FUNCTIONAL CHARACTERISATION OF THE N-TERMINAL DOMAIN OF POLYOMAVIRUS LARGE T ANTIGEN

by

KONSTANTIN KNOBLICH

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Cancer Sciences

College of Medical and Dental Sciences

The University of Birmingham

September 2009

Abstract

Although scientists have extensively researched the relationship between viral oncoproteins and cellular tumour suppressor proteins in recent years, the molecular interactions between these proteins is still poorly understood. It is the goal of this thesis to establish the key elements of specific interactions, in particular to characterise the interaction between the N-terminal part of the viral murine polyoma oncoprotein large T antigen (PyLTNT), and the cellular human regulator protein retinoblastoma (pRb)z. The homologous SV40 large T antigen protein has been studied thoroughly in recent decades, and has been associated with mesothelioma, osteosarcoma and brain tumours. However, the murine polyomavirus encodes for 154 additional amino acids that are rich in glycine and proline residues and could potentially play an important role towards cell transformation. Moreover, the polyoma virus protein has not been studied to this extent before, and structural and binding experiments conducted here reveal that it remains functional while natively unfolded. Nuclear Magnetic Resonance (NMR) spectroscopy was employed to characterise the protein's motional properties in its native state. A large part of the backbone residues was assigned, and regions interacting with pRb formed a localised structure. The determination of polyomavirus regions associated with retinoblastoma (PRAR) between residues 131 to 137 and 161 to 181 have never been observed before and represents a significant advance.

Acknowledgements

This thesis would not have been possible without the help and support of several people. I would firstly like to thank my supervisor Dr. Ulrich Günther for giving me the opportunity to work on a field that I had no prior knowledge of and has since broadened my horizons making me I would like to believe a better scientist. I would also like to thank him for giving me valuable guidance throughout my PhD. Secondly, I would like to thank my co-supervisor Prof. Michael Overduin who was always there to listen and advise. Much gratitude is owed to Cancer Research UK for providing the PhD studentship that made all this possible. A special thanks goes to Brian Schaffhausen and Tao Jiang that have provided me with more samples than I can count. I want to also thank Christian Ludwig and Sara Whittaker for providing help and teaching me NMR along the way. I want to also thank the former and current members of all my labs: Stefano, Ildefonso, John, Sue, Alessia, Mark, Paul, Tim, Mark, Rajesh, Pooja, Masae, Mike, Jas, Darren, Feli, and all those that I unintentionally forgot to mention, who have enriched my scientific background with their discussions, my day to day life in the lab with their personality, and helped me preserve my sanity when things became difficult. Finally, but not least, I would like to thank those that have believed in me and pushed me forwards towards my goals. Thank you Nuria for being there and lighting my way when it got dark. My parents Kurt and Georgia have been proud of me every step of this long journey and deserve a special thank you, I owe you the most.

Table of Contents

Index of Figures

Index of Tables

List of Abbreviations

AUC : *Analytical Ultracetrifugation*

β-OG : *Octyl-beta-D-glucopyranoside*

BKV : *BK Virus*

CD : *Circular Dichroism*

CDC : *Centre for Disease Control*

CHAPS : *3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate*

CSI : *Chemical Shift Index*

DSS : *4,4-dimethyl-4-silapentane-1-sulphonic acid*

FDA : *Federal Drug Administration*

FID : *Free Induction Decay*

FPLC : *Fast Protein Liquid Chromatography*

HetNOE : *Heteronuclear NOE*

HSQC : *Heteronuclear Single Quantum Coherence*

IPTG : *Isopropyl-β-D-1-thiogalactopyranoside*

JCV *: JC Virus*

LB : *Luria-Bertani*

LPPG : *1-palmitoyl-2-hydroxy-sn-glycero-3-(phospho-RAC-(1-glycerol))*

MCC: *Merkel Cell Carcinoma*

MWCO : *Molecular Weight Cut-off*

NIH : *National Institute of Health*

NLS : *Nuclear Localisation Signal*

NMR : *Nuclear Magnetic Resonance*

NOE : *Nuclear Overhauser Effect/Enhancement*

NOESY : *NOE spectroscopy*

ITC : *Isothermal Titration Calorimetry*

OD : *Optical Density*

O/N : *Overnight*

PI3K : *Phosphoinositide 3-kinase*

PML : *Progressive multifocal leukoencephalopathy*

PRAR : *Polyoma Region Associated with Retinoblastoma*

pRb : *Retinoblastoma protein*

PVN : *Polyomavirus-associated nephropathy*

PyV : *Polyomavirus*

PyLT : *Polyomavirus Large T antigen*

PyLTNT : *Polymavirus Large T antigen N-Terminal domain (residues 1-258)*

- **PyLTNTm1** : *residues 1-151*
- **PyLTNTm2** : *residues 1-200*
- **J-domain** : *residues 1-79*

SAXS : *Small Angle X-ray Scattering*

SV40 : *Simian Virus 40*

TMAO: *Trimethylamine N-oxide*

CHAPTER I - Introduction

“The most exciting phrase to hear in science, the one that heralds new discoveries, is not Eureka! (I found it!) but rather, “hmm.... that’s funny....” - Isaac Asimov 1920-1992

One of the great discoveries of the past fifty years is the role of oncogenes and tumour suppressors. In the course of this discovery the protein products of these tumour suppressor genes were studied intensely and yielded a large amount of information that furthered our knowledge of cell differentiation and cell death. For many discoveries viruses have paved the way to identify key proteins and signal transduction pathways.

From the simpler, yet potentially dangerous, retroviruses to the more complex and less well understood DNA viruses, we can pinpoint genes that are dedicated to differentiate cells for the benefit of the virus. Some gene products trigger a cascade of events that leads to cell differentiation with little or no viral benefit. Examples of such dedicated genes whose protein products cause cell differentiation can be found in the papilloma viruses i.e. E7 or E6 proteins. Other viruses carry genes for unintended cell differentiation, for example Burkitt's lymphoma that is caused by the Epstein-Barr virus. Although more than 95% of the population is carrying the Epstein-Barr virus, it can remain in a dormant state in the host, but can also facilitate the emergence of Burkitt's lymphoma.

In this thesis, the structural properties of a viral protein with strong oncogenic potential that belongs to the polyoma virus are characterised. The polyoma virus has a key role for the identification of cellular regulatory proteins that play a major part during cell transformation and its name arises from the fact that it is able to induce multiple tumours (Brodsky and Pipas, 1998; Benjamin, 2001). The discovery of many proteins including p53 and phosphoinositide 3-kinase (PI3K), the former a regulator

of apoptosis in the cell also known as the “guardian of the cell”, and the latter a signal transducer, can be credited to the polyoma virus.

Another protein that acts as a cell cycle regulator that will be discussed in more detail is the pRb protein. This protein has been shown to modulate the G1 to S phase transition in the cell cycle and is targeted by many viral proteins like EA1 from the adenoviruses, E7 from the papilloma viruses and also the large T antigen from the polyoma viruses.

Experimentation of these oncogenic proteins binding to pRb have shown that they change pRb’s structural conformation. The main theme of this thesis focuses on the interaction between the proteins polyoma large T antigen and pRb.

The starting point will be the history of the virus polyoma highlighting its significance for research and more specifically the significance it plays in this project. Further principles of the Nuclear Magnetic Resonance (NMR) methods and their relevance for exploring the interaction of the polyoma large T antigen with pRb will be introduced.

1.1. History of Polyomavirus

When Ludwik Gross first isolated a virus from mice in 1951 that could cause leukaemia and lymphomas when injected back into them (Gross, 1997), he could not foresee the controversy that would follow over his research in the next fifty years. In fact, almost five years had to pass for the world's first response to this discovery but unfortunately it was not a response of praise but that of fear.

The reason was the introduction of the first poliomyelitis (polio) vaccine. Children in those days became infected by the thousands by the crippling disease poliomyelitis, and when Jonas Salk discovered a vaccine against this terrifying disease in 1955, governments all over the world rushed to immunise the majority of the population.

While vaccination campaigns were under way in many parts of the world, two scientists in the National Institute of Health, Bernice Eddy and Sarah Steward injected the polio vaccine growth material into newborn hamsters and found that the animals were developing tumours (Dilworth, 2002). In fact, some of these mouse tumours can grow up to 25% of the host body mass in very short period of time (Benjamin, 2001). Further research revealed that they could attribute this phenomenon to a monkey virus found in the cells used to grow the vaccine. This virus was named the simian virus 40 (SV40). Furthermore this discovery showed that a virus from one host on which it had no detrimental effect, could infect another species and induce cancer. Later it was discovered that none of the viruses belonging to the same family as SV40 cause tumours in their natural hosts (Gottlieb and Villarreal, 2001).

The news immediately made many governments order the "cleanup" of the vaccine. By that time it was already controversially discussed, and fear of which effects the contaminated vaccine might have on millions of people who had already received it,

quickly spread. Follow-up studies were conducted over 20 years after vaccination e.g. Engels et al. who concludes that there is no direct link between the increase in cancer incidences and the vaccination (Engels et al., 2003). One bleak point in these studies was the failure to address some of the rare cancers that were observed in hamsters such as mesothelioma, osteosarcoma, non-Hodgkin's lymphoma and brain tumours. Another fact that was not addressed was the inability to distinguish between the exposed and non-exposed individuals since the virus in some instances might have been formalin inactivated during vaccine preparation. This was also observed by the Vaccine Safety Committee of the Institute of Medicine of the USA that concluded that these studies were 'sufficiently flawed' (Carbone et al., 2003).

For lack of better evidence this subject was silenced. In the coming years, research groups kept publishing results of SV40 link to tumours (Peden et al., 2008). This slowly divided the scientific community into those who thought the virus could induce tumours, and those who thought it had no effect on humans. These reports might have been individually anecdotal but if put together showed something beyond coincidental.

Finally in 1997 a meeting organized by the National Institute of Health (NIH), the Federal Drug Administration (FDA) and the Centre for Disease Control (CDC) was called to address the matter. In the preceding years to that meeting, SV40 sequences were found in human cancer cells of brain tumours, osteosarcomas and mesotheliomas, which were three out of four tumours that were spotted in hamsters. The meeting determined that a multi-laboratory study had to be conducted to determine if i) the virus could be classified as carcinogenic and ii) whether its presence was significantly higher in mesotheliomas compared to other cancer types (Carbone, 1999). The multi-laboratory study ruled out the concerns of contamination in PCR but remained indecisive concerning the other questions (Carbone, 1999).

In 2000 a second conference was held were it was established that SV40 is present in mesotheliomas and probably other tumours. In 2002 two large studies showed that SV40 was also present in non-Hodgkin's lymphoma (Ferber, 2002) which completed the correlation of hamster cancers with that of humans. In light of this evidence, we now understand why research in this field is becoming increasingly important. Nevertheless, some reports try to discredit previous studies on the SV40 association with the diseases it causes, by claiming contamination of their PCR results via plasmid homologues (Pearson, 2004). The most interesting discovery is the ability of the polyoma viruses to transform cells. Common assays used to assess transformation by SV40 in cells are their ability of the SV40 infected fibroblasts for immortalisation, survival and growth with little or no serum, high saturation densities in cultures, foci formation in multilayered cells and the ability to form tumours in animals after injection (Ahuja et al., 2005). A more detailed description of how they achieve tumourigenesis will follow.

1.2. Association with Disease

The polyomaviruses have only a few members that belong to the class of human pathogens. BKV and JCV cause polyomavirus-associated nephropathy (PVN) and progressive multifocal leukoencephalopathy (PML), respectively (White et al., 2005), and SV40 is a cross-species infecting pathogen associated with mesotheliomas, brain tumours, osteosarcomas and non-Hodgkin's lymphoma (Moens et al., 2007). PML is an opportunistic infection occurring in 5% of patients diagnosed with AIDS but the JCV virus can be also found in the urine of healthy people (Koralnik, 2006). PML represents the active state of a JCV infection which remains otherwise dormant in the majority of the population (White et al., 2005). In 2008, reports of drug induced PML made headlines implicating the drug natalizumab (Tysabri) which is used to treat multiple sclerosis and Crohn's disease. From the two cases reported one was fatal raising questions about the drug and its suitability since it might promote JCV reactivation. PML is a rare disease therefore there is controversy regarding natalizumab's effect, especially as in the reported cases immune-suppressants were administered at the same time and preceding to the natalizumab treatment. Moreover, natalizumab is reported to inhibit cytotoxic T-lymphocytes crossing the blood brain barrier which would include the anti-JCV lymphocytes (Carson et al., 2009).

Another recently discovered tumour that is associated with JCV is Merkel Cell Carcinoma (MCC). This skin tumour is particularly aggressive and the survival rate for patients that are diagnosed with the disease is 50%. Up to 75% of individuals have been found to carry antibodies of either BKV or JCV (Ravichandran and Major, 2006).

However, the mechanisms of infection are poorly understood even for SV40, which is the most extensively researched virus of the family. PyV is known to infect rodents but lacks the ability to cross-infect like SV40 which can be found in macaques and humans. This might be due to some crucial differences between the polyomaviruses in the sequence of their proteins.

SV40, JCV and BKV are phylogenetically closer related compared to PyV (Pérez-Losada et al., 2006), which might explain the inability of PyV to cross-infect. BKV and JCV are 72-75% homologous with each other and 70% homologous with SV40 (Eash et al., 2006; Ravichandran and Major, 2006). PyV displays only 46% DNA homology to SV40 and is thereby more distantly related. JCV's and BKV's ability to induce tumours is T antigen dependent, and while they are able to cross infect and cause tumours in a variety of animals, the tumours are localised according to the tissue and cell restricted tropism of the virus (Eash et al., 2006). It has been speculated that tropism influences LT antigens of JCV and BKV ability to bind less efficiently to pRb than SV40's LT (Bollag et al., 2000). The transforming capacity of JCV seems also to be greater than that of BKV (Eash et al., 2006). A variety of biochemical assays confirm however, the presence of JCV and BKV in a variety of tumour cells and specifically in brain tumours in humans (White et al., 2005). Interestingly, human brain tumour cells prevent JCV from replicating by inhibition of expression of viral capsid proteins, and in laboratory animals JCV only causes brain tumours when inoculated intracranially (White et al., 2005). For BKV there are conflicting reports of whether the virus should be associated with brain tumours (White et al., 2005).

Nevertheless, polyoma viruses follow a persistent scheme of infection in mammals (Pérez-Losada et al., 2006). They seem to cause mild primary infection followed by lifelong non-pathogenic persistent reparation of episomal viral DNA (Pérez-Losada et

al., 2006). Their primary target seems to be the kidneys of a single species which might indicate that they have evolved together with their hosts (Pérez-Losada et al., 2006).

Viruses are however, rarely capable of causing cancer on their own and viral infection does not necessarily result to oncogenesis (Carbone et al., 2003). With polyoma viruses, infection in their natural hosts rarely results in oncogenesis, but the likelihood increases when they cross species (Carbone et al., 2003). Malignant transformation of cells occurs very rarely by polyoma viruses and is only possible through integration into the host genome (Carbone et al., 2003).

Despite a lack of association with disease in humans PyV has provided an exceptional model to study cellular proteins that are involved in tumour prevention. PyV has also been of great importance in finding closely related functions with other oncogenic viruses.

1.3. Polyomavirus Characteristics

The polyoma virus has the ability to transform cells by employing a limited number of proteins to overcome the vast cellular machinery that is in place to prevent unintended cell transformation. It is therefore logical that some of these proteins require the ability to interact with multiple targets. Genome organisation in a virus is compact and evolutionarily optimised to include only the most necessary genes. Polyoma viruses have two life cycles, one lytic and one transforming (Gottlieb and Villarreal, 2001). A variety of polyomavirus diseases are associated with the cytolytic properties of the virus (Lednicky and Butel, 1999). A non-productive infection is initiated by the transforming cycle where the early genes will take over the host DNA synthesis (Gottlieb and Villarreal, 2001).

Going back to the overall picture, the polyomaviridae family is divided into two genera i) the polyomavirus where PyV belongs to ii) and papillomavirus. PyV is grouped together with the better known SV40, BKV, JCV that in contrast are associated with disease in humans. PyV and SV40 are the most widely studied viruses from this group (Lednicky and Butel, 1999). PyV has a non-enveloped circular double-stranded DNA genome which is 5297 base pairs long. The genome consists of an origin of replication that allows for bi-directional transcription thereby encoding the early and late proteins (Figure 1). The early proteins are involved in multiple cellular interactions but also in viral genome proliferation allowing the virus to remain undetected by the immune system. During the first 20h post infection, the viral machinery is dedicated to driving the cell into the cell cycle, initiating expression of host proteins needed for viral replication (Brodsky and Pipas, 1998). This first stage of the viral life cycle is where the early proteins play a major role as they are the only proteins expressed.

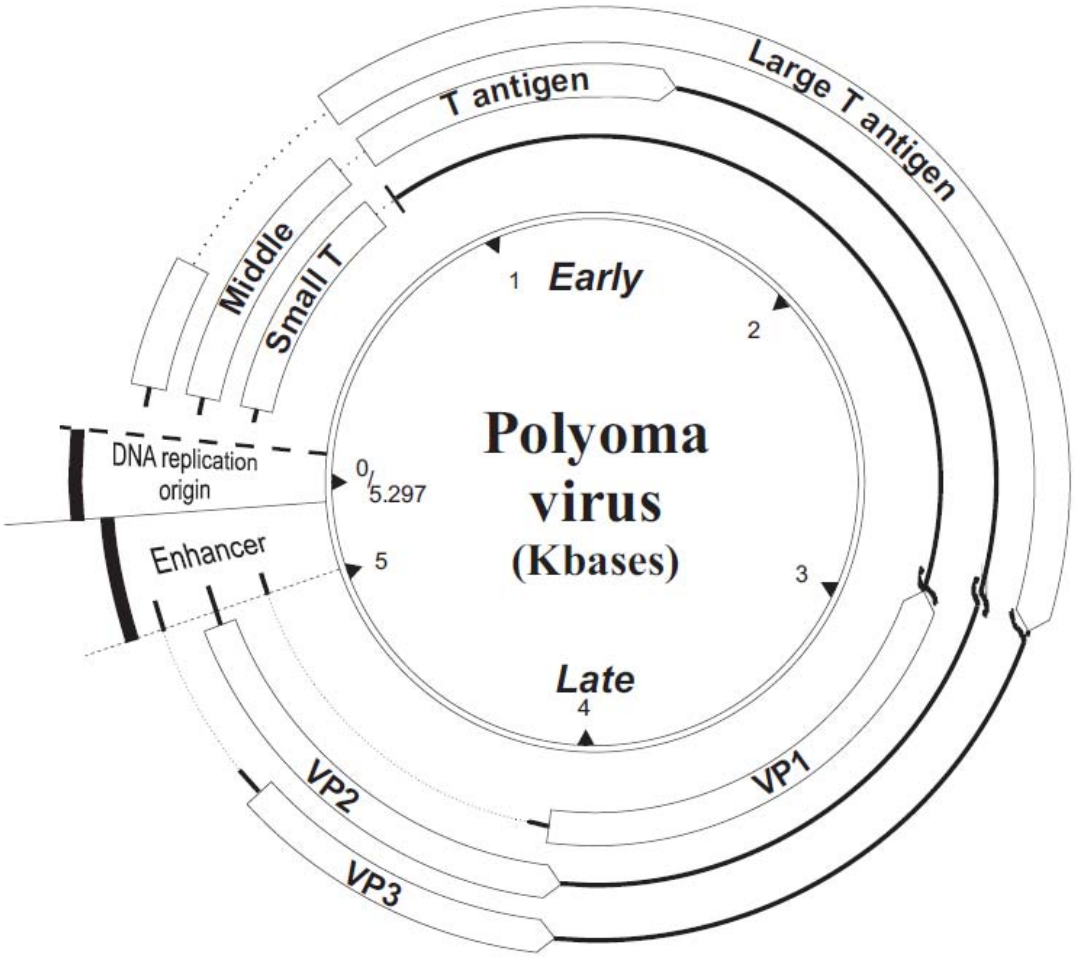


Figure 1. Polyomavirus DNA

The schematic representation of the genome indicates the early and late coding regions that are represented by boxes, with each protein produced written inside. The mRNA non-coding regions are represented by solid lines and the introns by dotted lines. The number of kilobases of DNA is marked in the inside. The early and late regions are marked on the inside of the circle. Enhancer region for early gene transcription and origin of DNA replication are also marked. Reprinted by permission from Macmillan Publishers Ltd: Oncogene (Ichaso and Dilworth, 2001), copyright 2001.

Early proteins are named according to their size as Large, Middle and Small T antigens often abbreviated as LT, MT and ST, respectively. A mechanism that is

employed for immune evasion, but only seen in the SV40 genome so far, is that of a miRNA that down-regulates the T antigen expression and thereby helps the virus escape the cytotoxic T lymphocytes and enhances successful infection (Sullivan et al., 2005).

The MT is unique to rodent viral strains and is the primary mediator of cell transformation (Scherneck et al., 2001). It was due to MT that discovery of p53, today heralded as the cell guardian, was made (Dilworth, 2002). The late proteins are the structural proteins VP1, VP2 and VP3 and are necessary for final assembly of the virus.

Early experiments on peptide fingerprinting and mutant analysis revealed all three T antigen sequences (Figure 2), showing that the first 79 N-terminal amino acids are common to all three T antigens. Furthermore, MT and ST share a central region of 112 amino acids (Wilson et al., 1986), whereas the C- terminus of each protein is unique (Dilworth, 2002). The additional coding capacity for the unique regions in MT and LT comes from the translation of separate reading frames found in the same ~700 base pairs of the genome (Dilworth, 2002).

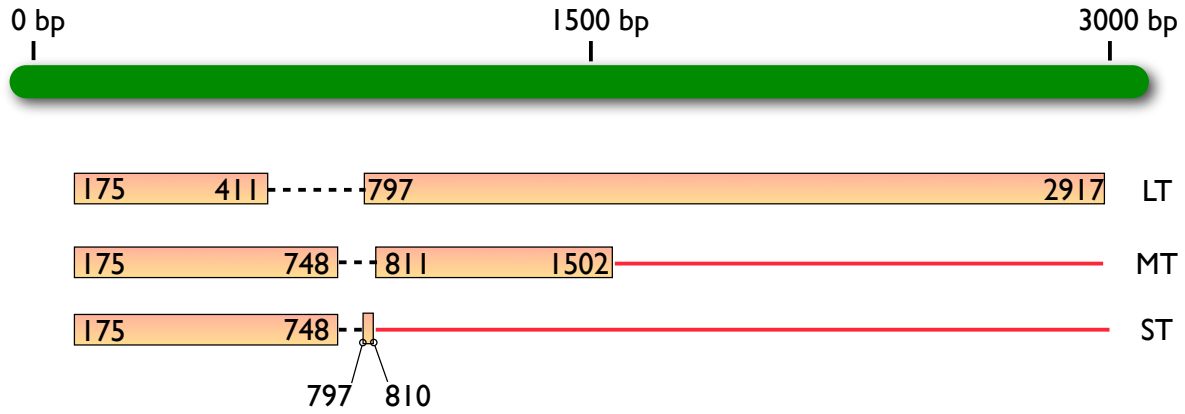


Figure 2. The polyomavirus early region coding capacity

The early region of polyomavirus is represented at the top in green. The mRNA sequences are present underneath. Solid boxes represent the coding sequences for

the three T antigens, the solid lines represent regions that are present in the mRNA, and dotted lines represent introns that are removed during splicing. Recreated from (Dilworth, 2002)

LT is a versatile multifunctional protein which the virus encodes to bypass cellular mechanisms and to integrate itself into the host DNA. It also binds to DNA and acts as a cell transforming agent conveying immortalisation ability to the cell (DeCaprio, 2009). It does so by melting and unwinding the DNA using its origin binding domain first as a monomeric form and then by forming into hexamers and acting as a helicase (Figure 3) (Bochkareva et al., 2006). This multi-functionality of the protein makes LT an intriguing target to study, as it interacts with key cellular tumours suppressors such as p53 and pRb and structural examinations show that it acts as a helicase in its aggregated form (VanLoock et al., 2002; Li et al., 2003; Gai et al., 2004).

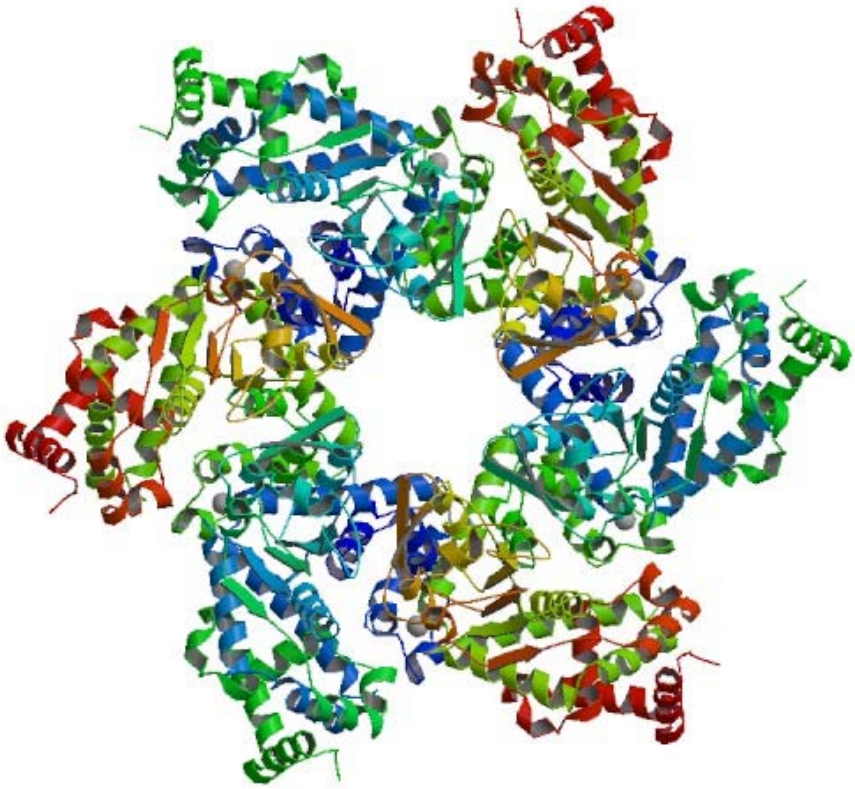


Figure 3. Hexameric form of the C-terminal SV40 large T antigen

The hexameric form of LT is comprised of an ensemble of six monomeric forms of the crystal structure representing residues 251-627. (Li et al., 2003)

The aggregated form of LT has been observed in electron micrographs showing the formation of hexameric complexes of LT which change conformation upon contact with DNA (VanLoock et al., 2002). Crystal structures of the hexamer complexes with p53 reveal conformational changes in the DNA binding area of p53 (Lilyestrom et al., 2006). The interaction between the two proteins leaves p53 unable to form functional tetramers because of the occupancy of the whole DNA binding surface by LT (Lilyestrom et al., 2006). However, the well structured hexameric form was observed for the SV40 C-terminus (residues 251-627) only and does not include the full length LT protein.

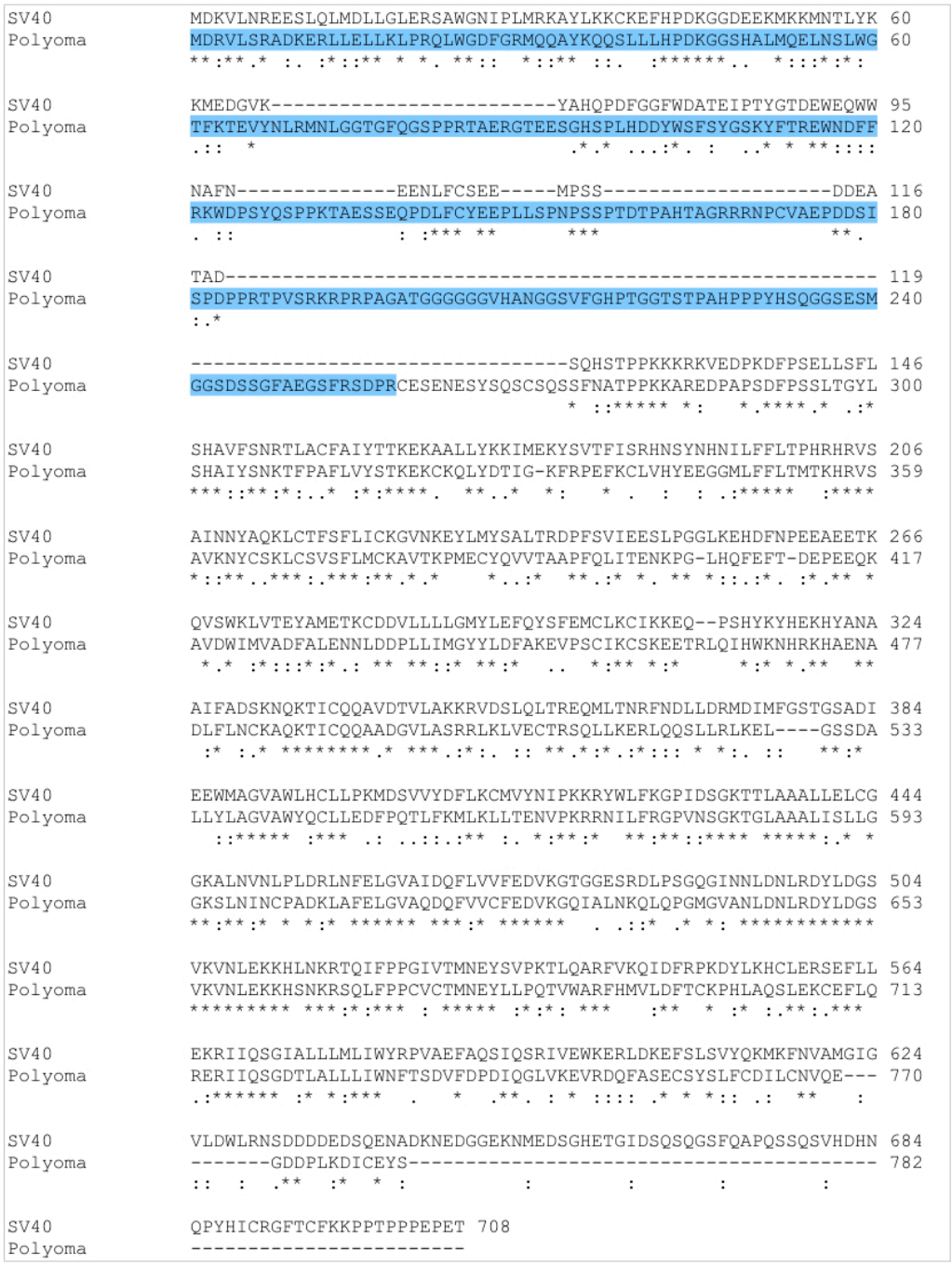


Figure 4. Alignment of SV40 and Polyomavirus LT sequences

CrustalW analysis (Larkin et al., 2007) *was used to align the two sequences. PyLTNT is highlighted in blue. The N-terminus is clearly different to the rest of the protein, owing to the additional 154 amino acids, that exhibits high homology.*

The PyV amino acid sequence reveals that there are important differences to the SV40 sequence (Figure 4). An additional 154 amino acids are encoded at the N-terminus of PyV which are not present in the SV40 (Pilon et al., 1996). These amino

acids might play a role in the host specificity of the protein and also in pRb interaction.

1.4. Large T Antigen Domains

As mentioned earlier, viral genomes are relatively small encoding limited number of proteins to establish a viral infection. LT is part of the viral machinery responsible for driving cell proliferation utilising host cell enzymes (Li et al., 2001). It is therefore important for LT to have the capacity to be multi-functional. Limited proteolysis of LT showed that the individual functionality of the N (1-258) and C-termini (259-728) is retained (Pipas, 1992; Gjørup et al., 1994). For SV40, the C-terminus contains a p53 interacting domain. p53 plays an important role in activating cell apoptosis, preventing DNA synthesis and interrupting cell cycle progression at the G2 and G1 phase (Sullivan and Pipas, 2002). Despite the fact that the C-terminus is important for efficient cell transformation, it is not required for immortalisation of the cells (Gjørup et al., 1994). In contrast, PyLT immortalises in a p53 independent manner (Pilon et al., 1996; Benjamin, 2001), unless it interacts at another point in an unidentified way in the cellular pathway. However, the function of the C-terminus is not known to date for the PyLT. The capability of PyV to immortalise cells through pRb inactivation is not essential for the induction of tumours as seen in studies with LT mutants that do not possess the ability to bind pRb (Benjamin, 2001). In addition, since PyV encodes for MT as well, p53 binding might be mediated by MT rather than LT (Gottlieb and Villarreal, 2001).

The N-terminus of LT has a number of important binding sites (Figure 5), including the J domain and the LxCxE motif, and also contains a nuclear localisation signal (NLS), spanning from residues 189-195, that has been shown to be vital for protein translocation to the nucleus and for pRb interaction (Howes et al., 1996). It has also been shown that there is a variety of not yet fully characterised proteins that co-immunoprecipitate with LT, and bind to the N-terminal region including Cul7, a

member of the cullins that promote ubiquitination (White and Khalili, 2005), a process that involves the ubiquitin molecule, that attaches itself to other proteins and promotes degradation.

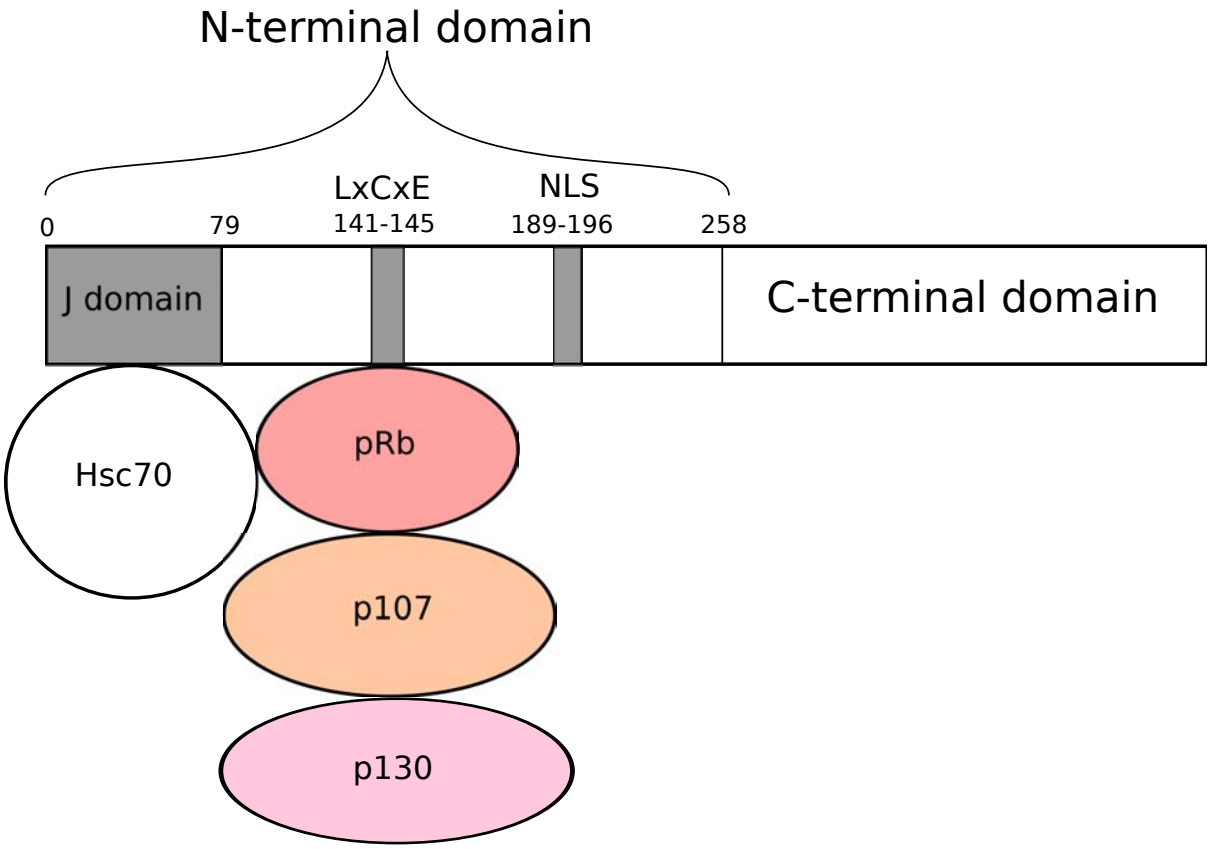


Figure 5. Cartoon representation of the LT protein with emphasis on the N-terminal domain and its binding sites

This schematic representation of the N-terminal domain shows in grey important interacting regions. Each region is annotated with its name. The appropriate binding partner proteins for each region is represented by circles containing their name. Hsc70 is also shown here to interact with pRb after binding to the J domain to facilitate pRb inactivation.

The J domain comprises of the first 79 amino acids and is conserved among T antigens. It contains an HPD motif that binds Hsc70 (Lin and DeCaprio, 2003), which in turn acts on the pRb-E2F complex (Caracciolo et al., 2006). The HPD motif (42-44

a.a.) is essential for the dissociation of the pRb-E2F complex (Whalen et al., 2005) and mutational substitutions can render the J domain ineffective for Hsc70 binding (DeCaprio, 1999). Additionally, it has been shown that a deletion of the amino acids 17-27 of the LT increases the rate of hepatic tumour formation in mice, a possible indicator of the importance of these residues in the control of cell proliferation (Bennoun et al., 1998).

PyLT also depends on the J domain (Figure 6) for pRb inactivation (Sheng et al., 1997; Zalvide et al., 1998; Sheng et al., 2000). However, modifications of the downstream functions linked to the phosphorylation state of pRb and p130 and to the transactivation of promoters are J domain independent (Sheng et al., 2000). In addition, the J domain of PyV differs in structure to that of SV40. There are three helices in the PyV J domain instead of the four in SV40 and also in the DnaJ, a human cell structural and functional homologue of the J domain (Pellecchia et al., 1996; Berjanskii et al., 2000; Lee and Cho, 2002). The J domain structure of PyV while resembling that of DnaJ, is not modified upon extension of its C-terminus (Huang et al., 1999). While it is possible to restore partial function of the J domain of all SV40 T antigens if their original J domain sequence is removed, by replacement with the JCV J domain, the same is not possible by exchanging the domain against the DnaJ (Genevaux et al., 2002). Furthermore, the J domain of MT and LT behave differently, as any mutation in their respective J domain renders LT more likely to be unstable than MT, by a factor of two (Whalen et al., 2005). The helices of the PyV J domain extend from residues 7-16 (Helix I), 27-41 (Helix II) and 49-70 (Helix III). Helix III has been found to be important for Hsc70 recruitment (Garimella et al., 2006). The three, instead of four helices, found in the PyV J domain, compared to the DnaJ, might also be responsible for functional differences. This is supported by the DnaJ acting as a promoter for folding of unfolded or partially folded proteins and

polypeptides (Greene et al., 1998), while the PyV J domain acts as a molecular chaperone (Garimella et al., 2006).

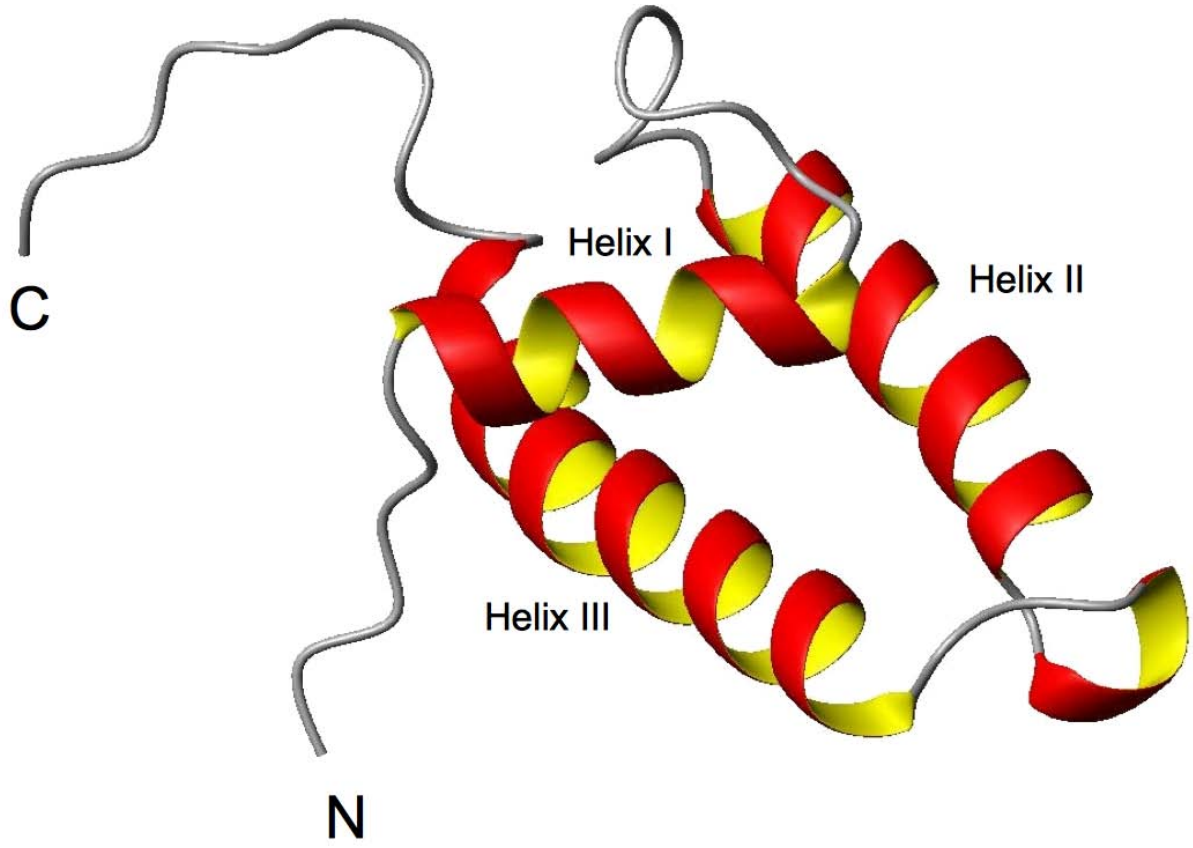


Figure 6. J domain structure of polyomavirus

This structure elucidated by NMR by Berjanskii et al. (Berjanskii et al., 2000) show the three helices that are present.

Amino acid substitutions reveal that the J domain might have multiple specificities for interaction with other proteins (Li et al., 2001). The segment between amino acids 13 and 19 seems to be of great importance for the initiation of viral DNA replication (Li et al., 2001), but there is no evidence that links the functions of the J domain in the LT with the J domain in the MT and ST even though the N-terminal sequence is shared among these proteins. Other binding sites of the J domain associate with more recently discovered proteins Bub1 and Cul7 (Nemethova et al., 2004).

Finally, the LxCxE motif which is part of LT but outside the J domain, is essential for binding the pRb protein family in particular pRb, which regulates G1/S phase progression via its phosphorylation and releases the E2F transcription factor that enables this transition (Lee and Cho, 2002; Nemethova et al., 2004). Despite the presence of the LxCxE motif, there is evidence that the J domain is required for efficient E2F release by pRb in the presence of PyLTNT (Sullivan et al., 2000a; Garimella et al., 2006). Further to this, both LxCxE and the J domain amino acid sequences must act in a cis sequential alignment suggesting that their presence is invaluable for the binding to pRb and for the efficient transformation of the cell (Srinivasan et al., 1997). Another sequence not yet identified on the protein but suggested by genetic complementation tests resides carboxy-terminal to amino acid residue 136 in the SV40 LT (Ahuja et al., 2005). This second independent function or element can immortalise primary cells and needs to act in a cis sequential arrangement with the other two elements of the protein, the LxCxE motif and the J domain (Ahuja et al., 2005). However, in Figure 4 the alignment shows that this element is not in the first 258 amino acids of PyLTNT. In fact, the element resides further downstream in homology with the SV40 sequence. Any mutation in the LxCxE domain would also disrupt the binding to the retinoblastoma family proteins (DeCaprio, 1999). In addition to the domains required, LT preferentially binds pRb in its hypophosphorylated form and might even be cell specific (Moens et al., 2007).

Taking all these findings into account, it becomes evident that the PyLTNT protein must be studied as a whole rather than individual sequences, or domains. For this reason investigations in this thesis were based on the smallest functional protein sequence that contains the above mentioned domains and is amenable to NMR analysis, here referred to as PyLTNT. This includes the J domain, the LxCxE motif and the NLS C-terminal residues. Later it will also be explained why the full length

PyLTNT is important from a structural point of view, as even minor conformational changes can contribute to the effectiveness of binding.

1.5. The Retinoblastoma Protein

The discovery of the retinoblastoma protein was spurred by research on a rare childhood tumour termed retinoblastoma, opening a new chapter of research and making the retinoblastoma protein one of the most studied cellular proteins known today. The retinoblastoma is a hereditary disease caused by tumourigenic retinal cells that contain a mutation on both retinoblastoma alleles in the locus at chromosome 13q14 (Whyte, 1995). The disease expresses early in children but is rare affecting 1 in 15,000 life births. The retinoblastoma gene product has a tumour suppressor function and many of the tumours found in man exhibit mutations on this retinoblastoma gene (Giacinti and Giordano, 2006).

Polyoma viruses have shown interaction with a range of the cell cycle machinery of their host that has been demonstrated to be vital for completion of their replication (Kim et al., 2001). The retinoblastoma family comprising of pRb, p107 and p130, exercises tight control over cell cycle regulation, especially over G1/S transition (Burkhart and Sage, 2008). It is therefore understandable that viral replication is dependent upon disruption of the control exercised by the retinoblastoma proteins.

The most significant protein of the retinoblastoma family is pRb, which has been found to be inactive in a wide variety of tumours although, in many tumours, it gets inhibited during progression of the disease rather than during the initial stages, suggesting a more complex role (Burkhart and Sage, 2008). This is a good indication that protein inactivation at later stages during tumour progression is very critical and external factors like viruses might enhance the chances of inactivation and progression. It has been noted that pRb can associate with over 100 protein partners and regulates transcription in a number of genes (Burkhart and Sage, 2008). In simple terms, pRb's main role seems to be the withdrawal from the cell cycle

progression in proliferating cells (Korenjak and Brehm, 2005). It is therefore not astonishing that the retinoblastoma proteins have been targeted by DNA viruses. Adenovirus, papilloma virus and SV40 are just a few of the well studied examples.

The main pathway retinoblastoma proteins are involved in is the progression of cells into the mitotic stage (Knudsen and Knudsen, 2008). They achieve this by inhibiting E2F transcription factors, thereby blocking proliferation of the cell. The typical pRb function involves control of the cell cycle by means of self hypo- and hyper-phosphorylation and thereby it controls E2F family members (Korenjak and Brehm, 2005) which comprise of at least eight transcription factors, each being cell cycle dependent and cell specific. E2F1, E2F2, E2F3a bind pRb and are transcription activators, while the E2F3b that also binds pRb is a transcription repressor (Cobrinik, 2005). E2F1 is specifically bound by pRb via pRb's distinctive C-terminal region, which is not found in p107 or p130 (Knudsen and Wang, 1998; Burkhart and Sage, 2008). E2F4 and E2F5 are also transcription repressors but interact with the other two proteins of the retinoblastoma family, p107 and p130. While p107 is predominantly expressed in proliferating cells p130 is found in higher concentration in arrested cells (Cobrinik, 2005). E2F6, E2F7 and E2F8 have no interacting domain but still function as repressors. More recent results show that the regulation of the E2Fs requires retinoblastoma family members via recruitment of histone deacetylases, chromatin remodelling factors and histone methyltransferases (Moens et al., 2007).

While the whole pRb has a size of 928 amino acids (Zhu, 2005), the minimum required sequence for efficient PyLTNT binding was mapped between residues 379 to 792 and is referred to as the small pocket (Kaelin et al., 1990). This small pocket consists of an A and a B domain connected via a small spacer region. This spacer region is vital for E2F binding. However, efficient E2F binding requires a larger region

that includes most of the B domain (Caracciolo et al., 2006). The residues in the pocket region with E2F that are in contact with pRb have been mapped by crystal structure determination (Xiao et al., 2003). The spacer region is important for the association with other proteins, and is homologous but differs slightly between the retinoblastoma proteins (Giacinti and Giordano, 2006). Tumours linked to inactive pRb often exhibit mutations in the pocket region (Lee et al., 1998). Viruses also exploit the considerable homology at the pocket domain between the retinoblastoma family proteins (Felsani et al., 2006).

The disruption of the pRb-E2F complex by PyLTNT is mediated by ATP hydrolysis (Sullivan et al., 2000a), and is targeted through the B domain (Figure 7) in a similar manner to the E7 and E1A proteins that use their LxCxE motif as a primary binding mode (Sullivan and Pipas, 2002). This motif seems to be essential for any binding to occur between the viral proteins and the host cell pRb protein. Curiously, despite detailed structural mapping of the pRb residue interaction with the LxCxE motif (Lee et al., 1998) there is to my knowledge, only one study of an attempt to map other interactions of the PyLTNT and pRb. Lee & Cho mention a DSSE motif with weak hydrogen bond interactions between SV40 112S and pRb (Lee and Cho, 2002). This motif is found not only in E1A and E7 proteins but also in the PyLTNT sequence in the form of a DSSI motif, between residues 177 to 180.

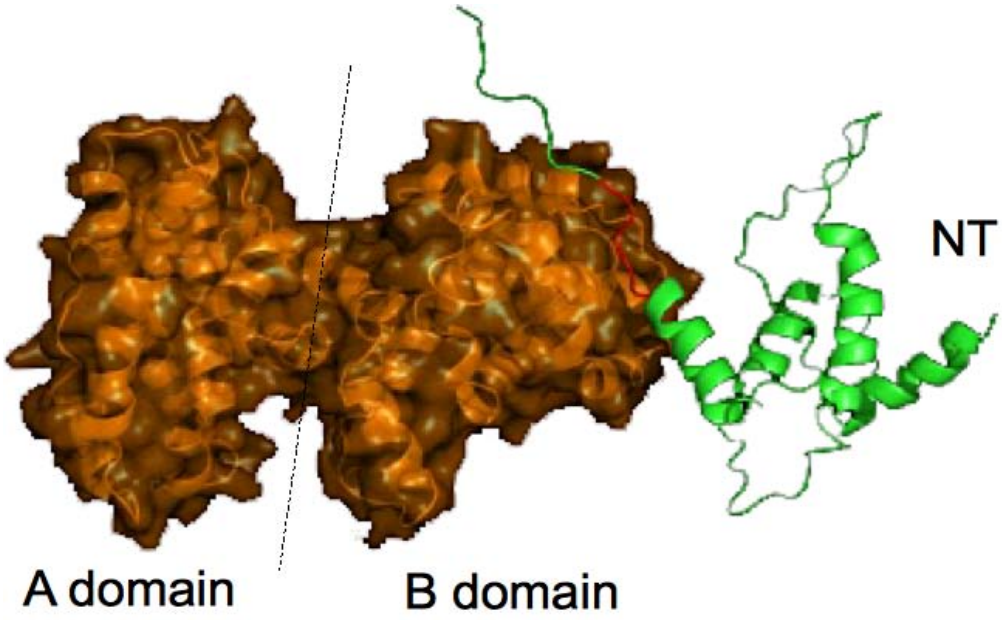


Figure 7. Crystal structure of pRb interacting with SV40 N-terminal domain

The pRb structure (residues 378-772) is highlighted in orange while the SV40 N-terminal domain (residues 7-117) is highlighted in green. The two proteins bind via the LxCxE motif shown in red at the B domain. The structure was determined by X-ray crystallography (Kim et al., 2001).

The shallow groove on the surface of the B region of pRb, formed by a number of non-consecutive amino acids, makes contact with the LxCxE motif (Lee et al., 1998). Interestingly, the B domain region where the LxCxE motif binds is nearly 30Å away from the spacer region which binds E2F (Liu et al., 2006). This leaves the spacer region free to bind to the E2Fs and to allow simultaneous interactions of the proteins. The pRb residues in the spacer region are extremely conserved in the retinoblastoma protein (Felsani et al., 2006), rendering them vulnerable to viral takeover. As already mentioned, while the LxCxE motif is vital for the specific binding of PyLTNT, the J-domain is also required (Sheng et al., 1997; Zalvide et al., 1998; Sheng et al., 2000).

Inactivation of pRb early during tumour development can be disadvantageous as alterations in pRb occur later during the progression of the disease (Burkhart and Sage, 2008). Although the evidence for its tumour suppressive function is overwhelming, its loss of function seems to be cell type or even tissue dependent (Burkhart and Sage, 2008). In spite of this, pRb inactivation has been shown to advance angiogenesis, increase cell death and DNA repair, as well as promote chromosomal instability and prevent cellular senescence (Burkhart and Sage, 2008). Therefore, disruption of the E2F-pRb complex is also vital for the cell and is responsible for more than just increase of proliferation (Cobrinik, 2005).

1.6. Natively Unfolded Proteins

The course of this thesis showed why the PyLTNT-pRb interaction was hardly studied from the side of PyLTNT. The protein is not only difficult to stabilise but is also largely unfolded making it a difficult but also intriguingly interesting target for biophysical analysis. This unfolded nature seems to be a common theme in viral oncoproteins, as will be evident further on. Natively unfolded proteins (NUPs) represent a major challenge for structural investigations and will therefore be introduced in the following section.

NUPs have been a controversial subject for some time now. Only during the last 10 years has there been enough evidence to support the theory that structured forms of proteins are not always a necessity for biological function. The structure-function paradigm has been a central dogma for over 100 years since the lock and key mechanism theory was proposed as an explanation for enzymatic activity (Uversky, 2002). Systematic analysis of amino acid sequences of proteins has revealed that large parts of the proteins, or even whole proteins, are lacking order in their structure. In fact, more than 15,000 proteins contain disordered regions, with over a 1000 having high scores of disorder (Uversky, 2002).

These NUPs can be further divided into premolten globules and intrinsic coils, with the first group exhibiting some residual structure, and the second having no secondary structure or showing hydrodynamic dimensions of a random coil (Uversky, 2002). Some studies confer that a NUP conformation is a result of the combined high net charge of the protein and its low mean hydrophobicity (Uversky et al., 2000). The low hydrophobicity allows for a less compact structure while the high net charge leads to charge-charge repulsion. Taken these two factors into account, a formula for likelihood of NUP conformation has been proposed:

$$C = (R+1.151)/2.785$$

where R is the net charge and if the value C is smaller than parameter H, which is the mean hydrophobicity, a protein is typically unstructured. This formula, predicts NUPs fairly reliably solely on the basis of the protein sequence.

Besides the two factors mentioned above, NUPs can also be predicted by low numbers of amino acid residues like I, L, V, W, F, Y, C, N, and higher amounts of other residues like E, K, R, G, Q, S, P (Fink, 2005). Prolines and glycines, contribute considerably to protein flexibility. In general, one could conclude that NUPs are characterised by a lack of a hydrophobic core and therefore have low numbers of the hydrophobic residues.

These proteins can take advantage of the fact that they contain little or no structure. Firstly, by allowing a considerable number of interactions with corresponding targets, and secondly, for disorder-order transitions, so called induced folding, as a way of regulating cellular processes, despite the entropic cost of folding (Uversky et al., 2000; Dyson and Wright, 2005; Receveur-Bréchot et al., 2006). The complexes formed have usually low affinity but high specificity, a feature commonly found in signal transduction (Dyson and Wright, 2005). This relies on the thermodynamic properties of unfolded proteins that fold upon binding. To have a fast rate the strength of the protein-protein interaction characterized by K_d must be much weaker than standard folded protein-protein K_d, unfolded proteins can also be thought of as a special case of extremely weak binding proteins, i.e., ordered. In such cases, the low affinity is exhibited in the unfolded protein. Such proteins also have extensive interfaces that they use during interactions. Otherwise, if these proteins were folded, to create the same interface, they would have to be larger in size and would make

the cell more crowded, or the cell itself would have to be 15-30% larger to accommodate for these extensions in their sequences (Fink, 2005).

Another interesting observation is that these proteins are more common in eukaryotes rather than prokaryotes. Presumably because prokaryotes do not have highly regulated degradation pathways, therefore have short life times, proteins are more costly, and the lack of compartmentalisation cannot protect NUPs from degradation (Fink, 2005). In contrast, the rapid degradation might be an advantage since they don't require ubiquitination (Fink, 2005). Their unfolded properties also provide them with an insensitivity to temperature increase (Receveur-Bréchot et al., 2006), which is advantageous concerning their stability. Examples of viral unfolded proteins have been found in the papillomaviruses, which belong to the same family as the polyomaviruses. The E7 protein, that interacts with p53 and pRb in a similar manner to PyLT, also shares the characteristic features of a partially unfolded protein (Uversky et al., 2006).

The study of these proteins has shown little progress in the matter of structural information. This is primarily because they have properties that make them unsuitable for commonly used structure determination techniques. They are highly flexible and lack conformational dispersion causing crowded NMR spectra, and are unable to crystallize for X-ray crystallography. However, the resonances observed in NMR are usually very intense and therefore useful for analysis, except for regions with altered flexibility, where signals may disappear owing to coalescence in intermediate exchange. Furthermore, NMR provides information on the protein dynamics, which can be used for localised structural information. Nevertheless, in combination with other techniques, such as small angle X-ray scattering (SAXS), localised structure formation can be elucidated for NUPs by NMR although at low

resolution, to determine basic biophysical properties with biological relevance. It is therefore apparent that NMR is one of the best methods for the analysis of NUPs.

The fact that these NUPs can be predicted by their amino acid sequence, gave rise to bioinformatic tools available online, as will be described later. In this thesis, a number of these have been used to allow initial understanding of the PyLTNT protein, as will be shown later. The following section addresses the characteristic NMR spectra that can be associated with these proteins.

1.7. NMR

NMR and X-ray crystallography are the predominant methods used to elucidate high resolution structures of proteins. X-ray crystallography has been favoured by many labs because of its relative ease of use and applicability to biomolecular samples with high molecular weight. Crystal forms of proteins can give good resolution structures. However, X-ray crystallography is hindered by the fact that not all proteins are able to crystallize, and also by being unable to determine unstructured regions in a protein. This is where NMR spectroscopy can supply additional information.

NMR has been in use for over 50 years, advancing rapidly in the last 20, enabling new applications in chemistry, biochemistry and medicine (Claridge, 1999; Cavanagh et al., 2006; Levitt, 2008). NMR is an extremely versatile tool for protein properties and is capable of measuring kinetics, dynamics and structure determination. NMR spectroscopy is now able to obtain structural information for proteins of molecular weights up to 30-35 kDa, in some cases up to 82kDa (Grishaev et al., 2008).

NMR is based on the intrinsic property that atoms have, a property termed “spin”. The spin behaves similarly to a rotating charged particle with an angular momentum. The net spin of each atom, denoted by I, depends on the nuclear subatomic composition. Additionally, nuclei with the same atomic number but different mass numbers are called isotopes. For some elements several isotopes exist where only some have a net spin. These isotopes occur in nature at various isotopic distributions. ^{1}H atoms are a naturally abundant isotope with a net spin, and NMR relies heavily on this atom for one-dimensional experiments. Many elements have a net spin for some isotope but the isotopic distribution in nature may be unfavourable. For example, the spin-active isotope of carbon (^{13}C) has only a natural abundance of 1.1%, and that of nitrogen (^{15}N) is only 0.37%. To assign NMR spectra in proteins,

isotopic labelling of ^{13}C and ^{15}N is required. This is readily possible as cells can incorporate sources of these isotopes into proteins from compounds like ^{13}C-glucose and $^{15}NH_4Cl$, which are commercially available. With the use of these isotopes, a new era of NMR started that enabled the development of multidimensional experiments as shown later. Many other atoms are used for NMR detection but ^{1}H, ^{13}C and ^{15}N are the ones used for protein assignment.

When spin active nuclei are subject to a magnetic field, they align in relation to the external field (B_o). For a simple introduction the Bloch vector model will be used to explain some basic properties of spins (Figure 8).

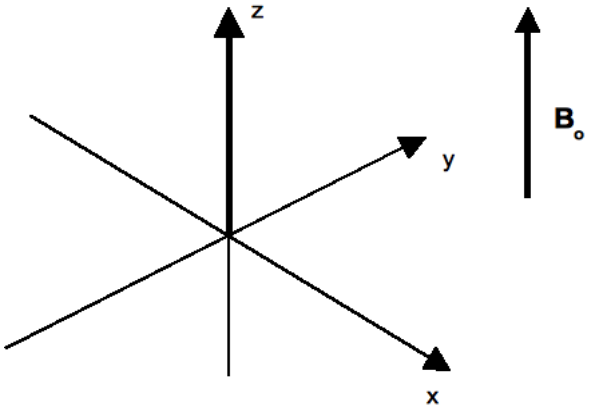


Figure 8. Bloch vector model

The axes are denoted as z, x and y and represent the orientation of the overall magnetisation arising from an ensemble of spins. The external magnetic field (Bo) is aligned along the z axis.

If the z axis represents the direction of the Bo field the spins begin to precess around it. The frequency of precession (ω_o), also termed the Larmor frequency is equal to:

$$\omega = -\gamma B_o$$

where γ is the gyromagnetic ratio and B_o the magnetic field. The oscillation frequency denoted as v can be measured in Hz and is given by $-\gamma B_o/2\pi$.

When spins are irradiated with a specific radiofrequency pulse, the magnetisation vector changes direction and reaches to what is called the transverse plane (x/y) for a 90° ($\pi/2$) pulse. The signal can be recorded in the form of a free induction decay (FID) that can be translated into an absorption signal of specific frequency using a Fourier transformation.

There are two relaxation processes the spin-lattice (longitudinal) relaxation and spin-spin (transverse) relaxation. Longitudinal relaxation rate (R_1) depends on interactions that originate from vibrational and rotational motions, which create a complex magnetic field. This field has multiple components that cause magnetisation to return to the equilibrium state. This is called the spin-lattice relaxation with a relaxation time T_1. T_1 also depends on the gyromagnetic ratio γ and the lattice. Spin-spin relaxation originates from the interaction between nuclei that are close to each other. The time constant T_2 describes this decay of magnetisation. The signal width (lw) is determined by the transverse relaxation rate ($R_2=1/T_2=\pi*lw$). If the transverse relaxation is fast, (large values of R_2) broad spectral signals are obtained while sharp signals correspond to slow relaxation.

Proteins as molecules with many interacting atoms cause complex NMR spectra with many resonances, too many to be resolved in a proton NMR spectrum. One dimensional (1D) proton spectra show signals at different frequencies for the different protons in the molecular system. The chemical shift dispersion of signals originates from varying the electron density around each nucleus that influences the Larmor frequency of each individual nuclear spin. This is a consequence of shielding or deshielding of the nuclei against the Bo field caused by induced magnetic fields involving surrounding electrons. Each proton signal can now be characterised

according to its electronic environment so that a methyl signal is distinct from an amide signal in a proton spectrum. Moreover, signals are split by the presence of adjacent spins in the same molecule, which can have different spin states causing the so called scalar coupling (*J*). These couplings and other properties of spins can be used in pulse programs to link different nuclei. Spectra can get extremely crowded for larger molecules, owing to degenerate frequencies for many different nuclei. To simplify spectra, large magnetic fields can be employed, because the frequency is field-dependent and therefore dispersion increases. In addition, spectral dimensions can be created by linking different nuclei exploiting scalar couplings or dipolar interactions between nuclei. The additional dimensions give rise to many types of NMR spectra which will be explained later.

One of the advantages of NMR is that it allows the investigation of proteins as close as possible to the native state. It also allows the observation of proteins under complex conditions or when they are natively unfolded. Moreover, NMR also has the ability to probe weak affinities between proteins providing an alternative to Biacore™.

Solution NMR relies heavily on the size of the protein, since rotational tumbling of the molecule affects resolution of spectra, with large proteins tumbling slowly and causing broad lines owing to faster transverse relaxation. Another problem is spectral crowding which is amplified in unfolded proteins, by lack of differentiation in chemical shift, even in multi-dimensional spectra. This can partly be resolved by the use of higher field magnets and specific pulse sequences. Deuterium labelling can decrease the relaxation rate of ^{13}C allowing for better resolution and sensitivity. Other methods to reduce linewidths include the use of higher temperatures to increase tumbling rates (Foster et al., 2007). This is only possible though for molecules that

are thermal sufficiently stable. To increase tumbling rates one must also use low viscosity solvents (Foster et al., 2007).

NMR has frequently been used to study protein interactions. This can be achieved by relatively simple two-dimensional spectra that can be used to map binding interfaces based on chemical shift perturbations (explained in more detail in section 3.4.2.). Protein-protein and protein-ligand interfaces can also be mapped using the Nuclear Overhauser Effect (NOE) arising from through space cross-relaxation of protons. Such effects are typically effective for distances up to 4-5 Å.

NMR has previously been used to study interactions of the LxCxE peptide with pRb. It was concluded that while the motif is essential for binding, higher affinity flanking residues extending the xLxCxExXx sequence are needed, where x should not be a positively charged amino acid, and X should preferentially be a hydrophobic residue (Singh et al., 2005). However, as evident from the sequence of PyV, none of the flanking residues are positively charged, and the X position has a proline residue. The residue 141D in PyLTNT is also of great importance, as it can form a hydrogen bridge with a residue of pRb and significantly enhance the interaction, especially since it is readily accessible in PyV (Lee and Cho, 2002).

There have been extensive studies of the protein structure of the SV40 LT but little is known of the structure of PyV LT. X-ray crystal structures include those of the N-terminal region of SV40 LT in complex with pRb (Kim et al., 2001).

A variety of LT domains have been examined by NMR. Most of these domains were part of the SV40 LT and include solved solution structures of the DNA binding domain of SV40 LT (Luo et al., 1996; Bochkareva et al., 2006), J domain structures of E. coli together with their glycine/phenylalanine rich regions (Pellecchia et al., 1996) and the PyV J domain solution structure (Berjanskii et al., 2000). The latter

study reveals a nicely formed structure of the J domain which included 3 helices (Berjanskii et al., 2000).

From this work it becomes evident that the protein, beyond the first 79 amino acids that form the J domain, becomes increasingly unfolded. NMR experiments on PyLTNT show that it is a partially unfolded protein with a structured J domain and an unstructured C-terminus. Partially unfolded proteins or NUPs have been considered for some time to play a major role in protein interactions. Unstructured regions are present in a variety of proteins, in particular in the absence of binding partners, as in some cases binding induces folding. The assignment of NMR resonances of NUPs is particularly challenging as the protein is flexible and has limited chemical shift in amide HN and C^{α} atoms (Dyson and Wright, 1998). Moreover, such proteins lack signals in NOESY spectra, not only for the amide region but also between side chain residues. It is therefore difficult to obtain determinants for secondary structure.

In contrast, the nitrogen and carbonyl carbon chemical shifts are somewhat better dispersed as a result of them being influenced more by the amino acid sequence than the secondary structure. Moreover, these resonances are very narrow owing to rapid fluctuations of the protein chain, allowing for good three-dimensional and two-dimensional experiments even at low concentrations (Dyson and Wright, 1998).

A study by Ohlenschläger et al. on the E7 protein, which has a similar functions as LT, has shown that structurally E7 is also a partially natively unfolded protein and that the full length E7 protein contains signals at a narrow chemical shift range (Ohlenschläger et al., 2006). To improve the dispersion Ohlenschläger studied the isolated C-terminus of the protein which folds into a structured domain (Ohlenschläger et al., 2006).

Another interesting application of NMR in NUPs is the possibility to measure dynamic properties of biomolecules. The flexibility of these proteins can be probed

by ^{15}N spin relaxation measurements. While the quantitative interpretation of ^{15}N relaxation rates and heteronuclear NOEs (hetNOE) is complex, valuable information can be extracted (Dyson and Wright, 1998), in particular regarding localised structure formation. Relaxation measurements will be discussed in sections 2.8.9 and 3.4.3 for the PyLTNT protein.

1.8. Objective

The objective of this thesis was to elucidate the biophysical properties related to activity of PyLTNT, in particular regarding conformational changes upon binding to pRb. For this, NMR was employed to obtain additional information on residual structure.

The hypothesis is based on the differences between SV40 and PyV, in particular in their N-terminal sequence which differs to a large extent in amino acid composition. Moreover, PyLTNT has an additional 154 amino acid residues which are likely to be important for yet unknown protein interactions, including those with Rb family members. It must be expected that the additional sequence is important to the virus as these residues have developed under evolutionary pressure and viruses have a highly compact genome. It was also expected to find determinants of localised structure related to activity in the protein.

To further analyse these ideas, the aim was to i) to assign the protein via NMR and ii) to determining the role of each residue towards binding with pRb. While assignment was very challenging, as will be described, it gave an insight into how different NMR experiments can be used to provide additional assignment information. Furthermore, by using kinetic information, the nature of binding and the overall complexity of the process that is pRb inactivation was further elucidated.

The hypothesis that the 154 additional amino acid residues might play a role towards pRb binding was investigated by NMR, and other biochemical methods, and is outlined in the following pages.

CHAPTER II – Materials & Methods

2.1. Materials

All chemicals used were purchased directly from Sigma-Aldrich or its partners (eg. Fluka Analytical). Lipids were purchased from Avanti Lipids Inc. (Alabaster, AL, USA). For small volume liquid centrifugation instruments by Fisher Scientific UK Ltd. were used. For larger volumes and higher speeds centrifuges from Beckman Coulter Inc. were used with corresponding rotors. CELTONE™ and Spectra-9™ media were purchased from Spectra Stable Isotopes (Columbia, MD, USA). Concentrators were either obtained from Millipore (Billerica, MA, USA) or from Sigma-Aldrich. The chromatographic instruments used were from GE Healthcare. More specifically the ÄKTA prime™ plus was used for isolating tagged proteins together with HisTrap™ and GSTrap™ columns, and the ÄKTA purifier™ for further purification using gel filtration size exclusion chromatography. Protein marker (Precision plus protein standards), gels and gel tanks were from Biorad. Complete® protease inhibitors were from Roche Applied Science. SYPRO® Orange for thermofluor experiments was purchased by Sigma-Aldrich.

2.2. Protein Expression

pET-30b and pET-15b DNA plasmids vectors that encoded residues 1-258 of PyLTNT were kindly provided by Brian Schaffhausen (Tufts School of Medicine, Boston, USA). pET-15b plasmid that encodes for residues 388-792 of pRb was also received from Brian Schaffhausen. Vector pET-30b of PyLTNT had a six histidine tag at the C-terminus with an incorporated Kanamycin resistance gene for validation of the insert, while the second vector had a GST tag at the N-terminus with an Ampicillin resistance gene for validation. Original plasmids were received dry on blotting paper. The paper was cut and inserted into two PCR tubes and 150µl of elution buffer from the Quiagen kit for plasmid purification was pipetted into each tube. From then on identical procedures were followed for production of protein from each of the PyLTNT vectors. For the pRb vector a 5 µl DNA solution was used in the exact same way as PyLTNT for cloning and expression. Furthermore, pET-30b vectors of two smaller constructs of PyLTNT were received, named PyLTNTm1 and PyLTNTm2 encoding residues 1-151 and 1-200 respectively with a histidine tag at the C-terminus, which were treated in the same way as PyLTNT.

10 µl of the PCR tube that contained the clone was mixed with 20µl of competent cells and incubated in an eppendorf tube on ice for 30 minutes. The DNA was frozen at -20 ^{o}C for future use. XL1-Blue cells were used initially, as they are suitable for creation of a stock cell culture because of their high copy numbers. After 30 minutes on ice the mixture of cells with DNA was heat-shocked for 45 seconds at 42^{o}C to allow the plasmids to be incorporated into the cells, and then the mixture was put back on ice for 5 minutes. The sample was then transferred into 400 µl of SOC medium and allowed to grow for 2 hours at 37 ^{o}C. 50µl of that medium was pipetted onto a Luria-Bertani (LB) agar plate with the appropriate antibiotic, depending on the

insert, and the rest of the medium was spread onto a second LB plate with antibiotic and both plates were left over night (O/N) to grow at 37 °C.

A distinct colony was picked the next day with a pipette tip and inserted into a tube with 2ml LB medium and the appropriate antibiotic (1 ml/L of 30 µg/ml Kanamycin or 1 ml/L of 100 µg/ml Ampicillin). The cells were allowed to grow O/N at 37 °C. The next day 1 µl from the medium was pipetted into another tube containing 5 ml of LB with antibiotic and was left to grow O/N. The rest of the 2 ml medium was made into stock cell cultures by addition of 15 % glycerol and frozen at -80 °C for future use.

The next day the cells from the 5ml LB culture were centrifuged at 4000 g for 10 minutes before removing the supernatant and treating the cell pellet with buffers from the Miniprep kit for DNA extraction (Quiagen). The protocol given with the kit was followed to finally elute 50 µl of DNA in H_2O. The transformation procedure was repeated only this time using BL21 (DE3) cells that are commonly used for protein purification and are required for expression of proteins from the T7 promoter in pET vectors. 5 µl of the DNA was taken and mixed with 10 µl BL21 (DE3) cells in a tube and put on ice for 30 minutes. The cells were heat-shocked for 45 seconds and put straight back on ice for 5 minutes before transferring into 400 µl of SOC medium and allowed to grow for 2 hours at 37 °C. 20 µl of that medium was pipetted onto a LB solid medium plate with antibiotic and let grow O/N at 37 °C. The rest of the SOC medium was spread onto a second LB plate with antibiotic and left O/N to grow at 37 °C.

The next day a colony was picked and inserted into a tube with 2ml LB medium and the appropriate antibiotic. The cells were allowed to grow for 2 hours at 37 °C. 1ml of the medium was pipetted into 1 Litre of LB medium with antibiotic in a 5 Litre flask to allow proper aeration of the culture. The flasks were then left at 37 °C at 180 rpm in a shaker incubator until Optical Density (OD) at 600 nm reached ~0.7. At this point the

culture was induced with 1 ml of 1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG). For PyLTNT the culture was left O/N to grow at 25 °C. For pRb the cells were grown for 24 hours at 18 °C. Stocks of the cells were prepared by addition of 15 % glycerol to uninduced 1 ml BL21 (DE3) cells containing the insert of interest and frozen at -80 °C.

2.2.1. Minimal Media

For production of a labelled sample the steps followed are identical to the procedure of preparing the cells for growth in LB medium, only after the two hours of growth in 2 ml LB, 1 ml of the sample was pipetted into a M9 minimal medium. To prepare the medium 6 g/L of Na_2HPO_4, 3 g/L of KH_2PO_4 and 1 g/L of NaCl were mixed in H_2O (D_2O if necessary for labelling), adjusted at pH 7.2 and autoclaved. Then 2 ml/L of a 1 M $MgSO_4$, 2 ml/L of a 50 mM $CaCl_2$, 0.4 ml/L of a 3 mM $FeSO_4$, 1 ml/L of a 20 mg/ml Thiamine and a 1 ml/L of a metal mix solution (4 mM $ZnSO_4$, 1mM $MnSO_4$, 4.7mM H_3BO_3 and 0.7mM $CuSO_4$) were added to the medium and mixed. Finally 1 g/L of ^{13}C labelled glucose and 1 g/L of ^{15}N labelled NH_4Cl were added to 10 ml H_2O before filtering and addition to the medium.

After several attempts of unsuccessful growth of E.coli cells in this medium, it became apparent that a modifications of the protocol was required. However, a variety of vitamin mixtures and even a 50 % mixture of rich CELTONE™ medium with M9 did not lead to satisfactory results. So, further production in M9 media had to be abandoned. Alternatively labelled semi-rich media Spectra 9™ and rich CELTONE™ media from Spectra Gases were used.

2.2.2. Isotopically labelled Rich Media

CELTONE™ is a uniformly labelled rich medium that gave similar results as LB media. For the production of PyLTNT protein in CELTONE™ the protocol used was identical to that used for LB media.

Spectra 9™ medium showed promising results initially in small scale cultures (100 ml) but in larger volumes (1 L) protein production was stunted. Further experimental cultures revealed that success of cell growth relied on good aeration. The culture had to be grown in custom made baffled glass flasks with a maximum medium ratio to flask volume of 1:6 and an optimum of 1:10. Furthermore, the shaker incubator was set at 37 ^{o}C at 120 rpm and cells were let to grow until an $O.D._{600}$ reached ~ 0.6 when they were induced. The cell density had to be below 0.65 before induction, to be able to collect any protein. After the induction the culture was let to grow O/N at 25 ^{o}C at 120 rpm. The next day the culture was collected and the protein purified in exactly the same manner as for the LB protein production.

2.2.3. Selective Labelling

Selective labelling schemes were used to identify resonances that were not possible to assign using traditional uniform labelling. Some experiments of this nature were tried during the course of this thesis but, unfortunately, cross-labelling to other residues was abundant. Several modifications of the protocol did not improve the results and attempts eventually discontinued.

In these experiments rich LB media were used to grow the cells, and after the cells reached mid-log phase, excess amounts of ^{15}N labelled phenylalanine (1 g/L) was solubilised directly into the medium. The idea behind this experiment is that the excess labelled phenylalanine would be more readily available to be taken up by the bacteria than the unlabelled one. The method was adapted from Englander et al. (Englander et al., 2006). However, results showed that the labelled amino acids were broken down and metabolised to aspartate and glutamate, although phenylalanine was an end product in the metabolic biochemical pathway of the biosynthesis of amino acids. This cross-labelling could be attributed to the fact that, because of the excess availability of the amino acid, it was metabolised by the bacteria and was broken down. An attempt was made to inhibit cross-labelling by addition of excess unlabelled aspartate and glutamate, at the same concentration as the labelled phenylalanine. Unfortunately, cross-labelling was still observed and the experiments were suspended.

2.3. Protein Purification

The cell culture was centrifuged in large 1 L volume bottles at 6,500 rpm for 17 minutes in the Avanti® J-20XP centrifuge with a JLA 8.1000 rotor. The supernatant was discarded and the cell pellet was resuspended with 30 ml of a buffer containing 5 mM imidazol, 50 mM potassium phosphate adjusted to pH 8.0, 150 mM KCl, 2 mM DTT, 2 % NaN_3 and 1 tablet of Complete® protease inhibitors. Cells were then lysed using a French press at 2000 psi. The lysis was repeated three times before the homogenised solution was centrifuged at 24,000 rpm for 1 hour in the Avanti® J-25 centrifuge with a JA 25.50 rotor. Alternatively, the cells were lysed using the Emulsiflex C-3™ homogenizer before centrifugation. The supernatant was obtained after centrifugation and the pellets discarded. The supernatant was filtered from debris before the first step of liquid chromatography.

The ÄKTA prime™ plus was used in a first step of purification. Two buffers differing in imidazol concentration were used for a gradient purification. Buffer A was the same as used for resuspension of cells after centrifugation without protease inhibitors. Buffer B was the same as buffer A only with 250 mM imidazol. The protein was loaded onto a HisTrap column (or GSTrap column depending on insert) and eluted with a gradient of 5 to 250 mM imidazol (or 10mM glutathione for the GSTrap column). For verification of the protein an SDS-PAGE gel was used with a protein marker. The expected protein was ~28.5 kDa. Following verification second step of purification with gel filtration was initiated.

A fast protein liquid chromatography (FPLC) ÄKTA purifier™ system using a Superdex 75 column was used for second step of purification for PyLTNT, whereas a Superdex 200 column was used for the pRb. Fractionation was at 3 ml for each tube at a flow rate of 1 ml/minute.

The buffer used for this was the same as for the ÄKTA prime™ plus but without imidazol. The column was washed prior to elution and run O/N. The protein fractions from the ÄKTA prime™ plus were concentrated to 5ml using Amicon or Vivaspin concentrators with a molecular weight cut-off (MWCO) of 10,000 Da and loaded onto the Superdex column. For both proteins the chromatograms showed both the aggregated form and monomeric form of the proteins. Confirmation of the monomeric form was done by SDS-PAGE gel using Precision plus standard markers. The confirmed fractions were concentrated to 1ml for PyLTNT and stored for further experiments. For pRb the fractions could not be concentrated without causing precipitation and aggregation. Other constructs prepared in the same way were PyLTNTm1 (amino acid residues 1-151) and PyLTNTm2 (amino acid residues 1-200) as well as J domain (amino acid residues 1-79). These three constructs were directly prepared by Brian Schaffhausen after an initial step of purification and were then further purified.

Protein concentration was determined using a Bradford assay. The red Coomasie reagent changes into a stable Coomasie blue upon biding the protein. Predetermined concentrations of protein dilutions in Bradford reagent were measured at an absorbance of 595 nm forming a linear graph. The linear graph was then used to estimate the unknown concentration of PyLTNT and pRb after diluting protein into Coomasie reagent and measuring the recorded absorbance.

Another method for protein concentration estimation was by using the Beer-Lambert equation after measuring the UV absorbance at 280 nm:

$$c = A/l\varepsilon$$

where A is the absorbance at 280 nm, c is the molar concentration, l is the cell path length (1 cm) and ε is the molar absorption coefficient of the protein in question (for PyLTNT ε= 38690 M^{-1} cm^{-1} and for pRB ε= 40680 M^{-1} cm^{-1}).

2.4. Protein Preparation for NMR Experiments

For PyLTNT expressed in LB, CELTONE™ or Spectra 9™ media, the protein was concentrated to a final volume of 600 µl containing 10 % of D_2O for the deuterium field-frequency lock. Typical NMR protein concentrations were between 0.3 to 0.8 mM. The buffers used were the same as described in section 2.3. For titration with pRb, the 600 µl PyLTNT was mixed with pRb at equimolar concentrations in the same buffer conditions, and then concentrated back to 600 µl containing 10 % D_2O.

2.5. Biochemical Characterisation of PyLTNT

In the process of preparing and analysing PyLTNT a variety of methods were used. The protein was isolated once a protocol for expression and purification was established, conditions were optimised using Thermofluor analysis studies and circular dichroism (CD) as well as analytical ultracetrifugation (AUC) and isothermal titration calorimetry (ITC). For all other experiments, the buffer for all protein samples was the same potassium phosphate buffer at pH 7.0 that was used for preparation.

2.5.1. Thermofluor

The ThermoFluor® assay was used to determine optimal buffer conditions for the PyLTNT protein and later also the pRb protein. The principle of this technique relies on a fluorescent dye (SYPRO® Orange), which gives rise to a fluorescence signal when bound to hydrophobic residues of a protein when this starts to unfold as a result of temperature elevation. Data from the assays of the fluorescent readings were normalised in Excel, and a selection of four distinct conditions were plotted. The first derivative was taken to determine the point of inflection of each melting curve. This can be considered as the half maximal unfolding fluorescence (UF_{50}).

For NUPs this method is less effective as the protein is already unfolded and the dye binds immediately to a multitude of hydrophobic residues. For partially folded proteins, the ThermoFluor® assay may work if the residual secondary structure elements unfold. The thermal limit of each protein is different and depends on the buffer conditions. ThermoFluor® is therefore often used to optimise buffer conditions for optimal protein stability, i.e. folding and the highest possible temperature.

Screening of buffer conditions involves diluting 1 µl of the protein in 18 µl of buffer and adding 1 µl of the fluorescent dye. 96 buffer conditions are picked to test in a 96 well plate. The plate is inserted in a RT-PCR machine to measure fluorescence as temperature is elevated. Since the amount of buffer conditions that can be tested is limited by the number of wells, the combination of the conditions that contributed towards the thermal stability of the protein can be used for buffer preparation. This was also the case with PyLTNT that responded well towards certain conditions but their combination proved to be more beneficial towards long term stability as realised by NMR experiments.

2.5.1. CD

CD spectropolarimetry uses circular polarised light, to provide information about secondary structure of biomolecules in solution. Circular polarised light has the property of moving in circular motion along its propagated direction. This circular motion can be either left- or right-handed. The difference in absorption, between the orientations of the circular polarised light, as it changes direction when it passes through the solution is what we observe in the spectrum. The far-UV CD spectrum of a protein reveals the fraction of the spectrum that has a particular secondary structure, but it cannot pinpoint the location of the secondary structure. Despite this limitation, it is a valuable tool to describe conformational changes of a molecule.

CD spectra were recorded in a JASCO J-810 spectropolarimeter with a 1 cm temperature controlled holder and a 1 mm pathlength quartz cell in the previously described phosphate buffer. Spectra were recorded at different pH. CD spectra were measured as previously described (Karlin et al., 2002) with number of residues, molecular mass and concentration for PyLTNT being 264, 28 kDa and 0.3 mM respectively. Data were analysed with the online available software K2D for secondary structure (Andrade et al., 1993).

2.5.2. AUC

AUC relies on real time monitoring of a sample as sedimentation occurs at large g-force. The system uses light to measure absorption and reports the sample concentration as it migrates through the sample cell. This is called a sedimentation velocity experiment and was carried out to determine the molecular weight of the protein, and its purity. It is also possible to determine the shape of a protein and its aggregation state. For the purpose of this thesis, AUC was used to determine whether the protein was in its monomeric form without forming dimers upon concentration. The sample cell and the reference cell contained the same buffer. A Beckman XL-I analytical ultracentrifuge using an 8 cell 50Ti rotor was used for the AUC studies. Sedimentation velocity experiments were carried out by centrifuging a 2 sector cell at 40,000 rpm for 17 h at 4 °C. 400 µl of the sample and 400 µl of the buffer were loaded into respective cells. The absorbance of the sample was measured at a wavelength of 280 nm throughout the cell. A total of 126 measurements were taken of the sample during the run. These data were then analyzed by applying the c(s) routine in SEDFIT (Schuck, 2004).

2.5.3. FPLC

An ÄKTA purifier™ system with an analytical Superdex 75 HR 10/30 column was used to conduct binding analysis of the the PyLTNT with pRb. The principle relied on the fact that upon complex formation a higher molecular weight would be established. This would be observed on the chromatogram as a peak eluting early in the collected fraction. PyLTNT and pRb were run separately and also, when mixed in equimolar amounts. The molar concentration of each protein was 0.5mM. The buffer used for the analysis was the same that was used for purification (pH 7.0). The experiment was run at a flow of 2ml/min and fractions were collected and run on SDS-PAGE for confirmation and analysis.

2.5.4. ITC

ITC is commonly used to determine thermodynamic parameters of interactions. It measures binding affinities (K_a), enthalpy changes (ΔH) and binding stochiometry (n) of the interaction between two or more molecules in solution. From these measurements Gibbs energy changes (ΔG) and entropy changes (ΔS) can be determined using the relationship:

$$\Delta G = -RT\ln K = \Delta H - T\Delta S$$

where R is the gas constant ,T is the absolute temperature, and K is the equilibrium constant.

In practice, a known amount of the ligand molecule is titrated into a sample cell where the protein is in solution at a known concentration. The temperature of the sample cell is constant. When the ligand is added the heat is measured in relation to a reference cell. This heat is actually the time dependent power that is required to maintain the temperature of the sample cell in respect to the reference cell.

This results in a series of curves, with every curve corresponding to the heat exchange during one titration step. The pattern of these curves yield the thermodynamic parameters of the interaction as a function of the molar ratio.

Experiments were carried out using a MicroCal VP-ITC instrument (MicroCal, UK) at 25 ^{o}C. For the PyLTNT-pRb ITC experiment, 5 µM of the PyLTNT protein was equilibrated in the sample cell. A 100 µM pRb protein was then titrated into the sample cell using 25 titration points, each point of 10 µl volume. Similarly, for the PyLTNTm2-pRb interaction, a 100 µM concentrated pRb protein was titrated using 25 titration points, each point being 10µl in volume into a sample cell containing 10

µM of PyLTNTm2. Results were processed using the software Origin and curves were fitted using over 100 iterations with a one set of sites simulation. The first point of the titration was removed as were some values along the final points of titration for better curve fitting. As a heat reference pRb protein was titrated into buffer and the heat reference was subtracted from the binding experiments.

2.6. NMR Instrumentation

Spectra were recorded using the following spectrometers: a Varian INOVA 600MHz with a HCN 5mm z-PFG (2001) probe, a Varian INOVA 900MHz with an HCN 5mm z-PFG cryogenically cooled probe, a Varian INOVA 800MHz with HCN 5mm cryogenically cooled z-PFG probe, and a Bruker DRX AVANCE 500MHz with HCN 5mm cryogenically cooled z-PFG probe. For the ^{13}C observed experiments that were recorded at the CERM magnetic resonance centre, a Bruker 700MHz with a TXO cryogenically cooled probe optimised for ^{13}C direct detection was used. All experiments were carried out at 25^{o}C, using standard pulse sequences from the Biopack (Varian, Inc.) or Bruker pulse sequence libraries. Internal referencing of chemical shifts was achieved using 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) for the absolute value of protons, and nitrogen and carbon were referenced indirectly using conversion factors derived from ratios of NMR frequencies (Wishart et al., 1995).

2.7. Data Analysis

NMR spectra were initially processed using either the NMRLab (Günther et al., 2000) or NMRpipe (Delaglio et al., 1995) software. Processing involved phase correction, addition of solvent filter and linear prediction. Manual and semi-automatic assignment of resonances was carried out using the CcpNmr Analysis (Vranken et al., 2005) software. The ^{1H}, ^{13}C and ^{15}N chemical shifts of PyLTNT were deposited in the BioMagResBank (http:\\www.bmrb.wisc.edu) under accession number 16044 (Knoblich et al., 2009).

2.8. NMR Experiments

For protein analysis by NMR a variety of experiments were conducted to obtain assignment of NMR signals. Some experiments were designed to work better in unfolded parts of the protein (e.g. HNN described in section 2.8.6.), others worked better on the folded part, the J domain. Experiments were optimised using one- and two-dimensional slices and projections. In the following sections NMR experiments used in this thesis will be summarised, with appropriate representative spectra given where necessary to further understand correlations of nuclei. Experimental parameters for protein dynamics are described in individual section. All other experimental parameters can be found in the appendix (Table 2).

2.8.1. 1D

One-dimensional proton experiments can provide limited information on the nature of folding or the aggregation state of a protein (Rehm et al., 2002). They were used mainly for initial tests in the optimisation of the purification protocol, whereas multi-dimensional experiments are required for resonance assignments and structure determination by NMR. One-dimensional experiments usually consist of a simple 90o pulse followed by signal acquisition. For efficient water signal suppression a number of additions to the pulse sequence can be used. The commonly used presaturation pulse sequence (Hoult, 1976) added at the beginning of a pulse sequence, provides reasonably good water signal suppression. This particular pulse sequence was used primarily to calibrate basic setup. Alternatively excitation sculpting was employed where the water resonance is selectively excited (Hwang and Shaka, 1995).

2.8.2. HSQC

Heteronuclear single quantum coherence (HSQC) spectra are primarily used for secondary structure assessment and for the optimisation of solvent conditions once labelled proteins are available. HSQC spectra can also be used in titration experiments with a secondary ligand to obtain information about binding sites and kinetics.

This pulse sequence is two-dimensional with the first dimension for the proton frequencies while its second dimension, for proteins, gives nitrogen or carbon frequencies. The cross-peaks observed in a ^{1}H-^{15}N HSQC spectrum of a protein represent the amide signals of each individual amino acid residue with the exception of prolines that do not have a proton attached to the amine. The dispersion of backbone amide signals allows an estimation of the number of resonances visible in the spectrum and therefore potential assignability. An example of the HSQC of PyLTNT is depicted in section 3.4.1.

2.8.3. NOESY-HSQC/TOCSY-HSQC

Three-dimensional NOESY experiments were used to identify the number of homonuclear ^{1}H-^{1}H NOEs signals for the protein sample. The NOESY experiment provides useful structural information in the form of ^{1}H-^{1}H distances thereby confirming the presence of a secondary structure. A shared mutual dipolar longitudinal relaxation is what causes cross-peaks in NOESY spectra. This dipolar coupling represents a direct, magnetic through-space interaction.

Since their discovery, NOESY experiments were used for structure determination based on the inter- and intra-residue proton-proton distances. The NOESY experiment allows the assignment of homonuclear ^{1}H-^{1}H NOEs “through-space”. It yields correlations between the amide protons and other protons located within 4-5 Å.

The amide proton chemical shifts are commonly detected in the direct dimension of the HSQC experiment, while the amide nitrogen chemical shifts and the chemical shifts of the hydrogens nearby in space are detected in the two indirect dimensions. Secondary structure information is readily available from this experiment as, for instance, α-helices have NOE cross-peaks between consecutive amide protons. This aids sequential assignment, mainly by NH-CH^{α} to the preceding amino acid. Additionally, structure restraints can be obtained for structure calculations. Longer range cross-peaks are seen for sequential amino acid residues in α-helices and across β-sheets. Long-range cross peaks across the protein are observed for proteins with tertiary structure.

For PyLTNT the three-dimensional NOESY experiment provided little structural information but helped to validate the unfolded nature of the protein as few NOESY

signals were observed confirming the lack of structure as already suggested by limited dispersion in HN frequencies.

The three-dimensional TOCSY experiment is a “through-bond” experiment, and was used to obtain H^{α} chemical shifts, which are valuable for the chemical shift index prediction (section 3.4.1.), which in turn is used to predict secondary structure. This experiment provides similar results to a three-dimensional NOESY but only gives intra-residue information. The three-dimensional TOCSY experiment uses an isotropic mixing scheme to transfer magnetisation between scalar coupled proton spins within amino acid residues followed by a heteronuclear INEPT transfer employing the H-N amide J coupling to link the resonances to the backbone HN shifts.

The experiments described so far can be performed using ^{15}N labelled samples. For small proteins the combination of the NOESY-HSQC and the TOCSY-HSQC experiments is often sufficient for complete assignments. However, for larger proteins, it is necessary to use both ^{13}C and ^{15}N labelling. Experiments that make use of doubly labelled protein samples are called triple resonance experiments, as opposed to the double resonance experiments that were described thus far. The experiments in the following sections are based on doubly labelled samples and are triple resonance experiments.

2.8.4. HNCO

The HNCO is a commonly used three-dimensional experiment that correlates backbone connectivities between the H-N pair of one residue (*i*) with the carbonyl (C') resonance of the preceding residue (*i-1*) (Figure 9). This provides a good indication of the number of expected resonances for the assignment process and also gives crucial sequential backbone connectivities. The experiment was initially described by Kay et al. (Kay et al., 1990). It is used primarily because it is the most sensitive of the heteronuclear three-dimensional experiments and provided useful data for the sequential assignment of PyLTNT.

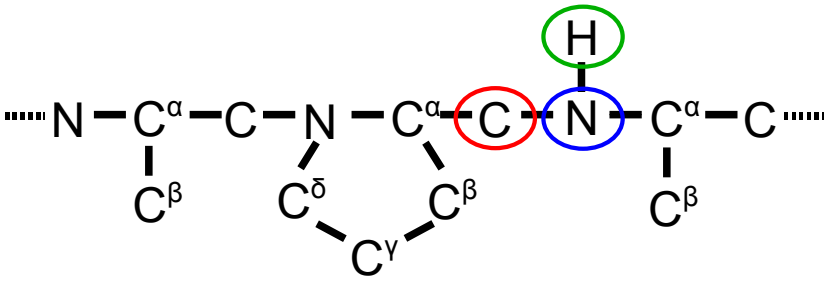


Figure 9. HNCO observable signals

The HNCO correlates the amide NH of the i residue with the C' of the i-1 residue. In green is the proton dimension in blue is the nitrogen dimension and in red is the carbon dimension of the spectrum.

2.8.5. HNCA/HN(CO)CA

The HNCA experiment was described together with the HN(CO)CA experiment by Bax & Ikura (Bax and Ikura, 1991). These two experiments complement each other and are commonly used to obtain sequential assignments. HNCA in particular, is a highly sensitive experiment giving information about the C^{α} in the *i* and *i-1* position whereas HN(CO)CA gives information on the individual spin system at the *i-1* position of the C^{α} (Figure 10). Both experiments are limited for usage with NUPs, mainly originating from a) spectral overlap and b) unfortunately fast C^{α} relaxation in folded/unfolded regions.

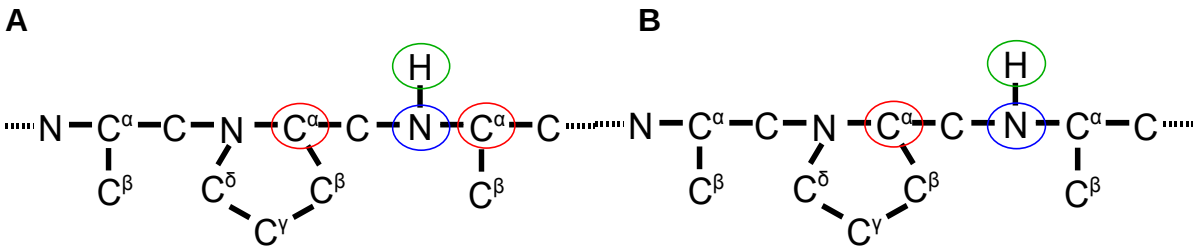


Figure 10. HNCA and HNCOCA observable signals

A) HNCA and B) HNCOCA signals observable in the spectra. Green is for the proton signal in the first dimension while blue is for the nitrogen signal in the second dimension and red is for the carbon signal/s observed in the third dimension.

2.8.6. HNN/HN(C)N

For unfolded proteins it is important to use pulse sequences that give good dispersion for assignment. The HNN and HN(C)N experiments described by Panchal et al. take advantage of the good chemical shift dispersion of the ^{15}N dimension (Panchal et al., 2001).

The HNN sequence is a three-dimensional experiment that correlates the proton dimension with the nitrogen dimensions of *i*, *i-1* and the succeeding (*i+1*) residues. Connections are then established through the backbone amides creating amino acid sequence dependent patterns due to the different signs of the self- and cross-peaks. These patterns can be tabulated as published by Panchal et al. (appendix; Figure 50) to include different cases of triplet amino acid sequences (Panchal et al., 2001). Special cases include the glycine residue, that does not have a C^{β} and therefore has a different evolution of magnetisation components, and the proline residue that does not have an exchangeable proton. Proline residues deprive the triplet spectrum from one resonance while glycine signals appear negative on the spectrum when in the i position.

While other traditional backbone assignment experiments would be interrupted in their sequential assignment whenever a proline is encountered, these experiments can use the prolines as a reference point to provide a start. They can also use the glycines as a reference for triplet identification as can be seen in Figure 11.

Modified HNN pulse sequences that use alanines and serines-threonines as checkpoints for easier assignment have been proposed (Chatterjee et al., 2006; Chugh et al., 2008). These pulse sequences use the distinct C^{α} of alanines to create different patterns of peaks in the same way as the C^{α} for glycines, only this time by using selective decoupling. This can be useful when there are no glycines in long

stretches in the sequence. The same applies for the C^{β} chemical shifts of serine and threonine which are considerably downfield compared to all other amino acids.

These pulse sequences are less sensitive than the common HNCA and HNCO experiments, but provide valuable information for unfolded proteins. Finally, the HN(C)N pulse sequence is selective for *i* and *i-1* amide residues, and can thereby provide, together with the HNN sequence an understanding of whether correlating amides belong to the *i-1* or *i+1* residue.

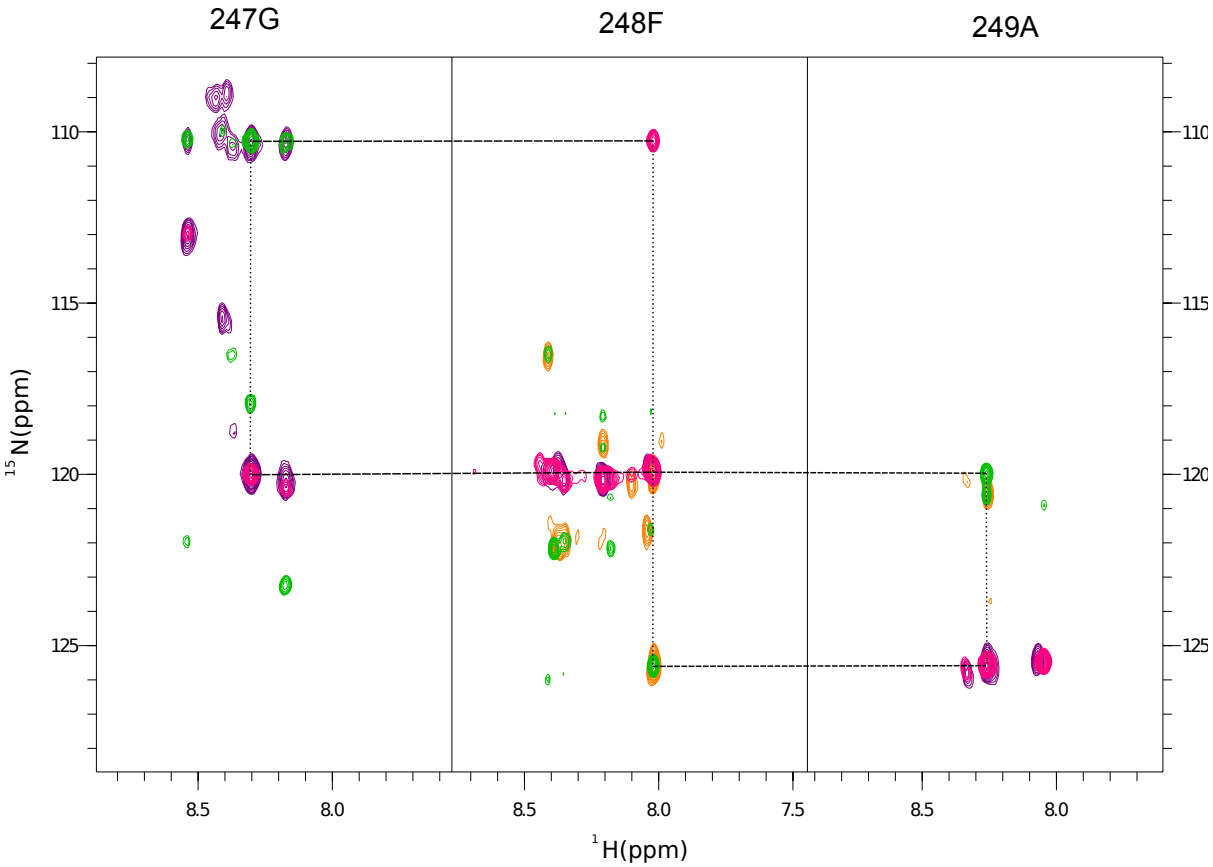


Figure 11. HNN/HN(C)N sequential assignment of PyLTNT

The sequential assignment was possible by superimposition of the two experiments. Colours for the HNN experiment are coloured pink for positive and green for negative signals whereas for the HNCN experiment purple is used for positive and yellow is for negative signals. The three residues shown here are connected via the cross-peaks of their amide nitrogen.

2.8.7. CBCA(CO)NH/HNCACB

The CBCA(CO)NH experiment was first described by Grzesiek and Bax in 1992 (Grzesiek et al., 1992). The CBCA(CO)NH pulse sequences correlates the C^{α} and C^{β} of the preceding residue with the amide H-N region. The CBCA(CO)NH experiment is perfectly supplemented by the HNCACB experiment (Wittekind, 1992) that correlates the H-N amide resonances with those of the *i* and *i-1* residue C^{α} and C^{β}. The HNCACB experiment is in some cases more sensitive than the analogous CBCANH experiment (Grzesiek et al., 1992), and preferable when proteins with short T_2 relaxation times are examined. For the unfolded part of the PyLTNT protein these pulse sequences gave superior results.

The advantage of these experiments is their sensitivity and the fact that they allow for partial side chain information. Which aids spin system identification and also allows better separation of resonances in crowded regions in the process of sequential assignments (Figure 12).

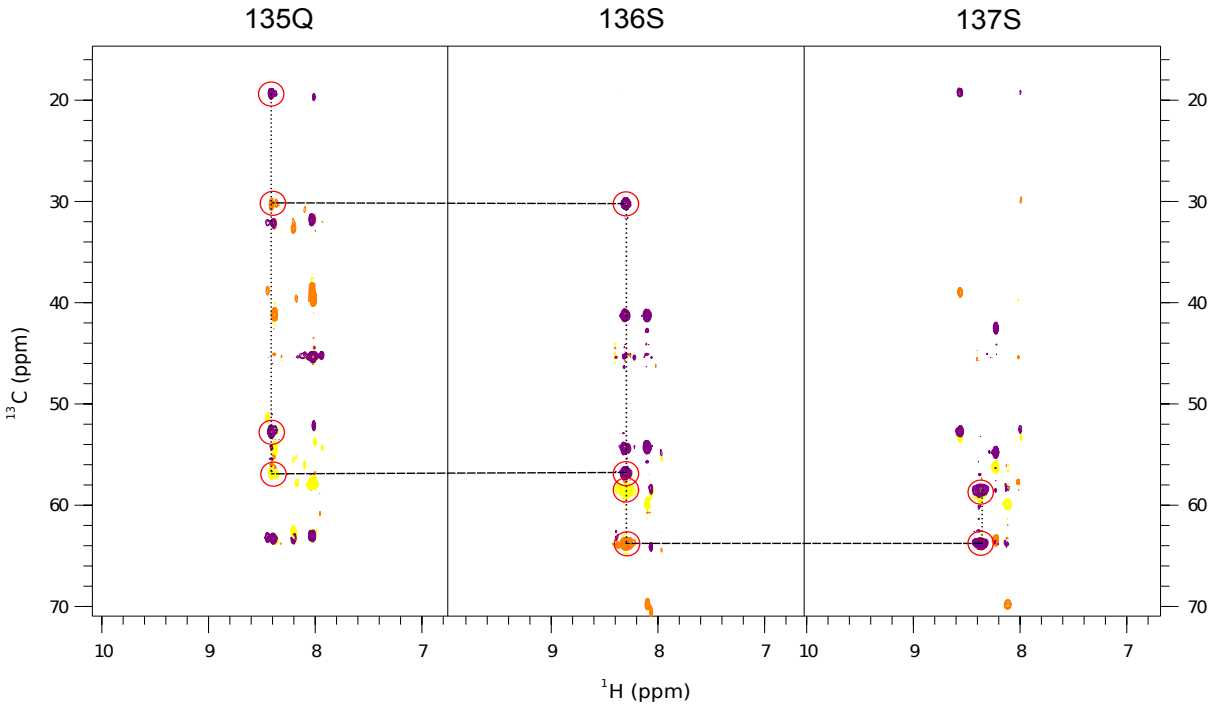


Figure 12. CBCACONH and HNCACB superimposed for sequential assignment of amino acid residues of PyLTNT

Purple coloured resonances belong to the CBCACONH spectrum and indicate the i-1 position of C^{α} and C^{β}. Yellow and orange coloured resonances indicate the i and i-1 positions of C^{α} and C^{β} in the HNCACB spectrum. For clarity resonances of each spin system have been circled in red.

2.8.8. CON/(H)CBCACON/(H)CBCANCO

These protonless experiments avoid to record protons in the direct or indirect dimension and instead use ^{13}C for direct detection. Protonless experiments have several advantages over proton based experiments because, i) they allow for sequential assignments even in proline and glycine rich sequences, ii) they are less sensitive to exchange owing to the lower frequency dispersion of ^{13}C compared to ^{1}H, and iii) they can be particularly useful for NUPs and provide better carbonyl chemical shift dispersion. All these advantages were key for the assignment of the PyLTNT protein, as it has a proline and glycine rich amino acid sequence.

The experiments used in this thesis were described in a recent publication (Bermel et al., 2009) and achieve a considerable advantage in sensitivity over previous protonless experiments, by using the proton polarisation as a starting point in the magnetisation transfer (Bermel et al., 2009), thereby reducing the experimental time. Such experiments also require a specific cryogenically cooled ^{13}C observe probe with enhanced sensitivity.

While an HSQC experiment is commonly used to evaluate the sample quality and can usually visualise most resonances, it is particularly inefficient in highly mobile and solvent accessible residues as well as in NUPs (Bermel et al., 2006), because the chemical shift dispersion in the proton dimension in NUPs is small. The two-dimensional CON experiment records only the N dimension of the *i* residue with the correlating C' of the *i-1* residue. This allows proline residues to be visualised, and provides better signal dispersion making it ideal for NUPs (Figure 13).

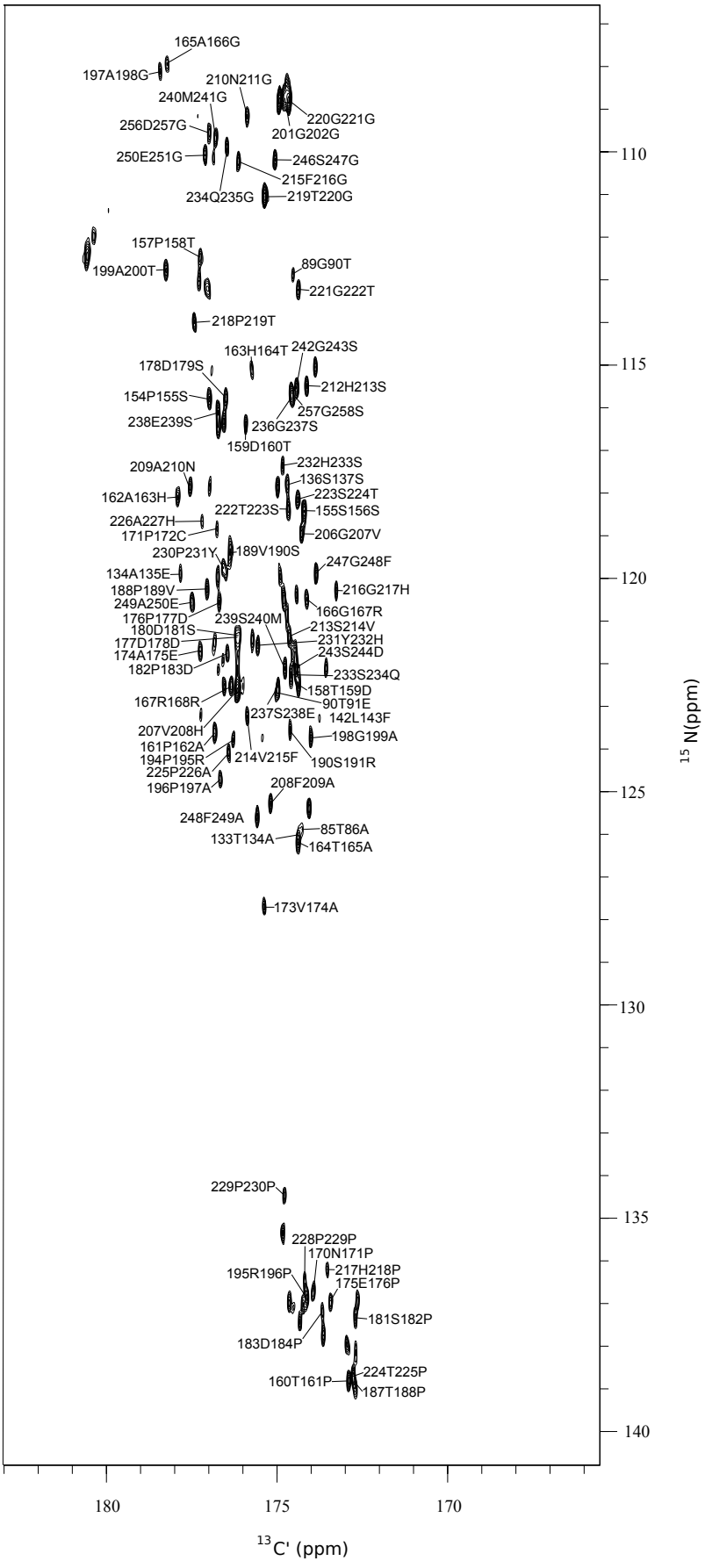


Figure 13. CON spectrum of PyLTNT

Resonances labelled according to amino acid position and type. The correlation of this spectrum provides the nitrogen resonance of residue i and the connecting C' resonance in position i-1. With kind permission from Springer Science+Business Media: (Knoblich et al., 2009)

The two heteronuclear 3D experiments (H)CBCACON and (H)CBCANCO benefit from sensitivity enhancement because they start from a proton polarisation. (H)CBCACON has three dimensions for $^{13}C'$, ^{13}C and ^{15}N and so does (H)CBCANCO, but the latter allows only for sequential linkage of the spin systems identified with the first experiment by encoding information of the $C^{\alpha}iN_{i+1}$, C'_i, $C^{\alpha}_{i+1}N_{i+1}$, C'_i and the C^{β}_i, C^{β}_{i+1} chemical shifts (Figure 14). The experiments are all in-phase anti-phase (IPAP) encoded to suppress the 55Hz C' line splitting arising from the C^{α} - C' scalar coupling (Bermel et al., 2005). All of these experiments together provide a method for a complete backbone assignment.

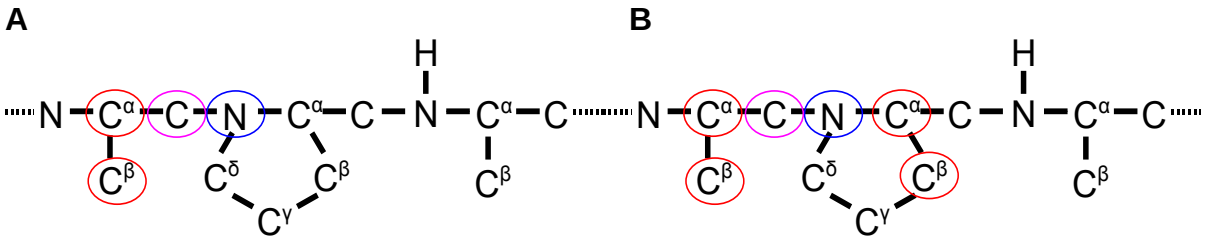


Figure 14. (H)CBCACON and (H)CBCANCO observable signals

A) (H)CBCACON and B) (H)CBCANCO observable signals in the spectra. Red is for the C^{α} and C^{β} observable signals in the first dimension while blue and pink are for the nitrogen and carbonyl signals in the other two dimensions respectively.

For the protonless experiments the samples were freeze-dried and then resuspended in 300µl of H_2O with 10% D_2O to increase concentration. The reduced

sample depth was then compensated by the use of a Shigemi tube, that has susceptibility matched glass beneath the level of the sample and above it to optimise the filling factor of the NMR coil.

2.8.9. Dynamics

Relaxation parameters related to residue specific dynamics have provided a wealth of information on localised structure formation. Relaxation parameters can detect dynamics of different time scales. NMR relaxation has been used as a powerful method to study and observe reduced conformational mobility due to binding another protein.

In proteins, relaxation rates, both longitudinal and transverse (R_1 and R_2) of backbone nitrogen nuclei and ^{1}H-^{15}N steady-state heteronuclear NOEs, give information about backbone dynamics and rotational tumbling of the molecule. The hetNOE experiment is primarily sensitive towards high frequency motions (ns-ps) (Claridge, 1999). However, transverse relaxation is also sensitive towards slow time scale motions (µs-ms), and conformational exchange.

Negative values observed in a hetNOE spectrum are indicative of large motional displacement on a sub-nanosecond time scale, which is likely in unfolded parts of proteins (Chatterjee et al., 2005). In accordance, conformational exchange contributions can be indicated by high R_2 values (Chatterjee et al., 2005). Mapping of reduced motions in the protein-protein interface can be achieved this way by comparing the bound and unbound relaxation rates. The difference of these rates should provide information on the interacting regions since slow exchange residues are affected the most.

The usefulness of the hetNOE is apparent but the drawback of this pulse sequence is that it starts by using the nitrogen magnetisation instead of the ten-fold stronger proton magnetisation which makes it more insensitive for detection than the other relaxation experiments that use protons as a starting point (Renner et al., 2002).

For PyLTNT hetNOE consisted of 5 seconds of saturation which was equal to the delay time used. The number of scans was 8 and with 1024x128 complex data points in t_1 and t_2 respectively. Two spectra were recorded. One with 5 seconds of saturation between scans and the other with five seconds of delay between them. The ratio of resonance intensity of the saturated state to the non saturated state were calculated. Errors were estimated by evaluating the standard deviation of the NOE, σ_{NOE}:

$$\sigma_{NOE}/NOE = ((\sigma I_{sat}/I_{sat})^2 + (\sigma I_{unsat}/I_{unsat})^2)^{1/2}$$

where σI_{sat} and σI_{unsat} are the standard deviations of the noise in the spectra.

R_2 relaxation informs on on residual structure of a protein. When proteins bind to each other it can lead to formation of secondary structure elements, and thereby tumbling rates which are slower on different time scales. This is sensed by the change in the R_2 relaxation rate. Therefore relaxation rates of PyLTNT were measured both in its free state and when bound to pRb.

^{15}N T_2 relaxation experiments of PyLTNT were acquired with the following ms relaxation delay times: 10, 30, 50, 70, 90, 110, 150, 190, 210 and 250 with 3 repeat points for error estimation. Each relaxation experiment was recorded with 8 scans with 1024x128 complex data points in t_1 and t_2 respectively. R_2 relaxation rate values were fitted to the equation:

$$I = I_o exp(-t/T_2)$$

where T_2 is the relaxation time, I is the resonance intensity at relaxation delay time t and I_o is the starting intensity. Curve fitting was then done using in-house scripts and resulting values were plotted.

2.9. SAXS

Small angle X-ray spectroscopy (SAXS) is a powerful method that can be used to elucidate protein dimensions. It can reveal information on the state of the protein at low resolution between 50 and 10 Å but unlike NMR and electron microscopy it is not restricted to size (Putnam et al., 2007). In combination with other techniques it has been used to model structures of macromolecules where other methods have failed. The scattering pattern recorded by SAXS is indicative of the conformation and its size of the protein, so that the shape of the protein can be determined *ab-initio*.

SAXS measures the intensity and the scattered X-rays after irradiation with X-rays. Scattering profiles must be obtained for the buffer to be subtracted from those of the protein (Tsutakawa et al., 2007). Unlike X-ray crystallography, SAXS is used for proteins in solution where the signals of all orientations of the sample are averaged giving an isotropic scattering. Although SAXS yields lower resolution structures they represent the native state unlike crystallography that might fix the structural conformation in a non-native form (Putnam et al., 2007).

Small angle X-ray scattering was recorded at the EMBL facility in Hamburg using the DORIS beam line. Samples were of varying concentrations between 1-7mg/ml and each sample was recorded with four exposures at three different concentrations with each exposure time being 30s. All samples were dialysed over night in the same phosphate buffer used for NMR preparation (see above).

SAXS data were processed using the ATSAS 2.2 software, provided by the biological small angle scattering group at EMBL Hamburg. More specifically, data were initially analysed with the software PRIMUS to perform averaging, merging and curve fitting. Secondly, the files generated from PRIMUS were imported into the software GNOM for regularisation and then models were generated using the software DAMMIF.

GNOM output was also exported to Excel to obtain plots of scaterring intensity, *P(R)* function, Kratky, and Guinier. The models generated with DAMMIF were averaged using the software DAMAVER. This averaging may not always be desirable as it can generate more globular models but has the advantage to allow for better fitting of existing structures as will be shown in section 3.5.

CHAPTER III - Results

3.1. Structure Prediction

Secondary structure prediction is a basic step in understanding notable characteristics about proteins. Bioinformatics aids biophysical analysis by predicting basic properties of proteins, thus guiding further analysis. For this thesis secondary structure prediction was used to assess the protein structure before and after a more thorough investigation. Bioinformatics software predicted the structured and unstructured regions of the PyLTNT protein from its amino acid sequence. NMR provided similar results to the calculated prediction. Further on, structure prediction helped to resolve discrepancies about PyLTNT and its ordered regions.

A variety of secondary structure prediction software were used. The reliability of the predictions was later judged upon comparison with NMR results. The specific software packages were PSIPRED (Bryson et al., 2005), PONDR (Iakoucheva et al., 2002), Globplot (Linding et al., 2003), FoldIndex (Prilusky et al., 2005), jPred (Cuff et al., 1998) and PreLink (Coeytaux and Poupon, 2005). With the exception of PSIPRED, that is used for predicting ordered secondary structure, these software are based on different algorithms to identify disordered regions in a protein sequence. PSIPRED was used initially to predict a global secondary structure of PyLTNT, and upon predicting very few structured elements, the other algorithms were compared to confirm the disordered regions. FoldIndex (Figure 15) and Globplot (Figure 16) gave the most consistent results for the PyLTNT protein when compared to the NMR data.

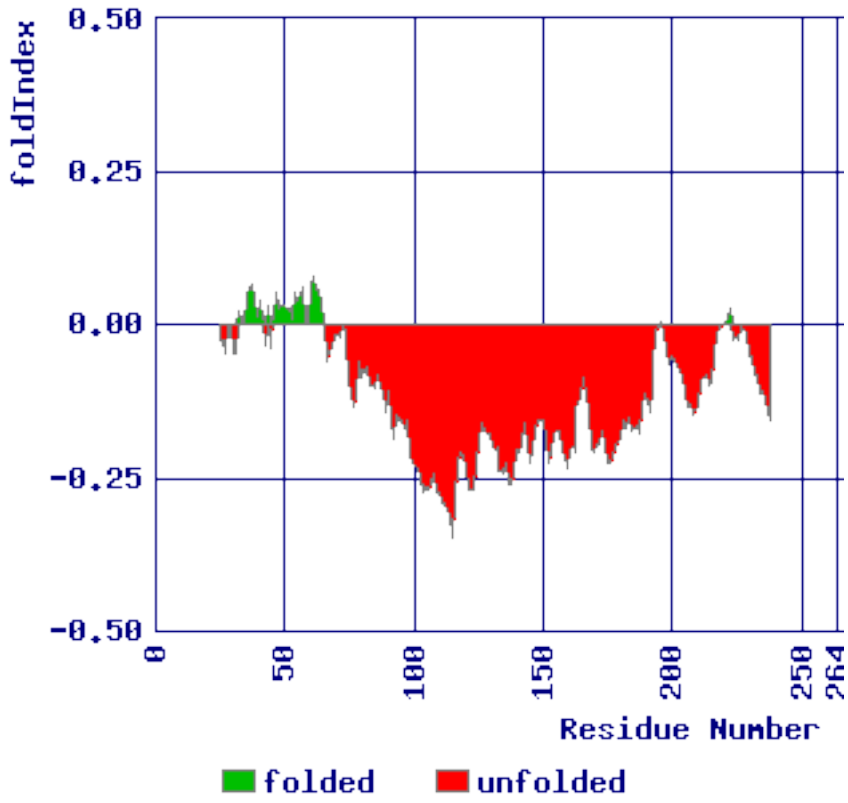


Figure 15. FoldIndex prediction of disordered regions

The y-axis shows the unfoldability. The C-terminal region after the J-domain is predicted to be completely unfolded. Minor structural elements might exist around residue 220 but NMR data did not confirm this.

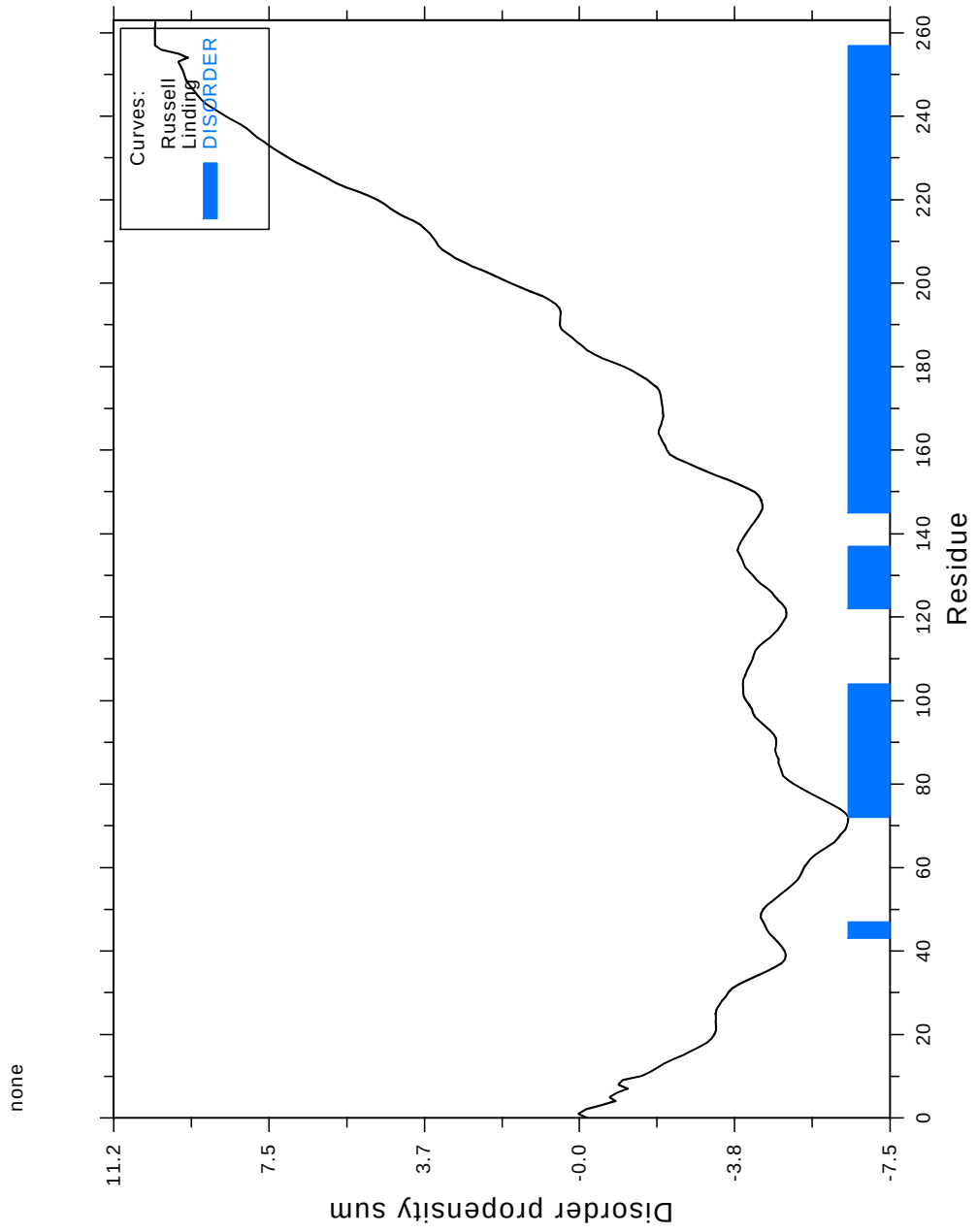


Figure 16. Globplot prediction of disordered regions

The regions predicted here to be unfolded are similar to the FoldIndex result. The tendency for disorder is large at the C-terminus of the protein. The blue regions

represent predicted disordered regions based on amino acid composition. The black line represents the disorder propensity sum which decreases because of the predicted order in the J domain region, and increases because of the predicted disorder at the C-terminus.

PyLTNT is proline (11 %) and glycine (12 %) rich and is known to interact with multiple targets giving it a high propensity towards an unfolded state (Receveur-Bréchot et al., 2006). Similarly the E7 protein has been predicted as being unfolded in its N-terminal region although it contains some functionally important interacting domains (Uversky et al., 2006). The putative purpose of the unfolded nature of the protein is its availability of short regions for multiple interactions.

3.2. Biochemical Methods

The following sections include results of the biochemical investigations carried out for the PyLTNT and the pRb proteins, in order to optimise buffer conditions for NMR analysis and to confirm the monomeric form of the protein.

3.2.1. Expression & Purification

Chromatographic separation of the two constructs of PyLTNT, one with a glutathione-S-transferase (GST) tag, and the other with a histidine tag (Histag), showed a low yield of the GST version when compared to the Histag version (Figure 17). Therefore, further experiments were conducted using the Histag protein that led to higher yields.

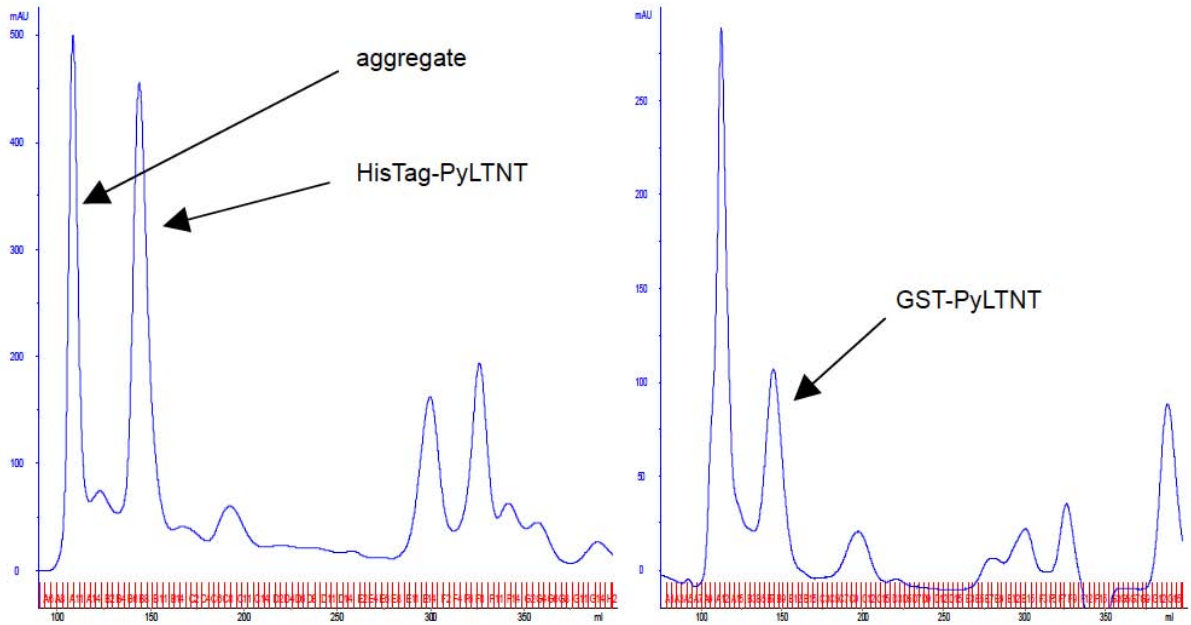


Figure 17. PyLTNT chromatograms

Absorbance vs elution volume are plotted in the two chromatograms for the two constructs. Left chromatogram shows PyLTNT with a HisTag; the right chromatogram shows PyLTNT with an N-terminal GST. Both proteins were purified using a Superdex 75 column on an ÄKTA purifier. The monomeric and aggregated forms of the proteins are labelled on the chromatogram.

The choice of isotopically labelled media was made after tests of protein expression with 100ml cultures (Figure 18). Among the screened lysed cultures, the most suitable media for expression was Spectra 9™ that gave similar yields to LB media. The optimised purification protocol gave ~20mg of PyLTNT protein suitable for NMR from 1L of culture. Similar amounts of pRb were obtained from the same volume of culture. PyLTNTm2 gave similar results while PyLTNTm1 had a very low expression with a total of only 8mg of protein out of 3L of culture. Only PyLTNT, pRb and PyLTNTm2 were used for NMR and other experiments, PyLTNTm1 construct prepared in this thesis were not sufficiently stable.

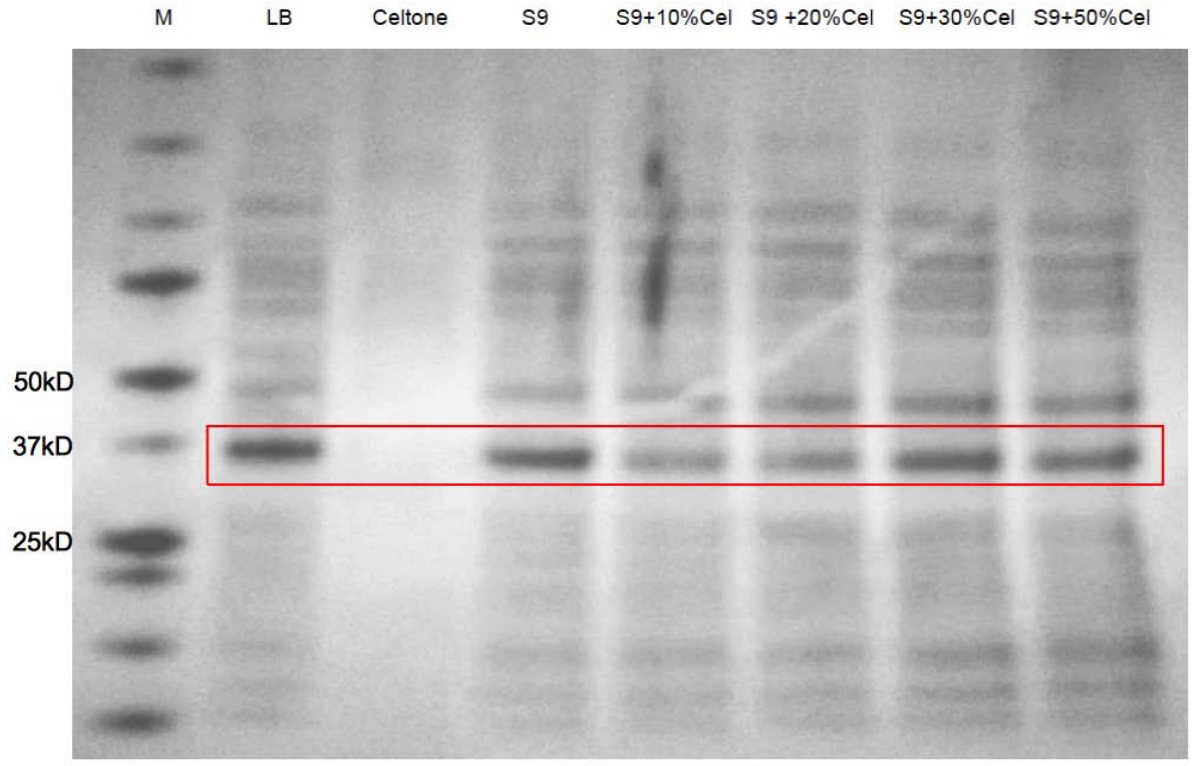


Figure 18. Comparison of media for expression of PyLTNT

100ml cultures were lysed by sonication and spun down. 20 µl of each of the supernatant was then used for SDS-PAGE and compared for different protein expression. The results show that Spectra 9™ medium (S9) on its own is sufficient to obtain the same level of expression as with the LB medium. The PyLTNT protein is

boxed in red. Addition of CELTONE™ proved advantageous only at higher concentrations.

3.2.2. Thermofluor

ThermoFluor® results showed that PyLTNT has a high thermal stability with highest melting point at ~62°C (Figure 19) as expected for a partially unfolded protein. This is similar to the DnaJ domain, although a folded protein, which has previously been found to have a melting point of 75°C (Szyperski et al., 1994). This suggests that the extension beyond the J domain does not substantially reduce the thermal stability of PyLTNT. This high thermal stability allows to perform NMR experiments at higher temperatures if desired. Precipitation over prolonged times is another problem at elevated temperatures which is not necessarily related to the thermal stability of the protein.

Variation of buffer conditions showed little difference, the only exception was that of higher salt concentrations while the zwitterionic detergent 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS) seemed to destabilise the protein. The detergent proved to promote cleavage of the PyLTNT protein. The ThermoFluor® assay showed that a buffer with a high salt concentration (150mM) and no other additives suffices for PyLTNT to remain stable without precipitating too quickly out of solution. NMR samples were typically stable over 2-3 weeks. The selection of buffer as previously mentioned for all subsequent experiments was a 50 mM potassium phosphate buffer with 150 mM KCl, 2 mM DTT and 2 % NaN_3. Protease inhibitors were always included.

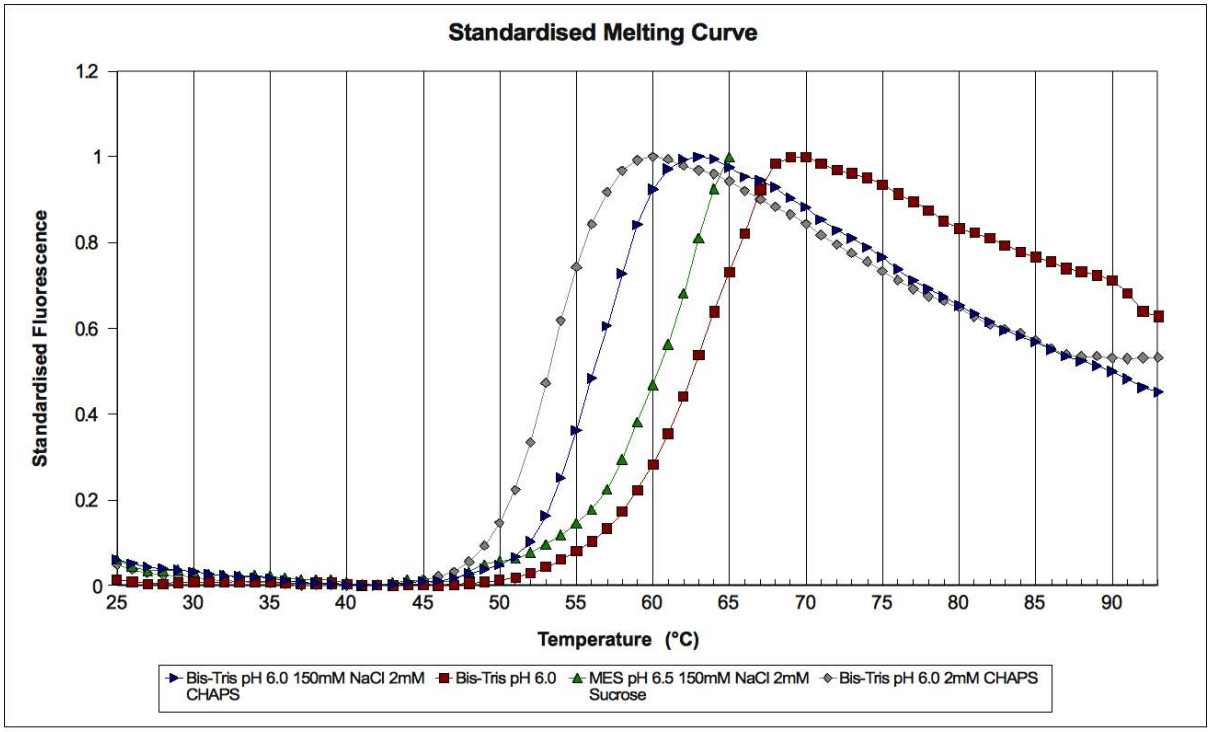


Figure 19. Thermofluor screening of PyLTNT

A variety of conditions were tested for optimum solubility and thermal stability. The averaged results showed that best stability could be achieved at pH ~6.0 in combination with a high salt concentration (150mM) when no additive is present (e.g. CHAPS). Here the comparison of 4 out of the 96 well conditions is made to show the change of thermal stability. The inflection point of each curve indicates the value of interest. The conditions for each curve are described below the graph. NMR screening tests showed that phosphate buffer at pH 7.0 sufficed for long term stability.

Thermofluor screening was also performed on the pRb protein, to determine if a different buffer than the one used for PyLTNT would be beneficial. However as shown in Figure 20, the best conditions were the same as for PyLTNT. This simplified conditions for the analysis of the complex, especially for titrations without the need for dialysis in each step.

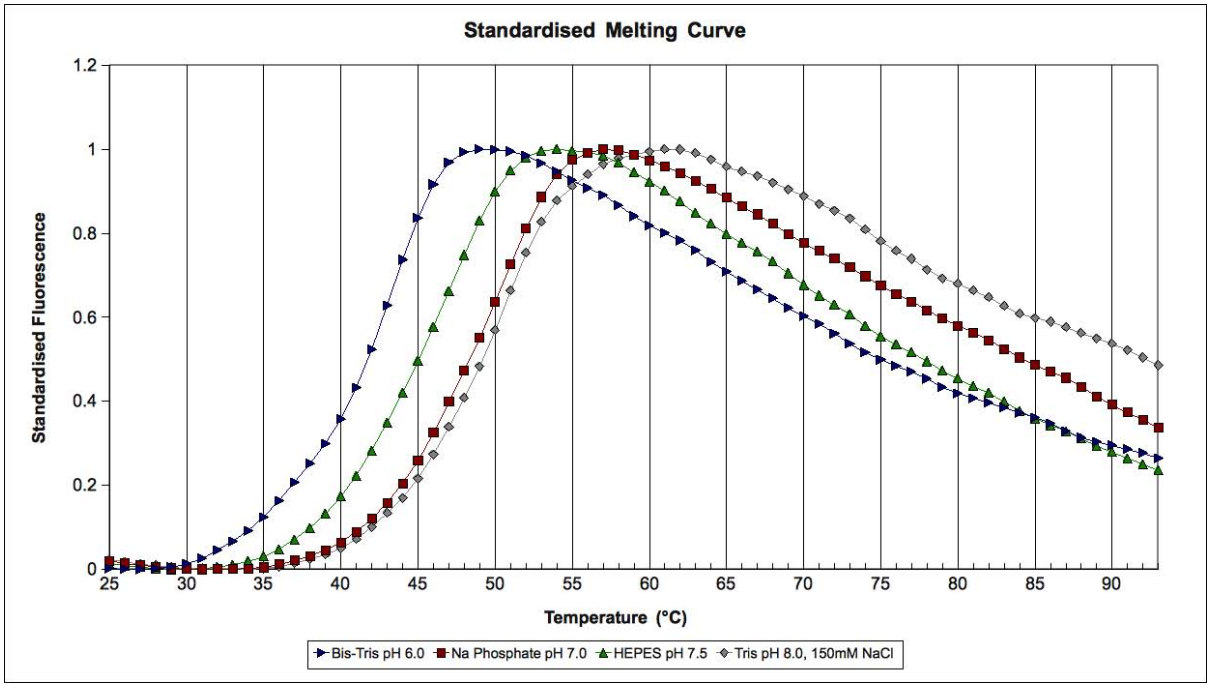


Figure 20. Thermofluor screening of pRb

A variety of conditions were tested for thermal stability. Best conditions were determined to be phosphate buffer at pH 7.0 and high salt concentration. This was conveniently also the choice of buffer for PyLTNT. A selection of four conditions that contributed towards buffer selection is shown here. Each condition is described underneath the graph.

3.2.3. CD

At any pH, PyLTNT showed a characteristic spectrum for a disordered protein with a large minimum around 200nm and a negative shoulder in the 222nm region indicative of residual ordered secondary structure (Figure 21). These spectra mirror results found for the E7 protein (Uversky et al., 2006).

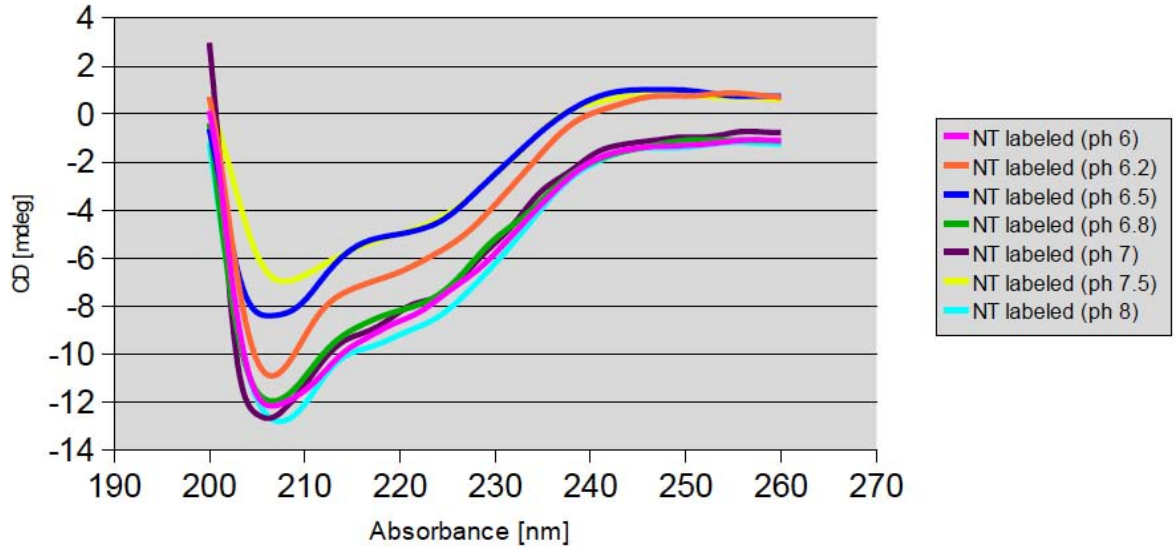


Figure 21. CD spectra of PyLTNT measured at different pH values

Different colours represent different pH values. Varying the pH of the buffer did not alter the secondary structure of the protein. This indicates that the unfolded nature of the protein is not pH dependent.

The CD spectrum as seen in Figure 22 is typical of NUPs, with a small structural element, the J domain. A negative ellipticity at 210nm indicates that the protein is largely unfolded. A small secondary “hump” at ~220nm is indicative of a structural element and is attributed to the J domain.

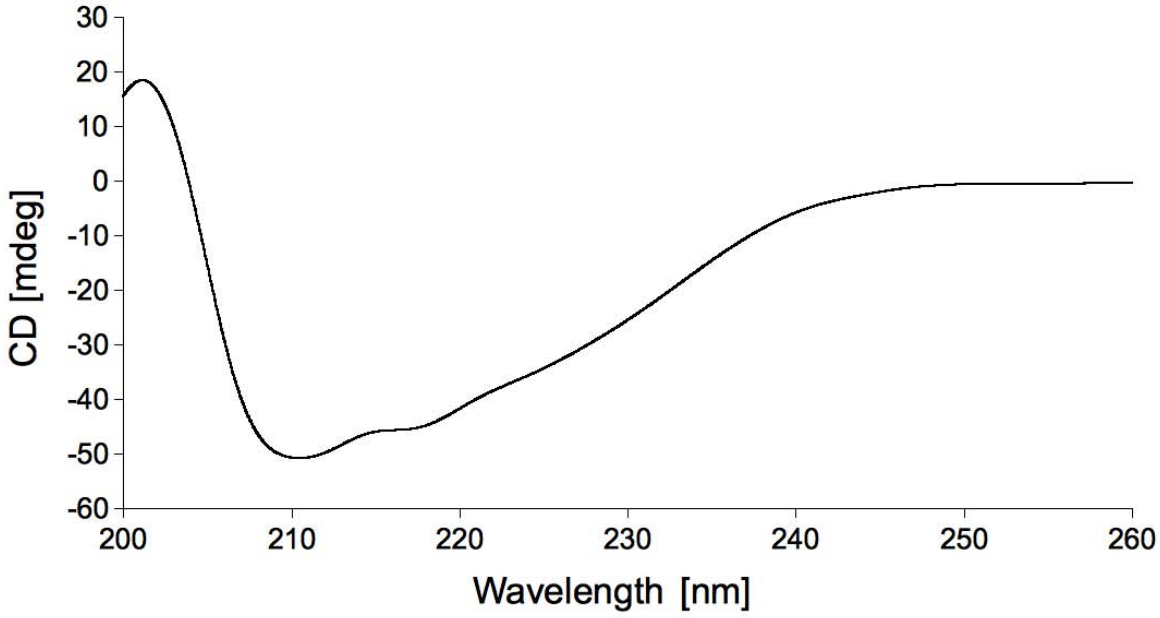


Figure 22. CD spectrum of PyLTNT

This single CD spectrum is characteristic of a partially unfolded protein containing elements of both a folded and an unfolded protein spectrum as described in the text.

3.2.4. AUC

The monomeric form of PyTLNT was confirmed by determining the molecular weight from the resulting AUC. Residual traces of larger molecular weight proteins can be seen in the spectrum arising from imperfect separation during gel filtration or a small amount of aggregation (Figure 23).

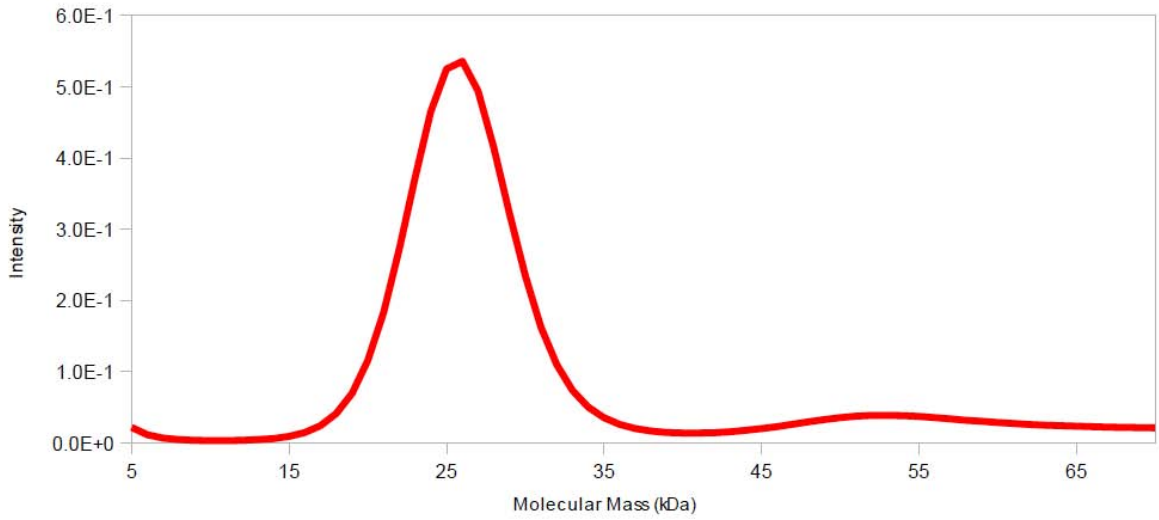


Figure 23. AUC analysis of PyLTNT

The monomeric form is identifiable at ~28kDa with some smaller residual impurities at a higher molecular weight.

3.3. Binding Studies

The characterisation of protein-protein interactions was of major importance in this thesis. Validation of the protein interaction between PyLTNT and pRb had to be established before proceeding with time consuming NMR experiments, especially considering the unfolded nature of the protein. Binding interactions with pRb showed that the unfolded protein was in its native state, and that it had not been unfolded during purification. A first assessment of the PyLTNT-pRb interaction was achieved by FPLC, showing that the two proteins form a complex. The complex binding efficiency was then further evaluated using ITC, that allowed a more systematic analysis of the nature of the binding.

3.3.1. FPLC

The FPLC results here determined that PyLTNT together with the pRb protein formed a complex upon binding. Each protein was first analysed individually and then as a complex to show that there was an interaction taking place (Figure 24). These measurements were repeated at two different pH values (7.0, 8.0) indicating a pH independent interaction.

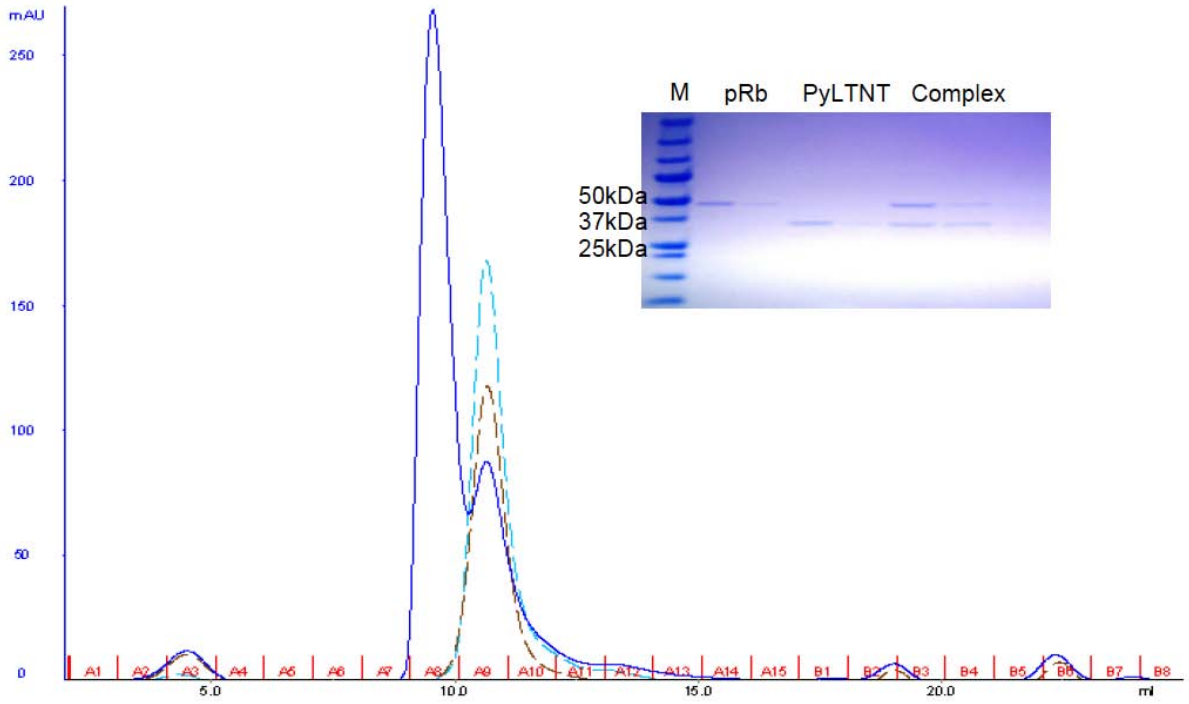


Figure 24. Interaction of PyLTNT and pRb using FPLC

The chromatogram of three FPLC analyses are superimposed here. The blue dashed line represents PyLTNT while the brown dashed line represents pRb, each protein was analysed individually. The blue solid line indicates two peaks with the first one being the complex of PyLTNT with pRb and the second being free PyLTNT. An SDS-PAGE is shown next to the FPLC chromatogram for the individual fractions collected. The pH value was 7.0.

3.3.2. ITC

ITC was used to estimate the binding affinity for the PyLTNT-pRb complex formation. Affinities could not be determined by NMR owing to slow exchange of the two proteins. As shown by the NMR titrations, the resonances seemed either to broaden beyond detection or to shift minimally, probably as a consequence of slow exchange. The small shift observed could also arise from aggregation, but FPLC results exclude this possibility.

ITC is the method of choice to estimate the binding affinity for the two proteins. Initial results showed that the affinity of the interaction between PyLTNT and pRb was reasonably strong with a K_d value of 124nM (±15nM). For a closer scrutiny, experiments were also performed between PyLTNTm2 and pRb. PyLTNTm2 is truncated by 58 amino acids compared to PyLTNT and showed a 3-fold reduction in binding efficiency with a K_d of 333nM (±54nM) compared with PyLTNT (Figure 25). Unfortunately, the further truncated PyLTNm1 could not, despite repeated attempts, be prepared in sufficient amounts for ITC analysis.

The results of the ITC must be carefully interpreted for NUPs, as there might be a contribution of concurrent induced folding resulting in thermal energy of folding and binding being mixed and giving a complex overall result. However, NMR results did not show any sign of induced folding.

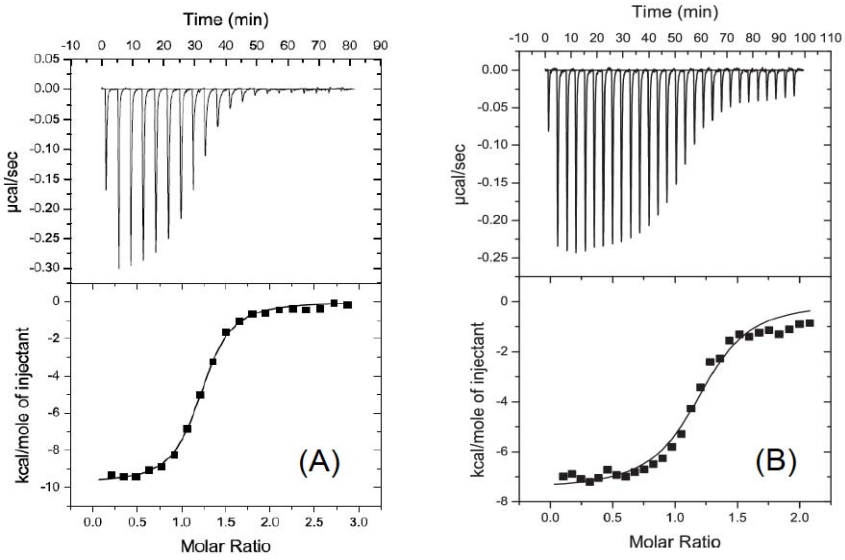


Figure 25. ITC results of pRb binding to PyLTNT and variant constructs

(A) pRb titrated to PyLTNT K_d = 124 ± 15nM, ΔH = -9823 ± 144.6, ΔS = -1.36, N = 1.18 ± 0.0120 (B) pRb titrated to PyLTNTm2 K_d = 401nM ± 71nM, ΔH = -7625 ± 204.2, ΔS = 3.69, N = 1.32 ± 0.0230. Each experiment was duplicated to ensure reproducibility, and each curve fitting was done with over 100 iterations. Heat referencing of pRb in buffer was also subtracted from the titrations.

3.4. NMR

NMR was used as a primary research tool to study the PyLTNT protein. The protein was examined under several conditions to optimise spectra acquisition. Some of the conditions tested were pH as well as additives to prevent aggregation, to improve S/N and to ensure the protein remained stable over an extensive period of time. Additives that were tested included 2mM of the zwitterionic detergent 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate (CHAPS), 2mM of the lipid 1-palmitoyl-2-hydroxy-sn-glycero-3-(phospho-RAC-(1-glycerol)) (LPPG), 2mM of the detergent octyl-beta-D-glucopyranoside (β-OG), a 50mM arginine/glutamate mixture (arg/glu) and 100mM trimethylamine N-oxide (TMAO).

Initial pH tests showed that samples in a phosphate buffer improved the number of observed resonances at pH lower that 7.0. A comparison of pH 8.0 and 6.0 shows that 80% of the resonances are detectable at pH 6.0, with the glycine region having the most notable effect (Figure 26). That is because the labile protons exchange at a faster rate with the solvent at high pH causing line broadening. For PyLTNT, at higher pH, residues that belong to the unfolded region that confers additional flexibility and thereby faster exchange cannot be seen. This effect is better observed in glycines of the HSQC that are have distinct chemical shifts.

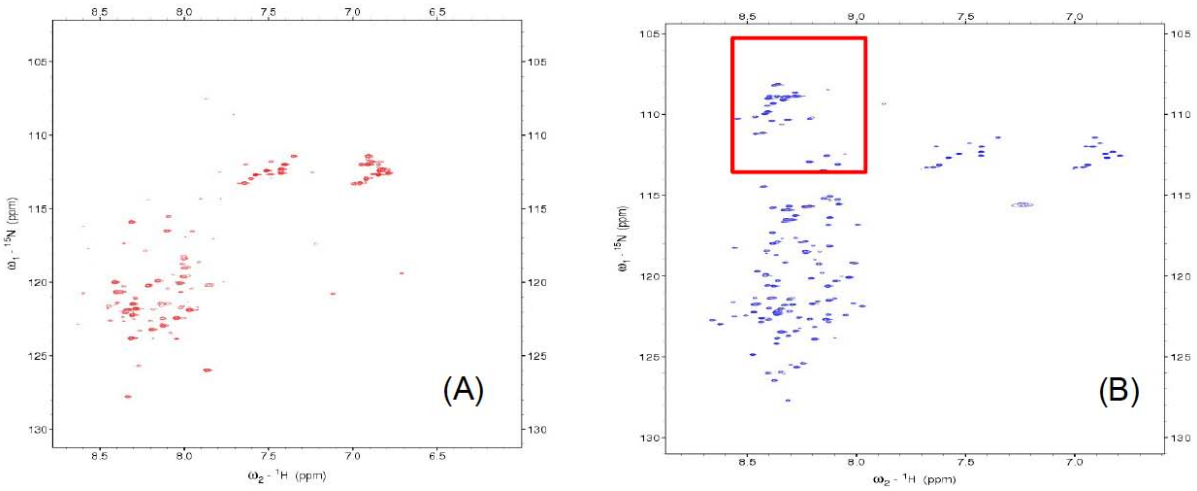


Figure 26. HSQC spectra of PyLTNT for different pH values

(A) pH 8.0 and (B) at pH 6.0. Red square highlights the additional glycine residues observed. NMR parameters are identical for the two spectra.

Additives extended the life time of the protein. Addition of 50mM arg/glu mixture prevented the protein from precipitation at room temperature over a period of 48h. In terms of spectral resolution, the arginine/glutamate mixture contributed to resonance broadening and was therefore deemed unsuitable for further experiments (Figure 27). The 2mM of LPPG, while extending protein life time, rendered some resonances undetectable (Figure 28). All other additives, failed to extend protein life time and inhibit precipitation over a longer period of time, and some even promoted degradation of the PyLTNT protein and were therefore not tested by NMR.

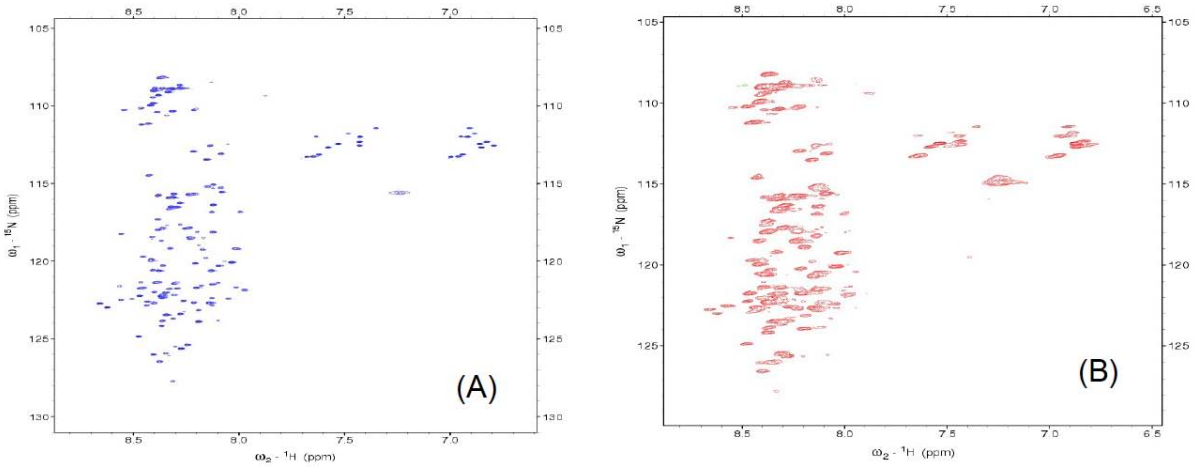


Figure 27. HSQC spectra of PyLTNT with and without the arg/glut additive

(A) with and (B) without addition of 50mM arg/glut at pH 6.0. The additive broadened the linewidth of the resonances as can be observed here. NMR parameters are identical for both spectra.

These results showed that a pH of 7.0 could be used and that no additives where adequate to be used for NMR to promote protein stability. The pH was chosen as a compromise of optimal stability. Most of the PyLTNT signals were detectable at pH 7.0. Furthermore, the pRb protein precipitated at pH lower than 7.0. To facilitate interaction experiments without the need for dialysis, both of the proteins were purified in the same buffer. Additionally, both proteins are stable at pH 7.0 as shown by ThermoFluor® assays.

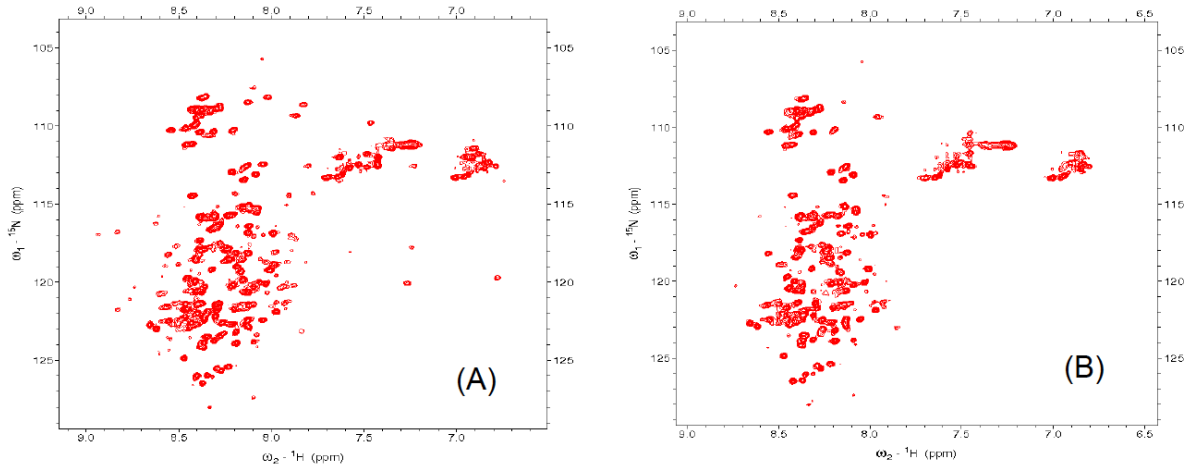


Figure 28. HSQC spectra of PyLTNT with additive

(A) without and (B) with addition of LPPG. Resonances are missing in spectrum B and are attributed to the J domain. NMR parameters for the two spectra are identical.

3.4.1. Assignments

Extensive amounts of NMR spectra were acquired for resonance assignment using both commonly used backbone assignment experiments, and experiments specifically designed for unfolded proteins. The nature of the protein necessitated this approach since it has both an unfolded C-terminus and a folded J domain. The experiments used for this thesis were already described in Chapter 2, where examples of different spectra were given in addition to their individual correlations between nuclei. An example of sequential assignment between residues 245 - 255 and residues 220 - 230 will be described to clarify the assignment process, and show how the various experiments used complemented each other.

The most useful experiments in originally assigning the PyLTNT protein were the CBCACONH and HNCACB. These experiments were used primarily because of the unfolded nature of the protein. By including information about the C^{α} and C^{β} of each residue it allowed us to clarify some of the ambiguous signals. More specifically, certain residues have characteristic signals of their carbon nuclei at specific ppm. An example is the C^{β} signal of alanines, that appears around 20ppm which is significantly different than any other C^{β} signal of any other residue. Another characteristic residue was glycine, this residue has no C^{β} signal and therefore can also be easily assigned in the spectrum. Additionally, serine and threonine C^{β} signals appear upfield of C^{α} signals making them unique and distinguishable from other residue types.

With these characteristic residues in mind, assignment of the two spectra can begin. Figure 28 is a preassigned section of the protein spanning from residues 245 to 255. It shows the superimposed spectra of CBCACONH and HNCACB. Without knowing the assignments it is immediately apparent that the 5th slice of the spectrum (249A)

represents an alanine residue because of the C^{β} signal on the HNCACB spectrum that appears around 20 ppm. This can be used as a starting point for the assignment. It is important to note that often there will be multiple starting points that will have to be used because of the high number of prolines that exist within such proteins. With the alanine as a starting point, the connections are of major importance to sequentially assign the signals. Therefore, from the alanine residue one has to go through the three-dimensional spectrum to find signals that connect to the C^{β} resonance. The C^{β} alanine signal of the *i* residue will appear as the *i-1* signal in the succeeding residue. In combination with the specific chemical shift of the C^{α} that will also appear as a *i-1* residue, accurate assignment of signals can be achieved.

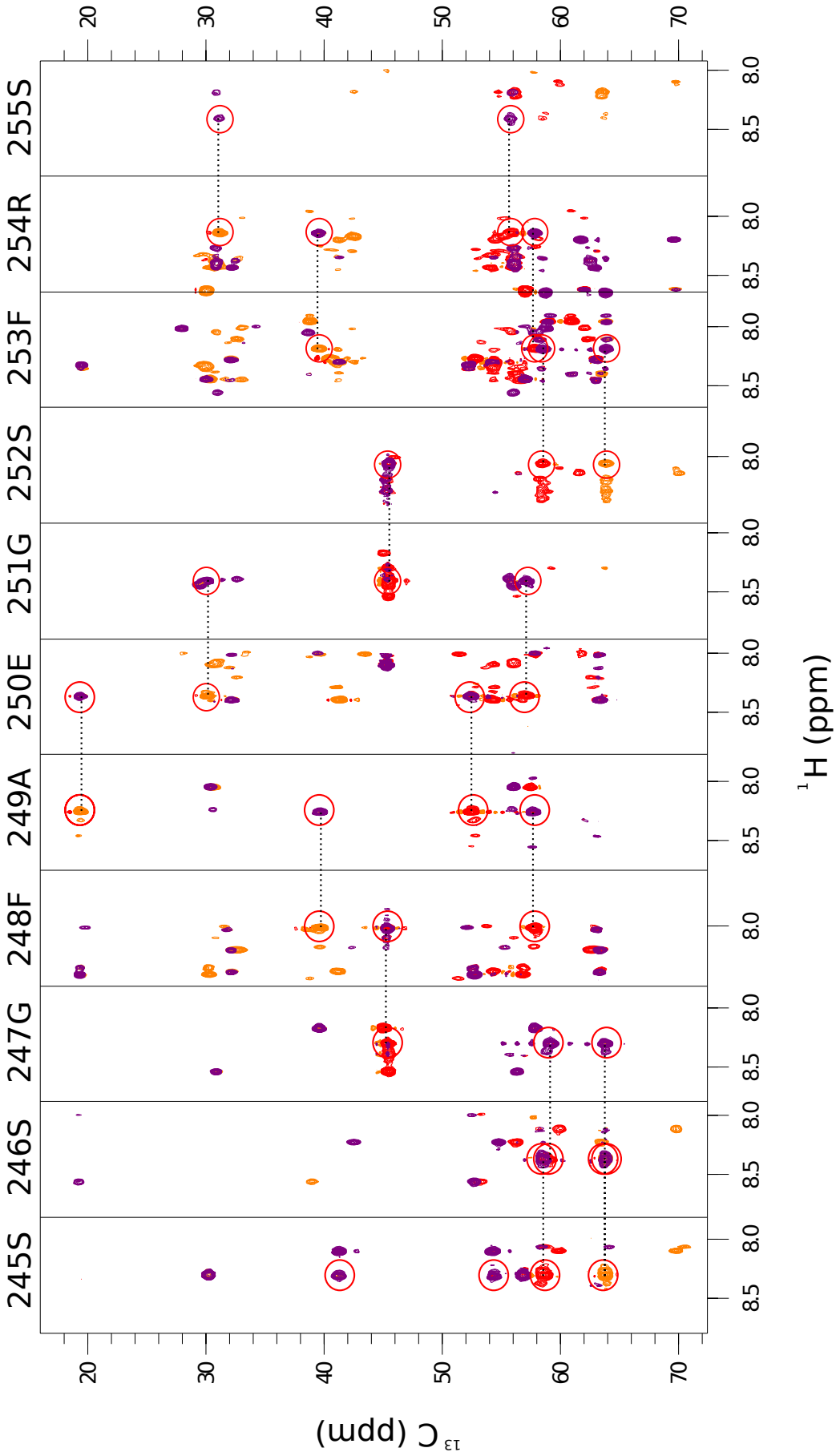


Figure 29. Sequential assignment of residues 245 to 255 using the superimposed CBCACONH and HNCACB spectra

CBCACONH experiment is represented by the purple resonances while the superimposed HNCACB is represented by the orange and red resonances. Red circles show the resonances associated with the residue represented on the top of each section. Dotted lines show how the resonances are interconnect with each other allowing sequential assignment.

In the specific example, the alanine residue seems to be connected with the signals of the glutamic acid residue in the *i+1* position. An alanine followed by a glutamic acid can be found 4 times in the protein sequence. To narrow down the possible residues that that these resonances belong to a further connection is needed. At the *i-1* position of the alanine residue, a phenylalanine can be found. Even if the residue is not immediately known because of ambiguous resonance assignment, one can move further to identify the next connection downstream. In this case, if it is assumed that the phenylalanine residue is unknown, the *i-1* residue it is connected to is a glycine. The glycine is distinct and easily recognisable. This allows us to know that we are looking now for a GxAE sequential assignment. In the sequence the only residues that match that description are residues 247 to 250. It is safe now to assume that the unknown resonances belong to a phenylalanine residue. From there moving further upstream is similar, with the glutamic acid being connected to a glycine and the glycine being connected to a serine. The serine is distinguishable by the upfield C^{β} resonance and cannot be confused by the threonine C^{β} resonance that is located even further upfield. Finally the phenylalanine and arginine follow the serine and are assigned based on the protein sequence as well as based on good connectivities between the resonances observed. This sequential assignment finishes with a serine at position 255. On the other end as well, the assignment

finishes with two serines in a row making it challenging but at the same time distinct in the assignment process.

Following the spectra mentioned, the sequential assignment in the HNN and HNCN spectra serves in this case to confirm the assignments of the resonances. As can be seen in Figure 30, the HNN and HNCN spectra exhibit certain type of triplet resonances. This makes it easy to distinguish and allocate each resonance in relation to the *i* residue and even on occasion the type, specifically if glycines and prolines are present in the triplet.

The assignment from the previous CBCACONH and HNCACB spectra help identify some of the resonances in the HNN/HNCN spectra. Furthermore the resonances are in most occasions well dispersed allowing to distinguish the different residues and confirm connectivities between them.

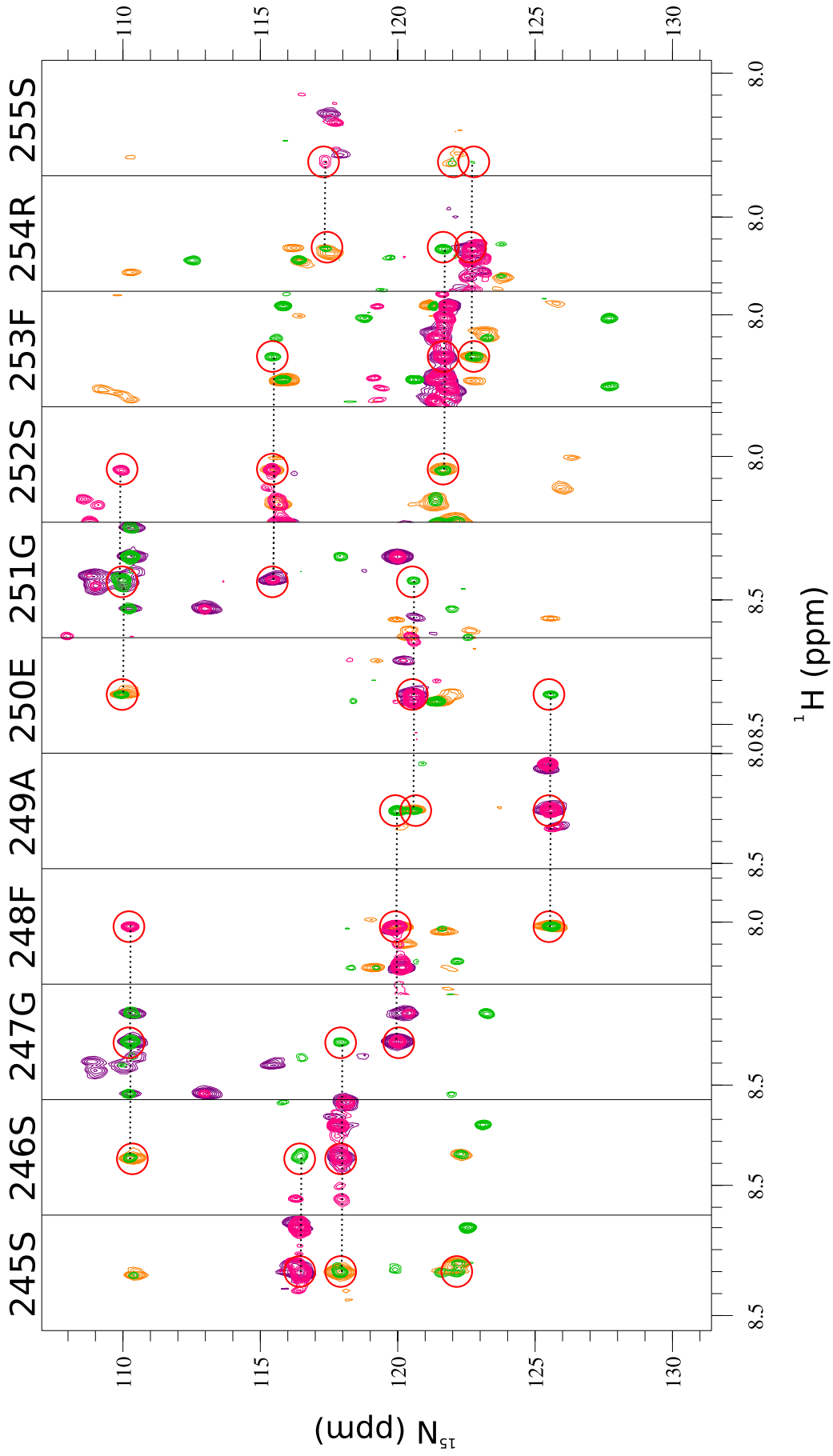


Figure 30. Sequential assignment of residues 245 to 255 using the superimposed HNN and HNCN spectra

The HNN spectrum is represented by the pink and green resonances while the HNCN spectrum is represented by the orange and purple resonances. Each red circle in each section points to a specific resonance that correlates to the NH resonance of the i residue. The connectivities are represented by dotted lines.

Starting for instance at the alanine residue at position 249, again like before, the *i* resonance is represented by a positive signal. In this case alanine has a positive pink signal in the HNN spectrum and two negative signals each representing *i+1* and *i-1* resonance correlations. To distinguish the preceding and succeeding resonances the HNCN can be used that only shows the resonances of the *i* residue as a positive signal in purple, and the *i+1* residue as a negative signal in orange. Immediately the assignment becomes easier because of the two spectra and the distinction of positive and negative signals. Added to this is the fact that when prolines are encountered in the preceding or succeeding position a resonance will be absent. This will be vital in the assignment process of residues 220 to 230 as will be shown. Furthermore, glycines are represented by negative signals when they are in the *i* position making them uniquely distinct.

The sequential assignment mentioned so far had no prolines present in the sequence. To demonstrate how assignment of a proline and glycine rich region was done, following are assignment spectra of residues 220 to 230. Figure 31 is another CBCACONH and HNCACB sequential assignment, however due to the prolines present additional information was needed to complete this sectional assignment.

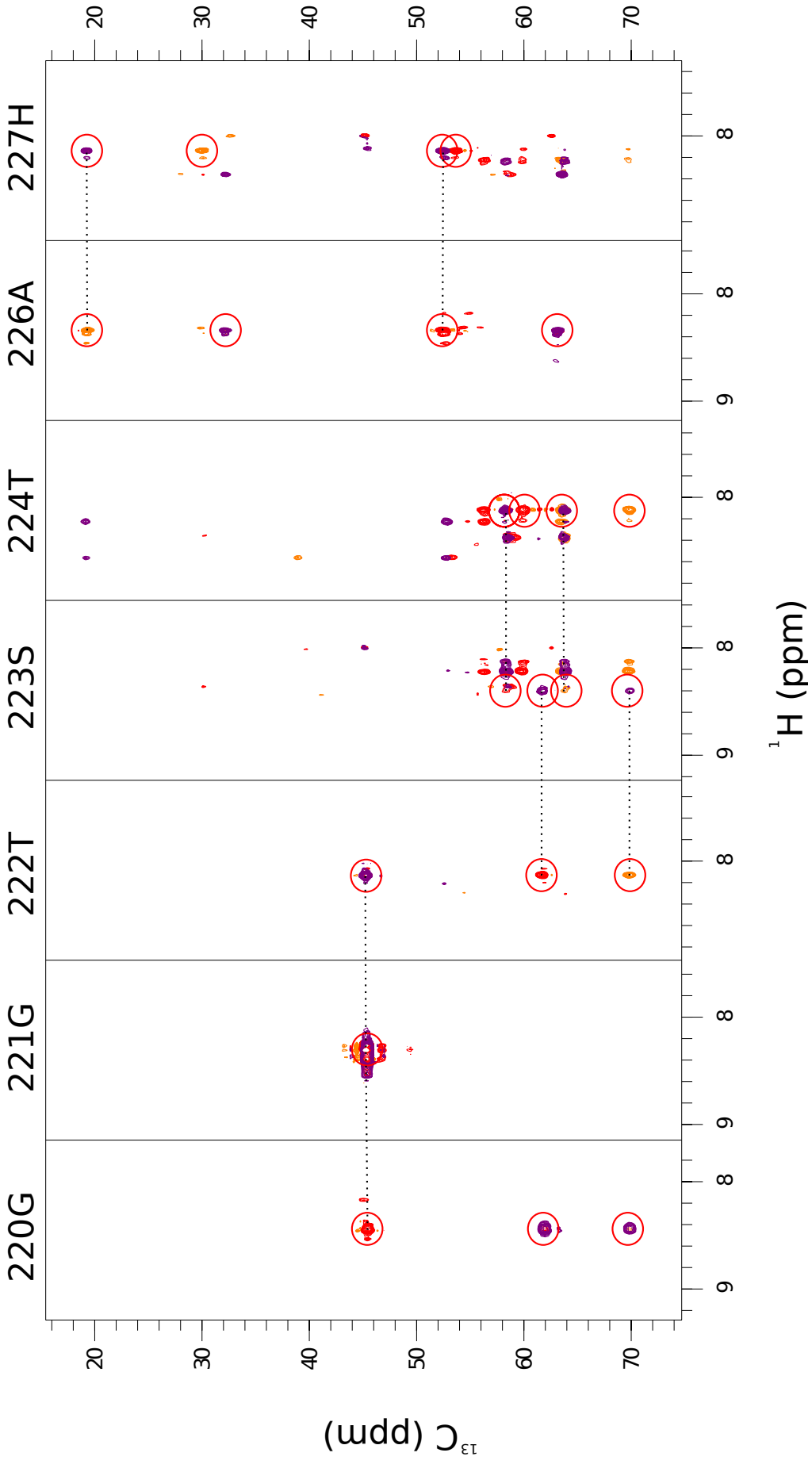

227H
226A
224T
223S
222T
221G
220G
20
30
40
50
60
70
8
9
^{13}C (ppm)
^{1}H (ppm)

Figure 31. Sequential assignment of residues 220 to 230 using the superimposed CBCACONH and HNCACB spectra

CBCACONH experiment is represented by the purple resonances while the superimposed HNCACB is represented by the orange and red resonances. Red circles show the resonances associated with the residue represented on the top of each section. Dotted lines show how the resonances are interconnect with each other allowing sequential assignment. The missing residues are prolines that cannot be seen in these experiments.

Furthermore, the CBCACONH and HNCACB resonances were confirmed by HNN and HNCN experiments. This was especially useful for residues 220 to 222 that connects two glycines with a threonine.

As seen in Figure 32 the HNN and HNCN experiment reconfirm the assignments in the CBCACONH and HNCACB experiments. Additionally, the 220 to 222 positioned residues are connected easily due to the triplet of resonances that allow sequential assignment. The same residues on the CBCACONH and HNCACB experiments are only connected via one connectivity due to the repeated glycine residues. This allows the distinction of the various glycines that are found in the sequence.

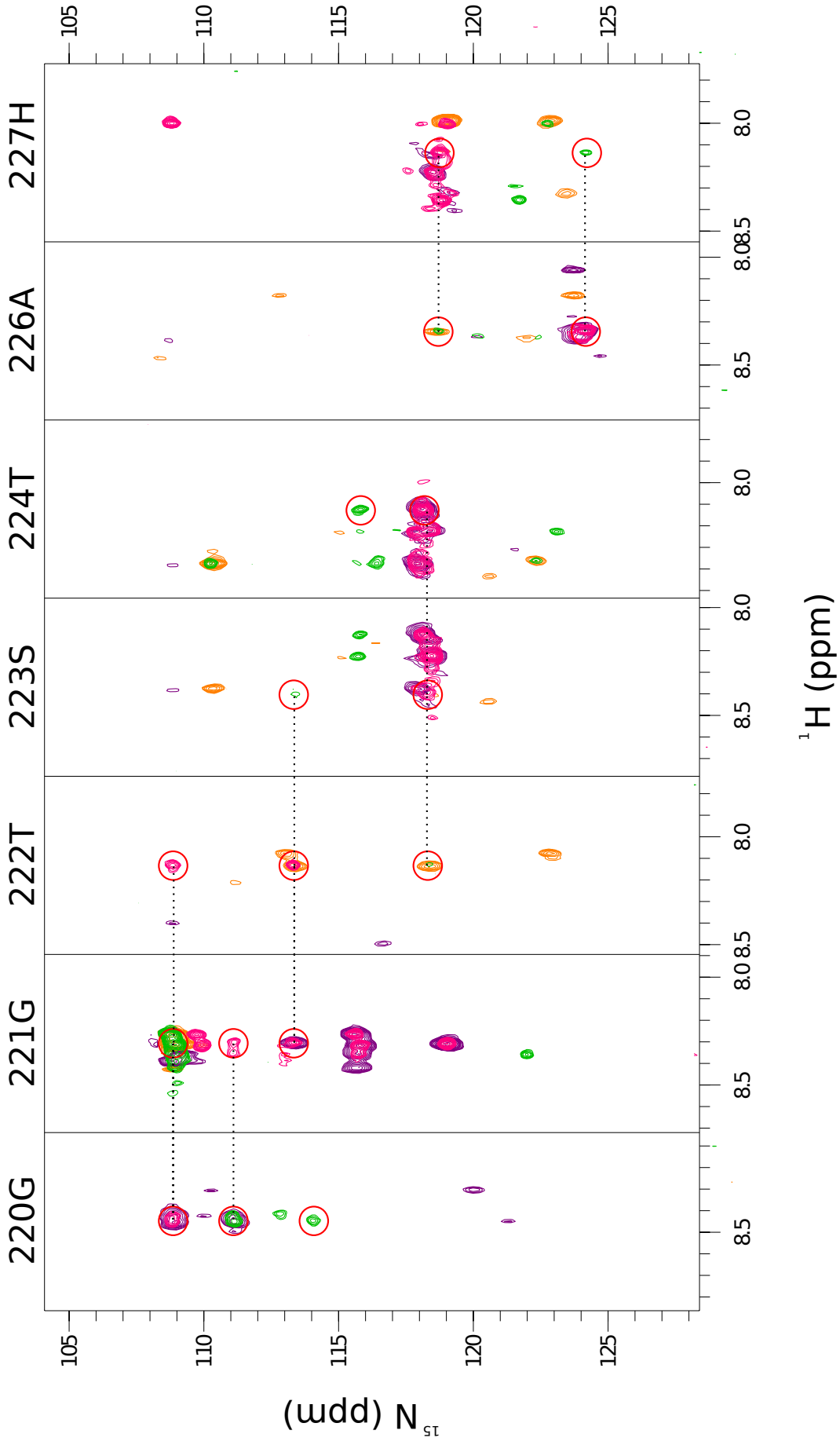

227H
226A
224T
223S
222T
221G
220G
105
110
115
120
125
8.0
8.5
^{15}N (ppm)
^{1}H (ppm)

Figure 32. Sequential assignment of residues 220 to 230 using the superimposed HNN and HNCN spectra

The HNN spectrum is represented by the pink and green resonances while the HNCN spectrum is represented by the orange and purple resonances. Each red circle in each section points to a specific resonance that correlates to the NH resonance of the i residue. The connectivities are represented by dotted lines. Missing resonances belong to proline residues.

The multitude of prolines found in this sequence makes it hard to sequentially assign the protein sequence and in addition creates gaps in the assignment. To overcome this obstacle the protonless experiments CBCACON and CBCANCO were used. These experiments, as described previously, show the proline resonance making them extremely valuable in sequences such as the one shown here. In combination with the HNN and HNCN experiments complete assignment of a glycine and proline rich region is possible.

In Figure 33 the CBCACON and CBCANCO experiments of the proline residues that are missing from the assignment process of residues 220 to 230 can be seen. The sequential assignment of the last three consecutive prolines allows the identification, in combination with the previously described experiments, of the position of the residues these resonances belong to.

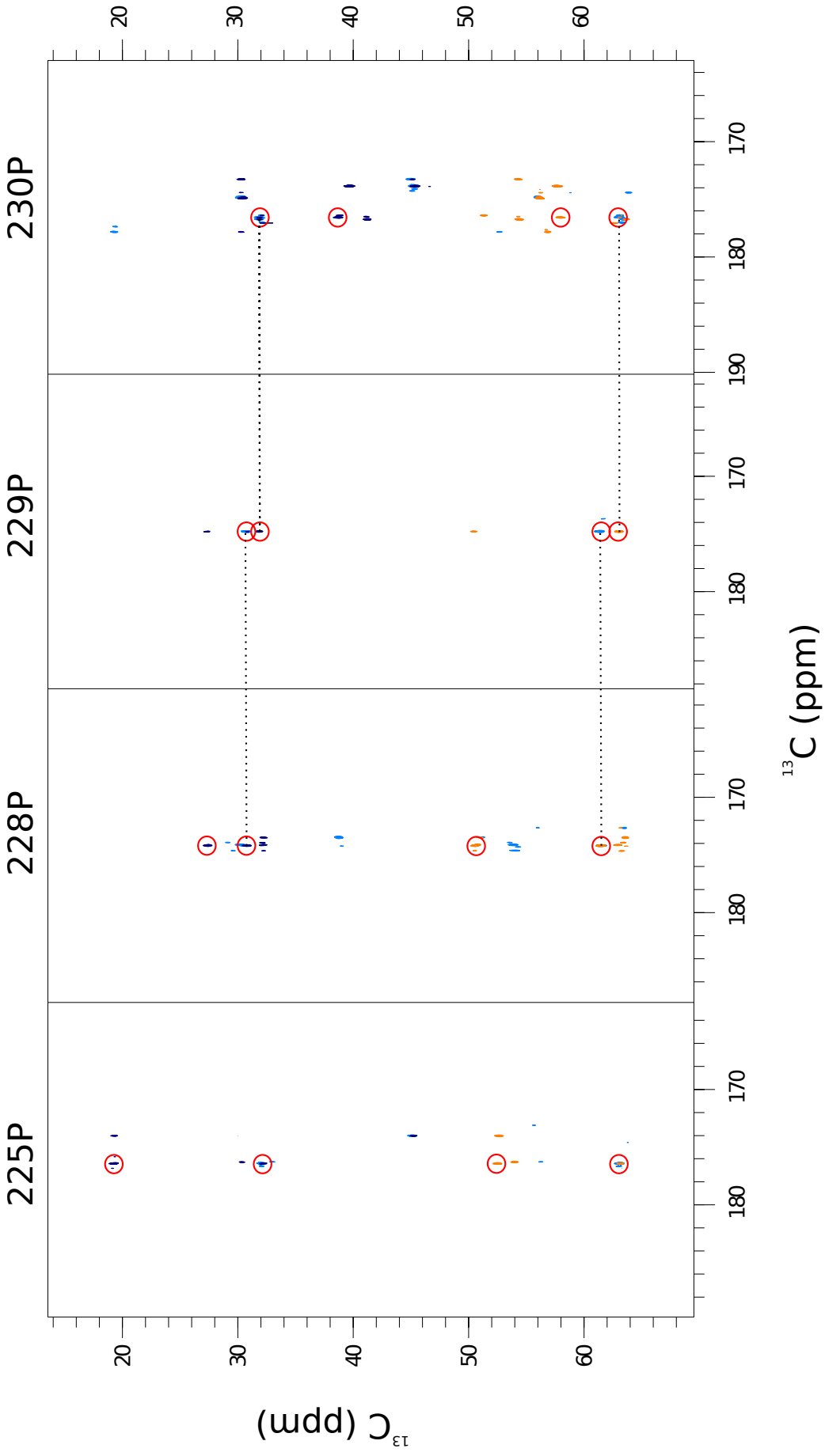

230P
229P
228P
225P
20
30
40
50
60
170
180
190
^{13}C (ppm)
^{13}C (ppm)

Figure 33. Sequential assignment of proline residues between region 220 to 230 using the superimposed CBCACON and CBCANCO spectra

CBCACON experiment is represented by the dark blue resonances while the superimposed CBCANCO is represented by the orange and light blue resonances. Red circles show the resonances associated with the residue represented on the top of each section. Dotted lines show how the resonances are interconnect with each other allowing sequential assignment. These slices show only the prolines.

The software CcpNmr Analysis was a pivotal tool in the assignment process. It allowed nuclei assignments to be transferred across the various experiments. Using these experiments assignment of 76% of the protein backbone was achieved (Figure 34) (Knoblich et al., 2009). The unfolded part of the protein was assigned completely aided by the strong S/N ratio the resonances exhibit, and supported by similar, although not identical, published assignments of the isolated J domain (Berjanskii et al., 2000). Residues 98-131 could not be assigned owing to either missing or unidentified resonances. Some C^{α}, C^{β} and H^{α} assignments are also missing for some of the N and C-terminal assigned residues. The ^{1}H, ^{13}C and ^{15}N chemical shifts of PyLTNT were deposited in the BioMagResBank (http://www.bmrb.wisc. edu) under accession number 16044. Intermittent missing residues limited the use of the chemical shift index (CSI) for secondary structure prediction (Wishart et al., 1992), as well as further dynamic data correlation such as heteronuclear NOEs.

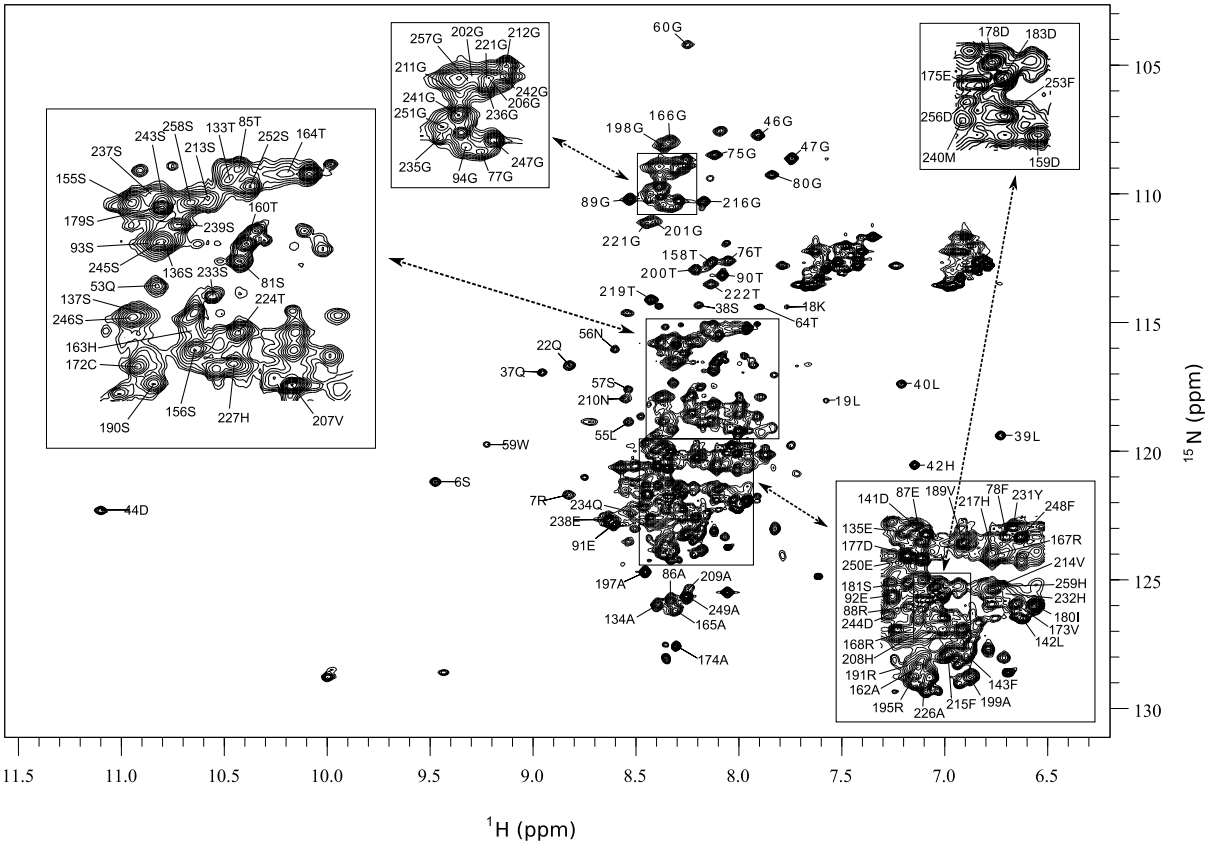


Figure 34. Assigned HSQC spectrum of PyLTNT

Individual amino acids are labelled. Parameters can be found in the appendix (Table 2).

The CSI secondary structure prediction by Wishart et al. is based on tabulated values of H^{α}, C^{α}, C^{β} and CO from existing structures that are used to predict the structure of a protein based on the H^{α}, C^{α}, C^{β} and CO of the protein of interest. Depending of whether the values are smaller, the same or larger than the predicted tabulated values, they are given values of -1, 0 and 1, respectively. Using these chemical shift indices patterns can arise. Four or more -1 values not interrupted by 1 are indicative of a helix, three or more values of 1 in a row not interrupted by -1 are indicative of a β-strand, and all other regions are designated as coil.

CSI (Figure 35) shows that the C-terminus of the protein remains predominantly unfolded, while the N-terminus has some helical conformation. While all residues were not available for assignment, the CSI is in agreement with data from the the

previously described J domain where three helices were described at the same amino acid sequence position.

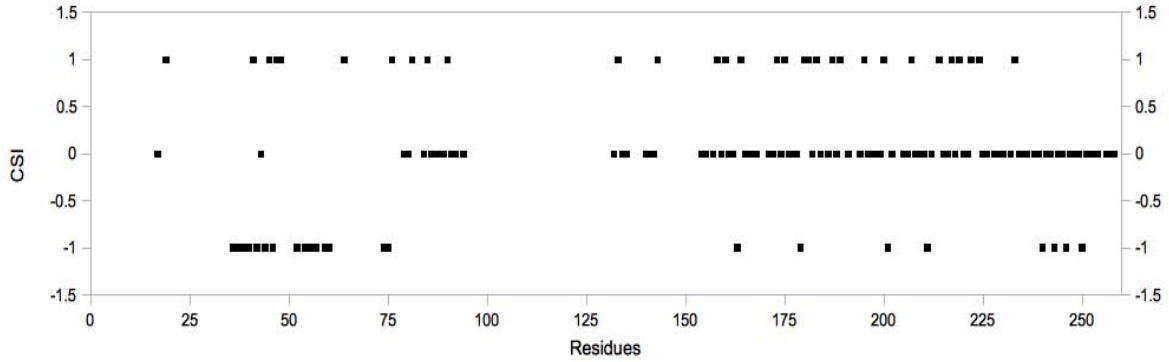


Figure 35. CSI of PyLTNT

Previously described helices (Berjanskii et al., 2000) *were found from residues 7-16 (Helix I), 27-41 (Helix II) and 49-70 (Helix III). True zero values are shown. Due to incomplete assignment not all residues could be used for CSI analysis.*

Chemical shift mapping of the differences between the isolated J domain and the PyLTNT protein (Figure 36) also highlights some important conformational changes in the J domain as part of PyLTNT. Most of these conformational changes occur in the C-terminal region of the isolated J domain. Significant changes occur mostly on the C-terminus, as extension of the J domain as part of PyLTNT cause change in conformation. Some minor chemical shift differences can also be observed between helix II and III, a region that has the HPD motif that forms a flexible loop and has been thought to be a functionally conserved feature (Berjanskii et al., 2002).

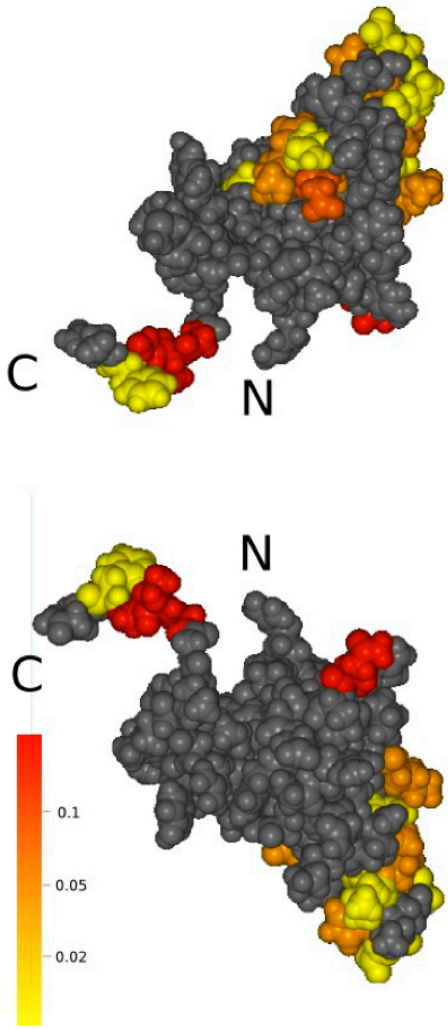


Figure 36. Chemical shift differences between PyLTNT and the individual J domain

The chemical shift differences between the isolated J domain and as part of PyLTNT are highlighted on the J domain structure. Colours represent the chemical shift perturbations calculated by $\Delta\delta_{total}=((\Delta\delta_H)^2+(\Delta\delta_N*0.2)^2)^{0.5}$. *With kind permission from Springer Science+Business Media:* (Knoblich et al., 2009)

3.4.2. Titrations

Titration of pRb into PyLTNT proved laborious as pRb was unable to bind to PyLTNT when highly concentrated for NMR experiments. This was observed after a series of titration points showed no chemical shift differences. To circumvent this, a dilute pRb sample equimolar in concentration to PyLTNT was added to PyLTNT, and then concentrated back to a volume of 600µl suitable for NMR analysis. Unfortunately, this procedure did not allow for a stepwise titration, which would have been useful for dynamic analysis. Instead a single addition of pRb to PyLTNT was used to identify changes in NMR parameters caused by the interaction.

Chemical shift changes cannot necessarily be used to identify residues involved in the interface of the protein because only part of the observed chemical shift changes arise from the direct interactions. Most of these chemical shift perturbations can be attributed to localised conformational changes. Nevertheless, chemical shift perturbations are useful for interaction site localisation, especially in a partially unfolded protein where long range interactions are unlikely. Usually interaction sites are supported by a consecutive row of amino acids showing chemical shift perturbations, whereas more isolated chemical shift perturbations are more likely to be caused by long-range conformational changes.

Unfortunately, a one step titration does not allow for an assessment of binding kinetics. Nevertheless, a rough estimation of the exchange regime is possible based on the behaviour of signals after the one step titration. Intermediate exchange was hard to be observed and was speculated after interaction for resonances had a significant lower intensity than before the addition of pRb. Additionally, resonances that completely disappeared after binding were assigned as slow exchange, on the basis that they a) were also essential for binding e.g. residues of the LxCxE motif or,

b) form additional clusters around the LxCxE motif. Resonances that shifted upon binding were assigned as fast exchange. Because of the limited shift experienced by most of the resonances, only those that formed clusters were attributed as being in fast exchange.

Results from the PyLTNT-pRb binding show a multitude of chemical shift perturbations (Figure 37). Examples of residues in fast exchange between the free and bound forms are 227H, 173V, 180I as well as others not shown in Figure 32. Examples of slowly exchanging signals are 142L and 141D in Figure 32.

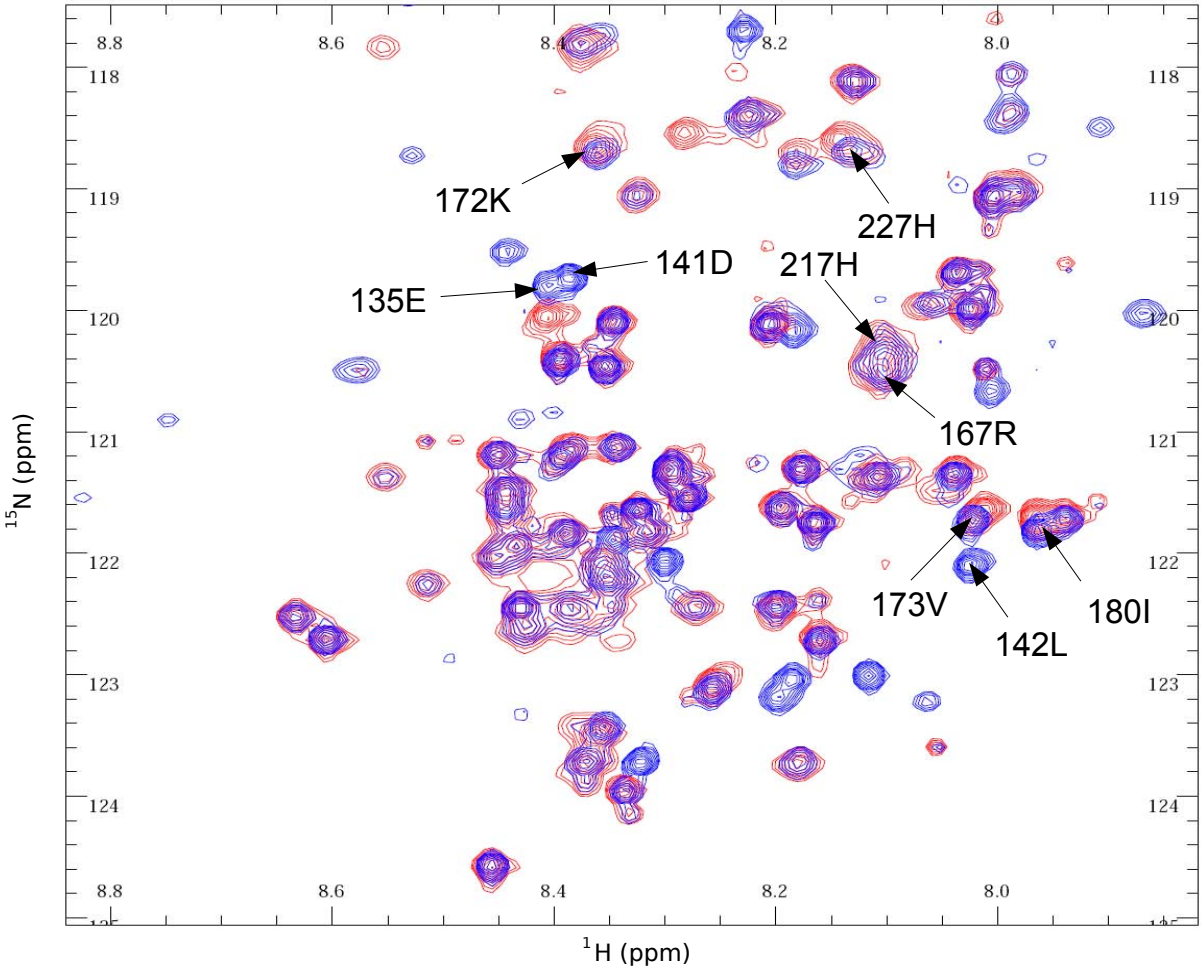


Figure 37. Enlarged region of superimposed HSQC spectra of PyLTNT before and after binding of pRb

The blue spectrum represents PyLTNT without pRb and the red spectrum represents an equimolar mixture of PyLTNT and pRb. Some of the interacting residues are labelled to show examples of potential fast (180I) and slow (142L) exchange.

In the PyLTNT sequence the interacting amino acids have been highlighted and categorised according to whether they can be found in the homologous SV40 sequence (Figure 38). Furthermore, the exchange regime of these amino acids has also be highlighted on the protein sequence (Figure 39).

```
1         10         20         30         40         50         60         70         80
MDRVLSRADKERLLELLKLPRQLWGDFGRMQQAYKQQSLLLHPDKGGSHALMQELNSLWGTFKTEVYNLRMNLGGTGFQGSPPRTAERG
90        100        110        120        130        140        150        160        170
TEESGHSPLHDDYWSFSYGSKYFTREWNDFFRKWDPSYQSPPKTAESSEQPDLFCYEEPLLSPNPSSPTDTPAHTAGRRRNPCVAEPDDS
180       190        200        210        220        230        240        250        260
ISPDPPRTPVSRKRPRPAGATGGGGGGVHANGGSVFGHPTGGTSTPAHPPPYHSQGGSESMGGSDSSGFAEGSFRSDGSHHHHH
```

Figure 38. PyLTNT sequence with interacting amino acids highlighted

Amino acids highlighted in green represent all that have been assigned. Amino acids highlighted in blue are those that are unique to the polyomavirus sequence, are interacting with pRb and have been assigned. Amino acids in red are those that are homologous to the SV40 sequence, are interacting with pRb and have been assigned. Underlined amino acid residues represent helices as predicted by the software PSIPRED (Bryson et al., 2005).

```
1         10         20         30         40         50         60         70         80
MDRVLSRADKERLLELLKLPRQLWGDFGRMQQAYKQQSLLLHPDKGGSHALMQELNSLWGTFKTEVYNLRMNLGGTGFQGSPPRTAERG
90        100        110        120        130        140        150        160        170
TEESGHSPLHDDYWSFSYGSKYFTREWNDFFRKWDPSYQSPPKTAESSEQPDLFCYEEPLLSPNPSSPTDTPAHTAGRRRNPCVAEPDDS
180       190        200        210        220        230        240        250        260
ISPDPPRTPVSRKRPRPAGATGGGGGGVHANGGSVFGHPTGGTSTPAHPPPYHSQGGSESMGGSDSSGFAEGSFRSDGSHHHHHH
```

Figure 39. PyLTNT sequence with exchange regime highlighted

Each amino acid is highlighted according to the interaction exchange regime with pRb. Red is for slow, blue is for fast and green is for intermediate exchange. These

are in respect to HSQC one step titration data. Underlined amino acid residues represent helices as predicted by the software PSIPRED (Bryson et al., 2005).

3.4.3. Dynamics

Although full structural analysis was not feasible for the PyLTNT protein, relaxation studies can indicate the formation of residual structure with reduced mobility caused by molecular interactions. The C-terminus of the free PyLTNT has no residual structure as supported by data of the hetNOE (Figure 40), and R_2 analysis. Secondary structure was predicted in the J domain region as was expected after the CSI analysis.

For residues in the J domain values above the theoretical maximum for NOEs of 1 were observed. This can arise from three different effects: i) HetNOEs of large molecules are difficult to quantitate because owing to the small difference between two large numbers, ii) The magnetisation transfer between the saturated water and the exchangeable amides can be exacerbated at high magnetic fields owing to the longer relaxation time of the water or, iii) It is a consequence of incomplete proton saturation that can lead to overestimation of errors (Caffrey et al., 1995).

The steady-state hetNOE values were calculated from the ratio of peak intensity of the saturated state to the non-saturated state. The saturation time was 5 seconds, equal to the repetition delay. This allows for complete saturation of the protons and for optimal relaxation. Therefore, overestimation of errors can not be explained by incomplete saturation. Nevertheless, these experiments were performed on a 800MHz magnet where the magnetisation transfer between water and the amides might be exacerbated. In addition, high error values were seen in residues involved in the J-domain as seen in Figure 40. The signals of those residues were close to the signal to noise threshold making it very difficult to get an accurate measurement that, in turn, lead to high error values.

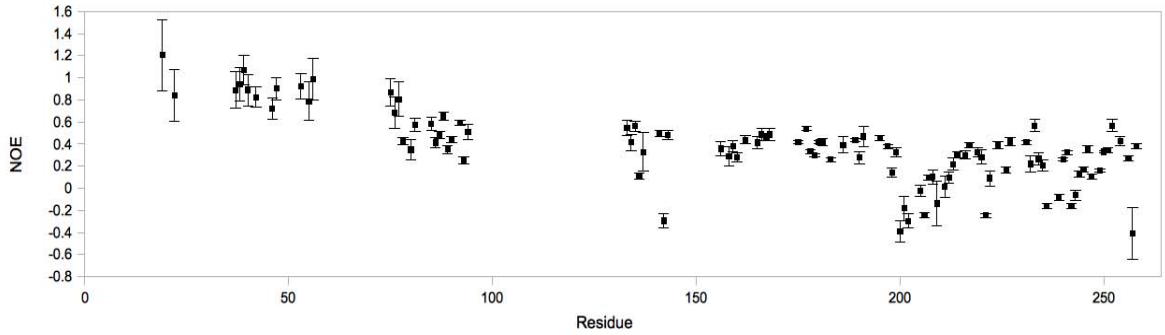


Figure 40. ^{1H}-^{15}N steady-state heteronuclear NOE analysis of PyLTNT

The low NOE values indicate lack of residual structure on the C-terminus while the N-terminus shows high values indicative of secondary structure. The J-domain shows clusters of secondary structure where the three helices have been predicted. Error bars are indicated for each value with maximum being 0.32 and minimum being 0.01.

Firstly, significant chemical exchange contributions were previously observed in the vicinity of disulphide bridges, for example in denatured lysozyme (Schwalbe et al., 1997). In this publication Schwalbe et al. noted, that the increased R_2 rate might be a consequence of the intrinsic effect of cross-linking on the conformational exchange around the disulphide bridges. The PyLTNT contains 2 cysteines and the increased R_2 rates are observed around the cycteine regions at 144C and 172C. However, recent work on the lysozyme protein has further revealed, that a single point mutation of the protein can diminish the R_2 values, although there is no direct effect on the disulfide bonding of the lysozyme cysteines (Klein-Seetharaman et al., 2009). These results indicated that there was a long-range tertiary interaction in lysozyme that had to be taken into account.

Additionally R_2 of PyLTNT was measured when the protein was in its free state and when bound to pRb (Figure 41). Upon binding residues involved in the interaction

become restricted in motion, leading to slower milli to microsecond and nano to picosecond timescale motions which are detected by the R_2 relaxation parameter. For the localised folding effects in PyLTNT it is very likely that the slower motion contribution is more pronounced than that at ps-ns time scale, although effects on a faster time scale cannot be excluded. Unfortunately, samples were not sufficiently stable to record both R_2 relaxation-dispersion spectra on the same sample. The ratio of pRb to PyLTNT was 1:10 so that the pRb protein is in sufficiently low concentration to allow rate measurements. The difference between those rates provides valuable information of which residues are involved in the binding (Figure 42).

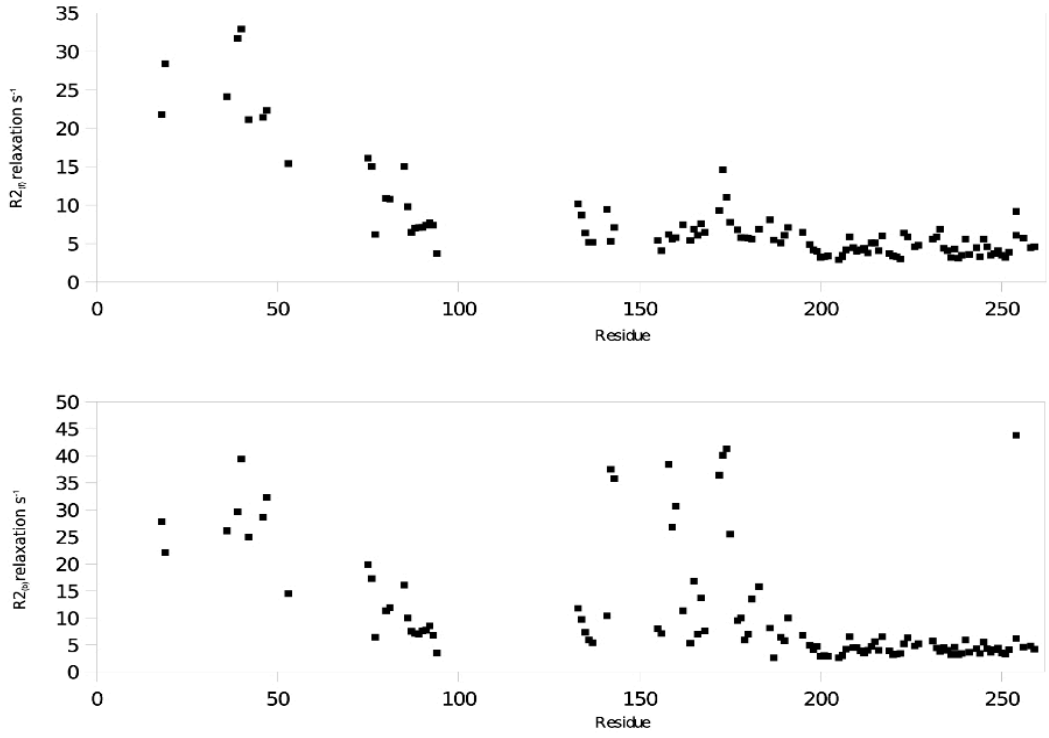


Figure 41. R_2 relaxation analysis of PyLTNT

(A) free PyLTNT and (B) bound to pRb. The comparison shows that there is no residual structure on the C-terminus prior to binding. Formation of residual structure can be seen upon binding in assigned residues on the unfolded part of the protein.

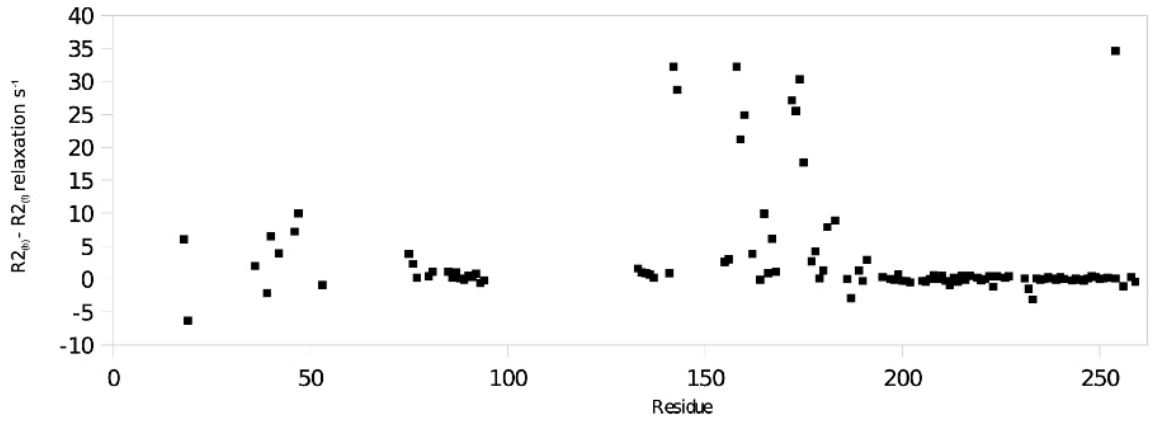


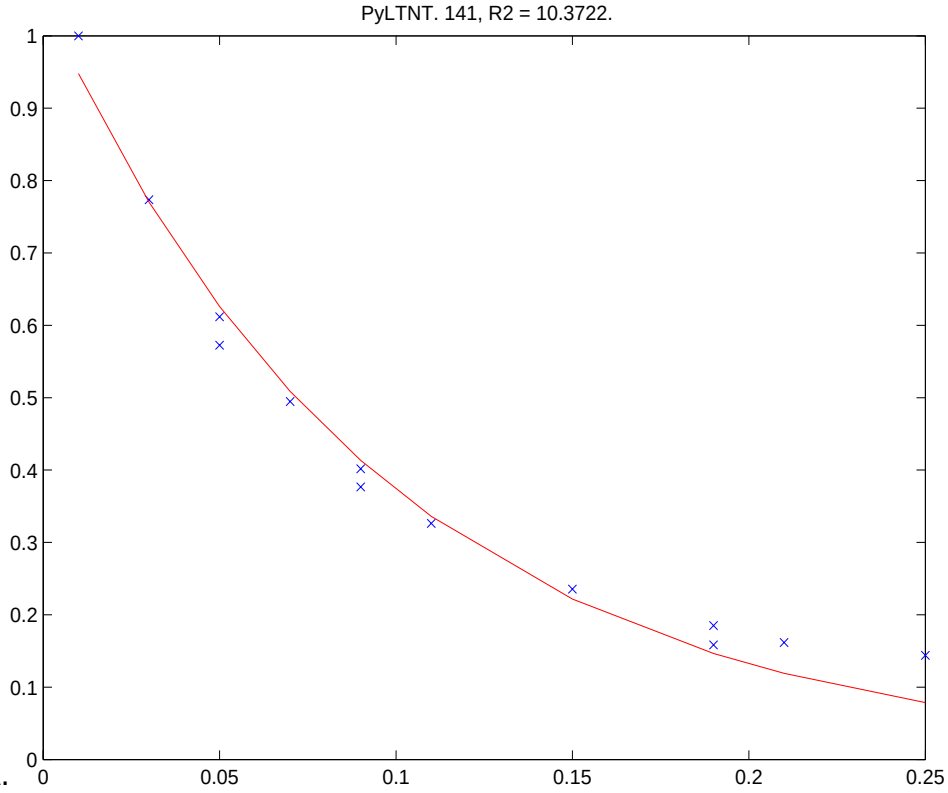
Figure 42. R_2 difference of the two states of PyLTNT

The binding of pRb allows for some residual structure to be formed around the LxCxE binding motif and the J domain. The LxCxE motif residues (141-145) as well other surrounding residues are in fast exchange as indicated by the the short T_2 (therefore large R_2 since $R_2=1/T_2$).

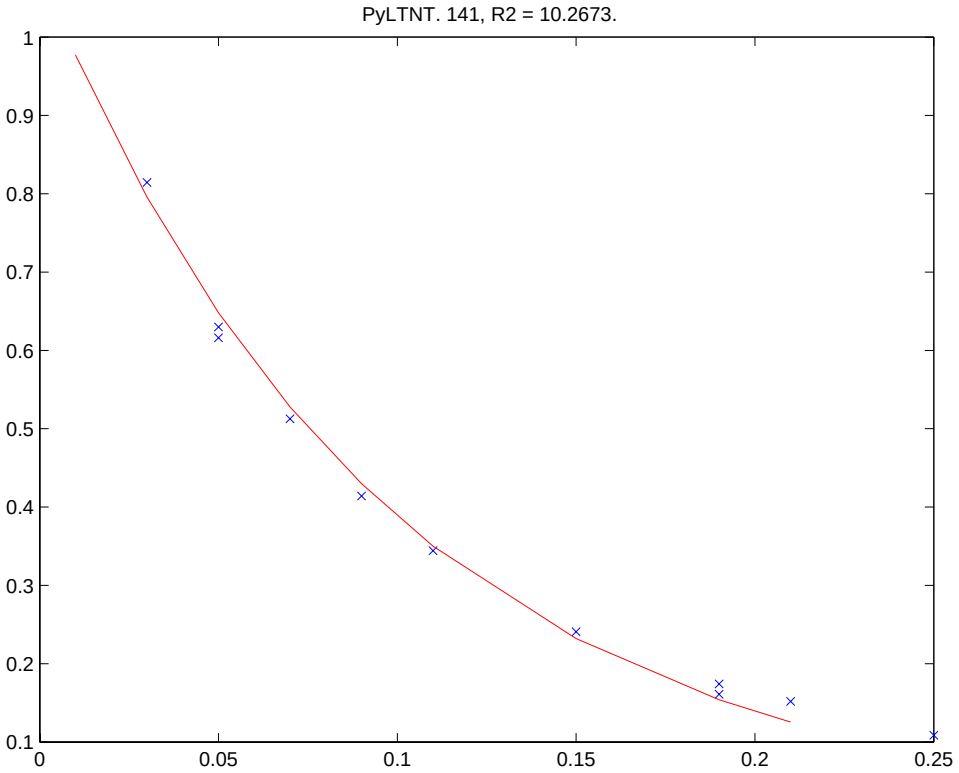
Secondly, when pRb interacts with PyLTNT, a different pattern appears in the unfolded region of the protein. Residues of the LxCxE motif that were found to be in slow exchange were expected to interact with the pRb protein. More importantly, there are interactions downstream of the LxCxE region. Residues 158 to 160 and 172 to 175 have a higher relaxation rate difference when bound to pRb. Specifically, the difference in relaxation rate between the free and bound protein is greatest in residues: 142L at 32.2 s^{-1}, 143F at 28.7 s^{-1}, 158T at 32.2 s^{-1}, 159D at 21.2 s^{-1}, 160T at 24.9 s^{-1}, 172C at 27.1 s^{-1}, 173V at 25.5 s^{-1}, 175E at 17.7s^{-1} and 254S at 34.6 s^{-1}. These results strongly support residual structure formation upon pRb binding in previously unexpected regions. Residue 254 that also exhibits a high relaxation rate can be an indirect consequence of long range interaction, as the C-terminal tail seems to fold back towards itself as observed in SAXS models. In Figure 43 representative T_2 fitting is shown for 2 residues that have different relaxation rates

after binding. The fitting while not optimal due to the crowding of resonances can serve to indicate the difference between the bound and unbound state. The relaxation of those resonances is fast and therefore fitting becomes difficult in the long relaxation delay times leading to resonances that have low signal or are non observable.

A.



B.



C.

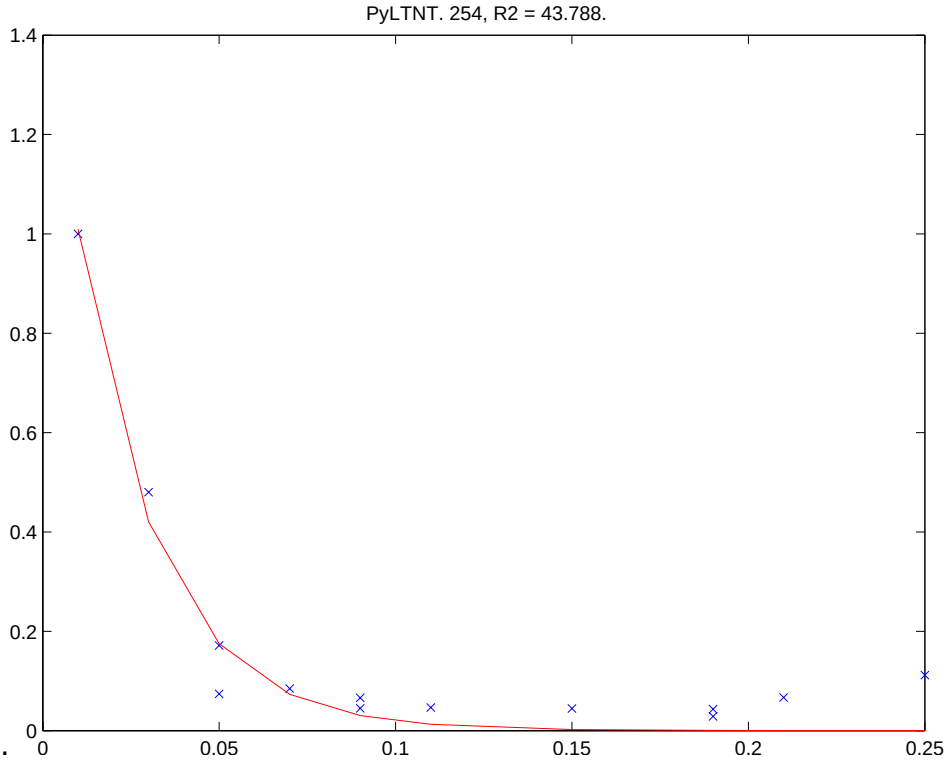


Figure 43. Curve fitting of residue 141 before and after complex formation with pRb for different relaxation delay values and an example of fast relaxation

The curve fitting can be seen by the red line. The three repeat points are seen by the additional three points in same values. A. Residue 141 relaxation rate before addition of pRb B. Residue 141 after addition of pRb exhibits only minimal reduction in its relaxation rate C. Example of fast relaxation in residue 254. Later points in the fast relaxing residue are inaccurate as due to relaxation the signal completely disappears.

Long relaxation delays led to resonances being too close to the signal to noise limit or even disappearing. Latter points in the curve fitting, for residues at fast relaxation rates were not considered accurate. For residues that had a slower relaxation rate as in residue 141 the fitting was more accurate even at longer relaxation delay times. Nevertheless, for fast relaxing residues as for residue 254 curve fitting was indicative of the relaxation process taking place. Relaxation rates calculated reflected the elevated rates that were expected.

Finally on the J domain region there appears to be some difference observed in relaxation rates, in support of initial analysis showing that the J domain needs to be present for effective pRb binding with the PyLTNT protein, as mentioned earlier in section 1.4. This difference is mainly observed for residues 46 and 47 and seems to be in a region that has been shown to have large R_2 values after complex formation with the Hsc70 protein (Berjanskii et al., 2002).

3.5. SAXS

Previous studies have shown that comparison of the crystal structures with the generated *ab-initio* models from SAXS are very accurate (Tsutakawa et al., 2007). Here SAXS was used to further analyse the unfolded nature of PyLTNT and to provide additional information on the protein complex. Models were fitted manually using existing pdb structures with the aid of the program PyMol.

Data manipulation for the final representation of the structures that are given here underwent through a series of evaluations. A number of parameters were examined including radius of gyration (*Rg*), scattering intensity at zero angle or *I(0)*, the pair-distance distribution function or *P(R)*, the scattering intensity or *I(q)*, and the momentum transfer (*q*). These parameters can be plotted with each other to give information on the molecular weight, molecular size, compactness, aggregation state, globularity, molecular shape and tertiary fold.

Initial information about the protein can be obtained by the data via the scattering plot that is derived by plotting log *I(q)* versus *q*. The scattering plot contains all the structural information for calculation of the structure.

The *P(R)* function depicts the paired-set of all distances between points within an object. This function can be used in SAXS to describe the paired-set of distances, within any given macromolecular structure, between all of the electrons allowing the detection of conformational changes. This implies that even small changes within the macromolecule can result in detectable changes in a *P(R)* distribution. Plotting *P(R)* versus *R* where *R* is the distance, can give information on the globularity of the protein.

The *Rg* and *I(0)* are pivotal structural parameters. *I(0)* gives the molecular weight in kDa, while *Rg* is a measure of a size or a molecular compactness. These two

parameters are obtained from a Guinier plot, ln *I(q)* versus q^2. However the Guinier plot utilises data from low scattering angles making it more sensitive towards small amounts of aggregation (Kataoka and Goto, 1996). The *P(R)* function is a better alternative for obtaining the *Rg* and *I(0)* values as it uses all data collected.

Additional information can be obtained from the Kratky plot that reports on the molecular shape of the protein. The shape of the plot can be obtained by plotting $I(q)*q^2$ versus *q*. NUPs appear in the diagrammatic representation of the scattering data (Kratky plots), as being disordered by the characteristic absence of a bell-shaped curve. Because these proteins have a low intermolecular packing density, only residual structure can be detected. This is observed in the Kratky plot as a “bump” where the bell-shape would be observed for folded proteins, which is followed by a plateau. In the case of PyLTNT the Kratky plot is characteristic of a partially folded protein as described in Putnam et al. (Putnam et al., 2007).

For each protein the scatering intensity, *P(R)* function, Kratky, and Guinier plots are given below in Figures 44-46. (Kataoka and Goto, 1996)

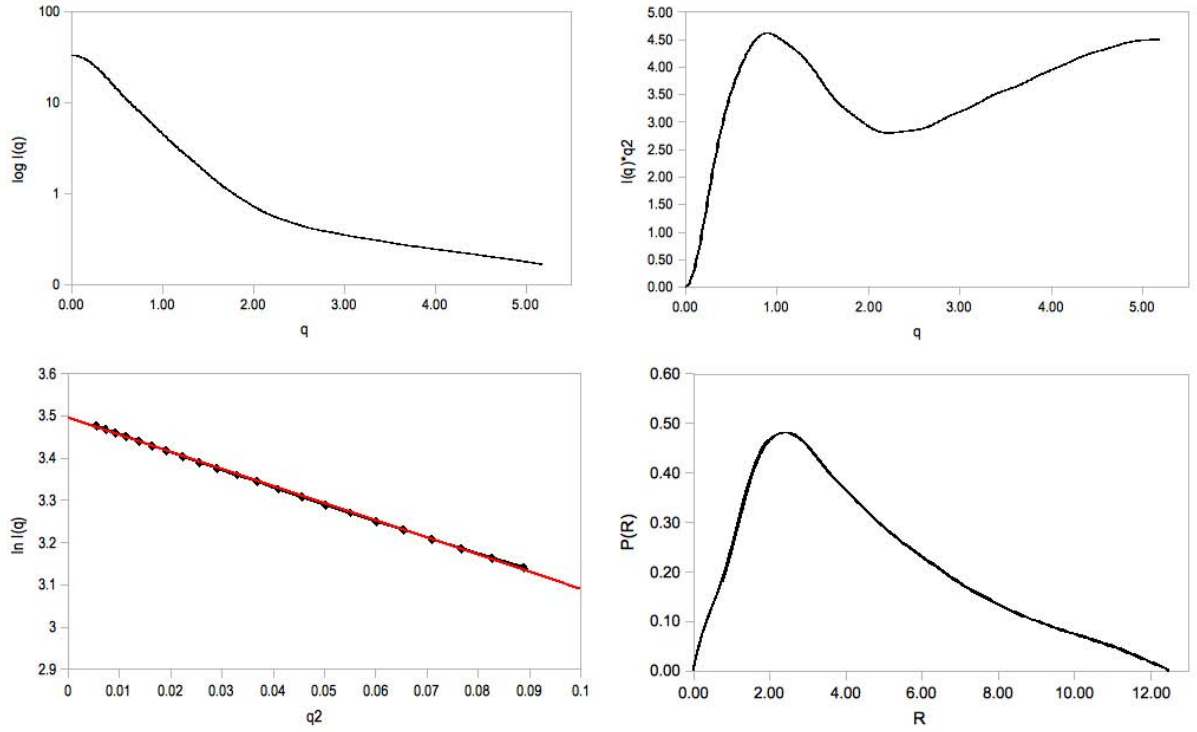


Figure 44. SAXS data plots for PyLTNT

(A) Scattering curve of the PyLTNT data given by logarithm of scattering intensity I(q) versus the momentum transfer q. Unit for q is in nm^{-1}. (B) Kratky plot of the PyLTNT protein given by $I(q)*q^2$ versus q. The partially unfolded nature of the protein is reflected in the increasing values in the larger q range. (C) Guinier plot of PyLTNT given by the natural logarithm I(q) versus q^2. Rg =3.62 ± 0.026, I(0)= 33.09 ± 0.1970. Slight overestimation of values was due to slight aggregation. However, Guinier plot values are more sensitive towards slight aggregation as they use data from low scattering angles. Values were estimated by the slope of the Guinier fitting (red line). (D) Pair-distance distribution function given by plotting of P(R) versus R given in nm. While the single peak is an indication of a globular protein the fact that the tail at a large r range is long indicates a elongated molecule. This can be a result of the globular J domain while the remaining protein is unfolded and therefore elongated.

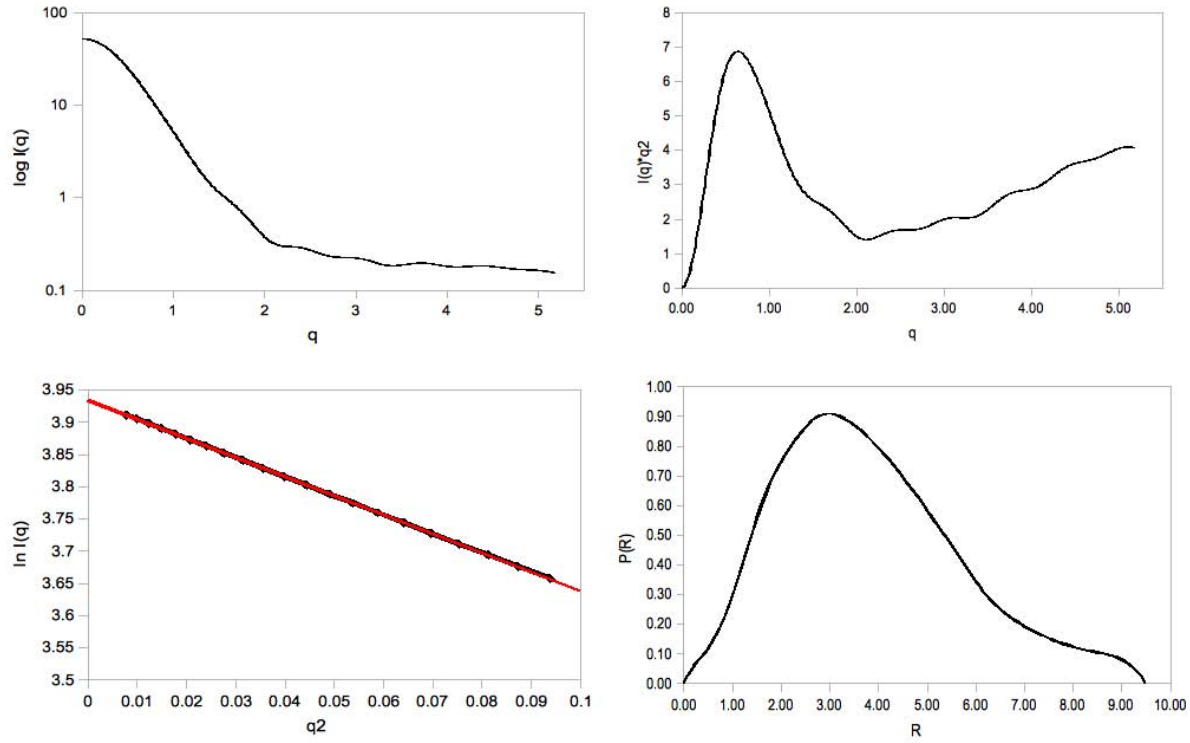


Figure 45. SAXS data plots for pRb

(A) Scattering curve of the pRb data given by logarithm of scattering intensity I(q) versus the momentum transfer q. Unit for q is in nm^{-1}. (B) Kratky plot of the pRb protein given by $I(q)*q^2$ versus q. The plot is indicative of a folded protein. (C) Guinier plot of pRb given by ln I(q) versus q^2. Rg = 3.01 ± 0.013, I(0)=51.16 ± 0.2258. Slight overestimation of values was due to slight aggregation. Values were estimated by the slope of the Guinier fitting (red line). (D) Pair-distance distribution function given by plotting of P(R) versus R given in nm. The single peak is an indication of a globular protein.

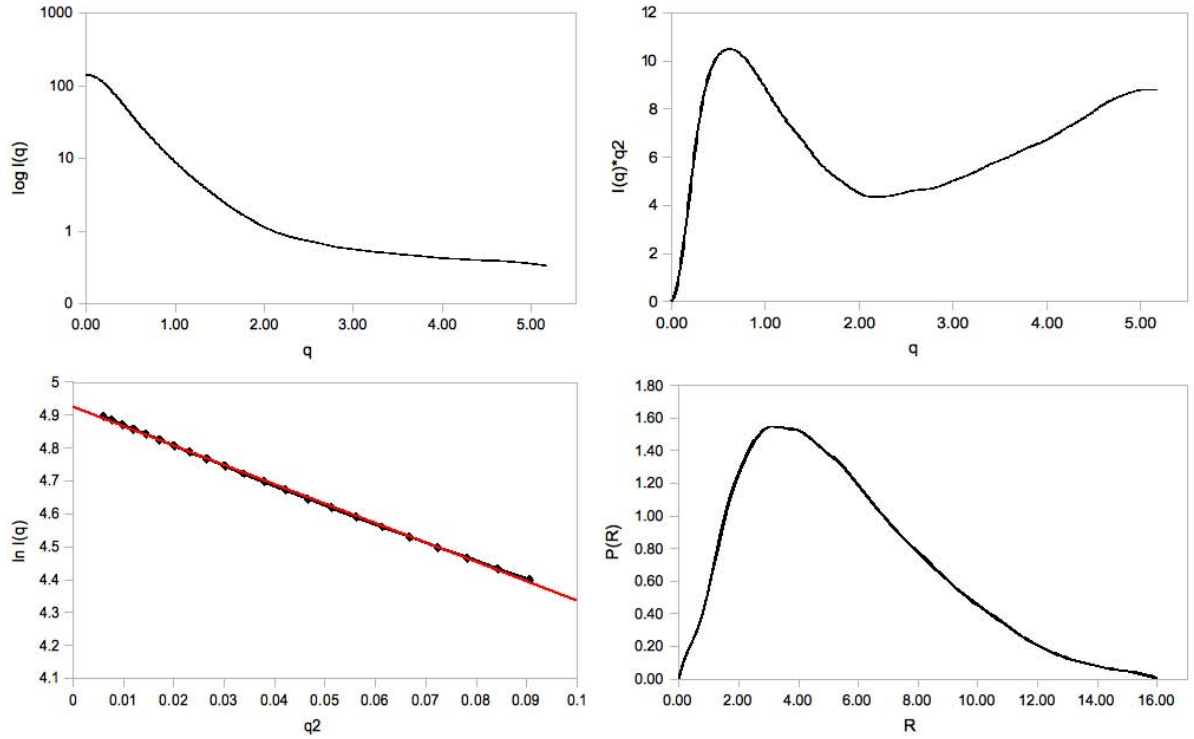


Figure 46. SAXS data plots for PyLTNT complex with pRb

(A) Scattering curve of the PyLTNT/pRb data given by log of scattering intensity I(q) versus q. Unit for q is in nm^{-1}. (B) Kratky plot of the PyLTNT/pRb protein given by I(q) $*q^2$ versus q. The partially unfolded nature indicated by the plot possibly arises from the complex of the folded pRb and largely unfolded PyLTNT/pRb. (C) Guinier plot of PyLTNT/pRb given by ln I(q) versus q^2. Rg =4.40 ± 0.027, I(0)=138.9 ± 0.6416. Overestimation of values was due to sample aggregation.Values were estimated by the slope of the Guinier fitting (red line). (D) Pair-distance distribution function given by plotting of P(R) versus R given in nm. The single peak is an indication of a globular protein attributed most likely to the pRb protein. However, the tail of the plot at larger R values is an indication of an elongated structure that most probably arises from the attached PyLTNT protein.

PyLTNT on its own was represented by an elongated model (Figure 47), the same approach was taken for the truncated PyLTNTm2. The J domain was fitted to the model from its existing NMR structure (pdb:1FAF). Additionally, pRb was represented

by a more globular structure that fitted well with the existing pRb structure (pdb: 1GH6) ignoring the HisTag in the protein Figure 48.

As seen in Figure 38 the J domain structure superimposed surprisingly well with the SAXS generated model of PyLTNT, especially considering that the protein is flexible. The SAXS models of PyLTNT/PyLTNTm2 with pRb were difficult to fit manually with the models of the J domain and pRb, owing to lack of connectivity between the two structures and lack of any reference point (Figure 49). The pictures seen here represent best manual fits of the structures into the SAXS models. Fortunately, the superimposition was facilitated by the difference in size of the molecular structures. Additionally, while values of the Guinier plots were slightly overestimated, predicted models obtained by SAXS data fitted well with existing X-ray and NMR structures.

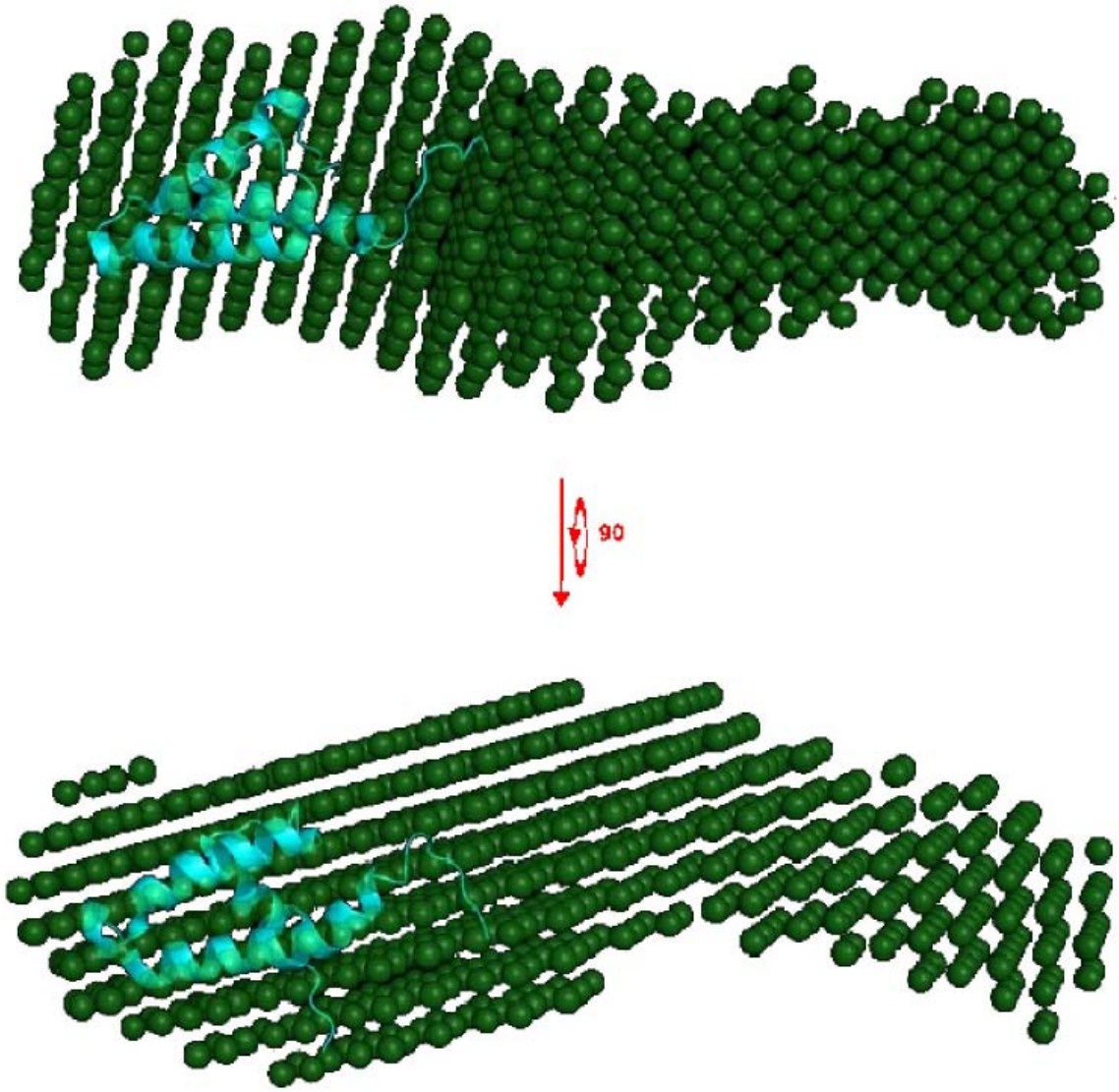


Figure 47. SAXS model of PyLTNT superimposed with the NMR J domain structure.

While the J domain is an N-terminal protein it was difficult to manually fit the structure without knowing at which end the N-terminus is located. Fortunately the J domain structure and the unfolded nature of the rest of the protein gave valuable hints regarding its position in the model. Another noticeable feature, is that the model is more compact than expected for a completely unfolded protein. This might be another indication of residual structure.

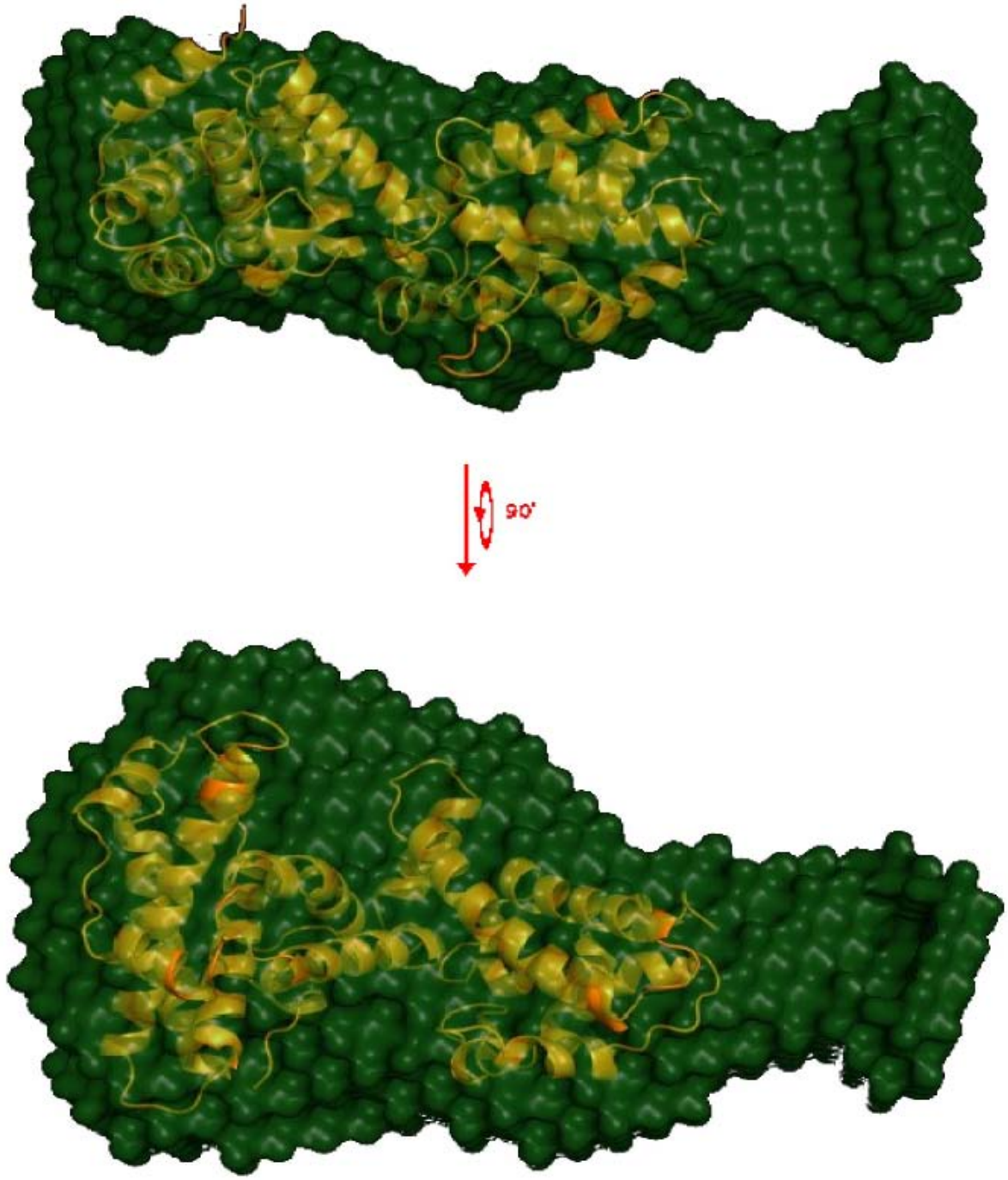


Figure 48. SAXS model of pRb

Superimposition of the pRb pdb structure with the SAXS generated model shows a good fit. The HisTag is expected to be in the unfilled area of the model.

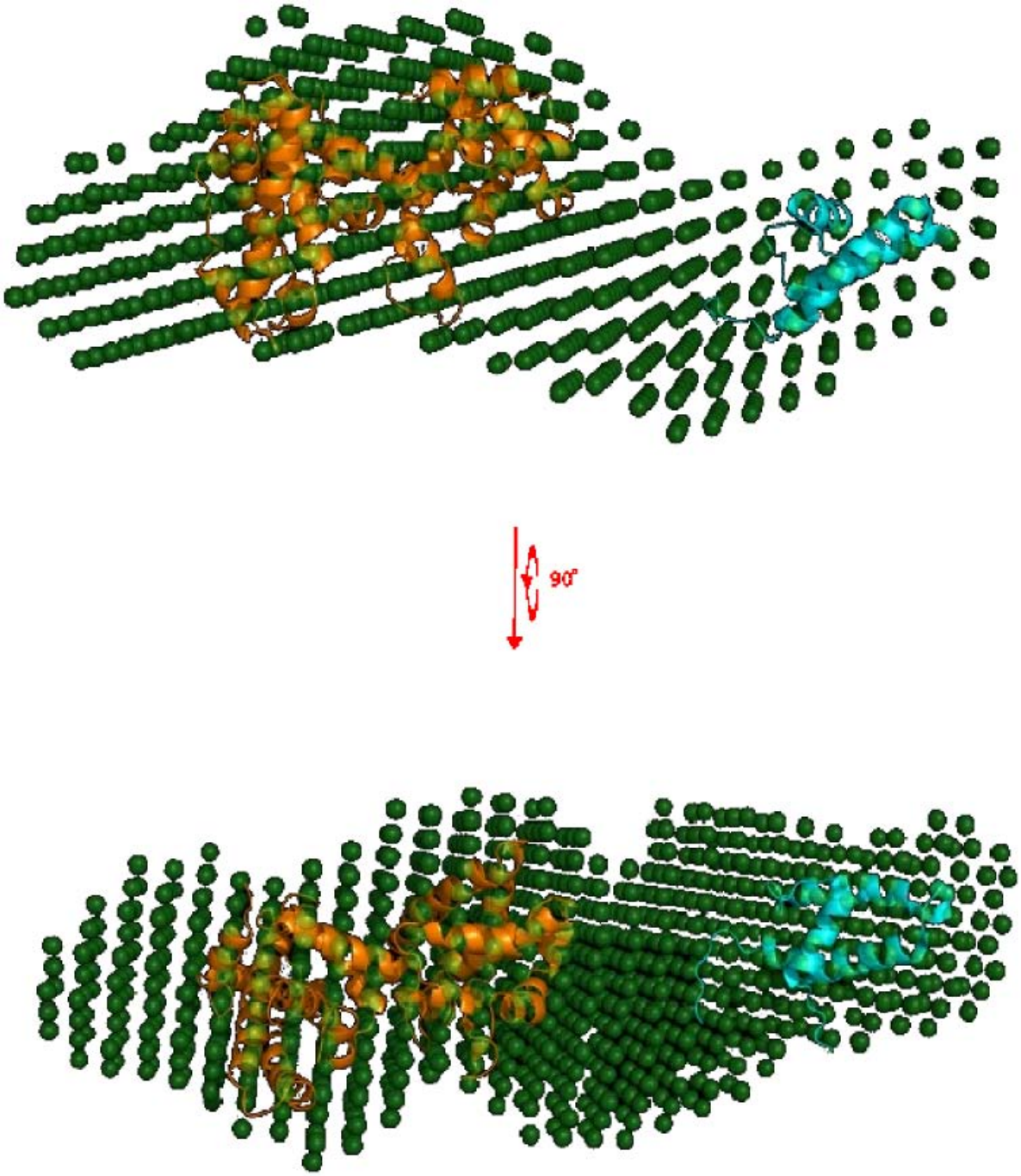


Figure 49. SAXS model of PyLTNT in complex with pRb

Structures of the J domain and pRb are superimposed and manually fitted.

Superimposition of existing pdb models with SAXS *ab-initio* structure reconstruction was able to resolve where the missing structural elements might be located. Alignment of the SAXS data with models generated from model prediction software

could potentially make for a better representation. Unfortunately this was beyond the scope of this thesis.

The 3D reconstruction of these maps with the corresponding molecular structures show a picture of the two molecules attached side by side like a sliding mechanism. As well, a more compact structure is seen for PyLTNT both in its free state and in the complex suggesting a greater tendency for structure than previously expected. The tail seems to attach to pRb and then twist towards itself.

CHAPTER IV – Discussion

4.1. Binding of pRb with PyLTNT

The relevance of the pRb/LT interaction for cell cycle control is well established and relevant interactions derived from biochemical investigations were previously reviewed by DeCaprio (DeCaprio, 2009). However, there has been a lack of structural data, and the current understanding of the role of large unstructured sequences in viral oncoproteins is limited. In this thesis ITC, NMR and SAXS analysis have been used to identify determinants of the PyLTNT-pRb interaction and to probe for localised structure formation triggered by the interaction.

Previous work with LxCxE peptides bound to pRb showed a 110 nM dissociation constant, and competition experiments show a 20-fold reduction in affinity compared with the full length E7 protein (Lee et al., 1998). The affinity of PyLTNT determined by ITC, was similar to that previously reported for the LxCxE peptide, but a smaller construct (PyLTNTm2) showed a 3-fold reduction in binding. The PyLTNT protein has a low binding affinity compared to the full length E7 protein, although they perform the same function by inhibiting pRb. However, the two proteins have a weak sequence homology and have limited conserved regions (Lee and Cho, 2002).

One explanation could be that only the LxCxE motif contributes to affinity. However, this assumption is contradicted by NMR results, that clearly show substantial changes in R_2 relaxation rates in a region around the LxCxE motif with an additional isolated but strong effect in the C-terminus of the protein.

The difference may also be linked to the balance between entropy and enthalpy owing to the interactions, the formation of secondary structure and furthermore, the formation of limited tertiary structure considering the compactness of the complex as determined by SAXS.

Taking all results into account a picture emerges where the PyLTNT and pRb interaction requires several clusters of residues in the protein to be present. The LxCxE motif and the J domain were previously reported to be contributing to the interaction. The NMR data presented in this thesis support the role of the LxCxE motif for pRb interaction, but not a strong contribution of the J domain as previously suggested (DeCaprio, 1999). However, transactivation of promoters is J domain independent, as previous modifications of the downstream functions linked to the phosphorylation state of pRb and p130 suggest (Sheng et al., 2000). This does not exclude the possibility of the J domain acting as a molecular chaperone, and recruiting Hsc70 for efficient binding of pRb. Only a few residues of the J domain are affected by binding and its residues exhibit only intermediate exchange indicative of a low affinity interaction.

SAXS confirmed the domain structure in the overall model of the protein complex. From this model the proteins seem to bind where expected at the B domain of pRb and the LxCxE motif of PyLTNT. What is also noticeable is that the PyLTNT unfolded part, downstream of the LxCxE motif, is folded back towards itself creating a loop that gives the protein a more compact structure. This was recognised by visually checking a completely unfolded model with the models produced by SAXS. Another aspect is that the proteins bind in a sideways manner possibly creating a sliding mechanism.

4.2. Identification of Additional Binding Sites

This project germinated from the question of whether there are additional, not yet discovered sites in the PyLTNT sequence, that contribute towards pRb binding. This thesis shows that PyLTNT is strongly associated with pRb and this involves additional interacting sites besides what was previously known. The NMR data showed that polyoma regions associated with retinoblastoma (PRARs) do exist but their importance remains enigmatic. In addition, it remains unknown whether these sites might play a role for other proteins in the retinoblastoma family. Efforts to prepare p130 in sufficient amounts for further NMR experiments failed as the protein proved very unstable *in vitro*.

The current data suggests a PRAR before the LxCxE motif that forms a tight association with pRb. Residues 133-137 seem to be indispensable for the polyoma LT association. Unfortunately truncated versions of the protein for the evaluation of the importance of these residues are not possible, because the LxCxE motif is needed for the interaction.

Other regions that showed localised structure in R_2 maps, include residues downstream of the LxCxE motif. Clusters of residues (161-166; 172-174; 177-181) were strongly affected, although HSQC spectra indicate fast exchange and therefore suggest a weak interaction. These PRARs have not yet been evaluated by mutational analysis to assess their binding efficiency, which was beyond the scope of this thesis.

Random interactions were observed at the C-terminus of the protein beyond residue 200 mainly from titration data, with one great R_2 effect at residue 254 where R_2 was already high in the free protein. Certain residues were in fast exchange in titration data when bound to pRb but, after removal of the last 58 residues the truncated

construct exhibited only a 3-fold reduction in binding efficiency which seems relatively insignificant. This suggests that these residues might randomly interact with pRb due to the flexibility of the unfolded protein tail, especially after residue 200 where six consecutive glycines attribute to additional conformational flexibility.

The SAXS models suggest that residues beyond position 200 are not attached to the pRb protein. This is in agreement with the R_2 data, that show most of the residues involved in secondary structure formation before residue 200.

4.3. Significance of New Binding Sites

Although the functional role of the additional binding sites was not further probed in this thesis, evidence suggests that the virus uses the unfolded nature of the LT for a multitude of interactions targeting different cellular proteins, some of which may still be unknown. One can speculate that the PyV virus was the ancestor strain from which the other viral homologues originate. This conclusion is supported by the fact the 154 additional residues are unique to polyoma, that the protein causes tumours only in rodents, and that SV40 that can infect humans has abolished these sites.

This means that some of the polyoma protein's original binding sites were lost in subsequent viral homologues to be replaced by more efficient ones. This view is supported by the lack of any direct binding with p53 in PyLT whereas all other variants of the polyoma family have a direct p53 binding site at their C-terminus of LT. In part, PyV has the additional MT protein which is not present in most other family members. MT binds to p21, a protein downstream the p53 cascade but, does also not directly bind p53.

The PRAR regions are composed of sequential residues forming clusters on the PyLTNT sequence. Some of the interactions appear less significant and might be attributed to the packing of the protein in the complex affecting localised flexibility rather than being directly responsible for interaction efficiency. The fact that the PyLTNTm2 shows modest but not a dramatic reduction in binding affinity indicates that the last 58 amino acids are not directly important for binding affinity for pRb. The relatively strong affinity (~120nM) exhibited by PyLTNT-pRb complex formation with pRb might also be partially attributed to amino acids 256 to 258. Previous work on mutants of PyLT with deletions of the region 256-272 resulted in a 3-fold increase in binding affinity of pRb by PyLT (Pilon et al., 1996). Considering that the protein used

in this thesis was lacking those amino acids, and that the last two residues of PyLTNT were substituted to add a HisTag for purification, it might have the same properties as the one previously reported.

The same deletion mutant has also been reported to bind only 22% of available p107 (Pilon et al., 1996). Similar results have been reported by substitution of the acidic residues in the 155-159 region, that exhibited a two-fold increase in binding capacity of the wild-type PyLT to pRb while binding only 23% of the available p107 (Pilon et al., 1996). In the amino acids involved in the pRb complex formation, residues 155S and 156S are shown, by NMR, to have no elevated relaxation rate indicating that they are not important for the interaction while residues 158T and 159D had increased relaxation rates. In Pilon et al., mutations replacing the acidic residues increased affinity for pRb. However, p107 binding significantly decreased to 20-25% of that of the wild-type PyLT. The virus has evolved towards an overall optimisation of interrupting cell cycle control and not for the optimal interaction with one specific protein.

In addition residues that are homologous to the SV40 sequence, just like 155S and 156S that are important for binding, reside downstream of these residues and form a new PRAR of their own (177-181). These residues were shown to have a low affinity for pRb interaction.

4.4. Future Prospects

A number of additional experiments could be envisioned to pinpoint the exact relation between the newly discovered PRARs and the efficiency of binding.

Future studies should address the interaction between the PyLTNT and its truncated forms, with the other members of the retinoblastoma family like p107 and p130. These experiments could potentially highlight different PRARs according to the different retinoblastoma proteins.

Future experiments using selective labelling might yield further assignments of the PyLTNT sequence. A method that might help to overcome scrambling is the addition of excess of all amino acid types in unlabeled form to the rich medium, with the exception of the one labelled amino acid of interest. Other alternative solutions to selective labelling include use of transaminase deficient strains, or strains that include rare tRNA codon genes (Tong et al., 2008). Another method includes inhibition of the feedback loops of amino acid metabolic pathways. Finally, a method that is becoming increasingly popular could be used that includes cell free expression for selective labelling.

On the topic of selective labelling, a popular method that could potentially be used for this protein is stereo-array isotope labelling (SAIL). The method relies on amino acid stereo-selective replacement as described by Kainosho et al. and is especially effective using cell-free expression (Kainosho et al., 2006). Briefly, one of the protons in methylene groups is replaced by a deuterated proton, two of the protons in methyl groups are replaced by deuterated protons, valine and leucine prochiral methyl groups are both deuterated such that one methyl is $-^{12}C(D)_3$ and the other is $-^{13}CH(D)_2$ and last, alternate labelling in the aromatic rings of phenylalanine, tyrosine

and tryptophan such that one is in the form of -^{12}CD and the other is in the form of -^{13}CH.

The advantage of this method is that the backbone connectivity information is preserved and linewidths are sharpened by decrease in dipolar relaxation pathways and long-range couplings. The main disadvantage of this method its prohibitive cost.

In conclusion while a number of possible methods and experiments have been performed, additional studies may be needed to assess the protein functionality. This may also include cell studies to determine the effect of PyLTNT after mutations of the PRARs.

4.5. Conclusions

> "I think that only daring speculation can lead us further and not accumulation of facts" – Albert Einstein 1879 – 1955

The results presented in this thesis lead to the conclusion that PyLTNT is natively unfolded to optimise interactions with many other proteins, among other pRb. Some of the localised folding induced by pRb binding, can be attributed to direct interactions whereas others may just be packing effects caused by the interaction. Specifically, it was possible to show that a sequence of amino acids between resides 151-200 and residues 133-137 is crucial for the pRb interaction. These sequential residues were termed PRARs because they were identified by R_2 analysis to be involved in localised structure formation upon pRb binding.

The work in this thesis was somewhat overwhelmed by the difficulty of assigning a 28kDa protein including the largest unfolded sequence assigned to date. This assignment provides the ground work for further analyses. As part of this thesis, an attempt was made to also characterise interactions with p130 and p107, but those failed because both proteins were too unstable. Future work should address these other binding partners as it remains a key question whether the observed PRARs would be different for the different retinoblastoma proteins, p107 or p130. The experiments by Pilon et al. strongly support such a hypothesis (Pilon et al., 1996).

The overall results of this thesis reveal an amazing picture regarding viral oncoproteins. Rather than optimising one interaction they contain sequences for many specific binding partners for which they reveal remarkably high affinities. Upon binding, the sequence shows features of localised structure formation for both direct interactions and, likely also, for regions which are more indirectly involved in the

interaction. The sequence of viral oncoproteins seems optimised to balance for promiscuous interactions with different binding partners.

Whilst previous studies of polyoma virus have mainly been driven by virologists, who used this virus as a model to decipher mechanisms of cellular transformation, this picture has radically changed as polyoma viruses have been associated with various diseases affecting humans. It is to be expected that the groundwork laid out in this thesis will help towards functional understanding with relevance for the development of novel treatments against diseases caused by polyoma viruses.

Appendix

	#	Dimension of acquired data (nucleus)			Spectral width, kHz		
Experiments		t_1	t_2	t_3	F_1	F_2	F_3
HSQC†*	16	256(^{15}N)	1024(^{1}H)		2.5	5.4	
HNCO*	4	1024(^{1}H)	64(^{13}C)	48(^{15}N)	6.5	3.8	2.5
HNCA*	16	1024(^{1}H)	42(^{13}C)	48(^{15}N)	6.5	6	2.5
HN(CO)CA*	16	1024(^{1}H)	42(^{13}C)	48(^{15}N)	6.5	6	2.5
HNN*	16	1024(^{1}H)	60(^{15}N)	68(^{15}N)	6.5	2.5	2.5
HN(C)N*	32	1024(^{1}H)	44(^{15}N)	44(^{15}N)	6.5	2.5	2.5
CBCA(CO)NH*	16	1024(^{1}H)	64(^{13}C)	48(^{15}N)	13	16	2.5
HNCACB*	16	1024(^{1}H)	64(^{13}C)	48(^{15}N)	6	16	2.5
CON‡	8	768(^{13}C)	52(^{15}N)		4	2.5	
(H)CBCACON‡	8	768(^{13}C)	84(^{15}N)	128(^{13}C)	2	2.5	11
(H)CBCANCO‡	8	768(^{13}C)	64(^{15}N)	124(^{13}C)	4	2.5	11
HSQC-NOESY*∂	8	1057(^{1}H)	48(^{15}N)	154(^{1}H)	6.5	13	2.5
HSQC-TOCSY*∂	8	945(^{1}H)	42(^{15}N)	128(^{1}H)	6	12	2.5

Table 1. Acquisition parameters of NMR experiments.

‡ data were acquired on an 700MHz spectrometer with a ^{13}C direct observe probe.

* data were acquired on an 800MHz spectrometer.

† data were acquired on an 900MHz spectrometer.

number of scans

∂ The mixing times for HSQC-NOESY and HSQC-TOCSY were 200 and 140 ms respectively.

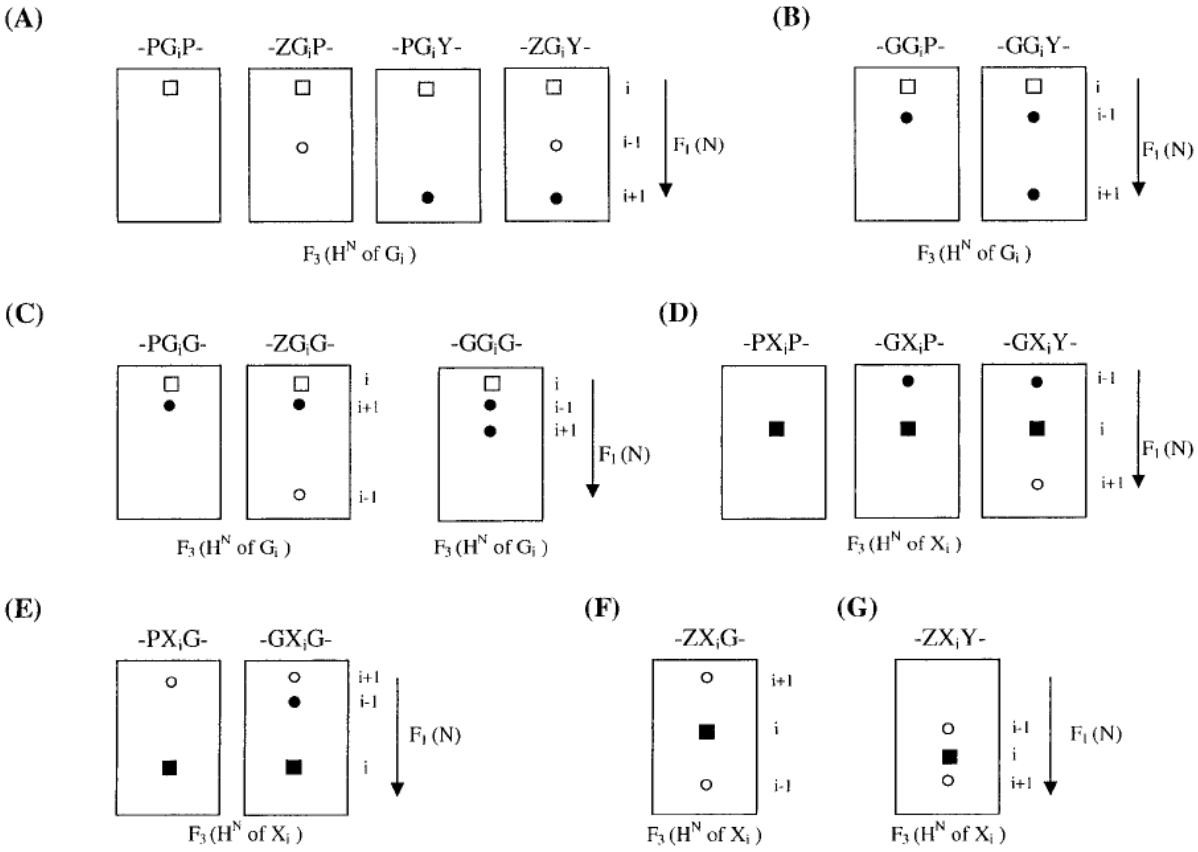


Figure 50. Peak patterns of various triplets for the HNN experiment in the F1(^{1H})-F3(^{15}N) planes

(A-C) F2 (^{15}N) chemical shift of the central glycine (i residue). (D-G) F2 chemical shift of central X amino acid (i residue). The diagonal (F1 = F2) peaks are indicated in squares and the sequential peaks in circles. Positive and negative signs are represented by filled and empty square/circles, respectively. The arrows indicate the direction of increase in the ^{15}N chemical shift values. The ^{15}N chemical shift of glycine is assumed to be upfield compared to that of the other amino acid residues. For the other residues the ^{15}N chemical shifts are arbitrary. With kind permission from Springer Science+Business Media: (Panchal et al., 2001).

References

Ahuja, D, Sáenz-Robles, MT, Pipas, JM (2005) SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene*, **24:** 7729-7745

Andrade, MA, Chacon, P, Merelo, JJ, Moran, F (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng*, **6:** 383-390

Bax, A, Ikura, M (1991) An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J Biomol NMR*, **1:** 99-104

Benjamin, TL (2001) Polyoma virus: old findings and new challenges. *Virology*, **289:** 167-173

Bennoun, M, Grimber, G, Couton, D, Seye, A, Molina, T, Briand, P, Joulin, V (1998) The amino-terminal region of SV40 large T antigen is sufficient to induce hepatic tumours in mice. *Oncogene*, **17:** 1253-1259

Berjanskii, MV, Riley, MI, Xie, A, Semenchenko, V, Folk, WR, Van Doren, SR (2000) NMR structure of the N-terminal J domain of murine polyomavirus T antigens. Implications for DnaJ-like domains and for mutations of T antigens. *J Biol Chem*, **275:** 36094-36103

Berjanskii, M, Riley, M, Van Doren, SR (2002) Hsc70-interacting HPD loop of the J domain of polyomavirus T antigens fluctuates in ps to ns and micros to ms. *J Mol Biol*, **321:** 503-516

Bermel, W, Bertini, I, Csizmok, V, Felli, IC, Pierattelli, R, Tompa, P (2009) H-start for

exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins. *J Magn Reson*, **198:** 275-281

Bermel, W, Bertini, I, Duma, L, Felli, IC, Emsley, L, Pierattelli, R, Vasos, PR (2005) Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew Chem Int Ed Engl*, **44:** 3089-3092

Bermel, W, Bertini, I, Felli, IC, Lee, Y-M, Luchinat, C, Pierattelli, R (2006) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J Am Chem Soc*, **128:** 3918-3919

Bochkareva, E, Martynowski, D, Seitova, A, Bochkarev, A (2006) Structure of the origin-binding domain of simian virus 40 large T antigen bound to DNA. *EMBO J*, **25:** 5961-5969

Bollag, B, Prins, C, Snyder, EL, Frisque, RJ (2000) Purified JC virus T and T' proteins differentially interact with the retinoblastoma family of tumor suppressor proteins. *Virology*, **274:** 165-178

Brodsky, JL, Pipas, JM (1998) Polyomavirus T antigens: molecular chaperones for multiprotein complexes. *J Virol*, **72:** 5329-5334

Bryson, K, McGuffin, LJ, Marsden, RL, Ward, JJ, Sodhi, JS, Jones, DT (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res*, **33:** W36-8

Burkhart, DL, Sage, J (2008) Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat Rev Cancer*, **8:** 671-682

Caffrey, M, Simorre, JP, Cusanovich, M, Marion, D (1995) Characterization of the dynamic properties of Rhodobacter capsulatus ferricytochrome c'--a 28 kDa

paramagnetic heme protein. *FEBS Lett*, **368:** 519-522

Caracciolo, V, Reiss, K, Khalili, K, De, F, G, Giordano, A (2006) Role of the interaction between large T antigen and Rb family members in the oncogenicity of JC virus. *Oncogene*, **25:** 5294-5301

Carbone, M (1999) Simian virus 40 and human tumors: It is time to study mechanisms. *J Cell Biochem*, **76:** 189-193

Carbone, M, Pass, HI, Miele, L, Bocchetta, M (2003) New developments about the association of SV40 with human mesothelioma. *Oncogene*, **22:** 5173-5180

Carson, KR, Focosi, D, Major, EO, Petrini, M, Richey, EA, West, DP, Bennett, CL (2009) Monoclonal antibody-associated progressive multifocal leucoencephalopathy in patients treated with rituximab, natalizumab, and efalizumab: a Review from the Research on Adverse Drug Events and Reports (RADAR) Project. *Lancet Oncol*, **10:** 816-824

Cavanagh, J, Fairbrother, WJ, Palmer III, AG, Skelton, NJ, Rance, M. (2006) *Protein NMR Spectroscopy: Principles and Practice*. London: Elsevier,

Chatterjee, A, Kumar, A, Chugh, J, Srivastava, S, Bhavesh, NS, Hosur, RV (2005) NMR of unfolded proteins. *Journal of Chemical Sciences*, **117:** 3-21

Chatterjee, A, Kumar, A, Hosur, RV (2006) Alanine check points in HNN and HN(C)N spectra. *J Magn Reson*, **181:** 21-28

Chugh, J, Kumar, D, Hosur, RV (2008) Tuning the HNN experiment: generation of serine-threonine check points. *J Biomol NMR*, **40:** 145-152

Claridge, TDW. (1999) *High-resolution NMR techniques in organic chemistry*.

London: Elsevier,

Cobrinik, D (2005) Pocket proteins and cell cycle control. *Oncogene*, **24:** 2796-2809

Coeytaux, K, Poupon, A (2005) Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics*, **21:** 1891-1900

Cuff, JA, Clamp, ME, Siddiqui, AS, Finlay, M, Barton, GJ (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics*, **14:** 892-893

DeCaprio, JA (1999) The role of the J domain of SV40 large T in cellular transformation. *Biologicals*, **27:** 23-28

DeCaprio, JA (2009) How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40. *Virology*,

Delaglio, F, Grzesiek, S, Vuister, GW, Zhu, G, Pfeifer, J, Bax, A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, **6:** 277-293

Dilworth, SM (2002) Polyoma virus middle T antigen and its role in identifying cancer-related molecules. *Nat Rev Cancer*, **2:** 951-956

Dyson, HJ, Wright, PE (1998) Equilibrium NMR studies of unfolded and partially folded proteins. *Nat Struct Biol*, **5 Suppl:** 499-503

Dyson, HJ, Wright, PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, **6:** 197-208

Eash, S, Manley, K, Gasparovic, M, Querbes, W, Atwood, WJ (2006) The human polyomaviruses. *Cell Mol Life Sci*, **63:** 865-876

Engels, EA, Katki, HA, Nielsen, NM, Winther, JF, Hjalgrim, H, Gjerris, F, Rosenberg, PS, Frisch, M (2003) Cancer incidence in Denmark following exposure to poliovirus vaccine contaminated with simian virus 40. *J Natl Cancer Inst*, **95:** 532-539

Englander, J, Cohen, L, Arshava, B, Estephan, R, Becker, JM, Naider, F (2006) Selective labeling of a membrane peptide with 15N-amino acids using cells grown in rich medium. *Biopolymers*, **84:** 508-518

Felsani, A, Mileo, AM, Paggi, MG (2006) Retinoblastoma family proteins as key targets of the small DNA virus oncoproteins. *Oncogene*, **25:** 5277-5285

Ferber, D (2002) Virology. Monkey virus link to cancer grows stronger. *Science*, **296:** 1012-1015

Fink, AL (2005) Natively unfolded proteins. *Curr Opin Struct Biol*, **15:** 35-41

Foster, MP, McElroy, CA, Amero, CD (2007) Solution NMR of large molecules and assemblies. *Biochemistry*, **46:** 331-340

Gai, D, Zhao, R, Li, D, Finkielstein, CV, Chen, XS (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell*, **119:** 47-60

Garimella, R, Liu, X, Qiao, W, Liang, X, Zuiderweg, ERP, Riley, MI, Van, D, Steven R (2006) Hsc70 contacts helix III of the J domain from polyomavirus T antigens: addressing a dilemma in the chaperone hypothesis of how they release E2F from pRb. *Biochemistry*, **45:** 6917-6929

Genevaux, P, Schwager, F, Georgopoulos, C, Kelley, WL (2002) Scanning mutagenesis identifies amino acid residues essential for the in vivo activity of

the Escherichia coli DnaJ (Hsp40) J-domain. *Genetics*, **162:** 1045-1053

Giacinti, C, Giordano, A (2006) RB and cell cycle progression. *Oncogene*, **25:** 5220-5227

Gjørup, OV, Rose, PE, Holman, PS, Bockus, BJ, Schaffhausen, BS (1994) Protein domains connect cell cycle stimulation directly to initiation of DNA replication. *Proc Natl Acad Sci U S A*, **91:** 12125-12129

Gottlieb, KA, Villarreal, LP (2001) Natural biology of polyomavirus middle T antigen. *Microbiol Mol Biol Rev*, **65:** 288-318 ; second and third pages, table of contents

Greene, MK, Maskos, K, Landry, SJ (1998) Role of the J-domain in the cooperation of Hsp40 with Hsp70. *Proc Natl Acad Sci U S A*, **95:** 6108-6113

Grishaev, A, Tugarinov, V, Kay, LE, Trewhella, J, Bax, A (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J Biomol NMR*, **40:** 95-106

Gross, L (1997) The role of viruses in the etiology of cancer and leukemia in animals and in humans. *Proc Natl Acad Sci U S A*, **94:** 4237-4238

Grzesiek, S, Dobeli, H, Gentz, R, Garotta, G, Labhardt, AM, Bax, A (1992) 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry*, **31:** 8180-8190

Günther, UL, Ludwig, C, Ruterjans, H (2000) NMRLAB-Advanced NMR data processing in matlab. *J Magn Reson*, **145:** 201-208

Hoult, DI (1976) Solvent peak saturation with single phase and quadrature Fourier transformation. *J. Magn. Reson*, **21:** 337

Howes, SH, Bockus, BJ, Schaffhausen, BS (1996) Genetic analysis of polyomavirus large T nuclear localization: nuclear localization is required for productive association with pRb family members. *J Virol*, **70:** 3581-3588

Huang, K, Flanagan, JM, Prestegard, JH (1999) The influence of C-terminal extension on the structure of the "J-domain" in E. coli DnaJ. *Protein Sci*, **8:** 203-214

Hwang, TL, Shaka, AJ (1995) Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A*, **112:** 275-279

Iakoucheva, LM, Brown, CJ, Lawson, JD, Obradovic, Z, Dunker, AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*, **323:** 573-584

Ichaso, N, Dilworth, SM (2001) Cell transformation by the middle T-antigen of polyoma virus. *Oncogene*, **20:** 7908-7916

Kaelin, WGJ, Ewen, ME, Livingston, DM (1990) Definition of the minimal simian virus 40 large T antigen- and adenovirus E1A-binding domain in the retinoblastoma gene product. *Mol Cell Biol*, **10:** 3761-3769

Kainosho, M, Torizawa, T, Iwashita, Y, Terauchi, T, Mei Ono, A, Guntert, P (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature*, **440:** 52-57

Karlin, D, Longhi, S, Receveur, V, Canard, B (2002) The N-terminal domain of the phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins. *Virology*, **296:** 251-262

Kataoka, M, Goto, Y (1996) X-ray solution scattering studies of protein folding. *Fold Des*, **1:** R107-14

Kay, LE, Ikura, M, Tschudin, R, Bax, A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson*, **89:** 496-514

Kim, HY, Ahn, BY, Cho, Y (2001) Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. *The EMBO journal*, **20:** 295

Klein-Seetharaman, J, Oikawa, M, Grimshaw, SB, Wirmer, J, Duchardt, E, Ueda, T, Imoto, T, Smith, LJ, Dobson, CM, Schwalbe, H (2009) Long-range interactions within a nonnative protein. *Science*, **295:** 1719-1722

Knoblich, K, Whittaker, S, Ludwig, C, Michiels, P, Jiang, T, Schaffhausen, B, Gunther, U (2009) Backbone assignment of the N-terminal polyomavirus large T antigen. *Biomol NMR Assign*, **3:** 119-123

Knudsen, ES, Wang, JY (1998) Hyperphosphorylated p107 and p130 bind to T-antigen: identification of a critical regulatory sequence present in RB but not in p107/p130. *Oncogene*, **16:** 1655-1663

Knudsen, ES, Knudsen, KE (2008) Tailoring to RB: tumour suppressor status and therapeutic response. *Nat Rev Cancer*,

Koralnik, IJ (2006) Progressive multifocal leukoencephalopathy revisited: Has the disease outgrown its name? *Ann Neurol*, **60:** 162-173

Korenjak, M, Brehm, A (2005) E2F-Rb complexes regulating transcription of genes important for differentiation and development. *Curr Opin Genet Dev*, **15:**

520-527

Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H, Valentin, F, Wallace, IM, Wilm, A, Lopez, R, Thompson, JD, Gibson, TJ, Higgins, DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23:** 2947-2948

Lednicky, JA, Butel, JS (1999) Polyomaviruses and human tumors: a brief review of current concepts and interpretations. *Front Biosci*, **4:** D153-64

Lee, C, Cho, Y (2002) Interactions of SV40 large T antigen and other viral proteins with retinoblastoma tumour suppressor. *Rev Med Virol*, **12:** 81-92

Lee, JO, Russo, AA, Pavletich, NP (1998) Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature*, **391:** 859-865

Levitt, MH. (2008) *Spin dynamics: basics of nuclear magnetic resonance*. West Sussex: Wiley,

Li, D, Zhao, R, Lilyestrom, W, Gai, D, Zhang, R, DeCaprio, JA, Fanning, E, Jochimiak, A, Szakonyi, G, Chen, XS (2003) Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature*, **423:** 512-518

Li, H, Söderbérg, K, Houshmand, H, You, ZY, Magnusson, G (2001) Effect on polyomavirus T-antigen function of mutations in a conserved leucine-rich segment of the DnaJ domain. *J Virol*, **75:** 2253-2261

Lilyestrom, W, Klein, MG, Zhang, R, Joachimiak, A, Chen, XS (2006) Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes Dev*, **20:** 2373-2382

Lin, JY, DeCaprio, JA (2003) SV40 large T antigen promotes dephosphorylation of p130. *J Biol Chem*, **278:** 46482-46487

Linding, R, Russell, RB, Neduva, V, Gibson, TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res*, **31:** 3701-3708

Liu, X, Clements, A, Zhao, K, Marmorstein, R (2006) Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. *J Biol Chem*, **281:** 578-586

Luo, X, Sanford, DG, Bullock, PA, Bachovchin, WW (1996) Solution structure of the origin DNA-binding domain of SV40 T-antigen. *Nat Struct Biol*, **3:** 1034-1039

Moens, U, Van, G, M, Johannessen, M (2007) Oncogenic potentials of the human polyomavirus regulatory proteins. *Cell Mol Life Sci*, **64:** 1656-1678

Nemethova, M, Smutny, M, Wintersberger, E (2004) Transactivation of E2F-regulated genes by polyomavirus large T antigen: evidence for a two-step mechanism. *Mol Cell Biol*, **24:** 10986-10994

Ohlenschläger, O, Seiboth, T, Zengerling, H, Briese, L, Marchanka, A, Ramachandran, R, Baum, M, Korbas, M, Meyer-Klaucke, W, Dürst, M, Görlach, M (2006) Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. *Oncogene*, **25:** 5953-5959

Panchal, SC, Bhavesh, NS, Hosur, RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. *J Biomol NMR*, **20:** 135-147

Pearson, H (2004) Monkey virus may be cleared of cancer link. *Nature*, **431:** 495

Peden, K, Sheng, L, Omeir, R, Yacobucci, M, Klutch, M, Laassri, M, Chumakov, K, Pal, A, Murata, H, Lewis, AMJ (2008) Recovery of strains of the polyomavirus SV40 from rhesus monkey kidney cells dating from the 1950s to the early 1960s. *Virology*, **370:** 63-76

Pellecchia, M, Szyperski, T, Wall, D, Georgopoulos, C, Wûthrich, K (1996) NMR structure of the J-domain and the Gly/Phe-rich region of the Escherichia coli DnaJ chaperone. *J Mol Biol*, **260:** 236-250

Pérez-Losada, M, Christensen, RG, McClellan, DA, Adams, BJ, Viscidi, RP, Demma, JC, Crandall, KA (2006) Comparing phylogenetic codivergence between polyomaviruses and their hosts. *J Virol*, **80:** 5663-5669

Pilon, AA, Desjardins, P, Hassell, JA, Mes-Masson, AM (1996) Functional implications of mutations within polyomavirus large T antigen Rb-binding domain: effects on pRb and p107 binding in vitro and immortalization activity in vivo. *J Virol*, **70:** 4457-4465

Pipas, JM (1992) Common and unique features of T antigens encoded by the polyomavirus group. *J Virol*, **66:** 3979-3985

Prilusky, J, Felder, CE, Zeev-Ben-Mordehai, T, Rydberg, EH, Man, O, Beckmann, JS, Silman, I, Sussman, JL (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21:** 3435-3438

Putnam, CD, Hammel, M, Hura, GL, Tainer, JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys*, **40:** 191-285

Ravichandran, V, Major, EO (2006) Viral proteomics: a promising approach for understanding JC virus tropism. *Proteomics*, **6:** 5628-5636

Receveur-Bréchot, V, Bourhis, J-M, Uversky, VN, Canard, B, Longhi, S (2006) Assessing protein disorder and induced folding. *Proteins*, **62:** 24-45

Rehm, T, Huber, R, Holak, TA (2002) Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. *Structure*, **10:** 1613-1618

Renner, C, Schleicher, M, Moroder, L, Holak, TA (2002) Practical aspects of the 2D 15N-[1h]-NOE experiment. *J Biomol NMR*, **23:** 23-33

Scherneck, S, Ulrich, R, Feunteun, J (2001) The hamster polyomavirus--a brief review of recent knowledge. *Virus Genes*, **22:** 93-101

Schuck, P (2004) A model for sedimentation in inhomogeneous media. I. Dynamic density gradients from sedimenting co-solutes. *Biophys Chem*, **108:** 187-200

Schwalbe, H, Fiebig, KM, Buck, M, Jones, JA, Grimshaw, SB, Spencer, A, Glaser, SJ, Smith, LJ, Dobson, CM (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry*, **36:** 8977-8991

Sheng, Q, Denis, D, Ratnofsky, M, Roberts, TM, DeCaprio, JA, Schaffhausen, B (1997) The DnaJ domain of polyomavirus large T antigen is required to regulate Rb family tumor suppressor function. *J Virol*, **71:** 9410-9416

Sheng, Q, Love, TM, Schaffhausen, B (2000) J domain-independent regulation of the Rb family by polyomavirus large T antigen. *J Virol*, **74:** 5280-5290

Singh, M, Krajewski, M, Mikolajka, A, Holak, TA (2005) Molecular determinants for the complex formation between the retinoblastoma protein and LXCXE sequences. *J Biol Chem*, **280:** 37868-37876

Srinivasan, A, McClellan, AJ, Vartikar, J, Marks, I, Cantalupo, P, Li, Y, Whyte, P, Rundell, K, Brodsky, JL, Pipas, JM (1997) The amino-terminal transforming region of simian virus 40 large T and small t antigens functions as a J domain. *Mol Cell Biol*, **17:** 4761-4773

Sullivan, CS, Cantalupo, P, Pipas, JM (2000a) The molecular chaperone activity of simian virus 40 large T antigen is required to disrupt Rb-E2F family complexes by an ATP-dependent mechanism. *Mol Cell Biol*, **20:** 6233-6243

Sullivan, CS, Tremblay, JD, Fewell, SW, Lewis, JA, Brodsky, JL, Pipas, JM (2000b) Species-specific elements in the large T-antigen J domain are required for cellular transformation and DNA replication by simian virus 40. *Mol Cell Biol*, **20:** 5749-5757

Sullivan, CS, Grundhoff, AT, Tevethia, S, Pipas, JM, Ganem, D (2005) SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature*, **435:** 682-686

Sullivan, CS, Pipas, JM (2002) T antigens of simian virus 40: molecular chaperones for viral replication and tumorigenesis. *Microbiol Mol Biol Rev*, **66:** 179-202

Szyperski, T, Pellecchia, M, Wall, D, Georgopoulos, C, Wûthrich, K (1994) NMR structure determination of the Escherichia coli DnaJ molecular chaperone: secondary structure and backbone fold of the N-terminal region (residues 2-108) containing the highly conserved J domain. *Proc Natl Acad Sci U S A*, **91:**

11343-11347

Tong, KI, Yamamoto, M, Tanaka, T (2008) A simple method for amino acid selective isotope labeling of recombinant proteins in E. coli. *J Biomol NMR*, **42:** 59-67

Tsutakawa, SE, Hura, GL, Frankel, KA, Cooper, PK, Tainer, JA (2007) Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. *J Struct Biol*, **158:** 214-223

Uversky, VN, Gillespie, JR, Fink, AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, **41:** 415-427

Uversky, VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, **11:** 739-756

Uversky, VN, Roman, A, Oldfield, CJ, Dunker, AK (2006) Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. *J Proteome Res*, **5:** 1829-1842

VanLoock, MS, Alexandrov, A, Yu, X, Cozzarelli, NR, Egelman, EH (2002) SV40 large T antigen hexamer structure: domain organization and DNA-induced conformational changes. *Curr Biol*, **12:** 472-476

Vranken, WF, Boucher, W, Stevens, TJ, Fogh, RH, Pajon, A, Llinas, M, Ulrich, EL, Markley, JL, Ionides, J, Laue, ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59:** 687-696

Whalen, KA, de, J, Rowena, Kean, JA, Schaffhausen, BS (2005) Genetic analysis of the polyomavirus DnaJ domain. *J Virol*, **79:** 9982-9990

White, MK, Gordon, J, Reiss, K, Del, V, Luis, Croul, S, Giordano, A, Darbinyan, A,

Khalili, K (2005) Human polyomaviruses and brain tumors. *Brain Res Brain Res Rev*, **50:** 69-85

White, MK, Khalili, K (2005) Expression of JC virus regulatory proteins in human cancer: potential mechanisms for tumourigenesis. *Eur J Cancer*, **41:** 2537-2548

Whyte, P (1995) The retinoblastoma protein and its relatives. *Semin Cancer Biol*, **6:** 83-90

Wilson, JB, Hayday, A, Courtneidge, S, Fried, M (1986) A frameshift at a mutational hotspot in the polyoma virus early region generates two new proteins that define T-antigen functional domains. *Cell*, **44:** 477-487

Wishart, DS, Bigam, CG, Holm, A, Hodges, RS, Sykes, BD (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR*, **5:** 67-81

Wishart, DS, Sykes, BD, Richards, FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31:** 1647-1651

Wittekind, M, Mueller, L (1992) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson*, **101:** 201-205

Xiao, B, Spencer, J, Clements, A, Ali-Khan, N, Mittnacht, S, Broceño, C, Burghammer, M, Perrakis, A, Marmorstein, R, Gamblin, SJ (2003) Crystal structure of the retinoblastoma tumor suppressor protein bound to E2F and the molecular basis of its regulation. *Proc Natl Acad Sci U S A*, **100:** 2363-2368

Zalvide, J, Stubdal, H, DeCaprio, JA (1998) The J domain of simian virus 40 large T

antigen is required to functionally inactivate RB family proteins. *Mol Cell Biol*, **18:** 1408-1415

Zhu, L (2005) Tumour suppressor retinoblastoma protein Rb: a transcriptional regulator. *Eur J Cancer*, **41:** 2415-2427