distinct phases. The first is a form of nucleation, in which oligomers form, but no definite CA conformational selection has occurred. The second is the transition from such oligomers to ones in which particular conformations are selected, which may be considered a transient phase for each particular oligomer. The third phase is the elongation of such oligomers in a reasonably defined sense, to yield the final capsids. In the third phase, a CA molecule incorporated into the growing structure is forced to select the conformation which corresponds to that adopted at the initiation of the second phase of the pathway. The 3 phases will be referred to as nucleation, selection and elongation respectively. This abstract picture of immature and mature assembly of course leaves the particular details of each phase undefined, but allows at least some conceptual unification of the separate assembly processes.

6.3. The purposes of immature and mature assembly

Immature assembly appears to be a very robust process, by which is meant that it can proceed with a high yield and high fidelity of capsid morphology across a broad range of protein and nucleic acid concentrations. This is presumably of great advantage to HIV-1, since good use will be made of most of the Gag which is expressed by the host cell, with very little going to waste. However, as argued, immature capsids are also difficult to dissociate on account of their insensitivity to protein and nucleic acid concentrations. That is, immature capsids have a hysteresis to dissociation. This is of no use to HIV-1 ultimately, since such stable capsids would not release their contents, namely the machinery for production of viral progeny. In some systems, spherical capsids in abundance are observed within the cell cytoplasm, for example expression of ΔMACANCSP2 in *E.Coli* (60), demonstrating immature capsid stability at least in prokaryotes. Thus, having a robust immature assembly system necessitates the incorporation of a second, mature, assembly system so as to produce virions which can easily release their genomes and various enzymes into their hosts, a prerequisite for replication.

The mature capsid must be inherently unstable in the host cell cytoplasm but stable in the virion before cell entry, and in order to meet these criteria, it is to be expected that it is difficult to assemble the mature capsid. It is indeed the case that mature assembly requires harsher conditions than immature assembly using *in vitro* assays, with high yields of mature-like structures requiring either very high NaCl concentrations (in the molar range) or the addition of a crowding agent such as ficoll or dextran. These conditions essentially provide a fictitiously high protein concentration, and in the small volume of the virion, packed with thousands of protein molecules in a highly confined space, such conditions are presumably met also. Hence maturation serves to produce a capsid which is stable when the local CA concentration is high, but which does not have the hysteresis associated with the immature capsid. Its purpose may essentially therefore be to prepare the virion to release its contents into a new host.

The purpose of immature assembly, on the other hand, is to assemble a large number of capsids, packaging the genome in the process, and allow exit of the immature virions from the host cell. The first requirement is met by the means already described. The second requirement, namely genome packaging, will be discussed separately. For the time being, it

suffices to say that the protein responsible for genome packaging, NC, inasmuch as being a component of Gag, is inherently present in the immature virion, leaving only the question of selective genome recognition.

The final requirement, cell exit, requires that immature capsids either be assembled at the cell membrane, or are targeted to the cell membrane, in order that membrane curvature be induced (54, 110). Immature assembly indeed takes place at the plasma membrane, since MA is myrsitoylated, though the process appears to start in the cytoplasm (87). However, MA alone appears not to result in membrane curvature, as it has been shown to assemble on membranes as flat 2D crystals (8). It may be the formation of the immature lattice, therefore, which allows a curved MA layer (ultimately the virion's matrix) to form immediately below the cell membrane, thereby contributing to the membrane curvature necessary for budding and release of fully assembled immature virions. Hence immature assembly should be considered a mechanism of inducing membrane curvature, which would require an independent mechanism if not inherent in immature assembly. That is to say, a single assembly step with a mature capsid would require the incorporation of a distinct mechanism into the retroviral lifecycle for the induction of membrane curvature, and also for targeting the unit of assembly to the membrane for incorporation into a virion. By incorporating the membrane targeting unit, capsid assembly unit and genome packaging unit into a single molecule (namely Gag), several separate tasks are accomplished by a single step in the retroviral lifecycle. This allows for economic use of the genome, which is characteristic of viruses and retroviruses.

6.4. Selective genome recognition and inhibiting assembly with NC binding molecules

Several possible means of treating HIV-1 infection by targeting NC have been suggested. In particular, the covalent modification of the cysteine residues involved in Zn^{2+} coordination has been explored, but no molecules found so far have been able to discriminate between NC and cellular zinc-binding proteins (69, 131).

It has been shown in chapter 5 that addition of a molecule which binds NC but doesn't promote assembly is not capable of inhibiting immature assembly, by using dACGCC and ΔMACANCSP2. This implies that inhibiting immature assembly by targeting NC may not be a viable antiviral mechanism. However, molecules with a higher affinity for NC may be effective at inhibiting immature assembly and should be sought nonetheless, since the

affinity of dACGCC for NC may simply be insufficient to fulfil this task. It was noted in the introduction that NC has other functions besides its role in assembly, assisting in reverse transcription (18, 130). Small molecules binding NC may be effective at inhibiting these processes, which may be reliant upon greater sequence specificity than immature assembly.

Given the variation of sequences which can be recognised by NC, which include homopolymers such as poly(G) and poly(A) amongst others (28), it is difficult to imagine how NC manages to mediate any process in the host cell cytoplasm, which at any instance contains many different DNA and RNA molecules, if it does not specifically recognise a subset of sequences present in the viral genome. The selection of genomic RNA appears to be reliant upon the presence of the ψ-site, as reviewed in chapter 1.7, which contains 4 stem-loops, and it may be that it is the presence of the stem-loop secondary structure, rather than the actual sequence, that facilitates high-affinity recognition by NC. It might be possible to apply similar arguments as to how the roles of NC in primer tRNA annealing and strand transfer during reverse transcription occur.

It is possible that HIV-1 simply relies on producing so much Gag that, although at any instant most NC molecules will not be bound to appropriate nucleic acid, sufficient do find an appropriate target that the lifecycle is completed. This argument is based on the number of Gag molecules contributing to a virion being in the range 1500-5000, but only 2 genomic RNA molecules need be recognised for production of an infectious virion. The strand annealing roles of NC are fulfilled by free NC, after entry of a mature virion into a new host. Again the argument can be applied that very few NC molecules actually need to involve themselves in reverse transcription. The vast majority of NC molecules may never encounter an HIV-1 nucleic acid, for in principle reverse transcription need take place only once per infection event for the production of many progeny, even though up to 5000 NC molecules will enter the host. It is difficult, if this argument is correct, to see that small molecules against NC could ever be effective as anti-virals unless they have a very high affinity for NC, and are present at intracellular levels in excess of the number of NC molecules which enter any host cell. Since the myriad of host cell nucleic acid molecules fail to out-compete the appropriate NC interactions, doing so with one other molecule could be extremely difficult.

An alternative is to design molecules which compete with NC for nucleic acid, that is, bind nucleic acid rather than NC itself. Tryptophan-rich hexapeptides have been shown to do so, but not yet *in vivo* (127) and discrimination of viral from cellular nucleic acid may be a problem.

6.5. Directions of future research

6.5.1. C-CA-NC as a solid state NMR model of the immature dimer

The similarity in behaviour between ΔMACANCSP2 and C-CA-NC when assembled in the presence of dSL3 raises the interesting possibility of studying such molecules as models for the immature state using solid-state NMR (ssNMR). Since ssNMR is not limited *per se* by the size of the system under study, but remains sensitive to both structure and dynamics, it has considerable potential in probing the nature of C-CA dimerisation in the immature state, and the behaviour of the linker. Both C-CA-NC and ΔMACANCSP2 could potentially be studied using ssNMR. The large number of residues in ΔMACANCSP2, however, could mean that assignments are difficult on account of spectral crowding, and for this reason C-CA-NC may make a better model. As a means of examining C-CA dimerisation when not a component of an assembled state, frozen crystals of C-CA could be used. However, the perturbations to the dimerisation interface resulting from (noon-physiological) crystallisation conditions could complicate interpretations.

The visible solution NMR chemical shifts of the C-CA domain of C-CA-NC may remain unchanged in the solid state, which would be indicative that the corresponding residues are not involved in an interaction in the immature state, provided C-CA-NC is indeed representative of the immature state (a point which will be returned to shortly). An obvious consequence of coincident C-CA chemical shifts between soluble C-CA-NC and aggregated C-CA-NC/dSL3 is that assignments can be carried across from solution NMR spectra, which would be of very great use. If the solution C-CA-NC chemical shifts change in the solid state, i.e. in aggregated C-CA-NC/dSL3, then a change in C-CA structure or its involvement in an interaction which doesn't occur in solution can be inferred. It may also be the case that the dimer interface becomes better defined, which is a prediction of the dimer conformation hypothesis. That is, the C-CA dimer may adopt a unique structure, in which one expects that peaks will appear for such residues. Hence ssNMR provides a means of testing the dimer conformation hypothesis.

In order to compare solution- and solid-state chemical shifts, carbon-carbon correlation experiments should be performed in the solid state, making use of cross-polarisation and carbon detection. An obvious choice of experiment is proton-driven spin diffusion, and such experiments could be performed with variable mixing times (during which spin diffusion occurs). The use of variable mixing times can report on the distances over which correlations occur.

It is a useful feature of cross-polarisation (CP) that it is ineffective if large-amplitude dynamics are present. Hence, in a CP-based experiment, flexible regions of protein are essentially 'filtered out' of the spectrum. The linker, which is highly flexible in solution, may therefore not contribute to the solid-state spectrum under CP conditions if it remains flexible, but will appear if it becomes structured. Assignments would be necessary to study it properly in this case, for example by determining secondary Cα shifts which could demonstrate the SP1 helix which has been suggested and sought, but never conclusively proved.

If the linker is flexible even in an assembled state, it may be visible if an INEPT transfer is used to transfer coherences from ^{1}H to ^{13}C rather than CP. Under such conditions, the rigid parts would not be visible since INEPT transfers are generally ineffective in the solid state. Hence an INEPT-based experiment essentially filters out all but the most flexible parts of a molecule from the ssNMR spectrum and can be used to demonstrate flexibility in the assembled state. If this is the case, a stable SP1 helix can be ruled out, but this observation would allow alternative roles of the linker to be explored.
A consistent observation when dSL3 was used for solution NMR was the loss of signal intensity in the NC domain. This made the study of dSL3/NC interactions intractable by NMR, but may not be a bad thing since if the same effects are observed in ssNMR, then the spectra will simply not be as crowded and will make the C-CA and linker regions easier to assign. A troublesome result may therefore become a useful feature in ssNMR.
Although the small size of C-CA-NC makes it a convenient model for the immature state, and there is some evidence to justify its use, controls to determine whether it is an appropriate model should be carried out. A comparison of C-CA-NC PDSD spectra to the same experiment for assembled ΔMACANCSP2 should be used, with coincident peaks in ΔMACANCSP2 for all C-CA-NC peaks being a reasonable justification of its use.

Several studies have implicated the formation of a β-hairpin in the N-terminus of N-CA as a determinant of capsid morphology. It is thought that the hairpin doesn't occur in full-length Gag and is not present in the immature capsid, but upon cleavage of MA from N-CA, the hairpin forms and results in the formation of a mature N-CA hexamer. It would be very interesting to determine whether formation of the β-hairpin actually directs the selection of the entire capsid morphology, or whether its formation renders it compatible only with a particular form of N-CA hexamer. If assembled C-CA-NC has a C-CA dimer conformation the same as that in assembled ΔMACANCSP2, then the β-hairpin is probably not responsible for selecting the assembly phenotype, but may nevertheless be a requirement of the N-CA hexamer.

6.5.2. *In silico* investigations of the immature dimer

It may be possible to determine the conformational space accessible to the C-CA dimer by means of molecular dynamics simulations. If an arbitrary dimer conformation is initially used, for example one of the various X-ray structures, then one can simulate how this structure evolves with time by means of free energy calculations. By this means, the energy of the dimer can be computed as a function of the dimer conformation. It may be observed that many conformations are explored, or alternatively that only very few conformations are stable, and it is amongst these conformations that the molecule exchanges. Both of these possibilities are consistent with the NMR data (or, rather, the lack of NMR data for the dimer interface). A potential problem is that it is impractical to use very long trajectories in MD simulations (50 ns is practical given sufficient computing power), so it can be difficult to detect slow dynamics. Hence, if exchange between dimer conformations is slow compared to the timescale of an MD simulation, one may not observe all the dimer conformations, for the initial structure may be stable in its starting conformation over the duration of the trajectory.

A much more ambitious task is the simulation of an entire immature capsid. It would of course be extremely interesting to do so, but given the complexity of the immature capsid, and the potential variability of timescales of process contributing to assembly, this task is probably beyond the practical capability of current MD and computing technology.

6.5.3. Further investigation of C-CA dimerisation in solution

The nature of C-CA dimerisation is by no means fully understood, and should be the subject of further experimental work as well as the computational studies suggested above.

This is of particular relevance from the standpoint of developing anti-HIV agents, since 'trapping' C-CA in any particular conformation should be inhibitory to one or both assembly steps. If the nature of dimerisation can be understood, it may be possible to develop agents which can do so, and at the same time develop means of determining the selected conformation.

The timescale of dimerisation has been very difficult to determine by the NMR experiments performed so far. However, some of the R_{ex} terms obtained for C-CA are sufficiently large to enable the use of relaxation dispersion experiments, which can determine rates of association (30).

It may be possible to obtain some information about the solution dimer by means of NOESY experiments with C-CA and C-CA-NC. If intra-molecular NOEs can be observed, then those residues which are always close in the dimer structural ensemble can be found, and this may begin to make details of the solution dimer ensemble tractable. Factors such as pH and temperature could also be varied, to see whether they influence the dimer ensemble. The variation of pH may be interesting, since the ΔMACANCSP2 assembly phenotype depends upon pH, with tubes forming at pH 6 but spheres at pH 8, although it is not clear whether a change in C-CA dimerisation or NC behaviour is responsible. Nevertheless, variation of pH may be useful not only in characterising particular dimer conformations but also in validating the dimer conformation hypothesis.

As highlighted in chapters 4.4 and 4.6, it is difficult to interpret the NMR data without prior knowledge of the timescale of exchange between monomeric and dimeric states, for this timescale determines the meaning of the mono-exponential functions used to interpret relaxation decay curves. It is generally the case that processes such as protein dimerisation result in, for example, chemical shift changes or some exchange broadening, either of which can accurately report upon the timescale of monomer/dimer exchange. In the case of C-CA, however, this was not the case. The chemical shifts did not depend upon protein concentration for any residues, and no signals at all were observable throughout the dimer interface. Thus there is no information in the NMR data for either C-CA or C-CA-NC which can reveal, even approximately, the timescale of monomer/dimer exchange. To be certain that the relaxation data for residues which are accessible to NMR has been interpreted correctly, therefore, it is necessary to find at least some approximation to the exchange timescale from another technique. An obvious choice is fluorescence quenching,

for the dimer interface contains a tryptophan residue which is important for dimerisation, but it was found that the fluorescence signal for C-CA-NC is independent of concentration (within experimental accuracy), so this tryptophan may not be a good reporter. However, such experiments were performed with C-CA-NC, and the W184 fluorescence signal may have been masked by a stronger signal from W282, so C-CA might be a better system to study.

It was suggested earlier in this discussion that the linker region may have a role in C-CA dimer conformational selection. It may be interesting to prepare a mutant of C-CA which terminates at residue G220 (rather than L231 as dictated by the position of the CA-SP1 cleavage site), thus lacking a flexible C-terminal tail, in order to see whether the absence of C-terminal flexibility reduces dimer conformational flexibility. It is noteworthy that a CA mutant terminating at residue G220 still assembles with a mature phenotype. It may be the case that a particular dimer conformation is preferred in this case, in which case peaks for residues in the dimer interface should become observable using solution NMR. If no such change occurs, then one can conclude that dimer conformational selection occurs only once larger intermediates have formed. A similar experiment is the preparation of C-CA mutant which includes the SP1 domain in addition to its usual 12 flexible C-terminal residues, in total spanning residues 146-245 of Gag. If it observed that this has reduced conformational flexibility around the dimer interface, then it is presumably in a dimeric state closer to that of the immature capsid.

6.5.4. Development of the cross-linked dimer model

Strong evidence has been presented in this thesis for a cross-linked dimer model of immature assembly, in which C-CA dimers are connected to other C-CA dimers by NC-nucleic acid interactions. However, the evidence is not conclusive, and further experiments to test the predictions of the model should be considered a priority. The possibility of electron microscopy as a test of the model has already been discussed in chapter 5. Several other predictions made about immature assembly on the basis of the model could also be tested. It was argued that a high nucleic acid concentration would, in the absence of an assembly process (for example using C-CA_{AA}-NC), result in a distribution of P_nL states in which the largest possible number of nucleic acid molecules are involved in an NC-nucleic acid interaction. This would saturate NC with nucleic acid in PL complexes, rather than the P_2L complexes predicted to be the minimum requirement of immature assembly. The

requirement for P_2L complexes as the minimum requirement of immature assembly could be tested. This could be accomplished by assembling an appropriate system, then determining the ratio of protein to nucleic acid in the assembled fraction.

The predicted inability of nucleic acid contributing to a particular assembled capsid to exchange with nucleic acid in solution could also be tested, for example by assembling an appropriate system, separating the assembled fraction, and then re-suspending with a nucleic acid solution which has been labelled in some way. Such a label could be a fluorescent tag, radiolabel or non-radioactive isotope label (e.g. ^{15}N) . Upon once more separating the assembled fraction, its nucleic acid content could be tested for labelled nucleic acid. The amount of labelled nucleic acid could be quantified, in principle providing rates of exchange of nucleic acid in solution and in the immature capsids. Of course, such rates are predicted to be very low.

Almost the same experiment could be performed to determine whether protein molecules in solution exchange with those in assembled capsids, by using a labelled protein instead of labelled nucleic acid. Such a label could be an isotope label, in which case solution NMR can be used. If unlabelled protein is used in the first assembly reaction, then labelled protein is added to the re-suspended assembled material, a deviation from the expected correlation time would imply that the labelled protein has been incorporated, at least to some extent, into assembled capsids.

6.5.5. Development of novel inhibitors of assembly

The development of inhibitors of assembly which bind C-CA or N-CA offer a promising route to new anti-retrovirals. It was remarked in the introduction that a small molecule, DSB, is capable of inhibiting assembly by binding at the CA-SP1 cleavage site. This raises the possibility of designing molecules which bind the various other PR cleavage sites within Gag. Such efforts could be guided by the known crystal structure of PR, by incorporating features of its active site into the new molecules, and by the known interactions of the various cleavage sites with the PR active site.

A potentially important field to research in the development of new anti-retrovirals is the extent to which non-specific nucleic acid recognition by NC plays a role in the retroviral lifecycle. An equally important field is whether NC recognises secondary structure rather than sequence, which could be investigated using a systematic study of different nucleic

acids of known structure, including stable stem-loops. This could impact on the type of molecules worthy of consideration as potential NC-binding anti-retrovirals.

Chapter 7 Appendix

7.1. Protein sequences

C-CA

Residues 146-231, as numbered from the N-terminus of N-CA.
MSPTSILDIR QGPKEPFRDY VDRFYKTLRA EQASQEVKNW MTETLLVQNA NPDCKTILKA LGPGATLEEM MTACQVGGPG HKARVL

C-CA$_{AA}$

The sequence is identical to C-CA, but with W184A and M185A mutations.
MSPTSILDIR QGPKEPFRDY VDRFYKTLRA EQASQEVKNA ATETLLVQNA NPDCKTILKA LGPGATLEEM MTACQVGGPG HKARVL

C-CA-NC

Residues 146-300 as numbered from the N-terminus of N-CA.

MSPTSILDIR QGPKEPFRDY VDRFYKTLRA EQASQEVKNW MTETLLVQNA NPDCKTILKA LGPGATLEEM MTACQGVGGP GHKARVLAEA MSQVTNPATI MIQKGNFRNQ RKTVKCFNCG KEGHIAKNCR APRKKGCWKC GKEGHQMKDC TERQAN

C-CA$_{AA}$-NC

MSPTSILDIR QGPKEPFRDY VDRFYKTLRA EQASQEVKNA ATETLLVQNA NPDCKTILKA LGPGATLEEM MTACQGVGGP GHKARVLAEA MSQVTNPATI MIQKGNFRNQ RKTVKCFNCG KEGHIAKNCR APRKKGCWKC GKEGHQMKDC TERQAN

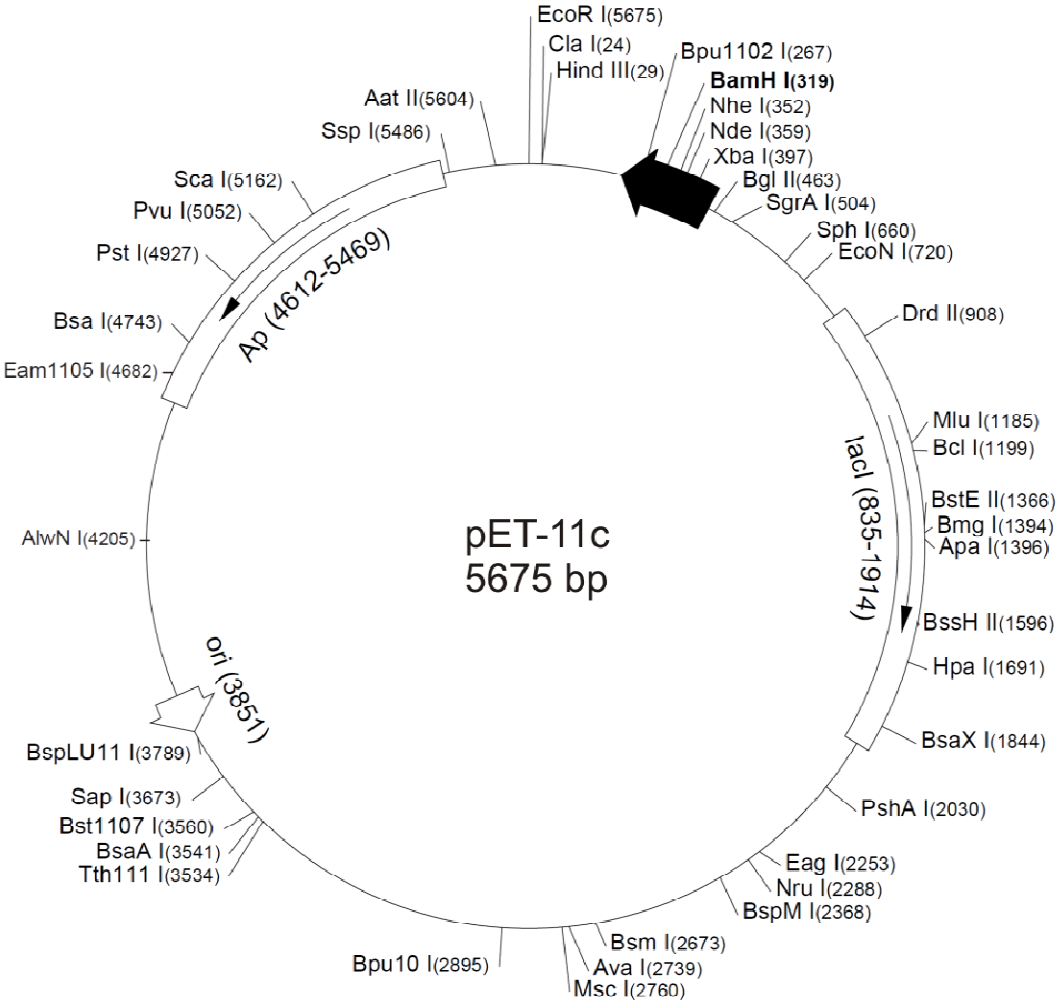
ΔMACANCSP2

This protein contains the MA domain with residues 15-99 deleted, the entire CA domain, along with SP1, NC and SP2.

GARASVLSGG ELDLDKIEEE QNKSKKKAQQ AAADTGHSSQ VSQNYPIVQN IQGQMVHQAI SPRTLNAWVK VVEEKAFSPE VIPMFSALSE GATPQDLNTM LNTVGGHQAA MQMLKETINE EAAEWDRVHP VHAGPIAPGQ MREPRGSDIA GTTSTLQEQI GWMTNNPPIP VGEIYKRWII LGLNKIVRMY SPTSILDIRQ GPKEPFRDYV DRFYKTLRAE QASQEVKNWM TETLLVQNAN PDCKTILKAL GPAATLEEMM TACQGVGGPG HKARVLAEAM SQVTNTATIM MQRGNFRNQR KMVKCFNCGK EGHTARNCRA PRKKGCWKCG KEGHQMKDCT ERQANFLGKI WPSYKGRPGN

7.2. Plasmid map for pET-11c vector

All proteins used in this thesis were expressed using the pET-11c vector (Novagen), as previously described (145). A map of this vector is given below.



7.3. Chemical shifts

This appendix lists all ^{1}H and ^{15}N shifts measured, and also the $^{13}C\alpha$ secondary shifts. The latter have been corrected for the presence of prolines (139).

Table 7.1: Chemical shifts for C-CA$_{AA}$-NC.

Residue	No.	`H (ppm)	^{15}N	2° Cα (ppm)
S	146	8.427	118.704	-0.127
P	147			
I	150	8.613	123.727	-0.025
L	151	7.595	119.345	1.228
D	152	7.7	116.141	1.221
I	153	7.47	122.276	-1.735
R	154	8.291	124.441	-2.269
Q	155	7.766	127.159	-0.221
G	156	9.881	116.818	-0.272
P	157			
K	158	8.584	115.083	-2.732
E	159	7.036	125.809	-2.5
P	160			
F	161	9.532	128.152	5.137
R	162	9.111	115.779	3.405
D	163	6.974	119.005	4.097
Y	164	7.613	124.572	1.216
V	165	8.308	120.255	4.037
D	166	7.555	117.688	4.719
R	167	7.95	119.21	3.716
F	168	9.039	124.619	4.204
Y	169	8.693	116.542	5.336
K	170	8.343	120.994	3.408
T	171	7.68	117.089	4.637
L	172	7.933	122.379	2.317
R	173	7.999	116.445	1.695
A	174	7.581	121.751	0.925
E	175	7.92	118.094	1.153
Q	176	8.206	118.803	0.323
A	177	8.014	122.837	0.148
S	178	8.174	114.349	-0.495
Q	179	8.557	121.323	0.867
E	180	8.493	119.801	2.007
V	181	7.844	120.813	1.084
K	182	8.359	123.229	0.843
N	183	8.492	119.394	0.881
A	184			0.925
A	185	8.325	122.182	0.614
T	186	8.197	114.778	3.393
E	187	8.597	120.096	3.64
T	188	7.761	112.72	3.005
L	189	7.873	123.388	2.083
L	190	7.639	119.526	2.239
V	191	7.513	116.879	3.493
Q	192	7.946	117.284	2.732
N	193	7.879	115.494	-0.518
A	194	7.337	124.839	-0.785

N	195	9.1	119.458	-1.683
P	196			-63.7
D	197	7.748	115.666	4.33
C	198	8.59	116.618	5.064
K	199	9.52	121.796	4.885
T	200	7.633	113.231	4.637
I	201	6.808	122.462	2.8
L	202	8.526	118.9	2.105
K	203	8.536	119.127	2.174
A	204	7.166	120.255	0.829
L	205	7.43	118.181	0.796
G	206	7.49	129.796	-0.902
P	207			-63.7
G	208	8.755	115.827	-0.197
A	209	7.162	123.4	-0.279
T	210	8.639	113.352	-1.717
L	211	8.941	122.592	2.81
E	212	8.612	117.743	4.304
E	213	7.79	120.619	3.398
M	214	8.542	120.182	4.422
M	215	8.782	117.085	3.718
T	216	8.378	116.999	4.524
A	217	8	122.908	2.037
C	218	7.389	110.881	1.562
Q	219	7.529	122.438	2.261
G	220	8.727	110.016	-0.097
V	221	7.377	120.473	0.904
G	222	8.802	109.761	-0.197
G	223	8.038	108.242	-0.6
P	224			
G	225	8.592	109.143	
H	226			
K	227	8.248	122.626	0.16
A	228	8.257	124.544	0.124
R	229	8.281	120.741	0.088
V	230	8.155	121.527	0.401
L	231	8.263	125.574	0.192
A	232	8.223	124.169	0.325
E	233	8.32	119.574	0.982
A	234	8.192	124.102	0.325
M	235	8.222	118.501	0.094
S	236	8.174	116.256	0.012
Q	237	8.309	121.746	-0.256
V	238	8.097	120.65	-0.102
T	239	8.193	117.809	-0.308
N	240	8.426	122.463	0.003
P	241			
A	242	8.283	122.278	0.325
T	243	7.951	112.687	0.498
I	244	7.919	122.686	-0.018
M	245	8.281	123.6	-0.007
I	246	8.067	122.355	-0.22
Q	247	8.389	124.115	-0.256
K	248			
G	249	8.36	109.384	
N	250	8.252	118.672	-0.386
F	251	8.191	120.713	-0.012
R	252	8.217	121.806	-0.113
N	253	8.217	121.806	0.017
P	254			
R	255	8.371	122.458	-0.214

K	256	8.411	123.225	-0.444
T	257	8.217	117.457	-0.408
V	258	8.404	123.244	-1.008
Q	259	8.158	124.639	-1.162
C	260	8.274	127.18	-3.672
F	261	8.705	128.803	0.995
N	262	9.607	121.691	3.44
C	263	8.932	117.457	0.153
G	264	7.919	113.419	1.011
K	265	8.143	120.998	-0.645
E	266	8.415	118.219	-0.025
G	267	8.656	107.544	0.407
H	268	7.198	114.111	-0.13
I	269	7.858	108.198	-2.334
A	270	8.84	124.735	3.446
K	271	8.276	117.123	2.174
N	272	7.928	114.189	-1.393
C	273	7.691	124.257	3.676
R	274	8.715	128.184	-56.48
A	275	8.761	128.194	-1.789
P	276			
R	277			-0.314
K	278	8.647	126.079	-0.343
K	279	8.444	123.5	-0.242
G	280	8.181	109.118	-0.801
C	281	8.205	124.14	
W	282			
K	283			
C	284	8.524	117.199	
G	285	8.159	113.84	
K	286	8.411	122.486	
E	287			
G	288			
H	289	7.169	112.942	-0.03
Q	290	8.907	117.021	-2.068
M	291	8.911	121.945	4.322
K	292	8.551	118.496	1.872
D	293	7.864	117.627	0.759
C	294	7.604	123.707	3.408
T	295	8.178	119.734	-0.207
E	296	8.659	125.532	
R	297			
Q	298	8.424	122.065	-0.256
A	299	8.388	126.219	-0.379
N	300	7.993	123.532	1.426

Table 7.2: Chemical shifts for C-CA-NC

Residue	No.	^{1}H (ppm)	^{15}N (ppm)
S	146		
P	147		
I	150		
L	151		
D	152	7.66525	114.8661
I	153	7.38367	122.2462
R	154		
Q	155	7.75093	127.5133
G	156	9.84443	116.789
P	157		
K	158	8.52754	114.9744

E	159	6.98762	125.7499
P	160		
F	161	9.4661	128.3304
R	162	9.04089	115.6846
D	163	6.93095	119.0384
Y	164	7.47388	124.2689
V	165	8.29945	120.1274
D	166	7.52273	118.2786
R	167	7.79597	118.9891
F	168	9.06395	124.1178
Y	169	9.06243	117.8227
K	170	8.22858	120.6572
T	171	7.55521	117.488
L	172		
R	173		
A	174		
E	175		
Q	176		
A	177	7.60635	125.131
S	178		
Q	179		
E	180		
V	181		
K	182		
N	183		
W	184		
M	185		
T	186		
E	187		
T	188		
L	189		
L	190		
V	191		
Q	192		
N	193	7.87441	115.1801
A	194	7.20061	124.7389
N	195	9.02485	119.3507
P	196		
D	197	7.69306	115.513
C	198	8.52183	116.6227
K	199	9.49585	121.7056
T	200	7.56338	113.1333
I	201	6.76368	122.4518
L	202	8.47938	119.0114
K	203	8.47512	119.0148
A	204	7.14637	120.3453
L	205	7.39798	118.3295
G	206	7.43928	104.6263
P	207		
G	208		
A	209	7.19574	123.4197
T	210	8.62702	113.4539
L	211	8.80902	122.6649
E	212	8.54252	117.6769
E	213	7.7485	120.6166
M	214	8.49337	120.1213
M	215	8.77498	117.0594
T	216	8.35266	117.0419
A	217	7.93195	122.9282
C	218	7.33086	110.879

Q	219	7.4774	122.4826
G	220		
V	221	7.30548	120.4625
G	222	8.68475	115.9523
G	223	7.97331	108.1726
P	224		
G	225		
H	226		
K	227	8.14969	122.5009
A	228	8.17559	124.1003
R	229	8.17637	120.6471
V	230	8.08608	121.3499
L	231	8.18265	125.2208
A	232	8.14286	123.9146
E	233	8.25928	119.4863
A	234	8.11273	123.9054
M	235	8.15867	118.315
S	236	8.10443	116.1418
Q	237	8.23678	121.6166
V	238	8.02723	120.4727
T	239	8.13322	117.6848
N	240	8.37603	122.2876
P	241		
A	242	8.22732	122.166
T	243	7.89589	112.6031
I	244	7.86165	122.6595
M	245	8.22923	123.5401
I	246	8.01617	122.3027
Q	247	8.33984	124.0506
K	248	8.3852	122.2528
G	249	8.287	109.322
N	250		
F	251		
R	252		
N	253	8.16445	121.7389
P	254		
R	255		
K	256	8.32475	122.4466
T	257	8.15864	117.4143
V	258	8.34796	123.1971
Q	259	8.0952	124.6301
C	260	8.19202	127.1372
F	261	8.57777	128.466
N	262	9.54819	121.607
C	263	8.84129	117.2936
G	264	7.84842	113.4169
K	265	8.06995	120.9047
E	266	8.35144	118.1146
G	267	8.59935	107.4936
H	268	7.1384	114.0866
I	269	7.80562	108.1896
A	270	8.74742	124.7057
K	271	8.23947	117.1775
N	272	7.87252	114.1381
C	273	7.64535	124.2951
R	274		
A	275	8.70922	128.1208
P	276		
R	277	8.42568	123.048
K	278	8.61672	126.2344

K	279	8.39653	123.2419
G	280	8.13032	109.0727
C	281	8.11787	124.2591
W	282	8.43995	129.5456
K	283	9.33247	122.9018
C	284	8.48288	117.1837
G	285	8.11496	113.8237
K	286	8.38628	122.2429
E	287		
G	288		
H	289	7.10974	112.8775
Q	290	8.86442	116.9114
M	291	8.88271	121.936
K	292	8.51209	118.4968
D	293	7.81599	117.5659
C	294	7.55529	123.7156
T	295	8.14163	119.7452
E	296	8.60942	125.5378
R	297	8.35336	122.7835
Q	298	8.41164	121.8204
A	299	8.35201	126.2822
N	300	7.96305	123.516

Table 7.3: Chemical shifts for C-CA$_{AA}$-NC + CAI

Residue	No.	^{1}H (ppm)	^{15}N (ppm)	2° Cα (ppm)
S	146			
P	147			
I	150			
L	151	7.393	119.71	1.27
D	152	7.493	114.622	1.049
I	153	7.273	121.988	-2.41
R	154	7.273	121.988	2.73
Q	155	7.733	127.364	-0.348
G	156	9.824	116.802	-0.134
P	157			
K	158	8.509	114.932	
E	159			
P	160			
F	161	9.467	128.483	
R	162	9.05	115.475	3.064
D	163	6.947	119.165	3.883
Y	164	7.497	123.983	4.401
V	165	8.085	119.834	4.077
D	166	7.405	118.657	4.751
R	167	7.643	117.794	3.664
F	168	8.849	125.446	3.991
Y	169	9.29	118.866	3.2
K	170	8.453	117.764	3.434
T	171	7.734	116.071	4.677
L	172	8.145	122.259	1.367
R	173	8.949	119.301	2.693
A	174	6.704	117.714	-0.256
E	175	8.273	120.994	0.413
Q	176	8.659	124.132	-1.186
A	177	7.517	125.83	-2.059

S	178	8.49	115.926	-1.299
Q	179	9.038	122.012	3.39
E	180			
V	181			
K	182	8.078	124.074	
N	183			
A	184			
A	185			
T	186			
E	187			
T	188			
L	189			
L	190			
V	191	8.206	116.254	3.076
Q	192	7.431	116.983	2.99
N	193	8.111	114.72	-0.596
A	194	7.249	124.588	-0.821
N	195	9.047	119.293	-1.798
P	196			2.817
D	197	7.685	115.457	4.18
C	198	8.539	116.505	4.986
K	199	9.49	121.84	4.77
T	200	7.551	113.354	4.61
I	201	6.758	122.603	2.797
L	202	8.552	118.89	2.138
K	203	8.496	119.487	2.366
A	204	7.122	120.559	0.915
L	205	7.224	127.805	0.802
G	206			-0.868
P	207			
G	208	8.696	109.518	-0.467
A	209	6.884	122.808	-0.42
T	210	8.393	112.873	-2.133
L	211	9.176	122.346	2.405
E	212	8.706	117.892	3.988
E	213	7.609	119.754	4.121
M	214	7.527	122.682	
M	215	8.671	116.897	2.839
T	216	8.272	116.142	4.21
A	217	7.976	122.904	1.983
C	218	7.358	110.638	1.715
Q	219	7.412	122.741	2.055
G	220	8.71	110.768	-0.134
V	221	7.308	120.34	0.806
G	222	8.745	115.912	-0.267
G	223	7.947	108.15	-0.801
P	224			
G	225	8.497	108.999	-0.267
H	226	8.263	118.648	0.482
K	227	8.396	122.99	
A	228	8.263	125	-0.153
R	229	8.239	120.866	-0.074
V	230	8.091	121.669	-0.395
L	231	8.215	125.789	0.001
A	232	8.158	124.178	0.114
E	233	8.228	119.456	0.783
A	234	8.124	124.111	0.114
M	235			0.035
S	236	8.1	116.208	-0.061
Q	237	8.235	121.669	-0.415

V	238	8.018	120.609	-0.196
T	239	8.113	117.756	-0.063
N	240	8.348	122.403	-0.132
P	241			
A	242	8.201	122.213	0.114
T	243	7.872	112.651	0.338
I	244	7.837	122.63	-0.073
M	245	8.201	123.54	-0.099
I	246	7.987	122.29	-0.274
Q	247	8.31	124.047	-0.348
K	248			
G	249			
N	250			
F	251	8.11	120.62	
R	252			
N	253	8.315	119.071	
P	254			
R	255			
K	256	8.293	122.42	-0.571
T	257	8.121	117.255	-0.33
V	258	8.303	123.054	
Q	259	8.058	124.522	-1.283
C	260	8.162	127.074	0.647
F	261	8.543	128.419	0.986
N	262	9.474	121.674	3.41
C	263	8.815	117.256	-0.088
G	264	7.81	113.365	0.734
K	265	8.152	120.928	
E	266			
G	267			
H	268	7.103	114.027	-0.319
I	269	7.766	108.174	-2.41
A	270	8.71	124.668	3.252
K	271	8.202	117.089	2.032
N	272	7.837	114.118	-1.464
C	273	7.606	124.213	3.584
R	274	8.625	128.155	-2.142
A	275	8.679	128.057	-1.822
P	276			
R	277			
K	278	8.578	126.096	-0.571
K	279	8.356	122.715	-0.371
G	280	8.093	109.08	-1.001
C	281			
W	282	8.372	129.455	0.942
K	283	9.269	122.831	2.166
C	284	8.45	117.146	0.046
G	285	8.076	113.761	0.734
K	286	8.036	120.859	-0.838
E	287	8.319	118.078	-0.218
G	288	8.565	107.475	-0.733
H	289	7.079	112.865	-0.186
Q	290	8.817	116.827	-2.151
M	291	8.841	121.927	4.308
K	292	8.478	118.458	1.899
D	293	7.792	117.543	0.612
C	294	7.528	123.658	3.985
T	295	8.082	119.621	-0.397
E	296	8.578	125.358	0.716
R	297			

Q	298			
A	299	8.311	126.161	-0.42
N	300	7.932	123.443	1.407

Table 7.4: Chemical shifts of C-CA$_{AA}$-NC + dACGCC

Residue	No.	^{1}H (ppm)	^{15}N (ppm)	2° Cα (ppm)
S	146	8.365	118.534	-1.103
P	147			
I	150	8.542	123.739	-0.044
L	151	7.511	119.353	1.277
D	152	7.606	116.03	1.177
I	153	7.371	122.197	-1.897
R	154	8.223	124.474	-2.344
Q	155	7.673	127.114	-0.253
G	156	9.791	116.732	-0.273
P	157			1.933
K	158	8.504	115.076	-2.702
E	159	6.944	125.762	-2.538
P	160			-0.313
F	161	9.468	128.208	4.856
R	162	9.024	115.752	3.422
D	163	6.884	118.972	4.04
Y	164	7.522	124.591	1.148
V	165	8.221	120.229	3.991
D	166	7.454	117.637	4.71
R	167	7.858	119.189	3.648
F	168	8.948	124.606	4.167
Y	169	8.604	116.469	5.265
K	170	8.256	121.041	3.399
T	171	7.596	117.113	4.83
L	172	7.835	122.284	2.295
R	173	7.919	116.463	1.8
A	174	7.506	121.746	0.992
E	175	7.821	117.992	1.053
Q	176	8.114	118.787	0.165
A	177	7.916	122.665	0.12
S	178	8.102	114.342	-0.488
Q	179	8.477	121.295	0.953
E	180	8.412	119.711	2.03
V	181	7.748	120.871	1.074
K	182	8.293	123.221	0.965
N	183	8.415	119.38	0.86
A	184			0.856
A	185	8.26	122.274	0.773
T	186	8.121	114.798	3.446
E	187	8.476	119.962	3.597
T	188	7.677	112.572	3.012
L	189	7.799	123.291	2.03
L	190	7.545	119.478	2.226
V	191	7.445	116.849	3.566
Q	192	7.866	117.258	2.779
N	193	7.799	115.449	-0.557
A	194	7.248	124.808	-0.918
N	195	9.007	119.428	-1.729
P	196			3.07

D	197	7.672	115.607	4.294
C	198	8.508	116.576	5.091
K	199	9.44	121.795	4.904
T	200	7.543	113.171	4.703
I	201	6.71	122.44	2.792
L	202	8.458	118.888	1.881
K	203	8.463	119.369	2.147
A	204	7.075	120.203	0.598
L	205	7.338	118.149	0.705
G	206	7.393	129.282	-0.931
P	207			-0.177
G	208	8.678	115.908	-0.165
A	209	7.079	123.42	-0.321
T	210	8.551	113.341	-1.7
L	211	8.862	122.566	2.788
E	212	8.522	117.729	4.318
E	213	7.701	120.541	3.302
M	214	8.467	120.202	4.554
M	215	8.697	117.044	3.755
T	216	8.296	116.966	4.579
A	217	7.895	122.865	2.107
C	218	7.308	110.854	1.679
Q	219	7.423	122.492	2.149
G	220	8.678	110.297	-0.049
V	221	7.293	120.477	0.952
G	222	8.72	109.761	-0.306
G	223	7.962	108.168	-0.691
P	224			0.086
G	225	8.502	109.1	-0.051
H	226	8.125	119.821	1.437
K	227	8.191	122.753	-0.061
A	228	8.171	124.38	0.431
R	229	8.172	120.774	0.176
V	230	8.069	121.469	0.344
L	231	8.178	125.447	0.09
A	232	8.12	124.013	0.356
E	233	8.223	119.484	0.995
A	234	8.106	124.017	0.242
M	235	8.134	118.322	0.044
S	236	8.089	116.186	0.084
Q	237	8.219	121.701	-0.312
V	238			-0.09
T	239	8.108	117.642	-0.047
N	240	8.345	122.428	0.023
P	241			0.05
A	242	8.197	122.289	0.208
T	243	7.866	112.724	0.507
I	244	7.833	122.706	-0.058
M	245	8.204	123.636	-0.134
I	246	7.998	122.374	-0.215
Q	247	8.317	124.16	-0.274
K	248	8.291	122.536	-0.118
G	249	8.291	109.404	-0.048
N	250	8.158	118.655	-0.27
F	251	8.104	120.625	-0.068
R	252	8.136	121.772	-0.11
N	253	8.238	119.038	0.119
P	254			-0.726
R	255	8.461	123.333	-0.408
K	256	8.007	120.611	-0.401

T	257	8.118	116.568	-0.363
V	258	8.269	123.735	-1.089
Q	259	7.994	124.317	-0.476
C	260	8.279	125.934	-3.994
F	261	8.453	128.664	-1.122
N	262	9.978	121.669	3.423
C	263	8.765	117.192	0.164
G	264	7.812	113.077	1.012
K	265	8.256	121.335	-0.671
E	266	8.319	118.301	0.043
G	267	8.689	107.492	0.479
H	268	7.025	113.824	-0.093
I	269	7.475	106.239	-2.49
A	270	8.986	124.959	3.71
K	271	8.914	118.481	2.055
N	272	7.819	114.324	-1.388
C	273	7.623	124.676	3.556
R	274	8.794	128.381	-0.866
A	275	8.72	128.013	-1.766
P	276			
R	277			2.635
K	278	8.909	125.3	-0.038
K	279	8.451	122.648	-0.136
G	280	8.299	108.17	-1.42
C	281	8.308	124.604	-3.725
W	282			
K	283			2.133
C	284	8.636	117.971	0.314
G	285	8.102	113.85	0.891
K	286	8.391	122.447	0.235
E	287			0.76
G	288	8.708	107.856	0.23
H	289	6.978	112.367	0.211
Q	290	8.734	117.564	-2.131
M	291	9.267	121.691	4.933
K	292	8.914	120.035	2.043
D	293	7.839	117.559	0.558
C	294	7.534	123.872	4.047
T	295	8.107	119.875	-0.312
E	296	8.625	125.421	0.89
R	297			-0.515
Q	298	8.359	121.72	-0.468
A	299	8.316	126.253	-0.353
N	300	7.928	123.525	1.44

Table 7.5: Chemical shifts of C-CA$_{AA}$-NC + dSL3

	No.	^{1}H (ppm)	^{15}N (ppm)	2° Cα (ppm)
S	146			
P	147			
I	150	8.584	123.799	-0.00544
L	151	7.506	119.145	1.27192
D	152	7.633	116.088	1.2548
I	153	7.378	122.205	-1.73275
R	154	8.227	124.413	-2.2533
Q	155	7.685	127.166	-0.17167
G	156	9.813	116.751	-0.1652
P	157			

K	158	8.51	115.032	-2.68662
E	159	6.948	125.764	-2.50745
P	160			
F	161	9.473	128.229	5.06078
R	162	9.044	115.743	3.47975
D	163	6.908	118.99	4.08875
Y	164	7.544	124.552	1.16949
V	165	8.239	120.235	3.93607
D	166	7.474	117.646	4.77765
R	167	7.881	119.185	3.68879
F	168	8.971	124.593	4.22037
Y	169	8.619	116.486	5.3561
K	170	8.273	120.98	3.42918
T	171	7.61	117.033	4.66175
L	172	7.862	122.342	2.32597
R	173	7.923	116.375	1.7372
A	174	7.508	121.718	1.05175
E	175	7.856	118.141	1.21412
Q	176	8.138	118.727	1.06616
A	177			
S	178	8.107	114.36	-0.4596
Q	179	8.485	121.286	0.91948
E	180	8.43	119.802	2.01186
V	181	7.769	120.812	1.02587
K	182	8.314	123.275	0.91227
N	183	8.435	119.432	0.83587
A	184	8.109	123.963	
A	185	8.268	122.201	0.81373
T	186	8.134	114.76	3.37908
E	187	8.527	120.076	3.57117
T	188	7.687	112.672	3.05006
L	189	7.81	123.336	2.073
L	190	7.544	119.451	2.19522
V	191	7.455	116.839	3.56253
Q	192	7.864	117.262	2.78756
N	193	7.817	115.457	-0.4681
A	194	7.244	124.833	-0.81781
N	195	9.03	119.428	-1.68506
P	196			
D	197	7.678	115.56	4.37081
C	198	8.516	116.571	5.08457
K	199	9.442	121.782	4.91363
T	200	7.562	113.197	4.64879
I	201	6.712	122.439	2.7956
L	202			2.12294
K	203	8.471	119.105	2.16024
A	204	7.077	120.161	0.59018
L	205	7.359	118.187	0.73953
G	206	7.437	104.595	-0.94735
P	207			
G	208	8.728	109.697	-0.12299
A	209	7.074	123.415	-0.20943
T	210	8.579	113.354	-1.67508
L	211	8.868	122.527	2.83264
E	212	8.54	117.708	4.3623
E	213	7.715	120.542	3.39184
M	214	8.496	120.187	4.58786
M	215	8.708	117.023	3.79947
T	216	8.32	117.047	4.52932
A	217	7.931	122.805	2.13967
C	218	7.309	110.827	1.69727

Q	219	7.438	122.526	2.24112
G	220	8.678	110.149	-0.06242
V	221	7.29	120.466	0.94939
G	222	8.664	116.001	-0.26837
G	223	7.963	108.171	-1.007
P	224			
G	225			
H	226			
K	227			
A	228	8.174	124.287	
R	229	8.182	120.656	0.34117
V	230	8.077	121.409	0.38074
L	231	8.178	125.335	
A	232	8.138	124.067	0.21392
E	233	8.239	119.46	1.10715
A	234	8.123	124.003	0.40189
M	235	8.143	118.336	0.12098
S	236	8.076	116.133	0.11934
Q	237	8.222	121.621	-0.27833
V	238	8.012	120.51	-0.08645
T	239	8.12	117.701	0.01779
N	240	8.356	122.354	0.061
P	241			
A	242	8.21	122.268	0.2519
T	243	7.875	112.715	0.43606
I	244	7.852	122.72	-0.08075
M	245	8.225	123.731	-0.11379
I	246	8.011	122.45	-0.22276
Q	247	8.331	124.269	-0.2596
K	248	8.321	122.644	-0.06876
G	249	8.3	109.448	0.08066
N	250	8.184	118.634	-0.18479
F	251	8.124	120.575	-0.04263
R	252	8.155	121.68	-0.00309
N	253	8.279	119.026	
P	254			
R	255			
K	256			
T	257	8.116	116.446	

Table 7.6: Chemical shifts of C-CA-NC + dACGCC

Residue	No.	^{1}H (ppm)	^{15}N (ppm)
S	146		
P	147		
I	150		
L	151		
D	152	7.48893	115.8643
I	153	7.37482	122.2365
R	154		
Q	155	7.74709	127.493
G	156	9.83242	116.7816
P	157		
K	158	8.52754	114.9744
E	159	6.97972	125.7287
P	160		

F	161	9.45904	128.3337
R	162	9.04452	115.7456
D	163	6.92422	119.0413
Y	164	7.46958	124.1157
V	165	8.2839	120.0117
D	166	7.52508	118.3692
R	167	7.80983	119.0077
F	168	9.07564	124.083
Y	169		
K	170	8.24629	120.5426
T	171	7.56128	117.3274
L	172		
R	173		
A	174		
E	175		
Q	176		
A	177	7.59024	125.2078
S	178		
Q	179		
E	180		
V	181		
K	182		
N	183		
W	184		
M	185		
T	186		
E	187		
T	188		
L	189		
L	190		
V	191		
Q	192		
N	193	7.86685	115.0066
A	194	7.20184	124.818
N	195	9.02485	119.3507
P	196		
D	197	7.68258	115.5603
C	198	8.50934	116.6237
K	199	9.47944	121.8712
T	200	7.55469	112.5931
I	201	6.75047	122.4354
L	202	8.4847	119.2082
K	203	8.47652	118.9987
A	204	7.14272	120.3506
L	205	7.38203	118.3746
G	206	7.45926	104.6254
P	207		
G	208		
A	209	7.18948	123.4207
T	210	8.62265	113.4483
L	211	8.80506	122.7473
E	212	8.53948	117.6987
E	213	7.74758	120.6242
M	214	8.482	120.104
M	215	8.79219	117.2337
T	216	8.35578	117.05
A	217	7.92506	122.8147
C	218	7.33116	110.8577
Q	219	7.47897	122.4733
G	220		

V	221	7.29573	120.4722
G	222	8.67207	115.9114
G	223	7.96568	108.1918
P	224		
G	225		
H	226		
K	227	8.13839	122.6197
A	228	8.17857	124.3016
R	229	8.17732	120.7206
V	230	8.08825	121.3203
L	231	8.18726	125.2933
A	232	8.15104	124.0336
E	233	8.26256	119.5348
A	234	8.12226	123.9585
M	235	8.1627	118.4143
S	236	8.10445	116.1289
Q	237	8.24864	121.6319
V	238	8.03615	120.5686
T	239	8.1327	117.7275
N	240	8.32081	122.5218
P	241		
A	242	8.22167	122.1547
T	243	7.90253	112.7023
I	244	7.86115	122.694
M	245	8.22612	123.5511
I	246	8.0214	122.3691
Q	247	8.33917	124.028
K	248	8.31234	122.4948
G	249	8.30727	109.4826
N	250		
F	251		
R	252		
N	253	8.2513	119.3633
P	254		
R	255		
K	256	8.61809	123.9891
T	257	8.16335	116.7569
V	258	8.32315	123.4972
Q	259	8.00174	124.4482
C	260		
F	261		
N	262	10.02693	121.7799
C	263	8.87462	117.2426
G	264	7.80996	113.2923
K	265	8.30135	121.2395
E	266	8.35061	118.3068
G	267	8.72829	107.4215
H	268	7.06141	113.9032
I	269	7.50969	106.1709
A	270	9.03107	125.1025
K	271	8.977	118.4132
N	272	7.87834	114.7251
C	273	7.7155	124.6404
R	274		
A	275	8.74952	128.0374
P	276		
R	277	8.47954	123.0081
K	278	8.945	125.2849
K	279	8.47726	122.5937
G	280	8.33247	108.2146

C	281	8.33576	124.4424
W	282	8.43816	129.3439
K	283		
C	284	8.63751	117.8355
G	285	8.12727	113.8424
K	286	8.39749	123.2805
E	287		
G	288		
H	289	7.0112	112.383
Q	290	8.78949	117.6646
M	291		
K	292		
D	293	7.86138	117.6096
C	294	7.55871	123.8167
T	295	8.13938	119.9467
E	296	8.6491	125.453
R	297	8.32423	122.5692
Q	298	8.39984	121.7058
A	299	8.34943	126.3016
N	300	7.95652	123.5386

7.4. Relaxation data

Table 7.7: List of T_1, T_2 and NOEs for C-CA_{AA}-NC

Res	T_1 (s)	ΔT_1 (s)	T_2 (s)	ΔT_2 (s)	NOE	ΔNOE
146	1.4035	0.0656	0.5088	0.0257	-1.25	0.07
148	0.7421	0.0125	0.1997	0.0029	0.11	0.03
149	0.6409	0.0243	0.1354	0.0046	0.39	0.05
150	0.6214	0.0192	0.0905	0.0030	0.68	0.04
151	0.6431	0.0119	0.0848	0.0014	0.78	0.03
152	0.6482	0.0125	0.0898	0.0016	0.67	0.03
153	0.6686	0.0090	0.0990	0.0013	0.75	0.02
154	0.6056	0.0143	0.1104	0.0023	0.85	0.03
155	0.5826	0.0135	0.0882	0.0015	0.86	0.02
156	0.5650	0.0178	0.0904	0.0021	0.85	0.03
158	0.8769	0.0530	0.0948	0.0034	0.79	0.03
159	0.6402	0.0125	0.0952	0.0015	0.82	0.02
161	0.6401	0.0197	0.0714	0.0018	0.94	0.03
162	0.6878	0.0097	0.0797	0.0010	0.81	0.02
163	0.6490	0.0108	0.0773	0.0012	0.87	0.02
164	0.6408	0.0176	0.0815	0.0021	0.76	0.03
165	0.6476	0.0120	0.0699	0.0012	0.93	0.02
166	0.6845	0.0083	0.0740	0.0009	0.87	0.02
167	0.6495	0.0117	0.0731	0.0011	0.90	0.02
168	0.6527	0.0120	0.0741	0.0013	0.82	0.02
169	0.6650	0.0166	0.0764	0.0017	0.69	0.02
170	0.6650	0.0140	0.1020	0.0016	0.71	0.02
171	0.6872	0.0153	0.0744	0.0015	0.76	0.03
172	0.6503	0.0163	0.0772	0.0019	0.83	0.03
173	0.6324	0.0096	0.0783	0.0010	0.71	0.02
174	0.6075	0.0080	0.0871	0.0011	0.79	0.02
175	0.6707	0.0162	0.0897	0.0018	0.65	0.03
176	0.6724	0.0121	0.1079	0.0014	0.66	0.02
177	0.6491	0.0089	0.1063	0.0011	0.75	0.01

178	0.6843	0.0107	0.1038	0.0013	0.57	0.02
179	0.7681	0.0176	0.1986	0.0034	0.43	0.02
180	0.6762	0.0094	0.1139	0.0011	0.50	0.02
181	0.6418	0.0061	0.1235	0.0012	0.54	0.01
182	0.7095	0.0087	0.2935	0.0037	0.44	0.01
183	0.7453	0.0260	0.1198	0.0027	0.60	0.02
184	0.7590	0.0272	0.1159	0.0026	0.72	0.02
185	0.7034	0.0050	0.2130	0.0013	0.28	0.01
186	0.6652	0.0121	0.0953	0.0014	0.75	0.02
187	0.6858	0.0152	0.0830	0.0015	0.69	0.02
188	0.6528	0.0127	0.0697	0.0010	0.69	0.02
189	0.6277	0.0107	0.0894	0.0014	0.74	0.02
190	0.6159	0.0098	0.0891	0.0012	0.84	0.02
191	0.6030	0.0110	0.0874	0.0015	0.79	0.02
192	0.6216	0.0075	0.0829	0.0010	0.87	0.02
193	0.6030	0.0112	0.0918	0.0015	0.79	0.02
194	0.5718	0.0072	0.0912	0.0010	0.80	0.01
195	0.6027	0.0124	0.0933	0.0014	0.84	0.02
197	0.6754	0.0072	0.1016	0.0009	0.80	0.01
198	0.6548	0.0127	0.0924	0.0017	0.82	0.02
199	0.5690	0.0065	0.0916	0.0011	0.83	0.01
200	0.6365	0.0065	0.0937	0.0010	0.86	0.01
201	0.5814	0.0074	0.0960	0.0012	0.61	0.28
203	0.5987	0.0061	0.0918	0.0009	0.80	0.02
204	0.6077	0.0054	0.0942	0.0007	0.80	0.01
205	0.6012	0.0050	0.1001	0.0008	0.77	0.01
206	0.5570	0.0062	0.1169	0.0010	0.75	0.01
208	0.7474	0.0505	0.1198	0.0055	0.53	0.03
209	0.6319	0.0044	0.1070	0.0007	0.80	0.01
210	0.7402	0.0112	0.0879	0.0011	0.79	0.02
211	0.6620	0.0086	0.0822	0.0010	0.78	0.02
212	0.6563	0.0065	0.0890	0.0009	0.78	0.02
213	0.6725	0.0099	0.0806	0.0012	0.82	0.02
214	0.6304	0.0098	0.0791	0.0013	0.94	0.02
215	0.6056	0.0085	0.0844	0.0010	0.80	0.02
216	0.6283	0.0078	0.0830	0.0011	0.78	0.02
217	0.6111	0.0065	0.0903	0.0009	0.79	0.01
218	0.6727	0.0110	0.0879	0.0011	0.84	0.02
219	0.5814	0.0046	0.0990	0.0008	0.76	0.01
220	0.8354	0.0711	0.1000	0.0050	0.78	0.04
221	0.6747	0.0073	0.1094	0.0010	0.61	0.01
222	0.6909	0.0184	0.1051	0.0017	0.77	0.02
223	0.6881	0.0064	0.1990	0.0013	0.36	0.01
225	0.9021	0.0985	0.2041	0.0105	0.25	0.04
227	0.8285	0.0501	0.2117	0.0061	0.16	0.03
228	0.7329	0.0133	0.1777	0.0025	0.27	0.02
229	0.7863	0.0116	0.2125	0.0029	0.14	0.02
230	0.6821	0.0050	0.2132	0.0016	0.08	0.01
231	0.7423	0.0063	0.1622	0.0012	0.11	0.01
233	0.7174	0.0044	0.2293	0.0012	-0.02	0.01
234	0.7375	0.0046	0.2526	0.0017	0.08	0.01
235	0.7738	0.0074	0.2383	0.0020	-0.06	0.01
236	0.9015	0.0192	0.2420	0.0036	-0.16	0.02
237	0.8464	0.0191	0.2879	0.0055	-0.22	0.02
239	0.8827	0.0161	0.2778	0.0043	-0.20	0.02
240	0.9892	0.0204	0.3823	0.0075	-1.65	0.03
242	0.6845	0.0043	0.2787	0.0022	0.14	0.01
243	0.7367	0.0075	0.2274	0.0021	0.11	0.01
244	0.6643	0.0039	0.3077	0.0020	0.52	0.19
245	0.7273	0.0090	0.1924	0.0018	-0.03	0.02

246	0.6977	0.0041	0.2817	0.0019	-0.10	0.01
247	0.7746	0.0108	0.2822	0.0035	-0.18	0.02
250	1.0536	0.0713	0.3317	0.0120	-0.10	0.04
251	0.9286	0.0317	0.3504	0.0084	-0.09	0.02
252	0.8793	0.0354	0.3117	0.0077	-0.03	0.02
255	0.8336	0.0190	0.3123	0.0050	-0.16	0.02
256	0.6945	0.0074	0.2065	0.0016	0.19	0.01
256	0.7884	0.0066	0.2205	0.0019	-0.27	0.01
257	0.9189	0.0291	0.2449	0.0053	0.17	0.03
258	0.6557	0.0086	0.1072	0.0012	0.52	0.17
259	0.5007	0.0077	0.1611	0.0020	0.49	0.02
260	0.4869	0.0111	0.1296	0.0023	0.71	0.03
261	0.4968	0.0239	0.1432	0.0055	0.76	0.07
262	0.4549	0.0110	0.1164	0.0023	0.60	0.03
263	0.5655	0.0142	0.1320	0.0025	0.59	0.03
264	0.4572	0.0112	0.1304	0.0021	0.64	0.03
265	0.5199	0.0086	0.1232	0.0016	0.66	0.02
266	0.5214	0.0066	0.1402	0.0014	0.57	0.02
267	0.5251	0.0107	0.1526	0.0021	0.57	0.02
268	0.4931	0.0048	0.1361	0.0009	0.61	0.01
269	0.5248	0.0079	0.1352	0.0018	0.59	0.02
270	0.4902	0.0092	0.1259	0.0021	0.73	0.03
271	0.5329	0.0095	0.1173	0.0015	0.73	0.02
272	0.5142	0.0086	0.1401	0.0016	0.69	0.02
273	0.5144	0.0052	0.1588	0.0013	0.62	0.01
275	0.5607	0.0076	0.1399	0.0012	0.58	0.01
277	0.6149	0.0188	0.1606	0.0038	0.61	0.04
278	0.8102	0.0649	0.1536	0.0090	0.70	0.09
279	0.6547	0.0450	0.2187	0.0096	0.40	0.06
280	0.6032	0.0297	0.2063	0.0068	0.40	0.05
282	0.4856	0.0242	0.1026	0.0043	0.58	0.06
283	0.4207	0.0272	0.1551	0.0082	0.54	0.08
284	0.5619	0.0268	0.1253	0.0042	0.56	0.07
285	0.4291	0.0237	0.1231	0.0049	0.50	0.06
289	0.4667	0.0283	0.1094	0.0051	0.79	0.06
290	0.5520	0.0403	0.1257	0.0067	0.51	0.07
291	0.5010	0.0324	0.1309	0.0064	0.63	0.07
292	0.4568	0.0302	0.1237	0.0053	0.57	0.06
293	0.5220	0.0272	0.1534	0.0056	0.78	0.07
294	0.4989	0.0158	0.1582	0.0044	0.58	0.04
295	0.7562	0.0675	0.1693	0.0091	0.50	0.05
296	0.5145	0.0195	0.1842	0.0061	0.57	0.05
297	0.6859	0.0317	0.2804	0.0097	0.16	0.04
298	0.8283	0.0279	0.3245	0.0100	-0.14	0.03
299	1.2325	0.0189	0.4649	0.0069	-3.81	0.04
300	1.3324	0.0054	0.5862	0.0033	-2.69	0.01

Table 7.8: List of T_1, T_2 and NOEs for C-CA_{AA}-NC + CAI

Res	T_1 (s)	ΔT_1 (s)	T_2 (s)	ΔT_2 (s)	NOE	ΔNOE
148	0.6974	0.0093	0.1483	0.0020	0.16	0.05
149	0.6077	0.0065	0.1026	0.0009	0.46	0.03
151	0.7199	0.0298	0.0639	0.0033	0.70	0.06
152	0.7274	0.0282	0.0769	0.0031	0.76	0.05
153	0.6993	0.0209	0.0742	0.0022	0.76	0.04
155	0.6446	0.0232	0.0850	0.0033	0.78	0.05

156	0.6833	0.0409	0.0882	0.0050	0.76	0.05
158	0.7654	0.0229	0.0674	0.0017	0.80	0.04
161	0.6859	0.0296	0.0659	0.0032	0.75	0.05
162	0.7565	0.0201	0.0596	0.0016	0.80	0.04
163	0.7208	0.0272	0.0709	0.0023	0.88	0.05
164	0.7257	0.0396	0.0731	0.0032	0.74	0.05
165	0.6932	0.0270	0.0540	0.0018	0.76	0.04
166	0.7063	0.0246	0.0589	0.0020	0.72	0.04
167	0.7279	0.0290	0.0677	0.0024	0.79	0.05
168	0.6718	0.0298	0.0646	0.0029	0.89	0.05
169	0.6685	0.0539	0.0586	0.0043	0.76	0.07
170	0.7862	0.0486	0.0519	0.0035	0.88	0.07
171	0.7670	0.0329	0.0609	0.0028	0.77	0.05
172	0.8296	0.0587	0.1182	0.0054	0.92	0.08
173	0.7738	0.0587	0.0562	0.0039	0.77	0.08
174	0.7999	0.0465	0.0540	0.0034	0.71	0.10
176	0.7656	0.0571	0.0361	0.0036	0.56	0.08
177	0.8815	0.0322	0.0773	0.0027	0.57	0.06
178	0.8072	0.0365	0.0769	0.0032	0.76	0.06
179	0.7389	0.0400	0.0754	0.0036	0.71	0.06
182	0.4950	0.0129	0.1118	0.0022	0.45	0.07
191	0.7574	0.0478	0.0696	0.0045	0.82	0.07
192	0.7195	0.0321	0.0676	0.0028	0.77	0.06
193	0.7100	0.0276	0.0696	0.0023	0.76	0.05
194	0.6038	0.0143	0.0702	0.0013	0.78	0.04
195	0.6305	0.0253	0.0729	0.0029	0.73	0.05
197	0.6859	0.0131	0.0781	0.0014	0.73	0.04
198	0.6710	0.0196	0.0769	0.0023	0.76	0.05
199	0.6566	0.0178	0.0758	0.0017	0.79	0.03
200	0.6321	0.0122	0.0760	0.0015	0.60	0.03
201	0.6528	0.0196	0.0772	0.0018	0.82	0.06
202	0.6431	0.0242	0.0718	0.0021	0.82	0.05
203	0.6494	0.0147	0.0680	0.0013	0.82	0.04
204	0.6551	0.0111	0.0752	0.0013	0.85	0.04
205	0.6712	0.0160	0.0771	0.0017	0.72	0.04
206	0.6536	0.0162	0.0948	0.0021	0.67	0.04
208	0.7222	0.0201	0.0758	0.0015	0.74	0.04
209	0.6627	0.0087	0.0852	0.0012	0.89	0.04
210	0.7791	0.0256	0.0662	0.0022	0.71	0.04
211	0.7653	0.0290	0.0719	0.0029	0.86	0.05
212	0.7537	0.0173	0.0662	0.0015	0.72	0.04
213	0.7435	0.0241	0.0699	0.0023	0.93	0.05
215	0.6795	0.0267	0.0701	0.0028	0.75	0.05
214	0.6954	0.0226	0.0693	0.0020	0.87	0.09
216	0.7127	0.0238	0.0653	0.0020	0.66	0.04
217	0.7140	0.0165	0.0658	0.0015	0.76	0.04
218	0.7586	0.0254	0.0803	0.0023	0.78	0.05
219	0.6598	0.0093	0.0802	0.0012	0.69	0.03
220	0.6534	0.0228	0.0755	0.0025	0.59	0.04
221	0.7175	0.0100	0.0917	0.0013	0.54	0.03
223	0.6319	0.0048	0.1778	0.0011	0.26	0.03
225	0.7004	0.0101	0.1743	0.0023	0.34	0.04
226	0.7038	0.0219	0.1614	0.0042	0.18	0.05
227	0.5679	0.0079	0.1210	0.0015	0.45	0.03
228	0.6346	0.0047	0.1606	0.0011	0.22	0.03
229	0.6473	0.0056	0.1872	0.0016	0.12	0.03
230	0.6141	0.0043	0.1760	0.0012	0.12	0.03
231	0.6045	0.0045	0.1665	0.0012	0.09	0.04
232	0.6498	0.0038	0.1802	0.0010	0.15	0.02
233	0.6740	0.0037	0.2060	0.0010	0.06	0.03

234	0.6313	0.0033	0.2324	0.0013	0.15	0.03
235	0.6610	0.0037	0.2524	0.0013	0.10	0.02
236	0.6612	0.0047	0.2779	0.0022	-0.12	0.04
237	0.6530	0.0050	0.2721	0.0023	-0.21	0.03
238	0.6710	0.0041	0.2598	0.0015	-0.32	0.03
239	0.6435	0.0049	0.2781	0.0022	-0.15	0.04
240	0.6600	0.0056	0.2954	0.0028	-0.22	0.03
242	0.6016	0.0036	0.2178	0.0012	0.27	0.03
243	0.6321	0.0039	0.2482	0.0015	0.22	0.03
244	0.6158	0.0030	0.2477	0.0014	0.22	0.03
245	0.5619	0.0038	0.2234	0.0019	0.12	0.04
246	0.5673	0.0031	0.2351	0.0015	-0.16	0.03
247	0.5789	0.0040	0.2428	0.0019	-0.19	0.03
249	0.6648	0.0054	0.2763	0.0023	-0.17	0.03
250	0.7069	0.0049	0.3187	0.0026	0.06	0.02
251	0.6638	0.0046	0.3205	0.0028	0.04	0.03
253	0.5825	0.0043	0.2766	0.0023	0.17	0.03
256	0.5976	0.0046	0.2557	0.0019	-0.10	0.02
257	0.6502	0.0064	0.2019	0.0019	0.17	0.04
258	0.6073	0.0037	0.2062	0.0011	0.10	0.02
259	0.5089	0.0072	0.1381	0.0015	0.43	0.04
261	0.4818	0.0121	0.1163	0.0021	0.66	0.06
262	0.6441	0.0145	0.1082	0.0020	0.78	0.03
263	0.5517	0.0149	0.1194	0.0025	0.52	0.05
264	0.4342	0.0094	0.1202	0.0019	0.73	0.04
265	0.5291	0.0085	0.1418	0.0019	0.43	0.04
266	0.5369	0.0081	0.1285	0.0016	0.55	0.04
267	0.4953	0.0084	0.1326	0.0018	0.60	0.04
268	0.4598	0.0047	0.1260	0.0011	0.60	0.03
269	0.4699	0.0085	0.1162	0.0015	0.67	0.04
270	0.4936	0.0102	0.1203	0.0021	0.64	0.05
271	0.4720	0.0056	0.0919	0.0008	0.70	0.04
272	0.5183	0.0074	0.1283	0.0016	0.70	0.04
273	0.4737	0.0055	0.1286	0.0012	0.58	0.03
274	0.6217	0.0159	0.1152	0.0024	0.53	0.05
275	0.5545	0.0068	0.1221	0.0010	0.56	0.03
278	0.5487	0.0131	0.1176	0.0026	0.43	0.06
279	0.5516	0.0072	0.1542	0.0019	0.21	0.04
280	0.5709	0.0133	0.1701	0.0034	0.49	0.07
282	0.4171	0.0145	0.0666	0.0020	0.55	0.07
283	0.4365	0.0200	0.1328	0.0053	0.66	0.09
284	0.4410	0.0201	0.1003	0.0032	0.47	0.08
285	0.4050	0.0153	0.0892	0.0024	0.51	0.06
287	0.5369	0.0077	0.1285	0.0014	0.55	0.04
289	0.3976	0.0187	0.0867	0.0028	0.67	0.08
290	0.5038	0.0194	0.0840	0.0026	0.55	0.06
291	0.5365	0.0136	0.0995	0.0018	0.54	0.06
292	0.4492	0.0117	0.0850	0.0016	0.46	0.05
293	0.4668	0.0136	0.1210	0.0031	0.54	0.06
294	0.4579	0.0104	0.1218	0.0023	0.56	0.06
295	0.5469	0.0153	0.1238	0.0030	0.50	0.05
296	0.4561	0.0089	0.1536	0.0021	0.50	0.05
298	0.6351	0.0068	0.3805	0.0048	-0.32	0.04
299	0.7868	0.0066	0.4637	0.0054	-0.91	0.04
300	1.1398	0.0040	0.6314	0.0037	-0.82	0.02

Table 7.9: List of T_1, T_2 and NOEs for C-CA_{AA}-NC + dACGCC

Res	T_1 (s)	ΔT_1 (s)	T_2 (s)	ΔT_2 (s)	NOE	ΔNOE
146	0.5592	0.0091	0.1208	0.0016	0.54	0.04
148	0.5439	0.0135	0.1864	0.0040	0.26	0.07
149	0.5068	0.0108	0.1274	0.0021	0.58	0.06
150	0.5746	0.0157	0.0878	0.0023	0.69	0.05
151	0.5755	0.0096	0.0786	0.0013	0.71	0.03
152	0.6093	0.0123	0.0836	0.0014	0.75	0.04
153	0.6135	0.0097	0.0879	0.0012	0.83	0.03
154	0.6126	0.0129	0.1105	0.0020	0.59	0.03
155	0.5997	0.0145	0.0869	0.0016	0.78	0.03
156	0.5334	0.0167	0.0918	0.0023	0.78	0.05
158	0.5204	0.0116	0.0828	0.0015	0.78	0.04
159	0.5762	0.0128	0.0868	0.0014	0.76	0.03
161	0.5972	0.0155	0.0718	0.0017	0.77	0.04
162	0.6345	0.0096	0.0735	0.0010	0.76	0.03
163	0.6828	0.0140	0.0791	0.0015	0.80	0.03
164	0.6145	0.0179	0.0751	0.0017	0.75	0.04
165	0.6138	0.0129	0.0801	0.0014	0.78	0.03
166	0.6447	0.0105	0.0724	0.0011	0.81	0.03
167	0.6585	0.0129	0.0717	0.0012	0.74	0.03
168	0.6226	0.0145	0.0724	0.0013	0.84	0.04
169	0.6061	0.0152	0.0704	0.0015	0.85	0.05
170	0.5079	0.0039	0.1722	0.0010	0.14	0.02
171	0.5463	0.0109	0.0751	0.0013	0.79	0.04
172	0.6567	0.0144	0.0893	0.0020	0.64	0.05
173	0.6485	0.0101	0.0746	0.0010	0.67	0.03
174	0.6287	0.0101	0.0782	0.0011	0.74	0.03
175	0.6126	0.0107	0.0844	0.0013	0.57	0.04
176	0.6071	0.0081	0.1574	0.0013	0.58	0.03
177	0.5182	0.0053	0.1091	0.0009	0.55	0.03
178	0.5210	0.0060	0.0926	0.0009	0.53	0.02
179	0.4538	0.0065	0.1137	0.0013	0.55	0.03
180	0.5326	0.0047	0.1082	0.0009	0.49	0.02
181	0.5788	0.0060	0.1079	0.0010	0.51	0.02
182	0.5373	0.0055	0.0942	0.0009	0.41	0.02
183	0.5809	0.0043	0.1001	0.0009	0.48	0.02
185	0.4795	0.0030	0.1545	0.0008	0.30	0.01
186	0.5372	0.0076	0.0851	0.0010	0.65	0.03
187	0.5783	0.0123	0.0798	0.0015	0.69	0.03
188	0.5656	0.0101	0.0633	0.0011	0.72	0.03
189	0.5832	0.0103	0.0866	0.0013	0.70	0.03
190	0.5739	0.0093	0.0826	0.0011	0.68	0.03
191	0.6229	0.0123	0.0844	0.0015	0.72	0.04
192	0.6087	0.0095	0.0798	0.0010	0.79	0.03
193	0.5721	0.0103	0.0883	0.0014	0.78	0.04
194	0.5210	0.0072	0.0856	0.0010	0.80	0.02
195	0.5312	0.0122	0.0893	0.0017	0.78	0.04
197	0.6146	0.0078	0.0953	0.0010	0.79	0.02
198	0.5491	0.0108	0.0885	0.0014	0.75	0.03
199	0.5633	0.0086	0.0873	0.0011	0.78	0.03
200	0.5459	0.0068	0.0921	0.0010	0.63	0.02
201	0.5613	0.0086	0.0938	0.0013	0.74	0.03
202	0.5529	0.0121	0.0947	0.0018	0.67	0.08
203	0.5767	0.0176	0.0989	0.0022	0.68	0.03
204	0.6054	0.0074	0.0905	0.0010	0.73	0.02
205	0.6135	0.0079	0.0989	0.0010	0.75	0.02
206	0.5707	0.0069	0.1157	0.0011	0.71	0.03
208	0.4517	0.0084	0.1060	0.0014	0.62	0.03

209	0.5902	0.0058	0.0967	0.0007	0.70	0.02
210	0.6838	0.0101	0.0832	0.0011	0.77	0.04
211	0.6019	0.0089	0.0819	0.0012	0.67	0.03
212	0.6625	0.0093	0.0831	0.0009	0.78	0.02
213	0.6029	0.0105	0.0759	0.0011	0.72	0.03
214	0.6234	0.0117	0.0799	0.0015	0.80	0.03
215	0.6409	0.0111	0.0801	0.0013	0.78	0.03
216	0.6146	0.0094	0.0794	0.0010	0.69	0.02
217	0.5863	0.0075	0.1557	0.0020	0.77	0.02
218	0.6138	0.0109	0.0836	0.0013	0.78	0.03
219	0.5515	0.0053	0.0949	0.0009	0.69	0.02
220	0.4078	0.0091	0.0881	0.0014	0.73	0.04
221	0.5665	0.0060	0.1002	0.0009	0.67	0.03
222	0.5221	0.0070	0.1029	0.0011	0.73	0.03
223	0.5032	0.0030	0.1802	0.0009	0.28	0.02
225	0.3966	0.0062	0.1519	0.0016	0.27	0.03
226	0.5068	0.0141	0.1201	0.0025	0.44	0.06
227	0.5225	0.0171	0.2344	0.0225	-0.18	0.07
228	0.5675	0.0051	0.1970	0.0015	0.10	0.01
229	0.4828	0.0216	0.2283	0.0093	0.13	0.02
230	0.5830	0.0040	0.1976	0.0012	0.12	0.02
231	0.5655	0.0041	0.1640	0.0010	0.07	0.02
232	0.5588	0.0027	0.1922	0.0008	0.06	0.01
233	0.5951	0.0035	0.2219	0.0012	-0.04	0.01
234	0.5654	0.0028	0.1902	0.0008	0.11	0.01
235	0.4661	0.0368	0.1725	0.0132	-0.02	0.02
236	0.4584	0.0036	0.1813	0.0011	-0.07	0.02
237	0.4474	0.0032	0.2193	0.0013	-0.09	0.02
239	0.4803	0.0031	0.1974	0.0010	-0.10	0.02
240	0.4090	0.0033	0.1883	0.0011	-0.03	0.03
242	0.5127	0.0027	0.2148	0.0009	0.08	0.02
243	0.4896	0.0032	0.1949	0.0011	0.08	0.02
244	0.5535	0.0031	0.2417	0.0013	0.10	0.02
245	0.5231	0.0039	0.1695	0.0010	0.04	0.02
246	0.5983	0.0032	0.2364	0.0011	0.00	0.02
247	0.4971	0.0038	0.2112	0.0013	-0.01	0.02
248	0.4421	0.0029	0.2127	0.0011	0.07	0.02
249	0.3545	0.0032	0.2153	0.0015	-0.02	0.02
250	0.4880	0.0022	0.1924	0.0006	0.02	0.01
251	0.3633	0.0026	0.1996	0.0011	0.09	0.02
252	0.4289	0.0034	0.1926	0.0010	0.12	0.02
253	0.3357	0.0038	0.1667	0.0011	0.17	0.02
256	0.5818	0.0038	0.1757	0.0018	0.20	0.02
257	0.5163	0.0081	0.1353	0.0016	0.35	0.03
258	0.5519	0.0125	0.1234	0.0024	0.43	0.04
259	0.5832	0.0457	0.1230	0.0088	0.65	0.16
260	0.5548	0.0131	0.3444	0.0085	0.57	0.12
261	0.9620	1.1026	0.0197	0.0309	0.29	0.28
262	0.6381	0.1015	0.1384	0.0197	0.56	0.24
263	0.5875	0.0323	0.1023	0.0047	0.56	0.10
264	0.4979	0.0281	0.0998	0.0038	0.60	0.07
265	0.5636	0.0068	0.1576	0.0014	0.51	0.02
266	0.6120	0.0304	0.1057	0.0048	0.55	0.08
267	0.5388	0.0224	0.1198	0.0041	0.71	0.08
268	0.4438	0.0114	0.1010	0.0020	0.73	0.04
269	0.6093	0.0573	0.0672	0.0063	0.64	0.11
270	0.4845	0.0305	0.1145	0.0062	0.76	0.10
271	0.6410	0.1512	0.1078	0.0240	0.41	0.22
272	0.5320	0.0202	0.1026	0.0033	0.61	0.09
273	0.5200	0.0271	0.1035	0.0053	0.79	0.09

274	0.4461	0.0206	0.0914	0.0035	0.69	0.06
275	0.5251	0.0099	0.0992	0.0015	0.69	0.04
278	0.6334	0.0976	0.0991	0.0126	0.89	0.25
279	0.5216	0.0207	0.0911	0.0031	0.45	0.07
280	0.4685	0.0239	0.1232	0.0053	0.57	0.09
281	0.4715	0.0311	0.2659	0.0177	0.43	0.08
284	0.6100	0.0709	0.0847	0.0089	0.68	0.17
285	0.5593	0.0264	0.1003	0.0036	0.68	0.08
286	0.4576	0.0112	0.1265	0.0026	0.23	0.03
289	0.3911	0.0115	0.0863	0.0019	0.75	0.05
290	0.5444	0.0241	0.0950	0.0031	0.64	0.08
292	0.6837	0.1671	0.1008	0.0261	0.52	-0.64
293	0.5525	0.0179	0.1012	0.0025	0.62	0.06
294	0.5262	0.0182	0.1059	0.0028	0.79	0.07
295	0.5068	0.0138	0.1201	0.0025	0.54	0.06
296	0.5141	0.0121	0.1334	0.0025	0.57	0.05
298	0.4332	0.0061	0.2197	0.0023	-0.04	0.03
299	0.5467	0.0065	0.3401	0.0038	-0.50	0.02
300	1.1876	0.0066	0.6488	0.0053	-1.29	0.02

Table 7.10 : List of T_1, T_2 and NOEs for C-CA_{AA}-NC + dSL3

	T_1 (s)	ΔT_1 (s)	T_2 (s)	ΔT_2 (s)	NOE	ΔNOE
148	0.368	0.007	0.139	0.002	0.244	0.030
149	0.425	0.013	0.107	0.002	0.516	0.040
150	0.516	0.016	0.082	0.002	0.674	0.050
151	0.533	0.018	0.078	0.002	0.867	0.022
152	0.524	0.011	0.085	0.001	0.715	0.033
153	0.582	0.008	0.094	0.001	0.716	0.024
154	0.541	0.010	0.107	0.002	0.768	0.026
155	0.596	0.014	0.090	0.001	0.778	0.029
156	0.507	0.014	0.091	0.002	0.886	0.051
158	0.395	0.023	0.080	0.003	0.864	0.054
159	0.532	0.010	0.090	0.001	0.835	0.028
161	0.481	0.014	0.073	0.002	0.904	0.041
162	0.624	0.010	0.077	0.001	0.827	0.024
163	0.656	0.012	0.086	0.002	0.884	0.026
164	0.586	0.014	0.081	0.002	0.865	0.036
165	0.622	0.011	0.076	0.001	0.877	0.027
166	0.666	0.010	0.077	0.001	0.912	0.025
167	0.631	0.010	0.080	0.001	0.967	0.028
168	0.607	0.010	0.074	0.001	0.849	0.030
169	0.666	0.015	0.077	0.001	0.803	0.035
170	0.607	0.009	0.112	0.001	0.832	0.028
171	0.550	0.011	0.074	0.001	0.782	0.034
172	0.653	0.013	0.117	0.002	0.877	0.046
173	0.630	0.009	0.077	0.001	0.724	0.026
174	0.572	0.008	0.083	0.001	0.792	0.026
175	0.537	0.012	0.085	0.002	0.709	0.032
176	0.429	0.006	0.105	0.001	0.660	0.022
178	0.330	0.006	0.087	0.001	0.632	0.022
179	0.297	0.013	0.103	0.003	0.693	0.046
180	0.384	0.005	0.096	0.001	0.550	0.020
181	0.476	0.004	0.105	0.001	0.585	0.017
182	0.403	0.006	0.089	0.001	0.483	0.021
183	0.311	0.009	0.091	0.002	0.641	0.034
184	0.365	0.004	0.117	0.001	0.229	0.013

185	0.359	0.004	0.096	0.001	0.576	0.015
186	0.391	0.007	0.086	0.001	0.872	0.028
187	0.566	0.008	0.083	0.001	0.824	0.025
188	0.464	0.009	0.064	0.001	0.762	0.027
189	0.499	0.009	0.088	0.001	0.780	0.029
190	0.563	0.007	0.086	0.001	0.850	0.020
191	0.621	0.012	0.088	0.001	0.939	0.037
192	0.655	0.010	0.087	0.001	0.868	0.026
193	0.588	0.009	0.090	0.001	0.907	0.027
194	0.553	0.007	0.087	0.001	0.802	0.020
195	0.558	0.012	0.091	0.001	0.834	0.031
197	0.557	0.008	0.097	0.001	0.807	0.019
198	0.547	0.010	0.090	0.001	0.932	0.031
199	0.570	0.007	0.089	0.001	0.817	0.022
200	0.543	0.006	0.092	0.001	0.877	0.021
201	0.559	0.007	0.100	0.001	0.877	0.031
203	0.580	0.004	0.090	0.001	0.899	0.014
204	0.610	0.006	0.092	0.001	0.853	0.021
205	0.617	0.006	0.099	0.001	0.733	0.020
206	0.556	0.007	0.120	0.001	0.727	0.016
208	0.339	0.007	0.089	0.001	0.763	0.027
209	0.560	0.005	0.099	0.001	0.788	0.013
210	0.611	0.009	0.091	0.001	0.857	0.026
211	0.565	0.008	0.083	0.001	0.936	0.028
212	0.603	0.007	0.085	0.001	0.909	0.025
213	0.579	0.009	0.079	0.001	0.839	0.028
214	0.593	0.008	0.083	0.001	0.842	0.024
215	0.627	0.008	0.085	0.001	0.890	0.026
216	0.602	0.008	0.082	0.001	0.807	0.023
217	0.532	0.004	0.087	0.001	0.719	0.013
218	0.606	0.010	0.089	0.001	0.858	0.026
219	0.514	0.005	0.093	0.001	0.739	0.018
220	0.322	0.036	0.093	0.006	0.728	0.107
221	0.462	0.005	0.096	0.001	0.779	0.022
222	0.263	0.018	0.098	0.007	0.539	0.056
223	0.327	0.003	0.129	0.001	0.370	0.014
228	0.366	0.008	0.130	0.003	0.446	0.022
229	0.334	0.008	0.122	0.002	0.352	0.023
230	0.399	0.007	0.134	0.002	0.233	0.018
231	0.394	0.008	0.137	0.003	0.268	0.023
232	0.377	0.004	0.136	0.001	0.254	0.013
233	0.398	0.005	0.173	0.002	0.188	0.017
234	0.365	0.003	0.117	0.001	0.201	0.013
235	0.342	0.007	0.100	0.002	0.142	0.022
236	0.284	0.010	0.112	0.003	0.094	0.030
237	0.298	0.010	0.139	0.004	0.065	0.027
238	0.318	0.006	0.076	0.001	-0.197	0.019
239	0.250	0.008	0.093	0.003	-0.050	0.028
240	0.276	0.014	0.105	0.004	-0.085	0.040
242	0.333	0.005	0.092	0.001	0.260	0.012
243	0.298	0.006	0.085	0.002	0.257	0.016
244	0.400	0.006	0.102	0.002	0.257	0.014
245	0.331	0.007	0.073	0.002	0.196	0.019
246	0.384	0.006	0.099	0.002	-0.028	0.015
247	0.302	0.007	0.110	0.002	0.064	0.026
248	0.304	0.006	0.086	0.001	0.188	0.021
249	0.283	0.016	0.094	0.004	0.182	0.045
250	0.344	0.011	0.165	0.004	0.260	0.057
251	0.314	0.011	0.102	0.002	0.326	0.032
252	0.273	0.010	0.091	0.003	0.454	0.043
253	0.266	0.025	0.087	0.006	0.419	0.085

257	0.325	0.014	0.128	0.003	0.329	0.117
263	0.640	0.051	0.068	0.005	0.575	0.168
267	0.715	0.101	0.060	0.010	0.980	0.495
268	0.641	0.027	0.056	0.002	0.802	0.094
270	0.569	0.107	0.098	0.017	0.456	0.284
273	0.623	0.036	0.068	0.004	1.038	0.120
275	0.698	0.044	0.060	0.003	0.740	0.106
292	0.563	0.050	0.054	0.004	1.072	0.204
293	0.644	0.047	0.073	0.005	0.863	0.146
294	0.630	0.034	0.071	0.003	0.621	0.099
295	0.403	0.024	0.075	0.004	0.574	0.065
296	0.477	0.013	0.084	0.002	0.527	0.053
298	0.297	0.013	0.142	0.004	-0.259	0.070
299	0.317	0.006	0.188	0.003	-1.239	0.047

Table 7.11: List of T_1, T_2 and NOEs for C-CA-NC and 500 μM

Res	T_1 (s)	ΔT_1 (s)	T_2 (s)	ΔT_2 (s)	NOE	ΔNOE
148	0.8597	0.0163	0.0886	0.0014	0.13	0.02
152	1.4816	0.2586	0.0523	0.0070	0.61	0.12
153	2.0801	1.7666	0.0257	0.0195	0.64	0.17
155	1.8333	0.4034	0.0360	0.0070	0.72	0.10
156	1.0833	0.1863	0.0394	0.0090	0.85	0.16
158	1.2603	0.3781	0.0303	0.0106	0.75	0.11
159	1.5385	0.1432	0.0367	0.0028	0.67	0.06
161	1.7556	0.2852	0.0377	0.0066	0.91	0.11
162	1.7259	0.1058	0.0341	0.0018	0.82	0.05
163	1.3469	0.0889	0.0373	0.0024	0.72	0.06
164	3.4185	3.1198	0.0014	0.0031	0.67	0.11
165	1.5832	0.1931	0.0323	0.0053	0.66	0.07
166	1.5138	0.1027	0.0300	0.0017	0.82	0.06
167	1.1027	0.1176	0.0319	0.0037	0.70	0.09
168	1.5574	0.3056	0.0170	0.0040	0.95	0.14
169	2.1830	1.9809	0.0428	0.0186	0.79	0.19
170	0.6904	0.0170	0.1755	0.0047	0.26	0.02
171	1.6925	0.2209	0.0293	0.0036	0.86	0.09
177	1.4891	0.1609	0.0583	0.0043	0.18	0.07
193	1.6372	0.2680	0.0522	0.0073	0.89	0.09
194	1.5095	0.0889	0.0352	0.0024	1.02	0.07
195	1.8837	0.3177	0.0328	0.0043	0.77	0.09
197	1.3930	0.0559	0.0367	0.0014	0.72	0.05
198	1.4326	0.1223	0.0321	0.0027	0.76	0.06
199	0.6536	0.0060	0.0914	0.0007	0.73	0.02
200	1.3955	0.0619	0.0413	0.0016	0.75	0.05
201	1.5716	0.1272	0.0338	0.0027	0.71	0.08
202	1.0162	0.0376	0.0641	0.0019	0.68	0.03
203	0.5752	0.0036	0.1270	0.0008	0.61	0.01
204	1.3720	0.0385	0.0336	0.0011	0.75	0.04
205	1.4123	0.0481	0.0369	0.0011	0.67	0.04
206	1.0173	0.0207	0.0531	0.0008	0.69	0.03
209	1.5569	0.0452	0.0367	0.0009	0.64	0.03
210	1.5941	0.1505	0.0370	0.0029	0.67	0.07
211	1.4905	0.1439	0.0438	0.0037	0.80	0.07
212	1.4488	0.0773	0.0322	0.0017	0.70	0.05
213	1.6788	0.1264	0.0300	0.0022	0.92	0.07

214	1.3539	0.1306	0.0280	0.0028	0.66	0.06
215	1.5094	0.1364	0.0339	0.0025	0.85	0.06
216	0.8976	0.0310	0.0651	0.0019	0.72	0.03
217	1.1930	0.0484	0.0844	0.0031	0.57	0.03
218	1.5121	0.0922	0.0391	0.0017	0.73	0.05
219	1.3253	0.0435	0.0350	0.0010	0.54	0.03
221	1.0324	0.0279	0.0420	0.0012	0.55	0.03
222	1.0710	0.1626	0.0639	0.0077	0.40	0.07
223	0.7661	0.0062	0.1015	0.0007	0.39	0.01
227	0.8105	0.0438	0.1014	0.0048	0.29	0.03
228	0.7159	0.0069	0.1042	0.0008	0.27	0.01
229	0.7443	0.0074	0.1215	0.0012	0.31	0.01
230	0.6987	0.0038	0.1162	0.0005	0.26	0.01
231	0.7018	0.0053	0.0825	0.0005	0.21	0.01
232	0.7159	0.0027	0.1287	0.0005	0.21	0.01
233	0.7178	0.0031	0.1362	0.0005	0.21	0.01
234	0.7129	0.0028	0.1365	0.0005	0.21	0.01
235	0.7205	0.0052	0.1350	0.0010	0.13	0.01
236	0.7502	0.0124	0.1293	0.0019	0.13	0.02
237	0.7356	0.0148	0.1731	0.0032	0.12	0.02
238	0.7102	0.0049	0.0992	0.0006	0.10	0.01
239	0.7201	0.0078	0.1611	0.0017	0.10	0.01
240	0.5586	0.0031	0.1131	0.0007	0.56	0.01
242	0.6594	0.0028	0.1829	0.0008	0.25	0.01
243	0.6955	0.0037	0.1334	0.0007	0.24	0.01
244	0.6594	0.0022	0.2008	0.0007	0.20	0.01
245	0.6478	0.0048	0.1015	0.0007	0.19	0.01
246	0.6656	0.0021	0.2054	0.0008	0.18	0.01
247	0.6485	0.0061	0.1947	0.0019	0.19	0.01
248	0.6505	0.0171	0.1573	0.0035	0.20	0.02
249	0.7405	0.0374	0.2424	0.0126	0.21	0.03
253	0.6147	0.0180	0.2207	0.0076	0.27	0.02
256	0.6526	0.0140	0.2179	0.0045	0.23	0.01
257	0.6906	0.0110	0.1679	0.0025	0.31	0.01
258	0.6461	0.0029	0.1621	0.0007	0.38	0.01
259	0.5972	0.0032	0.1259	0.0006	0.59	0.01
260	0.6116	0.0046	0.1078	0.0007	0.62	0.01
261	0.5646	0.0057	0.1020	0.0010	0.74	0.02
262	0.6266	0.0141	0.1005	0.0020	0.60	0.05
263	0.6601	0.0052	0.1056	0.0008	0.72	0.01
264	0.5724	0.0041	0.1044	0.0007	0.71	0.01
265	0.6602	0.0033	0.0984	0.0004	0.68	0.01
266	0.6503	0.0032	0.1135	0.0005	0.64	0.01
267	0.5989	0.0052	0.1154	0.0009	0.67	0.01
268	0.5695	0.0019	0.1102	0.0004	0.73	0.01
269	0.5990	0.0034	0.1169	0.0007	0.71	0.01
270	0.6069	0.0049	0.1010	0.0007	0.74	0.02
271	0.6030	0.0045	0.0920	0.0007	0.71	0.01
272	0.6057	0.0033	0.1131	0.0005	0.68	0.01
273	0.5931	0.0017	0.1192	0.0003	0.70	0.01
275	0.6362	0.0033	0.1069	0.0005	0.70	0.01
277	0.6354	0.0055	0.1238	0.0010	0.60	0.01
278	0.6649	0.0111	0.1152	0.0017	0.61	0.02
279	0.6023	0.0084	0.1493	0.0019	0.49	0.02
280	0.6038	0.0054	0.1840	0.0017	0.48	0.01
281	0.5795	0.0063	0.1309	0.0015	0.61	0.02
282	0.5216	0.0049	0.1044	0.0009	0.64	0.02
283	0.5539	0.0062	0.1212	0.0012	0.71	0.02
284	0.5711	0.0038	0.1182	0.0007	0.62	0.01
285	0.5307	0.0041	0.1036	0.0008	0.65	0.01

286	0.5735	0.0041	0.1150	0.0007	0.56	0.01
289	0.5291	0.0037	0.0973	0.0006	0.67	0.01
290	0.5572	0.0057	0.1087	0.0009	0.66	0.02
291	0.5674	0.0085	0.1115	0.0015	0.68	0.02
292	0.5397	0.0057	0.1057	0.0011	0.69	0.02
293	0.5693	0.0044	0.1179	0.0008	0.66	0.01
294	0.5676	0.0024	0.1330	0.0006	0.66	0.01
295	0.5721	0.0077	0.1309	0.0017	0.53	0.02
296	0.5486	0.0033	0.1465	0.0008	0.59	0.01
297	0.6188	0.0050	0.2052	0.0017	0.34	0.01
298	0.6576	0.0198	0.3028	0.0124	0.19	0.03
299	0.8169	0.0111	0.4387	0.0093	-0.21	0.01
300	1.1000	0.0021	0.7820	0.0042	-0.89	0.01

Table 7.12: List of T_1 and T_2 for C-CA-NC at 100 μM

Res	T_1 (s)	ΔT_1 (s)	T_2 (s)	ΔT_2 (s)
148	0.7670	0.0397	0.0901	0.0007
152	4.2235	8.9586	0.0696	0.0030
153	1.2832	0.5967	0.0015	0.0010
155	1.7923	0.4380	0.0475	0.0011
156	1.8438	0.3031	0.0384	0.0007
158	2.1412	0.6712	0.0603	0.0018
159	1.2998	0.1348	0.0349	0.0005
161	1.2322	0.2334	0.0299	0.0007
162	1.6032	0.1252	0.0398	0.0003
163	1.4440	0.1292	0.0419	0.0005
164	1.9118	0.6364	0.1276	0.0046
165	1.5705	0.1805	0.0405	0.0005
166	1.3591	0.1210	0.0453	0.0006
167	1.7093	0.3939	0.0378	0.0008
168	1.5419	0.5151	0.0673	0.0031
169	1.3423	1.5861	-0.0675	0.0044
170	0.6856	0.0352	0.1740	0.0014
171	1.3588	0.2603	0.0491	0.0008
177	1.2330	0.3624	0.0590	0.0017
193	1.5083	0.2730	0.0520	0.0009
194	1.3686	0.1944	0.0689	0.0012
195	1.3498	0.1415	0.0379	0.0005
197	1.4116	0.0623	0.0397	0.0002
198	1.1483	0.0855	0.0398	0.0004
199	0.6660	0.0163	0.0836	0.0003
200	1.2733	0.0777	0.0390	0.0002
201	1.3822	0.1162	0.0407	0.0004
202	1.2594	0.0528	0.0443	0.0002
203	1.2594	0.0540	0.0443	0.0002

204	1.3129	0.0674	0.0439	0.0003
205	1.3414	0.0727	0.0479	0.0003
206	0.9980	0.0424	0.0624	0.0004
209	1.5027	0.1126	0.0484	0.0004
210	1.7613	0.1474	0.0361	0.0004
211	1.5117	0.2309	0.0772	0.0016
212	1.3075	0.0799	0.0353	0.0003
213	1.4342	0.1058	0.0455	0.0004
214	1.7000	0.1377	0.0541	0.0006
215	1.4783	0.1357	0.0353	0.0003
216	1.2736	0.0634	0.0656	0.0004
217	1.3807	0.0743	0.0548	0.0003
218	1.5254	0.1217	0.0467	0.0004
219	1.3855	0.0531	0.0404	0.0002
221	1.0921	0.0495	0.0456	0.0003
222	0.8758	0.1243	0.0725	0.0013
223	0.8223	0.0156	0.0904	0.0002
227	0.8028	0.0519	0.1033	0.0010
228	0.6833	0.0140	0.0924	0.0003
229	0.7550	0.0166	0.1116	0.0003
230	0.7006	0.0090	0.1173	0.0002
231	0.7692	0.0115	0.0694	0.0002
232	0.6883	0.0064	0.1012	0.0001
233	0.7088	0.0064	0.1396	0.0003
234	0.7025	0.0072	0.1140	0.0001
235	0.7882	0.0121	0.1100	0.0003
236	0.7294	0.0224	0.1153	0.0006
237	0.6625	0.0206	0.1358	0.0008
238	0.6731	0.0118	0.0800	0.0002
239	0.6960	0.0137	0.1296	0.0004
240	0.6010	0.0106	0.1009	0.0003
242	0.6705	0.0066	0.1352	0.0002
243	0.6802	0.0105	0.1239	0.0003
244	0.6353	0.0053	0.1641	0.0003
245	0.6243	0.0119	0.0854	0.0002
246	0.6811	0.0059	0.1835	0.0003
247	0.6415	0.0141	0.1265	0.0003
248	0.6010	0.0107	0.1009	0.0002
249	0.7109	0.0515	0.1409	0.0011
253	0.6155	0.0278	0.1829	0.0013
256	0.6909	0.0173	0.1600	0.0005
257	0.6793	0.0251	0.1488	0.0007
258	0.6938	0.0086	0.1287	0.0002
259	0.5941	0.0084	0.1223	0.0003
260	0.6181	0.0118	0.0941	0.0002
261	0.5596	0.0165	0.1008	0.0004

262	0.7460	0.0622	0.0876	0.0009
263	0.6581	0.0144	0.0980	0.0003
264	0.5523	0.0093	0.1007	0.0002
265	0.6811	0.0090	0.1008	0.0002
266	0.6737	0.0080	0.1106	0.0002
267	0.6159	0.0113	0.1119	0.0003
268	0.5835	0.0071	0.1218	0.0002
269	0.6283	0.0088	0.1084	0.0002
270	0.5867	0.0150	0.0951	0.0003
271	0.6010	0.0143	0.0887	0.0002
272	0.6300	0.0087	0.1074	0.0002
273	0.5823	0.0070	0.1053	0.0002
275	0.6780	0.0118	0.0958	0.0002
277	0.6228	0.0135	0.1112	0.0004
278	0.6757	0.0313	0.0998	0.0006
279	0.6428	0.0247	0.1353	0.0006
280	0.6162	0.0128	0.1703	0.0006
281	0.6064	0.0078	0.1368	0.0002
282	0.4889	0.0110	0.0843	0.0003
283	0.5057	0.0143	0.1050	0.0004
284	0.5520	0.0101	0.1084	0.0002
285	0.5218	0.0108	0.0914	0.0002
286	0.6010	0.0099	0.1009	0.0002
289	0.5025	0.0146	0.0821	0.0003
290	0.5803	0.0191	0.0855	0.0003
291	0.4899	0.0202	0.1053	0.0005
292	0.5463	0.0150	0.0951	0.0003
293	0.5347	0.0086	0.0997	0.0002
294	0.5725	0.0087	0.1200	0.0002
295	0.5698	0.0179	0.1107	0.0005
296	0.5585	0.0110	0.1440	0.0004
297	0.5850	0.0102	0.1295	0.0003
298	0.6521	0.0301	0.2179	0.0018
299	0.7636	0.0194	0.1858	0.0007
300	1.0934	0.0069	0.3816	0.0006

7.5. Model-free fitted parameters

Table 7.13: List of C-CA$_{AA}$-NC model-free results.

	Res	S^2_s	ΔS^2_s	S^2_f	ΔS^2_s	τ_e	$\Delta\tau_e$	R_{ex}	ΔR_{ex}	$\chi 2$	$\chi 2(95\%)$
T	148	0.362	0.011	0.753	0.011	823	30			0.000	6.194
S	149	0.579	0.035	0.870	0.026	878	113			0.000	3.267
I	150	0.909	0.032			615	346			1.909	4.011
L	151	0.843	0.020	0.942	0.015	2256	797			0.000	3.452

D	152	0.860	0.022	0.950	0.014	756	199			0.000	3.503
I	153	0.851	0.008			29	8			0.428	6.535
Q	155	0.976	0.014							3.969	5.900
G	156	0.969	0.018							2.504	5.713
K	158	0.720	0.044					1.19	0.68	1.919	5.202
E	159	0.895	0.011							0.181	6.285
R	162	0.941	0.009							1.895	5.839
D	163	0.985	0.017							5.393	6.048
Y	164	0.895	0.022			1552	351			2.483	4.701
F	168	0.990	0.018					0.35	0.34	0.256	5.776
Y	169	0.949	0.010			275	326			2.145	6.300
K	170	0.718	0.019	0.864	0.014	1677	253			0.000	3.329
T	171	0.954	0.021					0.57	0.39	5.474	4.769
L	172	0.983	0.017							2.828	5.270
R	173	0.961	0.010			358	397			3.744	7.048
A	174	0.825	0.017	0.949	0.011	3277	1306			0.000	4.349
E	175	0.784	0.023	0.931	0.017	1159	188			0.000	3.083
Q	176	0.796	0.009			50	7			2.886	8.687
A	177	0.752	0.013	0.852	0.009	1871	295			0.000	4.814
S	178	0.801	0.008			79	7			0.234	8.967
Q	179	0.333	0.013	0.661	0.012	1325	53			0.000	6.263
E	180	0.602	0.011	0.867	0.009	1110	47			0.000	7.930
V	181	0.569	0.009	0.854	0.007	1298	48			0.000	7.960
N	183	0.609	0.024	0.781	0.017	1195	142			0.000	3.480
A	184	0.608	0.024	0.751	0.017	1942	327			0.000	2.498
T	186	0.868	0.010			34	9			0.139	6.220
E	187	0.912	0.013			99	26			4.039	5.697
T	188	0.900	0.018			89	23	3.08	0.31	0.000	6.826
L	189	0.858	0.019	0.949	0.013	1348	275			0.000	3.496
L	190	0.944	0.010							1.987	9.285
V	191	0.961	0.012							2.561	6.663
Q	192	0.972	0.008							5.753	7.346
N	193	0.936	0.012							4.438	7.278
A	194	0.965	0.008							5.646	6.673
N	195	0.931	0.012							0.734	6.548
D	197	0.844	0.006							3.586	7.126
C	198	0.899	0.012							2.083	6.485
I	201	0.885	0.014			438	149			0.011	6.412
K	203	0.942	0.006							3.048	5.837
A	204	0.923	0.006							1.216	6.596
L	205	0.814	0.011	0.908	0.008	2131	425			0.000	6.056
G	206	0.655	0.012	0.876	0.008	3019	450			0.000	5.736
G	208	0.706	0.027			56	10			0.035	7.980
A	209	0.771	0.009	0.849	0.005	3597	1275			0.000	7.729
T	210	0.864	0.013							3.542	7.074
L	211	0.909	0.013			32	12	0.79	0.21	0.000	7.635
E	212	0.872	0.009					1.00	0.15	6.214	7.549
E	213	0.919	0.014					0.75	0.25	0.011	7.270
M	215	0.976	0.009							1.471	6.144
T	216	0.928	0.012					0.95	0.22	6.182	7.253
A	217	0.829	0.012	0.934	0.008	3090	923			0.000	4.689
C	218	0.875	0.014					0.81	0.22	1.025	6.896
Q	219	0.780	0.009	0.928	0.007	2360	279			0.000	8.891
G	220	0.791	0.027			10	0			4.255	1.199
V	221	0.750	0.011	0.865	0.007	724	80			0.000	0.000
G	222	0.791	0.022	0.832	0.014	1249	609			0.000	0.577
G	223	0.331	0.005	0.731	0.005	1249	21			0.000	0.000
G	225	0.365	0.054	0.629	0.048	840	147			0.000	0.000
K	227	0.336	0.030	0.676	0.030	863	67			0.000	0.000
A	228	0.408	0.012	0.753	0.011	935	33			0.000	0.000

R	229	0.326	0.009	0.706	0.009	885	22			0.000	0.000
V	230	0.301	0.005	0.795	0.006	928	14			0.000	0.000
L	231	0.467	0.007	0.803	0.006	658	18			0.000	0.000
E	233	0.277	0.004	0.776	0.005	852	11			0.000	0.000
A	234	0.239	0.003	0.720	0.004	975	10			0.000	0.000
M	235	0.273	0.005	0.733	0.006	805	13			0.000	0.000
S	236	0.290	0.011	0.668	0.012	653	22			0.000	0.000
Q	237	0.214	0.010	0.685	0.013	749	19			0.000	0.000
T	239	0.233	0.009	0.665	0.010	722	17			0.000	0.000
A	242	0.186	0.004	0.729	0.005	1122	12			0.000	0.000
T	243	0.284	0.006	0.733	0.006	944	15			0.000	0.000
M	245	0.366	0.008	0.805	0.008	727	17			0.000	0.000
I	246	0.190	0.003	0.775	0.004	900	9			0.000	0.000
Q	247	0.208	0.007	0.729	0.009	806	13			0.000	0.000
N	250	0.194	0.023	0.541	0.029	789	49			0.000	0.000
F	251	0.162	0.012	0.588	0.017	884	24			0.000	0.000
R	252	0.190	0.015	0.620	0.020	897	30			0.000	0.000
R	255	0.183	0.009	0.670	0.013	830	16			0.000	0.000
K	256	0.340	0.005	0.700	0.005	908	16			0.000	0.000
T	257	0.397	0.013			48	3	0.62	0.14	0.000	7.408
V	258	0.661	0.009					3.76	0.13	2.558	6.763
Q	259	0.731	0.023	0.913	0.013	520	134			0.000	3.940
C	260	0.895	0.013			72	30			2.218	5.643
F	261	0.708	0.080	0.868	0.046	3000	4355			0.006	2.516
N	262	0.737	0.075			1301	286	1.29	0.43	0.000	1.168
C	263	0.805	0.013			82	16			6.157	7.162
G	264	0.791	0.074			1267	359			0.000	3.122
K	265	0.875	0.009			94	20			4.025	7.861
E	266	0.753	0.018	0.919	0.011	713	107			0.000	4.505
G	267	0.782	0.009			90	11			2.934	7.596
H	268	0.854	0.005			124	12			5.668	7.549
I	269	0.785	0.013			82	10	0.67	0.14	0.000	8.218
A	270	0.862	0.016					0.92	0.18	4.013	5.952
K	271	0.841	0.015			35	10	0.92	0.17	0.000	6.911
N	272	0.823	0.008			53	11			1.722	7.282
C	273	0.773	0.005			68	6			2.049	8.092
A	275	0.719	0.010			59	5	1.14	0.10	0.000	7.871
R	277	0.691	0.013			43	9			0.888	7.418
K	278	0.554	0.044					1.62	0.55	1.186	5.028
K	279	0.549	0.021			54	9			1.506	7.757
G	280	0.606	0.017			72	12			0.101	7.735
W	282	0.853	0.042			165	107	2.33	0.55	0.000	2.867
K	283	0.607	0.076			1604	479			0.357	3.711
C	284	0.763	0.039			82	30	1.03	0.44	0.000	5.045
G	285	0.426	0.172			1711	326	2.46	0.94	0.000	0.306
H	289	0.895	0.054					1.98	0.61	0.003	3.861
Q	290	0.708	0.053			78	29	2.12	0.60	0.000	4.808
M	291	0.896	0.035							6.531	6.741
K	292	0.899	0.041			329	1029			3.332	4.389
D	293	0.809	0.024							0.074	5.787
C	294	0.787	0.017			90	22			0.011	6.167
T	295	0.616	0.029			48	10			1.773	7.475
E	296	0.544	0.049	0.843	0.031	1490	323			0.000	0.813
R	297	0.308	0.034	0.718	0.029	923	90			0.000	1.695
Q	298	0.278	0.023	0.655	0.020	616	57			0.000	3.997

Table 7.14: List of C-CA$_{AA}$ + CAI model-free results

Res	S^2_s	ΔS^2_s	S^2_f	ΔS^2_s	τ_e	$\Delta\tau_e$	R_{ex}	ΔR_{ex}	$\chi 2$	$\chi 2(95\%)$

T	148	0.443	0.013	0.845	0.015	846	52			0.000	7.086
S	149	0.600	0.011	0.961	0.011	1201	61			0.000	8.349
L	151	0.933	0.035			971	530			1.442	4.193
D	152	0.929	0.026							2.010	6.203
I	153	0.956	0.021			106	86			0.103	4.939
Q	155	0.938	0.025							5.138	5.966
G	156	0.889	0.037							2.695	5.791
K	158	0.899	0.027					2.04	0.54	0.497	4.388
F	161	0.918	0.034			1726	838			1.001	3.494
R	162	0.985	0.027			894	1521	1.24	0.57	0.000	2.894
D	163	0.964	0.023							3.433	5.918
Y	164	0.893	0.049			1288	483			1.533	2.595
V	165	0.903	0.035			1918	710	3.92	0.75	0.000	0.862
D	166	0.902	0.033			1213	347	2.65	0.70	0.000	1.326
R	167	0.984	0.026							2.360	7.369
F	168	0.887	0.222			8173	27024			3.315	6.977
Y	169	0.946	0.061			2435	4304			3.554	4.181
K	170	0.959	0.060					4.33	1.59	0.674	4.889
T	171	0.980	0.042					1.19	0.99	1.198	5.159
R	173	0.950	0.072					3.34	1.66	0.585	4.409
A	174	0.922	0.054					4.44	1.43	1.538	4.949
Q	176	0.783	0.060			68	31	17.23	2.88	0.000	5.115
A	177	0.691	0.027			39	9	3.61	0.58	0.000	6.136
S	178	0.798	0.036					2.36	0.72	1.273	4.798
Q	179	0.928	0.040			638	540			1.737	4.188
K	182	0.491	0.016			1688	144			1.634	4.840
V	191	0.951	0.043							1.761	5.616
Q	192	0.997	0.030							1.212	6.857
N	193	0.993	0.025							7.325	6.541
A	194	0.917	0.032			2203	1127	1.73	0.44	0.000	2.118
N	195	0.972	0.032			395	1560			2.385	3.773
D	197	0.913	0.019			71	29	0.64	0.35	0.000	5.833
C	198	0.968	0.021							3.116	6.362
K	199	0.987	0.017							1.959	6.875
T	200	0.852	0.029			833	145	1.38	0.40	0.000	3.787
I	201	0.978	0.018							0.002	5.437
L	202	0.997	0.038					0.70	0.65	0.032	5.226
K	203	0.990	0.022					1.54	0.42	0.052	5.547
A	204	0.990	0.012							0.485	6.304
L	205	0.955	0.016			150	112			1.692	4.640
G	206	0.765	0.026	0.912	0.019	1299	262			0.000	2.617
G	208	0.923	0.019			778	275			6.427	4.875
T	210	0.952	0.022			155	135			0.256	4.316
L	211	0.936	0.026							0.728	7.350
E	213	0.964	0.022							3.870	5.753
M	214	0.923	0.031			1521	709			0.162	3.434
M	215	0.974	0.032					0.79	0.61	0.252	4.916
T	216	0.911	0.030			124	65	2.51	0.63	0.000	4.318
A	217	0.974	0.019			770	752			2.141	5.136
C	218	0.887	0.019							2.306	6.097
Q	219	0.937	0.010			172	66			0.006	6.801
G	220	0.879	0.029			738	234			2.401	2.667
V	221	0.736	0.015	0.912	0.011	878	83			0.000	0.000
G	223	0.294	0.005	0.822	0.010	1193	40			0.000	0.000
G	225	0.317	0.010	0.748	0.013	1262	68			0.000	0.000
H	226	0.358	0.019	0.809	0.021	990	54			0.000	0.000
K	227	0.484	0.013	0.938	0.015	1380	84			0.000	0.000
A	228	0.345	0.007	0.858	0.010	1090	40			0.000	0.000
R	229	0.278	0.006	0.837	0.011	1030	34			0.000	0.000
V	230	0.297	0.006	0.883	0.011	1032	35			0.000	0.000

L	231	0.321	0.007	0.917	0.012	982	34			0.000	0.000
A	232	0.295	0.004	0.834	0.008	1049	26			0.000	0.000
E	233	0.246	0.004	0.807	0.008	998	27			0.000	0.000
A	234	0.189	0.003	0.796	0.009	1153	33			0.000	0.000
M	235	0.170	0.003	0.768	0.006	1109	20			0.000	0.000
S	236	0.144	0.004	0.812	0.011	938	27			0.000	0.000
Q	237	0.150	0.004	0.847	0.010	877	21			0.000	0.000
V	238	0.169	0.003	0.862	0.009	795	18			0.000	0.000
T	239	0.141	0.004	0.838	0.011	922	26			0.000	0.000
N	240	0.128	0.004	0.830	0.010	884	19			0.000	0.000
A	242	0.201	0.004	0.800	0.010	1310	43			0.000	0.000
T	243	0.165	0.004	0.762	0.010	1261	44			0.000	0.000
I	244	0.162	0.003	0.778	0.011	1266	45			0.000	0.000
M	245	0.184	0.005	0.889	0.014	1140	43			0.000	0.000
I	246	0.174	0.004	0.961	0.011	909	22			0.000	0.000
Q	247	0.166	0.004	0.948	0.011	891	22			0.000	0.000
G	249	0.148	0.004	0.823	0.010	900	22			0.000	0.000
N	250	0.114	0.003	0.704	0.007	1109	23			0.000	0.000
F	251	0.103	0.003	0.745	0.009	1103	27			0.000	0.000
N	253	0.119	0.003	0.810	0.010	1241	33			0.000	0.000
K	256	0.154	0.003	0.889	0.009	957	17			0.000	0.000
T	257	0.562	0.005			102	6			0.060	7.203
V	258	0.469	0.007	0.858	0.006	623	20			0.000	7.973
Q	259	0.749	0.014			143	21	1.04	0.13	0.000	6.929
F	261	0.867	0.024			959	346			0.048	3.912
N	262	0.744	0.017					2.12	0.23	0.671	4.968
C	263	0.841	0.015			155	48			3.284	6.807
G	264	0.854	0.035			2980	1464			0.677	4.491
K	265	0.731	0.023	0.947	0.016	512	107			0.000	4.405
E	266	0.816	0.009			131	25			0.559	7.059
G	267	0.843	0.010			149	36			2.474	7.286
H	268	0.776	0.011			1419	132			2.980	5.821
I	269	0.885	0.019			1065	328			0.696	5.333
A	270	0.879	0.021			136	54	0.38	0.23	0.000	4.716
K	271	0.853	0.047			1420	224	2.31	0.33	0.000	3.406
N	272	0.780	0.024	0.911	0.017	1357	384			0.000	3.193
C	273	0.856	0.097			508	595			0.070	4.057
R	274	0.646	0.018			51	9	3.22	0.23	0.000	6.826
A	275	0.743	0.011			73	9	1.74	0.11	0.000	7.814
K	278	0.772	0.022			157	35	1.15	0.27	0.000	5.518
K	279	0.648	0.011			159	15	0.99	0.11	0.000	7.183
G	280	0.697	0.013			75	18			0.094	7.480
W	282	0.486	0.152			1881	300	9.19	0.88	0.000	0.380
K	283	0.724	0.059			1950	708			0.032	3.188
C	284	0.444	0.151			1570	249	4.16	0.93	0.000	0.570
G	285	0.290	0.141			2057	247	6.08	0.85	0.000	0.622
E	287	0.810	0.027	0.937	0.024	317	216			0.000	4.917
H	289	0.567	0.205			2742	962	5.14	0.98	0.000	0.506
Q	290	0.793	0.033			113	39	5.27	0.45	0.000	4.971
M	291	0.799	0.024			117	33	2.66	0.28	0.000	5.923
K	292	0.533	0.103			1341	187	5.83	0.59	0.000	1.264
D	293	0.913	0.024			296	426			7.984	4.834
C	294	0.753	0.117			988	373	1.16	0.65	0.000	1.771
T	295	0.778	0.024			117	27	0.86	0.29	0.000	6.283
E	296	0.594	0.085			1334	145	0.01	0.52	0.000	4.126
Q	298	0.135	0.008	0.849	0.012	832	26			0.000	6.562

Table 7.15: List of C-CA$_{AA}$-NC + dACGCC model-free results

	Res	S^2_s	ΔS^2_s	S^2_f	ΔS^2_s	τ_e	$\Delta\tau_e$	R_{ex}	ΔR_{ex}	$\chi 2$	$\chi 2(95\%)$
T	148	0.338	0.021	0.914	0.028	1174	101			0.000	4.403
S	149	0.577	0.016			1388	114			1.959	4.263
I	150	0.858	0.026			1485	376			0.260	3.964
L	151	0.817	0.029			1768	275	1.30	0.37	0.000	4.067
D	152	0.940	0.017			917	402			0.038	3.963
I	153	0.949	0.010							2.316	6.370
R	154	0.766	0.024	0.915	0.016	786	150			0.000	3.239
Q	155	0.934	0.022					0.82	0.33	1.473	4.640
G	156	0.912	0.032			2715	2165			0.039	3.929
K	158	0.812	0.036			3370	1808	1.38	0.42	0.000	3.176
E	159	0.988	0.013							3.961	6.457
F	161	0.916	0.036			1828	770	1.82	0.50	0.000	2.246
R	162	0.945	0.023			1051	343	0.95	0.31	0.000	6.578
D	163	0.951	0.013							1.732	6.738
Y	164	0.919	0.039			1299	496	1.08	0.49	0.000	2.481
V	165	0.940	0.018			1719	730			0.000	4.485
D	166	0.996	0.016					0.68	0.30	0.221	7.635
R	167	0.948	0.019			119	67	1.50	0.35	0.000	6.288
F	168	0.991	0.023					1.28	0.38	0.224	6.063
Y	169	0.922	0.045			7800	0	1.59	0.35	0.378	4.090
K	170	0.312	0.004			1113	17			5.298	7.519
T	171	0.829	0.038			4135	3642	1.87	0.37	0.000	4.041
L	172	0.838	0.027	0.953	0.019	812	218			0.000	2.429
R	173	0.935	0.016			365	262	1.05	0.25	0.000	6.158
A	174	0.950	0.015			988	383			1.275	4.892
E	175	0.810	0.031			906	109	0.82	0.37	0.000	3.808
Q	176	0.429	0.011	0.785	0.012	1805	136			0.000	7.494
A	177	0.646	0.008			1514	68			1.768	5.757
S	178	0.650	0.025			1318	60	2.14	0.25	0.000	6.663
Q	179	0.420	0.034			1869	102	1.56	0.35	0.000	5.354
E	180	0.626	0.007			1372	47			3.868	6.291
V	181	0.704	0.008			960	46			5.447	7.203
K	182	0.579	0.021			1120	36	2.49	0.22	0.000	6.900
N	183	0.417	0.027			1596	55	3.04	0.28	0.000	7.274
A	185	0.306	0.016			1306	17	0.81	0.15	0.000	7.209
T	186	0.750	0.026			1584	127	1.74	0.28	0.000	5.958
E	187	0.826	0.033			1477	219	1.33	0.40	0.000	2.965
T	188	0.887	0.031			1327	267	4.83	0.39	0.000	3.551
L	189	0.854	0.015			1515	234			0.256	4.544
L	190	0.866	0.028			1075	172	1.30	0.29	0.000	4.738
V	191	0.949	0.013			174	115			0.005	4.421
Q	192	0.997	0.010							4.135	6.922
N	193	0.931	0.017			1434	507			0.149	4.760
A	194	0.864	0.090			6099	9216	0.94	0.20	0.000	5.602
N	195	0.934	0.039			4093	6945			1.410	4.744
D	197	0.915	0.007							1.516	6.845
C	198	0.973	0.019			650	887			2.384	6.122
K	199	0.998	0.009							2.086	6.963
T	200	0.786	0.024			1248	97	0.98	0.24	0.000	8.905
I	201	0.922	0.016			890	273			0.671	3.722
L	202	0.901	0.022			944	349			0.224	5.242
K	203	0.853	0.024			1003	229			2.198	4.529
A	204	0.937	0.008			82	30			4.932	7.937
L	205	0.886	0.008			48	15			1.012	7.480
G	206	0.694	0.015	0.889	0.012	1974	327			0.000	5.167
G	208	0.520	0.043			2109	178	1.71	0.39	0.000	4.714
A	209	0.896	0.012	0.958	0.008	555	174			0.000	6.508
T	210	0.879	0.013					1.20	0.23	2.500	6.669
L	211	0.880	0.026			929	155	0.84	0.31	0.000	4.233

E	212	0.888	0.012					1.33	0.20	2.819	6.456
E	213	0.889	0.026			1318	240	1.33	0.33	0.000	4.238
M	214	0.988	0.013							2.532	7.022
M	215	0.940	0.016					0.89	0.29	2.016	5.876
T	216	0.941	0.010			302	231	0.90	0.19	0.000	6.280
A	217	0.272	0.067	0.734	0.007	5673	2051			0.000	3.594
C	218	0.977	0.012							4.084	6.562
Q	219	0.881	0.011			1121	130			0.530	6.292
G	220	1.000	0.141							1.980	5.667
V	221	0.902	0.063							3.311	4.646
G	222	0.895	0.076							3.396	4.981
G	223	0.315	0.032			1230	115			0.387	4.579
G	225	0.398	0.083			1255	277			1.983	2.878
H	226	0.780	0.134							3.355	5.507
K	227	0.188	0.400			897	387			0.013	3.152
A	228	0.294	0.043			1011	87			1.168	3.649
R	229	0.190	0.211			1174	217			0.000	3.206
V	230	0.301	0.035			961	91			2.997	3.933
L	231	0.408	0.042			865	106			0.080	3.950
A	232	0.310	0.024			938	66			1.126	4.122
E	233	0.257	0.036	0.908	0.048	912	68			0.000	0.000
A	234	0.316	0.026			974	64			2.613	3.699
M	235	0.350	0.497			867	548			0.043	3.493
T	239	0.280	0.031			964	103			3.063	4.590
A	242	0.234	0.024			1062	81			0.066	4.116
T	243	0.279	0.034			1063	94			0.571	4.395
I	244	0.200	0.032	0.899	0.049	1109	94			0.000	0.000
M	245	0.374	0.041			924	114			0.353	4.038
I	246	0.229	0.030	0.881	0.047	970	87			0.000	0.000
Q	247	0.244	0.036			1012	105			0.380	3.818
K	248	0.216	0.031			1182	112			2.713	3.891
N	250	0.293	0.020			999	59			2.865	3.794
R	252	0.269	0.033			1176	122			3.353	4.375
K	256	0.299	0.081			1505	554			0.032	5.343
T	257	0.664	0.012			802	65			0.679	6.399
V	258	0.805	0.013			199	50			1.072	5.039
Q	259	0.835	0.044							1.080	6.109
C	260	0.866	0.067			209	309	5.39	1.38	0.000	2.644
N	262	0.743	0.079							1.363	5.733
C	263	0.782	0.048			88	41	1.99	0.64	0.000	3.539
G	264	0.862	0.196			603	1330	1.48	1.33	0.000	2.045
K	265	0.514	0.012	0.860	0.011	1342	74			0.000	5.941
E	266	0.746	0.040			72	25	2.07	0.57	0.000	5.708
G	267	0.869	0.023							3.747	5.906
H	268	0.765	0.075			2727	944	1.60	0.51	0.000	2.458
I	269	0.785	0.074					7.35	1.57	2.243	3.343
A	270	0.917	0.038							6.070	5.664
K	271	0.842	0.136							3.588	5.562
N	272	0.951	0.023							3.200	5.880
C	273	0.961	0.035							0.044	5.261
R	274	0.683	0.116			2363	743	2.93	0.84	0.000	0.810
A	275	0.948	0.018			394	520			0.765	4.966
K	278	0.885	0.087							0.089	5.185
K	279	0.789	0.132			524	564	2.88	0.98	0.000	2.443
G	280	0.670	0.050			1633	420			0.345	3.824
C	281	0.132	0.051	0.862	0.061	1866	306			0.000	0.848
C	284	0.787	0.092					4.23	1.52	0.562	6.304
G	285	0.947	0.027							4.290	6.887
Q	290	0.979	0.026							4.867	6.104
K	292	0.830	0.148							0.522	7.253

D	293	0.914	0.018			239	222			0.666	4.364
C	294	0.954	0.020							2.994	5.472
T	295	0.630	0.021			1399	154			0.005	3.978
E	296	0.726	0.033	0.948	0.023	996	199			0.000	2.449

Table 7.16: List of C-CA$_{AA}$ model-free results.

	Res	S^2_s	ΔS^2_s	S^2_f	ΔS^2_s	τ_e	$\Delta\tau_e$	R_{ex}	ΔR_{ex}	$\chi 2$	$\chi 2(95\%)$
T	148	0.388	0.035	0.760	0.025	697	124			0.000	0.000
I	150	0.680	0.010					0.68	0.22	1.028	1.468
L	151	0.871	0.013			1315	184			1.292	2.733
D	152	0.915	0.011					0.36	0.16	2.807	3.096
I	153	0.889	0.005							3.801	5.023
R	154	0.694	0.020			283	68			0.000	0.000
Q	155	0.961	0.006							3.403	5.653
G	156	0.955	0.009							0.573	6.669
K	158	0.674	0.045					3.85	0.90	0.416	4.545
E	159	0.920	0.009			131	72			2.345	4.144
F	161	0.812	0.013					1.36	0.26	1.974	3.070
R	162	0.836	0.004							6.907	5.593
D	163	0.939	0.006							3.082	6.389
Y	164	0.969	0.007							0.743	5.390
V	165	0.892	0.005							0.005	3.159
R	167	0.956	0.005							0.301	5.344
F	168	0.953	0.006							6.638	4.850
Y	169	0.946	0.010					0.60	0.15	0.585	4.428
K	170	0.921	0.006							1.698	5.658
T	171	0.951	0.009					0.40	0.10	1.861	4.369
L	172	0.917	0.007							0.233	5.996
R	173	0.811	0.013			71	23	0.56	0.13	0.000	0.000
A	174	0.806	0.012			215	53			0.000	0.000
E	175	0.501	0.127			2224	377	1.58	0.82	0.000	0.000
Q	176	0.786	0.008							3.098	6.814
A	177	0.736	0.009			131	22			1.094	3.284
S	178	0.689	0.014			38	16			1.635	3.821
Q	179	0.631	0.038			54	28			3.121	3.825
E	180	0.569	0.025	0.814	0.021	1221	231			0.000	0.000
V	181	0.488	0.019	0.754	0.024	2131	622			0.000	0.008
K	182	0.672	0.011			85	13			3.036	3.675
N	183	0.619	0.038			46	22	2.46	0.59	0.000	0.000
A	184	0.646	0.032			42	20			2.582	4.546
A	185	0.672	0.032	0.825	0.022	1012	336			0.000	0.000
T	186	0.774	0.011					0.61	0.15	1.438	4.130
T	188	0.830	0.010					0.83	0.14	0.669	4.136
L	189	0.657	0.012			925	57			0.000	0.000
L	190	0.836	0.087			2206	415	0.84	0.44	0.000	0.233
V	191	0.852	0.006							7.274	6.062
Q	192	0.928	0.004							1.472	5.214
N	193	0.900	0.007							4.132	5.705
A	194	0.882	0.012			1282	221			0.171	3.476
N	195	0.888	0.017			1066	315			2.674	4.090
D	197	0.837	0.007			84	17			1.088	3.287
C	198	0.877	0.012			209	98			0.000	0.000
K	199	0.775	0.012			429	83			0.000	0.000
T	200	0.734	0.018	0.859	0.016	1957	773			0.000	0.189
I	201	0.955	0.005							3.838	5.360
K	203	0.843	0.003							9.022	5.594
A	204	0.939	0.005					0.40	0.07	0.019	3.536
L	205	0.860	0.006			181	32			0.272	2.744

G	206	0.655	0.010			543	48			0.000	0.000
G	208	0.672	0.021					1.24	0.32	1.676	3.446
A	209	0.708	0.007			542	40			0.000	0.000
T	210	0.661	0.013			227	27			0.000	0.000
L	211	0.831	0.007			28	17			10.069	3.505
E	212	0.860	0.005					0.35	0.05	1.323	3.703
E	213	0.870	0.007			49	23			1.187	3.122
M	214	0.787	0.011			206	38			0.000	0.000
M	215	0.862	0.025	0.926	0.023	2362	3150			0.000	1.849
T	216	0.739	0.008			44	10	0.58	0.10	0.000	0.000
A	217	0.583	0.056			1423	42	2.09	0.32	0.000	0.000
C	218	0.886	0.009					1.00	0.11	3.782	4.060
Q	219	0.857	0.010			3172	460			3.562	4.448

Table 7.17: List of C-CA model-free results at 1 mM.

Note that, since only models 1-3 were used on account of the absence of NOE data, the S^2_f column is excluded.

	Res	S^2_s	ΔS^2_s	τ_e	$\Delta\tau_e$	R_{ex}	ΔR_{ex}	$\chi 2$	$\chi 2(95\%)$
T	148	0.504	0.017	631	50			0.000	4.751
D	152	0.892	0.060			3.60	1.22	0.000	0.571
I	153	0.783	0.114			3.70	1.85	0.000	0.296
Q	155	0.945	0.022					0.258	4.806
G	156	0.874	0.025					0.598	4.184
K	158	0.865	0.031					0.008	3.916
E	159	0.752	0.024			4.20	0.45	0.000	2.827
F	161	0.965	0.025					3.670	3.936
R	162	0.805	0.018			2.00	0.33	0.000	4.471
D	163	0.906	0.015					0.108	4.292
Y	164	0.761	0.040			5.05	0.81	0.000	1.138
V	165	0.873	0.027			2.07	0.51	0.000	2.575
D	166	0.910	0.021			2.74	0.40	0.000	3.673
R	167	0.842	0.031			2.23	0.57	0.000	1.979
F	168	0.855	0.041			4.45	0.78	0.000	1.195
Y	169	0.703	0.083			5.10	1.95	0.000	0.224
K	170	0.958	0.021					0.167	4.020
T	171	0.861	0.032			2.83	0.65	0.000	1.670
A	177	0.883	0.040					1.746	3.882
N	193	0.988	0.022					1.230	4.521
A	194	1.000	0.014					0.439	6.245
N	195	0.952	0.040			1.26	0.68	0.000	1.854
D	197	0.778	0.016			0.99	0.29	0.000	4.451
C	198	0.894	0.015					0.372	4.776
K	199	0.931	0.013					0.733	4.366
T	200	0.842	0.012	291	87			0.000	5.250
I	201	0.920	0.013					3.238	4.536
L	202	0.903	0.011	190	85			0.000	4.304
K	203	0.881	0.013	328	118			0.000	4.282
A	204	0.923	0.008					3.884	4.558
L	205	0.876	0.007					1.787	5.461
G	206	0.542	0.009	826	51			0.000	5.579
A	209	0.829	0.011			0.82	0.21	0.000	5.631
T	210	0.789	0.019			2.09	0.35	0.000	3.365
L	211	0.846	0.017	156	64			0.000	3.782
E	212	0.955	0.011					3.374	4.809
E	213	0.901	0.021			0.95	0.40	0.000	3.155
M	214	0.916	0.015					0.516	4.216
M	215	0.873	0.016	568	213			0.000	2.606

T	216	0.901	0.011				1.483	4.591
A	217	0.927	0.010				0.345	4.304
C	218	0.831	0.016	174	59		0.000	4.520
Q	219	0.854	0.010	91	30		0.000	5.658

Table 7.18: List of C-CA-NC model-free results at 0.5 mM.

These results apply only to the linker and NC domains of C-CA-NC, since it was not possible to obtain data of sufficient quality to analyse the C-CA domain

	Res	S^2_s	ΔS^2_s	S^2_f	ΔS^2_s	τ_e	$\Delta\tau_e$	R_{ex}	ΔR_{ex}	$\chi 2$	$\chi 2(95\%)$
V	221	0.792	0.015	1.000	0.000	1007	62	1.65	0.77	0.000	0.001
G	222	0.548	0.077	0.826	0.080	948	124			0.000	0.000
G	223	0.310	0.004	0.804	0.006	1137	17			0.000	0.000
K	227	0.315	0.023	0.805	0.033	1002	39			0.000	0.000
A	228	0.297	0.004	0.872	0.007	998	12			0.000	0.000
R	229	0.248	0.004	0.798	0.007	1052	13			0.000	0.000
V	230	0.258	0.002	0.860	0.004	1006	9			0.000	0.000
L	231	0.393	0.004	0.979	0.006	906	9			0.000	0.000
A	232	0.229	0.002	0.836	0.004	965	10			0.000	0.000
E	233	0.213	0.002	0.824	0.004	964	8			0.000	0.000
A	234	0.211	0.002	0.826	0.004	970	10			0.000	0.000
M	235	0.216	0.003	0.847	0.006	895	7			0.000	0.000
S	236	0.231	0.006	0.831	0.011	891	12			0.000	0.000
Q	237	0.156	0.006	0.790	0.014	903	14			0.000	0.000
V	238	0.317	0.003	0.942	0.006	846	7			0.000	0.000
T	239	0.171	0.003	0.822	0.008	884	8			0.000	0.000
N	240	0.241	0.003	0.889	0.007	1556	37			0.000	0.000
A	242	0.135	0.001	0.809	0.003	1031	6			0.000	0.000
T	243	0.216	0.002	0.838	0.005	994	9			0.000	0.000
I	244	0.117	0.001	0.810	0.003	992	6			0.000	0.000
M	245	0.301	0.004	0.968	0.006	934	8			0.000	0.000
I	246	0.114	0.001	0.806	0.003	976	6			0.000	0.000
Q	247	0.122	0.003	0.831	0.008	976	10			0.000	0.000
K	248	0.169	0.008	0.861	0.019	973	18			0.000	0.000
G	249	0.093	0.011	0.708	0.032	1008	31			0.000	0.000
N	253	0.094	0.008	0.819	0.022	1072	22			0.000	0.000
K	256	0.101	0.005	0.795	0.015	1025	15			0.000	0.000
T	257	0.642	0.007			97	5			3.749	7.812
V	258	0.631	0.005	0.884	0.004	464	17			0.000	8.521
Q	259	0.835	0.003			123	7			1.520	8.422
C	260	0.850	0.007			110	10	0.90	0.09	0.000	7.120
F	261	0.936	0.031			488	333	0.44	0.25	0.000	5.602
N	262	0.828	0.021			102	26	1.76	0.28	0.000	4.755
C	263	0.818	0.007			40	5	1.36	0.10	0.000	7.952
G	264	0.931	0.008			467	117			3.428	6.840
K	265	0.829	0.005			57	4	1.70	0.07	0.000	7.439
E	266	0.826	0.005			76	4	0.45	0.06	0.000	8.165
G	267	0.860	0.008			85	8	0.46	0.10	0.000	8.045
H	268	0.910	0.003			87	7	0.50	0.04	0.000	7.800
I	269	0.862	0.005			63	6	0.42	0.07	0.000	8.791
A	270	0.865	0.008			48	8	1.65	0.10	0.000	7.196
K	271	0.913	0.007			111	16	1.56	0.11	0.000	7.619
N	272	0.886	0.003			101	8			0.128	7.059
C	273	0.872	0.003			77	5	0.10	0.04	0.000	7.711
A	275	0.829	0.005			50	3	1.30	0.06	0.000	7.748
R	277	0.782	0.011	0.917	0.007	499	62			0.000	6.673
K	278	0.830	0.009			88	11			5.784	6.714
K	279	0.655	0.016	0.919	0.011	707	50			0.000	4.697

G	280	0.533	0.011	0.861	0.007	892	35			0.000	6.539
C	281	0.698	0.014	0.947	0.010	1011	68			0.000	5.354
W	282	0.743	0.026			1116	71	1.89	0.19	0.000	4.815
K	283	0.865	0.018	0.965	0.011	729	170			0.000	4.043
C	284	0.830	0.007			622	47			1.770	6.715
G	285	0.715	0.020			1248	54	1.65	0.16	0.000	5.769
K	286	0.850	0.010			309	65	0.43	0.09	0.000	7.398
H	289	0.904	0.020			390	180	2.08	0.14	0.000	6.589
Q	290	0.880	0.009			129	22	1.22	0.11	0.000	6.816
M	291	0.918	0.008			228	115			0.637	6.013
K	292	0.835	0.024			1015	108	0.97	0.19	0.000	4.617
D	293	0.893	0.005			153	21			0.171	5.944
C	294	0.804	0.007	0.933	0.004	655	47			0.000	7.171
T	295	0.696	0.013			834	47			2.479	5.506
E	296	0.603	0.007	0.943	0.006	1214	43			0.000	7.137
R	297	0.370	0.007	0.867	0.006	956	17			0.000	8.739
Q	298	0.186	0.021	0.801	0.022	964	37			0.000	3.179

7.6. Residual dipolar couplings

Table 7.19: Residual dipolar couplings for C-CA$_{AA}$-NC

Residue	No.	RDC obs (Hz)	RDC calc (Hz)
S	146	-1.276	
P	147		
T	148	-1.155	-2.6363
S	149	-0.365	
I	150	2.734	4.3604
L	151	4.497	6.5729
D	152	-0.668	0.5488
I	153	1.701	1.9291
R	154	-0.365	-1.061
Q	155	-1.094	-1.6959
G	156	-4.557	-5.3081
P	157		
K	158	2.005	2.0107
E	159	-2.188	-1.411
P	160		
F	161	4.983	4.8639
R	162	7.413	6.2844
D	163	3.464	4.3756
Y	164	4.497	5.5019
V	165	6.502	6.4141
D	166	7.109	5.9349
R	167	5.53	4.4658
F	168	6.502	5.1128
Y	169	6.866	6.4312
K	170	6.38	
T	171	4.861	3.2665
L	172	6.319	6.4276
R	173	5.408	5.793
A	174	3.524	5.4639
E	175	1.944	
Q	176	1.884	1.3832
A	177	3.038	3.2159
S	178	0.122	-0.4258
Q	179	0.425	-0.8663

E	180	1.094	
V	181	-0.243	-0.1362
K	182	15.434	
N	183	-1.276	
A	184	-20.356	
A	185	0.243	
T	186	-1.033	-0.8126
E	187	1.58	2.7635
T	188	3.524	
L	189	1.337	
L	190	0.668	-0.3329
V	191	1.641	2.233
Q	192	2.309	2.4346
N	193	-2.734	-2.4607
A	194	-1.033	-1.4001
N	195	-4.557	-4.1031
P	196		
D	197	-1.155	-1.1448
C	198	-1.519	-1.0056
K	199	-2.856	-1.7552
T	200	-5.469	
I	201	-1.458	-1.1615
L	202		
K	203	-0.911	-1.3256
A	204	-1.155	-0.5251
L	205	-1.884	-0.9831
G	206	-3.403	-3.3132
P	207		
G	208	0.608	-1.1895
A	209	-4.497	-3.4791
T	210	4.132	3.7003
L	211	4.497	
E	212	2.188	
E	213	6.502	5.0231
M	214	2.127	2.3957
M	215	2.431	
T	216	2.917	2.0771
A	217	6.016	4.3937
C	218	3.524	
Q	219	-2.005	-0.1127
G	220	0.365	
V	221	2.431	
G	222	0.729	
G	223	-0.486	
P	224		
G	225	0	
H	226		
K	227	-4.74	
A	228	-0.425	
R	229	0.061	
V	230	-0.668	
L	231	-1.276	
A	232		
E	233	0.061	
A	234	3.038	
M	235	-0.668	
S	236	-0.122	
Q	237	9.236	
V	238		
T	239	-1.276	
N	240	-0.668	

P	241	
A	242	0
T	243	-0.304
I	244	-1.155
M	245	-14.887
I	246	-0.486
Q	247	-0.972
K	248	
G	249	
N	250	-1.337
F	251	-0.304
R	252	-0.061
N	253	
P	254	
R	255	-1.519
K	256	-1.033
T	257	-2.188
V	258	-2.674
Q	259	-3.22
C	260	-4.01
F	261	
N	262	1.944
C	263	2.248
G	264	-2.431
K	265	-2.431
E	266	-2.005
G	267	-2.127
H	268	-2.37
I	269	-3.281
A	270	-2.127
K	271	1.701
N	272	0.061
C	273	-3.038
R	274	
A	275	0.365
P	276	
R	277	-0.061
K	278	3.038
K	279	-4.375
G	280	-0.911
C	281	
W	282	
K	283	
C	284	
G	285	-1.701
K	286	
E	287	
G	288	
H	289	-0.547
Q	290	0.304
M	291	0.122
K	292	0.668
D	293	-0.243
C	294	-0.608
T	295	-0.122
E	296	0.243
R	297	
Q	298	-0.243
A	299	-0.243
N	300	0.851

Table 7.20: Residual dipolar couplings for C-CA_{AA}-NC + CAI

Residue	No.	RDC obs (Hz)	RDC calc (Hz)
S	146		
P	147		
T	148		
S	149	-1.884	
I	150		
L	151	7.474	6.6994
D	152	14.583	
I	153		
R	154	3.099	
Q	155	-1.823	-3.4153
G	156	-4.132	-4.0737
P	157		
K	158	0.729	-0.0386
E	159		
P	160		
F	161	5.529	5.642
R	162	8.264	7.5838
D	163	19.748	
Y	164	7.231	6.3379
V	165	21.084	
D	166	9.357	7.5986
R	167	-5.955	
F	168	6.198	6.5244
Y	169	9.904	7.0945
K	170	10.269	7.5937
T	171	6.684	6.4365
L	172	3.463	6.6114
R	173	7.048	6.33
A	174	6.015	6.9782
E	175	-1.337	
Q	176	-4.739	-2.4688
A	177	-1.398	-4.1489
S	178	-3.22	-4.5607
Q	179	2.795	5.1812
E	180		
V	181		
K	182		
N	183		
A	184		
A	185		
T	186		
E	187		
T	188		
L	189		
L	190		
V	191	4.132	3.3336
Q	192	6.076	4.9744
N	193	0.729	0.3128
A	194	-2.613	-2.3885
N	195	-4.436	-4.4641
P	196		
D	197	-3.038	-3.1069
C	198	-2.856	-1.7256
K	199	-0.911	-2.1243
T	200	-3.463	-0.4278

I	201	-1.641	-2.6598
L	202	-3.403	-2.5649
K	203	0.304	-1.1668
A	204	-1.762	-1.1389
L	205	-2.066	-2.8417
G	206		
P	207		
G	208	1.58	1.943
A	209	-7.838	-4.3601
T	210	5.469	5.7954
L	211	5.712	4.8681
E	212	3.767	2.623
E	213	5.104	5.4505
M	214	2.674	6.0992
M	215	4.01	2.6286
T	216	3.95	2.8787
A	217	6.076	6.7976
C	218	4.557	2.6902
Q	219	-1.519	-2.3147
G	220	-0.425	
V	221	3.828	
G	222	-0.122	
G	223	-1.337	
P	224		
G	225	-0.182	
H	226	-0.911	
K	227	-0.668	
A	228	-0.79	
R	229	-0.668	
V	230	-0.547	
L	231	1.823	
A	232	0.122	
E	233	-0.425	
A	234	-0.182	
M	235		
S	236	-0.608	
Q	237	-0.79	
V	238	-0.547	
T	239	-2.187	
N	240	14.218	
P	241		
A	242	0.547	
T	243	0.79	
I	244	-0.122	
M	245	-3.95	
I	246	0.547	
Q	247	1.398	
K	248		
G	249		
N	250		
F	251	-0.122	
R	252		
N	253		
P	254		
R	255		
K	256	-1.641	
T	257	-1.823	
V	258	-2.187	
Q	259	-3.342	
C	260	-2.977	
F	261	-0.911	

N	262	2.248
C	263	2.187
G	264	-2.734
K	265	
E	266	
G	267	
H	268	-1.762
I	269	-1.641
A	270	-1.701
K	271	
N	272	-0.79
C	273	-2.613
R	274	-0.182
A	275	1.215
P	276	
R	277	
K	278	
K	279	
G	280	-1.276
C	281	-1.094
W	282	
K	283	-0.243
C	284	-0.365
G	285	
K	286	0.365
E	287	-2.005
G	288	-2.37
H	289	-1.094
Q	290	-0.79
M	291	1.033
K	292	1.458
D	293	0.729
C	294	-0.486
T	295	-14.218
E	296	1.337
R	297	
Q	298	
A	299	-0.243
N	300	-0.061

Table 7.21: Residual dipolar couplings for C-CA_{AA}-NC + dACGCC

Residue	No.	RDC obs (Hz)	RDC calc (Hz)
S	146	12.684	
P	147		
T	148	-1.46	-2.9309
S	149	-0.836	
I	150	3.198	4.8973
L	151	2.669	7.2363
D	152	1.038	1.0506
I	153	1.802	2.321
R	154	-0.233	-1.1103
Q	155	0.297	-1.8062
G	156	-4.748	
P	157		
K	158	2.116	2.177
E	159	-1.707	-1.9615

P	160		
F	161	6.968	5.4456
R	162	8.589	7.0299
D	163	1.62	5.1757
Y	164	5.189	6.1474
V	165	9.592	6.8923
D	166	10.269	6.6876
R	167	6.958	5.2009
F	168	6.83	
Y	169	8.407	6.9296
K	170	8.991	3.6364
T	171	5.632	3.8019
L	172	4.994	6.9256
R	173	9.277	6.5306
A	174	4.857	6.2701
E	175	4.927	
Q	176	1.689	1.4897
A	177	2.141	3.7133
S	178	0.708	
Q	179	0.139	-0.501
E	180	2.18	4.255
V	181	0.15	0.1847
K	182	0.54	-1.8544
N	183	3.338	4.1432
A	184		
A	185	0.318	-2.114
T	186	-1.991	-0.445
E	187	-1.645	
T	188	5.722	
L	189	3.238	4.1341
L	190	-0.363	0.0193
V	191	2.805	2.8061
Q	192	4.227	
N	193	-1.559	
A	194	-0.878	-1.4697
N	195	-5.715	-4.8189
P	196		
D	197	-3.018	-1.143
C	198	-2.008	-1.0853
K	199	-4.397	-1.8382
T	200	-1.881	0.7993
I	201	-2.85	-1.1614
L	202	0.184	
K	203	0.487	-1.2285
A	204	0.951	-0.4467
L	205	-0.444	-1.0556
G	206	-3.362	-3.6779
P	207		
G	208	1.252	
A	209	-5.099	-4.0942
T	210	3.874	3.7523
L	211	7.445	
E	212	2.124	0.1001
E	213	6.64	5.7479
M	214	6.724	3.0886
M	215	2.048	-0.0438
T	216	3.832	2.6168
A	217	7.7	
C	218	5.237	
Q	219	-3.045	-0.2056
G	220	1.097	

V	221	5.604
G	222	0.196
G	223	-1.849
P	224	
G	225	0.418
H	226	1.431
K	227	-1.651
A	228	-1.692
R	229	-1.438
V	230	1.755
L	231	-1.457
A	232	3.637
E	233	-0.949
A	234	0.452
M	235	1.68
S	236	-0.413
Q	237	-0.52
V	238	-0.541
T	239	-3.884
N	240	-0.496
P	241	
A	242	-0.48
T	243	0.764
I	244	-1.478
M	245	0.941
I	246	-1.316
Q	247	-1.378
K	248	-8.753
G	249	-1.22
N	250	-3.15
F	251	0.312
R	252	5.222
N	253	-0.41
P	254	
R	255	1.008
K	256	-1.605
T	257	-3.241
V	258	-1.19
Q	259	2.665
C	260	0.001
F	261	-2.13
N	262	4.541
C	263	0.026
G	264	-1.292
K	265	6.062
E	266	-3.382
G	267	-2.359
H	268	-3.222
I	269	
A	270	-7.074
K	271	8.63
N	272	-1.387
C	273	-1.374
R	274	-0.696
A	275	3.564
P	276	
R	277	
K	278	-3.747
K	279	-2.941
G	280	-1.089
C	281	2.212

W	282	
K	283	
C	284	2.129
G	285	-0.761
K	286	-4.952
E	287	
G	288	1.007
H	289	6.713
Q	290	-1.885
M	291	0.879
K	292	0.97
D	293	0.122
C	294	3.002
T	295	6.522
E	296	-0.047
R	297	
Q	298	-1.875
A	299	0.664
N	300	-0.49

Reference List

1. Abragam, A. 1982. Citation Classic - the Principles of Nuclear Magnetism. Current Contents/Physical Chemical & Earth Sciences : 18 ISI:A1982NX54600001.

2. Accola, M. A., S. Hoglund, and H. G. Gottlinger. 1998. A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. Journal of Virology 72: 2072-2078 ISI:000071997600044.

3. Accola, M. A., B. Strack, and H. G. Gottlinger. 2000. Efficient particle production by minimal gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. Journal of Virology 74: 5395-5402 ISI:000087240700001.

4. Aiken, C., and C. H. Chen. 2005. Betulinic acid derivatives as HIV-1 antivirals. Trends in Molecular Medicine 11: 31-36 ISI:000226765300006.

5. Aldovini, A., and R. A. Young. 1990. Mutations of Rna and Protein Sequences Involved in Human-Immunodeficiency-Virus Type-1 Packaging Result in Production of Noninfectious Virus. Journal of Virology 64: 1920-1926 ISI:A1990CZ70400005.

6. Aldovini, A., and R. A. Young. 1990. Mutations of Rna and Protein Sequences Involved in Human-Immunodeficiency-Virus Type-1 Packaging Result in Production of Noninfectious Virus. Journal of Virology 64: 1920-1926 ISI:A1990CZ70400005.

7. Alfadhli, A., T. C. Dhenub, A. Still, and E. Barklis. 2005. Analysis of human immunodeficiency virus type 1 Gag dimerization-induced assembly. Journal of Virology 79: 14498-14506 ISI:000233279300004.

8. Alfadhli, A., D. Huseby, E. Kapit, D. Colman, and E. Barklis. 2007. Human immunodeficiency virus type 1 matrix protein assembles on membranes as a hexamer. Journal of Virology 81: 1472-1478 ISI:000243766800041.

9. Allain, B., M. Lapadattapolsky, C. Berlioz, and J. L. Darlix. 1994. Transactivation of the Minus-Strand Dna Transfer by Nucleocapsid Protein During Reverse Transcription of the Retroviral Genome. Embo Journal 13: 973-981 ISI:A1994MX77100028.

10. Amarasinghe, G. K., R. N. De Guzman, R. B. Turner, K. J. Chancellor, Z. R. Wu, and M. F. Summers. 2000. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Psi-RNA packaging signal. Implications for genome recognition. Journal of Molecular Biology 301: 491-511 ISI:000088831100019.

11. Amarasinghe, G. K., R. N. De Guzman, R. B. Turner, and M. F. Summers. 2000. NMR structure of stem-loop SL2 of the HIV-1 Psi RNA packaging signal reveals a novel A-U-A base-triple platform. Journal of Molecular Biology 299: 145-156 ISI:000087289400010.

12. Avilov, S. V., J. Godet, E. Piemont, and Y. Mely. 2009. Site-Specific Characterization of HIV-1 Nucleocapsid Protein Binding to Oligonucleotides with Two Binding Sites. Biochemistry 48: 2422-2430 ISI:000264272700016.

13. Bartonova, V., S. Igonet, J. Sticht, B. Glass, A. Habermann, M. C. Vaney, P. Sehr, J. Lewis, F. A. Rey, and H. G. Krausslich. 2008. Residues in the HIV-1 Capsid

Assembly Inhibitor Binding Site Are Essential for Maintaining the Assembly-competent Quaternary Structure of the Capsid Protein. Journal of Biological Chemistry 283: 32024-32033 ISI:000260760800084.

14. Bax, A. 2003. Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Science 12: 1-16 ISI:000179981200001.

15. Benjamin, J., B. K. Ganser-Pornillos, W. F. Tivol, W. I. Sundquist, and G. J. Jensen. 2005. Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography. Journal of Molecular Biology 346: 577-588 ISI:000226918800017.

16. Bennett, R. P., T. D. Nelle, and J. W. Wills. 1993. Functional Chimeras of the Rous-Sarcoma Virus and Human-Immunodeficiency-Virus Gag Proteins. Journal of Virology 67: 6487-6498 ISI:A1993MC01600018.

17. Berkhout, B., A. T. Das, and J. L. B. van Wamel. 1998. The native structure of the human immunodeficiency virus type 1 RNA genome is required for the first strand transfer of reverse transcription. Virology 249: 211-218 ISI:000076384600002.

18. Berkhout, B., R. Gorelick, M. F. Summers, Y. Mely, and J. L. Darlix. 2008. 6(th) international symposium on retroviral nucleocapsid. Retrovirology 5 ISI:000254441500001.

19. Berthet-Colominas, C., S. Monaco, A. Novelli, G. Sibai, F. Mallet, and S. Cusack. 1999. Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. Embo Journal 18: 1124-1136 ISI:000079184600004.

20. Bess, J. W., P. J. Powell, H. J. Issaq, L. J. Schumack, M. K. Grimes, L. E. Henderson, and L. O. Arthur. 1992. Tightly Bound Zinc in Human-Immunodeficiency-Virus Type-1, Human T-Cell Leukemia-Virus Type-I, and Other Retroviruses. Journal of Virology 66: 840-847 ISI:A1992GY96500026.

21. Bhattacharya, S., H. T. Zhang, A. K. Debnath, and D. Cowburn. 2008. Solution structure of a hydrocarbon stapled peptide inhibitor in complex with monomeric C-terminal domain of HIV-1 capsid. Journal of Biological Chemistry 283: 16274-16278 ISI:000256497100002.

22. Borsetti, A., A. Ohagen, and H. G. Gottlinger. 1998. The C-terminal half of the human immunodeficiency virus type 1 Gag precursor is sufficient for efficient particle assembly. Journal of Virology 72: 9313-9317 ISI:000076373700100.

23. Briggs, J. A. G., K. Gruenewald, B. Glass, F. Foerster, H. G. Kraeusslich, and S. D. Fuller. 2006. The mechanism of HIV-1 core assembly: Insights from three-dimensional reconstructions of authentic virions. Structure 14: 15-20 ISI:000234685100004.

24. Briggs, J. A. G., M. C. Johnson, M. N. Simon, S. D. Fuller, and V. M. Vogt. 2006. Cryo-electron microscopy reveals conserved and divergent features of Gag packing in immature particles of rous sarcoma virus and human immunodeficiency virus. Journal of Molecular Biology 355: 157-168 ISI:000234325600013.

25. Briggs, J. A. G., J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H. G. Krausslich. 2009. Structure and assembly of immature HIV. Proceedings of the National Academy of Sciences of the United States of America 106: 11090-11095 ISI:000267796100042.

26. Briggs, J. A. G., M. N. Simon, I. Gross, H. G. Krausslich, S. D. Fuller, V. M. Vogt, and M. C. Johnson. 2004. The stoichiometry of Gag protein in HIV-1. Nature Structural & Molecular Biology 11: 672-675 ISI:000222274100018.

27. Briggs, J. A. G., B. E. Watson, B. E. Gowen, and S. D. Fuller. 2004. Cryoelectron microscopy of mouse mammary tumor virus. Journal of Virology 78: 2606-2608 ISI:000189019300049.

28. Campbell, S., and A. Rein. 1999. In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. J. Virol. 73: 2270-2279 PM:9971810.

29. Campbell, S., and V. M. Vogt. 1995. Self-Assembly In-Vitro of Purified Ca-Nc Proteins from Rous-Sarcoma Virus and Human-Immunodeficiency-Virus Type-1. Journal of Virology 69: 6487-6497 ISI:A1995RU78400064.

30. Cavanagh, J., W. J. Fairbrother, A. G. Palmer, M. Rance, and N. J. Skelton. 2007. Protein NMR spectroscopy; principles and practice. Elsevier academic press.

31. Chen, Y., B. Wu, K. Musier-Forsyth, L. M. Mansky, and J. D. Mueller. 2009. Fluorescence Fluctuation Spectroscopy on Viral-Like Particles Reveals Variable Gag Stoichiometry. Biophysical Journal 96: 1961-1969 ISI:000266376500029.

32. Chou, J. J., S. Gaemers, B. Howder, J. M. Louis, and A. Bax. 2001. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. Journal of Biomolecular Nmr 21: 377-382 ISI:000172860700009.

33. Clavel, F., and J. M. Orenstein. 1990. A Mutant of Human-Immunodeficiency-Virus with Reduced Rna Packaging and Abnormal Particle Morphology. Journal of Virology 64: 5230-5234 ISI:A1990DZ52700081.

34. Clever, J., C. Sassetti, and T. G. Parslow. 1995. Rna Secondary Structure and Binding-Sites for Gag Gene-Products in the 5'-Packaging Signal of Human-Immunodeficiency-Virus Type-1. Journal of Virology 69: 2101-2109 ISI:A1995QL68300014.

35. Clever, J. L., and T. G. Parslow. 1997. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. Journal of Virology 71: 3407-3414 ISI:A1997WT18900006.

36. Clore, G. M., A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn. 1990. Deviations from the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins. Journal of the American Chemical Society 112: 4989-4991 ISI:A1990DH27700070.

37. Craven, R. C., A. E. Leuredupree, C. R. Erdie, C. B. Wilson, and J. W. Wills. 1993. Necessity of the Spacer Peptide Between Ca and Nc in the Rous-Sarcoma Virus Gag Protein. Journal of Virology 67: 6246-6252 ISI:A1993LX12000061.

38. Datta, S. A. K., J. E. Curtis, W. Ratcliff, P. K. Clark, R. M. Crist, J. Lebowitz, S. Krueger, and A. Rein. 2007. Conformation of the HIV-1 Gag protein in solution. Journal of Molecular Biology 365: 812-824 ISI:000243561500025.

39. De Guzman, R. N., Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, and M. F. Summers. 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 Psi-RNA recognition element. Science 279: 384-388 ISI:000071570800047.

40. Dorfman, T., J. Luban, S. P. Goff, W. A. Haseltine, and H. G. Gottlinger. 1993. Mapping of Functionally Important Residues of A Cysteine-Histidine Box in the Human-Immunodeficiency-Virus Type-1 Nucleocapsid Protein. Journal of Virology 67: 6159-6169 ISI:A1993LX12000052.

41. Dosset, P., J. C. Hus, D. Marion, and M. Blackledge. 2001. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. Journal of Biomolecular Nmr 20: 223-231 ISI:000170060800003.

42. Endres, D., M. Miyahara, P. Moisant, and A. Zlotnick. 2005. A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures. Protein Science 14: 1518-1525 ISI:000229497900014.

43. Endres, D., and A. Zlotnick. 2002. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. Biophys. J. 83: 1217-1230 PM:12124301.

44. Feng, Y. X., T. D. Copeland, L. E. Henderson, R. J. Gorelick, W. J. Bosche, J. G. Levin, and A. Rein. 1996. HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. Proceedings of the National Academy of Sciences of the United States of America 93: 7577-7581 ISI:A1996UY93000030.

45. Fuller, S. D., T. Wilk, B. E. Gowen, H. G. Krausslich, and V. M. Vogt. 1997. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. Current Biology 7: 729-738 ISI:A1997YB58300028.

46. Gamble, T. R., S. H. Yoo, F. F. Vajdos, U. K. vonSchwedler, D. K. Worthylake, H. Wang, J. P. McCutcheon, W. I. Sundquist, and C. P. Hill. 1997. Structure of the

carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science 278: 849-853 ISI:A1997YD47900043.

47. Ganser-Pornillos, B. K., A. Cheng, and M. Yeager. 2007. Structure of Full-Length HIV-1 CA: A Model for the Mature Capsid Lattice. Cell 131: 70-79 http://www.sciencedirect.com/science/article/B6WSN-4PTNRXP-F/2/d41f656383b0d8f899be6cc25a1b765b.

48. Gay, B., J. Tournier, N. Chazal, C. Carriere, and P. Boulanger. 1998. Morphopoietic determinants of HIV-1 Gag particles assembled in baculovirus-infected cells. Virology 247: 160-169 ISI:000075403000005.

49. Gelderblom, H. R. 1991. Assembly and Morphology of Hiv - Potential Effect of Structure on Viral Function. Aids 5: 617-638 ISI:A1991FR42400001.

50. Gelderblom, H. R., E. H. S. Hausmann, M. Ozel, G. Pauli, and M. A. Koch. 1987. Fine-Structure of Human-Immunodeficiency-Virus (Hiv) and Immunolocalization of Structural Proteins. Virology 156: 171-176 ISI:A1987F808200020.

51. Gentile, M., T. Adrian, A. Scheidler, M. Ewald, F. Dianzani, G. Pauli, and H. R. Gelderblom. 1994. Determination of the Size of Hiv Using Adenovirus Type-2 As An Internal Length Marker. Journal of Virological Methods 48: 43-52 ISI:A1994NX56200005.

52. Gheysen, D., E. Jacobs, F. Deforesta, C. Thiriart, M. Francotte, D. Thines, and M. Dewilde. 1989. Assembly and Release of Hiv-1 Precursor Pr55Gag Virus-Like Particles from Recombinant Baculovirus Infected Insect Cells. Cell 59: 103-112 ISI:A1989AU52300012.

53. Gitti, R. K., B. M. Lee, J. Walker, M. F. Summers, S. Yoo, and W. I. Sundquist. 1996. Structure of the amino-terminal core domain of the HIV-1 capsid protein. Science 273: 231-235 ISI:A1996UW78700039.

54. Gladnikoff, M., and I. Rousso. 2008. Directly monitoring individual retrovirus budding events using atomic force microscopy. Biophysical Journal 94: 320-326 ISI:000251615800036.

55. Gladnikoff, M., E. Shimoni, N. S. Gov, and I. Rousso. 2009. Retroviral Assembly and Budding Occur through an Actin-Driven Mechanism. Biophysical Journal 97: 2419-2428 ISI:000271454000007.

56. Gorelick, R. J., S. M. Nigida, J. W. Bess, L. O. Arthur, L. E. Henderson, and A. Rein. 1990. Noninfectious Human-Immunodeficiency-Virus Type-1 Mutants Deficient in Genomic Rna. Journal of Virology 64: 3207-3211 ISI:A1990DJ20300009.

57. Gottlinger, H. G. 2001. The HIV-1 assembly machine. Aids 15: S13-S20 ISI:000172834100003.

58. Gross, I., H. Hohenberg, C. Huckhagel, and H. G. Krausslich. 1998. N-terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles. Journal of Virology 72: 4798-4810 ISI:000073497600029.

59. Gross, I., H. Hohenberg, and H. G. Krausslich. 1997. In vitro assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. European Journal of Biochemistry 249: 592-600 ISI:A1997YC46100032.

60. Gross, I., H. Hohenberg, T. Wilk, K. Wiegers, M. Grattinger, B. Muller, S. Fuller, and H. G. Krausslich. 2000. A conformational switch controlling HIV-1 morphogenesis. Embo Journal 19: 103-113 ISI:000084668300012.

61. Guo, X. F., J. Hu, J. B. Whitney, R. S. Russell, and C. Liang. 2004. Important role for the CA-NC spacer region in the assembly of bovine immunodeficiency virus Gag protein. Journal of Virology 78: 551-560 ISI:000187957700001.

62. Harrison, G. P., and A. M. L. Lever. 1992. The Human-Immunodeficiency-Virus Type-1 Packaging Signal and Major Splice Donor Region Have A Conserved Stable Secondary Structure. Journal of Virology 66: 4144-4153 ISI:A1992HY08600018.

63. Hayashi, T., T. Shioda, Y. Iwakura, and H. Shibuta. 1992. Rna Packaging Signal of Human-Immunodeficiency-Virus Type-1. Virology 188: 590-599 ISI:A1992HU51100019.

64. Hayashi, T., Y. Ueno, and T. Okamoto. 1993. Elucidation of A Conserved Rna Stem-Loop Structure in the Packaging Signal of Human-Immunodeficiency-Virus Type-1. Febs Letters 327: 213-218 ISI:A1993LN69400018.

65. Hill, C. P., D. Worthylake, D. P. Bancroft, A. M. Christensen, and W. I. Sundquist. 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: Implications for membrane association and assembly. Proceedings of the National Academy of Sciences of the United States of America 93: 3099-3104 ISI:A1996UD37500089.

66. Hockley, D. J., M. V. Nermut, C. Grief, J. B. M. Jowett, and I. M. Jones. 1994. Comparative Morphology of Gag Protein Structures Produced by Mutants of the

Gag Gene of Human-Immunodeficiency-Virus Type-1. Journal of General Virology 75: 2985-2997 ISI:A1994PQ02800013.

67. Hoglund, S., A. Ohagen, J. Goncalves, A. T. Panganiban, and D. Gabuzda. 1997. Ultrastructure of HIV-1 genomic RNA. Virology 233: 271-279 ISI:A1997XJ07300003.

68. Houzet, L., J. C. Paillart, F. Smagulova, S. Maurel, Z. Morichaud, R. Marquet, and M. Mougel. 2007. HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. Nucleic Acids Research 35: 2695-2704 ISI:000247239600023.

69. Huang, M. J., A. Maynard, J. A. Turpin, L. Graham, G. M. Janini, D. G. Covell, and W. G. Rice. 1998. Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. Journal of Medicinal Chemistry 41: 1371-1381 ISI:000073346000004.

70. Huseby, D., R. L. Barklis, A. Alfadhli, and E. Barklis. 2005. Assembly of human immunodeficiency virus precursor Gag proteins. Journal of Biological Chemistry 280: 17664-17670 ISI:000228807200014.

71. Huthoff, H., and B. Berkhout. 2002. Multiple secondary structure rearrangements during HIV-1 RNA dimerization. Biochemistry 41: 10439-10445 ISI:000177435100009.

72. Ivanov, D., O. V. Tsodikov, J. Kasanov, T. Ellenberger, G. Wagner, and T. Collins. 2007. Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. Proceedings of the National Academy of Sciences of the United States of America 104: 4353-4358 ISI:000244972700022.

73. Johnson, M. C., H. M. Scobie, Y. M. Ma, and V. M. Vogt. 2002. Nucleic acid-independent retrovirus assembly can be driven by dimerization. Journal of Virology 76: 11177-11185 ISI:000178822400001.

74. Jowett, J. B. M., D. J. Hockley, M. V. Nermut, and I. M. Jones. 1992. Distinct Signals in Human-Immunodeficiency-Virus Type-1 Pr55 Necessary for Rna-Binding and Particle Formation. Journal of General Virology 73: 3079-3086 ISI:A1992KD37200003.

75. Kanevsky, I., F. Chaminade, D. Ficheux, A. Moumen, R. Gorelick, M. Negroni, J. L. Darlix, and P. Fosse. 2005. Specific interactions between HIV-1 nucleocapsid protein and the TAR element. Journal of Molecular Biology 348: 1059-1077 ISI:000229054400002.

76. Katen, S., and A. Zlotnick. 2009. The thermodynamics of virus capsid assembly. Methods Enzymol. 455: 395-417 PM:19289214.

77. Kay, L. E. 1998. Protein dynamics from NMR. Biochemistry and Cell Biology-Biochimie et Biologie Cellulaire 76: 145-152 ISI:000078073600002.

78. Kay, L. E. 2005. NMR studies of protein structure and dynamics. Journal of Magnetic Resonance 173: 193-207 ISI:000228076400003.

79. Kelly, B. N., B. R. Howard, H. Wang, H. Robinson, W. I. Sundquist, and C. P. Hill. 2006. Implications for viral capsid assembly from crystal structures of HIV-1 Gag(1-278) and CA(133-278)(N). Biochemistry 45: 11257-11266 ISI:000240575900002.

80. Khan, R., and D. P. Giedroc. 1992. Recombinant Human-Immunodeficiency-Virus Type-1 Nucleocapsid (Nc(P7)) Protein Unwinds Transfer-Rna. Journal of Biological Chemistry 267: 6689-6695 ISI:A1992HM05300037.

81. Khan, R., and D. P. Giedroc. 1992. Rna Double Helix Unwinding by Recombinant Hiv-1 Nucleocapsid (Nc) Protein. Faseb Journal 6: A359 ISI:A1992GY44002055.

82. Koradi, R., M. Billeter, and K. Wuthrich. 1996. MOLMOL: A program for display and analysis of macromolecular structures. Journal of Molecular Graphics 14: 51-& ISI:A1996UH51500008.

83. Krausslich, H. G., M. Facke, A. M. Heuser, J. Konvalinka, and H. Zentgraf. 1995. The Spacer Peptide Between Human-Immunodeficiency-Virus Capsid and Nucleocapsid Proteins Is Essential for Ordered Assembly and Viral Infectivity. Journal of Virology 69: 3407-3419 ISI:A1995QX93100019.

84. Lacowicz, j. r. 2009. Principles of fluorescence spectroscopy.

85. Lanman, J., T. T. Lam, S. Barnes, M. Sakalian, M. R. Emmett, A. G. Marshall, and P. E. Prevelige. 2003. Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. Journal of Molecular Biology 325: 759-772 ISI:000180544800013.

86. Lanman, J., T. T. Lam, M. R. Emmett, A. G. Marshall, M. Sakalian, and P. E. Prevelige. 2004. Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. Nature Structural & Molecular Biology 11: 676-677 ISI:000222274100019.

87. Larson, D. R., Y. M. Ma, V. M. Vogt, and W. W. Webb. 2003. Direct measurement of Gag-Gag interaction during retrovirus assembly with FRET and fluorescence correlation spectroscopy. Journal of Cell Biology 162: 1233-1244 ISI:000185794300006.

88. latt, S. A., and H. A. Sober. 1967. protein-nucleic acid interactions ii. oligopeptide-polyribonucleotide binding sites. Biochemistry 6: 3293-3306.

89. Layne, S. P., M. J. Merges, M. Dembo, J. L. Spouge, S. R. Conley, J. P. Moore, J. L. Raina, H. Renz, H. R. Gelderblom, and P. L. Nara. 1992. Factors Underlying Spontaneous Inactivation and Susceptibility to Neutralization of Human-Immunodeficiency-Virus. Virology 189: 695-714 ISI:A1992JF68400030.

90. Lee, B. M., R. N. De Guzman, B. G. Turner, N. Tjandra, and M. F. Summers. 1998. Dynamical behavior of the HIV-1 nucleocapsid protein. Journal of Molecular Biology 279: 633-649 ISI:000074358400011.

91. Lever, A., H. Gottlinger, W. Haseltine, and J. Sodroski. 1989. Identification of A Sequence Required for Efficient Packaging of Human Immunodeficiency Virus Type-1 Rna Into Virions. Journal of Virology 63: 4085-4087 ISI:A1989AK28400068.

92. Levitt, M. H. 2006. Spin dynamics: the basics of nuclear magnetic resonance. John Wiley and sons.

93. Liang, C., J. Hu, R. S. Russell, A. Roldan, L. Kleiman, and M. A. Wainberg. 2002. Characterization of a putative alpha-helix across the capsid-SP1 boundary that is critical for the multimerization of human immunodeficiency virus type 1 gag. Journal of Virology 76: 11729-11737 ISI:000178822400057.

94. Linial, M. L., and A. D. Miller. 1990. Retroviral Rna Packaging - Sequence Requirements and Implications. Current Topics in Microbiology and Immunology 157: 125-152 ISI:A1990EE44400005.

95. Lipari, G., and A. Szabo. 1982. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. Journal of the American Chemical Society 104: 4546-4559 ISI:A1982PC82900009.

96. Lipari, G., and A. Szabo. 1982. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. Journal of the American Chemical Society 104: 4559-4570 ISI:A1982PC82900010.

97. Maki, A. H., A. Ozarowski, A. Misra, M. A. Urbaneja, and J. R. Casas-Finet. 2001. Phosphorescence and optically detected magnetic resonance of HIV-1 nucleocapsid protein complexes with stem-loop sequences of the genomic Psi-recognition element. Biochemistry 40: 1403-1412 ISI:000166842200031.

98. Mandel, A. M., M. Akke, and A. G. Palmer. 1995. Backbone Dynamics of Escherichia-Coli Ribonuclease-H1 - Correlations with Structure and Function in An Active Enzyme. Journal of Cellular Biochemistry : 29 ISI:A1995QT86500079.

99. Marquet, R., J. C. Paillart, E. Skripkin, C. Ehresmann, and B. Ehresmann. 1994. Dimerization of Human-Immunodeficiency-Virus Type-1 Rna Involves Sequences Located Upstream of the Splice Donor Site. Nucleic Acids Research 22: 145-151 ISI:A1994MV14000005.

100. Marx, P. A., R. J. Munn, and K. I. Joy. 1988. Computer Emulation of Thin-Section Electron-Microscopy Predicts An Envelope-Associated Icosadeltahedral Capsid for Human Immunodeficiency Virus. Laboratory Investigation 58: 112-118 ISI:A1988L865000014.

101. Massiah, M. A., M. R. Starich, C. Paschall, M. F. Summers, A. M. Christensen, and W. I. Sundquist. 1994. 3-Dimensional Structure of the Human-Immunodeficiency-Virus Type-1 Matrix Protein. Journal of Molecular Biology 244: 198-223 ISI:A1994PU76100007.

102. Massiah, M. A., D. Worthylake, A. M. Christensen, W. I. Sundquist, C. P. Hill, and M. F. Summers. 1996. Comparison of the NMR and X-ray structures of the HIV-1 matrix protein: Evidence for conformational changes during viral assembly. Protein Science 5: 2391-2398 ISI:A1996WA11500002.

103. Mayo, K., D. Huseby, J. McDermott, B. Arvidson, L. Finlay, and E. Barklis. 2003. Retrovirus capsid protein assembly arrangements. Journal of Molecular Biology 325: 225-237 ISI:000180739500017.

104. Mely, Y., R. H. de, N. Morellet, B. P. Roques, and D. Gerad. 1996. Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy. Biochemistry 35: 5175-5182 PM:8611501.

105. Momany, G., L. C. Kovari, A. J. Prongay, W. Keller, R. K. Gitti, B. M. Lee, A. E. Gorbalenya, L. Tong, J. McClure, L. S. Ehrlich, M. F. Summers, C. Carter, and M. G. Rossmann. 1996. Crystal structure of dimeric HIV-1 capsid protein. Nature Structural Biology 3: 763-770 ISI:A1996VF28000010.

106. Morellet, N., R. H. de, Y. Mely, N. Jullian, H. Demene, M. Ottmann, D. Gerard, J. L. Darlix, M. C. Fournie-Zaluski, and B. P. Roques. 1994. Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR. J. Mol. Biol. 235: 287-301 PM:8289249.

107. Morellet, N., H. Demene, V. Teilleux, T. Huynh-Dinh, R. H. de, M. C. Fournie-Zaluski, and B. P. Roques. 1998. Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). J. Mol. Biol. 283: 419-434 PM:9769215.

108. Morellet, N., S. Druillennec, C. Lenoir, S. Bouaziz, and B. P. Roques. 2005. Helical structure determined by NMR of the HIV-1 (345-392)Gag sequence, surrounding p2: implications for particle assembly and RNA packaging. Protein Sci. 14: 375-386 PM:15659370.

109. Morellet, N., N. Jullian, H. DeRocquigny, B. Maigret, J. L. Darlix, and B. P. Roques. 1992. Determination of the Structure of the Nucleocapsid Protein Ncp7 from the Human-Immunodeficiency-Virus Type-1 by H-1-Nmr. Embo Journal 11: 3059-3065 ISI:A1992JE53700034.

110. Morita, E., and W. I. Sundquist. 2004. Retrovirus budding. Annual Review of Cell and Developmental Biology 20: 395-425 ISI:000225318200015.

111. Muriaux, D., and A. Rein. 2003. Encapsidation and transduction of cellular genes by retroviruses. Frontiers in Bioscience 8: D135-D142 ISI:000180266400022.

112. Nermut, M. V., C. Grief, S. Hashmi, and D. J. Hockley. 1993. Further Evidence of Icosahedral Symmetry in Human and Simian Immunodeficiency Virus. Aids Research and Human Retroviruses 9: 929-938 ISI:A1993MF84200001.

113. Nermut, M. V., D. J. Hockley, P. Bron, D. Thomas, W. H. Zhang, and I. M. Jones. 1998. Further evidence for hexagonal organization of HIV gag protein in prebudding assemblies and immature virus-like particles. Journal of Structural Biology 123: 143-149 ISI:000077145500006.

114. Newman, J. L., E. W. Butcher, D. T. Patel, Y. Mikhaylenko, and M. F. Summers. 2004. Flexibility in the P2 domain of the HIV-1 Gag polyprotein. Protein Science 13: 2101-2107 ISI:000222928800012.

115. Overton, H. A., Y. Fujii, I. R. Price, and I. M. Jones. 1989. The Protease and Gag Gene-Products of the Human Immunodeficiency Virus - Authentic Cleavage and Post-Translational Modification in An Insect Cell Expression System. Virology 170: 107-116 ISI:A1989U535800013.

116. Ozer, N., T. Haliloglu, and C. A. Schiffer. 2006. Substrate specificity in HIV-1 protease by a biased sequence search method. Proteins-Structure Function and Bioinformatics 64: 444-456 ISI:000238624300014.

117. Paillart, J. C., L. Berthoux, M. Ottmann, J. L. Darlix, R. Marquet, B. Ehresmann, and C. Ehresmann. 1996. Dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. Journal of Virology 70: 8348-8354 ISI:A1996VT70400011.

118. Paillart, J. C., M. Dettenhofer, X. F. Yu, C. Ehresmann, B. Ehresmann, and R. Marquet. 2004. First snapshots of the HIV-1 RNA structure in infected cells and in virions. Journal of Biological Chemistry 279: 48397-48403 ISI:000224957000115.

119. Paoletti, A. C., M. F. Shubsda, B. S. Hudson, and P. N. Borer. 2002. Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1. Biochemistry 41: 15423-15428 ISI:000180015100038.

120. Piatak, M., M. S. Saag, L. C. Yang, S. J. Clark, J. C. Kappes, K. C. Luk, B. H. Hahn, G. M. Shaw, and J. D. Lifson. 1993. High-Levels of Hiv-1 in Plasma During All Stages of Infection Determined by Competitive Pcr. Science 259: 1749-1754 ISI:A1993KT81000036.

121. Platt, E. J., and O. K. Haffar. 1994. Characterization of Human-Immunodeficiency-Virus Type-1 Pr55(Gag) Membrane Association in A Cell-Free System - Requirement for A C-Terminal Domain. Proceedings of the National Academy of Sciences of the United States of America 91: 4594-4598 ISI:A1994NL60800101.

122. Poon, D. T. K., G. D. Li, and A. Aldovini. 1998. Nucleocapsid and matrix protein contributions to selective human immunodeficiency virus type 1 genomic RNA packaging. Journal of Virology 72: 1983-1993 ISI:000071997600034.

123. Pornillos, O., B. K. Ganser-Pornillos, B. N. Kelly, Y. Z. Hua, F. G. Whitby, C. D. Stout, W. I. Sundquist, C. P. Hill, and M. Yeager. 2009. X-Ray Structures of the Hexameric Building Block of the HIV Capsid. Cell 137: 1282-1292 ISI:000267373400021.

124. Prabu-Jeyabalan, M., E. Nalivaika, and C. A. Schiffer. 2002. Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. Structure 10: 369-381 ISI:000174409600012.

125. Prats, A. C., L. Sarih, C. Gabus, S. Litvak, G. Keith, and J. L. Darlix. 1988. Small Finger Protein of Avian and Murine Retroviruses Has Nucleic-Acid Annealing

Activity and Positions the Replication Primer Transfer-Rna Onto Genomic Rna. Embo Journal 7: 1777-1783 ISI:A1988N729100029.

126. Provitera, P., R. El-Maghrabi, and S. Scarlata. 2006. The effect of HIV-1 Gag myristoylation on membrane binding. Biophysical Chemistry 119: 23-32 ISI:000234647700004.

127. Raja, C., J. Ferner, U. Dietrich, S. Avilov, D. Ficheux, J. L. Darlix, H. de Rocquigny, H. Schwalbe, and Y. Mely. 2006. A tryptophan-rich hexapeptide inhibits nucleic acid destabilization chaperoned by the HIV-1 nucleocapsid protein. Biochemistry 45: 9254-9265 ISI:000239238500028.

128. Ramboarina, S., N. Srividya, R. A. Atkinson, N. Morellet, B. P. Roques, J. F. Lefevre, Y. Mely, and B. Kieffer. 2002. Effects of temperature on the dynamic behaviour of the HIV-1 nucleocapsid NCp7 and its DNA complex. J. Mol. Biol. 316: 611-627 PM:11866521.

129. Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wongstaal. 1985. Complete Nucleotide-Sequence of the Aids Virus, Htlv-Iii. Nature 313: 277-284 ISI:A1985AAQ1500037.

130. Rein, A., L. E. Henderson, and J. G. Levin. 1998. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. Trends in Biochemical Sciences 23: 297-301 ISI:000075563400010.

131. Rice, W. G., J. A. Turpin, M. J. Huang, D. Clanton, R. W. Buckheit, D. G. Covell, A. Wallqvist, N. B. McDonnell, R. N. DeGuzman, M. F. Summers, L. Zalkow, J. P.

Bader, R. D. Haugwitz, and E. A. Sausville. 1997. Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. Nature Medicine 3: 341-345 ISI:A1997WL30500039.

132. Rodriguezrodriguez, L., Z. Tsuchihashi, G. M. Fuentes, R. A. Bambara, and P. J. Fay. 1995. Influence of Human-Immunodeficiency-Virus Nucleocapsid Protein on Synthesis and Strand Transfer by the Reverse-Transcriptase In-Vitro. Journal of Biological Chemistry 270: 15005-15011 ISI:A1995RE66600027.

133. Roques, B. P., N. Morellet, R. H. de, H. Demene, W. Schueler, and N. Jullian. 1997. Structure, biological functions and inhibition of the HIV-1 proteins Vpr and NCp7. Biochimie 79: 673-680 PM:9479450.

134. Rose, S., P. Hensley, D. J. Oshannessy, J. Culp, C. Debouck, and I. Chaiken. 1992. Characterization of Hiv-1 P24 Self-Association Using Analytical Affinity-Chromatography. Proteins-Structure Function and Genetics 13: 112-119 ISI:A1992HT89300003.

135. Royer, M., M. Cerutti, B. Gay, S. S. Hong, G. Devauchelle, and P. Boulanger. 1991. Functional Domains of Hiv-1 Gag-Polyprotein Expressed in Baculovirus-Infected Cells. Virology 184: 417-422 ISI:A1991GA34500047.

136. Saad, J. S., J. Miller, J. Tai, A. Kim, R. H. Ghanam, and M. F. Summers. 2006. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proceedings of the National Academy of Sciences of the United States of America 103: 11364-11369 ISI:000239353900047.

137. Sakaguchi, K., N. Zambrano, E. T. Baldwin, B. A. Shapiro, J. W. Erickson, J. G. Omichinski, G. M. Clore, A. M. Gronenborn, and E. Appella. 1993. Identification

of A Binding-Site for the Human-Immunodeficiency-Virus Type-1 Nucleocapsid Protein. Proceedings of the National Academy of Sciences of the United States of America 90: 5219-5223 ISI:A1993LF28500093.

138. Scarlata, S., and C. Carter. 2003. Role of HIV-1 Gag domains in viral assembly. Biochimica et Biophysica Acta-Biomembranes 1614: 62-72 ISI:000184483900007.

139. Schwarzinger, S., G. J. A. Kroon, T. R. Foss, J. Chung, P. E. Wright, and H. J. Dyson. 2001. Sequence-dependent correction of random coil NMR chemical shifts. Journal of the American Chemical Society 123: 2970-2978 ISI:000167806300005.

140. Shubsda, M. F., A. C. Paoletti, B. S. Hudson, and P. N. Borer. 2002. Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. Biochemistry 41: 5276-5282 ISI:000175223400024.

141. Skripkin, E., J. C. Paillart, R. Marquet, B. Ehresmann, and C. Ehresmann. 1994. Identification of the Primary Site of the Human-Immunodeficiency-Virus Type-1 Rna Dimerization In-Vitro. Proceedings of the National Academy of Sciences of the United States of America 91: 4945-4949 ISI:A1994NN21400071.

142. Song, R., J. Kafaie, L. Yang, and M. Laughrea. 2007. HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process; the DIS of stem-loop 1 initiates viral RNA dimerization. Journal of Molecular Biology 371: 1084-1098 ISI:000248809000019.

143. Spera, S., and A. Bax. 1991. Empirical Correlation Between Protein Backbone Conformation and C-Alpha and C-Beta C-13 Nuclear-Magnetic-Resonance Chemical-Shifts. Journal of the American Chemical Society 113: 5490-5492 ISI:A1991FU90000071.

144. Srinivasakumar, N., M. L. Hammarskjold, and D. Rekosh. 1995. Characterization of Deletion Mutations in the Capsid Region of Human-Immunodeficiency-Virus Type-1 That Affect Particle Formation and Gag-Pol Precursor Incorporation. Journal of Virology 69: 6106-6114 ISI:A1995RU78400021.

145. Sticht, J. 2005. HIV-1 capsid assembly and its inhibition. University of Heidelberg.

146. Sticht, J., M. Humbert, S. Findlow, J. Bodem, M. Muller, U. Dietrich, J. Werner, and H. G. Krausslich. 2005. A peptide inhibitor of HIV-1 assembly in vitro. Nature Structural & Molecular Biology 12: 671-677 ISI:000230956400016.

147. Summers, M. F., L. E. Henderson, M. R. Chance, J. W. Bess, T. L. South, P. R. Blake, I. Sagi, G. Perezalvarado, R. C. Sowder, D. R. Hare, and L. O. Arthur. 1992. Nucleocapsid Zinc Fingers Detected in Retroviruses - Exafs Studies of Intact Viruses and the Solution-State Structure of the Nucleocapsid Protein from Hiv-1. Protein Science 1: 563-574 ISI:A1992HW90500001.

148. Ternois, F., J. Sticht, S. Duquerroy, H. G. Krausslich, and F. A. Rey. 2005. The HIV-1 capsid protein C-terminal domain in complex with a virus assembly inhibitor. Nature Structural & Molecular Biology 12: 678-682 ISI:000230956400017.

149. Turner, B. G., and M. F. Summers. 1999. Structural biology of HIV. Journal of Molecular Biology 285: 1-32 ISI:000077968200001.

150. Urbaneja, M. A., B. P. Kane, D. G. Johnson, R. J. Gorelick, L. E. Henderson, and J. R. Casas-Finet. 1999. Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: Elucidation of the structural

determinants for function. Journal of Molecular Biology 287: 59-75 ISI:000079315400006.

151. Vogt, V. M., and M. N. Simon. 1999. Mass determination of Rous sarcoma virus virions by scanning transmission electron microscopy. Journal of Virology 73: 7050-7055 ISI:000081377400099.

152. von Schwedler, U. K., K. M. Stray, J. E. Garrus, and W. I. Sundquist. 2003. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. Journal of Virology 77: 5439-5450 ISI:000182297200041.

153. Vuilleumier, C., E. Bombarda, N. Morellet, D. Gerard, B. P. Roques, and Y. Mely. 1999. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. Biochemistry 38: 16816-16825 PM:10606514.

154. Wang, L. G., and F. J. Sigworth. 2006. Cryo-EM and single particles. Physiology 21: 13-18 ISI:000235301000003.

155. Watts, J. M., K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess Jr, R. Swanstrom, C. L. Burch, and K. M. Weeks. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460: 711-716 http://dx.doi.org/10.1038/nature08237.

156. Wiegers, K., G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, and H. G. Krausslich. 1998. Sequential steps in human Immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. Journal of Virology 72: 2846-2854 ISI:000072586900032.

157. Wilk, T., I. Gross, B. E. Gowen, T. Rutten, F. de Haas, R. Welker, H. G. Krausslich, P. Boulanger, and S. D. Fuller. 2001. Organization of immature human immunodeficiency virus type 1. Journal of Virology 75: 759-771 ISI:000166015000022.

158. Worthylake, D. K., H. Wang, S. H. Yoo, W. I. Sundquist, and C. P. Hill. 1999. Structures of the HIV-1 capsid protein dimerization domain at 2.6 angstrom resolution. Acta Crystallographica Section D-Biological Crystallography 55: 85-92 ISI:000078314000012.

159. Wright, E. R., J. B. Schooler, H. J. Ding, C. Kieffer, C. Fillmore, W. I. Sundquist, and G. J. Jensen. 2007. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. Embo Journal 26: 2218-2226 ISI:000245851500021.

160. Yeager, M., E. M. Wilson-Kubalek, S. G. Weiner, P. O. Brown, and A. Rein. 1998. Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy: implications for retroviral assembly mechanisms. Proc. Natl. Acad. Sci. U. S. A 95: 7299-7304 PM:9636143.

161. Yoo, S. H., D. G. Myszka, C. Y. Yeh, M. McMurray, C. P. Hill, and W. I. Sundquist. 1997. Molecular recognition in the HIV-1 capsid/cyclophilin a complex. Journal of Molecular Biology 269: 780-795 ISI:A1997XG93000012.

162. Yu, E. T., A. Hawkins, J. Eaton, and D. Fabris. 2008. MS3D structural elucidation of the HIV-1 packaging signal. Proceedings of the National Academy of Sciences of the United States of America 105: 12248-12253 ISI:000258905700029.

163. Zeffman, A., S. Hassard, G. Varani, and A. Lever. 2000. The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein (vol 297, pg 877, 2000). Journal of Molecular Biology 301: 1315 ISI:000089304600017.

164. Zhang, H., Q. Zhao, S. Bhattacharya, A. A. Waheed, X. Tong, A. Hong, S. Heck, F. Curreli, M. Goger, D. Cowburn, E. O. Freed, and A. K. Ebnath. 2008. A cell-penetrating helical peptide as a potential HIV-1 inhibitor. Journal of Molecular Biology 378: 565-580 ISI:000255965600008.

165. Zhou, J., X. Yuan, D. Dismuke, B. M. Forshey, C. Lundquist, K. H. Lee, C. Aiken, and C. H. Chen. 2004. Small-molecule inhibition of human immunodeficiency virus type 1 replication by specific targeting of the final step of virion maturation. Journal of Virology 78: 922-929 ISI:000187957700038.

166. Zhu, P., E. Chertova, J. Bess, J. D. Lifson, L. O. Arthur, J. Liu, K. A. Taylor, and K. H. Roux. 2003. Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. Proceedings of the National Academy of Sciences of the United States of America 100: 15812-15817 ISI:000187554600094.

167. Zlotnick, A. 2005. Theoretical aspects of virus capsid assembly. J. Mol. Recognit. 18: 479-490 PM:16193532.